PROCEEDINGS
OF THE
INTERNATIONAL CONGRESS
OF
MATHEMATICIANS
1954

Amsterdam

September 2—September 9

VOLUME III

ERVEN P. NOORDHOFF N.V., GRONINGEN
NORTH-HOLLAND PUBLISHING CO., AMSTERDAM
1956.

PROCEEDINGS OF THE

INTERNATIONAL CONGRESS OF MATHEMATICIANS

1954

held at Amsterdam

under the auspices of the

WISKUNDIG GENOOTSCHAP

EDITORIAL COMMITTEE

JOHAN C. H. GERRETSEN

JOHANNES DE GROOT

PRINTED IN THE NETHERLANDS

COPYRIGHT WISKUNDIG GENOOTSCHAP 1954

PUBLISHED WITH FINANCIAL ASSISTANCE FROM UNESCO

VOLUME III

STATED ADRESSES IN SECTIONS. SYMPOSIA

This Volume III contains all halfhour lectures given on invitation of the Congress Committee, arranged according to the sections. Further, it contains the halfhour lecture in section VII given on invitation of the International Committee on Mathematical Instruction. Finally, it contains all lectures given in symposium A *Stochastic Processes* and in symposium B *Algebraic Geometry* on invitation of the symposium committees, or given in joint sessions of these symposia and the Congress on invitation of the Congress Committee.

The one-hour lectures given in the Congress are to be found in Volume I of these Proceedings, which will be published soon.

Lectures given in symposium C *Mathematical Interpretation of Formal Systems* have been published separately under that title in the series "Studies in Logic and the Foundations of Mathematics", North-Holland Publishing Comp., Amsterdam, 1955. In symposium C there were no joint sessions with the Congress.

STATED ADDRESSES IN SECTIONS

SECTION I

ALGEBRA AND THEORY OF NUMBERS

STATED ADDRESSES

SIMULTANEOUS DIOPHANTINE APPROXIMATION

H. DAVENPORT

The simplest problems of Diophantine approximation relate to the approximation of a single irrational number θ by rational numbers p/q, and the principal question is how small we can make the error $\theta - p/q$ in relation to q for infinitely many approximations. It is well known that this question can be answered almost completely in terms of the continued fraction expansion of θ. It must be admitted that our knowledge of the relationship between the continued fraction expansion of θ and other possible representations of θ is very fragmentary, a striking enough example being the number $\sqrt[3]{2}$. However, the continued fraction theory gives us many general results which are best possible. Thus every θ admits an infinity of approximations satisfying

$$\left|\theta - \frac{p}{q}\right| < \frac{1}{\sqrt{5}\, q^2},$$

but if

$$\theta = a_0 + \frac{1}{a_1+}\,\frac{1}{a_2+}\,\cdots \qquad \text{and} \quad a_n = 1 \ \text{for all} \ n > n_0$$

then no more precise result is true. Again, if a_n is bounded then there is some C such that there are only finitely many approximations satisfying

$$\left|\theta - \frac{p}{q}\right| < \frac{C}{q^2},$$

and conversely. The class of numbers with bounded partial quotients is an extensive one, in the sense that it has the cardinal number of the continuum.

The problem of simultaneous Diophantine approximation is that of approximating to n given real numbers $\theta_1, \ldots, \theta_n$, not all rational, by rationals $p_1/q, \ldots, p_n/q$ with the same denominator. There are various ways of measuring the precision of the simultaneous approximation; perhaps the simplest is to take as a measure the maximum modulus of the n errors $\theta_r - p_r/q$. It is easily proved, by using Dirichlet's "Schubfachschluss", that there are infinitely many simultaneous approximations for which

$$\left|\theta_r - \frac{p_r}{q}\right| < \frac{1}{q^{1+1/n}} \qquad (r = 1, \ldots, n).$$

In any deeper investigation of the problem of simultaneous Diophantine approximation we are greatly handicapped by the absence of a full analogue of the continued fraction process. There *are* several analogues, but they all suffer from one of two defects: either they give much poorer approximations than those we know to exist, or they involve a sequence of operations which can be carried out on given $\theta_1, \ldots, \theta_n$ but cannot be used to define $\theta_1, \ldots, \theta_n$ because we do not know the limitations to which any such sequence is subject.

Such knowledge as we have concerning simultaneous approximation has been mostly deduced from the Geometry of Numbers. The errors (after multiplication by q) and the denominator q constitute a set of $n+1$ linear forms

$$p_1 - \theta_1 q, \ldots, p_n - \theta_n q, q$$

in $n+1$ variables of determinant 1, that is, a lattice in $n+1$ dimensional space of a special kind, and it is from this point of view that the geometry of numbers becomes relevant.

As an illustration, suppose we define a constant of best approximation C_n to be the lower bound of all numbers C with the property that every set $\theta_1, \ldots, \theta_n$ admits infinitely many approximations satisfying

$$| p_r - \theta_r q | < \left(\frac{C}{q}\right)^{\frac{1}{n}} \qquad (r = 1, \ldots, n).$$

I have proved recently [1]) that

$$C_n = \frac{1}{\Delta(K)},$$

where $\Delta(K)$ denotes the critical determinant of the region K in $n+1$ dimensional space defined by

$$\max\ (|x_1|, \ldots, |x_n|) \leqq |x_{n+1}|^{-1/n}.$$

This generalizes a result of Furtwängler [2]), and extends further to the case where the maximum modulus is replaced by a fairly general function of the n errors. An important result used in the proof is that any lattice in $n+1$ dimensions has a representation by linear forms which is such that a particular submatrix of order n is approximately a multiple of the unit matrix. This result is an advance on what was proved by Furtwängler in the same connection.

The value of C_1 is $1/\sqrt{5}$, corresponding to the fact that the critical determinant of the plane region $|xy| \leqq 1$ is $\sqrt{5}$. The exact value of C_2 is not known; the best estimates known are

$$\frac{2}{7} \leqq C_2 < \frac{1}{46^{1/4}} \left(= \frac{1}{2.60\ldots}\right),$$

due to Davenport [3]) and Cassels [4]). Both are derived from the geometry of numbers.

If, in the case of two numbers, we measure the effectiveness of the simultaneous approximation by the product of the errors instead of by the maximum, we are led to an outstandingly difficult unsolved problem first propounded by Littlewood. Given any two real numbers θ, φ do there exist integers x, y, z $(z \neq 0)$ which make

$$| (x - \theta z)(y - \varphi z)z |$$

arbitrarily small? Littlewood conjectured that such integers do exist. If this is so, we have a problem in which lattices of the special kind arising in Diophantine approximation behave differently from lattices in general. A sufficient condition for the truth of Littlewood's conjecture has been found recently by Cassels and Swinnerton-Dyer, and is of interest because it is expressed in terms of general concepts in the geometry of numbers. But it does not seem to bring the problem any nearer to solution.

I mentioned earlier that there is a wide class of single irrationals θ which are badly approximable in the sense that, for some C, there are only a finite number of approximations satisfying

$$\left| \theta - \frac{p}{q} \right| < \frac{C}{q^2}.$$

This suggests the problem of the pairs θ_1, θ_2 for which there are only a finite number of approximations satisfying

$$\left| \theta_1 - \frac{p_1}{q} \right|, \quad \left| \theta_2 - \frac{p_2}{q} \right| < \frac{C}{q^{3/2}}.$$

The existence of such pairs was first proved by Furtwängler [2]), but his construction was such that θ_1 and θ_2 are necessarily numbers in a cubic field. This is analogous to using quadratic irrationals, with periodic continued fractions, as examples of irrationals with bounded partial quotients, and raises the question of whether there is a wider class of pairs θ_1, θ_2, having the cardinal number of the continuum.

I have proved [5]) that this is the case. The construction which I use defines θ_1, θ_2 by means of an infinite sequence of unimodular substitutions. The substitutions in question are those provided by the operation of changing a basis of a cubic field by multiplying throughout by a unit. There is a choice of substitutions at each stage, and it is proved that two distinct sequences define distinct pairs θ_1, θ_2. The choice open at any stage depends, however, on the choices made at all the preceding stages. This is connected with the necessity of maintaining a balance between the two errors $\theta_1 - p_1/q$ and $\theta_2 - p_2/q$ under a

sequence of changes of variable, and somewhat delicate considerations are needed.

The result has recently been extended by Cassels to sets of n irrationals. The unimodular substitutions which he uses are found without reference to algebraic number-fields, and this results in a helpful simplification.

References

[1] H. Davenport. On a theorem of Furtwängler, J. London Math. Soc. 30 (1955), 186—195.

[2] P. Furtwängler. Uber die simultane Approximation von Irrationalzahlen, Math. Annalen 96 (1926), 169—175 and 99 (1928), 71—83.

[3] H. Davenport. Simultaneous Diophantine approximation, Proc. London Math. Soc. (3), 2 (1952). 406—416.

[4] J. W. S. Cassels. Simultaneous Diophantine approximation, J. London Math. Soc. 30 (1955), 119—121.

[5] H. Davenport. Simultaneous Diophantine approximation. Mathematika 1 (1954), 51—72.

ON ADDITIVE ARITHMETICAL FUNCTIONS AND APPLICATIONS OF PROBABILITY TO NUMBER THEORY

P. ERDÖS

The main topic of this lecture will be the study of the distribution function of additive arithmetical functions. Before proceeding with the subject I want to call attention to the excellent review of Kac [13] on this subject, were a very clear and detailed discussion of the interplay of probability and number theory is given. In the present review I try to emphasize those results which were not treated in detail by Kac.

A real valued function $f(n)$, $n = 1, 2, \ldots$ is called additive if

$$f(m.n) = f(m) + f(n) \text{ whenever } (m,n) = 1.$$

The function $g(n)$ $n = 1, 2, \ldots$ is called multiplicative if

$$g(m.n) = g(m).g(n) \text{ whenever } (m,n) = 1.$$

Additive and multiplicative functions are thus completely determined if one knows their values for all powers of primes p^{α}.

Examples for additive functions are $\log n$, $\nu(n)$ (the number of distinct prime factors of n); examples for multiplicative functions are n, Euler's function $\varphi(n)$, and $\delta(n)$ (the sum of divisors of n).

Let $h(n)$ be any real valued function defined for $n = 1, 2, \ldots$. We say that $h(n)$ has a distribution function $\psi(c)$ if for every c the density of integers satisfying $h(n) < c$ exists. Denoting this density by $\psi(c)$ we further must have $\psi(-\infty) = 1$, $\psi(+\infty) = +1$.

The question whether an arithmetic function has a distribution function received considerable attention. As far as I know the first result in this direction is due to Schoenberg [15] who proved that $\varphi(n)/n$ has a distribution function. A little later almost simultaneously Davenport[2], Behrend and Chowla proved that $\sigma(n)/n$ also has a distribution function. This result was especially interesting since for $c = 2$ we abtain the so called deficient numbers. (If $\sigma(n) < 2n$, n is called deficient, with $\sigma(n) \geqq 2n$, n is called abundant and with $\sigma(n) = 2n$, n is called perfect).

All these papers used the theory of Fourier transforms. I attacked the problem in a somewhat different way. An integer u is called primitive abundant if it is abundant and all its divisors are deficient. Clearly we obtain all abundant

numbers by considering the set of all multiples of all the primitive abundant numbers. Thus if we can show that the sum of the reciprocals of the primitive abundant numbers converges, it will follow by a simple argument that the density of the abundant numbers exists. Denote by $N(n)$ the number of primitive abundant numbers not exceeding n, then I proved that [4, 5]

$$(1) \qquad \frac{n}{e^{25\sqrt{\log n \operatorname{loglog} n}}} < N(n) < \frac{n}{e^{\frac{1}{8}\sqrt{\log n \operatorname{loglog} n}}}.$$

(1) easily implies that the sum of the reciprocals of the primitive abundant numbers converges and thus the density of the abundant numbers exists.

Unfortunately this method can be applied only in very special cases. Define n to be primitive α-abundant if $\sigma(n) \geqq \alpha n$ but for all divisors d of n $\sigma(d) < \alpha d$. It is not hard to prove that for almost all α the sum of the reciprocals of the primitive α-abundant numbers converges, but it is also easy to show that every interval contains $c\ \alpha - s$ for which the sum of the reciprocals of the primitive α-abundant numbers diverges (these results are unpublished). Further Besicovitch constructed a sequence of integers $a_1 < a_2 < \ldots$ so that the set of their multiples does not have a density. Thus it is clear that the above method does not lead to general results.

In a subsequent paper Schoenberg [14] deduced the existence of the distribution function of additive functions under very general conditions. Since the logarithm of a multiplicative function is additive it usually will suffice to consider additive functions.

Generalising several previous results [15, 6] I proved that the convergence of the two series

$$(2) \qquad \sum_p \frac{f^+(p)}{p}, \quad \sum_p \frac{(f^+(p))^2}{p}$$

where

$$(3) \qquad f^+(p) = \begin{cases} f(p) \text{ if } |f(p)| \leqq 1 \\ 1 \quad \text{if } |f(p)| > 1 \end{cases}$$

is a sufficient condition for the existence of the distribution function. In a subsequent paper Wintner and I [7]) proved that the convergence of (2) is the necessary and sufficient condition for the existence of the distribution function of $f(n)$.

Our principal tool was the following theorem of Kac [8]) and myself: Let $f(n)$ be additive. Put

$$(4) \qquad A_n = \sum_{p<n} \frac{f(p)}{p}, \quad B_n = \sum_{p<n} \frac{f^2(p)}{p}$$

Assume that

$$B_n \to \infty, \ f(p) = 0(1). \tag{5}$$

Denote by $K_n(\omega)$ the number of integers $m, 1 \leqq m \leqq n$ for which

$$f(m) < A_m + \omega B_m^{\frac{1}{2}}. \tag{5}$$

Then

$$\lim_{n\to\infty} \frac{k_n(\omega)}{n} = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\omega} e^{-u^2/2} du. \tag{6}$$

Our proof of this result uses a combination of probabilistic and number theoretic methods, we use the central limit theorem and Brun's method. As stated in the introduction Kac gives a more detailed discussion of the connections between probability and number theory. As far as I know Kac and Wintner were the first who called attention to the probabi listic nature of these theorems. I also omit here the connection of these results with those of Hardy-Ramanujan and Turan since Kac discusses them in detail. Recently new proofs were given to our theorem with Kac by Delange [1]) and Halberstam which avoid both the central limit theorem and Brun's method but uses instead the method of moments.

A simple argument shows that (6) indeed implies that $B_n \to \infty$ is incompatible with the existence of the distribution function. The fact that the convergence of $\sum_p \frac{f(p)}{p}$ is also necessary for the existence of the distribution function requires some new and complicated arguments.

Using Brun's method and the estimates of Cramer and Berry on the error term in the central limit theorem I [9]) proved the following result:
Let $f(n)$ be additive.

$$B_n \to \infty, \ f(p) \to 0.$$

Then the density of integers for which

$$0 < f(m) < c(\text{mod } 1)$$

equals c, and is thus independent of the function $f(n)$.

Let us now assume that the series in (2) converges. It follows from a result of Paul Levy that the necessary and sufficient condition for the continuity of the distribution function is that

$$\sum_{f(p)\neq 0} \frac{1}{p} = \infty \tag{7}$$

I obtained for this result a direct number-theoretical proof. In a subsequent

paper I extended this result considerably, in fact I proved the following theorem: [9])

Let $f(n)$ be any additive function which satisfies (7). Then to every ε there exists a δ so that if x is sufficiently large and if $a_1 < a_2 < \ldots < a_t \leqq x$ is a sequence of integers satisfying

$$| f(a_i) - f(a_j) | < \delta$$

then $t < \varepsilon x$. A slichtly inaccurate but suggestive formulation of this result is that the distribution function tries to be continuous even if it does not exist.

The above result is a consequence of the following one:

An additive function is said to be finitely distributed if there exist two constants c_1 and c_2 and a sequence $n_1 < n_2 < \ldots$ tending to infinity so that for every i there exists a sequence

$$a_1^{(i)} < a_2^{(i)} < \ldots < a_{t_i}^{(i)} < n_i$$

satisfying

$$|f(a_r^i) - f(a_s^i)| < c_1,\ t_i > c_2 n_i\ (1 \leqq r, s \leqq t_i).$$

If $f(n)$ has a distribution function it is clearly finitely distributed, also $\log n$ is finitely distributed.

The necessary and sufficient condition that $f(n)$ should be finitely distributed is that there should exist a constant c and an additive function $g(n)$ so that

$$f(n) = c \log n + g(n), \sum_p \frac{(g^+(p))^2}{p} < \infty. \tag{8}$$

The proof that (8) is sufficient is not difficult, but the proof of the necessity is rather involved.

(7) gives the necessary and sufficient condition for the continuity of the distribution function, it would be of interest to obtain necessary and sufficient conditions for the absolute sontinuity of the distribution function. Here I [10]) proved using results of Jessen and Wintner that if $| f(p) | < \frac{1}{p^\alpha}$ for some $\alpha > 0$ and (7) holds, then the distribution function is continuous but purely singular. On the other hand I gave examples of absolutely continuous distribution functions, in fact if

$$f(p^\alpha) = (-1)^{(p-1)/2} (\log\log p)^{-3/4}$$

the distribution function is an entire function. It does not seem easy to give a necessary and sufficient condition for the absolute continuity of the distribution function. Another difficult problem seems to be to decide which distribution functions can be the distribution functions of additive functions. It is not hard

to prove that the distribution function $\psi(x) = 0$ for $-\infty < x < 0$, $\psi(x) = x$ for $0 \leqq x \leqq 1$, $\psi(x) = 1$ for $1 < x < \infty$, can not be the distribution function of an additive function.

Some time ago Chowla conjectured that the density of the integers for which $d(n+1) > d(n)$ is $\frac{1}{2}$. It is not difficult to deduce from the results of Kac and myself [1]) that if $|f(p)| < c$ (7) is satisfied, the density of integers satisfying $f(n+1) > f(n)$ is $\frac{1}{2}$. It would be of interest to find out to what extent the condition $|f(p)| < c$ can be weakened. The additive function $f(n) = \log n$ shows that it can not entirely be omitted. If $f(p) = (\log p)^{\alpha}$, $\alpha \neq 1$, $\alpha > 0$ or $f(p) = p$ it presumably is true that the density of integers for which $f(n+1) > f(n)$ is $\frac{1}{2}$, but I am unable to prove these conjectures. The conjecture for the case $f(p) = p$ is easily seen to be equivalent with the following conjecture: The density of integers n for which the greatest prime factor of $n+1$ is greater than the greatest prime factor of n is $\frac{1}{2}$. Leveque [14]) proved that if $|f(p)| < c$ and $\sum_p \frac{f^+(p)^2}{p}$ diverges $f(n)$ and $f(n+1)$ are statistically independent. It would be interesting to know to what extent the condition $|f(p)| < c$ can be omitted here, again $f(n) = \log n$ shows that it is not entirely superfluous (the condition $\sum_p \frac{(f^+(p))^2}{p} = \infty$ can not be omitted).

The proof of all these conjectures seems to present great difficulties. The methods used here by Kac, Wintner and myself and some other investigators which combine the central limit theorem with Brun's method works very well if the effect of the very large primes (the primes greater than n^{ε}) on the value of $f(n)$ is negligible. But if this condition is not satisfied (like e.g. for $f(p) = p$) our method fails completely.

Some further problems. I [9]) proved that if $f(n+1) - f(n) \to 0$, then $f(n) = c \log n$. Let us now only assume that $|f(n+1) - f(n)| < c$. I conjecture then that $f(n) = c_1 \log n + g(n)$, $|(gn)| < c_2$.

Is it true that almost all integers n have two divisors d_1 and d_2 satisfying

(9) $\qquad d_1 < d_2 < 2d_1$ (or more generally $d_1 < d_2 < (1+\varepsilon)d_1$)?

I [11]) proved that the density of integers satisfying (9) exists, but I can not prove that it is 1.

Let $g(n)$ be a multiplicative function for which

$$\sum_p \frac{g(p) - 1}{p}$$

and

(11) $$\sum_{p} \frac{(g(p)-1)^2}{p}$$

converges. Then

(12) $$\lim_{n=\infty} \frac{1}{n} \sum_{k=1}^{n} g(k)$$

exists and is different from 0 and ∞ [12]). Does the limit (12) still exist if we only assume the convergence of (10)?

It might be remarked that the method used to prove the existence of the distribution function of additive functions often can be used to prove the existence of the distribution function of non additive functions e.g. Shapiro and I proved that (will be published in the Canadian Journal of Math.)

$$\sum_{k=1}^{n} \frac{\varphi(k)}{k} - \frac{6}{\pi^2} n$$

has a continuous and steadily increasing distribution function. Also

(14) $$f(a) = \sum \frac{1}{k l_a(k)}$$

where $l_a(k)$ is the exponent of $a (\mathrm{mod}\ k)$, has a distribution function.

Many aspects of this subject were not considered in this short note e.g. the interesting results of Van Kampen, Wintner, Kac and de Bruijn on connections between the theory of almost periodic functions and additive and multiplicative functions, perhaps I gave an undue amount of place to my own papers, but I may be excused since not unnaturally I am best acquainted with them. The references to the litterature are not complete and again I refer to the review article by Kac quoted above.

In closing I would like to call attention to a few other applications of probability to number theory. First of all Cramer used probability theory to make plausible conjectures about primes. Further Rényi by using probabilistic methods considerably strengthened the large sieve of Linnik. He himself derived from these results that every integer is the sum of a prime and an integer having a bounded number of prime factors. His results have several other applications. Finally I would like to call attention to the fact that one often can prove the existence of sequences with certain properties by probability methods. In fact one then proves that almost all sequences have the required property, perhaps without being able to construct a single sequence having the property.

FOOTNOTES.

[1]) A. S. BESICOVITCH, On the density of certain sequences of integers, Math. Annalen **110** (1934) 336—341.

[2]) H. DAVENPORT, Über die Numeri Abundantes, Preuss. Akad. Wiss. Sitzungsber. (1933) 830—837.

[3]) H. DELANGE, C. R. Acad. Sci. Paris **237**, (1953) 542—544. Halberstam's paper was presented to the international congress in Amsterdam and has not yet been published in detail. P. Erdos,

[4]) On the density of the abundant numbers, Journal London Math. Soc. **9** (1934) 278—282.

[5]) On primitive abundant numbers, ibid **10** (1935) 49—58.

[6]) On the density of some sequences of numbers, ibid. **10** (1935) 120—125, **12** (1937) 7—11, **13** (1938) 119—127.

[7]) (with A. WINTNER) Additive arithmetical functions and statistical independence, Amer. Journal of Math. **61** (1939) 713—722.

[8]) (with M. KAC) The Gaussian law of errors in the theory of additive arithmetical functions, ibid. **62** (1940) 738—742.

[9]) On the distribution function of additive functions, Annals of Math. **47** (1946) 1—20

[10]) On the smoothness of the asymptotic distribution function of additive arithmetical functions, Amer. Journal of Math. **61** (1939) 722—725.

[11]) On the density of some sequences of integers, Bull. Amer. Math. Soc. **54** (1948) 685—692.

[12]) Some asymptotic formulas for multiplicative functions, ibid. **53** (1947) 536—544.

[13]) M. KAC, Probability methods in some problems of analysis and number theory, Bull. Amer. Math. Soc. **55** (1949) 641—665.

[14]) W. LEVEQUE, Trans. Amer. Math. Soc. **66** (1949) 440—463. I. J. Schoenberg,

[15]) Über asymptotische Verteilung reeller Zahlen mod 1, Math. Zeitschrif **28** (1928) 171—200.

[16]) On the asymptotic distribution of arithmetical functions, Trans. Amer. Math. Soc. **39** (1936) 315—330.

DAS INHOMOGENE PROBLEM IN DER GEOMETRIE DER ZAHLEN

EDMUND HLAWKA

Der Fragenkreis, über welchen ich sprechen will, wurde in den vergangenen Jahren, besonders seit dem Bericht von Prof. H. Davenport [1]) auf dem letzten Kongreß, in vielen Arbeiten behandelt und es wurde dabei eine Reihe von wichtigen Ergebnissen erzielt. Es soll nun ein Überblick über die allgemeinen Begriffsbildungen und Sätze gegeben werden, ohne dabei den Anspruch auf Vollständigkeit zu erheben.

Zunächst einige Vorbemerkungen! Es sei $\mathfrak{G}$ ein n-dimensionales Gitter ($n \geqq 1$) im Vektorraum $\mathfrak{V}_n$ eines euklidischen Raumes R_n von gleicher Dimension über dem Körper der reellen Zahlen, besitzt also stets Basen, gebildet von freien Vektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$, so daß alle Gittervektoren $\mathfrak{g} \in \mathfrak{G}$ ganzzahlige Linearkomposita dieser Vektoren sind. Arithmetisch gesprochen, ist $\mathfrak{G}$ die Wertmenge von n homogenen reellen Linearformen $\xi_i = L_i(x) = \sum_{k=1}^{n} a_{ik}x_k$ mit $\text{Det}\,(a_{ik}) \neq 0$, wenn die $x_1, \ldots, x_n$ alle ganze Zahlen durchlaufen. Zwei Punkte p, p' im R_n heißen zueinander äquivalent modulo $\mathfrak{G}$ ($p \equiv p' \pmod{\mathfrak{G}}$), wenn sie durch eine Gittertranslation von $\mathfrak{G}$ auseinander hervorgehen. Die Borelmeßbaren Fundamentalbereiche $F(\mathfrak{G})$ von $\mathfrak{G}$, für die es zu jedem $p \in R_n$ genau einen äquivalenten Punkt in F gibt, haben alle das gleiche Maß, $|\text{Det}.(\mathfrak{a}_1, \ldots, \mathfrak{a}_n)|$, ($\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ Basis von $\mathfrak{G}$), das Maß $m(\mathfrak{G})$ von $\mathfrak{G}$. Für reelles $\lambda \neq 0$ ist $m(\lambda\mathfrak{a}) = |\lambda|^n m(\mathfrak{a})$, ($\lambda\mathfrak{G}$ das Gitter der $\lambda\mathfrak{a}$ ($\mathfrak{a} \in \mathfrak{G}$). Ist T eine beliebige Menge im R_n, $T\mathfrak{a}$ die Menge, welche aus T durch eine Gittertranslation $\mathfrak{a} \in \mathfrak{G}$ hervorgeht, so heißt die Menge aller Figuren $T\mathfrak{a}$ mit H. Hadwiger [2]), das Figurengitter (T, $\mathfrak{G}$). Ist T ein Punkt p, so ist (p, $\mathfrak{G}$) die Menge aller zu p äquivalenten Punkte, das Punktgitter zum Vektorgitter $\mathfrak{G}$, welches p enthält.

Das inhomogene Problem lautet nun so: Wann überdeckt die Vereinigungsmenge der Figuren von (T, $\mathfrak{G}$), (das Innere T^0 von T nicht leer), den R_n, d.h. wann gibt es zu jedem p ein $p' \in T$ mit $p' \equiv p \pmod{\mathfrak{G}}$, symbolisch $p \in T \pmod{\mathfrak{G}}$. Es heißt dann $\mathfrak{G}$ Überdeckungsgitter zur Menge T, (Kurz $\ddot{U}G$ zu T). Es ist $\mathfrak{G}$ stets $\ddot{U}G$ zu jedem Fundamentalbereich $F(\mathfrak{G})$. Die Überdeckung ist

---

[1]) Proceedings of the International Congress of Math. (1950) I, 166—174.

[2]) Comment Math. Helvetici 11 (1938) 221—233.

hier einfach, denn die Figuren haben keinen Punkt gemeinsam. Ist $\mathfrak{G}$ kein $\ddot{U}G$ zu T, dann „verkleinern" wir das Gitter, d.h. wählen $\lambda \neq 0$ so klein, bis $\lambda\mathfrak{G}$ ein $\ddot{U}G$ zu T wird. Es heißt $E(T, \mathfrak{G}) = 1/\inf |\lambda|$ ($\lambda \in l$, die Menge [3]) aller λ), ($-\infty < \lambda < \infty$), für die $\lambda\mathfrak{G}$ kein $\ddot{U}G$ zu T ist), die Exzentrität zu $(T, \mathfrak{G})$. Es ist $E(T, \sigma\mathfrak{G}) = \sigma E(T, \mathfrak{G})$ ($\sigma > 0$). Lassen wir jetzt $\mathfrak{G}$ fest und unterwerfen T von einem inneren Punkt z aus einer genügend großen Ähnlichkeitstransformation $T \to T_z(\lambda)$ (λ Vergrößerungsverhältnis), solange bis $\mathfrak{G}$ Überdeckungsgitter von $T_z(\lambda)$ wird! Es sei T aus der Klasse Σ_τ der Sternkörper S mit Mittelpunkt. Die S sind abgeschlossene Mengen mit einem Zentrum $z \in S^0$, so daß z Mittelpunkt ist und mit jedem Punkt p von S alle Punkte der Verbindungsstrecke zp (höchstens p ausgenommen), innere Punkte von S sind. Also $S \subset S_z(\lambda)$ für alle $\lambda \geqq 1$. Zu S gibt es eine stetige Indikatrix $f(p)$, so daß die Punkte von S^0 durch $f(p) < 1$ ($f(z) = 0$) und jene von $S_z(\lambda)$ durch $0 \leqq f(p) \leqq \lambda$ charakterisiert sind. Beispiele für Sternkörper S liefert jede homogene Form $F(\xi) = F(\xi_1, \ldots, \xi_n)$ von einem Grad $d(F(t\xi) = t^d F(\xi))$ mit der Indikatrix $f(\xi) = |F|^{1/d}$. Ist $\mathfrak{G}$ definiert durch n homogene Linear-formen $\xi_i = L_i(x_1, \ldots, x_n)$ dann ist $\mathfrak{G}$ genau dann $\ddot{U}G$ zu $S_z(\lambda)$, wenn es zu jedem n Tupel $c = (c_1, \ldots, c_n)$ reeller Zahlen, ein n Tupel $g = (g_1, \ldots, g_n)$ ganzer Zahlen gibt, so daß

$$f(L_1(g-c), \ldots, L_n(g-c)) \leqq \lambda \tag{1}$$

ist. Da die Linearformen $L_i(x_1-c_1, \ldots, x_n-c_n)$ inhomogen in $x_1, \ldots, x_n$ sind, heißt der ganze Fragenkreis inhomogenes Problem, im Gegensatz zum homogenen Problem, wo nach ganzzahligen Lösungen $g \neq 0$ von (1) für $c = 0$ gefragt wird.

Ist $p \in R_n$, dann sei $E(p) = E(p, S, \mathfrak{G}) = \inf \lambda$ ($\lambda \in L^*$, die Menge aller $\lambda > 0$, für die $p \in S_z(\lambda)$ (mod $\mathfrak{G}$) ist), also gleich $\inf f(p')$ ($p' \equiv p$ (mod $\mathfrak{G}$)) ist. Gehört $E(p)$ selbst zu L^*, so ist es ein angenommenes Minimum für p. Es ist $E(p) = E(p')$ für $p \equiv p'$ und ist in p oberhalb stetig. Die Exzentrizität $E = E(S, \mathfrak{G})$ ist dann nichts anderes, als sup $E(p)$ über alle p, also ist $\mathfrak{G}$, für jedes $\lambda > E$, $\ddot{U}G$ zu $S_z(\lambda)$. Ist dies auch für $\lambda = E$ der Fall, dann heißt E ein angenommenes Minimum zu $(S, \mathfrak{G})$. Es gibt dann [4]) ein p, so daß $E(p)$ ein angenommenes Minimum und gleich E ist. Ist P die Menge aller p für die $E(p) = E$ und ist $E_2 = \sup E$ ($p \notin P$) kleiner als E, dann heißt E isoliert, $E = E_1$ das erste Minimum, E_2 das zweite Minimum. Die Definition der Kette dieser Minima kann fortgesetzt werden, bis ein nicht isoliertes Minimum auftritt.

---

[3]) Ist T offen, dann ist das Komplement zu l in $-\infty < \lambda < \infty$ offen und enthält eine Umgebung von $\lambda = 0$. Vgl. H. P. F. Swinnerton-Dyer, Proc. Camb. Phil. Soc. 50 (1954) 20—25.

[4]) J. Heinhold, Math. Z. 44 (1939) 659—88.

Ist M die Menge aller $\mathfrak{G}$, welche $\dot{U}G$ für S sind (nicht notwendig aus Σ_τ) und ist M^* ihr Komplement, dann sei $\vartheta(S) = \inf m(\mathfrak{G})(\mathfrak{G} \in M^*)$, $\Theta(S) = \sup m(\mathfrak{G})(\mathfrak{G} \in M)$ ($\vartheta(S) = \infty$, wenn M^* leer ist). Es heißt $\mathfrak{G}$ minimales bzw. maximales Gitter, wenn $m(\mathfrak{G}) = \vartheta(S)$ bzw. $\Theta(S)$ ist. Ist S abgeschlossen und meßbar mit einem Maß $m(S) < \infty$, dann gibt es stets [5]) maximale Gitter, welche, wenn S beschränkt und abgeschlossen ist [6]), (S^0 nicht leer), $\dot{U}G$ von S sind; im allgemeinen Fall bedeckt aber doch $(S, \mathfrak{G})$ den R_n bis auf eine Menge vom Lebesqueschen Maße 0. Es ist

$$\vartheta(S) = E^{-n}(S),\ \Theta(S) = \varepsilon^{-n}(S),\ \varepsilon(S) = \inf E(S, \mathfrak{G})/m^{1/n},$$
$$E(S) = \sup E(S, \mathfrak{G})/m^{1/n}$$

($\mathfrak{G}$ durchläuft alle Gitter mit festem Maße m), (m beliebig). Wir nennen die Gitter $\mathfrak{G}$ mit Maß m, für welche $E(S, \mathfrak{G})$ angenommen und gleich $\varepsilon(S)$ ist, dünnste $\dot{U}G$ des R_n durch Mengen S. Ist S abgeschlossen beschränkt, S^0 nicht leer, dann gibt es nach dem Vorhergehenden solche Gitter. $D(S) = m(S)\varepsilon^n(S) = m(S)/\Theta(S)$ heißt Dichte der dünnsten regulären Überdeckung des R_n durch Mengen S. Ist nämlich B eine beliebige abgeschlossene beschränkte konvexe Menge mit Mittelpunkt und positivem Volumen, (die Klasse dieser Mengen bezeichnen wir mit K), so heißt sie regulär überdeckt durch Mengen S, wenn es ein Gitter $\mathfrak{G}$ gibt, so daß B von den Figuren aus $(S, \mathfrak{G})$ lückenlos überdeckt wird und es ist $d = Nm(S)/m(B)$ die Dichte dieser Überdeckung, wenn N die Anzahl der Figuren ist, welche an der Überdeckung teilhaben. $D(B, S) = \min d$ ist die Dichte der dünnsten regulären Überdeckung von B durch Mengen S. Ich habe gezeigt [6]), daß $\lim_{\lambda \to \infty} D(B_z(\lambda), S)$ existiert (z Mittelpunkt von B) und gleich $D(S)$ ist. Daraus folgt, daß $\Theta(S) \leqq m(S)$ ist.

Ist S aus der Klasse K(z Mittelpunkt von S), dann ist bekanntlich

$$\mu_n \leqq 2E(S, \mathfrak{G}) \leqq n\mu_n \leqq n2^n m(\mathfrak{G})/m(S)\mu_1^{n-1} \tag{2}$$

wo $\mu_i = \mu_i(S, \mathfrak{G}, z)$, $(i = 1, \ldots n)$ das i-te Minimum ist. Für beliebige Mengen S und Punkte p ist $\mu_i(S, \mathfrak{G}, p)$ $(1 \leqq i \leqq n) = \inf \lambda$ ($\lambda > 0$; $\dim((p, \mathfrak{G}) \cap S_p(\lambda)) \geqq i$). Aus (2) folgert K. Mahler [7]) mit Hilfe seines Übertragungssatzes die wichtige Abschätzung [8]): $E(S, \mathfrak{G}) \ll \tau_1^{-1}$, mit $\tau_1 = \mu_1(T, \mathfrak{G}^*, z(S))$ wo T in Bezug auf z polar zu S und $\mathfrak{G}^*$ reziprok zu $\mathfrak{G}$ ist. Der Kroneckersche Approxi-

---

[5]) R. P. Bambah, Journ. Ind. (1953) 447–59.

[6]) Für konvexe Mengen bei E. Hlawka, Monatsh. Math. 53 (1949) 81–131.

[7]) Casopis Mat. a. Fys. 68 (1939) 93–102.

[8]) $A \ll B$ soll heißen: Es gibt eine positive Konstante c_n, die nur von n abhängt, daß $A \leqq c_n B$. Analog ist $A \gg B$ erklärt. Für die c_n sind in allen zu besprechenden Formeln, Abschätzungen bekannt. Wir begnügen uns hier mit der qualitativen Darstellung.

mationssatz ist eine unmittelbare Folgerung. Folgende Ungleichung wurde von mir [9]) gefunden:

$$2\,E(S,\mathfrak{G}) \leqq \{2^{n-1}(\nu+1)\,m\,(\mathfrak{g})/m(S)\} \tag{3}$$

wo $\nu(S,\mathfrak{G},z)$ die (ungerade) Anzahl der Gitterpunkte des Gitters $(z,\mathfrak{G})$ im Innern von S ist. ($\{x\}$ nächst größere ganze Zahl von x). Für den klassischen Fall $m(S) = 2^n\,m(\mathfrak{g})$, $\nu = 1$ liefert sie den genauen Wert $E = \frac{1}{2}$. S ist dann bekanntlich ein Polyeder, begrenzt von $2^n - 1$ Paaren von parallelen Seitenflächen. Ein anderes Beispiel ist folgendes. Es sei S ein Würfel W mit der Kantenlänge L und $\mathfrak{G}$ ein Gitter mit den Basisvektoren $\mathfrak{a}_1, \ldots, \mathfrak{a}_n$ parallel zu den Kanten und den Längen $2^{n-2}(\nu+1)/L^{n-1}$, $L/(\nu+1)$, $\frac{1}{2}L, \ldots, \frac{1}{2}L$, ($\nu$ ungerade, $2^{n-2}(\nu+1)/L^n > 1$). Man sieht sofort:

$$\nu(W,\mathfrak{G},z) = \nu,\ m(W) = L^n,\ m(\mathfrak{G}) = 1,\ 2E(W,\mathfrak{G}) = 2^{n-1}(\nu+1)L^n$$

Da ν und L beliebig ist, so folgt: *Für jede beschränkte Menge S ist stets $E(S) = \infty$, also ist $\vartheta(S) = 0$.*

Nach A. M. Macbeath [10]), ist dies auch dann noch der Fall, wenn S nicht beschränkt, aber vom endlichen Maß ist. Es genügt sogar anzunehmen, daß es ein Paar paralleler Hyperebenen gibt, so daß nur der Teil von S außerhalb dieses Streifens endliches Maß hat. Es ist [6]) für $S \in K$, $\Theta(S) \geqq (2/n)^{-n}\Delta_z(S) \geqq n^{-n}\,m(S)$. Nach C. A. Rogers [11]) ist, sogar [12]) $\Theta \geqq 2^n/3^{n-1}\Delta \geqq 3^{-(n-1)}m$ ($\Delta_z(S)$ die Mahlersche Grenzdeterminante inf $m(\mathfrak{G})$ über alle $\mathfrak{g}$, für die $\nu(S,\mathfrak{G},z) = 1$ ist). Wenn man von dem ausführlich studierten Fall [13]) $n = 2$ absieht, so ist im wesentlichen für $n \geqq 3$ nur die Kugel behandelt. Der genaue Wert von Θ ist für $n = 3 : 32/5\sqrt{5}$, für $n \geqq 4$ ist bis jetzt nur bekannt [15]) $4/3 - \varepsilon_n < D(S) < (1{,}15)^n$ ($\varepsilon_n \to 0$ für $n \to \infty$). Diese Resultate werden mit Hilfe der konvexen Polyeder (Wabenzellen) Z erhalten, definiert als Menge aller Punkte

---

[9]) Für $\nu = 1$ bei E. Hlawka, Math. Ann. 125 (1952) 183—207. Es kann (3) noch verschärft werden. Vgl. auch E. Hlawka, Monatsh. Math. 58 (1954) 287—91. M. Kneser, Math. Z. 61 (1954) 429—34.

[10]) Proc. Camb. Phil. Soc. 47 (1951) 627—28.

[11]) Journ. Lond. Math. Soc. 25 (1950) 328—31.

[12]) Besitzt S, n aufeinander senkrecht stehende Symmetrieebenen durch $z(S)$, dann ist nach R. P. Bambah und K. F. Roth, Journ. Ind. Math. Soc. 16 (1952) 7—12, $D(S) \leqq \pi n^n/3\sqrt{3}\,n! \sim e^n(\pi/54n)^{\frac{1}{2}}$.

[13]) Vgl. das schöne Buch von L. Fejes Toth, Lagerungen in der Ebene, auf der Kugel und im Raum, Springer-Verlag, Berlin-Heidelberg-Göttingen 1953 X, 197 S. Hinzugefügt werde, J. Heinhold, Math. Z. (1942) 199—214.

[14]) Neuer Beweis bei R. P. Bambah, Proc. Nat. Inst. Sciences of India 20 (1954) 25—52.

[15]) R. P. Bambah and H. Davenport, Journ. Lond. Math. 27 (1952) 224—29, H. Davenport Rend. Palermo (2) 1 (1952) 92—107.

p mit $|p, z| \leqq |p, g|$, für alle Gitterpunkte g von $(z, \mathfrak{G})$, (z Mittelpunkt der Kugel). Es ist $\mathfrak{G}$, $\dot{U}G$ zu Z mit $D(Z) = 1$, also besitzt Z höchstens $2^n - 1$ parallele Seitenflächenpaare und es ist $E(S, \mathfrak{G}) = \max |e, z|$ (erstreckt über alle Eckpunkte e von Z). Ist S konvex und offen (braucht nicht beschränkt zu sein), p_0 ein Punkt $\in S \pmod{\mathfrak{G}}$, so wird im allgemeinen p_0 in mehreren Figuren liegen, aber es gibt Teilmengen T von S, so daß $p \in T \pmod{\mathfrak{G}}$, aber nur mehr einer Figur angehört. Nach C. A. Rogers [16]) ist dies für $T = S \cap H (\neq 0)$ der Fall, wo H ein gewisser abgeschlossener Halbraum ist. Es liegt also in T *genau* ein zu p_0 äquivalenter Punkt p (C. A. Rogers nennt ihn äußeren Punkt zu p_0 in S mod $\mathfrak{G}$), also auch im Durchschnitt von S und dem Spiegelbild $S^+(p)$ an p außer ihm kein weiterer Gitterpunkt von $(p, \mathfrak{G})$. Nach dem Minkowskischen Fundamentalsatz gilt also

$$2^n m(\mathfrak{g}) \geqq 2^n \Delta_p(S \cap S^+(p)) \geqq m(S \cap S^+(p)) = F(p) \tag{4}$$

(Satz von A. M. Macbeath [17])). Es ist $\vartheta(S)$ genau dann ∞, wenn für jedes ϱ, S eine Kugel vom Radius ϱ enthält bzw. in keinem Streifen, begrenzt von parallelen Hyperebenen, liegt.

Ist S nicht konvex, so liegen in der Ebene, Überdeckungssätze von L. J. Mordell und K. Rogers für Mengen S von allgemeineren Charakter vor [18]). Für $n \geqq 2$ liefert der Satz von Macbeath (4): Ist $\mathfrak{G}$, $\dot{U}G$ der konvexen Menge S, dann ist es auch $\dot{U}G$ von der Menge $F(p) \leqq 2^n m(\mathfrak{G})$.

Ist S definiert durch $\xi_1 \geqq 0, \ldots, \xi_n \geqq 0$, $(F(p) = 2^n |p_1| \ldots |p_n|, p = (p_1, \ldots p_n))$ so folgt der Satz von J. H. W. Chalk [19]): Jedes $\mathfrak{G}$ ist $\dot{U}G$ von $\xi_1 \ldots \xi_n \leqq m(\mathfrak{G})$ $\xi_1 \geqq 0, \ldots, \xi_n \geqq 0$. Die Menge S: $\xi_1 - (\xi_2^2 + \ldots + \xi_n^2) \geqq 0$ führt zu dem Satz, daß das gleiche gilt, für

$$0 \leqq \xi_1 - (\xi_2^2 + \ldots + \xi_n^2) \leqq (2^{n-2}(n+1) m(\mathfrak{G})/\omega_{n-1})^{2/n+1} \tag{5}$$

(ω_{n-1} Volumen der $(n-1)$—dim. Kugel).

Alle Mengen S, welche man genauer untersucht hat, sind automorph, d.h. 1) sie besitzen eine Gruppe $A = A_z$ von Automorphismen α in Bezug auf einen Punkt z, (der nicht zu S gehören braucht), d.h. von affinen Abbildungen α mit z als Fixpunkt, welche S in sich überführen.

---

[16]) Journ. London 29 (1954) 133—143.

[17]) Annals of Math. (2) 56 (1952) 269—93. Für $n = 2$ auch Quart. Journ. (2) 3 (1952) 268—81, Durch (4) ist der Inhalt dieser wichtigen Untersuchungen nicht erschöpft.

[18]) Duke Math. Journ. 19 (1952) 519—27, K. Rogers, Journ. London 28 (1953) 394—402. Dazu gehören die Untersuchungen von E. S. Barnes, Quart. J. (2) 1 (1950) 199—210, R. P. Bambah, Proc. Camb. Phil. Soc. 47 (1952) 457—60, J. H. W. Chalk ebenda 48 (1952).

[19]) Quart. J. Math. 18 (1947) 215—27. Ein weiterer Beweis bei A. M. Macbeath, Jour. London 23 (1948) 141—47; A. J. Cole, Quart. J. (2) 3 (1952) 56—62 zeigt dies für $\xi_1 > 0, \ldots, \xi_n > 0$, $|\xi_1 \ldots \xi_n| \leqq 1/2\, m(\mathfrak{G})$.

2) Es gibt eine beschränkte Menge $U(A)$, so daß es zu jedem p aus der Hülle von S ein $\alpha \in A$ gibt, so daß $\alpha p \in U$. Wichtige Beispiele sind:

$S_1 : |\xi_1^2 + \ldots + \xi_r^2 - (\xi_{r+1}^2 + \ldots + \xi_n^2)| \leqq 1 \quad (2 \leqq r < n)$, $S_2 : |\xi_1 \ldots \xi_n| \leqq 1$

$S(K_1, \ldots, K_t)$ (Verallgemeinerung von S_2): $f_1^{n_1} \ldots f_t^{n_t} \leqq 1$ $(n_1 + \ldots + n_t = n)$ wo f_i $(1 \leqq i \leqq t)$ die Indikatrix des konvexen Körpers K_i in Räumen L_i von der Dimension n_i ist, welche den Schnittpunkt z der t aufeinander senkrechten L_i zum Mittelpunkt hat. Ein Spezialfall mit $n_1 = l, n_2 = m$: $(\mathrm{Max}\,(|x_1|, \ldots |x_l|))^l$ $(\mathrm{Max}\,(|y_1|, \ldots |y_m|))^m$ (Diophantische Approximation!) Da die Einheitskugel E_z um z in S_1, S_2 und in $S(K_1, \ldots, K_t)$ liegt, wenn die K Kugeln sind, so erhält man nach (2) eine Abschätzung

$$E(S, \mathfrak{G}) \ll m(\mathfrak{G}) |\mu_1^{n-1}(E_z) \ll m(\mathfrak{G}) |\mu_1^{n-1}(S) \tag{6}$$

Da mit E_z auch αE_z in S liegt $(\alpha \in A)$, so kann in (6) $\mu_1(E_z)$ durch $\mu = \sup \mu(\alpha E_z)$ ersetzt werden. Je umfangreicher die Automorphismengruppe von S ist, desto größer wird μ sein. So lieferten die Methoden von C. L. Siegel und Tschebotarew (vgl. dazu [20]) für [21]) $S(K_1, \ldots, K_t)$: $E(S, \mathfrak{G}) \ll m(\mathfrak{G}) |\mu_1^{(n-t)}(S, \mathfrak{G})$: ($K_i$ Kugeln). In S_2 ist $t = n$, daher ist $\vartheta(S_2) > 0$. Das gleiche zeigte für S_1, H. Blaney [22]). Es gilt aber sogar:

$E(S, \mathfrak{G}) \ll \mu_1^{(n-2)/2(n-1)}(S_4, \mathfrak{G})\, m(\mathfrak{G})$, wo S_4: $0 < \xi_1^2 + \ldots + \xi_r^2 - (\xi_{r+1}^2 + \ldots + \xi_n^2) < 1$

Nun gibt es aber Gitter mit $m(\mathfrak{G}) = 1$ und beliebig kleinen μ_1; Z.B. für das Gitter: $\xi_i = x_i N^{-1/n}$ $(1 \leqq i < n)$, $\xi_n = N^{1-1/n} x_n$ (N natürliche Zahl), ist $\mu_1 = N^{-1/n}$. Daher ist für $n > 2$, $\Theta(S_1) = \infty$ also $\varepsilon(S_1) = 0$; d.h. es gibt für $n \geqq 3$ stets Überdeckungsgitter zu S_1 mit beliebig großem Maß [23]). Für $n = 2$ versagt der Schluß und es ist auch tatsächlich $\Theta(S) < \infty$. Allgemein läßt sich folgendes zeigen: Für $S(K_1, K_2)$ (K_1, K_2 nicht notwendigerweise Kugeln) ist $\Theta(S) \ll V \ll \Delta(S)$, wo V das Volumen von $\mathrm{Max}\,(f_1, f_2) \leqq 1$ ist. Die Fälle, wo K_1, K_2 Intervalle oder Kreise sind, wurden zuerst von H. Davenport [1]) behandelt und von J. W. S. Cassels [24]) bedeutend vereinfacht. H. P. F. Swinnerton-Dyer [3]) ist es zuerst gelungen, ein allgemeines Kriterium dafür zu finden,

---

[20]) Der Fall $n = 2$ wird genau untersucht in Proc. Camb. Phil. Soc. 47 (1952) 266—73.

[21]) Der Exponent kann in vielen Fällen noch verschärft werden. Vgl. L. E. Clarke, Quart. J. of Math. (Oxford) (2) 2 (1951) 308—15; H. Davenport ebenda (2) 3 (1952) 32—41; E. S. Barnes, Proc. Camb. Phil. Soc. 49 (1953) 360—62; H. Davenport, ebenda 110—93.

[22]) Jour. London 23 (1948) 153—160; vgl. dazu C. A. Rogers, ebenda 27 (1952) 314—19, wo der oben herausgestellte Gesichtspunkt deutlich hervorgehoben ist.

[23]) Für $n = 3$ hat dies H. Davenport, Journ. London 23 (1948) 199—202 hervorgehoben und gibt eine schärfere Abschätzung für $E(S, \mathfrak{G})$.

[24]) Proc. Camb. Phil. Soc. 48 (1952) 72—86 und die dort angegebene Literatur. Der allgemeine Fall bei E. Hlawka, Monatsh. 58 (1954) 292—305.

daß $\vartheta(S)$, also $E(S) < \infty$ ist: Ist S offen, automorph in Bezug auf z und enthält jede Gerade, welche einen Punkt von $Fr_z(S)$ enthält, auch einen Punkt von S, dann ist $\vartheta(S) > 0$ und es gibt minimale Gitter. Dabei ist $Fr_z(S)$ die Hülle der äußeren Randpunkte von S, d.h. der Randpunkte r, welche Häufungspunkte von $zr \cap S$ sind.

Wenn S Sternkörper ist, $\vartheta(S) > 0$ ist und minimale Gitter $\mathfrak{G}'$ existieren, dann kann $\vartheta(S)$ isoliert sein, d.h. es kann $\vartheta_2 = \inf m(\mathfrak{G})$ erstreckt über alle Gitter, welche nicht Überdeckungsgitter sind, aber auch nicht von der Gestalt $\lambda\mathfrak{G}'$ ($\mathfrak{G}'$ minimal), $> \vartheta(S)$ sein. Diese Kette kann mann bis zu einem nicht isoliertem Minimum fortsetzen. Für $n = 2$ hat H. Davenport [25]) und für $n \geqq 2$ J. W. S. Cassels [26]) gezeigt, daß S_2 nicht isoliert ist, wenn die Minkowskische Vermutung richtig ist. (Bewiesen für $n \leqq 4$). Höhere Minima untersuchte für $0 \leqq \xi_1 - \xi_2 2 \leqq 1$, A. M. Macbeath. [27]) H. P. F. Swinnerton-Dyer und E. S. Barnes [28]) haben für $|\xi_1 \xi_2| \leqq 1$ unter Benützung von Ideen J. W. S. Cassels gezeigt, wie $E(S, \mathfrak{G})$ und die weiteren Minima $E_k(S, \mathfrak{G})$ ($k = 2, \ldots$) bestimmt werden können, wenn S automorph ist in Bezug auf eine Gruppe $A_z(\mathfrak{G})$, welche aber auch $\mathfrak{G}$ in sich selbst überführt. Zu jedem p gibt es dann ein p_1, so daß $E(p) = E(p_1)$ und $E(p_1)$ angenommen wird, Die Wertmenge von $E(p)$ ist abgeschlossen.

Gibt es ein α aus $A_z(\mathfrak{G})$ mit folgenden Eigenschaften: 1) Es besitzt keine Eigenwerte vom Betrage 1. 2) Sind R, R^* Mengen (R beschränkt), $\mathfrak{g} \in \mathfrak{G}$ so daß für jedes $p \in R$, entweder $\alpha p \in R^*$ (mod $\mathfrak{g}$) oder $\alpha p (-\mathfrak{g}) \in R$, $\alpha^{-1} p \in R \cup R^*$ dann ist $p \in R$ Fixpunkt f von α, $\alpha f = f\mathfrak{g}$, wenn 3) $\alpha^n p \in R^*$(mod $\mathfrak{g}$) für alle ganzen $n \neq 0$ ist. (Lemma von Cassels). Es sei nur eine wichtige Folgerung aus diesem Lemma hervorgehoben (wenn es $\alpha \in A(\mathfrak{G})$ mit der Eigenschaft 1 gibt): Gibt es nur endlich viele inäquivalente p, für die $E(p) = E$ ist, dann ist E isoliert. Dieser Satz gilt auch für die höheren Minima. Bisher wurde nur der Fall betrachtet, daß die Überdeckung mindestens einfach ist. Es kann auch mehrfache Überdeckung [30]) verlangt werden. In der Theorie der diophanti-

---

[25]) Proc. Kon. Ned. Wet. Amsterdam 49 (1946) 815—21.

[26]) Journ. London 27 (1952) 485—92. Für $\Delta(S)$ gilt das Gegenteil, vgl. dazu G.S. Barnes, Proc. Camb. Phil. Soc. 49 (1953) 59—62.

[27]) Proc. Camb. Phil. Soc. 47 (1951) 266—73.

[28]) Acta Math. 87 (1952) 259—323), 88 (1952) 279—316. Acta Math. 92 (1954) 199—234, E. S. Barnes ebenda 235—64 (Weiter Arbeiten: H. J. Godwin, Journ. London 36 (1955) 114—119, Quart. Journ. of Math. V(II)(1954) 28—46; K. Jukeri, Ann. Akad. Feunicae, (1952) Ser. A. 136, 1—16). Ternäre quadr. Farmen E. S. Barnes Acta Math. 92 (1954) 13—33.

[29]) Für $|\xi_1 \xi_2 \xi_3| < 1$ ist die Methode angewendet bei P. A. Samet, Proc. Cam. 50 (1954) 372—390.

[30]) vgl. D. B. Sawyer Proc. Cam. Phil. S. 49 (1953) 156—57.

schen Approximationen interessiert man sich für die unendlich oftmalige Überdeckung. Hier hat eine Arbeit von L. J. Mordell [31]). Anlaß zu wichtigen Untersuchungen gegeben. Die Ergebnisse von J. M. H. Chalk [32]) und C. A. Rogers [33]) haben diesen Fragenkreis zu einem gewissen Abschluß gebracht.

---

[31]) J. London Math. Soc. 26 (1951) 93; vgl. auch Varnavides Proc. London Math. (3) 2 (1952), 234—244.

[32]) J. Lond. Math. Soc. 25 (1950) 46, Quart. J. of Math. (Oxford) 19 (1948) 67.

[33]) Journ. London (3) 4 (1954) 50—83.

WIEN.

SOME ASPECTS OF THE THEORY OF REPRESENTATIONS OF JORDAN ALGEBRAS [1]

N. JACOBSON

A *Jordan algebra* is a non-associative algebra whose multiplication composition satisfies

$$ab = ba \tag{1}$$

$$(a^2b)a = a^2(ba). \tag{2}$$

If $\mathfrak{A}$ is an associative algebra, $\mathfrak{A}$ defines a Jordan algebra $\mathfrak{A}_J$ relative to the composition $\{ab\} = ab + ba$. Such algebras and their subalgebras are called *special*. There exist also *exceptional* (non-special) algebras, the best known example being the algebra M_3^8 of three-rowed hermitian Cayley matrices relative to $\{ab\} = ab + ba$.

If the characteristic is $\neq 2, 3$, and we shall assume this from now on, then (1), (2) are equivalent to (1) and (2′) $abcd + adcb + a(bdc) = (ab)(cd) + (ac)(bd) + (ad)(bc)$ where, in general, $abc \ldots u = (((ab)c) \ldots u)$. Let R_a denote the mapping $z \to xa\ (= ax)$. Then (2′) is equivalent to either

$$R_aR_bR_c + R_cR_bR_a + R_{acb} = R_aR_{bc} + R_bR_{ac} + R_cR_{ab} \tag{3}$$

or

$$[R_aR_{bc}] + [R_bR_{ac}] + [R_cR_{ab}] = 0, \quad ([AB] = AB - BA). \tag{4}$$

A *representation* of a Jordan algebra $\mathfrak{J}$ is a linear mapping $a \to R_a$ of $\mathfrak{J}$ into the algebra of linear transformations of a vector space $\mathfrak{M}$ such that (3) and (4) hold. If R is such a representation then we can define two compositions of $\mathfrak{J}$ and $\mathfrak{M}$ into $\mathfrak{M}$ by setting $xa = xR_a$, x in $\mathfrak{M}$, a in $\mathfrak{J}$ and $ax = xR_a$. Then (3) and (4) guarantee that all the relations obtained from the fundamental identity (2) by taking one of the arguments in $\mathfrak{M}$ the other three in $\mathfrak{J}$ are satisfied. A space $\mathfrak{M}$ with multiplications $xa = ax$, a in $\mathfrak{J}$, satisfying these conditions is called a *Jordan bimodule*. If we drop the superfluous composition ax then we obtain a *Jordan module*. The concepts of representation, Jordan bimodule,

[1] The representation theory discussed here has been developed in two papers by the author: (1) *General representation theory of Jordan algebras*, Trans. Amer. Math. Soc. vol 70 (1951), pp. 509—530. (2) *Structure of alternative and Jordan bimodules*, Osaka Math. Jour., vol. 6 (1954), pp. 1—71.

Jordan module are fully equivalent. The concepts of equivalence, etc. for representations or modules are defined as usual.

If $\mathfrak{M}$ is a Jordan bimodule and $\mathfrak{E}$ is the space direct sum $\mathfrak{J} \oplus \mathfrak{M}$ then we can define a composition in $\mathfrak{E}$ by $(a + x)\ (b + y) = ab + ay + xb$, $a, b \in \mathfrak{J}$, $x, y \in \mathfrak{M}$. The conditions on $\mathfrak{M}$ are just the ones needed to insure that $\mathfrak{E}$ is a Jordan algebra. This Jordan algebra is called the *split null extension* of $\mathfrak{J}$ by $\mathfrak{M}$.

Let $a \to S_a$ be a linear mapping of $\mathfrak{J}$ into the space of linear transformations of $\mathfrak{M}$ such that

$$S_{ab} = \{S_a S_b\}. \tag{5}$$

Clearly such a mapping is a homomorphism of $\mathfrak{J}$ into the special Jordan algebra $\mathfrak{L}_J$ of linear transformations in $\mathfrak{M}$. It can be verified that $R_a = S_a$ satisfies (3) and (4). We call such a representation *special*. Next let $a \to S_a^{(1)}$, $a \to S_a^{(2)}$ be two special representations which *commute*: $[S_a^{(1)} S_b^{(2)}] = 0$. Then one can verify that $R_a = S_a^{(1)} + S_a^{(2)}$ defines a representation. We call this the *sum* of the two commuting special representations. In a sense one can form the sum of any two special representations. Thus suppose $S^{(i)}$, $i = 1, 2$, acts in $\mathfrak{M}_i$ and form the Kronecker product $\mathfrak{M}_1 \otimes \mathfrak{M}_2$. Then $a \to S_a^{(1)} \otimes 1_2$ and $a \to 1_1 \otimes S_a^{(2)}$ are commuting special representations. Hence $a \to S_a^{(1)} \otimes 1_2 + 1_1 \otimes S_a^{(2)}$ is a representation, the *Kronecker sum* of $S^{(1)}$ and $S^{(2)}$.

In order to study the various types of representations one introduces corresponding universal associative algebras. Thus one defines the *universal associative algebra* $\mathfrak{U}(\mathfrak{J})$ of all representations as $\mathfrak{F}/\mathfrak{K}$ where $\mathfrak{F}$ is the free algebra $\mathfrak{J} \oplus (\mathfrak{J} \otimes \mathfrak{J}) \oplus (\mathfrak{J} \otimes \mathfrak{J} \otimes \mathfrak{J}) \oplus \ldots$ determined by $\mathfrak{J}$, and $\mathfrak{K}$ is the ideal generated by the elements $a \otimes b \otimes c + c \otimes b \otimes a + acb - a \otimes bc - b \otimes ac - c \otimes ab$ and $a \otimes bc - bc \otimes a + b \otimes ac - ac \otimes b + c \otimes ab - ab \otimes c$. Every representation determines a representation of this associative algebra. Similarly we define the *universal associative algebra* $\mathfrak{U}_s(\mathfrak{J})$ *of special representations* as $\mathfrak{F}/\mathfrak{K}_s$ where $\mathfrak{K}_s$ is the ideal generated by the elements $a \otimes b + b \otimes a - ab$ Let a_s denote the coset of a in $\mathfrak{U}_s$. If $\mathfrak{J}$ has an identity element 1 then $u = 2.1_s$ is the identity in $\mathfrak{U}_s$, Consider now the algebra $\mathfrak{U}_s \otimes \mathfrak{U}_s$ and let $\mathfrak{U}_s^{(2)}(\mathfrak{J})$ be the subalgebra generated by the elements $a_s \otimes u + u \otimes a_s$. We call $\mathfrak{U}_s^{(2)}(\mathfrak{J})$ the *universal associative algebra* of sums of *commuting special representation,* for it can be shown that every such representation of $\mathfrak{J}$ determines one for the associative algebra $\mathfrak{U}_s^{(2)}(\mathfrak{J})$. If $\mathfrak{J}$ has an identity we can also define a *universal algebra* $\mathfrak{U}_1(\mathfrak{J})$ for *unital representations,* that is, for representations S such that $S_1 = 1$.

There are two fundamental relations connecting the various universal associative algebras for a Jordan algebra with an identity. The first of these is $\mathfrak{U}(\mathfrak{J}) \cong \mathfrak{U}_1(\mathfrak{J}) \oplus \mathfrak{U}_s(\mathfrak{J})$. Also we have a natural homomorphism of $\mathfrak{U}_1(\mathfrak{J})$ onto

$\mathfrak{U}^{(2)}(\mathfrak{J})$. In important cases this is an isomorphism. The tool for establishing this result is the following criterion:

Theorem 1. Let $\mathfrak{J}$ be a special Jordan algebra, $\mathfrak{M}$ a module for $\mathfrak{J}$ and $\mathfrak{E} = \mathfrak{J} \oplus \mathfrak{M}$ the corresponding split null extension. Then $\mathfrak{M}$ is equivalent to a submodule of a sum of two commuting special modules (corresponding to special representations) if and only if $\mathfrak{E}$ is a special Jordan algebra.

In a number of important cases the relation between $\mathfrak{U}_s(\mathfrak{J})$ and $\mathfrak{U}_s^{(2)}(\mathfrak{J})$ can be made more explicit as follows. Thus we have the automorphism $P: \Sigma a_i \otimes b_i \to \Sigma b_i \otimes a_i$ in $\mathfrak{U}_s \otimes \mathfrak{U}_s$. The elements $a_s \otimes u + u \otimes a_s$ which generate $\mathfrak{U}_s^{(2)}$ are evidently contained in the subalgebra $\mathfrak{V}$ of elements invariant under P. In a number of cases it turns out that $\mathfrak{U}_s^{(2)}$ coincides with the subalgebra $\mathfrak{V}$.

Our main aim in this paper is to sketch the theory of representations for finite dimensional semi-simple Jordan algebras. For the sake of simplicity we shall also assume that the base field is algebraically closed. It can be shown that if $\mathfrak{J}$ is finite dimensional then $\mathfrak{U}(\mathfrak{J})$ is finite dimensional. Our problem can be defined as that of determining the structure of this associative algebra. However, it does not seem easy to attack this directly. Instead it appears to be necessary to devise a number of other methods to obtain the representations. The one we shall indicate here leans heavily on Albert's structure theory. [2]) According to this every semi-simple finite dimensional algebra has an identity and is a direct sum of simple ones. One can classify the simple ones according to their degree: the maximum number of pairwise orthogonal idempotent elements contained in such an algebra. At the present time the structure of simple algebras of degree 1 over an algebraically closed field is not known. One might conjecture that it is trivial, namely, that every such algebra is one dimensional. This is known to be the case for special algebras and for algebras of characteristic 0. The structure of simple algebras of degree > 1 has been determined by Albert, as follows.

I. If $\mathfrak{J}$ is simple of degree 2 over an algebraically closed field then $\mathfrak{J}$ is *the Jordan algebra of a non-degenerate scalar product.* Such an algebra is obtained by starting with a vector space $\mathfrak{J}_0$ in which a non-degenerate scalar product (symmetric bilinear form) (x, y) is defined. One forms $\mathfrak{J} = \mathfrak{J}_0 + \Phi 1$ and defines a multiplication in $\mathfrak{J}$ by

$$(6) \qquad (x + \alpha 1)(y + \beta 1) = (\beta x + \alpha y) + ((x, y) + \alpha\beta)1$$

for x, y in $\mathfrak{J}_0$, α, β in Φ. One verifies that $\mathfrak{J}$ is a Jordan algebra and $\mathfrak{J}$ is simple if $\dim \mathfrak{J}_0 > 1$.

---

[2]) See Albert's papers: *A structure theory for Jordan algebras*, Ann. of Math., vol. 48 (1947), pp. 446–467 and *A theory of power associative commutative algebras*, Trans. Amer. Math. Soc. vol 69 (1950), pp. 503–527.

II. Let $\mathfrak{J}$ be simple of degree $\geqq 3$ over an algebraically closed field. Then there exists an alternative algebra [3]) $\mathfrak{D}$ with an involution whose self-adjoint elements are the multiples of 1 and whose norm form $x\bar{x}$ is non-degenerate, such that $\mathfrak{J}$ is isomorphic to the algebra of hermitian matrices of the n-rowed matrix algebra $\mathfrak{D}_n$ relative to the composition $\{ab\} = ab + ba$. Moreover, if $n \geqq 4$, then $\mathfrak{D}$ is necessarily associative.

To complete the determination of the algebras $\mathfrak{D}$ one has to determine the algebras $\mathfrak{D}$ and the involution. The following are the only possibilities:

(i) $\mathfrak{D} = \Phi 1$.

(ii) $\mathfrak{D} = \Phi(q)$, $q^2 = \mu 1 \neq 0$, $\bar{q} = -q$.

(iii) $\mathfrak{D}$ is a quaternion algebra, usual involution.

(iv) $\mathfrak{D}$ is the algebra of Cayley numbers, usual involution.

The structure theorem II for algebras of degree $\geqq 3$ is an easy consequence of the following result which plays a fundamental role also in the representation theory for these algebras.

Theorem 2. Let $\mathfrak{J}$ be a Jordan algebra containing an identity 1 and suppose $\mathfrak{J}$ contains $n \geqq 3$ orthogonal idempotent elements e_i such the $\Sigma e_i = 1$ and $n - 1$ elements u_{i1}, $i = 2, \ldots, n$, such that $e_i u_{i1} = \frac{1}{2} u_{i1} = u_{i1} e_1$, $u_{i1}^2 = 4(e_1 + e_i)$. Then $\mathfrak{J}$ is isomorphic to the subalgebra $\mathfrak{H}$ of all hermitian matrices of an algebra $\mathfrak{D}_n$ where $\mathfrak{D}$ is an algebra with an involution. Moreover, $\mathfrak{D}$ is associative if $n > 3$ and alternative if $n \geqq 3$ and its self-adjoint elements are in the mucleus.

We can now outline the main results of the representation of finite dimensional semi-simple Jordan algebras. Such an algebra $\mathfrak{J}$ has an identity. Hence any $\mathfrak{J}$-module $\mathfrak{M}$ can be decomposed as $\mathfrak{M} = \mathfrak{M}_0 + \mathfrak{M}_1 + \mathfrak{M}_{1/2}$ where $R_1 = 0$ in $\mathfrak{M}_0$, $= 1$ in $\mathfrak{M}_1$ and $= \frac{1}{2}$ in $\mathfrak{M}_{1/2}$. The submodule $\mathfrak{M}_0$ is trivial since $R_a = 0$ in it. Also it can be shown easily that $\mathfrak{M}_{1/2}$ is special. If $\mathfrak{J} = \mathfrak{J}_1 \oplus \ldots \oplus \mathfrak{J}_r$ is the direct decomposition of $\mathfrak{J}$ into simple ideals then one sees easily that $\mathfrak{M}_{1/2}$ is a direct sum of modules which are special for one of the $\mathfrak{J}_i$ and trivial for the others. Next one can decompose the unital module $\mathfrak{M}_1$ as a direct sum of two types of submodules denoted as $\mathfrak{M}_{ij}$, $i < j$, and $\mathfrak{M}^{(i)}$. The module $\mathfrak{M}_{ij}$ is a sum of two commuting special modules one of which is special for $\mathfrak{J}_i$, trivial for the other $\mathfrak{J}_i$, and the second of which is special for $\mathfrak{J}_j$ and trivial for the $\mathfrak{J}_k$, $k \neq j$. The module $\mathfrak{M}^{(i)}$ is unital for $\mathfrak{J}_i$ and trivial for the other $\mathfrak{J}_j$. In this way one reduces the considerations to those of special and unital modules for the simple components $\mathfrak{J}_i$.

---

[3]) An algebra $\mathfrak{M}$ is called *alternative* if the associator $(ab)\,c - a(bc)$ is a skew symmetric function of the arguments a, b, c. A vector space $\mathfrak{M}$ is an alternative *bimodule* for $\mathfrak{A}$ if bilinear compositions am, ma $\mathfrak{M}$ for a $\mathfrak{A}$, m $\mathfrak{M}$ are defined so that all associators with two arguments in A and one in M are skew-symmetric functions.

We assume now that $\mathfrak{J}$ is simple and the base field is algebraically closed. If $\mathfrak{J}$ is of degree 1 then we can proceed no further unless $\mathfrak{J}$ is 1 dimensional. In this case the representation theory is trivial. Next suppose $\mathfrak{J}$ is of degree $= 2$. Then $\mathfrak{J}$ is the Jordan algebra $\mathfrak{J}_0 + 1\Phi$ where for x, y in $\mathfrak{J}_0$, $xy = (x, y)\,1$ and (x, y) is a nondegenerate scalar product. To obtain the representations for this algebra one can proceed directly to determine the universal associative algebras. One considers first the universal associative algebra $\mathfrak{U}_s(\mathfrak{J})$ of special representations. It is clear from the definition that $\mathfrak{U}_s(\mathfrak{J})$ is the *Clifford algebra* of the scalar product (x, y). Its structure is well known: Let $\dim \mathfrak{J}_0 = n$. Then $\dim \mathfrak{U}_s(\mathfrak{J}) = 2^n$ and $\mathfrak{U}_s(\mathfrak{J})$ is the full matrix algebra Φ_{2^ν} if $n = 2\nu$ while $\mathfrak{U}_s(\mathfrak{J}) = \Phi_{2^\nu} \oplus \Phi_{2^\nu}$ if $n = 2\nu + 1$. In any case $\mathfrak{U}_s(\mathfrak{J})$ is semi-simple so that all special representations are completely reducible. There are either one or two distinct irreducible representations according as n is even or odd. We consider next the universal associative algebra $\mathfrak{U}_1(\mathfrak{J})$ of unital representations. One can show that this algebra can be obtained from the free algebra $\mathfrak{F}^*(\mathfrak{J}_0) = \Phi \oplus \mathfrak{J}_0 \oplus (\mathfrak{J}_0 \otimes \mathfrak{J}_0) \oplus (\mathfrak{J}_0 \otimes \mathfrak{J}_0 \otimes \mathfrak{J}_0) \oplus \dots$ by factoring out the ideal generated by the elements of the form

$$xyz + zyx - (x, y)z - (y, z)x. \tag{7}$$

This algebra has been considered in the theory of mesons and hence we shall call it the *meson algebra* of $\mathfrak{J}$ (or of the scalar product (x, y)). The principal facts on this algebra are contained in the following

Theorem 3. Let $\mathfrak{U}_1(\mathfrak{J})$ be the meson algebra of $\mathfrak{J}$ (of dim $n + 1$) then (1) $\dim \mathfrak{U}_1(\mathfrak{J}) = \binom{2n+1}{n}$, (2) $\mathfrak{U}_1(\mathfrak{J}) \cong \mathfrak{U}_s^{(2)}(\mathfrak{J})$, (3) $\mathfrak{U}_1(\mathfrak{J}) = \Phi_{\binom{n+1}{0}} +$

$$\Phi_{\binom{n+1}{1}} + \dots + \Phi_{\binom{n+1}{\nu}} \quad \text{if} \quad n = 2\nu, \quad \text{while} \quad \mathfrak{U}_1(\mathfrak{J}) =$$

$$\Phi_{\binom{n+1}{0}} = \Phi_{\binom{n+1}{1}} + \dots + \Phi_{\binom{n+1}{\nu-1}} + \Phi_{\frac{1}{2}\binom{n+1}{\nu}} + \Phi_{\frac{1}{2}\binom{n+1}{\nu}} \quad \text{if} \quad n = 2\nu - 1.$$

This result shows that $\mathfrak{U}_1(\mathfrak{J})$ is semi-simple. It follows that the unital representations of $\mathfrak{J}$ are completely reducible. Moreover, the result gives all the irreducible representations.

We consider next the algebras of degree $\geqq 3$. Let $\mathfrak{J}$ be such an algebra and suppose $\mathfrak{J}$ is isomorphic to the Jordan algebra of hermitian matrices of $\mathfrak{D}_n$ where $\mathfrak{D}$ is an alternative algebra with 1, having the structure indicated above. We recall that $\mathfrak{D}$ is either the Cayley algebra or $\mathfrak{D}$ is associative and the first alternative can occur only when $n = 3$. We consider first the special representations. The result here is

Theorem 4. If $\mathfrak{D}$ is associative and $\mathfrak{J}$ is the Jordan algebra of hermitian matrices of $\mathfrak{D}_n$, $n \geqq 3$, then $\mathfrak{D}_n$ is the universal associative algebra $\mathfrak{U}_s(\mathfrak{J})$ of the special representations of $\mathfrak{J}$.

This result is a special case of a more general one due to N. Jacobson and C. E. Rickart. It follows also from earlier results of F. D. Jacobson and N. Jacobson. [4])

If $\mathfrak{D}$ is the Cayley algebra then there are no special representations. Hence in this case $\mathfrak{U}_s(\mathfrak{J}) = 0$.

The determination of $\mathfrak{U}_1(\mathfrak{J})$ is fairly complicated and hence we can only give some general indications here. Let $\mathfrak{M}$ be a unital bimodule for $\mathfrak{J}$ and let $\mathfrak{E}$ be the corresponding split null extension $\mathfrak{J} \oplus \mathfrak{M}$. Then $\mathfrak{E}$ has an identity 1 and $\mathfrak{E}$ contains the e_i, $i = 1, \ldots, n$, u_{j1}, $j = 2, \ldots, n$ given in the key structure theorem. Hence there exists an alternative algebra $\mathfrak{F}$ with an involution such that $\mathfrak{E}$ is isomorphic to the algebra of hermitian matrices of $\mathfrak{F}_n$. If $n > 3$ $\mathfrak{F}$ is associative. In this case we can conclude by Theorem 1 that $\mathfrak{M}$ is a submodule of a sum of commuting special modules. This implies that $\mathfrak{U}_1(\mathfrak{J}) \cong \mathfrak{U}_s^{(2)}(\mathfrak{J})$ and one can determine the structure of this algebra. If $n = 3$, $\mathfrak{F}$ need not be associative. In this case one has to apply a second method which is also useful for $n > 3$. One shows that $\mathfrak{F} = \mathfrak{D} \oplus \mathfrak{N}$ where $\mathfrak{N}$ is an ideal such that $\mathfrak{N}^2 = 0$. This reduces the consideration to that of representations or bimodules for the alternativs algebra $\mathfrak{D}$. Actually, somewhat more is involved since the involution in $\mathfrak{F}$ has to be taken into account. We should remark also that to obtain the results with a minimum of calculation it seems best to combine the two methods.

The main results that one obtains are that $\mathfrak{U}_1(\mathfrak{J})$ is semi-simple and its structure can be given. The results for $\mathfrak{U}_s(\mathfrak{J})$ and $\mathfrak{U}_1(\mathfrak{J})$ taken together show that the representations are completely reducible and the irreducible representation can be given. If we combine these results with the earlier considerations on module decompositions we can see that the representations of known semi-simple algebras are completely determined and the irreducible representations can be deduced from the structure of $\mathfrak{J}$. In concluding we should remark also that similar results can be obtained for arbitrary base fields. For example complete reducibility holds for the representation if $\mathfrak{J}$ is a known separable algebra.

---

[4]) See the references in (2) of footnote 1.

DIFFERENTIALGLEICHUNGEN UND AUTOMORPHE FUNKTIONEN

HANS MAASS

Nachdem die Einführung von Differentialgleichungen in die Theorie der automorphen Funktionen durch die Siegelschen Untersuchungen über indefinite quadratische Formen [1] [1]) hinreichend gerechtfertigt ist, erscheint es angezeigt, einmal im Zusammenhang auf die allgemeinen Gesichtspunkte hinzuweisen, die hierbei zur Geltung gekommen sind. In einem Teilgebiet $\mathfrak{G}$ des Raumes der komplexen Variablen $(z) = (z_1, \ldots, z_n)$ sei eine positive hermitesche Metrik

$$ds^2 = \sum_{\mu,\nu=1}^{n} g_{\mu\nu}(z) dz_\mu d\bar{z}_\nu \quad (g_{\mu\nu} = \bar{g}_{\nu\mu}) \tag{1}$$

erklärt, sodaß $\mathfrak{G}$ als ein Riemannscher Raum angesehen werden kann. Unter einer Bewegung verstehen wir eine pseudokonforme Transformation, welche $\mathfrak{G}$ topologisch abbildet und ds^2 invariant läßt. Ist $f(z)$ in $\mathfrak{G}$ erklärt, σ eine beliebige Bewegung und α, β ein gegebenes Paar komplexer Zahlen, so werde

$$f(z)|\sigma = \left(\frac{\partial(\sigma z)}{\partial(z)}\right)^{\frac{\alpha}{m}} \left(\frac{\partial(\sigma z)}{\partial(z)}\right)^{\frac{\beta}{m}} f(\sigma z) \tag{2}$$

mit geeignetem $m = m(\mathfrak{G}) > 0$ gesetzt. Wir nennen $f(z)$ eine automorphe Form zur (diskontinuierlichen) Bewegungsgruppe G, zum Gewichtssystem α, β und zum Multiplikatorsystem v, wenn $f(z)|\sigma = v(\sigma) f(z)$ für $\sigma \subset G$ gilt und wenn $f(z)$ von gewissen vorgegebenen Differentialoperatoren annulliert wird. $\{G; \alpha, \beta, v\}$ bezeichne die lineare Schar dieser Formen. Eine vernünftige Theorie dieser Formenscharen $\{G; \alpha, \beta, v\}$ kann anscheinend nur dann erwartet werden, wenn die gegebenen Differentialoperatoren mit f auch stets $f|\sigma$ annullieren, wobei σ eine beliebige Bewegung bezeichnet. Solche Invarianzforderungen sind bei der Auswahl der Differentialoperatoren zu beachten. Wir können den Fall eines Systems reeller Koordinaten $(z_1, \ldots, z_n)$ mit reellen $g_{\mu\nu}$ und $\beta = 0$ in die Betrachtung mit einbeziehen.

In den zur Zeit vorliegenden Untersuchungen handelt es sich entweder darum, automorphe Funktionen $(\alpha = \beta = 0)$ zu finden, die der Wellengleichung $(\Delta + \lambda) f = 0$ genügen, wobei Δ den Beltramischen Differential-

[1]) Ziffern in eckigen Klammern verweisen auf das angefügte Literaturverzeichnis.

operator bezeichnet, oder darum, zu einem gegebenen Formentypus, der durch Eisensteinreihen der Art

$$(3) \qquad \mathfrak{E}(z;\alpha,\beta) = \sum_{\sigma} \omega(\sigma) \left(\frac{\partial(\sigma z)}{\partial(z)}\right)^{\frac{\alpha}{m}} \left(\overline{\frac{\partial(\sigma z)}{\partial(z)}}\right)^{\frac{\beta}{m}}$$

repräsentiert wird, ein System von kennzeichnenden Differentialgleichungen zu bestimmen. Überdies konnte in speziellen Fällen mit Hilfe von Differentialoperatoren eine lineare Korrespondenz zwischen den Formenscharen $\{G;\alpha,\beta,v\}$ und $\{G;\alpha \pm 1, \beta \mp 1, v\}$ hergestellt werden. Es ist bemerkenswert, daß immer, wenn Wellenfunktionen in der Gestalt

$$(4) \qquad f(z) = \sum_{\sigma} \omega(\sigma)\ (g(\sigma z)^{-\nu} \qquad \text{mit } g(z) = |\, g_{\mu\nu}(z)|$$

gefunden werden konnten, $\mathfrak{G}$ nicht beschränkt und der Fundamentalbereich der zugrunde liegenden Gruppe nicht kompakt ist.

Auf automorphe Wellenfunktionen der hyperbolischen Ebene $y > 0$ mit der Grundform $ds^2 = y^{-2}(dx^2 + dy^2)$ wird man geführt [2], wenn man auf die Funktionen, die eine Dirichletsche Reihenentwicklung besitzen und gewissen Funktionalgleichungen genügen, wie sie bei den Zetafunktionen reeller quadratischer Zahlkörper auftreten, die Mellinsche Integraltransformation anwendet. Von besonderer Bedeutung ist die Bestimmung der automorphen, im Fundamentalbereich quadratisch integrierbaren Wellenfunktionen der Modulgruppe. Aus derartigen Eigenfunktionen sind die Größencharaktere der binären quadratischen Formen abzuleiten, die bei der Zuordnung der Siegelschen Modulformen zweiten Grades zu Dirichletschen Reihen benötigt werden [3]. Das genannte Eigenwertproblem ist für die Modulgruppe und ihre Hauptkongruenzuntergruppen neuerdings von W. Roelcke (Heidelberg) vollständig gelöst worden. Es ergab sich die Existenz unendlich vieler diskreter Eigenwerte, dazu ein kontinuierliches Spektrum. Die zugehörigen Eigenpakete sind aus den Eisensteinreihen vom Typus (4) durch einen Integrationsprozeß zu gewinnen. Zu einer umkehrbar eindeutigen Korrespondenz zwischen Siegelschen Modulformen p-ten Grades und Dirichletreihen gelangt man, wenn man die in [3] und [4] entwickelten Methoden kombiniert. Die erforderlichen Größencharaktere der quadratischen Formen in p Variablen erhält man aus den Eigenfunktionen zur Gruppe der Bewegungen $Y \to UYU'$ des Riemannschen Raumes $Y > 0$, $|\, Y\, | = 1$ mit der Fundamentalform $ds^2 = Sp(Y^{-1}dY)^2$. Dabei bezeichnet Y eine p-reihige reelle symmetrische Matrix und U eine unimodulare Matrix. Diese Eigenfunktionen sind mit den in Y homogenen Eigenfunktionen von der Dimension 0 bezüglich des vollen Raumes $Y > 0$ identisch. Für den vollen Raum ist $\Delta = Sp(Y\mathsf{D}_y)^2$ mit $\mathsf{D}_y = \left(e_{\mu\nu}\frac{\partial}{\partial y_{\mu\nu}}\right)$,

$Y = (y_{\mu\nu})$, $e_{\mu\nu} = \frac{1}{2}(1 + \delta_{\mu\nu})$, $\delta_{\mu\nu}$ = Kroneckersymbol. Eine Verallgemeinerung der Ansätze in [2] auf den hyperbolischen Raum $x_n > 0$ mit der Fundamentalform $ds^2 = x_n^{-2}(dx_1^2 + dx_2^2 + \ldots + dx_n^2)$ ist in [5] ausgeführt worden. Die Bewegungen sind hier mit Hilfe Cliffordscher Zahlsysteme dargestellt worden.

Die Kennzeichnung Eisensteinscher Reihen vom Typus (3) durch Differentialgleichungen ist zunächst für Grenzkreisgruppen vorgenommen worden [6]. Der Operationsbereich ist hier wieder die obere z-Halbebene. Mit $m = 2$ ist nun

$$\mathfrak{E}(z; \alpha, \beta) = \sum_{\sigma} \omega(\sigma)(cz + d)^{-\alpha}(c\bar{z} + d)^{-\beta},$$

wenn $\sigma z = (az + b)(cz + d)^{-1}$ mit reellen a, b, c, d angesetzt wird. Eine Differentialgleichung für $\mathfrak{E}(z; \alpha, \beta)$ mit der erforderlichen Invarianzeigenschaft gewinnt man mit Hilfe der Operatoren

$$\mathsf{K}_\alpha = \alpha + (z - \bar{z})\frac{\partial}{\partial z} \text{ und } \Lambda_\beta = -\beta + (z - \bar{z})\frac{\partial}{\partial \bar{z}}. \tag{5}$$

Man findet

$$\mathsf{K}_\alpha \mathfrak{E}(z; \alpha, \beta) = \alpha \mathfrak{E}(z; \alpha + 1, \beta - 1),\ \Lambda_\beta \mathfrak{E}(z; \alpha, \beta) = -\beta \mathfrak{E}(z; \alpha - 1, \beta + 1), \tag{6}$$

also

$$\mathfrak{Q}_{\alpha\beta}\mathfrak{E}(z; \alpha, \beta) = 0 \quad \text{mit} \quad \mathfrak{Q}_{\alpha\beta} = \Lambda_{\beta-1}\mathsf{K}_\alpha + \alpha(\beta - 1). \tag{7}$$

Einfache Rechnungen ergeben für die durch $\mathfrak{Q}_{\alpha\beta}\{G; \alpha, \beta, v\} = 0$ gekennzeichneten Formenscharen

$$\mathsf{K}_\alpha\{G; \alpha, \beta, v\} \subset \{G; \alpha + 1, \beta - 1, v\},\ \Lambda_\beta\{G; \alpha, \beta\, v\} \subset \{G; \alpha - 1, \beta + 1, v\}, \tag{8}$$

$$\{G; \alpha, \beta, v\} \mid \sigma = \{\sigma^{-1} G\sigma; \alpha, \beta, v^\sigma\}, \tag{9}$$

wobei v^σ in gewisser Weise aus v und σ zu berechnen ist. Die Heckeschen Untersuchungen über Modulformen und Dirichletreihen können auf die diesbezüglichen Formenscharen $\{G; \alpha, \beta, v\}$ vollständig übertragen werden. Spezielle Formenscharen $\{G; \alpha, \beta, v\}$ wurden im Zusammenhang mit indefiniten quadratischen Formen bereits in [7] untersucht.

Die Differentialgleichungstheorie [6] läßt sich, in beschränktem Umfang zunächst, auf die Siegelsche Modulgruppe p-ten Grades verallgemeinern [8]. Es sei Z eine p-reihige komplexe symmetrische Matrix mit positivem Imaginärteil und $\sigma = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ eine Modulsubstitution p-ten Grades. Mit $m = p + 1$ nimmt dann (3) die Gestalt

$$\mathfrak{E}(Z; \alpha, \beta) = \sum_{\sigma} \omega(\sigma) \mid CZ + D \mid^{-\alpha} \mid C\overline{Z} + D \mid^{-\beta}$$

an. Ist $I = I_p$ die p-reihige Einheitsmatrix, $Z = (z_{\mu\nu})$, $\mathsf{D}_z = \left(e_{\mu\nu}\frac{\partial}{\partial z_{\mu\nu}}\right)$ und

(10) $$\mathsf{K}_\alpha = \alpha I + (Z - \overline{Z})\mathsf{D}_z, \quad \mathsf{\Lambda}_\beta = -\beta I + (Z - \overline{Z})\mathsf{D}_{\overline{z}},$$

so hat man in

(11) $$\mathfrak{\Omega}_{\alpha\beta} = \mathsf{\Lambda}_{\beta - \frac{p+1}{2}}\mathsf{K}_\alpha + \alpha\left(\beta - \frac{p+1}{2}\right)I = (Z-\overline{Z})\{((Z-\overline{Z})\mathsf{D}_{\overline{z}})'\mathsf{D}_z + \alpha\mathsf{D}_{\overline{z}} - \beta\mathsf{D}_z\}$$

ein System von p^2 Differentialoperatoren, die sämtlich $\mathfrak{G}(Z;\alpha,\beta)$ annullieren. Die durch $\mathfrak{\Omega}_{\alpha\beta}\{G;\alpha,\beta,v\} = 0$ gekennzeichneten Formenscharen transformieren sich wieder nach der Regel (9). Der Übergang von der Schar $\{G;\alpha,\beta,v\}$ zu den Scharen $\{G;\alpha \pm 1, \beta \mp 1, v\}$ kann jetzt natürlich nicht mehr durch die Operatoren K_α, $\mathsf{\Lambda}_\beta$ vermittelt werden. Man benötigt hierfür gewisse Differentialoperatoren M_α, N_β von p-ter Ordnung, die im Falle $p = 1$ mit K_α, $\mathsf{\Lambda}_\beta$ identisch sind. Jedoch konnten die erforderlichen Invarianzuntersuchungen bisher nur für $p \leqq 2$ vollständig durchgeführt werden. Enthält $\{G;\alpha,\beta,v\}$ periodische Formen, die eine Fourierentwicklung der Art

$$\sum_T a(Y,T)e^{2\pi i Sp TX} \quad (Z = X + iY,\ \overline{Z} = X - iY)$$

gestatten, so entsteht das Problem, die möglichen Funktionen $a(Y,T)$ als Lösungen gewisser Differentialgleichungen zu bestimmen. Das gelingt noch im Falle $p = 2$. Doch ist über das funktionentheroretische Verhalten der Funktionen $a(Y,T)$ nichts Nennenswertes bekannt.

Der Wirkungsbereich der Modulgruppe p-ten Grades ist bekanntlich ein unbeschränktes Modell eines der vier allgemeinen komplexen Haupttypen von E. Cartans irreduziblen beschränkten symmetrischen Räumen. Es zeigt sich nun, daß für die andern Haupttypen Differentialgleichungen in analoger Bedeutung aufgestellt werden können, wenn man sich jeweils ein unbeschränktes Modell in geeigneter Weise verschafft. Wir greifen den Raum der rechteckigen Matrizen vom Typus $W = W^{(p,q)}$ heraus. Er wird beschrieben durch $I_q - W'\overline{W} > 0$; seine Fundamentalform lautet

$$ds^2 = Sp\, dW'(I_p - \overline{W}W')^{-1}\, d\overline{W}(I_q - W'\overline{W})^{-1}.$$

Die Fälle reeller und komplexer W können gleichzeitig behandelt werden. Wir setzen $\varepsilon = 1$ im reellen Fall und $\varepsilon = i$ im komplexen Fall. Unter der Voraussetzung $p \geqq q$ werde die Matrix $Q^{(p,q)}$ durch $R^{(p,p-q)}$ zu einer unitären Matrix (Q, R) ergänzt. Wir setzen dann

$$Z = 2(I_p + \overline{Q}Q')^{-1}(W + \varepsilon\overline{Q})(-\overline{\varepsilon}Q'W + I_q)^{-1},$$
$$Y = \tfrac{1}{2}(\overline{\varepsilon}Q'Z + \varepsilon\overline{Z}'\overline{Q}) - \tfrac{1}{4}\overline{Z}'\overline{R}R'Z,$$
$$X = \tfrac{1}{2}(Q'Z - \varepsilon^2\overline{Z}'\overline{Q}),$$
$$T = \frac{\overline{\varepsilon}}{2}R'Z$$

Kennzeichnend für unsern Raum ist

$$\overline{Y}' = Y > 0, \qquad \overline{X}' = -\varepsilon^2 X, \qquad T \text{ beliebig.}$$

Für die Bewegungen $Z \to (AZ + B)(CZ + D)^{-1}$ ist

$$\sigma' \iota \overline{\sigma} = \iota \qquad \text{mit } \sigma = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad \iota = \begin{pmatrix} -\frac{1}{2} R \overline{R}' & \varepsilon Q \\ \overline{\varepsilon} \overline{Q}' & O \end{pmatrix}$$

charakteristisch.

Im *komplexen Fall* ist nun mit $m = p + q$

$$\mathfrak{E}(Z; \alpha, \beta) = \sum_{\sigma} \omega(\sigma) |CZ + D|^{-\alpha} |\overline{CZ + D}|^{-\beta}.$$

Diese Reihen genügen den q^2 Differentialgleichungen

$$\mathfrak{Q}_{\alpha\beta} \mathfrak{E}(Z; \alpha, \beta) = 0,$$

wobei

$$Z = (z_{\mu\nu}), \quad \mathsf{D}_z = \left(\frac{\partial}{\partial z_{\mu\nu}}\right)$$

$$\mathfrak{Q}_{\alpha\beta} = ((QZ' - \overline{Z}\overline{Q}' + 2iR\overline{R}')\mathsf{D}_z)'\mathsf{D}_{\overline{z}} + \alpha Q'\mathsf{D}_{\overline{z}} - \beta(\overline{Q}'\mathsf{D}_z)'$$

ist. Hervorzuheben ist hier der hermitisch-symplektische Fall für $p = q$. Wird $Q = I_p$ gewählt, so ergeben sich einfache Formeln, insbesondere auch $Z = X + iY$, $\overline{Z}' = X - iY$.

Im *reellen Fall* ist mit $m = p + q$ und $\beta = 0$

$$\mathfrak{E}(Z, \alpha) = \sum_{\sigma} \omega(\sigma) \, | \, CZ + D \, |^{-\alpha}.$$

Man erhält $\dfrac{q(q+1)}{2}$ unabhängige Differentialgleichungen

$$\mathfrak{Q}_{\alpha} \mathfrak{E}(Z, \alpha) = 0,$$

wobei

$$\mathfrak{Q}_{\alpha} = ((ZQ' + QZ' + 2RR')\mathsf{D}_z)'\mathsf{D}_z + (\alpha + 1)Q'\mathsf{D}_z + (\alpha + 1)(Q'\mathsf{D}_z)'$$

ist. Von besonderem Interesse ist der Spezialfall $q = 1$. Setzt man noch $T' = (t_2, t_3, \ldots, t_p)$, $Y = t_1^2, t_1 > 0$, so wird

$$ds^2 = t_1^{-2}(dt_1^2 + dt_2^2 + \ldots + dt_p^2).$$

Es liegt hier also der p-dimensionale hyperbolische Raum vor. Man erhält so über den Matrizenkalkül einen neuen Zugang zu der Untersuchung [5]. Schließlich sei noch auf den Fall $q = 2$ mit drei Differentialgleichungen hingewiesen. Es handelt sich hierbei um eine reelle Darstellung eines weiteren komplexen Haupttypus von E. Cartan.

[1] SIEGEL, Math. Ann. 124 (1951/52), 17—54 u. 364—387.

[2] MAASS, Math. Ann. 121 (1949), 141—183.

[3] MAASS, Math. Ann. 122 (1950), 90—108.

[4] KOECHER, Journal f. rein. u. angew. Math. 192 (1953), 1—23.

[5] MAASS, Abh. a. d. Math. Sem. d. Univ. Hamburg 16, Heft 3/4 (1949), 72—100.

[6] MAASS, Math. Ann. 125 (1953), 235—263.

[7] MAASS, Sitz.-ber. d. Heidelberger Akad. d. Wiss., math.-naturwiss. Klasse 1949, 1. Abh.

[8] MAASS, Math. Ann. 126 (1953), 44—68.

MATH. INST., HAUPTSTRASSE 47/51,
HEIDELBERG, DEUTSCHLAND.

PROPRIÉTÉS ARITHMÉTIQUES DE CERTAINES FAMILLES DE COURBES ALGÉBRIQUES

A. NÉRON

See Symposium B. Algebraic Geometry

SPECIALIZATION METHODS IN ALGEBRAIC GEOMETRY

D. G. NORTHCOTT

See Symposium B. Algebraic Geometry

SECTION II

ANALYSIS

STATED ADDRESSES

FUNKTIONENTHEORIE AUF KOMPLEXEN MANNIGFALTIGKEITEN

HEINRICH BEHNKE

Eine der wesentlichen Einsichten, die wir Riemann [1]) verdanken, ist die Tatsache, dass die Gesamtheit der algebraischen Funktionselemente irgendeiner analytischen Funktion einer Veränderlichen, das analytische Gebilde dieser Funktion genannt, eine komplexe Mannigfaltigkeit $\mathfrak{M}^1$ über der geschlossenen z-Ebene bildet. Auf Poincaré [2]), Klein, Koebe, Hermann Weyl [3]) und Radó [4]) geht die Erkenntnis zurück, dass jede komplexe Mannigfaltigkeit der Dimension 1 das konforme Bild eines analytischen Gebildes ist.

In höheren komplexen Dimensionen betreffen die beiden Begriffe analytisches Gebilde und komplexe Mannigfaltigkeit im allgemeinen Fall nicht mehr dieselben Dinge. Es gibt analytische Gebilde, die keine komplexen Mannigfaltigkeiten sind. So hat das analytische Gebilde von $w = \sqrt{z_1 z_2}$, also einer besonders einfachen algebraischen Funktion, über $z_1 = z_2 = 0$ keinen sphärischen Punkt mehr. Insbesondere also gibt es dort kein Paar holomorpher Funktionen, das eine Umgebung dieses Punktes eineindeutig in den Raum C^2 von zwei komplexen Veränderlichen abbildet. Ebenso gibt es komplexe Mannigfaltigkeiten, auf denen als einzige dort meromorphe Funktionen die Konstanten vorkommen [5]).

A. *Kritische Punkte und Riemannsche Gebiete.*

In der Literatur werden die analytischen Gebilde der Funktionen $f(z_1, \ldots, z_n)$ durchweg ohne die nicht uniformisierbaren Punkte, hier kurz

[1]) Vergl. B. Riemann, Grundlagen für eine allg. Theorie der Funktionen einer veränderlichen complexen Größe, Gesammelte Werke pp. 3-47.

[2]) Vergl. H. Poincaré, Mémoire sur les fonctions fuchsiennes, Act. Mathem., **1** (1882), pp. 207.

[3]) Vergl. H. Weyl, Die Idee der Riemannschen Fläche, Teubnerverlag 1913.

[4]) Vergl. T. Radó, Über den Begriff der Riemannschen Fläche, Acta Szeged **2**, pp. 101-121.

[5]) Der (kompaktifizierte) Fundamentalbereich der Periodenmatrix.

$$\begin{pmatrix} 1 & 0 & i\sqrt{2} & i\sqrt{5} \\ 0 & 1 & i\sqrt{3} & i\sqrt{7} \end{pmatrix}$$

im C^2 hat nach C. L. Siegel diese Eigenschaft. Vergl. C. L. Siegel, Several complex variables, Princeton.

kritische Punkte genannt, betrachtet (das entspricht dem naiven Standpunkt in der klassischen Theorie, die algebraischen Verzweigungspunkte herauszunehmen). Doch sind neuerdings einige Arbeiten erschienen, in denen begonnen wird, auch verzweigte Gebiete zu untersuchen.

F. Hirzebruch [6]) nimmt aus den analytischen Gebilden $\mathfrak{G}$ der Funktionen $f(z_1, z_2)$ die in diesem Falle isoliert liegenden kritischen Punkte heraus und ersetzt sie durch Bäume von kompakten analytischen Flächen, so dass das abgeänderte Gebilde $\mathfrak{G}'$ eine komplexe Mannigfaltigkeit wird, in der die eingesetzten Flächen singularitätenfrei liegen. Dieser Prozess entspricht der in der algebraischen Geometrie bekannten Auflösung der Singularitäten algebraischer Flächen. Wichtig ist dabei die in der algebraischen Geometrie bekannte quadratische Transformation, die von H. Hopf unter der Bezeichnung σ-Prozess für beliebige komplexe Mannigfaltigkeiten der (komplexen) Dimension 2 definiert wurde. Eine analoge Theorie für die analytischen Gebilde der Funktionen von $n (n > 2)$ komplexen Veränderlichen steht noch aus.

Sodann gibt es zwei Arbeiten, in denen der Begriff der komplexen Mannigfaltigkeit so zum Begriff des Riemannschen Gebietes (espace analytique général) erweitert wird, dass die analytischen Gebilde aller $f(z_1, \ldots, z_n)$ darunter fallen.

Die in der ersten Arbeit [7]) eingeführten Riemannschen Gebiete $\mathfrak{G}$ haben die folgende Eigenschaft: Jeder Punkt besitzt eine Umgebung $\mathfrak{U}$, die zu einem analytischen Repräsentanten über dem endlichen Raume C^2 von n komplexen Veränderlichen analytisch äquivalent ist. Dabei ist ein analytischer Repräsentant eine endlichfache, analytisch verzweigte Überlagerung $\mathfrak{R}_{\mathfrak{U}}$ einer offenen Hyperkugel $\mathfrak{K}_{\mathfrak{U}}$ im C^n. Analytisch verzweigt heisst, dass die Fusspunkte der Verzweigungspunkte von $\mathfrak{R}_{\mathfrak{U}}$ eine analytische Fläche $f = 0$ in $\mathfrak{K}_{\mathfrak{U}}$ ausmachen.

Bei H. Cartan [8]) sind die analytischen Repräsentanten anders definiert: Jeder analytische Repräsentant ist analytisches Gebilde einer durch eine Pseudopolynomgleichung $w^k + A_1(z_1, \ldots, z_n)\, w^{k-1} + \ldots + A_k(z_1, \ldots, z_n) = 0$ definierten Funktion. Der so aufgestellte Begriff des espace analytique général ist per definitionem spezieller als der vorher eingeführte Begriff des Riemannschen Gebietes. Zuletzt zeigt es sich aber doch, dass beide Begriffe äquivalent sind [9]).

---

[6]) Vergl. F. Hirzebruch, Über vierdimensionale Riemannsche Flächen mehrdeutiger Funktionen von zwei komplexen Veränderlichen, Math. Ann. **126** (1953), pp. 1-22.

[7]) Vergl. H. Behnke und K. Stein, Modifikationen komplexer Mannigfaltigkeiten und Riemannscher Gebiete, Math. Ann. **124** (1951/52), pp. 1-16.

[8]) Vergl. H. Cartan, Séminaire 1950-51, Exposé 13.

[9]) Die Äquivalenz konnte in einer demnächst erscheinenden Arbeit von R. Remmert und H. Röhrl nachgewiesen werden.

B. *Funktionen auf Riemannschen Flächen.*

Im folgenden werden wir uns weiterhin nur noch mit der Funktionentheorie auf komplexen Mannigfaltigkeiten [10]) beschäftigen (also wie üblich die kritischen Punkte herauslassen). Wir beschränken uns dabei fast immer auf nicht kompakte komplexe Mannigfaltigkeiten, weil die kompakten Mannigfaltigkeiten zu besonderen umfangreichen Theorien, nämlich der der Kählerschen und der der algebraischen Mannigfaltigkeiten Anlass gegeben haben.

Die Funktionentheorie auf nicht kompakten komplexen Mannigfaltigkeiten sei hier zunächst für $n = 1$ behandelt und zwar in der modernen Sprache, in der wir sie dann auf $n > 1$ verallgemeinern können. Dabei wird ausschliesslich von *eindeutigen* Funktionen auf den jeweils betrachteten Mannigfaltigkeiten gesprochen.

Seit Poincaré ist bekannt, dass auf jeder $\mathfrak{M}^1$ eine meromorphe Funktion existiert. Dass es aber zu jeder offenen Riemannschen Fläche eine auf ihr holomorphe Funktion gibt, die nicht zu den trivialen konstanten Funktionen zählt, ist eine Aussage, die erst neuerdings bewiesen ist [11]).

Das gelingt mit Hilfe der verallgemeinerten Integralformel:

$$(1) \qquad f(z) = \frac{1}{2\pi i} \int_{\partial \mathfrak{G}} A(\zeta, z)\, f(\zeta)\, d\zeta,$$

wobei $A(\zeta, z)\, d\zeta$ ein meromorphes Differential mit einem Pol erster Ordnung und dem Residuum 1 für $\zeta = z$ in der Produktmannigfaltigkeit $^*\mathfrak{G} \times {}^*\mathfrak{G}$ ist. Dabei stehen die Bereiche [12]) $\mathfrak{G}$, $^*\mathfrak{G}$ und die Riemannsche Fläche $\mathfrak{R}$ in der Beziehung $\mathfrak{G} \subset {}^*\mathfrak{G} \subset \mathfrak{R}$[13]). $\mathfrak{A} \subset \mathfrak{B}$ heisst, dass $\mathfrak{A}$ eine relativkompakte Teilmenge von $\mathfrak{B}$ ist.

Das Integral in (1) wird ebenso wie das Cauchysche Integral beim Beweis des Rungeschen Satzes behandelt. So kommen wir zu dem Satz über die Approximation von Funktionen, die holomorph in gegebenen Bereichen $\mathfrak{G}$ sind, durch Funktionen, die noch in $\mathfrak{G}$ umfassenden Bereichen holomorph bleiben.

(*Approximationssatz*). *Es sei f in $\mathfrak{G}$ holomorph, es sei $\mathfrak{G}$ in bezug auf* $^*\mathfrak{G}$

---

[10]) Der Begriff der komplexen Mannigfaltigkeit taucht, so weit bekannt ist, zum ersten Mal in einer Arbeit von O. Teichmüller über „Veränderliche Riemannsche Flächen" auf, vergl. Deut. Math. 1944, pp. 344-359. Die Aufmerksamheit der Fachwelt wurde auf diesen Begriff durch die Vorträge und Aufsätze der Herren Heinz Hopf und Charles Ehresmann 1947 und ff. Jahre gelenkt.

[11]) H. Behnke und K. Stein, Entwicklung analytischer Funktionen auf Riemannschen Flächen, Math. Ann. **120** (1947/49), pp. 430-461.

[12]) Unter einem Bereich sei hier eine offene, nicht notwendig zusammenhängende Punktmenge in $\mathfrak{R}$ verstanden.

[13]) Selbstverständlich wird vom Rande von $\mathfrak{G}$ vorausgesetzt, dass er rektifizierbar ist.

relativ einfach zusammenhängend, d. h. jeder Zyklus aus $\mathfrak{G}$, *der in* $^*\mathfrak{G}$ *berandet, berandet auch in* $\mathfrak{G}$; *dann ist* $f(z)$ *in* $\mathfrak{G}$ *der Limes einer dort gleichmässig konvergierenden Folge* $f_n(z)$, *wobei die* $f_n(z)$ *in* $^*\mathfrak{G}$ *holomorph sind.*

Als $^*\mathfrak{G}$ kann auch $\mathfrak{R}$ gewählt werden, wenn $\mathfrak{G}$ in bezug auf $\mathfrak{R}$ einfach zusammenhängend ist. Das trifft insbesondere zu, wenn wir als $\mathfrak{G}$ eine oder mehrere sich nicht schneidende Kreisscheiben in $\mathfrak{R}$ wählen.

Hieraus folgt sofort, dass es auf jeder offenen Riemannschen Fläche $\mathfrak{R}$ holomorphe, nicht konstante, eindeutige Funktionen gibt. Man wähle nämlich in einer Kreisscheibe $f \equiv 0$, in einer anderen nicht schneidenden Kreisscheibe von $\mathfrak{R}$ die Funktion $f \equiv 1$. Dann sind die $f_n(z)$ für $n > n_0$ auf ganz $\mathfrak{R}$ holomorphe, nicht konstante Funktionen.

Nun ergeben sich die folgenden Eigenschaften von $\mathfrak{R}$ unmittelbar.

1. *Die Trennbarkeit der Punkte auf* $\mathfrak{R}$ *durch die holomorphen Funktionen.* Sind P_1 und P_2 verschiedene Punkte auf $\mathfrak{R}$, so gibt es immer eine auf $\mathfrak{R}$ holomorphe Funktion $f(z)$, so dass $f(P_1) \neq f(P_2)$.

2. *Die Existenz von überall auf* $\mathfrak{R}$ *holomorphen lokalen Uniformisierenden.* Zu jedem $P_1 \in \mathfrak{R}$ gibt es eine auf $\mathfrak{R}$ holomorphe Funktion $t(P)$, die eine Umgebung von P_1 uniformisiert.

3. $\mathfrak{R}$ *ist holomorphkonvex* [14]). D. h. zu jedem relativ kompakten Bereich $\mathfrak{B}$ von $\mathfrak{R}$ ist die Menge der Punkte P, für die in bezug auf alle in $\mathfrak{R}$ holomorphen Funktionen f gilt:

$$|f(P)| \leqq \text{Max } |f(\mathfrak{B})|,$$

eine kompakte Menge in $\mathfrak{R}$.

4. *Es gilt die verallgemeinerte Aussage von Mittag-Leffler für* $\mathfrak{R}$ [15]). Das bedeutet: Wenn auf der Riemannschen Fläche $\mathfrak{R}$ zu einer sich auf $\mathfrak{R}$ nicht häufenden Punktmenge $P_\nu, \nu = 1, 2, \ldots$, Hauptteile $G_\nu(z)$ vorgegeben werden, so gibt es immer auf $\mathfrak{R}$ eine meromorphe Funktion $F_0(z)$, so dass $F_0(z) - G_\nu(z)$ in einer Umgebung $\mathfrak{U}(P_\nu)$ holomorph ist. Zum Beweise wird die Funktion $A(\zeta_0, z)$ benutzt, die auf $\mathfrak{R}$ meromorph ist und nur im Punkte $z = \zeta_0$ einen Pol erster Ordnung mit dem Residuum 1 hat. Die konvergenzerzeugenden Summanden, die zur Konstruktion der Mittag-Lefflerschen Reihe erforderlich sind, liefert der oben zitierte Approximationssatz.

---

[14]) H. Behnke und K. Stein, Elementarfunktionen auf Riemannschen Flächen als Hilfsmittel für die Funktionentheorie mehrer Veränderlichen, Canad. Journ. of Mathem. **2** (1950), pp. 152-165.

[15]) Zum Nachweis der Eigenschaften 4, 5 und 6 auf offenen Riemannschen Flächen vergl. H. Behnke und K. Stein, Die Sätze von Weierstrass und Mittag-Leffler auf Riemannschen Flächen, Die Jahresschr. d. naturforsch. Ges. Zürich **85** (1940),
ferner H. Florack, Reguläre und meromorphe Funktionen auf nicht geschlossenen Riemannschen Flächen, Schriftenreihe des Math. Inst. d. Univ. Münster (Westf.), Heft 1, (1944).

Das vorstehende lässt sich auch als Lösung des sog. ersten Cousinschen Problems auf $\mathfrak{R}$ auffassen. Man weise den Punkten P_ν die obigen Lokalfunktionen $G_\nu(z)$ und den übrigen Punkten von $\mathfrak{R}$ die Lokalfunktionen $G_P(z) \equiv 0$ zu. Als $\mathfrak{U}(P)$ wird dann für $P = P_\nu$ die Umgebung $\mathfrak{U}(P_\nu)$ gewählt, im anderen Falle eine so kleine Umgebung, dass in ihr keine der Punkte P_ν liegen. Zu dieser Verteilung meromorpher Lokalfunktionen, die hier selbstverständlich der Verträglichkeitsbedingung genügt, gibt es dann die Funktion $F_0(z)$, so dass $F_0(z) - G_P(z)$ in $\mathfrak{U}(P)$ bzw. $F_0(z) - G_\nu(z)$ in $\mathfrak{U}(P_\nu)$ holomorph ist. Man sagt, es gelte in $\mathfrak{R}$ die erste Cousinsche Aussage.

5. *Ganz Analoges gilt für die Aussage von Weierstraß über die Existenz von Funktionen zu vorgegebenen Nullstellen* [16]). Wir drücken sie hier nach *P.* Cousin noch auf eine zweite Art aus: Zu jedem Punkt $P \in \mathfrak{R}$ wird eine lokale holomorphe Funktion $G_P(z)$ in $\mathfrak{U}(P)$ gegeben. In $\mathfrak{U}(P) \cap \mathfrak{U}(Q)$ soll $\dfrac{G_P(z)}{G_Q(z)}$ holomorph und von 0 verschieden sein. Dann gibt es auf $\mathfrak{R}$ eine holomorphe Funktion $F_1(z)$, so dass in den ausgezeichneten Umgebungen $\mathfrak{U}(P)$ die Funktion $\dfrac{F_1(z)}{G_P(z)}$ holomorph und ungleich 0 ist. Wir sagen, auf $\mathfrak{R}$ gelte die zweite Aussage von Cousin.

6. *Es gilt auf* $\mathfrak{R}$ *die Aussage von Poincaré* [17]): Ist $F(z)$ eine auf $\mathfrak{R}$ meromorphe Funktion, so gibt es auf $\mathfrak{R}$ zwei holomorphe Funktionen $g_1(z)$, $g_2(z)$, die lokal teilerfremd sind, so dass $F(z) = \dfrac{g_1(z)}{g_2(z)}$ auf $\mathfrak{R}$ ist.

C. *Holomorph vollständige komplexe Mannigfaltigkeiten.*

Für die komplexen Mannigfaltigkeiten höherer Dimension ist vieles ganz anders. L. Calabi und B. Eckmann [18]) haben eine offene komplexe Mannigfaltigkeit $\mathfrak{M}^n$ vom Typ der $2n$-dimensionalen Zelle angegeben, bei der der Ring der auf $\mathfrak{M}^n$ holomorphen Funktionen nur aus den Konstanten besteht. Offenbar können auf einer solchen Mannigfaltigkeit die Eigenschaften 1) — 5) aus *B.* nicht erfüllt werden. Für $n = 1$ ist die Zelle nur zweier komplexer Strukturen fähig, nämlich der des Kreises und der der offenen Ebene. In beiden Fällen gibt es genügend viel nichtkonstante holomorphe Funktionen, so dass 1) — 5) erfüllt werden können. Für $n > 1$ hat die $2n$-dimensionale Zelle kontinuierlich

---

[16]) K. Weierstraß, Zur Theorie der eindeutigen analytischen Funktionen, Math. Werke 2, pp. 77-124.

[17]) H. Poincaré, Sur les fonctions de deux variables, Act. Math. **2** (1883).

[18]) L. Calabi und B. Eckmann, A class of compact complex manifolds, which are not algebraic, Ann. of Mathem **58** (1953), pp. 494-500.

viele komplexe Strukturen, darunter die von den eben zitierten Autoren entdeckten funktionsarmen Strukturen.

Somit müssen wir, um auf einer $\mathfrak{M}^n$ Funktionentheorie treiben zu können, besondere Forderungen an $\mathfrak{M}^n$ stellen. Hierzu ist von K. Stein der Begriff der holomorph vollständigen (komplexen) Mannigfaltigkeit eingeführt worden [19]). Eine komplexe Mannigfaltigkeit $\mathfrak{M}^n$ heisst eine holomorph vollständige Mannigfaltigkeit, wenn die Bedingungen 1) — 3) von *B.* erfüllt sind und ausserdem $\mathfrak{M}^n$ eine abzählbare Basis hat. Die Bedingung 2 ist dabei in folgendem Sinne auf mehrere Veränderliche zu übertragen: Zu jedem Punkte $P_1 \in \mathfrak{M}^n$ muss es n in $\mathfrak{M}^n$ holomorphe Funktionen geben, die eine Umgebung $\mathfrak{U}(P)$ eineindeutig auf ein Gebiet im C^n abbilden. Das Abzählbarkeitsaxiom können wir hier nicht vermeiden. T. Radó [20]) hatte ja in seiner viel beachteten Arbeit gezeigt, dass es auf einer $\mathfrak{M}^1$ entbehrlich ist. Andererseits haben H. Hopf und auch L. Calabi und M. Rosenlicht [21]) Beispiele von komplexen Mannigfaltigkeiten angegeben, die keine abzählbare Basis besitzen [22]).

Auf den holomorph vollständigen Mannigfaltigkeiten gilt wieder ein Approximationssatz:

Ist $\mathfrak{G}^n$ eine holomorph vollständige Mannigfaltigkeit, die Teil einer holomorph vollständigen Mannigfaltigkeit $\mathfrak{M}^n$ ist, und ist $\mathfrak{G}^n$ auf $\mathfrak{M}^n$ halbstetig über lauter holomorph vollständige Mannigfaltigkeiten ausdehnbar, so lässt sich jede in $\mathfrak{G}^n$ holomorphe Funktion im Innern von $\mathfrak{G}^n$ durch in $\mathfrak{M}^n$ holomorphe Funktionen approximieren [23]).

---

[19]) Vergl. K. Stein, Analytische Funktionen mehrerer komplexer Veränderlichen und das zweite Cousin'sche Problem, Math. Ann. **123** (1951), pp. 201-222,
und die Arbeiten von
H. Cartan, Seminaire 1951/52, Exposé IX,
H. Cartan, Variétés analytiques complexes et cohomologie, Colloque sur les fonctions de plusieurs variables, (Brüssel 1953),
J. P. Serre, Quelques problèmes globaux relatifs aux variétés de Stein, Colloque sur les fonctions de plusieurs variables, (Brüssel 1953).

[20]) vergl. T. Radó, a.a.O. [4]).
In dem neuen Buch von R. Nevanlinna wird ein weiterer von A. Pfluger angegebener Beweis durchgeführt.

[21]) E. Calabi und M. Rosenlicht, Complex analytic manifolds without countable base, Proc. of the American Math. Soc. **4**, 335-340 (1953).

[22]) Die konstruierten Gegenbeispiele sind keine holomorph vollständigen Mannigfaltigkeiten. Es ist zur Zeit noch offen, ob es solche gibt. Zus z. Korr.: Das Problem konnte inzwischen von H. Grauert gelöst werden. Vgl. Math. Ann. 129 (Juni 1955).

[23]) H. Behnke, Généralisation du théoreme de Runge pour les fonctions multiformes de variables complexes, Colloque sur les fonctions de plusieurs variables, (Brüssel, 1953).

Ferner gilt die erste Cousinsche Aussage. K. Oka [24]) hatte dieses schon 1936 für den speziellen Fall schlichter endlicher Holomorphiegebiete im C^n bewiesen. Bei diesem Beweis trat zum ersten Mal ein neuer sehr fruchtbarer Gedanke auf. Holomorphiegebiete können von innen approximiert werden durch analytische Polyeder $\mathfrak{A}$, das sind Gebiete im C^n, die charakterisiert werden können durch $f_j(z_1, \ldots, z_n) \in \mathfrak{R}_j, j = 1, \ldots, n$ ($\mathfrak{R}_j$ ein Gebiet im C^1, f_j holomorph in $\mathfrak{R}$, $\mathfrak{A} \in \mathfrak{R} \subset C^n$). Solche analytischen Polyeder $\mathfrak{A}$ werden als analytische Flächen in einen höherdimensionalen Zylinderbereich eingebettet. Dadurch gelingt es, die in $\mathfrak{A}$ gegebenen Funktionen in einfacher Weise zu approximieren. Dieses Prinzip ist später dann noch sehr ausgebaut worden. H. Cartan hat die Gültigkeit der ersten Aussage von Cousin auf allen holomorph vollständigen Mannigfaltigkeiten nachgewiesen [25]).

Die Frage, ob zu allen möglichen Nullstellenverteilungen, die den Verträglichkeitsbedingungen genügen, auf einer gegebenen holomorph vollständigen Mannigfaltigkeit holomorphe Funktionen existieren, ist schwerer zu beantworten. Schon 1917 hat T. Gronwall [26]) in einem ersten Gegenbeispiel gezeigt, dass es Zylindergebiete im C^n gibt, die zu vorgegebenen Nullstellenflächen keine holomorphen Funktionen aufweisen. Diese Zylindergebiete sind selbst holomorph vollständige Mannigfaltigkeiten. K. Oka hat dann später nachgewiesen, dass, wenn es zu einer vorgegebenen Nullstellenverteilung in einem endlichen Holomorphiegebiet (und das ist immer eine holomorph vollständige Mannigfaltigkeit) eine stetige Funktion gibt, auch dort eine holomorphe Funktion mit den vorgegebenen Nullstellen konstruiert werden kann [27]).

Daraus folgt, dass die zusätzlichen Bedingungen, die erforderlich sind, damit in einem gegebenen Holomorphiegebiet die zweite Aussage von Cousin für eine dort vorgegebene Nullstellenverteilung gilt, nur topologischer Natur sind. K. Stein hat solche topologischen Bedingungen angegeben, die teils notwendig und teils hinreichend sind [28]). H. Cartan und J. P. Serre haben neuer-

---

[24]) Vergl. K. Oka, Sur les fonctions de plusieurs variables. I-Domaines convexes par rapport aux fonctions rationelles, Ser. A, Vol. **6**, Nr. 3 (1936). II — Domaines d'holomorphie, Ser. A, Vol. **7**, Nr. 2 (1937). Beides: Journal of Science of the Hirosima University.

[25]) H. Cartan, a.a.O. [19]).

[26]) Vergl. T. H. Gronwall, On the expressibility of a uniform function of several complex variables as the quotient of two functions of entire character, Americ. M. S. Trans. **18**, 50-64 (1917).

[27]) Vergl. K. Oka, Sur les fonctions de plusieurs variables. III — Deuxième problème de Cousin, Journ. of Science of the Hiros. Univ. **19**, 7-19.

[28]) Vergl. K. Stein, Topologische Bedingungen für die Existenz analytischer Funktionen komplexer Veränderlichen zu vorgegebenen Nullstellen, Math. Ann. **117** (1940/41), pp. 727-757.

dings für alle holomorph vollständigen Mannigfaltigkeiten eine notwendige und hinreichende Kohomologiebedingung aufgestellt [29]).

Die (schwache) Aussage von Poincaré gilt auf jeder holomorph vollständigen Mannigfaltigkeit, d. h. jede dort meromorphe Funktion lässt sich als Quotient zweier holomorpher Funktionen darstellen. Dabei kann es aber vorkommen, dass Zähler und Nenner nicht lokal teilerfremd sind [30]).

Die beiden Cousinschen Aussagen und die Aussage von Poincaré hängen eng miteinander zusammen. Wenn auf einer $\mathfrak{M}^n$ die erste Aussage von Cousin (in einer verschärften Form) und die Aussage von Poincaré (in der scharfen Form) gelten, so ist auch dort die zweite Aussage von Cousin richtig. Von den acht formal verschiedenen Möglichkeiten der Gültigkeit dieser drei Aussagen können de facto nur fünf auftreten [31]).

Die Nullstellen einer holomorphen nicht identisch verschwindenden Funktion bilden Flächen von der komplexen Dimension $n - 1$. Betrachtet man die simultanen Nullstellen von k holomorphen Funktionen, so kommt man zu niederdimensionalen analytischen Flächen. Unter einer analytischen Menge in $\mathfrak{M}^n$ verstehen wir eine in $\mathfrak{M}^n$ abgeschlossene Punktmenge, bei der es zu jedem ihrer Punkte P endlich viele in einer Umgebung $\mathfrak{U}(P)$ holomorphe Funktionen gibt, deren gemeinsame Nullstellen in $\mathfrak{U}(P)$ die Punkte der analytischen Menge sind. Wieder tritt die Frage nach der globalen Darstellung analytischer Mengen auf. Damit haben sich H. Cartan und K. Oka beschäftigt [32]).

Eine analytische Menge $\mathfrak{N}$ in $\mathfrak{M}^n$ bestimmt in jedem Punkte $P \in \mathfrak{M}^n$ ein Ideal $\mathfrak{J}$ im Ringe der in P holomorphen Funktionen: $\mathfrak{J}$ besteht genau aus denjenigen in P holomorphen Funktionen, die in einer Umgebung von P auf $\mathfrak{N}$ verschwinden. Diese Ideale genügen, wie H. Cartan nachgewiesen hat [32a]), einer gewissen Kohärenzbedingung. Wird umgekehrt in jedem Punkte $P \in \mathfrak{M}^n$ ein Ideal im Ringe der in P holomorphen Funktionen vorgegeben und genügen

---

[29]) J. P. Serre, a.a.O. [19]).

[30]) J. P. Serre, a.a.O. [19]).

[31]) H. Behnke und K. Stein, Analytische Funktionen mehrerer Veränderlichen zu vorgegebenen Null- und Polstellenflächen, Jahresber. d. D.M.V. **47** (1937 .

[32]) H. Cartan, I. Sur les matrices holomorphes de n variables complexes, Journ. de Math. pura et appl. **19**, 1-26 (1940),

II. Idéaux de fonctions analytiques de n variables complexes, Ann. Ecole Norm. Sup. **61**, 149-197 (1944),

III. Idéaux et modules de fonctions analytiques de variables complexes, Bull. Soc. Math. de France **78**, 28-64 (1950).

K. Oka, Sur les fonctions analytiques de plusieurs variables, VII. — Sur quelques notions arithmétiques, Bull. Soc. Math. de France **78**, 1-27 (1950).

[32a]) Siehe H. Cartan, a.a.O. [32]) III.

wieder diese Ideale der erwähnten Kohärenzbedingung, so ergibt sich die Frage, ob es ein Ideal im Ringe der in ganz $\mathfrak{M}^n$ holomorphen Funktionen gibt, welches in jedem Punkte $P \in \mathfrak{M}^n$ das dort gegebene Ideal erzeugt. Für diesen Schluss vom Lokalen ins Globale hat H. Cartan die Theorie der analytischen Garben (faisceaux analytiques, engl. sheaf) aufgestellt. Aus seinen sehr weitreichenden Ergebnissen [33]) folgt speziell, dass in einer holomorph vollständigen $\mathfrak{M}^n$ ein globales Ideal der gesuchten Art stets existiert. Hieraus ergibt sich unmittelbar:

Jede analytische Menge einer holomorph vollständigen Mannigfaltigkeit ist in jedem relativ kompakten Teilgebiet die gemeinsame Nullstellenmenge von endlich vielen auf $\mathfrak{M}^n$ holomorphen Funktionen. H. Grauert hat darüber hinaus nachgewiesen, dass dieses sogar global für ganz $\mathfrak{M}^n$ gilt.

Die ältesten in der Literatur vorkommenden holomorph vollständigen Mannigfaltigkeiten sind die Holomorphiegebiete *über* dem C^n. Um einzusehen, dass alle diese Gebiete tatsächlich solche Mannigfaltigkeiten sind, muss nur noch gezeigt werden, dass sie die Eigenschaft 3 der Holomorphiekonvexität haben. Bei endlichblättrigen beschränkten Gebieten folgt dies unittmelbar aus der klassischen Arbeit von H. Cartan und P. Thullen [34]). Später ist es dann für alle fast-endlichblättrigen Gebiete bewiesen worden [35]).

In einer soeben erschienenen Arbeit von K. Oka [36]) wird es nun für alle beschränkten Gebiete nachgewiesen. Die Resultate dieser Arbeit reichen sogar noch sehr viel weiter. Schon E. E. Levi [37]) hatte lokal hinreichende Bedingungen für Holomorphiegebiete angegeben. Seitdem ist die Frage, ob das Erfülltsein dieser lokalen Bedingungen in jedem Randpunkt eines Gebietes $\mathfrak{G}$ dafür hinreicht, dass $\mathfrak{G}$ Holomorphiegebiet ist, in der Literatur nicht zur Ruhe gekommen. Schliesslich bewies K. Oka 1942 [38]), dass dieses für Gebiete im C^2

---

[33]) Die Hauptsätze von H. Cartan sind als Aussagen über die Kohomologiegruppen in $\mathfrak{M}$ mit Koeffizienten in einer kohärenten analytischen Garbe formuliert. Vergl. H. Cartan, a.a.O. [19]), Théorème A und B.

[34]) H. Cartan und P. Thullen, Regularitäts- und Konvergenzbereiche, Math. Ann. **106** (1932).

[35]) H. Behnke und K. Stein, Konvergente Folgen nicht schlichter Regularitätsgebiete, Ann. di Mat. pura et appl. (4), **28** (1949).

[36]) Vergl. K. Oka, Sur les fonctions de plusieurs variables. IX — Domaines finis, sans point critique intérieur, Jap. Journ. Math. 1953.

[37]) Vergl. E. E. Levi, Studii sui punti singolari essenziali delle funzioni anal. di due o più var. compl., Ann. Mat. pura et app. (3) **17** (1910).

[38]) Vergl. K. Oka, Sur les fonctions de plusieurs variables. VI — Domaines pseudoconvexes, Tôhoku Math. Journ. **49**, 15-52 (1942).

zutrifft. In den letzten Jahren ist von F. Norguet [39]) und H. Bremermann [40]) gleiches für die schlichten Gebiete im C^n bewiesen worden. Für beliebige beschränkte Gebiete über dem C^n wird der Beweis in der obigen Arbeit von K. Oka nunmehr geliefert. Damit ist mittelbar auch bewiesen:

Jede beliebige (unverzweigte) Überlagerung eines Holomorphiegebietes ist wieder Holomorphiegebiet.

D. *Analytische Projektionen.*

Sind auf einer komplexen Mannigfaltigkeit $\mathfrak{M}^n$, die nun wieder nicht mehr notwendig holomorph vollständig zu sein braucht, k unabhängige meromorphe Funktionen $f_1, \ldots, f_k$ gegeben und wird eine Funktion F betrachtet, die in jedem Punkte von $\mathfrak{M}^n$ von den $f_1, \ldots, f_k$ analytisch abhängig ist, so bedient man sich zweckmässig des von K. Stein eingeführten Begriffs der analytischen Projektion [41]).

Durch die Gleichungen $w = f_j(P)$, $j = 1, \ldots, k$, wird eine meromorphe Abbildung Φ von $\mathfrak{M}^n$ in den nach Osgood abgeschlossenen Raum $\overline{C}^k$ bestimmt. Diese Abbildung ordnet jedem Punkt von $\mathfrak{M}^n$, in dem alle f_j holomorph sind oder höchstens einen Pol besitzen, einen Punkt als Bild zu; jedem Punkt von $\mathfrak{M}^n$, der für wenigstens ein f_j Unbestimmtheitsstelle ist, entspricht dagegen vermöge Φ eine i. a. mehrpunktige Bildmenge, von der W. Thimm [42]) nachgewiesen hat, dass sie stets eine in $\overline{C}^k$ algebraische Menge ist. Es wird nun gefragt, ob es ein k-dimensionales Riemannsches Gebiet $\mathfrak{R}^k$ über dem $\overline{C}^k$ mit folgenden Eigenschaften gibt:

1. Es existiert eine meromorphe Abbildung Ψ von $\mathfrak{M}^n$ auf $\mathfrak{R}^k$, so dass $\Phi = p \cdot \Psi$ gilt, wo p die Projektion von $\mathfrak{R}^k$ in seine Grundpunkte im C^k bezeichnet.
2. Ist F irgendeine von den $f_1, \ldots, f_k$ auf $\mathfrak{M}^n$ abhängige meromorphe Funktion, so gibt es eine auf $\mathfrak{R}^k$ meromorphe Funktion $^*F$ mit $F = {}^*F \cdot \Psi$.

---

[39]) Vergl. F. Norguet, Sur les domaines d'holomorphie des fonctions uniformes de plusieurs variables complexes (Passage du local au global), Bull. Soc. Math. de France **82**, 137-159 (1954).

[40]) Vergl. H. Bremermann, Über die Äquivalenz der pseudokonvexen Gebiete und der Holomorphiegebiete im Raum von n komplexen Veränderlichen, Math. Ann. **128** (1954), pp. 63-91.

[41]) Vergl. K. Stein, Analytische Projektion komplexer Mannigfaltigkeiten, Colloque sur les fonctions de plusieurs variables, Brüssel 1953, pp. 97-157, sowie eine weitere zu diesem Thema in den Math. Ann. erscheinende Arbeit. — K. Koch, Die analytische Projektion, Schriftenreihe des Math. Inst. der Univ. Münster (Westf.), 1953. — R. Remmert, Holomorphe und meromorphe Abbildungen komplexer Räume. Erscheint in den Math. Ann.

[42]) W. Thimm, Über die Menge der singulären Bildpunkte einer meromorphen Abbildung, Math. Zeitschr. **57**, 456-480 (1953).

Falls ein solches $\mathfrak{R}^k$ existiert, ist es (bei geeigneter Normierung) bis auf analytische Homöomorphie eindeutig bestimmt. Im Falle der Existenz heisst $\mathfrak{R}^k$ die analytische Projektion von $\mathfrak{M}^n$ vermöge Φ. Ist $\mathfrak{M}^n$ nicht kompakt, so braucht die analytische Projektion nicht zu existieren. Sie ist aber stets dann vorhanden, wenn die Dimension der Fasern der durch $f_1, \ldots, f_k$ bestimmten Abbildung nie grösser als $n-k$ ist (d. h. wenn keine Entartungsstellen vorliegen). Für $k = 1$ ist diese Bedingung immer erfüllt.

Sehr ergiebig ist es, analytische Projektionen von kompakten komplexen Mannigfaltigkeiten zu betrachten. Diese Projektionen existieren ohne einschränkende Voraussetzung über die Entartung. Die Projektion $\mathfrak{R}^k$ ist dann kompakt und folglich endlichblättrig über C^k. Daraus folgt [43]):

1. *Auf kompakten komplexen Mannigfaltigkeiten sind analytisch abhängige Funktionen auch algebraisch abhängig.*
2. *Der Körper der meromorphen Funktionen auf einer kompakten komplexen Mannigfaltigkeit ist isomorph einer endlichfach algebraischen Erweiterung des Körpers der rationalen Funktionen in k Unbestimmten über dem C^1. Mit den Unbestimmten im Sinne der Algebra sind die Funktionen $f_1, \ldots, f_k$ (wobei k die maximale Anzahl der auf $\mathfrak{M}^n$ unabhängigen meromorphen Funktionen ist) zu identifizieren.*

Die weitergehenden Untersuchungen auf kompakten Mannigfaltigkeiten führen zu den Kählerschen und den algebraischen Mannigfaltigkeiten und damit zu den modernen Untersuchungen von W. V. D. Hodge, A. Weil, H. Cartan, J. P. Serre, K. Kodaira, D. Spencer, F. Hirzebruch.

E. *Modifikationen.*

Der in den vorstehenden Betrachtungen häufig auftretende Raum von n komplexen Veränderlichen wird meistens zweckmässig als abgeschlossener Raum betrachtet. Doch ist im Gegensatz zur klassischen Funktionentheorie die Abschliessung hier nicht eindeutig. In der Literatur kommen vor allem zwei Abschliessungen vor, der sog. komplex projektive Raum, der die Transformationen $z_j' = \dfrac{L_j(z_1, \ldots, z_n)}{L_0(z_1, \ldots, z_n)}$, $j = 1, \ldots, n$, zulässt wobei L_j ganze lineare Funk-

[43]) Die folgenden Resultate wurden früher von W. L. Chow und W. Thimm auf anderem Wege gewonnen. Vergl. zu 1): W. Thimm, Über die algebraischen Relationen zwischen meromorphen Funktionen in abgeschlossenen Räumen, Dissert. Königsberg 1939, und: Über meromorphe Abbildungen komplexer Mannigfaltigkeiten, Math. Ann. **128**, 1-48 (1954). Zu 2) vergl.: W. L. Chow u. K. Kodaira, On analytic surfaces with two independent meromorphic functions, Proc. of the Nat. Acad. of Sciences **38**, 319-325 (1952). — Wegen weiterer Beweise siehe R. Remmert, a.a.O. [41]) sowie K. Stein, Analytische Abbildungen allgemeiner analytischer Räume, Colloque de Topologie de Strasbourg, April 1954.

tionen sind und die Koeffizientenmatrix der L_j vom Range $n+1$ ist, und der Osgood'sche Raum, der in jeder einzelnen komplexen Veränderlichen getrennt abgeschlossen wird und dementsprechend die Transformationen $z_j' = \frac{a_j z_j + b_j}{c_j z_j + d_j}$, $j = 1, \ldots, n$, $\Delta_j \neq 0$, in sich zulässt. Schliesslich kommt man unter Benutzung eines Resultates von L. Bieberbach [44]) über Abbildungen des C^n auf sich zu einem weiteren Abschluss, bei dem überunendlichferne Punkte auftreten. Offenbar gibt es noch unendlich viele weitere Möglichkeiten, den C^n abzuschliessen.

Man wird danach trachten, von den Aussagen über das Verhalten der Funktionen in den unendlichfernen Punkten des einen Raumes auf das in den unendlichfernen Punkten des anderen Raumes schliessen zu können. Das hat zum Begriff der Modifikation [45]) geführt, dem in letzter Zeit zahlreiche Arbeiten gewidmet sind.

Wir sagen, eine komplexe Mannigfaltigkeit $\mathfrak{M}^n$ wird zu einer komplexen Mannigfaltigkeit $^*\mathfrak{M}^n$ modifiziert, wenn aus $\mathfrak{M}^n$ eine abgeschlossene Teilmenge $\mathfrak{N}$ einer analytischen Menge herausgenommen wird und stattdessen eine neue Menge $^*\mathfrak{N}$ so zugefügt wird, dass sich die unveränderte komplexe Struktur von $\mathfrak{M}^n - \mathfrak{N}$ zu der komplexen Struktur in $(\mathfrak{M}^n - \mathfrak{N}) \cup {}^*\mathfrak{N} = {}^*\mathfrak{M}^n$ fortsetzen lässt und dass es zu jeder Umgebung $\mathfrak{U}(\mathfrak{N})$ in $\mathfrak{M}^n$ eine Umgebung $\mathfrak{U}(^*\mathfrak{N})$ in $^*\mathfrak{M}^n$ gibt, so dass gilt: $\mathfrak{U}(\mathfrak{N}) - \mathfrak{N} \supset \mathfrak{U}(^*\mathfrak{N}) - {}^*\mathfrak{N}$ [46]).

Setzt man die Existenz einer in $\mathfrak{U}(\mathfrak{N})$ holomorphen, auf $\mathfrak{N}$ verschwindenden, jedoch nicht identisch verschwindenden Funktion voraus, so gilt:

1. *$^*\mathfrak{N}$ ist in einer $(n-1)$-dimensionalen analytischen Menge enthalten;*
2. *jede in einer Umgebung von $\mathfrak{N}$ holomorphe Funktion lässt sich nach der Modifikation in $^*\mathfrak{N}$ fortsetzen. Verschwindet sie auf $\mathfrak{N}$, so verschwindet sie auch auf $^*\mathfrak{N}$;*
3. *wenn $\mathfrak{N}$ charakterisiert wird durch s holomorphe Gleichungen $f_j = 0$, $j = 1, \ldots, s$, so ist $^*\mathfrak{N}$ eine analytische Menge und wird in $^*\mathfrak{M}^n$ durch dieselben Gleichungen charakterisiert.*

Wir sprechen von einer stetigen Modifikation, wenn die Identität $^*\mathfrak{M}^n - {}^*\mathfrak{N} \to \mathfrak{M}^n - \mathfrak{N}$ als stetige und damit holomorphe Abbildung in $^*\mathfrak{M}^n$ fortsetzbar ist.

---

[44]) Vergl. L. Bieberbach, Beispiel zweier ganzer Funktionen zweier komplexer Variablen, welche eine schlichte volumentreue Abbildung des R_4 auf einen Teil seiner selbst vermitteln, Sitzgsber. Preuß. Akad. Wiss., Phys. math. Kl. 1933.

[45]) Vergl. H. Behnke und K. Stein, Modifikationen komplexer Mannigfaltigkeiten und Riemannscher Gebiete, Math. Ann. **124** (1951), pp. 1-16.

[46]) Vergl. H. Behnke und K. Stein, a.a.O. [45]).

Eine solche Modifikation heisst eine eigentliche Modifikation, wenn die fortgesetzte Abbildung im Sinne von Bourbaki eigentlich ist, also die Urbilder kompakter Mengen kompakt sind. Dann gilt:

1. *Der Körper der meromorphen, sowie die Ringe der holomorphen Funktionen sind zueinander kanonisch isomorph* [47]).
2. *Ist* $\mathfrak{N}$ *singularitätenfrei,* $\mathfrak{N}$ *und* $\mathfrak{N}^*$ *irreduzibel, so ist jede eigentliche Modifikation ein verallgemeinerter* σ*—Prozess* [48]).

Von dem von H. Hopf [49]) für zweidimensionale komplexe Mannigfaltigkeiten $\mathfrak{M}^2$ definierten σ-Prozess war schon in A. die Rede. Er besteht darin, dass auf $\mathfrak{M}^2$ an stelle eines Punktes eine geschlossene projektive Ebene, Trägersphäre genannt, einzusetzen ist. Man spricht von einem verallgemeinerten σ-Prozess, wenn auf einer $\mathfrak{M}^n$ für die Punkte einer singularitätenfrei eingelagerten, k-dimensionalen analytischen Fläche $\mathfrak{N}(k \leqq n-2)$ ein $(n-1)$-dimensionaler Faserraum $^*\mathfrak{N}$ eingesetzt wird, dessen Basis die Fläche und dessen Faser der $(n-k-1)$-dimensionale komplexe projektive Raum ist [50]).

3. *Ist die eigentliche Modifikation so definiert, dass beim Übergang von* $\mathfrak{N}$ *zu* $^*\mathfrak{N}$ *kein Punkt von* $\mathfrak{N}$ *nur durch einen Punkt ersetzt wird, so ist* $^*\mathfrak{N}$ *rein* $(n-1)$*-dimensional und die Dimension von* $\mathfrak{N}$ *kleiner gleich* $n-2$ [51]).

Wie schon unter A. ausgeführt, hat F. Hirzebruch gezeigt, dass das analytische Gebilde einer meromorphen Funktion $f(z_1, z_2)$ durch eine eigentliche Modifikation in eine komplexe Mannigfaltigkeit überführt werden kann, derart dass auch keine Unbestimmtheitsstellen von f zurückbleiben. Schliesslich ist zu erwähnen, dass W. Stoll [52]) umfangreiche Untersuchungen über meromorphe Modifikationen komplexer Mannigfaltigkeiten $\mathfrak{M}^n$ im Falle $n = 2$ soeben durchgeführt hat. Damit hat der Begriff der Modifikation in den wenigen Jahren seit seiner Aufstellung schon zu vielseitigen Betrachtungen geführt.

---

[47]) Vergl. E. Kreyszig, Stetige Modifikationen komplexer Mannigfaltigkeiten, Math. Ann. 128, 479—492 (1955).

[48]) Vergl. H. Grauert und R. Remmert, Zur Theorie der Modifikationen I, Math. Ann. 129, 274—296 (1955).

[49]) In der algebraischen Geometrie ist dieser Prozess der Modifikation für die dort vorliegenden speziellen Fälle schon lange benutzt worden (Monoidale Transformation).

[50]) Umfangreiche Untersuchungen über die Natur des Faserraumes sind von C. Segre aufgestellt.

Vgl. ferner auch E. Kreyszig, a.a.O. [47]).

[51]) Vergl. H. Grauert und R. Remmert, a.a.O. [48]).

[52]) Vergl. W. Stoll, Über meromorphe Modifikationen, Habilitationsschrift, Tübingen 1954.

LES SOLUTIONS ÉLÉMENTAIRES ET LE PROBLÈME DE CAUCHY

F. J. BUREAU

1. Désignons par $(x) = (x_1, \ldots, x_p)$ des variables spatiales, par $t = x_{p+1}$, le temps et par u, une fonction inconnue. Considérons une équation linéaire aux dérivées partielles $Lu = 0$ à $p+1$ variables indépendantes (x, t) et proposons-nous de choisir parmi toutes les solutions possibles de cette équation, une solution particulière vérifiant certaines conditions accessoires (linéaires) $Au = 0$. Des exemples simples fournis par la Physique mathématique suggèrent d'étudier deux types particuliers d'équations linéaires, savoir les équations totalement elliptiques et les équations totalement hyperboliques. Les premières correspondent à des phénomènes d'équilibre, les secondes à des phénomènes de propagation par ondes.

Le problème type relatif aux équations elliptiques est le problème de Dirichlet dans lequel on propose de rechercher une solution de l'équation $\Delta_p u = 0$ (Δ_p: laplacien à p dimensions) régulière à l'intérieur d'un domaine V à p dimensions et prenant sur la frontière $\dot{V}$ de V des valeurs données. On sait que dans des conditions très générales, ce problème possède une solution unique d'ailleurs instable.

Le problème type relatif aux équations hyperboliques est le problème de la propagation des ondes sonores dans un milieu homogène. Dans ce problème, on se propose de rechercher une solution de l'équation $u_{tt} - \Delta_3 u = 0$ telle que u et u_t se réduisent pour $t = 0$ à des fonctions données de (x) (Problème de Cauchy) En général, ce problème possède une solution unique, stable.

2. De l'étude des problèmes particuliers que nous venons d'indiquer, on conclut que la solution u est une fonctionnelle linéaire des données. La dualité doit donc jouer un rôle important dans cette étude. Cette dualité peut s'exprimer de deux manières et se traduit dans la théorie des équations aux dérivées partielles par deux méthodes. Ou bien, on peut utiliser une formule de réciprocité généralisant la formule de Green; ou bien, on peut, en représentant une fonction par une intégrale de Fourier ou de Laplace ou encore, si elle est analytique, par une intégrale de Cauchy, chercher à généraliser le calcul symbolique de Cauchy et Heaviside. Cette seconde méthode s'applique à des équations à coefficients constants et à des données de Cauchy sur l'hyperplan $t = 0$; elle a fait l'objet dans ces dernières années de nombreuses et importantes recherches

(Fantappié, Herglotz, Petrowsky, Gårding, Leray). La première méthode, au contraire, peut aussi bien être utilisée pour des équations à coefficients variables et des données portées par une hypersurface ne se réduisant pas à un hyperplan.

Chacune de ces méthodes a conduit à définir des solutions dites „élémentaires"; mais ces diverses définitions n'étant pas équivalentes, il en résulte une certaine confusion.

Nous chercherons ici à indiquer brièvement les résultats obtenus en utilisant la première méthode, de manière à faciliter la comparaison avec les résultats obtenus plus récemment à partir de la seconde méthode.

3. Considérons

1) un opérateur différentiel linéaire $L = L_x$ d'ordre n à p variables indépendantes $x_1, \ldots, x_p$:

2) un domaine connexe V à p dimensions dont nous supposerons la frontière $\dot{V}$ assez régulière pour pouvoir appliquer le théorème de la divergence; nous désignerons par $\pi = (\pi_1, \ldots, \pi_p)$ la normale à $\dot{V}$ dirigée positivement vers l'intérieur de V;

3) des fonctions $u(x)$, $v(x)$ régulières (d'ordre s) dans $V + \dot{V}$ c'est-à-dire continues ainsi que leurs dérivées partielles jusqu'à un ordre s assez élevé pour justifier les opérations effectuées ultérieurement. Pour abréger, nous ne préciserons pas cet ordre s bien qu'il n'y ait là en général, aucune difficulté.

Désignons par L^* l'opérateur différentiel linéaire d'ordre n tel que l'on ait la relation

$$\int_V (vLu - uL^*v)dx = -\int_{\dot{V}} P(u, v; \pi)d\sigma \tag{1}$$

où $d\sigma$ désigne l'élément de surface de $\dot{V}$ et où $P(v) = P(u, v; \pi) = \sum_{i=1}^{p} \pi_i P_i(u, v)$ est une expression bilinéaire en les dérivées partielles de u et de v.

La relation précédente est la formule de réciprocité et généralise la formule bien connue de Green. L^* est l'adjointe de L et $L^{**} = L$.

Désignons maintenant par L_n les termes de L contenant les dérivées de u d'ordre le plus élevé. Dans L_n, remplaçons chacun des symboles $\dfrac{\partial^n}{\partial x_1^{n_1} \ldots \partial x_p^{n_p}}$, $n_1 + \ldots + n_p = n$, par le produit $\alpha_1^{n_1} \ldots \alpha_p^{n_p}$ des indéterminées $\alpha_1, \ldots, \alpha_p$. Nous associons ainsi à L (et à L_n) une forme algébrique $\mathfrak{L}_n(x; \alpha)$ d'ordre n en les α; dans $\mathfrak{L}_n(x; \alpha)$, la variable x indique que les coefficients peuvent dépendre de x.

L'opérateur L est, dans un domaine D,

1) totalement elliptique, si la forme $\mathfrak{L}_n(x; \alpha)$ est définie pour $x \in D$;

2) totalement hyperbolique par rapport à x_p, si l'hypersurface $\mathfrak{L}_n(x;$

$\alpha_1, \ldots, \alpha_{p-1}, 1) = 0$, $x \in D$, supposée sans composantes multiples[1]), se décompose en $\frac{n}{2}$ ovales réels emboîtés les uns dans les autres.

La condition de totale ellipticité n'appelle pas de commentaire particulier. Au contraire, la vérification de la condition de totale hyperbolicité est plus compliquée. Cependant, cette question a fait récemment un progrès important grâce à un théorème de P. Brusotti.

4. Il y a lieu maintenant de préciser ce qu'il faut entendre par „solution" des problèmes envisagés ici. En général, nous dirons qu'une fonction u est une solution de $Lu = 0$ si elle vérifie cette équation en tout point où elle est régulière d'ordre n au moins. Mais l'équation de propagation du son montre qu'il est indispensable d'admettre la présence de singularités constituées par des variétés dont la dimension peut atteindre $p - 1$. Bien plus, l'utilisation de la formule de réciprocité conduit à envisager systématiquement de telles solutions.

Enfin, il convient aussi de distinguer entre solution de l'équation $Lu = 0$ et solution du problème $Lu = 0$, $Au = 0$.

5. *Equations totalement elliptiques.* Il est connu depuis longtemps que l'on obtient une formule fondamentale de la théorie du potentiel en prenant pour v dans la formule de réciprocité correspondant à $\Delta_p u = 0$, la solution élémentaire (singulière) $v(x; y) = r_{xy}^{2-p}$, $r_{xy} = |x - y|$, $p \geqq 3$ de cette équation. Si Σ est l'hypersphère de centre x et de rayon R, on a

$$K_p = \lim_{R \to 0} \int_{\Sigma} \frac{dv(x; y)}{d\pi} d\sigma_y = (p - 2)\omega_p \tag{2}$$

où ω_p désigne la „surface" de l'hypersphère unité dans l'espace réel E_p à p dimensions. Si $u(x)$ est une fonction régulière en x, on a alors

$$K_p u(x) = \lim_{R \to 0} \int_{\Sigma} u(y) \frac{dv(x; y)}{d\pi} d\sigma_y. \tag{3}$$

Pour étendre cette méthode à l'équation plus générale, totalement elliptique $Lu = 0$, il faut obtenir une solution élémentaire de l'équation adjointe $L^* v = 0$. Ce problème est d'ailleurs entièrement déterminé et consiste à trouver une solution $v(x; y)$ de $L^* v = 0$ possédant les propriétés suivantes:

1) $v(x; y)$ est régulière d'ordre n en tout point $x \neq y$;

2) en $x = y$, les dérivées partielles d'ordre n de $v(x; y)$ deviennent infinies d'ordre p par rapport à r_{xy}^{-1}; les dériveés partielles d'ordre $\leqq n - 1$ deviennent infinies d'ordre $p - 1$ au plus ou restent finies.

---

[1]) Cette définition s'étend évidemment aussitôt au cas où Ω_n possède des composantes multiples.

Posons

$$\Theta = \lim_{R \to 0} \int_{\Sigma} P(1, v; \pi) d\sigma_y. \tag{4}$$

En isolant le point x par l'hypersphère Σ, en appliquant la formule de réciprocité au domaine restant $V - \Sigma$ et en faisant tendre R vers zéro, on obtient la formule de représentation

$$\Theta u(x) = \int_{V} vLu dx + \int_{\dot{V}} P(u, v; \pi) d\sigma. \tag{5}$$

Considérons d'abord des équations à coefficients constants ne contenant que des dérivées d'ordre $n (L = L_n, L = L^*)$. Pour ces équations, la solution élémentaire a été donnée, pour $p = 2$, $n > 2$ par Somigliana; pour $p = 3$, $n > p$ par I. Fredholm; pour p quelconque, $n \geqq p$ par F. Bureau [1]).

Pour $p = 2$, n quelconque, E. E. Levi a donné une méthode qui s'applique même si les coefficients de L ne sont pas analytiques mais seulement suffisamment réguliers. Cette méthode de E. E. Levi qui utilise comme ,,présolution'' ou ,,parametrix'', la solution obtenue pour une équation homogène $L_n u = 0$ à coefficients constants peut être étendue sans aucune difficulté, comme nous l'avons indiqué en 1936, aux équations $Lu = 0$ à coefficients variables.

Lorsque l'équation $L_n u = 0$ $(n \geqq p)$ est à coefficients constants, la solution élémentaire $v(x)$ et ses dérivées partielles des divers ordres s'expriment à l'aide d'intégrales attachées à la variété algébrique $L_n(\alpha_1, \ldots, \alpha_{p-1}, 1) = 0$. Pour calculer la constante Θ donnée par (4), nous avons d'abord utilisé une méthode de récurrence ramenant le calcul à $p = 2$ ou $p = 3$ suivant la parité de p. Cette constante peut aussi être calculée par une méthode directe.

D'ailleurs, comme nous l'avons déjà indiqué en 1936, la formule de représentation (5) permet de généraliser à l'équation $Lu = 0$, les résultats essentiels de la théorie classique du potentiel newtonien.

Observons enfin qu'en utilisant une transformation de Fourier, G. Herglotz a donné la solution de l'équation $L_n u = g(x)$, où L_n est à coefficients constants et $g(x)$, une fonction régulière, sous la forme $u(x) = \int_{y \in E_p} g(y) v(x - y) dy$.

Malheureusement, la méthode utilisée par G. Herglotz doit être considérée comme formelle et la formule fondamentale (5) n'y est pas obtenue.

Lorsque l'opérateur différentiel L a ses coefficients analytiques et contient des dérivées de u d'ordre $< n - 1$, la solution élémentaire a été déterminée pour p quelconque, $n = 2$, par J. Hadamard (sous la forme d'un développement en série); pour n et p quelconques, par F. John (sous la forme d'une combinaison

[1]) La condition n ≧ p peut être facilement écartée.

de dérivées d'une certaine intégrale et en utilisant le théorème de Cauchy-Kowalewsky).

6. Revenons maintenant au problème de Dirichlet pour l'équation de Laplace. On sait qu'en général, ce problème n'est pas résolu par la formule (5) qui exige la connaissance de données surabondantes sur $\dot{V}$. Mais on sait aussi qu'il existe (dans des conditions générales) une solution G de l'équation $\Delta_p u = 0$ prenant sur la frontière $\dot{V}$ la valeur 0 et se comportant en un point intérieur x (pôle de la solution) comme la solution r_{xy}^{2-p} ($p \geqq 3$) de cette équation. Dans ces conditions, le problème de Dirichlet est résolu par

$$u(x) = \int_{\dot{V}} P(u; G; \pi) d\sigma$$

l'intégrale ne dépendant que des valeurs de u sur $\dot{V}$ et non des valeurs de $\frac{du}{d\pi}$. Ce résultat s'étend sans peine, mutatis mutandis, aux équations aux dérivées partielles totalement elliptiques d'ordre supérieur à deux.

La solution singulière $v(x, y)$ dont il a été question plus haut apparaît comme la solution élémentaire d'un problème de Dirichlet pour l'espace entier E_p, si Lu se réduit par exemple à une „testing function", c'est-à-dire à une fonction de x indéfiniment dérivable à support compact.

Ainsi, on est très naturellement conduit à ne pas se limiter à considérer une solution élémentaire d'une équation, mais bien à envisager „*la*" solution élémentaire d'un *problème* associé à cette équation.

7. *Equations totalement hyperboliques.* Si nous passons aux équations totalement hyperboliques, nous rencontrons des faits analogues, cependant dans des circonstances plus difficiles. Une discussion minutieuse de la méthode utilisée par V. Volterra pour intégrer l'équation des ondes cylindriques et une certaine analogie formelle avec la théorie de l'équation de Laplace ont conduit J. Hadamard à déduire la solution du problème de Cauchy pour une équation totalement hyperbolique du second ordre à p variables spatiales, de la „solution élémentaire" de l'équation adjointe à l'équation donnée. La méthode développée par J. Hadamard peut être réduite à une méthode de singularités entièrement analogue à celle qui est utilisée dans la théorie des équations totalement elliptiques. Cependant, l'application de cette méthode conduit à introduire des opérateurs linéaires que nous désignons par pf. (partie finie) et pl. (partie logarithmique), associant à certaines intégrales divergentes, des quantités finies.

Considérons d'abord une équation $Lu = 0$ totalement hyperbolique du second ordre à $p + 1$ variables indépendantes (p variables spatiales (x), une

variable temporelle t) et à coefficients analytiques. Soit $\Gamma = 0$ l'équation du conoïde caractéristique Γ, de sommet (x, t) de cette équation.

La solution élémentaire utilisée par M. Hadamard s'écrit lorsque (y, τ) est intérieur à Γ,

1) Si p est pair

$$v(x, t; y, \tau) = \Gamma^{-\frac{p-1}{2}} U(x, t; y, \tau); \tag{6}$$

2) Si p est impair

$$v(x, t; y, \tau) = \Gamma^{-\frac{p-1}{2}} U(x, t; y, \tau) + \mathscr{K}(x, t; y, \tau) \lg \Gamma; \tag{7}$$

lorsque (y, τ) est extérieur à Γ, U et $\mathscr{K}$ sont identiquement nulles.

Les fonctions $U, \mathscr{K}$ sont des fonctions holomorphes de leurs arguments et $U(x, t; x, t) = 1$. D'autre part, $\mathscr{K}(x, t; y, \tau)$ est aussi une solution de l'équation.

Les formules indiquées montrent que les dérivées partielles d'ordre 1 de $v(x, t; y, \tau)$ deviennent au voisinage de $\Gamma = 0$, infinies comme $\Gamma^{-\frac{p+1}{2}}$ et que dans les mêmes conditions, v est infinie comme $\Gamma^{-\frac{p-1}{2}}$. Enfin, si nous désignons par Σ_η la section du conoïde Γ et de l'hyperplan $t - \tau = \eta$ ($\eta > 0$ et assez petit), nous avons

1) si p est pair

$$A_p = \lim_{\eta \to 0} pf \int_{\Sigma_\eta} P(1, v; \pi) d\sigma_{y\tau}; \tag{8}$$

2) si p est impair

$$B_p = \lim_{\eta \to 0} pl \int_{\Sigma_\eta} P(1, v; \pi) d\sigma_{y\tau}; \tag{9}$$

A_p et B_p sont des constantes non nulles.

Cela étant, supposons que les données de Cauchy soient portées par une hypersurface S délimitant avec le conoïde Γ, un domaine V simplement connexe En appliquant alors la formule de réciprocité (1) à un domaine $V_{\varepsilon, \eta}$ ($\varepsilon > 0$ et petit) intérieur à V et tendant vers V lorsque ε et η tendent vers zéro, on obtient des formules de *représentation* de la solution du problème de Cauchy.

D'une manière précise, si S^* désigne la portion de S intérieur à Γ, nous avons

1) si p est pair

$$A_p u(x, t) = pf \int_V vLu dy d\tau - pf \int_{S^*} P(u_0, u_1; v; \pi) d\sigma, \tag{10}$$

2) si p est impair

$$B_p u(x, t) = pl \int_V vLudyd\tau - pl \int_{S^*} P(u_0, u_1; v; \pi)\, d\sigma - \int_\Gamma P(u; \mathscr{K}; \pi) d\sigma. \tag{11}$$

Lorsque p est pair, la relation (10) nous donne une formule de représentation ne contenant que les données du problème posé, savoir une solution de l'équation adjointe donnée et les conditions initiales (u_0, u_1).

Il n'en est pas de même si p est impair, car dans la formule de représentation (11) interviennent aussi les valeurs inconnues de u et $\frac{du}{d\pi}$ sur le conoïde Γ. Il convient donc de modifier cette première formule de représentation de manière à faire disparaître ces valeurs inconnues. A cet effet, remarquons que $\mathscr{K}$ est une solution de l'équation adjointe à l'équation donnée. En appliquant la formule de réciprocité au domaine V et aux fonctions u et $\mathscr{K}$, on obtient

$$\int_V \mathscr{K} Ludyd\tau = \int_{S^*} P(u_0, u_1; \mathscr{K}; \pi) d\sigma + \int_\Gamma P(u; \mathscr{K}; \pi) d\sigma$$

d'où l'on déduit la valeur de $\int_\Gamma P(u; \mathscr{K}; \pi) d\sigma$. En substituant cette valeur dans (11), on obtient la seconde formule de représentation

$$B_p u(x; t) = pl \int_V vLudyd\tau - pl \int_{S^*} P(u_0, u_1; v; \pi) d\sigma - \int_V \mathscr{K} Ludyd\tau + \int_{S^*} P(u_0, u_1; \mathscr{K}; \pi) d\sigma \tag{12}$$

qui exprime u en fonction des seules données du problème de Cauchy.

Jusqu'ici les formules (10) (12) sont seulement des formules de représentation. Pour voir si elles constituent des solutions du problème posé, *il faut en effectuer la synthèse*, c'est-à-dire, vérifier si elles satisfont à toutes les conditions indiquées (équation indéfinie et conditions accessoires). Cette recherche montre qu'il y a lieu d'imposer à l'hypersurface S portant les données de Cauchy, la condition d'être partout d'orientation d'espace. Lorsque cette condition est remplie, les formules (10) et (12) donnent la solution unique du problème posé.

Remarquons encore que les opérateurs pf. et pl. permettent d'effectuer commodément la synthèse dont il vient d'être question.

8. On peut étendre les résultats précédents aux équations totalement hyperboliques d'ordre plus grand que 2, à plus de deux variables indépendantes et à coefficients constants.

A cet effet, considérons une équation $Lu \equiv L\left(\frac{\partial}{\partial x}\ ;\ \frac{\partial}{\partial t}\right) u = 0$ d'ordre n à $p+1$ variables indépendantes (p variables spatiales x; une variable temporelle t) telle que L ne contienne que des dérivées partielles d'ordre n de u. Nous supposerons de plus $n \geqq p$ et pour simplifier, nous considèrerons plus particulièrement les cas $p = 2$ et $p = 3$.

9. Supposons $p = 2$ et désignons par $P_\nu = (\alpha_\nu, \beta_\nu)$, $(\nu = 1, \ldots, n)$, les intersections de la courbe $L(\alpha, \beta; 1) = 0$ avec la droite $E \equiv \alpha x_1 + \beta x_2 + t = 0$. La solution élémentaire est donnée par

$$v(x; t) = i\pi \Sigma \varepsilon_\nu \int_{P_0}^{P} E^{n-3} \frac{d\alpha}{L_2} \tag{13}$$

où $L_2 = \frac{\partial f}{\partial \beta}$, $\varepsilon_\nu = \text{sign } R\left(\frac{\beta_\nu}{i\alpha_\nu}\right)$; $P_0 = (\alpha_0, \beta_0)$ est un point arbitraire. Dans (13), la somme est étendue à toutes les intersections imaginaires P_ν.

On peut montrer que $v(x; t)$ possède les propriétés essentielles indiquées au paragraphe 6 et que l'on a en outre avec les notations de ce paragraphe

$$\Theta = \lim_{\eta \to 0} pf \int_{\Sigma_\eta} P(1, v; \pi) d\sigma_{y\tau} \tag{14}$$

avec $\Theta = -(n-3)!2n\pi^2$.

La méthode des singularités peut donc être appliquée dans les conditions habituelles et donne la formule de représentation

$$\Theta u(x; t) = n! \int_V vLu dy d\tau - pf \int_S P(v) d\sigma. \tag{15}$$

La synthèse de cette formula montre que si S possède une orientation d'espace, on a effectivement la solution du problème de Cauchy.

10. Supposons $p = 3$ et admettons pour simplifier que la forme $L(\alpha, \beta, \gamma, 1)$ correspondant à l'équation étudiée soit telle que la surface $L(\alpha, \beta, \gamma, 1) = 0$ ne possède que des points elliptiques.

Posons $E \equiv \alpha x_1 + \beta x_2 + \gamma x_3 + t$ où t peut prendre des valeurs complexes et x_1, x_2, x_3 des valeurs réelles. Soient $d\omega$ l'élément de surface et ω l'ensemble des $\frac{n}{2}$ ovales ω_k de la surface $L = 0$ (nous supposerons ω_{k+1} complètement intérieur à ω_k). Posons encore

$$\mathscr{W}_k(x; t) = \int_{\omega_k} E^{n-4} \lg E d\omega. \tag{16}$$

Si le plan $E = 0$ (t réel) rencontre ω_j, mais non ω_{j+1}, la solution élémentaire est donnée par

$$v(x;t) = \sum_{k=j+1}^{\frac{n}{2}} \mathscr{R}\mathscr{W}_k. \tag{17}$$

Si $E = 0$ rencontre tous les ovales de la surface, nous avons $v(x;t) = 0$. Cette fonction $v(x;t)$ possède les propriétés essentielles indiquées au paragraphe 6 et l'on a

$$\Theta = \lim_{\eta \to 0} pl \int_{\Sigma_\eta} P(1, v; \pi) d\sigma_{y\tau} \tag{18}$$

où $\Theta = -4\pi^2 n!(n-4)!$

La méthode des singularités nous donne alors la première formule de représentation

$$\Theta u(x;t) = pl \int_S P^*(v) d\sigma + \int_\Gamma \mathfrak{M}^* d\sigma \tag{19}$$

où $P^*(v)$ est une expression dépendant seulement de v et de ses dérivées partielles du premier ordre; $\mathfrak{M}^*$ est déterminé par les valeurs de u et de ses dérivées partielles des premier et second ordres sur $\Gamma = 0$.

Pour obtenir la valeur de la seconde intégrale de (19), nous considérons une solution $\mathscr{K}(x;t)$ (solution auxiliaire) de l'équation donnée et déterminée comme il suit. Si le plan $E = 0$ rencontre ω_j mais non ω_{j+1}, nous posons

$$\mathscr{K}(x;t) = \frac{1}{\pi} \sum_{k=1}^{i} \mathscr{I}\mathscr{W}_k = \pm \int_\omega E^{n-4} \operatorname{sign} E d\omega; \tag{20}$$

on a aussi

$$\mathscr{K}(x;t) = \frac{1}{\pi} \mathscr{I} w(x;t). \tag{21}$$

On peut montrer que si $E = 0$ rencontre tous les ovales de la surface $L = 0$ $\mathscr{K}(x\,t)$ est identiquement nulle. En d'autres termes, $\mathscr{K}(x;t) \equiv 0$ pour tout, point $(x;t)$ extérieur au cône caractéristique. Enfin, $\mathscr{K}(x;t)$ et ses dérivées partielles d'ordre $\leqq n-1$ restent bornées sur le cône caractéristique.

Cela étant rappelé, on obtient en appliquant la formule de réciprocité,

$$n! \int_V \mathscr{K} L u dy d\tau = \int_S P(\mathscr{K}) d\sigma + \int_\Gamma \mathfrak{M}^* d\sigma. \tag{22}$$

En combinant (19) et (22), on déduit la seconde formule de représentation

$$\Theta u(x;t) = pl \int_S P^*(v) d\sigma + n! \int_V \mathscr{K} L u dy d\tau - \int_S P(\mathscr{K}) d\sigma \tag{23}$$

exprimant la solution $u(x;t)$ (supposée existante) du problème de Cauchy en fonction des données du problème.

La synthèse de la solution montre que si S possède une orientation d'espace, la formule précédente donne la solution du problème de Cauchy.

En conclusion, nous pouvons dire que „tous les résultats de la théorie du problème de Cauchy peuvent et doivent se déduire seulement de la solution élémentaire et de la solution auxiliaire".

Remarque. Les problèmes les plus intéressants rencontrés en Physique mathématique ont été étudiés par cette méthode, à partir des formules générales.

11. *Intégrales de Fourier et problème de Cauchy.* Considérons le problème de Cauchy pour une équation linéaire aux dérivées partielles à coefficients constants, totalement hyperbolique (par rapport à t), les données initiales étant connues sur $t = 0$. On sait qu'au moins formellement, on peut représenter la solution de ce problème à l'aide d'intégrales multiples de Fourier et que si l'équation est à p variables spatiales, la solution ainsi obtenue contient des intégrales $2p$-uples étendues à tout l'espace.

Cependant, la méthode des singularités donne la solution du problème de Cauchy à l'aide d'intégrales p-uples. Dès lors, il y a lieu de procéder à la réduction des intégrales $2p$-uples dont il vient d'être question.

Cette réduction est possible si on généralise convenablement la formule de Parseval en utilisant la solution élémentaire et la solution auxiliaire de l'équation et les pf. ou pl. d'intégrales divergentes.

12. Dans l'étude des équations totalement elliptiques, la solution singulière $v(x; t)$ est apparue comme la solution élémentaire permettant de résoudre l'équation $Lu = \varphi$ pour l'espace entier E_p, φ étant une testing function.

Par analogie, on a considéré comme le problème fondamental, le problème de Cauchy pour $Lu = \varphi$ avec des données nulles à l'infini. Sa solution peut être représentée par le produit de composition $\varphi * \mathcal{k}$ où $\mathcal{k}$ est une distribution résolvante entièrement déterminée par les solutions élémentaires $v(x; t)$ et auxiliaire $\mathcal{K}(x; t)$.

Cependant, la connaissance de $\mathcal{k}$ n'est pas équivalente à celle de v et de $\mathcal{K}$ et $\mathcal{k}$ seul ne permet pas de résoudre immédiatement et dans tous les cas, le problème général de Cauchy. On peut s'en assurer sur des exemples.

1) Equation des ondes; ici $\mathcal{K} = 0$ et $\mathcal{k}$ est déterminé par v.

2) Equation des ondes amorties; ici $\mathcal{K} \neq 0$ et $\mathcal{k}$ est déterminé par v et $\mathcal{K}$.

3) Equation homogène à coefficients constants avec $n \geqq p$.

a) si p est pair, $\mathcal{k}$ est déterminé par v;

b) si p est impair, $\mathcal{k}$ est déterminé par $\mathcal{K}$ seul; pour résoudre le problème de Cauchy dans toute sa généralité, il faut faire intervenir les solutions v et $\mathcal{K}$ (ou au moins les valeurs de v et de certaines de ses dérivées sur le cône Γ).

13. *Problèmes aux limites pour les équations totalement hyperboliques.*

La solution élémentaire joue encore un rôle essentiel dans la solution des

problèmes aux limites pour les équations totalement hyperboliques. A titre d'exemple, considérons l'équation

$$Du = u_{tt} - Lu = 0, \qquad L_x = v_{xx} - a(x)v$$

et les conditions aux limites suivantes:

1) En $t = 0$, on a

$$u(x, 0) = u_0(x) = 0, \qquad u(x, 0) = u_1(x);$$

2) Quel que soit $t \geqq 0$, on a $u(0, t) = u(l, t) = 0$.

On sait qu'il existe une solution $G(x, t; y, \tau)$ de l'équation (la fonction de Green-Riemann) présentant des singularités convenables le long des caractéristiques issues de $P = (x, t)$ et se réfléchissant sur $x = 0$ et $x = l$ et telle que la solution du problème s'écrive

$$2u(P) = \int_0^l G(x, t; y, 0)u_1(y)dy.$$

Admettons que $a(x)$ soit tel que les valeurs caractéristiques λ_n $(n = 1, 2, \ldots)$ déterminées par

$$L_x v + \lambda v = 0, \qquad v(0) = v(l) = 0$$

soient simples; soient $v_n(x)$ les fonctions caractéristiques correspondantes. Si on pose

$$K_s(x, y; t) = \sum_{k=1}^{\infty} \frac{v_k(x)v_k(y)}{\lambda_k^{s+\frac{1}{2}}} \sin \sqrt{\lambda_k}t,$$

On a

$$K_0(x, y; t) = \tfrac{1}{2}G(x, y; t, 0),$$

K_0 étant la valeur de K_s en $s = 0$ obtenue par prolongement analytique.

14. *Autres problèmes de Cauchy.* Les solutions élémentaires et auxiliaires permettent aussi de résoudre le problème de Cauchy pour des types d'équations ne rentrant pas dans les précédents.

A titre d'exemple, signalons que la théorie des fluides visqueux, peu compressibles, conduit pour des valeurs des constantes élastiques λ et μ vérifiant la relation $\lambda + 2\mu = 0$ à l'équation

$$D_1 Du = 0 \text{ où } D_1 = \frac{\partial}{\partial t} + \mu\Delta_3, D = \frac{\partial^2}{\partial t^2} - a^2\Delta_3,$$

a = constante, Δ_3 laplacien en trois variables spatiales.

La solution du problème de Cauchy pour cette équation fait intervenir une solution élémentaire

$$v(r, t-\tau) = \frac{1}{r}\int_{\tau}^{t} \left\{ e^{-\frac{[r-a(\alpha-\tau)]^2}{4\mu(t-\alpha)}} - e^{-\frac{[r+a(\alpha-\tau)]^2}{4\mu(t-\alpha)}} \right\} \frac{d\alpha}{\sqrt{t-\alpha}},$$

$$r^2 = (x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2$$

et une solution auxiliaire

$$\mathcal{K}(r; t-\tau) = \begin{cases} \dfrac{1}{2r} \lg \dfrac{a(t-\tau)-r}{a(t-\tau)+r} & \text{si } r \leqq a(t-\tau), \\ 0 & \text{si } r > a(t-\tau). \end{cases}$$

15. *Problèmes de Cauchy singuliers.* Dans les problèmes de Cauchy que nous venons de considérer, les coefficients de l'équation étudiée étaient réguliers sur l'hypersurface portant les données de Cauchy. Cependant, l'équation de Tricomi, dont on connaît le rôle important dans l'étude des fluides compressibles (domaine transonique) a attiré l'attention sur des problèmes de Cauchy singuliers.

Considérons l'équation d'Euler-Poisson-Darboux

$$u_{tt} + \frac{k}{t} u_t = \Delta_x u \tag{24}$$

avec les données initiales

$$u(x, 0) = u_0(x), \qquad u(x, 0) = u_1(x)$$

k étant une constante non nulle et $u_0(x)$, $u_1(x)$ des fonctions régulières. Pour $t = 0$, le coefficient variable kt^{-1} $(k \neq 0)$ devient infini et le problème de Cauchy correspondant est singulier. Ce problème a été résolu récemment par A. Weinstein en combinant une méthode de descente généralisée avec deux formules de récurrence. Une autre méthode utilisant le prolongement analytique de certaines intégrales définies a été utilisée par J. B. Diaz et H. F. Weinberger. Enfin, pour les valeurs exceptionnelles de k (entiers négatifs impairs), la solution a été donnée par E. K. Blum.

Au lieu de l'équation (24), nous avons envisagé l'équation plus générale

$$u_{tt} + kt^{-1}u_t = L_x u \tag{25}$$

L_x étant un opérateur elliptique quelconque. On peut montrer qu'il existe encore une solution „élémentaire" de cette équation permettant de résoudre le problème posé.

16. Signalons enfin que l'usage des solutions élémentaires n'est pas limité aux équations aux dérivées partielles mais que l'on peut aussi associer une solution élémentaire à certaines équations intégro-différentielles comme l'ont montré V. Volterra et M. Gevrey.

On trouvera une bibliographie étendue dans les articles suivants:

F. J. Bureau: *Quelques questions de Géométrie suggérées par la théorie des équations aux dérivées partielles totalement hyperboliques*. Colloque de Géométrie algébrique, Liège 1949.

F. J. Bureau: *Divergent integrals and partial differential equations*. Second symposium in Applied Mathematics. Chicago. 1954. Communications on pure and applied Mathematics, Vol. VIII, 1955, pp. 143—202.

SOME ASPECTS OF THE THEORY OF NON-LINEAR VIBRATIONS

M. L. CARTWRIGHT

Much of the early work on non-linear oscillations was motivated by an interest in the periodic solutions of the equations involved; stable periodic oscillations were desired in the corresponding physical systems and the most important problems were the determination of the conditions for the existence of such oscillations, the magnitudes of the amplitudes and the way in which the amplitude varied through resonance, and in the case of autonomous systems a very precise estimate of the period was required. These were the early problems of the radio engineers; their systems operated at high frequencies so that the oscillations built up to their steady state in a very short interval of time, and the transient features were of minor importance.

I wish now to focus attention on two other classes of problem; they are by no means new, but I believe them to be of increasing importance on account of their applications to automatic control theory. The first is the problem of finding conditions which exclude unwanted oscillations so that the engineer can operate the corresponding system safely. The other is the problem of determining how rapidly a given system will settle down to its final steady state, or rather, what form of non-linearity within the limits available will give the most satisfactory response to a given input signal. In the latter form the problem seems to demand a complete mastery of the mathematical form of the solutions of the equation as the non-linear function varies.

Let us first review the contributions of the existing theory to the problem of excluding unwanted oscillations. Autonomous second degree equations can be handled by phase plane and other methods, and present no great difficulty in this respect, but difficulties arise with forced oscillations. Topological methods are seldom precise enough to exclude all unwanted oscillations and approximations of nearly linear systems usually ignore the possibility of oscillations starting outside the range of validity of the approximations. However the conditions for synchronization of the system represented by

$$\ddot{x} - k\,(1 - x^2)\,\dot{x} + x = p\,k \cos \lambda t,\ \lambda \text{ near } 1,\ k \text{ small}, \tag{1}$$

were obtained by considering it as a nearly linear system; van der Pol and others obtained conditions for the suppression or attraction of the unwanted

so-called free oscillation by the forced by putting $x = b_1 \cos t + b_2 \sin t$ and supposing that b_1, b_2 vary slowly. The problem is reduced to the consideration of first order equations in b_1, b_2 (or corresponding equations in phase and amplitude) which are by no means nearly linear, and it is by no means easy to determine the qualitative behaviour of the solutions as λ varies. In fact Mr Gillies has shown that the results which I stated for a certain intermediate range of λ are not quite correct. The introduction of an unsymmetrical term αx, say, into the damping coefficient makes no difference unless it is very large, but Gillies has shown that for fixed $\alpha > 0$ the solutions of

$$\ddot{x} - k\,(1 + \alpha\, k^{-\frac{1}{2}}\, x - x^2)\,\dot{x} + x = p\, k \cos \lambda\, t,\ \lambda \text{ near } 1,\ k \text{ small},$$

behave very differently from those of (1).

The systems of the automatic control engineer usually have positive damping and at one time it was thought to be impossible for a stable periodic solution to exist in such systems with period different from that of the forcing term, but subharmonics of order 3 were observed in a physical system corresponding to an equation of the form

$$\ddot{x} + k\,\dot{x} + x - \beta x^3 = p \cos \lambda t,\ \lambda \text{ near } 3, \beta\, k, \text{ small}.$$

The existence of such solutions can be verified, but this is a negative and unsatisfactory answer to the problem. However Levinson dealt with the case of a linear restoring force and positive damping, and Littlewood and I obtained conditions under which all solutions of the equation

$$\ddot{x} + k f(x)\,\dot{x} + g(x, k) = k p(t),\ p(t) \text{ periodic}, \tag{2}$$

converge to a single solution. The essential features of these conditions are first that $k \geqq 1$, $f(x) \geqq 1$, $g(0) = 0$, $g'(x, k) \geqq 1$ which imply that $|x| < \xi_0$, $|\dot{x}| < \eta_0$, ξ_0, η_0 independent of k. Further

$$2\eta_0 \operatorname*{Max}_{|x| < \xi_0} |g''(x, k)| < k, \tag{3}$$

which means that if the restoring form is not too strongly non-linear in the range considered there is no stable oscillation other than the periodic one which is expected.

It is noteworthy that the proof depends on the fact that the damping $k f(x)$ is independent of $\dot{x}$ or t. Let

$$F(x) = \int_0^x f(x)\,dx,\ z = x_2 - x_1,\ \Delta f = f(\alpha_2) - f(x_1),\ \Delta g = g(x_2) - g(x_1),$$

where $x_2(t)$, $x_1(t)$ are any two solutions of (2). Then substituting x_1, x_2 in (2) and subtracting we have

$$\ddot{z} + k \frac{d}{dt}\left(\frac{\Delta F}{z} . z\right) + \frac{\Delta g}{z} . z = 0 \tag{4}$$

Multiplying by $\dot{z}$ and integrating, and then using integration by parts and substituting for $\ddot{z}$ in the result from (4) we obtain after some manipulation

$$\int_0^t z^2\,dt = O\ (1) \text{ as } t \to \infty$$

when (3) holds, and from this convergence of all solutions is deduced. Further results of this type seem highly desirable, but although the method works for more general $p\ (t)$ it seems to fail completely if $f\ (x)$ depends on $\dot{x}$ or t as well as x.

The theory of synchronous motors which was studied some years ago by Lyon and Edgerton, Tricomi and others, and again more recently, in particular by Hayes, involves the equation

$$\ddot{\theta} + \alpha\dot{\theta} + \sin\theta = \beta,$$

where $\alpha > 0, \beta > 0$. If $\beta < 1$ under suitable conditions the system will synchronize; that is to say that θ will tend to a value θ_0 such that $\sin\theta_0 = \beta$. The problem of the unwanted oscillation arises here in two distinct forms; first is given α and β determine β_0 if possible so that a sudden change from β to $\beta^1 < \beta_0$ will not give rise to an unwanted oscillation; the second is given $\beta < 1$ find a value α_0 such that for $\alpha < \alpha_0$ no unwanted oscillation is possible. In mathematical terms the first is a problem of determining a safe range of initial values. It was established early that α_0, β_0 exist and that they can be chosen so that for $\alpha > \alpha_0, \beta^1 > \beta_0$ the unwanted oscillations may occur. Since no synchronization is possible for $\beta > 1$ one might expect that α_0 would tend to infinity as $\beta \to 1$ but Hayes has shown that

$$((3\cos^2\theta_0 + 1)^{\frac{1}{2}} - 2\cos\theta_0)^{\frac{1}{2}} < \alpha_0 < 2\sin\tfrac{1}{2}\theta_0,$$

where $\beta = \sin\theta_0$. Like the other writers on this subject he puts $y = \dot{\theta}$ and uses the phase plane; he also replaces θ by $\theta + \pi - \theta_0$. His method depends on the use of a comparison function

$$\bar{y} = 2\left[(\tfrac{1}{4}\alpha^2 + \cos\theta_0)^{\frac{1}{2}} - \alpha f\ (\theta)\right]\sin\tfrac{1}{2}\theta,$$

where $f\ (0) = 0$ and $f\ (\theta)$ is antisymmetric about $\theta = \pi$. The function $g\ (\theta)$ defined by

$$\bar{y}\frac{d\bar{y}}{d\theta} + \alpha\bar{y} - y\frac{dy}{d\theta} - \alpha y = \sin\tfrac{1}{2}\theta\, g(\theta)$$

can be calculated when $f\ (\theta)$ is known and he shows that if $g\ (\theta)$ remains positive or negative throughout the interval $0 \leqq \theta < 2\pi$ then $\alpha > \alpha_0$ or $\alpha < \alpha_0$ as the case may be. Thus a remarkably precise quantitative result is obtained without any use of approximations other than the comparison function $\bar{y}$.

Turning now to the second problem, that of the rapidity of convergence we observe that whereas nearly linear systems converge slowly, solutions of equations which are not nearly linear may converge with great rapidity. For instance if a solution of

$$\ddot{x} - k\,(1 - x^2)\,\dot{x} + x = 0,\ k \text{ large},$$

starts at $t = 0$ with $x = x_0$, $\dot{x} = y_0 > 3/k$ I have shown that the solution reaches $x = 1$ with $\dot{x} > y_0 + \frac{2}{3}k - M\,(\log k)/k$, and therefore rises to a height $h > 3^{\frac{1}{2}}$ given by

$$|\,y_0 + k\,(h - \tfrac{1}{3}h^3)\,| < M\,(\log k)/k$$

where M is an absolute constant. From the maximum h the solution slowly descends to $x = 1$ from which point its behaviour differs little from that of the periodic solution.

In servomechanisms it is often possible to construct the element giving the non-linear effect in such a way that it corresponds to a specified function of x and $\dot{x}$ and some attempts have been made to improve response by intentional non-linearities. What seems to be needed here is a thorough grasp of the real variable methods and principles which are used in existence theorems and in work such as I have just described on relaxation oscillations and synchronous motors, and although such methods may be difficult in the abstract they correspond directly to physical principles which are often familiar to the engineer. Two simple examples of intentional non-linearities have been given by McDonald; the equations representing the first are

$$\begin{aligned} \ddot{x} &= -T \text{ for } x + L\,\dot{x} > 0, \\ &= T \text{ for } x + L\,\dot{x} < 0. \end{aligned}$$

The solutions move rapidly towards the origin but tend to overshoot it. The second system cures the tendency to overshoot by making the switch along the track of the solution through the origin so that

$$\begin{aligned} \ddot{x} &= -T \text{ for } |\dot{x}|\,\dot{x} + 2\,T\,x > 0, \\ &= \quad T \text{ for } |\dot{x}|\,\dot{x} + 2\,T\,x < 0. \end{aligned}$$

Unfortunately the physical system suffers from a tendency to other types of instability especially if the switch does not occur precisely at the point intended, so that some of the difficulties are engineering problems which I am not competent to discuss.

Many servomechanisms correspond to third order equations and then instead of the $x, \dot{x}$ phase plane we have to consider the $x, \dot{x}, \ddot{x}$ phase space. Switching effects such as the above would probably require non linear functions depending on $\ddot{x}$ as well as x and $\dot{x}$ which may be impossible, and even the basic

theory of oscillations in third order systems has hardly been touched. Van der Pol, Levinson, Rauch and others have discussed particular equations, but lack of characteristic common features in the third order equations of the engineers seems to have discouraged the creation of a general theory. One may however observe that, if $a_1\, a_2, a_3$ are positive constants, solutions of the equation

$$\dddot{x} + a_1\, \ddot{x} + a_2\, \dot{x} + a_3\, x = 0$$

all converge to $x = 0$ provided that

$$a_1, a_2 > a_3 \tag{5}$$

This condition may be interpreted topologically by writing

$$\dot{x} = y,$$
$$\dot{y} = z,$$
$$\dot{z} = -a_3\, x - a_2\, y - a_1\, z,$$

and using the theory of matrices to express this system in the form

$$\dot{\xi} = \lambda_1\, \xi,$$
$$\dot{\eta} = \lambda_2\, \eta,$$
$$\dot{\zeta} = \lambda_3\, \zeta,$$

where $\lambda_1, \lambda_2, \lambda_3$ are the roots of

$$\lambda^3 + a_1\, \lambda^2 + a_2\, \lambda + a_3 = 0$$

By (5) the real parts of all three roots are negative; one λ_3, say, is real and the other two may be conjugate complex numbers in which case writing $\xi = re^{i\varphi}$, $\eta = r^{-i\varphi}$ we reduce the system to

$$\dot{r} = -\tfrac{1}{2}\,(\lambda_1 + \lambda_2)\, r,\ \dot{\xi} = \lambda_3\, \xi$$

and an equation in φ. From which it follows that solutions cross all similar ellipses of a certain shape with centre at the origin inwards. It is natural to ask whether this is still the case when a_1, a_2, a_3 are functions of x and $\dot{x}$ provided that

$$\min\,(a_1\, a_2) > \max\, a_3 \tag{6}$$

for the range considered, and in particular whether this is so for equations which occur in connection with servomechanisms in which a_1 is constant and a_2, a_3 are functions of x only when (6) holds.

The theory of non-linear oscillations represented by equations of the fourth and higher orders is even less well established from the point of view of our two problems, and those few results which I have discussed for second and third order equations (and others which I have omitted either from ignorance or lack of time) are far from satisfying the demand of some engineers for a com-

plete theory comparable with the linear theory including Nyquist's criterion for stability. That demand, in view of the extraordinary behaviour of some non-linear systems, seems to me like asking for the moon, but I believe that there are still many openings for the application of real variable and topological methods especially for anyone who can appreciate the physical problems and physical insight of the engineers.

References

M. L. Cartwright, Journal Inst. Elec. Eng. (Radio Section) 95 (1948) 88—96.

M. L. Cartwright, Contributions to the theory of non-linear oscillations 2. Annals of Maths. Studies 29 (1952) 3—18.

M. L. Cartwright and J. E. Littlewood, Annals of Maths. 48 (1947) 472—494

A. W. Gillies, Quarterly Journal of Mech. and Applied Maths., 7 (1954) 152—167.

W. D. Hayes, Zeitschrift f. angewandte Math. und Physik (Zürich) 4 (1953) 398—401.

D. C. McDonald, Proc. of the symposium on non-linear circuit analysis (New York 1953) 402—411.

Addendum

Since the Congress I have found that the problem of third order equations satisfying (6) has to a large extent been solved by E. A. Barbashin, On the stability of the solution of a certain non-linear third order equation, *Akad. Nauk. S.S.S.R. Priklad. Mat. i Mekhan.* (1952) 16(5) 629—632 (See Math. Reviews 14 (1952) 376. An English translation has been made by the Department of Scientific and Industrial Research, C.T.S. No. 57, Technical Information Service, Charles House, 5—11 Regent St., London, S.W.1.) Further papers by Malkin and Erugin on related topics are contained in the same volume of the periodical.

RETRACTION, HOMOTOPY, INTEGRAL

LAMBERTO CESARI

1. Recent studies have shown that parametric surfaces in the ordinary space under the sole hypothesis that they be continuous and have finite area, have remarkable analytical and geometrical properties which have already proved to be adequate in questions of analysis and of the calculus of variations. Analogous results can be expected for k-dimensional parametric varieties V_k in the n-dimensional space E_n or in any metric space E.

Such studies seem to indicate a basis for differential geometry and the calculus of variations for continuous parametric varieties. Since no ordinary differential element necessarily exists, the basic elements like Jacobians, tangent planes and normals, etc., are introduced by topological and set-theoretical methods.

I intend to speak here of the integrals over a continuous parametric surface in E_3 or in E_n. Part of the results have been already extended to the integrals over a k-dimensional continuous parametric variety in E_n, but some important problems connected with this last extension have not yet been solved. In order to be able to present a more complete theory, I shall refer only to integrals over continuous surfaces in E_3.

By a continuous surface S in E_3

$$S = (T, A): p = p(w),\ w \in A,\ p = (x, y, z),\ w = (u, v),$$

I mean any continuous univocal (but not necessarily biunivocal) image of some 2-dimensional set A, say any plane open point set $A = G$, any finitely connected closed Jordan region $A = J$ (with or without identifications on the boundary), any set $A = G \subset J$ open in a region J, any 2-manifold A. In such a way continuous surfaces have here the same generality which is usual in Topology, though properties of them are studied which are analytical, or differential geometrical, or integral geometrical in nature. Thus this study combines methods and problems of Differential Geometry, of Topology, Analysis, and the Calculus of Variations.

2. *Properties of continuous surfaces.* (*a*) *Analytical properties.* Given a surface

$$(1) \qquad S = (T, A):\ x = x(u, v),\ y = y(u, v),\ z = z(u, v),\ (u, v) \in A,$$

that is a continuous mapping $S = (T, A): p = p(w),\ w \in A,\ w = (u, v)$,

$p = (x, y, z)$, from $A \subset E_2$ into E_3, I shall denote by T_1, T_2, T_3 the three plane mappings which are the projections of T onto the coordinate planes E_{21}, E_{22}, E_{23} of E_3:

$$(T_1, A)\colon\ x = 0, \qquad y = y(u, v),\ z = z(u, v),\ (u, v) \in A,$$
$$(T_2, A)\colon\ x = x(u, v),\ y = 0, \qquad z = z(u, v),\ (u, v) \in A,$$
$$(T_3, A)\colon\ x = x(u, v),\ y = y(u, v),\ z = 0, \qquad (u, v) \in A.$$

Concepts of bounded variation (BV) and absolute continuity (AC) of the plane mappings T_1, T_2, T_3 are usually introduced by topological and measure theoretical methods. More precisely concepts of absolute, positive, negative total variations $W(A, T_r)$, $W^+(A, T_r)$, $W^-(A, T_r)$, $r = 1, 2, 3$, and of absolute, positive, negative multiplicity functions $N(p; A, T_r)$, $N^+(p; A, T_r)$, $N^-(p; A, T_r)$, $r = 1, 2, 3$, are introduced, where these concepts are related by the equalities

$$W(A, T_r) = (E_{2r}) \int N(p; A, T_r),\ W^+ = (E_{2r}) \int N^+,\ W^- = (E_{2r}) \int N^-,\ r = 1, 2, 3.$$

Also the relative multiplicity $n(p; A, T_r) = N^+(p) - N^-(p)$ may be introduced. Then the following theorem holds: the Lebesgue area $L(S) = L(A, T)$ of the surface $S = (T, A)$ is finite if and only if all mappings (T_r, A), $r = 1, 2, 3$, are BV (L. Cesari 1942).

For the sake of simplicity let us suppose that A is a simple closed Jordan region. We shall denote by $\Gamma = \Gamma(A, T)$ the collection of all maximal continua of constancy for the mapping T. The collection Γ is a decomposition of A and we shall denote by $\{M\}$ the collection of all sets $M \subset A$, $M = \Sigma g$, which are sums of continua $g \in \Gamma$. Then the Lebesgue area $L(G, T)$, thought of as defined in the σ-ring of all sets $G \in \{M\}$ open in A, can be extended to a function $m(M)$, $M \in \{M\}$, which is a content function (A. Rosenthal, L. Cesari). Thus $m(M)$ is a regular ordinary measure, which is completely additive in the σ-field of all m-measurable sets $M \in \{M\}$, in particular in the σ-field of all Borel measurable sets $M \in \{M\}$.

Let $f(p)$, $p \in [S]$, be any continuous function in the set $[S] = T(A)$. Then for every real number t the compact set $M_t \subset A$, where $f[T(w)] = t$ can be thought of as a „contour" for the surface $S = (T, A)$. A concept of (generalized) Jordan length $l(t)$ of the image on S of the set M_t can be defined in various ways; for instance if we consider for each t the components G of the sets D^-, D^+ open in A where $f[T(w)] < t$, or $f[T(w)] > t$, and for each G the single components continua k of the compact set $F(G) = \overline{G} - G$, then for each k, Carathéodory's prime end theory can be used, where the ordered collection of Carathéodory ends replace the single points in the ordinary definition of length. Under the sole assumption that $f(p)$ is Lipschitzian in S ($| f(p) - f(p') | \leqq$

$K \mid p - p' \mid$ for all $p, p', \epsilon [S]$), the following relation integral geometrical in character (Cavalieri inequality) holds:

$$K \cdot L(S) \geqq \int_{-\infty}^{+\infty} l(t)dt \quad \text{(L. Cesari 1950).} \tag{3}$$

Thus the hypothesis $L(S) < +\infty$ implies that $l(t) < +\infty$ for almost all t, and a more detailed analysis shows that for almost all t the image on S of the contour M_t is a countable collection of continuous closed oriented curves C whose total length is $l(t)$ besides a set of single points which are the images of continua $g \subset A$, $g \epsilon \Gamma$, satisfying properties of total disconnection (L. Cesari 1950).

By a more detailed discussion R. E. Fullerton (1953) has given properties of slenderness of almost all contours M_t. In addition R. E. Fullerton has given a direct topological proof of the general possibility of dividing a non degenerate surface into small pieces with rectifiable boundaries (1953). Thus a property of surfaces needed in the approach to the integral over a surface $I(S)$ of G. M. Ewing (1951) has been obtained by a direct topological and set-theoretical process.

(*b*) *Geometrical properties.* If we denote by geometrical properties those concerning some differential elements, we have to put in this class of properties that each plane BV continuous mapping (T, A): $p = T(w)$, $w \epsilon A$, $w = (u, v)$ $p = (x, y)$, possesses almost everywhere in A^0 a generalized Jacobian $J(w)$, although the single components of the vector $T(w)$ may have ordinary partial derivatives at no point of A^0. The generalized Jacobian $J(w)$ is defined by purely topological and measure theoretical considerations, where Banach's theory for the derivation of normal (subadditive, or overadditive) interval functions is used. The generalized Jacobian $J(w)$ of any plane mapping (T, A) is an L-integrable function in A^0 and $W(A, T) \geqq (A)\int \mid J \mid dw$, where the equality sign holds if and only if T is AC. We have put $J(w) = 0$ everywhere in AA^*. By A^*, $\overline{A}$, A^0 we mean respectively the boundary, the closure, and the set of interior points of any set A.

The following theorem holds: If a continuous surface (1) has finite Lebesgue area, then T_1, T_2, T_3 are BV, the generalized Jacobians $J_r(w)$, $w \epsilon A$, $r = 1, 2, 3$, exist almost everywhere in A, are L-integrable functions in A, and

$$L(S) \geqq (A)\int (J_1^2 + J_2^2 + J_3^2)^{1/2}\, dw = (A)\int \mid J \mid dw, \tag{4}$$

where J is the vector $J = (J_1, J_2, J_3)$, and where the equality sign holds if and only if T_1, T_2, T_3 are AC (L. Cesari, T. Radó). In the last case the mapping (T, A) from A into E_3 is said to be absolutely continuous (AC).

The following theorem concerning the problem of representation also holds:

For each continuous mapping (1) from A into E_3 with finite Lebesgue area there is always another mapping (T', A) which is Fréchet-McShane equivalent to (T, A), which is given by functions x, y, z having partial ordinary derivatives $x_u, x_v, \ldots, z_v$ almost everywhere in A^0 and L^2-integrable in A^0. In addition the differential geometry relations $E = G$, $F = 0$ hold almost everywhere in A^0 where $E = x_u^2 + y_u^2 + z_u^2$, $G = x_v^2 + y_v^2 + z_v^2$, $F = x_u x_v + y_u y_v + z_u z_v$ (thus (T', A) is almost conformal), and the equality sign holds in (4); that is, the Lebesgue area is given by the classical area integral relative to (T', A).

Although S. Saks has shown that surfaces exist of finite area $z = z(u, v)$, $x = u, y = v$, having ordinary tangent plane at no point, it has been proved in different ways by L. Cesari that every continuous surface of finite Lebesgue area has almost everywhere a quasi-tangent plane defined by topological and measure theoretical considerations. Such a plane, say α, has the property that the absolute value of the generalized Jacobian of the projection of S on α is equal to the area dirivative, while the generalized Jacobians of the projections of S on two orthogonal planes β, γ are zero. Thus for what concerns those properties of S which are expressed by the generalized Jacobians it can be said that a surface of finite area has almost everywhere a tangent plane and therefore a normal n (quasi-tangent plane and quasi-normal).

3. *The integral $I(S)$ over a continuous surface.* An integral $I(S) = (S)\int F(p, t)$ over a surface $S = (T, A)$ of a function $F(p, t)$ of the point $\mathrm{p} = (x, y, z) \in [S]$ and the direction $t = (t_1, t_2, t_3)$, could be introduced as a Lebesgue Tonelli integral for AC mappings. Yet it is preferable to introduce it as a Weierstrass integral for every continuous mapping of finite Lebesgue area. This also in accordance with the direct method of the calculus of variations for curves and thus in line with L. Tonelli (1914), K. Menger (1937), C. Pauc, G. Bouligand N. Aronszajn.

Let $F(p, t)$, $p = (x, y, z)$, $t = (t_1, t_2, t_3)$ be any function continuous for all $p \in K$ and $t \neq 0$, where K is some closed subset of E_3, and positively homogeneous in t; i.e., $f(p, kt) = kf(p, t)$ for all $\mathrm{p} \in K$, $t \neq 0$, $k > 0$. Then $f(p, t)$ with $f(p, 0) = 0$ is continuous for all $p \in K$ and t.

Let (1) be any surface with $[S] \subset K$, $L(S) < +\infty$. For any simple closed polygonal region $q \subset A$, let $C, C_r, r = 1, 2, 3$, be the continuous closed oriented curves which are the images of the boundary q^* of q under T, T_r. Let $O(p; C_r)$ be the topological index of the point $p \in E_{2r}$ with respect to the curve C_r in E_{2r} where $O(p; C_r) = 0$ for every point $p \in [C_r]$. Thus $O(p; C_r)$ is an L-integrable function in E_{2r} and we denote by $u(q, T)$ the number

$$u(q, T) = (u_1^2 + u_2^2 + u_3^2)^{1/2} \geqq 0, \text{ where } u_r = u(q, T_r) = (E_{2r})\int O(p; C_r).$$

If $u(q, T) > 0$ then the quotients

$$a_r = u(q, T_r)/u(q, T),\ r = 1, 2, 3,\ a_1^2 + a_2^2 + a_3^2 = 1,$$

can be thought of as the direction cosines of the „average" normal to the piece of the surface S defined by T on q. We shall denote by u and a the vectors $u = (u_1, u_2, u_3)$, $a = (a_1, a_2, a_3)$.

Now, if D denotes any finite system of simple closed non-overlapping polygonal regions $q \subset A$, we may choose any point $\bar{w} \in q$ from each q and denote by $\bar{p} = T(\bar{w})$ the image of $\bar{w}$. In different ways it is possible to define an index δ measuring the degree of fineness of the system D. Then by the integral $I(S)$ of F on S we define the limit

$$I(S) = I(A, T) = (S)\int F(p, t) = \lim_{\delta \to 0} \sum_{q \in D} F(\bar{p}, u) = \lim_{\delta \to 0} \sum_{q \in D} F(\bar{p}, a)u(q, T).$$

I have proved the existence and finiteness of this limit in 1948 for every continuous mapping (T, A) with $L(S) < +\infty$. In addition, I have proved the following main properties of $I(S)$: (1) $I(S)$ is invariant with respect to the Fréchet-McShane equivalence of mappings; (2) If $S_n \to S$, and $L(S_n) \to L(S)$, then $I(S_n) \to I(S)$ (thus $I(S)$ can be approached by means of elementary surface integrals); (3) If the Lebesgue area is given by the classical area integral $L(S) = (A)\int |J|\, dw$, then $I(S)$ is given by a Lebesgue-Tonelli integral $I(S) = (A)\int F[p(w), J(w)]dw$, where $J = (J_1, J_2, J_3)$ is the vector of the generalized (or ordinary) Jacobians (L. Cesari 1948).

4. *The Gauss-Green and Stokes theorems.* Let $S = (T, A)$ be any continuous mapping from a simple closed Jordan region and C: (T, A^*) the oriented boundary of S. Let $A(p)$, $B(p)$, $C(p)$, $p \in E_3$, be functions continuous in E_3 with their partial derivatives A_y, A_z, B_z, B_x, C_x, C_y. Then the following extension of the Stokes theorem holds: If $L(S) < +\infty$ and $l(C) < +\infty$, then

(5) $(C)\int (At_1 + Bt_2 + Ct_3) = (S)\int [(C_y - B_z)t_1 + (A_z - C_x)t_2 + (B_x - A_y)t_3]$,

where the left hand member is a Weierstrass integral over the curve C and the right hand member is a linear integral $I(S)$ (J. Cecconi 1952). Cecconi has also given various extensions of the above statement to the case where A, B, C satisfy weaker conditions and to the case where T is a mapping from a finitely connected Jordan region.

Let $S = (T, A)$ be any closed surface (that is, T is constant on A^*), and let $O(p; S)$ be the topological index of any point $p \in E_3$ with respect to S [$O(p; S) = 0$ for every point $p \in [S]$]. Let $A(p)$, $B(p)$, $C(p)$ be functions continuous in E_3 with their partial derivatives A_x, B_y, C_z. Then the following extension of the Gauss-Green theorem holds: If $L(S) < +\infty$, $|[S]| = 0$, then

(6) $$(E_3)\int (A_x + B_y + C_z)O(p; S)dxdydz = (S)\int (At_1 + Bt_2 + Ct_3),$$

where the left hand member is a Lebesgue integral and the right hand member is a linear integral $I(S)$ (J. Cecconi 1951).

Extensions have been given by J. Cecconi to the case where A, B, C satisfy weaker conditions in E_3. The condition $|[S]| = 0$ cannot be omitted in (6) as J. Cecconi has shown by examples. This is in connection with the following theorem by R. G. Helsel concerning the convergence in measure of the topological index: If $S_n \to S$, $L(S_n) \to L(S)$, $|[S]| = 0$, then $(E_3)\int |O(p; S_n) - O(p; S)| \to 0$. Nevertheless, J. Cecconi has just announced that the same condition $|[S]| = 0$ can be omitted provided the index $O(p; S)$ is replaced by a more refined index, say $n(p)$ of the type of a relative multiplicity, and $n(p)$ coincides almost everywhere with $O(p; S)$ in $E_3 - [S]$. Other studies by Okamura (1950) and S. Mizohata (1950) have shown that the same condition $|[S]| = 0$ can be omitted in (6) under additional metric hypotheses on S.

5. *Semicontinuity and continuity properties of the integral* $I(S)$. If we suppose that $F(p, t)$ has partial derivatives with respect to t_1, t_2, t_3 continuous for all $p \in K$ and $t \neq 0$ then the definitions of regularity of the integral $I(S)$ can be introduced as usual in the calculus of variations by means of the Weierstrass function $E(p, t, t')$. The following theorem holds: Every positive semidefinite and semiregular integral $I(S)$ is lower semicontinuous in the class of all surfaces of finite Lebesgue area (L. Cesari 1948).

The condition $F \geqq 0$ on the boundary of S is a necessary condition for the lower semicontinuity of $I(S)$ on a surface S (E. J. McShane 1933). Also, a condition of regularity is necessary for the lower semicontinuity on a given surface S, namely $E[p(w), n(w), t'] \geqq 0$ at almost every point $w \in A^0$ where the generalized vector Jacobian $J(w) \neq 0$. Here $n(w)$ denotes the direction cosines of the quasi-normal $n = n(w)$ to the surface S and t' is any vector non-proportional to $n(w)$.

A theory of the semicontinuity of the integral $I(S)$ based merely on properties of convexity of the indicatrix $s = F(p, t)$, and hence in connection with Menger's approach for curves has been not yet discussed.

Under a condition of regularity somewhat stronger than usual, V. E. Bononcini has recently proved (1952) that if $I(S)$ is definite, positive, regular, then $I(S_n) \to I(S)$ implies $L(S_n) \to L(S)$. Thus if both integrals $I(S)$, $J(S)$ are definite positive and regular in the strong sense mentioned, then $I(S_n) \to I(S)$ implies $J(S_n) \to J(S)$. Other results were already known for surfaces given in canonical representations.

Finally, by using a particular refined theorem of approximation of the area of a surface (L. Cesari), V. E. Bononcini (1954) has proved the following theorem of continuity for linear integrals: If $I(S) = (S)\int (At_1 + Bt_2 + Ct_3)$

is any linear integral, and S_n, S are closed surfaces with $S_n \to S$, $L(S_n) \leqq M$, then $I(S_n) \to I(S)$. From here it follows that the surface integral

$$V(S) = (S)\int xt_1 = (S)\int yt_2 = (S)\int zt_3,$$

which can be thought of as a volume $V(S)$ enclosed in the (closed) surface S, has the following property of continuity: If $S_n \to S$, $L(S_n) \leqq M$, then $V(S_n) \to V(S)$ (V. E. Bononcini). J. Cecconi has recently that the volume $V(S)$ defined above as a surface integral can be thought of also as an integral in E_3, namely $V(S) = (E_3)\int n(p)$, where $n(p)$ is the same function mentioned in connection with the Gauss-Green theorem. In addition the same volume $V(S)$ satisfies the isoperimetric inequality

$$V^2(S) \leqq (36\pi)^{-1}L^3(S),$$

where S is any closed continuous surface. Also, the volume $V_R(S) = (E_3)\int |O(p; S)|$ considered by T. Radó (1947) satisfies the same inequality. The same holds for the volume studies by J. W. T. Youngs $V_Y(S) = (E_3)\int \varphi(p)$, where $\varphi(p) = \Sigma \mid O(p; S_i) \mid$ and S_i, $i = 1, 2, \ldots$ are the cyclic elements of S.

If $S = (T, A)$ is any continuous mapping from a simple closed Jordan region with $L(S) < +\infty$ and S_i, $i = 1, 2, \ldots$, are the cyclic elements of S, then for every integral $I(S)$ we have $I(S) = \Sigma I(S_i)$ (cyclic additivity theorem for the integral $I(S)$) (J. Cecconi 1953). This equality extends to every integral $I(S)$ the cyclic additivity theorem for the Lebesgue area of C. B. Morrey (1935).

6. *Retraction and homotopy.* Given a continuous mapping $S = (T, A)$ from a simple closed Jordan region and a continuum $K \subset A$ such that for every component G of $A - K$ the mapping T is constant on the subcontinuum $F(G) = \overline{G} - G \subset K$, let (T_0, A) denote the new mapping defined by $T_0 = T$ on K, and by $T_0(w) = T[F(G)]$ on every component G of $A - K$. The mapping (T_0, A) is called the retraction of T with respect to the continuum K. The cyclic elements S_i of S are particular retractions of S. We shall denote by θS the boundary curve $C = (T, A^*)$ of the surface $S = (T, A)$. We shall denote by $C_1 \sim C_2(M)$ the homotopy of two given closed continuous curves C_1, C_2 in a set M. Thus $C \sim 0(M)$, $C \not\sim 0(M)$ mean that the closed curve C is nulhomotop, or is not nulhomotop in the set M. We mention finally that a surface $S = (T, A)$ is said to be *base* (*open non-degenerate*) if for every maximal continuum of constancy $g \subset A$, $g \in \Gamma$, the set $E_2 - g\,[A - g]$ is connected. Then every surface S has a retraction S_0 which is a base surface (E. J. McShane, C. B. Morrey 1933). In addition, for every positive semidefinite integral $I(S)$ and every retraction of S we have $I(S_0) \leqq I(S)$. The following theorems hold:

(1) If C is any simple closed oriented curve in E_3 and ϵ any positive num-

ber, there is a number $d = d(\epsilon, C) > 0$ such that every surface $S = (T, A)$ with $|| \theta S, C || < d$ has a retraction S^0 which is open non-degenerate and $|| \theta S_0, C || < \epsilon$ (L. Cesari 1952).

(2) Given C as in (1) every surface S with $[\theta S] \subset [C]$, $\theta S \not\approx 0([C])$, has a retraction S_0 which is open non-degenerate and $[\theta S_0] \subset [C]$, $\theta S_0 \not\approx 0([C])$ (L. Cesari 1954).

(3) Given a simple closed curve C on a plane $a \subset E_3$ and any surface S whose projection S_a onto a has the property $[\theta S_a] \subset [C]$, $\theta S_a \not\approx 0([C])$, then there exists a retraction S_0 of S which is open non-degenerate and whose projection S_{0a} onto a has the properties $[\theta S_{0a}] \subset [C]$, $\theta S_{0a} \not\approx 0([C])$ (L. Cesari 1954).

The theorem (1) has been directly used in questions of the calculus of variations, the theorems (2), (3) have been used in a new recent proof of the general representation theorem mentioned above.

7. *Calculus of variations in homotopy classes.* On the basis of surface area theory (particularly of the concept of integral $I(S)$ introduced above and its properties of lower semicontinuity, §§ 3, 4, 5) the following existence theorem of the calculus of variations has been recently proved by J. Sigalow, J. M. Danskin, and L. Cesari (1950). If C is any continuous closed simple curve in a convex bounded closed set K, $C \subset K \subset E_3$, if $\{S\}$ is the class (non empty) of all surfaces $S = (T, A)$, $[S] \subset K$, with $\theta S = C$, then every positive definite regular integral $I(S)$ has an absolute minimum in $\{S\}$.

For closed surfaces the following theorem concerning classes of homotopic surfaces has been recently announced by J. Cecconi: If s is any open sphere of center o in a convex bounded closed set K, $s \subset K \subset E_3$, if $\{S\}$ is the class (non empty) ot all closed surfaces $S = (T, A)$ with $[S] \subset K - s$, $O(o; S) = n$ where n is any integer (thus $\{S\}$ is a class of surfaces homotopic to one another in $K - s$), then every positive definite regular integral $I(S)$ has an absolute minimum in $\{S\}$.

Extensions of both theorems can be given for semiregular positive definite integrals $I(S)$.

ON SOME PROBLEMS IN FOURIER ANALYSIS

K. CHANDRASEKHARAN

1. I shall consider here some results of recent research on Fourier series (F.s.) in several variables and other related expansions. It will appear that while in some respects there is an analogy between the behaviour of F.s. in one variable and in several variables, there is a marked contrast in other respects, which justifies the study. Let $f(X) \equiv f(x_1, \ldots, x_k) \in L^1$ in $0 \leqq x_1 < 2\pi, \ldots, 0 \leqq x_k < 2\pi$, and be periodic in each variable with period 2π. If its F.s. is denoted by $\sum a_{n_1,\ldots,n_k} \exp\{i(n_1x_1 + \ldots + n_kx_k)\}$, then one can sum this multiple series in many ways. One can consider the 'rectangular' partial sums

$$S_{N_1,\ldots,N_k} = \sum_{|n_r| \leqq N_r} a_{n_1,\ldots,n_k} \exp\{i(n_1x_1 + \ldots + n_kx_k)\},$$

or the 'square' partial sums $\{S_{N,\ldots,N}\}$, or the 'triangular' partial sums $\{S_M\}$ with $|n_1| + \ldots + |n_k| \leqq M$, or the 'circular' partial sums $\{S_R\}$ with $(n_1^2 + \ldots + n_k^2) \leqq R^2$. Each of these methods of summation will be seen to have its own interest. Thus the study of rectangular partial sums is connected with the differentiation of multiple integrals [14], while the study of the circular partial sums is connected with eigenfunction expansions and lattice-point problems [5, 6, 7].

2. The first problem I consider is the problem of 'localization'. A method of summation may be said to have the 'localization property' if, for any integrable function f which vanishes in a neighbourhood of a point X, the corresponding F.s. is summable at that point, by that method, to the sum zero. It is a fundamental theorem of Riemann that ordinary convergence has the localization property for F.s. in one variable. This is no longer true of several variables. It is known, however, that if $k = 2$, Cesáro's method (C, α) for $\alpha \geqq 1$ has the localization property for the square sums $\{S_{N,N}\}$, and the triangular sums S_M, while (C, α) for $\alpha < 1$ does not [11]. If $k \geqq 3$, then no (C, α) method for $\alpha \geqq 0$, nor even Abel's method, has the localization property either for square or for triangular partial sums. Further, in the case of rectangular sums, while the restricted (C, α, β) method has the localization property, for $k = 2$, if and only if $\alpha \geqq 1$, $\beta \geqq 1$, *no* restricted method $(C, \alpha, \beta, \ldots, \varkappa)$, $\alpha, \beta, \ldots, \varkappa \geqq 0$, can have the localization property if $k \geqq 3$ [12]. ['Restricted' here refers to the fact that in taking the multiple Cesáro limit for the sums $\{S_{N_1,\ldots,N_k}\}$ the indices

tend to infinity in such a way that $(N_i/N_j) \leqq \lambda$; $i, j = 1, \ldots, k$; $\lambda \geqq 1$]. This 'local' theorem of Herriot seems to be interesting when it is viewed against the 'global' theorem of Marcinkiewicz and Zygmund [14] that the restricted method $(C, \alpha, \beta, \ldots, \varkappa)$, $\alpha, \beta, \ldots, \varkappa > 0$ sums the rectangular F.s. of an integrable function almost everywhere, no matter what k is. Bochner [5] has shown, on the other hand, that if one considers circular partial sums, in any number of dimensions, then a variety of summation-processes can be associated, which do have the localization property. In particular, the method (C, α) for $\alpha > \frac{1}{2}(k-1)$ has the localization property, while (C, α) for $\alpha = \frac{1}{2}(k-1)$ does *not*. The proof of the latter fact, as given by Bochner, is by reductio ad absurdum. It would be of interest to have a concrete example of a function which vanishes in a neighbourhood of the origin, say, and such that

$$\limsup_{R\to\infty} | S_R^{\frac{1}{2}(k-1)}(0) | = \infty,$$

where $S_R^{\frac{1}{2}(k-1)}(0)$ is the $\left(C, \frac{1}{2}(k-1)\right)$ mean of the circular partial sum of the F.s. of f at the origin. (A similar remark applies to Herriot's result mentioned above). Bochner has also shown that if one considers multiple Fourier integrals (F.i.) instead of series, then (C, α) does have the localization property even for $\alpha = \frac{1}{2}(k-1)$. [That it fails for $\alpha < \frac{1}{2}(k-1)$ both for series and integrals is, of course, known]. The same is again true of the F.s. of functions in L^2, and this fact is an indirect consequence of the estimate for the number of lattice-points in a k-dimensional sphere [5]. It would be of interest to settle the question whether for functions in L^p $1 < p < 2$, the method $\left(C, \frac{1}{2}(k-1)\right)$ does or does not have the localization property. Allied with this is the problem of generalizing the fundamental theorem of Marcel Riesz on L^p classes. If $S_R^\delta(X)$ denotes the δ^{th} Cesáro-Riesz mean of the circular partial sum $S_R(X)$ at a point $X = (x_1, \ldots, x_k)$, the problem is to determine the smallest $\delta = \delta(p, k)$, such that for every $f \in L^p$, $p > 1$, and for all $R > 0$, we have

$$\int_0^{2\pi} \cdots \int_0^{2\pi} | S_R^\delta(X) |^p \, dX \leqq A_p \int_0^{2\pi} \cdots \int_0^{2\pi} | f(X) |^p \, dX.$$

For $k = 1$, M. Riesz has proved that $\delta = 0$. For general k it is not known whether we have such a result for all sufficiently large δ, and if we do, what the smallest such δ is. This theorem of Riesz is basic in the study of conjugate trigonometric series (t.s.) and conjugate functions. Some kind of parallel theory for several variables has been given in the case of rectangular sums [20], but neither the results nor even the right definitions seem to have been established in the case of circular means, where one can expect the basic results on conjugate series to go through. On the analogy of known theorems [13] in one variable, it would be interesting to give, for circular means, conditions on the

convergence or summability of a F.s. or t.s. which suffice to infer that the conjugate t.s. are convergent or summable almost everywhere.

Riemann's theorem on localization has an extension from Fourier series to trigonometric series. If T_1 and T_2 are two t.s. with coefficients tending to zero, and if $F_1(x)$, $F_2(x)$ are the sums of the series when integrated formally twice, if the difference $F_1(x) - F_2(x)$ vanishes, or is a linear function, in an interval $(x_0 - \varepsilon, x_0 + \varepsilon)$, $\varepsilon > 0$, then the series are equi-convergent at x_0; if $F_1(x) - F_2(x)$ is a linear function in the interval $a \leqq x \leqq b$, then the series T_1, T_2 are uniformly equiconvergent in every interval (a', b') interior to (a, b); further the series conjugate to T_1, T_2 are uniformly equi-convergent, in the wider sense, in (a', b'). This result of Riemann's has been extended by Rajchman and Zygmund [17, 24] much further, by the process of 'formal multiplication', to series whose coefficients do not tend to zero. Parallel results, in several variables, have been recently obtained [4, 19] in the case of circular and square sums which again show a certain difference, but further study seems to be necessary in the case of conjugate series. Here formal integration is replaced by the application of the anti-Laplacian.

Zygmund [25] has also extended the results on the localization of t.s. to trigonometric integrals (t.i.) by a general method which consists in directly tackling the difference between the respective partial sums. It has been shown [4b] that that method admits of an extension to double series and integrals if restricted convergence is used, and to several variables if circular convergence is used. However the case of conjugate integrals is yet to be studied."

3. The next problem I consider is that of uniqueness of multiple t.s. The basic result, in one variable, due to Cantor is that if a t.s. converges everywhere to zero, it vanishes identically. More generally, we have de la vallée Poussin's theorem that if a trigonometric series converges everywhere to a Lebesgue-integrable function f, then it is the (Lebesgue) Fourier series of f. We can go further and say that if a t.s. converges, except in an enumerable set E, to a finite and Lebesgue-integrable function f, then it is the (Lebesgue) F.s. of f. This can be generalized in many directions; one can seek to generalize (i) the exceptional set, or (ii) the concept of convergence, or (iii) the concept of integral, or combine the three in various ways. In the first direction, satisfactory results are not known since the structural characteristics of the exceptional set are not completely known. This leads to the well-known problem of characterizing sets of unicity from among sets of measure zero. But in the second direction, far-reaching results have been obtained by Verblunsky [21]. He has shown, in particular, that if a t.s. has coefficients a_n, $b_n = o(n)$, and its upper and lower Abel sums f^* and f_* are finite everywhere and Lebesgue integrable, then the series is a (Lebesgue) Fourier series. In particular, if the series has a_n,

$b_n = o(n)$, and is Abel summable to zero everywhere, then $a_n = b_n = 0$ for all n. From this one can deduce that a t.s. which is summable $(C, 1)$ to zero everywhere vanishes identically. Here no exceptional point can be admitted. Wolf has extended [22] Verblunsky's results to Cesáro means of general order. While there is scope for further work in the case of Abel means, even in one variable, a satisfactory generalization of the known results to several variables has not yet been obtained. Some preliminary work has, however, been done. M. T. Cheng [10] has shown that if one considers those double t.s. which satisfy the condition that

$$\sum_{n_1^2+\ldots+n_k^2 \leqq R^2} \frac{a_{n_1,\ldots,n_k}\, e^{i(n_1x_1+\cdots+n_kx_k)}}{(n_1^2+\ldots+n_k^2)}$$

converges uniformly in (x) as $R \to \infty$, then circular $(C, 1)$ summability everywhere to zero implies that the series vanishes identically. Here the condition imposed is very restrictive, and has been somewhat relaxed by Shapiro [18], but there is obviously scope for further work.

Before turning to generalizations in the third direction, we may consider trigonometric integrals (t.i.), the uniqueness problem for which was first tackled satisfactorily by Offord [16] who proved that if $f(x) \in L^1$ in every finite interval, and

$$\lim_{\omega\to\infty} \int_{-\omega}^{\omega} f(x)e^{i\alpha x}dx = 0,$$

for every real α, then $f(x) = 0$ almost everywhere. This may be considered as the integral analogue of Cantor's theorem for series. It has been generalized by Offord himself and Wolf [16, 23] on the lines of the earlier results of Verblunsky and Wolf on series. The problem of extending Offord's theorem, even in its simplest form, either to several variables, or to other types of integral in one variable, like the Hankel integral, has not yet been solved. The latter will be of interest in view of the fact that the Cesáro-Riesz mean of the circular f.s. of $f(X)$ and the 'circular mean' of the function $f(X)$, are formally Hankel transforms of each other. [9]

In the third direction of generalization, Burkill has recently given a new definition of an integral [8] called the SCP-integral which is more comprehensive then the first Denjoy integral and which seems to be the most appropriate in the context of uniqueness theorems. He has shown that if a t.s. converges everywhere to a sum $f(x)$, then $f(x)$ is SCP-integrable, and the coefficients a_n, b_n of the series can be written in the Fourier form as SCP-integrals. He has also shown [8] that the results of Verblunsky [21] and Offord [16] and Wolf [23], both for series and integrals, can be upheld if one works with SCP-integrable functions instead of Lebesgue- or Denjoy-integrable functions as those authors

had earlier done. The structure of the SCP-integral can become fully clear perhaps only by a study of its multi-dimensional analogue.

4. I shall now consider some generalizations and applications of the localization theory for circular Fourier series. If we consider the boundary-value problem $\Delta\omega(X) + \lambda\omega(X) = 0$, where $\Delta\omega$ is the Laplace operator with respect to the Euclidean metric, on the torus $0 \leqq x_1 < 2\pi, \ldots, 0 \leqq x_k < 2\pi$, then $\omega(X) = \exp\{i(n_1x_1 + \ldots + n_kx_k)\}$, where (n_k) are integers, form a complete set of eigenfunctions with the eigenvalues $\lambda = n_1^2 + \ldots + n_k^2$. Circular summation amounts to ordering the terms according to the magnitude of the eigenvalues λ. On this analogy, one can consider a somewhat different boundary-value problem $\Delta\omega(X) + \lambda\omega(X) = 0$ in D, where D is a bounded domain with a sufficiently regular boundary B, with the boundary condition $\omega(X) = 0$ for $X \in B$. There still exist an infinity of eigenfunctions $\omega_n(X)$, which are no longer bounded, with non-negative eigenvalues λ_n. One can then expand an arbitrary function $f \in L^2$ as a Fourier eigenfunction expansion:

$$f(X) \sim \Sigma\, a_n\omega_n(X), \quad a_n = \int_D f\omega_n dx.$$

It has been shown recently [2], on the basis of an estimate for these eigenfunctions due to Avakumovic' [3], that $(C, \frac{1}{2}(k-1))$ has the localization property for this class of expansions as well. Avakumovic''s estimate is that

$$\sum_{\lambda_n \leqq R} \omega_n^2(X) = c_kR^{k/2} + O\left(R^{\frac{1}{2}(k-1)}\right),$$

and this plays the same role here as the estimate for the number of lattice-points in a circle does in Bochner's theorem for F.s. of functions in L^2 [5]. It is probably true that for functions in L^2 such expansions are equisummable with the circular Fourier series almost everywhere, on the analogy of results on Sturm-Liouville series in one variable. As to the uniqueness of such expansions (not assumed to be 'Fourier') a preliminary result [15] formally similar to that of Cheng's [loc. cit.] but earlier than Cheng's is all that is known.

When we sum over circles, the number of terms that arise at each stage equals the number of lattice points inside a circle. Let $r(n)$ denote the number of lattice points *on* a circle of radius $\sqrt{n}$ with the origin as centre, and let $R(x) = \sum_{n \leqq x}{}' r(n)$, the prime indicating that the last term in the summation should be halved if x is an integer. More generaly let

$$R^\alpha(x) = \sum_{n \leqq x} (x-n)^\alpha r(n), \ \alpha > 0; \ R^0(x) \equiv R(x).$$

Then, if x is non-integral, the following identity due to G. H. Hardy is classical:

$$R^\alpha(x) = \frac{\pi x^{1+\alpha}}{1+\alpha} - x^\alpha - \pi^{-\alpha}\Gamma(1+\alpha)x^{\frac{1}{2}+\frac{1}{2}\alpha} \sum \frac{r(n)J_{1+\alpha}\{2\pi(nx)^{\frac{1}{2}}\}}{n^{\frac{1}{2}+\frac{1}{2}\alpha}}.$$

It has extensions to higher dimensional ellipsoids, not only for $\alpha \geqq 0$ but also for $\alpha < 0$, with summable, instead of convergent, series occurring on the right-hand side. The recognition that the right-hand side is the circular F.s. of the function on the left, allows us to invoke 'local' theorems on the point-wise convergence of such series, which are peculiar to the multi-dimensional case, and, leads [6, 7] not only to particularly simple proofs of many of the known results in all dimensions, but to further generalizations where the function $r(n)$ is replaced by

$$r_k(n, h) = \sum_{n_1^2+\ldots+n_k^2=n} e^{2\pi i(n_1h_1+\ldots+n_kh_k)},$$

h being a point in k-space, or even more generally by

$$r_k^P(\lambda_m, h) = \sum_{P(n,n)=\lambda_m} e^{2\pi i(n_1h_1+\ldots+n_kh_k)},$$

where λ_m is the m^{th} number occurring in the sequence of values of a positive definite symmetric form $P(n, n)$ when they are arranged in increasing order of magnitude. This involves the study of such series as

$$\sum r_k^P(\lambda_n, h) J_\mu\{2\pi\xi\lambda_n^{\frac{1}{2}}\}\lambda_n^{-s},$$

for all real s and for $\mu > -1$. The results obtained for positive-definite forms are not as sharp as those obtained for unimodular forms ($\lambda_n = n$) [6]. In the latter case it has been shown [1] that there is a distinction in behaviour according as h is or is not a lattice point. However not all of the results obtained so far are the best possible.

As a further generalization, one can consider the case where P is an indefinite quadratic form. The function $r_k^P(\lambda_n, h)$ will then have to be replaced by C. L. Siegel's ‚measure of representation'. Some results in this direction have been obtained by V. V. Rao, a student working in the Tata Institute. Here the Bessel function J_μ will be replaced, in case the determinant of P is negative, by Bessel functions of the second kind. Further work on these lines, such as an extension to matrix variables, is indicated.

References

[1] T. V. Avadhani: J. Indian Math. Soc. 16 (1952) 103—125.

[2] T. V. Avadhani: J. Indian Math. Soc. 18 (1954) 9—18.

[3] V. G. Avakumovic': Publ. de l'Institut Math. de l'academie Serbe des sciences, 4 (1952), 95—96.

[4] L. D. Berkovitz: Trans. Amer. Math. Soc. 70 (1951) 323—344; ibid. 73 (1952) 345—372.

[5] S. Bochner: Trans. Amer. Math. Soc. 40 (1936) 175—207.

[6] S. Bochner and K. Chandrasekharan: Quarterly J. Math. (Oxford) 19 (1948) 238—248; (2) 1 (1950) 80.

[7] S. Bochner and K. Chandrasekharan: Acta Szeged, 12 (1950) 1—15.

[8] J. C. Burkill: Proc. London Math. Soc. (3) 1 (1951) 46—57; 163—169
[9] K. Chandrasekharan: Proc. London Math. Soc. (2) 50 (1948) 210—222.
[10] M. T. Cheng: Annals of Math. 52 (1950) 403—416.
[11] J. G. Herriot: Trans. Amer. Math. Soc. 52 (1942) 72—94.
[12] J. G. Herriot: Duke Math. J. 11 (1944) 735—754.
[13] B. Kuttner: J. London Math. Soc. 12 (1935) 131.
[14] J. Marcinkiewicz and A. Zygmund: Fund. Math. 32 (1939) 122—132.
[15] S. Minakshisundaram: Proc. Nat. Acad. Sci. U.S.A. 33 (1947) 76—77.
[16] A. C. Offord: J. London Math. Soc. 11 (1936) 171—174; Proc. London Math. Soc. (2) 42 (1937) 422—80.
[17] A. Rajchman and A. Zygmund: Math. Zeit 25 (1926) 261—273.
[18] V. L. Shapiro: Duke Math. J. 20 (1953) 359—365; Proc. Amer. Math. Soc. 4 (1953) 692—695.
[19] V. L. Shapiro: Amer. J. Math. 75 (1953) 347—357.
[20] K. Sokol Sokolowski: Fund. Math. 34 (1946) 166—182.
[21] S. Verblunsky: Proc. London Math. Soc. (2) 31 (1930), 370—86.
[22] F. Wolf: Proc. London Math. Soc. 45 (1939) 328—356).
[23] F. Wolf: Univ. Calif. Pub. Math. 1 (1947) 159—228.
[24] A. Zygmund: Math. Zeit. 24 (1926) 47—104.
[25] A. Zygmund: *a.* Math. Annalen 99 (1928), 562—589 *b.* Annals of Math. 48(1947), 393—440.

Tata Institute of Fundamental Research
Bombay

ASYMPTOTIC SOLUTIONS OF DIFFERENTIAL EQUATIONS WITH TRANSITION POINTS*

A. ERDÉLYI

1. In the second one of his classical memoirs on what is now called the Sturm-Liouville problem, Liouville (1837, section III) investigated the behaviour of solutions of ordinary linear differential equations containing a large parameter. For the sake of simplicity, let

$$\frac{d^2y}{dx^2} + [\varrho^2 p(x) + q(x)]y = 0 \tag{1}$$

be the differential equation; assume that $q(x)$ is continuous, $p(x)$ is positive and twice continuously differentiable on the closed bounded interval I, and ϱ is a large parameter. Liouville transforms (1) by setting

$$\xi = \int \sqrt{p(x)}\,dx, \qquad \eta = p^{1/4}y \tag{2}$$

and obtains

$$\frac{d^2\eta}{d\xi^2} + \varrho^2\eta = r(\xi)\eta, \quad r(\xi) = \frac{1}{4}\frac{p''}{p^2} - \frac{5}{16}\frac{p'^2}{p^3} - \frac{q}{p}. \tag{3}$$

When ϱ is large, $r\eta$ can be neglected in the first approximation, so that solutions of (3) are approximately linear combinations of $\cos \varrho\xi$ and $\sin \varrho\xi$. More exactly, Liouville uses the method of variation of parameters to show that solutions of (3) satisfy what is now called a Volterra integral equation,

$$\eta(\xi) = c_1 \cos \varrho\xi + c_2 \sin \varrho\xi + \frac{1}{\varrho}\int_\alpha^\xi \sin \varrho(\xi - \tau) r(\tau)\eta(\tau)\,d\tau, \tag{4}$$

and proposes to solve this equation by the method of successive approximations.

It is quite easy to adapt Liouville's method for the case of a $p(x)$ which is negative on the whole interval, but the method breaks down if p changes its sign in I, for near a zero of p the transformation (2) is singular, and r cannot be neglected in (3).

For nearly a century Liouville's method was developed, and extended to equations in which the variables, functions, parameters are complex, to

*) Prepared under contract with the Office of Naval Research, U.S.A. Reproduction in whole or in part is permitted for any purpose of the United States Government.

equations of order n, to systems of differential equations, and to equations which depend on the large parameter in a more complicated manner than (1). The principal contributions are due to Horn, Blumenthal, Perron in Germany, and G. D. Birkhoff, M. H. Stone, Tamarkin, Trjitzinsky, Turritin in the U.S.A. In all this work it was assumed that the condition $p(x) \neq 0$, or the corresponding condition in more general cases, was satisfied.

Jeffreys (1953) has recently drawn attention to a paper by Green (1837) in which approximate solutions of a differential equation are obtained by methods very similar to those of Liouville.

2. About thirty years ago, several important problems arose in which $p(x)$ changes its sign in I. Oscillations of water in an elliptic lake lead to an equation of the form (1) with $p(x) = 1 - h^2 \cos 2x$, and if $h > 1$, $p(x)$ changes its sign. The one-dimensional Schrödinger equation is also of the form (1) with $p(x) = E - V(x)$, and for many potentials $p(x)$ changes its sign twice.

Jeffreys (1924—25) considered differential equations in which $p(x)$ has a zero. We shall again consider the simplest case,

$$\frac{d^2y}{dx^2} - \nu^2 p(x) y = 0 \tag{5}$$

where we assume that $p(x)$ has a simple zero at an interior point, x_0, of the interval I. For the sake of definiteness, we assume that $p(x)$ is positive when $x > x_0$, and negative when $x < x_0$. Let ε be a fixed small positive number, and let us break up the interval I in three parts, I_-, I_0, and I_+, the points of subdivision being $x_0 \pm \varepsilon$. Liouville's approximation holds in I_+ and in I_-. In I_0, Jeffreys approximates $p(x)$ by a linear function of x: the approximate differential equation can be solved explicitly in terms of Bessel functions of orders $\pm 1/3$, and these provide asymptotic solutions of (5) in I_0. The connection with the asymptotic solutions in I_+ and I_- is then established by means of the well-known asymptotic forms of Bessel functions. The solutions are „matched" at $x_0 \pm \varepsilon$, and it turns out that the „connection formulae" do not depend on the choice of ε. This method was later developed and generalized to zeros of higher orders by Goldstein and Jeffreys. Similar methods were also developed by theoretical physicists, especially Wenzel, Kramers, Brillouin, and Zwaan, and became known as the WKB method (now also called the JWKB method).

Assuming ν to be positive, we can describe the characteristic feature of this situation as follows. Every solution of (5) shows an oscillatory behaviour in I_-, and an exponential behaviour in I_+. In I_0, every solution passes from one type of behaviour to the other: this interval may be called a *transition region*, and x_0 a *transition* (or *turning*) *point*. (Note that the transition region is not sharply defined.) In the transition region the behaviour of the solutions of (5)

cannot be described in terms of elementary functions, and Bessel functions of orders $\pm 1/3$, or, what is more or less the same, Airy integrals, offer themselves as the simplest higher transcendental functions which may be used for comparison.

3. A thorough mathematical study of this problem was taken up by Langer. Instead of „piecing together" several asymptotic forms, Langer showed that it is possible to adapt Liouville's method to the present case so as to obtain a single asymptotic solution which is valid uniformly over the entire range of x. Langer has developed his theory for general second-order equations with a complex variable and complex parameter, and allowed $p(x)$ to have a single zero of any order, and he and his pupils applied his method to numerous special differential equations arising in mathematics and physics. We shall restrict ourselves to a brief indication of the salient features of Langer's method in the case of the differential equation (5). For the sake of simplicity, we assume that $p(x)$ has a simple zero at $x = 0$. Langer uses the transformation

$$\xi = \varphi_0(x) = \begin{cases} \left[\frac{3}{2}\int_0^x \sqrt{p(t)}\,dt\right]^{2/3} & x > 0 \\ -\left[\frac{3}{2}\int_x^0 \sqrt{-p(t)}\,dt\right]^{2/3} & x < 0 \end{cases} \tag{6}$$

$$\eta = \sqrt{\varphi_0'}\,y$$

which carries (5) into

$$\frac{d^2\eta}{d\xi^2} - \nu^2\xi\eta = r(\xi)\eta,\ r(\xi) = \frac{1}{2}\frac{\varphi_0'''}{\varphi_0'^3} - \frac{3}{4}\frac{\varphi_0''^2}{\varphi_0'^4}. \tag{7}$$

Under the circumstances envisaged here, φ_0' is bounded away from zero, and r is bounded, so that in the first approximation $r\eta$ can be neglected when ν is large. To obtain an estimate of the error, (7) may be turned into a Volterra integral equation

$$\eta(\xi) = c_1 H_1(\xi) + c_2 H_2(\xi) + \int_{\xi_0}^{\xi} H(\xi, \tau) r(\tau)\eta(\tau)\,d\tau \tag{8}$$

in which H_1 and H_2 are two linearly independent solutions of the differential equation

$$\frac{d^2H}{d\xi^2} - \nu^2\xi H = 0, \tag{9}$$

and

$$H(\xi, \tau) = \frac{H_1(\xi)H_2(\tau) - H_1(\tau)H_2(\xi)}{H_1'(0)H_2(0) - H_1(0)H_2'(0)}. \tag{10}$$

ξ_0 is an arbitrary point of the interval, and η and $c_1H_1 + c_2H_2$ satisfy the same initial conditions at ξ_0.

If the interval is finite, Langer proves from (8) that each η is approximated uniformly by that solution of (9) which has the same value and the same derivative at ξ_0 as η. The point ξ_0 is often chosen to be the transition point $\xi = 0$, but any other point of the interval will do. In the case of an infinite interval uniform approximation will hold in any finite sub-interval; it will hold in the infinite interval if $p(x)$ is subjected to an additional condition regarding its behaviour at infinity.

There are two ways for obtaining higher approximations. The first one of these goes back to Liouville, was used in Langer's earlier papers, it is simple in theory but rather cumbersome in practice: it consists in obtaining successive approximations to the solution of (8). The second one is more difficult to describe but simpler in practice. It is used in Langer's more recent work where comparison functions of the form

$$H_j(\xi) \sum_{n=0}^{m-1} \alpha_n(\xi)\nu^{-n} + H_j'(\xi) \sum_{n=1}^{m} \beta_n(\xi)\nu^{-n} \qquad j = 1, 2 \tag{11}$$

are employed, and the α_n and β_n are determined so as to make the differential equation satisfied by these comparison functions coincide with (7) up to terms of the order of ν^{-m}.

Cherry developed uniform approximations for unbounded regions, with special regard to solutions determined by their asymptotic behaviour at infinity. We shall again restrict ourselves to (5). Instead of (6), Cherry uses a more elaborate transformation,

$$\xi = \varphi(x, \nu) = \sum_{n=0}^{\infty} \varphi_n(x)\nu^{-n}, \quad \eta = \sqrt{\varphi'}\, y, \tag{12}$$

where φ_0 is the function given by (6), while r is determined by

$$r(\xi) = \frac{1}{2}\frac{\varphi'''}{\varphi'^3} - \frac{3}{4}\frac{\varphi''^2}{\varphi'^4} + \nu^2(p\varphi'^{-2} - \varphi). \tag{13}$$

Cherry now shows that under certain assumptions on p, the φ_n can be found so as to make

$$(1 + |\,\xi\,|^{7/2})r(\xi) = O(\nu^{-m})$$

for a given m, uniformly in x. Because r vanishes at infinity sufficiently strongly, $\xi_0 = \infty$ may be taken in (8), and it can then be proved that $c_1H_1(\xi) + c_2H_2(\xi)$ is an mth order approximation to $\eta(\xi)$, and also that these two functions show the same behaviour as $\xi \to \infty$ while ν is fixed (and sufficiently large). The connection with (11) can be exhibited by expanding

$$\sqrt{\varphi'(x, \nu)}\, H_j(\varphi(x, \nu))$$

in descending powers of ν. However, this expansion will in general fail to hold uniformly in the unbounded interval.

4. Having pointed out some of the characteristic features of several approaches to the problem, I shall now attempt to describe in somewhat greater detail the process of approximation, and shall indicate some of the results.

The differential equation to be discussed is

$$\frac{d^2y}{dx^2} - \nu^2 f(x, \nu)y = 0 \tag{14}$$

where ν is the large parameter. This differential equation will be compared with

$$\frac{d^2\eta}{d\xi^2} - \nu^2 F(\xi, \nu)\eta = 0 \tag{15}$$

the solutions of which are supposed to be known. The comparison is based on Liouville's transformation

$$\xi = \varphi(x, \nu), \qquad \eta = \sqrt{\varphi'}\, Y \tag{16}$$

which carries (15) into

$$\frac{d^2Y}{dx^2} - \left[\nu^2 F(\varphi, \nu)\varphi'^2 - \frac{1}{2}\{\varphi, x\}\right] Y = 0 \tag{17}$$

where

$$\{\varphi, x\} = \frac{\varphi'''}{\varphi'} - \frac{3}{2}\left(\frac{\varphi''}{\varphi'}\right)^2 \tag{18}$$

is the Schwarzian derivative of φ. Thus, for any φ, the solutions of (17) are known.

In order to make the *comparison equation* (17) resemble (14) closely, we seek to determine φ so that

$$\nu^2 f(x, \nu) - \nu^2 F(\varphi, \nu)\varphi'^2 + \tfrac{1}{2}\{\varphi, x\} = G(x, \nu) \tag{19}$$

is „small". Now, (14) may be written as

$$\frac{d^2y}{dx^2} - \left[\nu^2 F(\varphi, \nu)\varphi'^2 - \frac{1}{2}\{\varphi, x\}\right] y = Gy, \tag{20}$$

and by the method of variation of parameters this can be converted into a Volterra integral equation

$$y(x) = Y(x) + \int_x^x Y(x, t)G(t, \nu)y(t)dt \tag{21}$$

where

$$Y(x) = c_1 Y_1(x) + c_2 Y_2(x)$$

is some solution, and

$$Y(x, t) = \frac{Y_1(x)Y_2(t) - Y_1(t)Y_2(x)}{Y_1'(x)Y_2(x) - Y_1(x)Y_2'(x)} \tag{22}$$

is the Green's function, of (17).

We can now specify what we mean by G being „small". First let us select a positive function $\gamma(x)$ so that

$$\int_{x_0}^{x} | Y_j(t)Y(t) | \gamma(t)dt \qquad j = 1, 2$$

is bounded for all x, and then try to make

$$G(x, \nu) = O(\nu^{2-m}\gamma(x)) \tag{23}$$

uniformly for all x and all sufficiently large ν. The introduction of $\gamma(x)$ makes it possible to obtain uniform approximations in unbounded regions, and also to obtain approximations for solutions of (14) which are characterized in terms of their behaviour at a singular point or at infinity.

Thus we can distinguish two parts of our problem. First, φ is to be determined so as to satisfy (23), and then, the integral equation (21) must be solved. We shall discuss these two parts separately.

For the discussion of (23) we shall assume that f is of the form

$$f(x, \nu) = \sum_{n=0}^{\infty} f_n(x)\nu^{-n}, \tag{24}$$

the series being convergent for all sufficiently large ν and all x, and the f_n being sufficiently regular functions. F is supposed to have a similar expansion. It is plausible to try a similar expansion also for

$$\varphi(x, \nu) = \sum_{n=0}^{\infty} \varphi_n(x)\nu^{-n}. \tag{25}$$

Clearly, (23) does not determine φ completely, but on expansion in descending powers of ν it is found that $\varphi_n(x)$ is determined by (23) for $n < m$. In particular,

$$F_0(\varphi_0)\varphi_0'^2 = f_0(x). \tag{26}$$

In order that this differential equation should have a non-singular solution it is clearly necessary that the number and multiplicity of the zeros of f_0 and F_0 should be the same. Moreover, if a_k are the zeros of f_0, and α_k those of F_0, it is further necessary that

$$\int_{\alpha_k}^{\alpha_l} \sqrt{F_0(\xi)}d\xi = \int_{a_k}^{a_l} \sqrt{f_0(x)}dx. \tag{27}$$

For instance, if f_0 has no zero, then we may take F as a constant, equal to unity, and from (26),

$$\varphi_0(x) = \int \sqrt{f_0(x)}dx \tag{28}$$

which is Liouville's transformation. Again, if $f_0 = p$ has a single simple zero, $F_0(\xi)$ may be taken as ξ, and from (26) we obtain Langer's transformation (6). If f_0 has two simple zeros, at a_1 and a_2, the most plausible form of F_0 to use is $\xi^2 - \alpha^2$. The condition (27) now determines α^2 from

$$\int_{a_1}^{a_2} \sqrt{f_0(x)}dx = \int_{-\alpha}^{\alpha} \sqrt{\xi^2 - \alpha^2}\,d\xi = \frac{1}{2}\pi\alpha^2. \tag{29}$$

And so on.

Once φ_0 is determined, substitution of (25) in (19), and the condition (23) lead to linear differential equations of the first order for $\varphi_1, \ldots, \varphi_{m-1}$, and determine these functions explicitly. If x is a bounded variable, it is usually permissible to take $\gamma(x) = 1$ and $\varphi_m, \varphi_{m+1}, \ldots = 0$. In case of an unbounded variable x, there are different ways of satisfying (23). Cherry modifies the recurrent differential equations satisfied by the φ_n for $n \geqq m$ by omitting all those terms which arise from the Schwarzian in (19), while Bohnenblust determines φ from the differential equation

$$\nu^2 f(x, \nu) - \nu^2 F(\varphi, \nu)\varphi'^2 + \frac{1}{2}\left\{ \sum_{n=0}^{m-1} \varphi_n(x)\nu^{-n}, x \right\} = 0. \tag{30}$$

In either case certain restrictions must be imposed on the f_n to satisfy (23).

Having determined φ we turn to the integral equation (21). In general this will be a *singular* Volterra integral equation, and the existence and uniqueness of the solution must be established by the method of successive approximations. By estimating the integral in (21), it is then proved that Y represents y asymptotically for large ν. Asymptotic expansions (rather than asymptotic approximations of the mth order) of y may then be obtained by expanding $\varphi'^{-\frac{1}{2}}\eta(\varphi)$ (where φ is given by (25)) in descending powers of ν. However, it may be noted that the region of uniform validity of the asymptotic expansion is frequently smaller than the region of uniform validity of the asymptotic approximation from which it arose, a typical situation being that the asymptotic approximation (for large ν) is valid uniformly in an unbounded region of x, while the asymptotic expansion holds uniformly in every bounded subregion of that region.

5. We shall now attempt to indicate briefly some of the results proved by the comparison technique.

Langer in his earlier papers (1931, 1932) studied the differential equation

$$\frac{d^2y}{dx^2} + p(x)\frac{dy}{dx} - [\nu^2 f_0(x) + f_2(x)]y = 0 \tag{31}$$

in which f_0 was assumed to be of the form $x^\alpha g_0(x)$, where α is a positive number and g_0 is a function without zeros. He considered this equation either on a real interval with g_0 twice continuously differentiable, or else in a simply connected region of the complex plane with g_0 an analytic function, and compared (31) with the differential equation satisfied by Bessel functions of order $\frac{1}{\alpha + 2}$. His results are partly summarised in a symposium lecture (Langer 1934a). Further work (Langer 1934b) concerns the differential equation (14) where f has the form (24), f_0 has a double zero, and parabolic cylinder functions are used as comparison functions, and (Langer 1935) the differential equation

$$\frac{d^2y}{dx^2} - [\nu^2 f_0(x) + \nu f_1(x) + r(x, \nu))y = 0 \tag{33}$$

in which r is allowed to have a pole of order at most two at the transition point. Langer's general results have been applied by himself and by his pupils to a large number of special differential equations arising in mathematics and physics.

More recently Langer (1949) studied the differential equation (14) with f of the form (24) on a real interval, assuming all the f_n to be infinitely differentiable, and f_0 to have a single simple zero. An active group continues to work on these problems under Langer's direction. On the whole, their methods are designed for bounded regions of x, and give asymptotic expansions of solutions determined by initial values at an ordinary point of the differential equation, but under certain conditions on the coefficients the results hold in unbounded regions.

Cherry (1950 and other papers) developed a comparison technique for obtaining uniform asymptotic approximations, in an unbounded region, for solutions determined by their behaviour at infinity. He studied (14) with

$$f(x, \nu) = x + \sum_{r=1}^{\infty} g_r(x)\nu^{-2r} \tag{34}$$

in a region which contains several sectors of the complex x-plane, assuming that each of the g_r is of the form $\alpha_r x + \beta_r + o(1)$ as $x \to \infty$ in any one of the sectors contained in the region, the constants α_r and β_r being possibly different in different sectors. The point x_0 in the integral equation (21) is always a suitable point at infinity. The general results were applied to Bessel functions, and to certain hypergeometric functions occurring in gas-flow theory.

At present Cherry's work is being continued by a group consisting of H. F. Bohnenblust, the present writer, and several younger mathematicians. We study the equation (14) either on an unbounded real interval or in an un-

bounded „sector-region" of the complex x-plane, but our assumptions differ from Cherry's in two respects. Firstly, we do not assume $f(x, \nu)$ to be an analytic function of ν^{-2}, but take it to be of the form

$$f(x, \nu) = x + \sum_{r=1}^{n} g_r(x)\nu^{-2r} + g(x, \nu)\nu^{-2n-2}$$

where $g_1, \ldots, g_n$, and g (the latter for each fixed sufficiently large ν) are analytic functions of x, and secondly, instead of $g_r = \alpha_r x + \beta_r + o(1)$ we merely assume that $g_1, \ldots, g_n$, g (the latter for each sufficiently large ν) are $O(1+|x|)$.

A similar method has also been introduced by Fubini and developed by Tricomi (see, for instance, Tricomi, 1953, section 29) who applied it to the investigation of asymptotic properties of Bessel functions, Laguerre polynomials, etc.

An extension of the foregoing results to differential equations of higher orders would seem to be highly desirable but it appears to meet with formidable difficulties. Beyond the work of Meksyn (1947) and Wasow (1947, 1950, 1952, and more recent, unpublished, work) on differential equations of the fourth order, very little is known about equations of higher orders.

References

The following list contains only the papers referred to in this lecture. A more complete survey of the literature on asymptotic solutions of differential equations with turning points has been prepared at the California Institute of Technology under contract with the Office of Naval Research, U.S.A.

T. M. Cherry, 1950: Uniform asymptotic formulae for functions with transition points. Trans. Amer. Math. Soc. 68, 224—257.

George Green, 1837: On the motion of waves in a variable canal of small depth and width. Trans. Cambridge Phil. Soc. 6, 457—462.

Harold Jeffreys, 1924—25: On certain approximate solutions of linear differential equations of the second order. Proc. London Math. Soc. (2) 23, 428—436.

Harold Jeffreys, 1942: Asymptotic solutions of linear differential equations. Philos. Mag. (7) 33, 451—456.

Harold Jeffreys, 1953: On approximate solutions of linear differential equations. Proc. Cambridge Phil. Soc. 49, 601—611.

R. E. Langer, 1931: On the asymptotic solutions of ordinary differential equations, with an application to the Bessel functions of large order. Trans. Amer. Math. Soc. 33, 23—64.

R. E. Langer, 1932: On the symptotic solutions of differential equations with an application to the Bessel functions of large complex order. Trans. Amer. Math. Soc. 34, 477—480.

R. E. Langer, 1934a: The asymptotic solutions of ordinary linear differential equations of the second order with special reference to the Stokes' phenomenon. Bull. Amer. Math. Soc. 40, 545—582.

R. E. Langer, 1934b: The asymptotic solutions of certain linear ordinary differential equations of the second order. Trans. Amer. Math. Soc. 36, 90—106.

R. E. LANGER, 1935: On the saymptotic solutions of ordinary differential equations, with reference to the Stokes' phenomenon about a singular point. Trans. Amer. Math. Soc. **37**, 397—416.

R. E. LANGER, 1949: The asymptotic solutions of ordinary linear differential equations of the second order, with special reference to a turning point. Trans. Amer. Math. Soc. **67**, 461—490.

JOSEPH LIOUVILLE, 1837: Second mémoire sur le développement des fonctions ou parties de fonctions en séries dont divers termes sont assujettis à satisfaire a une même équation différentielle du second ordre contenant un paramètre variable. J. Math. Pure et Appl. **2**, 16—35.

D. MEKSYN, 1947: Asymptotic integrals of a fourth order differential equation containing a large parameter. Proc. London Math. Soc. (2) **49**, 436—457.

F. G. TRICOMI, 1953: Equazioni differenziali. Einaudi. Torino.

WOLFGANG WASOW, 1947: On the asymptotic solution of the differential equation for small disturbances in a laminar flow. Proc. Nat. Acad. Sci. U.S.A. **33**, 232—234.

WOLFGANG WASOW, 1950: A study of the solutions of the differential equation $y^{(4)} + \lambda^2(xy'' + y) = 0$ for large values of λ. Ann. of Math. (2) **52**, 350—361.

WOLFGANG WASOW, 1952: Asymptotic solution of the differential equation of hydrodynamic stability in a domain containing a transition point. National Bureau of Standards. U.S.A., Report 1620.

THE COEFFICIENTS OF SCHLICHT AND ALLIED FUNCTIONS.

W. K. HAYMAN

Let S be the class of functions

$$w = f(z) = z + a_2 z^2 + \dots$$

regular and schlicht (univalent) in $|z| < 1$, i.e. assuming distinct values w for distinct values of z. The functions $f(z) \in S$ map $|z| < 1$ 1 : 1 conformally onto a simply connected domain D in the w plane. During the last international Congress at Harvard Professors Schaeffer and Spencer [12] reported on the progress, largely due to their own work, on the so-called coefficient problem for S, i.e. the problem of characterising the region V_n in complex space of $n - 1$ dimensions occupied by the points $(a_2, a_3, \dots, a_n)$ for $f(z) \in S$.

To-day I should like to discuss the behaviour of the a_n for large n and the related problem of the behaviour of $f(z)$ near $|z| = 1$. Many of the results extend to more general classes of functions.

It was first shown by Koebe [9], that the distance d of $w = 0$ from the complement of D is greater than an absolute constant, and that for fixed z $|f'(z)|$ and $|f(z)|$ are bounded below and above by positive constants for all $f(z) \in S$. The exact inequalities $|a_2| \leqq 2$, $d \geqq \frac{1}{4}$ and the bounds for $|f'(z)|$, $|f(z)|$ and $|f'(z)/f(z)|$ were obtained in (1916) by Bieberbach [2], Faber, Pick, Gronwall and others. In each case the bounds are attained only for the functions

$$f(z) = \frac{z}{(1 - ze^{i\theta})^2} = \sum_{n=1}^{\infty} n z^n e^{i(n-1)\theta},$$

which we shall therefore call the greatest schlicht functions (g.s.f.). They map $|z| < 1$ onto the w plane cut along a radial slit from $-\frac{1}{4}e^{-i\theta}$ to ∞. These results made plausible the famous Bieberbach conjecture, that $|a_n| \leqq n$ holds for all positive integers n and $f(z) \in S$.

In (1923) Löwner [11] obtained a partial differential equation satisfied by a class of functions dense in S. By means of this he was able to show that $|a_3| \leqq 3$ and also to obtain the exact upper bounds for all the coefficients of the inverse function $z = f^{-1}(w)$, again with equality only for the g.s.f.'s. His method has also led to numerous other results, notably the exact bounds for $\arg(f'(z))$ and $\arg\left\{\frac{f(z)}{z}\right\}$. In these latter cases the extreme functions are not

the g.s.f.'s but map $|z|<1$ onto the plane cut along certain slits. Garabedian and Schiffer have just proved that $|a_4| \leqq 4$.

In (1925) Littlewood [10] proved that

$$I(r, f) = \frac{1}{2\pi} \int_0^{2\pi} |f(re^{i\theta})| \, d\theta < \frac{r}{1-r}, \; 0 < r < 1.$$

Hence taking $r = 1 - 1/n$,

$$|a_n| = \left| \frac{1}{2\pi i} \int_{|z|=r} \frac{f(z)dz}{z^{n+1}} \right| < \frac{1}{r^{n-1}(1-r)} < en.$$

The g.s.f's. have $I(r, f) = \dfrac{r}{1-r^2}$. The sharpest estimates to date are

$$I(r, f) < \frac{r}{1-r^2} + .55 \text{ and hence } |a_n| < \tfrac{1}{2}en + 1.51,$$

due to Bazilevic [1].

Let us consider next the more general class of functions $f(z) = \sum_0^\infty a_n z^n$, p-valent in $|z| < 1$, i.e. such that the equation $f(z) = w$ has at most p roots in $|z| < 1$. In (1935) Cartwright [5] showed that for such functions

$$M(r, f) = \max_{|z|=r} |f(z)| < A(p)\mu_p (1-r)^{-2p}, \; 0 < r < 1, \tag{1}$$

where $\mu_p = \max_{n \leqq p} |a_n|$ and $A(p)$ is a constant depending on p. Hence, using a generalisation of Littlewood's method, Biernacki [3] deduced that

$$I(r, f) < A(p)\mu_p (1-r)^{1-2p}, \; 0 < r < 1, \tag{2}$$

and so that

$$|a_n| < A(p)\mu_p n^{2p-1}, \; n \geqq p. \tag{3}$$

The orders of magnitudes in these results are best possible, as is shown by the p'th powers of the g.s.f's. The results were extended by Spencer [13] to mean p-valent $f(z)$, i.e. such that the equation $f(z) = w$ has on the average at most p roots, when w ranges over the disk $|w| < R$, $0 < R < \infty$. Here p need no longer be an integer. Then (1) remains true for $p > 0$, (2) for $p > \frac{1}{2}$ and (3) for $p > \frac{1}{4}$.

It will be convenient to use a weaker notion of mean valency due to Biernacki [4]. We shall say that a regular functions $f(z)$ is mean p-valent in a domain Δ, if the average number, $p(R)$ of roots of the equation $f(z) = w$, when w ranges over the circumference $|w| = R$ always satisfies $p(R) \leqq p$. Here $2\pi R p(R)$ is the total length of the arcs in the Riemann surface of $w = f(z)$, which lie over the circle $|w| = R$. Alternatively

$$p(R) = \frac{1}{2\pi}\int_0^{2\pi} n(Re^{i\varphi})d\varphi$$

where $n(Re^{i\varphi})$ is the number of roots of the equation $f(z) = Re^{i\varphi}$ in Δ.

Further we shall consider functions $f(z)$ analytically continuable with one-valued modulus throughout an annulus $r_0 < |z| < 1$. If $f_2(z)$ is the branch obtained from $f_1(z)$ by continuing once around the annulus, then $|f_2(z)/f_1(z)| \equiv 1$, and so by the maximum modulus theorem $f_2(z) = f_1(z)e^{i\mu}$, where μ is a real constant. Thus $f(z)/z^\mu$ remains one-valued in the annulus and possesses a Laurent expansion

$$f(z) = z^\mu \sum_{-\infty}^{+\infty} a_n z^n, \quad r_0 < |z| < 1. \tag{4}$$

We define $p(R)$ as above for a fixed branch of $f(z)$ in the annulus cut along a radius. Since replacing $f(z)$ by a different branch gives rise to a rotation in the w plane, $p(R)$ is independent of the radial cut and the particular branch taken, so that our definition of mean p-valency is meaningful.

The following remarks may justify these somewhat artificial looking hypotheses. In the first instance if $f(z_0) = 0$, then $f(z)$ takes all sufficiently small values w near z_0. Hence if $f(z)$ has q zeros in the annulus, then $p(R) \geqq q$ for all small R and so $q \leqq p$. Thus $f(z)$ has only a finite number of zeros in the annulus $r_0 < |z| < 1$, and so we can find a smaller annulus $r_1 < |z| < 1$, in which $f(z)$ is still mean p-valent and $f(z) \neq 0$. Next, if ν is any positive number, the transformation $W = w^\nu$ maps arcs of angular length α on $|w| = R$ onto arcs of angular length $\nu\alpha$ on $|w| = R^\nu$. Thus if $f(z)$ is mean p-valent and not zero in $r_1 < |z| < 1$, then $[f(z)]^\nu$ is mean νp-valent. For $[f(z)]^\nu$ can be analytically continued throughout $r_1 < |z| < 1$, but need not be one-valued. Thus our definitions enable us to discuss positive powers of $f(z)$ together with $f(z)$. This is very useful in view of the following [1]

Regularity Theorem. Suppose $f(z)$ is mean p-valent in $r_0 < |z| < 1$. Then the limit

$$\alpha = \lim_{r\to 1} (1-r)^{2p} M(r, f) \tag{5}$$

exists finitely. Further if $p > \frac{1}{2}$

$$\lim_{r\to 1} (1-r)^{2p-1} I(r, f) = \frac{\alpha\Gamma(p-\frac{1}{2})}{2\Gamma(\frac{1}{2})\Gamma(p)}, \tag{6}$$

and if $p > \frac{1}{4}$

$$\lim_{n\to+\infty} \frac{|a_n|}{n^{2p-1}} = \frac{\alpha}{\Gamma(2p)}. \tag{7}$$

Let us call α the *limit constant* of $f(z)$.

One of the main advantages of this theorem lies in the close link between the maximum modulus $M(r, f)$, which is usually not too difficult to deal with and the means $I(r, f)$ and coefficients a_n, which are frequently less tractable. Thus by our previous remark $[f(z)]^\nu$ is mean (νp)-valent in the zerofree annulus $r_1 < |z| < 1$ of $f(z)$ and by (5) $[f(z)]^\nu$ clearly has limit constant α^ν. Thus we obtain at once information about the means and coefficients of $[f(z)]^\nu$ from our theorem, by replacing α by α^ν and p by $p\nu$.

Suppose for instance that $f(z) \in S$, $p = 1$. Then we have

$$\left|\frac{f'(re^{i\theta})}{f(re^{i\theta})}\right| \leqq \frac{1+r}{r(1-r)} = \frac{d}{dr}\log\frac{r}{(1-r)^2},$$

and it follows that except for the g.s.f's $(1-r)^2 r^{-1} M(r, f)$ decreases strictly with increasing r, $0 < r < 1$. These results remain true if $f(z)$ is merely mean 1-valent in $|z| < 1$ [6]. Hence if $f_1(z) = z + a_2 z^2 + \ldots$ is mean 1-valent in $|z| < 1$

$$\lim_{n\to\infty} \frac{|a_n|}{n} = \alpha \leqq 1$$

and $\alpha = 1$ only for the g.s.f's. In particular the Bieberbach conjecture holds for all sufficiently large n and any fixed $f(z) \in S$.

More generally let

$$f_p(z) = z^p(1 + b_1 z + b_2 z^2 + \ldots.)$$

be mean p-valent in $0 < |z| < 1$, where $p > \frac{1}{4}$. Then $f_p(z) \neq 0$ in $0 < |z| < 1$ since the immediate neighbourhood of $z = 0$ contributes p to $p(R)$ for small positive R. Thus

$$f_1(z) = [f_p(z)]^{1/p} = z\left(1 + \frac{b_1}{p} z + \ldots\right)$$

is mean 1-valent in $0 < |z| < 1$, and so in $|z| < 1$, since $f_1(z)$ is evidently singlevalued. Thus if α is the limit constant of $f_1(z)$, the limit constant of $f_p(z)$ is $\alpha^p \leqq 1$, and we have

$$\lim_{n\to\infty} \frac{|b_n|}{n^{2p-1}} = \frac{\alpha^p}{\Gamma(2p)} \leqq \frac{1}{\Gamma(2p)},$$

with equality only for the p'th powers of the g.s.f's. An interesting special case occurs, when p is a positive integer and $f_p(z)$ is p-valent in $|z| < 1$.

More generally suppose that p, k are positive integers, $1 \leqq k < 4p$, and that

$$f_{p,k}(z) = z^p(1 + b_1 z^k + b_2 z^{2k} + \ldots)$$

is mean p-valent in $|z|<1$. The transformation $z^k = Z$ maps the annulus $0<|z|<1$ described once, onto $0<|Z|<1$ described k times. Hence

$$f_{p,k}(Z^{1/k}) = Z^{p/k}(1 + b_1 Z + b_2 Z^2 + \ldots)$$

is mean (p/k)-valent in $0<|Z|<1$ and so

$$\lim_{n\to\infty} \frac{|b_n|}{n^{2p/k-1}} = \frac{\alpha^{p/k}}{\Gamma(2p/k)} \leqq \frac{1}{\Gamma(2p/k)}.$$

Equality holds in this case only if $f_{p,k}(z) = z^p(1 - z^k e^{i\theta})^{-2p/k}$.

If $p = 1$ and $f_{1,k}(z) \in S$ we may go further. In this case the function

$$f_1(Z) = f_{1,k}(Z^{1/k}) = Z(1 + kb_1 Z + \ldots)$$

also belongs to S, and so $|kb_1| < 2$, except when $f_1(Z) = Z(1 - Ze^{i\theta})^{-2}$. $f_{1,k}(z) = z(1-z^k e^{i\theta})^{-2/k}$. If $|kb_1| = c$, where $0 \leqq c < 2$, it follows from a recent result of Jenkins [8] that the limit constant α of $f_1(Z)$ satisfies

$$\alpha \leqq \alpha(c) = 4[2 - (2-c)^{\frac{1}{2}}]^{-2} \exp\{2 - 4/[2 - (2-c)^{\frac{1}{2}}]\},$$

and that this inequality is sharp for every c, $0 \leqq c < 2$. Hence if $k = 1$, 2 or 3 and

$$f_{1,k}(z) = z + a_{k+1}z^{k+1} + a_{2k+1}z^{2k+1} + \ldots$$

belongs to S and $|a_{k+1}| = c/k$, then we have

$$\lim_{n\to\infty} \frac{|a_{nk+1}|}{n^{2/k-1}} \leqq \frac{[\alpha(c)]^{1/k}}{\Gamma(2/k)},$$

and this is sharp for every c, with $0 \leqq c \leqq 2$.

Nevertheless if $k = 2$, the upper bound of $|a_5|$ for varying odd $f(z) \in S$ is $e^{-\frac{2}{3}} + \frac{1}{2} = 1.01\ldots$ and $|a_n| > 1$ is possible for every fixed odd n greater than three and a corresponding function $f(z)$.

I should like to conclude by indicating how one may prove the Regularity theorem [1]. We may suppose that $f(z)$ is mean p-valent and not zero in an annulus $1 - 2\delta < |z| < 1$. One can then prove the basic inequality

$$\left|\frac{f'(re^{i\theta})}{f(re^{i\theta})}\right| \leqq \frac{4p\delta}{(1-r)(r+2\delta-1)}, \quad 1-2\delta < r < 1, \quad |\theta| \leqq \pi.$$

If we write $f(re^{i\theta}) = Re^{i\lambda}$, this becomes for θ fixed

$$\left|\frac{d}{dr}\log R + i\lambda'(r)\right| \leqq \frac{4p\delta}{(1-r)(r+2\delta-1)} = 2p\frac{d}{dr}\log\left(\frac{r+2\delta-1}{1-r}\right). \quad (8)$$

Hence $\left(\frac{1-r}{r+2\delta-1}\right)^{2p} R$ decreases with increasing r and a fixed θ.

Suppose now that

$$\alpha = \overline{\lim_{r\to 1}} (1-r)^{2p} M(r, f) > 0 \tag{9}$$

(If $\alpha = 0$, our theorem is easily proved by classical methods.) We can then find $z_n = r_n e^{i\theta_n}$, with $r_n \to 1$, $|\theta_n| \leqq \pi$, such that

$$\beta_n = \left(\frac{1-r_n}{r_n + 2\delta - 1}\right)^{2p} |f(r_n e^{i\theta_n})| \to \frac{\alpha}{(2\delta)^{2p}}.$$

It follows from the above monotonicity property that

$$\left(\frac{1-r}{r+2\delta-1}\right)^{2p} |f(re^{i\theta_n})| \geqq \beta_n, \quad 1-2\delta < r < r_n.$$

Hence if θ_0 is a limit point of the sequence θ_n we deduce

$$\left(\frac{1-r}{r+2\delta-1}\right)^{2p} |f(re^{i\theta_0})| \geqq \frac{\alpha}{(2\delta)^{2p}}, \quad 1-2\delta < r < 1,$$

In view of (9) this implies

$$M(r, f) \sim |f(re^{i\theta_0})| \sim \frac{\alpha}{(1-r)^{2p}}, \text{ as } r \to 1, \tag{10}$$

and this gives (5).

Next one can deduce from (8) that

$$(1-r)[\lambda'(r)]^2 \leqq 8p \frac{d}{dr} \log\left[\frac{(r+2\delta-1)^{2p}}{(1-r)^{2p} R}\right].$$

On taking $\theta = \theta_0$ and integrating from $1-\delta$ to 1 we deduce

$$\int_{1-\delta}^{1} (1-r)[\lambda'(r)]^2 dr \leqq \log\left\{\frac{(2\delta)^{2p} |f[(1-\delta)e^{i\theta_0}]|}{\alpha}\right\} < \infty$$

in view of (10). If r_1 is so near 1 that $\int_{r_1}^{1} (1-r)[\lambda'(r)]^2 dr < \varepsilon$ and $r_1 < r_2 < 1$ it follows that

$$|\lambda(r_2) - \lambda(r_1)| \leqq \int_{r_1}^{r_2} |\lambda'(r) dr \leqq \left\{\int_{r_1}^{r_2} (1-r)[\lambda'(r)]^2 dr \int_{r_1}^{r_2} \frac{dr}{1-r}\right\}^{\frac{1}{2}}$$

$$\leqq \varepsilon^{\frac{1}{2}} \left\{\log \frac{1-r_1}{1-r_2}\right\}^{\frac{1}{2}}$$

From this and (10) it follows that if $r_1, r_2 \to 1$ so that $(1-r_1)/(1-r_2)$ remains bounded, $r_1 < r_2 < 1$, then

$$f(r_2 e^{i\theta_0}) \sim f(r_1 e^{i\theta_0}) \left(\frac{1-r_1}{1-r_2}\right)^{2p}.$$

By a "Normal Families" argument one can then deduce that

$$f(z) \sim f(re^{i\theta_0}) \left(\frac{1-r}{1-ze^{-i\theta_0}} \right)^{2p}$$

as $r \to 1$, uniformly while z remains in a region

$$\Delta(r, \varepsilon): \varepsilon(1-r) \leqq \left| 1 - ze^{-i\theta_0} \right| \leqq \frac{1-r}{\varepsilon}, \quad \left| \arg\left(1 - ze^{-i\theta_0} \right) \right| \leqq \frac{\pi}{2} - \varepsilon$$

for a fixed ε, and that the area of the image by $f(z)$ of the part of the annulus $1 - 2\delta < |z| < 1$ which lies outside $\Delta(r, \varepsilon)$ is relatively small. In fact the area over the w plane of the image of that part of $\Delta(r, \varepsilon)$ by $f(z)$, where $|f(z)| < R = M(r, f)$, will already be nearly $p\pi R^2$, the maximum permissible amount for mean p-valent functions. From these facts (6) and (7) can be deduced. In fact we obtain the asymptotic formula

$$na_n \sim \frac{f\left[\left(1 - \frac{1}{n} \right) e^{i\theta_0} \right] e^{-i(n+\mu)\theta_0}}{\Gamma(2p)}, \text{ as } n \to +\infty,$$

and in view of (10), we have (7).

References

[1] I. E. Bazilevic, 'On distortion theorems and coefficients of univalent functions', (Russian) Math. Sbornik N.S. 28 (70) (1951) 147—164.

[2] L. Bieberbach, 'Über die Koeffizienten derjenigen Potenzreihen, welche eine schlichte Abbildung des Einheitskreises vermitteln', Berliner Berichte (1916), pp. 940—955.

[3] M. Biernacki, 'Sur les fonctions multivalentes d'ordre p', Comptes rendus 203 (1936), 499—451.

[4] M. Biernacki, 'Sur les fonctions en moyenne multivalentes', Bull. Sci. Math. (2), 70 (1946), 51—76.

[5] M. L. Cartwright, 'Inequalities in the theory of functions', Math. Annalen, 111 (1935), 98—118.

[6] W. K. Hayman, 'Some applications of the transfinite diameter to the theory of functions', J. d'Analyse Math. I (1951), 155—179.

[7] W. K. Hayman, 'The asymptotic behaviour of p-valent functions', Proc. Lond. Math. Soc. (3) 5 (1955), 251—254.

[8] J. A. Jenkins, 'On a problem of Gronwall', Ann. of Math. 59 (1954) ,490—504.

[9] P. Koebe, 'Über die Uniformisierung der algebraischen Kurven II', Math. Ann. 69 (1909), 1—81.

[10] J. E. Littlewood, 'On inequalities in the theory of functions', Proc. Lond. Math. Soc. (2) 23 (1925), 481—519.

[11] K. Löwner, 'Untersuchungen über schlichte konforme Abbildungen des Einheitskreises I.' Math. Ann. 89 (1923), 103—121.

[12] A. C. Schaeffer and D. C. Spencer, 'Coefficient regions for schlicht functions.' Proc. Int. Cong. Math., Harvard, 1950. II, 224—232.

[13] D. C. Spencer, 'On finitely mean valent functions. II', Trans. Amer. Math. Soc. 48 (1940), 418—435.

SOME ASPECT OF CAUCHY'S PROBLEM [1]

EINAR HILLE

1. *Introduction.*

The subject matter of this paper is closely related to that of Professor K. Yosida's hour addres to which I refer for back ground, further problems, and literature. Professor Yosida having kindly placed a copy of his manuscript at my disposal, I shall try to avoid topics which are covered in his addres, but naturally semi-group theory will crop up steadily in my talk and I shall have to discuss the one dimensional case of the temporally homogeneous diffusion equations.

2. *The abstract Cauchy problem.*

In a number of publications (see [7] — [12] in the References at the end of this paper) the author has proposed and discussed an abstract Cauchy problem which may be formulated as follows:

ACP^1. Given a complex (B)-space X and a linear operator U with domain $D[U]$ and range $R[U]$ in X and given an element $y_0 \in X$, find a function $y(t) = y(t; y_0)$ such that

(i) *$y(t)$ is strongly absolutely continuous and continuously differentiable in each finite subinterval of $[0, \infty)$ [or $(0, \infty)$];*

(ii) *for each $t > 0$, $y(t) \in D[U]$ and*

$$U[y(t)] = y'(t); \tag{2.1}$$

$$\text{(iii)} \quad \lim_{t \to 0+} ||y(t; y_0) - y_0|| = 0.$$

This is the formulation given by R. S. Phillips [14] and differs somewhat from earlier versions of mine in the phrasing of condition (i), but it has certain advantages, in particular, it leads to a much stronger form of Theorem 4 below than I was able to prove. The reader will note that there are two alternatives in condition (i); the corresponding problems will be denoted by ACP_1^1 and ACP_2^1 respectively. The symbol ACP^n will be used in section 8 in referring to corresponding problems for the equation

$$U^n[y(t)] = y^{(n)}(t). \tag{2.2}$$

[1]) This paper was written under the sponsorship of the Air Research and Development Command [Contract AF 18 (600)—1127]. Credit should also be given to the National Science Foundation whose generous support made it possible for the author to deliver the paper in person.

These problems are evidently suggested by the classical Cauchy problem especially for diffusion equations, where U is a second order differential operator, and so far the most important applications have been to the integration problem of these equations. But in principle the considerations apply to much wider classes of differential and functional equations (see for instance Joanne Elliott [1]). In these applications the space X is one of the common function spaces of analysis, C or L_p, with the usual metric. R. S. Phillips (personal communication) has pointed out that in such cases numerically-valued representations can be found for the abstract valued function $y(t)$ so that an abstract solution becomes a solution of the differential equation in the ordinary sense.

3. *Uniqueness and non-uniqueness.*

If for a given y_0 and a given operator U the corresponding ACP^1 has two distinct solutions their difference is a *nul solution* $y(t; 0)$. The existence of such solutions is regulated by the two following theorems (see E. Hille [11, pp. 95—97]) where a solution is said to be of *normal type* ω if

$$\limsup_{t\to+\infty} t^{-1} \log ||y(t)|| = \omega < \infty. \tag{3.1}$$

Theorem 1. *For each $y_0 \in X$ the ACP^1 has at most one solution of normal type if U is a closed operator whose point spectrum is not dense in any right halfplane.*

Theorem 2. *If U is closed a necessary and sufficient condition that the ACP^1 shall have a nul solution of normal type $\leqq \omega$ is that the characteristic equation*

$$U[x(\lambda)] = \lambda\, x(\lambda) \tag{3.2}$$

have a solution $x(\lambda) \not\equiv 0$, bounded and holomorphic in each halfplane $\Re(\lambda) \geqq \omega + \varepsilon$, $\varepsilon > 0$.

Since a solution of (3.2) may be multiplied by any numerically-valued bounded holomorphic function, we may assume that $(1 + |\nu|)^3\, ||x(\omega + \varepsilon + i\nu)|| \in L(-\infty, \infty)$. The desired nul solution is then given by

$$y(t) = \frac{1}{2\pi i} \int_{\gamma-i\infty}^{\gamma+i\infty} e^{\lambda t} x(\lambda)\, d\lambda,\ \gamma > \omega. \tag{3.3}$$

If the point spectrum of U covers the whole complex plane, "*explosive*" solutions may exist. Thus if X is an L-space, if $x(\lambda)$ is an entire function of order < 1, positive for $\lambda < 0$, while $||\, x(\lambda)|| \notin L(0, \infty)$, then to every real t_0 there exist solutions of (2.1) which are entire functions of $(t - t_0)^{-1}$ but belong to X for every $t \neq t_0$. Here $||\, y(t)\, || \to +\infty$ as $t \to t_0 -$ but may very well tend to zero as $t \to t_0 +$. Cf. [6] and [11, pp. 97—99].

4. *Applications to the heat equation.* Theorem 2 may be used to construct nul solutions of the ordinary heat equation of a nature different from those found by A. Tychonoff. Here (2.1) and (3.2) take the form

$$\frac{\partial^2 y}{\partial s^2} = \frac{\partial y}{\partial t}\ ,\quad \frac{d^2x}{ds^2} = \lambda x \tag{4.1}$$

respectively. We consider the interval $(-\infty, \infty)$ and the L-space of measurable function $f(s)$ with finite norm defined by

$$\|f\|_\varrho = \int_{-\infty}^{\infty} |f(s)| \exp(-|s|^\varrho)\, ds \tag{4.2}$$

where $\varrho \geqq 0$. For $0 \leqq \varrho \leqq 2$ one finds that the characteristic equation has no solution which is holomorphic and bounded in a right half-plane but for $\varrho > 2$ suitable multipliers may be found and nul solutions are constructed with the aid of (3.3). For $\varrho > 2$ we have also explosive solutions belonging to the space, an example being given by the classical integral

$$\frac{1}{\sqrt{\pi}} \int_0^\infty e^{\lambda(t_0-t)} ch(\sqrt{\lambda} s)\, d\lambda = \frac{1}{t_0 - t} \sum_{n=0}^{\infty} \frac{1}{\Gamma(n + \frac{1}{2})} \left[\frac{s^2}{4(t_0-t)}\right]^n.$$

5. *Relations to semi-group theory.*

Let $\{S(t);\ 0 < t\}$ be a semi-group of linear bounded transformations on X to itself so that $S(t_1 + t_2) = S(t_1)\, S(t_2)$ for $t_1, t_2 > 0$. We assume $S(t)$ to be continuous in the strong operator topology for $t > 0$. In this case $S(t)$ is of normal type in the sense of (3.1). A semi-group is said to be of class (C_0) if

$$\lim_{t \to 0+} \| S(t)x - x \| = 0 \tag{5.1}$$

for each $x \in X$. $S(t)$ is of class $(0, A)$ in the sense of R. S. Phillips if $\int_0^1 \| S(t)x \| dt < \infty$ for each $x \in X$ and if the linear bounded operator

$$R(\lambda)x \equiv \int_0^\infty e^{-\lambda t}\, S(t)x\, dt, \tag{5.2}$$

defined for $\Re(\lambda) > \omega$, has the property

$$\lim_{t \to +\infty} \| \lambda\, R(\lambda)x - x \| = 0 \tag{5.3}$$

for each x. For semi-groups of both classes

$$\lim_{t \to 0+} t^{-1}[S(t)x - x] \equiv A_0 x \tag{5.4}$$

exists for x in a dense set $D[A_0]$. A_0 is called the *infinitesimal operator* of $S(t)$; A_0 is *closed* if $S(t)$ is of class (C_0), otherwise A_0 has a smallest closed extension A called the *infinitesimal generator of* $S(t)$. In either case the resolvent $R(\lambda; A)$ exists and $R(\lambda; A) = R(\lambda)$ for $\Re(\lambda) > \omega$. If $S(t)$ is of class $(0, A)$ relation (5.1) holds for $x \in D[A]$. Further

$$\frac{d}{dt}[S(t)x] = A\, S(t)\, x = S(t)\, A\, x, \qquad x \in D[A]. \tag{5.5}$$

From these relations one obtains, see R. S. Phillips [14]:

Theorem 3. *If U is the infinitesimal generator of a semi-group $\{S(t)\}$ of*

class $(0, A)$, *then the corresponding* ACP_2^1 *has a unique solution* $y(t; y_0) \equiv S(t)[y_0]$ *for each* $y_0 \in D[U]$.

Phillips has also proved the converse proposition:

Theorem 4. *Let* U *be a closed operator with dense domain and non-vacuous resolvent set* [*and* $\lambda||R(\lambda; U)|| = O(1)$ *as* $\lambda \to +\infty$]. *Suppose that for each* $y_0 \in D[U]$ *there is a unique solution to* ACP_1^1 [*to* ACP_2^1]. *Then* U *generates a semi-group* $\{S(t)\}$ *of class* (C_0) [*of class* $(0, A)$] *and* $y(t; y_0) = S(t)[y_0]$ *for all* $y_0 \in D[U]$.

For the case in which the ACP^1 corresponds to a genuine physical problem. Theorem 4 gives the sharpest known formulation of the implications of what J. Hadamard once called the *major premise of Huygens' principle.*

The two theorems show that solving the ACP^1 for the operator U and finding if U generates a semi-group are practically equivalent problems. In view of this situation there is evidently an urgent call for criteria enabling one to recognize whether or not a given operator U is the infinitesimal generator of a semi-group $\{S(t)\}$. For the investigations of Hille, Yosida, and Phillips on this question we refer to Yosida's address [16]. It is evident that U must be closed, its domain dense, and its resolvent $R(\lambda; U)$ must exist in a half-plane. Formula (5.2) shows that a necessary condition is that $R(\lambda; U)x$ be the Laplace transform of $S(t)x$, where $S(t)$ is strongly continuous for $t > 0$, is of normal type, and satisfies (5.1) for $x \in D[U]$. Conversely, if this condition is satisfied one verifies easily that $S(t) = S(t; U)$ is the semi-group generated by U and $S(t)$ is of class $(0, A)$. This fact serves to vindicate the time-honored device of applying the Laplace transformation to the characteristic equation.

It may very well happen, however, that a given operator U cannot generate a semi-group of class $(0, A)$. Before giving up this mode of attack as useless, one had better look into the reason for failure. A common reason for failure is that the spectrum of U covers the right half-plane. If this is the case, some restriction U_0 of U may still have all the desirable properties and the theory should be built around U_0 instead. This simple observation underlies the work of W. Feller in [2] of which more below.

6. *Diffusion equations with one degree of freedom.*

The first partial differential equations with variable coefficients to which the newly created theory of semi-groups was applied was the Fokker-Planck equation of one degree of freedom, studied by K. Yosida in [15]. Reading his work in manuscript led me to the investigations [4, 5, 11] on diffusion equations as well as to the work on the abstract Cauchy problem. Indirectly it also started Feller on his basic work in [2]. Yosida proceeded to develop the theory of diffusion equations with more degrees of freedom and for these investigations the reader is referred to his address. My results and those of Feller will be briefly summarized below.

Following the notation of [11] we write the formally adjoint diffusion equations

$$C[S] \equiv b(\xi)\frac{\partial^2 S}{\partial \xi^2} + a(\xi)\frac{\partial S}{\partial \xi} = \frac{\partial S}{\partial t}, \tag{6.1}$$

$$L[T] \equiv \frac{\partial}{\partial x}\left\{\frac{\partial}{\partial x}[b(x)T] - a(x)T\right\} = \frac{\partial T}{\partial t}, \tag{6.2}$$

where $a(x)$, $b(x)$ are continuous in (α, β) and $b(x) > 0$. We assume $-\infty \leqq \alpha < 0 < \beta \leqq +\infty$. Feller's notation is different, in particular, $a(x)$ and $b(x)$ are interchanged, and his equations involve terms in $c(\xi)S$ and $c(x)T$. We introduce the following auxiliary functions

$$W(x) = \exp\left\{-\int_0^x \frac{a(s)}{b(s)}\,ds\right\},\quad q(x) = \int_0^x [b(s)\,W(s)]^{-1}\,ds, \tag{6.3}$$

$$W_1(x) = \int_0^x W(\xi)\,d\xi,\quad W_2(x) = \int_0^x W_1(s)\,dq(s),\quad \Omega(x) = \int_0^x [W_1(x) - W_1(s)]dq(s).$$

We are concerned with the Cauchy problems for the operator C in $C(\alpha, \beta]$ and the operator L in $L(\alpha, \beta)$. The corresponding characteristic equations

$$C[u] - \lambda u = 0,\quad L[y] - \lambda y = 0 \tag{6.4}$$

are ordinary differential equations and for the decision problem it suffices to take λ positive. It turns out that C will generate a semi-group in $C[\alpha, \beta]$ if and only if the first equation under (6.4) has no solution in $C[\alpha, \beta]$ and similarly for L. The resulting semi-groups are made up of positive contraction operators. In terms of the functions (6.3) we can state the result as

Theorem 5. *The Cauchy problem for* (6.1) *in* $C[\alpha, \beta]$ *has a unique solution for every* $y_0 \in D[C]$ *if and only if* $\Omega(\alpha) = \Omega(\beta) = +\infty$. *The Cauchy problem for* (6.2) *in* $L[\alpha, \beta]$ *has a unique solution for every* $y \in D[L]$ *if and only if* $W_2(\alpha) = W_2(\beta) = +\infty$. *The solution of* (6.2) *is defined by a transition operator if and only if both conditions are satisfied.*

Equivalent results have been found by Feller, but the latter's main contribution to this problem is the study of the cases in which C or L or both fail to generate semi-groups. This situation arises when the point spectrum of the operator in question covers the whole complex plane. It then becomes necessary to restrict the operator so that its point spectrum releases the positive real axis. This is done by imposing *lateral conditions* and Feller set himself the task of determining all lateral conditions leading to semi-groups of positive contraction operators. This he based upon a classification of the boundaries into four types. In our notation the point β is

a *regular* boundary if $\Omega(\beta) < \infty$, $W_2(\beta) < \infty$,
an *exit* boundary if $\Omega(\beta) < \infty$, $W_2(\beta) = \infty$,
an *entrance* boundary if $\Omega(\beta) = \infty$, $W_2(\beta) < \infty$,
a *natural* boundary if $\Omega(\beta) = \infty$, $W_2(\beta) = \infty$.

Lateral conditions have to be prescribed if at least one boundary is not natural. If both boundaries are regular, the lateral conditions are immediate generalizations of the classical boundary conditions, but if exit or entrance boundaries occur great complications are possible. Thus for an exit boundary at β and a natural boundary at α, the solution $S(t; x)$ of (6.1) has to satisfy a condition of the form

$$p_2 S(t; \beta) = p_1 S(t; \alpha) + \tau \int_\alpha^\beta S(t; x)\, dp(x) - \sigma \lim_{x \to \beta} \left[b(x) \frac{\partial^2 S}{\partial x^2} + a(x) \frac{\partial S}{\partial x} \right]$$

where p_1, p_2, σ, τ are given constants and $p(x)$ a given function of bounded variation, all subject to certain conditions. In this case the true adjoint of (6.1) is no longer (6.2) but a more general type of functional equation.

In spite of the formidable nature of the lateral conditions, the solutions of (6.1) and (6.2) appear to have simpler properties when the conditions of Theorem 5 are not satisfied than when they are. For the case of two entrance boundaries Hille [11] has carried through a detailed analysis of the unique solution $S(t)[g]$ of (6.1). Here C has a pure point spectrum $\{\lambda_n\}$, $\lambda_0 = 0$, $\lambda_n < 0$, with corresponding characteristic functions $\omega_n(\xi)$ in terms of which the solution may be represented by the "theta series"

$$S(t)[g] = \sum_0^\infty g_n \omega_n(\xi)\, e^{\lambda_n t}, \quad g(\xi) \sim \sum_0^\infty g_n \omega_n(\xi). \tag{6.5}$$

Here $g(\xi)$ is any element of $C[\alpha, \beta]$ and $S(t)$ is analytic in $\Re(t) > 0$.

The reason for the relative simplicity of this case appears to be that the solutions of $C[u] = \lambda u$ tending to $+1$ as $\xi \to \alpha$ and β respectively, which are used in constructing the resolvent, are entire functions of λ and the characteristic values λ_n are simply the zeros of their Wronskian at $\xi = 0$, again an entire function of λ, usually of order $\frac{1}{2}$. If $a(x) \equiv 0$, $\alpha = -\infty$, $\beta = +\infty$, this case will arise if and only if

$$x[b(x)]^{-1} \in L(-\infty, \infty). \tag{6.6}$$

Cases of two regular boundaries with generalized classical boundary conditions have also been examined by Hille [11] with similar results. The case of exit boundaries has been considered by Joanne Elliot [see Trans. Amer. Math. Soc., 78 (1955) 406—425; added in proof]. Here again theta series arise.

7. *Equations of higher order with coincident characteristics.* The attack on partial differential equations of higher order than the second has not proceeded very far. Nevertheless the Cauchy problem for the equations

$$L_{2n}[T] \equiv \frac{\partial^n}{\partial x^n} \left\{ b(x) \frac{\partial^n T}{\partial x^n} \right\} = (-1)^{n-1} \frac{\partial T}{\partial t}, \quad b(x) > 0, \tag{7.1}$$

in $L(-\infty, \infty)$ seems to be within grasp, at least for the case

$$x^{2n-1}[b(x)]^{-1} \in L(-\infty, \infty) \tag{7.2}$$

which is the analogue of (6.6). Here again the n solutions of $L_{2n}[y] + (-1)^n \lambda y = 0$ which are in $L(-\infty, 0)$ and the n solutions in $L(0, \infty)$ used in constructing the resolvent, are entire functions of λ, apparently of order $1/(2n)$. The Wronskian at the origin is also an entire function of low order and all its zeros are real and negative. One would expect L_{2n} to have a pure point spectrum and the solutions of (7.1) to be given by theta series analogous to (6.5). In this discussion one has to use the known results for self-adjoint operators in $L_2(-\infty, \infty)$ (see K. Kodaira [13] and I. M. Glazman [3]) which apply since the resolvents involve the same solutions in both spaces.

For the case $n = 2$ some primary results have been found by J. L. Howell [unpublished Yale dissertation]. Among other things he proved, without imposing integrability conditions on $b(x)$, that the equation $L_4[y] + \lambda y = 0$ has no solutions in $L(-\infty, \infty)$ or $L_2(-\infty, \infty)$.

8. *Cauchy problems of higher order.* The problem formulated in section 2 may be generalized as follows:

ACP^n. Given a complex (B)-space X and a linear operator U with domain $D[U]$ and range $R[U]$ in X and given n elements $y_0, y_1, \ldots, y_{n-1}$ in X, find a function $y(t) = y(t; y_0, y_1, \ldots, y_{n-1})$ such that

(i) *$y(t)$ is strongly absolutely continuous and n times continuously differentiable in each finite subinterval of $[0, \infty)$;*

(ii) *for each $t > 0$, $y(t) \in D[U^n]$ and*

$$U^n[y(t)] = y^{(n)}(t); \tag{8.1}$$

(iii) $\lim_{t \to 0+} ||y^{(k)}(t; y_0, y_1, \ldots, y_{n-1}) - y_k|| = 0, \quad k = 0, 1, \ldots, n-1.$

For the theory of this problem see [8, 9]. There are uniqueness theorems and theorems on nul solutions analogous to Theorems 1 and 2 above. For $n = 2$ there is an interesting analogue of Theorem 3:

Theorem 6. *If $n = 2$ and U generates a group $\{T(t; U); -\infty < t < \infty\}$, strongly continuous at the origin with $T(0; U) = I$, then the corresponding ACP^2 has a unique solution for each $y_0 \in D[U^2]$ and $y_1 \in D[U] \cap R[U]$, namely*

$$y(t) = \tfrac{1}{2}[T(t; U)(y_0 + z_1) + T(-t; U)(y_0 - z_1)], \quad Uz_1 = y_1. \tag{8.1}$$

This formula reduces to the solution of the equation for the vibrating string when $U = d/ds$.

For $n > 2$ the ACP^n does not seem to be solvable for arbitrary initial values. If U generates a semi-group $T(t; U)$, holomorphic in a sector having its vertex at the origin and if $T(t; U)$ is strongly continuous on the boundary, then one can satisfy $m \leqq [\frac{1}{2} n] + 1$ of the conditions (iii) if the closed sector contains m n^{th} roots of unity, but the remaining $n - m$ conditions are no longer at our disposal.

References

[1] Joanne Elliott, The boundary value problems and semi-groups associated with certain integro-differential equations. Trans. Amer. Math. Soc., 76 (1954) 300–331.

[2] W. FELLER, The parabolic differential equations and the associated semi-groups of transformations. Ann. of Math., (2) 55 (1952) 468—519.

[3] I. M. GLAZMAN, On the theory of singular differential equations (Russian). Uspehi Mat. Nauk, (N.S.) 5, no. 6 (40) (1950) 102—135. Amer. Math. Soc. Translation No. 96 (1953).

[4] E. HILLE, On the integration problem for Fokker-Planck's equation in the theory of stochastic processes. Onzième Congrès des Math. Scandinaves, Trondheim 1949, (1952) 183—194.

[5] E. HILLE, Les probabilités continues en chaîne. Comptes Rendus Acad. Sci. Paris, 230 (1950) 34—35.

[6] E. HILLE "Explosive" solutions of Fokker-Planck's equation. Proc. Internat. Congres Math. Cambridge, Mass. 1950, I. p. 435.

[7] E. HILLE, A note on Cauchy's problem. Ann. Soc. Polonaise de Math., 25 (1952) 56—68.

[8] E. HILLE, Une généralisation du problème de Cauchy. Ann. Inst. Fourier, 4 (1953) 31—48.

[9] E. HILLE, Sur le problème abstrait de Cauchy. Comptes Rendus Acad. Sci. Paris, 236 (1953) 1466—1467.

[10] E. HILLE, Le problème abstrait de Cauchy, Rendiconti Sem. Mat. Università di Torino, 12 (1953) 95—103.

[11] E. HILLE, The abstract Cauchy problem and Cauchy's problem for parabolic differential equations. Journ. d'Analyse Math., 3 (1954) 81—196, Corrigenda.

[12] E. HILLE, An abstract formulation of Cauchy's problem. Proc. Twelfth Congress Scand. Math. Lund 1953, (1954) 79—89.

[13] K. KODAIRA, On ordinary differential equations of any even order and the corresponding eigenfunction expansions. Amer. Journal of Math. 72 (1950) 502—544.

[14] R. S. PHILLIPS, A note on the abstract Cauchy problem. Proc. Nat. Acad. Sci. U.S.A., 40 (1954) 244—248.

[15] K. YOSIDA, An operator theoretical treatment of temporally homogeneous Markoff process. Journal Math. Soc. Japan, 1 (1949) 244—363.

[16] K. YOSIDA, Semi-group theory and the integration problem of diffusion equations. These Proceedings, I.

[Added in proof: For further development of this Cauchy problem see § 23.3 of E. Hille and R. S. Phillips, Functional Analysis and Semi-Groups, Amer. Math. Soc. Coll., Vol. XXXI, second edition, to appear. When $n > 2$ the existence of a unique solution of the ACP^n for certain restricted classes of prescribed data implies that U is a bounded operator.]

YALE UNIVERSITY.
NEW HAVEN, CONN.,
U.S.A.

SOME RESULTS IN THE TRANSCENDENTAL THEORY OF ALGEBRAIC VARIETIES

K. KODAIRA

See Symposium B. Algebraic Geometry

ÜBER AUTOMORPHE FUNKTIONEN

P. J. MYRBERG

1. Ich werde in meinem Vortrag automorphe Funktionen behandeln, die einer gegebenen fuchsschen Gruppe Γ angehören. Es handelt sich also um analytische Funktionen, die den reellen homographischen Substitutionen $z' = \dfrac{az+b}{cz+d} = S(z)$ der Gruppe Γ gegenüber ungeändert bleiben. Wir nehmen der Kürze halber an, dass keine parabolischen Substitutionen vorhanden sind und dass ferner der Existenzbereich der Funktionen mit der hyperbolischen Ebene, d.h. mit der oberen z-Halbebene, zusammenfällt.

Es ist bekannt, dass vermittels solcher Funktionen jede algebraische Riemannsche Fläche $P(x, y) = 0$ beliebigen Geschlechtes uniformisiert werden kann. Ferner ist es u.a. möglich, die Lösungen jeder homogenen linearen Differentialgleichung mit algebraischen Koeffizienten durch Einführung einer geeigneten z-Variablen in eindeutige Funktionen zu verwandeln. Man kann also behaupten, dass durch die neuen Transzendenten im allgemeinen Falle dasselbe geleistet werden kann, was für die Flächen vom Geschlecht Null und Eins vermittels rationaler bzw. elliptischer Funktionen möglich ist.

2. Während die Uniformisierungstheorie in ihren wichtigsten Teilen schon ein vollendetes Kapitel der Analysis bildet, hat man betreffs der analytischen Darstellung der automorphen Funktionen verhältnismässig wenig geleistet. Poincaré hat in seinen grundlegenden Arbeiten über automorphe Funktionen seine Theorie auf unendliche Reihen gegründet, auf die später nach ihm benannten Reihen der Form

$$\Theta(z) = \sum R(S) \left(\frac{dS}{dz}\right)^m, \quad R \text{ rational}, \ m > 1.$$

Die Summe der Reihe definiert eine im allgemeinen meromorphe Funktion, die sich den Substitutionen von Γ gegenüber multiplikativ gemäss der Formel

$$\Theta(S) = \Theta(z) \left(\frac{dS}{dz}\right)^{-m}$$

transformiert. Hieraus folgt, dass der Quotient zweier Poincaréschen Thetafunktionen eine automorphe Funktion der gegebenen Gruppe Γ darstellt. Die grosse prinzipielle Bedeutung der Poincaréschen Thetareihen und ihrer Verall-

gemeinerungen besteht in der Tatsache, dass es mit ihrer Hilfe möglich ist, für diskontinuierliche Gruppen sehr allgemeiner Art die Existenz der zugehörigen automorphen Funktionen nachzuweisen. Ihre Abhängigkeit von den Elementen der zu bildenden Funktion ist allerdings sehr kompliziert — im Gegensatz zu den bei den elliptischen Funktionen geltenden Entwicklungen.

Zweck unseres Vortrages ist es, eine Klasse von ganzen, d.h. in der hyperbolischen Ebene regulären analytischen Funktionen einzuführen, die als eine direkte Verallgemeinerung der elliptischen Thetafunktionen angesehen werden können. Wir werden nämlich sämtliche ganzen Funktionen $g(z)$ bestimmen, die für jede Substitution von Γ eine Gleichung der Form

$$g(S) = e^{\lambda_s u(z)+\eta_s}\, g(z) \tag{1}$$

befriedigen, wo $u(z)$ ebenfalls ganz ist. Unsere automorphe Thetafunktion $g(z)$ soll insbesondere primitiv heissen, wenn sie im Fundamentalbereich B von Γ eine einzige einfache Nullstelle besitzt. Es wird sich zeigen, dass unsere Thetafunktionen zugleich Primfunktionen sind, indem man aus ihnen durch Multiplikation und Division einen Ausdruck für jede automorphe Funktion von Γ gewinnen kann, wenn ihre Nullpunkte und Pole gegeben sind.

3. Wir beginnen mit unserer hypothetischen u-Funktion. Indem man die Gleichung (1) nacheinander auf zwei Substitutionen von Γ anwendet, ergibt sich für jede S aus Γ für $u(z)$ eine Gleichung der Form

$$u(S) = \alpha_s u(z) + \beta_s. \tag{2}$$

Die Funktion $u(z)$, die sich bei den elliptischen Funktionen auf eine lineare Funktion $u(z) = Az + B$ reduziert, gehört hier zur Klasse der Poincaréschen Zetafunktionen. Auf die zur Gruppe Γ gehörige Riemannsche Fläche F übertragen, geht $u(z)$ in eine unendlich vieldeutige Funktion $u(x)$ über, die sich auf geschlossenenen Wegen L von F gemäss der Formel

$$u(x') = \alpha u(x) + \beta, \quad L(x) = x' \tag{3}$$

transformiert. Wir nennen α den zu L gehörigen Multiplikator und β die zu L gehörige Periode von $u(x)$. Aus der entsprechenden Gleichung

$$g(x') = e^{\lambda u(x)+\eta}\, g(x)$$

der transformierten Thetafunktion $g(x)$ ergibt sich für die daraus abgeleitete Funktion

$$\varphi(x) = \frac{1}{u'(x)} \frac{d \log g(x)}{dx}$$

die Formel

$$\varphi(x') = \frac{1}{\alpha}\varphi(x) + \frac{\lambda}{\alpha}, \tag{4}$$

die der Formel (3) analog ist. Unsere Hauptaufgabe besteht nun in der allgemeinen Untersuchung von Funktionen, die Gleichungen der Form (3) und (4) genügen.

In dem Spezialfall, wo Γ keine elliptischen Substitutionen enthält und wo also $z(x)$ eine Hauptuniformisierende von F ist, sind auch die Funktionen $u(x)$ und $\varphi(x)$ relativ zur Fläche F unverzweigt. Sie reduzieren sich dann auf die. von Prym und Appell eingeführten Integrale von multiplikativen Funktionen und speziell auf Abelsche Integrale, wenn alle Multiplikatoren gleich Eins sind.

4. Wir bezeichnen im allgemeinen Falle mit $e_1, e_2, \ldots, e_q$ die relativen Verzweigungspunkte von $z(x)$ bezüglich F und mit $n_1, n_2, \ldots, n_q$ ihre endlichen Ordnungszahlen. Einem Umkreis $\mathfrak{C}_\nu$ um e_ν auf F entspricht in der z-Ebene ein System von äquivalenten elliptischen Substitutionen T_ν von Γ mit den Perioden n_ν. In der entsprechenden Gleichung (3) ist der Multiplikator α dann gleich einer Einheitswurzel $\alpha = e^{2\pi i \frac{\tau_\nu}{n_\nu}}$. Wird nun mit n die kleinste Vielfache der Zahlen n_ν bezeichnet, so sind die Potenzen $u'^n(x), \varphi'^n(x)$ der Ableitungen von $u(x)$ und $\varphi(x)$ relativ zur Fläche F unverzweigte multiplikative Funktionen. Ihre Logarithmen sind aber dann Abelsche Integrale dritter Gattung, deren Residuen in den logarithmischen Singularitäten, den Nullstellen und Polen von $u'(x)$ und $\varphi'(x)$, ganze Zahlen sind.

Um einen expliziten Ausdruck für das zu u gehörige Integral

$$I_3 = \log u'(x) = \int r(x, y)dx \tag{5}$$

zu gewinnen, führen wir auf F ein kanonisches Schnittsystem

$$A_\nu, A'_\nu = A_{\nu+p}, \quad (\nu = 1, 2, \ldots, p) \tag{6}$$

ein, wo A'_ν den zu A_ν konjugierten Schnitt bezeichnet. Angenommen, dass die Ableitung von $u(z)$ in den zu den Verzweigungspunkten e'_ν von F entsprechenden Punkten von Null verschieden ist, ergibt sich aus dem Residuensatz für die Anzahl N der von den Punkten e_ν verschiedenen endlichen Nullpunkte x_ν von $u'(x)$ der Ausdruck

$$N = 2p-2 + q-\sigma.$$

Hier bezeichnet p das Geschlecht von F und σ die Summe $\sigma = \sum_{\nu=1}^{q} \frac{\tau_\nu}{n_\nu}$, die eine ganze Zahl ist, vorausgesetzt, dass sämtliche Verzweigungspunkte von F endlich sind. Wird ferner mit $\omega(a, b)$ das bezüglich (A_ν) normierte Abelsche Integral dritter Gattung mit den logarithmischen Singularitäten a (Residuum $+1$) und b (Residuum -1) bezeichnet und das System der bezüglich (A_ν) normierten Abelschen Integrale erster Gattung

$$W_1, W_2, \ldots, W_p$$

eingeführt, so ergibt sich für (5) der Ausdruck

$$(7) \qquad I_3 = \sum_{\nu=1}^{N} \omega(x_\nu, t_\nu) + \sum_{\nu=1}^{N'} \omega(x'_\nu, t'_\nu) + \sum_{\varkappa=1}^{p} b_\varkappa W_\varkappa + C,$$

wo t_ν, t'_ν die Punkte e_ν, e'_ν und x'_ν die im Unendlichen liegenden N' Nullpunkte von $u'(x)$ bezeichnen. Unser Ausdruck (7) enthält insgesamt $3p - 1 + q - \sigma$ Parameter, nämlich x_ν, $b_\varkappa$, C.

5. Es ist möglich, diese Parameter so zu wählen, dass die Multiplikatoren von $u(x)$ für die A_ν-Wege beliebig vorgeschriebene Werte α_ν annehmen werden. Es soll dies in dem allgemeinen Fall gezeigt werden, wo die Multiplikatoren von den der vermittels eines Abelschen Integrals erster Gattung $W(x)$ gebildeten Exponentialfunktion $e^{W(x)}$ verschieden sind.

Was zuerst die p ersten A-Wege betrifft, so werden die zugehörigen α_ν-Werte stets durch die Wahl $b_\nu = \frac{1}{2\pi_i} \log \alpha_\nu$ erreicht. Für die übrigen p Multiplikatoren gelten nun die Gleichungen

$$\sum_{\nu=1}^{N} W_\mu(x_\nu) + C_\mu = \log \alpha'_\mu, \qquad (\mu = 1, 2, \ldots, p)$$

wo C_μ gewisse Konstanten bezeichnen. Im vorliegenden allgemeinen Fall lassen sich hieraus p von den Punkten x_ν, dies seien $x_1, x_2, \ldots, x_p$, durch die $N-p$ übrigen bestimmen. Der allgemeine Ausdruck für eine u-Funktion mit den gegebenen $2p$ Multiplikatoren enthält somit insgesamt $N(u) = p-1+q-\sigma$ Parameter, nämlich $x_{p+1}, \ldots, x_N$ und C.

Man kann zeigen, dass der Ausdruck von $u(x)$ darstellbar ist in der Form,

$$u(x) = \sum_{\nu=1}^{N(u)} q_\nu u_\nu(x) + C$$

wo $u_\nu(x)$ gewisse unter den u-Funktionen und q_ν, C unbestimmte Parameter bezeichnen. Durch eine geeignete Wahl di ser Parameter lässt sich erreichen, dass die Perioden von $u(x)$ für gewisse $N(u)$ unter den Wegen A_ν, A'_ν, C_μ verschwinden werden und ferner, dass $u'(x_0) = 1$ für einen gegebenen von den Punkten x_ν verschiedenen Punkt x_0 wird.

6. Wir gehen jetzt zur Bestimmung unserer hypothetischen Funktion $\varphi(x)$ über, die wie $u(x)$ in den Punkten e_ν relativ zur Fläche F verzweigt ist und die ferner ihre Pole in dem gegebenen Nullpunkt x_0 von $g(x)$ und in den Nullpunkten $x_1, x_2, \ldots, x_N$ von $u'(x)$ hat. Ferner sollen die Multiplikatoren von $\varphi(x)$ den von u inverse Grössen sein. Durch Betrachtungen, auf die wir hier nicht näher eingehen können, kann man elementare Integrale zweiter Gattung φ_ν mit den inversen Multiplikatoren $\frac{1}{\alpha}$ konstruieren, welche einen

einzigen Pol in einem gegebenen Punkt x_ν mit der Entwicklung $\frac{1}{x-x_\nu} + \ldots$ besitzen. Durch solche elementaren Integrale lässt sich der allgemeine Ausdruck von φ in der Form

$$\varphi(x) = \varphi_0(x) + \sum_{\nu=1}^{N} c_\nu \varphi_\nu(x) + K \tag{8}$$

schreiben, wo c_ν, K unbestimmte Parameter bezeichnen. Durch Subtraktion eines Ausdrucks der Form

$$\sum_{\nu=1}^{N(\tilde{u})} d_\nu u_\nu(x), \quad N(\tilde{u}) = p - 1 + \sigma$$

wo $\tilde{u}_\nu(x)$ ein System von linear unabhängigen Integralen erster Gattung bezeichnet, die den inversen Multiplikatoren $\frac{1}{\alpha}$ angehören, kann man gewisse $N(\tilde{u})$ unter den Perioden von $\varphi_\nu(x)$ von ν unabhängig zum Verschwinden bringen.

Wir führen jetzt die vermittels der Abelschen Normalintegrale erster Gattung $W_\nu(x)$ gebildeten multiplikativen Funktionen $\frac{W_\nu'(x)}{u'(x)}$ ein, die linear durch die Funktionen $\varphi_1, \varphi_2, \ldots, \varphi_N$ darstellbar sind. Dadurch lassen sich gewisse p unter den Funktionen φ_ν in (8) eliminieren, dies seien $\varphi_1, \varphi_2, \ldots, \varphi_p$, wodurch der Ausdruck (8) die Form

$$\varphi(x) = \varphi_0(x) + \sum_{\nu=p+1}^{N} c_\nu' \varphi_\nu(x) + K + \sum_{\mu=1}^{p} b_\mu \frac{W_\mu'(x)}{u'(x)} \tag{8'}$$

annehmen wird. Auf die Perioden von $\varphi(x)$ hat die letzte Summe auf der rechten Seite von (8)' keinen Einfluss. Was die übrigen Parameter c_ν', K in (8)' betrifft, kann man sie stets so wählen, dass $N-p+1$ unter den noch nicht normierten Perioden von $\varphi(x)$ verschwinden werden. Dadurch bleiben aber nur zwei nichtverschwindende Perioden übrig. Diese Perioden, sie seien A_p, A_p', sind von x_0 unabhängig.

Wir bilden nun die Funktion $\psi(x) = \int \varphi u' dx$, die sich auf geschlossenen Wegen auf F gemäss der Formel

$$\psi(x') = \psi(x) + \lambda u(x) + \eta \tag{9}$$

transformiert. Wird $\psi(x)$ durch $\psi(x) - \sum_{\nu=1}^{p} b_\nu W_\nu(x)$ ersetzt, so kann man, durch eine geeignete Wahl der Konstanten b_ν erreichen, dass gewisse p unter den

Grössen η verschwinden werden, dies seien etwa die den Wegen $A_1, \ldots, A_p$ entsprechenden Konstanten. Es erübrigt noch

$$g(x) = e^{\psi(x)}$$

zu setzen, um schliesslich eine Funktion zu erhalten, die als Funktion $g(z)$ von z allen für die primitiven Thetafunktionen aufgestellten Bedingungen genügt. Sie ist in der hyperbolischen Ebene eine ganze Funktion, die im Fundamentalbereich B von Γ eine einzige einfache Nullstelle z_0 besitzt. Ferner genügt sie für jede Substitution von Γ einer Gleichung der Form (1). Speziell für diejenigen hyperbolischen Substitutionen S_ν, S'_ν von Γ, die einem Umlauf längs den Schnitten des kanonischen Systems (6) entsprechen, lauten die Gleichungen (1):

$$\begin{gathered} g(S_\nu) = g(z), \quad g(S'_\nu) = e^{\eta_\nu} g(z), \quad (\nu = 1, 2, \ldots, p-1) \\ g(S_p) = e^{\lambda_p u(z)}\, g(z), \quad g(S'_p) = e^{\eta_p + \lambda'_p u(z)} g(z). \end{gathered} \tag{10}$$

Für die elliptischen Substitutionen T_ν bleibt $g(z)$ im vorliegenden Falle invariant:

$$g(T_\nu) = g(z). \tag{10'}$$

7. Es sei jetzt $f(z)$ eine beliebige automorphe Funktion von Γ, die ihre Nullpunkte und Pole in den resp. mit den Punkten

$$a_\nu, b_\nu, \qquad (\nu = 1, 2, \ldots, l)$$

bezüglich Γ äquivalenten Punkten hat. Wir bilden nun vermittels der in z_0 verschwindenden primitiven Thetafunktion $g(z, z_0)$ den Ausdruck

$$f^*(z) = \prod_{\nu=1}^{l} \frac{g(z, a_\nu)}{g(z, b_\nu)}$$

und erhalten eine in der hyperbolischen Ebene meromorphe Funktion, deren Pole und Nullpunkte mit denjenigen von $f(z)$ zusammenfallen und die sich den Substitutionen von Γ gegenüber multiplikativ mit konstanten Faktoren verhält. Dies ist dann auch bei dem Quotienten

$$\Gamma(z) = f(z) : f^*(z)$$

der Fall, der eine nirgends verschwindende ganze Funktion darstellt. Der Logarithmus von $\Gamma(z)$ ist dann eine additive Funktion und als Funktion auf F ein Abelsches Integral erster Gattung $W_0(x)$, weil er keine Singularitäten besitzt. Weil ferner ihre p ersten Perioden $A_1, \ldots, A_p$ sämtlich Multipla von $2\pi i$ sind, hat $W_0(x)$ den Ausdruck

$$W_0(x) = \sum_{\nu=1}^{p} k_\nu W_\nu(x),$$

wo k_ν ganze Zahlen bezeichnen. Aus dem Obigen folgt für unsere automorphe Funktion der Ausdruck

$$f(z) = e^{W_0(z)} \prod_{\nu=1}^{l} \frac{g(z, a_\nu)}{g(z, b_\nu)} \tag{11}$$

vermittels der primitiven Thetafunktion $g(z)$. Hieraus geht hervor, dass unsere primitive Thetafunktion die Rolle einer Primfunktion spielt.

Wir haben im Vorhergehenden den speziellen Fall ausgeschlossen, wo die Multiplikatoren von $u(z)$ gleich denjenigen einer Exponentialfunktion von der Form $e^{W(z)}$ sind. In diesem ausgeschlossenen Falle lassen sich die Gleichungen (10) noch so vereinfachen, dass alle Grössen η, bis auf eine einzige, gleich Null werden. Wenn speziell alle Multiplikatoren gleich Eins gewählt werden, in welchem Falle $u(x)$ sich auf ein Abelsches Integral erster Gattung reduziert, kann man p verschiedene primitive Thetafunktionen einführen, deren Produkt gleich einer Riemannschen Thetafunktion von p Variablen zusammenfällt, wo die Variablen durch Abelsche Normalintegrale erster Gattung ersetzt worden sind.

8. Es erübrigt noch, unsere Thetafunktionen in ihrer Abhängigkeit von z durch analytische Ausdrücke explizite darzustellen.

Nach dem Obigen genügt die Funktion $u(x)$ der Differentialgleichung

$$u''(x) = r(x, y)u'(x). \tag{12}$$

Aus (9) geht ferner hervor, dass die Funktion $\psi(x) = \log g(x)$ mit $u(x)$ zusammen der Differentialgleichung dritter Ordnung

$$\zeta''' - (R + 2r)\zeta'' + (r^2 + rR - r')\zeta' = 0$$

genügt, wo $R(x)$ die durch $R = \frac{d}{dx} \log \varphi'(x)$ definierte rationale Funktion von F bezeichnet. Die Funktionen

$$G(z) = \frac{\psi'(z)}{f'(z)}, \quad v(z) = \frac{u'(z)}{f'(z)},$$

wo $f(z)$ eine beliebige automorphe Funktion von Γ bezeichnet, genügen aber dann den Gleichungen

$$G(S) = G(z) + \alpha v(z), \quad v(S) = \alpha v(z)$$

und bilden somit ein System von Poincaréschen Zetafunktionen. Durch Einführung einer Poincaréschen Zeta- und Thetareihe $Z(z)$ bzw. $\Theta(z)$ ergibt sich hieraus für $g(z)$ schliesslich der Ausdruck $g(z) = e^{\int \frac{Z(z)}{\Theta(z)} dz}$.

9. Wir wollen zum Abschluss auf eine Verallgemeinerung der obigen Entwicklungen hinweisen, wodurch es wenigstens in sehr allgemeinen Fällen möglich ist, die Poincaréschen Reihen zu vermeiden. Unsere Verallgemeinerung besteht darin, dass wir die Gleichungen (1) durch die allgemeineren

$$g(S) = e^{\sum_{\nu=1}^{M} \alpha_\nu u_\nu(z) + \beta} \, g(z) \tag{14}$$

ersetzen, wo die Exponenten linear von endlich vielen ganzen Funktionen $u_\nu(z)$ abhängen. Die neuen u-Funktionen genügen dabei für jede S aus Γ einem Gleichungssystem von der Form

$$u_\nu(S) = \sum_{\mu=1}^{M} a_{\nu\mu} u_\mu(z) + b_\nu,$$

und sie bilden also ein System von Poincaréschen Zetafunktionen. Wir werden uns im folgenden der Kürze halber auf Gruppen vom Geschlecht Null beschränken.

Es seien e_ν die Verzweigungspunkte der polymorphen Funktion $z(x)$ und n_ν die zugehörigen Verzweigungszahlen. Wie wir in einer schon vor ein paar Jahrzehnten veröffentlichten Arbeit gezeigt haben, ist es stets möglich, eine Riemannsche Fläche F_1 vom Geschlecht Eins zu konstruieren, deren sämtliche Verzweigungspunkte über den Punkten e_ν liegen, und zwar derart, dass die Ordnung jedes Verzweigungspunktes gleich einem Teiler der entsprechenden Zahl n_ν ist. Hieraus folgt, dass das zu F_1 gehörige elliptische Integral erster Gattung v eine eindeutige und reguläre Funktion von z in der hyperbolischen Ebene ist.

Die zu der Fläche F_1 gehörigen eindeutigen Funktionen sind automorphe Funktionen einer Untergruppe Γ_1 von Γ, deren Index gleich der Blätterzahl M von F_1 ist. Für jede S aus Γ_1 verhält sich $v(z)$ additiv gemäss der Formel

$$v(S) = v(z) + m_1\omega_1 + m_2\omega_2.$$

Diejenigen Substitutionen von Γ_1, für welche $v(z)$ ungeändert bleibt, bilden eine Untergruppe Γ_v von Γ_1, die eine fuchsoide Gruppe vom Geschlecht Null ist. Durch Betrachtungen, auf die wir hier nicht eingehen können, bekommt man für $v(z)$ den kanonischen Ausdruck einfachster Art

$$\frac{v(z) - v(a)}{v(z) - v(b)} : \frac{v(z_0) - v(a)}{v(z_0) - v(b)} = \prod_{\Gamma_v} \left[\frac{S(z) - a}{S(z) - b} : \frac{S(z_0) - a}{S(z_0) - b} \right], \tag{15}$$

welcher die Form eines Schottkyschen Produktes hat, dessen Konvergenz aber nicht absolut, sondern bedingt ist. Durch Zusammensetzung ergibt sich aus (15) und aus den bekannten Ausdrücken der elliptischen Funktionen für die automorphen Funktionen der gegebenen fuchsschen Gruppe Γ eine ähnliche

Darstellung. Damit wird ein allgemeines von F. Klein aufgestelltes Problem für den vorliegende Fall gelöst.

11. Es sei jetzt $\vartheta(v)$ eine elliptische Thetafunktion mit den Perioden ω_1 und ω_2. Durch Zusammensetzung der Funktionen $\vartheta(v)$ und $v(z)$ gelangt man zu einer ganzen Funktion $\vartheta(z, z_0)$, die eine Nullstelle z_0 im Fundamentalbereich B_1 von Γ_1 hat und die als eine primitive Thetafunktion von Γ_1 für jede S aus Γ_1 einer Gleichung der Form (1) genügt.

Es seien nun x_ν die über dem Punkt $x_\nu(z_0)$ liegenden Punkte der Fläche F_1 und z_ν ihre Bilder im Bereich B_1. Wir bilden das Produkt

$$\vartheta(z) = \prod_{\nu=1}^{M} \vartheta(z, z_\nu)$$

und bekommen eine primitive Thetafunktion der gegebenen Gruppe Γ_1, die für jede S aus Γ der verallgemeinerten Gleichung (14) genügt, wo $u_\nu(z)=v(T_\nu)$ und T_ν die Substitutionen der Faktorgruppe Γ_1/Γ bezeichnen. Vermittels solcher verallgemeinerten Thetafunktionen lässt sich dann jede automorphe Funktion von Γ analytisch darstellen. Die analytische Darstellung wird hierdurch vermittels des kanonischen Ausdruckes (15) und der für die elliptischen Thetafunktionen geltenden Ausdrücke bewerkstelligt.

Wir haben uns im Vorhergehenden auf die Gruppen vom Geschlecht Null begeschränkt. Was den allgemeinen Fall betrifft, bemerken wir hier nur, dass bei jeder gegebenen Riemannschen Fläche beliebigen Geschlechtes eine analoge Darstellung möglich ist, wenn man sich einer geeigneten Uniformisierung bedient. In welchem Umfang unsere Methode auf alle fuchsschen Gruppen übertragen werden kann, ist eine noch offene Frage.

HELSINKI.

CONTRIBUTIONS À LA THÉORIE DE LA DÉRIVATION DE FONCTIONS D'ENSEMBLE

CHR. PAUC

Les contributions à la Théorie de la Dérivation de Fonctions d'Ensemble qui vont être signalées ont leur origine dans le *Théorème de Radon-Nikodym* et un mémoire de *R. de Possel* (Journal Math. p. et app., IX s. 15 (1936), 391-409).

$[R, \mathscr{M}, \mu]$: *espace mesuré*, $R = \cup M_n^0$ où $M_n^0 \in \mathscr{M}$, $M_{n+1}^0 \supset M_n^0$, $\mu(M_n^0) < \infty$, $n = 1, 2, \ldots$, $\mathscr{N}$: famille des ensembles de μ-mesure nulle, $\mathscr{N}^*$: famille des sous-ensembles des ensembles de $\mathscr{N}$. Un ensemble est dit *borné* s'il est contenu dans un M_n^0. Par *distribution de Radon* nous entendons une fonction ψ à valeurs réelles ($\pm \infty$ compris) définie sur $\mathscr{M}$, finie pour M borné et σ-additive. Une telle distribution admet une représentation

$$\psi(M) = \psi(M.N_0) + \int_M f \, d\mu \quad \text{où } N_0 \in \mathscr{N}.$$

f est l'intégrant de Radon-Nikodym.

Nous observons que cet intégrant est défini (1) sans processus de dérivation ponctuelle, (2) à un ensemble *indéterminé* de mesure nulle près. Il s'agit donc non d'une fonction mais d'un „fonctionoide" selon O. M. Nikodym, „Ortsfunktion" selon Carathéodory.

Première Partie. De Possel a introduit l'instrument permettant de donner à l'intégrant une *représentation concrète* sous forme de dérivé, la *base de dérivation*: A tout point x d'un ensemble $E \subset R$ sont attachées des suites de Moore-Smith M_ι d'ensembles de $\mathscr{M}$ de mesure finie positive qui sont dites *converger sur* x, toute sous-suite confinale d'une suite convergeant sur x convergeant elle-même sur x. De manière précise (O. Haupt-Chr. Pauc, Propriétés de mesurabilité de bases de dérivation, en cours d'impression aux Port. Math.) la base de dérivation $\mathfrak{B}$ est la relation au sens de Hahn entre points de R et suites M_ι. Les ensembles intervenant dans les suites convergentes sont dits *constituants* et leur famille désignée par $\mathscr{S}$ („spread" des bases de A. P. Morse). Si ψ est une application de $\mathscr{S}$ dans un espace vectoriel V, par dérivé $D\psi(x)$ de ψ en x (*dérivé ponctuel*) nous entendons toute limite de

$$\psi(M_\iota)/\mu(M_\iota)$$

pour toutes les suites M_ι convergeant sur x. Si V est la droite numérique achevée, les bornes supérieure et inférieure des dérivés en x sont les *dérivés extrêmes* $\overline{D}\psi(x)$ et $\underline{D}\psi(x)$ en x.

A propos de cette notion de base de dérivation, il est intéressant de noter que la convergence des ensembles M_t sur un point ne présuppose ni métrique, ni topologie. Le terme „convergence" (ou „contraction") rappelle simplement l'origine de la notion. La mesure des M_t dans le cas général peut même tendre vers ∞.

Les hypothèses utilisées par de Possel pour assurer la représentation concrète de l'intégrant f comme dérivé ponctuel sont des *hypothèses vitaliennes*. Nous entendons par *recouvrement* $\mathfrak{B}$-*fin* d'un ensemble X une famille $\mathscr{V}$ de constituants telle que, en presque tout point x de X (c'est-à-dire pour $x \in X—N^*$ où $X \supset N^* \in \mathscr{N}^*$), il existe (au moins) une suite convergeant sur x dont tous les constituants appartiennent à $\mathscr{V}$. Une *hypothèse vitalienne* consiste à demander l'existence, pour tout recouvrement $\mathfrak{B}$-fin d'un ensemble quelconque X, d'une sous-famille dénombrable $\{C_n\}$ recouvrant X à un ensemble de $\mathscr{N}^*$ près et satisfaisant à une *condition d'empiètement*; ainsi l'*hypothèse classique de Vitali*, dite *forte*, exige la *disjonction* des C_n. Nous disons que la base $\mathfrak{B}$ dérive une classe de fonctions $\psi|\mathscr{M}$ admettant un théorème de Radon-Nikodym si, pour toute fonction ψ de cette classe, $D\psi$ est défini univoquement presque partout (c'est-à-dire ici à un ensemble *déterminé* de $\mathscr{N}^*$ près) et représente l'intégrant RN. Ainsi une base $\mathfrak{B}$ satisfaisant à Vitali fort dérive les μ-intégrales de Radon. La *condition faible de Vitali* tolère un μ-*empiètement* arbitrairement petit des C_n. De Possel a montré l'*équivalence entre Vitali faible et la dérivabilité des* μ-*intégrales lipschitziennes*. Ce remarquable résultat a été le point de départ vers des théorèmes d'équivalence analogues obtenus par C. A. Hayes et Chr. Pauc (Full individual and class differentiation theorems for set functions in their relations to halo and Vitali properties, à paraître au Canad. J. of Math., Janvier 1955) Dans l'un d'eux la dérivabilité concerne les *distributions de Radon*, l'empiètement correspond aux mesures de Radon Dans un autre il s'agit de la dérivabilité des *intégrales de fonctions de* $L^{(p)}$, $p > 1$, la définition de l'empiètement fait intervenir la q *ième* $(1/p + 1/q = 1)$ *puissance* de la fonction „*excès de recouvrement*"; $\mathfrak{B}$ doit en outre être une base de type D, la base euclidienne d'intervalles en est une illustration.

Dans le cas classique de la base euclidienne de cubes, la propriété forte de Vitali est établie à partir d'une *propriété de halo* de ces cubes: Si pour tout cube C_0 (*noyau*), $H(C_0)$ (*halo*) représente l'union des cubes $C \leqq C_0$ rencontrant C_0, la *dilatation en halo*

$$\varrho(C_0) = \mu(H(C_0))/\mu(C_0)$$

est uniformément bornée pour tous les cubes C_0 (en fait $= 3^n$ où $n =$ dimension de l'espace) Dans le cas général les C sont des constituants de la base, C_0 peut être un constituant, une union finie de constituants, un ensemble de $\mathscr{M}$ de mesure finie positive. Les propriétés de halo diffèrent aussi par les *conditions*

d'incidence entre C_0 et les C agrégés à C_0. En outre la condition de limitation uniforme de $\varrho(C_0)$ peut être remplacée par une condition de finitude ponctuelle ou locale, A. P. Morse (Trans. Amer. Math. Soc., **55** (1944), 205—235) dans le cas de ses bases métriques („blankets"), O. Haupt et Chr. Pauc (Sitz. Ber. Bay. Ak. Wiss., (1950), 187—207; Arch. d. Math. **4** (1953), 107—114) dans le cas général ont étudié l'implication de propriétés vitaliennes par des propriétés de halo. S'inspirant de H. Busemann et W. Feller (Fund. Math., **22** (1934), 226—256), C. A. Hayes et Chr. Pauc (loc. cit.) ont obtenu un théorème d'équivalence entre une propriété de dérivabilité et une propriété de halo.

Seconde Partie. Dans ce qui suit nous indiquons les grandes lignes d'une théorie de Burkill-Ward-Denjoy pour fonctions d' *„intervalle abstrait"* ou *„cellule"*. Les dérivés à définition ponctuelle sont remplacés par des intégrants RN ou par des dérivés définis globalement grâce à la convergence en moyenne.

(Axiomes $A1$) $\mathscr{I}$ désigne une famille non vide d'ensembles de $\mathscr{M}$ de mesure finie positive, appelés *cellules*. Par *partition cellulaire* de R nous entendons toute famille de cellules non empiètantes I_n, d'union $= R$, toute cellule I ne rencontrant qu' un nombre fini de I_n. Nous supposons que l'ensemble $\mathfrak{N}$ (*réseau*) des partitions cellulaires de R est un *système dirigé* (ou *filtre*) en finesse de partition, chaque cellule apparaissant dans une partition au moins. Dans l'étude des fonctions de cellules, seules interviennent les mesures des cellules, aussi restreignons-nous μ à la plus petite σ-algèbre booléenne $\mathscr{B}$ incluant $\mathscr{I}$, ce qui revient à supposer $\mathscr{M} = \mathscr{B}$ (*adaptation de la mesure au réseau*). Nous transposons implicitement les notions de la théorie des fonctions d'intervalle aux fonctions de cellule. Le cadre décrit est d'une grande généralité, c'est ainsi que pour $\mu(R)$ fini, $\mathscr{I}$ peut être $= \mathscr{B}$. Toutefois nous avons obtenu les théorèmes suivants (Chr. Pauc, C. R. Paris **236** (1953), 1937—1939): (T1) Si la fonction de cellule ψ est absolument continue, l'intégrale supérieure de partition β_s et l'intégrale inférieure de partition β_i de ψ selon Burkill peuvent être représentées comme μ-intégrales de fonctions f_s et f_i. (T2) Toute fonction additive ψ de variation bornée peut être représentés d'une manière et d'une seule comme la somme d'une fonction additive absolument continue ψ_r et d'une fonction additive ψ_s singulière selon Saks. (T2) a pu être déduit par H. Bauer de son théorème de décomposition d'une valuation de treillis en partie continue et partie purement discontinue (Sitz. Ber. Bay. Ak. Wiss., (1953), 89—117). Pour toute partition $\mathscr{P} = \{I_n\} \in \mathfrak{N}$ nous désignons par $\mathscr{P}$-*dérivé* de ψ la fonction

$$D_{\mathscr{P}}\psi(x) = \psi(I_j)/\mu(I_j),$$

I_j représentant une des cellules de $\mathscr{P}$ contenant x. Le $\mathscr{P}$-dérivé est défini univoquement p.p. Par $L^{(p)}$-dérivé de partition de ψ nous entendons la limite sur $\mathfrak{N}$, quand elle existe pour chaque cellule, de $D_{\mathscr{P}}\psi(x)$ relativement à la

convergence dans $L^{(p)}$. (T3) Si ψ est l'intégrale d'une fonction $f \in L^{(p)}$ (c'est-à-dire $\int_I |f|^p \, d\mu < \infty$ pour tout $I \in \mathscr{I}$), alors le $L^{(p)}$-dérivé existe $= f$ (mod $\mathscr{N}$). (T4) Une condition nécessaire et suffisante pour que la fonction de cellule ψ absolument continue soit sommable sur tout $I \in \mathscr{I}$ au sens de l'intégrale de partition de Burkill est que ψ soit $L^{(1)}$-dérivable. S'il en est ainsi $\beta(I)$ est l'intégrale de ce dérivé sur I. K. Krickeberg (C. R. Paris, **238** (1954), 764—766) a montré que la possibilité d'interprétation pour les fonctions lipschitziennes ψ de l'intégrant f_s de (T1) comme $L^{(1)}$-dérivé supérieur de partition implique une condition de Vitali forte alors que d'après (T4) aucune n'est nécessaire quand $\beta_s = \beta_i$.

(Axiomes A2) Nous postulons maintenant une fonction δ positive finie, définie sur $\mathscr{I}$ et non décroissante que nous appelons *norme*. Nous supposons que pour toute cellule existent des partitions cellulaires de norme arbitrairement petite, et en outre (Axiome U): A presque tout point x d'une cellule quelconque I correspond $\delta_0 = \delta(x, I)$ tel que toute cellule contenant x et de norme $< \delta_0$ soit incluse dans I. Cet axiome, dans le cas où R est un espace métrique et δ le diamètre, est vérifié quand les cellules sont mesurables Jordan (O. Haupt-Chr. Pauc, C. R. Paris, **230** (1950), 711—712), ce qui est le cas pour les décompositions simpliciales classiques. Sous les hypothèses énoncées, pour toute fonction ψ absolument continue les notions d'intégrale de Burkill et de $L^{(1)}$-dérivé définies suivant la finesse de partition et la finesse de norme sont équivalentes. Si outre (A1) et (A2) nous postulons Vitali fort les intégrants et $L^{(p)}$-dérivés intervenant dans (T, 1—4) admettent une représentation concrète comme δ-dérivés (O. Haupt-G. Aumann-Chr. Pauc, Integralrechnung, 10.1.—10.2, Berlin (1954)). D. Rutovitz et Chr. Pauc ont étudié les fonctions de cellule additives mais non de variation bornée, adjoignant à (A1) et (A2) un axiome dit *axiome de grille* dû à D. Rutovitz. *Définitions*: Un complexe $\mathscr{K}$ est une famille finie de cellules non empiètantes, $\mathscr{K}^u$ est la famille des cellules obtenues par union de cellules de $\mathscr{K}$. Une cellule J *morcelle* une cellule I si $I = J \cap J'$ où $J' \in \mathscr{I}$, $\mu(J.J') = 0$. Une *chaîne de morcellement p-régulière* $\mathscr{C}$ entre les cellules J et I est une suite finie de cellules $J_0, \ldots, J_n$ ($n \geqq 1$) telle que $J_0 = J$, $J_n = I$, J_k morcelle J_{k+1} et $\mu(J_{k+1})/\mu(J_k) \leqq p$ pour $k = 0, 1, \ldots, n-1$. Une *grille p-régulière $\mathscr{G}$ sur une cellule I_0* est une partition cellulaire non triviale de I_0 telle que: (i) si $J_0, \ldots, J_n$ est une chaîne de morcellement $\subset \mathscr{G}^u$ dont le premier élément J_0 morcelle J_n, alors tout J_k ($k = 0, 1, \ldots, n-1$) morcelle J_n, (ii) pour tout I, $J \in \mathscr{G}^u$ avec $J \neq I$ et $J \subset I$, il existe une chaîne de morcellement p-régulière $\mathscr{C} \subset \mathscr{G}^u$ entre J et I. *Axiome de Grille* (A3): Il existe un nombre p^0 fini telle que pour toute cellule I et tout complexe $\mathscr{K}$ à cellules incluses dans I, il existe sur I une grille $\mathscr{G}$ p^0-régulière telle que $\mathscr{K} \subset \mathscr{G}^u$.

Théorème de Structure: Si ψ est une fonction de cellule additive à δ-dérivés p.p.finis, il existe une fonction μ-mesurable f („*intégrant Denjoy*") possédant la propriété suivante: tout ensemble S de mesure extérieure positive inclut un sous-ensemble L de mesure positive et borné (c'est-à-dire inclus dans une union finie de cellules) tel que f soit borné mod $\mathcal{N}$ sur une enveloppe jordanienne (relativement à $\mathfrak{R}$) $\tilde{L}$ de L et que

$$\psi(I) = \int_{\tilde{L}I} f\, d\mu + \int_I \psi(J||L)$$

où, pour $J \in \mathcal{I}$,

$$\psi(J||L) = 0 \text{ si } J \text{ rencontre } L, = \psi(J) \text{ si } J.L \text{ est vide.}$$

Nos axiomes sont par exemple vérifiés pour les intervalles fermés d'une puissance dénombrable de la droite numérique $\mathbf{R}$ mesurée suivant Lebesgue, la norme δ étant compatible avec la topologie produit.

Si nous voulons interpréter l'intégrant Denjoy f comme un dérivé ponctuel nous devons postuler Vitali faible. Nous obtenons entre autres une généralisation d'un théorème de Ward car les intervalles de $\boldsymbol{R}^m$ pour m fini vérifient Vitali faible. Il n'en est pas de même pour ceux de $\boldsymbol{R}^{\aleph^\circ}$ (J. Dieudonné, Fund. Math., **37** (1950), 242—248).

NANTES.

SUR UNE MÉTHODE TOPOLOGIQUE DE L'EXAMEN DE L'ALLURE ASYMPTOTIQUE DES INTÉGRALES DES ÉQUATIONS DIFFÉRENTIELLES

T. WAŻEWSKI

Introduction.

La méthode dont nous allons parler date depuis 1947 [7]. Appliquée aux différents problèmes asymptotiques elle a permit de généraliser les résultats de plusieurs auteurs (S. Faedo, E. Levi, N. Levinson, O. Perron, I. G. Petrovsky, H. Poincaré, A. Wintner, V. A. Yacoubovitch et des autres). Elle permet de résoudre certains problèmes asymptotiques relatifs aux équations différentielles très peu régulières ou très régulières dans les conditions où l'on ne sait pas appliquer les méthodes classiques (par ex. la méthode des approximations successives). Nous chercherons de caractériser son mécanisme en nous bornant aux exemples les plus simples possible. — La bibliographie insérée à la fin n'est pas complète.

Hypothèse H.

Nous nous bornons à l'espace des points réels et au cas des fonctions réelles. Nous supposons que les deuxièmes membres du système d'équations différentielles

$$\frac{dx_i}{dt} = f_i(t, x_1, \ldots, x_n), \quad (i = 1, \ldots, n) \tag{1}$$

soient définis et continus dans un ensemble *ouvert* W et que par chaque point de W passe une intégrale *unique* de ce système.

Introduisons les notations

$$X = (x_1, \ldots, x_n), \quad (t, X) = (t, x_1, \ldots, x_n), \quad F = (f_1, \ldots, f_n).$$

Le système (1) peut être écrit sous la forme

$$X'(t) = F(t, X). \tag{1}$$

Notations.

Nous désignerons l'intégrale de ce système issue du point $P \in W$ par $X = X(t, P)$. Cette intégrale sera envisagée comme „*saturée*", c'est-à-dire elle sera considérée dans le plus grand intervalle (ouvert) $D(P)$ dans lequel elle existe. Nous désignerons par t_P la première coordonée du point P. Posons

$$Y(t, P) = (t, X(t, P)).$$

Pour $P \in W$ nous désignerons par $L^+(P)$ la classe des points $Q = Y(t, P)$ pour les quels $t_P \leqq t$, $t \in D(P)$.

Demi-intégrale saturée.

L'ensemble $L^+(P)$ sera dit demi-intégrale (droite) saturée issue de P ou tout court, demi-intégrale issue de P.

Tuyau T. Demi-intégrale asymptotique par rapport à T.

Soit $T \subset W$ un ensemble ouvert. Cet ensemble sera appelé dans la suite tuyau. La partie commune de W avec la frontière de T constitue la frontière de T relative à W. Elle sera désignée par $Fr(T, W)$. Nous posons $\overline{T} = T + Fr(T, W)$.

La demi-intégrale $L^+(P)$ sera dite asymptotique par rapport à T lorsque

$$L^+(P) \subset T.$$

Conséquent $C(P)$de P. Ensemble de sortie S du tuyau T. Ombre gauche G de S.

Soit $P \in T$. Deux cas sont possibles. Ou bien $L^+(P) \subset T$ ou bien il existe un point $Q = Y(t^*, P)$, tel que $Q \in Fr(T, W)$ et que $Y(t, P) \in T$ lorsque $t_P \leqq t < t^*$. Ce point sera dit *conséquent* de P (relatif à T) et sera désigné par $C(P)$ (terminologie de Poincaré). Chaque point qui est conséquent d'un point $P \in T$ sera dit *point de sortie* (relatif à T). La classe de tous les points de sortie sera désignée par S et sera dite *ensemble de sortie* (relatif à T). La classe de tous les points $P \in T$ pour lesquels $C(P)$ existe sera dite *ombre gauche* de S et sera désignée par G.

Problème A.

Soit $P \in T$. Il s'agit d'indiquer les conditions pour que $L^+(P) \subset T$ c'est-à-dire pour que cette demi-intégrale soit asymptotique par rapport à T.

Analyse de ce problème. Il est à peu près évident que la condition nécessaire et suffisante pour que $L^+(P) \subset T$ consiste en ce que $P \in T - G$. Si $T - G \neq 0$ il existe au moins une demi-intégrale asymptotique par rapport à T. Si la section de $T - G$ par un hyperplan $t = t_0$ constitue un ensemble à p dimensions au sens de Urysohn et Menger, la classe des demi-intégrales asymptotiques par rapport à T dépend de p paramètres essentiels au moins. Il se pose donc la question de savoir si il est possible d'obtenir le renseignement sur le nombre de dimension de $T - G$ (ou au moins si cet ensemble n'est pas vide) et si c'est possible *sans effectuer la solution du système* (1). Or pour déterminer l'ensemble G il faudrait, en général, savoir calculer effectivement toutes les intégrales du système (1) qui rencontrent l'ensemble T.

Il est cependant évident que chaque intégrale rencontrant G rencontre forcément l'ensemble S. Or si T est donné il est, en général, facile de répondre à la question si un point $p \in Fr(T, W)$ appartient à S ou non et cela sans résoudre le système (1). On peut donc déterminer effectivement S sans résoudre ce système. Si S est vide alors G est évidemment vide et $T - G = T$.

Dans ce cas banal chaque demi-intégrale issue de T est asymptotique par rapport à T. Supposons que $S \neq 0$. L'ensemble G est parfaitement déterminé par les ensembles S et T et par le système (1). Il peut arriver que les propriétés topologiques de S et de T suffisent pour établir certaines propriétés topologiques de G et de $T - G$ sans qu'il soit nécessaire de connaître l'allure des intégrales à l'intérieur de T c'est-à-dire sans résoudre effectivement le système (1).

Cette analyse conduit à formuler le suivant problème B, ce qui constitue un résultat ayant une certaine importance pour orienter la recherche visant aux théorèmes de caractère topologique qui fournissent les renseignements sur le nombre de paramètres essentiels dont dépend la famille des intégrales possédant les propriétés asymptotique de differents types.

Problème B.

Indiquer tels conditions suffisantes I de caractère topologique qui se rapportent aux ensembles T et S et qui fournissent les renseignements sur le caractère topologique de l'ensemble $T - G$ et en particulier sur son nombre de dimensions (ou le renseignement que $T - G$ n'est pas vide), sans solution effective du système (1). Voici encore un problème analogue.

Problème C.

Soit $\overline{T} = T + Fr(T, W)$ et soit $Z \subset \overline{T}$. Il s'agit d'indiquer tels conditions topologiques K relatives aux ensembles T, S et Z qui sont suffisantes à ce que $Z(T - G) \neq 0$. Il s'agit des conditions qui ne nécessitent pas la solution effective du système (1).

Les conditions K intervenant dans le problème C seront évidemment suffisantes à ce qu'il existe un point $P \in ZT$, tel que $L^+(P) \subset T$, c'est-à-dire tel que la demi-intégrale $L^+(P)$ issue de P soit asymptotique par rapport à T. Si l'on indique sur un hyperplan $t = t_0$ deux ensembles disjoints Z_1 et Z_2 satisfaisant (avec S et T) aux conditions K il en résultera l'existence de deux demi-intégrales asymptotiques différentes entre elles au moins. Grâce à cette remarque on peut obtenir le renseignement sur le nombre de telles demi-intégrales et, dans certaines hypothèses, les renseignemens sur la structure de l'ensemble engendré par ces demi-intégrales.

Dans l'énoncé du Théorème D fournissant une condition K en question (v. plus bas) interviennent encore deux notions dont voici les définitions.

Rétraction. Rétracte.

Soit $A \subset B$. Une transformation $Q = U(P)$ continue dans B et telle que $U(P) \in A$ pour $P \in B$ et $U(P) = P$ pour $P \in A$ est dite rétraction de B en A. Si une telle transformation existe A est dit rétracte de B au sens de Borsuk.

La frontière d'une boule fermée à p dimensions n'est pas un rétracte de cette boule. Cette proposition (résultant facilement d'un théorème de Brouwer sur le point fixe d'une transformation continue) intervient fréquemment dans les applications du Théorème D qui suit.

Ensemble de sortie stricte S^*.

Nous dirons qu'un point $P \in Fr(T, W)$ est un point de sortie stricte (relatif à T et au système (1)) lorsqu'il est un point de sortie et lorsqu'il existe un $\epsilon > 0$, tel que $Y(t, P) \in W - \overline{T}$ pour $t_P < t < t_P + \varepsilon$. L'ensemble de tous les points de sortie stricte sera dit ensemble de sortie stricte et sera désigné par S^*. Cette notion constitue une modification de la notion de surface sans contact introduite par Poincaré.

Théorème D. Si $S = S^*$, $Z \subset T + S$, ZS est un rétracte de S et ZS n'est pas un rétracte de Z alors il existe un point $P \in ZT$, tel que $L^+(P_0) \subset T$, c'est-à-dire tel que la demi-intégrale $L^+(P_0)$ est asymptotique relativement à T et au système (1) (cf. [7]).

Démonstration. $C(P)$ désignant le conséquent de P la transformation $K(P) = C(P)$ pour $P \in G$ et $K(P) = P$ pour $P \in S$ est continue dans $G + S$, ce qui resulte de ce que $S = S^*$ (cf. [7]). Soit $R = V(Q)$ une rétraction de S en ZS. Supposons qu'un point en question P_0 n'existe pas. Il est facile de constater que la transformation $R = V(K(P))$ constitue une rétraction de Z en ZS, ce qui est contraire à l'hypothèse que ZS n'est pas un rétracte de Z.

Voici un exemple illustrant la liaison de ce théorème avec les problèmes asymptotiques au sens classique.

Exemple. Considérons le système de deux équations

$$x'(t) = f(t, x, y), \quad y'(t) = g(t, x, y) \tag{2}$$

Supposons que f et g soient continues partout et que par chaque point passe une intégrale unique de (2). Considérons le tuyau T défini par les inégalités

$$|x| < u(t), \ |y| < v(t), \ 0 < t < +\infty$$

où les fonctions $u(t)$ et $v(t)$ sont continues et positives. Supposons que les points (t, x, y) satisfaisant aux conditions

$$|x| = u(t), \ |y| < v(t), \ 0 < t < +\infty$$

soient points de sortie stricte (par rapport à T et au système (2)) et que les points (t, x, y) satisfaisant aux conditions

$$|x| \leqq u(t), \ |y| = v(t), \ 0 < t < +\infty$$

ne soient pas points de sortie.

Ces hypothèses sont en particulier satisfaites lorsque $u(t)$ et $v(t)$ possèdent les dérivées premières continues et lorsque

$$\left.\begin{aligned} xf(t,x,y) &> u(t)u'(t) \text{ pour } |x| = u(t),\ |y| \leqq v(t),\ 0 < t < +\infty \\ yg(t,x,y) &< v(t)v'(t) \text{ pour } |x| \leqq u(t),\ |y| = v(t),\ 0 < t < +\infty \end{aligned}\right\} \quad (3)$$

Comme Z nous prenons le segment $I(b)$ reliant les points

$$t = t_0 > 0,\ x = -u(t_0),\ y = b \text{ et } t = t_0,\ x = u(t_0),\ y = b$$

où $|b| < v(t_0)$. Dans ces hypothèses les prémisses du Théorème D se trouvent vérifiées. Sur chaque segment $I(b)$ existe donc au moins un point $P(b)$ tel que la demi-intégrale $x = x(t), y = y(t)$ passant par ce point existe dans l'intervalle $t_0 \leqq t < +\infty$ et parcourt totalement dans le tuyau T, c'est-à-dire satisfait aux inégalités

$$|x(t)| < u(t),\ |y(t)| < v(t) \text{ pour } t_0 \leqq t < +\infty.$$

En faisant varier la coordonée b de $P(b)$ on voit que l'ensemble B de tous ces points $P(b)$ a la puissance du continu au sens de Cantor.

Une liaison avec la théorie des dimensions de Urysohn et Menger.

Il résulte d'un théorème plus spécial que le Théorème D (et démontré au moyen d'un lemme prouvé à cet effet par M. M. K. Borsuk et K. Kuratowski) que le nombre de dimensions de l'ensemble B intervenant dans l'exemple ci-dessus est au moins égal à l'unité. La famille de demi-intégrales asymptotiques par rapport au tuyau T et au système (2) dépend donc d'un paramètre essentiel au moins [9].

Renseignement sur le nombre de paramètres essentiels dont dépend la famille des intégrales asymptotiquement bornées (au sens classique) et des intégrales tendant asymptotiquement vers un point limite. Il résulte de la remarque précédente qu'au cas où les fonctions $u(t)$ et $v(t)$ sont bornées (ou bien si elles tendent vers zéro lorsque $t \to +\infty$) alors il existe une famille de demi-intégrales dépendant d'un paramètre essentiel au moins qui existent et sont asymptotiquement bornées dans l'intervalle $t_0 \leqq t < +\infty$ (ou bien qui tendent vers le point (0,0) lorsque $t \to +\infty$).

Application à la solution d'un problème aux limites. Soit $J(c)$ le segment reliant les points

$$t = t_1 > t_0,\ x = c,\ y = -v(t_1) \text{ et } t = t_1,\ x = c,\ y = v(t_1)$$

où $|c| < u(t_1)$. En s'appuyant sur le théorème D on peut démontrer (cf. [7]) que pour tous les b et c pour lesquels $|b| < v(t_0)$, $|c| < u(t_1)$ il *existe* au moins une intégrale du système (2) rencontrant les segments $I(b)$ et $J(c)$ et telle que l'arc partiel de cette intégrale correspondant à l'intervalle $t_0 \leqq t \leqq t_1$ est situé dans le tuyau T.

Une difficulté liée avec l'application du Théorème D aux problèmes asymptotiques classiques.

L'exemple précédent est manifestement artificiel. En effet, le tuyau T était donné d'avance et ce n'est qu'ensuite (en un certain sens) que l'on a choisi le système (2) de façon que les prémisses du Théorème D soient remplies. Au cas des applications la situation est inverse: le système d'équations différentielles est donné d'abord et on doit ensuite construire le tuyau T d'une façon adaptée à la nature du problème asymptotique en question et à la forme particuliére du système envisagée. Cette tâche représente la difficulté la plus essentielle en ce qui concerne l'application du Théorème D et exige une analyse préalable bien attendive du problème (cf. par exemple [2]). Nous passons à une classe de problèmes où l'on peut procéder d'une façon plus méthodique pour construire un tuyau convenable T.

Méthode M de la comparaison de l'allure asymptotique des intégrales de deux systèmes d'équations différentielles. Application au cas des équations perturbées.

La construction d'un tuyau permettant d'appliquer le Théorème D se réduit, en général, à la recherche de certaines fonctions auxiliaires satisfaisant aux *inégalités différentielles* convenables. Ces fonctions jouent un rôle analogue que les fonctions $u(t)$ et $v(t)$ intervenant dans les inégalités différentielles (3). Supposons que le système (2) soit assez simple (par exemple linéaire aux coefficiens constants) et que les fonctions $u(t)$ et $v(t)$ vérifiant les inégalités (3) soient faciles à trouver. En ajoutant aux deuxièmes membres de (2) les perturbations p et q nous obtiendrons le système

$$x'(t) = f(t, x, y) + p(t, x, y), \quad y'(t) = g(t, x, y) + q(t, x, y) \tag{4}$$

Formons avec les deuxièmes membres de ce système les inégalités différentielles analogues aux inégalités (3) et cela avec *les mêmes fonctions* $u(t)$ et $v(t)$

$$x(f(t, x, y) + p(t, x, y)) > u(t)u'(t) \quad \text{pour} \quad |x| = u(t),\ |y| \leqq v(t),\ 0 < t < +\infty$$
$$y(g(t, x, y) + q(t, x, y)) < v(t)v'(t) \quad \text{pour} \quad |x| \leqq u(t),\ |y| = v(t),\ 0 < t < +\infty$$

Il est clair que ces inégalités auront lieu pour p et q suffisamment petits. Les prémisses du Théorème D seront donc aussi vérifiées relativement au système perturbé (4) pour les mêmes ensembles S,S^*, T et Z. Le système (4) admettra donc aussi une famille dépendant d'un paramètre essentiel au moins de demi-intégrales asymptotiques par rapport au même T.

De là résulte la suivante méthode M. Etant donné un système de deux èquations différentielles aux deuxièmes membres $F(t, x, y)$ et $G(t, x, y)$, on cherche de mettre F et G sous la forme $F = f + p$, $G = g + q$ de façon que le système (2) soit suffisamment simple. On construit ensuite T pour (2) et l'on examine si l'on peut procéder ensuite de la façon présentée ci-dessus. On peut

appliquer cette méthode pour comparer l'allure asymptotique d'un système de n èquations non linéaires avec celle d'un système linéaire (cf.p.ex. [5], [8]). On peut obtenir les limitations effectives pour la grandeur des perturbations admissibles. La méthode marche dans les conditions de régularité bien restreintes (par exemple unicité et l'existence de la différentielle de Stolz seulement au point singulier [8]).

Application à l'examen de la structure de l'ensemble engendré par les intégrales possédant une propriété asymptotique.

Il s'agit de savoir si cet ensemble a la forme d'une surface assez regulière. Pour appliquer le Théorème D on a besoin de trouver des conditions suffisantes pour que par les ensembles Z assez simples passe une demi-intégrale unique asymptotique par rapport à T (cf. [6]).

Théorème analogues au Théorème D. M. F. Albrecht en s'inspirant d'une sugestion de M. M. Ganea et Halanay a appliqué la notion du rétracte par déformation (Borsuk) pour construire un théorème analogue permettant de prouver l'existence des solutions périodiques au cas de deux équations. Son théorème qui n'est pas équivalent au Théorème D, n'est ni plus général ni plus particulier que le théorème D. Or M. A. Pliś a introduit une spécification de la notion du rétracte (rétracte par déformation quasi-isotope) et a obtenu un théorème [1]) englobant le Théorème D et celui de M. Albrecht comme cas particuliers. Il s'applique même au cas où Z est composé de deux circonférences entrelacées. Les articles en question sont en train d'impression ([1], [4]).

BIBLIOGRAPHIE.

[1] ALBRECHT F. Remarques sur un théorème de T. WAŻEWSKI relatif à l'allure asymptotique des intégrales des équations différentielles, (Bull. Acad. Polon. des Sciences Cl. III—Vol. II, No. 7, 1954.)

[2] MIKOŁAJSKA Z. Sur les mouvements asymptotiques d'un point matériel mobile dans le champ des forces repoussantes, (Ibidem, Vol. I, Nos 1—2, 1953.)

[3] MIKOŁAJSKA Z. Sur une propriété asymptotique des intégrales d'une équation différentielle du second ordre, (Ibidem, Vol. II, No 3, 1954.)

[4] PLIŚ A. On a topological method of the study of the behaviour of the integrals of ordinary differential equations, (Ibidem, Vol. II, No 9, 1954.)

[1]) Désignons par K l'intervalle fermé $0 \leqq u \leqq 1$ et par $B \times K$ le produit cartésien de B et K. Nous dirons que A est un $RQI(B, E)$(A est un rétracte par déformation quasi-isotope de B dans E) lorsqu'il existe une fonction $G(P, u)$ telle que $G(P, u) \in E$ pour $(P, u) \in B \times K$, $G(P, 0) = P$ pour $P \in B$, $G(P, u) = P$ pour $(P, u) \in A \times K$, $G(P, 1) \in A$ pour $P \in B$ et lorsque la transformation $N = V_u(P) = G(P, u)$ constitue un homéomorphisme pour P variable dans B lorsque $0 \leqq u < 1$ (le cas $u = 1$ exclu). Voici l'énoncé du théorème en question de M. A. Pliś: Si l'on admet l'Hypothèse H relativement au système (1) et si $S = S^*$, $Z \subset T + S$, $Z \neq 0$, S n'est pas un $RQI(Z + S, T + S)$ alors il existe un $P_0 \in ZT$ tel que $L^+(P_0) \subset T$.

[5] SZMYDTÓWNA Z. Sur l'allure asymptotique des intégrales des équations différentielles ordinaires, (Ann. Soc. Polon. Math. T. XXIV, année 1951.)

[6] SZMYDT Z. Sur la structure de l'ensemble engendré par les intégrales tendant vers le point singulier du système d'équations différentielles, (Bull. Acad. Polon. des Sciences Cl. III—Vol. I, No 6, 1953.)

[7] WAŻEWSKI T. Sur un principe topologique de l'examen de l'allure asymptotique des intégrales des équations différentielles ordinaires, (Ann. Soc. Polon. de Math. T. XX, année 1947.)

[8] WAŻEWSKI T. Sur les intégrales d'un système d'équations différentielles tangentes aux hyperplans caractéristiques issues du point singulier, (Ibidem, T. XXI, année 1948.)

[9] WAŻEWSKI T. Sur l'évaluation du nombre des paramètres essentiels dont dépend la famille des intégrales d'un systéme d'équations différentielles ayant une propriété asymptotique, (Bull. Acad. Polon. des Sciences Cl. III—Vol. I, Nos 1—2, 1953.)

BOHATERÓW STALINGRADU 77.M.8.,
KRAKÓW, POLOGNE.

HILBERT TRANSFORMS IN E^n

A. ZYGMUND

In this lecture I present certain results obtained by Professor A. P Calderón and myself within the last three years. [1])

1. *Hilbert transforms.* Let $f(x)$ be a function defined for all real x and belonging to one of the classes L^p $(-\infty, +\infty)$, $1 \leqq p < \infty$. By the *Hilbert transform* of f we mean the function

$$\tilde{f}(x) = \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{f(t)}{x-t}\, dt, \tag{1.1}$$

where the integral is taken in the principal — value sense:

$$\frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{f(t)}{x-t}\, dt = \lim_{\varepsilon \to 0} \int_{|x-t|>\varepsilon} \frac{f(t)}{x-t}\, dt.$$

It is very well known that the integral (1.1) exists almost everywhere. For $1 < p < \infty$ we also have the M. Riesz inequality

$$\|\tilde{f}\|_p \leqq A_p \|f\|_p, \tag{1.2}$$

where A_p depends on p only, and $\|f\|_p = \left(\int_{-\infty}^{+\infty} |f|^p dx\right)^{1/p}$. For $p = 1$, (1.2) does not hold.

The inequality (1.2) can be strengthened in a way which will be useful later. Let

$$\tilde{f}_*(x) = \sup_{\varepsilon} \left| \frac{1}{\pi} \int_{|x-t|>\varepsilon} \frac{f(t)}{x-t}\, dt \right|.$$

Then again

$$\|\tilde{f}_*\|_p \leqq A_p \|f\|_p \tag{1.3}$$

for $1 < p < \infty$. [2])

We shall now discuss an extension of the notion of Hilbert's transform to functions of several variables.

---

[1]) Some of the results have already been published; see papers **1**, **2**, **3**, **4** of the bibliography. For older literature see **5**.

[2]) The corresponding result for periodic functions is proved, by complex methods, in **6**, pp. 249—250, and the proof for f nonperiodic is similar. Another proof is given in **1**, p. 116.

We consider the n-dimensional Euclidean space E^n. The points of this space will be denoted by $x = (\xi_1, \xi_2, \ldots, \xi_n)$, $y = (\eta_1, \eta_2, \ldots, \eta_n)$, ... etc. By definition, $0 = (0, 0, \ldots, 0)$. We shall also treat E^n as a vector space by identifying the point x with the vector $0x$. Addition of vectors and multiplication of vectors by scalars are defined in the usual way, and we write $|x| = (\xi_1^2 + \xi_2^2 + \ldots + \xi_n^2)^{\frac{1}{2}}$. By L^p we mean the set of measurable functions $f(x)$ such that $\|f\|_p = \left(\int_{E^n} |f|^p dx\right)^{1/p} < \infty$, $dx = d\xi_1 \ldots d\xi_n$ denoting the element of volume in E^n.

By Σ we mean the unit sphere $|x| = 1$. By x' we systematically denote the projection of x onto Σ, that is the intersection of Σ with the ray $0x$.

The Hilbert transform (1.1) on the straight line is the convolution of f wish the kernel $K(x) = 1/\pi x$. The Hilbert transform of $f(x)$ in E^n is defined as the convolution of f with a certain kernel K:

$$\tilde{f} = f * K = \int_{E^n} f(y)\, K(x-y)dy. \tag{1.4}$$

The kernels we consider are suggested by various problems of the theory of the potential and, in the most important cases, are of the form

$$K(x) = \frac{\Omega(x')}{|x|^n},$$

where Ω is a function defined on the unit sphere Σ. The kernel K is in general non-integrable near $x = 0$, and the Hilbert transform (1.4) is taken in the principal-value sense

$$\tilde{f}(x) = \lim_{\varepsilon \to +0} \int_{|x-y|>\varepsilon} f(y)\, K(x-y)dy.$$

Of course the integral under the limit sign on the right is improper, but under very general conditions (for example, if Ω is bounded) converegs absolutely for each $f \in L^p$, $1 \leqq p < \infty$ and each x.

We impose two conditions on the kernel K:

(i) The integral of Ω extended over Σ is zero;

(ii) Ω satisfies a Lipschitz condition of positive order, *i. e.*

$$|\Omega(x') - \Omega(y')| \leqq A\, |x' - y'|^{\alpha}.$$

Condition (ii) can in some cases be considerably relaxed (see § 2 below), but this matter, though of considerable intrinsic interest, is not of primary importance in applications, where Ω is usually a spherical harmonic. On the other hand, condition (i) is essential. If it is not satisfied, then we easily see that for any function f constant but not zero in a sphere D the integral (1.4) diverges to ∞ at every point of D. Condition (i) is satisfied in the case $n = 1$,

$K(x) = 1/x$, since the sphere Σ then consists of the points $x' = \pm 1$, and $\Omega(x') = \text{sign } x'$.

While for $n = 1$ we have essentially only one Hilbert transform, in higher dimensions we have infinitely many, each characterized by the corresponding Ω. If $n = 2$, x' is an angle between 0 and 2π, and condition (i) means that the constant term of the Fourier series of $\Omega(x')$ is zero.

A few special cases deserve mention. If $n = 2$ and $\Omega(\varphi) \sim \Sigma' c_n e^{in\varphi}$ (the prime indicating that $c_0 = 0$), then

$$K(z) = \Sigma' c_k \frac{e^{ik\varphi}}{r^2},$$

where $z = r\, e^{i\varphi}$ is a complex number. This shows the particular importance of the special kernels

$$\frac{e^{ik\varphi}}{r^2} = \frac{z^k}{|z|^{k+2}}, \tag{1.5}$$

and among them of the kernels

$$z/|z|^3 \tag{1.6}$$

and

$$1/z^2. \tag{1.7}$$

The latter kernel is the only one among (1.6) which is analytic. It was considered in unpublished work of Beurling and we shall call it the *Beurling kernel.*

The kernel (1.7) is a special case of a kernel considered for general n by M. Riesz and Horwath, and defined by the formula

$$K(x) = \frac{x}{|x|^{n+1}}$$

This K is a vector kernel.

Theorem 1. *Suppose that K satisfies conditions (i) and (ii) above and write*

$$\tilde{f}_\varepsilon(x) = \int_{|x-y|>\varepsilon} f(y)\, K(x-y)dy.$$

Suppose that $f \in L^p$, $1 \leqq p < \infty$. Then

$$\tilde{f}(x) = \lim_{\varepsilon \to 0} \tilde{f}_\varepsilon(x)$$

exists almost everywhere. If $1 < p < \infty$, then also

$$||\tilde{f}_\varepsilon||_p \leqq A_p\, ||f||_p, \quad ||\tilde{f}||_p \leqq A_p\, ||f||_p, \tag{1.8}$$

$$\lim_{\varepsilon \to +0} ||f - \tilde{f}_\varepsilon||_p = 0, \tag{1.9}$$

(1.10) $$\|\tilde{f}_*\|_p \leqq A_p \|f\|_p,$$

where $$\tilde{f}_*(x) = \sup_{\varepsilon>0} \left| \int_{|x-y|>\varepsilon} f(y)\, K(x-y)dy \right|.$$

The constants A_p depend on p and the kernel K only.

The proof of this and of other results of a similar nature can befound in **1**. Here we only mention (and the result is of importance in the theory of the potential) that if $F(e)$ is any countably additive mass distribution in E^n, then the Stieltjes integral

$$\int_{E^n} K(x-y)\, F(dy)$$

converges almost everywhere.

2. *A partial generalization of Theorem* 1. We call the function $\Omega(x')$ *even* if $\Omega(-x') = \Omega(x')$, and *odd* if $\Omega(-x') = -\Omega(x')$. The formula

$$\Omega(x') = \tfrac{1}{2}\{\Omega(x') + \Omega(-x')\} + \tfrac{1}{2}\{\Omega x') - \Omega(-x')\}$$

decomposes the kernel K into a sum of two analogous kernels with numerators respectively even and odd.

Theorem 2. *Suppose that Ω is odd that $\Omega(x')$ is integrable over Σ (in particular, the integral of Ω over Σ is zero). Suppose that $f \in L^p$, where p is strictly greater than* 1. *Then the integral*

$$\tilde{f}(x) = f*K$$

converges almost everywhere. Moreover the function $\tilde{f}_*$ of Theorem 1 *satisfies the inequality* (1.10).

Hence for Ω odd and $f \in L^p$, $p > 1$, condition (ii) is superfluous. The proof is a comparatively simple deduction from results in the one dimensional case. We give it here.

Let t be apoint on Σ; thus t is also a unit vector. Let L be any straight line parallel to t. Let $\epsilon = \epsilon(x)$ be a positive (and measurable) function of x. For any point x on L consider the expression

$$g_\varepsilon(x,\ t) = \int_{|\xi|>\varepsilon} \frac{f(x-\xi t)}{\xi}\, d\xi = \int_{\varepsilon}^{\infty} \frac{f(x-\xi t) - f(x-\xi t)}{\xi}\, d\xi,$$

which, for $\epsilon \to 0$, tends to the one-dimensional Hilbert transform of the function f confined to the line L. By (1.3),

$$\int_L |g_\varepsilon(x,\ t)|^p dx \leqq A_p^p \int_L |f(x)|^p dx.$$

Imagine that the direction t, and so also line L, is parallel to one of the co-

ordinate axes. Integrating with respect to the remaining coordinates we get

$$\int_{E^n} |g_\varepsilon(x,t)|^p dx \leqq A_p^p \int_{E^n} |f(x)|^p dx$$

or, briefly,

$$||g_\varepsilon(x,t)||_p \leqq A_p^p \, ||f(x)||_p.$$

where t is any fixed unit vector.

It is easy to relate the functions $\tilde{f}_\varepsilon(x)$ and $g_\varepsilon(x,t)$. For, first of all,

$$\tilde{f}_\varepsilon(x) = \int_{|y|>\varepsilon} f(x-y)\, K(y) dy = \int_{|y|>\varepsilon} f(x-y)\, \frac{\Omega(y')}{|y|^n}\, dy.$$

We now set $y = t\xi$, where t is an arbitrary unit vector and $\xi > 0$. In the last integral we integrate first on concentric spheres with radii ξ, and then integrate with respect to ξ over the interval (ϵ, ∞). The integral becomes

$$\int_\varepsilon^\infty \int_\Sigma f(x-t\xi)\, \frac{\Omega(t)}{\xi^n}\, \xi^{n-1} d\xi dt = \int_\Sigma \Omega(t) \left\{ \int_\varepsilon^\infty \frac{f(x-t\xi)}{\xi}\, d\xi \right\} dt. \tag{2.1}$$

If we replace here t by $-t$ and use the fact that Ω is odd we obtain

$$-\int_\Sigma \Omega(t) \left\{ \int_\varepsilon^\infty \frac{f(x+t\xi)}{\xi}\, d\xi \right\} dt \tag{2.2}$$

Taking the semi-sum of the expressions we obtain

$$\tilde{f}_\varepsilon(x) = \tfrac{1}{2} \int_\Sigma \Omega(t) g_\varepsilon(x,t) dt,$$

a result which may be summarized as follows: if Ω is odd, then for fixed x the integral $\tilde{f}_\varepsilon(x)$ is the average of $g_\varepsilon(x,t)$ over the unit sphere Σ, with weight function Ω. From Minkowski's inequality

$$|| \int \lambda || \leqq \int || \lambda ||$$

we deduce

$$||\tilde{f}_\varepsilon(x)||_p \leqq \tfrac{1}{2} || \int_\Sigma \Omega(t) g_\varepsilon(x,t) dt || \leqq \tfrac{1}{2} \int_\Sigma |\Omega(t)|\, ||g_\varepsilon(x,t)||_p dt \leqq \tfrac{1}{2} A_p || f ||_p \int_\Sigma |\Omega|\, dt,$$

or briefly

$$||\tilde{f}_\varepsilon(x)||_p \leqq B_p \, ||f||_p \, ||\Omega||_1,$$

where B_p depends on p only.

By taking suitable $\epsilon = \epsilon(x)$ we are immediately led to (1.10). The existence of $\tilde{f}$ for almost all x is a consequence of (1.10), (and so are the inequalities (1.8) and the formula (1.9)).

The conclusion of Theorem 2 no longer holds if Ω is even. The sitution is then somewhat different and the proof much less simple. The following theorem is stated without proof.

Theorem 3. *Suppose that* $\Omega(x')\log^+|\Omega(x')|$ *is integrable over* Σ *and that the integral of* Ω *over* Σ *is zero. The conclusions of Theorem* 1 *then hold for any* $f \in L^p$, $p > 1$.

Examples show that the integrability of $|\Omega|\log^+|\Omega|$ is the best possible assumption and cannot be relaxed. In particular, the mere integrability of Ω is not enough.

The proofs of Theorems 2 and 3 break down when $p = 1$, so that these theorems are not a complete generalization of Theorem 1.

In certain problems we have to consider convolutions

$$\int_{E_n} f(x-y)\frac{\Omega(y')}{|y|^n}dy$$

where Ω depends not only on y' but also on x, *i.e.* $\Omega = \Omega_x(y')$. Going through the proof of Theorem 2 we see that it holds in this case if $\Omega_x(y')$ is odd in y', and if

$$\sup_x |\Omega_x(y')|$$

is integrable over Σ (in particular, if Ω is bounded in x and y').

3. *The periodic case.* The notion of a Hilbert transform can be extended, properly modified, to the case of periodic functions. In the one-dimensional case we have the *conjugate function*

$$f^*(x) = \frac{1}{\pi}\int_{-\pi}^{\pi} f(y)\ \tfrac{1}{2}\cot\ \tfrac{1}{2}(x-y)dy,$$

defined for f of period 2π and integrable.

Consider in E^n an orthogonal system of axes, and let $e_1, e_2, \ldots e_n$ be vectors of length 2π situated on the axes. Let

$$x_0 = 0,\ x_1, \ldots,\ x_k, \ldots$$

be the sequence of all lattice points generated by these vectors:

$$x_k = \mu_1 e_1 + \mu_2 e_2 + \ldots + \mu_n e_n,$$

where $k = 0, 1, \ldots$, and the μ's run through all integers, positive, negative and zero. Given any kernel $K(x)$ satisfying conditions (i) and (ii) above, we set

$$K^*(x) = K(x) + \sum_{k=1}^{\infty}\{K(x+x_k) - K(x_k)\}. \tag{3,1}$$

The series on the right is absolutely and uniformly convergent over any finite

sphere. If $n = 1$, $K(x) = 1/x$, we obtain for $K^*(x)$ the function $\frac{1}{2}\cot\frac{1}{2}x$; if $n = 2$ and $K(z)$ is the Beurling kernel $1/z^2$, $K^*(z)$ is the Weierstrass function $\wp(z)$, etc.

The function $K^*(x)$ is periodic and has all the lattice vectors x_k as periods. Let $f(x) = f(\xi_1, \xi_2, \ldots, \xi_n)$ be any locally integrable function, of period 2π in each ξ_j. The convolution

$$(3.2) \qquad (2\pi)^{-n} \int_R f(y)\, K^*(x - y)dy$$

of f and K^* will be called the *conjugate function of* f, with respect to the kernel K^*, and denoted by $f^*(x)$. R denotes the fundamental cube $|\xi_j| \leqq \pi$, $j = 1, 2, \ldots, n$.

We see from (3.1) that $K^* - K$ is bounded in R. It follows from the corresponding results for $\tilde{f}$ that f^* exists almost everywhere. The following result is an easy consequence of Theorem 1.

Theorem 4. *If f is in $L^p(R)$ and $p > 1$, then f^* is also in $L^p(R)$ and*

$$\int_R |f(x)|^p dx \leqq A_p^p \int_R |f(x)|^p dx.$$

Consider the Fourier coefficients

$$c_m = (2\pi)^{-n} \int_R f(x) e^{-i(m,x)}\, dx,$$

and the Fourier series

$$\Sigma\, c_m e^{i(m,x)}$$

of $f(x)$. Here $m = (\mu_1, \mu_2, \ldots, \mu_n)$ is an integral lattice point and $(m, x) = \mu_1\xi_1 + \ldots + \mu_n\xi_n$. Should the conjugate function f^* be integrable we denote its Fourier coefficients by c_m^*, so that

$$f^*(x) \sim \Sigma\, c_m^*\, e^{i(m,x)}$$

Denote by γ_m the Fourier coefficients of the (periodic) fuction K^*:

$$\gamma_m = (2\pi)^{-n} \int_R K^*(x) e^{-i(m,x)} dx.$$

These integrals exist in the principal — value sense, and a simple and classical argument shows that for $m \neq (0, 0, \ldots, 0)$ the Fourier coefficients of K^* coincide with the Fourier transform

$$\hat{K}(m) = (2\pi)^{-n} \int_R K(x) e^{-i(m,x)}\, dx$$

of K.

Since f^* is a convolution of K^* and f, we can anticipate that $c_m^* = \gamma_m c_m$, and the result is actually not difficult to prove. In particular,

$$c_m^* = \hat{K}(m)c_m \qquad (m \neq 0$$

From this and Theorem 4 we deduce the following result:

Theorem 5. *Let $\Sigma\, c_m e^{i(m,x)}$ be the Fourier series of an f in L^p, $p > 1$. Then $\Sigma'\, c_m \hat{K}(m) e^{i(m,x)}$, where the prime denotes that the term $m = 0$ is omitted in summation, is the Fourier series of a function g such that*

$$\int_R |g|^p dx \leqq A_p^p \int_R |f|^p dx.$$

4. *Discrete Hilbert transforms.* A discrete analogue of the transform (1.1) can be formulated as follows. Let

$$X = (\ldots, \xi_{-1}, \xi_0, \xi_1, \ldots, \xi_\nu, \ldots)$$

be a two-way infinite sequence of complex numbers, and let us define the sequence

$$\tilde{X} = (\ldots, \tilde{\xi}_{-1}, \tilde{\xi}_0, \tilde{\xi}_1, \ldots, \tilde{\xi}_\mu, \ldots)$$

by the equations

$$\tilde{\xi}_\mu = \sum_{\nu \neq \mu} \frac{\xi_\nu}{\mu - \nu}.$$

Then

$$(\Sigma\, |\tilde{\xi}_\mu|^2)^{\frac{1}{2}} \leqq A\, (\Sigma |\xi_\nu|^2)^{\frac{1}{2}}, \tag{4.1}$$

or

$$|| \tilde{X} ||_2 \leqq A\, || X ||_2,$$

if we adopt the abbreviation

$$|| X ||_p = (\Sigma |\xi_\nu|^p)^{1/p}.$$

This result of Hilbert and Toeplitz was generalized by M. Riesz who showed that

$$||\tilde{X}||_p \leqq A_p ||X||_p \tag{4.2}$$

for every $p > 1$, where A_p depends on p only. Theorem 1 leads to a generalization of M. Riesz' result. We state this generalization without proof.

Suppose that the kernel K satisfies conditions (i) and (ii) above. Let $e_1, e_2, \ldots, e_n$ be a system of n linearly independent vectors in E^n, and let $x_0 = 0, x_1, x_2, \ldots$ be the sequence of all lattice points in E^n generated by this system. Let $X = (\xi_0, \xi_1, \ldots)$ be any sequence of complex numbers. We define the transform $\tilde{X} = (\tilde{\xi}_0, \tilde{\xi}_1, \ldots)$ of X by the equations

$$\tilde{\xi}_\mu = \sum_{\nu \neq \mu} \xi_\nu K(x_\mu - x_\nu). \tag{4.3}$$

Theorem 6. *The transform* (4.3) *of* X *satisfies* (4.2), *where* A_p *depends on on* p *and* K.

5. *Radial kernels.* Certain problems of Analysis lead to kernels which are of a different type from Hilbert's though have close relation with the latter. Such kernels occur already in the one-dimensional case.

Theorem 7. *Let* $h(t)$ *be a Fourier-Stieltjes transform on the straight line,*

$$(h(t) = \int_{-\infty}^{+\infty} e^{itu} d\gamma(u), \tag{5.1}$$

where

$$\int_{-\infty}^{+\infty} |d\gamma(u)| = V < \infty.$$

Then for every f *of the class* $L^p(-\infty, +\infty)$, $p > 1$, *the integral*

$$F(x) = \int_{-\infty}^{+\infty} f(x+t) \frac{h(t)}{t} dt$$

converges almost everywhere and

$$||F||_p \leqq A_p V ||f||_p. \tag{5.2}$$

By (1.3) we have

$$\left\| \int_{|t|>\varepsilon} f(x+t) \frac{dt}{t} \right\|_p \leqq A_p ||f||_p,$$

where ϵ depends on x. Apply this not to $f(x)$ but to $f(x)\, e^{ixu}$, where u is fixed. We obtain

$$\left\| \int_{|t|>\varepsilon} f(x+t) \frac{e^{itu}}{t} dt \right\|_p \leqq A_p ||f||_p,$$

and, by Minkowski's inequality,

$$\left\| \int_{-\infty}^{+\infty} d\gamma(u) \int_{|t|>\varepsilon} f(x+t) \frac{e^{itu}}{t} dt \right\|_p \leqq A_p V ||f||_p.$$

If ε stays above a positive number ε_0, we can interchange the order of summation and integration and get

$$\left\| \int_{|t|>\varepsilon} f(x+t) \frac{h(t)}{t} dt \right\|_p \leqq A_p V ||f||_p.$$

Making now ε_0 tend to 0 we get the inequality when ε is any positive measurable function of x. From this the existence of $F(x)$, as well as the inequality (5.2), follows in a straightforward fashion.

To the integral $\xi(x)$ so obtained we may apply the argument of Section 2. It works if we assume that $h(t)$ is an *odd* function of t, and we get the following result.

Theorem 8. *Suppose that $f \in L^p(E^n)$, $p > 1$. Let $h(t)$, given by (5.1), be an odd function of t, and let $\Omega(x')$ be even and intebrable over Σ. Then the integral*

$$\int_{E^n} f(x+y) \frac{h(|y|)}{|y|^n} \Omega(y')dy$$

converges almost everywhere and its value $\tilde{f}(x)$ satisfies

$$||\tilde{f}||_p \leqq A_p V ||\Omega||_1 \, ||f||_p.$$

The same conclusion holds if Ω is odd and h even. The case $\Omega \equiv 1$ has some interesting applications. To arrive at them we first consider a problem in the one-dimensional case.

Suppose that $f(x) \in L^p(-\infty, +\infty)$, $p > 1$. The problem of the representation of f by its Fourier integral reduces to showing that the Dirichlet integral

$$f_w(x) = \frac{1}{\pi} \int_{-\infty}^{+\infty} f(x+t) \frac{\sin wt}{w} dt$$

converges to $f(x)$ as $w \to \infty$. We may consider here either pointwise convergence or convergence in L^p. As regards the former, no general results can be obtained unless we consider the limit of $f_w(x)$ by the method of arithemetic means of positive order. It is therefore quite remarkable (the result is due to M. Riesz) that, as regards convergence in L^p, the last integral does actually tend to f:

$$||f(x) - f_w(x)||_p \to 0,$$

as $w \to \infty$. That the norm $||f_w(x)||_p$ remains bounded as $w \to \infty$, follows immediately from Theorem 8. For if we set $h(t) = \sin wt$, then $h(t)$ is a Fourier-Stieltjes transform:

$$2i \sin wt = \int_{-\infty}^{+\infty} e^{itu} d\gamma(u),$$

where $\gamma(u)$ is a step-function having the jumps ± 1 at the points $\pm w$ as the only discontinuities. Thus the total variation of γ is 2 and, by Theorem 8, $||f_w(x)||_p \leqq A_p ||f||_p$. The refinement to $||f - f_w||_p \to 0$ is straightforward.

The corresponding problem in n dimensions is more difficult. As regards the so called spherical means of the Fourier integral, it was shown by Bochner that for pointwise representations we must apply not the ordinary convergence but summability (C, α) where

$$\alpha > \lambda = \tfrac{1}{2}(n-1),$$

and, as in the case $n = 1$, this result cannot be improved. The question arises *whether the* (C, λ) *means converge to* f *in* L^p. It turns out that for n odd, $n = 1, 3, 5, \ldots$, we are led to integrals covered by the last theorem, and the problem does admit of an affirmative answer. Unfortunately the kernels which occur for n even are of somewhat different nature and the problem in that case is still open.

6. *Applications to potentials.* Suppose for simplicity that $n \geqq 3$ and consider the potential

$$u(x) = u(\xi_1, \ldots, \xi_n) = \int_{E^n} |x—y|^{2-n} f(y) dy$$

generated by a function f which, say, vanishes in the neighbourhood of infinity. The problem of the existence of the second derivatives, ordinary or generalized, of $u(x)$ is classical and has close connection with the existence of Hilbert transforms in E^n. It turns out that for the existence of *ordinary* second derivatives of u it is enough to assume that $|f| \log^+ |f|$ is integrable. The following theorem gives a somewhat stronger result.

Theorem 9. *If* $|f| \log^+ |f|$ *is integrable, then in almost every subspace* $\xi_3 = \xi_3^0, \ldots, \xi_n = \xi_n^0$, *the function* $u(\xi_1, \xi_2, \xi_3^0, \ldots, \xi_n^0)$ *is a continuous, indeed absolutely continuous, function of the variables* ξ_1, ξ_2, *and has a second differential almost everywhere in* ξ_1, ξ_2.

We say that a function $v(\xi_1, \xi_2)$ has a second differential at a point ξ_1^0, ξ_2^0 if $v(\xi_1^0 + h, \xi_2^0 + k) - v(\xi_1^0, \xi_2^0)$ is equal to

$$(\alpha_1 h + \alpha_2 k) + \tfrac{1}{2}(\alpha_{11} h^2 + 2\alpha_{12} hk + \alpha_{22} k^2) + o\{(h^2 + k^2)^{\frac{1}{2}}\}$$

for h and k tending to 0, $\alpha_1, \ldots, \alpha_{22}$ denoting constants.

Theorem 9 implies that under its hypotheses all second derivatives $u_{\xi_i \xi_k}$ exist almost everywhere, and it can easily be shown that they are given by the classical formulas. Results analogous to Theorem 9 but pertaining to first derivatives can be obtained for potentials of single or double layer of masses distributed on hyperplanes in E^n.

Bibliography

[1] A. P. Calderón and A. Zygmund, On the existence of certain integrals, Acta Math. 88 (1952), 85—139.

[2] A. P. Calderón and A. Zygmund, On a problem of Mihlin, Transactions of the American Math. Soc. 78 (1955), 209—224.

[3] A. P. Calderón and A. Zygmund, Singular integrals and periodic functions, Studia Math., 142 (1955).

[4] A. ZYGMUND. On the existence and properties of certain singular integrals, Lectures at the Kingston meeting of the American Math. soc., August 1953, mimeographed notes.

[5] S. MIHLIN, Singular integral equations, Uspekhi Mat. Nauk, 3 (1948) 29—112 (in Russian).

[6] A. ZYGMUND, Trigonometrical series, Warszawa, 1935.

Section III

GEOMETRY AND TOPOLOGY

STATED ADDRESSES

REGULAR HONEYCOMBS IN HYPERBOLIC SPACE

H. S. M. COXETER

1. *Introduction.* Schlegel (1883, pp. 444, 454) made a study of honeycombs whose cells are equal regular polytopes in spaces of positive, zero, and negative curvature. The spherical and Euclidean honeycombs had already been described by Schläfli (1855), but the only earlier mention of the hyperbolic honeycombs was when Stringham (1880, pp. 7, 12, and *errata*) discarded them as "imaginary figures", or, for the two-dimensional case, when Klein (1879) used them in his work on automorphic functions. Interest in them was revived by Sommerville (1923), who investigated their metrical properties.

The honeycombs considered by the above authors have finite cells and finite vertex figures. It seems desirable to make a slight extension so as to allow infinite cells, and infinitely many cells at a vertex, because of applications to indefinite quadratic forms (Coxeter and Whitrow 1950, pp. 424, 428) and to the close packing of spheres (Fejes Tóth 1953, p. 159). However, we shall restrict consideration to cases where the fundamental region of the symmetry group has a finite content, like that of a space group in crystallography. This extension increases the number of three-dimensional honeycombs from four to fifteen, the number of four-dimensional honeycombs from five to seven, and the number of five-dimensional honeycombs from zero to five.

A further extension allows the cell or vertex figure to be a *star*-polytope, so that the honeycomb covers the space several times. Some progress in this direction was made in two earlier papers: one (Coxeter 1933), not insisting on finite fundamental regions, was somewhat lacking in rigour; the other (Coxeter 1946) was restricted to two dimensions. The present treatment is analogous to § 14.8 of *Regular Polytopes* (Coxeter 1948, p. 283). We shall find that there are four regular star-honeycombs in hyperbolic 4-space, as well as two infinite families of them in the hyperbolic plane.

2. *Two-dimensional honeycombs.* In the Euclidean plane, the angle of a regular p-gon, $\{p\}$, is $(1 - 2/p)\pi$. In the hyperbolic plane it is smaller, gradually decreasing to zero when the side increases from 0 to ∞. Hence, if p and q are positive integers satsifying

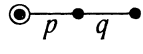
$$(p - 2)(q - 2) > 4, \tag{2.1}$$

we can adjust the size of the polygon so as to make the angle $2\pi/q$. Then q

such $\{p\}$'s will fit together round a common vertex, and we can add further $\{p\}$'s indefinitely. In this manner we construct a two-dimensional honeycomb or *tessellation* $\{p, q\}$, which is an infinite collection of regular p-gons, q at each vertex, filling the whole hyperbolic plane just once (Schlegel 1883, p. 360). We call $\{p\}$ the *face*, $\{q\}$ the *vertex figure*. The centres of the faces of $\{p, q\}$ are the vertices of the *reciprocal* (or dual) tesselation $\{q, p\}$, whose edges cross those of $\{p, q\}$. The simplest instances, $\{7, 3\}$ and $\{3, 7\}$, are shown conformally in Figs. 1 and 2.

As limiting cases, we admit $\{\infty, p\}$, whose faces $\{\infty\}$ are inscribed in horocycles instead of finite circles, and its reciprocal $\{p, \infty\}$, whose vertices are all at infinity (i.e., on the absolute conic).

The lines of symmetry, in which $\{p, q\}$ reflects into itself, are its edges (produced) and the lines of symmetry of its faces. They form a network of congruent triangles whose angles are π/q (at a vertex of $\{p, q\}$), $\pi/2$ (at the mid-point of an edge), and π/p (at the centre of a face). The symmetry group is generated by reflections in the sides of such a *characteristic* triangle $P_0P_1P_2$. Alternatively, we may begin with the triangle and derive $\{p, q\}$ by Wythoff's construction (Coxeter 1948, p. 87) as indicated by the symbol



This means that the vertices of $\{p, q\}$ are the images of the vertex P_0 (where the angle is π/q) in the kaleidoscope formed by mirrors along the three sides of the triangle. (The nodes of the graph represent mirrors, those not directly joined being at right angles.)

For some instances of the network of characteristic triangles, see Klein (1879, p. 448), Fricke (1892, p. 458), and Coxeter (1939, pp. 126, 127).

The classical formulae for a right-angled triangle (Coxeter 1947, p. 238) enable us to compute the sides

$$\varphi = P_0P_1, \qquad \chi = P_0P_2, \qquad \psi = P_1P_2$$

in the form

$$\cosh\varphi = \cos\frac{\pi}{p} \Big/ \sin\frac{\pi}{q}, \quad \cosh\chi = \cot\frac{\pi}{p}\cot\frac{\pi}{q}, \quad \cosh\psi = \cos\frac{\pi}{q} \Big/ \sin\frac{\pi}{p}$$

(Sommerville 1923, p. 86; cf. Coxeter 1948, pp. 21, 64). Then we may describe $\{p, q\}$ as a tesselation of edge 2φ, whose faces are p-gons of circum-radius χ and in-radius ψ.

3. *Three-dimensional honeycombs.* When

$$(p - 2)(q - 2) < 4,$$

the symbol $\{p, q\}$ denotes a Platonic solid (e.g., $\{4, 3\}$ is a regular hexahedron)

which exists not only in Euclidean space but equally well in hyperbolic space. When the edge increases from 0 to ∞, the dihedral angle decreases from its Euclidean value

$$2 \arcsin \left(\cos \frac{\pi}{q} / \sin \frac{\pi}{p} \right)$$

to $(1 - 2/q)\pi$. For, the solid angle at a vertex resembles a Euclidean solid angle in that its section by a sphere is a spherical q-gon, whose angle must exceed $(1 - 2/q)\pi$.

Thus the dihedral angle can take the value $2\pi/r$ whenever p, q, r are integers, greater than 2, satisfying

3.1 $$\sin \frac{\pi}{p} \sin \frac{\pi}{r} < \cos \frac{\pi}{q}$$

and $(q - 2)(r - 2) < 4$. Then r such $\{p, q\}$'s will fit together round a common edge, and we can add further $\{p, q\}$'s indefinitely. In this manner we construct a three-dimensional honeycomb $\{p, q, r\}$, which is an infinite collection of $\{p, q\}$'s, r at each edge, filling the whole hyperbolic space just once. The arrangement of cells round a vertex is like the arrangement of faces of the polyhedron $\{q, r\}$, which is called the *vertex figure*. The centres of the cells of $\{p, q, r\}$ are the vertices of the *reciprocal* honeycomb $\{r, q, p\}$, whose edges cross the $\{p\}$'s of $\{p, q, r\}$.

The actual instances are

$$\{3, 5, 3\}, \quad \{4, 3, 5\}, \quad \{5, 3, 4\}, \quad \{5, 3, 5\},$$

in which the cells are respectively: an icosahedron of angle $2\pi/3$, a hexahedron of angle $2\pi/5$, and dodecahedra of angles $\pi/2$ and $2\pi/5$ (Schlegel 1883, p. 444). The existence of the "hyperbolic dodecahedron space" (Weber and Seifert 1933, pp. 241—243) reveals an interesting property of the self-reciprocal honeycomb $\{5, 3, 5\}$: its symmetry group has a subgroup (of index 120) which is transitive on the vertices and has the whole cell for a fundamental region. This resembles the translation group of the Euclidean honeycomb of cubes, $\{4, 3, 4\}$, only now instead of translations we have screws.

Returning to the general discussion, we allow the dihedral angle of the Platonic solid $\{p, q\}$ to take its minimum value $(1 - 2/q)\pi$, so as to obtain the further honeycombs

$$\{3, 4, 4\}, \quad \{3, 3, 6\}, \quad \{4, 3, 6\}, \quad \{5, 3, 6\},$$

whose vertices are all at infinity (i.e., on the absolute quadric). We naturally consider also the respective reciprocals

$$\{4, 4, 3\}, \quad \{6, 3, 3\}, \quad \{6, 3, 4\}, \quad \{6, 3, 5\},$$

whose cells are inscribed in horospheres instead of finite spheres (Coxeter and Whitrow 1950, p. 426), as well as the three self-reciprocal honeycombs

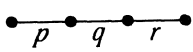
$$\{6, 3, 6\}, \qquad \{4, 4, 4\}, \qquad \{3, 6, 3\},$$

which suffer from both peculiarities at once.

In other words, necessary and sufficient conditions for the existence of a hyperbolic honeycomb $\{p, q, r\}$ are 3.1 and

3.2 $$(p-2)(q-2) \leqq 4, \qquad (q-2)(r-2) \leqq 4.$$

The honeycomb can be derived by Wythoff's construction from its characteristic simplex, which is the quadrirectangular hyperbolic tetrahedron



(Coxeter 1948, p. 139). The inequalities 3.2 ensure that this tetrahedron has a finite volume, being entirely accessible except that it may have one or two vertices at infinity. From relations between the edges and angles of the tetrahedron, we find the edge-length of $\{p, q, r\}$ to be 2φ while its cell $\{p, q\}$ has circum-radius χ and in-radius ψ, where

$$\cosh\varphi = \cos\frac{\pi}{p}\sin\frac{\pi}{r} \Big/ \sin\frac{\pi}{h_{q,r}}, \qquad \cosh\psi = \sin\frac{\pi}{p}\cos\frac{\pi}{r} \Big/ \sin\frac{\pi}{h_{p,q}},$$

$$\cosh\chi = \cos\frac{\pi}{p}\cos\frac{\pi}{q}\cos\frac{\pi}{r} \Big/ \sin\frac{\pi}{h_{p,q}}\sin\frac{\pi}{h_{q,r}},$$

$h_{p,q}$ being given by

$$\cos^2\frac{\pi}{h_{p,q}} = \cos^2\frac{\pi}{p} + \cos^2\frac{\pi}{q}$$

(Coxeter 1948, p. 19). For the particular cases, see Table III, the first part of which was given earlier by Sommerville (1923, p. 96), whose e, R, r are our 2φ, χ, ψ.

4. *Four-dimensional honeycombs.* When

4.1 $$\sin\frac{\pi}{p}\sin\frac{\pi}{r} > \cos\frac{\pi}{q},$$

the symbol $\{p, q, r\}$ denotes a regular four-dimensional polytope (Coxeter 1948, p. 135), which has smaller angles in hyperbolic space than in Euclidean. A discussion analogous to that of § 3 shows that infinitely many such cells $\{p, q, r\}$ can be fitted together, s round each plane face, to make a four-dimensional honeycomb

$$\{p, q, r, s\},$$

whenever both $\{p, q, r\}$ and $\{q, r, s\}$ are finite polytopes or Euclidean honeycombs such that

$$4.2 \qquad \frac{\cos^2 \pi/q}{\sin^2 \pi/p} + \frac{\cos^2 \pi/r}{\sin^2 \pi/s} > 1$$

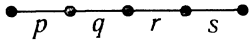
(cf. Coxeter 1948, p. 136). In this manner, the six polytopes

$$\{3, 3, 3\}, \quad \{3, 3, 4\}, \quad \{4, 3, 3\}, \quad \{3, 4, 3\}, \quad \{3, 3, 5\}, \quad (5, 3, 3\},$$

and the cubic honeycomb $\{4, 3, 4\}$, yield the seven hyperbolic honeycombs

$$\{3, 3, 3, 5\}, \quad \{4, 3, 3, 5\}, \quad \{5, 3, 3, 5\}, \quad \{5, 3, 3, 4\}, \quad \{5, 3, 3, 3\},$$
$$\{3, 4, 3, 4\}, \qquad \{4, 3, 4, 3\}.$$

Each of these can be derived from the appropriate characteristic simplex



by Wythoff's construction (Coxeter 1948, p. 199).

The edge-length 2φ, and the circum- and in-radii of a cell, χ and ψ, are given by the formulae

$$\operatorname{sech}^2 \varphi = \frac{(-1, 1)(0, 5)}{(1, 5)}, \qquad \operatorname{sech}^2 \psi = \frac{(-1, 4)(3, 5)}{(-1, 3)},$$
$$\operatorname{sech}^2 \chi = (-1, 4)(0, 5)$$

(Coxeter 1948, p. 161), where the symbols (j, k) are derived by the recurrence formula

$$(j, k) = \frac{(j, k-1)(j+1, k) - 1}{(j+1, k-1)}, \quad (j, j+1) = 1,$$

from any convenient sequence of numbers

$$(-1, 1), \qquad (0, 2), \qquad (1, 3), \qquad (2, 4), \qquad (3, 5)$$

satisfying

$$(-1, 1)(0, 2) = \sec^2 \frac{\pi}{p}, \quad (0, 2)(1, 3) = \sec^2 \frac{\pi}{q},$$
$$(1, 3)(2, 4) = \sec^2 \frac{\pi}{r}, \quad (2, 4)(3, 5) = \sec^2 \frac{\pi}{s}.$$

In any particular case, the values are most easily computed by arranging the (j, k)'s in a triangular table:

$$\begin{array}{ccccccccc}
(-1, 1) & & (0, 2) & & (1, 3) & & (2, 4) & & (3, 5) \\
& (-1, 2) & & (0, 3) & & (1, 4) & & (2, 5) & \\
& & (-1, 3) & & (0, 4) & & (1, 5) & & \\
& & & (-1, 4) & & (0, 5) & & & \\
& & & & (-1, 5) & & & &
\end{array}$$

e.g., the respective tables for $\{5, 3, 3, 5\}$ and $\{4, 3, 4, 3\}$ are

$$\begin{array}{ccccccccc} 2\tau^{-2} & & 2 & & 2 & & 2 & & 2\tau^{-2} \\ & \sqrt{5}\tau^{-3} & & 3 & & 3 & & \sqrt{5}\tau^{-3} & \\ & & 2\tau^{-4} & & 4 & & 2\tau^{-4} & & \\ & & & \tau^{-6} & & \tau^{-6} & & & \\ & & & & -2\sqrt{5}\tau^{-6} & & & & \end{array} \qquad \begin{array}{ccccccccc} 1 & & 2 & & 2 & & 1 & & 4 \\ & 1 & & 3 & & 1 & & 3 & \\ & & 1 & & 1 & & 2 & & \\ & & & 0 & & 1 & & & \\ & & & & -1 & & & & \end{array}$$

where $\tau = \frac{1}{2}(\sqrt{5} + 1)$, so that $\tau^{-1} = \frac{1}{2}(\sqrt{5} - 1)$. The negative value of $(-1, 5)$ provides a verification that the honeycomb is hyperbolic, and the zero for $(-1, 4)$ indicates that the cell of $\{4, 3, 4, 3\}$ is infinite. (For the results of this computation, see Table IV.)

5. *Five-dimensional honeycombs.* Similarly, the cell $\{p, q, r, s\}$ and vertex figure $\{q, r, s, t\}$ of a five-dimensional honeycomb $\{p, q, r, s, t\}$ must occur among the finite polytopes

$$\{3, 3, 3, 3\}, \qquad \{3, 3, 3, 4\}, \qquad \{4, 3, 3, 3\}$$

or among the Euclidean honeycombs

$$\{3, 3, 4, 3\}, \qquad \{3, 4, 3, 3\}, \qquad \{4, 3, 3, 4\}$$

(Coxeter 1948, p. 136). Since

$$\{3, 3, 3, 3, 3\}, \qquad \{3, 3, 3, 3, 4\}, \qquad \{4, 3, 3, 3, 3\}$$

are finite, while $\{4, 3, 3, 3, 4\}$ is Euclidean, the only hyperbolic honeycombs $\{p, q, r, s, t\}$ are

$$\{3, 3, 3, 4, 3\}, \quad \{4, 3, 3, 4, 3\}, \quad \{3, 3, 4, 3, 3\}, \quad \{3, 4, 3, 3, 4\} \quad \{3, 4, 3, 3, 3\}$$

all of which have either infinite cells or all their vertices at infinity.

The edge-length and radii are now given by the formulae

$$\operatorname{sech}^2 \varphi = \frac{(-1, 1)(0, 6)}{(1, 6)}, \quad \operatorname{sech}^2 \psi = \frac{(-1, 5)(4, 6)}{(-1, 4)},$$

$$\operatorname{sech}^2 \chi = (-1, 5)(0, 6),$$

with $(3, 5)(4, 6) = \sec^2 \pi/t$. The triangular tables for $\{3, 3, 3, 4, 3\}$ and $\{3, 3, 4, 3, 3\}$ are

$$\begin{array}{cccccccc} 2 & & 2 & & 2 & & 2 & & 1 & & 4 \\ & 3 & & 3 & & 3 & & 1 & & 3 & \\ & & 4 & & 4 & & 1 & & 2 & & \\ & & & 5 & & 1 & & 1 & & & \\ & & & & 1 & & 0 & & & & \\ & & & & & -1 & & & & & \end{array} \qquad \begin{array}{cccccccc} 2 & & 2 & & 2 & & 1 & & 4 & & 1 \\ & 3 & & 3 & & 1 & & 3 & & 3 & \\ & & 4 & & 1 & & 2 & & 2 & & \\ & & & 1 & & 1 & & 1 & & & \\ & & & & 0 & & 0 & & & & \\ & & & & & -1 & & & & & \end{array}$$

yielding $\varphi = \chi = \infty$, $\psi = \log \tau$ for the former, and $\varphi = \chi = \psi = \infty$ for the latter. (See Table V.)

This is the end of the story, so far as honeycombs of density 1 are concerned. For, if $n > 5$, the only finite polytopes and Euclidean honeycombs that might serve as cells and vertex figures are

$$\alpha_n = \{3, 3, \quad ., 3, 3\}, \qquad \beta_n = \{3, 3, \ldots, 3, 4\},$$
$$\gamma_n = \{4, 3, \ldots, 3, 3\}, \qquad \delta_n = \{4, 3, \ldots, 3, 4\};$$

and these yield only α_{n+1}, β_{n+1}, γ_{n+1}, and δ_{n+1}. Hence (Schlegel 1883, p. 455)

There are no regular honeycombs in hyperbolic space of six or more dimensions.

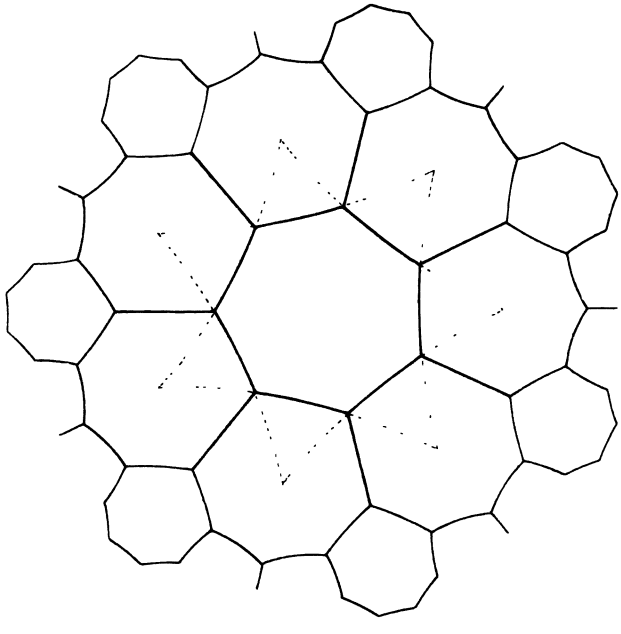


Figure 1

6. *Two-dimensional star-honeycombs.* If n/d is a fraction in its lowest terms, whose value is greater than 2, the symbol $\{n/d\}$ denotes a regular star-polygon whose n sides surround its centre d times: briefly, a regular n-gon of density d, such as the pentagram $\{\frac{5}{2}\}$. It is natural to ask whether the symbol $\{p, q\}$ for a hyperbolic tessellation remains valid when p or q is fractional. We can obviously *begin* to construct such a star-tessellation whenever 2.1 is satisfied. The question is whether it will cover the plane a finite number of times, i.e., whether its *density* is finite.

The symmetry group of such a tessellation $\{p, q\}$ is still generated by reflections in the sides of its characteristic triangle, whose angles are π/q, $\pi/2$, and π/p (Coxeter 1948, p. 109). If p or q is fractional, this triangle is dissected into smaller triangles by "virtual mirrors" (Coxeter 1948, pp. 75, 76), and the process of subdivision will continue until we come to a triangle all of whose angles are submultiples of π (or possibly zero). There might conceivably be

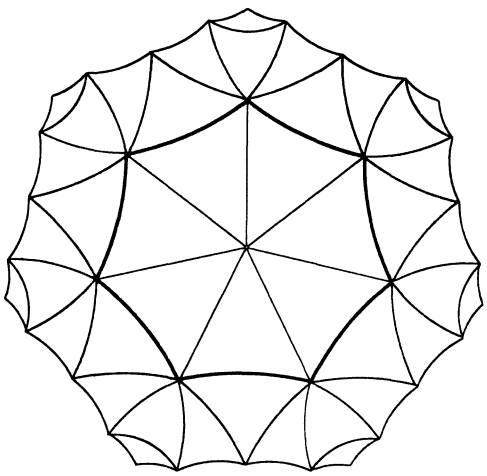


Figure 2

several different triangles of this kind, but the smallest of them will serve as a fundamental region for the group. We denote this smallest triangle by $(l\ m\ n)$ to indicate that its angles are π/l, π/m, π/n. The number of repetitions of it that fill the characteristic triangle $(2\ p\ q)$ is an integer $D > 1$, which is equal to the density of the tessellation (Coxeter 1948, p. 110).

Since each angle of $(2\ p\ q)$ must be a multiple of one of the angles of $(l\ m\ n)$, a zero angle of the former implies a zero angle of the latter. But since the subdivision of an asymptotic triangle $(2\ p\ \infty)$ yields at least one finite piece, this is impossible; both p and q must be finite.

Since the triangle $(2\ p\ q)$ is filled with D triangles $(l\ m\ n)$, of area $(1 - l^{-1} - m^{-1} - n^{-1})\pi$, we have

$$\tfrac{1}{2} - p^{-1} - q^{-1} = D(1 - l^{-1} - m^{-1} - n^{-1}),$$

where the numerators of the rational numbers p and q are divisors of one or two of the integers l, m, n.

The numerators of p and q cannot divide two *different* integers among l, m, n, say m and n; for then we would have

$$\tfrac{1}{2} - p^{-1} - q^{-1} \leqq \tfrac{1}{2} - m^{-1} - n^{-1} \leqq 1 - l^{-1} - m^{-1} - n^{-1},$$

implying $D \leqq 1$. Thus we may assume that these numerators both divide m, so that $p = m/x$ and $q = m/y$, where x and y are positive integers. Since p and q are not both integers,

$$x + y \geqq 3.$$

Moreover, $D \geqq 3$, since the triangle $(2\ p\ q)$ obviously cannot be *bisected* by a line through one of its acute vertices. Hence

$$\frac{1}{2} = \frac{x}{m} + \frac{y}{m} + D\left(1 - \frac{1}{l} - \frac{1}{m} - \frac{1}{n}\right)$$

$$\geqq \frac{3}{m} + 3\left(1 - \frac{1}{l} - \frac{1}{m} - \frac{1}{n}\right) = 3\left(1 - \frac{1}{l} - \frac{1}{n}\right),$$

i.e.,

$$\frac{1}{l} + \frac{1}{n} \geqq \frac{5}{6}.$$

Since also

$$\frac{1}{l} + \frac{1}{n} < \frac{1}{l} + \frac{1}{m} + \frac{1}{n} < 1,$$

the integers l and n must be 2 and 3, and

$$x + y = D = 3.$$

Thus the only possibility is the triangle $(2\ m\ m/2)$ dissected into three triangles $(2\ m\ 3)$. In other words,

The only regular star-tessellations in the hyperbolic plane are

$$\left\{\frac{m}{2}, m\right\} \text{ and } \left\{m, \frac{m}{2}\right\},$$

of density 3, *where* m *is any odd number greater than* 5.

Such star-tessellations exist also when $m = 5$, but then they are not hyperbolic but spherical. In fact, they are essentially the small stellated dodecahedron of Kepler and the great dodecahedron of Poinsot (Coxeter 1948, p. 95). But the other two Kepler-Poinsot polyhedra have no hyperbolic analogues.

We may describe the faces of $\{m/2, m\}$ as the stellated faces of $\{m, 3\}$. (One face of $\{\frac{7}{2}, 7\}$ is indicated by broken lines in Fig. 1.) Dually $\{m, m/2\}$ is derived from $\{3, m\}$ by regarding the same vertices and edges as forming m-gons (such as the heptagon of $\{7, \frac{7}{2}\}$ which is emphasized in Fig. 2) instead of triangles. The density 3 can be observed in the fact that each triangle of $\{3, m\}$ lies within three m-gons of $\{m, m/2\}$. In ascribing the same density to $(m/2, m\}$, we regard the middle part of each $\{m/2\}$ as being covered twice over. The relationship of $(m/2, m\}$ and $\{m, m/2\}$ is such that any face of either has the same

vertices as a corresponding face of the other. Both have the same vertices as $\{3, m\}$, and the same face-centres as $\{m, 3\}$. Thus

$$\chi(m, m/2) = 2\varphi(m, m/2) = 2\varphi(3, m),$$

in agreement with Table II.

7. *Four-dimensional star-honeycombs.* One might expect to find a three-dimensional star-honeycomb $\{p, q, r\}$ whose cell $\{p, q\}$ or vertex figure $(q, r\}$ is one of the Kepler-Poinsot polyhedra

$$\{\tfrac{5}{2}, 5\}, \qquad \{5, \tfrac{5}{2}\}, \qquad \{\tfrac{5}{2}, 3\}, \qquad \{3, \tfrac{5}{2}\}.$$

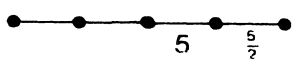
However, in every instance (Coxeter 1948, p. 264) the values of p, q, r satisfy 4.1, not 3.1. Hence

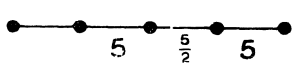
There are no regular star-honeycombs in hyperbolic 3-space.

Hoping for a more positive result in four dimensions, we seek a honeycomb $\{p, q, r, s\}$ in which both $\{p, q, r\}$ and $\{q, r, s\}$ occur among the sixteen regular polytopes (Coxeter 1948, pp. 293—294) while at least one of p, q, r. s has the fractional value $\frac{5}{2}$. A list of the twenty-three possibilities reveals that only four satisfy 4.2:

$$\{3, 3, 5, \tfrac{5}{2}\}, \qquad \{3, 5, \tfrac{5}{2}, 5\}, \qquad \{5, \tfrac{5}{2}, 5, 3\}, \qquad \{\tfrac{5}{2}, 5, 3, 3\}$$

(Coxeter 1948, p. 264: 14.15). That these four are genuine hyperbolic honeycombs, of finite density, may be seen by verifying that the reflections in the bounding hyperplanes of their characteristic simplexes





generate discrete groups. In Table I (cf. Coxeter 1948, p. 283) these simplexes appear as X and Z, and we see how they are dissected, by virtual mirrors (bisecting their dihedral angles $2\pi/5$, which are indicated by the mark $\frac{5}{2}$ in the graphical symbols), into smaller simplexes T and W, W and Y. Similarly W and Y, having angles $2\pi/3$, are further subdivided, until the process ends with the "quantum" T, which is thus seen to be the fundamental region for both these groups, as it is also for the symmetry group of the ordinary honeycombs $\{3, 3, 3, 5\}$ and $\{5, 3, 3, 3\}$.

The accuracy of Table I can be checked by observing that, in each graph, the branch with a fractional mark forms, with any third node, the symbol for a spherical triangle whose dissection is obvious; e.g., the dissection $Y = 2V$ embodies

$$(5\ 5\ \tfrac{3}{2}) = (5\ 2\ 3) + (2\ 5\ 3) \text{ and } (3\ 2\ \tfrac{3}{2}) = (3\ \tfrac{3}{2}\ 3) + (3\ 2\ 3)$$

(Coxeter 1948, p. 113). Since $X = 5T$ and $Z = 10T$, the densities of the star-honeycombs are 5 and 10.

Thus the six hyperbolic honeycombs

$$\{5, 3, 3, 3\}, \quad \{\tfrac{5}{2}, 5, 3, 3\}, \quad \{5, \tfrac{5}{2}, 5, 3\}, \quad \{3, 5, \tfrac{5}{2}, 5\}, \quad \{3, 3, 5, \tfrac{5}{2}\}, \quad \{3, 3, 3, 5\}$$

all have the same symmetry group, and their densities are the binomial coefficients

$$1, \quad 5, \quad 10, \quad 10, \quad 5, \quad 1.$$

It is interesting to compare them with the five spherical honeycombs (or Euclidean polytopes)

$$\{5, 3, 3\}, \quad \{\tfrac{5}{2}, 5, 3\}, \quad \{5, \tfrac{5}{2}, 5\}, \quad \{3, 5, \tfrac{5}{2}\}, \quad \{3, 3, 5\},$$

whose densitities are 1, 4, 6, 4, 1, and with the four spherical tessellations (or Euclidean polyhedra)

$$\{5, 3\}, \quad \{\tfrac{5}{2}, 5\}, \quad \{5, \tfrac{5}{2}\}, \quad \{3, 5\},$$

whose densities are 1, 3, 3, 1. We can summarize these results by saying that, when the Schläfli symbol for an n-dimensional honeycomb has $\frac{5}{2}$ in the r^{th} place with 5 before and after and 3 everywhere else, the density is $\binom{n}{r}$.

The dissection of characteristic simplexes can be translated into a direct derivation of the honeycombs from one another. The first stage is to derive $\{\frac{5}{2}, 5, 3, 3\}$ from $\{5, 3, 3, 3\}$ by stellating each cell $\{5, 3, 3\}$ to form a $\{\frac{5}{2}, 5, 3\}$ (Coxeter 1948, p. 264). The second stage is to replace each $\{\frac{5}{2}\}$ of $\{\frac{5}{2}, 5, 3, 3\}$ by the pentagon that has the same vertices, and consequently each $\{\frac{5}{2}, 5\}$ by a $\{5, \frac{5}{2}\}$ and each $\{\frac{5}{2}, 5, 3\}$ by a $\{5, \frac{5}{2}, 5\}$. In this process the vertex figure gets stellated from $\{5, 3, 3\}$ to $\{\frac{5}{2}, 5, 3\}$, and the result is $\{5, \frac{5}{2}, 5, 3\}$. The third stage is to replace each $\{5, \frac{5}{2}\}$ by the icosahedron that has the same edges, and consequently each $\{5, \frac{5}{2}, 5\}$ by a $\{3, 5, \frac{5}{2}\}$. The vertex figure $\{\frac{5}{2}, 5, 3\}$ is changed into $\{5, \frac{5}{2}, 5\}$, and the result is $\{3, 5, \frac{5}{2}, 5\}$. The fourth stage is to replace each $\{3, 5, \frac{5}{2}\}$ by the $\{3, 3, 5\}$ that has the same faces, so that the vertex figure $\{5, \frac{5}{2}, 5\}$ is changed into $\{3, 5, \frac{5}{2}\}$ (which has the same edges), and we obtain $\{3, 3, 5, \frac{5}{2}\}$. The fifth and last stage is to replace the cell $\{3, 3, 5\}$ by a cluster of 600 regular simplexes $\{3, 3, 3\}$ surrounding a common vertex, so as to obtain $\{3, 3, 3, 5\}$. Each simplex belongs to five of the clusters, as we would expect from the fact that $\{3, 3, 5, \frac{5}{2}\}$ has density 5.

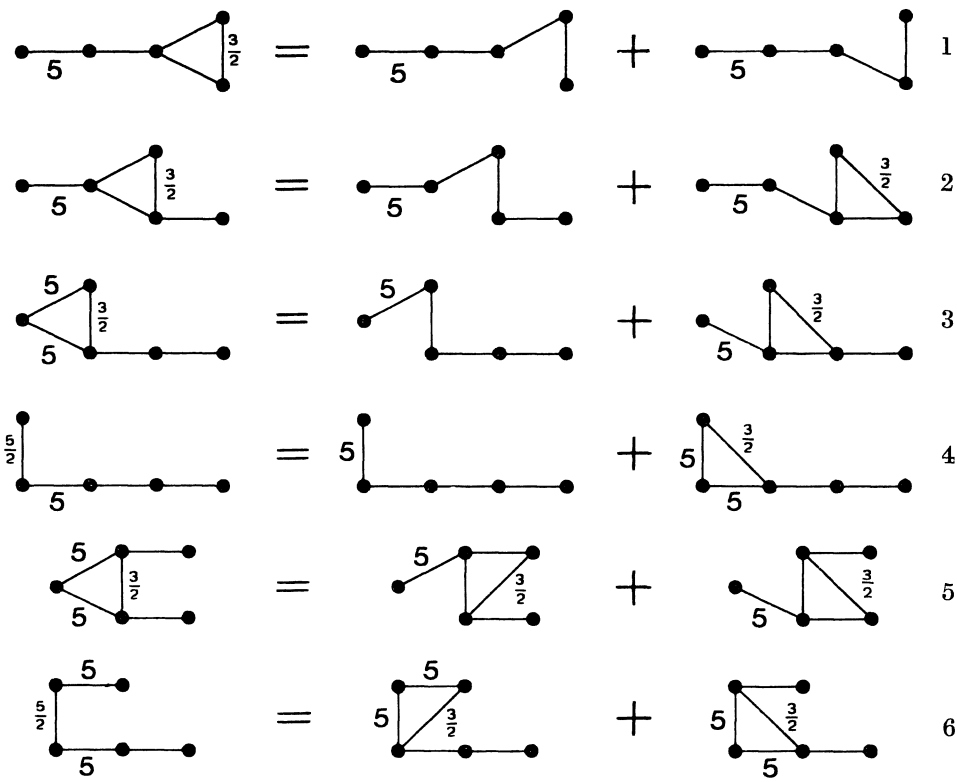
Taking the same stages in the reverse order, we may say that the vertices, edges, faces and cells of $\{3, 3, 3, 5\}$ belong also to $\{3, 3, 5, \frac{5}{2}\}$; the vertices, edges and faces belong also to $\{3, 5, \frac{5}{2}, 5\}$; the vertices and edges belong also to $\{5, \frac{5}{2}, 5, 3\}$; and the vertices belong also to $\{\frac{5}{2}, 5, 3, 3\}$, which is the "stellated" $\{5, 3, 3, 3\}$.

The formulae for φ, χ, ψ (§ 4) apply to star-honeycombs without any alteration. We see from Table IV that, since $\tau^3 = 2\tau^2 - 1$, the value of χ for each of them is equal to the value of 2φ for $\{3, 3, 3, 5\}$.

Finally, since there are no regular star-polytopes in five or more dimensions (Coxeter 1948, p. 278) to serve as cell or vertex figure,

There are no regular star-honeycombs in hyperbolic space of five or more dimensions.

Table I. The dissection of characteristic simplexes in hyperbolic 4-space.



1	$U = 2T$	3	$W = T + V = 4T$	5	$Y = 2V = 6T$
2	$V = T + U = 3T$	4	$X = T + W = 5T$	6	$Z = W + Y = 10T$

Table II. Regular honeycombs in the hyperbolic plane

Tessellation	Density	$\cosh \varphi$	$\cosh \chi$	$\cosh \psi$
$\{p, q\}$ (integers)	1	$\cos\frac{\pi}{p}\operatorname{cosec}\frac{\pi}{q}$	$\cot\frac{\pi}{p}\cot\frac{\pi}{q}$	$\operatorname{cosec}\frac{\pi}{p}\cos\frac{\pi}{q}$
$\left\{\frac{m}{2}, m\right\}$ (m odd)	3	$\cos\frac{2\pi}{m}\operatorname{cosec}\frac{\pi}{m}$	$\cot\frac{\pi}{m}\cot\frac{2\pi}{m}$	$\frac{1}{2}\operatorname{cosec}\frac{\pi}{m}$
$\left\{m, \frac{m}{2}\right\}$		$\frac{1}{2}\operatorname{cosec}\frac{\pi}{m}$		$\cos\frac{2\pi}{m}\operatorname{cosec}\frac{\pi}{m}$

Table III. The fifteen regular honeycombs in hyperbolic 3-space

eycomb	Density	$\cosh^2\varphi$	$\cosh^2\chi$	$\cosh^2\psi$
5, 3}	1	$\frac{3}{4}\tau^2$	$\frac{1}{4}\tau^6$	$\frac{3}{4}\tau^2$
3, 5}	1	$\frac{1}{2}\sqrt{5}\,\tau$	$\frac{1}{2}\tau^4$	$\frac{1}{2}\tau^2$
3, 4}		$\frac{1}{2}\tau^2$		$\frac{1}{2}\sqrt{5}\,\tau$
3, 5}	1	$\frac{1}{4}\sqrt{5}\,\tau^3$	$\frac{1}{4}\tau^8$	$\frac{1}{4}\sqrt{5}\,\tau^3$
3, 6}	1	∞	∞	$\frac{9}{8}$
3, 3}		$\frac{9}{8}$		∞
4, 4}	1	∞	∞	$\frac{3}{2}$
4, 3}		$\frac{3}{2}$		∞

Honeycomb	Density	$\cosh^2\varphi$	$\cosh^2\chi$	$\cosh^2\psi$
{3, 6, 3}	1	∞	∞	∞
{4, 3, 6}	1	∞	∞	$\frac{3}{2}$
{6, 3, 4}		$\frac{3}{2}$		∞
{4, 4, 4}	1	∞	∞	∞
{5, 3, 6}	1	∞	∞	$\frac{3}{4}\sqrt{5}\,\tau$
{6, 3, 5}		$\frac{3}{4}\sqrt{5}\,\tau$		∞
{6, 3, 6}	1	∞	∞	∞

$\left(\tau = \frac{\sqrt{5}+1}{2}\right)$

Table IV. The eleven regular honeycombs in hyperbolic 4-space

Honeycomb	Density	$\cosh \varphi$	$\cosh \chi$	$\cosh \psi$
$\{3, 3, 3, 5\}$	1	τ	$\sqrt{\frac{1}{5}}\,\tau^3$	$\sqrt{\frac{2}{5}}\,\tau$
$\{5, 3, 3, 3\}$		$\sqrt{\frac{2}{5}}\,\tau$		τ
$\{4, 3, 3, 5\}$	1	$\sqrt{2}\,\tau$	τ^3	$\sqrt{\frac{1}{2}}\,\tau$
$\{5, 3, 3, 4\}$		$\sqrt{\frac{1}{2}}\,\tau$		$\sqrt{2}\,\tau$
$\{5, 3, 3, 5\}$	1	τ^2	τ^6	τ^2

Honeycomb	Density	$\cosh \varphi$	$\cosh \chi$	cosh
$\{3, 4, 3, 4\}$	1	∞	∞	$\sqrt{2}$
$\{4, 3, 4, 3\}$		$\sqrt{2}$		∞
$\{\frac{5}{2}, 5, 3, 3\}$	5	$\sqrt{2}\,\tau$	τ^3	τ
$\{3, 3, 5, \frac{5}{2}\}$		τ		$\sqrt{2}$
$\{3, 5, \frac{5}{2}, 5\}$	10	τ	τ^3	τ
$\{5, \frac{5}{2}, 5, 3\}$				

Table V. The five regular honeycombs in hyperbolic 5-space

Honeycomb	Density	φ	χ	ψ
$\{3, 3, 3, 4, 3\}$	1	∞	∞	$\log \tau$
$\{3, 4, 3, 3, 3\}$		$\log \tau$		∞
$\{3, 3, 4, 3, 3\}$	1	∞	∞	∞
$\{3, 4, 3, 3, 4\}$	1	∞	∞	∞
$\{4, 3, 3, 4, 3\}$				

REFERENCES

H. S. M. COXETER 1933. The densities of the regular polytopes, Part 3, Proc. Camb. Phil. Soc. **29**, 1—22.

H. S. M. COXETER 1939. The abstract groups $G^{m,n,p}$, Trans. Amer. Math. Soc. **45**, 73—150.

H. S. M. COXETER 1946. The nine regular solids, Proc. 1st Canadian Math. Congress, 252—264.

H. S. M. COXETER 1947. Non-Euclidean geometry, Toronto (University Press).

H. S. M. COXETER 1948. Regular polytopes, London (Methuen).

H. S. M. COXETER and G. J. WHITROW 1950. World-structure and non-Euclidean honey combs, Proc. Roy. Soc. A, **201**, 417—437.

L. FEJES TÓTH 1953. On close-packings of spheres in spaces of constant curvature, Publ. Math. Debrecen, **3**, 158—167.

R. FRICKE 1892. Ueber den arithmetischen Charakter der zu den Verzweigungen (2, 3, 7) und (2, 4, 7) gehörenden Dreiecksfunctionen, Math. Ann. **41**, 443—468.

F. KLEIN 1879. Ges. Math. Abh. **3** (Berlin, 1923), 90—136.

L. SCHLÄFLI 1855. Ges. Math. Abh. **2** (Basel, 1953), 164—190.

V. SCHLEGEL 1883. Theorie der homogen zusammengesetzten Raumgebilde, Nova Acta Leop. Carol. **44**, 343—459.

D. M. Y. SOMMERVILLE 1923. Division of space by congruent triangles and tetrahedra, Proc. Roy. Soc. Edinburgh, **43**, 85—116.

W. I. STRINGHAM 1880. Regular figures in n-dimensional space, Amer. J. Math. **3**, 1—14.

C. WEBER and H. SEIFERT 1933. Die beiden Dodekaederräume, Math. Z, **37**, 237—253.

ZUR COHOMOLOGIETHEORIE VON RÄUMEN UND GRUPPEN

BENO ECKMANN

Das Grenzgebiet zwischen Topologie und Algebra, das etwa mit dem Namen „homologische Algebra" bezeichnet wird, hat sich in der letzten Zeit nach beiden Seiten hin ausgedehnt, derjenigen der Räume und derjenigen der algebraischen Strukturen. Die Anwendungen und die gegenseitigen Anregungen erstrecken sich von Homotopiefragen bis zur Klassenkörpertheorie; verschiedenartige geometrische und algebraische Fragenstellungen haben sich als verwandt und denselben Methoden zugänglich erwiesen.

In diesem Vortrag handelt es sich um einen methodisch und inhaltlich eng begrenzten Ausschnitt aus diesem weit verzweigten Gebiet: nämlich um einen grundlegenden Satz über Komplexe mit Operatoren [1]) und verschiedene Folgerungen bezüglich der additiven Cohomologie; diese sind teils topologisch, teils gruppentheoretisch — es liegt wohl in der Natur der Sache, dass hier eine strenge Trennung nicht angebracht wäre.

1. Den natürlichen Ausgangspunkt bilden die *Ueberlagerungsräume.* Es sei U eine reguläre Ueberlagerung des Raumes S, mit der Decktransformationengruppe G, die in U fixpunktfrei operiert; U zerfällt in Aequivalenzklassen, die den Punkten von S eineindeutig entsprechen. Eine Funktion in S lässt sich auffassen als eine solche in U, die auf jeder Aequivalenzklasse konstant ist, eine „automorphe" Funktion. Analog können Cohomologieeigenschaften von S *in* U untersucht werden, wenn man sich dort auf Coketten beschränkt, die automorph sind, d.h. bei den Decktransformationen $x \in G$ in sich übergehen. Es sei hier etwa die singuläre Theorie zugrundegelegt. Für ein singuläres p-Simplex σ_p bezeichne $x\sigma_p$ sein Bild bei $x \in G$; dadurch wird G zu einem Automorphismus der Kettengruppe C_p von U, welcher randtreu ist: $\partial(xc_p) = x\partial c_p$ für alle $x \in G$, $c_p \in C_p$. M.a.W.: G ist eine Automorphismengruppe des singulären Komplexes $C(U)$ von U. Die automorphen Coketten f^p sind die, welche bei diesen Automorphismen in sich übergehen.

Etwas allgemeiner sei angenommen, die Abelsche Gruppe J, in welcher die Werte der f^p liegen, besitze G als Gruppe von Linksoperatoren, und es gelte

[1]) Vgl. [1] und die dort zu findenden Literaturhinweise. Als übersichtliche Einführung in die vorliegenden Fragestellungen verweisen wir ferner auf [3]. — Zahlen in eckiger Klammer [ ] beziehen sich auf das Literaturverzeichnis am Schluss des vorliegenden Berichtes.

$f^p(x\,c_p) = x.f^p(c_p)$ für alle $x \in G$, $c_p \in C_p$); solche Coketten heissen *zulässig* oder äquivariant (die automorphen erhält man im Spezialfall, wo die Operationen von G in J alle trivial, d.h. gleich der Identität von J sind: $f^p(xc_p) = f^p(c_p)$). Die mit solchen zulässigen Coketten gebildeten Cohomologiegruppen seien im Folgenden mit H^p bezeichnet; im Gegensatz dazu die ,,gewöhnlichen'', auf keine Automorphismengruppe bezogenen, mit $\mathbf{H}^p$. Offenbar gilt die Isomorphie $\mathbf{H}^p(S, J) \cong H^p(C(U), J)$, wenn G in J trivial operiert — aber auch sonst, falls unter $\mathbf{H}^p$ die Cohomologiegruppen von S mit *lokalen Koeffizienten* im Sinne von Steenrod verstanden werden.

R_G sei der ganzzahlige Gruppenring von G. Vermöge linearer Fortsetzung fassen wir die C_p und J als R_G-Modulen auf, die Randbildung ∂ als R_G-Homomorphismus von C_p in C_{p-1}, ($p \geqq 0$; C_{-1} sei die additive Gruppe Z der ganzen Zahlen, in welcher G trivial operiert, und $\partial\sigma^0 = 1$ für jedes 0-Simplex σ^0), und die zulässigen Coketten f^p als R_G-Homomorphismen von C_p in J.

2. Durch diese Situation wird der Begriff des ,,*Komplexes über dem Ring R*'' (kurz ,,R-Komplex'') nahegelegt, wo R ein beliebiger Ring mit Einselement e ist; darunter verstehen wir einen Komplex C, in dem die Kettengruppen C_p, $p = -1, 0, 1, 2, \ldots$ R-Moduln [2]) sind, und die Randoperationen ∂ R-Homomorphismen [2]) von C_p in C_{p-1}, $p \geqq 0$ (für $p = 0$ *auf* C_{-1}). Die zulässigen Cohomologiegruppen $H^p(C, J)$ von C sind so definiert: J sei ein R-Modul, $C^p = \mathrm{Hom}_R(C_p, J)$ die Gruppe der R-Homomorphismen von C_p in J, δ der zu ∂ duale Homomorphismus von C^p in C^{p+1} (für $f^p \in C^p$ ist $\delta f^p \in C^{p+1}$ durch $\delta f^p(c_{p+1}) = f^p(\partial C_{p+1})$ erklärt); wegen $\partial\partial = 0$ ist $\delta\delta = 0$; Z^p sei der Kern von δ in C^p und $H^p(C, J) = Z^p/\delta C^{p-1}$ für $p \geqq 1$, $= Z^0$ für $p = 0$. Soll der Ring hervorgehoben werden, auf welchen sich ,,zulässig'' bezieht, so schreiben wir H_R^p statt H^p. Offenbar kann man i.A. einen Komplex C gleichzeitig über verschiedenen Ringen betrachten, insbesondere stets über Z (ganze Zahlen) statt R; für eine beliebige Abelsche Gruppe J ist dann $H_Z^p(C, J)$ die gewöhnliche Cohomologiegruppe $\mathbf{H}^p(C, J)$ von C.

Die Tatsache, dass in C gleichzeitig *verschiedene Arten zulässiger Coketten* f^p auftreten, sei nun in folgender Weise formuliert und verallgemeinert. Es ist stets $f^p \in \mathrm{Hom}_Z(C_p, J)$, aber $f^p(rc_p)$ als Funktion von $r \in R$ von Fall zu Fall verschieden. Wir setzen $\Phi = \mathrm{Hom}_Z(R, J)$, für eine Abelsche Gruppe J; Φ ist ein R-Modul vermöge $(r'\varphi)(r) = \varphi(rr')$, für alle $r, r' \in R$. Ferner sei ψ ein Teilmodul von Φ. Eine Cokette $f^p \in \mathrm{Hom}_Z(C_p, J)$ heisst ,,vom Typus ψ'',

---

[2]) Ein R-Modul ist eine additive Abelsche Gruppe, welche R als Ring von Linksoperatoren besitzt, wobei e der identische Operator ist; ein R-Homomorphismus ist ein bezüglich dieser Operatoren zulässiger Homomorphismus.

wenn für jedes feste $c_p \in C_p$ die Funktion $f^p(rc_p) = \psi(r)$ zu Ψ gehört. In üblicher Weise, aber unter *durchgehender Beschrankung auf Coketten vom* Typus Ψ, erhält man dann Cohomologiegruppen $H^p_{\text{Typ }\Psi}(C, J)$.

Beispiele: (1) $\Psi = \Phi = \text{Hom}_Z(R, J)$. Dann ist $H^p_{\text{Typ }\Phi}(C, J) = H^p_Z(C, J) = \mathbf{H}^p(C, J)$ die gewöhnliche Cohomologiegruppe von C.

(2) $\Psi = \text{Hom}_R(R, J)$, wo J ein R-*Modul* ist. $H^p_{\text{Typ }\Psi}(C, J) = H^p_R(C, J)$ ist die bezüglich R zulässige Cohomologiegruppe von C. — Man sieht leicht ein, dass in diesem Falle $\Psi \cong J$ ist: die Zuordnung $\psi \to \psi(e)$ vermittelt den Isomorphismus.

(3) $\Psi = \text{Hom}_Q(R, J)$, wo Q ein Teilring von R ist und J ein Q-Modul. $H^p_{\text{Typ }\Psi}(C, J)$ ist die bezüglich Q zulässige Cohomologiegruppe $H^p_Q(C, J)$.

In den folgenden Beispielen sei $R = R_G$ der *Gruppenring* der Gruppe G. Die Elemente $\varphi \in \Phi$ sind durch ihre Werte $\varphi(x)$, $x \in G$, bestimmt, also beliebige Funktionen von G nach J. — Wir schreiben kurz statt Hom_{R_G} stets Hom_G und statt $H^p_{R_G}$ stets H^p_G.

(4) Ψ bestehe aus den Funktionen $\psi(x) = x.\psi(e)$, wo J ein R_G-Modul und e das Neutralelement von G ist. Z.B. $\psi(x) =$ konstant. Dieser Typus liefert im singulären Komplex $C(U)$ des Ueberlagerungsraumes U von S (vgl. § 1) die Gruppen $H^p(S, J)$ des Raumes S.

(5) G sei unendlich, und Ψ bestehe aus den Funktionen ψ, welche nur für endlich viele $x \in G$ einen Wert $\neq 0$ haben. $H^p_{\text{Typ }\Psi}$ heisst in diesem Falle die G-endliche Cohomologiegruppe von C, $H^p_{G\text{-endl.}}(C, J)$. — Für $J = Z$ ist Ψ, wie leicht zu sehen, zu R_G selbst R_G-isomorph.

(6) g sei eine Untergruppe von G, und $\Psi = \text{Hom}_g(R_G, J)$, wo J ein R_G-Modul ist. $H^p_{\text{Typ }\Psi}(C, J) = H^p_g(C, J)$ ist die bezüglich $g \subset G$ zulässige Cohomologiegruppe von C. — Wir nennen Ψ den *durch den R_g-Modul J induzierten R_G-Modul*; ist J der Modul einer linearen Darstellung von g, so ist Ψ derjenige der induzierten linearen Darstellung von G, im üblichen Sinne.

Aus diesen Beispielen ist zu ersehen, dass in topologischen und algebraischen Anwendungen Cohomologiegruppen verschiedener Typen Ψ auftreten.

3. Alle diese Typen lassen sich nun auf *einen einzigen, den „zulässigen" zurückführen*, und zwar bezüglich eines einzigen Operatorenringes R. Es gilt nämlich, für alle p

$$H^p_{\text{Typ }\Psi}(C, J) \cong H^p_R(C, \Psi).$$

Zum Beweis ordnet man jeder Cokette f^p vom Typus Ψ eine Cokette F^p mit Werten in Ψ zu: $F^p(c_p) = f^p(rc_p)$, als Funktion ψ von $r \in R$ aufgefasst. F^p ist zulässig bezüglich R, und $f^p \to F^p$ liefert den gewünschten Isomorphismus. — Aus den Beispielen in § 2 ergibt sich unmittelbar:

(1') $H_Z^p(C, J) \cong H_R^p(C, \Phi)$

(2') $H_R^p(C, J) \cong H_R^p(C, \Psi)$. — Wegen $\Psi \cong J$ ist dies nichtssagend. Analog bei Beispiel (4).

(3') $H_Q^p(C, J) \cong H_R^p(C, \Psi)$, wo Q ein Teilring von R, J ein Q-Modul und $\Psi = \mathrm{Hom}_Q(R, J)$ ist.

(5') $H_{G\text{-endl.}}^p(C, J) \cong H_G^p(C, R_G)$.

(6') $H_g^p(C, J) \cong H_G^p(C, \Psi)$, wo g eine Untergruppe von G, J ein R_g-Modul und Ψ der induzierte R_G-Modul ist.

Diese formalen Aussagen, gültig ohne weitere Voraussetzungen über den Komplex C, werden nun geometrisch und algebraisch angewendet.

4. *Ueberlagerungen.* Es sei $\overline{S}$ eine Ueberlagerung von S, die *nicht regulär* zu sein braucht, U die universelle Ueberlagerung von S, also auch von $\overline{S}$; die Decktransformationengruppe G von U über S enthält diejenige g von U über $\overline{S}$ als Untergruppe (g ist dann und nur dann Normalteiler von G, wenn $\overline{S}$ über S regulär ist). Gemäss § 1 ist $\mathbf{H}^p(S, J) \cong H^p(C(U), J)$, wo $C(U)$ der singuläre Komplex von U ist und J ein R_G-Modul. Ebenso gilt, für einen R_G-Modul J, $\mathbf{H}^p(\overline{S}, J) \cong H_g^p(C(\mathrm{U}), J)$; nach § 3, (6') ist diese Gruppe $\cong H_G^p(C(U), \Psi)$, also $\cong \mathbf{H}^p(S, \Psi)$. Wir erhalten somit die Isomorphie

$$\mathbf{H}^p(\overline{S}, J) \cong \mathbf{H}^p(S, \Psi),$$

gültig für eine beliebige Ueberlagerung $\overline{S}$ *von* S, *zur Untergruppe* g *von* G *gehörend, und für einen* R_g*-Modul* J; Ψ *ist der von* J *induzierte* R_G*-Modul* (G operiert darin i.A. nichttrivial, selbst wenn g in J trivial operiert).

Ist hierbei J sogar ein R_G-Modul (z.B. der triviale), so existiert eine natürliche *Einbettung* q von $J \cong \mathrm{Hom}_G(R_G, J)$ in $\Psi = \mathrm{Hom}_g(R_G, J)$. Sie induziert einen Homomorphismus $\mathfrak{Q}$ von $\mathbf{H}^p(S, J)$ in $\mathbf{H}^p(S, \Psi) \cong \mathbf{H}^p(\overline{S}, J)$; man verifiziert leicht, dass $\mathfrak{Q}$ die *zur Ueberlagerungsprojektion* u *von* $\overline{S}$ *auf* S *duale Abbildung* ist. Nach geläufigen Schlussweisen gibt q: $J \to \Psi$ sogar Anlass zu einer exakten Sequenz, welche $\mathbf{H}^p(S, J)$, $\mathbf{H}^p(S, J)$ und $\mathbf{H}^p(S, \Psi/J)$ verknüpft. Solche explizite Zusammenhänge zwischen induzierten Darstellungen und Ueberlagerungsräumen gestatten Anwendungen in beiden Richtungen.

Ein wichtiger Fall ist der, wo die Ueberlagerung $\overline{S}$ von S von *endlicher* Blätterzahl n (= Index von g in G) ist. Dann lässt sich der Einbettung q von J in Ψ eine geometrisch und algebraisch naheliegende Abbildung t von Ψ in J zur Seite stellen; man wählt Repräsentanten $x_1, \ldots, x_n$ der Rechtsrestklassen von G mod g und setzt für $\psi \in \Psi$, $t(\psi) = \Sigma x_i^{-1} \psi(x_i)$. Man sieht leicht ein, [3]) dass $t(\psi)$ von den Repräsentanten nicht abhängt, dass t ein R_G-Homomorphis-

[3]) vgl. [2], wo auch verschiedene algebraische Anwendungen zu finden sind.

mus von Ψ in J ist, und dass $tq\ (\tau) = n\tau$ ist für $\tau \in J$, kurz $tq = M_n$ geschrieben[4]). Die durch t induzierte Abbildung $\mathfrak{T}$ von $\mathbf{H}^p\ (\overline{S},\ J)$ in $\mathbf{H}^p\ (S,\ J)$ heisst *Verlagerung* (oder *Transfer*), und es gilt: $\mathfrak{T}\mathfrak{Q} = M_n$ (in $\mathbf{H}^p\ (S,\ J)$). Ein Anwendungsbeispiel[3] Ist M_n in $\mathbf{H}^p(S,\ J)$ *eineindeutig*, dann auch $\mathfrak{Q}$ von $\mathbf{H}^p(\overline{S},\ J)$, also ist dann $\mathbf{H}^p(S,\ J)$ *einer Untergruppe von* $\mathbf{H}^p(\overline{S},\ J)$ *isomorph* — Folgerungen bezüglich der Bettischen Zahlen, Torsionskoeffizienten usw. liegen auf der Hand.

5. *Freie azyklische Komplexe; algebraischer Aspekt.* Wir nehmen an, der R-Komplex C sei (a) frei bezüglich R, (*b*) azyklisch in allen Dimensionen. (*a*) besagt, dass alle C_p für $p \geqq 0$ freie R-Moduln sind, (*b*) dass in C_p, $p \geqq 0$, jeder Zyklus ein Rand ist. Offenbar hat (*b*) mit dem Operatorenring nichts zu tun. Ist R frei über Z, was wir hier zur Vereinfachung *) annehmen wollen, so ist wegen (*a*) der Komplex C auch frei über Z (Beispiel: $R = R_G$). — Nach bekannten Schlüssen gilt für die Cohomologie von C:

I. Zu jeder *exakten Sequenz* von R-Moduln $0 \to J \xrightarrow{j} J^0 \xrightarrow{\pi} J' \to 0$ gehört die exakte Sequenz $H^0(J) \to \ldots \to H^p(J) \xrightarrow{j_*} H^p\ (J_0) \xrightarrow{\pi_*} H^p(J') \xrightarrow{\Delta} H^{p+1}\ (J) \to \ldots$ Dabei handelt es sich um die zulässigen Cohomologiegruppen $H^p_R\ (C,\ J)$ usw; das Argument C und den Index R lassen wir jetzt meistens weg. j_* und π_* sind die durch j bzw. π induzierten Homomorphismen, und Δ ist in geläufiger Weise durch die Operation δ definiert.

II. $H^p_Z(J) = 0$ für $p \geqq 1$.

III. $H^0(J) = \mathrm{Hom}_R(C_{-1},\ J)$, und für einen R-Homomorphismus h von J in J_1 ist $h_*H^0\ (J) \to H^0\ (J_1)$ durch Zusammensetzen der Elemente von $\mathrm{Hom}_R(C_{-1},\ J)$ mit h gegeben.

Nach § 3, (1′) ist $H^p_Z\ (J) \cong H^p\ (\Phi)$, wo $\Phi = \mathrm{Hom}_Z(R,\ J)$. Bilden wir nun vermöge der durch $J \cong \mathrm{Hom}_R\ (R,\ J)$ gegebenen Einbettung j von J in Φ die exakte Folge $0 \to J \xrightarrow{j} \Phi \xrightarrow{\pi} \Phi/J \to 0$, so folgt:

$H^p\ (\Phi/J) \cong H^{p+1}\ (J)$, für $p \geqq 1$,

$H^0\ (\Phi/J)/\pi_*H^0\ (\Phi) \cong H^1\ (J)$.

Dadurch sind alle $H^p(J)$ auf C_{-1} zurückgeführt, also *bestimmt durch* R, C_{-1} *und* J, *und im übrigen vom Komplex* C *unabhängig*. Die Zurückführung ist ganz explizit und ergibt für die $H^p(J)$ Ausdrücke in Iterationen von $\Phi = \mathrm{Hom}_Z(R, J)$, Φ/J und $\mathrm{Hom}_R(C_{-1},\ J)$. Ferner gibt es *zu jedem* R-*Modul* M einen freien azyklischen R-Komplex mit $C_{-1} = M$.

---

[4]) Für jede additive Abelsche Gruppe sei M_n der durch Multiplikation mit der ganzen Zahl n definierte Endomorphismus.

*) Die folgenden Betrachtungen gelten auch ohne diese Voraussetzung, wenn der Begriff eines „R-injektiven" Moduls verwendet wird; $\Phi = \mathrm{Hom}_Z\ (R,\ J)$ ist dann durch eine R-injektive Erweiterung von J (vgl. B. Eckmann u. A. Schopf, Arch. Math. IV (1953), 75—78) zu ersetzen.

Einem R-Modul M ist also, für jeden R-Modul J, eine Folge von Gruppen $H^p(J)$, $p = 0, 1, 2, \ldots$ zugeordnet, für $h: J \to J_1$ eine Folge induzierter Homomorphismen h_*: $H^p(J) \to H^p(J)_1$ und für $J_0/J = J'$ eine Folge von Homomorphismen $\Delta : H^p(J') \to H^{p+1}(J)$. Dabei gilt I. Exaktheit der Cohomologiesequenz, II. $H^p(\Phi) = 0$ für $p \geqq 1$ und $\Phi = \mathrm{Hom}_Z(R, J)$. III. $H^0(J) = \mathrm{Hom}_R(M, J)$, mit dem entsprechenden h_*. Offenbar kann man I. — III. als *Axiome* der *„Cohomologietheorie des R-Moduls M"* betrachten; sie *lassen sich stets erfüllen, und bis auf Isomorphie der $H^p(J)$ nur auf eine Art* (sollen auch die induzierten Homomorphismen eindeutig festgelegt sein, d.h. bei dieser Isomorphie einander entsprechen, so braucht es weitere, naheliegende Axiome).

Ist R der Gruppenring R_G einer Gruppe G und $M = Z$, als trivialer R_G-Modul aufgefasst, so spricht man von den *Cohomologiegruppen $H^p(G, J)$* der *Gruppe G*, bezüglich des R_G-Moduls J. Offenbar braucht dabei G keine Gruppe zu sein (nur ein „Monoid)"

Ist R eine *Algebra A* über einer Körper K, so betrachtet man gewöhnlich durchwegs beidseitige A-Moduln und setzt $M = A$ (als 2-seitiger A-Modul). Die zugehörigen Gruppen $H^p(J)$, für einen 2-seitigen A-Modul J, heissen die Cohomologiegruppen der Algebra A. Der Körper K hat in diesem Falle in allen unsern Betrachtungen die Rolle von Z zu übernehmen.

6. Die Bedeutung der Abelschen Gruppen $H^p(G, J)$, die in dieser Weise einer beliebigen Gruppe G zugeordnet worden sind, lässt sich auf verschiedene Arten untersuchen: 1. Die durch die Axiome I. — III. gegebene explizite Beschreibung vermittelst Φ, Φ/J usw. liefert u.a. folgende Aussagen: $H^0(G, J)$ ist der Teilmodul aller bei G invarianten Elemente von J; $H^1(G, J)$ ist die sog. Crosshomomorphismengruppe von G nach J; $H^2/(G, J)$ ist die Gruppe der Erweiterungen von J durch G, die zu den Operatoren in J gehören. Operiert G trivial in J, so ist $H^0(G, J) = J$; $H^1(G, J) = \mathrm{Hom}(G, J)$; $H^2/(G, J)$ die Gruppe der zentralen Erweiterungen von J durch G. — 2. Jeder freie azyklische R_G-Komplex C mit $C_{-1} = Z$ (trivial) ist ein *„Modell" für die Cohomologietheorie von G*, da $H^p(G, J) \cong H^p(C, J)$ ist. Solche Modelle lassen sich in sehr einfacher Weise rein algebraisch konstruieren, z.B. durch mehrfache Tensorprodukte von R_G; man erhält dann eine Beschreibung [5]) der $H^p(G, J)$ durch *Funktionen mehrerer Variablen $x \in G$* mit Werten in J, die bestimmten Relationen genügen. Aehnlich verhält es sich in der Cohomologie der Algebren.—3. Es gibt in manchen Fällen *geometrische Modelle*, nämlich dann, wenn G die Fundamentalgruppe eines zusammenhängenden Raumes S mit *azyklischer* universeller Ueberlagerung U ist; der singuläre Komplex $C(U)$ ist ein Modell für die $H^p(G, J)$. Hieraus

[5]) vgl. [1] und [3], sowie die dort erwähnten Publikationen von Eilenberg-MacLane und von Eckmann.

ergeben sich die bekannten *algebraischen Beziehungen zwischen der Fundamentalgruppe G von S und den* $H^p(S, J)$ sowie weitern Homologieeigenschaften von S. — Wenn man sich nur für die H^p, $p \leqq N$ interessiert, so braucht U nur in diesen Dimensionen $p \leqq N$ azyklisch zu sein (was für $N = 1$ stets der Fall ist).

Ist g eine *Untergruppe* von G, so ist jedes Modell C für G auch eines für g; aus $H_g^p(C, J) = H_G^p(C, \Psi)$, wo J ein R_g-Modul und $\Psi = \mathrm{Hom}_g(R_G, J)$ ist, erhalten wir die Relation

$$H^p(g, J) \cong H^p(G, \Psi) \qquad \text{für alle } p \geqq 0.$$

Genau wie im Falle der Ueberlagerungen $\overline{S}$ von S (§ 4) ist auch hier, für einen R_G-Modul J, die Abbildung $\mathfrak{Q}$ von $H^p(G, J)$ in $H^p(g, J)$ erklärt und, wenn der Index n von g in G endlich ist, die *Verlagerung* $\mathfrak{T}$ von $H^p(g, J)$ in $H^p(G, J)$. Am geeigneten Modell verifiziert man [6]): $\mathfrak{Q}$ ist gegeben durch die *Beschränkung* der Funktionen mehrerer Variablen $x \in G$ auf g; $\mathfrak{T}$ ist, für $p = 1$ und bei trivialen Operatoren in J, eine Abbildung von Hom (g, J) in Hom (G, J) die zur klassischen *„Verlagerung von G in die Untergruppe g"* dual ist. Es gilt wiederum $\mathfrak{T}\mathfrak{Q} = \mathrm{M}$, und die Anwendungen [6]) sind analog denen bei Ueberlagerungen. $\mathfrak{T}$ verallgemeinert also die klassische gruppentheoretische Verlagerung dual auf Dimensionen $p \geqq 1$ und auf nichttriviale Operatoren, und stellt sie in den Rahmen weniger Axiome und Festsetzungen, die einer einfachen geometrischen Situation entspringen.

7. Zum Abschluss eine Bemerkung über die Gruppen $H^p(G, R_G)$ einer Gruppe G ($J = R_G$, als R_G-Modul betrachtet). Sie sind einerseits verknüpft mit den *„Derivationen* in einer Gruppe G" im Sinne von Fox [7]): $H^1(G, R_G)$ ist die Gruppe $\mathfrak{D}$ aller Derivationen in G modulo der trivialen Derivationen, und $H^p(G, R_G)$ eine natürliche Verallgemeinerung dieses Begriffs, der Aussagen über G liefert. Andererseits ist, in einem Modell C für G, $H^p(G, R_G) \cong \mathrm{H}^p(C, R) = H^p_{G\text{—endl.}}(C, Z)$, nach (6'). Es sei z.B. C die universelle Ueberlagerung U eines endlichen Polyeders S mit der Fundamentalgruppe G (ein solches existiert, wenn G von endlich vielen Elementen erzeugt ist), als simplizialer Komplex aufgefasst, und U sei azyklisch. Dann ist $H^p_{G\text{—endl.}}(U, J) = \mathbf{H}^p_{\text{endl.}}(U, J)$ die mit *endlichen* Coketten gebildete gewöhnliche Cohomologiegruppe von U, und es gilt

$$H^p(G, R_G) \cong \mathbf{H}^p_{\text{endl.}}(U, Z).$$

Die linke Seite hängt nur von G, die rechte nur von U ab: man erhält eine Be-

---

[6]) Ausführliche Darstellung und Anwendungen s. [2].

[7]) Hinweise auf die hier verwendeten Begriffe und Sätze, insbesondere von Hopf und Specker, s. [1], Abschnitt 7.

ziehung zwischen den „Derivationen" in der Fundamentalgruppe des endlichen Polyeders S und topologischen Eigenschaften seiner universellen Ueberlagerung U (falls diese azyklisch ist).

Beschränken wir uns auf $p = 1$: $\mathbf{H}^1_{\text{endl.}}(U, Z)$ ist [7]) eine freie Abelsche Gruppe vom Rang $e - 1$, wo e die *Endenzahl* [7]) von U ist (Rang 0, wenn U endlich ist). Somit ist e durch G bestimmt, $= 0$ wenn G endlich, $=$ Rang von $H^1(G, R_G) + 1$, wenn G unendlich ist. e heisst die Endenzahl der Gruppe G; sie ist übrigens [7]) stets $= 0, 1, 2$ oder ∞. Die Derivationsgruppe $\mathfrak{D}$ ist somit eine freie Abelsche Gruppe vom Rang $e - 1 = 0,\ 1$ oder ∞.

LITERATUR

[1] B. ECKMANN, On complexes with operators, Proc. Natl. Acad. Sci. U.S.A., **39**, 35—42 (1953).

[2] B. ECKMANN, Cohomology of groups and transfer, Ann. Math. **58**, 481—493 (1953).

[3] S. EILENBERG, Topological methods in abstract algebra. Cohomology of groups. Bull. Am. Math. Soc. **55**, 3—37 (1949).

LA TOPOLOGIE DANS LES FONDEMENTS DE LA GÉOMETRIE

HANS FREUDENTHAL

En 1847, dans l'oeuvre de von Staudt, la topologie frappa à la porte de la géométrie. Après cette rencontre il s'écoulera presque un demi-siècle, avant qu'elle ne soit vraiment admise.

Vous connaissez l'artifice de von Staudt qui définit la projectivité comme une transformation conservant la relation d'harmonicité, artifice fort élégant et en avance sur son temps, qui n'avait pas encore cultivé le goût des définitions implicites.

Pour von Staudt cette définition était le moyen indispensable pour démontrer le théorème fondamental de la géométrie projective, sans traduire les notions géométriques dans le langage algébrique. Von Staudt considère l'ensemble des points invariants de la projectivité donnée, mais il le fait avec l'oeil du précantorien. Il ne voit que deux possibilités: un ensemble est ou continu ou discontinu. Dans le premier cas il contient un segment, dont la fermeture harmonique donne déjà la droite entière. Dans l'autre cas l'ensemble doit posséder deux points successifs a et b; mais s'il y avait alors un troisième point invariant c, hors du segment ab, son conjugé harmonique serait située à l'intérieur de ab, ce qui refute l'existence d'un troisième point invariant et prouve le théorème fondamental.

Felix Klein a jeté le premier doute sur ce raisonnement. Ce fait étonnant s'est passé en 1873 — étonnant parce que le critique s'appelait Klein et non par exemple Weierstrass. L'analyse logique n'était pas le fort de Klein, et ce qu'il a écrit sur cette question pendant les années suivantes, était aussi confus que possible.

L'analyse du continu arithmétique réalisée par Cantor et Dedekind (1872) n'avait pas encore atteint les géomètres, pour lesquels le continu géométrique restait une donnée intuitive, qu'on n'analysait pas.

En outre, les analyses de Cantor et Dedekind devaient être approfondies pour les besoins de la géométrie. On ne s'était pas encore attaqué à la notion d'ordre, car on était toujours parti de l'arithmétique et des nombres. Un événement important fut la redécouverte de l'axiome d'Archimède par O. Stolz (1881—3), qui par la lecture des Eléments d'Euclide réussissait à regagner le

niveau d'Eudoxe. Le premier système axiomatique où l'ordre et la continuité sont bien distingués, est celui de P. Veronese (1890).

Mais en vérité on s'était engagé dans une impasse. Par la notion d'ordre on avait fondé la version réelle de la géométrie et on en avait bloqué la version complexe, c'est-à-dire justement celle qu'on avait l'habitude d'explorer. En 1899 Hilbert découvrit la connexion qui existe entre les axiomes d'incidence et les lois algébriques des corps. La géométrie projective sur un corps quelconque était née. On apprend à traduire des axiomes géométriques par des axiomes algébriques et inversement, que Pappus-Pascal correspond à la commutativité du corps, que les axiomes d'ordre et de continuité peuvent s'énoncer dans deux langages parallèles.

Grâce au théorème de Pappus-Pascal on peut se passer des notions de continuité, qui dans la conception de Dedekind dépendent de l'ordre linéaire, obstacle sur la voie à la géométrie complexe. Mais la recherche des corps topologiques inaugurée par D. van Dantzig (1931) a ouvert de nouvelles perspectives. Le postulat de la compacité locale conduit à une réduction effective des corps possibles (van Dantzig, Pontryaguine, 1932). La notion de continuité s'introduit dans la géométrie sous un nouvel aspect: celui de la compacité. Il n'y a que trois géométries projectives compactes et connexes à dimension positive: réelle, complexe et quaternionienne. Dans ces géométries les droites sont des sphères à 1, 2 et 4 dimensions, les plans et hyperplans y sont des variétés topologiques, dont la dimension topologique est égale à la dimension géométrique multipliée par 1, 2 ou 4.

Jusqu'ici nous avons supposé tacitement le théorème de Desargues, qui dans la géométrie spatiale est une conséquence des axiomes triviaux d'incidence. L'indépendance du théorème de Desargues dans le plan était déjà connue de Klein ou même de Beltrami. Le premier exemple formel d'une géométrie non-arguésienne a été donné par Hilbert (1899). En généralisant cet exemple, on peut dire que les rayons lumineux d'un plan assez arbitraire à indice de réfraction variable (mais où il n'y a pas de réflexion totale) peuvent être considérés comme les droites d'une géométrie non-arguésienne.

Donc même les postulats topologiques n'ont pas le pouvoir d'imposer au plan projectif le carcactère arguésien. On peut aller plus loin, en passant à la topologie différentielle et en postulant que la géométrie projective plane soit d'une classe différentielle k — cela veut dire que ses systèmes locaux de coordonnées dépendent l'un de l'autre par des fonctions différentiables jusqu'à l'ordre k et que la relation d'incidence est aussi différentiable de l'ordre k. Mais cela ne suffit pas: on peut construire des géométries non-arguésiennes, qui possèdent ce degré de régularité; il est vraisemblable qu'on peut même admettre la valeur $k = \infty$. Au contraire on ne sait pas, s'il y a des géométries planes non-

arguésiennes, dont le degré de régularité est celui de l'analyticité réelle.

A Guy Hirsch (1946), qui a orienté les recherches dans cette direction, nous devons une conception plus large: On peut maintenir le postulat topologique que le plan projectif soit une variété topologique et que ses droites soient des sous-variétés situées assez régulièrement, tout en admettant des dimensions topologiques supérieures à 2 et 1 pour le plan et ses droites. Sans entrer dans les détails quant aux conditions de régularité, je résume le résultat: la dimension d'un plan projectif est toujours une puissance de 2; la droite projective est une sphère, dont la dimension est la moitié de celle du plan. La démonstration de ce théorème fait appel à la théorie topologique des espaces fibrés.

Le désir de se libérer des conditions de régularité dans l'énoncé de ce théorème, est assez naturel. On chercherait toutes les géométries qui satisfont aux axiomes d'incidence triviaux du plan projectif et aux postulats topologiques qui suivent: le plan est un espace compact, les droites en sont des sous-ensembles fermés et connexes, la droite joignant deux points et le point d'intersection de deux droites sont fonctions continues de ces deux arguments. Tout ce qu'on peut déduire jusqu'ici de ces données, c'est que les droites de ces géométries jouissent d'une propriété très forte de contractilité, à savoir: chaque vrai sous-ensemble fermé de la droite est contractile en un seul point sur la droite. Ce théorème peut être obtenu avec des méthodes bien connues de la théorie des bouts de groupes. Il y a un système de projectivités laissant invariants deux points (0 et ∞) et transformant un point donné 1 dans un point quelconque $a(\neq 0, \neq \infty)$, tel que le produit ab qui donne l'image de b par la transformation appliquant 1 en a, est continu et biunivoque. Les 0 et ∞ sont alors les bouts du reste de la droite projective (qui est un „Schiebraum"), d'où l'on conclut à la contractilité que je viens d'énoncer.

En outre nos droites projectives sont d'une homogénéité excessive. Est-il trop téméraire de conjecturer que ces plans projectifs sont nécessairement des plans de Hirsch et donc des variétés dont la dimension est une puissance de deux? Mais dans l'état actuel de la topologie, nous ne disposons pas d'outils pour étudier effectivement des espaces homogenes, si cette homogénéité ne se manifeste pas dans un groupe localement compact.

Quant aux cas spéciaux on connaît des géométries planes projectives à 2, 4, 8 et 16 dimensions, les plans projectifs sur le corps réel, complexe, quaternionien et sur le corps alternatif des octaves. La construction habituelle des trois premiers, où l'on représente les points et les droites par des triples à un facteur à gauche ou à droite près et la relation d'incidence par une équation $\xi_1\omega_1 + \xi_2\omega_2 + \xi_3\omega_3 = 0$, ne fonctionne plus pour les octaves qui n'obéissent pas à la loi associative. Cependant on peut parvenir au plan projectif des octaves en ajoutant au plan affin des octaves de Mlle. Moufang une droite à l'infini ou en appliquant une

méthode topologique élaborée par G. Hirsch. Une construction plus symétrique qui revèle aussi le groupe projectif de ce plan, est celle où on emploie l'algèbre exceptionnelle de P. Jordan. On considère des matrices, 3×3, hermitiennes, avec des octaves pour coefficients, et on en fait une algèbre commutative en symétrisant le produit matriciel ordinaire. Les points et les droites du plan sont les idempotents de cette algèbre, la relation d'incidence est $X \circ U = 0$. Le groupe projectif de ce plan est une forme réelle du E_6 exceptionnel, ses groupes elliptique et hyperbolique sont des formes du F_4. Cette géométrie est non-arguésienne, mais elle obéit à la loi de l'unicité du conjugué harmonique, loi plus faible que le théorème de Desargues. Elle est analytique et même algébrique — il y a des raisons de conjecturer que, pour un nombre quelconque de dimensions, l'analyticité implique l'harmonicité. Reste encore à confesser que nous ne savons rien sur l'existence possible ou l'impossibilité de géométries projectives planes à un nombre plus élevé de dimensions. La réponse dépend de problèmes non résolus de la théorie des espaces fibrés. Une autre question est celle de la signification géometrique des E_7 et E_8; elle a trouvé une réponse partielle, mais pas encore satistaisante.

A cette occasion il convient de signaler que quelques résultats de la géométrie des octaves que j'ai publiés au cours de ces dernières années, ont été déjà annoncés par P. Jordan dans un article que je n'ai connu que beaucoup plus tard.

La méthode de recherche fondamentale géométrique, où le point de départ est la relation d'incidence, n'est pas la seule. L'autre est celle où le premier rôle revient aux transformatiosn géométriques.

Riemann, dans son discours fameux de 1854, cherchait parmi les variétés auxquelles on a donné son nom, celles dans lesquelles les corps rigides peuvent mener une existence indépendante de leur place, et il trouvait les espaces à courbure constante. Helmholtz, au début indépendemment de Riemann, précisait la notion de la mobilité des figures (1868); d'autre part il partait d'un invariant assez arbitraire d'un couple de points au lieu de la métrique riemannienne. Les définitions toujours vagues de Helmholtz ont été améliorées par Lie (1890), qui y introduisit la notion de groupe: L'espace est une variété à n dimensions assez régulière régi par le groupe des transformations compatibles avec l'invariant prescrit. La libre mobilité veut dire que, $k \leqq n - 2$ points généraux étant fixés, un $(k + 1)$-ème peut occuper grace au groupe toutes les positions compatibles avec l'invariant, tandis que $n - 1$ points étant fixés, il n'y a plus de mouvement. En vérité au lieu du problème proposé, Helmholtz résolvait celui que Lie a appelé „libre mobilité infinitésimale"; Lie les a traités tous les deux, le problème fini avec des méthodes qui ne sont pas à l'abri de toute reproche du point de vue moderne. En tout cas les espaces qui résultent

du problème, sont de nouveau ceux à courbure constante. Helmholtz croit son analyse supérieure à celle de Riemann, parce qu'il réussit à démontrer le caractère pythagoricien de l'invariant, introduit comme hypothèse par Riemann. Qu'il admette l'existence des déplacements dès le début, cette limitation ne lui semble pas sérieuse, parce que d'apès lui, sans cette notion de congruence, il n'y a pas non plus de comparaison de longueurs, importante pour l'existence d'une métrique. Il est clair que Helmholtz se trompe sur ce point. Pour comparer des longueurs on doit disposer, non de corps rigides, mais seulement d'étalons, dont la dimension longitudinale est rigide, et dont les dimensions transversales peuvent être négligées. C'est justement l'idée de Riemann.

En 1902, Hilbert a repris ces recherches en leur donnant une tournure purement topologique. En 1928 dans une note importante, mais dépourvue de démonstrations, Kolmogorov a généralisé les idées de Hilbert à n dimensions. Des recherches plutôt partielles ont été faites par Brouwer, R. L. Moore, Lubben, Süss, Kerékjártó, Montgomery et Zippin. La solution du cinquième problème de Hilbert, même quand elle n'était pas encore complète, a influencé le traitement du problème de Helmholtz-Lie. En effet, ensemble avec des idées algébriques elle peut servir à caractériser topologiquement certains groupes géométriques. Un des plus beaux résultats a été obtenu par J. Tits (1951) qui précédé par Kérékjarto et suivi par le conférencier, a pu classifier les groupes triplement transitifs: un tel groupe opérant dans un espace localement compact, métrisable et non totalement discontinu est essentiellement le groupe projectif de la droite réelle ou complexe. J. Tits a aussi classifié les groupes doublement transitifs, qui sous ls mêmes conditions coincident avec les groupes des transformations linéaires et entières de la droite réelle, complexe ou quaternionienne. Tout récemment A. Borel a trouvé de vastes généralisations de ces résultats, qui se trouvent partiellement aussi dans un mémoire inédit de J. Tits.

Sous une nouvelle forme, celle de la métrique, Garret Birkhoff (1944) a ranimé l'invariant des recherches classiques. Pour caractériser les espaces à courbure constante, il a étudié les isométries des espaces métriques. Continuant des travaux de Busemann, H. C. Wang (1951) a réussi à classifier les espaces métriques compacts et connexes où chaque couple de points peut être transformé par isométrie dans chaque couple congruent. Pour les espaces localement compacts il a fait de même sous des conditions supplémentaires, parmi lesquelles celle d'une dimensions impaire est la plus essentiele. Outre les espaces classiques du problème de Helmholtz-Lie et leurs analogues complexes et quaternioniens, il trouve le plan elliptique des octaves.

La plus grande clarté a été atteinte par J. Tits (1953), qui est retourné au système de Kolmogorov, tout en embrassant en même temps les résultats de Wang. Cependant il faut remarquer que sa liste est très incomplète.

D'après ces recherches, l'état de choses est le suivant. Il est assez naturel de postuler la compacité locale, la connexité et la métrisabilité de l'epace où le groupe opère, et la transitivité du groupe même. Ce sont des postulats triviaux.

Puis il faut demander une espèce de rigidité de l'espace. Cela peut être l'existence de l'invariant d'un couple de points semblable à celle postulée par Helmholtz-Lie, ou, postulat plus fort, qui n'est pas strictement nécessaire, l'existence d'une métrique invariante. Le même résultat s'obtient par un dispositif ingénieux de Hilbert, qu'on retrouve chez Kolmogorov sous le nom de continuité uniforme du groupe, et auquel Zippin, sous une forme affaiblie a donné le nom de rigidité topologique: Deux points a et b étant donnés, il y a un ensemble ouvert, dont les transformés ne rencontrent pas en même temps des voisinages de a et b. Il est remarquable que cette condition permet de pourvoir l'espace d'une mesure invariante.

Cette condition garantit aussi la compacité du groupe de stabilité ou d'isotropie d'un point quelconque. D'autre part on a besoin d'un postulat qui fait que sous ce groupe l'orbite d'un point soit assez étendue. L'homogénéité métrique à deux points de Wang veut dire que cette orbite est précisément une sphère métrique. La condition de Kolmogorov que deux orbites quelconques sont situées comme deux sphères au sens de la séparation topologique, a les mêmes conséquences, mais des conditions beaucoup plus faibles suffisent toujours, par exemple celle qu'une seule orbite partage l'espace (ou le partage localement). La condition d'homéomorphie de deux orbites quelconques conduit à des résultats assez voisins. Dans le cas de deux dimensions Hilbert a pu se contenter d'exiger que l'orbite soit infinie.

L'analyticité du groupe n'exige pas de nouvelles suppositions, grâce à la solution récente du cinquième problème de Hilbert. On peut alors continuer l'analyse avec les méthodes des groupes de Lie. Le groupe de stabilité est linéaire et compact, donc orthogonal. Il est transitif sur la sphère. La théorie de Cartan permet de classifier les groupes orthogonaux transitifs sur la sphère. Puis il faut chercher les groupes transitifs qui possèdent un de ces groupes comme groupe de stabilité. Ce problème devient plus clair si l'on cherche tous les groupes qui possèdent un groupe orthogonal irréductible comme groupe de stabilité. On démontre qu'ils sont ou semi-simples ou que leur radical est un groupe de translations pareil à celui de l'espace euclidien. Le dernier cas est le plus facile, le premier revient à la recherche des sous-groupes semi-simples de groupes semi-simples, achevée par Malcev et Dynkin, mais plus élémentaire au cas présent.

Ainsi on arrive à une classification des variétés riemanniennes, dont le groupe d'isotropie est irréductible, et qui admettent aussi une définition pure-

ment topologique. Pour caractériser les groupes classiques du problème de Helmholtz-Lie, on introduira des conditions supplémentaires. Les plus naturelles sont celles qui augmentent le degré de transitivité (ou de mobilité). On voit aisément que la libre mobilité est beaucoup trop forte. La mobilité double marque les groupes de Tits, qui sont toujours assez nombreux. La mobilité triple engendre justement les espaces à courbure constante, car elle permet de transformer chaque élément plan dans chaque autre et donc de constater la constance de la courbure. C'est là le postulat qui a été étudié par Busemann et qui dit que chaque triangle peut être transformé dans chaque triangle congruent par une isométrie de l'espace.

TOPOLOGICAL TRANSFORMATION GROUPS

DEANE MONTGOMERY

1. *Introduction*

This note will summarize some of the recent work on topological groups and discuss a few topics in transformation groups mainly in S^3 and S^4. In one aspect of this subject, namely the relation of general locally compact groups to Lie groups, information is now fairly complete. However in most other aspects the situation is far less satisfactory and in many areas known results are fragmentary. In particular the study of compact and locally compact groups acting on manifolds is filled with unexplored areas and this is true in both the differentiable and non-differentiable cases.

2. *Locally compact groups*

Turning first to the best known situation, that is the one concerning the relation of locally compact groups to Lie groups, let G denote a topological group. By definition, G is a set of elements which satisfies a) G is a group, b) G is a T_0 space, c) the group operations of multiplication and the taking of inverses are continuous in the given topology.

If the space of G is locally euclidean and if coordinates at the identity may be so chosen that the local group operations are analytic then G is called a Lie group.

Hilbert [3] asked whether the following is true:

Theorem 1. *Every locally euclidean group is a Lie group.*

His question was more general and somewhat indefinite being concerned first with how far the theory of Lie groups and transformation groups can be developed without differentiability, and beyond that with functional equations. Some of the questions of the next section are suggested by Hilbert's remarks or are natural extensions of them. However the question of the truth of Theorem 1 above has often been called Hilbert's fifth problem. Theorem 1 was first proved to be true by the contents of two papers taken together, the first by Gleason [2], the secnd by Zippin and the author [7].

After topological groups had been studied it began to be felt that locally compact groups, which can in fact be more general than Lie groups, must nevertheless be closely related to them. For G compact this was verified by von Neu-

mann in 1934, for G abelian by Pontrjagin, and for G solvable by Chevalley and Malcev. These results also proved Theorem 1 for G with these same restrictions. For references see for example [7]. To be more precise about what is meant consider the definition below.

Definition. *A topological group G is said to be approximated by Lie groups if for every neighborhood U of the identity there is a compact invariant subgroup H, $H \subset U$, such that G/H is a Lie group.*

It was conjectured from about 1930 onwards by a number of mathematicians, not necessarily in the same terminology, that the following is true (if G is connected, $G = G'$).

Theorem 2. *Every locally compact group G contains an open and closed subgroup G' which is approximated by Lie groups.*

When G has finite dimension this theorem was first proved in the papers [2] and [7] taken together. By an ingenious method Gleason proved that if a locally compact finite dimensional G has no small subgroups it is a Lie group. Zippin and the author proved that if locally compact groups of dimension less than n can be approximated by Lie groups then a connected locally connected n-dimensional group contains a proper invariant subgroup F such that G/F has no small subgroup. By Gleason's theorem G/F is a Lie group; it can then be shown that G is Lie group. From this Theorem 1 and for finite dimensions Theorem 2 follow readily. Thus [2] and [7] taken together give an inductive proof of Theorem 1 and the finite-dimensional part of Theorem 2.

In the general (infinite-dimensional) case, Theorem 2 and the fact that a locally compact group without small subgroups is a Lie group were proved by Yamabe [12, 13] who used and made important improvements in the methods of Gleason. The methods of [7] are geometric and related to transformation groups; they are also related to methods used by Zippin and the author to prove the theorems in low dimensions. Kuranishi [5] had a very informative paper on these questions.

3. *Transformation groups.*

In this section let G be a compact topological group which acts as a transformation group on a separable manifold M, that is for each g, a homeomorphism $g(x) = f(g; x)$ of G onto itself is given such that 1) $g_1(g_2(x)) = (g_1g_2)(x)$and 2) $f(g; x)$ is simultaneously continuous. If only e, the identity of G, leaves every x in M fixed, G is called effective. It is known under the conditions above that an effective G is separable metric.

It is unknown whether such a compact G acting affectively on the manifold M must be a Lie group. This is known however in certain special cases, for example a) when each transformation is differentiable [5] and b) when G is

connected and M is three-dimensional [9]. To establish this conjecture in general it would be sufficient to establish that if such a compact effective G is zero-dimensional then it must be finite. From now on we shall consider the case where G is a compact Lie group.

To begin let G be a cyclic group of order two. Bing [1] has shown that G can act on E^3 or S^3 without being topologically equivalent to an orthogonal transformation. He constructs on orientation reversing homeomorphism T of period two of E^3 onto itself whose set of fixed points is homeomorphic to a plane, but whose two complementary domains are not simply connected. The example is constructed by identifying two solid horned spheres along a common boundary and showing that the resulting space is E^3; the transformation is a "reflection" across this boundary. His construction can be modified slightly [8] to obtain an *orientation preserving* homeomorphism T^* of period two of E^3 onto E^3. The set of fixed points of T^* is a wildly imbedded line. It would be interesting to have a similar example with prime period other than two.

The pathology in such examples can persist even when G is connected as has been shown in [8]. In that paper the transformation T^* above is constructed and is used to define an action of the circle group C on E^4 which is such that the group could not be differentiable in any differentiable structure of E^4; even less could this transformation group be orthogonal in any coordinate system of E^4. The action is such that the elelment 1/2 (taking C tot be he reals mod one) has a wildly imbedded plane of fixed points. It is this fact which shows the impossibility of differentiability or linearity properties in any coordinates for E^4.

While all of these examples fail to be linear, they nevertheless possess a great deal of regularity and the problem remains of determining how closely the action of a compact Lie group G resembles the differentiable and linear cases. In this connection see [9, 10, 11]. Quite apart from this, of course, is the question of determining geometric properties in the large when G is known to act differentiably or simplicially, for here too very little is known.

Now let T be an orientation preserving simplicial homeomorphism of S^3 of period 2. Smith [11] has shown even without the assumption that T is simplicial that F, the set of fixed points of T, must be a simple closed curve. However it is unknown whether T is equivalent to an orthogonal transformation and moreover it is unknown whether F can be knotted.

Samelson and the author have shown that at any rate F can not belong to a certain class of knots including as a special case the cloverleaf knot.

To explain more precisely the nature of the class of knots considered, a definition will be convenient. It is always assumed that S^3 is divided simplicially by planes through the origin in E^4 and that, when necessary, subdivisions of a

given subdivision are taken. Let C_1 and C_2 be two disjoint simple closed polygons in S^3. Then C_1 is called a parallel knot of order two if there exists a polyhedral Möbius band in S^3 which has C_1 as edge and C_2 as a "middle line". The theorem proved is as follows.

Theorem *Let T be an orientation preserving simplicial homeomorphism of period 2 of S^3 on itself. Let the set of fixed points F be a parallel knot of order two of the curve C. Then the fundamental group of $S^3 - C$ is infinite cyclic and the linking number of F and C is ± 1.*

As remarked, this shows that F can not be in a certain class of knots including the cloverleaf knot. This is because the cloverleaf knot bounds a Möbius band and the edge and center line do not have linking number ± 1. The methods show that if F is unknotted then T is equivalent to an orthogonal transformation which has also been proved by Moise.

When a compact Lie group G acts on a manifold the possibility for pathology is least when G is transitive and increases when

$$\dim M - \dim \text{(highest dimensional orbit)}$$

increases. Samelson and the author have found that a great deal of regularity exists when the difference above is at most two.

BIBLIOGRAPHY

[1] BING, A homeomorphism between the 3-sphere and the sum of two solid horned spheres, Annals of Math. 56 (1952), 354—362.

[2] GLEASON, Groups without small subgroups, Annals of Math. 56 (1952), 193—212.

[3] HILBERT, Mathematische Probleme, Göttingen Nachr. (1900), 253—297.

[4] IWASAWA, On some types of topological groups, Annals of Math. 50 (1949), 507—557.

[5] KURANISHI, On conditions of differentiability of locally compact groups, Nogoya Math Jour. 1 (1950), 71—81.

[6] MONTGOMERY and SAMELSON, A theorem on fixed points of involutions in S 3, to appear.

[7] MONTGOMERY and ZIPPIN, Small subgroups of finite-dimensional groups, Annals of Math 56 (1952), 213—241.

[8] MONTGOMERY and ZIPPIN, Examples of transformation groups, Proc. Am. Math. Soc. 5 (1954), 460—465.

[9]. MONTGOMERY and ZIPPIN, Non-abelian compact connected transformation groups of three-space, Amer. J. Math. 61 (1939), 375—387.

[10] NEWMAN, A theorem on periodic transformations of space, Quart. J. of Math. 2 (1931), 1—8.

[11] SMITH, Fixed points of periodic transformations, Appendix B, 352—373 of hefschetz, Algebraic topology, New York 1942.

[12] YAMABE, On the conjecture of Iwasewa and Gleason, Annals of Math. 58 (1953), 48—54.

[13] YAMABE, Generalization of a theorem of Gleason, Annals of Math. 58 (1953), 351—365.

COHOMOLOGIE ET GÉOMÉTRIE ALGÉBRIQUE

J. P. SERRE

See Symposium B. Algebraic Geometry

ON PSEUDO-HERMITIAN AND PSEUDO-KÄHLERIAN MANIFOLDS

KENTARO YANO

§ 1. *Pseudo-compex manifolds.*

In a complex manifold, we can define a mixed tensor $F_i^{\cdot h}$ which has the numerical components

$$F_\lambda^{\cdot \kappa} = +i\delta_\lambda^\kappa\,,\quad F_{\bar\lambda}^{\cdot \kappa} = 0,\quad F_\lambda^{\cdot \bar\kappa} = 0,\quad F_{\bar\lambda}^{\cdot \bar\kappa} = -i\delta_{\bar\lambda}^{\bar\kappa} \tag{1.1}$$

with respect to any complex coordinate system $z^\kappa = \xi^\kappa + i\xi^{\bar\kappa}$, $z^{\bar\kappa} = \xi^\kappa - i\xi^{\bar\kappa}$ which gives the complex structure to the manifold, where $\kappa, \lambda, \mu, \ldots = 1, 2, \ldots, n$ and $\bar\kappa, \bar\lambda, \bar\mu, \ldots = \bar 1, \bar 2, \ldots, \bar n$. This tensor field satisfies the relations

$$F_j^{\cdot h} F_i^{\cdot j} = -A_i^h, \tag{1.2}$$

where $h, i, j, \ldots = 1, 2, \ldots, n, \bar 1, \bar 2, \ldots, \bar n$.

When a $2n$-dimensional manifold of class C^r admits a mixed tensor field $F_i^{\cdot h}$ satisfying (1.2), we call the manifold an almost complex manifold (C. Ehresmann [1, 2]) [1]. If there exists a complex coordinate system in which $F_i^{\cdot h}$ has the components (1.1), then we say that the almost complex structure $F_i^{\cdot h}$ gives a complex structure to the manifold. The following theorem is now well known.

Theorem 1.1. *If an almost complex structure $F_i^{\cdot h}(\xi)$ gives a complex structure to the manifold, then we have*

$$N_{ji}^{\cdot\cdot h} \overset{def}{=\!=} 2F_{[j}^{\cdot l}(\partial_{|l|}F_{i]}^{\cdot h} - \partial_{i]}F_l^{\cdot h}) = 0, \tag{1.3}$$

where $\partial_l = \partial/\partial\xi^l$, and if an almost complex structure $F_i^{\cdot h}$ of class C^ω satisfies (1.3), *then it gives a complex structure to the manifold.*

Theorems essentially the same as this were stated and proved by G. de Rham, C. Ehresmann, P. Libermann [1], B. Eckmann, A. Frölicher [1], E. Calabi, D. C. Spencer [1], H. Guggenheimer [1, 2] and W. V. D. Hodge [1]. The tensor $N_{ji}^{\cdot\cdot h}$ was first introduced by A. Nijenhuis [1] [2]) and the equation (1.3) was given by B. Eckmann and A. Frölicher [1].

We remark here that the Nijenhuis tensor $N_{ji}^{\cdot\cdot h}$ satisfies the identities (B. Eckmann [1]):

$$N_{jl}^{\cdot\cdot h}F_i^{\cdot l} = -N_{jl}^{\cdot\cdot l}F_l^{\cdot h} = -N_{il}^{\cdot\cdot h}F_j^{\cdot l}. \tag{1.4}$$

[1]) See the Bibliography at the end of the paper.

[2]) See also J. A. Schouten [3].

We call pseudo-complex structure an almost complex structure which satisfies (1.3) but is not necessarily of class C^ω. We call pseudo-complex manifold a manifold with pseudo-complex structure.

It is possible to introduce, in a complex manifold, an affine connexion $\Gamma^{\cdot\cdot h}_{ji}$ such that $\nabla_j F^{\cdot h}_{i} = 0$ [3]) where ∇_j denotes the covariant differentiation with respect to $\Gamma^{\cdot\cdot h}_{ji}$. It is also possible to introduce, in an almost complex manifold, an affine connexion such that $\nabla_j F^{\cdot h}_{i} = 0$. In fact the two following connexions have this property.

$$\Gamma^{\cdot\cdot h}_{ji} = \overset{\circ}{\Gamma}{}^{\cdot\cdot h}_{ji} - \tfrac{1}{2}(\overset{\circ}{\nabla}_j F^{\cdot l}_{i}) F^{\cdot h}_{l}, \tag{1.5}$$

$$\Gamma^{\cdot\cdot h}_{ji} = \overset{\circ}{\Gamma}{}^{\cdot\cdot h}_{ji} - \tfrac{1}{2}(\overset{\circ}{\nabla}_{(j} F^{\cdot l}_{i)}) F^{\cdot h}_{l} - \tfrac{1}{2}(\overset{\circ}{\nabla}_{[l} F^{\cdot h}_{j]}) F^{\cdot l}_{i}, \tag{1.6}$$

where the $\overset{\circ}{\Gamma}{}^{\cdot\cdot h}_{ji}$ are the components of an arbitrary symmetric affine connexion and where $\overset{\circ}{\nabla}_j$ denotes the covariant differentiation with respect to $\overset{\circ}{\Gamma}{}^{\cdot\cdot h}_{ji}$.

On the other hand, the Nijenhuis tensor $N^{\cdot\cdot h}_{ji}$ can be written also in the form

$$\begin{aligned} N^{\cdot\cdot h}_{ji} &= 2F^{\cdot l}_{[j}(\nabla_{|l|} F^{\cdot h}_{i]} - \nabla_{i]} F^{\cdot h}_{l}) \\ &+ 2(S^{\cdot\cdot h}_{ji} - F^{\cdot m}_{j} F^{\cdot l}_{i} S^{\cdot\cdot h}_{ml} + F^{\cdot m}_{j} F^{\cdot h}_{l} S^{\cdot\cdot l}_{mi} + F^{\cdot m}_{i} F^{\cdot h}_{l} S^{\cdot\cdot l}_{jm}), \end{aligned} \tag{1.7}$$

where ∇_j denotes the covariant differentiation with respect to an arbitrary affine connexion $\Gamma^{\cdot\cdot h}_{ji}$ and $S^{\cdot\cdot h}_{ji}$ is its torsion tensor.

Thus if the manifold is pseudo-complex and if we introduce an affine connexion such that $\nabla_j F^{\cdot h}_{i} = 0$, then the torsion tensor satisfies

$$S^{\cdot\cdot h}_{ji} - F^{\cdot m}_{j} F^{\cdot l}_{i} S^{\cdot\cdot h}_{ml} + F^{\cdot m}_{j} F^{\cdot h}_{l} S^{\cdot\cdot l}_{jm} + F^{\cdot m}_{i} F^{\cdot h}_{l} S^{\cdot\cdot l}_{jm} = 0. \tag{1.8}$$

Conversely, if we can introduce, in an almost complex manifold, an affine connexion such that $\nabla_j F^{\cdot h}_{i} = 0$ and that the torsion tensor satisfies (1.8), then $N^{\cdot\cdot h}_{ji} = 0$ and the manifold is pseudo-complex. Thus we have

Theorem 1.2. *In order that an almost complex manifold be a pseudo-complex manifold, it is necessary and sufficient that we can introduce an affine connexion such that* $\nabla_j F^{\cdot h}_{i} = 0$ *and that* (1.8) *holds.*

Furthermore, if the manifold is pseudo-complex we can introduce a symmetric affine connexion such that $\nabla_j F^{\cdot h}_{i} = 0$, because the affine connexion (1.6) satisfies

$$S^{\cdot\cdot h}_{ji} = \tfrac{1}{4} N^{\cdot\cdot h}_{ji} = 0. \tag{1.9}$$

---

[3]) For the geometrical meaning of this equation, see J. A. Schouten [4], p. 391. We follow the notations of this book. Note that

$$\nabla_j F^{\cdot h}_{i} = \partial_j F^{\cdot h}_{i} + \Gamma^{jh}_{l} F^{\cdot l}_{i} - \Gamma^{l}_{ji} F^{\cdot h}_{l}.$$

Conversely, if we can introduce, in an almost complex manifold, a symmetric affine connexion such that $\nabla_j F_i^{\cdot h} = 0$, then $N_{ji}^{\cdot\cdot h} = 0$ and the manifold is pseudo-complex. Thus we have

Theorem 1.3. *In order that an almost complex manifold be a pseudo-complex manifold, it is necessary and sufficient that we can introduce a symmetric affine connexion such that* $\nabla_j F_i^{\cdot h} = 0$. (W. V. D. Hodge [1], E. M. Patterson [1], B. Eckmann [1]).

§ 2. *Almost and pseudo-Hermitian manifolds and almost and pseudo-Kählerian manifolds.*

In a Hermitian manifold with metric $ds^2 = 2g_{\bar{\lambda}\varkappa}\, dz^{\bar{\lambda}} dz^{\varkappa}$, the tensors g_{ih} and $F_i^{\cdot h}$ are related by

$$F_i^{\cdot m} F_h^{\cdot l} g_{ml} = g_{ih}. \tag{2.1}$$

When an almost (pseudo-) complex manifold has a positive definite metric $ds^2 = g_{ih} d\xi^i d\xi^h$ which satisfies (2.1), the manifold is called an almost (pseudo-) Hermitian manifold. In this case, the tensor $F_{ih} = F_i^{\cdot l} g_{lh}$ is antisymmetric in i and h. A. Lichnerowicz [1] showed that it is always possible to introduce a Hermitian metric in an almost complex manifold.

Equation (2.1) and the antisymmetry of the tensor F_{ih} show that the transformation $v^h \to F_i^{\cdot h} v^i$ changes the vector v^h into a vector orthogonal to it and does not change its length.

In a Kählerian manifold, the metric tensor satisfies

$$\partial_\mu g_{\bar{\lambda}\varkappa} = \partial_\lambda g_{\bar{\mu}\varkappa} \tag{2.2}$$

and the tensor F_{ih} satisfies

$$F_{jih} \overset{def}{=\!=} 3\partial_{[j} F_{ih]} = 0. \tag{2.3}$$

When an almost (pseudo-) Hermitian manifold satisfies $F_{jih} = 0$, it is called an almost (pseudo-) Kählerian manifold.

If $\overset{\circ}{\nabla}_j F_{ih}$ vanishes in an almost Hermitian manifold, where $\overset{\circ}{\nabla}_j$ denotes the covariant differentiation with respect to the Christoffel symbols $\{{}^h_{ji}\}$, then the tensors $N_{ji}^{\cdot\cdot h}$ and F_{jih} vanish also and consequently the manifold is pseudo-Kählerian.

Conversely, if $N_{ji}^{\cdot\cdot h}$ and F_{jih} vanish, then the Nijenhuis tensor can be written in the form

$$N_{jih} = -2\,(F_{[j}^{\cdot l} F_{i]hl} - F_j^{\cdot l} \overset{\circ}{\nabla}_h F_{il}) \tag{2.4}$$

which shows that $\overset{\circ}{\nabla}_j F_{ih}$ vanishes also. Thus we have

Theorem 2.1. *In order that an almost Hermitian manifold be pseudo-Kählerian, it is necessary and sufficient that* $\overset{\circ}{\nabla}_j F_{ih} = 0$. (K. Yano and I. Mogi [1], B. Eckmann [1], K. Yano [3]).

In an almost Hermitian manifold, the three following connexions are known (G. Legrand [1], K. Yano [4]):

$$\overset{1}{\Gamma}{}^{h}_{ji} = \{^{h}_{ji}\} - \tfrac{1}{2}(\overset{\circ}{\nabla}_j F^{\cdot}_{il})F^{lh}, \tag{2.5}$$

$$\overset{2}{\Gamma}{}^{h}_{ji} = \{^{h}_{ji}\} - \tfrac{1}{2}(\overset{\circ}{\nabla}_j F_{il} + \overset{\circ}{\nabla}_i F_{jl} - \overset{\circ}{\nabla}_l F_{ij})F^{lh}, \tag{2.6}$$

$$\overset{3}{\Gamma}{}^{h}_{ji} = \{^{h}_{ji}\} - \tfrac{1}{2}(\overset{\circ}{\nabla}_j F_{il} - \overset{\circ}{\nabla}_i F_{jl} + \overset{\circ}{\nabla}_l F_{ij})F^{lh}. \tag{2.7}$$

The first, introduced by A. Lichnerowicz [2], satisfies $\overset{1}{\nabla}_j F_i^{\cdot h} = 0$ and is metric. The second, introduced by J. A. Schouten and D. van Dantzig [1], and the third satisfy $\overset{2}{\nabla}_j F_i^{\cdot h} = 0$, $\overset{3}{\nabla}_j F_i^{\cdot h} = 0$ and are metric when the manifold is pseudo-Hermitian.

§ 3. *Curvature in a pseudo-Kählerian manifold.*

In a pseudo-Kählerian manifold, we can prove

Theorem 3.1. *If a pseudo-Kählerian manifold is of constant curvature, then it is of zero curvature* (S. Bochner [2]).

Theorem 3.2. *If a pseudo-Kählerian manifold is conformally flat, then it is of zero curvature.*

Theorem 3.3. *A projective correspondence between two pseudo-Kählerian manifolds is necessarily affine* (S. Bochner [2], W. J. Westlake [2], K. Yano [2]).

Theorem 3.4. *A conformal correspondence between two pseudo-Kählerian manifolds is necessarily a trivial one.*

Theorem 3.5. *A necessary and sufficient condition that a $2n$-dimensional pseudo-Hermitian manifold be conformal to a pseudo-Kählerian manifold is that for $2n > 4$.*

$$C_{jih} \overset{def}{=\!=} F_{jih} - \frac{1}{2(n-1)}(F_{ji}F_h + F_{ih}F_j + F_{hj}F_i) = 0 \tag{3.1}$$

and for $2n = 4$

$$C_{ji} \overset{def}{=\!=} 2\,\partial_{[j}F_{i]} = 0, \tag{3.2}$$

where $F_j = F_{jih}F^{ih}$ (W. J. Westlake [1], K. Yano [2]).

§ 4. *Pseudo-analytic vector fields* (S. Bochner [1]).

In a Kählerian manifold, an analytic covariant vector field v_i satisfies

$$F_j^{\cdot l}\nabla_i v_l - F_i^{\cdot l}\nabla_l v_j = 0 \tag{4.1}$$

where ∇_i denotes the covariant differentiation with respect to $\{^{h}_{ji}\}$.

In a pseudo-Kählerian manifold, we call a covariant vector field satisfying (4.1) a pseudo-analytic covariant vector field. From (4.1), we can deduce

$$g^{ji}\nabla_j\nabla_i v_h - K_h^{\cdot l}v_l = 0, \tag{4.2}$$

and using this, we can prove

Theorem 4.1. *In order that, in a compact pseudo-Kählerian manifold, a covariant vector field be pseudo-analytic, it is necessary and sufficient that it be* harmonie

In a Kählerian manifold, an analytic contravariant vector field v^h-satisfies

$$\underset{v}{\pounds} F_i^{\cdot h} \overset{def}{=\!=} -F_i^{\cdot l}\nabla_l v^h + F_l^{\cdot h}\nabla_i v^l = 0, \tag{4.3}$$

where $\underset{v}{\pounds}$ denotes the Lie derivative with respect to v^h (J. A. Schouten [4], K. Yano [1]).

In a pseudo-Kählerian manifold, we call a contravariant vector field satisfying (4.3) a pseudo-analytic contravariant vector field.

From (4.3) we can see that if v^h is pseudo-analytic, then $F_i^{\cdot h}v^i$ is so also. Moreover, if u^h and v^h are both pseudo-analytic, then $\underset{u}{\pounds} F_i^{\cdot h} = 0$, $\underset{v}{\pounds} F_i^{\cdot h} = 0$, from which $(\underset{u}{\pounds}\underset{v}{\pounds}) F_i^{\cdot h} = 0$, where $(\underset{u}{\pounds}\underset{v}{\pounds})$ denotes the Lie derivative with respect to the vector

$$[u, v]^h = u^l\nabla_l v^h - v^l\nabla_l u^h. \tag{4.4}$$

Thus we have

Theorem 4.2. *If, in a pseudo-Kählerian manifold, u^h and v^h are both pseudo-analytic contravariant vector fields, then $F_i^{\cdot h}u^i$, $F_i^{\cdot h}v^i$, $[u, v]^h$, $[Fu, v]^h$, $[u, Fv]^h$ and $[Fu, Fv]^h$ are all pseudo-analytic vector fields.*

In an almost complex manifold, we have the identity

$$[u, v]^h + F_i^{\cdot h}[Fu, v]^i + F_i^{\cdot h}[u, Fv]^i - [Fu, Fv]^h = -N_{ji}^{\cdot\cdot h}u^j v^i, \tag{4.5}$$

from which we have

Theorem 4.3. *In order that an almost complex manifold be pseudo-analytic, it is necessary and sufficient that the left hand side of* (4.5) *vanishes for any vectors u^h and v^h* (B. Eckmann [1]).

Now from (4.3) we can deduce

$$g^{ji}\nabla_j\nabla_i v^h + K_i^{\cdot h}v^i = 0, \tag{4.6}$$

from which we have the following theorems which hold in a compact pseudo-Kählerian manifold.

Theorem 4.4. *A contravariant pseudo-analytic vector v^h* satisfying $\nabla_h v^h = 0$ *is a Killing vector* (K. Yano and S. Bochner [1]).

Theorem 4.5. *If a vector field is at the same time contravariant and covariant pseudo-analytic, then it is a parallel vector field.*

Theorem 4.6. *If a pseudo-analytic contravariant vector v^h satisfies $F_i^{\cdot h}\nabla_h v^i = 0$, then $F_i^{\cdot h}v^i$ is a Killing vector field.*

Theorem 4.7. *If a pseudo-analytic contravariant vector field v^h satisfies $\nabla_h v^h = 0$, $F_i^{\cdot h}\nabla_h v^i = 0$, then it is a parallel vector field.*

§ 5. *Pseudo-Kählerian manifolds of constant holomorphic curvature.*

We call holomorphic sectional curvature a sectional curvature determined by two orthogonal vectors u^h and $F_i^{\cdot h}u^i$. We can prove

Theorem 5.1. *If a pseudo-Kählerian manifold has constant holomorphic sectional curvature at any point, then the curvature tensor of the manifold is of the form*

$$(5.1) \qquad K_{kjih} = \frac{k}{4}[(g_{kh}g_{ji} - g_{jh}g_{ki}) + (F_{kh}F_{ji} - F_{jh}F_{ki}) - 2F_{kj}F_{ih}],$$

where k is an absolute constant (K. Yano and I. Mogi [1]) [4].

Theorem 5.2. *In a pseudo-Kählerian manifold of constant holomorphic curvature, a sectional curvature determined by two orthogonal unit vectors u^h and v^h is given by $\varkappa = \frac{k}{4}(1+3a^2)$ where $a = F_{ih}u^iv^h$ is the cosine of the angle between two unit vectors $F_i^{\cdot h}u^i$ and v^h and consequently $a^2 \leqq 1$. Thus*

$$(5.2) \qquad \frac{k}{4} \leqq \varkappa \leqq k \text{ for } k > 0 \text{ and } k \leqq \varkappa \leqq \frac{k}{4} \text{ for } k < 0.$$

(S. Bochner [2], K. Yano and S. Bochner [1], K. Yano and I. Mogi [1]).

Theorem 5.3. *In order that, in a pseudo-Kählerian manifold, there exist always a 2-dimensional totally geodesic subspace passing through an arbitrarily given point and being tangent to an arbitrarily given holomorphic plane element at this point, it is necessary and sufficient that the manifold be of constant holomorphic curvature* (K. Yano and I. Mogi [1]).

Theorem 5.4. *In order that a pseudo-Kählerian manifold admit a group of motions which carries any two vectors u^h and $F_i^{\cdot h}u^i$ at a point P into any two vectors $'u^h$ and $'F_i^{\cdot h}{}'u^i$ at any point $'P$, it is necessary and sufficient that the manifold be of constant holomorphic curvature.* (K. Yano and I. Mogi [1]).

We may call the assumptions of Theorems 5.3 and 5.4 the axioms of holomorphic planes and of the holomorphic free mobility respectively.

Theorem 5.5 *In a pseudo-Kählerian manifold of positive constant holomorphic curvature k, the distance between two consecutive conjugate points on a geodesic is constant and is equal to $2\pi/\sqrt{k}$.* (K. Yano and I. Mogi [1]).

---

[4]) See also J. A. Schouten [1], where he has introduced Hermitian space of constant curvature.

BIBLIOGRAPHY

BOCHNER, S..

[1] Vector fields and Ricci curvature. Bull. Amer. Math. Soc. 52 (1946), 776—797.

[2] Curvature in Hermitian manifolds. Ibidem, 53 (1947), 179—195.

CALABI, E. and D. C. SPENCER:
[1] Completely integrable almost complex manifolds. Ibidem, 57 (1951), 254—255.

CHERN, S. S.:
[1] Characteristic classes of Hermitian manifolds. Annals of Math. 47 (1946), 85—121.

ECKMANN, B.:
[1] Sur les structures complexes et presque complexes. Colloques Internationaux de C.N.R.S. Géométrie Différentielle. Strasbourg (1953), 151—159.

ECKMANN, B. and A. FRÖLICHER:
[1] Sur l'intégrabilité des structures presque complexes. C.R. Acad. Sci. Paris, 232 (1951) 2284—2286.

EHRESMANN, C.:
[1] Sur la théorie des espaces fibrés. Coll. de Topol. Alg. C.N.R.S. Paris (1947), 3—15.
[2] Sur les variétés presque complexes. Proc. Int. Congr. Math. (1952), 412—419.

GUGGENHEIMER, H.:
[1] Über Kählersche und symplektische Differentialalgebren. Tohoku Math. Journal, 4 (1952), 157—171.
[2] Geometria pseudo-kähleriana. Rend. Acad. Naz. dei Lincei, 14 (1953), 220—222.

HODGE, W. V. D.:
[1] Structure problems for complex manifolds. Rend. Mat. e Appl. Ser. V, 11 (1952), 101—110.

LEGRAND, G.:
[1] Connexions définies sur une variété presque hermitique. C.R. Acad. Sci. Paris, 237 (1953), 1626—1627.

LIBERMANN, P.:
[1] Sur le problème d'équivalence de certaines structures infinitésimales. Thèse (1953).

LICHNEROWICZ, A.:
[1] Généralisations de la géométrie kählérienne globale. Coll. de Géom. Diff. tenu à Louvain (1951), 99—122.
[2] Un théorème sur les espaces homogènes complexes. Archiv der Math. 5 (1954), 207—215.

NIJENHUIS, A.:
[1] X_{n-1}-forming sets of eigenvectors. Indigationes Mathematicae. 8 (1951), 200—212.

PATTERSON, E. M.:
[1] A characterisation of Kähler manifolds in terms of parallel fields of planes. The Journal of the London Math. Soc., 28 (1953), 260—269.

SCHOUTEN, J. A.:
[1] Über unitäre Geometrie. Proc. Kon. Akad. v. Wetensch. Amsterdam, 32 (1929), 457—465.
[2] Über unitäre Geometrien konstanter Krümmung. Ibidem 34 (1931), 1293—1304.
[3] Sur les tenseurs de V_n aux directions principales V_{n-1}-normales. Coll. Géom. Diff. tenu à Louvain (1951), 67—70.
[4] Ricci Calculus 2nd edition (1954).

SCHOUTEN, J. A. and D. VAN DANTZIG:
[1] Über unitäre Geometrie. Math. Ann., 103 (1930), 319—346.

WESTLAKE, W. J.:
[1] Conformally Kähler manifolds. Proc. Cambridge Phil. Soc., 50 (1954), 16—19.
[2] Hermitian spaces in geodesic correspondence. Proc. Amer. Math. Soc., 5 (1954), 301—303.

YANO, K.:
[1] Groups of transformations in generalized spaces, Tokyo (1949).
[2] Geometria conforme in varietà quasi hermitiane. Rend. Acad. Naz. dei Lincei (to appear).
[3] Quelques remarques sur les variétés à structure presque complexe. Bull. Soc. Math. France 83 (1955) 57—80.
[4] On three remarkable affine connexions in an almost Hermitian space. Indigationes Mathematicae 17 (1955) 24—32.

YANO, K. and S. BOCHNER:
[1] Curvature and Betti Numbers. Annals of Mathematics Studies, 32 (1953).

YANO, K. and I. MOGI:
[1] Sur les variétés pseudo-kählériennes à courbure holomorphique constante. C. R. Acad. Sc. Paris, 237 (1953), 962—964.

SECTION IV

PROBABILITY AND STATISTICS

STATED ADDRESSES

CONTROLLED RANDOM WALKS

D. BLACKWELL

See Symposium A. Stochastic Processes

INTERRELATIONS BETWEEN BROWNIAN MOTION AND POTENTIAL THEORY

J. L. Doob

The transition probabilities, and related quantities fundamental in the theory of stochastic processes of diffusion, satisfy parabolic differential equations. This fact suggests that there may be an intimate connection, not only between the solutions of the differential equations and the probability distributions involved in the corresponding stochastic processes, but between the solutions of the differential equations and the sample trajectories of the corresponding stochastic processes.

It the parabolic equation is the heat equation in N space variables, the corresponding stochastic process is N dimensional Brownian motion. To simplify the discussion, the solutions of the heat equation will be taken at first to be independent of the time variable, that is, we shall consider harmonic functions of N variables. The most important results obtained can be outlined as follows.

(1) Let u be a function defined and subharmonic on an open set D, and let z be any point of D. Then u is continuous, as a function of the trajectory parameter, on almost every Brownian trajectory from z (as long as the trajectory remains in D), except that, if $u(z) = -\infty$, the initial point of the trajectories must of course be excluded.

(2) Let u be a function defined and subharmonic on an open set D, and suppose that there is a positive harmonic function v, also with domain D, such that $u \leqq v$. To avoid trivialities, we suppose that the complement of D has positive capacity. Then, if z is any point of D, u has a finite limit along almost every Brownian trajectory from z to the boundary of D.

(3) Let D be any open set whose complement has positive capacity, let f be a function defined on the boundary C, and suppose that f is in the L_1 class as defined in terms of the harmonic measure μ_z relative to the point z of D. It is well known that this L_1 class does not depend on z, and the generalized Dirichlet (first boundary value) problem of finding a harmonic function corresponding to the boundary function f has the solution u given by

$$u(z) = \int_C f(s)\, \mu_z(ds).$$

In probability language, this formula is shown to become

$$u(z) = E\{f(\zeta)\}.$$

Here $E\{-\}$ denotes expectation, and ζ is a random point on C. It is the first point in which a Brownian trajectory from z meets C. In other words, the distribution of the point ζ on C is precisely the harmonic measure relative to the initial point z. Moreover, it is proved that, if z_0 is any point of D, u has the specified boundary value as a limit along almost every Brownian trajectory from z_0 to C.

The results (2) and (3) are simple generalizations of Fatou's theorem on the existence of radial limit values of harmonic functions defined on a plane disc, as developed by later writers. A difficulty met at once in attempting to generalize these theorems is the choice of the proper family of curves of approach to the boundary, for general domains. It is noteworthy that the family of Brownian trajectories used here does not depend on the domain.

The fundamental tool used in proving these theorems is the theory of certain stochastic processes, semimartingales and martingales. The fundamental fact which dominates the treatment is that, if u is subharmonic [harmonic] on N space, and not too large, then, if $\{z(t),\ 0 \leqq t < \infty\}$ is a Brownian motion process in N dimensions, the stochastic process $\{u[z(t)],\ 0 \leqq t < \infty\}$ is a semimartingale [martingale]. If u is only defined on an open subset D of N space, if the $z(t)$ process trajectories have initial point in D, and if the trajectories are stopped when they meet the boundary, so that $z(t)$ is now constant for sufficiently large t, with probability 1, if, as we suppose now, the complement of D has positive capacity, the preceding result can be extended. In fact, for example, if u is a generalized Dirichlet solution, as defined in (3), and if u is defined on the boundary C as the specified boundary function f, then the $u[z(t)]$ process is a martingale. In general, if u is defined on $D \cup C$, and is subharmonic [harmonic] on D, and sufficiently smooth near C, the $u[z(t)]$ process is a semimartingale [martingale]. Full details of this work have now been published elsewhere [Trans. Amer. Math. Soc. 77, 86-121 (1954)]. The explicit formulas and the existence of boundary limits obtained in this theory simplify considerably the discussion of such potential-theoretic concepts as the equilibrium potential and the sweeping out procedure.

Essentially the same theorems can be proved relating solutions of the heat equation and the corresponding subsolutions to a somewhat different type of Brownian trajectory. The details will be published elsewhere. This approach makes natural the study of the first boundary value problem for parabolic equations in spite of the tradition that this problem is meaningless for these equations. The fact is, however, that the usual first boundary value problem

ideas remain appropriate. For example, a sufficiently smooth boundary function determines a unique first boundary value problem solution, which has the assigned boundary value as a limit at each regular boundary point at which a bounded boundary function is continuous. The regular and irregular boundary points are defined as usual, although the characteristics of a regular boundary point for the heat equation are very different from those of a regular boundary point for Laplace's equation. The class of irregular boundary points for the heat equation is, however, small in an appropriate measure-theoretic sense, and the values of the given boundary function at the irregular boundary points are irrelevant to the first boundary value problem solution. Using these results, much of the classical potential theory can be obtained as a specialization of the more general potential theory of the heat equation, in which a potential is generated by a distribution of heat sources in time and space.

LOIS DES GRANDS NOMBRES POUR DES ÉLÉMENTS ALÉATOIRES GÉNÉRAUX

R. M. FORTET

See Symposium A. Stochastic Processes

SECTION V

MATHEMATICAL PHYSICS AND APPLIED MATHEMATICS

STATED ADDRESSES

FEHLERMASZPRINZIPIEN IN DER PRAKTISCHEN ANALYSIS

L. COLLATZ

§ 1. *Verschiedene Fehlermaßprinzipien.*

Es liege die Aufgabe vor, eine in einem Bereich B des $x_1 \ldots x_n$-Raumes gegebene Funktion

$$f(x_1, \ldots, x_n)$$

durch eine Linearkombination Q_p aus ebenfalls gegebenen Funktionen $u_\nu(x_j)$ $(\nu = 0, 1, 2, \ldots, p)$

$$Q_p(x_j) = \sum_{\nu=0}^{p} a_\nu u_\nu(x_j) \tag{1}$$

möglichst gut anzunähern, d.h. der Fehler

$$\varepsilon_p(x_j) = Q_p(x_j) - f(x_j) \tag{2}$$

soll in B „möglichst klein'' werden. Die bekanntesten der Prinzipien für die Auswahl von Q_p lassen sich in zwei Klassen einteilen:

I. Man legt eine Norm zugrunde. Es sei in dem Raum $\mathfrak{R}$ der betrachteten Funktionen für jede Funktion g eine Norm $\| g \| \geqq 0$ definiert, wobei $\| g \| = 0$ genau für $g = 0$ in B gilt. Man stellt dann die Forderung auf

$$\| \varepsilon_p \| = \text{Min.}, \tag{3}$$

und bestimmt daraus die gesuchten a_ν. Die bekanntesten Arten sind:

Ia. Fehlerquadratmethode. Mit Hilfe einer in B positiven integrablen Funktion $\varphi(x_j)$ verwendet man die Norm

$$\| g \| = + \sqrt{\int_B \varphi\, g^2\, d\tau} \tag{4}$$

Dabei werden nur in B (im Riemannschen oder Lebesgue'schen Sinne) quadratisch integrable Funktionen g betrachtet und in (4) bedeutet $d\tau$ das Volumenelement. Zur Bestimmung der a_ν dienen dann die linearen Normalgleichungen:

$$\frac{1}{2} \frac{\partial \| \varepsilon_p \|^2}{\partial a_\varrho} = \int_B \varepsilon_p \varphi u_\varrho \, d\tau = 0 \quad (\varrho = 0, \ldots, p). \tag{5}$$

Ib. Prinzip vom kleinsten Fehlerbetragmaximum. Man legt etwa den Raum der in B stetigen Funktionen zugrunde, setzt B als abgeschlossenen Bereich voraus, wählt dort eine positive stetige Funktion φ und verwendet die Norm

$$\| g \| = \operatorname*{Max}_{\text{in } B} \frac{|g|}{\varphi}. \tag{6}$$

Die a_ν sind wieder aus (3) zu ermitteln, was allerdings jetzt analytisch wesentlich komplizierter in der Durchführung wird. Nach § 2 ist aber für viele Anwendungen dieses Prinzip *Ib* wichtiger als das Prinzip *Ia*, welches in den früheren Jahrzehnten fast ausschliesslich verwendet wurde. Die Unterschiede der Ergebnisse bei den beiden Prinzipien *Ia* und *Ib* können beträchtlich sein, wie das Beispiel zeigt:

Beispiel: Die Funktion $f(x) = \sqrt{x}$ möge im Intervall $\langle 0, 1\rangle$ durch eine lineare Funktion $Q_1 = a_0 + a_1 x$ angenähert werden. Bei $\varphi \equiv 1$ führt die Fehlerquadratmethode zu der Näherung $Q_{1F} = \frac{4}{15} + \frac{4}{5} x$ mit der Maximalabweichung

$$|\varepsilon|_{\max} = \varepsilon_F = \tfrac{4}{15} \approx 0{,}2667;$$

dagegen führt die Methode vom kleinsten Betragmaximum bei $\varphi \equiv 1$ zu $Q_{1M} = \frac{1}{8} + x$ mit der wesentlich kleineren Maximalabweichung

$$|\varepsilon|_{\max} = \varepsilon_M = \tfrac{1}{8} = 0{,}125.$$

II. Es werden gewisse lineare homogene Ausdrücke L_ϱ zugrunde gelegt und die a_ν aus

$$L_\varrho(\varepsilon_p) = 0, \qquad (\varrho = 0, \ldots, p) \tag{7}$$

ermittelt. Wir beschränken uns auf die Aufzählung einiger Prinzipien aus der grossen Zahl verschiedener Möglichkeiten.

IIa. Kollokation.

In B wählt man $p + 1$ Punkte P_ϱ und setzt $L_\varrho(\varepsilon) = \varepsilon(P_\varrho)$; sofern die Determinante

$$\Delta = \det u_\nu(P_\varrho) = \begin{vmatrix} u_0(P_0) \ldots u_0(P_p) \\ \ldots\ldots\ldots\ldots \\ u_p(P_0) \ldots u_p(P_p) \end{vmatrix} \tag{8}$$

$\neq 0$ ist, sind die a_ν eindeutig bestimmt. Die Kollokation löst in diesem Fall zugleich die Interpolationsaufgabe, die gegebene Funktion $f(x)$ durch eine Linearkombination Q darzustellen, welche mit f an den gegebenen Punkten P_ϱ übereinstimmt.

IIb. Orthogonalitätsmethode.

Man wählt $p + 1$ linear voneinander unabhängige Funktionen $g_\varrho(x_j)$ und

setzt $L_\varrho(\varepsilon) = \int_B \varepsilon g_\varrho \, d\tau$. Die Fehlerquadratmethode lässt sich auch als Spezialfall $g_\varrho = \varphi \, u_\varrho$ auffassen.

IIc. Teilgebietsmethode. Der Bereich B wird in $p + 1$ Teilbereiche B_ϱ aufgeteilt und $L_\varrho(\varepsilon) = \int_{B_\varrho} \varepsilon d\tau$ gesetzt.

IId. Methoden der Ausgleichsrechnung.

Man wählt mehr als $p + 1$ Punkte P. Für die Fehler $\varepsilon(P_\varrho)$ lassen sich dann die bekannten Methoden der Ausgleichsrechnung anwenden.

Wir stellen zunächst in einem Ueberblick zusammen, welche Fehlerabschätzungen auf einem wichtigen Teilgebiet, nämlich dem der linearen partiellen Differentialgleichungen 2. Ordnung, vorliegen und sehen, daß für diese Anwendungen das Prinzip *Ib* besonders brauchbar ist.

§ 2. *Stand der Fehlerabschätzungen bei den klassischen Problemen partieller Differentialgleichungen.*

2*a*) Elliptische Differentialgleichungen.

Vorgelegt sei die Differentialgleichung für eine Funktion $u(x_1, \ldots, x_n)$

$$Lu = -\sum_{j,k=1}^{n} a_{jk} \frac{\partial^2 u}{\partial x_j \partial x_k} - \sum_{j=1}^{n} b_j \frac{\partial u}{\partial x_j} + cu = r \tag{9}$$

Dabei sind a_{jk}, b_j, c und r gegebene stetige Funktionen von $x_1, \ldots, x_n$. Die Matrix der a_{jk} sei in jedem Punkte eines betrachteten abgeschlossenen beschränkten einfach zusammenhängenden Bereiches B positiv definit. B werde von einer $(n-1)$-dimensionalen Hyperfläche Γ begrenzt, die sich aus endlich vielen Flächenstücken mit stetiger Tangentialhyperebene zusammensetzen lasse. Es sei $c \geqq 0$.

Der Rand Γ bestehe aus 2 Teilen Γ_1 und Γ_2, von denen auch ein Teil leer sein darf. Dann seien als Randbedingungen vorgegeben:

$$\left.\begin{array}{ll} A_1 u + A_2 L^* u = A_3 & \text{auf } \Gamma_1 \text{ mit } A_2/A_1 < 0 \\ u = A_4 & \text{auf } \Gamma_2 \end{array}\right\} \tag{10}$$

(A_1 bis A_4 gegebene Funktionen auf dem Rande). Die Randbedingungen sind hier der Kürze halber als Gleichungen geschrieben und so zu verstehen, daß die Grenzwerte der linken Seiten bei Annäherung von innen her an den Rand die Werte A_3 bzw. A_4 haben. In $L^* u = A \dfrac{\partial u}{\partial \sigma}$ berechnen sich A und die Konormale [1]) σ nach (innere Normale ν)

[1]) Vgl. etwa A. G. Webster, G. Szegö: Partielle Differentialgleichungen der mathematischen Physik, Leipzig und Berlin, 1930, S. 311.

$$\sum_{j=1}^{n} a_{jk} \cos(\nu, x_j) = A_k^* = A \cos(\sigma, x_k) \quad (k = 1, \ldots, n). \tag{11}$$

Bei der Potentialgleichung ist $L^*u = \dfrac{\partial u}{\partial \nu}$.

Nun sei v eine Näherung mit $Lv = r$. Dann ist [2]) auf Γ die Funktion

$$\gamma = \begin{cases} v - u + \dfrac{A_2}{A_1} L^*(v - u) & \text{auf } \Gamma_1 \\ v - u & \text{auf } \Gamma_2 \end{cases} \tag{12}$$

bekannt und es gilt die Fehlerabschätzung: Nimmt γ auf Γ Werte beiderlei Vorzeichens oder wenigstens den Wert Null an, so ist

$$\gamma_{\min} \leqq v - u \leqq \gamma_{\max} \text{ in } B. \tag{13}$$

Hat γ ein festes Vorzeichen, so hat auch $v - u$ dasselbe Vorzeichen und es gilt $|v - u| \leqq |\gamma|_{\max}$. Verwendet man für v einen linearen Ansatz, etwa $v = \sum_{\nu=0}^{p} c_\nu v_\nu$ mit $Lv_0 = r$, $Lv_\varrho = 0$ für $\varrho = 1, \ldots, p$, so entspricht γ der Fehlerfunktion ε_p in (2) und die Fehlerabschätzung fällt am günstigsten aus, wenn man die c_ν nach dem Prinzip *Ib* auswählt.

Der Beweis für (13) benutzt wesentlich eine von E. Hopf [3]) gegebene Verallgemeinerung des klassischen Randmaximumsatzes für Potentialfunktionen. Hopf hat bereits Erweiterungen auf gewisse Typen nichtlinearer elliptischer Differentialgleichungen gegeben; damit lässt sich z.B. eine Fehlerabschätzung für die 1. Randwertaufgabe bei der Gleichung für die Minimalflächen (Plateau'sches Problem) durchführen.

Es liegen Ansätze vor, die Fehlerabschätzungen auf die Plattengleichung zu übertragen [4]). Eine Fehlerabschätzung macht keine Schwierigkeit, wenn die Randwertaufgabe von monotoner Art ist [5]). Das ist z.B. der Fall bei gestütztem stückweise geradlinigem Rande der Platte; dagegen bei Einspannung der Platte liegen die Verhältnisse wesentlich komplizierter.[6])

---

[2]) Vgl. GRÜNSCH, H. J.: Eine Fehlerabschätzung bei der 3. Randwertaufgabe der Potentialtheorie. Z. angew. Math. Mech. Bd. 32 (1952). S. 279—281.

[3]) E. HOPF: Elementare Bemerkungen über die Lösungen partieller Differential gleichungen 2. Ordnung vom elliptischen Typus. Sitzungsber. Preuss. Akad. der Wiss. Math. Phys. Klasse S. 147—152, Berlin 1927.

[4]) Diaz, J. B. u. H. J. GREENBERG: Upper and lower bounds for the solution of the first biharmonic boundary value problem. J. Math. Phys. Bd. 27 (1948) S. 193—201.

[5]) L. COLLATZ: Aufgaben monotoner Art. Arch. Math. Bd. 3 (1952). S.373.

[6]) MIRANDA, C.: Formule di maggiorazione e teorema di esistenza per le funzioni biarmoniche di due variabili. Giornale di Matematiche di Battaglini Bd. 78 (1948/49). S. 98—118.

2*b*. Parabolische Differentialgleichungen.

Ein sehr allgemeiner Abschätzungssatz stammt von Westphal [7]). B sei der offene Bereich in der x-y-Ebene: $0 < y < Y$; $x_0(y) < x < x_1(y)$, (Y eine Konstante, $x_0(y)$ und $x_1(y)$ zwei gegebene stetige Funktionen von y mit $x_0(y) < x_1(y)$ für $0 \leqq y \leqq Y$). Γ_y sei die Randstrecke mit $y = Y$, der übrige Rand von B sei Γ. Bei dem Operator T

$$Tu = u_y - f(x, y, u, u_x, u_{xx}), \tag{14}$$

sei f eine gegebene Funktion ihrer Argumente und überdies bei fester Wahl von x, y, u, u_x monoton nichtfallend in u_{xx}; u sei in $\tilde{B} = B + \Gamma + \Gamma_y$ stetig und u_y und u_{xx} mögen in B existieren. Dann gilt der Satz über die monotone Art [8]) des Operators T: Aus $Tv < Tu$ in $B + \Gamma_y$ und $v < u$ auf Γ folgt $v < u$ in B.

Görtler [9]) verallgemeinert die Abschätzungssätze auf das System für zwei Funktionen $u(x, y)$, $v(x, y)$

$$u u_x + v u_y = f(x, u, u_y, u_{yy}); \; u_x + v_y = 0.$$

2*c*. Hyperbolische Differentialgleichungen.

Bei hyperbolischen Differentialgleichungen ist das Bild noch lange nicht so abgerundet wie bei elliptischen und parabolischen Differentialgleichungen. Ein gewisses Analogon zu den Randmaximumsätzen wurde kürzlich bewiesen[10]).

Für eine Funktion $u(\xi, \eta)$ sei

$$Lu = u_{\xi\eta} + a u_\xi + b u_\eta. \tag{15}$$

B sei ein offener, von einem Stück der positiven ξ-Achse, einem Stück der positiven η-Achse und einer stetigen Kurve C begrenzter Bereich, wobei C von jeder Parallelen zur ξ- oder η-Achse in höchstens einem Punkte geschnitten werde. $\tilde{B}$ entstehe aus B durch Hinzunahme der Randpunkte. In B sei $a \leqq 0$, $a_\xi + ba - c \leqq 0$, $c \leqq 0$. Wenn ferner $u_\eta(0, \eta) \geqq 0$ in allen zu $\tilde{B}$ gehörigen Randpunkten der η-Achse, $Lu + cu \geqq 0$ in B gilt und das Maximum von u in $\tilde{B}$ positiv ist, dann wird das Maximum von u auf C angenommen. Im

---

[7]) Westphal, H.: Zur Abschätzung der Lösungen nichtlinearer parabolischer Differentialgleichungen. Math. Z. Bd. 51 (1949) S. 690—695.

[8]) Ein numerisches Beispiel für die hierdurch gegebene Möglichkeit einer Fehlerabschätzung bei L. Collatz, Numerische Behandlung von Differentialgleichungen, 2. Aufl. Springer. Berlin Göttingen Heidelberg 1955, S. 311.

[9]) Görtler, H.: Ueber die Lösungen nichtlinearer partieller Differentialgleichungen vom Reibungsschichttypus. Z. angew. Math. Mech. Bd. 30 (1950) S. 265—267.

[10]) Agmon, S., Nirenberg, L. and Protter, M. H.: A maximum principle for a class of hyperbolic equations and applications to equations of mixed elliptic-hyperbolic type. Comm. Pure Appl. Math. 6, S. 455—470 (1953).

Falle $c = 0$ ist die Voraussetzung, daß das Maximum von u positiv sei, entbehrlich.

Dieser Satz ist z.B. bei der Telegraphengleichung

$$u_{xx} = u_{tt} + Eu_t + Fu \tag{16}$$

im Falle $J = -\frac{E^2}{8} + \frac{F}{2} \leqq 0$ anwendbar, wenn es sich um einen Bereich zwischen einer Charakteristik und einer keine Charakteristik berührenden Kurve handelt, z.B. bei dem Bereich $t \geqq x \geqq 0$. Wenn die Invariante J positiv ausfällt, sind die oben über a, b, c getroffenen Voraussetzungen nicht erfüllt.

Liegt z.B. eine Randwertaufgabe vor, bei der die Werte von u und u_t auf dem Stück $0 \leqq x \leqq s$ der x-Achse und die Werte von u auf den Geraden $x = 0$ und $x = s$ vorgegeben sind, so lässt sich der obige Satz anwenden, wenn man noch die Werte von u in dem Dreieck $x \geqq t \geqq 0$, $t \leqq s - x$ kennen würde. Aber für dieses Dreieck kann man das Picardsche Iterationsverfahren benutzen mit einer Fehlerabschätzung mit Hilfe des Fixpunktsatzes für das Iterationsverfahren in allgemeinen Räumen.

In der genannten Arbeit wird auch eine Fehlerabschätzung für ein gemischtes elliptisch-hyperbolisches Problem der Tricomi-Gleichung

$$u_{xx} + f(y)u_{yy} = 0 \text{ mit } \operatorname{sgn} f(y) = \operatorname{sgn} y \tag{17}$$

durchgeführt. Auch hierbei erhält man die günstigste für den ganzen Bereich geltende Fehlerabschätzung, wenn man am Rande eine Annäherung nach dem Prinzip *Ib* vornimmt.

§ 3. *Annäherung nach dem Prinzip des kleinsten Fehlerbetragmaximums.*

Es werde daher nach der bereits von Tschebyscheff aufgestellten Methode der Annäherung einer gegebenen Funktion $f(x_j)$ nach dem Prinzip *Ib* eine Linearkombination (1) aus gegebenen Funktionen $u_\nu(x_j)$ mit der Forderung (3) (6) gesucht. Dabei sei der Einfachheit halber $\varphi \equiv 1$. Ist wieder B ein abgeschlossener beschränkter Bereich und sind f und u_ν dort stetig, so ist das Fehlerbetragmaximum eine stetige Funktion M der Koeffizienten a_ν

$$\operatorname*{Max}_B |\, \varepsilon_p(x_j) \,| = M\,(a_0, \ldots, a_p) \tag{18}$$

und m sei die untere Grenze von $M(a_\varrho)$ bei variierenden a_ϱ.

Voraussetzung I: In B gibt es $p + 1$ Punkte $P_0, \ldots, P_p$, so daß die Determinante (8) $\neq 0$ ist. Dann lässt sich die Existenz von mindestens einer „Minimallösung", d.h. einer Linearkombination (1), für welche M den Wert m annimmt, beweisen.

Voraussetzung II: Zu dem System der u_ν gehört ein Graf G mit den Eigenschaften:

1) Die Knoten lassen sich so in zwei Sorten einteilen, daß jede Strecke als Endpunkte Knoten verschiedener Sorten hat;

2) wenn $\sum_{\nu=0}^{p} a_\nu u_\nu$ im Innern jeder Strecke an mindestens einem Punkt verschwindet, so sind alle $a_\nu = 0$.

Unter den Voraussetzungen I, II gilt der folgende Abschätzungssatz: Zu der Linearkombination Q_p von (1) wird ε_p nach (2) an allen Knoten K_σ des Grafen ermittelt: $\varepsilon_p\,(K_\sigma) = \varepsilon'_\sigma$. Es sei ε'_σ in allen Knoten der einen Sorte positiv und in allen Knoten der anderen Sorte negativ. Mit

$$m' = \operatorname*{Min}_{\sigma} |\, \varepsilon'_\sigma \,| \text{ und } M' = \operatorname*{Max}_{B} |\, \varepsilon_p \,| \tag{19}$$

gilt

$$m' \leqq m \leqq M'. \tag{20}$$

Die bekannten Fälle der Annäherung einer Funktion von einer reellen Veränderlichen durch Polynome oder trigonometrische Ausdrücke [11])[12]) sind hier als Spezialfälle enthalten. Es seien aber auch einige weitere Beispiele genannt:

1) Bei Annäherung einer Funktion $f(x, y)$ durch in x und y lineare Funktionen kann man als Graf ein beliebiges konvexes Viereck in der x-y-Ebene verwenden. Z.B. werde

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

durch

$$Q = \tfrac{11}{12} - \tfrac{1}{3}\,(x + y)$$

in dem Quadrat $0 \leqq x,\ y \leqq 1$ angenähert. Mit dem Rand dieses Quadrates als Graf gilt $m' = M' = \frac{1}{12}$. Nach (20) weiss man daher sofort, daß auch $m = \frac{1}{12}$ und Q eine Minimallösung ist.

2) Es werde eine Funktion $f(x, y)$ durch ein in x und y quadratisches Polynom angenähert. Dann lässt sich als Graf u.a. die folgende Konfiguration aus 7 Knoten verwenden: Die drei Ecken P_1, P_2, P_3 eines beliebigen Dreiecks, drei Punkte P_4, P_5, P_6 auf den Seiten dieses Dreiecks und ein Punkt P_0 im Innern des Dreiecks P_4, P_5, P_6. Der Graf besteht aus allen Strecken des Dreiecks P_1, P_2, P_3, dem Dreieck P_4, P_5, P_6 und den drei Verbindungsstrecken von P_0 mit P_4, P_5, P_6.

---

[11]) C. DE LA VALLÉE POUSSIN: Leçons sur l'Approximation des Fonctions d'une variable Réelle. Paris (1952) Kap. VI and VII.

[12]) N. L. ACHIESER: Vorlesungen über Approximationstheorie, Berlin 1953, vgl. insbesondere Kap. II.

(24a) HAMBURG 13, HARVESTEHUDERWEG 81.

METHODS OF FUNCTIONAL LINEAR ANALYSIS IN MATHEMATICAL PHYSICS

"A priori" estimates for the solutions of boundary value problems

GAETANO FICHERA

Let D be a connected domain (open set) of the Euclidean space R_n. Let $E(u)$ be an arbitrary linear differential operator, whose coefficients are defined in D. Let B_1 be a set contained in the boundary B of D. We denote by S the space of all admissible functions and by $L(u)$ a linear operator that transforms u into a function defined on B_1. We deal with the following boundary value problem:

$$E(u) = f \quad \text{in } D, \qquad L(u) = g \quad \text{on } B_1. \tag{1}$$

The interest connected with the problem of giving upper and lower bounds for the unknown solution and its derivatives is well known. This problem was considered both in ancient and recent papers from several points of view. The simplest statement of the problem requires the construction of a pair of numbers b' and b'' such that:

$$b' \leqq u(x) \leqq b''$$

for every fixed point x of D, with a similar requirement for the derivatives of u. Many researches were performed in this connection, however such a solution of the proposed problem fails to bring into evidence the functional dependence of u on the data f and g.

We shall adopt a different enunciation of the problem. Let us suppose that the space S of all admissible functions is a Banach space and that (f,g) is a vector of a Banach space S'. Let us consider the linear transformation:

$$T(u) \equiv (E(u), L(u))$$

which maps S on a linear subset of S'. Let ω be the zero-element in S and consider the functional:

$$I(u) = \frac{\| u \|}{\|T(u)\|} \qquad u \subset S - \omega.$$

Our problem is the following:

a) to prove that $I(u)$ is bounded;

b) to compute an upper bound for $I(u)$.

In the terminology of linear abstract Analysis our problem consists in showing the boundedness of the inverse transformation T^{-1} and in computing its norm.

The solution of the stated problem is interesting for attaining several aims, for example in the existence theory for boundary value problems. In this particular connection only question a) is to be answered, since the only information required on T^{-1} is its boundedness. But in many other applications and, especially, in Mathematical Physics, the numerical value of $\|T^{-1}\|$ or, at least, bounds for such a number are required.

This leads to the consideration of problem b). In other words, by solving both a) and b), not only it is proved that the following inequality:

$$\| u \| \leqq K \| T(u) \|$$

holds, but also the value of K is determined.

The quantitative aspect of the above problem will be mainly considered in this paper.

This aspect is connected — as observed above — with many problems of applied Mathematics.

Let us suppose — for instance — that $E(u)$ be a 2nd order positive elliptic self-adjoint differential operator

$$E(u) \equiv \sum_{h,k}^{1,n} \frac{\partial}{\partial x_h}\left(a_{hk}\frac{\partial u}{\partial x_k}\right) + cu \qquad (c \leqq 0)$$

and let us put $L(u) \equiv u$ on B, and $B_1 \equiv B$.

We consider the following particular case of problem (1):

$$E(u) = f \text{ in } D, \qquad L(u) \equiv u = 0 \text{ on } B. \tag{2}$$

Let us put:

$$\|u\| = \left(\int_D u^2 d\tau\right)^{\frac{1}{2}}, \quad \|T(u)\| = \left(\int_D |E(u)|^2\, d\tau\right)^{\frac{1}{2}} \qquad (d\tau = dx_1 dx_2 \ldots dx_n) \tag{3}$$

and therefore:

$$I(u) = \frac{\left(\int_D u^2\, d\tau\right)^{\frac{1}{2}}}{\left(\int_D |E(u)|^2\, d\tau\right)^{\frac{1}{2}}}.$$

If we now define:

$$\varkappa(D) = l.u.b.\ I(u),$$

it is well known that $[\varkappa(D)]^{-1}$ is the lowest eigenvalue λ, of the following homogeneous boundary value problem:

$$E(u) + \lambda u = 0 \text{ in } D, \qquad u = 0 \quad \text{on } B.$$

In this particular case now considered, our general problem coincides with the classical problem which consists in assigning lower bounds for λ_1.

The fact that $\varkappa(D)$ is an increasing function of D [1]) gives a rough approach to the solution of the problem.

An upper bound for $\varkappa(D)$ will be furnished by the knowledge of a particular domain $D' \supset D$, for which λ_1 is known.

But this simple procedure fails in analogous problems. With the same meaning for $E(u)$, $L(u)$ and B_1, let us consider the following boundary value problem:

$$E(u) = 0 \quad \text{in } D, \qquad L(u) \equiv u = g \quad \text{on } B$$

and define:

$$\|u\| = \left(\int_D u^2 \, d\tau\right)^{\frac{1}{2}}, \qquad \|T(u)\| = \left(\int_B u^2 \, d\sigma\right)^{\frac{1}{2}},$$

$$I(u) = \frac{\left(\int_D u^2 \, d\tau\right)^{\frac{1}{2}}}{\left(\int_B u^2 \, d\sigma\right)^{\frac{1}{2}}}, \qquad \mu(D) = l.u.b.\ [I(u)]^2.$$

Even in this case $\mu(D)$ is connected with an eigenvalue problem, namely:

$$EE(v) = 0 \text{ in } D; \quad \frac{\partial v}{\partial \nu} + \mu E(v) = 0, \quad v = 0 \text{ on } B$$

$\left(\frac{\partial v}{\partial \nu}\right.$ denotes the inner conormal derivative associated with $\left.E\right)$.

More precisely, $\mu(D)$ is the greatest eigenvalue of the above problem, but it appears very complicated to investigate the dependence of $\mu(D)$ on D and it is probably not true that $\mu(D)$ is an increasing function of D, as in the preceding example. On the other hand, even in problem (2), if we abandon the Hilbert normalisation (3) and employ, for instance, either an $L^{(p)}$-norm or the norm of continuous functions, then the dependence of $l.u.b.\ I(u)$ on D is no longer helpful in order to solve our problem. Thus a different approach has to be attempted.

In this paper a rather general procedure is outlined in connection with straight-forward problems of applications. The underlying idea is very simple. It consists in associating to the functional $I(u)$ and to the Banach space S a functional $F(w)$ and a class Γ of vectors w, satisfying the following conditions:

I) Γ is not empty if $I(u)$ is bounded.

---

[1]) That is to say $D \subset D'$ implies $\varkappa(D) \leqq \varkappa(D')$.

II) For every $u \subset S - \omega$ and $w \subset \Gamma$ the inequality:

$$I(u) \leqq F(w)$$

holds.

We shall now employ the $L^{(p)}$-norm in S and in S'. Then a very fortunate and rather unexpected circumstance holds. In some problems the class Γ does not depend upon p, while the functional $F(w)$, for some fixed $w \subset \Gamma$, is a bounded function of p. This implies that from the inequality:

$$\left(\int_D |u|^p \, d\tau\right)^{\frac{1}{p}} \leqq F(w) \, \|T(u)\| \leqq K \, \|T(u)\|$$

the following one can be derived, when $p \to \infty$:

$$\max_{D+B} |u| \leqq K \max |T(u)|.$$

That is to say, pointwise bounds for u are, in a certain sense, obtainable as limiting cases of integral bounds.

In this paper the principal results connected with 2nd order parabolic and elliptic partial differential equations are presented. Systems of linear partial differential equations are also considered. Results with respect to the problem of giving "a priori" bounds for the solutions of systems of equations seem to have not been too much considered in the mathematical literature. The system of elasticity will be particularly considered and bounds for displacements and stresses will be derived.

1. *Parabolic equations.* — Let us consider in Euclidean $(n+1)$-space $x_1, x_2, \ldots, x_n, t$ the differential operator:

$$E(u) \equiv \sum_{h,k}^{1,n} a_{hk}(x,t) \frac{\partial^2 u}{\partial x_h \partial x_k} + \sum_{h=1}^{n} b_h(x,t) \frac{\partial u}{\partial x_h} + c(x,t)u - \frac{\partial u}{\partial t} \quad [x \equiv (x_1, x_2, \ldots, x_n)]$$

and suppose that:

$$E_0(u) \equiv \sum_{h,k}^{1,n} a_{hk}(x,t) \frac{\partial^2 u}{\partial x_h \partial x_k} + \sum_{h=1}^{n} b_h(x,t) \frac{\partial u}{\partial x_h} + c(x,t)\, u$$

is positive elliptic and that $c(x,t) \leqq 0$ for any fixed t. The coefficients a_{kk} are supposed of class two in $D + B$ (viz. continuous in $D + B$ with their first and second derivatives), b_h of class one and c continuous. The domain D is supposed to be included between the two hyperplanes $t = 0$, $t = T > 0$. For each t_0 in the closed interval $(0,T)$, the intersection of $D + B$ with the hyperplane $\pi(t_0)$, having equation $t = t_0$, is a regular n-dimensional domain $D(t_0)$ lying in $\pi(t_0)$. Let $B[D(t_0)]$ be the boundary of $D(t_0)$. We suppose the following conditions satisfied:

$$\pi(0) \cdot B = D(0), \qquad \pi(T) \cdot B = D(T), \qquad \pi(t_0) \cdot B = B[D(t_0)] \qquad (0 < t_0 < T).$$

We sketch briefly the main results connected with parabolic equations,

by indicating: a) the different possible assumptions for the space S of admissible functions; b) the definitions of $\|u\|$, $\|T(u)\|$ and, consequently, of $I(u)$; c) the class Γ and the functional $F(w)$ corresponding to S and $I(u)$.

Proofs will not be given here.

Any function u belonging to S is supposed to be continuous with its first derivatives in $D + B$. Second derivatives are supposed to be continuous in D and to satisfy convenient hypotheses, in order that $\|T(u)\|$ make sense. For instance, when in defining $\|T(u)\|$ we use $\int_D |E(u)|^p d\tau$, we shall suppose that u possesses such second derivatives that the above integral is finite.

It has to be observed that the assumption on the continuity of u in $D + B$ could be weakened.

I.

a) S is the space of solutions of the homogeneous equation $E(u) = 0$;

b)
$$\|u\| = \left(\int_D u^p \, d\tau\right)^{\frac{1}{p}}, \quad \|T(u)\| = \left(\int_{D(0)+B_1} u^p \, d\sigma\right)^{\frac{1}{p}},$$

where p is a positive even integer [2]) and B_1 the part of the boundary B verifying the conditions: $x \subset B[D(t)]$, $0 < t < T$;

c) Γ is the class of the functions w of class two in $D + B$ and verifying the conditions:

$$E^*(w) > 0 \qquad \text{in } D + B, \qquad w = 0 \qquad \text{on } B_1;$$

$$F(w, p) = \left| \frac{\max\limits_{B_1} \left| \dfrac{\partial w}{\partial \nu} \right| + \max\limits_{D(0)} |w|}{\min\limits_{D+B} E^*(w)} \right|^{\frac{1}{p}},$$

$\dfrac{\partial w}{\partial \nu}$ is the conormal derivative associated with the elliptic operator E_0 and E^* is the adjoint operator of E.

Γ is not empty and for $u \subset S - \omega$ and $w \subset \Gamma$ the inequality

(4)
$$I(u) \leqq F(w,p)$$

holds.

That is to say:

$$\boxed{\left(\int_D u^p \, d\tau\right)^{\frac{1}{p}} \leqq F(w, p)\left(\int_{D(0)+B_1} u^p \, d\sigma\right)^{\frac{1}{p}}}.$$

[2]) This hypothesis on p will be used throughout this paper. However we can suppose p an arbitrary real number not less than 1, provided that we restrict ourselves to consider non-negative functions of S.

Since $\lim_{p\to\infty} F(w, p) = 1$, the limiting process gives:

$$\max_{D+B} |u| \leqq \max_{D(0)+B_1} |u|.$$

This is the known maximum principle for parabolic equations [3]).

II.

a) S is the space of function verifying the above specified qualitative properties and vanishing on $D(0) + B_1$;

b) $$\|u\| = \left(\int_D u^p \, d\tau\right)^{\frac{1}{p}}; \quad \|T(u)\| = \left(\int_D |E(u)|^p d\tau\right)^{\frac{1}{p}};$$

c) Γ is defined by: $E^*(w) > 0$ in $D + B$, $w \leqq 0$ in D,

$$F(w, p) = p \frac{\max_{D+B} |w|}{\min_{D+B} E^*(w)}.$$

Inequality (**4**) holds, that is to say:

$$\left(\int_D u^p \, d\tau\right)^{\frac{1}{p}} \leqq F(w, p) \left(\int_D |E(u)|^p \, d\tau\right)^{\frac{1}{p}}.$$

In this case, the limiting process does not apply. However, if we assume the additional hypothesis $c(x,t) < 0$ in $D + B$, the following inequality can be proved for p large enough and by a proper choice of w:

$$\boxed{\left(\int_D u^p \, |c| \, d\tau\right)^{\frac{1}{p}} \leqq \frac{p}{p + 1 - q_0} \left(\int_D \left|\frac{E(u)}{c}\right|^p |c| \, d\tau\right)^{\frac{1}{p}}} \tag{5}$$

where q_0 is the least integer such that:

$$q_0 \, |c| > c + \sum_{h,k}^{1,n} \frac{\partial^2 a_{hk}}{\partial x_h \partial x_k} - \sum_{h=1}^{n} \frac{\partial b_h}{\partial x_h}.$$

The limiting process gives:

$$\boxed{\max_{D+B} |u| \leqq \max_{D+B} \left|\frac{E(u)}{c}\right|}. \tag{6}$$

Inequalities (5), (6) hold even if we substitute the boundary condition $u = 0$ on B_1 by $\frac{\partial u}{\partial l} = 0$ on B_1, or $u = 0$ on B_1', $\frac{\partial u}{\partial l} = 0$ on B_1'';

[3]) See [16]. Numbers in brackets refer to Bibliography at the end of the paper.

[$B_1 = B_1' + B_1''$, B_1'. B_1'' = empty set; where $\frac{\partial}{\partial l}$ denotes differentiation along an arbitrary direction lying in $D(t)$ and penetrating into the interior of $D(t)$].

2. *Elliptic equations.* — Analogous results to those obtained for parabolic equations can be established for elliptic equations, under the same hypotheses for coefficients and function u (see pag. 4 and 5). Suppose that:

$$E(u) \equiv \sum_{h,k}^{1,n} a_{hk}(x) \frac{\partial^2 u}{\partial x_h \partial x_k} + \sum_{h=1}^{n} b_h(x) \frac{\partial u}{\partial x_h} + c(x)u, \quad (c \leqq 0)$$

and that D is a regular bounded domain of R_n.

Let us consider some examples.

I.

a) S is the space of the solutions of $E(u) = 0$;

b)
$$\|u\| = \left(\int_D u^p \, d\tau\right)^{\frac{1}{p}}, \quad \|T(u)\| = \left(\int_D u^p \, d\sigma\right)^{\frac{1}{p}};$$

c) Γ is the class of functions w such that: $E^*(w) > 0$ in $D + B$, $w = 0$ on B;

$$F(w, p) = \left(\frac{\max_B \left| \frac{\partial w}{\partial \nu} \right|}{\min_{D+B} E^*(w)} \right)^{\frac{1}{p}}$$

$\left(\frac{\partial}{\partial \nu}\right.$ is the conormal differentiation associate with $\left.E\right)$.

We are led to the inequality:

$$\boxed{\left(\int_D u^p \, d\tau\right)^{\frac{1}{p}} \leqq F(w, p) \left(\int_B u^p \, d\sigma\right)^{\frac{1}{p}}}.$$

The limiting process furnishes the well known inequality:

$$\max_{D+B} |u| \leqq \max_B |u|.$$

II.

a) S is defined as in I but with the additional condition $u = 0$ on B_1 [B_1 is a subset of B];

b)
$$\|u\| = \left(\int_{B_2} u^p \, d\sigma\right)^{\frac{1}{p}}; \quad \|T(u)\| = \left(\int_{B_2} \left| \frac{\partial u}{\partial \nu} \right|^p d\sigma\right)^{\frac{1}{p}}, \quad [B_2 = B - B_1]$$

c) Γ is defined by: $E^*(w) \geqq 0$ in D, $w \leqq 0$ in D, $\dfrac{\partial w}{\partial \nu} - bw > 0$ on B_2

$$[b = \sum_{h=1}^{n} \left(b_h - \sum_{k=1}^{n} \frac{\partial a_{hk}}{\partial x_k} \right) \cos\,(x_h, N);\ N \text{ inner normal}];$$

$$F(w, p) = p \frac{\max\limits_{B_2} |w|}{\min\limits_{B_2} \left(\dfrac{\partial w}{\partial \nu} - b \right)};$$

$$\boxed{\left(\int_{B_2} u^p d\sigma \right)^{\frac{1}{p}} \leqq F(w, p) \left(\int_{B_2} \left| \frac{\partial u}{\partial \nu} \right|^p d\sigma \right)^{\frac{1}{p}}}.$$

The limiting process cannot be applied in this case. But, if we replace the condition $u = 0$ on B_1 by the condition $\dfrac{\partial u}{\partial \nu} = 0$ on B_2, the following inequality can be obtained:

$$\boxed{\max_{D+B} |u| \leqq \max_{B_1} |u|}.$$

III.

a) S is the space of function satisfying the conditions $u = 0$ on B_1, $\dfrac{\partial u}{\partial \nu} = 0$ on B_2;

b) $$\|u\| = \left(\int_D u^p \, d\tau \right)^{\frac{1}{p}}, \quad \|T(u)\| = \left(\int_D |E(u)|^p d\tau \right)^{\frac{1}{p}};$$

c) Γ is defined by: $E^*(w) > 0$ in $D + B$, $w \leqq 0$ in D, $\dfrac{\partial w}{\partial \nu} - bw \geqq 0$ on B_2

$$F(w, p) = p \frac{\max\limits_{D} |w|}{\min\limits_{D} E^*(w)};$$

$$\left(\int_D u^p \, d\tau \right)^{\frac{1}{p}} \leqq F(w, p) \left(\int_D |E(u)|^p \, d\tau \right)^{\frac{1}{p}}.$$

The case $p = 2$, $n = 2$, $E = E^*$ is particularly interesting, since $[F(w,2)]^{-1}$ gives a lower bound for the least eigenvalue of an anisotropic membrane free along a part B_2 of its boundary and fixed along the remaining part B_1. As far as I know, such bounds, connected with eigenvalue problem with mixed boundary

conditions, had not been given before, even in the case of the isotropy [4]). In order to give a simple example, let us consider more particularly the case of isotropy. The eigenvalue problem is now the following:

$$\Delta_2 u + \lambda u = 0 \text{ in } D, \qquad u = 0 \text{ on } B_1, \qquad \frac{\partial u}{\partial N} = 0 \text{ on } B_2$$

($\Delta_2 \equiv \sum_{k=1}^{n} \frac{\partial^2}{\partial x_k^2}$, and $\frac{\partial}{\partial N}$ denotes differentiation along the inner normal).
Let us suppose that B_2 is expressible by $x_2 = f(x)$ $(a \leqq x_1 \leqq b)$. Suppose that no point of D satisfies the conditions $a \leqq x_1 \leqq b$, $x_2 < f(x_1)$ and let the interval (a_2, b_2) be the projection of D on the x_2 - axis. We can then choose $w = [x_2 - \min f(x_1)]^2 - \delta^2$ where δ in the greatest of the two numbers $|a^2 - \min f(x)|$, $|b_2 - \min f(x)|$. For the lowest eigenvalue λ_1 we then obtain the following lower bound:

$$\boxed{\lambda_1 \geqq \frac{1}{\delta^2}}.$$

The two sets of real numbers, which are given by $I(u)$ and $F(w,p)$ for u vanring in $S - \omega$ and w vanring in Γ, are not adjacent, in the cases considered above for elliptic as well as for parabolic equations. However in the particular case $p = 2$ many of the results can be improved in such a way that the above mentioned classes can be adjacent for a proper choice of Γ and $F(w)$. This seems to be very useful in practical applications, where an upper bound for $I(u)$ very close to *l.u.b.* $I(u)$ is needed. Results of such kind were discussed in a previous paper [5]).

3. *System of partial differential equations.* — Let us consider m-vectors $u \equiv (u_1\ u_2, \ldots, u_m)$, whose components are functions defined in the same domain D of Euclidean n-space. We shall consider boundary value problems of the following kind:

$$E_r(u) \equiv \sum_{s=1}^{m} E_{rs}(u_s) \text{ in } D, \quad L_r(u) = g_r \text{ on } B,$$

where $E_{rs}(u)$ is defined as follows:

$$E_{rs}(u_s) \equiv \sum_{h,\,k}^{1,\,n} a_{hk}^{(rs)} \frac{\partial^2 u_s}{\partial x_h \partial x_k} + \sum_{h=1}^{n} b_h^{(rs)} \frac{\partial u_s}{\partial x_h} + c^{(rs)} u_s \qquad (E_{rs} = E_{sr}),$$

and $L_r(u)$ is a linear operator transforming u into a function defined on B.

---

[4]) For the *uniform* case (viz. u = 0 on whole B) see [1], [5].

[5]) See [5].

We shall assume that the following quadratic forms are respectively positive, and negative, semidefinite for any x in $D + B$:

$$\sum_{r,s}^{1,m} \sum_{h,k}^{1,n} a_{hk}^{(rs)} \lambda_h^{(r)} \lambda_k^{(s)}, \qquad \sum_{r,s}^{1,m} c^{(rs)} \lambda_r \lambda_s.$$

Some results, already established for a single equation can be extended to systems of the above mentioned type, under analogous hypotheses for coefficients and unknown functions as above (see pag. 4 and 5).

I.

a) S is the space of m vector $u \equiv (u_1, u_2 \ldots, u_m)$ vanishing on B;

b) $$\|u\| = \left(\int_D \sum_{r=1}^{m} |u_r|^2 \, d\tau \right)^{\frac{1}{2}}, \quad \|T(u)\| = \left(\int_D \sum_{r=1}^{m} |E_r(u)|^2 \, d\tau \right)^{\frac{1}{2}};$$

c) Γ is defined by: $w \leqq 0$ in D, $\sum_{r,s}^{1,m} E_{rs}^* (w) \lambda_r \lambda_s > 0$ in $D + B$ (this means that the quadratic form on the left member is positive definite for any $x \subset D + B$),

$$F(w) = 2m \max_{D+B} |w| \max_{D+B} |A(w)|.$$

$A(w)$ is the function which in any point x of $D + B$ coincides with the greatest element in the principal diagonal of the inverse matrix of the following matrix: $(E_{rs}^* (w))$. For any $u \subset S$ and $w \subset \Gamma$ the inequality holds

$$I(u) = \frac{\|u\|}{\|T(u)\|} \leqq F(w).$$

That is to say:

(7) $$\left(\int_D \sum_{r=1}^{m} |u_r|^2 \, d\tau \right)^{\frac{1}{2}} \leqq F(w) \left(\int_D \sum_{r=1}^{m} |E_r(u)|^2 \, d\tau \right)^{\frac{1}{2}}.$$

In the particular case $E_{rs} \equiv E_{rs}^*$, $[F(w)]^{-1}$ is a lower bound for the lowest eigenvalue of the following problem:

$$E_r(u) + \lambda u_r = 0 \text{ in } D, \qquad u_r = 0 \text{ on } B.$$

II.

a) S is the space of the solutions of the system $E_r(u) = 0$;

b) $$\|u\| = \left(\int_D \sum_{r=1}^{m} |u_r|^2 \, d\tau \right)^{\frac{1}{2}}, \quad \|T(u)\| = \left(\int_B \sum_{r=1}^{m} |u_r|^2 \, d\sigma \right)^{\frac{1}{2}};$$

c) Γ is defined by: $w \leqq 0$ in D, $w = 0$ on B, $\sum_{r,s}^{1,m} E^*(w) \lambda_r \lambda_s > 0$ in $D + B$;

$$F(w) = 2m\left[\max_{B}\left|\sum_{r,s}^{1,m}\frac{\partial w}{\partial \nu^{(rs)}}\right|\max_{D+B}|A(w)|\right]^{\frac{1}{2}},$$

$\frac{\partial w}{\partial \nu^{(rs)}}$ is the inner conormal derivative associated with E_{rs}, and $A(w)$ is defined as above.

The final inequality is:

$$\boxed{\left(\int_D \sum_{r=1}^{m}|u_r|^2\,d\tau\right)^{\frac{1}{2}} \leqq F(w)\left(\int_B \sum_{r=1}^{m}|u_r|^2\,d\sigma\right)^{\frac{1}{2}}}. \tag{8}$$

III.

a) S is defined by the condition $u = 0$ on B_1 (B_1 is a proper subset of B), $\sum_{s=1}^{m}\frac{\partial u_s}{\partial \nu^{(rs)}} = 0$ on $B_2 = B - B_1$ $(r = 1, 2, \ldots, m)$;

b) and *c*) $\|u\|$, $\|T(u)\|$, $F(w)$ are defined as in I, and the same for Γ with the additional condition

$$\sum_{rs}^{1,m}\left(\frac{\partial w}{\partial \nu^{(rs)}} - b^{(rs)}\right)\lambda_r\lambda_s \geqq 0 \text{ in } B_2 \quad \left[b^{(rs)} = \sum_{h=1}^{n} b_h^{(rs)} - \sum_{k=1}^{n}\frac{\partial a_{hk}^{(rs)}}{\partial x_k}\cos(x_h, N)\right].$$

Even in this case inequality (7) holds.

When $E_{rs} = E_{rs}^*$, $(F(w))^{-1}$ is a lower bound for the lowest eigenvalue of the eigenvalue problem with mixed boundary conditions.

Several other examples could be considered in connection with the general systems introduced in this section. However, we prefer to consider some particular cases which are of interest in the mathematical theory of elasticity.

4. *System of elasticity.* — We are now concerned with the following particular case:

$$n = m, \quad E_{rs}(u_s) = \delta_r^s \Delta_2 u_s + k\frac{\partial^2 u_s}{\partial x_r \partial x_s}$$

δ_r^s is the Kronecker's delta, k a positive constant.

We are thus led to consider bounds for the solutions of the following vectorial differential equation:

$$\Delta_2 u + k\ \text{grad div}\ u = f.$$

It is well known that for $m = 3$, the above equation expresses the condition for the equilibrium of an isotropic elastic body D, subjected to body-forces f. The vector $u(x)$ represents the displacement of a point x of D.

Let us suppose that the body is fixed along its boundary. That is to say $u = 0$ on the boundary B of D.

Let δ be the radius of any circular domain Σ containing D and let us

choose $w = \varrho^2 - \delta^2$ where ϱ denotes the distance between x and the center of Σ. The inequality (7) gives:

$$\int_D |u|^2 \, d\tau \leqq \left(\frac{m\delta^2}{m+k}\right)^2 \int_D |\Delta_2 u + k \,\mathrm{grad}\, \mathrm{div}\, u|^2 \, d\tau \,. \tag{9}$$

In connection with the boundary value problem: $\Delta_2 u + k$ grad div $u = 0$ in D, $u = g$ on B, we have the following inequality:

$$\int_D |u|^2 \, d\tau \leqq m(m+k) \max_B |\mathrm{grad}\, w| \max_{D+B} |A(w)| \int_B |u|^2 \, d\sigma \,. \tag{10}$$

The above inequality (10) is a particular case of inequality (8) and w is therefore an arbitrary non-negative function, vanishing on B and such that the quadratic form $\sum_{r,s}^{1,m} E_{rs}(w)\, \lambda_r \lambda_s$ is positive definite for any $x \subset D + B$. We can choose $w = \varrho^2 - \delta^2$ if D is a circular domain having radius δ, or $w = \prod_{s=1}^{m} (a_s - x_s)(x_s - b_s)$ in the case of a rectangular domain, defined by $a_s < x_s < b_s$ $(s = 1, 2, \ldots, m)$.

In order to derive pointwise bounds from the integral bounds, we shall use the following inequalities:

$$|u(x)| \leqq \frac{m}{\omega^{\frac{1}{2}} [\delta(x)]^{\frac{m}{2}}} \frac{\alpha^m}{\alpha^2 - 1} \left(1 + \alpha^{-\frac{m}{2}+2}\right) \left(\int_D |u|^2 \, d\tau\right)^{\frac{1}{2}} , \tag{11}$$

$$|u_{x_k}(x)| \leqq \frac{m+1}{\omega^{\frac{1}{2}} [\delta(x)]^{\frac{m}{2}+1}} \frac{\alpha^{m+1}}{\alpha^3 - 1} \left(1 + \alpha^{\frac{m}{2}+3}\right) \left(\int_D |u|^2 d\tau\right)^{\frac{1}{2}} , \tag{12}$$

where $\delta(x)$ is the distance of x from B, ω is the measure of the unit m-sphere; α an arbitrary real number, greater than 1.

The above inequalities hold for any solution of the homogeneous equation (1).

By combining (10) with (11), (12) we attain pointwise bounds for $u(x)$ and its first derivatives, that is to say for displacements and stresses.

Pointwise bounds can be given even in the case when body forces are present. The approach is very simple. Let $u^{(0)}$ be a particular solution of the equation $\Delta_2 u + k$ grad div $u = f$; $u^{(0)}$ is easily obtained by using the funda-

mental solutions given by Somigliana. By applying (12) to the vector $u - u^{(0)}$ bounds of following kind

$$|u_{x_k}(x)| \leqq G[\delta(x)]\left\{\left(\int_D |u|^2\, d\tau\right)^{\frac{1}{2}} + \max_{D+B} |f|\right\}$$

are obtainable for the derivatives of u, where $G[\delta(x)]$ is a known function depending on the distance of x from B. Analogous bounds can be derived for the displacements. By using (9) and (10) the desired bounds are easily obtained.

From the results of the preceding section it follows that similar bounds can be given for different boundary conditions.

BIBLIOGRAPHY

[1] BARTA, Comptes Rendus, 204, 472—1947.
[2] R. CACCIOPPOLI, Giorn. di Mat., 1950—'51.
[3] J. B. DIAZ—H. J. GREENBERG, Jour. of Math. and Phy., 1948.
[4] J. B. DIAZ—H. J. GREENBERG, Quaterly of Appl. Math., 1948.
[5] G. FICHERA, Annali di Mat. pura e appl., 1954.
[6] K. O. FRIEDRICHS, Duke Math. Jour., 1947.
[7] K. O. FRIEDRICHS, Annals of Math., 1947.
[8] M. GEVREY, Jour. de Math., 1913 and 1914.
[9] G. GRIOLI, Atti IV Congresso Un. Mat. Ital., 1951.
[10] G. GRIOLI, Annali di Mat. pura e appl., 1952.
[11] E. E. LEVI, Ann. di Mat. pura e appl., 1908.
[12] C. G. MAPLE, Doctoral Thesis-Carnegie Inst. of Technology, 1948.
[13] C. MIRANDA, Rend. Acc. Naz. Lincei, 1951.
[14] A. PARAF, Ann. Fac. Toulouse, 1892.
[15] M. PICONE, Rend. Acc. Lincei, 1927.
[16] M. PICONE, Ann. di Mat. pura e appl., 1929.
[17] C. PUCCI, Rend. Acc. Lincei, 1952.
[18] C. PUCCI, Jour. of Mechanics and Analysis, 1953.
[19] A. SIGNORINI, Annali Scuola Norm. Sup. Pisa, 1953.
[20] J. L. SYNGE, Proc. of the R. Soc., 1951.
[21] J. L. SYNGE, Proc. of the R. Irish Ac., 1950.
[22] K. WASHIZU, Jour. of Math. and Phy., 1953.

HILBERT SPACE METHODS IN VARIATIONAL THEORY AND NUMERICAL ANALYSIS

MAGNUS R. HESTENES

Introduction. Calculus of Variations is one of the older branches of mathematics. It was developed almost simultaneously with the calculus. Both Newton and Leibnitz concerned themselves with variational problems. Since its beginning variational theory has enjoyed a continual development in many directions. From time to time efforts have been made to unify the various aspects of variational theories. These efforts have been reasonably successful. It is hoped that the present considerations will be a further step towards this unification.

One of the important problems in the calculus of variations is the study of functionals which are quadratic in character. These are not only important in themselves but also because the study of the second variation of any functional is of this character. A relative complete development of the theory of quadratic functionals is accordingly an important chapter in Calculus of Variations. The author has long felt that this theory could be developed in a unified manner by studying quadratic forms in a Hilbert space. The present paper is a report on the author's activities [1]) in this direction, together with remarks concerning the extension of these ideas to more general problems in the Calculus of Variations. Some connections with numerical analysis are brought forth in the last section.

2. *Typical quadratic Variational problems.*

In the present section we shall describe two variational problems which are sufficiently general to include most of the interesting quadratic problems in the calculations of variations.

Problem I. Let $\mathfrak{A}$ be the class of all arcs x defined by absolutely continuous functions

$$x: \qquad x^j(t) \qquad (a \leqq t \leqq b, j = 1, \ldots, n)$$

with square integrable derivates $\dot{x}^j(t)$ in the finite interval $a \leqq t \leqq b$. We shall be interested in the study of the properties of a functional

$$J(x) = K(x) + \int_a^b 2\,\omega(t, x, \dot{x})\,dt$$

[1]) See M. R. Hestenes, "Application of the theory of quadratic forms in Hilbert space to the calculus of variations", Pacific Journal of Mathematics, Vol 1 (1950) pp. 525-582.

on subspaces of $\mathfrak{A}$. Here $K(x)$ is a quadratic form in the endvalues of x and

$$2\,\omega = P_{jk}x^j x^k + 2Q_{jk}x^j\dot{x}^k + R_{jk}\dot{x}^j\dot{x}^k,$$

repeated indices denoting summation with respect to that index. The class $\mathfrak{A}$ with

$$(x,\ y) = x^j(a)y^j(a) + \int_a^b \dot{x}^j(t)\dot{y}^j(t)dt$$

or the inner product is a Hilbert space over the field of reals.

Almost all variational problems involving simple integrals for which adequate theory exists are equivalent to the study of $J(x)$ on a subclass of $\mathfrak{A}$. For example, if one is concerned with the general problem of Bolza of quadratic type, one is lead to study of $J(x)$ on the subclass $\mathfrak{B}$ of all arcs in $\mathfrak{A}$ satisfying conditions of the forms

$$a_{\mu j}x^j(a) + b_{\mu j}x^j(b) + \int_a^b \{A_{\mu j}x^j + B_{\mu j}\dot{x}^j\}\,dt = 0 \quad (\mu = 1, \ldots p),$$

$$N_{\alpha j}x^j + N_{\alpha j}\dot{x}^j = 0 \qquad\qquad (\alpha = 1, \ldots, m < n).$$

The integral $J(x)$ is said to satisfy the *condition of Legendre* on $\mathfrak{A}$ (in its weak form) in case

$$R_{jk}(t)\Pi^j\Pi^k \geqq h\Pi^i\Pi^i$$

holds with $h = 0$ for almost all t on $a \leqq t \leqq b$ and all $(\Pi) \neq (0)$. If h can be chosen to be positive then $J(x)$ satisfies the *condition of Legendre* on $\mathfrak{A}$ *in its strengthened form.* If the arcs are restricted to lie in $\mathfrak{B}$, then the above conditions of Legendre must be modified by restricting the Π's to satisfy the conditions $N_{ij}(t)\Pi^j = 0$.

Problem II. Let $\mathfrak{A}$ be the class of all functions

$$x: \qquad x^j(t) \qquad\qquad (t \in T,\ j = 1, \ldots\ n)$$

of class $C^{(r)}$ on a bounded open set T in $(t^1, \ldots, t^m)$ — space.
Let

$$(x, y) = \int_T \{x^j(t)y^j(t) + x^j_k(t)y^j_k(t)\}\,dt$$

where the last terms is to be the sum of corresponding derivatives of order $\leqq r$. This space when completed with respect to (x, y) as an inner product forms a Hilbert space $\mathfrak{A}$. We shall be interested in the study of a quadratic form of the type

$$J(x) = K + \int_T R^{\alpha_1 \cdots \alpha_r\ \beta_1 \cdots \beta_r}_{jk}(t)x^j_{\alpha_1 \ldots \alpha_r}\, x^k_{\beta_1 \ldots \beta_r}\,dt$$

where $K(x)$ is continuous in the weak topology. Here

$$x^j_{\alpha_1 \cdots \alpha_r} = \frac{\partial^r x^j}{\partial t^{\alpha_1} \cdots \partial t^{\alpha_r}};$$

for example, $K(x)$ could be a quadratic form in $x^j(t)$ and its derivatives of order $< r$ plus terms involving the r-*th* plus derivative linearly.

It is clear that Problem II continues Problem I for the case $m = 1$. However it is convenient to treat them separately. In problem II for continuous coefficients $R^{\alpha\beta}_{jk} = R^{\beta\alpha}_{kj}$ the condition of Lengendre states that

$$R^{\alpha_1 \cdots \alpha_r \beta_1 \cdots \beta_r}(t)\eta^j\eta^k \xi_{\alpha_1} \cdots \xi_{\alpha_r} \xi_{\beta_1} \cdots \xi_{\beta_r} \geqq 0$$

for all $(\xi) \neq 0$ and $(\eta) \neq 0$, the equality being excluded in the strengthened form.

3. *Fundamental concepts.*

The present section will be devoted to a description of certain concepts in the theory of quadratic forms which are useful in the theory of the calculus of variations.

Consider now a *bounded* symmetric bilinear form $J(x, y)$ in a Hilbert space $\mathfrak{A}$. We shall restrict all our consideration to the field of reals. Let $J(x) = J(x, x)$ be the associated quadratic form. In the terminology of the calculus of variations, $J(x, y)$ is one half the first variation of $J(x)$ at x. Associated with each subspace $\mathfrak{B}$ of $\mathfrak{A}$ is the *J-orthogonal complement* $J\mathfrak{B}$ of $\mathfrak{B}$, namely, all x in $\mathfrak{A}$ such that $J(x, y) = 0$ for all y in $\mathfrak{B}$. If, for example, in Problem I above we select $\mathfrak{B}$ to be the class of all arcs that vanish at $t = a$ and $t = b$, then its J-orthogonal complement is the class of all extremals, that is, all solutions of the Euler equations

$$\frac{d}{dt}\omega_{\dot{x}^j} = \omega_{x^j} \tag{3 : 1}$$

A similar remark holds for Problem II, and for subproblems of I and II. Thus, the theory of homogeneous linear differential equations is closely connected with the concept of J-orthogonolization.

It may happen that the intersection $\mathfrak{B}_0$ of a subspace $\mathfrak{B}$ and its J-orthogonal complement contains an element $x \neq 0$. The dimension n of $\mathfrak{B}_0$ is called the *nullity* of J on $\mathfrak{B}$. If $n = 0$, then $J(x)$ is said to be *nondegenerate* on $\mathfrak{B}$. Classes $\mathfrak{B}_0$ of this type are of particular interest in the calculus of variations, since they may describe solutions of differential equations having special properties. In the illustration cited above, the nullity n of $J(x)$ on $\mathfrak{B}$ is the

number of linearly independent solutions of (3 : 1) in a maximal class that vanish $t = a$ and $t = b$ and hence describes the order of conjugacy of these points.

A quadratic form $J(x)$ will be said to be *nonsingular* on a subspace $\mathfrak{B}$ of $\mathfrak{A}$ in case $\mathfrak{A}$ is the direct sum of $\mathfrak{B}$ and its J-orthogonal complement, or equivalently, in case every bounded linear functional $L(x)$ on $\mathfrak{B}$ is expressible on $\mathfrak{B}$ uniquely in the form $L(x) = J(x, y)$, where y is in $\mathfrak{B}$. If $J(x)$ is nonsingular on every subclass $\mathfrak{B}$ of $\mathfrak{A}$ on which it is nondegenerate, then $J(x)$ is said to be *quasinonsingular* on $\mathfrak{A}$.

It is not difficult to show that quasinonsingularity is a property that is enjoyed by most variational integrals. Moreover, most existence theorems are easy consequences of this property. In examples I and II, the integral $J(x)$ is normally nonsingular on the class $\mathfrak{B}$ of all functions that vanish on the boundary. This means that every function $\mathfrak{z}$ in $\mathfrak{A}$ is expressible uniquely in the form $\mathfrak{z} = x + y$ where x is in $\mathfrak{B}$ and y is a solution of the Euler equations. Thus, $\mathfrak{z}$ and y have the same boundary values, thereby, establishing a solution to one form of the Dirichlet problem. Furthermore, in Problem I, by taking $L(x) = \int_a^b f_j(t)x^j(t)dt$, one sees that the equation

$$\frac{d}{dt}\omega_{xj} - \omega_{xj} = f_j$$

has a unique solution vanishing at a and b, provided the endpoints are not conjugate; assuming that the nonsingularity of $J(x)$ on $\mathfrak{B}$ has been established.

Quasinonsingularity of $J(x)$ is a consequence of ellipticity of $J(x)$. In order to define ellipticity, let us say that a quadratic form $D(x)$ is *positive definite* in case there is a positive number h such that $D(x) \geqq h \| x \|^2$. A quadratic form $K(x)$ will be said to be *completely continuous* if it is continuous in the weak topology. A quadratic form $J(x)$ will be said to be of *elliptic* type on $\mathfrak{A}$ if it is expressible as the sum

$$J(x) = K(x) + D(x) \tag{3:2}$$

where $K(x)$ is completely continuous and $D(x)$ is positive definite. It can be shown that $J(x)$ or $-J(x)$ is of elliptic type if and only if given a sequence $\{x_q\}$ converging weakly to x_0 and having $J(x_q) \to J(x_0)$, then $\{x_q\}$ converges strongly to x_0. If $J(x)$ is of elliptic type it is *weakly lower semi-continuous*. Every weakly lower semi-continuous quadratic form $J(x)$ is of the type

$$J(x) = K(x) + P(x) \tag{3:3}$$

where $P(x) \geqq 0$ and $K(x)$ is completely continuous.

In applications to the calculus of variations, the concept of weak lower

semi-continuity is equivalent to the condition of Legendre in its weak form. Similarly the concept of ellipticity is equivalent to the condition of Legendre in its strengthened form. For multiple integral problems this fact has been established only recently. [1]) This holds for problems of the Bolza type as well as for the simpler problems. Since we are interested mainly in problems of elliptic type, the integrals $J(x)$ are quasinonsingular at least in the class of functions which vanish on the boundary.

A quadratic form $J(x)$ of elliptic type on $\mathfrak{A}$ is of finite nullity on each subspace of $\mathfrak{A}$. It is also of finite index. By the *index* of $J(x)$ on $\mathfrak{A}$ will be meant the dimension of the maximal linear subclass of $\mathfrak{A}$ on which $J(x) < 0$ whenever $x \neq 0$. The index of $J(x)$ on the class of arcs (in Problem I) that vanish at a and b is equal to the sum of the orders of the conjugate points of $t = a$ and $a \leqq t \leqq b$. From the theory of indices one can obtain readily the Sturm oscillation and comparison theorems for differential equations.

From the point of view of applications to the calculus of variations one of the most useful aspects of the theory of Hilbert space is the study of the properties of a quadratic form $J(x)$ on a family of expanding subspaces $\mathfrak{A}(\lambda)$ $(\lambda' \leqq \lambda \leqq \lambda'')$. This is useful in the study of ellipticity, in index theorems, in the theory of conjugate and focal points, that is, in oscillation theory, as well as in spectral theory. To illustrate its use, suppose in Problem I we restrict $\mathfrak{A}$ to the class of arcs vanishing at $t = a$ and $t = b$ and let $\mathfrak{A}(\lambda)$ $(a \leqq \lambda \leqq b)$ be the class of arcs vanishing identically on $\lambda \leqq t \leqq b$. The points of discontinuity of the index $i(\lambda)$ of $J(x)$ on $\mathfrak{A}(\lambda)$ are the conjugate points of $t = a$ and the jump $i(\lambda + o) - i(\lambda)$ is the order of λ as a conjugate point. Note that $i(\lambda - o) = i(\lambda)$ and that $i(a + o) = o$. The index i of $J(x)$ on $\mathfrak{A}$ is equal to the sum $i(b)$ of the orders of conjugate points of $t = a$ on $a \leqq t \leqq b$. These results have been extended to multiple integral problems in a forthcoming thesis by Rene Dennemeyer. This is only one of the many applications of this theory.

As an interesting by product of this theory one can obtain Lagrange multiplier rule from the concept of quasinonsingularity.

It should be noted that we have emphasized the quadratic form $J(x)$ rather than the associated linear operator Ax. This is done because the theory is then presented in a form which appears to be most convenient in application to the calculus of variations.

4. *Extensions to nonlinear problems.*

In the last section it was noted that a quadratic form was of elliptic type if

---

[1]) Remarks in reference to this fact have been made by K. O. Friedrichs, "On the Differentiability of the Solutions of Linear Elliptic Differential Equations," Comm. on Pure and Applied Mathematics, vol. VI, No. 3 (1953), pp. 299-326.

and only if it is the sum of a positive definite form and a completely continuous form. I should like to call your attention to the fact that a similar decomposition appears in the non-linear problems in the calculus of variations. By way of illustration consider an integral

$$J(C) = \int_a^b f(t, x, \dot{x})dt$$

defined over a class of arcs

$$c: \qquad x^j(t) \qquad (a \leqq t \leqq b)$$

with suitable continuity properties. Let

$$J^*(C) = \int_C \{f(t, x, r) + (\dot{x}^j - r^j)f_{\dot{x}^j}(t, x, r)\}\, dt$$

where $r^j(x)$ are suitably chosen slope functions. Then

$$J(C) = J^*(C) + E^*(C), \tag{4:1}$$

where $E^*(C)$ is the integral of the Weierstrass E-function. The strengthened condition of Weierstass is normally equivalent to the condition that

$$E^*(C) \geqq hE_L^*(C) \qquad (h > 0),$$

where E_L is the integral of the E-function for the length integral. This condition is analogous to the condition of positive definiteness. The Hilbert integral $J^*(C)$ has strong continuity properties.

The decomposition (4:1) plays an important role in the calculus of variations. It is the basis for the theory of fields and can be shown to play an important role in the proof of existence theorems. It is the basis for all sufficiency theorems for maxima and minima.

In case $J(c)$ is quadratic in x and $\dot{x}$, then the decomposition (4:1) yields one of the type (3:2) for the case when $\mathfrak{A}$ is the class of arcs that vanishes at $t = a$ and $t = b$. Otherwise further adjustments must be made in the choice of $E^*(c)$.

5. *A minimax principle.*

The study of the integral $J^*(C)$ described in the last section leads to a new approach to the calculus of variations. In order to describe this approach let us write this integral in terms of the Hamiltonian $H(t, q, p)$ commonly used in mechanics. Then the integral takes the form

$$F(q, p) = \int \{p_i dq^i - H(t, q, p)dt\} \tag{5:1}$$

where now q denotes the arc and p an arbitrary function. If we assume that $H(t, q, p)$ is positively regular in p, then for a fixed path q, the functional

$F(q, p)$ attains its maximum when p_i is such that

$$\dot{q}^i = H_{p_i}$$

that is, in case p_i are the momenta along the path. The maximum of $F(q, p)$ is the integral

$$\int L(t, q, \dot{q})dt$$

of the associated Lagrangian L. The trajectories of the dynamical system minimize these integrals (at least over short intervals of time). Thus, these trajectories can be obtained from a minimax principle applied to $F(q, p)$.

Turning now to the parametric problem the ideas suggested in the last paragraph can be expressed as follows:

Consider a line integral

$$\int p_i(x)dx^i \tag{5:2}$$

where the functions p_i are to be normalized by a condition of the form

$$H(x, p) = 1$$

where H is positively homeogeneous of order one in p and has suitable regularity properties. Then for a fixed path x we have

$$\sup \int p_i dx^i = \int f(x, \dot{x})dt \tag{5:3}$$

where $f(x, \dot{x})$ is positively regular. In fact all integrals of this type can be obtained by maximinizing a line integral in this manner. Thus extremals of (5:3) can be obtained from (5:2) by a minimax principle.

The remarks described above indicate that one can develop a large part of the theory of the calculus of variations from the theory of line integrals together with a minimax principle. If one starts from this point of view one is led directly to the theory of fields and to the decomposition (4:1).

6. *Applications to numerical analysis.*

A large class of computational methods are based on constructive existence theorems. These can be developed, for the most part, with the help of a minimizing principle. Thus, calculus of variations, and, for linear problems, Hilbert space theory can play an important role in the development of the theory of numerical analysis. By way of illustration, consider the conjugate gradient method for solving the linear system

$$Ax = k$$

This method was developed independently by E. Stiefel and the author [1]). This method can be put in various forms, all of which can be derived from the case in which A is a positive definite hermitian operator. In this event algorithm takes the following simple form:

$$x_{i+1} = x_i + a_i p_i$$
$$r_{i+1} = r_i - a_i A p_i = k - A x_{i+1}$$
$$p_{i+1} = r_{i+1} + b_i p_i$$
$$a_i = \| r_i \|^2 / (p_i, A p_i)$$
$$b_i = \| r_{i+1} \|^2 / \| r_i \|^2.$$

We have the fundamental relations

$$(r_i, r_j) = 0, \quad (p_i, A p_j) = 0 \quad (i \neq j).$$

These formulas can be applying a minimizing principle to this error function

$$E(x) = (r, A^{-1} r), \qquad r = k - Ax.$$

The residual r is the gradient of $E(x)$, apart from a scalar factor, and the correction vector p_i is called the conjugate gradient of $E(x)$, a terminology suggested by geometrical considerations. This method is applicable to integral equations as well as differential equations. Convergence of this method is very rapid, more rapid than any geometric series.

A discussion of these results can be found in a dissertation by R. M. Hayes [2]).

---

[1]) M. R. Hestenes and E. Stiefel, "Method of conjugate gradients for solving linear systems", Journal of Research, National Bureau of Standards, Vol 49 (1952) pp. 409-436.

[2]) R. M. Hayes, "Iterative methods for solving linear problems in Hilbert space", Dissertation, University of California at Los Angeles, soon to appear in Journal of Research, National Bureau of Standards.

THE UNIVERSITY OF CALIFORNIA LOS ANGELES, CALIFORNIA

AUGUST 15, 1954.

PROBLÈMES MATHÉMATIQUES POSÉS PAR LA MÉCANIQUE STATISTIQUE DE LA TURBULENCE

J. KAMPÉ DE FÉRIET

1. La notion de moyenne joue un rôle fondamental dans la Mécanique Statistique de la Turbulence; les équations, données par O. Reynolds, en 1895, pour les valeurs moyennes des grandeurs définissant le mouvement d'un fluide incompressible, sont la base de presque tous les développements théoriques; mais, si l'on examine les calculs par lesquels Reynolds déduit ses équations de celles de Navrier [7], on voit immédiatement qu'il postule pour la moyenne $\tilde{f}$ d'une fonction f des propriétés très particulières. Il fallut attendre 1930 [6] pour le premier examen vraiment critique de ce problème fondamental. Nous avons, en 1949, [8] donné une formulation abstraite générale du problème et appliqué notre méthode à l'anneau des fonctions ne prenant qu'un nombre fini de valeurs; à G. Birkhoff [2] sont dues plusieurs notions nouvelles notamment celle des ensembles T-réduisants et l'étude du cas des fonctions continues sur un espace compact, repris ensuite par J. Sopka [13]. Mme Dubreil-Jacotin [3] a introduit la notion importante de transformation régulière et sur un point (notre Théorème 6 dans [8]), rectifié nos résultats. Mme Shu-Teh Chen Moy [11] a établi que, si l'ensemble des fonctions considérées est celui des fonctions mesurables sur un espace de probabilité, une transformation de Reynolds n'est autre chose qu'une probabilité conditionnelle: malgré l'élégance de ce résultat, la restriction aux mesures finies est gênante dans beaucoup d'applications; nous nous proposons de montrer, aprés avoir rappelé quelques résultats généraux, comment on peut construire des transformations de Reynolds opérant dans un ensemble de fonctions mesurables lorsque la mesure est seulement σ-finie.

2. Soit X un ensemble quelconque et $\mathscr{R}$ un anneau de fonctions $f(x)$ à valeurs réelles [1] sur X, contenant les constantes [2]

$$f = \alpha c_X \quad \alpha \text{ réel}$$

---

[1]) Le cas des valeurs réelles que nous traitons ici est évidemment le plus important pour les applications, mais le problème se formule d'une façon identique si les f prennent leurs valeurs dans un corps $\mathfrak{K}$.

[2]) Nous désignons par $c_E(x)$ la fonction caractéristique de la partie E de X.

Définition 1. — Une application $f \to Tf$ de $\mathscr{R}$ sur une partie de lui-même est une transformation de Reynolds si elle vérifie:

$$(T_1) \qquad T(f+g) = Tf + Tg$$

$$(T_2) \qquad T(\alpha f) = \alpha Tf$$

$$(T_3) \qquad T(fTg) = TfTg.$$

En outre lorsqu'une topologie est donnée sur $\mathscr{R}$, en désignant un voisinage de f par $V(f)$

$$(T_4) \qquad g \in V(f) \Rightarrow \quad Tg \in V(Tf).$$

La transformation de Reynolds est *normalisée* si:

$$(T_5) \qquad Tc_X = c_X$$

Ne considérant ici que les transformations normalisées, nous passerons toujours cette qualification sous silence. De (T_3) et (T_5) résulte que T est une projection:

$$T^2 f = Tf$$

Soit $\mathfrak{M}_T$ et $\mathfrak{N}_T$ respectivement l'ensemble des fonctions m et n telles que

$$Tm = m \qquad Tn = 0$$

Théorème 1. — [8] p. 168. — $\mathfrak{M}_T$ est un sous-anneau de $\mathscr{R}$.

Théorème 2. — [8] p. 168. — $m \in \mathfrak{M}_T$, $n \in \mathfrak{N}_T \Rightarrow mn \in \mathfrak{N}_T$.

Théorème 3. — [8] p. 168. — Toute $f \in \mathfrak{M}$ se décompose d'une et d'une seule manière en:

$$f = m + n$$

(Décomposition de Reynolds: m = moyenne de f, n = fluctuation).

Définition 2. — On appelle T-idempotent tout idempotent $c \in \mathfrak{M}_T$. Si c_E est un T-idempotent, on dit aussi que la partie E est T-idempotente. Exemple: X = droite réelle, $\mathscr{R}$ anneau de toutes les fonctions réelles: $Tf = f(|x|)$ est une transformation de Reynolds; les T-idempotents sont les parties de X symétriques par rapport à 0.

3. Donnons nous une σ-algèbre $\mathfrak{F}$ de parties de X; nous nous proposons d'étudier les *transformations de Reynolds de l'anneau $\mathscr{R}$ des fonctions mesurables* par rapport à $\mathfrak{F}$. Ordonnons $\mathscr{R}$ en écrivant $f \leqq g$ si $f(x) \leqq g(x)$ pour *tout* $x \in X$; *en guise de condition* topologique, introduisons la condition [3].

$$(T_4') \qquad f \geqq 0 \Rightarrow Tf \geqq 0$$

[3]) C'est G. Birkhoff [2] qui en a montré l'intérêt et l'importance.

Grâce à la décomposition:

$$f = f_+ - f_- \qquad f_+ = \mathrm{Sup}\,(f, 0) \qquad f_- = \mathrm{Sup}\,(-f, 0)$$

on définira [4] pour tout $f \in \mathscr{R}$ sa transformée par:

$$Tf = Tf_+ - Tf_-$$

et on est ramené à étudier la transformée des $f \geqq 0$. Soit donc $\mathfrak{M}$ *l'ensemble des fonctions* f *non négatives mesurables par rapport à* $\mathfrak{F}$; (T_4') signifie que

$$f \in \mathfrak{M} \Longrightarrow Tf \in \mathfrak{M}$$

Théorème 4. — Dans $\mathfrak{M}$, $f \leqq g \Longrightarrow Tf \leqq Tg$.

Si on a $f < +\infty$ la proposition est évidente; si $f(x) = +\infty$ sur un ensemble non vide, une démonstration a été donnée par Mme Moy [11] p. 51.

Cette proposition donne un sens à la condition:

$$(T_4'') \qquad f_n \uparrow f \Longrightarrow Tf_n \uparrow Tf$$

qui nous servira de condition topologique (T_4), puisque (T_4') est maintenant satisfaite d'elle-même. Les transformations que nous considérons sont donc les applications $f \to Tf$ de $\mathfrak{M}$ sur une partie de lui même $\mathfrak{M}_{\mathfrak{T}}$, satisfaisant les conditions (T_1), (T_2), (T_3), (T_4'') et (T_5).

Soit $\mathfrak{F}_{\mathfrak{T}}$ l'ensemble des parties T-idempotentes de X: $F \in \mathfrak{F}_{\mathfrak{T}}$ si et seulement si $Tc_F = c_F$.

Théorème 5. — Mme Moy [11] p. 54: $\mathfrak{F}_{\mathfrak{T}} \subset \mathfrak{F}$ est une σ-algèbre.

Mme Moy suppose, dès le début, que X est un espace de probabilité, c'est à dire qu'on a défini sur $\mathfrak{F}$ une mesure μ telle que $\mu(X) = 1$; mais ceci ne joue aucun rôle dans la démonstration qu'elle donne de cette proposition, ainsi que de la suivante:

Théorème 6. — Mme Moy [11] p. 55: Pour tout $f \in \mathfrak{M}$, Tf est mesurable par rapport à $\mathfrak{F}_{\mathfrak{T}}$.

4. On nomme *fonction simple*, toute fonction de la forme

$$s(x) = \sum_1^n \alpha_j c_{E_j}, \qquad E_j \in \mathfrak{F}, \quad \alpha_j > 0, \quad n \text{ fini}$$

Toute fonction $f \in \mathfrak{M}$ est la limite d'une suite croissante de fonctions simples s_n et réciproquement; comme:

$$Ts = \sum_1^n \alpha_j Tc_{E_j}$$

il est clair que T est définie pour tout $f \in \mathfrak{M}$ quand on connait Tc_E pour tout $E \in \mathfrak{F}$.

---

[4]) Comme nous admettons que Tf_+ et Tf_- prennent la valeur $+\infty$, Tf ne sera définie qu'aux points où, au moins, l'un des deux nombres sera fini.

Théorème 7. — [9] p. 789. Pour que l'ensemble des Tc_E, $E \in \mathfrak{F}$ définisse une transformation de Reynolds dans $\mathfrak{M}$ il faut et il suffit que:

$$(T_5) \qquad Tc_X = c_X$$

et que pour tout $E \in \mathfrak{F}$ et tout $F \in \mathfrak{F}_{\mathfrak{T}}$:

$$(C_1) \qquad 0 \leqq Tc_E \leqq 1$$

$$(C_2) \qquad Tc_E = \Sigma_1^{+\infty} Tc_{E_n} \qquad E = U_1^{+\infty} E_n, \quad E_n \text{ disjoints}$$

$$(C_3) \qquad Tc_{E \cap F} = c_F \, Tc_E$$

5. Le théorème 7 permet de construire effectivement des transformations de Reynolds dans des cas très généraux, en particulier lorsque X est un espace de mesure $(X, \mathfrak{F}, \mu)$ σ-fini [5]

Théorème 8. — Soit une σ-algèbre $\mathfrak{F}_{\mathfrak{T}} \subset \mathfrak{F}$; à toute mesure ν définie sur $\mathfrak{F}$ et telle que:

$$(1) \qquad \nu(F) = \mu(F) \qquad \text{pour tout } F \in \mathfrak{F}_{\mathfrak{T}}$$

nous pouvons faire correspondre une transformation de Reynolds dont l'ensemble des T-idempotents est $\mathfrak{F}_{\mathfrak{T}}$.

En effet choisissons un ensemble $E \in \mathfrak{F}$ et considérons:

$$\nu_E(F) = \nu(E \cap F)$$

ν_E est une mesure définie sur $\mathfrak{F}$ et absolument continue par rapport à μ sur $\mathfrak{F}_{\mathfrak{T}}$; car, sur $\mathfrak{F}_{\mathfrak{T}}$:

$$\mu(F) = 0 \Longrightarrow \nu(F) = 0 \Longrightarrow \nu_E(F) = 0.$$

D'après le théorème de Radon-Nikodym [6], pour chaque $E \in \mathfrak{F}$, il existe donc une fonction de x, $\lambda_x(E) \geqq 0$, mesurable par rapport à $\mathfrak{F}_{\mathfrak{T}}$ telle que [7]

$$(2) \qquad \int_F \lambda_x(E) d\mu = \nu_E(F), \text{ pour tout } F \in \mathfrak{F}_{\mathfrak{T}}.$$

Lorsque E est donné, la fonction de x, $\lambda_x(E)$ est unique, à une équivalence près modulo μ. En posant:

$$Tc_E = \lambda_x(E) \quad \text{pour tout } E \in \mathfrak{F},$$

nous définissons une transformation de Reynolds T dans $\mathfrak{M}$; pour le prouver, il suffit de démontrer que Tc_E satisfait les conditions du théorème 7.

---

[5]) $\mu(E)$ est une fonction d'ensemble définie pour tout $E \in \mathfrak{F}$, non négative, complètement additive et telle qu'il existe un recouvrement dénonbrable $X \subset U_1^{+\infty} E_n$, avec $\mu(E_n) < +\infty$.

[6]) On a le droit de l'appliquer parce que μ est σ-finie.

[7]) Dans cette intégrale et dans toutes celles qui suivent, E est fixe, la variable d'intégration est x.

Pour (C_1), (C_2) et (T_5) (qui peuvent s'interprêter en disant que pour chaque x, $\lambda_x(E)$ définit une mesure de probabilité sur $\mathfrak{F}$) la démonstration est immédiate; il suffit donc d'établir que (C_3) est satisfaite. Or de (2) nous tirons:

$$\int_X \lambda_x(E)d\mu = \nu(E)$$

d'où:

$$\int_X \lambda_x(E \cap F)d\mu = \nu(E \cap F) \tag{3}$$

En remplaçant E par $E \cap F$ dans (2), on obtient:

$$\int_F \lambda_x(E \cap F)d\mu = \nu(E \cap F) \quad F \in \mathfrak{F}_{\mathfrak{T}} \tag{4}$$

Le rapprochement de (3) et (4) montre que:

$$\lambda_x(E \cap F) = 0 \quad \text{si } x \in F' \tag{5}$$

D'autre part en retranchant (4) de (2):

$$\int_F [\lambda_x(E) - \lambda_x(E \cap F)]d\mu = 0$$

mais comme pour chaque x, $\lambda_x(E)$ est une mesure: $\lambda_x(E) \geqq \lambda_x(E \cap F)$ donc:

$$\lambda_x(E \cap F) = \lambda_x(E) \quad \text{si } x \in F. \tag{6}$$

Les relations (5) et (6) sont bien équivalentes à (C_3).

Remarque: Lorsque $\lambda_x(E)$ est connue, on a pour toute $f \in \mathfrak{M}$

$$Tf = \int_X f(y)d\lambda_x(y)$$

6. Exemple: Soit X la droite réelle, μ la mesure de Lebesgue; posons:

$$F_k = \{x : k \leqq x < k + 1\}, \quad k \text{ entier } -\infty < k < +\infty;$$

soit $\mathfrak{F}_{\mathfrak{T}}$ l'ensemble des parties de X qui sont des unions finies ou dénombrables de F_k. Donnons nous une fonction $\varphi(x)$ telle que: (a) $\varphi \geqq 0$ (b) φ est intégrable dans tout intervalle fini, (c) $\int_k^{k+1} \varphi(x)dx = 1$ pour tout entier k. Si nous prenons:

$$\nu(E) = \int_E \varphi(x)dx \quad E \in \mathfrak{F}$$

ν satisfait à (1); la transformation de Reynolds correspondante est donnée par:

$$Tf = \Sigma_{-\infty}^{+\infty} \left[\int_k^{k+1} f(x)\varphi(x)dx \right] c_{F_k}$$

C'est une transformation régulière d'après la terminologie de Mme Dubreil-Jacotin [3].

[1] J. ARBAULD. Comptes-Rendus Paris T. 239. 1954. p. 858—860.

[2] G. BIRKHOFF. Colloque Algèbre et Théorie des Nombres Paris. 1949. p. 143—153.

[3] Mme DUBREIL—JACOTIN. Comptes-Rendus Paris T. 236. 1953. p. 1136—1138.

[4] Mme DUBREIL—JACOTIN. Comptes-Rendus Paris. T. 236. 1953. p. 1950—1951.

[5] Mme DUBREIL—JACOTIN. Comptes-Rendus Paris T. 239. 1954. p. 856—858.

[6] A. A. ISAKSON. (exposé de J. M. BURGERS) 3ème Congrès Inter. Mec. Ap. Stockholm 1930. p. 18—21.

[7] J. KAMPÉ DE FÉRIET. La Science Aérienne. T. 3. 1934. p. 9—34. T. 4. 1935. p. 12—52.

[8] J. KAMPÉ DE FÉRIET. Annales Sté Scientifique de Bruxelles T. 63. 1949. p. 156—172.

[9] J. KAMPÉ DE FÉRIET. Comptes-Rendus Paris T. 239. 1954. p. 787—789.

[10] J. KAMPÉ DE FÉRIET. Comptes-Rendus Paris T. 239. 1954. p. 934—936.

[11] Mme SHU-TEH CHEN MOY. Pacific Journ. of Math. T. 4. 1954 No I, p. 47—63.

[12] O. REYNOLDS. Philos. Trans. R. Soc. (A.) T. 186. Partie I p. 123—164. 1895.

[13] J. SOPKA. Harvard Thesis. 1951.

LILLE.

HALBBESCHRÄNKTE DIFFERENTIALOPERATOREN HÖHERER ORDNUNG

FRANZ RELLICH

1. *Der Begriff der Halbbeschränktheit.* Ein Hermitescher Operator A, erklärt in einem Teilraum $\mathfrak{A}$ eines Hilbertschen Raumes $\mathfrak{H}$ heisst *halbbeschränkt nach unten*, wenn es eine reelle Zahl γ gibt, mit der

$$(u, Au) \geqq \gamma(u, u)$$

gilt für alle u aus $\mathfrak{A}$; im Folgenden werden wir kürzer „halbbeschränkt" sagen an Stelle von „halbbeschränkt nach unten".

Interessiert man sich für das Eigenwertproblem $Au = \lambda u$ und setzt man ausserdem A in $\mathfrak{A}$ als selbstadjungiert (hypermaximal) voraus, so bedeutet Halbbeschränktheit, dass das Spektrum von A in $\mathfrak{A}$ nicht bis $-\infty$ reicht; kein Punkt des Spektrums liegt unterhalb von γ.

Wir betrachten das Eigenwertproblem

$$L(u) \equiv \sum_{\alpha=0}^{m} (-1)^{m-\alpha}\left(p_\alpha(x)u^{(m-\alpha)}\right)^{(m-\alpha)} = \lambda\, k(x)u, \quad a < x < b$$

In $a < x < b$ seien die Funktionen $k(x)$, $p_\alpha(x)$ reell und stetig, die $p_\alpha(x)$ seien mit stetigen Ableitungen bis zur Ordnung $m - \alpha$ einschließlich versehen und es gelte $k(x) > 0$, $p_0(x) > 0$. Mit dem inneren Produkt

$$(v, u) = \int_a^b \bar{v}\, u k(x)dx \text{ und } Du = \frac{1}{k(x)} L(u) \text{ wird}$$

$$(u, Du) = \int_a^b \left\{p_0(x) \mid u^{(m)} \mid^2 + p_1(x) \mid u^{(m-1)} \mid^2 + \ldots + p_m(x) \mid u \mid^2\right\} dx,$$

wenn wir als Definitionsbereich von D den Raum $\dot{\mathfrak{D}}$ aller in $a < x < b$ komplexwertigen Funktionen wählen, die dort stetige Ableitungen bis zur Ordnung $2m$ einschließlich besitzen und bei $x = a$ und bei $x = b$ identisch verschwinden. Halbbeschränktheit dieses Operators D in $\dot{\mathfrak{D}}$ heisst also genau, dass die untere Grenze des Quotienten

$$\int_a^b \{p_0(x)|u^{(m)}|^2 + p_1(x)|u^{(m-1)}|^2 + \ldots + p_m(x)|u|^2\}\, dx / \int_a^b |u|^2 k(x)dx$$

für alle $u \not\equiv 0$ in $\dot{\mathfrak{D}}$ endlich ist. Dementsprechend sind alle Operatoren, die aus einem sinnvollen Variationsproblem hergeleitet sind, halbbeschränkt. Bei den meisten Eigenwertproblemen, welche in den Anwendungen interessieren, liegen halbbeschränkte Operatoren vor.

Begrifflich ist die Halbbeschränktheit Hermitescher Operatoren zuerst durch A. Wintner formuliert worden, der in [9] auch die Bezeichnung „halbbeschränkt" zum ersten Mal verwendet. Später hat K. Friedrichs ([1] und [2]) die abstrakte Theorie der halbbeschränkten Operatoren weitergeführt. Er hat insbesondere erkannt, dass die Fortsetzung eines halbbeschränkten Operators zu einem selbstadjungierten Operator einfach möglich ist. Diese „Friedrichs'-sche Fortsetzung" ermöglichst es, die wichtige Frage nach den zulässigen bzw. notwendigen Randbedingungen rasch zu beantworten. In diesem Zusammenhang möchte ich folgende Frage stellen. Es sei $q(x, y, z)$ in $-\infty < x, y, z < \infty$ stetig und es sei $\mathfrak{A}$ der Raum aller in $-\infty < x, y, z < \infty$ zweimal stetig differenzierbaren Funktionen, die in einer individuellen Umgebung von ∞ identisch verschwinden. Es sei der Operator $Au = -\Delta u + q(x, y, z)u$ in $\mathfrak{A}$ mit dem inneren Produkt $(v, u) = \int_{-\infty}^{+\infty}\int\int \bar{v}\, u\, dx\, dy\, dz$ *halbbeschränkt.*

Frage: Ist A in $\mathfrak{A}$ bereits wesentlich selbstadjungiert, d.h. ist seine Abschließung selbstadjungiert? Meines Wissens ist diese Frage offen. Wenn die Antwort „ja" heisst, dann braucht man im Unendlichen keine zusätzlichen Randbedingungen zu formulieren um $-\Delta u + q(x, y, z)u = \lambda u$ zu einem sinnvollen Eigenwertproblem zu machen.

Die Randbedingungen spielen für die Halbbeschränktheit eines Differentialoperators eine wesentlich verschiedene Rolle je nachdem ob es sich um einen gewöhnlichen oder einen partiellen Differentialoperator handelt. Betrachten wir den oben definierten gewöhnlichen Differentialoperator D in $\dot{\mathfrak{D}}$, so wird er im allgemeinen nicht wesentlich selbstadjungiert sein. Der Raum $\dot{\mathfrak{D}}$ muss vielmehr durch einen umfassenden Raum $\mathfrak{D}$ ersetzt werden, in dem anstelle des identischen Verschwindens der Funktionen $u(x)$ an den Intervallenden $x = a$ und $x = b$ weniger einschneidende Randbedingungen gestellt werden, so zwar, dass nunmehr D in $\mathfrak{D}$ wesentlich selbstadjungiert ist.

Wenn D in $\mathfrak{D}$ halbbeschränkt ist, muss wegen $\dot{\mathfrak{D}} \subseteq \mathfrak{D}$ auch D in $\dot{\mathfrak{D}}$ halbbeschränkt sein. Es gilt aber auch das Umgekehrte: *Wenn D in $\dot{\mathfrak{D}}$ halbbeschränkt ist, dann muss D in $\mathfrak{D}$ halbbeschräkt sein.* Das folgt so: Weil es sich um einen

gewöhnlichen Differentialoperator handelt ist die Abschließung von D in $\mathfrak{D}$ ein hermitescher Operator mit endlichen Defektindices (die beide gleich sind, weil die Koeffizienten $p_\alpha(x)$, $k(x)$ reell sind). Die Halbbeschränktheit von D in $\mathfrak{D}$ ist also eine Folge eines Satzes van E. Heinz, der besagt (4):

Wenn ein halbbeschränkter (abgeschlossener) Operator gleiche und endliche Defektindices hat, dann ist jede selbstadjungierte Fortsetzung halbbeschränkt. Ob unser Differentialoperator halbbeschränkt ist oder nicht, hängt also nicht von den Randbedingungen ab (sondern nur von den Koeffizienten $p_\alpha(x)$, $k(x)$). Diese bedeutende Vereinfachung wird man nicht mehr für partielle Differentialoperatoren erwarten, weil diese im allgemeinen unendliche Defektindices besitzen und dann der Satz von Heinz nicht mehr anwendbar ist. Es ist aber interessant zu sehen, bei wie einfachen Randbedingungen die Halbbeschränktheit verloren gehen kann. Das lehrt folgendes von H. O. Cordes angegebene Beispiel:

Man betrachte das Eigenwertproblem $-\Delta u = \lambda u$ in $x^2 + y^2 \leqq 1$ mit der Randbedingung $\alpha u_r + \beta u = 0$ auf $x^2 + y^2 = r^2 = 1$; α und β sind reelle stetig differenzierbare Funktionen des Polarwinkels φ, die nirgends gleichzeitig verschwinden. Bezeichnet $\mathfrak{A}$ die Gesamtheit aller in $x^2 + y^2 \leqq 1$ zweimal stetig differenzierbaren Funktionen, die auf dem Rande des Einheitskreises der Beziehung $\alpha u_r + \beta u = 0$ genügen und bezeichnet $\mathring{\mathfrak{A}}$ den Teilraum, der aus allen Funktionen von $\mathfrak{A}$ besteht, die in der Umgebung des Randes $r = 1$ identisch verschwinden, so ist durch $Au = -\Delta u$ in $\mathring{\mathfrak{A}}$ ein halbbeschränkter Operator erklärt, wenn $(v, u) = \iint\limits_{x^2+y^2\leqq 1} \bar{v} u \, dx \, dy$ als inneres Produkt gewählt wird. Aber $Au = -\Delta u$ in $\mathfrak{A}$ ist nicht notwendig halbbeschränkt, wie H. O. Cordes durch geeignete Konstruktion von α und β zeigt.

2. *Halbbeschränktheit am linken bzw. rechten Intervallende.* Wenn der in Abschnitt 1 erklärte Operator D in $\mathring{\mathfrak{D}}$ nicht halbbeschränkt ist, dann kann dies durch das Verhalten der Koeffizienten $p_\alpha(x)$, $k(x)$ am linken Ende oder am rechten Ende verursacht sein. Daher empfiehlt sich folgender Begriff: Wenn $a < c < b$ ist, so bezeichnen wir die Menge aller $u(x)$ aus $\mathring{\mathfrak{D}}$, die in $c \leqq x < b$ $(a < x \leqq c)$ verschwinden mit $\mathring{\mathfrak{D}}_{ac}$ $(\mathring{\mathfrak{D}}_{cb})$. Wir nennen D in $\mathring{\mathfrak{D}}$ *halbbeschränkt am linken (rechten) Ende,* wenn es ein c mit $a < c < b$ gibt, so dass D in $\mathring{\mathfrak{D}}_{ac}$ $(\mathring{\mathfrak{D}}_{cb})$ halbbeschränkt ist. *Es ist D in $\mathring{\mathfrak{D}}$ dann und nur dann halbbeschränkt, wenn D in $\mathring{\mathfrak{D}}$ sowohl am linken Ende als auch am rechten Ende halbbeschränkt* ist (H. Krumhaar [5]).

3. *Hinreichendes Kriterium für Halbbeschränktheit von D in $\mathring{\mathfrak{D}}$ am linken Ende.* Wenn die Differentialgleichung von der zweiten Ordnung ist $(m = 1)$, dann gilt der Satz: *D in $\mathring{\mathfrak{D}}$ ist am linken Ende halbbeschränkt, wenn es eine reelle Zahl λ gibt, so dass die Differentialgleichung $L(u) = \lambda k(x) u$ eine reelle*

Lösung $u = y(x)$ besitzt, deren Nullstellen sich in $a < x < b$ nicht bei $x = a$ häufen. (Ph. Hartmann [3] sowie [6]).

Um diesen Satz auf den Fall $m > 1$ zu übertragen benötige ich den Begriff der „conjugierten Lösungen", den es für $m = 1$ nicht gibt. Er ist so erklärt. Aus $\int_{x_1}^{x_2} \{vL(u) - uL(v)\}\, dx = [v, u]_{x=x_2} - [v, u]_{x=x_1}$ mit

$$[v, u] = \sum_{\alpha=0}^{m-1} (-1)^{m-\alpha} \sum_{\beta=0}^{m-1-\alpha} (-1)^{\beta} \left(p_{\alpha}(x) u^{(m-\alpha)}\right)^{m-1-\alpha-\beta} v^{(\beta)}$$

$$- \sum_{\alpha=0}^{m-1} (-1)^{m-\alpha} \sum_{\beta=0}^{m-1-\alpha} (-1)^{\beta} \left(p_{\alpha}(x) v^{(m-\alpha)}\right)^{m-1-\alpha-\beta} u^{(\beta)}$$

folgt $[v, u] = \text{const.}$ in $a < x < b$ falls dort $\mathfrak{L}(u) = \lambda k u$ und $L(v) = \lambda k v$ ist.

Wir nennen nun zwei Lösungen $u = y_1$ und $u = y_2$ von $L(u) = \lambda k u$ *conjugiert*, wenn sie linear unabhängig sind und wenn $[y_2, y_1] = 0$ ist.

Für $m = 1$ ist $[v, u] = -p_0(x)(u'v - uv')$ immer von Null verschieden, sobald u, v zwei linear unabhängige Lösungen sind; also hat die Differentialgleichung zweiter Ordnung überhaupt keine conjugierten Lösungen.

Für $m > 1$ gibt es m (aber nicht mehr) linear unabhängige Lösungen, die paarweise conjugiert sind.

Es gilt (7) der Satz: *D in $\mathfrak{D}$ ist am linken Ende halbbeschränkt, wenn es eine reelle Zahl λ gibt, so dass die Differentialgleichung $L(u) = \lambda k u$ linear unabhängige reelle Lösungen $y_1, y_2, \ldots, y_m$ besitzt, die den beiden Forderungen genügen:*

1) *Die Nullstellen der Wronskischen Determinante $\Delta(x) = \det\left((y_k^{(i-1)}(x))\right)_{i,k=1,\ldots,m}$ haben in $a < x < b$ keinen Häufungspunkt bei $x = a$.*

2) *Wenn $m > 1$ ist, so sind die $y_1, \ldots, y_m$ paarweise conjugiert.*

Dass die zweite Forderung nicht entbehrlich ist, zeigt das Eigenwertproblem $u^{(4)} - \frac{9}{16}\frac{65}{16} x^{-4} u = \lambda u$, $0 < x < \infty$. Für $\lambda = 0$ hat diese Differentialgleichung die Lösungen

$$y_k = x^{\varrho_k} \text{ mit } \varrho_1 = 3 + \tfrac{1}{4},\ \varrho_2 = -\tfrac{1}{4},\ \varrho_3 = \tfrac{3}{2} + i\,\tfrac{3}{4},\ \varrho_4 = \tfrac{3}{2} - i\,\tfrac{3}{4}.$$

Die beiden Lösungen $y_1 = x^{3+\frac{1}{4}}$, $y_2 = x^{-\frac{1}{4}}$ sind linear unabhängig und reell und $\Delta(x) = y_1 y_2' - y_2 y_1' = -\frac{7}{2} x^2$ verschwindet nirgends in $0 < x < \infty$. Der Operator $Du = u^{(4)} - \frac{9}{16}\frac{65}{16} x^{-4} u$ im Raum $\mathfrak{D}$ aller in $0 < x < \infty$ viermal stetig differenzierbaren Funktionen, die bei $x = 0$ und $x = \infty$ identisch verschwinden, ist aber nicht halbbeschränkt. Sonst gäbe es eine Konstante γ mit der

$$(u, Du) = \int_0^\infty \left\{ |u''|^2 - \tfrac{9}{16}\tfrac{5}{16} x^{-4} |u|^2 \right\} dx \geqq \gamma \int_0^\infty |u|^2 dx$$

für alle u aus $\mathfrak{D}$ wäre. Das ist aber nicht der Fall, wie man leicht direkt einsieht oder dem Satz des nächsten Abschnittes entnehmen kann, da ϱ_3 und ϱ_4 einfache Wurzeln des zu $x = 0$ gehörigen charakteristischen Polynoms auf der kritischen Symmetriegeraden sind. — In der Tat sind die beiden Lösungen y_1 und y_2 nicht conjugiert, es ist vielmehr

$$[y_2, y_1] = y_1''' y_2 - y_1'' y_2' - y_1 y_2''' + y_1' y_2'' = (3 + \tfrac{1}{4})(2 + \tfrac{1}{4})(1 + \tfrac{1}{4}) + \\ + (3 + \tfrac{1}{4})(2 + \tfrac{1}{4})\tfrac{1}{4} + \tfrac{1}{4}(1 + \tfrac{1}{4})(2 + \tfrac{1}{4}) + (3 + \tfrac{1}{4})\tfrac{1}{4}(1 + \tfrac{1}{4})$$

von Null verschieden.

4. *Halbbeschränktheit an einer Stelle der Bestimmtheit.* Wir wählen das linke Intervallende $a = 0$ und machen über die Koeffizienten $p_\alpha(x)$, $k(x)$ der Differentialgleichung $L(u) = \lambda k u$ die zusätzliche Voraussetzung: Es sei $x = 0$ (identisch in λ) eine Stelle der Bestimmtheit d.h. es sei

$$p_\alpha(x) = \sum_{n=0}^\infty p_{\alpha n} x^{\sigma - 2\alpha + n}, \quad k(x) = \sum_{n=0}^\infty k_n x^{\sigma - 2m + n}$$

mit reellem σ in einer (rechtseitigen) Umgebung von $x = 0$. Das zu $x = 0$ gehörige charakteristische Polynom ist dann

$$P(\varrho) = \sum_{\alpha=0}^m p_{\alpha 0} \binom{\varrho}{m-\alpha} \binom{2s-\varrho}{m-\alpha} [(m-\alpha)!]^2 - \lambda k_0 \text{ mit } 2s = 2m - 1 - \sigma.$$

Seine Wurzeln liegen in der komplexen ϱ-Ebene symmetrisch zur reellen Achse (die Differentialgleichung hat reelle Koeffizienten), aber auch symmetrisch zur Geraden $\varrho = s + it$, $-\infty < t < \infty$, (Die Differentialgleichung ist selbstadjungiert); die zweite Symmetrie ist zuerst bei H. Krumhaar [5] verwendet worden. Ich konnte folgenden Satz beweisen: *Wenn $k_0 \neq 0$ ist, dann ist D in $\mathfrak{D}$ bei $x = 0$ halbbeschränkt. Wenn $k_0 = 0$ ist, dann ist für die Halbbeschränktheit von D in $\mathfrak{D}$ bei $x = 0$ notwendig und hinreichend, dass jede Wurzel des charakteristischen Polynoms $P(\varrho)$, die in der komplexen Ebene auf der kritischen Symmetriegeraden $\varrho = s + it$, $-\infty < t < \infty$ liegt, eine gerade Vielfachheit besitzt.*

Wir wenden diesen Satz an, um herauszufinden für welches grösste c die Ungleichung $c \int r^{-4} |u|^2 d\tau \leqq \int |\Delta u|^2 d\tau$ für alle u aus $\mathfrak{A}$ gelten könnte, wobei $\mathfrak{A}$ alle Funktionen $u(x_1, x_2, \ldots, x_h)$ umfasst, die im gesamten $x_1, x_2, \ldots, x_h$-Raum beliebig oft differenzierbar sind und in einer individuellen Umgebung des Ursprungs und von Unendlich identisch verschwinden; die Integrale sind über den Gesamtraum zu erstrecken, $r = (x_1^2 + x_2^2 + \ldots + x_h^2)^{1/2}$. Eine solche Ungleichung müsste erst recht in den Teilraum von $\mathfrak{A}$ gelten,

der aus allen Funktionen besteht, die nur von r abhängen. Dann würde es sich darum handeln zu wissen, für welches grösste c die Ungleichung

$$I = \int_0^\infty \left\{ | u'' + \frac{h-1}{r} u' |^2 - cr^{-4} | u |^2 \right\} r^{h-1} dr \geqq 0$$

besteht, die wegen $u \equiv 0$ bei $r = 0$ und $r = \infty$ mit

$$I = \int_0^\infty \left\{ r^{h-1} | u'' |^2 + (h-1) r^{h-3} | u' |^2 - cr^{h-5} | u |^2 \right\} dr \geqq 0$$

gleichwertig ist. Für unsere Frage macht es nichts aus, ob die zugelassenen Funktionen unendlich oft oder viermal stetig differenzierbar sind. Deshalb nennen wir, in Übereinstimmung mit unserer früheren Bezeichnungsweise, $\dot{\mathfrak{D}}$ die Gesamtheit aller in $0 < r < \infty$ viermal stetig differenzierbaren Funktionen, die bei $r = 0$ und $r = \infty$ identisch verschwinden. Wir fragen: Für welche c ist der Operator

$$Du = (r^{h-1} u'')'' - (h-1)(r^{h-3} u')' - cr^{h-5} u$$

in $\dot{\mathfrak{D}}$ mit dem inneren Produkt $(v, u) = \int_0^\infty \bar{v} u r^{h-1} dr$ nicht mehr halbbeschränkt?

Das charakteristische Polynom der Differentialgleichung

$$(r^{h-1} u'')'' - (h-1)(r^{h-3} u')' - cr^{h-5} u = \lambda r^{h-1} u$$

ist $P(\varrho) = \varrho(\varrho - 2)(\varrho + h - 4)(\varrho + h - 2) - c$. Auf der Symmetriegeraden $\varrho = s + it = 2 - \frac{h}{2} + it$ (nach obiger Bezeichnung ist $m = 2$, $\sigma = h - 1$,

$$p_0(r) = r^{h-1},\ p_1(r) = (h-1) r^{h-3},\ p_2(r) = -cr^{h-5},\ k(r) = r^{h-1})$$

wird

$$P(\varrho) = \left[t^2 + \left(2 - \frac{h}{2}\right)^2 \right] \left(t^2 + \frac{h^2}{4} \right) - c \geqq \frac{h^2 (h-4)^2}{16} - c$$

und die Schranke rechts wird für $\varrho = 2 - \frac{h}{2}$ erreicht. Wenn also $c \leqq \frac{h^2(h-4)^2}{16}$ ist, dann hat das charakteristische Polynom $P(\varrho)$ auf der kritischen Symmetriegeraden keine Nullstelle ungerader Vielfachheit, es ist D in $\dot{\mathfrak{D}}$ halbbeschränkt. Wenn aber $c > \frac{h^2(h-4)^2}{16}$ ist, dann nimmt $P(\varrho)$ auf der kritischen Symmetriegeraden negative Werte an, also muss $P(\varrho)$ auf dieser Geraden mindestens

eine Nullstelle ungerader Vielfachheit haben, also ist D in $\mathfrak{D}$ nicht halbbeschränkt. Es kann also für $c > \frac{h^2(h-4)^2}{16}$ nicht $I \geqq 0$ sein (was ja sogar Halbbeschränktheit von D in $\mathfrak{D}$ mit der unteren Schranke $\gamma = 0$ bedeuten würde). Erst recht kann für solche c nicht die obige in $\dot{\mathfrak{A}}$ formulierte Ungleichung gelten.

Zu beantworten bleibt, ob $\frac{h^2(h-4)^2}{16} \int r^{-4} \mid u \mid^2 d\tau \leqq \int \mid \Delta u \mid^2 d\tau$ richtig ist für alle u aus $\dot{\mathfrak{A}}$. Dazu habe ich ([8] Seite 121—134) bewiesen: 1. *Es sei* $u(x_1, \ldots, x_h) \not\equiv 0$ *eine Funktion aus* $\dot{\mathfrak{A}}$ *und es sei im Falle* $h = 2$ *ausserdem* $\int_0^{2\pi} u(r\cos\varphi, r\sin\varphi)\cos\varphi\, d\varphi = 0$, $\int_0^{2\pi} u(r\cos\varphi, r\sin\varphi)\sin\varphi d\varphi = 0$ *für* $0 < r < \infty$. *Dann ist*

$$\frac{h^2(h-4)^2}{16} \int_{0<r<\infty} r^{-4} \mid u \mid^2 d\tau < \int_{0<r<\infty} \mid \Delta u \mid^2 d\tau$$

2. *Es sei* a *eine beliebige reelle Zahl,* R *eine beliebige positive Zahl und* $c > 0$ *für* $h = 2$ *aber* $c > \frac{h^2(h-4)^2}{16}$ *für* $h = 1, 3, 4, \ldots$ *Dann gibt es zu diesen Zahlen* a, c, R *eine Funktion* u_0 *aus* $\dot{\mathfrak{A}}$, *die identisch verschwindet in* $r > R$ *und eine Funktion* u_∞ *aus* $\dot{\mathfrak{A}}$, *die identisch verschwindet in* $r < R$ *mit*

$$c \int_{0<r<\infty} r^{-4} \mid u_0 \mid^2 d\tau > \int_{0<r<\infty} \mid \Delta u_0 \mid^2 d\tau + a \int_{0<r<\infty} \mid u_0 \mid^2 d\tau$$

$$c \int_{0<r<\infty} r^{-4} \mid u_\infty \mid^2 d\tau > \int_{0<r<\infty} \mid \Delta u_\infty \mid^2 d\tau.$$

[1] K. Friedrichs, Spektraltheorie halbbeschränkter Operatoren I. und II. Math. Ann. Bd. 109 (1934) 465—487, 685—713 und Bd. 110 (1935) 777—779.

[2] K. Friedrichs, Über die ausgezeichnete Randbedingung in der Spektraltheorie der halbbeschränkten gewöhnlichen Differentialoperatoren zweiter Ordnung. Math. Ann. Bd. 112 (1936) 1—23.

[3] Ph. Hartmann, Differential equatons with non-oscillatory eigenfunctions. Duke math. J. 15 (1948) 697—709.

[4] E. Heinz, Zur Theorie der Hermiteschen Operatoren des Hilbertschen Raumes. Nachrichten d. Akad. d. Wissenschaften Göttingen, 1951, 1—4.

[5] H. Krumhaar, Zur Theorie der gewöhnlichen selbstadjungierten Differentialoperatoren gerader Ordnung. Math. Ann. erscheint demnächst.

[6] F. RELLICH, Halbbeschränkte gewöhnliche Differentialoperatoren zweiter Ordnung. Math. Ann. 122 (1951), 343—368.

[7] F. RELLICH, Halbbeschränktheit gewöhnlicher Differentialoperatoren höherer Ordnung. Math. Ann. erscheint demnächst.

[8] F. RELLICH, Perturbation theory of eigenvalue problems. Institute of Mathematical Sciences New York University 1953.

[9] A. WINTNER, Spektraltheorie der unendlichen Matrizen, Leipzig 1929.

GÖTTINGEN, BUNSENSTRASZE 3/5.

SOME RECENT PROGRESS IN THE THEORY OF SURFACE WAVES IN WATER

By J. J. STOKER

1. *Introduction.*

In talks of this kind at International Congresses, the common expectation is that a survey of the literature in a restricted field will be given. That is not the speaker's intention on this occasion, principally because such a survey would, because of the large amount of work that has been done recently, necessarily be a dull catalogue of many items, no one of which could be adequately described because of time limitations. Consequently, the speaker has decided to confine the discussion to four different problems covering various aspects of the field, with the hope that they can be made understandable and interesting. In addition, the four problems chosen are all cases dealt with by the speaker and his colleagues and friends during the past few years.

The problems to be discussed belong to the classical theory of the hydrodynamics of non-viscous, incompressible fluids. In addition, the flows (in general, non-steady flows) are assumed to be irrotational. Thus a velocity potential $\Phi(x, y, z; t)$ exists, and the velocity vector (u, v, w) is given by

$$(u, v, w) = \operatorname{grad} \Phi. \tag{1.1}$$

In the region R occupied by the fluid, Φ of course satisfies the Laplace equation

$$\Phi_{xx} + \Phi_{yy} + \Phi_{zz} = 0. \tag{1.2}$$

The region R is in general bounded above by a free surface:

$$y = \eta(x, z; t) \tag{1.3}$$

as indicated in Fig. 1.1, and below by a rigid bottom surface

$$y = -h(x, z). \tag{1.4}$$

The pressure $p(x, y, z; t)$ in the fluid is determined by Bernoulli's law

$$\Phi_t + \frac{1}{2}\left(\Phi_x^2 + \Phi_y^2 + \Phi_z^2\right) + \frac{p}{\varrho} + gy = 0. \tag{1.5}$$

---

[1]) All of this work, except the work on flood waves in rivers in section 5, has had the support and cooperation of the mathematics ans mechanics branches of the Office of Naval Research of the U. S. Navy.

The peculiarity of surface wave problems which makes them difficult, and interesting, is the existence of a free surface on which the pressure is, say, zero; this leads to the nonlinear boundary condition

$$\Phi_t + \frac{1}{2}\left(\Phi_x^2 + \Phi_y^2 + \Phi_z^2\right) + \frac{p}{\varrho} + g\eta = 0 \text{ at } y = \eta, \tag{1.6}$$

and to a kinematic condition

$$\Phi_x \eta_x - \Phi_y + \Phi_z \eta_z + \eta_t = 0. \tag{1.7}$$

At the rigid bottom surface one has, of course, the condition

$$\Phi_n = 0 \text{ at } y = -h. \tag{1.8}$$

A completely formulated mathematical problem, for the unique determination of the velocity potential $\Phi(x, y, z; t)$ and the surface elevation $\eta(x, z; t)$ would require, in addition to the differential equation (1.2), and the boundary conditions (1.6), (1.7), (1.8), also the imposition of initial conditions at the time $t = 0$ and conditions at ∞. Obviously, such nonlinear boundary problems

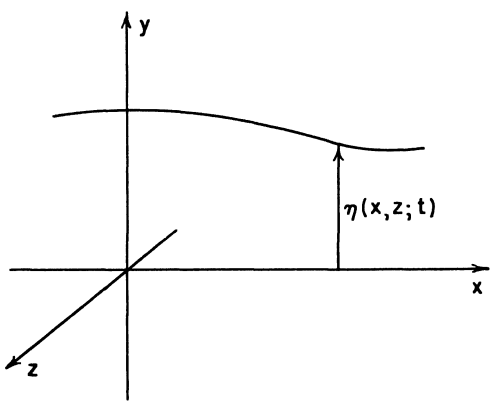


Fig. 1.1

would be very difficult to solve, and consequently only a little progress has been made in applying the exact theory. However, some notable results have been obtained as, for example, the proof of the existence of steady progressing waves of small amplitude by Levi-Civita [10][1] and Struik [20], and of the existence of the so-called solitary wave by Lavrentiev [9] and by Friedrichs and Hyers [4].

Because of the difficulty of the exact theory, it is not strange that most of the literature dealing with water waves is concerned with two simpler approximate theories which result from physical assumptions which lead to the

[1]) Numbers in square brackets refer to the bibliography at the end of this address.

presence of small parameters of two different kinds in the mathematical formulation. One then develops the solution of the basic problem with respect to such a parameter, usually stopping with the lowest order terms. The remainder of this talk is concerned with problems involving these two theories which I proceed to formulate briefly. Both of these theories seem to have been derived by Lagrange, but they are commonly attributed to others.

The first of the two approximate theories is a linear theory based on the assumption that the surface waves have small amplitudes — in fact, the theory is a theory of small oscillations of the water about an equilibrium position (which may be either the state of rest or a uniform flow); this theory is usually given the names of Cauchy and Poisson, who worked out the first significant consequences of it. The great simplifications which this theory brings about result from the linearization of the free surface conditions (1.6) and (1.7) which comes about on rejecting quadratic terms in Φ, η, and their derivatives in comparison with linear terms; at the same time it is correct and consistent to satisfy the free surface conditions at the undisturbed equilibrium position $y = 0$, rather than on the unknown free surface $y = \eta$. One has therefore to determine a harmonic function $\Phi(x, y, z; t)$ in the space between the plane $y = 0$ and any given boundary surfaces, which satisfies the free surface conditions

$$\left\{\begin{array}{l} \Phi_t + g\eta + \dfrac{p}{\varrho} = 0 \\ \qquad\qquad\qquad\qquad \text{at } y = 0. \\ \eta_t - \Phi_y = 0 \end{array}\right. \tag{1.9}$$

or, upon elimination of η through differentiation

$$\Phi_{tt} + g\Phi_y + \frac{p_t}{\varrho} = 0 \text{ at } y = 0. \tag{1.10}$$

It is this boundary condition of mixed type — involving the normal derivative Φ_y as well as Φ_{tt} — which gives this theory, which otherwise belongs to the classical theory of boundary problems for harmonic functions, its peculiarities and mathematical interest. In the following sections, two applications of this type of theory will be discussed.

The second of the approximate theories commonly employed in studying water waves is based on the assumption that the *depth* of the water is small in an appropriate sense, rather than the amplitude of the waves. Such theories are commonly referred to as shallow water theories, or as long wave theories. Various derivations of the theory can be given; the one devised by Friedrichs [3] is explained briefly here. Two typical lengths, d and k, one in the depth, the

other in the horizontal direction are introduced. An essential feature of the derivation of the theory by Friedrichs is the introduction of dimensionless variables in the space coordinates in such a way that the horizontal and the vertical coordinates are stretched in different ways; thus one sets

$$\bar{x} = {}^{x}/_{k},\ \bar{y} = {}^{y}/_{d},\ \bar{z} = {}^{z}/_{k}. \tag{1.11}$$

The vertical and horizontal velocity components are also treated differently:

$$\bar{u} = (gd)^{-\frac{1}{2}}u,\ \bar{v} = \frac{k(gd)^{-\frac{1}{2}}}{h}v,\ \bar{w} = (gd)^{-\frac{1}{2}}w. \tag{1.12}$$

Finally, the dimensionless time τ, and other quantities are introduced as follows:

$$\tau = \frac{\sqrt{gd}}{k}t,\ \bar{p} = \frac{1}{\varrho gd}p,\ \bar{\eta} = {}^{\eta}/_{d},\ \bar{h} = {}^{h}/_{d}. \tag{1.13}$$

Once the new quantities have been introduced, it turns out that the lengths d and k occur in all of the basic equations only in the combination ${}^{d^2}/_{k^2}$; hence one introduces the essential parameter σ by the formula:

$$\sigma = {}^{d^2}/_{k^2}. \tag{1.14}$$

Typical equations which result are the continuity equation

$$\sigma\bar{u}_{\bar{x}} + \bar{v}_{\bar{y}} + \sigma\bar{w}_{\bar{z}} = 0, \tag{1.15}$$

and the Euler equation of motion for the x-direction, for example:

$$\sigma[\bar{u}_{\tau} + \overline{u}\overline{u}_{\bar{x}} + \overline{u}\overline{u}_{\bar{z}} + \overline{P}_{\bar{x}}] + \overline{v}\overline{u}_{\bar{y}} = 0. \tag{1.16}$$

The shallow water theory is now obtained by developing all dependent quantities in powers of σ, which is of course assumed to be small and thus, in view of (1.14), representative of a configuration in which the water is shallow:

$$\begin{cases} \bar{u} = \bar{u}^{(0)} + \sigma\bar{u}^{(1)} + \dots \\ \bar{v} = \bar{v}^{(0)} + \sigma\bar{v}^{(1)} + \dots \\ \bar{\eta} = \bar{\eta}^{(0)} + \sigma\bar{\eta}^{(1)} + \dots \\ - - - - - - - \\ - - - - - - - - \end{cases} \tag{1.17}$$

These developments are inserted in all equations of the problem, and conditions for the coefficients of the perturbation series are obtained. The zero order terms yield, among others, the following results:

$$\begin{cases} v^{(0)} = 0 \\ p^{(0)} = \eta^{(0)} - y \\ u^{(0)} = u^{(0)}(x, z; t), \end{cases} \tag{1.18}$$

which state that the vertical component of the velocity is identically zero, that the pressure is given by the hydrostatic pressure law, and that the horizontal velocity component is independent of the vertical coordinate. The usual derivations of the shallow water theory begin with one or the other of the latter two conditions on $p^{(0)}$ and $u^{(0)}$ as basic assumptions. In two dimensional problems one then finds next as equations for the horizontal velocity component and the free surface displacement the equations

$$\begin{cases} u_t + uu_x = -g\eta_x \\ \eta_t + [u(\eta + h)]_x = 0 \end{cases} \tag{1.19}$$

on reverting to the original variables, dropping superscripts, and assuming that the bottom surface is horizontal. It is this pair of equations which constitutes the basic formulation of the long wave — shallow water theory. As one sees, it is a nonlinear theory. The equations are of hyperbolic type, and, in fact, they are exactly analogous to the differential equations for a one-dimensional unsteady flow of a compressible gas which obeys an adiabatic law in which the adiabatic exponent has the fixed value two (cf. Riabouchinsky [17]). Of course, initial and boundary conditions must be given in connection with (1.19) in order that a problem with a unique solution is formulated.

The fourth and fifth sections below will describe two applications of this theory.

2. *Diffraction of waves by a vertical plane barrier.*

(Sommerfeld's diffraction problem.)

In water of infinite depth, the diffraction of simple harmonic plane waves by the vertical rigid plane barrier $x > 0$, $y < 0$, $z = 0$ is to be studied in terms of the theory of waves of small amplitude outlined above. We set

$$\Phi(x, y, z; t) = e^{my}e^{i\sigma t} f(x, z), \tag{2.1}$$

and observe that the free surface condition at $y = 0$ given by (1.10) will be satisfield for $p = 0$ on the free surface if

$$\sigma^2 = gm. \tag{2.2}$$

The function $f(x, z)$ satisfies the reduced wave equation

$$\nabla^2 f + m^2 f = 0, \tag{2.3}$$

but for simplicity we set $m = 1$ in the following. The boundary condition on the rigid barrier is satisfied by requiring

$$f_\theta = 0 \text{ for } \theta = 0 \text{ and } 2\pi, \tag{2.4}$$

in polar coordinates (ϱ, θ), as shown in Fig. 2.1. The condition at ∞ for f results

from the assumption of an incoming plane wave; we prescribe that

$$(2.5)\quad f \sim g \equiv \begin{cases} e^{i\varrho\cos(\theta-\alpha)} + e^{i\varrho\cos(\theta+\alpha)}, & 0 < \theta < \pi - \alpha \\ e^{i\varrho\cos(\theta-\alpha)}, & \pi - \alpha < \theta < \pi + \alpha \\ 0, & \pi + \alpha < \theta < 2\pi, \end{cases}$$

i.e. that at ∞ f behaves in a discontinuous way corresponding to a division of the plane into regions in which a) the incoming wave and its reflection from the barrier coexist, b) the region of the transmitted wave past the edge of the barrier, and c) the shadow region. A. S. Peters and the speaker [13] have proved the following uniqueness theorem (actually, the theorem is more general in that it allows barriers of a more general character):[1]) *A solution* $f(\varrho, \theta)$ *of* (2.3) *which can be written in the form* $f = g + h$, *with* g *defined as in* (2.5), h *a function satisfying the Sommerfeld radiation condition*

$$(2.6)\qquad \lim_{\varrho\to\infty} \varrho^{\frac{1}{2}}\left(\frac{\partial h}{\partial \varrho} + ih\right) = 0 \text{ (uniformly in } \theta),$$

and with f *bounded at the origin while* $f_\varrho \sim \frac{c}{\varrho^k}$, $k < 1$, *at the origin, is uniquely determined.*

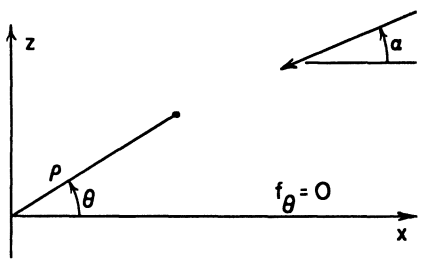


Fig. 2.1

Many solutions of this classical old problem have been given, starting with Sommerfeld himself [19] in 1896, but these solutions could hardly be called elementary since they require either very special insight into special solutions of the wave equation, or techniques such as that of the Wiener-Hopf integral equation. However, once it is realized that the decomposition of the solution into a sum of two discontinuous functions

$$(2.7)\qquad f = g + h,$$

with the properties stated in the uniqueness theorem is feasible, an elementary

[1]) Rellich [16] has previously given uniqueness theorems for solutions of the reduced wave equation which employ devices similar to the ones used in [13].

method of solution suggests itself. One simply writes the solution $f(\varrho, \theta)$ as a Fourier series:

$$f(\varrho, \theta) = \Sigma a_{\frac{n}{2}} J_{\frac{n}{2}}(\varrho) \cos \frac{n\theta}{2} = \Sigma \alpha_{\frac{n}{2}}(\varrho) \cos \frac{n\theta}{2} + h, \tag{2.8}$$

with $\alpha_{\frac{n}{2}}$ the Fourier coefficients of the function g. The operator $\varrho^{\frac{1}{2}}\left(\frac{\partial}{\partial \varrho} + i\right)$, which occurs in the radiation condition, is applied to the equation, followed by a passage to the limit $\varrho \to \infty$. Since $J_{\frac{n}{2}}(\varrho)$ and its derivatives behave like $1/\sqrt{\varrho}$, and since $\alpha_{\frac{n}{2}}(\varrho)$ is also easily seen to behave in the same way (by use of the principle of stationary phase, for example), and since the operator applied to h vanishes when $\varrho \to \infty$, it follows that the Fourier coefficients $a_{\frac{n}{2}}$ of f are determined. In fact, they are determined very easily. Afterwards, it is also not difficult to sum the series in such a way as to obtain the well-known integral representations of the solution. For the details, and for references to other literature, the paper [13] cited above should be consulted.

3. *The Motion of Ships as Floating Rigid Bodies.*

A general theory for the motion of ships has recently been derived by A. S. Peters and the speaker [12] in which the only assumptions made are those necessary to linearize the problem, together with the assumption of a perfect fluid in irrotational motion. The theory is a generalization of the classical theory of Michell [11] and Havelock [6], which assumes the ship to be held rigidly at rest in a stream, and leads to formulas for the wave resistance in its dependence on the shape of the hull of the ship. An extensive literature dealing with applications of this theory exists; for a bibliography, the paper of Weinblum [23] is recommended.

The problem of the motion of a ship as a freely floating rigid body has, of course, also been considered by others. One approach which has been employed has been to use a combined theoretical and empirical formulation which involves writing down equations of motion of the ship with coefficients which should in part be determined by experiment; it is assumed in addition that there is no coupling between the different degrees of freedom involved in the general motion of the ship; for examples of this type of theory the papers by St. Denis and Weinblum [18], and by Pierson and St. Denis [14] should be consulted. Haskind [5] attacks the problem in the same degree of generality as Peters and the speaker, but derives his theory completely in the end only in a certain special case; Haskind's theory is obtained by a procedure different from that used in [12], also, and leads to different basic equations.

The theory I wish to outline assumes that the hull of the ship is given by an equation of the form (cf. Fig. 3.1):

$$z' = \pm \beta h(x', y'), \quad z' \lesseqgtr 0, \tag{3.1}$$

with respect to coordinates rigidly fixed to the hull; in this equation β is a dimensionless parameter, of the nature of a typical breadth-length ratio, which is assumed to be small. The theory is derived by developing systematically all quantities in powers of this parameter; the resulting perturbation theory is a linear theory. For example, one writes

$$\begin{cases} \Phi(x, y, z; t) = \beta\Phi_1(x, y, z, t) + \beta^2\Phi_2 + \dots \\ \eta(x, z; t) = \beta\eta_1(x, z, t) + \beta^2\eta_2 + \dots \\ \theta_i(t; \beta) = \beta\theta_{i1}(t) + \beta^2\theta_{i2} + \dots, \; i = 1, 2, 3 \end{cases} \tag{3.2}$$

for the velocity potential, the free surface elevation, and the angular displacements of the ship, all relative to a moving coordinate system (x, y, z) so chosen that the x, z-plane lies in the undisturbed free surface, and the y-axis contains

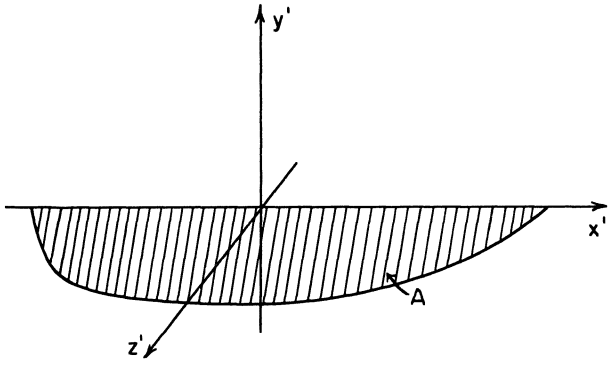


Fig. 3.1

the c.g. of the ship. As boundary condition for the ship's hull the following is obtained, to first order in β:

$$\Phi_{1z} = \pm s_0 h_x - s_0\theta_{21} + (w_1 + \dot{\theta}_{21})x - \dot{\theta}_{11}(y - y_c') \text{ on } A\pm. \tag{3.3}$$

This condition is to be satisfied on the vertical for-and-aft midsection of the ship, rather than on the curved hull — a consequence of the assumption of a slender ship inherent in assuming the parameter β to be small. The quantity s_0 is the speed of the ship in its course. As one sees, some of the unknown time-dependent displacements of the ship's hull occur in this boundary condition. In addition to (3.3) there are two free surface conditions:

$$(3.4)\qquad \begin{cases} g\eta_1 + s_0\Phi_{1x} - \Phi_{1t} = 0 \\ \\ \Phi_{1y} - s_0\eta_{1x} + \eta_{1t} = 0 \end{cases} \quad \text{at } y = 0.$$

Finally, it is necessary to develop the equations of motion of the ship, assumed to have oscillations of small amplitude relative to a uniform forward speed in the course; it is necessary to consider the terms of zero, first, and second orders in β. Those of zero order yield Archimedes' law, for instance. The terms of first order furnish conditions on those displacements which, like the rolling displacement θ_1, are displacements out of the vertical plane, involving the pressure on the ship's hull determined by Φ_1, and which figure in the boundary condition (3.3). The terms of second order yield ordinary differential equations of second order for the displacements in the vertical plane, i.e. the heaving, pitching, and surging oscillations; the ship resistance is also determined, as well as the so-called trim of the ship. For the heaving displacement $y_1(t)$, for example, the result is:

$$(3.5)\qquad M_1y_1=k_1y_1+k_2\theta_{31}+\varrho\int_A h_y\{[\Phi_{1t}-s_0\Phi_{1x}]^+ + [\Phi_{1t}-s_0\Phi_{1x}]^-\}dA$$

Here k_1 and k_2 are constants fixed by the geometry of the hull. As one sees, there is coupling between the heaving and pitching oscillations, and also with others through Φ_1 which in its turn depends on the roll, yaw, and sway. All of these relations, together with initial conditions, the specification of the propeller thrust as a force of order β^2, and conditions at ∞ describing the character of the seaway, lead to a completely formulated problem which doubtlessly would have a unique solution. However, linear though the problem is, it permits no explicit solutions except for the case of motions of the ship confined to the vertical plane in a sea-way with crests at right-angles to the ship's course; in all other cases an integral equation of unconventional type for Φ must be solved before the motion of the ship can be determined.

Some interesting conclusions from the theory are possible without solving concrete problems. For example, the motion of the water is influenced, up to terms of first order in β, only by the displacements out of the vertical plane (i.e. the roll, yaw, and sway) and not at all by the three parallel to the vertical plane (the heave, pitch, and surge). A little reflection makes this result quite plausible: our theory is based on the assumption that the ship is a thin disk, with a thickness of order β, disposed vertically in the water; thus displacements in this plane of order β cannot create waves of order β, but waves of order β^2 or higher. On the other hand, first order displacements of the disk at right angles to itself create waves also of first order. The situation is describable also in the following way: a knife blade placed vertically in the water creates no

waves of noticeable magnitude unless displacements perpendicular to the plane of the blade are produced. This fact has an interesting consequence: strict resonances are possible in heaving and pitching, since these modes create no waves to first order and hence are not subject to damping due to energy being carried off by progressing waves to infinity.

4. *The Solitary Wave.*

The famous old problem of the solitary wave has been solved recently in a rigorous way by Lavrentiev [9] and by Friedrichs and Hyers [4]. The problem is to find a free surface wave which moves with uniform velocity and without change of form; a wave periodic in the direction of travel is not sought, but rather a wave which dies out at ∞, as indicated in Fig. 4.1.

Approximate treatments of the problem leading to formulas for the shape, propagation speed, etc. were devised long ago by Boussinesq [2] and Rayleigh [15], among others; these treatments are all, significantly enough, essentially nonlinear in character. The speaker thought for some years that the rigorous existence of the solitary wave was doubtful for the following reasons. Such waves with arbitrarily small amplitudes exist, hence one should be able to

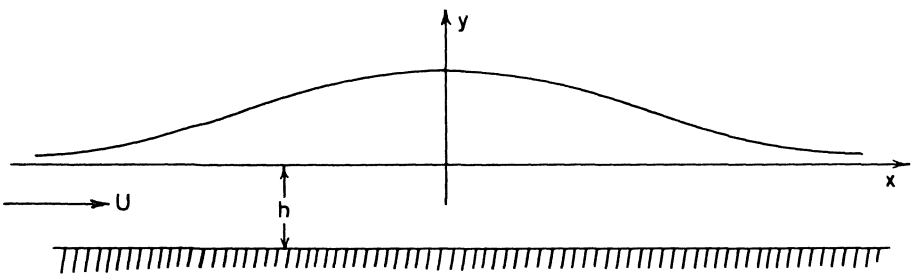


Fig. 4.1

determine them by a development with respect to amplitude in the manner of Levi-Civita; however, the coefficients of the perturbation series can easily be seen to vanish because of the assumption that the wave amplitude dies cut at ∞, by using a uniqueness theorem due to Weinstein [24]. However, the fact is that the solitary wave is not analytic in its amplitude in the neighborhood of the zero value.

Lavrentiev apparently attacks the problem by taking, roughly speaking, the solutions of Levi-Civita for periodic waves and making an appropriate passage to the limit corresponding to permitting the wave length to become infinite; this treatment seems to be difficult. Friedrichs and Hyers solve the problem by making use of the nonlinear shallow water theory sketched out in

the first section. A formal solution of the problem along these lines had already been worked out by Keller [8] (who attained essentially the same results as earlier writers) by carrying out the perturbation series of the shallow water theory to terms of second order. To first order, the shallow water theory permits only a "trivial" wave motion which progresses unchanged in form, i.e. the uniform flow with undisplaced free surface. However, this manifold of exact solutions of the problem has a bifurcation when the velocity U and the depth h are such that the velocity has the so-called critical value $U = \sqrt{gh}$, and the nonlinear shallow water theory when carried to second order yields non-trivial waves of the desired type when the parameter U^2/gh is near to unity (but always greater than unity). Friedrichs and Hyers prove rigorously that the solitary wave exists by showing that iterations with respect to the parameter $\varepsilon = U^2/gh - 1$ converge for small values of ε to a solution of a nonlinear integral equation which is derived from the shallow water theory.

The fact of the existence of the solitary wave has been considered a paradox by Birkhoff [1] and by Ursell [22], because of the fact that the shallow water theory to lowest order gives the result that every wave, no matter what its form, would always change its shape and eventually break. On the other hand, Levi-Civita has proved the existence of periodic progressing waves, and the solitary wave has long been observed in nature. This seeming paradox has already been resolved above: the solitary wave is a second order phenomenon in terms of the shallow water theory which results by bifurcation from a definite uniform flow as solution of the problem to first order.

5. *River Regulation and Flood Prediction Problems.*

In these problems the basic idea is to use the nonlinear partial differential equations of the shallow water theory, with appropriate changes to account for the variable cross section areas of a river valley, and with the addition of an important term furnishing a resistance to the flow due to roughness of the river bed, to study flows in rivers. The problem of carrying out flood predictions by using a digital computer (in this case the UNIVAC) has been studied at the Institute of Mathematical Sciences under a contract with the Corps of Engineers of the U. S. Army. Two reports [21], and [7], have been issued on the subject. Thirty-six hour flood predictions have been made for 400 miles of the Ohio River, and checked against the observations for the case of the big flood of 1945. The method seems to work well, and hence has many implications for the future, e.g. in relation to the building of expensive models. Studies of flood prediction at the junction of the Ohio and Mississippi Rivers, and of regulation problems at the Kentucky dam at the mouth of the Tennessee River are also under way.

BIBLIOGRAPHY

[1] BIRKHOFF, G. Hydrodynamics, Princeton University Press (1950).

[2] BOUSSINESQ, J. Essai sur la théorie des eaux courantes. Mémoires presentés par divers savants, Academie des Sciences de l'Institute de France, Volume 23, 1877.

[3] FRIEDRICHS, K. O. On the Derivation of the Shallow Water Theory. Appendix to The Formation of Breakers and Bores by J. J. Stoker, Communications on Pure and Applied Mathematics, New York University, Volume 1 No. 1, January 1948.

[4] FRIEDRICHS, K. O. and D. H. HYERS. The Existence of Solitary Waves, Communications on Pure and Applied Mathematics, Volume 7, 1954.

[5] HASKIND, M. D. The Hydrodynamic Theory of the Oscillation of a ship in Waves (in Russian), Prik. Mat. i. Mek., Volume 10, no. 1, 1946.

[6] HAVELOCK, T. H. The Wave-making Resistance of Ships, Proc. Roy. Soc. London (A), 81, 1909.

[7] ISAACSON, E., J. J. STOKER and A. TROESCH. Numerical Solution of Flood Problems in Simplified Models of the Ohio River and the Junction of the Ohio and Mississippi Rivers. A report submitted to the Ohio River Division of the U. S. Army Corps of Engineers, 1954. IMM-NYU 205.

[8] KELLER, J. B. The Solitary Wave and Periodic Waves in Shallow, Water, Communications on Pure and Applied Math., Volume 1, no. 4, pp. 323-339, December 1948.

[9] LAVRENTIEV, M. On the Theory of Long Waves, Akad. Nauk Ukrain. R.S.R., zb. Proc. Inst. Math., Volume 1, 1948 (in Ukranian).

[10] LEVI-CIVITA, T. Détermination rigoureuse des ondes permanentes d'ampleur finie, Mathematische Annalen, Volume 93, pp. 264-314, 1925.

[11] MICHELL, J. H. The Wave Resistance of a Ship, Phil. Mag., Volume 45, 1898.

[12] PETERS, A. S. and J. J. STOKER. The Motion of a Ship, as a Floating Rigid Body, in a Seaway, Report by Institute of Mathematical Sciences, New York University, IMM-NYU 203, 1954.

[13] PETERS, A. S. and J. J. STOKER. A Uniqueness Theorem and a New Solutions for Sommerfeld's and Other Diffraction Problems, Communications on Pure and Applied Mathematics, Volume VII, No. 3, August 1954, p. 565.

[14] PIERSON, W. J. and M. ST. DENIS. On the Motion of Ships in Confused Seas, Soc. Nav. Arch. and Mar. Eng., 1953.

[15] RAYLEIGH, LORD. On the Theory of Long Waves and Bores, Proc. of the Royal Society, Series A, 1914.

[16] RELLICH, F. Über des asymptotische Verhalten der Lösungen von $\triangle^2 u + u = 0$, in unendlichen Gebieten, Jahresbericht der deutschen Mathematiker-Vereinigung, Volume 53, 1943.

[17] RIABOUCHINSKY, D. Sur l'analogie hydraulique des mouvements d'un fluide comm. pressible, Comptes Rendus Academie des Sciences, Paris, Tome 195, 1932, p. 998.

[18] ST. DENIS, M. and G. WEINBLUM. On the Motions of Ships at Sea, Soc. Nav. Arch, and Mar. Eng., 1905.

[19] SOMMERFELD, A. Mathematische Theorie der Diffraktion, Math. Annalen, Bd. 47 p. 317, 1896.

[20] STRUIK, D. J. Détermination rigoureuse des ondes irrotationelles périodiques dans un canal à profondeur finie, Mathematische Annalen, Volume 95, pp. 595-634, 1926.

[21] STOKER, J. J. Numerical solution of flood prediction and river regulation problems I: Derivation of basic theory and formulation of memerical methods of attach. A report submitted to the Ohio River Division of the U. S. Army Corps of Engineers, IMM-NYU 200, 1953.

[22] URSELL, F. The Long-Wave Paradox in the Theory of Gravity Waves, Proc. Cambridge Phil. Soc., Volume 49, 1953.

[23] WEINBLUM, G. P. Analysis of Wave Resistance, Report 710 issued by the David W. Taylor Model Basin (Washington, D. C.) September 1950.

[24] WEINSTEIN, A. Sur un problème aux limites dans une bande indéfinite. Comptes Rendus de l'Academie des Sciences, Volume 184, p. 497, 1927.

ELLIPTIC AND HYPERBOLIC AXIALLY SYMMETRIC PROBLEMS [1]

A. WEINSTEIN [2]

1. *Introduction.* It is well known that among ordinary differential equations with variable coefficients those with singular coefficients, like Bessel's equation, are the most important. While partial differential equations with variable coefficients have been extensively studied in recent times, the usual assumption is that the coefficients remain bounded. However, there is an important class of partial differential equations with unbounded coefficients which in special cases have been considered since Euler's time and which became of increasing importance for the general theory as well as for applications.

In the present lecture a short review will be presented about the various problems concerning the equation

$$\frac{\partial^2 u}{\partial y^2} + \frac{k}{y}\frac{\partial u}{\partial y} = L[u] \tag{1.1}$$

for a function $u = u(x_1, \ldots, x_m, y)$ where L is a linear differential operator in the variables $x_1, x_2, \ldots, x_m$ with constant or variable regular coefficients. k denotes here a real parameter, $-\infty < k < \infty$. We shall be mainly concerned with two cases in which L is either $-\Delta u$ or $+\Delta u$, where $\Delta u = \sum_{i=1}^{m} \frac{\partial^2 u}{\partial x_i^2}$.

In the first case our equation (1.1) is elliptic. In the second case it is hyperbolic. For $k = 1, 2, 3, \ldots$ our equation admits the following interpretation. Let us consider the function U depending on the $m + k + 1$ variables $x_1, \ldots, x_m; y_1, \ldots, y_{k+1}$ and satisfying the equation

$$\sum_{i=1}^{k+1} \frac{\partial^2 U}{\partial y_i^2} = L[U]$$

Let us consider the solutions $U = u$ of this equation, which depend only on the $m + 1$ variables $x_1, \ldots, x_m$ and $y = \sqrt{y_1^2 + \ldots + y_{k+1}^2}$. Such solutions

[1]) This research was supported in part by the United States Air Force through the Office of Scientific Research of the Air Research and Development Command.

[2]) Guggenheim fellow and Research Professor, Institute for Fluid Dynamics and Applied Mathematics, University of Maryland.

will satisfy the equation (1.1) and we may say that they are obtained by introducing cylindrical or axial symmetry in the original $m + k + 1$ dimensional space. This fact gives useful hints for the investigation of the general case when k is any number.

2. *Two Fundamental Recursion Formulas.* Let us denote by u^k a solution of (1.1). Sometimes we write $u^{(k)}$ for u^k. Then we have the following two fundamental recursion formulas

$$u_y^k(x, y) = yu^{k+2}(x, y) \tag{2.1}$$

$$u^k(x, y) = y^{1-k}\, u^{2-k}(x, y) \tag{2.2}$$

Here x is an abbreviation for $(x_1, x_2, \ldots, x_m)$. These fundamental formulas are simple consequences of elementary properties of the expression

$$u_{yy} + \frac{k}{y} u_y$$

where u is considered to be a function of y alone. This fact was recognized only recently. Previously the recursion formulas were used in isolated cases as properties of special differential equations. The recursion formulas have an amazing range of applications, most of them new and others which simplify considerably many classical proofs.

3. *Generalized Axially Symmetric Potential Theory.* This theory, abbreviated by *GASPT*, deals with the equation [1],

$$\frac{\partial^2 u}{\partial y^2} + \frac{k}{y}\frac{\partial u}{\partial y} + \sum_{i=1}^{m} \frac{\partial^2 u}{\partial x_i^2} = 0, \quad -\infty < k < \infty \tag{3.1}$$

While most of the following results are valid for any value of m, the case $m = 1$ is of particular interest as it includes for $k = 1$ the axially symmetric Laplace equation and above all because it leads to a generalization of the Cauchy-Riemann equations. In fact the equations

$$y^k u_x^{(k)} = u_y^{(-k)} \tag{3.2}$$

and

$$y^k u_y^{(k)} = -u_x^{(-k)}, \; x = x_1, \tag{3.3}$$

introduce to every $u^{(k)}$ an associated function $u^{(-k)}$, a fact discovered by Stokes for $k = 1$ and developed by Beltrami in his celebrated memoirs on symmetric potentials. In this case $u^{(+1)}$ and $u^{(-1)}$ are called the velocity potential and the Stokes stream function.

Various and seemingly disconnected problems such as the theory of Weber's discontinuous integrals, the torsion of beams and the theory of Tricomi's equation and its application to transonic flow recently gave an

impetus for the development of the theory for all values of k. Among the results thereby obtained let us mention explicit integral formulas for all values of m for the fundamental solutions of (3.1) in the large, valid for all values of $y \geqq 0$ which for $k < 1$ lead to the Green's function. The fact (overlooked by Beltrami) that for $m = 1$ the corresponding associate function is many valued leads to a considerable extension of the Weber-Shafheitlin discontinuous integrals. A fundamental point of the theory is the uniqueness problem (Weinstein [1], Hyman [2] and Huber [3]) for functions u^k which vanish on the singular line, $y = 0$.

The formula (2.2) which plays already an important part in the previous consideration becomes essential in the classical problem of the determination of the stream function of the flow of an incompressible fluid about a body of revolution. In fact this formula, which we call a correspondence principle, reduces the hydrodynamical problem to the electrostatic problem for a body of the same meridian profile but in a space of five dimensions. By using results of electrostatics, new flows have been explicitly computed [4].

The same correspondence principle connects the virtual mass of an axially-symmetric body with the electrostatic capacity of the corresponding 5-dimensional body [5]. As the influence of symmetrization on the capacity is known by the investigations of Polya and Szego our principle leads to corresponding results for the virtual mass. Incidentally, *GASPT* leads to a natural generalization of the concept of symmetrization which includes the classical cases of Steiner and Schwarz. The behavior of the virtual mass under symmetrization is of importance in view of a fundamental minimum property of this quantity which was discovered by Riabouchinsky [6] and which opened a new approach to the Helmholtz problem. This problem had already been solved for jets by the present author [7] using the method of continuity. Recently, Riabouchinsky's result was successfully used in the Helmholtz wake problem by Garabedian and Spencer [8]. Among other applications of *GASPT* we mention the extension of the method of sources and sinks which yielded flows about new types of bodies.

4. *Hyperbolic problems.* By putting in (1.1)

$$L[u] = \sum_{i=1}^{m} \frac{\partial^2 u}{\partial x_i^2}$$

we obtain the general case of the Euler-Poisson-Darboux equation (E.P.D. equation). This equation has a long history and names like Euler, Poisson, Riemann, and Darboux are connected with its theory. One of the basic problems is the singular Cauchy problem with the data $u(x, 0) = f(x)$, $u_y(x, 0) = 0$. This problem was formerly considered only in some special cases, seemingly unrelated

to each other. The most outstanding classical investigations are those of Darboux who considered the cases $m = 1, 0 < k < 2$, and those of Poisson who besides being a precursor of Darboux, considered the case $m = 3$, $k = 2$. While, therefore, non-integer values of k were introduced in this theory earlier than in *GASPT*, it is only recently that all values of k and m have been investigated [9], [10]. The results are the following: for $k \geqq m - 1$ the solution of the Cauchy problem can be explicitly given. In fact, for $k = m - 1$ it is given by a Mean Value formula previously used by Poisson for $m = 2$ and rewritten for any value of m by Asgeirsson. From this follows the solution for $k = m$, $m + 1$, $m + 2, \ldots$, by Hadamard's method of descent. The resulting formula is analytic in k and remains valid for all $k > m - 1$. One could call this result an interpolation or the generalized method of descent. For $k < m - 1$ but $k \neq -1, -3, -5, \ldots$, the solution can be obtained either by the use of the recursion formula (2.2) alone [9], or in a more elegant form by using both formulas (2.1) and (2.2) which leads to the equation

$$u^k = y^{1-k}\left(\frac{\partial}{y\partial y}\right)^n (y^{k+2n-1}\, u^{k+2n})$$

where n is chosen such that $k + 2n \geqq m - 1$ and u^{k+2n} is determined by the Cauchy conditions

$$u^{k+2n}(x, 0) = \frac{f(x)}{(k+1)(k+3)\ldots(k+2n-1)},$$

$$u_y^{k+2n}(x, 0) = 0$$

This formula [10] contains as a special case the classical solution of the three-dimensional wave equation given by Poisson, and also the celebrated formula of Parsival-Volterra for cylindrical waves. One can say that Poisson used *implicitly* in his method the solution u^2 in order to find the solution u^0 of the wave equation in the three dimensional case. This idea has strangely enough been completely abondoned even for the wave equation while it is general and simple, being based, for small values of k, on the recursion formulas.

While the *EPD* equation is singular for all values of k, it is only for $k = -1, -3, \ldots$, that we obtain at $y = 0$ a singular behavior. Unless the initial values are polyharmonic in $x_1, x_2, \ldots, x_m$, the derivatives of u of a certain order with respect to y will become infinite for $y = 0$, [9], [11], [12]. In this sense one could say that the solution does not depend continuously on the initial data.

The E.P.D. equation with $k = -1, -2, -3, \ldots$ is useful in the solution of the radiation problem for the wave equation in any number of dimensions [13]. In this case the corresponding problem for the E.P.D. equation can be generaliz-

ed for all values of $k < 1$. This problem coincides, for $k = 1/3$, with a problem of Tricomi and Germain-Bader [14] for the Tricomi equation.

The singular Cauchy problem for the general hyperbolic equation of type (1.1) has recently been solved by F. Bureau. Again our recursion formulas play an important part in his theory, [15].

As we have mentioned, the equations of type (1.1), important in themselves, are sometimes used to solve other equations with regular coefficients which do not possess any kind of axial symmetry. This fact is emphasized by some recent results on the polyharmonic equation and the iterated wave equation. As is well known, a solution of the biharmonic equation $\Delta\Delta u = 0$ admits the classical decomposition due to Almansi and Goursat, $u = u_1^0 + yu_2^0$ where u_1^0 and u_2^0, according to our notations, are harmonic functions of x and y. In this decomposition the function yu_2^0 is already a biharmonic function. The methods of axial symmetry lead to a new decomposition, namely $u = u^0 + u^{-2}$ where both functions, u^0 and u^{-2}, satisfy equations of the second order of the type considered in our general theory. This remark admits, obviously, various applications and extends considerably the theory of generalized axially symmetric equations.

Bibliography

[1] A. Weinstein, *Generalized axially symmetric potential theory*, Bull. Amer. Math. Soc., Vol. 59 (1953) pp. 20-38.

[2] M. Hyman, Abstracts (Preliminary Reports), Bull. Amer. Math. Soc., Vol. 54 (1948) p. 1065; vol. 55 (1949) pp. 284-285. Proc. K. Nederl. Akad. Wetens. A, 57, (1954) pp. 408-413.

[3] A. Huber, *On the uniqueness of generalized axially symmetric potentials*, Annals of Mathematics, 1954.

[4] L. E. Payne, *On axially symmetric flow and the method of generalized electrostatics*, Quarterly of Appl. Math., vol. 10 (1952) pp. 197-204.

[5] L. E. Payne and A. Weinstein, *Capacity, virtual mass and generalized symmetrization*, Pacific Journal of Math., vol. 2 (1952) pp. 633-641.

[6] D. Riabouchinsky, *Sur un problème de variation*, C. R. Acad. Sci. Paris, vol. 185 (1927) pp. 840-841.

[7] A. Weinstein, *Sur le theorème d'existence des jets fluides*, R. Accademia dei Lincei, 157 (1927).

[8] P. R. Garabedian and D. C. Spencer, *Extremal methods in cavitational flow*, Journal of Rational Mechanics and Analysis, vol. 1 (1952) pp. 359-409.

[9] A. Weinstein, *Sur le problème de Cauchy pour l'équation de Poisson et l'équation des ondes*, C. R. Acad. Sci. Paris, vol. 234 (1952) pp. 2584-2585.

[10] A. Weinstein, *The singular solutions and the Cauchy problem for generalized Tricomi equations*, Communications on Pure and Applied Mathematics, vol. VII, no. 1 (1954), pp. 105-116.

[11] J. B. Diaz and H. F. Weinberger, *A solution of the singular initial value problem for the Euler-Poisson-Darboux equation*, Proc. of the Amer. Math. Soc., vol. 4 (1953) pp. 703-715.

[12] E. K. Blum, *The Euler-Poisson-Darboux equation in the exceptional cases*, Proc. of the Amer. Math. Soc., vol. 5 (1954), pp. 511-520.

[13] A. Weinstein, *The generalized radiation problem and the Euler-Poisson-Darboux, equation*, Summa Brasiliensis Mathematicae (in print).

[14] P. Germain and R. Bader, *Sur le problème de Tricomi*, Rendiconti del Circulo Matematico di Palermo, series 2, t. 2 (1953), p. 53.

[15] F. Bureau, *Divergent Integrals and Partial Differential Equations*, Communications on Pure and Applied Mathematics, vol. VII (1954), pp. 143-202.

SECTION VI

LOGIC AND FOUNDATIONS

STATED ADDRESSES

DIE FIKTION DER ÜBERABZÄHLBARKEIT

PAUL LORENZEN

Die Aufforderung der Kongressleitung zu einem Referat habe ich mit Freude und in großer Dankbarkeit für diese Ehre angenommen. Als Thema habe ich die Fiktion der Überabzählbarkeit gewählt. Darf ich dazu als erstes bemerken, daß mit diesem Titel nicht behauptet sein soll, der Satz: „Es gibt überabzählbare Mengen" sei falsch. Eine solche Behauptung hätte ja nur Sinn, wenn eine Definition des Mengenbegriffs vorläge, die einen Beweis (oder evtl. eine Widerlegung) der Behauptung ermöglicht.

Gerade hier liegt aber die Schwierigkeit. Daß aus der Cantorschen Defintion der Menge als einer „Zusammenfassung" jedenfalls nichts streng bewiesen werden kann, dürfte wohl ebenso klar sein wie bei der Euklidischen Definition des Punktes als eines Etwas, das keine Teile hat.

Um eine Grundlage für das Folgende zu haben, werde ich als unproblematisch nur den Umgang mit Figuren voraussetzen, die aus endlich vielen Atomfiguren zusammengesetzt sind, so wie es z.B. der Metamathematiker gewohnt ist, mit den Formeln einer formalisierten Theorie als bloßen Figuren zu operieren.

Was eine endliche Menge von Figuren ist, ist dann leicht zu definieren. Denn wir können zunächst endlich viele Figuren zusammensetzen, etwa durch Kommata getrennt:

$$x_1,\ x_2,\ \ldots,\ x_n.$$

Wenn ich diese neuen Figuren kurz „Systeme" nenne, und jetzt zwei solche Systeme schon dann als gleich betrachte, wenn sie aus denselben Figuren, (also unabhängig von der Reihenfolge und Vielfachheit) zusammengesetzt sind, so erhalten wir durch diese Abstraktion die endlichen Mengen:

$$\{x_1,\ x_2,\ \ldots,\ x_n\}.$$

Zwei endliche Mengen haben gleiche Kardinalzahl, wenn es eine eineindeutige Zuordnung zwischen ihnen gibt. Dieser Satz ist der Ausgangspunkt für die Cantorschen transfiniten Kardinalzahlen. Für endliche Mengen M und N gilt aber genauer, daß sie schon dann gleiche Kardinalzahl haben, wenn eine eineindeutige Zuordnung zwischen ihnen logisch möglich ist.

Für $M = \{x_1, \ldots, x_m\}$ und $N = \{y_1, \ldots, y_n\}$ heißt diese „logische Möglichkeit" folgendes. Mit einem Formelsymbol $\pi(x, y)$ wird die Konjunktion

von Aussagen gebildet, die aussagen, daß π eine eineindeutige Zuordnung zwischen M und N ist, also

(1) $\underset{M}{\wedge}_x \underset{N}{\vee}_y \pi(x, y)$ d.h. $(\pi(x_1, y_1) \vee \pi(x_1, y_2) \vee \ldots) \wedge (\pi(x_2, y_1) \vee \pi(x_2, y_2)$ $\vee \ldots) \wedge \ldots$

(2) $\underset{N}{\wedge}_y \underset{M}{\vee}_n \pi(x, y)$

(3) $\underset{M}{\wedge}_x \underset{N}{\wedge}_{y', y''} (\pi(x, y') \wedge \pi(x, y'') \rightarrow y' = y'')$ d.h. $(\pi(x_1, y_1) \wedge \pi\ (n_1, y_2) \rightarrow$ $\rightarrow y_1 = y_2) \wedge \ldots$

(4) $\underset{N}{\wedge}_y \underset{M}{\wedge}_{x', x''} (\pi(x', y) \wedge \pi(x'', y) \rightarrow x' = x'')$

M und N haben genau dann gleiche Kardinalzahl, wenn die Konjunktion von (1) — (4) unwiderlegbar ist, d. h. wenn ihre Negation keine aussagenlogische Tautologie ist.

Dieser Satz hat bei unendlichen Mengen kein Analogon. Mit einem Formelsymbol $\pi(x)$ ist $\underset{M}{\wedge}_x \pi(x)$ für unendliche Mengen M keine Formel der Aussagenlogik mehr, auch nicht der Prädikatenlogik. Würde man also den Kardinalzahlbegriff für endliche Mengen auf die logische Möglichkeit einer eineindeutigen Zuordnung statt auf die „Existenz" einer solchen gründen, so käme man von vornherein gar nicht in die Versuchung „transfinite" Kardinalzahlen definieren zu wollen. Es bliebe bei der schlichten Unterscheidung zwischen „endlich" und „unendlich".

Unendliche Mengen von Figuren können wir nicht durch Zusammensetzung ihrer Elemente erhalten, sondern nur durch Aussageformen $A(x)$ über die Figuren x. Betrachten wir zwei Aussageformen $A(x)$ und $B(x)$ schon dann als gleich, wenn

$$\wedge_x (A(x) \leftrightarrow B(x))$$

gilt, so entstehen durch diese Abstraktion die „Mengen". Jede Menge wird durch eine Aussageform dargestellt.

Die Schwierigkeit der Definition des Begriffs Menge ist dadurch verschoben auf die Definition von „Aussageformen". Diese Verschiebung hat den großen Vorteil, daß die „Aussageformen" — wie auch ihre Definition im einzelnen gewählt werden mag — jedenfalls wieder Figuren sein werden, etwa zusammengesetzt aus logischen Partikeln, wie $\rightarrow, \wedge \vee \neg$, aus Variablen, dem Gleichheitszeichen $=$ usw. Die, wie man sagt, „naive" Annahme, man wisse, was eine Aussageform ist (weil man ja die Umgangssprache beherrsche), es sei also keine Definition hierfür erforderlich, ist bekanntlich durch die mengentheoretischen Paradoxien widerlegt. Seitdem suchen die Kritiker nach einem

„Fehler", der zu diesen Widersprüchen geführt hat. Der einzige Fehler, der bei dieser Suche bisher gefunden ist, ist die Benutzung unendlicher Regresse bei der Definition der widerspruchsvollen Mengen. Diese Entdeckung führte Poincaré zum Verbot der imprädikativen Definitionen und Russell stellte daraufhin seine verzweigte Stufenlogik auf. Auch Weyl befand sich zunächst (1918 in seinem Buche „Das Kontinuum") genau auf diesem Wege.

Mit rühmlicher Ausnahme der Intuitionisten, die eigene Wege einschlugen, hat man ausschließlich deshalb weil die verzweigte Stufenlogik zu kompliziert sei, und weil sie den Satzbestand der modernen Mathematik nicht in befriedigendem Maße „retten" könne, nur deshalb hat man diesen Ansatz aufgegeben zu Gunsten einer unverzweigten Stufenlogik. Die Imprädikativität dieser Logik ist aber — jedenfalls bisher — nicht zu rechtfertigen gewesen: der Fehler eines unendlichen Regresses ist dadurch also nicht vermieden. Die verzweigte Stufenlogik lässt sich jedoch statt durch die Benutzung unverzweigter Stufen auch durch die Einführung von so etwas wie ungestufter Zweige sehr vereinfachen. Und hierbei kann man „prädikativ" bleiben, d. h. man vermeidet unendliche Regresse.

In grober Skizzierung lässt sich ein solcher Aufbau der Mengenlehre folgendermaßen beschreiben. Mit dem zu Grunde gelegten Fundament des Operierens mit Figuren sind auch die Grundzahlen 1, 2, 3, ... gegeben. Man kann sie ja etwa als die Figuren definieren, die sich aus einem einzigen Atom zusammensetzen lassen. Auf diesem operativen Fundament — und es dürfte wohl das einzige Fundament sein, das von niemandem ernsthaft in Frage gestellt wird, — erhält man dann durch induktive Definitionen die ersten Relationen und Funktionen, z.B. die Ordnung $\leqq$, die Addition $+$, und damit die ersten Aussageformen. Die beiden Prinzipien: (1) Induktive Definition, (2) Zusammensetzung mit den logischen Partikeln

$$\rightarrow,\ \wedge,\ \vee,\ \wedge_x,\ \vee_x,\ \neg,$$

erzeugen in iterierter Anwendung eine „1. Schicht" S_1 von Aussagen und Aussageformen. Die Grundzahlen und die Formeln dieser ersten Schicht bilden eine abzählbare Menge: durch induktive Definition lässt sich eine Abzählung definieren. Die Aussageform, die diese Abzählung darstellt, gehört aber zu einer neuen „2. Schicht" S_2 von Formeln. Diese wird mit Hilfe induktiver Definitionen und der logischen Partikeln über der 1. Schicht S_1 genau so konstruiert wie S_1 über S_0 d. h. über den Grundzahlen.

Dieser Prozess der Schichtenbildung lässt sich iterieren:

$$S_0 \subseteqq S_1 \subseteqq S_2 \subseteqq \ldots$$

sogar über ω hinaus, etwa bis ω^ω. Für eine Limeszahl θ sei S_θ die Vereinigung aller S_ϑ mit $\vartheta < \theta$.

Ohne auf irgendwelche Einzelheiten eingehen zu müssen, ist klar, daß man vom operativen Fundament aus niemals zu etwas im Cantorschen Sinne Überabzählbaren gelangen kann. „Überabzählbares" kann es nur in einem relativen Sinne geben, nämlich derart, daß für eine Menge von Figuren keine Abzählung in einer vorgegebenen Schicht darstellbar ist.

Nach Konstruktion einer Schicht S_ϑ beweist das Cantorsche Diagonalverfahren, daß es in S_ϑ keine Aussageform gibt, die eine Abzählung aller in S_ϑ darstellbaren Mengen von Grundzahlen darstellt. In $S_{\vartheta+1}$ ist dagegen eine Abzählung der in S_ϑ darstellbaren Mengen darstellbar. Das Diagonalverfahren liefert daher eine in $S_{\vartheta+1}$ darstellbare Menge von Grundzahlen, die in S_ϑ nicht darstellbar ist. Die Überabzählbarkeit der Menge „aller" Mengen von Grundzahlen lässt sich also so interpretieren, daß in keiner Schicht alle Mengen von Grundzahlen darstellbar sind, jede Schicht liefert neue Mengen.

Ich möchte im Folgenden darüber berichten, wie weit diese Interpretation die Ergebnisse der modernen Mathematik zu rechtfertigen gestattet. Diese hängen ja wesentlich, z. B. in der Integrationstheorie und in der Topologie, von der Unterscheidung: abzählbar — überabzählbar ab.

Es entsteht also die Aufgabe, innerhalb der skizzierten Konstruktion von Sprachschichten einen „Ersatzbegriff" für die Überabzählbarkeit zu finden. Wir fixieren dazu willkürlich eine erste Limeszahl θ_1 und eine zweite Limeszahl $\theta_2 > \theta_1$ und denken uns alle Schichten bis S_{θ_2} konstruiert. Die Schichten bis zum Index θ_1 mögen „primär" heißen, die höheren Schichten bis θ_2 „sekundär".

Diese Unterscheidung überträgt sich auf Mengen: Eine Menge heißt primär, wenn sie durch eine Aussageform einer primären Schicht darstellbar ist, sonst sekundär.

Zunächst sieht es so aus, als ob das Begriffspaar primär — sekundär nichts zu tun haben kann mit dem Cantorschen: abzählbar — überabzählbar. Während jede Untermenge einer abzählbaren Menge nämlich selbst abzählbar ist, kann man zeigen, daß jede unendliche primäre Menge sekundäre Untermengen besitzt. Die Begriffe primär und sekundär sind zudem nicht auf Mengen beschränkt, sondern auf Grund der darstellenden Aussageformen sind auch primäre und sekundäre Relationen, primäre und sekundäre Funktionen (speziell primäre und sekundäre Folgen) zu unterscheiden.

Ich möchte trotzdem behaupten, daß „sekundär" als ein Ersatz für die Überabzählbarkeit der modernen Mathematik benutzt werden kann.

Als ein erstes Beispiel führe ich die Lebesguesche Integrationstheorie an. Von den Grundzahlen gehe man zunächst zu den rationalen Zahlen über, was ja keine Schwierigkeiten bietet. Mit Hilfe der primären konvergenten (im Cauchyschen Sinne) Folgen rationaler Zahlen führe man danach die reellen Zahlen ein. Jede reelle Zahl ist also primär. Jedes Intervall reeller Zahlen enthält aber

reelle Zahlen beliebig hoher primärer Schichten und ist daher eine sekundäre Menge. Das entspricht der üblichen Überabzählbarkeit. Ebenso ist auch jede offene Menge sekundär.

Natürlich haben jetzt nicht alle Cauchy-konvergenten Folgen reeller Zahlen eine reelle Zahl als Grenzwert, sondern nur die primären. Die „absolute" Vollständigkeit der reellen Zahlen wird aber in der modernen Analysis auch nur scheinbar gebraucht. Hat man z. B. eine offene Überdeckung des Intervalles [0, 1], so hat man auch eine Überdeckung mit rationalen Intervallen u_1, u_2, u_3, ... Wäre jetzt $[0, 1] \subsetneq u_1 \cup u_2 \cup \ldots \cup u_n$ für jedes n, dann gäbe es — so schließt man mit dem Auswahlprinzip — eine reelle Folge x_1, x_2, x_3, ... mit $x_n \in [0, 1]$ und $x_n \notin u_1 \cup u_2 \cup \ldots \cup u_n$. Die Existenz einer konvergenten Teilfolge von x_1, x_2, ... führt dann zum Widerspruch.

Dieser Beweis lässt sich folgendermaßen präzisieren. Aus $[0, 1] \subsetneq u_1 \cup \ldots \cup u_n$ folgt die Existenz einer rationalen zahl r mit $r \in [0, 1]$ und $r \notin u_1 \cup \ldots \cup u_n$. Man erhält also ohne Auswahlprinzip eine rationale Folge r_1, r_2, r_3, ... und wenn die Endpunkte der überdeckenden Intervalle u_n primäre Folgen bilden, ist auch die Folge der r_n primär, besitzt also eine primäre Cauchy-konvergente Teilfolge. Dieser Beweis benutzt bei Beschränkung auf primäre Überdeckungen nur die primäre Vollständigkeit der reellen Zahlen, nicht die absolute.

Die Borelsche Definition einer reellen Zahl als eines Masses für offene Mengen, lässt sich entsprechend zwar nicht für alle offenen Mengen durchführen, wohl aber für alle diejenigen Vereinigungen von Intervallen, bei denen die Intervallendpunkte durch primäre Mengen gegeben sind. Diese offenen Mengen sind also durch ein primäres „Gerüst" von Endpunkten gegeben. Ich nenne sie primär-offene Mengen. Wörtlich wie in der modernen Mathematik lässt sich für die primär-offenen Mengen ein Maß definieren, und es lassen sich dann auch „messbare" Mengen einführen: Es sind nur die offenen Mengen stets durch primär-offene zu ersetzen. Die messbaren Mengen bilden bekanntlich einen abzählbar-vollständigen Verband, hier bilden sie einen primär-vollständigen Verband: für jede primäre Familie von messbaren Mengen (d. h. für solche Familien, für die ein geeignet zu definierendes primäres „Gerüst" vorhanden ist) sind auch Durchschnitt und Vereinigung messbar. Insbesondere hat jede primäre Menge das Maß 0, was dem Lebesgueschen Satz, daß jede abzählbare Menge das Maß 0 hat, genau entspricht.

Die Integrationstheorie lässt sich — wörtlich parallel zur modernen Behandlung — weiter ausbauen. Neben die messbaren Mengen treten die messbaren Funktionen. Sie sind integrierbar. Die quadratisch-integrierbaren Funktionen bilden einen metrischen Raum L^2, der zum Hilbertschen Raum H isomorph ist. H wird hier definiert als der Raum der quadratisch-summierbaren

primären Folgen reeller Zahlen. Topologisch wichtig ist, daß der Hilbertsche Raum eine abzählbare Umgebungsbasis besitzt. Die Punkte mit rationalen Koordinaten, die außerdem fast alle 0 sind, bilden nämlich eine abzählbare Punktmenge, die im Hilbertschen Raum dicht ist. Diese Punktmenge ist in der 1. Schicht darstellbar, also jedenfalls primär. Auch im L^2 lässt sich entsprechend eine primäre Umgebungsbasis definieren.

Auf etwas größere Schwierigkeiten stößt man, wenn man in der abstrakten Topologie das „Abzählbarkeitsaxiom" durch das Axiom: „Es gibt eine primäre Umgebungsbasis" ersetzen will. Die Voraussetzungen der topologischen Sätze müssen dann häufig präziser formuliert werden. Zur Übertragung des Urysohnschen Satzes, daß jeder normale topologische Raum mit abzählbarer Basis in den Fundamentalquader des Hilbertschen Raums topologisch abgebildet werden kann, wird z. B. nicht nur gebraucht, daß die Basis primär ist, sondern auch daß der ganze Raum als Vereinigung primärer Mengen darstellbar ist. Auf die Einzelheiten aller dieser Untersuchungen kann ich hier nicht eingehen. Da verständlicherweise noch nicht der Versuch gemacht worden ist, die gesamte Mathematik auf die eben skizzierte Methode umzustellen, kann bisher niemand mit Sicherheit voraussagen, daß bei der Umstellung von „überabzählbar" auf „sekundär" nichts Wesentliches verloren geht. In manchen Fällen könnte es ja auch auf den subjektiven Geschmack ankommen, ob etwas für „wesentlich" gehalten wird oder nicht.

Ich möchte hier einen Fall dieser Art anführen. Es ist in der modernen Mathematik üblich, manche Beweise, die die Abzählbarkeit einer Menge erfordern, dadurch auf überabzählbare Mengen auszudehnen, daß man die Abzählbarkeit durch die „Wohlordenbarkeit" ersetzt. Ich wähle ein algebraisches Beispiel. K sei ein Körper und es werde nach der algebraisch abgeschlossenen Hülle von K gefragt. Wenn K abzählbar wäre, so würde man in der modernen Algebra sagen, dann wäre es auch der Polynomring $K[x]$. Mit einer Abzählung $f_1(x), f_2(x), \ldots$ dieses Ringes wäre eine Kette algebraischer Oberkörper $K \subseteq K_1 \subseteq K_2 \subseteq \ldots$ zu konstruieren, so daß $f_n(x)$ in K_n zerfällt. Die Vereinigung $\overline{K} = U_n K_n$ wäre die algebraisch abgeschlossene Hülle und $\overline{K}$ wäre auch abzählbar. Ist aber K überabzählbar, so muß anstelle der Abzählung eine Wohlordnung verwendet werden:

$$f_1(x),\ f_2(x), \ldots,\ f_\omega(x), \ldots$$

Eine entsprechende transfinite Körperkette $K \subseteq K_1 \subseteq K_2 \subseteq \ldots \subseteq K_\omega \subseteq \ldots$ liefert dann durch ihre Vereinigung die algebraisch abgeschlossene Hülle. Der erste Teil dieser Überlegungen lässt sich wörtlich übernehmen. Ist K primär, dann gibt es auch eine primäre Abzählung von K (das ist eine Eigenschaft, die nur davon abhängt, daß die sprachlichen Mittel jeder Schicht genügend stark

sind). Die Körperkette $K \subseteqq K_1 \subseteqq K_2 \subseteqq \ldots$ wird primär und daher auch die algebraisch abgeschlossene Hülle $\bar{K}$. Ist aber K sekundär, so tritt — im Gegensatz zur modernen Algebra — eine sekundäre Abzählung an die Stelle der primären Abzählung. Es bleibt bei einer ω-Folge $K \subseteqq K_1 \subseteqq K_2 \subseteqq \ldots$, die jetzt nur sekundär statt primär ist. Auch die algebraisch abgeschlossene Hülle $\bar{K}$ ist sekundär. Es bietet sich hier keine Notwendigkeit an, zu Wohlordnungen überzugehen. Denn wozu sollte man sekundäre Wohlordnungen betrachten, wenn es sekundäre Abzählungen gibt? Und primäre Wohlordnungen der sekundären Menge K sind genau so wenig wie primäre Abzählungen vorhanden.

Ich wage zu bezweifeln, daß der Verlust dieser Art von „transfiniten" Schlüsse für die Algebra wesentlich ist.

Für die, die in dieser Geschmacksfrage anderer Meinung sind, möchte ich jedoch noch auf folgendes hinweisen. Bei sekundären Körpern K ist häufig, wie beim reellen Zahlkörper und allen seinen endlichen Erweiterungen, der Körper K eine Vereinigung von primären Körpern $K^1 \subseteqq K^2 \subseteqq \ldots$ etwa derart, daß K^ϑ genau die in der Schicht S_ϑ darstellbaren Elemente von K enthält. Setzt man hier die primären Abzählungen von K^1, von $K^2 - K^1$, von $K^3 - K^2$, ... hintereinander, so erhält man auf zwanglose Weise eine (sekundäre) Wohlordnung vom Typ $\theta_1 \cdot \omega$. Die Konstruktion der algebraisch abgeschlossenen Hülle $\bar{K}$ mit Hilfe dieser Wohlordnung liefert auch für $\bar{K}$ eine Darstellung als Vereinigung primärer Körper. Diese Darstellung erhält man nicht, wenn man von einer (sekundären) Abzählung ausgeht. Insofern könnten also auch in dem hier dargestellten Aufbau der Mathematik transfinite Ordinalzahlen Verwendung finden.

Zusammenfassend darf ich wohl behaupten, daß die Heranziehung der sekundären Schichten (natürlich nicht ins Uferlose, sondern nur bis zu unproblematischen Ordinalzahlen, wie etwa ω^ω) in Verbindung mit den in jeder Schicht zur Verfügung stehenden sprachlichen Mitteln (schon in der ersten Schicht sind Relationen darstellbar, die nicht mehr „arithmetisch" im Sinne von Kleene und Mostowski sind) genügend Bewegungsfreiheit liefert, um den Gedankengängen der modernen Mathematik weitestgehend zu folgen.

Sollte man trotzdem einmal an eine Grenze stoßen, so braucht man daraus nicht unbedingt zu schließen, daß das operative Fundament zu schmal ist und daß also irgendwo ein „Sprung ins Ungewisse" erst den wahren Mathematiker ausmacht. Es bleibt auch dann die Möglichkeit, die Mahnung zu beherzigen, die Herr Skolem uns auf dem letzten Internationalen Kongress nahelegte: „We ought not to regard all that's written in the traditional textbooks as something sacred".

Bonn.

DEVELOPMENT AND APPLICATIONS OF THE "PROJECTIVE" CLASSIFICATION OF SETS OF INTEGERS.

ANDRZEJ MOSTOWSKI

The aim of this paper is to review some recent results belonging to the field of mathematical logic and connected with the well-known classification of sets of integers due to Kleene [3].

1. *Classes* $\mathbf{P}_n$ *and* $\mathbf{Q}_n$. By "sets" we mean in the sequel sets of non-negative integers as well as sets of ordered n-tuples $\mathfrak{x} = (x_1, x_2, \ldots, x_n)$ of non-negative integers. We denote by $\mathbf{P}_0$ or by $\mathbf{Q}_0$ the class of recursive sets, i.e., of sets the characteristic functions of which are general recursive. A set X belongs to the class $\mathbf{P}_{n+1}$ (or to the class $\mathbf{Q}_{n+1}$) if there is a set Y such that $\mathfrak{x} \in X \equiv (\exists x)[(\mathfrak{x}, x) \in Y]$ and $Y \in \mathbf{Q}_n$ (or $\mathfrak{x} \in X \equiv (x)[(\mathfrak{x}, x) \in Y]$ and $Y \in \mathbf{P}_n$).

The classification of sets determines a similar classification of functions the arguments and the values of which run over the set of non-negative integers (or more generally over the sets of ordered k-tuples and l-tuples of integers). A function f belongs to the class $\mathbf{P}_n$ or $\mathbf{Q}_n$ if the set X of pairs $(\mathfrak{x}, \mathfrak{y})$ satisfying the equation $\mathfrak{y} = f(\mathfrak{x})$ belongs to the class $\mathbf{P}_n$ or $\mathbf{Q}_n$.

2. *Properties of classes* $\mathbf{P}_n$ *and* $\mathbf{Q}_n$. Many properties of these classes have been established in papers by Kleene, Markwald, and myself. I shall give below some theorems which I believe to be new. Some of them illustrate once again the striking analogy between the classes $\mathbf{P}_n$ and $\mathbf{Q}_n$ and the projective sets of Lusin.

I. *A function of the class* $\mathbf{P}_n$ *belongs at the same time to the class* $\mathbf{Q}_n$.

II. *For every* $n > 0$ *there are functions which belong to* $\mathbf{Q}_n$ *but do not belong to* $\mathbf{P}_n$.

III. *Each function* $f \in \mathbf{P}_n$, $n > 0$, *has the form* $f(\mathfrak{x}) = s_1(g(\mathfrak{x}))$ *where* $g \in \mathbf{Q}_{n-1}$ *and* s_1, s_2 *are arbitrary pairing functions.*

In the next two theorems we assume that f and g are functions with numerical values.

IV. *If* $f \in \mathbf{Q}_1$, $g \in \mathbf{P}_0$, *and* $f(\mathfrak{x}) \leqq g(\mathfrak{x})$ *for every* $\mathfrak{x}$, *then* $f \in \mathbf{P}_0$.

V. *If* $f \in \mathbf{Q}_1$, *then the set* X *of pairs* $(\mathfrak{x}, y)$ *satisfying the inequality* $y < f(\mathfrak{x})$ *belongs to* $\mathbf{P}_1$.

These theorems find applications to the so-called problem of uniformization. Let X be a plane set which has a non-empty intersection with every

straight line $x = n$, $n = 0, 1, 2, \ldots$. A function f is said to uniformize X if $f(x) \in X$ for every x.

VI. *A* $\mathbf{P}_1$*-set can be uniformized by means of a general recursive function.*

VII. *If* $f \in \mathbf{Q}_1 - \mathbf{Q}_0$ *and* X *is the set of pairs* (x, y) *satisfying the inequality* $y \geqq f(x)$, *then* $X \in \mathbf{Q}_1$ *and* X *cannot be uniformized by means of general recursive functions.*

Final theorems of this section will be concerned with the so-called problems of separation which play the prominent role also in the theory of projective sets. These problems arose in the following way: Kleene [3] has shown that the intersection $\mathbf{P}_1 \cap \mathbf{Q}_1$ is equal to $\mathbf{P}_0$. An analogous result is also true in the theory of projective sets and is proved there by means of a theorem which states that two disjoint A-sets can be separated by a pair of disjoint Borel-sets. Now Kleene [4] has shown that disjoint $\mathbf{P}_1$-sets cannot, in general, be separated by means of $\mathbf{P}_0$-sets. Other important examples of non-separable $\mathbf{P}_1$-sets were given by Trachténbrot and Uspénskij has applied them to an analysis of incompleteness theorem of Gödel (cf. [18] and [8]). We shall give below other applications of non-separable $\mathbf{P}_1$-sets (cf. theorem XVII and the construction described on p. 5).

It is easy to generalize Kleene's construction to higher classes:

VIII. *For every* $n \geqq 1$ *there exist disjoint sets* X, $Y \in \mathbf{P}_n$ *such that if* A, $B \in \mathbf{P}_n \cap \mathbf{Q}_n$, $X \subset A$, $Y \subset B$, then $A \cap B \neq 0$.

Thus the separation-theorem does not hold for sets of the class $\mathbf{P}_n$, $n \geqq 1$. It is curious that it does hold for sets of class $\mathbf{Q}_n$:

IX. *If* X, $Y \in \mathbf{Q}_n$, $X \cap Y = 0$, *then there are disjoint sets* A, $B \in \mathbf{P}_n \cap \mathbf{Q}_n$ *such that* $X \subset A$ *and* $Y \subset B$.

It follows in particular from this theorem that disjoint $\mathbf{Q}_1$-sets are always separable by means of recursive sets.

3. *The logical form of some definitions from analysis.* A number of interesting problems are suggested by the study of definitions of fundamental notions of analysis.

Let $f(x, y)$ be a function of two numerical arguments with numerical values and let $\mathfrak{X}$ be a class of real numbers. We denote by $Z(\mathfrak{X}, f)$ the set of integers y such that $\lim_{x \to \infty} 10^{-x} f(x, y)$ exists and is in $\mathfrak{X}$. We assume that f runs over the class of general (or primitive) recursive functions (or a subclass thereof) and ask what is the class of the sets $Z(\mathfrak{X}, f)$.

In order to express briefly our theorems we denote by $\mathfrak{X}_0$ the class of all non-negative real numbers, by $\mathfrak{X}_1$ the class with the unique element 0, by $\mathfrak{X}_2$ the class of non-negative integers, by $\mathfrak{X}_3$ the class of non-negative irrational

numbers, and by $\mathfrak{X}_4$ the class of non-negative rational numbers. We have then the following theorems:

X. *If f runs over the class $\mathbf{P}_0$ (or over the class of primitive recursive functions), then the sets $Z(\mathfrak{X}_i, f)$ $(i = 0, 1, 2, 3)$ coincide with the $\mathbf{Q}_3$-sets and the sets $Z(\mathfrak{X}_4, f)$ coincide with the differences $X - Y$ where $X, Y \in \mathbf{Q}_3$.*

The theorem remains true if we take instead of $\mathfrak{X}_1$ a class which contains exactly one number having the form $\lim [g(n)/h(n)]$ where g and h are general (or primitive) recursive functions and $\lim g(n) = \lim h(n) = \infty$.

XI. *If f runs over the class of general (or primitive) recursive functions such that $\lim_x 10^{-x} f(x, y)$ exists for every y, then the sets $Z(\mathfrak{X}_1, f)$ and the sets $Z(\mathfrak{X}_2, f)$ coincide with the $\mathbf{Q}_2$-sets; the sets $Z(\mathfrak{X}_3, f)$ coincide with the $\mathbf{Q}_3$-sets, and the sets $Z(\mathfrak{X}_4, f)$ coincide with the $\mathbf{P}_3$-sets.*

The comparison of theorems X and XI shows that if we restrict the variability of f to the narrower class described in theorem XI we reduce the classes to which belong the sets $Z(\mathfrak{X}_1, f)$, $Z(\mathfrak{X}_2, f)$, and $Z(\mathfrak{X}_4, f)$; on the contrary the class to which belong the sets $Z(\mathfrak{X}_3, f)$ remains unaltered.

XII. *If f runs over the class of general (or primitive) recursive functions such that $\lim_x 10^{-x} f(x, y) = \lambda(y)$ exists for every y and the set $\{\lambda(0), \lambda(1), \ldots\}$ is finite, then the sets $Z(\mathfrak{X}_1, f)$ coincide with the $\mathbf{P}_2 \cap \mathbf{Q}_2$-sets.*

Instead of representing real numbers in the form $\lim 10^{-x} f(x, y)$ we can start with an arbitrary other limiting process, e.g.,

$$\sigma(f, y) = \sum_{x=0}^{\infty} \left(10^{-x}(10 \dot{-} f(x, y))\right)$$

and ask about the exact class to which belong the sets $Z'(\mathfrak{X}, f) = \{y : \sigma(f, y) \in \mathfrak{X}\}$.

XIII. *If f runs over the class of general (or primitive) recursive functions, then the sets $Z'(\mathfrak{X}_1, f)$ and the sets $Z'(\mathfrak{X}_2, f)$ coincide with the $\mathbf{Q}_1$-sets, the sets $Z'(\mathfrak{X}_3, f)$ coincide with the $\mathbf{Q}_2$-sets, and the sets $Z'(\mathfrak{X}_4, f)$ coincide with the $\mathbf{P}_2$-sets.*

Theorems X—XIII were obtained with the help of the following two lemmas which may prove useful also in other investigations of this type (the first of these lemmas was also obtained independently by Markwald [10]).

XIV. *A set Z belongs to the class $\mathbf{P}_2$ if and only if there is a primitive recursive function $f(x, y, z)$ such that $z \in Z \equiv (\exists x)(y)[f(x, y, z) = 0]$.*

XV. *A set Z belongs to the class $\mathbf{P}_3$ if and only if there is a primitive recursive function $f(x, y, z)$ such that $z \in Z \equiv (\exists x, a)(u)(\exists y)\{[f(x, y, z) = a](y > u)\}$.*

Theorems X—XIII and possibly some other theorems of a similar type can be applied along with the well-known method due to Kuratowski and Tarski [7] to proofs that it is impossible to simplify a given mathematical definition. I shall consider only one example.

A sequence $\{a_n\}$ of real numbers is convergent if and only if

$$(*) \qquad (n)(\exists k)(l_1, l_2)\ \{[(l_1 > k).\ (l_2 > k)] \supset (|a_{l_1} - a_{l_2}| < 1/n)\}.$$

Is it possible to simplify this definition so as to use a smaller number of quantifiers?

To answer this question let us call a real-valued function φ of n real variables recursive if there are two recursive functions f, g with integral values and with $2n$ integral arguments such that

$$\varphi(p_1/q_1, \ldots, p_n/q_n) = f(p_1, \ldots, p_n, q_1, \ldots, q_n)/g(p_1, \ldots, p_n, q_1, \ldots, q_n)$$

for arbitrary $p_1, \ldots, p_n, q_1, \ldots, q_n$. We have then the following theorem (which results easily from theorem X):

XVI. *If F is an expression (with $p + q + r$ free variables) built up from symbols of real recursive functions, the identity sign, and the propositional connectives, then the class of sequences $\{a_n\}$ which satisfy the formula*

$$(*_*) \qquad (\exists n_1, \ldots, n_p)(k_1, \ldots, k_q)(\exists l_1, \ldots, l_r)\ F(a_{n_1}, \ldots, a_{n_p}, a_{k_1}, \ldots, a_{k_q}, a_{l_1}, \ldots, a_{l_r})$$

does not coincide with the class of convergent sequences.

In the case considered in the theorem XVI the method of Kuratowski and Tarski based on the use of continuous variables yields a better result insofar as it shows the impossibility of reducing $(*)$ to the form $(*_*)$ for more general expressions F, in fact for expressions built up from symbols for quite arbitrary functions. It is an interesting problem to examine whether the method of Kuratowski—Tarski always yields better results than the method based on the theory of recursive functions.

I shall mention still one result concerning limits of fractions $10^{-x} f(x, y)$ with recursive f. According to S. Mazur [12] we call a sequence $\{a_n\}$ of real numbers recursive if there is a general recursive function f such that $|a_n - 10^{-k} f(k, n)| < 10^{-k+1}$ for arbitrary k and n.

XVII. *There is a recursive sequence $\{a_n\}$ such that if $a_n = \sum_{k=0}^{\infty} 10^{-k} g(k, n)$, then the function g is not recursive.*

I give below the construction of the sequence $\{a_n\}$ in order to indicate the role played in it by the non-separable $\mathbf{P}_1$-sets.

Let X_1, X_2 be two such sets. One shows easily that there exist general recursive functions h_1, h_2 such that, for arbitrary x, y, and for $i = 1, 2$,

$$0 \leqq h_i(y, x) \leqq h_i(y + 1, x) \leqq 1,$$
$$x \in X_i \equiv (\exists z)\{[h_i(z, x) = 1](z \geqq 1)\}.$$

Let $p_i(y, x) = (\mu z)_{1 \leqq z \leqq y}[h_i(z, x) = 1]$. Then the required function f is defined as follows:

$$f(n, k) = (10^n - 1) + h_1(n, k)[1 + 10^{n - p_1(n,k)}] + h_2(n, k)[1 - 10^{n - p_2(n,k)}].$$

4. *Applications to the completeness-theorem.* The classification of sets into classes $\mathbf{P}_n$ and $\mathbf{Q}_n$ has found important applications in the study of different classes of formulas (which for the sake of this study can conveniently be identified with their Gödel-numbers). It were in fact these applications which suggested the definitions of classes $\mathbf{P}_n$ and $\mathbf{Q}_n$.

An analysis of Gödel's incompleteness-theorem along these lines was carried out in previous publications (cf. [3], [13]). I shall now describe some applications to the completeness theorem of Gödel.

A well-known corollary to this theorem states that every consistent set of first-order axioms possesses an arithmetical model, i.e., a model consisting of relations defined in the set of integers. Since an n-termed relation can be identified with a set of n-tuples $\mathfrak{x} = (x_1, x_2, \ldots, x_n)$, we can also say that an arithmetical model consists of sets. If all these sets belong to the class $\mathbf{P}_n$ or $\mathbf{Q}_n$, then we say that the model belongs to this class.

Kleene [5] and Hasenjaeger [2] have shown that a consistent set of axioms possesses always a model of class $\mathbf{P}_2 \cap \mathbf{Q}_2$. Examples of axioms which have no model in $\mathbf{P}_1 \cup \mathbf{Q}_1$ were found by Kreisel [6] and myself [14]. I shall describe here another simplified set of axioms with this property.

Let G be a free semigroup (with cancellation) with the generators a, b, c. Words not containing c will be called proper. The word $a \,.\, a \ldots a$ (n-times) will be identified with the integer n.

Let $B = \{\alpha, \gamma_1, \gamma_1', \ldots, \gamma_k, \gamma_k'\}$ be a sequence of $2k + 1$ proper words. A word $\xi = c\alpha_1 c\alpha_2 c\alpha_3 \ldots c\alpha_l c$ is called B-generating if $\alpha_1, \alpha_2, \ldots, \alpha_l$ are proper words and each α_j is either identic with α or arises from an α_i, $i < j$, by a "production" of the form $\alpha_i = \gamma_s\beta \to \beta\gamma_s' = \alpha_j$. The word α_l is called the last ingredient of ξ.

For a given B we denote by S_B the set of integers n such that there is a B-generating word ξ with the last ingredient n. It is known from the works of Post [16] that each $\mathbf{P}_1$-set can be represented in the form S_B for a suitable B. Let $X_i = S_{B_i}$ $(i = 1, 2)$ be two non-separable $\mathbf{P}_1$-sets.

We consider now a set of first-order axioms with the undefined terms

$A(x)$, $B(x)$, $C(x)$, $D_1(x)$, $D_2(x)$, $E(x, y)$, $F(x, y, z)$. The intended interpretation of the axioms is as follows: The universe of discourse is the semigroup G described above; $E(x, y)$ is the identity-relation between the elements of G, and $F(x, y, z)$ is the relation: x is the result of multiplication of y and z; $A(x)$, $B(x)$, $C(x)$ are classes consisting of a, b, c alone; finally D_1 is the class X_1 and D_2 is the complement of X_1.

The axioms fall in two parts: the first part contains axioms stating the general properties of semigroups and the axiom stating that if $\alpha\beta = \gamma\delta$, then either there is a ξ such that $\alpha = \gamma\xi$ and $\delta = \xi\beta$ or there is an η such that $\gamma = \alpha\eta$ and $\beta = \eta\delta$. The axioms of the second group state that each integer belongs either to D_1 or to D_2 but not to D_1 and D_2 simultaneously and further that if α is an integer which is the last ingredient of a B_i-generating word, then α is in D_i, $i = 1, 2$.

The consistency of these axioms is evident. It is also easy to see that if they had a $\mathbf{P}_1$-model, sets X_1 and X_2 would be separable. By adjoining the new primitive terms D_i' together with the axioms $(x)[D_i'(x) \equiv \sim D_i(x)]$, $i = 1,2$, we obtain a set of axioms no model of which is in $\mathbf{P}_1 \cup \mathbf{Q}_1$.

It is not yet known whether there exist axioms no model of which would belong to the smallest field F of sets generated by $\mathbf{P}_1 \cup \mathbf{Q}_1$. This problem seems to be connected with the problem whether there is a consistent and complete extension of the axiomatic system of arithmetic of integers such that the set of its Gödel-numbers belongs to F. It is known from Rosser's work [17] that no such set belongs to $\mathbf{P}_1 \cup \mathbf{Q}_1$ whereas it is easy to show that there are such sets in $\mathbf{P}_2 \cap \mathbf{Q}_2$.

Further interesting problems are connected with a generalization of logical quantifiers. The ordinary quantifiers (x), $(\exists x)$ are operators which correlate one of the truth-values T, F with each propositional function of one argument. Let us consider quite generally an infinite set J and operators Q such such that if φ is a subset of J (i.e. a propositional function of one argument defined in J), then $Q\varphi$ is either T or F. An operator Q will be called a quantifier limited to J if it satisfies the invariance condition: $Q\varphi = Q\psi$ for every one-one mapping π of J onto itself and every propositional functions $\varphi(x)$, $\psi(x)$ satisfying the equivalence $(x)\,[\varphi(x) \equiv \psi(\pi(x))]$. (Since we wish to treat quantifiers as logical operators, it seems natural to assume that they satisfy the invariance condition; cf. Mautner [11]).

Many problems of mathematical logic pertaining to quantifiers can be reformulated in such a way that they remain meaningful if quantifiers are taken in the general meaning of this word explained above. We illustrate this on the example of the completeness-problem.

We consider expressions built up from individual variables, predicate

variables, propositional connectives, the sign of identity, quantifiers (x), $(\exists x)$, and symbols for other quantifiers $Q_1, Q_2, \ldots, Q_k$ limited to J. Let Z be the set of (Gödel-numbers of) expressions which are true in J. The generalized completeness-problem for quantifiers $Q_1, Q_2, \ldots, Q_k$ is this: for which quantifiers $Q_1, Q_2, \ldots, Q_k$ is Z a $\mathbf{P}_1$-set?

XVIII. In case when the set J is denumerable we have the following theorem. *Z is a $\mathbf{P}_1$-set if and only if the quantifiers $Q_1, Q_2, \ldots, Q_k$ are definable in the predicate-calculus with identity. If at least one quantifier Q_i is not definable in the predicate calculus with identity, then the set Z does not belong to any class $\mathbf{P}_n$.*

In addition to this theorem we give a complete description of all those quantifiers limited to an arbitrary infinite set J which are definable in the predicate calculus with identity.

Let $n_1, n_2, \ldots, n_q$ be non-negative integers and define the quantifiers $Q^*_{J, n_1, \ldots, n_q}$, $Q^{**}_{J, n_1, \ldots, n_q}$ (limited to J) as follows:

$\{Q^*_{J, n_1, \ldots, n_q} \varphi = T\} \equiv$ {the cardinal number of φ is either n_1 or $n_2 \ldots$ or n_q},

$\{Q^{**}_{J, n_1, \ldots, n_q} \varphi = T\} \equiv$ {the cardinal number of J—φ is either n_1 or $n_2 \ldots$ or n_q}.

XIX. *If I is an arbitrary infinite set, then $Q^*_{J, n_1, \ldots, n_q}$ and $Q^{**}_{J, n_1, \ldots, n_q}$ (where $n_1, \ldots, n_q$ are arbitrary non-negative integers are the only non-constant quantifiers limited to J which are definable in the predicate calculus with identity.*

It follows from XIX that neither of the quantifiers

$$\{Q^s_J \varphi = T\} \equiv \{\text{the set } \varphi \text{ is finite}\}$$

$$\{Q^p_J \varphi = T\} \equiv \{\text{the set } \varphi \text{ is at most denumerable}\}$$

is definable in the predicate calculus with identity, whereas from XVIII it follows that if J is denumerable, then the completeness problem for the quantifier Q^s_J has a negative solution. (The quantifier Q^p_J is constant if J is denumerable).

In case of a non-denumerable J the completeness-problem becomes much more difficult. I was unable to establish the class of the set Z even in the case $Q_1 = Q_2 = \ldots = Q_k = Q^p_J$. The only results I know of in this direction is the (trivial) theorem XX and the (slightly less trivial) theorem XXI given below:

XX. *If J is an arbitrary infinite set and Q^s_J occurs among the quantifiers $Q_1, \ldots, Q_k$, then set Z does not belong to any $\mathbf{P}_n$.*

XXI. *The quantifier Q^s_J is not definable by means of the quantifier Q^p_J and notions occurring in the ordinary predicate calculus with identity.*

To conclude this section I mention still one problem: It is well-known that if V is the set of (Gödel-numbers of) formulas of the predicate calculus which

are valid in the set of all integers, then the definition of V has the form

$$(*_*^*) \quad n \in V \equiv (X)\, F\,(X, n)$$

where F is a formula built up from symbols of recursive functions and predicates and expressions of the form $y \in X$ by means of propositional connectives and quantifiers limited to the set of integers. The variable X runs here over the set of arbitrary sets of integers. Thus $(*_*^*)$ is an example of what is called a non-elementary definition. Yet we know from the completeness theorem that V is a $\mathbf{P}_1$-set. We are thus led to the problem: what are the necessary and sufficient conditions for F in order that the set V defined by $(*_*^*)$ be recursively enumerable?

5. *Classes* $\mathbf{P}_\alpha$ *and* $\mathbf{Q}_\alpha$ *for transfinite values of* α. Many problems of logic lead to sets of formulas which do not belong to any $\mathbf{P}_n$ (cf. theorem XVIII). It is therefore desirable to extend our classification so as to include as many sets as possible.

One such extension was indicated in my paper [15] where I defined classes $\mathbf{P}_\alpha$ and $\mathbf{Q}_\alpha$ for constructive ordinals α. I accepted for $\mathbf{P}_{\alpha+1}$ and $\mathbf{Q}_{\alpha+1}$ the same definitions in terms of $\mathbf{P}_\alpha$ and $\mathbf{Q}_\alpha$ as for $\alpha < \omega$. For $\alpha = \lim \alpha_n$ I took as $\mathbf{P}_\alpha$ the class of the unions $\bigcup_n F_n(f(n))$ and as $\mathbf{Q}_\alpha$ the class of the intersections $\bigcap_n F_n(f(n))$ where F_n is a universal function for the class $\mathbf{P}_{\alpha_n}$ and f is a recursive function. This definition is meaningful only if we determine a law according to which a sequence α_n is correlated with α. It is known that such a law exists for constructive ordinals.

It seems to me now that it is more natural to adopt another definition modelled on the notion of constructible set (Gödel [1]).

Let $F(\mathfrak{x})$ be a formula with n free variables built up from symbols of recursive predicates, recursive functions, expressions of the form $z_i \in X_j$, logical connectives, and quantifiers. Let α be a limit-number and assume that classes $\mathbf{P}_\xi$, $\mathbf{Q}_\xi$, $\xi < \alpha$, have already been defined. We define $\mathbf{P}_\alpha$ as the class of all sets having the form $\{\mathfrak{x}: F_\alpha(\mathfrak{x})\}$ where $F_\alpha(\mathfrak{x})$ is got from $F(\mathfrak{x})$ by the relativization of the quantifiers (X_j) and $(\exists X_j)$ to the class $\bigcup_{\xi<\alpha} \mathbf{P}_\xi$. For an arbitrary β we define $\mathbf{Q}_\beta$ as the class of the complements of $\mathbf{P}_\beta$-sets and $\mathbf{P}_{\beta+1}$ as the class of the projections of $\mathbf{Q}_\beta$-sets.

Classes $\mathbf{P}_\alpha$ and $\mathbf{Q}_\alpha$ are thus defined for arbitrary ordinals α. According to a result of Gödel there is an ordinal $\alpha_0 \leqq \Omega$ such that $\mathbf{P}_{\alpha_0} = \mathbf{P}_\xi = \mathbf{Q}_\xi$ for $\xi \geqq \alpha_0$; it is therefore sufficient to consider only the classes $\mathbf{P}_\alpha$ and $\mathbf{Q}_\alpha$ with $\alpha \leqq \Omega$.

The transfinite classification gives rise to many open and difficult problems. It would for instance be very interesting to know which place in this classification occupy the set of true sentences of arithmetic of real numbers and the sets Z defined on p. 7.

The foregoing survey of results connected with the classification of sets of integers into the $\mathbf{P}_n$- and $\mathbf{Q}_n$-sets shows that we may expect almost always to face an interesting problem when we try to determine the place which any individually given set of integers occupies in this classification.

LITERATURE

[1] K. GÖDEL, Consistency of the axiom of choice and of the generalized continuum-hypothesis with the axioms of set-theory. Princeton 1940.

[2] G. HASENJAEGER, Eine Bemerkung zu Henkin's Beweis für die Vollständigkeit des Prädikatenkalküls der ersten Stufe. Journal of Symbolic Logic 18 (1953), pp. 42-48.

[3] S. C. KLEENE, Recursive predicates and quantifiers. Transactions of the American Mathematical Society 53 (1943), pp. 41-73.

[4] S. C. KLEENE, A symmetric form of Gödel's theorem. Indagationes Mathematicae 12 (1950), pp. 244-246.

[5] S. C. KLEENE, Introduction to metamathematics, Groningen 1952.

[6] G. KREISEL, Note on arithmetic models for consistent formulae of the predicate calculus II. Proceedings of the XIth International Congress of Philosophy, vol 14. Bruxelles and Amsterdam 1953, pp. 39-49.

[7] K. KURATOWSKI and A. TARSKI, Les opérations logiques et les ensembles projectifs. Fundamenta Mathematicae 17 (1931), pp. 240-248.

[8] V. A. USPÉNSKIJ, Téoréma Gedela i téorija algorifmov. Doklady Akademii Nauk SSSR 91 (1953), pp. 737-740.

[9] W. MARKWALD, Zur Theorie der konstruktiven Wohlordnungen. Mathematische Annalen 127 (1954), pp. 135-149.

[10] W. MARKWALD, to appear in Fundamenta Mathematicae 42.

[11] F. J. MAUTNER, An extension of Klein's Erlanger Program: logic as invariant-theory. American journal of mathematics 68 (1946), pp. 345-384.

[12] S. MAZUR, Computable analysis. Rozprawy Matematyczne (to appear).

[13] A. MOSTOWSKI, On definable sets of positive integers. Fundamenta Mathematicae 34 (1947), pp. 81-112.

[14] A. MOSTOWSKI, On a system of axioms which has no recursively enumerable arithmetic model. Fundamenta Mathematicae 40 (1953), pp. 58-61.

[15] A. MOSTOWSKI, A classification of logical systems. Studia Philosophia, Poznan, vol. 4 (1951), pp. 237-274.

[16] E. L. POST, Formal reductions of the general combinatorial decision problem. American Journal of Mathematics 65 (1943), pp. 197-215.

[17] J. B. ROSSER, Extensions of some theorems of Gödel and Church. Journal of Symbolic Logic 1 (1936), pp. 87-91.

[18] B. A. TRACHTÉNBROT, O rékursivnoj otdélimosti, Doklady Akadémii Nauk SSSR, 88 (1953), pp. 569-572.

THE UNIVERSITY OF WARSAW.

THE RELATIVE STRENGTH OF ZERMELO'S SET THEORY AND QUINE'S NEW FOUNDATIONS

J. BARKLEY ROSSER

The present note is a summary of some of the results that have so far been obtained in an investigation that is still proceeding. Many of the technical points involved are as yet in the process of being published, having appeared only in the unpublished doctoral theses of Firestone and Orey (see [1] and [2]).

We shall refer here generically to any of the considerable variety of systems that stem from the original Zermelo concept as a Zermelo set theory (we shall use merely "*ZST*" henceforth). In [3], Chap. III and Chap. IV, is given a survey of the better known *ZST*'s, with references to the literature. We shall focus attention on three *ZST*'s as being sufficiently representative for the present summary, and shall present them in forms convenient for the present exposition without regard whether these are their familiar forms.

The term "Quine's new foundations" (we shall use merely "*NF*" henceforth) usually refers to a specific system, namely that proposed in [5] and expounded at length in [6]. However, one consequence of the comparison of *NF* with the *ZST*'s is the discovery of a variety of systems which parallel the main kinds of *ZST*'s while retaining the characteristic features of *NF*. Specifically, corresponding to each of the forms of *ZST* which we shall consider, there is a form of *NF*; these forms of *NF* differ among themselves in a way quite analogous to that of their corresponding *ZST*'s; also, in each of the *NF's* one can form a model of the corresponding *ZST*, so that if some of these *NF's* are consistent, then so are their corresponding *ZST's*. The study (if any) of relative consistency proofs in the reverse direction has not met with any success. Most likely, we do not yet have the best forms of *NF* corresponding to the various forms of *ZST*, and a search is continuing for more satisfactory forms of *NF*.

In some forms of *ZST*, there are two categories of classes, those which are also members (that is, the "sets") and those which are not (see [4]). Wang (see [7]) has pointed out that Quine's Mathematical Logic (see [8]) can be considered as consisting of *NF* with the addition of classes that are non-members. This suggests that *ZST's* with two categories of classes should be compared with Mathematical Logic, rather than with *NF*. In [2], such a comparison is carried out, and proceeds quite analogously to the simpler comparison of a single-category *ZST* with *NF*. Accordingly, we shall here restrict

attention to *ZST's* in which all classes are also members. The question of finding the appropriate forms of *NF* to compare with *ZST's* containing individuals has not been considered.

We herewith consider the most uncomplicated version of *ZST*, in which there is a single category of object, namely that of sets, and a single predicate, namely that of class membership. There are roughly two sorts of axioms: those that specify the existence of certain particular sets, and those that permit the generation of new sets from those already given. Seeking convenience rather than elegance, we list the primary generation axioms:

The pair axiom: Given two sets, α and β, there exists a set of which α and β are the sole members. (It is not required that α and β be distinct).

The sum axiom: Given a set α, there exists the set consisting of the sum of all the members of α.

The power axiom: Given a set α, there exists the set consisting of all the subclasses of α.

The subclass axiom: Given a set α and a statement $F(x)$, there exists the subset of α consisting of all x's in α which satisfy $F(x)$.

Since each generation axiom merely produces a new set from some given set (or sets) these axioms alone do not suffice to prove the existence of any set, and they must be supplemented by some axiom that explicitly states the existence of some specific set (from which one then derives many other sets by the generation axioms). By taking different existence axioms, one can get different *ZST's*.

One possibility is to assume an axiom stating the existence of the null set. Then the stated generation axioms suffice to generate only sets of finite cardinality. Consequently the resulting *ZST* is of limited interest. A much more popular course is to assume an axiom stating the existence of what is in effect the set of non-negative integers. By repeatedly applying the power-class axiom, one can generate classes of high enough cardinality for many mathematical purposes. However, one cannot infer the existence of a class with the cardinality corresponding to an infinite number of applications of the power-class axiom to the set of non-negative integers.

If one wishes to deal with very high cardinalities, one can either assume an axiom explicitly stating the existence of some set of adequately high cardinality or one can introduce a new generation axiom, namely:

The replacement axiom: Given a set α and a statement $F(x, y)$ such that for each x in α there is at most one y such that $F(x, y)$ is satisfied, there exists the set of all y's for which there is an x in α such that $F(x, y)$ is satisfied.

With this added generation axiom and the assumption of the existence of the set of non-negative integers, one can develop a full-fledged theory of

cardinals and ordinals. One could get still stronger systems, or variations o those already indicated, but these suffice for our present purpose. Indeed, let us denote by S_1, S_2, and S_3 the three following systems. For S_1, we use the four primary generation axioms and assume the existence of the null set. For S_2, we add to S_1 the assumption of the existence of the set of non-negative integers. For S_3, we add to S_2 the replacement axiom.

For NF, there is also a single category of object, namely that of sets which are also members, and a single predicate, namely that of class membership. Sets are generated by the single principle

$$(P) \qquad (Ey)(x)(x\varepsilon y \equiv F(x)),$$

where $F(x)$ is required to be stratified. For the technical definition of "stratified," see [5] or [6]. However, in effect, it signifies that one can assign types to the variables of $F(x)$ in such a way that $F(x)$ would become a legitimate formula of simple type theory. One can replace the single principle (P) by an existence axiom and a set of generation axioms having no reference to stratification (see [9]), but these lack the naturalness of the Zermelo formulation, and it is the principle (P) which is really characteristic for NF.

By letting $F(x)$ contain the names of sets already given, one can use the principle (P) to generate new sets from given ones. Alternatively, one can use the principle (P) to prove the existence of various specified sets. Thus, by (P) one easily infers the existence of the null set, and one can even infer the existence of the set of non-negative integers (see [10]).

If one tries to identify the sets of NF as sets of a ZST in a straightforward way, one immediately runs into difficulty. For example, in NF there is a universal set. Not only is there no universal set for a ZST, but the universal set of NF fails to have a number of properties that sets are expected to have in mathematics and which sets do have in the various ZST's. We indicate three such properties.

(1) The power class of a set α (that is, the set of all subsets of α) should have higher cardinality than α. However, the power class of the universal set of NF is just the universal set itself.

(2) A set α should have the same cardinality as the set of all its unit subsets. Indeed, intuitively one sets up the one-to-one correspondence by letting each member x of α correspond to the unit subset containing just x. However, this correspondence is not stratified, and so one is not entitled to infer its existence by invoking principle (P). Moreover, in NF, one can prove that the universal set has higher cardinality than the set of all its unit subsets.

(3) It is very popular among mathematicians to assume that every set can be well ordered. One can add such an assumption to a typical ZST without risk

(see [4]). However, Specker has shown that it is a theorem of NF that the universal set cannot be well ordered (see [10]).

One interpretation for this state of affairs is that the principle (P) is too generous, and supplys us with various pathological sets as well as with various sets that we wish. Accordingly we seek a criterion for distinguishing the pathological sets. Then we can seek a correspondence between the non-pathological sets of NF and the sets of a ZST.

We shall indicate one such possible criterion, and summarize some results of adopting it. We wish to emphasize that there are other possible criteria, and we invite attention to the problem of devising such other criteria. Much remains to be learned about the present criterion, and it is too early to decide whether it is wholly adequate.

To indicate the criterion that has been studied so far, let us follow [6] and say that a set α is strictly Cantorian if α can be put into one-to-one correspondence with its set of unit subsets in such a way that each member x of α is made to correspond with the unit set which contains just x. Let us agree to consider a set as pathological if it is not strictly Cantorian. If attention be restricted to strictly Cantorian sets, one gets in NF a theory of cardinals and ordinals that follows the classical theory much more satisfactorily than when one allows sets to appear that are not strictly Cantorian.

We now seek a correspondence between the sets of some ZST and the strictly Cantorian sets of NF. Thus we exclude from consideration such pathological sets as the universal set of NF. However, there still arise difficulties. For strictly Cantorian sets the power axiom holds (see [6], Thm. XIII.2.7), but the sum axiom fails to hold, as one can see by taking α to be the unit set whose sole member is the universal set. It thus appears that if we wish a strict correspondence, we must use same stronger criterion than that of being strictly Cantorian. So far, attention has been focussed on an alternative approach.

In seeking a correspondence between sets of a ZST and some collection of sets of NF, let us abandon the requirement that the relation of class membership in the ZST must correspond to the relation of class membership in NF. If this is done, it is possible to get a correspondence between the sets of S_1 and some (or all) of the non-negative integers of NF. This is done as follows. One can enumerate all finite sets of non-negative integers (see [6], Thm. XI.4.18). Let $\varnothing_0, \varnothing_1, \ldots$ be the finite sets of non-negative integers. Define ε_1 by saying that for non-negative integers m and n, $m\varepsilon_1 n$ holds if and only if m is a member of $\varnothing_n$. Then with non-negative integers playing the role of sets and ε_1 playing the role of class membership, the axioms of S_1 are readily verified as theorems of NF. Thus the sets whose existence is guaranteed in S_1 must correspond to some (or all) of the non-negative integers in NF.

To model S_2 or S_3 in NF in an analogous manner certainly requires that one be able to discover in NF sets of high cardinalities. Accordingly it seems likely that we must add to NF axioms insuring the existence of suitable sets of high cardinalities before we can model S_2 or S_3. However, we may as well make use of a device for keeping the cardinalities involved relatively low.

In [4], a model Δ is set up for a certain strong ZST. Suitable portions of Δ can be used as models for S_2 and S_3. As Δ is well ordered, we can readily set up a model Δ_0 in the ordinal numbers which is isomorphic (in a loose but obvious sense) to Δ. By using the relation ε_2 between ordinals, that serves in Δ_0 as the image of class membership in Δ, we can get models for S_2 and S_3 by using suitable fragments of Δ_0. We observe that by paralleling the definition of Δ, we can make a definition of ε_2 by transfinite induction over the ordinals. Thus we can get models for S_2 and S_3 in any formal logic which:

(a) contains a sufficiently developed theory of ordinals to carry out the definition of ε_2 by transfinite induction;

(b) contains sufficiently large segments of the ordinals so that adequately high cardinalities are available.

Let us first consider point (a) in NF. If we restrict attention to the ordinals of those well ordered sets which are strictly Cantorian, we get a fairly decent ordinal theory. In [2], Orey has shown that by adding a special axiom dealing with transfinite induction for such ordinals, one gets the full classical theory of ordinals in its familiar form. However, this axiom is almost certainly stronger than is needed for the definition of ε_2. More likely, we can define ε_2 without addition of any extra axioms to NF, but this point needs further study before a firm statement is justified.

Suppose we have added to NF an axiom (or perhaps no axiom) sufficient for the definition of ε_2. Then let us consider point (b). Here is where we find a strong analogy with the ZST's. To get S_2 from S_1, we added the assumption that there is an infinite set. Correspondingly, to extend NF (supposed adequate for ordinal theory) enough to have a model for S_2, we add the assumption that there is an infinite strictly Cantorian set. With NF extended as indicated above to contain a model of S_2, we can insure a model for S_3 by adding an axiom identical in wording with the replacement axiom except that α is required to be strictly Cantorian.

Looking at S_2 and S_3 as successive extensions of S_1 got by adding specified axioms, we see that exactly analogous axioms dealing with strictly Cantorian sets of NF will suffice to enlarge a suitable ordinal theory in NF to provide models for S_2 and S_3.

Actually, the version of the replacement axiom indicated for use in NF is stronger than necessary. Since S_3 is to be modelled in the ordinals, it suffices in

the replacement axiom to require that α be a strictly Cantorian class of ordinals. This indeed can be still further weakened (see [2]). Probably the weakest axiom that would suffice would be a simple assumption that there is a well-ordered strictly Cantorian set of sufficiently high cardinality.

Probably one cannot get a model of S_2 with any weaker assumption than that there is an infinite set which is strictly Cantorian. However, this assumption can be given various alternative forms. For example, it is equivalent to the axiom of counting (see [6], p. 485).

To summarize, it appears that NF is considerably stronger than the weak ZST which we have called S_1. However, the extra strength of NF appears in its ability to generate sets like the universal set, which are pathological, though useful and interesting. If we restrict attention to the non-pathological sets of NF, then we find NF to be about on a par with S_1. If we strengthen S_1 by adding an axiom of infinity to get S_2, then we can keep NF at least on a par by adding to it an axiom of infinity for non-pathological sets. Moreover, the same situation holds for the axiom of replacement. This does highlight the fact (which is sometimes discounted) that it is primarily the presence of the strong axiom of replacement that makes S_3 a very powerful system of logic.

References

[1] C. D. Firestone, Sufficient conditions for the modelling of axiomatic set theory. Doctoral Thesis at Cornell University, September, 1947.

[2] Steven Orey, Formal development of ordinal number theory and applications to consistency proofs. Doctoral Thesis at Cornell University, September, 1953.

[3] H. Wang and R. McNaughton, Les systèmes axiomatiques de la théorie des ensembles. Gauthier-Villars, Paris, 1953.

[4] Kurt Gödel, The consistency of the axiom of choice and of the generalized continuum-hypothesis with the axioms of set theory. Princeton University Press, Princeton, 1940.

[5] W. V. Quine, New foundations for mathematical logic. The American Mathematical Monthly, vol. 44 (1937), pp. 70-80.

[6] J. Barkley Rosser, Logic for mathematicians. McGraw-Hill Book Co., New York, 1953.

[7] Hao Wang, A formal system of logic. The Journal of Symbolic Logic, vol. 15 (1950), pp. 25-32.

[8] W. V. Quine, Mathematical Logic. Third edition. Harvard University Press, Cambridge 1951.

[9] Theodore Hailperin, A set of axioms for logic. The Journal of Symbolic Logic, vol. 9 (1944), pp. 1-19.

[10] E. P. Specker, The axiom of choice in Quine's New Foundations for mathematical logic. Proceedings of the National Academy of Sciences, vol. 39 (1953), p. 972-974.

Cornell University.

SECTION VII

PHILOSOPHY, HISTORY AND EDUCATION

STATED ADDRESSES

SELF-EDUCATION BY CHILDREN IN MATHEMATICS USING GESTALT METHODS - I.E. LEARNING-THROUGH-INSIGHT

C. T. DALTRY

Mathematicians and Teachers

I suppose that this Congress has two aims — to create mathematics and to *communicate* mathematics, new and old. The first aim concerns the creative mathematicians, the second the creative teachers of that vast and growing subject. In this talk I hope to persuade members of my audience whether they regard themselves as mathematicians rather than teachers, or the other way about, that these two aims are one: that to communicate mathematics most effectively the teacher should show the mind of the creative mathematician at work.

At a time when more and more people need mathematics it is necessary to consider the efficient communication of the subject and particularly the causes of failure to learn mathematics. As a teacher of mathematics I have always been more concerned with my failures than with my successes: failures are a challenge to a good teacher, as to a good doctor. I suggest that the clue to good teaching of mathematics lies in the study of failure to learn mathematics.

A Fundamental Conflict

It is natural to assume that because the final form of a piece of mathematics is an ordered structure built up brick upon brick that it must be communicated to human beings in the same logical fashion. This assumption ignores the nature of the learner, in particular his whole emotional response to mathematics and to his teacher. Everything we have discovered about the art of teaching suggests that the overriding consideration is the formation of a close personal relationship between pupil and teacher. For this, teacher and pupil must engage on a joint enterprise. This cannot arise, with non-mathematical pupils, in a situation in which the teacher has to impose mathematics from without. There is a permanent, inherent tension or conflict in the teaching of mathematics which does not occur in the teaching of any other subject — a tension between logical and psychological. To resolve this tension the teacher needs to study the pupil's attitude to mathematics, in the hope that failure will be replaced by success and confidence.

The Fear of Mathematics

The prevailing attitude to mathematics in many children is that of fear, which inhibits and paralyses. The origin of this fear is loss of confidence, of self-esteem through failure to achieve success. To overcome failure we must remember that „nothing succeeds like success" and arrange for the child to be successful up to his own level of attainment. We must first accept the pupil *as he is* and adjust our mathematical curriculum to his needs and capacities. We must arouse and maintain his interest, we must simplify and select our material. This process of communication is not easy for a creative mathematician because it is concerned with personal values — with the pupil, or learner — and not with impersonal values like consistency or generality of mathematical knowledge. Many a good mathematician seeking communicate to the non mathematical has often felt this conflict within himself — the feeling that one is telling half truths, or circulating doubtful coinage.

Begin with a problem!

Assuming that we agree on interesting a pupil in his learning it follows that he must be involved personally, emotionally, in his mathematics. The best way to do this is to pose a problem, a challenge that is just difficult enough for the pupil to think his way through. If the challenge arises from the pupil's own situation — "how can I measure this?", "Why does this work?" so much the better. Many pupils will learn mathematics best through the workshop and laboratory. The teacher can always contrive a situation which will lead to a problem. I should define a problem as "a difficulty, *appreciated* by the pupil, and *awakening* in him a desire for its solution". Notice that the pupil's situation is comparable to that of a researcher in mathematics.

The guidance of history

This approach follows the historical, the research line of development in which the tools, techniques, processes and principles of mathematics are devised in response to the challenge of problems. To use this method calls for especial skill on the part of the teacher since it is so much easier to start with the principle or the process and then to apply these to solve the problem! However the trouble entailed is always rewarded by increased satisfaction on the part of teacher and pupil. This approach, from problem to principle is probably already familiar to many of my audience who may have used it in the teaching of geometry, mechanics, caculus, logarithms, etc., etc. My own master T. P. Nunn summed up the method in the phrase "A lesson in mathematics should be a voyage of discovery". I believe this method is fundamental for successful teaching.

Discovery in Mathematics

Unfortunately he did not go on to show how discovery in mathematics takes place. In elementary work little more than trial and error, enlightened commonsense may suffice. The published work of mathematicians rarely helps us — we see only the final structure and we do not know how it came to be built up. It is tempting to suppose that creative thinking in mathematics must be done by stating axioms, definitions and looking around for theorems — to proceed on traditional lines of formal logic. But this is certainly erroneous Have we any other guides?

Creative thinking

Some years ago under gifted teachers of music and painting, I had the experience of personal achievement in playing the piano and in painting. These are forms of creative activity and I attended very closely to what went on in my behaviour (thinking and feeling) when painting a picture. I felt the force of tensions driving me to create a picture and realised how I had to follow whatever intuition or "hunch" came into my mind. It was essential to respond freely: to create from the deeper levels of the personality. Shortly afterwards I read Ronald Morris' "The Quality of Learning" and was lead to Max Wertheimer's "Productive Thinking". (I should say that my own experience of creative mathematics is within the very modest field of school work and I cannot speak with much confidence of creation in mathematics beyond that field. No doubt many others can, and will.) In this book I found a firm basis for my own observation and insights as a teacher and a learner. I believe that all who are concerned with mathematics would gain insight from its study: it is essential reading for those who wish to appreciate what I am trying to say.

Gestalt views on problem solving

Wertheimer, as many of you may know, was one of the founders of the Gestalt school of psychologists who regard our perceptions not as haphazard aggregations but as wholes-of-related-parts. These psychologists suggest that a problem should be viewed as a whole, that our desire to solve the problem sets up tensions which cause us to restructure the parts in search of their deeper relationships until in the experience of "the flash of insight" we suddenly perceive the solution. Three points need emphasis: first the contemplation of the problem as a whole, second the following of "hunches" or tensions, third the flash of insight. To contemplate the problem as a whole implies the conscious use of sketches, models, original forms of diagrammatic representation, devision of new notations, symbols. In seeking insight we may invent a simpler, a more symmetrical, even a more general problem. We may invent

simpler numerical data, we may assume an answer and work back from it, we may work from both ends, but we ask ourselves what are the tensions that are present, in our minds: what do we feel we understand, what is not clear, in what direction are we impelled to move by our desire to solve the problem?

And of course insight may be deceptive: but the genuine flash of insight is so familiar to every member of my audience that I need not describe the phenomenon in detail.

Perhaps the analogy of doing a crossword puzzle is helpful: by comparison many of the suggested traditional problem-solving routines in mathematics resemble a game of dominoes — the search for the right beginning, the unique next step, the linear progression, — or the extension of a railway into a jungle.

The method proposed

I therefore suggest that children be encouraged to learn matehmatics by problem solving using *as a beginning* the suggestions of Wertheimer. These seem to cover the field of school mathematics, and no doubt a wider region. The essence of the method is to pose a suitable challenge and to say in effect, "Look, this difficulty was overcome by ordinary human beings, using their native wit. You know some of the methods they used. Here is nothing mysterious about elementary mathematics. Ues your ingenuity, invent any method that seems likely: it can all be discovered using commonsense! Believe in your own powers, follow your hunches. But don't spend too long on it, and do be prepared to try different methods. Be flexible, not rigid. There are usually lots of ways of seeing through a piece of mathematics". Elementary arithmetic and geometry can be built up in this way using such a method of discovery, and of course all ways of perceiving insight should be shown and discussed so that children become conscious of the depth of their own powers. *Every* approach and every method of solution deserves some consideration and explicit recognition. If precise rules or formulae are desired these should be *explicitly* selected for good reason.

Children would then learn by doing, their education would be self-education — the most enduring. There would be much working and discussion in groups. One might describe the process as creative thinking and as such it would transfer to other fields. Conscious recognition of the method would be essential and would make education in mathematics a form of personal creative activity. It would then be possible truly to claim that mathematics teaches you to think. In particular I suggest that a frank recognition of the "flash of insight" could *revolutionise* the whole of mathematical education. Children should be encouraged to seek it and not to rest until some measure of it has

been attained The flash of insight is a release from tension: it is a universal experience. We all are familiar with its expression, — "I see it," "I've got it!" Without this release our understanding is incomplete. To attain such expressions of release should be an immediate aim in every mathematical experience.

Furthermore I suggest a check on the depth of insight. Children should be encouraged, once insight has been attained to offer their own evidence of learning, to make up their own exercises or illustrations, to explain to their fellows. This notion of children constructing their own exercises is a fruitful test of the quality of their learning.

On tools and drill

Some teachers will feel uneasy about the teaching of the tools and drill-procedures in mathematics. There is no difficulty. The development of a process such as the solution of a linear equation, can be a piece of creative thinking in itself as one advances, step-by-step from simple to complex instances building up consciously the complex from the simple. As for drill one may recall the advice to pianists of the great Tobias Matthay "Never sit at the piano unless you make music". The growth of skill in practising a tool whether it is an element of pianistic or of mathematical skill should be undertaken consciously and not blindly — much as an athlete may consciously study his performances when practising for a race, and by studying them seek to improve them.

Some cautions

Teachers of mathematics need to guard against misleading analogies derived from comparisons between mathematical structures and the growth of knowledge in youth. For instance the view that mathematics is built up like a wall, brick by brick each item firmly cemented to its neighbour is a dangerous guide: I prefer the analogy of crystals forming and growing in a solution. Nor is the insistence on firm foundations necessary: as we study mathematics we deepen our insight into foundations. Firm foundations are often only verbal knowledge, memorised statements that have no content or insight to support them, mere mechanical rules of procedure. Such are worthless. Nor is the insistence on mathematics as a collection of tools entirely helpful: a tool has mathematical value only for some purposeful job of work. Pupils should see the job before they devise and practice the tool.

And all teachers need to guard against forms of education that inhibit creative activity: to insist continually on the right beginning, the best method, the *only* way, the current fashionable notation can be most dangerous. The nature of mathematics continually tempts us to hurry on to a final form of

a demonstration, ignoring the stages of thinking through which this demonstration has been attained. For instance generalisation in mathematical is a noble concept: as an aim in the classroom it is insidiously dangerous. Insistence on set methods — for instance in the written expression or forms of solution — can be utterly inhibiting. Children think that the form is the essence of the mathematics and that the thinking must only be in that form. A teacher of a former generation insisted that when manipulating trigonometrical identities one must always begin with the left hand side and transform it into the right hand side. Superficially this seems so sound, so mathematical. But every creative teacher or mathematician knows that the secret of success in this kind of manipulation lies in freedom to transform according to the innate structure of the particular identity — and to previous experience.

Width of experience

It is clear that ability to think about problem solving will increase with experience (successful experience!) and it is one of the benefits of the Gestalt view that all varieties of problem solving are relevant since they show varying ways of perceiving relationships. Note also that the Gestalt approach justifies our familiar procedures for encouraging insight by using colours, shading, models, dynamic patterns, films. I believe that in all mathematical education we need to attend more closely to individual *emotional* responses to proofs or processes or notations so that we discover more about individual insight. Wertheimer refers to the satisfaction we experience when we find a direct proof using the materials of the problem as given. Many of us may have felt dissatisfaction with certain logically valid forms of demonstration such as reductio ad absurdum or even induction. These do not give depth or insight into the relationships involved. Hence our dissatisfaction.

Two difficulties

First we must recognise that this method of learning mathematics by individual discovery runs counter to traditional methods of teaching mathematics in which every step is determind by the knowledge and authority of the teacher. I believe this method accords better with the way in which human beings learn mathematics and therefore it may have some universal validity.

Second we must admit that it makes great demands on the teacher. So do all methods of individual education and these demands may be greater than we can expect teachers to sustain — just as the advances in modern medicine make great demands on the practising doctor. Nevertheless this method may be used incidentally by teachers who wish to try it. I believe

that great pyschological security on the part of the teacher is essential if the method is to be developed to the full.

Some practical suggestions

For many children in English secondary schools the Binomial Theorem represents the highwater mark of mathematical attainment. To discuss this theorem we have available many methods: and to compare these from the "insight" point of view is most revealing. The simplest is to pose the problem of the generalisation of $(a + b)^n$, but first to consider the simple case of $(1 + x)^n$ using successive multiplications to find the coefficients. It is easy to infer the coefficient $\frac{n(n-1)}{2}$ as the result of addition $1 + 2 + \ldots + (n-1)$. Once when I was teaching in this way a pupil pointed out that the ratios of terms in the column were in a simple order

$$\begin{array}{ccccccc} & & & 1 & 1 & & \\ & & & 1 & 2 & 1 & \\ & & & 1 & 3 & 3 & 1 \\ & & & 1 & 4 & 6 & 4 & 1 \end{array}$$

e.g. 2 1 gives $\frac{1}{2}$
3 3 gives $\frac{3}{3}$ or 1
4 6 gives $\frac{6}{4}$ or $1\frac{1}{2}$

from which the coefficient $\frac{n(n-1)(n-2)}{2\times 3}$ may be inferred. A good illustration of "insight". Many interesting consequences follow from this pattern of coefficients and if we develop the expansion of $(1 + x)^n$ with $n = \frac{1}{2}$ or -1 we pose other problems which follow the historical development of mathematics. We may refer to Newton's work on the expansion of roots of equations, to Wallis' on integration. Wallis' treatment of the area under the graphs of x^2, x^3, ... is a further illustration of a problem solving approach to mathematical learning. This kind of discussion makes the expansion "come alive" to pupils.

The method and the curriculum

It would be a considerable contribution to mathematical education and indeed to education in general if we had some general line of procedure in tackling any topic of mathematical education — and I am bold enough to invite you to consider the claims of the arguments I have put before you. If these claims were substantiated we could give our attention to the perplexing

problem of the content and order of the mathematical *curriculum*, that is of the topics we desire our pupils to explore and learn. The advance of knowledge makes this an acute problem. We may, for instance, know *how* to teach statistics but do we know when? Here again I agree with the Gestalt insistence on the concept of wholeness. Exclusiveness in education is always a mistake and so are restrictions in our views of the content of mathematics. I look forward to a time when mathematics and the sciences and arts shall be studied as parts of a related whole. The determination and ordering of those parts is a task for us in the future. (In my own country one of the most urgent needs is to restore to the curriculum the kind of problem-solving activity that is fundamental in mathematical thinking but which has almost disappeared from school teaching through the insistence on routine procedures, drill, revision, solving, factorising, reproduction of proofs etc. etc. in order to secure safe passes in examinations.)

Looking ahead

I am conscious of the inadequacy of the outline sketch which I have offered to you this afternoon. May I recall that I am concerned to give confidence and success through discovery, through problem solving using restructuring and seeking insight. Self-education in mathematics would become a dynamic experience, it would touch the heart as well as the head. I believe it would raise the level of respect for mathematicians and for their work and that it would produce more effective research workers. The free use of oneself in creative activity undoubtedly releases deep sources of energy — from stores below the level of consciousness. Such release brings achievement and happiness.

Every mathematician present is aware of the experience of the flash of insight, and all who teach in classrooms are aware from time to time of a sudden deep stillness when some stage is reached that absorbs the full attention of the children. I call this experience (and it is not confined to classrooms!) the hush of interest.

Is it too much to hope that in our future teaching whether we are mathematicians or teachers we may seek and achieve two goals — the hush of interest, and the flash of insight?

INTERNATIONAL INQUIRY OF THE INTERNATIONAL MATHEMATICAL INSTRUCTION COMMISSION (IMIC) ON

THE ROLE OF MATHEMATICS AND MATHEMATICIAN AT PRESENT TIME [1]

G. KUREPA

1. *Origin of the Inquiry.* The IMIC was of the opinion that the instruction of mathematics at a given time is closely connected with the role of mathematics and mathematicians at the same epoch. Therefore the IMIC decided at the meetings in Geneva 20.10.1952 and Paris 21.02.1953 to undertake an inquiry on this theme in order to get an insight in some contemporary characteristic items and results which in particular are relevant for a close examination of the mathematical education considered as a whole [2]). Sure, such an inquiry could be done at every time and relative to every science or art. The reason that the IMIC initiated just such an inquiry at our time consists in the fact that we are present to very revolutionary discoverings both in mathematics and in Science generally (cf no 7).

2. *Circular and Questionnaire concerning the Inquiry.* Consequently, the Executive Committee of IMIC issued a circular letter at the beginning of 1954 [3]) and asked directly the national subcommittees of IMIC and the countries adhering to the International mathematical union to take an active part on the Inquiry; one composed a Questionnaire (cf. annex I) containing more than twenty questions covering almost the whole problematics of connections of mathematics with other human activities. The questionnaire was disseminated in a quite restrictive manner; but from now on it should be disseminated in order to incite still more a collaboration of various organizations and specialists. It is to be hoped that in the coming years we should get still a lot of valuable contributions and suggestions. A book is planed to collect the results of the Inquiry.

---

[1]) This report was presented 08.9.1954 at the International mathematical Congress in Amsterdam; in french it will be published in Enseignement Mathématique — the official organ of IMIC.

[2]) The second inquiry of IMIC concerned more particularly the problematic of mathematical teaching in the critical jontion region of the medium and superior levels i.e. for 16—21 year pupils.

[3]) Cf. International Mathematical News 27/8 p. 8 (1953), 3 1/2, p. 3—4 (1954).

3. *General reporter. National reporters. Future contributions.* As genera l reporter I am glad to point out that the Inquiry was accepted very favorably by many national subcommittees of IMIC, many organizations and people. Several very known scientists gave their support for the inquiry; the answers are still arriving.

On behalf of IMIC I have the privilege to thank all those who contributed so far or who shall still contribute in the future to the more profound elaboration of the Inquiry. Here is the list of national subcommittees of IMIC which took an active part at the Inquiry:

Country:	Reporter:	Exact title of report [4]):
1. Austria	O. Weinberger	Über die Anwendung der Mathematik auf Staatswissenschaften.
2. France	G. Darmois	Le rôle du mathématicien dans la vie contemporaine, pp. 8.
3. Germany	E. Kamke	Die Rolle der Mathematik im heutigen Leben, pp. 32.
4. Italy	G. Ascoli	Le rôle de la mathématique et du mathématicien dans la vie contemporaine, pp. 3.
5. Netherlands	D. van Dantzig	The Function of Mathematics in Modern Society and its consequences for the teaching of Mathematics, pp. 12.
6. U.S.A.	A. M. Gleason	The expanding role of Mathematics, pp. 4.

In the meantime was published: W. Leontieff, Mathematics in Economics (Bull. Amer. Math. Soc. 60, 215—233 (1954).

While the reports of Professors Ascoli, Darmois and van Dantzig shall be now presented by their authors respectively, we are regretting that Professors Gleason, Kamke and Weinberger were unable to attend this Congress and present their reports.

4. *Definition of Mathematics.* One is anxious to define the Mathematics. Now, it seems to us that it is illusive to attempt to give an explicit definition of what is Mathematics. Analogously, it is impossible, at present time at least, to give an explicit definition of what are: logic, number, electricity, matter, arts, justice, beauty etc. In this connection we are anxious to say this: Since the greek epoch we like to appreciate the deductive method in mathematics as the most important one. Probably one exaggerates at this point. Pure and complete deductions are rather very scarce in the Nature and in the Life. What

[4]) Maybe, the most appropriate title shoud read ,,The role of Mathematics and Mathematician at the beginning of Atomic Age."

occurs is a concatenation of induction and deduction. In law sciences and in retrograde chess one is dealing with a retrograde logic. Long deductions may lead us to rather paradoxical results. Therefore it is necessary to give a correct position and not an exaggerative one — to the pure deductive method in mathematics. Maybe, it should be better to ask "Not to do illogically" than "Do (act) strictly logically". In the life one encounters a very complicated logic. On the other hand, doing mathematics, one defines them (implicit and descriptive definitions).

Personnally, I consider that the mathematics contains: 1) Logic, 2) Arithmetics (in the large sense; Analysis and Algebra included), 3) Geometry, 4) Mechanics and theoretical (mathematical) physics.

The Set Theory and the Logic constitute two approaches to Mathematics. In particular, the notions of set and processus (transformation) are basical for the contemporary mathematics. The notion of organization or structure of a set occurs everywhere in mathematics (e.g. algebraic structure, topological structures, relations etc.) reflecting so what happens in the natural or social phenomena. There are very many mathematical branches so far.

5. *Definition of mathematician.* Approximatively, mathematicians are people dealing with mathematics (theoretical researches, applications, teaching etc). At our time, in the middle of the 20th century, one becomes a mathematician officially, by attending various schools during 16—20 years and by performing several examinations on mathematics. One becomes so: mathematician, statistician, professor, master, ingineer, doctor of mathematics; probably there are still other denominations for professional mathematicians in various countries.

A small aspect of how one becomes mathematician is visible also at synoptical panels of the Pedagogical Exposition of this Congress. It is particularly to be notified that various schools, colleges, faculties, universities, academies, mathematical institutes etc. deliver various degrees of mathematical qualifications.

6. *Scope of Mathematics.* The number of mathematicians and mathematical institutions as well as the number of institutions where mathematics is teached has increased considerably during the 20 last years, both relatively and absolutely. One can say that nowadays about 5000 mathematicians publish their works in about 1000 periodicals. At this moment we have no exact data about the number of mathematicians and the number of hours teached on mathematics in various countries. If one considers the chess-ordinary and fairy — as a kind of mathematics in the large sense, the number of mathematicians at large sense (the ingeneers as practical mathematicians included) is extremely high. From this point of view it is instructive to mention at least the

following societies, in order to show the diversity of subjects examined by mathematicians:

a) The International Association for Symbolic Logic publishing The Journal of Symbolic Logic; in 1954 the 19th volume is to be published.

b) The International Econometric Society (since 1932) with 2500 members; this Society is publishing the Econometrica.

Moreover, various societies are founded by the existence of which many people are surprised; so e.g. in U.S.A. one has:

c1) Industrial Mathematical Society (since 1949 "to bring together persons who use Mathematics in engineering in order to learn more about and to extend the field of mathematics as well as to *develop* new procedures for the solution of problems that arise in modern industrial research");

c2) The Biometric Society (since 1947);

c3) The Psychometric Society (by 1930);

c4) The Society for Quality Control (since 1946) numbering about 6000 members in 1953.

Moreover to notify the well known societies in U.S.A.: The American statistical Association, The American Mathematical Society, The Mathematical Association of America etc.

6.2. *Periodicals.* The dissemination of mathematical knowledge is done by means of books and periodicals whose number is steadily increasing; so e.g. in Mathematical Reviews or in Zentralblatt für Mathematik are mentionned about 900 periodicals. Almost every country has an university, Science Academy, Scientific societies etc. working on mathematics and publishing mathematical results (original or expository ones). Moreover, the number of periodicals concerning elementary and medium mathematics is quite great. At present time one has the following 5 journals informing about the entire mathematical activity:

Bulletin Analytique, Paris;

Zentralblatt für Mathematik (since 1931), Berlin;

Mathematical Reviews (since 1939)

Boletin del Centro de documentacion cientifica y tecnica (since 1952) Mexico;

Referativni Žurnal: Matematika (since 1954) Moscow.

6.3. *Mathematics Laboratories.* In great industrial, economical and commercial enterprises one founded mathematics laboratories. This fact is without precedence in the history of mankind. In particular, in connection with various calculating devices there are: ingineers, mathematicians, technicians. The sublime role of mathematician is, besides others, that of programming: to prepare the material to be used by machine in order to get from the machine a

palpable result in form of a number or curve containing a law or data aimed by scientist, ingeneers, legislator etc. Here is an example. By 1935 the Bell company introduced the quality control for the first time; during the world war II this system was used in other industries also. In USA one had in 1952 50 times more statisticians for the Quality Control than in 1945, and in 1946 a corresponding Quality Control Society was founded which had in 1953 about 6000 members.

6.4. *Mathematics Institutes.* In various countries or towns one founded mathematics institutes or mathematics centres, independently (as in Austria, U.S.A.) or jointly with other institutions as Academies (SSSR, Germany, Poland, Yugoslavia), research centres (Italy, France, Netherlands), Physics institutes (England, Sweden) etc. Let us remind also that the first international institution projected by UNESCO was the International Mathematics Centre; it should be build in Rome, but the statutes of the Centre are not ratified so far by a sufficient number of states.

6.5. *Employment of mathematicians.* Still by 1930, rarely a mathematician worked except in teaching or doing researches in mathematics or in insurance companies. At present time it is otherwise. It should be extremely interesting to know the exact distribution of mathematics jobs in an industrialized country. Here is an extract of Professor E. Kamke's report concerning 31 "diplomed mathematicians" (from 42 in all) who 1946—1953 studied at Technische Hochschule in Stuttgart: [5])

1) Teaching in high schools 7
2) Assistant professor 3
3) Mathematician in electricity enterprises 6
4) Insurance 2

One by one were employed in each of this divisions: Professor in a machine school, optical industry, officer in an ingineer office of electricity, civil constructions, automobilism, steel construction, statician in a steel office, owner of a building office, gum industry, tax direction, statistics, fiduciary society, astronautic institute.

---

[5]) In the same report one learns that in 1942 when the Germany was in shortage for various specialists one introduced the diploma examinations parallelly to "promotions" and Lehramtsprüfung.

"Die Ausbildung der Diplom-Mathematiker hat das Ziel, die Studierenden mit dem mathematischen Rüstzeug zu versehen, das zur Lösung der in den Anwendungsgebieten vorkommenden Aufgaben entwickelt worden ist. Die Studierenden sollen ferner so weit mit einigen dieser Anwendungsgebieten selbst vertraut werden, dass sie im Stande sind, deren Probleme mathematisch zu fassen und zu behandeln." (Aus der Diplomprüfungsordnung der mathem.-naturwissenschaftlichen Fakultät der Universität Tubingen).

7. *New fundamental acquisitions.*

7.1. *Set-Correspondence (function) - Logic.* From everyday's considerations evolved the fundamental notion of set becoming the very basis of mathematics. The basic notions as: function, correspondence, relation etc. are backed upon those of set and organization of set [6]) (various structures). The social life is now a rich source of inspiration for mathematician (cf. notions of: order, relation, group, class etc.). The Set Theory and the Logic constitute two approaches to Mathematics. The set-structural-granular-concept of energy and light give rise to the notion of quanta and photons, respectively. Here one has to revise very profoundly much of our knowledge and the manner of cognition. One is acquainted with a great number of quantum mechanics problems, several ones of which are purely mathematical ones.

7.2. *Optimum and Maximum. Stochastic processes.* The stochastic dependence arose; it generalizes the classical functional dependence and is now an extremely powerful tool for explanation of natural, economical, social and human phenomena. To extremal progressing of macrophysics one added the optimal progressing of microphysics, biology, economics, social sciences etc. In addition to strict causality and Aristotle's logic we have now the probabilistic and statistical truth, indeterminism, multivalent logics etc.

7.3. *New calculating devices.* One succeeded to represent digits by means of various physical quasi-imponderable effects (electrical pulses, electrical and magnetical charges etc); hence the possibility of imponderable representation of numbers, calculations etc. This started after 1930 and one refers to the names of Aiken, Couffignal, Eckert, Williams etc. All this will imply one of the greatest revolutions in the human thought. The mathematics became again very closely tied with physics and with psychology and one is anxious to examine the manner of thinking, calculating and of other psychical and intellectual functions. One performed very fast calculating machines and machines sensible to forms (Gestalt) and size. As a corollary, a new numerical science, Numerical Analysis, is arising with its proper problems and its proper rules; so e.g. if one operates with a given number of decimal digits, the ordinary association law does not hold necessarily. The arithmetical operations are mechanically executable in some microseconds. As illustration here is a table of how lasts the multiplication of 2 decimal numbers, 8 digits each:

Pencil and paper	5 minutes (the exactness being required);
Table machine	10 seconds;
Holeryt machines	2 seconds;
Electronic devices	100 msec—200 μsec.

[6]) It is to be remarked that according to J. von Neumann (Mathematische Zeitschrift, 27, 669—752 (1928)) the notion of set could, conversely, be deduced from that of function.

7.4. *Individual-Collective. Concrete-General.* In the contemporary mathematics this problem is examined in two ways particularly: statistically and topologically. In statistics, each individual case plays a theoretical role but materially one may nothing say as to prevision of this case: accident has neither memory nor conscience. Statistics became a huge mathematical science with applications in every human activity: from physics to clinical and human sciences and arts. In the study of man, its role is important still for the reason that the experimentation on men is not allowed. In our inquiry the majority of people stressed the particular importance of statistics (and statistical methods), of differential equations (ordinary and partial) and linear systems.

The relation: individual-collective give rise, from the topological point of view, to various existence problems (fix point theorems with several applications; we can note that in 1952 one succeeded to deduce the fundamental theorem of algebra from the Brouwer's fix point theorem for circle). The topology is in full florescence and is melted even with statistics. The conception of individual in the modern atomistic is one of the most complex and requires a profound changing of thinking in some areas (e.g. in quantum theory). The interconnection of logical quantifiers is of extreme importance and belongs to the same problematic: individual-collective, particular & concrete-general.

7.5. *General Atomistic.* Quantum mechanics, quantification of our knowledge and the general Atomistic are amidst the highest acquisitions of man. A revision of our cognition is required.

7.6. *Cybernetics* [7]). The theory of communications and control has arisen and contains the immense realm of robots and servomechanisms on the one hand and on the other hand: the theory of nervous systems, of brain and of psychical actions. Here the feed-back is a universal phenomenon: the existence of a difference between the goal and the instantaneous state serves as a reason for diminuation of this difference. The processus of human activity: of scientific, artistic, economical etc. activities, is a collossal feed-back processus.

A practical conclusion: the preceding acquisitions and the general interconnection of mathematics with other activities in all domains have to be reflected also on our working and teaching in mathematics: the content of mathematics and mathematician is in close correlation with the whole our activity.

Therefore arises the necessity of collaboration of mathematician and mathematics with other specialists and other sciences respectively (*interscientific fecundation*).

---

[7]) cf. N. Wiener, Cybernetics or Control and Communication in the Animal and the Machine, Paris 1948 (ed. Hermann) pp. 194. v. also Les machines à calculer et la pensée humaine (Colloque international, Paris 8—13.01.1951) Paris, 30+570 (1953).

8. *Joint activities. Examples.* As succeeded examples of joint activity we mention especially: IBM, Cowles Commission and Cybernetics.

8.1. *IBM* in New York was founded in 1914 and 1924; its activity is well known. Calculating machine Mark 1 of Aiken and Watson computing laboratory issued from IBM.

8.2 *Cowles Commission* was founded in 1932 at the same time as Econometrica and Econometrical Society [8]). As this Commission is a model of a contemporary joint activity, let us indicate an extract of its statutes.

"The Cowles Commission for Research in Economics is an organized group of scholars dedicated to research in economic theory and measurement. It seeks to make additions to fundamental knowledge about society, through theory construction, through measurement for the testing of theory, through development of methods of measurement, and through application of results in specific areas. Its research staff blends many different educational and cultural backgrounds and unites economists engaged in theoretical and empirical inquiries with statisticians and methamaticians into a balanced cooperative research effort."

Here follows a description of working in such a complex body.

"In seeking a systematic approach, Koopmans hit upon the principles of an analytic method that was first known as "linear programming" but has now come to be called more accurately "activity analysis of production." It has grown up from several converging sources, including analyses by Wald and von Neumann of the Walrasian general equilibrium theory, discussion in welfare economics, Leontief's interindustry analysis, and the programming activities of government administrators as studied in the U.S. Air Forces by G. B. Dantzig and M. K. Wood. It is similar to the traditional economic theory of production in that it seeks first to establish the technological relation between imputs and outputs, i.e. to answer questions like these: if the quantities of all imputs and all outputs but one are held at specified levels, what is the maximum quantity of the remaining output that can be produced? Alternatively, if the quantities of all outputs and all inputs but one are to be held at specified levels, what is the minimum quantity of the remaining input that is required?" (loc. cit. p. 51).

8.3. *Cybernetics* resulted from free discussions between specialists of various sciences, particularly between mathematicians, ingineers, biologists etc. (cf. Wiener, loc. cit). Professor G. Darmois will you describe now the genesis of the *Institut de Statistique de l'Université de Paris* — a good example of instruction and performance; professor D. van Dantzig will describe the *Mathematisch*

---

[8]) cf. Economic Theory and Measurement (A twenty year Research Report 1932—1952). Chicago 1952, pp. 180.

Centrum of the town whose hospitality we are enjoying and will show its activity and its role particularly in connection with great floods occured in this country in 1952. Both institutions should serve as model-examples for many countries. I refer to the report of A. Walther [9]) concerning the *Institut der praktischen Mathematik* in Darmstadt as a synthetical work using: numerical calculations, diagrams and machines. [9]).

8.4. *Industry and Production.* We are present to a profound change in the Industry (Production): Besides the well known departments: administrative, technical, commercial ones a new one: Scientific Department is emmerging; in this Section mathematics and mathematician play an important and active role. The mathematician is active here by several ways: he takes part in projects, construction, fabrication, sale, purchase, distribution; sampling method and quality control are highly appreciated methods. The productivity and rendement of industry, efficacity of a legislature or of a military operation, optimal researches concerning working, education, teaching etc. are subject to mathematical and more particularly to statistical investigations.

It is extremely interesting to know what happens in a big industrial, commercial or cultural body (organization, enterprise etc.) as e.g. in a factory of cars, planes, ships, various apparatus and tools, gums etc. For instance, in the fabrication of planes one can distinguish following divisions: [10])

1) Lofting group for determination of loft lines, forms and choice of material;

2) Aerodynamic group for configuration and detailed forms;

3) Dynamic group;

4. Stress group;

5) Equipment group to determine the optimal conditions for radar location etc.

In every of these divisions one operates with various mathematical means; here as well as else one encounters many mathematical problems not resolved as yet.

The architect Le Corbusier is constructing, actually, a whole town in India with special climatical references (sun, shadow, light etc.); this should be an opportunity to compose a very book of mathematics if (assisted by a professional mathematician) one indicated all of mathematics occuring in this case.

The particular and the personality of the past are now replaced by the collective personality called: working group (staff), laboratory, school etc; in

[9]) pp. 260—274 in the book: H. Behnke, Der mathematische Unterricht fur 16—21 jährige Jugend in der B.R. Deutschland, Göttingen 1954 pp. 332.

[10]) cf. Viall, J., Mathematics in the Aircraft Industry (The Mathematical Teacher, 46, 145—151 (1953)).

every such collection mathematics is duly represented by one or more mathematicians; it is to be remarked that mathematicians are here requested, appreciated and duly paid.

9. *Fundamental proposition.* One is familiarized with the connection of mathematics with astronomy, physics, geophysics, meteorology (here we call attention to Milankovic's Mathematical Climatology), chemistry, biology etc. One is acquainted how the origin of various mathematical fields and methods are attached to various actions or phenomena, examples:

a) Geometry — measurement of soils in the Niles region;

b) differential calculus — considerations about: speed, tangents, casks, extrema etc.

c) Topology — the breadges on the Spree.

d) Probability — theory of simple plays (as die).

e) Calculus of variations: Dido's anecdote, the brachystochrone etc.

f) Set Theory — various everyday's considerations.

Less are known the connections of mathematics with other human activities; such connections exist: mathematics is linked with arts, linguistic (cf. semantic and syntax), military science, all kinds of plays etc. Bearing this fact in mind and knowing not only the high degree but also the great rythm of development of our cognition and activity, the following fondamental proposition arises:

The activity in each domain is reflected in mathematics too; and reciprocally; hence the necessity of a close contact and collaboration of mathematicians and specialists in other fields with the aim on the one hand to find new pulses and on the other hand to apply mathematical results. In the realm and interconnections of present natural, artistic, social and human phenomena there are certainly such ones which could originate new mathematical investigations. The results of these investigations could serve not only for examinations of phenomena where they originated but also of lot of other things: The same mathematical phenomenon may occur in various and very disparate forms [11]. (*diversity of realization of a mathemactical model*-great enigma resulting may be from the immense numerousness and complexity of the explored events and things on the one hand, and of non-numerousness and the anatomic-physiological simplicity of exploring organs on the other hand).

10. *Mathematics as a crossing-point of human activities.* Since every activity reflects on mathematics, one concludes indirectly that there exists a correlation between sciences, arts, and other human activities. In the future one would say

---

[11]) cf. M. Petrovic, Mécanismes communs aux phénomènes disparates, Paris 1921, pp. 279.

more on these interrelations when shall be examined more accurately: the functionning of our sense and brain as well as the mechanism of other activities of various living beings. Yet, we can request already now: to every phenomenon (theory) of one activity field associate the set (vacuous or non vacuous) of its analogs in each other activity field (*Problem of the hierarchy of a phenomenon or of a theory*). The transfer of a theory into a different ambient can involve very appreciable results (*hybrid fertilization, duality considerations* etc). The mathematics contributes to such a transfer; even the possibility and necessity of mathematics is linked with the presence of same or similar regularities in circumstances which are apparently less or more disparate. In this connection let us say the following:

There are objects (things) whose mutual relations are independent from the exploring human being; there are others too, the interconnections of which are changing during the examination or which are changed by a long chain of transformations on the way to reach the human brain. One is dealing here with a feed-back processus which is changing when exploring it. In the atomic physics there are many examples of this kind; Consequently, an association of human beings is like a feed-back machinery, especially when this association is guided either directly (directed economy) or indirectly (industrialized free economy; here occur as determining items: profit, comfort, work efficacity etc). Between these two (psychologically extremal) cases are located other cases.

11. *Moral duty of mathematician.* The mathematician is dealing with the most essential relations, without prejudices as to country, religion, race, nationality, education etc. This does not mean that a mathematician does not possess of all that. But his role is to stress all what is essential in a situation and so contribute to a mutual understanding between individuals, collectives, domain of sciences, arts etc.

At our time, middle of the 20th century while various omni-potent robots are starting to act, this spirit of mutual comprehension is to become a leading idea in our education and our Weltanschauung.

In conclusion, Mathematics ($\mu\acute{\alpha}\vartheta\eta\mu\alpha$ = science) is like a creative conscience of mankind. Without Mathematics, the mankind, our cognition and our civilization would be as an organism suffering of a general ataxy [12]). Our mathematics is correlative to the possibility that we get an organized, ordered, view on the creative chaos of the infinitely varied nature; the mathematics is a less or more approximative reflect of the reality.

12. *Consequences for IMIC.* Of what precedes one is aware of the essential

[12]) To remind of the considerable mathematical effort which was to be done in order to organize such a congress like this one. And what to say in this direction as to the organizing of a highly industrialized productive mankind!

role of mathematics in the entire human activity and in particular in our civilization. Now, during last decades several results of primary importance are reached. Various working methods are formed, they interact mutually. Mankind is awaiting a help from mathematics and mathematician. New points of view, more direct, are appeared permitting to operate more directly, more efficaciously. At present time, both: the education and cultural-scientific effort became massive; therefore we have to proceed to a considerable revision of methods of work, and instruction and to an appropriate selection of matters to be teached. The effect of teaching (education) in general and that one of mathematics in particular are to be examined mathematically too.

For this reason and owing to the capital importance of Science and Technique, the IMIC as an international organism concerned with scientific education is faced with very great and highly responsible problems and duties. The moment has arrived to solve it in an active manner.

BIBLIOGRAPHY

The bibliography relative to the subject of the Inquiry is very great; a part of it is indicated in The Mathematics Teacher, especially in issues 1953—4.

cf. also: The Annals of the Computation Laboratory of Harvard University. So far issued more than twenty volumes starting in 1946 with Vol. 1.

An International Inquiry on
THE ROLE OF MATHEMATICS AND MATHEMATICIANS IN CONTEMPORARY LIFE.

Questionnaire

1. Your name and address
2. What is your main field of activity?
3. In what direction (scientific, logical, aesthetic, applications, graphs, models, projections, educational, social, economic, financial, as tool, linguistic etc) do you particularly stress the importance of mathematics?
4. If you are (not) a mathematician, do you ever consult non-mathematicians (mathematicians) in your activity?
5. Indicate at least one case in which you personally or somebody you know personally have applied successfully mathematics or whose mathematical investigations were implied by applications?
6. What kind of mathematical possibilities (functions, graphs, tables, models, kinetic models, projections, nomographical methods, statistical methods, probability, infinitesimetal methode, deduction, high speed computing devices etc) do you consider as the most appropriate in your field?
7. What is your opinion of the role of high speed computing devices and robots in general?

8. Indicate the appropriate measure (methodical, educational etc) to assure that the social importance of mathematics and that of mathematician should coincide as much as possible.
9. Indicate at least one problem from the domain of your activity which it would be worth while to examine mathematically and by mathematicians respectively.
10. What are the jobs or opportunities connected with your activity which require, in order to be well conducted, a pronounced mathematical education and which consequently, would be better managed by a mathematician than by a non-mathematician?
11. What is your opinion on the future interconnection between your field of activity and (other fields of) mathematics? In what branches of mathematics are you especially (actively or passively) interested?
12. What is your opinion about the interconnection between mathematics and exact (natural) sciences, in particular with astronomy, physics, chemistry, geophysics, mineralogy etc?
13. The same for biological sciences, both pure and applied (e.g. medicine, agriculture etc).
14. The same for engineering sciences.
15. The same for social sciences.
16. The same for economic sciences.
17. The same for philosophical sciences.
18. The same for industry.
19. The same for arts in general and in particular with regard to music, painting, sculpture, architecture etc.
20. The same for linguistic.
21. The same for war sciences.
22. The same for sports, games, chess etc.
23. Other problems you could add or suggest in connection with the role of mathematics and mathematicians.—

Note. It is desirable that the answers should be as concrete as possible; indicate: references, characteristic items, quotations etc.

Let everyone answer the question(s) in which he (she) is most interested. The answers are to be addressed to

G. Kurepa
vice-president of
International Mathematics Instruction Committee
Zagreb, Mathematics Institute
Yugoslavia

P.s. Please diffuse the above inquiry.

SCHOOL MATHEMATICS FOR UNIVERSITIES AND FOR LIFE

KAY PIENE

We will assume that *mathematics* as a school subject should at least include *a*) the use of letters for variable or unknown numbers (quantities), *b*) a certain use of proofs.

We will further assume that a child must have reached a certain maturity to be capable of learning mathematics.

A *school* is an institution for teaching *pupils* who possess the above-mentioned capacity, whether mathematics is taught or not.

In the word *universities* we also include other institutions of equal standard preparing *students* for liberal professions.

Life can be defined as all human activity outside schools and universities.

The chief problem we are going to discuss is: How can we organize the teaching of mathematics in schools in order to prepare in the best possible way for universities as well as for life?

This problem has interested me personally because, for some years, I have had the privilege of teaching mathematics in a grammar school, of being an examiner in mathematics at a university and of instructing future teachers of mathematics, the problem, however, ought to interest colleagues all over the world.

Strangely enough the interest for problems of this kind — in Europe, perhaps not in other parts of the world — was greater formerly — say 50 years ago. It is sufficient to mention one person: Felix Klein, and one organization: IMUK, the International Commission on the teaching of mathematics. Since Klein died we have had no man of his dimensions in this field, and most of IMUK's activities were done before — say 1920.

Since then, life in general has become more and more mathematical, but the teaching of mathematics has not advanced in the same tempo. This development has several causes. First, schools have a tendency to be stable and conservative. Further, as far as I can judge, 50 years ago university professors, more often than is the case to-day, were ex-school teachers. It is also possible that university professors to-day are more interested in preparing for research scholars than in preparing future teachers, since more and more frequently the best students do not go to schools as teachers.

According to Felix Klein the teaching of mathematics should be guided

by the concept of function and by the idea developing the ability to visualize the (three-dimension) space.

In my opinion this way was a wise program, and certainly the first principle has been put into practice to a great extent in various countries.

The second principle, however, has not been realized to such a degree as Klein hoped. Too often solid geometry means cutting solids with planes and thus going down to the safer *two* dimensions — the result is: computations instead of geometry. The reasons for the strong position given to the concept of function is obvious. Not so evident is the second idea. Let me mention such phenomena as 3-dimensional films, the new methods for making maps, the use of n-dimensional geometry in psychology, economics etc. Far too often problems from practical life are supposed to be linear, which they are not, and it is dangerous to go on to stick to that one dimension, thereby making the problem meaningless.

Consequently, we need more people who can see and think in three and more dimensions.

Can we still use Klein's program or is it necessary to consider the changes that have taken place in the world since the time?

We have seen the progress made by mathematical statistics, which have been introduced into and made profitable in all sorts of sciences and even humanistic activities. At the University of Oslo there are now courses in statistics for students of education, economics, medicine, psychology and for future secondary teachers — that is in all faculties except that of divinity. We might also mention the revolution made by electrical or electronic computational machines — problems can be solved in minutes instead of months.

Therefore, new demands will be made to the schools asking for deeper and wider knowledge, especially in the "newer" disciplines of mathematics.

On the other hand schools are not to-day what they were 50 years ago. Schools — always in our meaning of the word — were at that time grammar schools (Gymnasien, lycées, collèges), chiefly with the task of preparing for university studies, taking practically only those pupils who were supposed to form the élite of the next generation. Since then, new masses have entered schools. The schools take a much larger proportion of the corresponding age groups, and they are not only foundations for special studies at universities.

The situation might seem hopeless, but fortunately a certain tendency has weakened the consequences of the development: Schools by degrees have become more differentiated, divided into sections in such a manner that pupils may choose the type of school that best suits their abilities, interests and probable future career.

We must therefore operate with a set of schools and for each set find

a suitable program of mathematics, and we must remember that a certain topic of mathematics cannot wait and be treated in the higher learning institutions, if it is useful also for those going directly out in life, or those going through studies where no mathematics is taught.

Usefulness must be a central point of view. Often another point of view is stressed, that years ago Bacon expressed in the sentence: Mathematics makes men subtle. In other words: may teaching of mathematics give values to and have good influence on reasoning *outside* mathematics? If there is such a process, I do not believe it is automatic. The mathematics *program* in schools must in our days be determined according to what is useful later on — in studies or in life. But the *methods of teaching* should be such as to provoke values also outside mathematics itself, in other school subjects as well as in activities in life.

If we accept that utilitarian point of view, we are still faced with certain difficulties: How to decide which topies are useful? Do universities and life generally agree on that point? It is necessary to get information on the needs of industry, business, administration etc. and then let the programmes be set up in co-operation between schools, universities and life.

The maximum program will be for future students of mathematics and natural sciences, for civil engineers, officers etc. Students of medicine, dentistry, veterinary sciences, agricultural sciences etc. may be satisfied with a somewhat limited programme. It is possible that a third and smaller programme might cover all other needs, so that three programmes in all would be sufficient.

Let me, however, first express the opinion that *all schools* (cf. our definition) *should teach some amount of mathematics*. The minimum program should comprise arithmetic, algebra, geometry and trigonometry, but these subjects should as much as possible be taught as an entirety.

Arithmetic. It is an essential task to see to it that the pupils make an honest use of numbers. Computation with approximate numbers and (mental) methods of estimating and controlling the solutions of problems must be considered. Problems must be practical, if possible within the pupil's world of interest ("social mathematics"). Calculation with percentages and ratios must be well taught.

Algebra. Essential here is the understanding of the use of letters in formulas and equations. We must avoid artificial drill and especially "riders" that are formally solvable, but in reality impossible because they involve calculations with numbers not introduced. Example: Let a be a positive integer and assume that only fractions of the type t/n, where t and n are positive integers, be defined. Then $(a-3)/a + (4-a)/a + (a-5)/a$ can

be evaluated formaliter, but not realiter. Mechanical drill should be replaced by complete discussions of the type: If . . ., then we get the following solution, no solution etc. Graphs, logarithms and slide rule must not be introduced too late, while directed numbers can wait.

Geometry. Essential here is the understanding of the structure and methods The starting point may be more practical, with the use of compasses and other instruments, but when the pupils are sufficiently mature, they should be well familiar with the meaning of a definition, an axiom, a theorem and a proof. Better to let the pupils find proofs for new theorems than to let them struggle with constructions that will be more or less solved by an accident. Better introduce vectors and drop useless theorems which owing to tradition linger on in the textbooks.

Trigonometry can be started early and a good introduction in the use of mathematical tables can be given. For those going on with mathematics it is necessary to master a certain collection of formulas. In fact, this collection is not large, and many formulas can and should be found at once by the use of the definitions.

This ought to be a minimum program and a basis for those going further. Algebra and trigonometry must be continued for the last mentioned, perhaps not so far as in some countries to-day. (Complex numbers and spherical trigonometry can be omitted).

Solid geometry should — as mentioned before — stress the ability to visualize in three dimensions. Axiomatic points of view should not be considered to any great extent.

Analytical geometry has in many countries had a strong position — and in the opinions of many teachers — a deserved position. It is now time for reduction. The concept of locus, the (1,1) correspondence between curve and equation must be kept, likewise the equation of the straight line and the conics in Cartesian co-ordinates, but as an independent discipline analytical geometry will disappear and what is left must be treated with the (elementary).

Theory of functions, that should have a central position and a sound foundation — practically with graphs as background — theoretically in a sober definition.

Calculus should comprise differentiation and integration. The Norwegian plans are, I think, right when they say that "the first elements will be worked organically in the algebra, and the purpose is above all to lead the pupil into the infinitesimal manner of thinking, to make him familiar with the idea of limit value. A new mathematical calculating technique is not the chief aim". Curve discussions (variation of functions) and maximum and minimum must be central themes. The concept of differential should be used to continue

the treatment of approximate numbers, errors etc., that started in the lowest classes in the schools. Integration should not go too far; well founded ideas and full understanding are better than mechanical drill technique.

What now has been mentioned is more or less the normal, "classic" curriculum, but I have proposed some reductions, dropping themes that are less useful and thus making it possible to introduce more modern ones. Most important today is of course *statistics*, and I feel it urgent that all interested groups should discuss the place of statistics in schools.

In the begin the treatment should be practical: classification, distribution tables, absolute and relative frequences and histograms. Then follow with some measures of central tendency (means, averages). The range and the average deviation are easy to compute, but the measures are primitive. The computation of the far more useful standard deviation is complicated, though the exact development of the formula is not outside the capacities of the pupils.

The importance of the Gaussian, or normal distribution, is so great to-day that this theme should be a part of the "general education" given by schools. Even if it is not possible on that level to give a complete treatment, I still think it can be a satisfactory treatment. — Some ideas of correlation might well be given. (Scatterdiagram and regression).

More difficult is the question: in addition to the mentioned elements of the descriptive statistics, should some ideas of statistical induction, sampling theory, be given? I am inclined to answer yes.

Sampling theory requires a certain reasoning different from that needed in descriptive statistics, and it will take some time before the pupils will become accustomed to it. It also requires the elements of the probability calculus, but that is not too difficult for schools, if the classical definition is used (the limit-definition may be mentioned) and at the same time it is a subject which is valuable from a "general education" point of view.

I would however limit statistical induction to one case: mean and standard error in simple sampling of either attributes or variables. Again, it is more important to let the pupils get the sampling idea than to inform them of a lot of standard-error formula, or of a technique acquired on a "recipe-pattern", that is according to the principle: 1. do that, 2. do that, 3. do that — and so on.

We are again touching the chief problem: where to draw the borderline between school mathematics and university mathematics? The interests are different. The professors like to get students with well-founded basic knowledge, preferring themselves to handle the more subtle topics (the concept of limit, differential quotient etc.), in order to be sure that they are thoroughly and soundly treated.

School teachers prefer not to spend too much time on drill, learning of formulas, tools and technique, but want also to give their pupils some insight into the different possibilities of "modern" mathematics — even if they cannot do it at the same level of exactness as the universities.

How can we get over this conflict? Both parts must show a certain moderation. School teachers must not include in their programme topics that obviously can be better treated at the universities. University professors must understand the problems in the schools, not expect that all sorts of formulas are present, nor expect the same degree of exactness that they practise themselves.

This does not at all mean that schools should teach a sort of second-rate mathematics. Mathematics *is* mathematics in schools as in life. It is not permissible to swindle in schools, introducing tacit suppositions, using false or incomplete proofs, or breaches in the logic of the subject.

If the correct procedure is too difficult or complicated, the pupils must be told: „It is possible to prove the following problem". Plausible reasons might be given, but no false, pretended „proof". Of course we must avoid allowing mathematics to degenerate into a collection of unproved theorems, a cooke-book full of recipes. So often we see the misuse of statistics; statistical procedures or formulas are applied without insight into those necessary suppositions that must be valid in every single case.

If a special technique is highly important, professors and teachers must co-operate and try to transform the proof or the deduction of the formulas down to a level suited to the school pupils. Let us remember that many topics taught at universities a hundred years ago belong to the present-day curriculum in schools all over the world.

We have not profited from the vast research results of psychology during the last fifty years, and we have not invited psychology to assist us in solving new problems. The teachers of mathematics must consider the research work within psychology, arrange the curriculum according to maturity, find the optimal age for every topic, find out when children are mature for understanding the meaning of a proof, directed numbers, irrational numbers etc, further try to find methods of learning which best lead to the aims, or which make it possible to reach those aims with less effort. We must also motivate the pupils in their learning and thus make mathematics more popular.

I am well aware that much has been done in this direction in some countries, but I also know that much more is left to be done. The teaching of mathematics can and must be more effective and more useful than it is to-day. But this is only one side of the teaching problem. It should also give clear insight into the nature of mathematics — ideas of the rich possibilities of

the subject, but also of its limitations. Teaching should promote fair, honest and sober attitudes concerning mathematics amongst the pupils, and it should give real and lasting values for those who never in further studies nor in life meet or need mathematics — if such persons exist.

SYMPOSIA

SYMPOSIUM A

STOCHASTIC PROCESSES

under the care of D. VAN DANTZIG

Honorary President: R. M. FRÉCHET

OPENING SPEECH BY D. VAN DANTZIG

Ladies and gentlemen,

It is a great pleasure for me to welcome you all on behalf of the „Wiskundig Genootschap" in Amsterdam at the symposium on Stochastic Processes, the first of the three Symposia it has organized on the occasion of the International Mathematical Congress.

We are happy that a number of so prominent mathematicians have accepted our invitation to give a conference before the symposium or to participate in the discussion.

Professor Fréchet, who has done so much work of great importance on the subject of this symposium, thus continuing the great tradition of France, the country which since the birth of probability theory in 1653 has offered the world an unbroken line of probabilists of the highest order, has honoured us by accepting the honorary presidency of this symposium, and has kindly offered us a presidential address. For this we thank him sincerely, and we regret greatly that he is prevented to attend the symposium. We also thank professor Lévy, who kindly consented to read Frichet's presidential address.

As the „Wiskundig Genootschap" has placed this symposium under my care, so that I do not act as a chairman, but only as a general "careman", it has invited Professor Loève to take the chair during this morning's meeting, Professor Blackwell this afternoon, and Professor Lévy on Saturday morning. We offer them our thanks for having accepted this invitation.

We are well aware of the fact that not all stochasticists of the first rank have been invited to participate in this symposium. In those cases, where mathematicians, who, according to their scientific merits doubtless should have been invited, either to offer a paper, or to participate in the discussion, were not asked to do so, I beg them and you all to believe that we did our utmost to minimize the number of such cases, inasfar as limited funds, and before all, lacking time and organizational difficulties, allowed us.

The „Wiskundig Genootschap" wants to thank UNESCO for its financial support, without which this meeting would not have become possible, and the Organizing Committee of the International Mathematical Congress for its organizational aid, without which it would never have become real. Moreover

it wishes to thank the speakers for the trouble they have taken in preparing their lectures, Professor Doob who has kindly accepted the invitation to give a general survey of the subject, and the discussiants for their willingness to further the exchange of knowledge and the mutual understanding, for these mathematicians will have the merit of having made this symposium success ful, as we hope and trust it will be. Finally we thank the other participants for their presence, for the interest they show thereby in the subject, providing the symposium with substance, and with the resonance which will cause the conferences given here to be heard outside these four walls.

The choice of a subject for this symposium — a subject from the field of probability theory and statistics — was not a very difficult one. There were a number of possibilities, several of which were attractive. The place, however, taken by the theory of stochastic processes, is of such importance and generality, that this subject seemed to be especially suitable for a symposium. From a purely mathematical point of view it ranks among the fields of greatest beauty. On the other hand, the applications of the theory of stochastic processes are so numerous, so well known to all of you, that I need not try to enumerate them.

It has even been said that whatever happens is part of a process and contains some stochastic element, so that application of the theory of stochastic processes is by no means a privilege of mathematicians, statisticians, physicists, biologists, econometrists, etc., etc., but that no one can do anything at all without applying this theory in a practical sense.

But let us not draw out and inflate the meaning of a word like bubble-gum, but keep it neatly wrapped up.

Even then it can be said that the activity in this field has greatly increased during the last 10 years. There is e.g. a marked difference in this respect between the contents of the Proceedings of the first and the second Berkeley Symposium on Mathematical Statistics and Probability, held in 1945—1946 and 1950 respectively. Of the 16 theoretical papers of the first symposium, which are not concerned with special fields of application, only two deal with stochastic processes. Of the 27 papers of similar kind of the second symposium, some 15 are concerned with decision functions and stochastic processes. Although these numbers are a little uncertain, due to the unexact definitions of the fields, the increase is spectacular. Looking through the litterature of the last few years, a similar increase is found and the appearance of the books of Fréchet in second edition, of Lévy, of Doob and of Blanc-Lapierre and Fortet are a proof of the intensification of interest in large circles of mathematicians.

The subject, however, is not an easy one. A profound knowledge of it is

necessary even for the applications. A personal contact between those mathematicians, who have worked on the subject already and those who may profit from the results hitherto reached, seems of paramount inportance for its further development and applicability. The intention of the present symposium is to further these aims and, hoping that it will be a success in this respect, I have the honour to declare this symposium opened.

ALLOCUTION DU PRÉSIDENT D'HONNEUR R. M. FRÉCHET

Mes chers collègues,

Permettez-moi, d'abord, d'exprimer publiquement à la Société Mathématique néerlandaise et au Comité d'Organisation de ce Congrès, combien j'ai été touché par la marque d'estime qu'ils m'ont témoignée en me nommant Président d'honneur de ce colloque — l'un des trois colloques internationaux qui vont compléter de façon si heureuse les travaux du Congrès International des Mathematiciens.

Vous aurez surtout ici à présenter ou à discuter des contributions à la théorie des processus stochastiques.

A cette occasion, il est intéressant de noter qu'on a été conduit simultanément à cette théorie par deux voies entièrement différentes; d'une part, poussé par la tendance mathématique traditionelle à la généralisation, d'autre part en vue de répondre aux besoins des applications.

Ce colloque offre une occasion de rendre hommage aux précurseurs de la théorie des processus stochastiques. Parmi ceux-ci, plusieurs noms seraient à citer; nous nous limiterons à trois d'entre eux, parce que l'importance de leurs initiatives semble souvent méconnue dans les historiques ou les listes bibliographiques récentes.

C'est sans doute Slutsky, qui a le premier introduit pour une fonction aléatoire quelconque, la notion de dérivée stochastique en généralisant des notions fondamentales du Calcul des Probabilités. Et, d'autre part, les besoins de la météorologie, auxquels avaient profondément réfléchi Wehrlé et Dedebant, les ont conduit, à peu près à la même époque que Slutsky, à formuler une définition analogue à la sienne. Réfléchissons maintenant sur les fondements de la théorie. Dans les applications, le processus stochastique représente généralement l'évolution d'une nombre aléatoire X dans le cours du temps t. Cependant la théorie mathématique correspondante garde son sens quand t est un paramètre quelconque. Quand on porte son attention sur la suite des épreuves qui se déroule dans le temps, on a à scruter une suite de valeurs de X qui peut varier très irrégulièrement quand le temps croît. Autrement dit, cette suite d'épreuves définit une fonction $X(t)$ qui n'est généralement, ni dérivable, ni continue. C'est la principale difficulté de la théorie des processus stochas-

tiques. Elle a conduit à abandonner la continuité et la dérivabilité classiques *qui disparaissent ici*, par la continuité et la dérivabilité ,,stochastiques''.

Il semble, au premier abord, que la théorie des ,,processus stochastiques'' ne soit qu'un autre nom donné à la théorie des ,,fonctions aléatoires'' dont nous allons maintenant nous occuper. Pourtant deux différences séparent ces théories. D'une part, dans les applications les plus fréquentes des processus stochastiques, le paramètre t est le temps, qui est irréversible et, par suite, la loi de probabilité du nombre aléatoire $X(t_0)$ ne dépend que des valeurs prises par $X(t)$ aux époques antérieures à t_0. Au contraire, pour la fonction aléatoire $X(t)$ la plus générale, la loi de probabilité de $X(t_0)$ peut dépendre des valeurs de $X(t)$ prises pour toutes les valeurs de t autre que t_0.

Mais il y a une autre différence, plus essentielle encore. Si l'on considère la suite d'épreuves correspondant aux valeurs de t, comme une seule épreuve, alors, à chacune de ces épreuves d'une nouvelle catégorie, correspondra non pas un nombre mais une fonction. On peut ainsi considérer $X(t)$ comme une fonction choisie *toute entière d'un seul coup*, mais, au hasard, dans chaque épreuve. Dans les applications qui correspondent naturellement à cette façon de voir, on fait un choix au hasard dans une famille de fonctions *donnée d'avance*. Ces fonctions, de nature concrète, pourront généralement être considérées comme continues, souvent comme dérivables, parfois même comme analytiques ou même comme plus simples encore. Les problèmes qui se posent seront de nature différente de ceux qui se présentent pour les processus stochastiques. Les fonctions considérées auront une régularité, connue d'avance, qui ne soulèvera pas de difficultés. Ce qu'il faudra faire, c'est définir les fonctions moyennes, les fonctions typiques, la dispersion, les fonctions caractéristiques, etc. Mais au moment d'aborder ces questions, on pourra se demander s'il ne serait pas aussi simple et plus profitable de se placer dans un cas plus général.

On a étudié successivement les nombres aléatoires, les points aléatoires dans l'espace à deux ou trois dimensions; on étudie maintenant les fonctions aléatoires. Le cycle est-il fermé? N'y a-t-il pas déjà d'autres éléments aléatoires dont l'étude devrait s'imposer dans la Science et dans la Technique? De nombreux exemples s'offrent à l'esprit pour répondre par l'affirmative.

Parmi les plus simples, citons les courbes et les surfaces aléatoires qui abondent dans la Nature. C'est ainsi que nous avons fait relever et analyser deux collections de sections craniennes (dont les différentes formes sont d'un grand intérêt pour l'anthropologie). De même, nous avons fait jeter cent fois un fil, au hasard. Et nous avons fait relever et analyser les formes très diverses prises par le fil en tombant sur le sol. En choisissant au hasard un oeuf dans une corbeille d'oeufs, on a un example de surface aléatoire. Etc.

On pourrait peut-être objecter: Mais toute courbe, toute surface peut-être représentée au moyen de fonctions, il n'y a rien de nouveau à les considérer. Pourtant, observons que si l'on sait bien ce qu'est la somme de deux fonctions, on n'a pas encore, jusqu'ici, donné une définition satisfaisante de la somme de deux courbes, de deux surfaces. Cela suffit pour se rendre compte que l'étude des courbes aléatoires constitue bien un problème tout à fait nouveau.

Mais alors, faudra-t-il faire séparément et successivement une théorie des nombres aléatoires, une autre des vecteurs aléatoires, une autre des fonctions aléatoires, une autre des courbes aléatoires, etc., comme on a fait autrefois une théorie des forces, une autre des moments, une autre des vitesses, etc. Ou bien s'inspirera-t-on de l'exemple de l'Analyse vectorielle, laquelle a su grouper les propriétés communes de ces différentes grandeurs, en ne laissant aux théories particulières que ce qu'elles avaient de réellement particulier, et en simplifiant *d'avance* les théories de nouvelles grandeurs dont les applications allaient plus tard introduire la considération?

Ce sont ces observations qui nous ont amené à nous demander s'il ne serait pas possible d'édifier *une théorie générale* des éléments aléatoires *de nature quelconque*?

La question est maintenant résolue. Cela est possible et, même, nous avons été étonné de voir combien sont nombreuses les définitions et les propriétés classiques des nombres aléatoires qu'on peut étendre de façon raisonnable et sans complications excessives aux éléments aléatoires de nature quelconque. Il ne saurait être question de retarder vos travaux en expliquant, même brièvement, comment cela a pu être réalisé. Nous nous contenterons donc de renvoyer, pour plus de details, aux mémoires originaux dont on trouvera une liste à la suite d'un résumé de la théorie des éléments aléatoires de nature quelconque au Chapitre VI de la *Deuxieme* Edition de notre ouvrage *„Recherches théoriques modernes sur le Calcul des Probabilités"*.

Nous tenons cependant à dire que nos travaux sur cette nouvelle théorie ont été très heureusement poursuivis et complétés par quelques jeunes mathématiciens parmi lesquels je nommerai dans l'ordre chronologique approximatif de leurs contributions: Mlle Mourier, M. M. Shafik Doss, Fortet, Nasr, Badrekian. Récemment le professeur Van Dantzig, s'est aussi — à coté de tant de travaux dans de si multiples directions — intéressé efficacement au même domaine de recherches.

Mes chers collègues, je m'excuse d'avoir donné des explications qui étaient inutiles pour plusieurs d'entre vous. Mais la théorie des éléments aléatoires de nature quelconque est encore peu connue et il m'a semblé utile de commencer ce Colloque en attirant l'attention des autres sur ce nouveau

Chapitre du Calcul des Probabilités, Chapitre qui est en connexion étroite avec les processus stochastiques aux statisticiens qu'aux théoriciens du Calcul des Probabilités.

Je n'ai plus maintenant que l'agréable devoir d'exprimer, non pas l'espoir mais la certitude, que l'objet de ce Colleque recevra de vos communications une impulsion nouvelle.

CONTROLLED RANDOM WALKS

DAVID BLACKWELL

1. *Introduction.* Let $M = ||m_{ij}||$ be an $r \times s$ matrix whose elements m_{ij} are probability distributions on the Borel sets of a closed bounded convex subset X of k-space. We associate with M a game between two players, I and II, with the following infinite sequence of moves, where $n = 0, 1, 2, \ldots$:

Move $4n + 1$: I selects $i = 1, \ldots, r$.
Move $4n + 2$: II selects $j = 1, \ldots, s$ not knowing the choice of I at move $4n + 1$.
Move $4n + 3$: a point x is selected according to the distribution m_{ij}.
Move $4n + 4$: x is announced to I and II.

Thus, a mixed strategy for I is a function f, defined for all finite sequences $a = (a_1, \ldots, a_n)$ with $a_k \in X$, $n = 0, 1, 2, \ldots$, with values in the set P_r of r-vectors $p = (p_1, \ldots, p_r)$, $p_i \geqq 0$, $\sum p_i = 1$: the ith coordinate of $f(a_1, \ldots, a_n)$ specifies the probability of selecting i at move $4n + 1$ when $a_1, \ldots, a_n$ are the X-points produced during the first $4n$ moves. A strategy g for II is similar, except that its values are in P_s. For a given pair f, g of strategies, the X-points produced are a sequence of random vectors $x_1, x_2, \ldots$, such that the conditional distribution of x_{n+1} given $x_1, \ldots, x_n$ is $\sum_{i,j} f_i(x_1, \ldots, x_n)\, m_{ij} g_j(x_1, \ldots, x_n)$, where f_i, g_j are the ith and jth coordinates of f, g.

The problem to be considered in this paper is the following: To what extent can a given player control the limiting behavior of the random variables $\bar{x}_n = (x_1 + \ldots + x_n)/n$? For a given closed nonempty subset S of X, we shall denote by $H(f,g)$ the probability that $\bar{x}_n$ approaches S as $n \to \infty$, i.e., the distance from the point $\bar{x}_n$ to the set S approaches zero, where $x_1, x_2, \ldots$ is the sequence of random variables determined by f, g. We shall say that S is *approachable by* I *with* f^* (II *with* g^*) if $H(f^*,g) = 1$ ($H(f, g^*) = 1$) for all $g(f)$, and shall say that S is *approachable by* I (II) if there is an $f(g)$ such that S is approachable by I with f (II with g). We shall say that S is *excludable by* I *with* f if there is a closed T disjoint from S which is approachable by I with f. *Excludability by* II *with* g, *excludability by* I, and *excludability by* II are defined in the obvious way.

It is clear that no S can be simultaneously approachable by I and excludable by II. The main result to be described below is that every convex S is

either approachable by I or excludable by II; a fairly simple necessary and sufficient condition for a convex S to be approachable by I is given, a specific f which achieves approachability is described, and an application is given. Finally, an example of a (necessarily nonconvex) S which is neither approachable by I nor excludable by II is given, and some unsolved problems are mentioned.

2. *The main result.* For any $p \in P_r (q \in P_s)$ denote by $R(p)$ $(T(q))$ the convex hull of the $s(r)$ points $\sum_i p_i \overline{m}_{ij}$, $j = 1, \ldots, s$ $(\sum_j \overline{m}_{ij} q_j, i = 1, \ldots, r)$ where $\overline{m}_{ij}$ is the mean of the distribution m_{ij}. By selecting i with distribution q at a given stage, I forces the mean of the vector x selected at that stage into $R(p)$, and no further control over the mean of x is possible. It is intuitively plausible, and true, that $R(p)$ $(T(q))$ is approachable by I (II) with $f \equiv p$ $(g \equiv q)$. Thus, unless S intersects every $T(q)$, it is excludable by II and hence not approachable by I. It turns out that any convex S which intersects every $T(q)$ is approachable by I; a more complete statement is

Theorem 1. *For any closed convex S, the following conditions are equivalent:*

(a) *S is approachable by* I.

(b) *S intersects every $T(q)$.*

(c) *For every supporting hyperplane H of S, there is a p such that $R(p)$ and S are on the same side of H.*

If S is approachable by I, it is approachable by I with f defined as follows. For any $a = (a_1, \ldots, a_n)$ for which $\bar{a} = (a_1 + \ldots + a_n)/n \in S$, $f(a)$ is arbitrary. If $\bar{a} \notin S$, $f(a)$ is any $p \in P_r$ such that $R(p)$ and S are on the same side of H, where H is the supporting hyperplane of S through the closest point s_0 of S to $\bar{a}$ and perpendicular to the line segment joining $\bar{a}$ and s_2.

Theorem 1 is proved in [1]; equivalence of (b) and (c) is an immediate consequence of the von Neumann minimax theorem [2], while the proof of the rest of the theorem is complicated in detail, though the main idea is simple.

3. *An application.* As an application of Theorem 1, we deduce a result of Hannan and Gaddum. This result concerns the repeated playing of a zero-sum two person game with $r \times s$ payoff matrix $A = ||a_{ij}||$. If the game is to be played N times (N large), and I knows in advance that the number of times II will choose j is Nq_j, $j = 1, \ldots, s$, he can achieve the average amount $h(q) = \max_i \sum a_{ij} q_j$. Hannan and Gaddum show that, without knowing q in advance I can play so that, for any q, I's averge income is almost $h(q)$; in our terminology, this result is the following:

Let M be the $r \times s$ matrix with $m_{ij} = (\delta_j, a_{ij})$, where δ_j is the jth unit vector in s-space. The set S consisting of all (q, y) such that $y \geqq h(q)$ is approachable by I.

This follows immediately from condition (b) of Theorem 1, for $T(q)$ is the

convex hull of the r points $(q, \sum a_{ij}q_j)$, and one of these is the point $(q, h(q))$, so that $T(q)$ intersects S.

4. *An example.* If $k = 1$, every closed S is either approachable by I or excludable by II. For $k = 2$, there are sets which are neither; an example is:

$$M = \begin{Vmatrix} (0, 0) & (0, 0) \\ (1, 0) & (1, 1) \end{Vmatrix},$$

$S = A\ B$, where A is the line segment joining $(\frac{1}{2}, 0)$, $(\frac{1}{2}, \frac{1}{4})$ and B is the line segment joining $(1, \frac{1}{2})$ and (1,1). The strategy g with $g(a_1, \ldots, a_n) = 1$ for $u_{2n} \leqq n < u_{2n+1}$, $g = 2$ otherwise, where $\{u_n\}$ is a sequence of integers becoming infinite so fast that $(u_1 + \ldots + u_n)/u_{n+1} \to 0$ forces $\bar{x}_n$ to oscillate between the lines $y = 0$ and $y = x$, so that $\bar{x}_n$ cannot converge to S, and S is not approachable by I. On the other hand, I can force $\bar{x}_n$ to come arbitrarily near S infinitely often as follows. By choosing 2 successively a number of times large in comparison with the number of previous trials, I forces an $\bar{x}_n$ near $(1, a)$ for some a, $0 \leqq a \leqq 1$. If $a \geqq \frac{1}{2}$, $\bar{x}_n$ is near S; if $a < \frac{1}{2}$, by choosing 1 n times in succession, I forces $\bar{x}_{2n}$ to be approximately $\left(\frac{1}{2}, \frac{a}{2}\right)$, which is in S. Thus S is neither approachable by I nor excludable by II.

5. *Some unsolved problems.*

A. Find a necessary and sufficient condition for approachability. This problem has not been solved even for the example of section 4.

B. Call a closed S *weakly approachable* by I if there is a sequence of strategies f_n such that for every $\varepsilon > 0$,

$$\sup_g \text{Prob } \{\varrho(\bar{x}_n(f_n, g), S) > \varepsilon\} \to 0$$

as $n \to \infty$, where $\varrho(x, S)$ is the distance from x to S.

Define weak approachability by II similarly, and call S *weakly excludable* by II if there is a closed T disjoint from S which is weakly approachable by II. Is every S either weakly approachable by I or weakly excludable by II? For the example of section 4, the answer is yes.

C. Does the class of (weakly) approachable sets for a given M depend only on the matrix of mean values of M?

REFERENCES.

[1] DAVID BLACKWELL, "An analog of the minimax theorem for vector payoffs," to appear in the *Pacific Journal of Mathematics*.

[2] J. VON NEUMANN and O. MORGENSTERN, *Theory of Games and Economic Behavior*, Princeton, 1944.

UNIVERSITY OF CALIFORNIA, BERKELEY

SUR QUELQUES APPLICATIONS DE LA THÉORIE DES FONCTIONS ALÉATOIRES À L'OPTIQUE

A. BLANC-LAPIERRE

Introduction.

Les problèmes de bruit de fond en électricité ont fourni un champ d'application très riche à la théorie des fonctions aléatoires. Les ondes lumineuses étant aussi de nature électromagnétique, cette théorie doit également pouvoir rendre des services dans la description statistique des grandeurs lumineuses.

Un exemple précisera les liens physiques existant entre ces deux champs d'applications. On sait qu'une résistance R, à la température T, est le siège de fluctuations de potentiel que l'on peut décrire en introduisant une force électromotrice de fluctuations (fictive) laplacienne en série avec R, et de densité spectrale constante $4\,kRT$ (théorème de *Nyquist*). On connaît d'autre part la loi de *Planck* sur le rayonnement isotherme. En fait le théorème de Nyquist est une conséquence directe de la *loi de Rayleigh* à laquelle se réduit la loi de *Planck* dans le domaine des fréquences radioélectriques pour lesquelles $\frac{h\nu}{kT} \ll 1$.

Cela découle de l'étude des échanges d'énergie entre le rayonnement situé dans une enceinte à la température T et la résistance R convenablement couplée à ce rayonnement par un adaptateur et une antenne [1].

Je me propose de rassembler ici quelques calculs et quelques résultats, dont certains sont d'ailleurs connus, relatifs à des applications de la théorie des fonctions aléatoires à l'optique. La variable lumineuse sera traitée comme une *grandeur scalaire* dans la première partie, puis, dans la seconde partie, comme une *grandeur vectorielle.*

I. Variable lumineuse scalaire

A. Nature laplacienne de la Variable lumineuse.

Soit $v(M, t)$ la valeur au point M et à l'instant t de la variable lumineuse; *si on suppose que M n'est pas situé sur la source lumineuse* (incohérente à une échelle fine par rapport à nos moyens d'observation) et *si le flux lumineux n'est pas trop faible,* on déduit, des lois de l'addition des variables aléatoires indépendantes, que $v(M, t)$ est une *fonction aléatoire laplacienne*; nous la supposerons *stationnaire en t* (source macroscopiquement constante).

En liaison avec le caractère laplacien de $v(M, t)$, on peut faire quelques remarques simples dans le cadre de l'ancienne théorie classique du rayonnement, qui assimile l'action d'une source lumineuse à celle d'oscillateurs amortis, relancés à des instants distribués au hasard. On associe alors à la source une vibration $\sigma(t) = \Sigma_{(t_j)} R(t - t_j)$ où les t_j sont répartis dans le temps suivant une distribution de Poisson de densité ϱ; R représente un train d'oscillations amorti. Loin de la source, et dans un domaine pas trop grand, la variable lumineuse est proportionnelle à $\sigma(t - x/c)$ [x = distance à la source; c = vitesse de propagation]. Soit θ la constante de temps de $R(t)$. *Il est bien connu que σ est laplacien si $\varrho\theta \gg 1$, c'est-à-dire si les trains d'oscillations interfèrent profondément.*

Considérons *l'énergie emmagasinée* dans le volume du parallélépipède centré sur x, limité par les plans $x_0 - l$ et $x_0 + l$ et de surface de base unité normalement à l'axe des x. A l'instant t, cette énergie est égale à l'intégrale $W(l, t)$ de $\sigma^2(x, t)$ dans le domaine $x_0 - l < x < x_0 + l$. Calculons son espérance mathématique $a[l]$ et sa fluctuation $\delta[l]$. Pour $l \gg c\theta$, on trouve

$$a[l] = 2l\varrho\mu_1(0) \tag{1.A.1}$$

$$\delta^2(l) = 2lc[\varrho \int_{-\infty}^{+\infty} \mu_2(\theta)d\theta + 2\varrho^2 \int_{-\infty}^{+\infty} \mu_1^2(\theta)d\theta] \tag{1.A.2}$$

avec

$$\mu_1(\theta) = \int_{-\infty}^{+\infty} R(t)R(t - \theta)dt \text{ et } \mu_2(\theta) = \int_{-\infty}^{+\infty} R^2(t)R^2(t - \theta)dt \tag{1.A.3}$$

Il est intéressant de voir ce que donne la formule (1.A.3) en prenant pour $R(t)$ un paquet d'ondes

$$R(t) = \int_{-\infty}^{+\infty} \cos\ 2\pi\nu t\,.\,e^{-\frac{(\nu-\nu_0)^2}{2\Delta^2}} d\nu = \sqrt{2\pi}\,\Delta \cos 2\pi\nu_0 t\,.\,e^{-2\pi^2 t^2\Delta^2} \tag{1.A.4}$$

On trouve alors, pour $\Delta \ll \nu_0$ (ce que je supposerai):

$$a(l) = l\varrho\Delta\sqrt{\pi} \text{ et } \delta^2(l) = 2lc\{\varrho\Delta^2\frac{\pi}{4} + \frac{\Delta}{8}\varrho^2\sqrt{2\pi}\} \tag{1.A.5}$$

δ^2 est la somme de deux termes: l'un, prépondérant pour ϱ petit, varie comme $a(l)$; l'autre, prépondérant pour ϱ grand, varie comme $a^2(l)$. Seul, ce dernier est à considérer si σ est approximativement laplacien. La formule (1.A.5) est analogue à celle qui, en mécanique statistique quantique, donne le carré moyen $\overline{[\delta W\nu]^2}$ des fluctuations de l'énergie $W\nu$ dans le rayonnement isotherme de fréquence ν (à $d\nu$ prés) pour un élément de volume dv.

On a alors:

$$\overline{[\delta W\nu]^2} = \overline{(W\nu - \overline{W\nu})^2} = h\nu\overline{W\nu} + \frac{\overline{W\nu}^2}{g_t} \tag{1.A.6}$$

où g_i est le nombre d'états distincts correspondant au domaine $dv \times dv$. Ce rapprochement est saisissant si on introduit *l'énergie moyenne* ε_0 et l'étalement Δx de chaque train d'oscillations auxquels on peut donner les valeurs

$$\varepsilon_0 = c \int_{-\infty}^{+\infty} R^2(t)dt = c\Delta \frac{\sqrt{\pi}}{2}, \quad \Delta x = c/\Delta\sqrt{2\pi} \tag{1.A.7}$$

(le coefficient $1/\sqrt{2\pi}$ a été choisi pour que l'on ait $\Delta v.\Delta t = 1$).

On peut dire que la longueur $2l$ du domaine considéré peut contenir un nombre de trains d'oscillations successifs égal à $g_i' = 2\,l/\Delta x$ (1.A.5) devient alors:

$$\delta^2(l) = \varepsilon_0 a(l) + \frac{a^2(l)}{g_i'} \tag{1.A.8}$$

C'est exactement la formule (1.A.6) si on remplace g_i par g_i' et ε_0 par hv. Il n' y a pas à insister sur cette analogie de formules et *il est bien évident que* (1.A.8) *ne saurait constituer une démonstration de* (1.A.6) mais cette ressemblance suggère une certaine correspondance entre les termes des seconds membres des relations (1.A.6) et (1.A.8). Dans le modèle classique, le caractère laplacien correspond à la possibilité de négliger le terme en a devant celui en a^2; par suite de l'analogie signalée, il peut paraître raisonnable de penser que cette situation correspond à la possibilité de négliger le terme en $\overline{Wv}$ devant celui en $\overline{W_v^2}$ ce qui est possible s' il y a une grande densité de photons dans le volume considéré [2].

B. Cohérence des sources.

La plupart des résultats pratiques relatifs au degré de cohérence des sources *en lumière monochromatique* sont contenus dans deux articles de H. H. Hopkins (3) qui traite cette question indépendamment de tout lien avec la théorie des fonctions aléatoires. En fait, les problèmes de cohérence ne sont que des cas particuliers de questions liées à la théorie de la corrélation et, en les replaçant dans leur véritable cadre mathématique, on rend leur exposé plus systématique; il n' y a d'ailleurs aucune difficulté pour étendre la théorie au cas de *sources non monochromatiques* [2].

Soit une source plane $A[\alpha \times \beta]$. A un petit élément $(d\alpha.d\beta) = dM$ d'aire $d\varphi = d\alpha d\beta$ entourant un point M, associons la vibration élémentaire

$$ds[\alpha, \beta; t] = ds[M; t]$$

ds est une valeur réelle fonction aléatoire de l'élément $dM = d\alpha d\beta$ et du temps t. Il est possible qu'il existe une fonction aléatoire $\sigma[M, t] = ds/dM$; cela n'est pas obligatoire. En particulier, σ n'existe pas pour une source incohérente. Pour ne pas restreindre la généralité, il suffit de raisonner sur l'intégrale $s[\alpha, \beta; t] = s[M; t]$ de $ds(\alpha', \beta'; t)$ sur un domaine rectangulaire $D(\alpha, \beta)[\alpha_0 < \alpha' \leqq \alpha; \beta_0 < \beta' \leqq \beta]$ avec une convention de signe convenable.

Sous des conditions très générales, on écrira alors:

$$s(M, t) = \int_{-\infty}^{+\infty} e^{2\pi i\nu t}\, d_\nu\, S(M, \nu) = \int_{-\infty}^{+\infty} \int_{D(\alpha, \beta)} e^{2\pi i\nu t}\, d_{\nu\times M'}\, S'(\nu, M') \quad \text{(1.B.1)}$$

Tous les problèmes pratiques relatifs à l'état de cohérence des divers éléments de la source A se ramènent à des questions de distribution de puissance moyenne dans des phénomènes d'interférences; ils ne font donc intervenir que des *moyennes du second ordre* et ne dépendent que de la donnée des éléments différentiels suivants, dont les connaissances sont d'ailleurs équivalentes:

$$\Delta[M_1, M_2; t_1, t_2] = E\,\{ds[M_1, t_1]ds[M_2, t_2]\} \quad \text{(1.B.2)}$$

$$\Delta'[\nu_1, \nu_2; M_1, M_2] = E\,\{d_{\nu\times M}\, S'(\nu_1, M_1) d^*_{\nu\times M}\, S'(\nu_2, M_2)\} \quad \text{(1.B.3)}$$

ou E représente l'espérance mathématique.

Le caractère stationnaire en t implique d'ailleurs que Δ ne dépend que de $t_1 - t_2$, et que $\Delta' = 0$ pour $\nu_1 \neq \nu_2$. Il serait trop long de décelopper ici toutes les utilisations que l'on peut faire du formalisme qui précède, bornons-nous à montrer comment il permet une définition précise de la cohérence d'une source. Soient deux petits éléments dM_1 et dM_2 de A, $ds(M_1, t)$ et $ds(M_2, t)$ les vibrations correspondantes *quand dirons-nous que ces deux sources élémentaires sont cohérentes*? Pour répondre à cette question, comparons les figures d'interférences obtenues en faisant interférer:

a. $ds(M_1, t)$ avec lui-même [cas I];

b) $ds(M_2, t)$ avec lui-même [cas II];

c) $ds(M_1, t)$ avec $ds(M_2, t)$ [cas III].

Nous dirons qu'il y a cohérence si, compte tenu des précisions suivantes, ces 3 *figures d'interférences sont identiques.* Nous supposons que, grâce à des écrans absorbants neutres, nous avons ramené nos deux sources élémentaires à avoir même intensité. Le schéma mathématique de nos expériences est le suivant: A chaque point d'un domaine (ϱ) de l'espace, nous associons une certaine *différence de marche l.* Organisons nos expériences dans un *milieu non dispersif* de sorte qu'à une valeur donnée de l corresponde un *retard* τ bien défini. Nous admettons que des miroirs peuvent exister dans notre dispositif interférentiel ce qui pourrait remplacer ds par $-ds$. Dans l'un ou l'autre cas, la figure d'interférence peut être caractérisée par la densité $I(\tau)$ de la puissance moyenne en fonction de τ ou mieux par la fonction $J[\tau] = (I[\tau] - I[\infty])/I(\infty)$. Si J_1, J_2, J_3 sont les fonctions J correspondant aux cas I, II, III, la condition de cohérence sera qu'il existe une valeur τ_0 telle que:

$$|J_1(\tau)| = |J_2(\tau)| = |J_3(\tau + \tau_0)| \quad \text{(1.B.4)}$$

On peut alors établir le résultat suivant: *La condition nécessaire et suffisante pour que les deux éléments* dM_1 *et* dM_2 *soient cohérents est qu'il existe deux nombres*

réels k et τ_0 indépendants de t tels que les deux conditions équivalentes suivantes soient remplies presque sûrement:

$$ds(M_1, t) = k\, ds(M_2, t - \tau_0) \tag{1.B.5}$$

$$d_{\nu \times M}\, S'(\nu, M_1) \equiv k\, e^{-2\pi i \nu \tau_0}\, d_{\nu \times M}\, S'(\nu, M_2) \tag{1.B.6}$$

Le coefficient

$$\mu = \frac{E\{d_{\nu \times M}\, S'(\nu, M_1) d^*_{\nu \times M}\, S'(\nu, M_2)\}}{\sqrt{E\{|d_{\nu \times M}\, S'(\nu, M_1)|^2\}\; E\{|d_{\nu \times M}\, S'(\nu, M_2)|^2\}}} \tag{1.B.7}$$

qui joue le rôle d'un coefficient de corrélation généralisé satisfait alors à

$$\mu \equiv \pm\, e^{-2\pi i \nu \tau_0} \tag{1.B.8}$$

Mais (1.B.8) ne suffit pas pour entraîner la validité de (1.B.5) et (1.B.6) avec *k indépendant de ν*. Si, pour une valeur de ν, on a $|\mu| = 1$ (coefficient de Hopkins égal à 1), les deux sources présentent toutes les caractéristiques de la cohérence *à condition de filtrer le rayonnement de façon à ne laisser passer que la fréquence ν*. Nous dirons qu'*il y a alors cohérence monochromatique pour la fréquence ν; il faut noter que, s'il y a cohérence monochromatique pour toutes les fréquences ν d'un certain ensemble D, il n' y a pas, en général, cohérence au sens de* (1.B.4) *pour l'ensemble des fréquences D.*

Ne nous limitons plus maintenant aux deux éléments dM_1 et dM_2 de A, mais considérons dans leur ensemble tous les éléments de A. Nous dirons que la source A est cohérente si tous ses éléments sont cohérents deux à deux. Nous dirons que la source A présente la cohérence monochromatique, si tous ses éléments présentent deux à deux la cohérence monochromatique pour toutes les fréquences.

On peut alors énoncer les résultats suivants:

1°. Pour que A présente la cohérence monochromatique, il faut et il suffit que $s(M, t)$ puisse se mettre sous la forme

$$s(M, t) = \int_{-\infty}^{+\infty} e^{2\pi i \nu t} K[M, \nu] dy(\nu) \tag{1.B.9}$$

où: $K[M, \nu]$ est une fonction certaine *à valeur complexe* définie par la somme de ses accroissements $dK_M(M, \nu)$

$$\text{sur } D(\alpha, \beta) \text{ et telle que } dK_M[M, \nu] = dK^*_M[M, -\nu]$$

$y(\nu)$ est une fonction aléatoire à accroissements non corrélés telle que $dy(\nu) = dy^*(-\nu)$.

2°. Pour que A soit cohérente, il faut et il suffit que, en plus des conditions précédentes, on ait:

$$dK_M[M, \nu] = d_M K'(M)\, e^{-2\pi i \nu \tau(M)} \tag{1.B.10}$$

où: $K'(M)$ et $\tau(M)$ sont deux fonctions *réelles* de M.

3^0. Il y aura incohérence si, pour deux éléments dM_1 et dM_2 disjoints quelconques, on n'observe aucun phénomène d'interférence pour le cas III. La condition, pour qu'il en soit ainsi, est que, se reportant à (1.B.2) ou (1.B.3) on ait $\Delta = \Delta' = 0$ pour $M_1 \neq M_2$. La fonction $S'(\nu, M)$ intervenant dans (1.*B*.1) est alors à accroissements non corrélés dans $\nu \times M$.

4^0. S'il n' y a ni cohérence, ni incohérence, on dira qu'il y a cohérence partielle. Tous les problèmes d'imagerie correspondants se traitent à partir des propriétés des éléments différentiels introduits en (1.B.2) ou (1.B.3).

II. Variable lumineuse vectorielle

Donnons un exemple d'application où on prend en considération le caractère vectoriel de la variable lumineuse. Il s'agit de la description des *propriétés statistiques du vecteur de Fresnel correspondant à une onde plane d'état de polarisation quelconque* (le cas de la lumière naturelle a été traité par Hurwitz (4) dès 1945). On peut, à chaque instant, représenter le vecteur de Fresnel par un nombre complexe $Z(t)$. $Z(t)$ est une fonction aléatoire stationnaire laplacienne [5]. L'étude de $Z(t)$ est intéressante dans le cas du *quasi monochromatisme*, c'est-à-dire lorsque l'on peut écrire [6]:

$$Z(t) = \Delta X_+(t)\, e^{2\pi i \mu_0 t} + \Delta X_-(t)\, e^{-2\pi i \mu_0 t} \tag{2.1}$$

avec

$$\Delta X_+(t) = \int_{\mu_0 - \Delta\mu}^{\mu_0 + \Delta\mu} e^{2\pi i(\mu - \mu_0)t}\, dx_+(\mu);$$

$$\Delta X_-(t) = \int_{\mu_0 - \Delta\mu}^{\mu_0 + \Delta\mu} e^{-2\pi i(\mu - \mu_0)t}\, dx_-(\mu) \tag{2.2}$$

$$\text{et } \Delta\mu \ll \mu_0 \tag{2.3}$$

Le temps de corrélation d'un tel phénomène étant de l'ordre de $1/\Delta\mu$, si $\Delta\mu \ll \mu_0$, l'image de $Z(t)$ dans le plan complexe paraîtra décrire une ellipse correspondant à une pulsation μ_0 et se déformant lentement au cours du temps. On pourra parler, à chaque instant, d'une *vibration elliptique tangente* ou d'une *ellipse instantanée* $C(t, \varepsilon)$ qui sera un être aléatoire fonction du temps t et de l'épreuve ε. Il est intéressant d'étudier:

a. *les propriétés statistiques de l'ensemble* $C(t, \varepsilon)$ *pour* t *donné* (t *fixe et* ε *décrivant l'ensemble des épreuves*).

b. *l'évolution de* $C(t, \varepsilon)$ *au cours du temps sur chaque épreuve* (t *variable et* ε *fixe*). Let résultats s'expriment simplement en utilisant la *représentation sphérique de Poincaré* des vibrations elliptiques.

A la vibration

$$Fe^{2\pi i \mu t} + Ge^{-2\pi i \mu t} \tag{2.4}$$

nous associons le point de la sphére de rayon ϱ, de longitude 2α et de latitude 2β tels que

$$\left.\begin{aligned} \varrho = ||F| + i\,|G|| \quad ; \quad 2\alpha = \operatorname{Arg} FG \quad ; \quad 2\beta = \frac{\pi}{2} - 2\lambda \\ \lambda = \operatorname{Arg}\,[|F| + i\,|G|] \end{aligned}\right\} \tag{2.5}$$

La puissance associée à cette vibration est égale à $|F|^2 + |G|^2$.

Une rotation autour de l'axe polaire de la représentation sphérique traduit l'effet d'un *pouvoir rotatoire*. Une rotation autour d'un diamètre équatorial correspond à une *biréfringence*. Une rotation autour d'un diamètre quelconque traduit l'effet simultané d'un *pouvoir rotatoire et d'une biréfringence*.

Si, dans (2.5), *F et G sont aléatoires*, nous dirons que la vibration correspondante est *polarisée ou du type elliptique*, si, *sur chaque épreuve*, la trajectoire est une ellipse, l'orientation du grand axe, l'excentricité et le sens de parcours étant le même pour toutes les épreuves. La condition pour qu'il en soit ainsi, est que l'on ait

$$E[|F|^2]\,E[|G|^2] = |E[FG]|^2 \tag{2.6}$$

a cause du caractère stationnaire, on a évidemment

$$E[FG^*] = 0$$

Nous dirons que la vibration (2.5) est *du type non polarisé* ou *du type lumière naturelle* si on a

$$E[|F|^2] = E[|G|^2] \quad \text{et} \quad E[FG] = 0 \tag{2.7}$$

D'une façon générale, une composante (2.5) *où F et G sont aléatoires, peut toujours etre considérée comme la somme de deux composantes* $[F_1, G_1]$ et $[F_2, G_2]$ *indépendantes, dont l'une est elliptique, et l'autre, du type non polarisé; les propriétés du second ordre de ces composantes sont définies par*:

$$\begin{aligned} &E[F_1G_1] = E[FG]; \quad E[|F_1|^2] = E[|F|^2] - df; \quad E[|G_1|^2] = E[|G|^2] - df \\ &E[F_2G_2] = 0; \quad E[|F_2|^2] = E[|G_2|^2] = df \end{aligned} \tag{2.8}$$

df étant la racine comprise entre 0 et la plus petite des valeurs $E[|F|^2]$ et $E[|G|^2]$ de l'équation:

$$|E(FG)|^2 = (E[|F|^2] - df)(E[|G|^2] - df) \tag{2.9}$$

Le taux d'ellipticité associe à la vibration $(E.G)$, c'est-à-dire le quotient de la puissance moyenne correspondant à la composante elliptique par la puissance moyenne totale vaut:

$$\eta = 1 - \frac{2df}{E[|F|^2] + E[|G|^2]} \tag{2.10}$$

A. *Propriétés statistiques de $C(t, \varepsilon)$ pour t donné [t fixe et ε variable].* Pour t fixé, $Z(t)$ défini par (2.1) est du type (2.4) avec F et G aléatoires. La vibration correspondante peut être décomposée en deux vibrations indépendantes dont l'une est du type non polarisée, et l'autre, du type polarisée. Grâce à une rotation autour d'un diamètre de la représentation sphérique, on pourra toujours se ramener au cas où la vibration du type polarisée est une vibration circulaire gauche [$G_1 = 0$]. Nous donnerons les résultats en supposant qu'il en est ainsi. Désignons alors par dP la puissance moyenne totale et par η le taux d'ellipticité. Rapportons l'espace aux coordonnées sphériques ϱ, θ, φ[φ longitude, θ colatitude] liées à la représentation sphérique. Soit $f[\varrho, \theta]\, dv$ la probabilité pour le point représentatif se trouve dans un élément de volume dv dont le centre a pour rayon polaire ϱ et pour colatitude θ, on peut, à partir du caractère laplacien de $Z(t)$, établir que l'on a:

$$f[\varrho, \theta] = \frac{2\varrho}{\pi[1-\eta^2](dP)^2} e^{-\frac{2\varrho^2\left[(1-\eta)\cos^2\frac{\theta}{2} + (1+\eta)\sin^2\frac{\theta}{2}\right]}{dP(1-\eta)^2}} \tag{2.11}$$

Si on ne s'intéresse qu'à la distribution des valeurs de ϱ on trouve la densité de probabilité

$$f_1[\varrho] = \frac{2\varrho}{\eta dP}\left\{e^{-\frac{2\varrho^2}{dP(1+\eta)}} - e^{-\frac{2\varrho^2}{dP(1-\eta)}}\right\} \tag{2.12}$$

Si, dans (2.12), on fait tendre $\eta \to 0$ on trouve une formule donnée par Hurwitz pour la lumière naturelle

$$f_1[\varrho] = \frac{8\varrho^3}{(dP)^2} e^{-\frac{2\varrho^2}{dP}} \tag{2.13}$$

B. *Evolution de $C(t, \varepsilon)$ au cours du temps sur chaque épreuve (t variable et ε fixe).*

Sous des conditions relativement générales, on peut établir le théorème ergodique suivant:

Théorème. Soit D un domaine quelconque de l'espace de la représentation sphérique; $P(m \in D)$ la probabilité pour que le point représentatif m soit dans D, $2\Delta T$ la mesure de l'ensemble des instants compris entre $-T$ et $+T$ et tels que $m \in D$, on a presque sûrement

$$\lim_{T\to\infty} \frac{\Delta T}{T} = P(m \in D) \tag{2.14}$$

On peut d'autre part préciser les probabilités relatives à l'évolution de $C(t, \epsilon)$.

Par exemple, dans le cas d'une vibration correspondant à de la lumière naturelle à spectre uniforme sur le domaine $\mu_0 - \Delta\mu$, $\mu_0 + \Delta\mu$, on trouve le résultat suivant:

Soient deux domaines élémentaires dV_1 et dV_2 de l'espace de représentation sphérique, correspondant à des valeurs ϱ_1 et ϱ_2 et séparés par un angle au centre $\theta = 2\chi$, la probabilité pour que, en t_1, le point figuratif soit dans dV_1 et, pour que, en $t_2 = t_1 + \tau$ il soit dans dV_2, est donnée par la relation:
$P(\varrho_1, \varrho_2, \theta) dV_1 dV_2$

$$= \frac{1}{64\pi^2[a^2(0)-a^2(\tau)]} e^{-\frac{a(0)\left[\varrho_1^2+\varrho_2^2\right]}{2[a^2(0)-a^2(\tau)]}} \varrho_1 \varrho_2 I_0 \left\{ \frac{\varrho_1 \varrho_2 a(\tau) \cos \chi}{a^2(0)-a^2(\tau)} \right\} dV_1 dV_2 \qquad (2.15)$$

où I_0 est la fonction de Bessel d'ordre zero
avec

$$a(\tau) = \frac{dP}{4} \frac{\sin 2\pi\tau\Delta\mu}{2\pi\tau\Delta\mu}$$

La liaison entre le fait que l'on a $M(t_1) \in dV_1$ et le fait que l'on a $M(t_2) \in dV_2$ disparait pour tous les zéros de a (τ), c'est-à-dire pour

$$|2\pi\tau\Delta\mu| = K\pi \quad (K \neq 0)$$

BIBLIOGRAPHIE.

[1] On trouvera des renseignements sur cette question et des indications bibliographiques dans: A. BLANC-LAPIERRE et R. FORTET. Théorie des Fonctions aléatoires - Masson - Editeur - Paris - 1953 p. 538.

[2] VOIR A. BLANC-LAPIERRE et P. DUMONTET. Sur la notion de cohérence en optique. A paraître dans la Revue d'Optique (1955).

[3] H. H. HOPKINS. Proceedings of the Royal Society A. Volume 208 (1951) p. 283 et Volume 217 (1953) p. 408.

[4] H. HURWITZ. Journal of the Optical Society of America 35, 1945 p. 525.

[5] A. BLANC-LAPIERRE et R. FORTET. Théorie des fonctions aléatoires p. 471—521.

[6] A. BLANC-LAPIERRE. Comptes rendus de l'Académie des Sciences 232, 1951. p. 934 et 1070. On trouvera des Calculs détaillés dans A. BLANC-LAPIERRE et R. FORTET (Théorie des fonctions aléatoires) p. 478—503.

Note ajoutie à la correction des épreuves.

Depuis que cet article a été écrit, nous avous eu connaissance des tres intéressants articles de E. WOLF sur les problèmes de cohérence. Le lecteur pourra s' y reporter avec beaucoup de profit (E. WOLF, Proc. Royal Soc. [A] t. 225 1954 p. 96 et t. 230 p. 246 1955).

PRESENT STATE AND FUTURE PROSPECTS OF STOCHASTIC PROCESS THEORY

J. L. Doob

The theory of stochastic processes has developed sufficiently in the past two decades so that one can now properly give a survey of its present state and hazard some guesses as to its future development. It is clear that no mathematician can reliably predict what the mathematics of the next twenty years will be, even in his own field. In fact, if he could know what it would be, he would negate his own foreknowledge by developing the new theories himself, long before the twenty years were up! Thus the guesses to be given below are simply the guesses of someone who is now interested in stochastic processes, and is considering the areas in which he would like to obtain new results.

Probability theory advances in two ways. On the one hand, probability problems lead to problems in other fields, in differential and integral equations, for example, which can be formulated and solved with little or no knowledge of the probability background. On the other hand, there are the peculiarly probabilistic problems, say on the convergence of mutually independent random variables, or on the continuity of sample functions of stochastic processes, which must be attacked by the methods peculiar to probability. The measure theoretic formalization of probability concepts twenty-one years ago by Kolmogorov made possible further progress in the latter direction, progress which had been woefully retarded up to that time. Using this formalization, it is now generally accepted that the most useful mathematical definition of a stochastic process is simply that a stochastic process is a family of random variables. Although this definition is so general as to seem pointless, anything less general appears to be insufficient. In most studies, the random variables of a stochastic process have been numerical-valued, and have been indexed by a real-valued parameter. This case will be called the *standard case* below. Our survey will first cover the standard case, going into its general definition, and describing the types of standard stochastic processes that have been studied intensively. An indication of the present state of the theory, and the possibilities for future development, will also be given. Finally, a few remarks will be made on new possibilities in the standard case, and on the non-standard case.

Before proceeding to the body of the discussion, we remark that, in view of the wide definition of a stochastic process, anyone can define a type of stochastic process. In fact, anyone with a rudimentary knowledge of probability can hope in five minutes or so to define a type of stochastic process that has not appeared in the literature. The real problem is to devise some relationship between random variables leading to a property of a family of random variables that will prove interesting and important, and the only properties discovered up to this time are the property of mutual independence, the Markov property, the stationarity property, the martingale property. We have mentioned here only properties specific enough to lead to intersting results, excluding, for example, the property that the random variables of the process have finite second moments. This meager list of four properties indicates that probabilists are sadly in need of some new ideas!

1. *Basic concepts involved in discussions of standard processes.*

Let $\{x(t),\ t \in T\}$ be a standard stochastic process, that is, a family of measurable functions, defined on a (probability) measure space, with linear parameter set T. The random variable $x(t)$ is a function of the point w of the specified measure space, with value $x(t, w)$ at the point w. Fixing w, $x(t, w)$ defines a function of t, and the functions of t obtained in this way are called the *sample functions* of the process. The given measure of w sets determines a measure of sets of sample functions. By definition of random variable, the joint distributions of the finite sets of the random variables of the process are known. These distributions are called the *finite dimensional distributions* of the process. In the past, stochastic processes have been characterized by the properties of these finite-dimensional distributions. For example, if T is the whole line, the process is called *stationary* if, for every finite parameter set $t_1, \ldots, t_n$, the joint distribution of the random variables $x(t_1 + h), \ldots, x(t_n + h)$ does not depend on the number h. It remains to be seen whether other methods of classification will ever prove fruitful. According to a classical measure theorem of Kolmogorov, corresponding to any specification of mutually consistent finite dimensional distributions, there is a stochastic process with the specified distributions. Moreover, associated with this existence theorem, and with the method of classification by finite-dimensional distributions is the fact that the finite dimensional distributions have been the *only* specifications of the process. Thus, any properties of the process have necessarily been obtainable from these joint distributions, either directly using the properties of measure functions, or by the intermediary of some additional principle which need only be consistent with the properties of measure functions. Such additional principles were introduced by the speaker in 1937 and later, in order to make

possible the discussion of such properties as boundedness, continuity, and measurability, of the sample functions of a stochastic process with a non-denumerable parameter set.

Thus, with the help of certain additions to the basic measure-theoretic formalization of probability, there is now an adequate formalization of standard stochastic processes, and it does not appear that much more work will be done here except in the direction of alternative approaches. For example, a more algebraic approach is now gaining in popularity, in which events are interpreted not as sets of a measure space but as elements of a Boolean algebra. The sample functions must be defined indirectly, in such an approach, and this is a disadvantage. However such an approach has the advantage that it automatically avoids some of the difficulties involved in studying sample functions, and this is an advantage if one is not interested in the sample functions for their own sake.

2. *Standard stochastic processes with mutually independent random variables.*

Stochastic process theory has been limited for much of its life, in fact during the centuries before this one when the term *stochastix process* was invented, to the study of finite or denumerably infinite sequences of mutually independent random variables. The study of the distributions of sums of such random variables appears inexhaustible, not only from the point of view of general theory but of refinement of known results. One prediction it is safe to make is that this study will continue, and will remain fruitful.

The continuous parameter analogue of sums of mutually independent random variables is a process $\{x(t),\ t \in T\}$ of independent increments, with parameter set T an interval. That is, if $t_1 < \ldots < t_n$ are parameter points, it is supposed that the random variables

$$x(t_2) - x(t_1), \ldots, x(t_n) - x(t_{n-1})$$

are mutually independent. The general study of these processes, initiated by de Finetti in 1929, has now finished its first task. That is, the general character of the sample functions is known, the (infinitely divisible) class of distributions involved has been fully described, and many special properties have been discovered. The Brownian motion process, in which the above differences have normal distributions with zero means and variances a constant multiple of the parameter differences, has been the most studied special case, and the fact that new and important properties of this process are still being discovered suggests that there will be much further general work on these processes, besides the obvious refinement in detail. In fact, as is remarked below, there is even

now no satisfactory treatment of diffusion processes, which depend essentially on Brownian motion processes.

3. Standard Markov processes.

The study of Markov processes was initiated by Markov (in a special case) at the beginning of this century. The stochastic process $\{x(t),\ t \in T\}$ is a Markov process if, when $t_1 < \ldots < t_n$ are parameter values, and A is a linear Borel set, the equality

$$P\{x(t_n, w) \in A \mid x(t_1), \ldots, x(t_{n-1})\} = P\{x(t_n, w) \in A \mid x(t_{n-1})\}$$

is true with probability 1. That is, somewhat roughly, the $x(t)$ process is a Markov process if the conditional probability of a future event, given present and past states, only depends on the present state.

Standard Markov processes have been studied most intensively in two cases, the stationary chain case, in which $x(t)$ only has integer values, so that the probability of a transition from state i to state j in time s is given by a number which we can write in the form $p_{ij}(s)$, and the diffusion case, which is defined below. When the parameter set T is the set of positive integers, the basic properties of Markov chains, essentially the asymptotic properties of $p_{ij}(t)$ for large t, have been known for years, but, when T is the interval $(0, \infty)$, new problems arise, for example the continuity properties of the sample functions and regularity properties of the transition probability functions must now be investigated. These problems have been solved in the non-pathological cases, but not even all the obvious fundamental problems in this area have been solved in the general case, in spite of recent important work by Lévy. There remains the systematic study of the non-stationary case, which has been barely begun.

The standard Markov processes of diffusion type can be roughly described as follows. The parameter set is an interval, and there are functions m, σ, called *diffusion coefficients*, for which

$$x(t + h) - x(t) = m[t, x(t)]h + \sigma(t, x(t))\xi_{t,h} + \ldots.$$

Here the remainder is negligible in comparison with the terms displayed, as $h \to 0$, and $\xi_{t,h}$ is a random variable which is normally distributed, independent of the class of random variables $\{x(s),\ s \leqq t\}$, and has expectation 0, dispersion h. In the simplest non-trivial special case, m vanished identically, σ is identically a positive constant, and the corresponding process is the Brownian motion process. This process was first studied (unrigorously) by Bachelier at the turn of the century, before either Markov processes or processes with independent increments, to both of which classes this process belongs, had received general definitions.

The class of diffusion processes is not rigorously defined; that is one of the difficulties of this subject. There are almost as many different approaches to these processes as there are investigators, and in fact it is not yet known under what conditions the various approaches lead to the same processes. There appears to be general agreement that the definitions should be phrased in such a way that almost all sample functions of a diffusion process are continuous (under the appropriate conventions necessary for such a statement). Under various regularity conditions, the transition probability functions of a diffusion process satisfy parabolic differential equations involving the diffusion coefficients, first discussed systematically by Kolmogorov in 1931, but the exact role of these equations is not yet clear. Conversely, given a pair of diffusion coefficients, there are various ways of defining the corresponding diffusion process: by solving the corresponding parabolic differential equations to find the transition probabilities (Feller); by semi-group methods, at least in the stationary case (Yosida); by rewriting the above difference equation in the form of an integral equation, replacing $\xi_{t,h}$ by the differential element of a Brownian motion process, and solving for $x(t)$ in terms of this Brownian motion process (Ito). As already remarked, it is not clear when these methods, for a specified pair of diffusion coefficients, yield the same stochastic process, or even when the transition probability functions of the processes obtained in any way other than the first satisfy the parabolic partial differential equations characteristic of diffusion processes. In addition, there are various operations commonly performed on diffusion processes leading to new ones, such as stopping the trajectories when they reach a specified closed set, and it is not yet clear when the transition probability functions of the new processes satisfy these same parabolic equations, with the same diffusion coefficients but different initial conditions. In short, there is obviously much work to be done in this area, and much is being done. If $x(t)$ is an n-dimensional vector-valued random variable, with $n > 1$, it is to be expected that the corresponding problems may become considerably more difficult. The corresponding diffusion processes are non-standard in that case, of course.

4. Standard stationary processes.

These stochastic processes have already been defined. Since the definition is so general, it is not to be expected that many specific results can be obtained for these processes, and in fact the general theorems applicable to these processes are centered around one theorem, the ergodic theorem (law of large numbers) in its various forms. Further developments must be based on more specialized hypothesis.

The processes called *stationary in the wide sense*, and characterized by the

condition that $E\{x(t)\bar{x}(s)\}$ depend only on $t - s$ (in the real case, the conjugate sign is of course unnecessary) have received considerable attention in the past and under various special hypotheses, can be expected to receive considerable further study, but, just as for stationary processes, further specialization is undoubtedly necessary to obtain new results. The study of these processes has been and will continue to be intensified by the interest they have in statistics.

5. *Standard martingales.*

Martingales and semimartingales are relative newcomers. The first general martingale theorems were proved by Lévy about twenty years ago, but these processes were first studied systematically by the speaker, in 1940. A stochastic process $\{x(t),\ t \in T\}$ is a martingale if, when $t_1 < \ldots < t_n$ are parameter values, the equality

$$E\{x(t_n) \mid x(t_1), \ldots, x(t_{n-1})\} = x(t_{n-1})$$

is true with probability 1. That is, roughly, the conditional expectation of future values, given present and past values, is the present value. For a semimartingale, the equality is replaced by inequality ($\geqq$). It is likely that the most important general theorems on these processes are now known, that is, the convergence properties of martingale and semimartingale sequences, the continuity properties of sample functions of continuous parameter processes of these types, and the invariance properties of these types under various frequently used transformations. However there remain many fruitful possibilities for research in the applications of martingale theory. For example, the martingale equation above states that a certain value is an integral average. This suggests the possibility of applying martingale theory to the study of the first boundary value (Dirichlet) problem for elliptic, or more generally, for parabolic partial differential equations. The most general such application would be the following. Suppose that u is a function, defined on some topological space, with the property that the value of u at a point z of an open set D is a weighted average of the values of u on the boundary of D. Then it is natural to try fo find a system of continuous probability trajectories (sample trajectories of a stochastic process whose random variables have values in the topological space on which u is defined) such that $u(z)$ is the average value of u over the boundary of D, average here meaning with the weighting defined by the distribution of the first point in which the probability trajectories from z meet the boundary of D. This program has been carried out in detail by the speaker for u the solution of a second order linear differential equation in one dimension, for u harmonic in n dimensions, and more generally for u the solution of the heat equation in $n + 1$ dimensions. In each case, if $z(t)$ is the position

of the trajectory issuing from the point z, at time t, where the trajectory is stopped when it reaches the boundary of D, the $u[z(t)]$ stochastic process is a martingale. The properties of martingales can then be used to show that functions in the class of u, under mild restrictions, have boundary limits along probability trajectories, these limits being the assigned values if u is a generalized first boundary value problem solution. The general abstract problem has not yet been studied, however, although a positive solution would have the greatest interest. This problem illustrates the importance, for a class of functions u defined on a topological space, of finding the stochastic processes $\{z(t),\ t \in T\}$, whose random vairables have values in the space, for wich the stochastic process $\{u[z(t)],\ t \in T\}$ is a martingale. The properties of martingales can then be used to derive new properties of the functions in the given class. Somewhat more generally, the familiar subfunctions of differential equation theory, such as the subharmonic functions, lead to semimartingales.

6. New standard stochastic processes.

In the previous sections, we have discussed the most important types of standard stochastic processes. An obvious question is: Can one expect that important new types will be discovered? No scientist with any historical sense would give a negative answer to such a question, but progress has certainly been slow in the development of new types. What has been sought is a simple propety with interesting implications, and we have already noted that only the independence, Markov, stationarity, and martingale properties (or closely related ones) have satisfied this criterion.

It is possible that the successful study of non-linear relations in other parts of mathematics will suggest new types of stochastic processes. One example of such a possibility is the following. Let L be a differential operator, and consider the differential equation $L(f) = A$. If f is the x coordinate of a particle subject to a stochastic driving force, this equation can be interpreted as the equation of motion of a particle, when, if the particle is a molecule, the equation is called Langevin's equation. This differential equation has been studied in detail if L is a linear differential operator, with constant coefficients, and if A is the sample function of a stationary process, $f(t)$ can be found explicitly in terms of A, and the $f(t)$ process, with appropriate initial conditions, is stationary. There is no corresponding theory if L is a non-linear differential operator, although such a theory might be very illuminating in many applications, for example in the analysis of the rocking of a ship exposed to wave action. Here again one would expect, under reasonable conditions, to find a stationary solution, obtaining in this way a new class of stationary processes.

7. Non-standard processes.

An obvious generalization of the standard processes is to suppose that the random variables of the process have values in an n-dimensional Euclidean space, for $n > 1$. Some work has been done in generalizing the results for $n = 1$ to this case, but much remains to be done even at this elementary level of generalization. Recently there has been work on processes whose random variables take on values in a Riemannian manifold, and this is obviously just the beinning of such work. There has also been some work on random variables with values in a Banach space, but no significantly new types of stochastic process with such random variables have been discovered as yet.

Finally, one can generalize the standard processes by allowing other than linear parameter sets. An example is the Poisson process in which events are distributed in n-space instead of in time, the expected number of events in an open set of the space being proportional to the volume of the open set. Here the most natural parameter set is the class of all Lebesgue measurable sets in n dimensions, and the random variable corresponding to such a Lebesgue measurable set is the number of events in the set. Only the barest beginnings have been made in discussing types of processes with non-linear parameter sets, although it would seem at first that here is the easiest context in which new and interesting types of stochastic processes could be defined.

LES FONCTIONS ALÉATOIRES COMME ÉLÉMENTS ALÉATOIRES DANS UN ESPACE DE BANACH

ROBERT M. FORTET

Il est fréquemment possible de considérer une fonction aléatoire [numérique, réelle ou complexe, d'une variable numérique, pour fixer les idées: réelle sur un intervalle (t_1, t_2)] comme un élément aléatoire dans un espace de Banach; cela est très utile si on doit étudier par exemple une fonctionnelle de la fonction aléatoire faisant intervenir celle-ci globalement sur tout l'intervalle (t_1, t_2). Divers faits suggèrent alors d'établir, en un sens convenable, un théorème central limite pour les éléments aléatoires dans un espace de Banach.

Exemple 1. On sait l'importance pratique des fonctions aléatoires dérivées d'un processus de Poisson $N(t)$, par exemple stationnaire de densité ϱ; si $X_\varrho(t)$ est une telle fonction, convenablement normée, on sait que lorsque $\varrho \to +\infty$, la loi temporelle de $X_\varrho(t)$ tend vers la loi temporelle laplacienne d'une fonction aléatoire laplacienne $Y(t)$; il est alors intuitif, mais jusqu'à présent on n'en avait pas de démonstration, que la fonction de répartition d'une fonctionnelle telle par exemple que:

$$\int_{t_1}^{t_2} |X_\varrho(t)|dt$$

tend vers la fonction de répartition de la même fonctionnelle prise sur $Y(t)$, soit:

$$\int_{t_1}^{t_2} |Y(t)|dt.$$

Or $X_\varrho(t)$ apparait comme la somme (normée) d'un grand nombre (d'ailleurs aléatoire) de fonctions aléatoires indépendantes, c'est à dire d'un grand nombre d'éléments aléatoires indépendants dans un espace de Banach: on rencontre bien le problème du théorème central limite, c'est à dire de la convergence (en loi) vers un élément laplacien.

Exemple 2. $Y(t)$ désignant la fonction aléatoire de Wiener-Lévy classique, on a eu besoin, pour une méthode de Monte-Carlo, de „réalisations empiriques" de la variable aléatoire

$$\int_{t_1}^{t_2} V[Y(t)]dt,$$

où $V(x)$ est une fonction certaine; la méthode est d'obtenir une réalisation de $Y(t)$, sur laquelle on prendra la fonctionnelle

$$\int_{t_1}^{t_2} V[Y(t)]dt;$$

malheureusement $Y(t)$, processus permanent, n'est en toute rigueur réalisable qu'avec une infinité de tirage au sort; on substitue alors à $Y(t)$ une fonction $X_n(t)$, exigeant seulement n tirages au sort, dont la loi temporelle tend vers celle, laplacienne, de $Y(t)$; il est facile de trouver une telle $X_n(t)$ et naturellement ce sera une somme (normée) de n fonctions aléatoires simples, indépendantes, autrement dit de n éléments aléatoires indépendants dans un espace de Banach.

Mais le procédé n'est justifié que si la fonction de répartition de

$$\int_{t_1}^{t_2} V[X_n(t)]dt$$

tend vers celle de

$$\int_{t_1}^{t_2} V[Y(t)]dt.$$

On possédait déjà des résultats à cet égard, les plus étendues étant ceux de D. Donsker [1]; mais la connexion intuitive avec un théorème central limite généralisé n'avait pas été suffisamment mise en lumière.

Exemple 3. Soit $X_1, X_2, \ldots, X_j, \ldots$ une suite indéfinie de variables aléatoires indépendantes, de fonction de répartition continue $F(x)$; supposons, sans diminuer la généralité, les X_j équiréparties sur l'intervalle (0,1), de sorte que $F(x) = x(0 \leqq x \leqq 1)$; soit $X_j(x)$ la fonction aléatoire définie par

$$X_j(x) = \begin{cases} 0 \text{ si } x < X_j \\ 1 \text{ si } x \geqq X_j \end{cases}$$

et posons:

$$F_n(x) = \frac{1}{n} \sum_{j=1}^{n} X_j(x),$$

de sorte que $F_n(x)$ est une somme de n fonctions aléatoires, si on veut de n éléments aléatoires dans un espace de Banach (on peut prendre par exemple l'espace de Hilbert L_2).

Or $F_n(x)$ est la fonction de répartition empirique que fournira la réalisation de $X_1, X_2, \ldots, X_n$; le problème de sa convergence vers $F(x)$ pour $n \to +\infty$ est resolu par le classique théorème de Glivenko-Cantelli d'après lequel $F_n(x)$

tend uniformement en x vers $F(x)$; en d'autres termes: pour une fonction $f(x)$ de x sur (0,1), désignons par $\mu(f)$ la fonctionnelle:

$$\mu(f) = p.p.b.s.\ f(x), \quad 0 \leqq x \leqq 1$$

cette fonctionnelle n'ayant évidemment un sens que pour une certaine classe de fonctions f; le théorème de Glivenko-Cantelli est que:

$$\text{p.s. } \lim_{n \to +\infty} \mu[F_n - F] = 0$$

J'ai montré ailleurs que cela est un cas particulier d'un théorème beaucoup plus général (cf. R. Fortet et E. Mourier [1]).

Pour préciser, on peut remarquer que $\sqrt{n}\,[F_n(x) - F(x)]$ doit naturellement tendre en loi vers un élément laplacien limite $Y(x)$ (facile à trouver) et que vraisemblablement la fonction de répartition de $\mu[\sqrt{n}(F_n - F)]$ doit tendre vers celle de $\mu[Y]$; c'est en effet ce qu'établissent les classiques théorèmes de Kolmogoroff et Smirnov.

Méthode et résultats. Soit $\mathfrak{X}$ un espace de Banach d'éléments x, $\mathfrak{X}^*$ son dual d'éléments x^*; désignons par θ l'élément nul de $\mathfrak{X}$, par $<x^*, x>$ le nombre obtenu en appliquant à $x \in \mathfrak{X}$ la fonctionnelle linéaire $x^* \in \mathfrak{X}^*$. Soient $X_1, X_2, \ldots X_j, \ldots$ une suite indéfinie d'éléments aléatoires dans $\mathfrak{X}$, mutuellement indépendants, de même loi de probabilité; on supposera en outre: $E(X_j)=\theta$, $E(||X_j||^2) < +\infty$. Pour prouver que

$$Z_n = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j$$

tend en loi, lorsque $n \to +\infty$, vers un élément aléatoire laplacien Y dans $\mathfrak{X}$, il est naturel d'utiliser la méthode de la caractéristique qui réussit pour les variables aléatoires; on voit sans difficulté que pour $n \to +\infty$ la caractéristique $\varphi_n(x^*)$ de $\mathfrak{Z}_n$ tend vers une limite $\varphi(x^*)$; mais il n'est pas prouvé en général que $\varphi(x^*)$ est une caractéristique.

Disons que $\mathfrak{X}$ est du type $\mathfrak{G}$ s'il existe un nombre positif A et une application G de $\mathfrak{X}$ dans $\mathfrak{X}^*$: $x^* = G(x)$, tels que:

α) $||G(x)|| = ||x||$; $< G(x), x > = ||x||^2$

β) $||G(x) - G(y)|| \leqq A\,||x - y||$ quels que soient $x, y \in \mathfrak{X}$; les espaces de Hilbert, et toute une catégorie d'espaces uniformément convexes, en particulier les espaces L_α pour $\alpha \geqq 2$, sont du type G.

On démontre que si $\mathfrak{X}$ est séparable, réflexif et du type $\mathfrak{G}$, $\varphi(x^*)$ est une caractéristique, nécessairement d'un élément aléatoire laplacien Y tel que $E(||Y||^2) < +\infty$, $E(Y) = \theta$.

Il n'en résulte pas évidemment que, en loi, Z_n tend vers Y; mais soit $f(x)$

une fonctionnelle (numérique) de $x \in \mathfrak{X}$, uniformément continue sur toute boule de rayon fini de $\mathfrak{X}$, quelconque par ailleurs. Si *en outre* $\mathfrak{X}$ possède une base, on démontre que la fonction de répartition de $f(Z_n)$ tend vers celle de $f(Y)$.

On peut élargir la définition du type $\mathfrak{G}$ (de façon en particulier à englober les espaces L_α pour $1 < \alpha < 2$); et d'autre part traiter le cas où les X_j n'ont pas forcément la même loi de probabilité.

Ces résultats permettent de traiter les Exemples 1 et 2 ci-dessus; mais non, jusqu'à nouvel ordre, l'Exemple 3. Ils ont été obtenus en collaboration par Mlle E. Mourier et moi-même; un mémoire exposant les démonstrations détaillées est en cours d'impression dans Studia Mathematica; un résumé analogue à celui-ci paraitra dans les Actes du Colloque sur le Calcul des Probabilités et la Statistique Mathématique de l'Université Humboldt (Berlin, 19—21 octobre 1954).

BIBLIOGRAPHIE

[1] D. DONSKER, An invariance principle for certain probability limit theorem, Mem. of the Am. Math. Soc., no 6, 1951.

[1] R. FORTET et E. MOURIER, La convergence de la répartition empirique vers la répartition théorique Ann. Sc. de l'Ecole Normale Sup., 70, 1953, p. 267.

LOIS DES GRANDS NOMBRES POUR DES ÉLÉMENTS ALÉATOIRES GÉNÉRAUX

ROBERT M. FORTET

1. *Préambule.* Dans le Calcul des Probabilités classique, nous entendons par là celui qui concerne les variables aléatoires numériques, on appelle *loi des grands nombres* (du moins c'est le sens précis que nous donnerons ici à cette expression) un énoncé affirmant, bien entendu sous certaines conditions plus ou moins larges, qu'étant donnée une suite indéfinie $X_1, X_2, \ldots, X_j, \ldots$, de variables aléatoires, et si $n \to +\infty$, la variable aléatoire $Y_n = \frac{1}{n}(X_1 + X_2 + \ldots + X_n)$ converge stochastiquement vers une limite L (en général aléatoire); le cas typique est celui où les X_n sont en chaine strictement stationnaire, et plus particulièrement celui où elles sont mutuellement indépendantes. Suivant la nature de la convergence stochastique impliquée dans l'énoncé, on peut distinguer les lois des grands nombres en probabilité, presque-sûres, en moyenne d'ordre α, etc ...

Les lois des grands nombres sont sans doute un des éléments sur lesquels repose le raisonnement inductif; il est donc philosophiquement important d'étendre les lois des grand nombres à des éléments aléatoires $X_1, X_2, \ldots, X_j, \ldots$ de nature aussi générale que possible, c'est à dire prenant leurs valeurs non plus dans l'ensemble des nombres réels, mais dans un espace $\mathfrak{X}$ de nature aussi générale que possible.

De toutes façons une telle extension est une idée qui vient naturellement à l'ésprit, et qu'impose aujourd'hui la théorie des fonctions aléatoires (éléments aléatoires à valeurs dans un espace fonctionnel), et certaines applications statistiques (définition de l',,homme moyen'' de Quetelet et M. Fréchet, etc. ...).

Toutefois il ne peut être question d'étendre les lois des grands nombres, au sens un peu étroit que nous avons adopté plus haut, pour un espace $\mathfrak{X}$ tout à fait quelconque: la formation de Y_n exige que $\mathfrak{X}$ soit un espace vectoriel, et les convergences stochastiques qui constituent les lois des grands nombres exigent que $\mathfrak{X}$ soit muni d'une topologie. En laissant de coté le cas assez évident des espaces vectoriels à un nombre fini de dimensions, les espaces vectoriels topologiques où la topologie est la plus facile à manier tout les espaces de Banach; aussi les recherches dans cette direction ont commencé, et jusqu'à présent se sont presque limité, au cas où $\mathfrak{X}$ est un espace de Banach; et naturellement les premiers

résultats ont concerné les espaces de Hilbert [1]): l'existence d'un produit scalaire permet alors de reprendre avec succès des modes de raisonnement classiques dans le cas des variables aléatoires (cf. N. Glivenko [1], une allusion dans A. Kolmogoroff [1], et M. Fréchet [1]).

2. *Eléments aléatoires dans un espace de Banach.* Supposons donc que $\mathfrak{X}$ est un espace de Banach, réel pour fixer les idées, d'élément nul θ; deux topologies, la topologie dite *forte* et celle dite *faible*, peuvent être considérées dans $\mathfrak{X}$; nous désignerons:

par x^* une fonctionnelle linéaire quelconque sur $\mathfrak{X}$, continue avec la topologie forte dans $\mathfrak{X}$;

pat $\mathfrak{X}^*$ le dual (fort) de $\mathfrak{X}$, c'est à dire l'ensemble des x^*;

par $\langle x^*, x\rangle$ le nombre obtenu en appliquant à l'élément x de $\mathfrak{X}$ la fonctionnelle linéaire x^*;

par $\mathfrak{U}$ l'ensemble des épreuves;

par u une épreuve ou élément quelconque de F.

On suppose défini sur $\mathfrak{U}$, dans les conditions habituelles, un corps de Borel $\mathfrak{B}$ et, sur ce corps de Borel, une mesure de probabilité μ.

Un élément aléatoire X à valeur dans $\mathfrak{X}$ n'est autre qu'une fonction $x(u)$ de $u \in \mathfrak{U}$, dont la valeur $x(u) \in \mathfrak{X}$. Nous ne considèrerons d'ailleurs que des éléments X tels que, quelle que soit $x^* \in \mathfrak{X}^*$, $\langle {}^*x, X\rangle$ [c'est à dire $\langle x^*, x(u)\rangle$] soit une fonction de u mesurable-μ; rappelons qu'on appelle alors *espérance mathématique* de X l'élément $E(X)$ de $\mathfrak{X}$, unique s'il existe, tel que: $\langle x^*, E(X)\rangle = E(\langle x^*, X\rangle)$ pour tout $x^* \in \mathfrak{X}^*$; c'est à dire que $E(X)$ est l'intégrale de Pettis

$$\int_{\mathfrak{U}} x(u)d\mu.$$

Si $\| X \|$ est mesurable-μ (ce qui a nécessairement lieu si $\mathfrak{X}$ est séparable), et si $E(\| X \|) < +\infty$, $E(X)$ existe; ce fait, connue par ailleurs, peut aussi se déduire des lois des grands nombres que nous allons indiquer.

On peut assez facilement obtenir des lois des grands nombres relatives à la topologie faible dans $\mathfrak{X}$; mais comme dans $\mathfrak{X}$ la convergence forte implique la convergence faible, à notre connaissance les résultats établis dans cet ordre d'idée sont conséquences immédiates des lois des grands nombres relatives à la topologie forte, les seules dont nous allons parler.

I. *Loi des grands nombres presque-sûre.* Si $\mathfrak{X}$ est séparable, si X_0, $X_{\pm 1}$, $X_{\pm 2}$,

---

[1]) Nous appelons espace de Hilbert un espace de Banach, separable ou non, muni d'un produit scalaire ou hermitique avec lequel il est normé.

$\ldots, X_{\pm j}, \ldots$ est une suite strictement stationnaire [2]) d'éléments aléatoires à valeurs dans $\mathfrak{X}$, et si $E(|| X_j ||) < + \infty$, il y a une probabilité 1 pour que Y_n tende fortement vers un élément aléatoire limite L, tel d'ailleurs que $E(|| L ||) < + \infty$, $E(L) = E(X_j)$; si en outre les X_j sont mutuellement indépendants, L se réduit presque-certainement à l'élément certain $E(X_j)$

Ce théorème est une extension du théorème ergodique de Birkhoff au cas de fonctions à valeurs dans $\mathfrak{X}$

II. *Loi des grands nombres en moyenne d'ordre* α. Si $\mathfrak{X}$ est séparable, si $X_0, X_{\pm 1}, X_{\pm 2}, \ldots, X_{\pm j}, \ldots$ est une suite strictement stationnaire d'éléments aléatoires à valeurs dans $\mathfrak{X}$, et si $E(|| X_j ||^\alpha) < + \infty$ avec $1 \leqq \alpha < + \infty$, il existe un élément aléatoire L à valeurs dans $\mathfrak{X}$, tel que $E(|| L ||^\alpha) < + \infty$, $E(L) = E(X_j)$ et que:

$$\lim_{n \to + \infty} E(|| Y_n - L||^\alpha) = 0;$$

si en outre les X_j sont mutuellement indépendants, $L = E(X_j)$ presque-surement.

Pour $\alpha = 1$, cet énoncé II peut d'établir directement; pour $1 < \alpha < + \infty$, on peut utiliser un théorème ergodique de Yosida et Kakutani, après avoir déterminé la forme des fonctionnelles linéaires continues sur les espaces $\mathfrak{X}_{(\alpha)}$: nous désignons par $\mathfrak{X}_{(\alpha)}$ l'espace de Banach des éléments aléatoires X à valeurs dans $\mathfrak{X}$ pour lesquels $E(|| X ||^\alpha) < + \infty$; ce problème est d'ailleurs interessant par lui-même.

3. *Précisions à l'énoncé II.* Placons-nous dans les hypothèses de l'énoncé II, en supposant en outre que les X_j sont mutuellement indépendants, et (ce qui ne restreint pas la généralités) que $E(X_j) = L = \theta$; alors, et en écartant le cas evident où $X_j = \theta$ presque-sûrement, on montre que:

Si $\mathfrak{X}^*$ est séparable et si $\alpha \geqq 2$, il existe une constante positive ϱ telle que:

$$E(|| Y_n ||^\alpha) \geqq \varrho n^{-\frac{\alpha}{2}} \text{ pour tout } n.$$

Mais pour obtenir pour $E(|| Y_n ||^\alpha)$ une limitation dans l'autre sens, il semble nécessaire de faire sur $\mathfrak{X}$ des hypothéses d'une autre sorte. Disons d'un espace de Banach $\mathfrak{X}$ (séparable ou non) que c'est un espace $\mathfrak{G}$ s'il existe un nombre positif A et une application $G[x^* = g(x)]$ de $\mathfrak{X}$ dans $\mathfrak{X}^*$, dite canonique, tels que:

a) $|| g(x) || = || x ||$; $\langle g(x), x \rangle = || x ||^2$ pour tout $x \in \mathfrak{X}$;

b) $|| g(x) - g(y) || \leqq A || x - y ||$ quels que soient $x, y \in \mathfrak{X}$.

---

[2]) Nous passons sur la definition des suites d'éléments aléatoires dans X strictement stationnaires, ou des suites d'éléments mutuellement indépendants; bien qu'il y ait lieu de regarder ces définitions de pres, elles ne soulèvent pas de véritables difficultés.

Nous ne connaissons pas de caractérisation générale des espaces $\mathfrak{G}$; mais il est prouvé que des espaces uniformément convexes à norme uniformément régulière formant une classe assez large sont des espaces $\mathfrak{G}$; parmi eux se trouvent les espaces de Hilbert, et les espaces L_α pour $\alpha \geqq 2$, ce qui ouvre déjà un vaste champ d'applications; pour L_α, le meilleur choix de A est:

$$A = (\alpha - 1)2^{\alpha-2} \qquad (\alpha \geqq 2);$$

il semble bien d'après cela que la propriété, pour un espace de Banach, d'être un espace $\mathfrak{G}$ est une propriété de convexité.

On peut prouver que:

III. Si $\mathfrak{X}$ est un espace $\mathfrak{G}$, si les X_j sont mutuellement indépendants, si pour tout j $E(\| X_j \|^2) < +\infty$ et $E(X_j) = \theta$, on a:

$$E(\| Y_n \|^2) \leqq \frac{A}{n^2} \sum_{j=1}^{n} E(\| X_j \|^2) \text{ pour tout } n, \tag{3,1}$$

où le A est celui intervenant dans b) ci-dessus; de sorte que, si:

$$\sum_{j=1}^{n} E(\| X_j \|^2) = O(n^s), \text{ avec } s < 2, \text{ pour } n \to +\infty,$$

α) presque-sûrement Y_n tend (fortement) vers θ lorsque $n \to +\infty$;

β) Y_n tend vers θ en moyenne d'ordre 2 lorsque $n \to +\infty$, dans les conditions précisées par (3,1).

On peut remarquer, par comparaison avec I et II, que dans III $\mathfrak{X}$ n'est pas nécessairement séparable, ni les X_j de même loi de probabilité.

Les résultats exposés dans ces §§ 2 et 3 sont établis dans E. Mourier [1], R. Fortet et E. Mourier [1], [2]; il conviendrait de les préciser et de les étendre, mais tels quels nous avons déjà pu dans R. Fortet et E. Mourier [2], [3] en déduiré d'interessantes applications, et nous comptons en développer encore beaucoup d'autres.

4. *Un résultat de S. Doss.* Nous connaissons un seul exemple de loi des grands nombres concernant un élément aléatoire à valeurs dans un espace $\mathfrak{X}$ qui n'est pas forcément de Banach, ni même vectoriel, à savoir le résultat suivant de S. Doss (cf. S. Doss [1]). Comme $\mathfrak{X}$ peut ne pas être vectoriel, il faut substituer à Y_n un élément aléatoire Z_n dont la définition n'exige pas l'existence d'une addition dans $\mathfrak{X}$ — et par suite on déborde un peu le cadre des lois des grands nombres au sens strict adopté plus haut.

Supposons $\mathfrak{X}$ distancié; désignons par (x, y) la distance de 2 éléments x, y de $\mathfrak{X}$; ne donsidérons que des éléments aléatoires $X[x(u)]$ tels que, pour tout $y \in \mathfrak{X}$ et pour tout nombre positif ϱ, l'ensemble des u pour lesquels $(x(u), y) \leqq \varrho$ appartient à $\mathfrak{B}$. S. Doss appelle *moyenne* de X tout $a \in \mathfrak{X}$, s'il en existe, tel que: $(a, y) \leqq E[(X, y)]$ pour tout $y \in \mathfrak{X}$.

Soit alors $X_1, X_2, \ldots, X_j, \ldots$ une suite indéfinie d'éléments aléatoires dans $\mathfrak{X}$, mutuellement indépendants, de même loi de probabilité, et tels que X_j possède une moyenne et une seule a.

S. Doss a montré que si les sphères de $\mathfrak{X}$ sont compactes (ce qui implique que $\mathfrak{X}$ est séparable), si presque-sûrement il existe pour tout n un élément aléatoire Z_n dans $\mathfrak{X}$ tel que:

$$(Z_n, y) \leqq \frac{1}{n} \sum_{j=1}^{n} (X_j, y) \text{ pour tout } y \in \mathfrak{X};$$

alors, presque-sûrement, Z_n tend vers a [c'est à dire que presque-sûrement (Z_n, a) tend vers 0].

BIBLIOGRAPHIE

N. GLIVENKO [1]: Sur quelques formes générales de la loi des grands nombres, Rend. dei Lincei, 9, 1929, p. 466.

A. KOLMOGOROFF [1]: Sur la loi des grands nombres, Rend. dei Lincei, 9, 1929, p. 470.

M. FRÉCHET [1]: L'intégrale abstraite d'une fonction abstraite d'une variable abstraite et son application à la moyenne d'un élément aléatoire de nature quelconque, Rev. Scient. 82, 1944, p. 483.

E. MOURIER [1]: Eléments aléatoires dans un espace de Banach, Ann. Scient. de l'Institut H. Poincaré, 13, 1953, p. 161.

R. FORTET et E. MOURIER [1]: Résultats complémentaires sur les éléments aléatoires prenant leurs valeurs dans un espace de Banach, Bull. Sc. Math., 78, 1954, p.14.

[2]: Les fonctions aléatoires comme éléments aléatoires dans des espaces de Banach (à l'impression).

[3]: Convergence de la répartition empirique vers la répartition théorique, Ann. Sc. de l'Ec. Norm. sup., 70, 1953, p. 267.

S. Doss [1]: Sur la moyenne d'un élément aléatoire dans un espace distancie, Bull. des Sc. Math., 73, 1949, p. 48.

PARIS.

PERTURBATION METHODS IN THE STUDY OF KOLMOGOROFF'S EQUATIONS [1])

EINAR HILLE

1. *Introduction.*

In a recent study of Kolmogoroff's equations for the case of a countable number of possible states, I have found perturbation methods quite useful [2]). My renewed interest in these questions had three sources of inspiration: the investigation of Tosio Kato [7] of the semi-groups generated by Kolmogoroff's equations, R. S. Phillips's application of perturbation methods to the theory of semi-groups in [8], and the simple existence proof for Kolmogoroff's equations found by G. E. H. Reuter and W. Ledermann [9]. The methods of these writers will not be used explicitly in the following, but their ideas have influenced my thinking along various lines.

In the following a brief account will be given of the solution by perturbation methods of two basic problems:

(1) *the integration problem,*

(2) *the problem of constructing nul solutions.*

The case of temporally homogeneous processes will be handled in more detail, but it will be indicated below to what extent the considerations apply to more general situations.

2. *The integration problem for the backward equation.*

Let $A = (a_{jk})$ be a given matrix of constants such that

$$a_{jk} \geqq 0,\ j \neq k,\ a_{jj} \leqq 0,\ \sum_{k=1}^{\infty} a_{jk} \leqq 0,\ j, k = 1, 2, 3, \ldots \tag{2.1}$$

With this matrix we form the differential equation

$$Y'(t) = A\,Y(t),\ t > 0, \tag{2.2}$$

[1]) This research was supported by the United States Air Force, through the Office of Scientific Research of the Air Research and Development Command.

[2]) This is not a novel observation. After this paper was presented to the Symposium, David Kendall called my attention to the fact that Feller's discussion of the general Kolmogoroff equations in 1940 is based on perturbation methods. Actually, in reading Feller's argument in [1] I find that his proof for the case considered here would be essentially the same as mine. In view of the extreme generality and complexity of Feller's proof, it seems desirable to present the special case of countable states with the least possible machinary. This task is attempted in the following. See also the Addendum on the last page.

taking

$$Y(0) = I \tag{2.3}$$

as the initial condition. For the time being, (2.2) is merely a convenient abbreviation for

$$y'_{jk}(t) = \sum_{m=1}^{\infty} a_{jm}\, y_{mk}(t), \qquad j,\ k = 1,\ 2,\ 3, \ldots \tag{2.4}$$

A (weak) solution $Y(t) = (y_{jk}(t))$ is a set of differentiable functions $y_{jk}(t)$, satisfying (2.4) for $t > 0$, and (2.3) is understood as

$$\lim_{t \to 0+} y_{jk}(t) = \delta_{jk}. \tag{2.5}$$

In particular, all the infinite series in (2.4) shall converge. Actually we shall be interested mainly in positive solutions; $Y(t) > 0$ shall then mean

$$y_{jk}(t) \geqq 0 \tag{2.6}$$

for all j, k. In this case all terms on the right in (2.4) are positive except for the term with $m = j$ arising from the diagonal terms of the matrix A.

This suggests separating the diagonal terms from the rest and writing

$$A = D + T, \quad D = (\delta_{jk} a_{jk}), \quad T = ((1 - \delta_{jk}) a_{jk}) \tag{2.7}$$

so that D is a negative and T a positive matrix. Now the system

$$Y'(t) = DY(t) = Y(t)D, \ Y(0) = I \tag{2.8}$$

has a unique solution

$$Y(t) = E(t) = (e_{jk}(t)), \ e_{jk}(t) = \delta_{jk} e^{a_{jj} t}, \tag{2.9}$$

which is a positive diagonal matrix.

Using $E(t)$ we shall now try to solve (2.1) which we write in the form

$$Y'(t) - DY(t) = TY(t), \ Y(0) = I.$$

Treating this as a nonhomogeneous linear differential equation in the classical manner we are led to the integral equation

$$Y(t) = E(t) + \int_0^t E(t - s)\, TY(s)\, ds \tag{2.10}$$

which is to be understood simply as a convenient way of writing the infinite system

$$y_{jk}(t) = e_{jk}(t) + \int_0^t e^{a_{jj}(t-s)} \left\{ \sum_{m=1}^{\infty}{}' a_{jm} y_{mk}(s) \right\} ds, \quad j,\ k = 1, 2, 3, \ldots, \tag{2.11}$$

where the prime indicates that the value $m = j$ is to be omitted. Actually we

may obtain (2.11) directly from (2.4) by observing that if the latter has an "admissible" solution $\{y_{jk}(t)\}$ then

$$\frac{d}{ds}\left[e^{a_{jj}(t-s)}y_{jk}(s)\right] = e^{a_{jj}(t-s)}\sum_{m=1}^{\infty}{}' a_{jm}y_{mk}(s), \quad 0 < s < t, \tag{2.12}$$

the integrated form of which is (2.11). Here a solution will be considered admissible if the series on the right in (2.12) converges for $0 < s$, the sum being in $L(0, \omega)$ for every finite ω. Thus every admissible solution of (2.4) + (2.5) satisfies (2.11). The converse will be proved below. Let us observe in passing that a positive admissible solution satisfies

$$E(t) \leqq Y(t) \tag{2.13}$$

as is shown by anyone of the equations (2.10) to (2.12).

The equation (2.10) = (2.11) can now be solved by the method of successive approximations. We introduce a sequence of matrices $Y_n(t) = \{y_{jk}^n(t)\}$ defined by

$$Y_0(t) = E(t), \quad Y_n(t) = E(t) + \int_0^t E(t-s)\,TY_{n-1}(s)\,ds. \tag{2.14}$$

It is obvious that for $n = 0$ one has

$$0 \leqq y_{jk}^n(t) < 1, \ \sum_{k=1}^{\infty} y_{jk}^n(t) \leqq 1. \tag{2.15}$$

Using (2.1) and (2.14) one proves by induction that (2.15) holds for every n and the same method shows that

$$Y_{n-1}(t) < Y_n(t) \quad \text{or} \quad y_{jk}^{n-1}(t) \leqq y_{jk}^n(t) \tag{2.16}$$

for each j, k, and n. It follows that

$$\lim_{n\to\infty} y_{jk}^n(t) \equiv p_{jk}(t) \tag{2.17}$$

exists for each j and k and

$$0 \leqq p_{jk}(t) \leqq 1, \ \sum_{k=1}^{\infty} p_{jk}(t) \leqq 1. \tag{2.18}$$

Setting $P(t) = (p_{jk}(t))$ we see that $P(t)$ is a solution of (2.10) = (2.11). Differentiation of (2.11) with respect to t, setting $y_{jk}(t) = p_{jk}(t)$, shows that $p_{jk}(t)$ satisfies (2.4) + (2.5) so that $P(t)$ is a weak solution of (2.2) + (2.3).

3. *The minimal property and its implications.*

The solution $P(t)$ of (2.2) + (2.3) is not necessarily unique but it has the following minimal property:

If $Y(t)$ *is a positive solution of* (2.2) + (2.3) *then*

$$Y(t) \geqq P(t) \quad or \quad y_{jk}(t) \geqq p_{jk}(t) \tag{3.1}$$

for each j, k.

Here (3.1) follows from $Y(t) > Y_n(t)$ by letting $n \to \infty$ and this relation, which is true for $n = 0$ by (2.13), is shown to hold for all n by complete induction using (2.10) and (2.14).

This minimal property has important consequences. First it implies the *semi-group property*

$$P(s + t) = P(s)\, P(t), \quad 0 \leqq s, t < \infty. \tag{3.2}$$

To prove this, we observe first that if $B = (b_{jk})$ is a given matrix bounded in the sense that

$$||B|| \equiv \sup_j \sum_{k=1}^{\infty} |b_{jk}| < \infty, \tag{3.3}$$

then $P(t)\ B$ is a solution of the system

$$Y'(t) = A\, Y(t), \quad Y(0) = B, \tag{3.4}$$

and if B is positive, the argument used in proving (3.1) may be used to show that $P(t)\ B$ is the minimal solution of (3.4). If s is fixed positive, it follows that

$$P(s + t) = P(t)\ \ P(s)$$

for both sides represent solutions of (3.4) with $B = P(s) > 0$ and the right member is the least such solution. But the left member also has a minimal property; it is the least solution as function of t tending to I as $t \to 0+$, $s \to 0+$ and, since the right member has the same limit property, the inequality may be reversed so that (3.2) holds.

We have consequently

$$p_{jk}(t + s) = \sum_{m=1}^{\infty} p_{jm}(t) p_{mk}(s), \quad j, k = 1,\ 2,\ 3, \ldots \tag{3.5}$$

where the series have positive terms and converge uniformly with respect to s and t in any finite closed interval $[0, \omega]$. It follows that for t fixed positive, $h > 0$,

$$\frac{1}{h}\,[p_{jk}(t + h) - p_{jk}(t)] + p_{jk}(t)\,\frac{1}{h}\,[1 - p_{kk}(h)] = \sum_{m=1}^{\infty}{}' p_{jm}(t)\,\frac{1}{h}\,p_{mk}(h), \tag{3.6}$$

where the prime indicates that the value $m = k$ is omitted. We note that

$$\lim_{h \to 0+} \frac{1}{h}\, p_{jk}(h) = a_{jk},\ j \neq k, \quad \lim_{h \to 0+} \frac{1}{h}\,[1 - p_{kk}(h)] = -\,a_{kk}. \tag{3.7}$$

This follows readily from (2.11) since the infinite series in this formula converges uniformly to a_{jk} as $s \to 0+$. Consequently, letting $h \to 0+$ in (3.6) we see that for any fixed n

$$p'_{jk}(t) - a_{kk}\, p_{jk}(t) \geqq \sum_{m=1}^{n}{}' p_{jm}(t)\, a_{mk}, \tag{3.8}$$

so that the series

$$\sum_{m=1}^{\infty}{}' p_{jm}(t)\, a_{mk}$$

converges and its sum is at most equal to the left member of (3.8). Actually we have equality so that

$$p'_{jk}(t) = \sum_{m=1}^{\infty} p_{jm}(t)\, a_{mk}, \quad j, k = 1, 2, 3, \ldots \tag{3.9}$$

and $P(t)$ is also a solution of the adjoint equation

$$Z'(t) = Z(t)A, \quad Z(0) = I. \tag{3.10}$$

This is immediately evident if the infinite series in (3.6) are actually finite as is the case when A and hence also $P(t)$ are triangular matrices. More generally, the conclusion may be drawn if the series converge uniformly in h for fixed $t > 0$. Since by (2.11) $\sum'_k p_{mk}(h) \leqq |a_{mm}|h$, uniform convergence will hold if, for instance,

$$\sum_{m=1}^{\infty} |a_{mm}| p_{jm}(t) < \infty.$$

This is perhaps always the case, if so the proof escapes me. For a simple proof that $P(t)$ satisfies (3.10), see Reuter and Ledermann [9].

The adjoint equation may be studied directly by the same method. The equivalent integral equation

$$Z(t) = E(t) + \int_0^t Z(s)\, T\, E(t - s)ds \tag{3.11}$$

or

$$z_{jk}(t) = e_{jk}(t) + \int_0^t \left\{ \sum_{m=1}^{\infty}{}' z_{jm}(s)\, a_{mk} \right\} e^{a_{kk}(t-s)}\, ds \tag{3.12}$$

is introduced and the successive approximations $Z_n(t)$ with $Z_0(t) = E(t)$. The matrices $Z_n(t)$ are positive, $||Z_n(t)|| \leqq 1$, and they form a monotone increasing sequence converging to a limit $Q(t)$. This matrix $Q(t)$ satisfies (3.10) and is shown to be the minimal solution. As such it also has the semi-group property and, the analogue of (3.6) being uniformly convergent in h, satisfies the adjoint of the adjoint equation, that is, (2.2) + (2.3). Since $P(t)$ and $Q(t)$ satisfy both

equations and each is minimal with respect to one equation it follows that they are identical:

The two equations (2.2) + (2.3) *and* (3.10) *have the same minimal solution, namely* $P(t)$.

4. *Extensions.*

The methods used above apply also to the case in which the coefficients of the differential equations are no longer constants. We have then the two adjoint Kolmogoroff equations

$$-\frac{\partial}{\partial s} Y(s,t) = A(s)\, Y(s,t), \quad s \leqq t, \quad Y(t,t) = I, \tag{4.1}$$

$$\frac{\partial}{\partial t} Z(s,t) = Z(s,t)\, A(t), \quad s \geqq t, \quad Z(s,s) = I. \tag{4.2}$$

Here $A(s) = (a_{jk}(s))$ where the $a_{jk}(s)$ are continuous functions of s satisfying the inequalities

$$a_{jk}(s) \geqq 0, \; j \neq k, \; a_{jj}(s) \leqq 0, \; \sum_{k=1}^{\infty} a_{jk}(s) \leqq 0. \tag{4.3}$$

We separate the diagonal elements into a matrix $D(s)$ and set

$$A(s) = D(s) + T(s), \tag{4.4}$$

$$E(s,t) = \exp\left\{\int_s^t D(\sigma)\, d\sigma\right\} = \left\{\delta_{jk} \exp\left[\int_s^t a_{jj}(\sigma)\, d\sigma\right]\right\} = \left\{e_{jk}(s,t)\right\}. \tag{4.5}$$

The integral equation associated with (4.1) is then

$$Y(s,t) = E(s,t) + \int_s^t E(s,u)\, T(u)\, Y(u,t)\, du \tag{4.6}$$

or

$$y_{jk}(s,t) = e_{jk}(s,t) + \int_s^t e_{jj}(s,u) \left\{ \sum_{m=1}^{\infty}{}' a_{jm}(u)\, y_{mk}(u,t) \right\} du. \tag{4.7}$$

The successive approximations $Y_n(s,t)$ with $Y_0(s,t) = E(s,t)$ form a monotone increasing sequence of positive matrices with $||Y_n(s,t)|| \leqq 1$, converging to a limit $P(s,t) \equiv \{p_{jk}(s,t)\}$. This matrix is a solution of (4.1), in fact the minimal solution.

The minimal property gives the Chapman-Kolmogoroff equation

$$P(s,t) = P(s,u)\, P(u,t), \quad s \leqq u \leqq t, \tag{4.8}$$

for the right member as a function of s tends to $P(u,t)$ as $s \to u-$ and is the least solution of (4.1) having this property while the left member is the least solution tending to I as $s \to t-$. The two resulting inequalities together imply (4.8).

Using (4.8) we can then verify that $P(s, t)$ qua function of t satisfies (4.2). We have merely to observe that

$$P(s, t + h) = P(s, t)\, P(t, t + h)$$

whence

$$\frac{1}{h}\,[p_{jk}(s, t + h) - p_{jk}(s, t)] + p_{jk}(s, t)\,\frac{1}{h}\,[1 - p_{kk}(t, t + h)]$$

$$= \sum_{m=1}^{\infty}{}' p_{jm}(s, t)\,\frac{1}{h}\,p_{mk}(t, t + h)$$

and passing to the limit with h we obtain, assuming uniform convergence,

$$\frac{\partial}{\partial t}\,p_{jk}(s, t) = \sum_{m=1}^{\infty} p_{jm}(s, t)\, a_{mk}(t) \tag{4.9}$$

as asserted. Just as in the constant case, $P(s, t)$ is shown to be the minimal solution also of the adjoint equation.

5. *Uniqueness.*

We shall consider briefly questions of uniqueness in the temporally homogeneous case. To fix ideas we consider only solutions belonging to the Markoff algebra $\mathbf{M}$ of matrices $B = (b_{jk})$ normed by (3.3) in which the algebraic operations are defined in the usual manner. The matrix $P(t)$ of section 2 is an element of $\mathbf{M}$ for $t \geqq 0$, on the other hand, the Kolmogoroff matrix $A = (a_{jk})$ with properties (2.1) normally is not in $\mathbf{M}$. If exceptionally $A \in \mathbf{M}$ then $P(t) = \exp\,(t\,A)$,

$$\lim_{t \to 0+} ||P(t) - I|| = 0 \tag{5.1}$$

and $P(t)$ is the only solution of (2.2) + (2.3) as well as of (3.10). Conversely, by Theorem 8.4.2 of [2], if a matrix function $P(t) \in \mathbf{M}$ satisfies (3.2) and (5.1), then $P(t)$ satisfies equations of the type (2.2) + (2.3) and (3.10) with a matrix $A \in \mathbf{M}$.

Now if the initial condition holds only in the weak sense of (2.3) we may very well have several solutions. If this is the case the system

$$Y'(t) = A\,Y(t), \tag{5.2}$$

$$\lim_{t \to 0+} y_{jk}(t) = 0 \tag{5.3}$$

has a solution which is not identically zero. We speak of $Y(t)$ as a *weak nul solution* of (5.2). Usually we can go further, however. Suppose that (5.2) + (5.3) has a solution $Y_0(t) \in \mathbf{M}$ such that $||Y_0(t)|| \in L(0, 1)$. We can then form

$$Y_1(t) \equiv \int_0^t Y_0(s)\, ds \equiv \left(\int_0^t y_{jk}(s)\, ds\right)$$

and it is not difficult to show that the operations of integration and left hand multiplication by the matrix A commute. It follows that $Y_1(t)$ is also a nul solution and here (5.3) is replaced by the stronger condition

$$\lim_{t\to 0+} ||Y(t)|| = 0 \tag{5.4}$$

so we speak of a *strong nul solution*. We have $Y_1'(t) \in \mathbf{M}$ and $||Y_1'(t)|| \in L(0, 1)$. The process may be repeated and integrals of fractional order may be used. In this manner strong nul solutions may be constructed such that $||Y(t)|| \to 0$ faster than any preassigned power of t. New nul solutions may also be obtained by multiplying a given nul solution $Y_0(t)$ on the right by a matrix B in $\mathbf{M}$.

If the adjoint system

$$Z'(t) = Z(t)\, A \tag{5.5}$$

$$\lim_{t\to 0+} z_{jk}(t) = 0 \tag{5.6}$$

should possess a non trivial solution $Z_0(t)$, multiplication on the left by a bounded matrix B leads to new nul solutions. Fractional integration may also be attempted, but here the justification of the interchange of integration and right hand multiplication by A is usually more delicate to carry through. There is no difficulty if $Z_0(t)$ is positive for instance. Whether or not the existence of a nul solution implies the existance of a positive nul solution, appears to be an open question.

We conclude these rambling remarks by giving a necessary and sufficient condition for the existence of a nul solution of (5.2) satisfying a mild restriction as to its growth.

A necessary and sufficient condition that (5.2) + (5.4) *have a non trivial solution* $Y(t) \in \mathbf{M}$ *such that*

$$\limsup_{t\to+\infty} t^{-1} \log ||Y(t)|| \leqq \omega < \infty, \tag{5.7}$$

is the existence of a matrix $X(\lambda) \in \mathbf{M}$, *defined, holomorphic and bounded in every half-plane* $\mathbf{R}(\lambda) \geqq \omega + \varepsilon$, $\varepsilon > 0$, *such that.*

$$A\, X(\lambda) = \lambda X(\lambda). \tag{5.8}$$

The proof uses an argument employed by the writer in many similar situations (see [4], Theorem 2.2). If $Y(t)$ exists with the stated properties, then the Bochner integral

$$X(\lambda) = \int_0^\infty e^{-\lambda t}\, Y(t)\, dt \tag{5.9}$$

exists and is bounded and holomorphic in $\mathbf{R}(\lambda) \geqq \omega + \varepsilon$. Here

$$A\,X(\lambda) = \Big(\sum_{m=1}^{\infty} a_{jm} \int_0^{\infty} e^{-\lambda t}\, y_{mk}(t)dt\Big) = \Big(\int_0^{\infty} e^{-\lambda t} \Big[\sum_{m=1}^{\infty} a_{jm} y_{mk}(t)\Big] dt\Big)$$

for

$$\sum_{m=1}^{\infty} |a_{jm} y_{mk}(t)||e^{-\lambda t}| \leqq 2\,||Y(t)||\, e^{-\mathbf{R}(\lambda)t}|a_{jj}|,$$

so that integration and summation may be interchanged by Lebesgue's theorem. Hence

$$A\,X(\lambda) = \Big(\int_0^{\infty} e^{-\lambda t}\, y'_{jk}(t)dt\Big) = \Big(\lambda \int_0^{\infty} e^{-\lambda t}\, y_{jk}(t)dt\Big) = \lambda\, X(\lambda).$$

Thus the condition is necessary.

Conversely, if it is satisfied and $X_0(\lambda)$ is a solution of (5.8) holomorphic and bounded in $\mathbf{R}(\lambda) \geqq \omega + \varepsilon$, then $(\lambda - \omega + 1)^{-3} X_0(\lambda) \equiv X(\lambda)$ is also a solution and we can form the Bochner integral

$$Y(t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{\lambda t}\, X(\lambda)\, d\lambda,\ \gamma > \omega. \tag{5.10}$$

With the derivative taken in the strong sense, we have

$$Y'(t) = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{\lambda t}\, \lambda X(\lambda) d\lambda = \frac{1}{2\pi i} \int_{\gamma - i\infty}^{\gamma + i\infty} e^{\lambda t}\, A\, X(\lambda) d\lambda = A\, X(t)$$

where the last step is justified by the estimate

$$\sum_{m=1}^{\infty} |a_{jm}\, x_{mk}(\lambda)||e^{\lambda t}| \leqq 2\,|a_{jj}|\,||X_0(\gamma + i\tau)||\, e^{\gamma t}\,(1 + \tau^2)^{-3/2}.$$

Thus $Y(t)$ is a solution of (5.2), it satisfies (5.7) since $\gamma - \omega$ can be made as small as we please. Further $Y(t) \equiv 0$ for $t \leqq 0$ by (5.10) because the integral is independent of γ, $\gamma > \omega$, and γ may be taken as large as we please, making the integral arbitrarily small when $t \leqq 0$. Thus $Y(t)$ is a strong nul solution so the condition is also sufficient.

A necessary and sufficient condition that (5.2) have a positive strong nul solution satisfying (5.7) is consequently that among the bounded solutions of (5.8) there exist a matrix $X(\lambda)$ the elements of which are completely monotone on the positive real axis for $\lambda > \omega$.

It is usually harder to get a hold on the adjoint equation (5.5), where the interchange of the operations of integration and right hand multiplication by A is apt to be a delicate process. But if (5.5) has a positive strong nul solution

$Z(t)$ satisfying (5.7), then its Laplace transform $X(\lambda)$ is a holomorphic bounded solution of

$$X(\lambda)\, A = \lambda\, X(\lambda) \tag{5.11}$$

in $\mathbf{R}(\lambda) \geqq \omega + \varepsilon$, $\varepsilon > 0$, so the existence of such a matrix $X(\lambda)$ is at least a necessary condition.

6. *The existence of nul solutions.*

We proceed now to the proposed second application of perturbation methods now to the existence of nul solutions. Suppose that we know the existence of nul solutions for a particular Kolmogoroff matrix U, suppose that V is another Kolmogoroff matrix, does $U + V$ admit nul solutions? The following result lies almost on the surface.

If the system

$$Z'(t) = Z(t)\, U, \lim_{t \to 0+} ||Z(t)|| = 0, \tag{6 1}$$

has a non-trivial solution and if $V \in \mathbf{M}$, *then*

$$Q'(t) = Q(t)(U + V), \lim_{t \to 0+} ||Q(t)|| = 0, \tag{6.2}$$

also has a non-trivial solution.

In order to prove this, set

$$Q(t) = Z(t) + S(t). \tag{6.3}$$

Proceeding formally, one sees that

$$\frac{d}{ds}\,[S(s)] - S(s)U = Z(s)\, V + S(s)V.$$

Here we multiply on the right by $P(t - s; U)$ where $P(t; U)$ is the existing positive minimal solution of

$$P'(t) = U\, P(t), \lim_{t \to 0+} P(t) = I.$$

The result is an exact derivative which we integrate from 0 to t. Since by assumption $S(0) = 0$ we get

$$S(t) = G(t) + \int_0^t S(s)\, V\, P(t - s; U)ds \tag{6.3}$$

with

$$G(t) = \int_0^t Z(s)\, V\, P(t - s; U)ds. \tag{6.4}$$

Here we can interpret the integrals as taken in the sense of Bochner and (6.3) becomes simply an integral equation of the Volterra type with a known strongly

continuous function $G(t)$ and a bounded, strongly continuous kernel $V\,P(t-s;\,U)$. The solution may be obtained by the method of successive approximations. If

$$\max_t ||V\,P(t;\,U)|| = K,\ \int_0^t ||Z(s)\,V\,P(t-s;\,U)||ds \equiv \gamma(t), \tag{6.5}$$

one finds that

$$||S(t)|| \leqq \gamma(t) + K\int_0^t e^{K(t-s)}\gamma(s)ds. \tag{6.6}$$

Since

$$\gamma(t) = o[||Z(t)||] \text{ as } t \to 0+, \tag{6.7}$$

one sees that $Q(t) \equiv Z(t) + S(t)$ is not identically zero.

Equation (6.3) may be differentiated with respect to t where, however, the resulting integrals have to be interpreted in the weak sense used above, that is, the integral of a matrix is the matrix of the integrals of the elements, and from the resulting relations one readily concludes that $Q(t)$ satisfies (6.2).

This procedure is not restricted to the case in which $V \in \mathbf{M}$. It suffices that

$$V\,P(t;\,U) \in \mathbf{M},\quad 0 < t,\ ||V\,P(t;\,U)|| \in L(0,\,a),\quad 0 < a < \infty. \tag{6.8}$$

Equation (6.3) now has an integrable kernel, $G(t)$ is still strongly continuous and (6.7) holds, so that the classical method applies and the resulting solution $Q(t)$ is not identically zero.

It is not an easy matter to characterize the class of matrices V which satisfy (6.8). At least in some interesting cases V does not have to be bounded. On the other hand, if $U \notin \mathbf{M}$, U does not seem to satisfy (6.8), for one can show (see [3], Theorem 3, for the method) that $\lim\sup_{t\to 0+} t||U\,P(t;\,U)|| \geqq e^{-1}$.

In special cases one may allow a wider class of perturbation matrices than those admitted by (6.8). In a recent note [6] the following result was proved:

If U is a triangular matrix, if $Z(t)$ is a positive solution of (6.1), *satisfying* (5.7) *and such that* $Z'(t) \in \mathbf{M}$. $0 \leqq t$, *if the resolvent* $R(\lambda;\,U)$ *of* U *satisfies* $||\lambda\,R(\lambda;\,U)|| \leqq K$ *for* $\mathbf{R}(\lambda) > 0$, *if* $B \in \mathbf{M}$ *and* $(1+K)\,||B|| < 1$, *and if* $V = \beta B - B\,U$, $\beta \geqq 0$, *then* (6.2) *also has a solution.*

The proof involves the techniques of section 5 and several delicate interchanges of integration and matrix multiplication. The class of perturbation matrices, $V = \beta B - BU$, admissible in this theorem, appears to be as large as one can expect, but it would be desirable to remove the many restrictions on U. An earlier theorem of this type was given with insufficient proof in [5].

Bibliography

[1] W. Feller, On the integro-differential equations of purely discontinuous Markoff processes. Trans. Amer. Math. Soc., 48 (1940) 488—515.

[2] E. Hille, Functional Analysis and Semi-Groups. Amer. Math. Soc. Colloquim Publ., Vol. 31, xi + 528 pp., New York, 1948.

[3] E. Hille, On the differentiability of semi-group operators. Acta Sci. Math. Szeged, 12 B (1950) 19—24.

[4] E. Hille, The abstract Cauchy problem and Cauchy's problem for parabolic differential equations. J. d'Analyse Math., Jerusalem, 3 (1953—54) 81—196.

[5] E. Hille, Quelques remarques sur les équations de Kolmogoroff. Bull. Soc. Math. Phys. R.P. Serbie, 5 (1953—54) 3—14.

[6] E. Hille, Sur un théorème de perturbation. Univ. e Politecnico di Torino, Rendiconti del Sem. Mat., 13 (1954) 169—184.

[7] T. Kato, On the semi-groups generated by Kolmogoroff's differential equations. J. Math. Soc. Japan, 6 (1954) 1—15.

[8] R. S. Phillips, Perturbation theory for semi-groups of linear operators. Trans. Amer. Math. Soc., 74 (1953) 199—221.

[9] G. E. H. Reuter and W. Ledermann, On the differential equations for the transition probabilities of Markov processes with enumerably many states. Proc. Cambridge Phil. Soc., 49 (1953) 247—262.

Addendum

Further examination of the literature has revealed that the method used in section 2 above is essentially that of J. L. Doob, Markoff chains — denumerable case, Trans. Amer. Math. Soc., 455—473, esp. pp. 458—460. The discussion of the adjoint equation may be much simplified by observing that $Z_1(t) = Y_1(t)$ and hence, by complete induction, $Z_n(t) = Y_n(t)$ and $Q(t) = P(t)$. This observation also goes back to Doob. [Added in proof October, 1955.]

Yale University.

SOME PATHOLOGICAL MARKOV PROCESSES WITH A DENUMERABLE INFINITY OF STATES AND THE ASSOCIATED SEMIGROUPS OF OPERATORS ON l

D. G. KENDALL and G. E. H. REUTER

1. *Introduction.* It is known (see, for example, Doob [3]) that a temporally homogeneous Markov process $\{x_t : t \geqq 0\}$ with a denumerable infinity of possible states can be characterised by the transition-probabilities

$$p_{ij}(t) \equiv \boldsymbol{P}\{x_{u+t}(\omega) = j \mid x_u(\omega) = i\} \tag{1}$$

and the initial distribution $\boldsymbol{P}\{x_0(\omega) = i\}$ (here $i, j \geqq 1$, $t \geqq 0$, ω is the "sample point" in the basic probability-space and the left-hand side of (1) does not depend on u, for $u \geqq 0$). The array of functions $\{p_{ij} : i, j \geqq 1\}$ will by a harmless convention be loosely referred to as "the process". It necessarily satisfies the conditions (2) to (5) below and in order to eliminate the grosser pathologies we shall also always assume that the continuity condition (6) holds:-

$$p_{ij}(t) \geqq 0; \tag{2}$$

$$\sum_{j=1}^{\infty} p_{ij}(t) = 1; \tag{3}$$

$$p_{ij}(u+v) = \sum_{\alpha=1}^{\infty} p_{i\alpha}(u)p_{\alpha j}(v); \tag{4}$$

$$p_{ij}(0) = \delta_{ij}; \tag{5}$$

$$\lim_{t \downarrow 0} p_{ii}(t) = 1. \tag{6}$$

In all of (2) to (6) the following ranges of the variables are to be understood:- $i, j \geqq 1$; $t, u, v \geqq 0$.

Conversely, if (2) to (6) are satisfied then there exists a Markov process associated with the array $\{p_{ij} : i, j \geqq 1\}$ in the manner indicated at equation (1). In the analytical study of such systems it proves essential to consider the effects of weakening condition (3) to

$$\sum_{j=1}^{\infty} p_{ij}(t) \leqq 1. \tag{3*}$$

When (3*) holds instead of (3), and (2), (4), (5) and (6) are satisfied, we shall follow Jensen [10] in referring to the array $\{p_{ij} : i, j \geqq 1\}$ as a *quasi-process*.

For quasi-processes (and so *a fortiori* for processes) it is easy to show that $p_{ij}(.)$ is (for each fixed i) a uniformly continuous function of t on $0 \leqq t < \infty$, uniformly with regard to j, and that $p_{ii}(.)$ is (for each fixed i) a positive super-multiplicative function of t. The differentiability properties of $p_{ij}(.)$ at $t = 0$ (though *not* for $t > 0$) are now completely understood (see Doob [2] and Kolmogorov [13]; for the extension of their results from processes to quasi-processes see Kendall [12]). The limits

$$q_{ij} \equiv \lim_{t \downarrow 0} \frac{p_{ij}(t)}{t} \qquad (i \neq j) \tag{7}$$

and

$$q_i \equiv \lim_{t \downarrow 0} \frac{1 - p_{ii}(t)}{t} \tag{8}$$

always exist, q_{ij} is always finite and

$$\sum_{j \neq i} q_{ij} \leqq q_i \leqq \infty. \tag{9}$$

It is known in addition that there exist *processes* (so that (3) is satisfied) for which $(=, =)$, $(=, <)$ and $(<, <)$ hold in (9). The two non-trivial examples were discovered by Kolmogorov [13]; they show that the statement (9) cannot be improved upon. [1])

If we write $\boldsymbol{Q}$ for the matrix with elements

$$Q_{ii} \equiv -q_i, \; Q_{ij} \equiv q_{ij} \qquad (i \neq j), \tag{10}$$

then a superficial assessment from the intuitive standpoint (from which q_i appears as "the rate at which probability flows out of the state i", and q_{ij} appears as "the rate at which probability is transferred from state i to state j") might lead one to suppose that the functions $p_{ij}(\cdot)$ (and so the matrix $\boldsymbol{P}(t)$ of which $p_{ij}(t)$ is the (i, j)th element) would be uniquely determined by the matrix $\boldsymbol{Q}$. In many of the simplest cases this is true (see, e.g. Feller [4]) and $\boldsymbol{P}(t)$ can be found for all $t \geqq 0$ as the unique solution to the so-called "backward" and "forward" equations of Kolmogorov:

$$\begin{gathered} \frac{d}{dt}\boldsymbol{P}(t) = \boldsymbol{Q}\boldsymbol{P}(t) = \boldsymbol{P}(t)\boldsymbol{Q}, \\ \boldsymbol{P}(0) = \boldsymbol{I}. \end{gathered} \tag{11}$$

(Here $\boldsymbol{I}$ denotes the identity matrix and d/dt denotes differentation element-

[1]) We do not know of an example for which $(<, =)$ hold in (9), although there is no reason to suspect that this cannot occur. It would be useful if this niche in the rogues' gallery could be filled. [*Note added in proof.* It now has been. (See a forthcoming paper by one of us in *The Quarterly Journal of Mathematics.*)]

by-element.) But the situation is in fact much more complicated. Doob [2] discovered that the "forward" equations may be false, and that the "backward" equations may have more than one solution. Thus there may be several processes and quasi-processes associated with one matrix $\boldsymbol{Q}$; a particularly drastic example of such non-uniqueness has been given by Ledermann and Reuter [14]. The Kolmogorov examples to be discussed in the present paper are sufficient to show that the equations (11) can be meaningless, and that even when meaningful they can be false. For these reasons the emphasis has now shifted from the matrix $\boldsymbol{Q}$ and the equations (11) to the semigroups of positive contraction operators on l which are associated with processes and quasi-processes of the type under discussion.

Let l be the real Banach space whose elements x are sequences $x_1, x_2, \ldots$ of real numbers such that

$$||x|| \equiv \sum_{j=1}^{\infty} | x_j | < \infty,$$

and let P_t denote the linear bounded operator which maps l into itself and is defined by

$$(P_t x)_j \equiv \sum_{i=1}^{\infty} x_i p_{ij}(t) \tag{12}$$

for each $t \geqq 0$. Then (Hille [6]) if (2), (3*) and (4) to (6) are satisfied (that is, if the p_{ij} constitute a quasi-process) it is known that $\{P_t : t \geqq 0\}$ is a *contraction semigroup* [2]) on l. That is to say, the operators P_t satisfy the following conditions:-

$$\left.\begin{aligned} P_u P_v &= P_{u+v} && (u, v \geqq 0); \\ || P_t || &\leqq 1 && (t \geqq 0); \\ P_t x &\geqq 0 \text{ when } x \geqq 0 && (t \geqq 0); \\ P_0 &= I; \\ \lim_{t \downarrow 0} ||P_t x - x|| &= 0 && (x \in l). \end{aligned}\right\} \tag{13}$$

(Here $x \geqq 0$ means that each coordinate x_j is non-negative.)

Conversely it is known that any contraction semigroup on l admits a unique representation (12) in terms of a quasi-process $\{p_{ij} : i, j \geqq 1\}$. In what follows, "contraction semigroup on l" will be abbreviated to "semigroup". With this understanding, then, semigroups and quasi-processes have been put into (1, 1) correspondence with one another. Moreover, (3) will be satisfied

[2]) Whenever we refer to a contraction semigroup we shall mean a family of operators satisfying (13). This is not quite the customary usage and the insistence on strong continuity at $t = 0$ should be noted.

and the quasi-process will be a *process* if and only if the corresponding semigroup is a *transition semigroup*; i.e., if and only if

$$|| P_t x || = || x || \text{ when } x \geqq 0 \qquad (t \geqq 0). \tag{14}$$

From this point of view the problem of characterising a process or quasi-process in terms of its behaviour near $t = 0$ admits of a satisfactory solution. It is known that the domain of existence $\mathfrak{D}(\Omega)$ of the limit

$$\Omega x \equiv \text{strong } \lim_{t \downarrow 0} \frac{P_t x - x}{t} \tag{15}$$

is a linear set dense in l, and further that when the operator Ω (and so, by an important convention, $\mathfrak{D}(\Omega)$) is given then the semigroup (and so the quasi-process) is uniquely determined. The operator Ω is called the *infinitesimal generator* of the semigroup (or of the quasi-process). In the sense that the matrix $\boldsymbol{Q}$ and the operator Ω are both differential operators acting at $t = 0$ we can say that $\boldsymbol{Q}$ is an historical approximation to Ω, and a knowledge of $\boldsymbol{Q}$ may, as we shall see, be a useful clue (though rarely a sufficient one) for inferring the correct form of Ω in a particular problem. Similarly we can say that the Kolmogorov differential equations (11) are an historical approximation to the following set of rules for calculating the matrix representation of the semigroup (and so the individual functions $p_{ij}(.)$ constituting the quasi-process) when Ω is given.

Rule 1: Given $x \in l$ and $\lambda > 0$, solve the equation

$$\lambda \xi - \Omega \xi = x \tag{16}$$

to obtain the (necessarily existing and necessarily unique) solution,

$$\xi = R_\lambda x,$$

in $\mathfrak{D}(\Omega)$. R_λ will be a bounded linear operator on l with the matrix representation

$$(R_\lambda x)_j = \sum_{i=1}^{\infty} x_i r_{ij}(\lambda) \tag{17}$$

where each $r_{ij}(\lambda) \geqq 0$.

Rule 2: Find the (necessarily existing and necessarily unique) bounded continuous functions $p_{ij}(\cdot)$ having the Laplace transforms

$$r_{ij}(\lambda) = \int_0^\infty e^{-\lambda t} p_{ij}(t) dt \qquad (\lambda > 0). \tag{18}$$

Then $\{p_{ij} : i, j \geqq 1\}$ is the quasi-process and the semigroup operators are defined by (12).

To these rules we must add the following ones.

Rule 3: Let $\mathfrak{D}(\Omega)$ be any linear set in l and let Ω map $\mathfrak{D}(\Omega)$ linearly into l. Then Ω (with domain $\mathfrak{D}(\Omega)$) will be the infinitesimal generator of a semigroup associated with a quasi-process if (and only if):-

(a) $\mathfrak{D}(\Omega)$ is dense in l;

(b) the equation (16) has (for each $\lambda > 0$ and for each $x \in l$) a unique solution $\xi \equiv R_\lambda x$ in $\mathfrak{D}(\Omega)$;

(c) $R_\lambda x \geqq 0$ when $x \geqq 0$ $\quad (\lambda > 0)$;

(d) $\| R_\lambda x \| \leqq \| x \| / \lambda$ $\quad (x \in l,\ \lambda > 0)$.

Rule 4: The semigroup generated by Ω will be a transition semigroup and the quasi-process will be a process (i.e., (3) will be satisfied) if (and only if)

(e) $\| R_\lambda x \| = \| x \| / \lambda$ when $x \geqq 0$ $\quad (\lambda > 0)$.

Condition (c) is equivalent to

(c') $r_{ij}(\lambda) \geqq 0$ $\quad (i,\ j \geqq 1;\ \lambda > 0)$,

and if this holds (d) and (e) are respectively equivalent to

(d') $\sum_{j=1}^{\infty} r_{ij}(\lambda) \leqq 1/\lambda$ $\quad (i \geqq 1,\ \lambda > 0)$,

(e') $\sum_{j=1}^{\infty} r_{ij}(\lambda) = 1/\lambda$ $\quad (i \geqq 1,\ \lambda > 0)$.

The rules 1 to 4 summarise in the form most suited to our purpose the celebrated theorem of Hille [6, 7] and Yosida [19, 20].

We have seen that the formulation in terms of semigroups should make possible a completely satisfactory theory of the most general Markov process defined by (2), (3) and (4) to (6). The foundations of such a theory are now being constructed, notably by Hille [8, 9], Feller [3]), and Kato [11]. However, these writers have for the most part confined their attention to the processes and quasi-processes for which the relation

$$\sum_{j \neq i} q_{ij} = q_i < \infty$$

holds for each $i \geqq 1$. That this is a genuine limitation was demonstrated by Kolmogorov [13] who produced two *processes* (satisfying (3)) with the following exceptional properties.

First example (K1): $\quad q_1 = \infty.$ $\qquad (19)$

Second example (K2): $\quad \sum_{j \neq 1} q_{1j} < q_1 < \infty.$ $\qquad (20)$

His procedure was to define sets of functions $p_{ij}(\cdot)$ and then to verify (2),

[3]) Feller's work on denumerable Markov processes is as yet unpublished, but the authors of this paper are particularly indebted to it.

(3), (4) to (6) and (19) (or (20)). We propose to "re-manufacture" his examples via the theory of semigroups, and we hope that the detailed analytical study from this point of view will be of value in guiding the general theory towards completeness by showing how such pathologies in behaviour as (19) and (20) appear when transferred to the semigroup formulation.

We shall also investigate the way in which the "backward" and "forward" differential equations break down in K1 and K2, and we shall find equations which replace them.

Finally we shall explore the actual stochastic motions in K1 and K2 by synthesising the sample functions for each process in terms of independent negative-exponential random variables. In this last part of the work we shall make close contact with the work of Lévy [15, 16] who, independently of Kolmogorov, showed the possibility of infinite values for the diagonal elements of $\boldsymbol{Q}$.

2. *Kolmogorov's first example as a semigroup of operators.*

Kolmogorov [13] introduced K1 as a system of functions $p_{ij}(\cdot)$ $(i, j \geqq 1)$ satisfying (2), (3) and (4) to (6) and having the following $\boldsymbol{Q}$-matrix:-

$$\begin{Vmatrix} -\infty & 1 & 1 & 1 & . & . & . \\ a_2 & -a_2 & 0 & 0 & & & \\ a_3 & 0 & -a_3 & 0 & & & \\ a_4 & 0 & 0 & -a_4 & & & \\ . & . & . & & & & \end{Vmatrix} \tag{21}$$

where

$$a_\alpha > 0 \ (\alpha \geqq 2) \text{ and } \sum_{\alpha=2}^{\infty} \frac{1}{a_\alpha} < \infty. \tag{22}$$

When thought of as an "approximation" to Ω, $\boldsymbol{Q}$ acts by right multiplication on the row-matrix $\boldsymbol{x} \equiv [x_1, x_2, \ldots]$, and so a first attempt to set up the infinitesimal generator would lead us to write

$$(\Omega x)_1 \sim -\infty x_1 + \sum_{\alpha \geqq 2} a_\alpha x_\alpha$$

and

$$(\Omega x)_\beta \sim x_1 - a_\beta x_\beta \qquad (\beta \geqq 2).$$

We shall in fact take as our starting point the following *definitions* of $\mathfrak{D}(\Omega)$ and Ω.

(1) *A vector x from l is in $\mathfrak{D}(\Omega)$ if and only if*

$$\sum_{\alpha \geqq 2} | a_\alpha x_\alpha - x_1 | < \infty. \tag{23}$$

(2) *If* $x \in \mathfrak{D}(\Omega)$ *then*

$$\left.\begin{aligned} (\Omega x)_1 &\equiv \sum_{\alpha \geqq 2} (a_\alpha x_\alpha - x_1) \\ (\Omega x)_\beta &\equiv x_1 - a_\beta x_\beta \qquad (\beta \geqq 2). \end{aligned}\right\} \tag{24}$$

We shall find it a useful convention to agree that summations over i or j imply $i, j = 1, 2, 3, \ldots$ and that summations over α or β imply $\alpha, \beta = 2, 3, 4, \ldots$ (the first state being now excluded). This will be adopted without further comment. We must now show that Ω as just defined satisfies the conditions of Rules 3 and 4 of § 1. It will then follow that Ω generates a *transition* semi-group and so a *process* (satisfying (3)).

Let us notice first that any x satisfying (23) is automatically in l, because

$$x_\alpha = (a_\alpha x_\alpha - x_1)/a_\alpha + x_1/a_\alpha ,$$

and that $\mathfrak{D}(\Omega)$ is a linear subset of l. If u^i is the ith coordinate-vector in l, defined by $(u^i)_j \equiv \delta_{ij}$, then

$$u^1 \notin \mathfrak{D}(\Omega) \text{ and } u^\alpha \in \mathfrak{D}(\Omega) \qquad (\alpha \geqq 2) \tag{25}$$

Also

$$[\mathbf{1},\ 1/a_2,\ 1/a_3, \ldots] \in \mathfrak{D}(\Omega). \tag{26}$$

It follows from (25) and (26) that $\mathfrak{D}(\Omega)$ is dense in l, and it is trivial to verify that Ω is linear and maps $\mathfrak{D}(\Omega)$ into l.

On taking coordinates the equation (16) becomes

$$\lambda\xi_1 - \sum_\alpha (a_\alpha \xi_\alpha - \xi_1) = x_1$$

and

$$(\lambda + a_\beta)\xi_\beta - \xi_1 = x_\beta .$$

The second equation determines a unique ξ in $\mathfrak{D}(\Omega)$ for each arbitrary choice of the real number ξ_1. Thus ξ_1 must be chosen to satisfy the first equation, and it will be found that there is exactly one solution. Thus conditions (a) and (b) of Rule 3 are satisfied and we find that

$$\xi_1 = \Phi(\lambda)\left\{ x_1 + \sum_\alpha \frac{a_\alpha x_\alpha}{\lambda + a_\alpha} \right\},$$

and

$$\xi_\beta = \frac{\xi_1 + x_\beta}{\lambda + a_\beta},$$

where $\Phi(\lambda)$ is defined by

$$\Phi(\lambda)\left\{ 1 + \sum_\alpha \frac{1}{\lambda + a_\alpha} \right\} = \frac{1}{\lambda} \qquad (\lambda > 0). \tag{27}$$

The matrix representation of R_λ is thus

$$\left\| \begin{matrix} \Phi(\lambda) & \dots & \Phi(\lambda)/(\lambda + a_j) & \dots \\ \dots & \dots & \dots & \dots \\ a_i\Phi(\lambda)/(\lambda + a_i) & \dots & a_i\Phi(\lambda)/[(\lambda + a_i)(\lambda + a_j)] + \delta_{ij}/(\lambda + a_i) & \dots \\ \dots & \dots & \dots & \dots \end{matrix} \right\| \tag{28}$$

and it is now easy to verify the remaining conditions (c') and (e') of Rules 3 and 4.

The semigroup theory (see (18) of Rule 2) now tells us without further calculation that if $\varphi(t) \equiv p_{11}(t)$ then

$$\Phi(\lambda) = \int_0^\infty e^{-\lambda t}\varphi(t)dt, \qquad (\lambda > 0)$$

and so (from (27)) that

$$\varphi(t) + \int_0^t \varphi(u)k(t-u)du = 1 \qquad (t \geqq 0), \tag{29}$$

where

$$k(v) \equiv \sum_\alpha e^{-a_\alpha v} \qquad (v \geqq 0); \tag{30}$$

it further tells us that the integral equation (29) necessarily has a positive continuous bounded solution. (By Lerch's theorem there is at most one such solution). This shows very well an incidental advantage of the semigroup approach; in Kolmogorov's construction of K1, the integral equation (29) played an essential part, and this required a non-trivial existence theorem which was not established in all the necessary detail. A study of the integral equation (29) for a class of kernels $k(v)$ which includes (30) has now been made by one of us, and will be published elsewhere (Reuter [18]).

The Laplace transform $r_{ij}(\lambda)$ of $p_{ij}(.)$ is shown explicitly in (28). Thus we can obtain the matrix representation of the semigroup and the transition probabilities of the Markov process: if φ is the unique bounded continuous solution to (29), if $*$ denotes convolution, and if

$$\psi_\alpha(t) \equiv e^{-a_\alpha t} \qquad (t \geqq 0), \tag{31}$$

then

$$\begin{aligned} p_{11} &= \varphi \\ p_{1\beta} &= \varphi * \psi_\beta \\ p_{\alpha 1} &= a_\alpha \varphi * \psi_\alpha \\ p_{\alpha\beta} &= a_\alpha \varphi * \psi_\alpha * \psi_\beta + \delta_{\alpha\beta}\psi_\alpha. \end{aligned} \tag{32}$$

It is quite easy to check by direct differentiation that $(\boldsymbol{Q})_{ij}$ will have the value required by (21) except perhaps when $i = j = 1$. In this case,

$$\frac{1-p_{11}(t)}{t} = \frac{1-\varphi(t)}{t} = t^{-1}\int_0^t \varphi(u)k(t-u)du$$

$$\geqq \sum_{\varrho=2}^{N} t^{-1}e^{-a_\varrho t}\int_0^t \varphi(u)\,e^{a_\varrho u}\,du,$$

and so, since $\varphi(0) = 1$,

$$q_1 \geqq \sum_{\varrho=2}^{N} 1 = N-1 \qquad \text{(all } N \geqq 2\text{)},$$

whence $q_1 = \infty$. On referring to Kolmogorov's construction it will be found that his process K1 is identical with that defined by equations (32); this is *not* a trivial remark because for all we know there may exist more than one process associated with the matrix (21). When the functions $p_{ij}(.)$ are *defined* by the equations (32) it is possible but very laborious to verify (2) to (6) by direct calculation; such verifications are unnecessary for us because we know from the semigroup theory that the conditions (2) to (6) *must* be satisfied.

In the subsequent discussion we shall need some special properties of the function φ, defined as the unique bounded continous solution to (29); for the proof of these the reader is referred to Reuter [18].

(i) *The function φ is analytic for $0 < t < \infty$.* (33)

(ii) *φ is a strictly decreasing function and*

$$\lim_{t\to\infty}\varphi(t) = (1+\sum_{\alpha} 1/a_\alpha)^{-1}. \tag{34}$$

Both these follow from another result which may have some probabilistic interest:-

(iii) *φ is completely monotonic in $0 < t < \infty$ and has the representation*

$$\varphi(t) = \varrho_1 + \sum_{\alpha} \varrho_\alpha e^{-b_\alpha t} \qquad (t \geqq 0), \tag{35}$$

where

$$\varrho_i > 0, \quad \sum_i \varrho_i = 1, \quad \varrho_1 = \left(1+\sum_\alpha \frac{1}{a_\alpha}\right)^{-1};$$

$$0 < b_2 < b_3 < \ldots, \quad \sum_\alpha \frac{1}{b_\alpha} < \infty.$$

3. *Miscellaneous properties of K1.* A glance at the formulae (32) reveals the interesting fact that

$$p_{\alpha 1} = a_\alpha p_{1\alpha} \quad \text{and} \quad a_\beta p_{\alpha\beta} = a_\alpha p_{\beta\alpha}. \tag{36}$$

We shall make frequent use of these identities.

The result (34) concerning the limiting value of $\varphi(t) \equiv p_{11}(t)$ when

$t \to \infty$ can easily be extended to give the asymptotic behaviour of $p_{ij}(t)$ for all values of i and j. The formulae (32) show at once that

$$p_{i1}(t) \to \varphi(\infty) \text{ and } p_{i\beta}(t) \to \varphi(\infty)/a_\beta$$

when $t \to \infty$, where $\varphi(\infty)$ is given by (34). Let us write

$$\pi_1 \equiv \varphi(\infty) \text{ and } \pi_\beta \equiv \varphi(\infty)/a_\beta, \tag{37}$$

so that $\{\pi_j : j \geqq 1\}$ is a probability row-vector π with positive components. Then

$$\lim_{t\to\infty} p_{ij}(t) = \pi_j \qquad (i, j \geqq 1). \tag{38}$$

Some simple calculations based on (32) show that

$$P_t\pi = \pi, \ \pi \in \mathfrak{D}(\Omega) \text{ and } \Omega\pi = 0, \tag{39}$$

so that π represents a stationary distribution. Thus we can turn K1 into a stationary Markov process by postulating that

$$\boldsymbol{P}\{x_{t_1}(\omega) = i \text{ and } x_{t_2}(\omega) = j\} \equiv \pi_i p_{ij}(t_2 - t_1) \tag{40}$$

when $-\infty < t_1 \leqq t_2 < \infty$, and by defining the higher-order joint distributions in a similar way. The identities (36) can be written

$$\pi_i p_{ij} = \pi_j p_{ji},$$

and so they represent the curious fact that the process defined as at (40) is *identical with its reversal in time.*

We shall now establish two inequalities for the functions $p_{ij}(\cdot)$ associated with K1 which will be needed in § 4. They are:-

(iv) *If* $\alpha \geqq 2$ *then*

$$a_\alpha \mid p_{\alpha 1}(t) - p_{11}(t) \mid \leqq K_1(t) < \infty \tag{41}$$

for each $t > 0$.

(v) *If* $\alpha, \beta \geqq 2$ *then*

$$a_\alpha a_\beta \mid p_{\alpha\beta}(t) - p_{1\beta}(t) \mid \leqq K_2(t) < \infty \tag{42}$$

for each $t > 0$.

To prove (iv) we write

$$\begin{aligned} a_\alpha(p_{\alpha 1} - p_{11}) &= a_\alpha^2 \varphi * \psi_\alpha - a_\alpha \varphi \\ &= \int_0^t [\varphi(t-u) - \varphi(t)]\, a_\alpha^2 e^{-a_\alpha u}\, du - a_\alpha \varphi(t)\, e^{-a_\alpha t}, \end{aligned}$$

and split the range of integration into $(0, \frac{1}{2}t)$ and $(\frac{1}{2}t, t)$. If

$$A(t) \equiv \sup_{\frac{1}{2}t \leqq \tau \leqq t} |\varphi'(\tau)| \qquad (t > 0),$$

we then find that

$$a_\alpha \,|\, p_{\alpha 1} - p_{11} \,| \leqq A(t) \int_0^{a_\alpha t/2} x e^{-x}\, dx + a_\alpha e^{-a_\alpha t/2} + a_\alpha e^{-a_\alpha t}$$

$$\leqq A(t) + \frac{3}{t} \sup_{x \geqq 0} (x e^{-x}) \equiv K_1(t) < \infty.$$

This proves (iv). It is to be noted that $K_1(t)$ is bounded in any finite interval which excludes $t = 0$. To prove (v) we write (using (36))

$$a_\alpha a_\beta (p_{\alpha\beta} - p_{1\beta}) = a_\alpha (a_\alpha p_{\beta\alpha} - p_{\beta 1})$$

$$= \delta_{\alpha\beta} a_\alpha^2 e^{-a_\alpha t} + \int_0^t [p_{\beta 1}(t-u) - p_{\beta 1}(t)]\, a_\alpha^2 e^{-a_\alpha u}\, du - a_\alpha p_{\beta 1}(t)\, e^{-a_\alpha t}.$$

We now observe (from (32)) that

$$\frac{d}{dt} p_{\beta 1}(t) = - a_\beta (p_{\beta 1} - p_{11})$$

for all $t \geqq 0$, so that

$$B(t) \equiv \sup_{\frac{1}{2}t \leqq \tau \leqq t} |\, p'_{\beta 1}(\tau) \,|$$

$$\leqq \sup_{\frac{1}{2}t \leqq \tau \leqq t} K_1(\tau) < \infty \qquad (t > 0).$$

We then find, exactly as before, that

$$a_\alpha a_\beta \,|\, p_{\alpha\beta} - p_{1\beta} \,| \leqq \frac{1}{t^2} \sup (x^2 e^{-x}) + B(t) + \frac{3}{t} \sup (x e^{-x})$$

and the expression on the right-hand side is to be identified with $K_2(t)$ in (v).

From (iv) and (v) it follows that, for each $j \geqq 1$ and for all $t > 0$,

$$p_{1j}(t) = \lim_{i \to \infty} p_{ij}(t). \tag{43}$$

4. *The adjoint semigroup and the "backward" and "forward" equations. Applications to K*1.

We may expect to gain useful information from a study of the "correct" forms (in so far as these exist) of the "backward" and "forward" differential equations for K1. As a preliminary to this discussion it will be necessary to say something about the *adjoint* semigroup. The importance of the adjoint semigroup was first made clear by Feller [5]; our treatment of this question is based on recent work by R. S. Phillips which has at the moment of writing only been published in summary form [17] [4]). We have, however, been able to verify all of his results which we shall need.

[4]) Professor Phillips informs us that a detailed account of this will shortly appear in the *Pacific Journal of Mathematics*.

Suppose we are given a semigroup $\{T_t : t \geqq 0\}$ of bounded linear operators, mapping a real Banach space X into itself, such that T_0 is the identity, $\| T_t \| \leqq M < \infty$ for all $t \geqq 0$ and $\| T_t x - x \| \to 0$ when $t \downarrow 0$ for each $x \in X$. Let Ω be the infinitesimal generator for this semigroup and let J_λ be the resolvent operator $(\lambda I - \Omega)^{-1}$ (defined for each $\lambda > 0$). (The semigroup constructed in § 1 from a quasi-process is of this type, with $T_t \equiv P_t$, $J_\lambda \equiv R_\lambda$ and $M = 1$.) Let X^* be the adjoint space of X, and let T_t^*, J_λ^* and Ω^* be the adjoints of T_t, J_λ and Ω, respectively. The operators T_t^* and J_λ^* are defined everywhere and are bounded, but Ω^* is defined only on a domain $\mathfrak{D}(\Omega^*)$ such that:-

$x^* \in \mathfrak{D}(\Omega^*)$ if and only if $\exists\, x_1^* \in X^*$ such that

$$(x^*, \Omega x) = (x_1^*, x)$$

for all $x \in \mathfrak{D}(\Omega)$.

If $x^* \in \mathfrak{D}(\Omega^*)$ then x_1^* is unique and is defined to be the value of $x^*\Omega^*$. Now $\mathfrak{D}(\Omega^*)$ may not be dense in X^*, but we shall have

$$\mathfrak{D}(\Omega^*) = \mathfrak{R}(J_\lambda^*) \qquad (\lambda > 0), \tag{44}$$

and the closure of $\mathfrak{D}(\Omega^*)$ in the strong topology of X^* is called $X^\oplus$, a subspace of X^*. Phillips shows that $X^\oplus$ is mapped into itself by each T_t^* $(t \geqq 0)$, and that if $T_t^\oplus$ is the operator obtained by contracting the domain of T_t^* from X^* to $X^\oplus$, then $\{T_t^\oplus : t \geqq 0\}$ is a semigroup of operators on $X^\oplus$ possessing all the properties postulated for the original semigroup $\{T_t : t \geqq 0\}$.

He also shows that the domain $X^\oplus$ of the new semigroup is maximal with respect to strong continuity at $t = 0$; in fact if $x^* \in X^*$ and if

$$\| x^* T_t^* - x^* \| \to 0 \text{ when } t \downarrow 0,$$

then $x^* \in X^\oplus$ (so that this provides another characterisation of $X^\oplus$). The resolvent operator $J_\lambda^\oplus$ for the new semigroup is obtained by contracting the domain of J_λ^* to $X^\oplus$, and the infinitesimal generator $\Omega^\oplus$ is obtained by contracting the domain of Ω^* so that its range lies in $X^\oplus$. That is,

$x^* \in \mathfrak{D}(\Omega^\oplus)$ if and only if both $x^* \in \mathfrak{D}(\Omega^*)$ and $x^*\Omega^* \in X^\oplus$.

The operation of passing from $\{T_t\}$ on X to $\{T_t^\oplus\}$ on $X^\oplus \subseteqq X^*$ can be repeated to give a sequence of adjoint semigroups on the spaces X, $X^\oplus$, $X^{\oplus\oplus}$, etc. Phillips proves that the original space X is canonically represented by an isomorph $\tilde{X}$ which is a closed subspace of $X^{\oplus *}$ such that

$$\tilde{X} \subseteqq X^{\oplus\oplus} \subseteqq X^{\oplus *}.$$

If $x \rightleftarrows \tilde{x}$ is the canonical $(1:1)$ mapping between X and $\tilde{X}$, then we have

$$\| x \| / M \leqq \| \tilde{x} \| \leqq \| x \|, \tag{45}$$

so that, if the original semigroup is derived from a process or a quasi-process, then the relationship between X and $\tilde{X}$ will be that of isometric equivalence. Phillips proves that the second adjoint semigroup on $X^{\oplus\oplus}$ is always an extension of the canonical image of the original semigroup. If $\tilde{X} = X^{\oplus\oplus}$, so that this extension is not a proper one, he calls the original semigroup *reflexive*. He shows that the semigroup will be reflexive if J_λ is (for any one $\lambda > 0$) a weakly compact operator, and in particular if X is reflexive. A necessary and sufficient condition for the reflexivity of the semigroup is the weak compactness of J_λ (for any one and then for all $\lambda > 0$) in the weak topology defined on X by the range of J_λ^*.

We turn now to a discussion of the differential equations. It is known that

(a) *if* $x \in \mathfrak{D}(\Omega)$, *then*

$$DT_t x = T_t \Omega x \qquad (t \geqq 0), \tag{46}$$

where D denotes strong differentiation with respect to t, and that

(b) *if* $T_t x \in \mathfrak{D}(\Omega)$ *for all positive* t, *then*

$$DT_t x = \Omega T_t x \qquad (t > 0). \tag{47}$$

From the theory of the adjoint semigroup we get at once

(c) *if* $x^* \in \mathfrak{D}(\Omega^\oplus)$, *then*

$$Dx^* T_t^\oplus = x^* \Omega^\oplus T_t^\oplus \qquad (t \geqq 0) \tag{48}$$

and now instead of (b) we obtain

(d) *if* $x^* T_t^* \in \mathfrak{D}(\Omega^\oplus)$ *for all positive* t, *then*

$$Dx^* T_t^* = x^* T_t^* \Omega^\oplus \qquad (t > 0). \tag{49}$$

It is of some importance that in (d) we do *not* have to assume that $x^* \in X^\oplus$.

When $X \equiv l$ and the semigroup is of the type described in § 1, we define u^i as in § 2 and we define $v^j \in X^* \equiv m$ by $(v^j)_i \equiv \delta_{ij}$. Any linear expression for p'_{ij} in terms of $p_{\cdot j}$ may be called a "backward" differential equation, and any linear expression for p'_{ij} in terms of $p_{i\cdot}$ a "forward" differential equation. If we adopt this terminology then we obtain two forms for a "backward" equation B_{ij} as follows:-

$$p'_{ij} = \sum_{k=1}^{\infty} (\Omega u^i)_k p_{kj} \qquad (t \geqq 0) \tag{50}$$

if $u^i \in \mathfrak{D}(\Omega)$, and

$$p'_{ij} = (p_{\cdot j} \Omega^\oplus)_i \qquad (t > 0) \tag{51}$$

if $v^j P_t^* \equiv p_{\cdot j} \in \mathfrak{D}(\Omega^\oplus)$ for all positive t.

Similarly we obtain "forward" equations F_{ij} as follows:-

$$p'_{ij} = (\Omega p_{i\cdot})_j \qquad (t > 0) \tag{52}$$

if $P_t u^i \equiv p_{i\cdot} \in \mathfrak{D}(\Omega)$ for all positive t, and

$$p'_{ij} = \sum_{k=1}^{\infty} p_{ik}(v^j \Omega^{\oplus})_k \qquad (t \geqq 0) \tag{53}$$

if $v^j \in \mathfrak{D}(\Omega^{\oplus})$. It may be remarked that all the derivatives in (46)—(53) are two-sided save at $t = 0$.

One might at first expect that other versions of the differential equations would come from the second adjoint semigroup, but this is not so; we get nothing new because $\tilde{X}$ is invariant under $T_t^{\oplus\oplus}$ and

$$\mathfrak{D}(\Omega^{\oplus\oplus}) \cap \tilde{X}$$

is the canonical image of $\mathfrak{D}(\Omega)$ in $X^{\oplus\oplus}$.

In order to apply these ideas to K1 we must first calculate the adjoint semigroups. Let $x^* \in X^* \equiv m$ and as usual let $\alpha \geqq 2$, so that $u^\alpha \in \mathfrak{D}(\Omega)$. Then

$$(x^*, \Omega u^\alpha) = a_\alpha(x_1^* - x_\alpha^*), \tag{54}$$

and so, if $x^* \in \mathfrak{D}(\Omega^*)$, the right-hand side of (54) must be bounded for all α. On the other hand, if this boundedness condition is satisfied and if $x \in \mathfrak{D}(\Omega)$, we shall have

$$\begin{aligned}(x^*, \Omega x) &= x_1^* \sum (a_\alpha x_\alpha - x_1) - \sum x_\alpha^*(a_\alpha x_\alpha - x_1) \\ &= \sum (x_1^* - x_\alpha^*)(a_\alpha x_\alpha - x_1) \\ &= x_1 \sum (x_\alpha^* - x_1^*) + \sum x_\alpha a_\alpha (x_1^* - x_\alpha^*) \\ &= (z^*, x)\end{aligned}$$

where

$$z_1^* \equiv \sum (x_\alpha^* - x_1^*) \tag{55}$$

and

$$z_\alpha^* \equiv a_\alpha(x_1^* - x_\alpha^*),$$

and $z^* \in m$. Thus $\mathfrak{D}(\Omega^*)$ *consists of those vectors* $x^* \in m$ *for which*

$$a_\alpha \mid x_1^* - x_\alpha^* \mid \leqq K < \infty, \tag{56}$$

and $x^*\Omega^* \equiv z^*$ *is given by the equations* (55).

It is now easily shown that $X^{\oplus} \equiv \overline{\mathfrak{D}(\Omega^*)}$ *is the subspace of vectors* x^* *such that*

$$x_\alpha^* \to x_1^* \text{ as } \alpha \to \infty. \tag{57}$$

The theory sketched above then shows that $\mathfrak{D}(\Omega^{\oplus})$ *consists of all* $y \in X^{\oplus}$ *such that*

$$\lim_{\alpha \to \infty} a_\alpha(y_\alpha - y_1) = -\sum (y_\alpha - y_1). \tag{58}$$

(The limit on the left-hand side is required to exist; this then implies the

absolute convergence of the series on the right-hand side, since $\sum_{\alpha} 1/a_\alpha < \infty$.) On inspection we find that for K1,

$$\left.\begin{aligned} u^1 &\notin \mathfrak{D}(\Omega), \qquad u^\alpha \in \mathfrak{D}(\Omega), \\ v^1 &\notin X^\oplus, \qquad v^\beta \in X^\oplus, \end{aligned}\right\}$$

and that

$$\left.v^j \notin \mathfrak{D}(\Omega^\oplus).\right\}$$

It follows at once that we can obtain the equations $B_{\alpha j}$ from (50); these will be found in the appropriate places in Table I. From (53) we can obtain nothing, but in a few moments we shall show that for K1,

$$P_t u^i \in \mathfrak{D}(\Omega) \qquad (t > 0), \tag{59}$$

and from this and (52) we can obtain a complete set of "forward" equations F_{ij}; these will also be found in Table I.

On making use of $B_{\alpha j}$ and F_{ij} and the identities (36) we shall show that

$$v^j P_t^* \in \mathfrak{D}(\Omega^\oplus) \qquad (t > 0), \tag{60}$$

and this in conjunction with (51) will then give us the missing equation B_{1j}.

First we must prove (59). Now $P_t u^i \equiv p_{i\cdot}(t) \in \mathfrak{D}(\Omega)$ for all positive t if and only if

$$\sum_{\alpha} | a_\alpha p_{1\alpha} - p_{11} | < \infty \tag{61}$$

and

$$\sum_{\alpha} | a_\alpha p_{\beta\alpha} - p_{\beta 1} | < \infty \tag{62}$$

for each β and all $t > 0$, and the truth of (61) and (62) follows easily from the identities

$$a_\alpha p_{1\alpha} - p_{11} = p_{\alpha 1} - p_{11}\,, \quad a_\alpha p_{\beta\alpha} - p_{\beta 1} = a_\beta(p_{\alpha\beta} - p_{1\beta})$$

(which follow from (36)) and the inequalities (41) and (42). Thus all the equations $B_{\alpha j}$ and F_{ij} of Table I have been established. We shall have to make use of these in order to show that $v^j P_t^* \equiv p_{\cdot j}(t) \in \mathfrak{D}(\Omega^\oplus)$ for each j and all positive t; this will then enable us to complete the set of differential equations by establishing B_{1j}. It will be convenient to treat the cases $j \equiv \beta \geqq 2$ and $j = 1$ separately. However, before dividing the argument we note that $v^j P_t^* \in X^\oplus$ for each j and all positive t; this statement which is equivalent to

$$\lim_{\alpha \to \infty} p_{\alpha j}(t) = p_{1j}(t) \qquad (j \geqq 1,\ t > 0),$$

was proved above at equation (43).

We shall have $p_{\cdot\beta} \in \mathfrak{D}(\Omega^\oplus)$ for $t > 0$ if and only if

$$\lim_{\alpha \to \infty} a_\alpha(p_{\alpha\beta} - p_{1\beta}) = -\sum_{\alpha} (p_{\alpha\beta} - p_{1\beta}) \qquad (t > 0). \tag{63}$$

We first transform (63) with the aid of the identities (36) to the equivalent relation,

$$\lim_{\alpha\to\infty} a_\alpha(a_\alpha p_{\beta\alpha} - p_{\beta 1}) = -\sum_\alpha (a_\alpha p_{\beta\alpha} - p_{\beta 1}) \qquad (t > 0), \tag{64}$$

and then we note that because of the equations $B_{\beta j}$ and $F_{\beta j}$ we can replace (64) by the equivalent relation

$$\lim_{\alpha\to\infty} a_\alpha(p_{\beta\alpha} - p_{1\alpha}) = -(p_{11} - p_{\beta 1}) \qquad (t > 0),$$

and this in turn is equivalent to

$$\lim_{\alpha\to\infty} \{a_\beta(p_{\alpha\beta} - p_{1\beta}) - (p_{\alpha 1} - p_{11})\} = 0 \qquad (t > 0) \tag{65}$$

(on using (36) again). But (65) is an immediate consequence of (41) and (42).

In the same way $p_{\cdot 1} \in \mathfrak{D}(\Omega^{\oplus})$ for all positive t if and only if

$$\lim_{\alpha\to\infty} a_\alpha(p_{\alpha 1} - p_{11}) = -\sum_\alpha (p_{\alpha 1} - p_{11}) \quad (t > 0).$$

Because of (36) this is equivalent to

$$\lim_{\alpha\to\infty} a_\alpha(a_\alpha p_{1\alpha} - p_{11}) = -\sum_\alpha (a_\alpha p_{1\alpha} - p_{11}) \quad (t > 0),$$

and because of F_{1j} this is equivalent to

$$\lim_{\alpha\to\infty} a_\alpha p'_{1\alpha} = p'_{11} \ (t > 0)$$

or (using (36) again)

$$p'_{\alpha 1}(t) \to p'_{11}(t) \quad (t > 0) \tag{66}$$

as $\alpha \to \infty$.

Now

$$p_{\alpha 1}(t) = a_\alpha e^{-a_\alpha t} \int_0^t \varphi(u) e^{a_\alpha u} du$$

and so

$$p'_{\alpha 1}(t) = -a_\alpha^2 e^{-a_\alpha t} \int_0^t \varphi(u) e^{a_\alpha u} du + a_\alpha \varphi(t)$$

$$= a_\alpha \int_{0+}^t \varphi'(u) e^{-a_\alpha(t-u)} du + a_\alpha e^{-a_\alpha t};$$

thus (66) is equivalent to

$$\lim_{\alpha\to\infty} \int_0^{t-} \varphi'(t-u) e^{-a_\alpha u} a_\alpha du = \varphi'(t) \quad (t > 0). \tag{67}$$

Let t and ε be fixed and positive and let a fixed number δ be chosen in the range $0 < \delta < t$ so that

$$|\varphi'(t-u) - \varphi'(t)| < \varepsilon \text{ if } 0 \leqq u \leqq \delta.$$

Then

$$\left| \int_\delta^{t-} \varphi'(t-u) e^{-a_\alpha u} a_\alpha du \right| \leqq a_\alpha e^{-a_\alpha \delta} \left[1 - \varphi(t-\delta) \right]$$

(because of the monotony of φ), and the expression on the right-hand side tends to zero when $\alpha \to \infty$. Also

$$\left| \int_0^\delta \left[\varphi'(t-u) - \varphi'(t) \right] e^{-a_\alpha u} a_\alpha du \right| \leqq \varepsilon$$

and

$$\int_0^\delta \varphi'(t) e^{-a_\alpha u} a_\alpha du \to \varphi'(t) \quad (\alpha \to \infty).$$

The relation (67) (and so the inclusion of $p_{\cdot 1}$ in $\mathfrak{D}(\Omega^\oplus)$) now follows on combining these results.

A complete set of "backward" and "forward" differential equations has now been obtained for K1; it is exhibited in Table I.

	$j = 1$	$j = \beta \geqq 2$
$i = 1$	$p'_{11} = \sum_\alpha (p_{\alpha 1} - p_{11})$ $p'_{11} = \sum_\beta (a_\beta p_{1\beta} - p_{11})$	$p'_{1\beta} = \sum_\alpha (p_{\alpha\beta} - p_{1\beta})$ $p'_{1\beta} = p_{11} - a_\beta p_{1\beta}$
$i = \alpha \geqq 2$	$p'_{\alpha 1} = a_\alpha (p_{11} - p_{\alpha 1})$ $p'_{\alpha 1} = \sum_\beta (a_\beta p_{\alpha\beta} - p_{\alpha 1})$	$p'_{\alpha\beta} = a_\alpha (p_{1\beta} - p_{\alpha\beta})$ $p'_{\alpha\beta} = p_{\alpha 1} - a_\beta p_{\alpha\beta}$

Table I. *The "backward" and "forward" differential equations for* K1. [In the cell (i, j) the first equation is B_{ij} and the second is F_{ij}. All equations hold for $t \geqq 0$ and all derivatives are two-sided save at $t = 0$.]

It will be observed that all eight equations give correct statements at $t = 0$, even where the derivation applies only for $t > 0$. The most interesting are those involving infinite series. In each case the series is absolutely convergent for $t > 0$, in virtue of (36), (41) and (42), and the forms are not wholly unexpected. Thus B_{11} and F_{11} should be compared with the forms,

$$p'_{11} \simeq -\infty p_{11} + \sum_\alpha p_{\alpha 1},$$

$$p'_{11} \simeq -\infty p_{11} + \sum_\beta a_\beta p_{1\beta},$$

which would have been suggested by (11). In fact, for this example, all the equations suggested by (11) prove to be correct in so far as they make sense.

5. *Kolmogorov's second example as a semigroup of operators.*

We now turn to a discussion of Kolmogorov's second example, K2; this is a system of function $p_{ij}(.)$ $(i, j \geqq 1)$ satisfying (2), (3) and (4) to (6) and having the following $\boldsymbol{Q}$-matrix:-

$$\begin{Vmatrix} -1 & 0 & 0 & 0 & 0 & . & . & . \\ 0 & 0 & 0 & 0 & 0 & & & \\ 0 & a_3 & -a_3 & 0 & 0 & & & \\ 0 & 0 & a_4 & -a_4 & 0 & & & \\ 0 & 0 & 0 & a_5 & -a_5 & & & \\ . & . & . & & & & & \end{Vmatrix} \tag{68}$$

Once again it is the first state which is anomalous, and this time

$$0 = \sum_{j \neq 1} q_{1j} < q_1 = 1 < \infty.$$

The second state is an *absorbing* one ($q_2 = 0$). Concerning the constants $\{a_\varrho : \varrho \geqq 3)\}$ we assume that *they are all positive and such that*

$$\sum_{\varrho=3}^{\infty} 1/a_\varrho < \infty. \tag{69}$$

Where appropriate we shall write

$$a_1 = 1, \ a_2 = 0. \tag{70}$$

Our construction of K2 will start from the following definitions of Ω and $\mathfrak{D}(\Omega)$.

(1) *A vector* $x \in l$ *is in* $\mathfrak{D}(\Omega)$ *if and only if both*

$$\text{(a)} \sum_{j=1}^{\infty} | a_{j+1}x_{j+1} - a_j x_j | < \infty \tag{71}$$

and

$$\text{(b)} \lim_{j \to \infty} a_j x_j = x_1. \tag{72}$$

(2) *If* $x \in \mathfrak{D}(\Omega)$ *then*

$$\begin{aligned} (\Omega x)_1 &\equiv -x_1, \\ (\Omega x)_2 &\equiv a_3 x_3, \\ (\Omega x)_\sigma &\equiv a_{\sigma+1}x_{\sigma+1} - a_\sigma x_\sigma \quad (\sigma \geqq 3). \end{aligned} \tag{73}$$

It will be noticed that Ω is a *contraction* of the matrix operator (68) (when this acts by right multiplication). Of the conditions at (1) the first is inevitable if Ωx is to be in l but the second is of a form which may be found surprising. Condition (1) (a) implies in any case that $\lim a_j x_j$ exists; thus the extra condition (1) (b) merely assigns a value to this limit and it can be replaced by

$$(1)(\text{b}'): \sum_{j=1}^{\infty} (a_{j+1}x_{j+1} - a_j x_j) = 0. \tag{74}$$

It is also to be noted that if x satisfies (1) (a) then $| x_j | \leqq K/a_j$ for $j \geqq 3$ and so necessarily $x \in l$.

The domain assigned to Ω is linear, and

$$u^1 \notin \mathfrak{D}(\Omega), \quad u^\alpha \in \mathfrak{D}(\Omega) \quad (\alpha \geqq 2) \tag{75}$$

and

$$[1, \ 0, \ 1/a_3, \ 1/a_4, \ \ldots \ ] \in \mathfrak{D}(\Omega);$$

thus $\mathfrak{D}(\Omega)$ is dense in l, and Ω maps $\mathfrak{D}(\Omega)$ linearly into l. We now have to show that for each $\lambda > 0$ and for each $x \in l$ there is exactly one solution in $\mathfrak{D}(\Omega)$ to the equation (16), which we can write as the following set of simultaneous equations:-

$$\begin{aligned} &(\lambda + 1)\xi_1 = x_1, \\ &\lambda\xi_2 - a_3\xi_3 = x_2, \\ &(\lambda + a_\sigma)\xi_\sigma - a_{\sigma+1}\xi_{\sigma+1} = x_\sigma \quad (\sigma \geqq 3). \end{aligned} \tag{76}$$

These equations determine ξ_1 uniquely and they associate a unique set of values of $\xi_3, \ \xi_4, \ \ldots$ with each choice of ξ_2. We first show that (76) cannot have more than one solution in $\mathfrak{D}(\Omega)$. For, if there are two solutions ξ' and ξ'' in $\mathfrak{D}(\Omega)$, then

$$\eta \equiv \xi' - \xi'' \in \mathfrak{D}(\Omega), \ \eta_1 = 0, \ \eta_3 = \lambda\eta_2/a_3$$

and

$$\eta_\sigma = \frac{\lambda\eta_2}{a_\sigma} \prod_{\tau=3}^{\sigma-1} \frac{\lambda + a_\tau}{a_\tau} \quad (\sigma \geqq 4).$$

Thus

$$a_\sigma\eta_\sigma = \lambda\eta_2 \prod_{\tau=3}^{\sigma-1} \left(1 + \frac{\lambda}{a_\tau}\right) \quad (\sigma \geqq 4)$$

and so (letting $\sigma \to \infty$)

$$0 = \eta_1 = \lambda\eta_2 \prod_{\tau=3}^{\infty} \left(1 + \frac{\lambda}{a_\tau}\right).$$

But the infinite product is convergent (because of (69)), and so $\eta_2 = 0$ and $\eta = 0$ and $\xi' = \xi''$.

We now show that there is a solution to (76) in $\mathfrak{D}(\Omega)$ by constructing it explicitly. We first write

$$F_2(\lambda) \equiv \frac{1}{(\lambda + 1)\lambda} \prod_{\tau=3}^{\infty} \frac{a_\tau}{\lambda + a_\tau}, \tag{77}$$

$$F_\sigma(\lambda) \equiv \frac{1}{(\lambda + 1)(\lambda + a_\sigma)} \prod_{\tau=\sigma+1}^{\infty} \frac{a_\tau}{\lambda + a_\tau} \quad (\sigma \geqq 3), \tag{78}$$

and note that

$$F_\sigma(\lambda)/F_2(\lambda) = \frac{\lambda}{a_\sigma} \prod_{\tau=3}^{\sigma-1} \left(1 + \frac{\lambda}{a_\tau}\right) \quad (\sigma \geqq 3) \tag{79}$$

and that

$$\sum_{\alpha=2}^{\sigma} F_\alpha(\lambda)/F_2(\lambda) = \prod_{\tau=3}^{\sigma}\left(1+\frac{\lambda}{a_\tau}\right) < \prod_{\tau=3}^{\infty}\left(1+\frac{\lambda}{a_\tau}\right) < \infty \quad (\sigma \geqq 3). \tag{80}$$

We now put

$$\xi_2 = x_1 F_2(\lambda) + \frac{1}{\lambda}\sum_{m=2}^{\infty} x_m \prod_{2<\tau\leqq m} \frac{a_\tau}{\lambda + a_\tau}$$

(this series is absolutely convergent) and the equations (76) then give

$$\xi_1 = \frac{x_1}{\lambda + 1}$$

and

$$\xi_\sigma = x_1 F_\sigma(\lambda) + \frac{1}{\lambda + a_\sigma}\sum_{m=\sigma}^{\infty} x_m \prod_{\sigma<\tau\leqq m} \frac{a_\tau}{\lambda + a_\tau} \tag{81}$$

when $\sigma \geqq 3$ (these series also are absolutely convergent). We must prove that this vector ξ is in $\mathfrak{D}(\Omega)$.

Because the equations (76) are satisfied we have

$$a_{j+1}\xi_{j+1} - a_j\xi_j = \lambda\xi_j - x_j$$

for all $j \geqq 1$. Thus ξ will be in $\mathfrak{D}(\Omega)$ if and only if (i) it is in l and (ii) $\sum \xi_j = \sum x_j/\lambda$. Now when all the x's are non-negative the same will be true of the ξ's and we find after some computation that then

$$\sum_{\beta=2}^{\infty} \xi_\beta = \frac{1}{\lambda}\left(\frac{x_1}{\lambda+1} + \sum_{\alpha=2}^{\infty} x_\alpha\right).$$

If we now add the initial term $\xi_1 = x_1/(\lambda+1)$ it follows that $\sum\xi_j = \sum x_j/\lambda$ whenever the vector x has non-negative components. In general let $x = x^+ - x^-$ be the canonical decomposition of x, so that $[\,|x_1|, |x_2|, \ldots] = x^+ + x^-$. Then if we write $R_\lambda x$ for ξ, it is clear that $\sum |\xi_j|$ will be dominated by $\sum(R_\lambda(x^+ + x^-))_j$ and this is finite (by the argument just concluded). Thus ξ is in l, and

$$\sum \xi_j = \sum (R_\lambda(x^+ - x^-))_j = \sum (R_\lambda(x^+))_j - \sum (R_\lambda(x^-))_j$$

$$= \frac{1}{\lambda}(\sum x^+{}_j - \sum x^-{}_j) = \sum x_j/\lambda,$$

as required.

The calculation of the last paragraph also show that $R_\lambda x \geqq 0$ when $x \geqq 0$, that

$$\| R_\lambda x \| \leqq \| x \|/\lambda$$

and that equality holds in this last relation whenever $x \geqq 0$. Thus, by Rules

3 and 4 of § 1, Ω on $\mathfrak{D}(\Omega)$ is the infinitesimal generator of a *transition* semigroup and the Markov *process* associated with this semigroup is (by definition) K2.

Exactly as in § 2 the semigroup theory (together with Lerch's theorem) now tells us without further calculation that, for the K2 process,

$$
\begin{aligned}
&p_{11}(t) = e^{-t},\\
&p_{i1}(t) = 0, \quad (i \geqq 2),\\
&p_{22}(t) = 1,\\
&p_{2j}(t) = 0 \quad (j \geqq 3),\\
&p_{ij}(t) = 0 \quad (3 \leqq i < j),\\
&p_{ij}(t) = e^{-a_j t} * \prod_{j<\tau\leqq i}{}^{*} \{a_\tau e^{-a_\tau t}\} \quad (2 \leqq j \leqq i),
\end{aligned}
$$

(the formula for p_{22} being a special case of this) and

$$
p_{1j}(t) = e^{-t} * e^{-a_j t} * \prod_{\tau>j}{}^{*} \{a_\tau e^{-a_\tau t}\} \quad (j \geqq 2) \tag{82}
$$

The stars in the above formulae denote convolutions, $\prod^{*}$ denotes a multiple convolution of all the "factors" named, and the infinite convolution in (82),

$$
\prod_{\tau>j} {}^{*} \{a_\tau e^{-a_\tau t}\}, \tag{83}
$$

is best defined as the unique bounded continuous function with Laplace transform

$$
\prod_{\tau>j} \frac{a_\tau}{\lambda + a_\tau}. \tag{84}
$$

(We need not appeal here to the theory of infinite convolutions because the semigroup theory assures us of the existence of a unique bounded continuous function having as its Laplace transform

$$
a_3 F_3(\lambda) \equiv \frac{1}{\lambda + 1} \prod_{\tau=3}^{\infty} \frac{a_\tau}{\lambda + a_\tau}
$$

provided that each a_τ is positive and that $\sum 1/a_\tau < \infty$. By an obvious change of variable we can conclude from this that

$$
\prod_{n=1}^{\infty} \frac{b_n}{\lambda + b_n}
$$

is the Laplace transform of a unique bounded continuous function whenever b_n is positive for all $n \geqq 1$ and $\sum 1/b_n < \infty$.)

These formulae are identical with those of Kolmogorov except that in place of (82) he put [5])

$$
p_{1j}(t) \equiv \int_0^t \lim_{\alpha\to\infty} p_{\alpha j}(u) e^{-(t-u)} du \quad (j \geqq 2) \tag{85}
$$

[5]) As printed, Kolmogorov's formula differs from (85) by an obvious misprint.

Equation (85) was part of Kolmogorov's *definition* of K2; for us it is a *consequence* of (71) to (73), and we shall deduce it from the semigroup theory in a moment. (For this, see the argument at (92) below.)

First we must show that K2 as we have defined it actually does have (68) as its $\boldsymbol{Q}$-matrix. The verification is elementary save for the proof that

$$q_{1j} = 0 \quad (j \geqq 2), \tag{86}$$

and we give this in detail. We have only to put

$$\theta_j(t) \equiv e^{-a_j t} * \prod_{\tau>j}^{*}\{a_\tau e^{-a_\tau t}\},$$

so that

$$p_{1j}(t) = \int_0^t \theta_j(u) e^{-(t-u)} du$$

$$= e^{-t} \int_0^t \theta_j(u) e^u du,$$

and then it will be seen that

$$q_{1j} = p'_{1j}(0+) = \theta_j(0) = 0,$$

as required.

Next we construct *the semigroup adjoint to* K2. It has already been pointed out that $u^i \in \mathfrak{D}(\Omega)$ for $i \geqq 2$, and so, if $x^* \in \mathfrak{D}(\Omega^*)$, we shall have

$$|(x^*, \Omega u^i)| = |(x^*\Omega^*, u^i)| \leqq ||x^*\Omega^*|| \cdot ||u^i||$$

$$\leqq ||x^*\Omega^*||,$$

and so

$$a_i |x^*_{i-1} - x^*_i| \leqq K(x^*) < \infty \quad (i \geqq 2). \tag{87}$$

We shall show that this is in fact *the necessary and sufficient condition for* x^* *to be in* $\mathfrak{D}(\Omega^*)$. For let (87) be satisfied, and let x be any element of $\mathfrak{D}(\Omega)$. Then

$$(x^*, \Omega x) = \sum_{i=1}^{\infty} x_i^*(a_{i+1}x_{i+1} - a_i x_i)$$

$$= \lim_{N\to\infty} \sum_{i=1}^{N} x_i^*(a_{i+1}x_{i+1} - a_i x_i)$$

$$= \lim_{N\to\infty} \left\{ \sum_{i=2}^{N} a_i x_i (x^*_{i-1} - x^*_i) - x_1 x_1^* + a_{N+1}x_{N+1}x_N^* \right\}.$$

Now by hypothesis $a_{N+1}x_{N+1} \to x_1$, and from (87) $\sum(x^*_N - x^*_{N-1})$ is absolutely convergent, so that

$$\lim_{N\to\infty} x_N^*$$

exists. Thus

$$(x^*,\ \Omega x) = \sum_{i=2}^{\infty} a_i x_i (x^*_{i-1} - x^*_i) + x_1(\lim_{N\to\infty} x^*_N - x^*_1).$$

Also we can write

$$\lim_{N\to\infty} x^*_N - x^*_1 = \sum_{i=2}^{\infty} (x^*_i - x^*_{i-1}),$$

and so

$$(x^*, \Omega x) = x_1 \sum_{i=2}^{\infty} (x^*_i - x^*_{i-1}) + \sum_{i=2}^{\infty} a_i x_i (x^*_{i-1} - x^*_i) = (z,\ x),$$

where

$$z_1 \equiv \sum_{i=2}^{\infty} (x^*_i - x^*_{i-1}) = \lim_{N\to\infty} x^*_N - x^*_1,$$

and

$$z_i \equiv a_i(x^*_{i-1} - x^*_i) \quad (i \geqq 2). \tag{88}$$

The required conclusion follows from the fact that $z \in m$, and *the formulae* (88) *define* $z \equiv x^*\Omega^*$ *when* $x^* \in \mathfrak{D}(\Omega^*)$.

We have already noticed that $\lim_{i\to\infty} x^*_i$ exists if $x^* \in \mathfrak{D}(\Omega^*)$, so that $\mathfrak{D}(\Omega^*) \subseteqq c$, and *we shall now show that*

$$X^{\oplus} \equiv \overline{\mathfrak{D}(\Omega^*)} = c. \tag{89}$$

For if x^*_i is constant for $i \geqq i_0$ then $x^* \in \mathfrak{D}(\Omega^*)$, and such vectors are dense in c. It follows that $\mathfrak{D}(\Omega^{\oplus})$ *consists of all vectors* $x^*(\in m)$ *such that*

$$\lim_{i\to\infty} a_i(x^*_{i-1} - x^*_i) \tag{90}$$

exists. The operator $\Omega^{\oplus}$ is defined for $x^* \in \mathfrak{D}(\Omega^{\oplus})$ by $z = x^*\Omega^{\oplus}$ where z is given by (88).

We can now determine which of the coordinate vectors lie in $\mathfrak{D}(\Omega)$ or $\mathfrak{D}(\Omega^{\oplus})$; we find that

$$\left.\begin{aligned} u^1 \notin \mathfrak{D}(\Omega),\ u^\alpha \in \mathfrak{D}(\Omega) \text{ for } \alpha \geqq 2;\\ v^j \in \mathfrak{D}(\Omega^{\oplus}) \text{ for } j \geqq 1. \end{aligned}\right\} \tag{91}$$

Thus a complete set of "backward" and "forward" equations can be constructed from (51) and (53) in conjunction with (91). These equations are set out in Table II. It will be observed that the "backward" equation B_{1j} differs by a term

$$\lim_{\alpha\to\infty} p_{\alpha j}$$

from the "classical" equation

$$p'_{1j} = -p_{1j}$$

which would be obtained from (11). This "classical" equation is false for $j \geqq 2$, since it would then imply that $p_{1j}(t) \equiv 0$. From the semigroup standpoint, this breakdown is associated with the fact $u^1 \notin \mathfrak{D}(\Omega)$.

	$j = 1$	$j \geqq 2$
$i = 1$	$p'_{11} = \lim_{\alpha\to\infty} p_{\alpha 1} - p_{11}$ $p'_{11} = -p_{11}$	$p'_{1j} = \lim_{\alpha\to\infty} p_{\alpha j} - p_{1j}$ $p'_{1j} = a_{j+1}p_{1,j+1} - a_j p_{1j}$
$i \geqq 2$	$p'_{i1} = a_i(p_{i-1,1} - p_{i,1})$ $p'_{i1} = -p_{i1}$	$p'_{i,j} = a_i(p_{i-1,j} - p_{i,j})$ $p'_{ij} = a_{j+1}p_{i,j+1} - a_j p_{ij}$

Table II. *The "backward" and "forward" differential equations for* K2. [In the cell (i, j) the first equation is B_{ij} and the second is F_{ij}. All equations hold for $t \geqq 0$ and all derivatives are two-sided save at $t = 0$].

By comparing B_{1j} with F_{1j} we obtain the identity

$$\lim_{\alpha\to\infty} p_{\alpha j} = (a_{j+1}p_{1,j+1} - a_j p_{1j}) + p_{1j} \quad (j \geqq 2), \tag{92}$$

and formulae (82) to (84) then show that $\lim_{\alpha\to\infty} p_{\alpha j}$ has the Laplace transform

$$\frac{1}{\lambda + a_j} \prod_{\tau > j} \frac{a_\tau}{\lambda + a_\tau},$$

so that, using Lerch's theorem again,

$$p_{1j}(t) = \int_0^t \lim_{\alpha\to\infty} p_{\alpha j}(u) e^{-(t-u)} du \quad (j \geqq 2).$$

This completes the identification of our process K2 with Kolmogorov's process satisfying (85).

The results listed at (91) leave open the question: does $P_t u^1 \in \mathfrak{D}(\Omega)$ for $t > 0$? The answer is, *yes*, and although we do not think that the mapping of the coordinate-vectors into $\mathfrak{D}(\Omega)$ by the operators of the semigroup (for $t > 0$) is likely to be a universal phenomenon, the connexion with the Kolmogorov differential equations is sufficiently interesting to make the proof of this fact for K2 worth recording. We have $P_t u^1 = p_{1\cdot}(t)$ and so we must prove (i) and (ii) below:

$$\text{(i)} \sum_{j=1}^{\infty} | a_{j+1}p_{1,j+1} - a_j p_{1j} | < \infty \quad (t > 0);$$

$$\text{(ii)} \sum_{j=1}^{\infty} (a_{j+1}p_{1,j+1} - a_j p_{1j}) = 0 \quad (t > 0).$$

It should be noticed that if (i) and (ii) hold for a particular value of t, say $t = t_1 > 0$, then $P_{t_1}u^1 \in \mathfrak{D}(\Omega)$. Now from the differential equations B_{1j} and F_{1j} we know that

$$a_{j+1}p_{1,j+1} - a_jp_{1j} = p'_{1j} = \lim_{\alpha\to\infty} p_{\alpha j} - p_{1j} \quad (j \geqq 1), \tag{93}$$

and so (i) is satisfied and indeed

$$\sum_{j=1}^{\infty} | a_{j+1}p_{1,j+1} - a_jp_{1j} | = \sum_{j=1}^{\infty} | p'_{1j} | \leqq 2 \tag{94}$$

for all $t \geqq 0$, because $\sum p_{1j} = 1$ and

$$\sum_{j=1}^{\infty} \lim_{\alpha\to\infty} p_{\alpha j} \leqq \liminf_{\alpha\to\infty} \sum_{j=1}^{\infty} p_{\alpha j} = 1.$$

Also (93) shows that the derivatives p'_{1j} are bounded and continuous for $j \geqq 1$ and $t \geqq 0$; thus, if $\lambda > 0$,

$$\int_0^{\infty} e^{-\lambda t}p'_{1j}(t)dt = -\delta_{1j} + \lambda \int_0^{\infty} e^{-\lambda t}p_{1j}(t)dt.$$

Now the expression on the right-hand side is equal to $-1/(\lambda + 1)$ when $j = 1$ and to $\lambda F_j(\lambda)$ when $j \geqq 2$. Thus, using (94) and (80), we get

$$\int_0^{\infty} e^{-\lambda t} \sum_{j=1}^{\infty} p'_{1j}(t)dt = -\frac{1}{\lambda + 1} + \sum_{j=2}^{\infty} \lambda F_j(\lambda)$$

$$= -\frac{1}{\lambda + 1} + \lambda F_2(\lambda) \prod_{\tau=3}^{\infty} \left(1 + \frac{\lambda}{a_\tau}\right) = 0.$$

It follows from Lerch's theorem that

$$\sum_{j=1}^{\infty} (a_{j+1}p_{1,j+1} - a_jp_{1j}) = \sum_{j=1}^{\infty} p'_{1j}(t) = 0$$

for almost all $t \geqq 0$, and so there exists a sequence $\{t_n : n \geqq 1\}$ such that $t_n \downarrow 0$ when $n \to \infty$ and $P_tu^1 \in \mathfrak{D}(\Omega)$ whenever $t = t_n$ for some $n \geqq 1$. But then (by a simple semigroup argument) $P_tu^1 \in \mathfrak{D}(\Omega)$ whenever $t \geqq t_n$, and so for all $t > 0$.

6. *The reflexivity of the semigroups.* We have already mentioned Phillips' concept of a reflexive semigroup and his theorem that the weak compactness of the resolvent operator J_λ (for any one $\lambda > 0$) is a sufficient condition for reflexivity. These ideas are of analytic but not perhaps of probabilistic interest, and the reader who is only interested in the Markov processes may prefer to omit this section of the paper.

Each of the two semigroups K1 and K2 provides an interesting example of a reflexive semigroup acting on a non-reflexive space, l. It is an easy and

instructive exercise to verify this directly. One finds, for example, that $\tilde{X} = X^{\oplus\oplus} = X^{\oplus *}$ for the K1 semigroup, and that $\tilde{X} = X^{\oplus\oplus} \subset X^{\oplus *}$ for K2. Here however we shall establish the reflexivity by a compactness argument, since Phillips' sufficient condition becomes particularly easy to apply when $X \equiv l$. If T is a bounded linear operator mapping a Banach space X into itself, and S is the strongly closed unit sphere in X, then T is called strongly (weakly) compact if TS is relatively compact in the strong (weak) topology of X. Now relative compactness and relative sequential compactness are equivalent properties for subsets of X, in both the strong and weak topologies. Moreover if $X \equiv l$, the equivalence of strong and weak convergence implies the equivalence of strong and weak relative sequential compactness. Thus *an operator T on l is weakly compact if and only if it is strongly (sequentially) compact.*

Now the (strongly) compact operators on l have been completely characterised by Cohen and Dunford [1]. If the operator T on l is defined by

$$(T x)_j \equiv \sum_{i=1}^{\infty} x_i t_{ij} \tag{95}$$

then their theorem asserts that T *is (strongly) compact if and only if*

$$\sup_{i \geqq 1} \sum_{j=n}^{\infty} | t_{ij} | \to 0 \quad (n \to \infty). \tag{96}$$

Accordingly, (96) applied to the operator defined by

$$t_{ij} \equiv \int_0^{\infty} e^{-\lambda t} p_{ij}(t) dt,$$

for any one $\lambda > 0$, is a *sufficient* condition for the semigroup associated with the process or quasi-process $\{p_{ij} : i, j \geqq 1\}$ to be reflexive.

Now for K1 we have

$$\sum_{j=n}^{\infty} | t_{1j} | = \Phi(\lambda) \sum_{j=n}^{\infty} \frac{1}{\lambda + a_j}, \quad (n \geqq 2)$$

and

$$\sum_{j=n}^{\infty} | t_{ij} | \leqq \frac{a_i}{\lambda + a_i} \Phi(\lambda) \sum_{j=n}^{\infty} \frac{1}{\lambda + a_j} + \frac{1}{\lambda + a_i}$$

if $n \geqq 2$, $i \geqq 2$ (the last term being absent if $n > i$). Thus

$$\sup_{i \geqq 1} \sum_{j=n}^{\infty} | t_{ij} | \leqq \Phi(\lambda) \sum_{j=n}^{\infty} \frac{1}{a_j} + \sup_{j \geqq n} \frac{1}{a_j},$$

$$\to 0 \text{ as } n \to \infty.$$

Similarly for K2 we find that

$$\sum_{j=n}^{\infty} | t_{1j} | = \sum_{j=n}^{\infty} F_j(\lambda) \quad (n \geqq 2),$$

and

$$\sum_{j=n}^{\infty} | t_{ij} | \leqq \sum_{j=n}^{\infty} \frac{1}{\lambda + a_j} \quad (i \geqq 2,\ n \geqq 2).$$

Thus

$$\sup_{i \geqq 1} \sum_{j=n}^{\infty} | t_{ij} | \leqq \sum_{j=n}^{\infty} \left(F_j(\lambda) + \frac{1}{a_j} \right),$$

$$\to 0 \text{ as } n \to \infty,$$

(because $\sum F_j(\lambda) < \infty$). It follows that *both semigroups are reflexive.*

7. *A representation of K1.* In the preceding sections of this paper we have accumulated a great deal of information (for each of K1 and K2) concerning both the array $\{p_{ij} : i, j \geqq 1\}$ of transition-functions and the associated semigroup $\{P_t : t \geqq 0\}$ of operators on l, but none of this is of the slightest assistance when we wish to form a mental picture of the kind of stochastic motion which is being described. Indeed it is plain that the methods of linear analysis have done all that they can for us, and that we must now enter into the realm of measure theory. Two methods of attack are open to us. In the first place we could use the array $\{p_{ij} : i, j \geqq 1\}$ of transition-functions to construct a representation of the Markov process by defining the appropriate measure on function space in the manner indicated by Doob [3] (page 89), and we could then attempt in the spirit of Doob's book to obtain a qualitative characterisation of the sample-functions valid save in a set of measure zero. The complete implementation of this programme for K1 would, we think, be exceedingly difficult, although it might be possible to carry it through for K2. The method we prefer, however, proceeds in the opposite direction. We shall start by constructing a very simple probability-space [6]) $(\Omega, \mathscr{F}, \boldsymbol{P})$ (which will be the direct product of countably many copies of the real line, each copy carrying a certain negative-exponential probability measure), and the stochastic process $\{x_t : t \geqq 0\}$ will then be introduced by giving for each $t \geqq 0$ an explicit definition of the $\mathscr{F}$-measurable integer-valued functions $x(t,.)$ over Ω. In order to identify the process as a representation of K1 (or of K2) we would then have to prove (i) that the constructed process has the correct transition-functions and (ii) that it is a Markov process.

[6]) We follow Doob [3] in writing Ω for the fundamental probability-space, ω for a typical point of Ω, $\mathscr{F}$ for the Borel field of measurable subsets of Ω and $\mathbf{P}(E)$ for the probability-measure of a set $E \in \mathscr{F}$. There will be no risk of confusing the *space* Ω with the *operator* Ω of the first part of this paper.

We shall prove (i) by showing that

$$\boldsymbol{P}\{x(t, \omega) = j\} = p_{ij}(t) \quad (t \geqq 0,\ j \geqq 1), \tag{97}$$

and that

$$\boldsymbol{P}\{x(t_1, \omega) = j_1 \text{ and } x(t_2, \omega) = j_2\} = p_{ij_1}(t_1)p_{j_1j_2}(t_2 - t_1); \tag{98}$$

it will then follow that

$$\boldsymbol{P}\{x(t_2, \omega) = j_2 \mid x(t_1, \omega) = j_1\} = p_{j_1j_2}(t_2 - t_1). \tag{99}$$

Here $0 \leqq t_1 \leqq t_2$; $j_1, j_2 \geqq 1$; i is the initial state and *the functions p are those appropriate to K1 (or to K2).*

To prove (ii) would then be equivalent to showing that, for each $n \geqq 1$,

$$\boldsymbol{P}\{x(t_r, \omega) = j_r \text{ for } 1 \leqq r \leqq n\} = p_{ij_1}(t_1) \prod_{r=1}^{n-1} p_{j_r,j_{r+1}}(t_{r+1} - t_r), \tag{100}$$

for $0 \leqq t_1 \leqq t_2 \leqq \ldots \leqq t_n$ and $j_r \geqq 1$ $(1 \leqq r \leqq n)$; for $n = 1$, $n = 2$, this reduces to (97) and (98), respectively.

From (97) to (99) it will follow that the constructed process is a representation of K1 (or of K2) *in the sense that* each *pair* of random variables x_{t_1} and x_{t_2} has the desired joint distribution.

We shall *not* attempt to write out a proof of (100) for a general value of n, partly because to do so would be excessively laborious and partly because the desired result (ii) (the Markovian character of the constructed process) is on the fringe of what is intuitively obvious and may even be regarded as self-evident by some readers after the construction has been presented. We do not accept this view ourselves, but we consider that the effort required to elaborate a direct proof of (100) for a general value of n would be better employed in an attempt to envisage the most general construction of the present kind, for it may be that the most natural method of establishing the Markovian character of such a process will reveal itself more readily when one is not obsessed with the particularity of a special case. At all events it is clear that some way round this sort of difficulty will have to be found if the more complicated pathologies recently discovered by Lévy (in the papers cited below) are to find their proper place in the measure-theoretic formulation of the subject.

It will perhaps have struck the reader that the programme just outlined for the construction of processes representing K1 and K2 presupposes that one can guess successfully both the appropriate fundamental probability-spaces and the measurable functions which are to be defined upon them. This would hardly have been possible without the help provided by two recent papers of Lévy [15, 16] to which we are heavily indebted. A study of these

made it seem very plausible that the constructions outlined by Lévy to illustrate his processes of the fifth and third types must when suitably modified lead directly to K1 and K2, respectively, and this opinion proved to be correct.

We turn now to the construction of the representation of K1. Let $\{a_\alpha : \alpha \geqq 2\}$ be a set of real numbers such that

$$a_\alpha > 0 \quad (\alpha \geqq 2) \text{ and } \sum_{\alpha=2}^{\infty} 1/a_\alpha < \infty. \tag{101}$$

As in § 2 we shall only use α, β and γ as indices or suffices if the allowed range of values is 2, 3, 4, ... We label a countable family of copies of the *positive* real line $0 < u < \infty$ (carrying the Borel sets) with the labels

$$\left.\begin{array}{ll} & (\alpha), \\ & (\alpha,\ r) \\ \text{and} & [\alpha,\ r], \end{array}\right\} \quad (\alpha = 2, 3, 4, \ldots;\ r = 1, 2, 3, \ldots)$$

and we endow these real lines with negative-exponential probability-measures as shown below:-

on (α) : $e^{-a_\alpha u} a_\alpha du$;

on $(\alpha,\ r)$: $e^{-a_\alpha u} a_\alpha du$;

on $[\alpha,\ r]$: $e^{-u} du$.

We then form the cartesian product of all these probability-spaces; the resulting probability-space will be $(\Omega, \mathscr{F}, \boldsymbol{P})$ and the coordinate-variables will be labelled as follows:-

on (α): v^α;

on $(\alpha,\ r)$: v_r^α;

on $[\alpha,\ r]$: u_r^α.

We now define the random variables

$$U_0^\alpha \equiv 0 \quad (\alpha \geqq 2)$$

and

$$U_r^\alpha \equiv u_1^\alpha + u_2^\alpha + \ldots + u_r^\alpha \quad (\alpha \geqq 2,\ r \geqq 1), \tag{102}$$

and we denote by $m(\beta \mid \alpha,\ r)$ the unique integer $m \geqq 0$ such that

$$U_m^\beta < U_r^\alpha \leqq U_{m+1}^\beta \quad (\beta \neq \alpha); \tag{103}$$

such an integer certainly exists save perhaps in the ω-set Z of $\boldsymbol{P}$-measure zero in which $\lim_{s\to\infty} U_s^\beta$ is finite for some β. We may now define, if $\omega \notin Z$,

$$T_r^\alpha \equiv U_r^\alpha + \sum_{s=1}^{r-1} v_s^\alpha + \sum_{\beta\neq\alpha} \sum_{s=1}^{m} v_s^\beta \quad (\alpha \geqq 2,\ r \geqq 1), \tag{104}$$

where $m \equiv m(\beta \mid \alpha, r)$, empty sums being interpreted as zero; if $\omega \in Z$, we put $T_r^\alpha \equiv \infty$.

We shall now show that

$$\boldsymbol{E}\{\exp(-\lambda T_r^\alpha)\} = \left(\frac{a_\alpha}{\lambda + a_\alpha}\right)^{r-1} \left[1 + \lambda + \sum_{\beta \neq \alpha} \frac{\lambda}{\lambda + a_\beta}\right]^{-r} \tag{105}$$

when $\lambda > 0$. To prove this we first integrate over all the v's (note that m does not depend on these); this reduces the expectand in (105) to

$$\exp(-\lambda U_r^\alpha)\left(\frac{a_\alpha}{\lambda + a_\alpha}\right)^{r-1} \prod_{\beta \neq \alpha} \left(\frac{a_\beta}{\lambda + a_\beta}\right)^{m(\beta|\alpha,r)}.$$

We then integrate over all the variables u_s^β with $\beta \neq \alpha$; during this integration, the $m(\beta \mid \alpha, r)$ act independently like Poisson variables of expectation U_r^α, and we obtain

$$\left(\frac{a_\alpha}{\lambda + a_\alpha}\right)^{r-1} \exp\{-\lambda U_r^\alpha - \sum_{\beta \neq \alpha} \frac{\lambda}{\lambda + a_\beta} U_r^\alpha\}.$$

Finally, integration over the variables $u_1^\alpha, u_2^\alpha, \ldots, u_r^\alpha$ leads to (105).

Now

$$\boldsymbol{P}\{T_r^\alpha < \infty\} = \lim_{\lambda \downarrow 0} \boldsymbol{E}\{\exp(-\lambda T_r^\alpha)\} = 1,$$

so that each T_r^α is finite with probability one. We therefore augment Z by adding to it the ω-set of $\boldsymbol{P}$-measure zero on which $T_r^\alpha = \infty$ for some pair (α, r). We then easily verify that, if $\omega \notin Z$,

$$0 < T_1^\alpha < T_2^\alpha < \ldots \quad (\alpha \geqq 2)$$

and

$$T_l^\gamma + v_l^\gamma < T_n^\delta \text{ whenever } U_l^\gamma < U_n^\delta;$$

the latter fact is most easily checked by rewriting (104) as

$$T_r^\alpha \equiv U_r^\alpha + \sum_{(\beta,s)} v_s^\beta, \tag{106}$$

the summation being over all pairs (β, s) such that $U_s^\beta < U_r^\alpha$. If we now augment Z still further by adding to it the set of $\boldsymbol{P}$-measure zero on which

$$U_l^\gamma = U_n^\delta \text{ for some } \gamma, \delta, l, n$$

with $\gamma, \delta \geqq 2$; $l, n \geqq 1$; and $(\gamma, l) \neq (\delta, n)$, then the half-open intervals

$$[T_l^\gamma, T_l^\gamma + v_l^\gamma) \text{ and } [T_n^\delta, T_n^\delta + v_n^\delta)$$

are disjoint when $\omega \notin Z$ and $(\gamma, l) \neq (\delta, n)$, lie in the same order on the real line as do the points U_l^γ and U_n^δ, and do not contain the real number zero. [7])

[7]) In the preceding work, we have taken the $\mathscr{F}$-measurability of the functions $m(\beta \mid \alpha, r)$ and T_r^α, and of the set Z, for granted. It is easily verified, and we shall continue to ignore similar trivial measurability questions in what follows.

We are now able to define a set of fundamental random functions $x_i(t,.)$ on Ω (where $t \geqq 0$ and $i \geqq 1$) as follows.

(i) If $\omega \in Z$, then $x_1(t, \omega) \equiv 1$.

(ii) If $\omega \notin Z$, then

$$x_1(t, \omega) \equiv \beta \quad (\beta \geqq 2)$$

whenever, for some integer $r \geqq 1$,

$$T_r^\beta \leqq t < T_r^\beta + v_r^\beta,$$

and

$$x_1(t, \omega) \equiv 1$$

for all other values of t.

(iii) If $0 \leqq t < v^\alpha$ then

$$x_\alpha(t, \omega) \equiv \alpha \quad (\alpha \geqq 2)$$

and if $t \geqq v^\alpha$ then

$$x_\alpha(t, \omega) \equiv x_1(t - v^\alpha, \omega).$$

With these definitions, $\{x_i(t,.): t \geqq 0\}$ *will be the representation of* K1 *when* i *is the initial state*; it will be noticed that, as a consequence of (i) — (iii),

$$x_i(0, \omega) = i \quad (\omega \in \Omega, \ i \geqq 1).$$

In attempting to visualise the stochastic motion defined by the random functions $\{x_i(t,.) : t \geqq 0\}$, one should observe that (with probability one):

v^α is the time spent in the initial state if $i \equiv \alpha \geqq 2$

and, if the initial state is 1, that:

T_r^α is the epoch at which the system begins its rth visit to the state α;

v_r^α is the duration of this rth visit;

u_r^α is the time spent in the state 1 between the $(r-1)$st and rth visits to the state α.

In fact, if $\omega \notin Z$, we have

$$T_r^\alpha - T_{r-1}^\alpha = u_r^\alpha + \sum_{(\beta,s)} v_s^\beta \tag{107}$$

where the summation runs over pairs (β, s) such that

$$T_{r-1}^\alpha \leqq T_s^\beta < T_r^\alpha \text{ (and we define } T_0^\alpha \equiv 0).$$

Equation (107) describes the character of the motion very conveniently.

We shall now prove (97) in the form

$$q_{ij}(t) \equiv P\{x_i(t, \omega) = j\} = p_{ij}(t),$$

where the $p_{ij}(t)$ are those defined at (32) for K1, and we take the five possible identifications of the pair (i, j) in the following order:

(i) $(i, j) = (1, \beta)$.
(ii) $(i, j) = (1, 1)$,
(iii) $(i, j) = (\alpha, \beta)$, with $\alpha \neq \beta$,
(iv) $(i, j) = (\alpha, \alpha)$,
(v) $(i, j) = (\alpha, 1)$.

(i) We have

$$q_{1\beta}(t) = \sum_{r=1}^{\infty} \boldsymbol{P}\{T_r^{\beta} \leqq t < T_r^{\beta} + v_r^{\beta}\}$$
$$= \sum_{r=1}^{\infty} \int_{T_r^{\beta} \leqq t} e^{-a_{\beta}(t - T_r^{\beta})} d\boldsymbol{P}$$
$$= \sum_{r=1}^{\infty} \int_0^t e^{-a_{\beta}(t-\tau)} dF_r^{\beta}(\tau)$$

where F_r^{β} is the distribution-function of the random variable T_r^{β}. We note that $q_{1\beta}$ is non-negative, bounded and Borel measurable as a function of t. Now if $\lambda > 0$,

$$Q_{1\beta}(\lambda) \equiv \int_0^{\infty} e^{-\lambda t} q_{1\beta}(t) dt$$
$$= \frac{1}{\lambda + a_{\beta}} \sum_{r=1}^{\infty} \boldsymbol{E}\{\exp(-\lambda T_r^{\beta})\}$$
$$= \Phi(\lambda)/(\lambda + a_{\beta}) = \int_0^{\infty} e^{-\lambda t} p_{1\beta}(t) dt,$$

in virtue of (105) and (32). Lerch's theorem now assures us that $q_{1\beta}(t) = p_{1\beta}(t)$ for almost all $t \geqq 0$. Moreover, we have

$$e^{a_{\beta} t} q_{1\beta}(t) = \sum_{r=1}^{\infty} \int_0^t e^{a_{\beta}\tau} dF_r^{\beta}(\tau)$$

so that

$e^{a_{\beta} t} q_{1\beta}(t)$ is weakly increasing for $t \geqq 0$,

$e^{a_{\beta} t} p_{1\beta}(t)$ is continuous for $t \geqq 0$

and

these two functions are equal for $t = 0$ and for almost all $t > 0$. A simple argument then shows that $q_{1\beta} \equiv p_{1\beta}$.

(ii) This follows from (i) and the fact that

$$\sum_{j=1}^{\infty} q_{1j}(t) = 1 = \sum_{j=1}^{\infty} p_{1j}(t).$$

(iii) We have

$$q_{\alpha\beta}(t) = \boldsymbol{P}\{t \geqq v^{\alpha} \text{ and } x_1(t - v^{\alpha}, \omega) = \beta\}$$
$$= \int_0^t q_{1\beta}(t - v) e^{-a_{\alpha} v} a_{\alpha} dv = a_{\alpha} p_{1\beta} * \psi_{\alpha}(t) = p_{\alpha\beta}(t),$$

from (32).

(iv) If $\alpha = \beta$ then to the above expression we must add a term

$$\boldsymbol{P}\{t < v^{\alpha}\} = e^{-a_{\alpha}t},$$

so

$$q_{\alpha\alpha}(t) = a_{\alpha}p_{1\alpha} * \psi_{\alpha}(t) + \psi_{\alpha}(t) = p_{\alpha\alpha}(t).$$

(v) This now follows as at (ii) above, and (97) has been established.

We shall now sketch a proof of (98). Evidently we may assume without real loss of generality that $0 < t_1 < t_2$. We first suppose that

$$i = 1,\ j_1 \equiv \alpha \geqq 2 \text{ and } j_2 \equiv \beta \geqq 2, \quad \alpha \neq \beta,$$

and in this case we must prove that

$$\sum_{r=1}^{\infty}\sum_{s=1}^{\infty} \boldsymbol{P}\{T_r^{\alpha} \leqq t_1 < T_r^{\alpha} + v_r^{\alpha} \text{ and } T_{m+s}^{\beta} \leqq t_2 < T_{m+s}^{\beta} + v_{m+s}^{\beta}\}$$
$$= p_{1\alpha}(t_1)p_{\alpha\beta}(t_2 - t_1), \tag{108}$$

where $m \equiv m(\beta \mid \alpha, r)$. In order to transform the probabilities occurring in the left-hand member of (108), we introduce the following notations (in which $r \geqq 1$ and $s \geqq 1$ are to be regarded as *fixed*);

$$\tilde{v}^{\alpha} \equiv [v_r^{\alpha} - (t_1 - T_r^{\alpha})^{+}]^{+},$$

where x^{+} denotes max $(x, 0)$;

$$\tilde{U}_s^{\beta} \equiv U_{m+s}^{\beta} - U_r^{\alpha},$$

where m denotes $m(\beta \mid \alpha, r)$;

$$\tilde{m}(\gamma \mid \beta, s) \equiv m(\gamma \mid \beta, m(\beta \mid \alpha, r) + s) - m(\gamma \mid \alpha, r),$$

where $\gamma \neq \alpha, \beta$; and

$$\tilde{m}(\alpha \mid \beta, s) \equiv m(\alpha \mid \beta, m(\beta \mid \alpha, r) + s) - r.$$

It will be noticed that all the m's and $\tilde{m}$'s are finite when $\omega \notin Z$. Lastly, we define

$$\tilde{v}_l^{\gamma} \equiv v_{m(\gamma \mid \alpha, r)+l}^{\gamma} \qquad (\gamma \neq \alpha),$$
$$\tilde{v}_l^{\alpha} \equiv v_{r+l}^{\alpha},$$

and

$$\tilde{T}_s^{\beta} \equiv \tilde{U}_s^{\beta} + \sum_{l=1}^{s-1} \tilde{v}_l^{\beta} + \sum_{\gamma \neq \beta} \sum_{l=1}^{\tilde{m}} \tilde{v}_l^{\gamma},$$

where $\tilde{m}$ denotes $\tilde{m}(\gamma \mid \beta, s)$.

The reader may now verify that the (r, s) — term in the double sum on the left-hand side of (108) may be written as

$$\boldsymbol{P}\{T_r^{\alpha} \leqq t_1 < T_r^{\alpha} + v_r^{\alpha} \text{ and } \tilde{v}^{\alpha} + \tilde{T}_s^{\beta} \leqq t_2 - t_1 < \tilde{v}^{\alpha} + \tilde{T}_s^{\beta} + \tilde{v}_s^{\beta}\}, \tag{109}$$

and that, apart from the explicit occurrence of v_r^{α} and $\tilde{v}^{\alpha}$, no v_l^{γ} enters *both*

the statements whose conjunction defines the set whose $\boldsymbol{P}$-measure has to be calculated in (109). Accordingly, in evaluating this $\boldsymbol{P}$-measure, we begin by integrating over v_r^α. Now if v is a negative-exponential random variable with distribution $e^{-\lambda v}\lambda dv$ $(\lambda > 0)$, if $\varphi(.)$ is a bounded Borel-measurable function and if $c \geqq 0$, then the identity,

$$\int_{v>c} \varphi([v - c]^+)e^{-\lambda v}\lambda dv = \int_c^\infty e^{-\lambda v}\lambda dv \int_0^\infty \varphi(v')e^{-\lambda v'}\lambda dv',$$

shows that the integral on the left can be evaluated by first taking an expectation with respect to $(v - c)^+$ *as if it were an independent random variable* v' *having the same distribution as* v, and then taking the expectation with respect to v itself. [8]) Thus, in evaluating (109), we may replace $\tilde{v}^\alpha$ by any variable independent of the other coordinate variables and distributed like v^α; since v^α does not otherwise enter into the present calculation, we may and shall replace $\tilde{v}^\alpha$ by v^α.

The v-variables are now all independent and have the indicated negative-exponential distributions. Now consider the successive steps in the evaluation of (109) by integrating over the coordinate variables. First integrate over all the v's; the effect of these integrations will be visible only in the "enumerating" variables

$$m(\gamma \mid \alpha,\ r) \quad (\gamma \neq \alpha), \text{ and } \tilde{m}(\gamma \mid \beta, s) \quad (\gamma \neq \beta)$$

(*cf.* the calculations following (105)). Next, integrate over all the u_n^δ with $\delta \neq \alpha, \beta$; during these integrations, in which the u^α's and u^β's are fixed, the above m's and $\tilde{m}$'s (for $\gamma \neq \alpha, \beta$) act independently as Poisson variables of respective mean values U_r^α and $\tilde{U}_s^\beta$. Finally, we integrate over the u-variables carrying α or β as an index: first over all u_l^α with $l \geqq r + 1$, which makes $\tilde{m}(\alpha \mid \beta, s)$ act as a Poisson variable of mean value $\tilde{U}_s^\beta$; then over all the u^β's, $m(\beta \mid \alpha, r)$ and $\tilde{U}_s^\beta$ here acting independently like a Poisson variable of mean value U_r^α and like a $\frac{1}{2}\chi_{2s}^2$ respectively; lastly over all u_l^α with $1 \leqq l \leqq r$, the last remaining random variable U_r^α acting like a $\frac{1}{2}\chi_{2r}^2$.

This completes the survey of the integrations involved in the evaluation of (109). But a similar survey for the evaluation of

$$\boldsymbol{P}\{T_r^\alpha \leqq t_1 < T_r^\alpha + v_r^\alpha\}\boldsymbol{P}\{v^\alpha + T_s^\beta \leqq t_2 - t_1 < v^\alpha + T_s^\beta + v_s^\beta\}$$

will show that this expression has the same value as that occurring in (109): thus the left-hand side in (108) equals

$$\sum_{r=1}^\infty \sum_{s=1}^\infty \boldsymbol{P}\{T_r^\alpha \leqq t_1 < T_r^\alpha + v_r^\alpha\}\boldsymbol{P}\{v^\alpha + T_s^\beta \leqq t_2 - t_1 < v^\alpha + T_s^\beta + v_s^\beta\}$$

---

[8]) Compare the similar use of this property of the negative-exponential distribution in Doob [3], p. 253.

$$= \boldsymbol{P}\{x_1(t_1, \omega) = \alpha\}\, \boldsymbol{P}\{x_\alpha(t_2 - t_1, \omega) = \beta\}$$
$$= p_{1\alpha}(t_1)\, p_{\alpha\beta}(t_2 - t_1),$$

and this establishes (108).

It only remains to explain how (98) is to be established for other values of the triple (i, j_1, j_2). If $i = 1$ and $j_1 = j_2 = \alpha \geqq 2$, then we have

$$\boldsymbol{P}\{x_1(t_1, \omega) = x_1(t_2, \omega) = \alpha\}$$
$$= \sum_{r=1}^{\infty} \boldsymbol{P}\{T_r^\alpha \leqq t_1 < t_2 < T_r^\alpha + v_r^\alpha\} +$$
$$+ \sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \boldsymbol{P}\{T_r^\alpha \leqq t_1 < T_r^\alpha + v_r^\alpha \text{ and } T_{r+s}^\alpha \leqq t_2 < T_{r+s}^\alpha + v_{r+s}^\alpha\}. \quad (110)$$

The first sum in (110) can be written

$$\sum_{r=1}^{\infty} e^{-a_\alpha(t_2 - t_1)} \boldsymbol{P}\{T_r^\alpha \leqq t_1 < T_r^\alpha + v_r^\alpha\},$$

and this is equal to

$$e^{-a_\alpha(t_2 - t_1)} p_{1\alpha}(t_1).$$

The second (double) sum in (110) can be shown by a slight simplification of the previous argument to be equal to

$$\sum_{r=1}^{\infty} \sum_{s=1}^{\infty} \boldsymbol{P}\{T_r^\alpha \leqq t_1 < T_r^\alpha + v_r^\alpha\} \boldsymbol{P}\{v^\alpha + T_s^\alpha \leqq t_2 - t_1 < v^\alpha + T_s^\alpha + v_s^\alpha\}$$
$$= p_{1\alpha}(t_1)\{p_{\alpha\alpha}(t_2 - t_1) - e^{-a_\alpha(t_2 - t_1)}\},$$

and the proof of the corresponding case of (98) now follows easily.

So far we have established (98) when $i = 1$, $j_1 \geqq 2$ and $j_2 \geqq 2$. Its truth for $j_2 = 1$ now follows at once from the known fact that

$$\sum_{j=1}^{\infty} p_{\alpha j}(t_2 - t_1) = 1,$$

and its truth for $j_1 = 1$ follows similarly from the known fact that

$$\sum_{j_1=1}^{\infty} p_{1j_1}(t_1) p_{j_1 j_2}(t_2 - t_1) = p_{1j_2}(t_2).$$

It only remains for us to prove (98) (for general j_1 and j_2) when $i \equiv \alpha \geqq 2$. This time we have

$$\boldsymbol{P}\{x_\alpha(t_1, \omega) = j_1 \text{ and } x_\alpha(t_2, \omega) = j_2\}$$
$$= \boldsymbol{P}\{v^\alpha \leqq t_1,\ x_1(t_1 - v^\alpha, \omega) = j_1 \text{ and } x_1(t_2 - v^\alpha, \omega) = j_2\}$$
$$+ \delta_{\alpha j_1} \boldsymbol{P}\{t_1 < v^\alpha \leqq t_2 \text{ and } x_1(t_2 - v^\alpha) = j_2\}$$
$$+ \delta_{\alpha j_1} \delta_{\alpha j_2} \boldsymbol{P}\{v^\alpha > t_2\}$$
$$= \int_0^{t_1} p_{1j_1}(t_1 - v) p_{j_1 j_2}(t_2 - t_1) e^{-a_\alpha v} a_\alpha dv$$
$$+ \delta_{\alpha j_1} \int_{t_1}^{t_2} p_{1j_2}(t_2 - v) e^{-a_\alpha v} a_\alpha dv + \delta_{\alpha j_1} \delta_{\alpha j_2} e^{-a_\alpha t_2}$$

$$= a_\alpha(\psi_\alpha * p_{1j_1})(t_1) \cdot p_{j_1j_2}(t_2 - t_1) + \delta_{\alpha j_1}\psi_\alpha(t_1) \cdot a_\alpha(\psi_\alpha * p_{1j_2})(t_2 - t_1)$$
$$+ \delta_{\alpha j_1}\delta_{\alpha j_2}\psi_\alpha(t_1)\psi_\alpha(t_2 - t_1)$$
$$= p_{\alpha j_1}(t_1)p_{j_1j_2}(t_2 - t_1), \text{ by (32).}$$

8. *A representation of K2.* Let $\{a_\varrho : \varrho \geqq 3\}$ be a set of real numbers such that

$$a_\varrho > 0 \ (\varrho \geqq 3) \text{ and } \sum_{\varrho=3}^{\infty} 1/a_\varrho < \infty. \tag{111}$$

Whenever the suffix ϱ is used it is to be understood that $\varrho \geqq 3$. Let us form the cartesian product of a sequence of copies of the positive real line (carrying the Borel sets), the factor-lines being endowed with probability-measures and the coordinate-variables being labelled as follows:

Real line	*Probability-measure*	*Coordinate-variable*
1	$e^{-u}du$	u_1
ϱ	$e^{-a_\varrho u}a_\varrho du$	u_ϱ

There is no coordinate-variable u_2, but u_2 will be introduced as a constant and set equal to $+\infty$. This product-space with its product-measure will be identified with the fundamental probability-space $(\Omega, \mathscr{F}, \boldsymbol{P})$ of the problem, and we now define the random variables

$$T_1 \equiv u_1,$$
$$T_\varrho \equiv u_1 + \sum_{n \geqq \varrho} u_n. \tag{112}$$

Where appropriate, T_2 will be defined by a formula analogous to that defining T_ϱ, but it will be observed that $T_2 \equiv \infty$. Intuitively, the states will be described in the *anti*-well-ordered sequence

$$1, \ldots, j+1, j, j-1, \ldots, 4, 3, 2;$$

u_j will be the time spent in state j, and T_j will be the epoch of departure from state j. Thus after leaving state 1 (i.e. for $t > T_1$) the system goes through an ordinary death process (with mean lifetime a_j^{-1} in state j) and after reaching the absorbing state 2 remains there (for $t \geqq T_3$).

Since $\sum a_\varrho^{-1} < \infty$, the series $\sum u_\varrho$ converges except on an ω-set Z of $\boldsymbol{P}$-measure zero. We can now define the fundamental random functions as follows:

(i) If $\omega \in Z$, then $x_i(t, \omega) \equiv i \quad (t \geqq 0)$.

(ii) If $\omega \notin Z$, then

$$x_1(t, \omega) \equiv 1 \quad (0 \leqq t \leqq T_1),$$

and

$$x_1(t, \omega) \equiv j \quad (T_{j+1} \leqq t < T_j, \ j \geqq 2);$$

also

$$x_2(t,\ \omega) \equiv 2 \quad (t \geqq 0)$$

and

$$x_\varrho(t,\ \omega) \equiv x_1(T_{\varrho+1} + t,\ \omega) \quad (t \geqq 0).$$

It will be observed that $x_i(t,\ \omega)$ is now uniquely defined for all $i \geqq 1$, $t \geqq 0$ and $\omega \in \Omega$ because (when $\omega \notin Z$) each $t > T_1$ lies in *exactly one* of the half-open intervals $[T_\varrho,\ T_{\varrho-1})$. With these definitions $\{x_i(t, .) : t \geqq 0\}$ *will be the representation of* K2 *when* i *is the initial state*; it will be noticed that

$$x_i(0,\ \omega) = i \qquad (\text{all } \omega \in \Omega).$$

We proceed to the proof of (97) and (98), where now the array $\{p_{ij} : i,\ j \geqq 1\}$ refers to K2. The proof that

$$\boldsymbol{P}\{x_i(t,\ \omega) = j\} = p_{ij}(t) \tag{113}$$

is elementary if $i \geqq 2$ or if $i = j = 1$, and so it will be enough to consider the case $i = 1$, $j \geqq 3$; the remaining case $i = 1$, $j = 2$ will then follow from the fact that

$$\sum_{j=1}^{\infty} p_{ij}(t) = 1.$$

Now

$$\begin{aligned} \boldsymbol{P}\{x_1(t, \omega) = \varrho\} &= \boldsymbol{P}\{T_{\varrho+1} \leqq t < T_\varrho\} \\ &= \boldsymbol{P}\{T_{\varrho+1} \leqq t < T_{\varrho+1} + u_\varrho\} \\ &= \int_0^t e^{-a_\varrho(t-T)} dF_{\varrho+1}(T) \end{aligned} \tag{114}$$

where F_ϱ is the distribution-function of T_ϱ. Now the t-function at (114) is non-negative, bounded and Borel measurable, and if $\lambda > 0$ we shall have

$$\begin{aligned} \int_0^\infty e^{-\lambda t} \boldsymbol{P}\{x_1(t, \omega) = \varrho\} dt &= \frac{1}{\lambda + a_\varrho} \boldsymbol{E}(e^{-\lambda T_{\varrho+1}}) \\ &= \frac{1}{(\lambda+1)(\lambda+a_\varrho)} \prod_{\sigma=\varrho+1}^{\infty} \frac{a_\sigma}{\lambda + a_\sigma} \\ &= \int_0^\infty e^{-\lambda t} p_{1\varrho}(t) dt, \end{aligned}$$

by (82). Thus

$$\int_0^t e^{a_\varrho T} dF_{\varrho+1}(T) = e^{a_\varrho t} p_{1\varrho}(t)$$

for almost all $t > 0$, and in particular for $t = 0$; moreover the left-hand side increases with t and the right-hand side is a continuous function of t. It follows that they are equal for all $t \geqq 0$, and so (113) is established.

We now have to prove (98) in the form

$$\boldsymbol{P}\{x_i(t_1, \omega) = j \text{ and } x_i(t_2, \omega) = k\} = p_{ij}(t_1)p_{jk}(t_2 - t_1). \tag{115}$$

We can suppose that $0 < t_1 < t_2$ and we can omit consideration of *one* value of j because

$$\sum_{j=1}^{\infty} p_{ij}(t_1)p_{jk}(t_2 - t_1) = p_{ik}(t_2);$$

we choose to omit the value $j = k$. For the moment let $i = 1$. Now each side of (115) vanishes if $j \neq 1$ and $j < k$, or if $k = 1$ and $j > 1$. Thus we are left with the cases

(a) $i = 1,\ j > k > 1$

and

(b) $i = j = 1,\quad k > 1.$

In case (a) we have

$$\begin{aligned}
&\boldsymbol{P}\{x_1(t_1, \omega) = j \text{ and } x_1(t_2, \omega) = k\} \\
&= \boldsymbol{P}\{T_{j+1} \leqq t_1 < T_j \leqq T_{k+1} \leqq t_2 < T_k\} \\
&= \boldsymbol{P}\{T_{j+1} \leqq t_1 < T_{j+1} + u_j \text{ and } \tilde{u}_j + \sum_{k+1}^{j-1} u_n \leqq t_2 - t_1 < \tilde{u}_j + \sum_{k+1}^{j-1} u_n + u_k\},
\end{aligned}$$

where $\tilde{u}_j = (u_j - (t_1 - T_{j+1})^+)^+$. The splitting of the last probability into the two factors

$$\boldsymbol{P}\{T_{j+1} \leqq t_1 < T_j\}\, \boldsymbol{P}\{T_{k+1} \leqq T_{j+1} + t_2 - t_1 < T_k\}, = p_{1j}(t_1)p_{jk}(t_2 - t_1),$$

now follows as in § 7.

In case (b) we have

$$\begin{aligned}
&\boldsymbol{P}\{x_1(t_1, \omega) = 1 \text{ and } x_1(t_2, \omega) = k\} \\
&\qquad = \boldsymbol{P}\{t_1 \leqq u_1 \text{ and } T_{k+1} \leqq t_2 < T_k\} \\
&\qquad = \boldsymbol{P}\{t_1 \leqq u_1 \text{ and } \tilde{u}_1 + \sum_{k+1}^{\infty} u_n \leqq t_2 - t_1 < \tilde{u}_1 + \sum_{k}^{\infty} u_n\}
\end{aligned}$$

where $\tilde{u}_1 = (u_1 - t_1)^+$, and then in the same way as before we find that this probability is equal to

$$\boldsymbol{P}\{t_1 \leqq u_1\}\, \boldsymbol{P}\{T_{k+1} \leqq t_2 - t_1 < T_k\} = p_{11}(t_1)p_{1k}(t_2 - t_1).$$

To complete the proof of (115) we have now only to deal with the case $i > 1$, but this is entirely elementary.

It will be noticed that (apart from the complications associated with the absorbing state) K2 is the reversal in time of a process with well-ordered jumps of a type rather similar to that introduced by Doob [2].

References

[1] L. W. Cohen and N. Dunford, "Transformations on sequence spaces", Duke Math. J., **3** (1937), 689—701.

[2] J. L. Doob, "Markoff chains-denumerable case", Trans. Amer. Math. Soc., **58** (1945), 455—473.

[3] J. L. Doob, Stochastic Processes (New York, 1953).

[4] W. Feller, An Introduction to Probability Theory and its Applications, Vol. I (New York, 1950).

[5] W. Feller, "The parabolic differential equations and the associated semi-groups of transformations", Annals of Math., **55** (1952), 468—519.

[6] E. Hille, Functional Analysis and Semigroups (New York, 1948).

[7] E. Hille, "On the generation of semi-groups and the theory of conjugate functions", Kungl. Fys. Sälls. Lund Förh., **21** (1952), No. 14.

[8] E. Hille, "On the integration of Kolmogoroff's differential equations", Proc. Nat. Acad. Sci. U.S.A., **40** (1954), 20—25.

[9] E. Hille, "Quelques remarques sur les equations de Kolmogoroff", Bull. Soc. Math. Phys. Serbie, **5** (1953), 3—14.

[10] A. Jensen, A Distribution Model (Copenhagen, 1954).

[11] T. Kato, "On the semi-groups generated by Kolmogoroff's differential equations", J. Math. Soc. Japan, **6** (1954), 1—15.

[12] D. G. Kendall, "Some analytical properties of continuous stationary Markov transition functions", Trans. Amer. Math. Soc., **78** (1955), 529—540.

[13] A. N. Kolmogorov, "On some problems concerning the differentiability of the transition-probabilities in a temporally homogeneous Markov process having a denumerable set of states", Učenye Zapiski (Matem.) Moskov. Gos. Univ. (4), **148** (1951), 53—59.

[14] W. Ledermann and G. E. H. Reuter, "Spectral theory for the differential equations of simple birth and death processes", Philos. Trans. Roy. Soc. London Ser. A, **246** (1954), 321—369.

[15] P. Lévy, "Systèmes markoviens et stationnaires; cas dénombrable", Ann. Sci. École Norm. Sup. (3), **68** (1951), 327—381.

[16] P. Lévy, "Complément à l'étude des processus de Markoff", ibid., **69** (1952), 203—212.

[17] R. S. Phillips, "The adjoint semi-group", Bull. Amer. Math. Soc., **60** (1954), 158.

[18] G. E. H. Reuter, "Über eine Volterrasche Integralgleichung mit totalmonotonem Kern", Archiv. der Math. (to appear).

[19] K. Yosida, "On the differentiability and the representation of one-parameter semi-groups of linear operators", J. Math. Soc. Japan, **1** (1948), 15—21.

[20] K. Yosida, "An operator-theoretical treatment of temporally homogeneous Markoff processes", ibid., **1** (1949), 244—253.

Magdalen College, Oxford.
Victoria University, Manchester.

PROCESSUS SEMI-MARKOVIENS

PAUL LÉVY

I. *Introduction.*

1. L'idée de processus stochastique est, au moins pour un déterministe, liée à celle de l'existence de paramètres cachés, qui n'interviennent pas dans la description de l'état apparent actuel du système étudié, et qui cependant influent sur son évolution future. L'ignorance où nous sommes de leurs valeurs nous oblige à ne parler pour l'avenir que d'un ensemble d'évolutions possibles, et dans certains cas nous pouvons définir dans cet ensemble une loi de probabilité, sans cesse modifiée par la connaissance de données nouvelles.

Parmi les paramètres cachés, il peut en exister qui présentent à la fois les deux caractères suivants: d'une part le passé nous renseigne sur leurs valeurs, ou du moins nous conduit à admettre pour eux une loi de probabilité autre que celle qui résulterait de la seule connaissance de l'état apparent actuel; d'autre part cette circonstance nous conduit à modifier nos prévisions concernant l'avenir. Si aucun paramètre caché ne présente à la fois ces deux caractères, le processus est markovien. Nous dirons qu'il est *semi-markovien* s'il y a un et un seul paramètre caché de cette nature, le temps u écoulé depuis le dernier changement d'état, qui serait donc marqué par une horloge ramenée au zéro à chaque changement d'état apparent.

Naturellement, si u est maintenu au zéro par des changements d'états ininterrompus, rien ne distingue l'évolution semi-markovienne d'une évolution markovienne. Nous éliminerons ici ce cas, et aussi celui où on aurait des alternances d'intervalles de repos et d'intervalles de mouvement ininterrompu (continu ou non) [1]). D'une manière précise, nous ferons l'hypothèse suivante: *les intervalles de repos* (que nous appellerons intervalles i_ν), *forment presque sûrement sur l'axe des t un ensemble partout dense.*

Il suffit évidemment de considérer l'horloge qui marque le temps u comme faisant partie du système étudié pour ramener un processus semi-markovien à un processus markovien régi par une équation de Chapman-Kolmogoroff. Mais il nous paraît préférable d'employer une autre méthode, inspirée par nos

[1]) Pour des processus plus généraux, il y aurait lieu d'étudier séparément les possibilités de mouvement continu, celles d'un mouvement ininterrompu, mais discontinu, et celles du mouvement intermittent seul considéré ici. On étudierait ensuite la probabilité du passage de chacun de ces régimes aux autres.

travaux antérieurs sur les processus markoviens dans le cas dénombrable, c'est-à-dire celui où il n'y a qu'une infinité dénombrable de cas possibles [2]). Nous serons ainsi conduits à étudier séparément les états à durées positives, et les états instantanés. Pour ces derniers, nous n'aurons qu'à rappeler ce que nous avons déjà dit, et le compléter sur certains points, notamment en ce qui concerne l'extension de nos résultats au cas non dénombrable. C'est par l'étude des autres états que nous commencerons, mais après quelques remarques préliminaires sans lesquelles le problème ne serait pas nettement posé.

2. Chaque fois que le système étudié arrive à un état A_h, la durée U_h de son séjour dans cet état est une variable aléatoire non négative, dont la loi ne dépend que de h. Il peut arriver que les trois éventualités

$$U_h = 0, \quad 0 < U_h < \infty, \quad U_h = \infty$$

aient des probabilités toutes positives, ou que deux d'entre elles le soient. Nous supposerons ce cas éliminé par un artifice très simple, qui consiste à effectuer immédiatement une expérience préliminaire qui déterminera, non seulement si U_h est positif, ou nul, mais s'il est fini ou infini, et à dire alors qu'on est dans le cas A'_h, ou A''_h, ou A'''_h suivant le résultat de cette expérience [3]). Cela fait, nous pouvons dire que: *chaque état A_h est, soit sûrement instantané* (c'est-à-dire que $U_h = 0$), *soit sûrement final* ($U_h = \infty$), *soit enfin à durée sûrement positive ou finie*; nous dirons dans ce dernier cas qu'il est *normal*.

Naturellement, au sujet des états finaux, nous n'avons rien à dire, sinon que l'évolution est terminée quand le système arrive à un tel état. Remarquons aussi, au sujet des états normaux, qu'une épreuve particulière ne peut en réaliser qu'un nombre fini ou une infinité dénombrable. Mais il peut y en avoir une infinité non dénombrable qui soient possibles. Même s'il n'y a jamais que deux états qui puissent succéder à un état donné, il peut arriver qu'après une infinité de changements en un temps fini il y ait une infinité non dénombrable d'états possibles.

Si un état A_h est instantané, l'ensemble E_h des instants où il est réalisé ne comprend aucun intervalle. Il peut être de types très variés; il peut en particulier être non mesurable. Nous ferons une deuxième hypothèse restrictive, qui consiste à supposer que *tous les E_h sont presque sûrement mesurables*. Cela élimine

---

[2]) Ann. Ec. Norm. sup., 68 (1951), p. 327—381, et 69 (1952), p. 203—212. Voir aussi C. R. Acad. Sc., 236 (1953), p. 1630—1632.

[3]) Précisons bien qu'une réduction analogue peut être nécessaire si U_h n'est que presque sûrement nul. On pourrait en effet avoir confondu un état A'_h sûrement instantané mais susceptible d'être réalisé une infinité non dénombrable de fois, et un état normal A''_h. A chaque arrivée à l'un ou l'autre de ces états, il est presque sûr que $U_h = 0$; cela n'exclut pas la possibilité de réalisations intermittentes de A''_h.

les processus des sixième et septième types de la classification introduite dans nos travaux antérieurs relatifs au cas dénombrable.

3. Rappelons le principe de cette classification. Nous désignerons par $\mathscr{E}'$ la réunion des intervalles *ouverts* i_ν, et par $\mathscr{E}$ son complément sur l'axe des t, qui est nécessairement un ensemble fermé. D'après l'hypothèse restrictive faite au n°. 1, $\mathscr{E}$ ne comprend aucun intervalle, et peut être de cinq types possibles:

Type $\mathscr{E}_1$, ou *type fini*; $\mathscr{E}$ n'a pas de point d'accumulation à distance finie.

Type $\mathscr{E}_2$, ou *type transfini*; $\mathscr{E}$ a au moins un point d'accumulation à distance finie, et est bien ordonné.

Type $\mathscr{E}_3$: $\mathscr{E}$ est dénombrable, et n'est pas bien ordonné.

Type $\mathscr{E}_4$: $\mathscr{E}$ est non dénombrable, mais encore de mesure nulle.

Type $\mathscr{E}_5$: $\mathscr{E}$ a une mesure positive.

Un processus est alors du type ν ($\nu = 1, \ldots, 5$) si, au moins pour certains états initiaux, $\mathscr{E}$ peut (avec une probabilité positive) être du type $\mathscr{E}_\nu$, mais ne peut jamais être d'un type d'indice plus élevé. Les indices 6 et 7, qui introduisent des ensembles E_h non mesurables, sont ici exclus.

Dès le second type, $H(t)$ désignant l'indice de l'état réalisé à l'instant t (cet *indice* pouvant d'ailleurs être une variable continue, ou même un point d'un espace à plusieurs dimensions), apparaissent des instants pour lesquels $H(t-0)$ n'est pas déterminé; dès le troisième type, $H(t+0)$ peut aussi ne pas l'être. Il peut être commode, comme nous l'avons fait dans nos travaux cités, de dire qu'à ces instants $H(t)$ est l'indice d'un état *fictif*. Cela semble même indispensable dans certains cas, dont nous avons donné des exemples, où il y a plusieurs états fictifs qu'il faut distinguer les uns des autres. Naturellement, non seulement chaque état fictif ne peut être réalisé que sur un ensemble de t-mesure nulle, mais il en est de même de leur réunion. On ne saurait considérer comme fictif un ensemble d'états qui occupe sur l'axe des t un ensemble de mesure positive, et ait donc une probabilité positive d'être réalisé à un instant choisi au hasard.

Nous supposerons aussi qu'il n'a de même aucune chance d'être réalisé à un instant donné d'avance. Cette hypothèse, équivalente à la précédente dans le cas markovien, est ici un peu plus restrictive.

II. *Etude des états normaux.*

4. Soit A_h un de ces états; soit K l'indice de l'état (réel ou fictif) qui lui succèdera. Pour chacun de ces états, nous pouvons nous donner d'une manière absolument quelconque la loi dont dépend l'ensemble des variables aléatoires U_h et K, à cela près que U_h est presque sûrement positif et fini. Nous inspirant, comme dans nos travaux antérieurs, de l'idée qu'il est commode d'étudier,

d'abord les différentes successions d'états possibles et leurs probabilités, ensuite seulement la rapidité de l'évolution, nous nous donnerons d'abord, pour chaque h, la loi de probabilité de K, ensuite, en fonction de h et de la valeur k de K, la fonction de répartition conditionnelle $F_{h,k}(u)$ de U_h. La fonction de répartition inconditionnelle est alors

$$F_h(u) = E\{F_{h,K}(u)\}. \tag{1}$$

L'idée d'introduire $F_h(u)$ est due à K. L. Chung, qui a observé qu'en introduisant ainsi une fonction de répartition arbitraire, on obtient des processus qui comprennent à la fois les processus markoviens pour lesquels $F_h(u)$ est de la forme $1 - e^{-\lambda_h u}$, et les chaînes de Markoff pour lesquelles U_h est nécessairement multiple d'un certain quantum de temps θ; $F_h(u)$ est alors de la forme $1 - q^{-n}$, n étant la partie entière de u/θ. C'est cette idée de Chung qui est l'origine du présent travail. Du moment que le système garde le souvenir du temps écoulé depuis le dernier chagement d'état, il nous a paru naturel d'admettre que la loi de probabilité de K peut dépendre de ce temps. Mais au lieu d'étudier d'abord U_h, et ensuite K, nous avons préféré l'ordre inverse, qui est plus commode.

5. L'étude des différentes successions d'états possibles s'effectue exactement comme dans le cas markovien. Il faut d'ailleurs remarquer que, même si ces successions ne sont pas des suites bien ordonnées, chacune d'elles est définie sans ambiguïté, et cela même s'il y a des états instantanés. Il y a au plus une infinité dénombrable d'intervalles i_ν; on peut les dénombrer, et, pour chacun d'eux, définir ses relations d'ordre avec ceux dénombrés avant lui et l'état A_h qui lui correspond. L'ensemble $\mathscr{E}$ étant discontinu, chacun de ses points correspond à une coupure entre les intervalles ainsi dénombrés.

On peut, et même d'une infinité de manières, associer à chaque processus semi-markovien un processus markovien qui corresponde à la même répartition de la probabilité dans le même ensemble de successions d'états possibles. On est donc sûr de pouvoir obtenir n'importe quel processus semi-markovien en formant d'abord tous les processus markoviens, et en les modifiant ensuite par l'introduction des fonctions $F_{h,k}(u)$, substituées aux fonctions $1 - e^{-\lambda_h u}$ pour tous les états qui ne sont ni instantanés ni finaux.

6. En ce qui concerne le temps T que dure une succession donnée d'états, nous allons d'abord montrer que:

Théorème. T est toujours, soit presque sûrement fini, soit presque sûrement infini.

D'après la convention faite au n°. 2, il en est ainsi pour chacun des termes U_ν dont T est la somme. Ces différents termes sont indépendants (puisqu'au début de chaque intervalle i_ν tout le passé est oublié). Le théorème énoncé est alors un corollaire du théorème connu sur les séries à termes aléatoires indé-

pendants, qui sont toujours presque sûrement convergentes ou presque sûrement divergentes.

Ce théorème n'est bien entendu démontré pour le moment que pour les quatre premiers types, dans lesquels les états instantanés ne peuvent être réalisés que sur un ensemble de mesure nulle. Nous verrons au n°. 12 que, pour un processus du cinquième type, l'existence de cet ensemble conduit à ajouter à T un terme qui n'est pas aléatoire. Le théorème considéré reste donc vrai.

Il en résulte que, dans une succession à durée infinie, on peut définir une coupure non aléatoire qui est presque sûrement la limite exacte des états réalisés en un temps fini.

Remarquons aussi que, comme dans le cas markovien, un même état normal ne peut presque sûrement pas être réalisé une infinité de fois en un temps fini. Cela est évident si l'on examine ce problème indépendamment de l'étude des différentes successions possibles. Chaque fois qu'un état A_h est réalisé, sa durée $U_h^{(v)}$ dépend de la fonction de répartition de Chung, $F_h(u)$, qui ne varie pas avec v. Par suite, le temps nécessaire à n réalisations augmente presque sûrement indéfiniment avec n.

7. D'autres résultats relatifs au cas markovien cessent au contraire d'être applicables. Ainsi, dans le cas markovien, T est presque sûrement fini ou infini en même temps que sa valeur probable, et, pour les processus des quatre premiers types, c'est une variable aléatoire à densité de probabilité continue et positive de zéro à l'infini.

Au contraire, dans le cas semi-markovien, tous ces énoncés peuvent être en défaut pour une succession réduite à un seul état A_h. Alors $T = U_h$, et cette variable aléatoire peut être sûrement finie tout en ayant une valeur probable infinie, et peut aussi dépendre d'une loi totalement discontinue. La continuité s'améliore naturellement par l'addition de plusieurs termes U_h. Mais on peut définir des exemples de suites infinies d'états pour lesquelles la durée T dépende d'une loi discontinue.

Si par exemple tous les états sont réalisés dans l'ordre des h croissants ($h = 1, 2, \ldots$), et que U_h ait les valeurs possibles u_h et $u_h + 1$, avec les probabilités $1 - \alpha_h$ et α_h, si $S = \Sigma u_h$ et $\alpha = \Sigma\alpha_h$ sont finis, T est presque sûrement la somme de S et d'un entier fini, dont la valeur probable est α.

Dans le cas où U_h a deux valeurs possibles et également probables $u_h = q^{-h}$ et $2u_h$, avec $q > 2$, les valeurs possibles pour le temps T forment un ensemble parfait discontinu, de mesure nulle. La loi dont il dépend n'est donc pas absolument continue.

On peut multiplier ces exemples à l'infini. Signalons encore le cas où chaque U_h est sûrement rationnel, et où chacune des valeurs $2^{-h!}$ et $2^{1-h!}$ a la probabilité $\frac{1-\alpha_h}{2}$, la série $\Sigma\alpha_h$ étant convergente. Alors ΣU_h est presque sûrement un

nombre transcendant de Liouville. Cette somme dépend encore d'une loi singulière, c'est-à-dire à répartition continue sur un ensemble de mesure nulle.

8. L'étude des théorèmes asymptotiques comporte évidemment beaucoup de possibilités qui n'existent pas dans le cas markovien. Plaçons-nous par exemple dans le cas d'un groupe final ergodique, contenant les états A_0 et A_1, et désignons respectivement par T_0 et T_1 les temps passés dans ces deux états jusqu'à un instant t (nous les supposons, comme dans le cas dénombrable, *exactement* réalisés une infinité de fois). Pour comparer T_0 et T_1, nous pouvons faire abstraction des autres états, de sorte que $t = T_0 + T_1$, et si plusieurs intervalles où un même état soit réalisé deviennent ainsi consécutifs, les réunir en un seul intervalle. On est ramené au cas d'une alternance non aléatoire des états A_0 et A_1, les temps U_0 et U_1 étant seuls aléatoires.

Si leurs valeurs probables μ_0 et μ_1 sont finies, T_0/T_1 tend presque surement vers μ_0/μ_1. Ce résultat subsiste si μ_1 est seul infini. Mais si μ_0 et μ_1 sont tous les deux infinis, le rapport T_0/T_1 n'a presque sûrement pas de limite. Il peut même osciller indéfiniment entre des minima qui tendent vers zéro et des maxima qui augmentent indéfiniment.

Voici un exemple de cette circonstance. Donnons-nous une suite d'entiers n_p qui croissent assez rapidement pour que

$$\sum \frac{p n_p}{n_{p+1}} < \infty, \tag{2}$$

et une variable aléatoire X, ayant une infinité dénombrable de valeurs possibles x_p, de probabilités $\alpha_p \sim 1/n_p$ $(p \to \infty)$. Soit S_n la somme des n premières déterminations de X, au cours d'une suite illimitée d'épreuves indépendantes. Pour $n = n'_p = p n_p$, désignons respectivement par N_p et N'_p les nombres de termes de S_n égaux respectivement à x_p et x_{p+1}. D'après des théorèmes connus, il est presque sûr que tous les N'_p d'indices p assez grands sont nuls, tandis que N_p est un infiniment grand équivalent à p. Si d'ailleurs les x_p croissent assez vite pour que

$$\frac{n_p}{n_{p-1}} x_{p-1} = o(x_p) \qquad (p \to \infty)\ ^4), \tag{3}$$

la somme $N_p x_p$ des termes égaux à x_p est prépondérante, de sorte que, pour $n = n_p$, S_n est un infiniment grand presque sûrement équivalent à $p x_p$ [5]).

---

[4]) Les conditions (2) et (3) sont realisées si on prend, par exemple

$$n_p = (3p)!,\ x_p = (4p)!$$

[5]) Nous ne parlons ici que des valeurs n'_p de n. Pour être assuré, quel que soit n assez grand, de la prépondérance du plus grand terme effectivement réalisé, il suffit de rem-

Prenons maintenant pour U_0 et U_1 des variables ayant des valeurs possibles u_p et v_p, de probabilités α_p, et liées aux x_p par les formules

$$u_{2p} = u_{2p+1} = x_{2p}, \quad v_{2p-1} = v_{2p} = x_{2p-1}. \tag{4}$$

Après n alternances des états A_0 et A_1, T_0 et T_1 sont des sommes du type S_n, comprenant l'une n termes du type U_0, l'autre n termes du type U_1. Pour $n = n'_{2p}$, on a alors presque sûrement

$$T_0 \sim 2px_{2p}, \quad T_1 \sim 2p \frac{n_{2p}}{n_{2p-1}} x_{2p-1} = o(T_0), \tag{5}$$

tandis que le contraire a lieu pour $n = n'_{2p+1}$, à cause de l'apparition dans T_1 des termes égaux à x_{2p+1}.

III. *Le rôle des états instantanés.*

9. Soit A_0 un état instantané, pouvant être réalisé sur un ensemble E_0 de mesure positive. Soit τ la mesure de $E_0 \cap (0, t)$. C'est une fonction continue de t, à dérivée nulle en dehors de E_0, presque partout définie et égale à l'unité sur cet ensemble. Il en est donc ainsi presque sûrement pour n'importe quel $t \in E_0$ si l'on n'a aucun renseignement sur ce qui se passe à droite de ce point. Désignons par I_ν, ou par (t'_ν, t''_ν), un quelconque des intervalles qui constituent le complément de E_0 (sans nous demander s'il comprend un ou plusieurs i_ν). Comme il n'y en a qu'une infinité dénombrable, il est presque sûr que, pour tous les points t''_ν, $\tau = \tau(t)$ a une dérivée à droite égale à un; il en est de même de la dérivée à gauche pour les points t'_ν. Des points isolés ne peuvent donc jouer aucun rôle, et, pourvu bien entendu qu'elle soit $< \tau(\infty)$, une valeur positive donnée de τ correspond, soit (exceptionnellement) aux deux extrémités d'un intervalle I_ν, soit à un point et un seul de la fermeture $\overline{E}_0$ de E_0. Nous reviendrons plus loin sur la différence entre $\overline{E}_0$ et E_0.

10. Le nombre des I_ν de longueurs $> u$ réalisés pendant que τ varie de τ' à τ'' est évidemment une variable de Poisson, de valeur probable $(\tau'' - \tau')N_1(u)$, $N_1(u) = -N(u)$ étant une fonction bornée pour tout $u > 0$, et non croissante. On a évidemment $N(\infty) = 0$, et aussi $N(0) = -\infty$ (autrement il n'y aurait qu'un nombre fini de I_ν, et l'état A_0 ne serait pas instantané). Donc, comme nous l'avions déjà indiqué dans nos travaux cités, la différence $t - \tau$ (somme des

---

placer la condition (3) par la condition un peu plus restrictive

$$\frac{n_p}{n_{p-1}} x_{p-1} \log p = o(x_p).$$

En effet, pour $n \sim 2n_p \log p$, et p assez grand, il y a au moins un terme égal à x_p, et la somme des termes égaux à x_{p-1} est encore $o(x_p)$.

longueurs des I_ν) dépend, pour chaque valeur donnée de τ, de la loi indéfiniment divisible définie par la formule

$$\log E\{e^{iz(t-\tau)}\} = \tau \int_0^\infty (e^{izu} - 1) dN(u), \tag{6}$$

où la fonction $N(u)$ doit vérifier les conditions

$$N(+0) = -\infty,\ dN(u) \geqq 0 \text{ (pour } u \text{ et } du > 0),\ \int_0^\infty \frac{u}{1+u} dN(u) < \infty. \tag{7}$$

La condition $N(\infty) = 0$ n'est pas essentielle, puisque $dN(u)$ intervient seul. Elle donne seulement une forme canonique.

N'importe quelle fonction $N(u)$ vérifiant ces conditions peut être obtenue pour définir dans un processus semi-markovien la relation stochastique qui existe entre τ et $t - \tau$. On peut en effet associer à chaque $u > 0$ un état A_u tel qu'après un temps u passé dans cet état le système revienne sûrement à l'état A_0. La fonction $N(u)$ définit les probabilités des réalisations de ces différents états. D'après la condition $N(+0) = -\infty$, les valeurs de τ correspondant à leurs réalisations forment un ensemble partout dense, et A_0 est bien un état instantané. Si $N(+0) > -\infty$, la formule (6) est encore acceptable, mais A_0 serait un état à durée U_0 positive, et de type markovien, c'est-à-dire que cette variable U_0 serait indépendante de l'état qui suit A_0, et que sa fonction de répartition serait de la forme $1 - e^{-\lambda u}$. Sans ces restrictions, la loi dont dépend $t - \tau$ (pour chaque valeur donnée de τ) ne serait en général pas indéfiniment divisible.

Il faut remarquer que n'importe quelle loi du type (6) peut être obtenue même dans le cas dénombrable. On peut en effet grouper tous les états A_u en une infinité dénombrable de groupes A'_h correspondant chacun à une variation finie de $N(u)$, et ne pas distinguer les différents A_u d'un même groupe. En définissant convenablement, d'abord les probabilités associées aux différents A'_h, puis leurs durées U_h, on peut reconstituer les conditions des choix définis par $N(u)$.

Au contraire, si on se restreint au cas markovien, il faut imposer à $N(u)$ de nouvelles conditions restrictives. Il faut notamment que cette fonction soit absolument continue, et à dérivée continue et positive de zéro à l'infini; mais ces conditions ne sont pas suffisantes. Le résultat obtenu pour les processus semi-markoviens est à ce point de vue plus satisfaisant que celui relatif au cas markovien.

11. En supprimant sur l'axe des t le temps réservé à l'état A_0, c'est-à-dire en prenant $t' = t - \tau$ comme nouveau temps, on obtient, du moins s'il n'y a pas d'autre état instantané, un processus du quatrième type; E_0 est remplacé

par un ensemble E_0', non dénombrable, et de mesure nulle. Inversement, en dilatant cet ensemble, on peut retrouver, non seulement le processus initial pour lequel $t = t' + \tau$, mais une infinité de processus analogues pour lesquels $t = \tau + ct'$.

Cette dilatation de E_0' implique qu'on sache définir τ, à un facteur constant près. On y arrive aisément par la remarque suivante: si, t variant de t' à t'', τ varie de τ' à τ'', le nombre $\Phi(l)$ des intervalles I_ν intérieurs à (t', t'') et de longueurs $> l$, est, quand l tend vers zéro, un infiniment grand presque sûrement équivalent à $(\tau'' - \tau')N(l)$. Il suffit de choisir arbitrairement un intervalle pour lequel on conviendra que $\tau'' - \tau' = 1$; on en déduira l'allure asymptotique de $N(l)$, et la τ-mesure de n'importe quel autre intervalle.

En énonçant cette remarque, nous avons admis que la formule (6) était applicable. Il reste à montrer qu'elle n'implique aucune restriction.

Nous supposerons l'état A_0 instantané (cette hypothèse est d'ailleurs inutile si le processus est markovien). Considérons en effet la section de E_0 commencant à un point quelconque t_0 de cet ensemble et se terminant à l'extrémité t du premier intervalle I_ν de longueur supérieure à l_0 situé à droite de t_0 [6]). Si $l > l_0$, la probabilité que cette longueur soit supérieure à l a une valeur bien déterminée, de la forme $N(l)/N(l_0)$, et, quel que soit l positif, le nombre probable des I_ν intérieurs à l'intervalle (t_0, t) et de longueurs supérieures à l est $N(l)/N(l_0)$. Ce nombre probable deviendra $\tau N(l)$, avec $\tau = n/N(l_0)$, si on s'arrête après le $n^{ième}$ intervalle I_ν de longueur supérieure à l_0, et, par un passage à la limite, on arrive à la même forme pour une section quelconque de E_0. La formule (6) en resulte, et les remarques relatives à la détermination de τ d'après l'allure de $\Phi(l)$ pour l très petit restent valables [7]).

Au lieu de l'infiniment grand $\varphi(l)$, nous aurions pu considérer l'infiniment petit $\Psi(l)$, longueur totale des I_ν intérieurs à (t', t'') et de longueurs $\leqq l$, ou bien la mesure $\Psi(2l) + 2l\Phi(2l)$ de l'ensemble des points de l'intervalle (t', t'') dont la distance à E_0 est > 0 et $\leqq l$. Toutes ces expressions conduisent à des *mesures du voisinage de* E_0, invariantes pendant la dilatation et la contraction de cet ensemble, et qui sont proportionnelles à τ.

12. Au lieu de se donner les longueurs des intervalles I_ν, nous pouvons nous donner une succession d'états, comportant une infinité de passages,

---

[6]) Nous supposons l_0 assez petit pour que cette circonstance soit possible; elle finira alors presque sûrement par être réalisée s'il y a une infinité de retours à l'état A_0.

[7]) Nous avons écarté les processus des sixième et septième types, dans lesquels l'état A_0 et d'autres états A_0', A_0'', . . ., peuvent être associés de manière que chacun d'eux puisse être réalisé sur un ensemble presque sûrement partout dense dans $\overline{E}_0$. Les raisonnements du texte s'appliquent même à ces processus, en substituant $\overline{E}_0$ à E_0, et l'ensemble de ces états à A_0.

nécessairement instantanés, à l'état A_0. Le problème est un peu différent de celui du no 11, parce que les longueurs des I_ν ne sont pas connues. Mais, par des raisonnements analogues à ceux du n°. 11, on arrive encore à ce résultat que le temps τ passé dans l'état A_0 n'est pas aléatoire. Ainsi, même pour les processus markoviens, pour lesquels le temps consacré à une succession donnée d'états normaux, s'il est fini, peut être arbitrairement petit, le temps consacré aux états instantanés n'est pas aléatoire, de sorte que, pour les processus du cinquième type, la durée totale a en général une borne inférieure positive.

Cela n'est pas en contradiction avec le fait que le temps nécessaire pour passer de l'état A_0 à un autre état A_h peut toujours, pour un processus markovien, être arbitrairement petit. C'est que la succession des états intermédiaires n'est pas ici donnée d'avance, et la période des retours possibles à l'état A_0 peut être arbitrairement petite. Dans ces conditions, et pour tous les processus semi-markoviens, si l'état A_0 est instantané et susceptible d'être réalisé sur un ensemble de mesure positive, et si le passage de A_0 à A_h est possible, l'augmentation de τ avant la première réalisation de A_h a une valeur probable l/λ toujours positive et finie, et sa densité de probabilité est $\lambda e^{-\lambda t}$.

13. Occupons-nous maintenant du cas où il y a deux états instantanés A_0 et A_1, pouvant au cours d'une même épreuve être réalisés sur des ensembles non dénombrables E_0 et E_1. Nous pouvons toujours, en dilatant au besoin ces ensembles, supposer que leurs mesures τ_0 et τ_1 sont positives. Cette opération préalable une fois effectuée, considérons le processus auxiliaire obtenu cette fois en supprimant les intervalles I_ν extérieurs à la fois à E_0 et E_1. Le nouveau temps sera $\tau = \tau_1 + \tau_2$, et les mesures de E_0 et E_1 seront conservées dans cette représentation sur l'axe des τ.

Le processus auxiliaire sera ici un processus markovien à deux états possibles. Un tel processus pourrait être du sixième type. Mais cette possibilité est exclue puisque les τ-images de E_0 et E_1 sont mesurables, et toutes les deux à mesures positives. Il n'y a alors qu'une seule possibilité: sur l'axe des τ, les états A_0 et A_1 ne sont plus instantanés. En revenant à l'axe des t, on y trouve alternativement des intervalles favorables à A_0, contenant chacun une section de E_0 et ne contenant aucun point de E_1, et des intervalles favorables à E_1. Ces intervalles ont tous des longueurs positives, supérieures à celles de leurs τ-images, et en un temps fini il n'y en a qu'un nombre fini.

Les fermetures de $\overline{E}_0$ et $\overline{E}_1$ n'ont donc en un temps fini qu'au plus un nombre fini de point commun. Il peut d'ailleurs arriver qu'il n'y en ait aucun. Un point où, sur l'axe des τ, on passe de A_0 à A_1, ou inversement, peut correspondre à un intervalle I_ν de l'axe des t, ou à un passage direct d'un état à l'autre sur l'axe des t. On peut aisément définir des processus pour lesquels ces deux circonstances sont effectivement possibles.

14. Ces remarques, déjà indiquées dans notre note de 1953, permettent une conclusion pour le cas dénombrable. Les fermetures $\overline{E}_h$ et $\overline{E}_k$ de deux ensembles E_h et E_k n'ont qu'un nombre fini de points communs sur une portion finie de l'axe des t (nous venons de le démontrer pour les états instantanés; cela est évident s'il s'agit d'états dont un au moins soit normal). Donc il y a au plus une infinité dénombrable de points qui appartiennent à la fermeture de plus d'un E_h, et aucun point ne peut appartenir à la fermeture à droite (ou à gauche) de deux ensembles E_h.

D'ailleurs l'ensemble des points pour lesquels le système n'est pas dans un état déterminé est évidemment de mesure nulle (ce n'est qu'un autre énoncé d'une remarque faite au n°. 3 à propos des états fictifs). Donc un ensemble E_h ne peut différer de sa fermeture que par un ensemble de mesure nulle.

15. Naturellement le cas non dénombrable donne lieu à des possibilités nouvelles. Il peut exister des états réalisables chacun au plus sur un ensemble de mesure nulle, et dont la réunion n'ait pas ce caractère. Chacun est alors fictif, mais leur réunion ne peut pas être considérée comme fictive.

Nous avons déjà indiqué l'exemple d'un système dépendant d'un paramètre u, jamais décroissant, et marquant un temps d'arrêt à chaque valeur rationnelle du paramètre. Comme il s'agissait d'une étude réservée au cas dénombrable, nous ne pouvions réserver aucun temps aux valeurs irrationnelles, sauf peut-être à une infinité dénombrable de ces valeurs). Nous pouvons ici lui réserver un temps $\tau = \tau(u)$, cette fonction étant non décroissante, à cela près quelconque, et, si elle est continue, n'étant pas aléatoire. Nous pouvons généraliser cet exemple, en prenant pour t une somme de deux termes, l'un étant une fonction non aléatoire continue et non décroissante $\tau(u)$, et l'autre une somme d'arrêts correspondant à un ensemble dénombrable de valeurs données de u, indépendants les uns des autres, les lois dont ils dépendent n'étant restreintes que par une condition de convergence relative à une partie au moins de l'axe des u. Si cet ensemble dénombrable est partout dense, il s'agit bien d'un ensemble du type étudié ici. Dans le cas contraire, il y aurait des périodes de mouvement continu, d'ailleurs non aléatoires; ce serait le cas écarté par l'hypothèse faite au n°. 1.

THE ESTIMATION OF VELOCITY DISTRIBUTIONS FROM COUNTS

D. V. LINDLEY

1. *Introduction.*

Chandrasekhar (2) has given an account of Smoluchowski's development of the theory of probability after-effect. The theory has been used by Fürth (3, 4) to estimate the speed of pedestrians and by Rothschild (5) to estimate the speed of spermatazoa. The basic ideas behind these uses of the theory are as follows. Counts are made of the number of particles (e.g. pedestrians or spermatazoa) in a given region (in one, two or three dimensions) at successive equally-spaced instants of time. If the particles move independently one of another and if, loosely speaking, the motion is stationary then Smoluchowski shewed that

$$\mathscr{E}[(n_{i+1} - n_i)^2] = 2\nu(1 - p), \tag{1}$$

where $\{n_i\}$ is the observed sequence of counts, ν is the average number of particles in the region and p is the probability that a particle, inside the region at one time instant is inside at the next time instant. The left-hand side of (1) may be estimated by

$$\sum_{i=1}^{k-1} (n_{i+1} - n_i)^2/(k - 1), \tag{2}$$

where k is the total number of instants at which counts were made. ν may be estimated by $\bar{n} = \sum_{i=1}^{k} n_i/k$ and therefore an estimate of p may be found. p is related to the speed of the particles and hence information may be gained about their movement. The advantage of the method is that it is not necessary to follow the motion of individual particles, indeed it is not even necessary to be able to identify particles from one instant to the next. Rothschild has suggested a modification of the method wherein the region is divided into sub-regions and counts made of the numbers in the sub-regions.

In the present paper we extend Smoluchowski's result to obtain the probability distribution of the stochastic process $\{n_i\}$ (section 2). In section 3 the relation between p and the motion is studied and leads to a method of estimation of the mean speed of the particles valid for a very wide class of types of motion (the only essential restrictions are independence and stationarity). In section 4 the distribution theory is used to obtain the standard error of the mean speed estimator for the case of one dimension and it is shewn

that for many series of observations the coefficient of variation is approximately $\sqrt{2/k}$: more general approximations are given for undeviated motion. In section 5 uniform undeviated motion is considered in more detail and the results compared with Rothschild's. Other applications are mentioned. It is hoped in a future paper to discuss the estimation of other characteristics of the distribution of speed besides the mean, and to describe some results that have been obtained with some data on the movement of pedestrians.

I am extremely grateful to Lord Rothschild for first drawing my attention to the problem, for many discussions about the work and for allowing me to see, in advance of publication, some results of his and Dr. Ruben's. Some of the results in this paper overlap those obtained by Ruben. This was not realized at the time the paper was read (September 1954).

2. *Distribution theory*

We consider particles in motion throughout a region which, in the theory, we shall suppose infinite: in practice it will be large compared with a region A wherein the particles are to be observed. For the region A, counts are to be made of the total number of particles within it at k equally spaced intervals of time $t, t + \tau, t + 2\tau, \ldots t + (k-1)\tau$. From these counts $\{n_i\}$ $(i = 1, 2, \ldots k)$ and the dimensions of the region A, and from no other observations, but utilizing certain assumptions about the motions of the particles, it is desired to estimate the speed of the particles. That this is possible may be seen intuitively by the consideration that if the particles move slowly the n_i will change slowly, if they move rapidly the n_i will change rapidly.

We assume the particles move independently of one another. One other assumption about their motion will be introduced later.

We find the cumulant generating function (c.g.f.) of $\{n_i\}$ $(i = 1, 2, \ldots k)$. In the exposition we confine ourselves to the case $k = 3$: the general case follows similarly.

We introduce notation exemplified by the following:

$n_{100} \equiv$ the number of particles inside A, outside A and outside A at three successive instants of observation.

$n_{1\cdot 0} \equiv$ the number of particles inside A at one time instant and outside A two time instants later, no statement being made about where the particles were at the intermediate time instant.

Thus a suffix 1, refers to inside A; a suffix 0 to outside A; and a dot in place of a suffix means that no statement is made about the position. We require the c.g.f. of $n_{1\cdot\cdot}, n_{\cdot 1\cdot}, n_{\cdot\cdot 1}$, previously denoted by n_1, n_2, n_3.

Our remaining assumption about the motion is that the distributions of n_{ijk} (suffixes either 0 or 1 but not all 0) are all stationary Poisson distributions. Sufficient conditions for this to be true have been given by Doob (6, VIII,

§ 5) for $n_{1}..$, similar ideas extend to the other quantities. By the former assumption these quantities are independent. Since there are supposed to be a large number of particles, moving independently, and the probability of any one of them, for example staying in A during three successive observations, is small, the Poisson distribution follows as the usual limit from the binomial. The joint motion of the particles must be such as to keep this stationary. The individual particles do not need to have stationary distributions, for example they may have a steady drift, but if so there must be a steady replacement of particles drifting in to replace those drifting out.

Let $p(v)$ $\equiv$ probability that a particle, inside A at time t, will be inside at time $t+v$;

and $p(v, w)$ $\equiv$ probability that a particle, inside A at time t, will be inside at times $t+v$ and $t+v+w$.

By the stationarity these quantities do not depend on t.

It is important to note that these definitions only refer to the position of the particles at certain times, and not to their motions between these times. Thus a particle may have left A and returned to it in $(t, t+v)$ and yet the event, whose probability is $p(v)$ will have then occurred. Our $p(\tau)$ is Chandrasekhar's $1-P$ though his definition of P ("the probability that a particle somewhere inside A will have emerged from it during the time τ") is obscure, if not misleading. He uses it correctly. Rothschild's definition and use is incorrect though it is probably a good approximation. We discuss it in detail in section 5.

Now $n_{1}.. = n_{111} + n_{110} + n_{101} + n_{100}$, and $n_{\cdot 1}.$, $n_{\cdot\cdot 1}$ similarly.

The c.g.f. required is

$$\log \mathscr{E}\{\exp\,(n_{1}..z_{1} + n_{\cdot 1}.z_{2} + n_{\cdot\cdot 1}z_{3})\},$$

and the expression in brackets therefore equals

$$\begin{aligned} &n_{111}(z_{1}+z_{2}+z_{3}) + n_{110}(z_{1}+z_{2}) + n_{101}(z_{1}+z_{3}) \\ &\quad + n_{011}(z_{2}+z_{3}) + n_{100}z_{1} + n_{010}z_{2} + n_{001}z_{3}. \end{aligned} \tag{3}$$

The random variables in this expression are independent Poisson variables and therefore the c.g.f. required is the sum of their c.g.f.'s. It only remains to find the means of these Poisson variables.

For fixed $n_{1}..$, n_{111} is binomial with index $n_{1}..$ and parameter $p(\tau, \tau)$, but $n_{1}..$ is Poisson with mean ν, say, the average number of particles in A.

Hence $$\mathscr{E}(n_{111} \mid n_{1}..) = n_{1}..\, p(\tau, \tau)$$

and $$\mathscr{E}(n_{111}) = \nu p(\tau, \tau).$$

Similarly

$$\mathscr{E}(n_{110}) = \mathscr{E}(n_{11\cdot} - n_{111}) = \nu\{p(\tau) - p(\tau, \tau)\},$$

$$\mathscr{E}(n_{101}) = \mathscr{E}(n_{1\cdot 1} - n_{111}) = \nu\{p(2\tau) - p(\tau, \tau)\},$$

$$\mathscr{E}(n_{100}) = \mathscr{E}(n_{1}.. - n_{110} - n_{101} - n_{111}) = \nu\{1 - p(\tau) - p(2\tau) + p(\tau, \tau)\}.$$

By the stationarity

$$\mathscr{E}(n_{110}) = \mathscr{E}(n_{011}),\ \mathscr{E}(n_{100}) = \mathscr{E}(n_{001}),$$

and

$$\mathscr{E}(n_{010}) = \mathscr{E}(n_{\cdot 1\cdot} - n_{111} - n_{110} - n_{011}) = \nu\{1 - 2p(\tau) + p(\tau, \tau)\}.$$

The c.g.f. of a Poisson variable with mean μ is $\mu(e^z - 1)$. Substitution of these values in (3) gives, after a little elementary algebra which reveals the general form of the expressions, the c.g.f. as

$$\begin{aligned}&\nu\{p(\tau, \tau)(e^{z_1} - 1)(e^{z_2} - 1)(e^{z_3} - 1) + p(\tau)(e^{z_1} - 1)(e^{z_2} - 1)\\ &+ p(\tau)(e^{z_2} - 1)(e^{z_3} - 1) + p(2\tau)(e^{z_1} - 1)(e^{z_3} - 1) + e^{z_1} + e^{z_2} + e^{z_3} - 3\}.\end{aligned} \tag{4}$$

The substitution $\theta_i = e^{z_i} - 1$ gives the logarithm of the factorial moment generating function as

$$\begin{aligned}&\nu\{p(\tau, \tau)\theta_1\theta_2\theta_3 + p(\tau)\theta_1\theta_2 + p(\tau)\theta_2\theta_3\\ &\quad + p(2\tau)\theta_1\theta_3 + \theta_1 + \theta_2 + \theta_3\}.\end{aligned} \tag{5}$$

The general form of (5) is ν times a linear function of products of θ_i's (all different). For any such product the coefficient is determined by writing it as $\theta_{i_1}\theta_{i_2}\ldots\theta_{i_k}$ with $i_1 < i_2 < \ldots < i_k$, whence the coefficient may be written $p((i_2 - i_1)\tau,\ (i_3 - i_2)\tau, \ldots (i_k - i_{k-1})\tau)$, where

$p(\tau_1, \ldots \tau_{k-1})$ $\equiv$ probability that a particle, inside A at time t, will be inside at times $t+\tau,\ t+\tau_1+\tau_2, \ldots t+\tau_1+\tau_2+\ldots+\tau_{k-1}$.

The distribution, whose c.g.f. is given by (4), is a multivariate Poisson [1]) distribution. Though it does not seem possible to write down the general form of the distribution at all succinctly, any cumulant can be written down at sight, thus

$$\varkappa_{rostooouov} = \nu p(2\tau,\ \tau,\ 4\tau,\ 2\tau) = \nu p_{2142}, \tag{6}$$

say, where r, s, t, u, v are positive. The number of variables in $p(.)$, and their values, depending on the distance apart of successive non-zero suffixes of the cumulant.

In particular

$$\varkappa_{11} = \nu p(\tau),\ \varkappa_{101} = \nu p(2\tau),\ \varkappa_{1001} = \nu p(3\tau) \text{ etc.},$$

and

$$\varkappa_1 = \nu,$$

so that $p_k = p(k\tau)$ is the autocorrelation function of the stochastic process. Smoluchowski's result (1) follows immediately.

---

[1]) The name is justified because not only are the marginal distributions Poisson, but it shares with the univariate Poisson distribution the property of being infinitely divisible.

By considering the joint distribution of the standardized random variables, $(n_i - \nu)/\nu^{\frac{1}{2}}$ $(i = 1, 2, \ldots k\}$, it follows immediately from (5) that their joint distribution tends, as $\nu \to \infty$, to the normal form with zero means, unit variances and autocorrelation function $p(k\tau)$. Thus for large ν, ν and $p(k\tau)$ $(k = 1, 2, \ldots)$ describe the process completely.

It is worth noting that the process is not, in general, Markovian. Thus if the number of particles within A is known at one time instant, it will help the prediction of the future development of the process to know in addition the number at the previous time instant: for if the latter is, for example, smaller than the former it means that several particles have just entered A and, with most types of motion, this will influence their prospects of future departure. For the special type of motion considered below, section 5, the process will be Markovian of high order if there is a lower positive bound on the velocity, because then a particle cannot remain in A for more than a certain maximal period of time, and in a further special case will be Markov of order one.

3. *The Autocorrelation Function*

So far no assumptions have been made about the motion of the particles other than that they move independently and that certain distributions are stationary Poisson. The resulting joint distribution of the number of particles counted at successive time instants depends on $p(t)$, $p(t, t), \ldots$ and the evaluation of these functions depends on more definite assumptions about the motion. The most important of these functions is $p(t)$, the autocorrelation function of the continuous time stochastic process $\{n(t)\}$, where $n(t)$ is the number of particles inside A at time t. It has been shewn that for large ν, ν and $p(t)$ completely specify the process. We can extend the definition of $p(t)$ to negative argument, whence $p(t) = p(-t)$.

The expression for $p(t)$ is easily seen to be

$$p(t) = \int_A dx \int_A dy \,.\, \varrho(x) f\left(\frac{y - x}{t},\ t;\ x\right) / t, \tag{7}$$

where x and y denote points of A,

$\varrho(x)$ $\equiv$ frequency distribution of a randomly observed particle within A,

$f(u, t; x)$ $\equiv$ frequency distribution of the average velocity, u, of a particle at x, during a subsequent time interval of length t.

By our assumption of stationarity neither ϱ nor f will depend on the time at which the interval in question begins.

As before, let $\{n_i\}$ denote the observed sequence of counts. From observations on n_i over several instants, the correlogram can be calculated as an estimator of $p(t)$, and hence some information about f is obtained from (7).

This is the basic idea behind the present paper. It is remarkable, however, that the mean speed of the particles is related very simply to the derivative of the autocorrelation function at the origin, $p'(0)$. Strictly this is the derivative to the right $p'(+0) = -p'(-0)$; the notation will always refer to this derivative. The estimation of this derivative therefore enables estimation of what is perhaps the most important characteristic of the motion, the mean speed, to be made. We proceed to exhibit this relationship in the case of one-dimensional motion. The motion can be supposed to take place along a line and we suppose the region A to be an interval $(0, a)$ on this line.

(7) may then be written

$$p(t) = \int_0^a dx \int_0^a dy \,.\, \varrho(x) f\left(\frac{y-x}{t},\ t;\ x\right)/t \tag{8}$$

The substitution $u = (y-x)/t$ gives,

$$p(t) = \left\{\int_0^{a/t} du \int_0^{a-ut} dx + \int_{-a/t}^{0} du \int_{-ut}^{a} dx\right\} \varrho(x) f(u,\ t;\ x).$$

If
$$g(u,\ t;\ z) = \int^z dx \,.\, \varrho(x) f(u,\ t;\ x),$$

$$p(t) = \int_0^{a/t} du[g(u,\ t;\ a-ut) - g(u,\ t;\ 0)]$$
$$+ \int_{-a/t}^{0} du[g(u,\ t;\ a) - g(u,\ t;\ -ut)].$$

Differentiation with respect to t, yields

$$p'(t) = \int_0^{a/t} du[g^{(2)}(u,\ t;\ a-ut) - g^{(2)}(u,\ t;\ 0) - ug^{(3)}(u,\ t;\ a-ut)]$$
$$+ \int_{-a/t}^{0} du[g^{(2)}(u,\ t;\ a) - g^{(2)}(u,\ t;\ -ut) + ug^{(3)}(u,\ t;\ -ut)],$$

where $g^{(i)}(.,\ .;\ .)$ denotes the result of differentiation with respect to the i^{th} variable in $g(.,\ .;\ .)$.

If now $t \to 0$, and the functions are continuous,

$$p'(0) = \int_{-\infty}^{\infty} du[g^{(2)}(u,\ 0;\ a) - g^{(2)}(u,\ 0;\ 0)]$$
$$- \int_0^{\infty} du \,.\, u\varrho(a) f(u,\ a) + \int_{-\infty}^{0} du \,.\, u\varrho(0) f(u;\ 0), \tag{9}$$

where $f(u;\ x) = \lim_{t \to 0} f(u,\ t;\ x)$, the probability distribution of the velocity of a particle at x.

The first integral in (9) is, with $t = 0$,

$$\int_{-\infty}^{\infty} du \int_{0}^{a} dx \,.\, \varrho(x) \frac{d}{dt} f(u,\ t;\ x)$$

$$= \int_{0}^{a} dx \,.\, \varrho(x) \frac{d}{dt} \int_{-\infty}^{\infty} du \,.\, f(u,\ t;\ x),$$

which vanishes since the integral with respect to u is 1. Consequently

$$p'(0) = \varrho(0) \int_{-\infty}^{0} du \,.\, uf(u,\ 0) - \varrho(a) \int_{0}^{\infty} du \,.\, uf(u,\ a). \tag{10}$$

Consider one of the ends of the interval, say that at 0. Analogously to (8) we can write down the average number of particles which are initially to one side, say $x < 0$, and after a lapse of time t, are on the other side, $x > 0$, as

$$\int_{0}^{\infty} dy \int_{-\infty}^{0} dx \,.\, \nu(x) f\left(\frac{y - x}{t},\ t;\ x\right) / t \tag{11}$$

where $\nu(x)$ is the density of particles at x. $\nu(x)$ and $\varrho(x)$ are related by the equation

$$\varrho(x) = \nu(x) / \int_{0}^{a} dx \,.\, \nu(x) \tag{12}$$

Manipulating (11) in a similar manner to (8), allowing t to tend to 0, we obtain for the rate of passage of particles *into* A across the end at $x = 0$, the expression

$$\nu(0) \int_{0}^{\infty} du \,.\, uf(u,\ 0).$$

Similar expressions obtain for the rate out of A at $x = 0$ and both rates at $x = a$. The rates in and out must balance for stationarity and we therefore have [1])

$$\begin{aligned} &\nu(0) \int_{0}^{\infty} du \,.\, uf(u,\ 0) - \nu(a) \int_{-\infty}^{0} du \,.\, uf(u,\ a) \\ &= -\nu(0) \int_{-\infty}^{0} du \,.\, uf(u,\ 0) + \nu(a) \int_{0}^{\infty} du \,.\, uf(u,a). \end{aligned} \tag{13}$$

(10), (12) and (13) give the result,

---

[1]) Equation (13) is a generalization of a well-known result in the kinetic theory of gases. It says that the density at a point is inversely proportional to the mean velocity.

$$-2p'(0)\int_0^a dx\,.\,v(x) = v(0)\int_{-\infty}^{\infty} du\,.\,|u|f(u,\,0) + v(a)\int_{-\infty}^{\infty} du\,.\,|u|f(u,\,a)$$

or
$$-2p'(0)\int_0^a dx\,.\,v(x) = v(0)c(0) + v(a)c(a), \tag{14}$$

where $c(x)$ is the mean *speed* of the particles at x.

This is the desired relationship. If the density and mean speed do not depend on x it simplifies to give,

$$p'(0) = -c/a. \tag{15}$$

It is not possible to obtain an estimate of $p'(0)$ from obsevrations at discrete times, but if τ is small we can use the approximation based on a Maclaurin series of

$$p(\tau) = 1 + \tau p'(0) \tag{16}$$

for small τ. A heuristic derivation of (15) has been obtained earlier by Ruben.

There are now several ways of proceeding. Smoluchowski's original equation (1) can be used and we have the estimator

$$\hat{c} = \frac{a}{2\tau}\left\{\frac{1}{k-1}\sum_{i=1}^{k-1}(n_{i+1} - n_i)^2 \,\middle|\, \frac{1}{k}\sum_{i=1}^{k} n_i\right\},$$

or if

$$C_s = \frac{1}{k-1}\sum_{i=1}^{k-s}(n_i - \bar{n})(n_{i+s} - \bar{n}),$$

the autocovariance of lag s, this is approximately

$$\hat{c} = \frac{a}{\tau}\,.\,\frac{C_0 - C_1}{\bar{n}} \tag{17}$$

where
$$\bar{n} = \sum_{i=1}^{k} n_i/k.$$

An alternative method is to estimate $p(\tau)$ by the auto-correlation coefficient $r_1 = C_1/C_0$, whence

$$\hat{c} = \frac{a}{\tau}\left(1 - \frac{C_1}{C_0}\right). \tag{18}$$

A third possibility is not to use C_0 but $\bar{n}$; these estimate the same quantity, ν, in virtue of the Poisson property. In (17) or (18) this yields the estimator

$$\hat{c} = \frac{a}{\tau}\left(1 - \frac{C_1}{\bar{n}}\right). \tag{19}$$

The merits of these three estimators, as judged by their standard errors will now be discussed.

Note that, apart from the assumptions of independence and stationarity, the only restrictions placed on the motion are that the distributions involved are sufficiently respectable for the analytical manipulations made above to be valid. Use of (16), though, involves the introduction of a bias depending on τ and $p''(0)$.

4. *Standard Errors*

The standard errors of the three estimates given above may be found from the formulae given by Bartlett (1) for the variances and covariances of the autocovariances. Throughout the discussion we shall only consider the variances to order k^{-1}, that is our expressions will hold for large k. To this order it is immaterial whether ν or $\bar{n}$ appears in the expression for C_s; though it does matter which appears in the denominator of (17).

Bartlett gives (equation (5) of (1))

$$\operatorname{cov}(C_s, C_{s+t}) = \frac{1}{k}\sum_{i=-\infty}^{+\infty}[\nu^2(p_i p_{i+t} + p_{i-s}p_{i+s+t}) + \varkappa_{i,s,t}],$$

where $\varkappa_{i,s,t}$ is the first cumulant involving the four variables n_1, n_{1+s}, n_{1+i} and $n_{1+i+s+t}$. In deriving this expression another assumption has been introduced, namely that the total period of observation, $k\tau$, is large compared with the effective length of the correlogram; that is the interval within which $p(t)$ is markedly non-zero. This assumption is needed in replacing expressions like

$$\sum_{s=0}^{k}\left(1 - \frac{s}{k}\right)p_s,$$

which occur in the calculation of the covariances,

by

$$\sum_{s=0}^{\infty} p_s.$$

In most types of movement there will exist a value t_0, such that the probability that a particle stays in A for a time longer than t_0 is small. For the application of Bartlett's formulae, and hence ours, it is essential that $k\tau$ be large compared with t_0.

Now, to the same order of approximation

$$\begin{aligned}\operatorname{var}\left(\frac{C_0 - C_1}{\bar{n}}\right)\nu^2 = {} & \operatorname{var}(C_0) - 2\operatorname{cov}(C_0, C_1) + \operatorname{var}(C_1) \\ & - 2(1 - p_1)(\operatorname{cov}(C_0, \bar{n}) - \operatorname{cov}(C_1, \bar{n})) \\ & + (1 - p_1)^2 \operatorname{var}(\bar{n}).\end{aligned}$$

By arguments similar to Bartlett's we obtain

$$\operatorname{cov}(C_0, \bar{n}) = \frac{\nu}{k} \sum_{-\infty}^{+\infty} p_i,$$

$$\operatorname{cov}(C_1, \bar{n}) = \frac{2\nu}{k}\left(p_0 + \sum_{i=1}^{\infty} p_{1i}\right), \text{ and}$$

$$\operatorname{var}(\bar{n}) = \frac{\nu}{k} \sum_{-\infty}^{\infty} p_i.$$

After some algebra we have the result,

$$\operatorname{var}(\hat{c})\tau^2/a^2 = \frac{1}{k}\left\{\sum_{-\infty}^{\infty}\left(3p_i^2 - 4p_ip_{i+1} + p_ip_{i+2}\right)\right\} \tag{20}$$

$$+ \frac{1}{k\nu}\left\{p_1^2 \sum_{-\infty}^{\infty} p_i - 4p_1\left(p_1 + \sum_{i=1}^{\infty} p_{1i}\right) + \left(p_1 + 2p_{11} + 2\sum_{i=1}^{\infty} p_{1i1}\right)\right\}$$

for the variance of the estimator (17) employing Smoluchowski's original result.

This expression involves knowledge of the whole of the correlogram, and more if the second term is to be used. We can, however, shew that the dependence on the correlogram away from the origin is very slight when τ is small. To do this we obtain an alternative expression for (20) when τ is small.

Consider the first term in (20). Let

$$\varphi(\tau) = \int_{-\infty}^{\infty} dt \,.\, p(t)p(t+\tau),$$

and suppose that $p(t)$ has continuous derivatives up to the third order in $(0, \infty)$. Note that the odd derivatives are necessarily discontinuous at the origin since $p(-t) = p(t)$.

Now

$$\varphi(\tau) = 2\int_0^{\infty} dt \,.\, p(t)p(t+\tau) + \int_0^{\tau} dt \,.\, p(t)p(\tau - t), \tag{21}$$

and both integrands now have continuous derivatives up to the third order over their ranges of integration. Hence the Euler-Maclaurin expansion of $\varphi(\tau)$ may be found at intervals distant τ apart.

After some algebra we obtain

$$\varphi(\tau) = \tau \sum_{-\infty}^{\infty} p_ip_{i+1} + \tfrac{1}{3}\tau^2 p(\tau)p'(0) + O(\tau^4)$$

$$= \tau \sum_{-\infty}^{\infty} p_ip_{i+1} + \tfrac{1}{3}\tau^2 p'(0) + \tfrac{1}{3}\tau^3 p'(0)^2 + O(\tau^4).$$

Similarly,

$$\varphi(0) = \tau \sum_{-\infty}^{\infty} p_i^2 + \tfrac{1}{3}\tau^2 p'(0) + O(\tau^4),$$

$$\varphi(2\tau) = \tau \sum_{-\infty}^{\infty} p_i p_{i+2} + \tfrac{1}{3}\tau^2 p'(0) + \tfrac{2}{3}\tau^3 p'(0)^2 + O(\tau^4).$$

These expressions for integrals in terms of sums may be used to replace the sums in the first member of (20) by integrals. The expression in braces is τ^{-1} times

$$3\varphi(0) - 4\varphi(\tau) + \varphi(2\tau) + \tfrac{2}{3}\tau^3 p'(0)^2 + O(\tau^4).$$

An expansion of $\varphi(\tau)$ by Maclaurin's theorem gives, on observing that $\varphi'(0) = 0$,

$$\tfrac{2}{3}\tau^3 \varphi'''(0) + \tfrac{2}{3}\tau^3 p'(0)^2 + O(\tau^4).$$

$\varphi'''(0)$ can be evaluated from (21), and equals $2p'(0)^2$.

Hence ignoring the second term in (20), that is supposing ν is large, and dropping the term $O(\tau^4)$,

$$\text{var } (\hat{c}) = \frac{a^2}{\tau^2}\frac{1}{k\tau} 2\tau^3 p'(0)^2 = 2c^2/k, \tag{22}$$

by (15). This formula is only valid, we repeat, when k is large, τ small, but $k\tau \gg t_0$ (see above) and ν is large. This is possibly the most important case. The coefficient of variation is then $\sqrt{2/k}$, independent of c. A series of 200 observations is therefore necessary in order to obtain a coefficient of variation of 10 %. This gives an easily understood idea of the accuracy of the method.

If ν is small, that is the region contains few particles, the second term in (20) will have to be included. This involves the functions $p(t, u)$ and $p(t, u, v)$ in addition to $p(t)$. An approximation for this expression can be found, involving the derivatives of these functions, along similar lines to those used above. For the present we shall treat of a special case only, where the motion of a particle is always in the same direction on the line, though this direction may vary between particles. In this situation a particle can visit the interval only once and therefore $p(t, u) = p(t + u)$ and $p(t, u, v) = p(t + u + v)$. Consequently, even if ν is not large, the variance of $\hat{C}$, in (20), still depends on the correlation function alone. We shall refer to this as "undeviated" motion. The second term in (20) is then

$$p_1^2 \sum_{-\infty}^{+\infty} p_i - 2p_1\Big(\sum_{-\infty}^{\infty} p_i - 1\Big) + \Big(\sum_{-\infty}^{\infty} p_i - 1 - p_1\Big) \tag{23}$$

divided by $k\nu$. This expression equals

$$\Big\{\sum_{-\infty}^{\infty} p_i(1 - p_1) - 1\Big\}(1 - p_1). \tag{24}$$

By the Euler-Maclaurin expansion,

$$P \equiv \int_{-\infty}^{\infty} p(t)dt = \tau \sum_{-\infty}^{\infty} p_i + O(\tau^2),$$

and by Maclaurin's theorem

$$1 - p_1 \equiv 1 - p(\tau) = -\tau p'(0) - \tfrac{1}{2}\tau^2 p''(0) + O(\tau^3).$$

Thus (24) is

$$[(P + O(\tau^2))(-\tau p'(0) - \tfrac{1}{2}\tau^2 p''(0) + O(\tau^3)) - \tau] \times [-p'(0) - \tfrac{1}{2}\tau p''(0) + O(\tau^2)]$$

which gives

$$\tau p'(0)(1 + Pp'(0) + \tfrac{1}{2}\tau^2 p''(0)[1 + 2Pp'(0)] + O(\tau^3).$$

The complete expression for the variance of $\hat{C}$, derived from (20), with k large, τ small, but $k\tau \gg t_0$, but without ν being large, is, for undeviated motion

$$\frac{2c^2}{k} + \frac{c}{k\tau\nu}(Pc - a) + \frac{a}{2k\nu} p''(0)(a - 2Pc). \tag{25}$$

The method of estimation has ignored $p''(0)$ in using (16) so that the bias is at least comparable with the last term in (25). Although in many applications $p''(0) = 0$ the magnitude of this bias is worth noting. However large k is, there is a bias due to the omission of the second derivative in the Maclaurin expansion: this is of magnitude $-a\tau p''(0)/2$, aud is always negative. There will also be a bias resulting from the ratio of two quantities, namely $C_0 - C_1$ and $\bar{n}$. This can be found, to order k^{-1}, on exactly similar lines to those employed in finding the variance. We have

$$\mathscr{E}\left(\frac{a}{\tau} \cdot \frac{C_0 - C_1}{\bar{n}}\right) = \frac{a}{\tau}\left[(1 - p_1) - \frac{1}{\nu^2}\operatorname{cov}(C_0 - C_1, \bar{n}) + \frac{(1 - p_1)}{\nu^2}\operatorname{var}(\bar{n})\right].$$

The second and third terms give $a/k\tau\nu$ times

$$2(p_1 + \sum_{i=1}^{\infty} p_{1i}) - p_1 \sum_{-\infty}^{\infty} p_i,$$

which, for undeviated motion where $p_{1i} = p_{1+i}$, is

$$\sum_{-\infty}^{\infty} p_i(1 - p_1) - 1,$$

which has been considered before (24). Retaining only the dominant terms we have

$$\mathscr{E}\left(\frac{a}{\tau} \cdot \frac{C_0 - C_1}{n}\right) = c - a\tau p''(0)/2 + \frac{1}{k\tau\nu}(Pc - a). \tag{26}$$

Other terms may, of course, be calculated if required. In any application this bias should be considered in addition to the variance.

Consider now the second estimator (18) based directly on the autocorrelation coefficient. By an exactly similar argument to that used before we obtain in place of (20) the result

$$\text{var } (\hat{c})\tau^2/a^2 = \frac{1}{k}\left\{ \sum_{-\infty}^{\infty} (p_i^2(1 + 2p_1^2) + p_{i-1}p_{i+1} - 4p_1p_ip_{i+1}) \right\}$$

for the dominant term: the second term is unaltered. By the Euler-Maclaurin expansion the expression in braces is τ^{-1} times

$$\varphi(0)(1 + 2p_1^2) - 4p_1\varphi(\tau) + \varphi(2\tau) + O(\tau^2),$$

and with Maclaurin's theorem we obtain the result

$$2\varphi(0)(1 - p_1)^2 + O(\tau^2).$$

Hence, under the same conditions as used in deriving (22),

$$\text{var } (\hat{c}) = \frac{a^2}{\tau^2} \cdot \frac{1}{k\tau} \cdot 2\varphi(0)\tau^2p'(0)^2 = \frac{2c^2}{k} \cdot \frac{\varphi(0)}{\tau}$$

For the range of values of τ for which this expression is valid, $\varphi(0)$ is approximately $\tau \sum_{-\infty}^{\infty} p_i^2 = \tau + 2\tau \sum_{1}^{\infty} p_i^2 \gg \tau$. Hence this estimator is worse than Smoluchowski's because of its larger variance, and should not be used. Similar considerations apply to the estimator (19) which an argument along the same lines shows to have even larger variance.

We have not investigated whether Smoluchowski's estimate is the 'best' in the sense of having least variance, nor have we investigated the optimal choice of τ to obtain the least variance. Both these questions would presumably have answers dependent on the exact form of $p(t)$ which is generally unknown: indeed if it was well-known there would be little point in the analysis. It is important to note that the estimator and its standard error do not depend, to the first order of approximation, on any knowledge of the autocorrelation function. For values of k and τ for which our expression (24) is valid it is clear that the choices of τ or a are not critical, since the dominant term in the variance is not dependent on them.

5. *Undeviated Uniform Motion.*

The case of undeviated motion, where a particle does not change its direction, has been mentioned. We now consider a special case of such motion where not only is the direction unaltered but the velocity of a particle is fixed for that particle though it varies from particle to particle with frequency distribution

$f(V)$ [1]). We refer to this as undeviated uniform motion. The function f appearing in (8) does not depend on t or x. By (13) the density will be uniform and so (8) becomes

$$p(t) = \int_0^a dx \int_0^a dy \,.\, f\left(\frac{y-x}{t}\right)/at,$$

and if $v = (y-x)/t$,

$$p(t) = \left\{\int_0^{a/t} dv \int_0^{a-vt} dx + \int_{-a/t}^0 dv \int_{-vt}^a dx\right\} f(v)/a$$

$$= \int_0^{a/t} dv \,.\, (a - vt) f(v)/a + \int_{-a/t}^0 dv \,.\, (a + vt) f(v)/a.$$

Hence if $g(w)$ denotes the distribution of the particles *speed*, i.e. $g(w) = f(w) + f(-w)$, $w > 0$,

$$p(t) = \int_0^{a/t} dw \left(1 - \frac{wt}{a}\right) g(w). \tag{27}$$

Then

$$p'(t) = -\int_0^{a/t} dw \,.\, wg(w)/a, \tag{28}$$

and in particular

$$p'(0) = -\int_0^{\infty} dw \,.\, wg(w)/a = -c/a$$

confirming our previous analysis.
We also have on differentiating again,

$$p''(t) = ag(a/t)/t^3, \tag{29}$$

which is the inversion formula corresponding to (27).

A useful result is that

$$P = \int_{-\infty}^{\infty} p(t)dt = a\int_0^{\infty} dw \,.\, w^{-1}g(w). \tag{30}$$

Hence $P = a\mathscr{E}(w^{-1})$ is approximately $a\left(\frac{1}{c} + \frac{\mathrm{var}(w)}{c^3}\right)$ and the second term in the error, (25), is $a \,.\, \mathrm{var}\,(w)/k\tau\nu c$, to the same order of approximation.

Suppose that there is a lower bound w_0 on the speed of the particles; then from (27) $p(t) = 0$ for $t > a/w_0$. Hence the quantity a/w_0 plays the role of t_0 mentioned above and our formula for the error will be valid provided that the

---

[1]) This is the frequency distribution of velocity of a random observed particle within A.

total time of observation is considerably greater than t_0, the time taken for the slowest particle to pass through the region.

If, on the other hand, there is an upper bound to the speed w_1; then, again from (27),

$$p(t) = 1 - ct/a$$

for $t < a/w_1$. Hence the approximation based on a Maclaurin expansion (16) is exact provided $\tau \leqq a/w_1$, the time taken for the fastest particle to pass through the region. It follows from (29) that provided $g(w)$ has close contact with the axis $g = 0$ for large w the value of $p''(0)$ will be zero: often $g(w) = 0(e^{-bw})$, $b > 0$ and this condition is well-satisfied. The bias is therefore only likely to cause trouble when some particles have large speeds.

We now consider some special cases. The simplest such is where all particles have the same speed [1]) c.

We have

$$\begin{aligned} p(t) &= 1 - ct/a, && \text{for } |t| \leqq a/c, \\ &= 0 \quad , && \text{for } |t| \geqq a/c. \end{aligned}$$

and $w_0 = w_1 = c$. It is immediately apparent that $p(t)$ has not got continuous derivatives up to the third order in $(0, \infty)$; a condition that we have assumed in deriving our error formulae. It is necessary to take into account the discontinuities at $t = a/c$, but a detailed analysis including these shews that the same formulae persist. It is necessary that $\tau < a/c$. When this holds the bias is zero, and since $P = a/c$ and $p''(0) = 0$ the two terms of order ν^{-1} in (25) disappear and the statement that the coefficient of variation is $\sqrt{2/k}$ is valid even for small ν. This seems to be the situation considered by Fürth.

Suppose next that

$$g(w) = \frac{1}{r}\frac{1}{s!}\left(\frac{w}{r}\right)^s e^{-\frac{w}{r}}, \text{ for } w \geqq 0,$$

a Pearson Type III distribution. The mean speed is $(s+1)r = c$, $P = a/rs$, from (30), and

$$s!\, p(t) = \int_0^{a/rt} dx \,.\, \left(1 - \frac{xrt}{a}\right) x^s e^{-x}. \tag{31}$$

For small t the exponential term, $e^{-a/rt}$ dominates the behaviour of $p(t)$; $p''(0) = 0$ and any bias will be small. For large t, $p(t) = 0(t^{-(s+1)})$, so that there is not exponential contact with $p = 0$. For small s a long series of observations

[1]) But not necessarily the same velocity, they may move in either direction, that is with velocities $\pm c$.

will be necessary: this corresponds to the presence of slow-moving particles. The formula (25) for the variance is

$$\frac{2c^2}{k} + \frac{ca}{k\tau vs}.$$

As a final possibility consider, with $s > 0$

$$g(w) = \frac{r}{s!}\left(\frac{1}{rw}\right)^{s+2} e^{-\frac{1}{wr}}, \text{ for } w \geqq 0.$$

This corresponds to assuming that the time that a particle stays in A has a Type III distribution. The mean speed is $c = 1/rs$, $P = ar(s + 1)$, from (30), and

$$s!p(t) = \int_{t/ra}^{\infty} \left(1 - \frac{t}{rax}\right) x^s e^{-x} dx,$$

which gives

$$p(t) = P_{s-1}\left(\frac{t}{ra}\right) e^{-t/ra}, \tag{32}$$

where $P_s(x)$ is a polynomial in x of degree s. The behaviour is opposite to the previous case, near the origin the polynomial term dominates and $p(t) = 0(t^{s-1})$. For $s \leqq 1$ there will be a marked bias, since $p''(0) \neq 0$, and for moderate $s > 1$, even though $p''(0) = 0$, there will be a slight bias unless τ is very small. For large t the exponential term dominates and the correlogram rapidly dies away, so that a shorter series of observations would be needed than in the previous case: this corresponds to the presence of fast moving particles. The following rule would seem worth following: if there are some fast particles (relative to the mean speed) try to make τ small, if there are slow particles concen trate on expending the total time, $k\tau$, of observation, without of course increasing τ so much that the bias becomes important.

The formula (25) for the variance is, for $s > 1$,

$$\frac{2c^2}{k} + \frac{ca}{k\tau vs}.$$

The case $s = 1$ is of special interest and worth discussing separately. We have

$$g(w) = w^{-3} e^{-\frac{1}{wr}} r^{-2}, \tag{33}$$

and

$$p(t) = e^{-t/ra}. \tag{34}$$

The process can be shewn, by arguments developed in fhe next paragraph, to be Markov of order one. The inversion formula (29) shews that this is the only

distribution of speed yielding a Markov process of order one, though other motions, not of this type, might possibly yield such a process. It is therefore the distribution appropriate to the theory developed by Rothschild. The ideas of this theory are, briefly, that the motion is such that the particles arrive in a Poisson process (as we have assumed), but that any particle has a probability $\lambda\delta t + 0(\delta t)$ of leaving the region in any time interval of length δt, where λ is a constant. It is then easy to show that $p(t) = e^{-\lambda t}$. Rothschild assumes that this description of the motion and its consequent autocorrelation will provide a reasonable approximation to most types of motion. In fact our analysis shews that for linear motion [1]) the velocity distribution (33), which has $p(t) = e^{-\lambda t}$, is one in which there are particles with high speeds. Other types of motion produce autocorrelation functions which are not of exponential type (e.g. (31) and (32)), whilst in the extreme case of no variability in speed the function is far from exponential.

The unusual nature of the distribution (33) may be appreciated by the following considerations [2]). The assumptions in Rothschild's theory amount to using the model in which the particles arrive at random and take a time t to cross the region, where t has a probability distribution $\lambda e^{-\lambda t}$. Now $t = a/w$ so that we could equivalently assume a distribution for w equal to

$$\lambda a w^{-2} e^{-\lambda a/w}, \tag{35}$$

which differs from (33). The explanation is that (35) is the distribution of speed of particles entering A: (33) is the distribution of a randomly observed particle within A. Fast and slow moving particles are counted equally in arriving at (35): the fast-moving particles have less chance of being observed than the slow-movers in arriving at (33). Indeed the chance of their being observed is proportional to their time of passage, consequently the distribution of speed of a randomly observed particle may be obtained by multiplying (35) by $t = a/w$ and modifying the constant: this gives (33). Notice that although (33) has a finite mean, (35) has not; this emphasizes the importance of the fast-moving particles. The special case is atypical in that $p''(0) \neq 0$. This introduces a bias and complicates the error formulae. The bias is partly avoided by Rothschild because he uses $(C_0 - C_1)/\bar{n}$ to estimate $1 - e^{-\lambda\tau}$, not $\lambda\tau$, where $\lambda = c/a$: but this introduces a possible bias when the autocorrelation function is not of this form. The approximate formula for the standard error can be avoided in this special case (at least as far as term $0(k^{-1})$) because the series occurring in (20) are summable. For example the first term in (20) is

$$(3 + x)(1 - x)^2/(1 + x)k,$$

[1]) Rothschild is concerned with the two-dimensional case.

[2]) The ideas in this and the next paragraph arose out of a conversation with Dr. Cox.

where $x = e^{-\lambda\tau}$. The second term is

$$x\,(1 - x)/kv$$

These results agree with Rothschild's.

This is perhaps an appropriate point to mention that the theory we have developed here has other applications. Consider an immigration-emigration process in which the particles arrive at a constant random rate and have independent lifetimes t with probability distribution $h(t)$. The arrival corresponds to entering the region A, the lifetime corresponds to the time of passage across A and $g(w)$ and $h(t)$ are connected by the relation

$$g(w) \propto w^{-3}h(a/w).$$

The above was the special case $h(t) = \lambda e^{-\lambda t}$. Our theory is the stationary theory of this process representing the number of particles. Another application is to textile research. Consider a length of material made up of several fibres. Suppose the left-hand ends of the fibres occur at a regular random rate along the length, and that their lengths t have a distribution $h(t)$. Then the correspondence is established by associating the left-hand ends with the arrivals and the lengths with the passage times. $n(t)$ is then the number of fibres at a point t along the length of material. The case $h(t) = \lambda e^{-\lambda t}$ is of special importance and Rothschild's theory may be more applicable than ours.

References

[1] Bartlett, M. S., Supp. J. roy. statist. Soc., **8**, 27—41 (1946).

[2] Chandrasekhar, S., Rev. mod. Phys. **15**, 1—89, Chapter III, (1943).

[3] Fürth, R., Physik. Zeits., **19**, 421—426, (1918).

[4] Fürth, R., Physik. Zeits., **20**, 21, (1919).

[5] Rothschild, Lord, J. expt. Biol., **30**, 178—199, (1953).

[6] Doob, J. L., Stochastic Processes. Wiley, New York, (1954).

Statistical Laboratory,
University of Cambridge.

SYMPOSIUM B

ALGEBRAIC GEOMETRY

under the care of H. D. KLOOSTERMAN

Honorary President: F. SEVERI

ÜBER DIE IDEALTHEORETISCHE GRUNDLEGUNG DER ALGEBRAISCHEN GEOMETRIE

W. GRÖBNER

Die Idealtheorie der Polynomideale ist durch die Arbeiten einer großen Anzahl hervorragendster Mathematiker der letzten 100 Jahre begründet worden, so vor allem durch Kronecker, Dedekind, Hilbert, E. Lasker, F. S. Macaulay, Emmy Noether und viele andere. Sie wurde zwar gewöhnlich im Anschluß an die Begriffe und Erfordernisse der algebraischen Geometrie entwickelt, aber sie ist in dieser Zeit zu einer so imposanten und in sich festgefügten Theorie erwachsen, daß ihr auch selbständig ohne Rücksicht auf ihre Anwendungen in der Geometrie ein großes Interesse gebühren würde.

Sicherlich ist die Idealtheorie der Polynomringe eine Theorie, die auf völlig klaren und eindeutigen Begriffen gegründet und mit klaren und wirkungsvollen Methoden ausgestattet ist; sie benötigt daher für ihr Bestehen und Gelten keiner Stütze von seiten einer anderen Wissenschaft. Man kann ihren Inhalt von der Geometrie völlig abstrahieren und auch ihr Hauptproblem kann ganz unabhängig von geometrischen Anschauungen so formuliert werden:

Es ist die Aufgabe gestellt, die Struktur der durch endlich viele transzendente oder algebraische Erweiterungen aus einem Grundkörper K ableitbaren Körper zu studieren, oder, was dasselbe bedeutet, die Struktur der Restklassenkörper der Primideale in Polynomringen [1]).

Die Schwierigkeiten dieses Problems kommen vor allem darin zum Ausdruck, daß die genauen Erweiterungen, die wir an einem gegebenen Körper K vollziehen müssen, um einen gegebenen Erweiterungskörper $\overline{K}$ zu konstruieren, in keiner Weise feststehen, sondern auf unübersehbare Weise variiert werden können. Wir können zu K der Reihe nach irgendwelche Elemente aus $\overline{K}$ adjungieren und erhalten nach endlich vielen Adjunktionen $\overline{K}$, wobei die Auswahl der zu adjungierenden Elemente völlig willkürlich ist.

Der rationale Körper $\overline{K} = K(x, y)$, wo x, y unabhängige Transzendente über K sind, kann z.B. durch sukzessive Adjunktion von x und y zu K konstruiert werden, aber auch etwa durch sukzessive Adjunktion von $u_1 = x^3y^5$

---

[1]) Vgl. einen ähnlichen Standpunkt bei E. Kähler, Sur la théorie des corps purement algébriques. Deuxième Colloque de Géométrie algébrique tenu à Liège les 9, 10, 11, et 12 juin 1952, CBRM, p. 69—82.

und $u_2 = x^2y^3$, oder auch durch sukzessive Adjunktion der 3 Größen $u_1 = x^2$, $u_2 = y^3$, $u_3 = x + y$ [2]).

Man kann unser Problem auch so formulieren: *Es sind alle Polynomideale zu bestimmen, deren Restklassenkörper zu einem gegebenen Körper $\overline{K}$ isomorph sind*: das ist eine Klasse von *birational äquivalenten Primidealen.*

Wenn man genau zusieht, erkennt man, daß diese Formulierung auch das eigentliche Anliegen der algebraischen Geometrie mit umschließt, denn diese ist ja auf das Ziel hin gerichtet, die „Geometrie auf einer algebraischen Mannigfaltigkeit" zu studieren, d.h. die charakteristischen Eigenschaften zu ermitteln, welche einem algebraischen Gebilde an sich zukommen und unabhängig sind von dem zufällig vorliegenden projektiven Modell dieser Mannigfaltigkeit. Damit dürfte ein von Emmy Noether immer wieder betonter Standpunkt eingenommen sein, daß man nämlich grundsätzlich die mathematischen Probleme von ihren unwesentlichen Zusammenhängen loslösen und ihr abstraktes logisches Gefüge bloßlegen soll, um den sicheren Weg zur Lösung zu finden.

Das soll aber nun nicht bedeuten, daß die geometrischen Veranschaulichungen als wertlos und überflüssig beiseite geworfen werden müssen; sie bilden eine Umkleidung der abstrakten logischen Theorie, durch welche uns diese gleichsam erst zu vollem Leben erweckt und mit jener eigentümlichen Schönheit ausgestattet wird, die unseren Geist gefangen nimmt: etwa so wie wir bei der Erforschung der geologischen Struktur einer Landschaft von der Schönheit ihrer Natur abstrahieren, ohne damit zu sagen, daß wir in unserem Leben darauf zu verzichten gewillt wären.

Wenn wir noch einen Augenblick bei diesem Bilde bleiben dürfen, so könnten wir im vorliegenden Fall unsere Aufgabe dahin umschreiben, daß wir für die abstrakt logischen Begriffe und Methoden, die in der Idealtheorie ausgebildet vorliegen, die entsprechenden geometrischen Begriffe und Veranschaulichungen ausforschen müssen, die dazu passen. Der wichtigste Begriff aber, von dem das meiste abhängt, ist derjenige der *algebraischen Mannigfaltigkeit.*

Der Begriff der algebraischen Mannigfaltigkeit.

Wir beschränken uns hier grundsätzlich auf den Standpunkt der klassischen algebraischen Geometrie, d.h. wir setzen den Grundkörper K als *algebraisch abgeschlossen* (vorzugsweise, aber nicht notwendig von der Charakteristik null) voraus. Das ist selbstverständlich eine Vereinfachung unserer Probleme, die jedoch durch die Bemerkung gerechtfertigt werden kann, daß die

---

[2]) In der Tat ist $K(u_1, u_2, u_3) = K(x, y)$, weil

$$x = \frac{u_3^3 + 3u_1u_3 - u_2}{u_1 + 3u_3^2}, \qquad y = \frac{2u_3^3 - 2u_1u_3 + u_2}{u_1 + 3u_3^2}$$

gilt.

Probleme mit nicht abgeschlossenen Grundkörper K eigentlich weniger geometrischer als zahlentheoretischer Natur sind.

In meinem 1949 erschienen Buch [3]) bin ich konsequent vom Prinzip ausgegangen, daß *jedem Polynomideal umkehrbar eindeutig eine algebraische Mannigfaltigkeit entspreche.* Dieser Standpunkt hat zwar vielfach Bedenken und Kritik erregt, es scheint mir aber, daß sie zum größten Teil auf Mißverständnissen beruhen.

Zunächst scheint es mir nämlich gar nicht so wichtig zu sein, unsere geometrische Anschauung, die ja von vorneherein sehr unvollkommen ist, zu zwingen, die komplizierten Gebilde, die irgendwelchen Primäridealen höherer Multiplizität entsprechen, etwa durch Anlagerung von „unendlich benachbarten" Punkten auch wirklich vorzustellen; das geht klaglos in den einfachsten Fällen, dürfte aber bei höheren Multiplizitäten im allgemeinen nicht mehr zu machen sein.

Daß die einer abstrakten Definition entsprechenden Gebilde zum Teil unser Anschauungsvermögen übersteigen, sieht man auch an anderen Beispielen: Wenn wir z.B. definieren, daß eine Parameterdarstellung

$$x = f(t), \qquad y = g(t),$$

wo $f(t)$ und $g(t)$ in einem Intervall stetige Funktionen von t sind, eine stetige Kurve darstellt, so müssen wir dabei in Kauf nehmen, daß neben allen unserer Anschauung entsprechenden stetigen Kurven in dieser Definition auch Kurven, wie die bekannte *Peano-Kurve*, eingeschlossen sind, die unsere Anschauung unmöglich als Kurven akzeptieren kann.

Der vorliegende Fall ist übrigens gar nicht so extrem. F. Severi hat vor kurzem in einer sehr eingehenden Untersuchung [4]) über die in der algebraischen Geometrie der italienischen Schule verwendeten Mannigfaltigkeitsbegriffe an dritter Stelle den Begriff der „*Basismannigfaltigkeit*" dargelegt, der dadurch zustande kommt, daß man in jedem Punkt einer gewöhnlichen Mannigfaltigkeit noch ein gewisses „Verhalten" vorschreibt. Dieser Begriff kommt schon sehr nahe an den idealtheoretischen Mannigfaltigkeitsbegriff heran; in der Tat zeigt Severi, daß hier eine umkehrbar eindeutige Zuordnung zwischen Basismannigfaltigkeiten und einer sehr umfassenden Klasse von H-Idealen, die er als „komplett" bezeichnet, besteht.

---

[3]) Moderne algebraische Geometrie, Springer Wien 1949.

[4]) F. Severi. Le diverse concezioni di varietà nella geometria algebrica. Rend. Accad. Naz. dei XL, Serie IV, Vol. 2 (1952), 155—181. Ferner vom selben Verfasser: Les images géométriques des idéaux de polynomes, C. r. Acad. Sci. Paris **232** (1951), 2395—2396; Propriétés des images géométriques des idéaux de polynomes, C. r. Acad. Sci. Paris **233** (1951), 15—17.

Dieser von Severi geprägte Begriff einer Basismannigfaltigkeit V besteht also aus den Punkten einer algebraischen Mannigfaltigkeit im gewöhnlichen Sinne (Nullstellengebilde), in denen darüber hinaus ein besonderes Verhalten vorgeschrieben ist, d.h. eine V enthaltende Hyperfläche f muß nicht nur in den Punkten des Nullstellengebildes verschwinden, sondern dort noch eingewisses besonderes „Verhalten" zeigen. Wenn man dieses vorgeschriebene Verhalten etwas anders als Severi in der zitierten Arbeit definieren würde, nämlich durch die Vorschrift, daß eine Hyperfläche f, welche V enthalten soll, nicht nur in diesen Punkten verschwinden muß, sondern daß ihre Ableitungen dort gewisse lineare Bedingungen [5]) erfüllen müssen, so würde man genau den allgemeinen Begriff einer algebraischen Mannigfaltigkeit erhalten, den wir vom Standpunkt der Idealtheorie aus fordern müssen.

Diese Begriffsbestimmung einer algebraischen Mannigfaltigkeit läuft also auf das Postulat hinaus, daß eine *algebraische Mannigfaltigkeit dann und nur dann eindeutig festliegt, wenn es möglich ist, alle Hyperflächen anzugeben, die sie enthalten,* oder mit anderen Worten: *Eine algebraische Mannigfaltigkeit ist durch die Angabe aller sie enthaltenden Hyperflächen eindeutig definiert.*

Enthalten zwei Hyperflächen f und g die algebraische Mannigfaltigkeit V, so auch sämtliche Hyperflächen $af + bg$, wo a und b irgendwelche Polynome sind. *Daher ist die Menge aller, eine algebraische Mannigfaltigkeit enthaltenden Hyperflächen immer ein Polynomideal und umgekehrt definiert jedes Polynomideal eine ganz bestimmte algebraische Mannigfaltigkeit.*

Es darf hier bemerkt werden, daß der Begriff des „Zykels" im allgemeinen (wenn nämlich die Dimensionen der Komponenten nicht die höchsten sind) nicht diesem Postulat entspricht und daher von unserem Standpunkt aus nicht als algebraische Mannigfaltigkeit bezeichnet werden kann. Denn es ist im allgemeinen nicht möglich, die Hyperflächen anzugeben, welche einen vorgegebenen Zykel enthalten; es ist z.B. nicht möglich, die Kurven in der Ebene anzugeben, welche einen bestimmten, mit der Multiplizität 2 gerechneten Punkt der Ebene enthalten. Es ist wohl möglich alle Kurven der Ebene anzugeben, welche einen bestimmten, auf einer bestimmten Kurve gelegenen und dort mit der Multiplizität 2 gerechneten Punkt ausschneiden, denn in diesem Fall ist das zugehörige Primärideal durch die Multiplizität eindeutig bestimmt; dagegen gibt es zum Primideal (x, y) in $K[x, y]$ unendlich viele voneinander verschiedene Primärideale der Multiplizität 2; auch geometrisch ist ja ein in der Ebene mit der Multiplizität 2 zu rechnender Punkt erst eindeutig bestimmt, wenn man zum Punkte noch eine *Richtung* hinzufügt. Ähnlich ist zwar eine mit einer

[5]) Diese Bedingungen sind von Macaulay aus dem „inversen System" abgeleitet worden („modular equations"); vgl. F. S. Macaulay, The algebraic Theory of modular Systems. Cambridge University Press 1916.

gewissen (positiven) Multiplizität gerechnete Fläche im R_3, oder eine mit einer gewissen (positiven) Multiplizität gerechnete Kurve auf einer Fläche eine algebraische Mannigfaltigkeit, aber nicht ein mit einer Multiplizität behafteter Punkt auf einer Fläche usw.

Eine kleine Änderung ergibt sich, wenn man mit homogenen Idealen (H-Idealen) arbeitet. Da dann das Primideal $(x_0, x_1, \ldots, x_n)$ und alle zugehörigen Primärideale keine Nullstellen besitzen, so sind ihre algebraischen Mannigfaltigkeiten im *projektiven* Raum leer. Wenn wir nun eine zu diesem Primideal gehörende Primärkomponente eines H-Ideals kurz als „T-Komponente" bezeichnen, so können wir unseren Satz für H-Ideale so aussprechen:

Jedem H-Ideal entspricht eindeutig eine bestimmte algebraische Mannigfaltigkeit, aber umgekehrt entsprechen einer algebraischen Mannigfaltigkeit unendlich viele H-Ideale, die sich alle nur um eine T-Komponente unterscheiden; dasjenige unter diesen, das überhaupt keine T-Komponente besitzt, ist durch die algebraische Mannigfaltigkeit eindeutig bestimmt [6]).

Operationen mit Idealen und algebraischen Mannigfaltigkeiten.

Aus unserer Begriffsbestimmung der algebraischen Mannigfaltigkeit folgt unmittelbar:

Der Summe zweier Polynomideale entspricht der Durchschnitt der zugehörigen algebraischen Mannigfaltigkeiten; dem Durchschnitt zweier Polynomideale entspricht die (mengentheoretische) Summe ihrer algebraischen Mannigfaltigkeiten:

$$AM(\mathfrak{a} + \mathfrak{b}) = AM(\mathfrak{a}) \cap AM(\mathfrak{b}),$$
$$AM(\mathfrak{a} \cap \mathfrak{b}) = AM(\mathfrak{a}) + AM(\mathfrak{b}).$$

Der bekannte *Lasker-Noethersche Zerlegungssatz* für Polynomideale, wonach jedes Polynomideal als Durchschnitt größter Primärkomponenten dargestellt werden kann, von denen die isolierten eindeutig sind, liefert nun in Anwendung auf die algebraischen Mannigfaltigkeiten das fundamentale Ergenbnis:

Jede algebraische Mannigfaltigkeit ist eine Summe von primären Mannigfaltigkeiten, das sind irreduzible Mannigfaltigkeiten im gewöhnlichen Sinne, für die noch ein gewisses „Verhalten" vorgeschrieben ist. Die isolierten Komponenten sind dabei eindeutig bestimmt, während die eingebetteten Komponenten der geometrischen Anschauung entsprechend nicht eindeutig festgelegt sind.

Ist $\mathfrak{a}$ ein H-Ideal des homogenen Polynomringes $K[x_0, x_1, \ldots, x_n]$, so heißt

$$\bar{\mathfrak{a}} = \mathfrak{a} \cap K[x_0, x_1, \ldots, x_{n-1}]$$

das *Eliminationsideal* von $\mathfrak{a}$, das der Elimination der Variablen x_n entspricht.

---

[6]) Wenn man also mit H-Idealen arbeitet, so wird man zweckmäßig die T-Komponenten jeweils weglassen.

Es hat sich gezeigt, daß nur mit Hilfe dieses Begriffes, also idealtheoretisch, eine einwandfreie Eliminationstheorie aufgebaut werden kann [7]). Dem Eliminationsideal kommt auch eine wichtige geometrische Bedeutung zu, denn $AM(\bar{a})$ ist die *Projektion* von $AM(\mathfrak{a})$ aus dem unendlich fernen Punkt der x_n-Achse auf die Koordinaten-Hyperebene $x_n = 0$. In der Tat kann jede Nullstelle von $\bar{\mathfrak{a}} : \{\xi_0, \xi_1, \ldots, \xi_{n-1}\}$ auf mindestens eine Weise durch Bestimmung der letzten Koordinate ξ_n zu einer Nullstelle $\{\xi_0, \xi_1, \ldots, \xi_n\}$ von $\mathfrak{a}$ ergänzt werden [8]).

Der Multiplizitätsbegriff.

Schon um den sogenannten Fundamentalsatz der Algebra, daß jede algebraische Gleichung vom Grade k in einer Unbekannten genau k Wurzeln besitzt, aussprechen zu können, benötigt man den Multiplizitätsbegriff; es ist also nicht erstaunlich, daß dieser Begriff eine grundlegende Rolle sowohl in der Idealtheorie, wie auch in der algebraischen Geometrie spielt.

Es gilt zunächst die *Multiplizität eines Primärideals* zu definieren. Diese Definition ist schon sehr alt und geht meines Wissens auf Macaulay und Schmeidler zurück. Sie ergibt sich ganz ungezwungen aus dem gruppentheoretischen Begriff der Kompositionsreihe eines Primärideals; man definiert *die Länge dieser Kompositionsreihe als Multiplizität des Primärideals.*

Dabei ist wiederum zu betonen, daß ein Primärideal im allgemeinen durch das zugehörige Primideal und durch seine Multiplizität *nicht eindeutig bestimmt* ist; das ist nur bei Hauptidealen der Fall. Zur vollen geometrischen Charakterisierung etwa eines isolierten Schnittpunktes zweier algebraischen Mannigfaltigkeiten genügt also im allgemeinen nicht die bloße Angabe des Schnittpunktes und seiner Multiplizität, wenn man sich auch bei vielen Fragen der abzählenden Geometrie mit der Angabe der Multiplizitäten begnügt.

Auf Grund der idealtheoretischen Definition der Multiplizität ist die Frage nach der Multiplizität einer isolierten irreduziblen Komponente eines Schnittes von zwei (oder mehr) algebraischen Mannigfaltigkeiten $AM(\mathfrak{a})$ und $AM(\mathfrak{b})$ in größter Allgemeinheit beantwortet: Man bildet die Schnittmannigfaltigkeit $AM(\mathfrak{a} + \mathfrak{b})$ und zerlegt sie nach dem *Lasker-Noetherschen Satz* in ihre Primärkomponenten; die gefragte isolierte Komponente ist eindeutig bestimmt (gleichgültig welcher Dimension sie ist) und die Länge des zugehörigen Primärideals ist ihre Multiplizität.

Schon *Macaulay* [9]) hat nachgewiesen, daß diese Definition der Multiplizität in allen Fällen mit dem klassischen Begriff übereinstimmt, wo dieser überhaupt

---

[7]) Vgl.: W. Gröbner, Über die Eliminationstheorie. Monatsh. f. Math. **54** (1950), 71—78, sowie die dort angeführte Literatur. Eine ausführliche Darstellung dieser Eliminationstheorie habe ich in meinem oben zitierten Buche gegeben.

[8]) Vgl. mein oben zitiertes Buch, S. 45.

[9]) F. S. Macaulay, l.c., p. 77.

bereits in einer sicheren Art definiert ist. Insbesondere gilt der gewöhnliche *Bézoutsche Satz*, wobei die *idealtheoretische Multiplizität genau mit der von der u-Resolvente gelieferten Multiplizität übereinstimmt.*

Der *verallgemeinerte Bézoutsche Satz*, daß zwei algebraische Mannigfaltigkeiten $AM(\mathfrak{a})$ und $AM(\mathfrak{b})$ der Dimension d und $n - d$ in einem R_n sich in so vielen Punkten schneiden, als das Produkt ihrer Ordnungen beträgt, gilt allerdings nicht uneingeschränkt. In meinem oben zitierten Buch habe ich aber gezeigt, daß auch dieser verallgemeinerte Satz sicher immer dann gilt, wenn das Ideal $\mathfrak{a}$ perfekt und das Ideal $\mathfrak{b}$ Hauptklassenideal ist. Das ist z.B. im R_2 und im R_3 notwendig immer erfüllt: daher gibt es erst einigermaßen komplizierte Beispiele im R_4, wo dieser verallgemeinerte Bézoutsche Satz nicht bestätigt wird.

So weit ich sehe sind in allen wirkllch verläßlichen Beispielen von Anwendungen des verallgemeinerten Bézoutschen Satzes in der klassischen algebraischen Geometrie die Voraussetzungen für seine Gültigkeit auch nach den idealtheoretischen Begriffen erfüllt. In solchen Fällen aber, wo diese idealtheoretischen Voraussetzungen nicht erfüllt sind, wäre auch dann, wenn man den idealtheoretischen Multiplizitätsbegriff ablehnt, zumindestens Vorsicht geboten.

Man hat m.E. bei der Kritik des idealtheoretischen Multiplizitätsbegriffes diese Umstände zu stark dramatisiert und mit Unrecht das Postulat von der unbedingten Gültigkeit des verallgemeinerten Bézoutschen Satzes in den Vordergrund gestellt. Viel wichtiger dürfte das Postulat erscheinen, daß der *Multiplizitätsbegriff einfach und durchsichtig, und vor allem ohne jedwede Ausnahme überall anwendbar ist.* Die algebraische Geometrie darf nicht in Verlegenheit geraten, wenn es sich etwa einmal darum handelt, die Multiplizität eines gemeinsamen Schnittpunktes von drei Kurven im R_3, oder von drei Flächen im R_4 anzugeben usw.

Ebenso wichtig dürfte aber auch sein, daß der Multiplizitätsbegriff durchwegs *assoziativ* ist, d.h. daß die Multiplizität eines Schnittes von drei Mannigfaltigkeiten U, V, W nicht von der Reihenfolge abhängt, in welcher diese Mannigfaltigkeiten etwa zusammengefaßt werden:

$$(U \cap V) \cap W \text{ oder } U \cap (V \cap W).$$

Diese Assoziativität ist sowohl aus geometrischen, wie aus abstrakt logischen Erwägungen unbedingt erforderlich, sie wird aber von manchen modernen Multiplizitätsbegriffen nicht erfüllt [10]).

---

[10]) Vgl. E.—A. Behrens, Zur Schnittmultiplizität uneigentlicher Komponenten in der algebraischen Geometrie. Math. Z. **55** (1952), 199—215. Der idealtheoretische Multiplizitätsbegriff ist nach Severi „statisch", dem die von der italienischen Geometrie be-

Syzygientheorie und Hilbertfunktion.

In der Arbeit von 1890 [11]), die für die Theorie der Polynomideale grundlegend und richtunggebend war, definiert *Hilbert* die Syzygienmoduln und die charakteristische Funktion (= *Hilbertfunktion*) eines Moduls, d.h. eines H-Ideals. Schon Hilbert bemerkt in dieser Arbeit, daß die Koeffizienten dieser Funktion in engem Zusammenhang mit den von M. Noether eingeführten Geschlechtszahlen stehen[12]). Das kann aber nur bei der hier vertretenen idealtheoretischen Interpretation der algebraischen Geometrie der Fall sein, und es muß als Vorzug dieser Interpretation gewertet werden, daß sie die hier zufließenden Erkenntnisse unmittelbar geometrisch auswerten kann.

Es ist z.B. eine interessante Aufgabe, die Änderungen der Hilbertfunktion bei birationalen Transformationen eines H-Ideals zu verfolgen. Man sieht leicht ein, daß jede birationale Transformation sich zusammensetzen läßt aus einer *Veroneseschen Transformation* und einer *Projektion*. Wird ein H-Ideal $\mathfrak{a}_x$ durch eine Veronesesche Transformation des Grades μ in $\mathfrak{a}_y$ transformiert, so stehen die Hilbertfunktionen in dem einfachen Zusammenhang:

$$H(t;\, \mathfrak{a}_y) = H(\mu t;\, \mathfrak{a}_x).$$

Es ist leicht, die Invarianten für diese Transformation anzugeben[13]). Gegenüber *Projektionen*, welche die Singularitäten einer Mannigfaltigkeit ungeändert lassen, ist die Hilbertfunktion invariant. Es bleibt also nur übrig, diejenigen Projektionen, welche Singularitäten wegschaffen oder neue hinzubringen, genauer zu untersuchen, um die *„absoluten arithmetischen Invarianten"* aus der Hilbertfunktion zu gewinnen *).

Auch die Syzygientheorie hat eine unvermutete geometrische Bedeutung und Anwendung gewonnen. Es sei $\mathfrak{a}$ ein H-Ideal in $K[x_0, x_1, \ldots, x_n]$, unter dessen Primärkomponenten diejenige höchsten Ranges (kleinster Dimension) den Rang r habe. Bedeutet k die Länge der Syzygienkette von $\mathfrak{a}$, so gilt nach Hilbert immer [13]).

---

vorzugten „dynamischen" Begriffe gegenüberstehen. Das ist aber kein grundsätzlicher Einwand, denn jeder dynamische Begriff setzt einen statischen voraus und umgekehrt kann jeder statische Begriff durch Stetigkeitsüberlegungen in einen dynamischen übergeführt werden.

[11]) D. Hilbert, Über die Theorie der algebraischen Formen. Math. Ann. **36** (1890), 473—534.

[12]) Vgl. auch F. Severi, Fondamenti per la Geometria sulle varietà algebriche. Rend. Circolo mat. Palermo **28** (1909) und Ann. Mat. pura appl. IV, **32** (1951), 1—81. Hier wird dieser Zusammenhang mit den Koeffizienten der Postulationsformel untersucht und das „virtuelle arithmetische Geschlecht" definiert.

*) Vgl. meine demnächst in den Monatsheften erscheinende Arbeit: Die birationalen Transformationen der Polynomideale.

[13]) D. Hilbert, l.c. p. 492, Theorem III.

$$k \leqq n + 1;$$

es ist mir gelungen[14]), für k auch eine untere Schranke anzugeben, nämlich die eben definierte Zahl r, sodaß die Ungleichung vollständig lautet:

$$r \leqq k \leqq n + 1.$$

Besonders ausgezeichnet sind diejenigen Ideale, wo $k = r$ ist[15]); sie stellen sich als identisch mit den von Macaulay[16]) eingeführten „perfekten" Idealen heraus oder mit den von P. Dubreil[17]) untersuchten Idealen „absolument de première espèce".

Die geometrische Anwendung des Begriffes „perfekt", z.B. auf Raumkurven, hat gezeigt, daß perfekte Raumkurven mit den von Severi und Gaeta[18]) untersuchten Kurven „mit endlichem Residuum" identisch sind. Die interessanten Eigenschaften dieser Kurven, die geometrisch nur unter einschränkenden Voraussetzungen (Irreduzibilität und Singularitätenfreiheit) abgeleitet werden konnten, waren mit Hilfe der idealtheoretischen Begriffe und Methoden viel einfacher und ohne einschränkende Voraussetzungen beweisbar.

---

[14]) W. Gröbner, Über die Syzygientheorie der Polynomideale. Monatsh. f. Math. **53** (1919), 1—16.

[15]) Die obere Grenze $k = n + 1$ ist nicht interessant, weil dann notwendig auch $r = n + 1$ sein muß; d.h. $\mathfrak{a}$ enthält dann notwendig eine T-Komponente. Vgl. die eben zitierte Arbeit, S. 9, Satz 7.

[16]) F. S. Macaulay, l.c. p. 87 ss.

[17]) P. Dubreil, Quelques propriétés des variétés algébriques se rattachant aux théories de l'algèbre moderne. Hermann & Cie Paris 1935; Sur la dimension des idéaux de polynomes. Journ. de Math. **15** (1936), 271—283. Die Begriffe: „de première espèce, „de seconde espèce", „absolument (complètement) de première espèce", „multiplement de première espèce" von P. Dubreil lassen sich einfach von der Syzygientheorie aus folgendermaßen interpretieren:

Es sei $\mathfrak{a}$ ein (ungemischtes) H-Ideal vom Range r (Dimension $n - r < 0$), dessen Syzygienkette die Länge k habe. Dann ist sicher

$$r \leqq k \leqq n.$$

Ist $k = r$, so ist $\mathfrak{a}$ nach unserer Terminologie „perfekt", nach P. Dubreil „complètement de première espèce";
ist $k = n$, so heißt $\mathfrak{a}$ nach P. Dubreil „de seconde espèce";
ist $k = n - \delta$, $\delta = 1, 2, \ldots$, so heißt $\mathfrak{a}$ nach P. Dubreil „de première espèce", „doublement de première espèce, . . ., „multiplement de première espèce" ecc.

[18]) F. Gaeta, Sulle curve sghembe algebriche di residuale finite. Ann. Mat. pura appl. IV, **27** (1948), 177—241; Nuove ricerche sulle curve sghembe algebriche di residuale finito e sui gruppi di punti del piano. Ann. Mat. pura appl. IV, **31** (1950), 1—64; Détermination de la chaîne syzygétique des idéaux matriciels parfaits et son application à la postulation de leurs variétés algébriques associées. C. r. Acad. Sci. Paris **234** (1952), 1833—1835.

Zum Abschluß begnüge ich mich darauf hinzuweisen, daß mit diesen wenigen Beispielen die Anwendung der Idealtheorie auf die algebraische Geometrie nicht erschöpft ist. Vor allem läßt sich die Theorie der *äquivalenten Ideale* und der *linearen Scharen* aufbauen[19]); auch die allgemeine Theorie der algebraischen Mannigfaltigkeiten, so wie sie Severi in den oben zitierten Abhandlungen vom geometrischen Standpunkt aus dargestellt hat, läßt sich ohne prinzipiellen Schwierigkeiten ins idealtheoretische übertragen, und man darf hoffen, daß mit der Zeit auch hier wirklich neue Erkenntnisse mit Hilfe der Idealtheorie gewonnen werden können.

---

[19]) W. Gröbner, Idealtheoretischer Aufbau der algebraischen Geometrie. Hamburger math. Einzelschr. Heft 30, (1941).

INNSBRUCK.

DER SATZ VON RIEMANN-ROCH IN FAISCEAU-THEORETISCHER FORMULIERUNG EINIGE ANWENDUNGEN UND OFFENE FRAGEN

FRIEDRICH HIRZEBRUCH

In diesem Vortrag möchte ich in einem ersten § die Resultate über den Riemann-Rochschen Satz und seine Verallgemeinerungen zusammenfassen, die ich in einer Note [1]) bereits veröffentlicht habe und über welche sich eine ausführliche Arbeit in Vorbereitung befindet. In § 2 soll dann als Anwendung u.a. eine Formel für die $h^{p,q}$ (= Anzahl der komplex-linear-unabhängigen harmonischen Formen vom Typ (p, q)) eines vollständigen Durchschnitts $V_n^{(a_1, a_2, \ldots, a_r)}$ von Hyperflächen der Grade $a_1, \ldots, a_r$ im komplexen projektiven Raum $\boldsymbol{P}_{n+r}$ bewiesen werden. In § 3 wird u.a. die spezielle Gestalt des Riemann-Rochschen Satzes für algebraische Mannigfaltigkeiten mit verschwindenden Pontrjaginschen Klassen betrachtet. In einer gemeinsamen Arbeit mit A. Borel soll darauf näher eingegangen werden. Im letzten § werden einige Bemerkungen zur Theorie des virtuellen arithmetischen Geschlechtes gemacht und in diesem Zusammenhang auf ein ungelöstes Problem über kompakte komplexe Mannigfaltigkeiten hingewiesen.

1. *Der Hauptsatz.*

1.1. Es sei V_n eine kompakte fast-komplexe Mannigfaltigkeit [1a]) der komplexen Dimension n mit den Chernschen Klassen $c_i \in H^{2i}(V_n, \boldsymbol{Z})$, $(c_0 = 1, c_i = 0$ für $i > n)$. Genauer gesagt sind die c_i die Chernschen Klassen des kontravarianten Tangentialbündels von V_n, das den komplexen Vektorraum $\boldsymbol{C}_n$ als Faser und die lineare Gruppe $\boldsymbol{GL}(n, \boldsymbol{C})$ als Strukturgruppe hat. $c_n[V_n]$, der Wert von c_n auf dem zur natürlichen Orientierung von V_n gehörigen ($2n$-dimensionalen) Fundamentalzyklus, ist gleich der Euler-Poincaréschen Charakteristik von V_n. Ein Bündel W über V_n mit dem komplexen Vektorraum $\boldsymbol{C}_q$ als Faser und mit $\boldsymbol{GL}(q, \boldsymbol{C})$ als Strukturgruppe wird kurz Vektorraum-Bündel und für $q = 1$ Geradenbündel genannt. Es seien $d_i \in H^{2i}(V_n, \boldsymbol{Z})$ die Chernschen Klassen von W, $(d_0 = 1, d_i = 0$ für $i > q)$. Wir schreiben formal unter Ver-

---

[1]) F. Hirzebruch, Proc. Nat. Acad. Sci., U.S.A., **40**, 110—114 (1954).

[1a]) Mannigfaltigkeiten werden nicht notwendigerweise als zusammenhängend vorausgesetzt.

wendung einer Unbestimmten x

$$(1)\quad 1 + c_1x + c_2x^2 + \ldots c_nx^n = (1 + \gamma_1x)(1 + \gamma_2x)\ldots(1 + \gamma_nx) \text{ und}$$
$$1 + d_1x + d_2x^2 + \ldots + d_qx^q = (1 + \delta_1x)(1 + \delta_2x)\ldots(1 + \delta_qx).$$

Wir verabreden, jede Potenzreihe, die symmetrisch in den beiden Variablenreihen γ_i und δ_i ist, als Potenzreihe in den elementar-symmetrischen Funktionen c_i, d_i aufzufassen. Dem Bündel W über V_n wird ein Polynom $T_y(V_n, W)$ vom Grade n und mit rationalen Koeffizienten zugeordnet.

Definition:

$$(2)\quad T_y(V_n, W) = \varkappa_n\Big[\Big(\sum_{j=1}^{q} e^{(1+y)\delta_j}\Big).\ \prod_{i=1}^{n}\big((1 + y)\gamma_i(1 - e^{-(1+y)\gamma_i})^{-1} - y\gamma_i\big)\Big]$$

$\varkappa_n[\ldots]$ ist immer so definiert: Der Ausdruck in $[\ldots]$ ist ein Element von $H^*(V_n, \boldsymbol{Z}) \otimes A$, ($A$ ist ein Koeffizientenbereich, im vorliegenden Fall ist $A = \boldsymbol{Q}(y)$ = Polynomring in einer Unbestimmten über den rationalen Zahlen. H^* bedeutet Cohomologiering). Von dem Ausdruck in $[\ldots]$ wird der Teil von der topologischen Dimension $2n$ genommen, dessen Wert auf dem Fundamentalzyklus der orientierten V_n ein bestimmtes Element von A ist, das mit $\varkappa_n[\ldots]$ bezeichnet wird. Wenn betont werden soll, dass $\varkappa_n$ in Bezug auf die Mannigfaltigkeit V_n zu verstehen ist, dann wird $\varkappa_n[\ldots]_{V_n}$ an Stelle von $\varkappa_n[\ldots]$ geschrieben. Wir setzen

$$(3)\quad T_y(V_n, W) = \sum_{p=0}^{n} T^p(V_n, W)y^p,$$
$$T_0(V_n, W) = T^0(V_n, W) = T(V_n, W).$$

Offenbar ist

$$(4)\quad T(V_n, W) = \varkappa_n\Big[\Big(\sum_{j=1}^{q} e^{\delta_j}\Big).\ \prod_{i=1}^{n} \gamma_i/(1 - e^{-\gamma_i})\Big]$$
$$= \varkappa_n\left[\left(\sum_{j=1}^{q} e^{\delta_j}\right)e^{c_1/2}\ .\ \prod_{i=1}^{n}\frac{\gamma_i/2}{\sinh \gamma_i/2}\right]$$

Ferner ist

$$(5)\quad T^p(V_n, W) = T(V_n, \mathfrak{T}^{(p)} \otimes W),$$

wo $\mathfrak{T}^{(p)}$ das Bündel der kovarianten p-Vektoren ist. Die Faser von $\mathfrak{T}^{(p)}$ ist der komplexe Vektorraum der Dimension $\binom{n}{p}$. Mit $\otimes$ wird das Tensorprodukt von Vektorraum-Bündeln bezeichnet.

1.2. In diesem Abschnitt sei V_n eine kompakte komplexe Mannigfaltigkeit der komplexen Dimsenion n und W sei ein *komplex-analytisches* Vektorraum-Bündel über V_n mit $\boldsymbol{C}_q$ als Faser und $\boldsymbol{GL}(q, \boldsymbol{C})$ als Strukturgruppe. Dann ist der Faisceau $\Omega(W)$ der Keime von lokalen holomorphen Schnittflächen von W wohldefiniert. Die Cohomologiegruppe $H^i(V_n, \Omega(W))$, die der Kürze wegen mit

$H^i(V_n, W)$ bezeichnet werde, ist ein Vektorraum über den komplexen Zahlen, der endliche Dimension hat [2]). $H^i(V_n, W)$ verschwindet für $i > n$. $H^0(V_n, W)$ ist der komplexe Vektorraum aller globalen holomorphen Schnittflächen von W. Die ganze Zahl $\chi(V_n, W)$ wird definiert durch

$$\chi(V_n, W) = \sum_{i=0}^{n} (-1)^i \dim H^i(V_n, W). \tag{6}$$

Es sei $\mathfrak{T}^{(p)}$ das (komplex-analytische) Bündel der kovarianten p-Vektoren von V_n. Das Bündel $\mathfrak{T}^{(n)}$ ist ein Geradenbündel, wird mit K bezeichnet und kanonisches Geradenbündel von V_n genannt. Die erste Chernsche Klasse von $\mathfrak{T}^{(n)}$ ist gleich $-c_1$, wo c_1 die erste Chernsche Klasse von V_n bezeichnet. $\mathfrak{T}^{(0)}$ ist das triviale Geradenbündel.

Definition:

$$\chi^p(V_n, W) = \chi(V_n, W \otimes \mathfrak{T}^{(p)}) \tag{6*}$$

$$\chi_y(V_n, W) = \sum_{p=0}^{n} \chi^p(V_n, W) y^p$$

Nach Serre [3]) ist

$$H^p(V_n, W) \simeq H^{n-p}(V_n, K \otimes W^*) \text{ und} \tag{7}$$

$$H^p(V_n, W \otimes \mathfrak{T}^{(q)}) \simeq H^{n-p}(V_n, W^* \otimes \mathfrak{T}^{(n-q)}).$$

Hierbei ist W^* das duale Bündel von W.

Wenn W das triviale Geradenbündel ist, dann wird $\chi_y(V_n, W)$ gleich $\chi_y(V_n)$ gesetzt usw. $\chi(V_n) = \chi_0(V_n) = \chi^0(V_n)$ heisst arithmetisches Geschlecht von V_n. Nach Dolbeault [4]) ist für eine kählersche Mannigfaltigkeit V_n

$$\dim H^q(V_n, \mathfrak{T}^{(p)}) = h^{p,q}, \tag{8}$$

wo $h^{p,q}$ die Anzahl der komplex-linear-unabhängigen harmonischen Formen vom Typ (p, q) ist. Also ist

$$\chi^p(V_n) = \sum_{q=0}^{n} (-1)^q h^{p,q} \tag{9}$$

Insbesondere gilt für eine *kählersche* Mannigfaltigkeit V_n

$$\chi(V_n) = \sum_{q=0}^{n} (-1)^q g_q, \tag{10}$$

wo g_q die Anzahl der komplex-linear-unabhängigen holomorphen Formen

---

[2]) H. Cartan et J. P. Serre, C. R. Acad. Sci., Paris, **237**, 128—130 (1953). — H. Cartan, Séminaire E.N.S., 1953—54. — K. Kodaira, Proc. Nat. Acad. Sci., U.S.A., **39**, 865—868 (1953).

[3]) J. P. Serre, Un théorème de dualité, Comm. math. Helv., **29**, 9—26 (1955).

[4]) P. Dolbeault, C. R. Acad. Sci., Paris, **236**, 175—177 (1953).

(1. Gattung) vom Grade q ist. Die Serreschen Dualitätsformeln (7) implizieren

(11) $\chi(V_n, W)=(-1)^n\chi(V_n, K\otimes W^*)$ und $\chi^p(V_n, W)=(-1)^n\chi^{n-p}(V_n, W^*)$.

1.3. *Hauptsatz*: *Es sei V_n eine algebraische Mannigfaltigkeit, (das soll hier heissen: V_n sei eine kompakte komplexe Mannigfaltigkeit, die komplex-analytisch und singularitätenfrei in einen komplexen projektiven Raum eingebettet werden kann). W sei ein komplex analytisches Vektorraum-Bündel über V_n. Dann ist*

(12) $$\chi(V_n, W) = T(V_n, W).$$

Aus dem Hauptsatz folgt mit Hilfe von (5) und (6*):

Es gilt unter den Voraussetzungen des Hauptsatzes

(12*) $$\chi_y(V_n, W) = T_y(V_n, W),$$

also insbesondere

(13) $$\chi_y(V_n) = \varkappa_n\left[\prod_{i=1}^{n}\left((1+y)\gamma_i(1-e^{-(1+y)\gamma_i})^{-1}-y\gamma_i\right)\right] \quad \text{und}$$

(14) $$\chi(V_n) = \varkappa_n\left[\prod_{i=1}^{n}\gamma_i/(1-e^{-\gamma_i})\right]$$

Die Formel (14) besagt, dass das arithmetische Geschlecht $\chi(V_n)$ gleich dem Toddschen Geschlecht $T(V_n)$ ist [5]).

An den Hauptsatz schliessen sich eine Reihe von Fragen an. Zum Beispiel:

1) Ist die Gleichung $\chi(V_n, W) = T(V_n, W)$ auch richtig für eine kompakte komplexe Mannigfaltigkeit V_n und ein komplex-analytisches Vektorraum-Bündel W über V_n?

2) Ist $T(V_n, W)$ immer eine ganze Zahl? (V_n und W wie in 1.1.). Zu diesen Fragen vgl. man [6]).

1.4. Der vorstehende Hauptsatz ist eine weitgehende Verallgemeinerung des klassischen Riemann-Rochschen Satzes:

Für eine algebraische Mannigfaltigkeit V_n stehen die Divisorenklassen (in Bezug auf lineare Äquivalenz) in eineindeutiger Beziehung mit den (Isomorphieklassen von [7])) komplex-analytischen Geradenbündeln über V_n. Die

---

[5]) J. A. Todd, Proc. London Math. Soc. (2), **43**, 190—225 (1937). — F. Hirzebruch, Proc. Nat. Acad. Sci., U.S.A., **39**, 951—956 (1953).

[6]) F. Hirzebruch, Some problems on differentiable and complex manifolds, Ann. of Math., **60**, 213—236 (1954).

[7]) Im folgenden identifizieren wir stets zwei komplex-analytische Geradenbündel, wenn sie komplex-analytisch isomorph sind. Ein komplex-analytisches Geradenbündel über V_n ist dann aufzufassen als ein Element der Cohomologiegruppe $H^1(V_n, \mathfrak{h})$, wo $\mathfrak{h}$ der Faisceau der Keime von lokalen nicht-verschwindenden holomorphen Funktionen von V_n ist. Vgl. J. P. Serre, Centre Belge Rech. math., Colloque sur les fonctions de plusieurs variables, 57—68 (1953). — Für den Satz, dass auf algebraischen Mannigfaltigkeiten die

Geradenbündel bilden in Bezug auf das Tensorprodukt, das der Multiplikation der Divisoren entspricht, eine (multiplikativ geschriebene [8])) abelsche Gruppe, die der Gruppe der Divisorenklassen in natürlicher Weise isomorph ist. Das inverse Bündel F^{-1} eines Geradenbündels F ist gleich dem dualen Bündel F^* von F. Die erste Chernsche Klasse $f = d_1(F) \in H^2(V_n, \boldsymbol{Z})$ des Geradenbündels F wird einfach Cohomologieklasse von F genannt. Wenn F zu dem Divisor D gehört, dann ist f dual zu der durch den Divisor repräsentierten $(2n-2)$-dimensionalen ganzzahligen Homologieklasse von V_n. Die Untergruppe derjenigen komplex-analytischen Geradenbündel, deren Cohomologieklasse verschwindet, ist gleich der Picardschen Mannigfaltigkeit von V_n. Nach dem Hauptsatz gilt für ein Geradenbündel F über einer algebraischen Mannigfaltigkeit V_n

$$(15)\quad \chi(V_n, F) = \sum_{i=0}^{n} (-1)^i \dim H^i(V_n, F) = \varkappa_n \left[e^{f+\frac{1}{2}c_1} \prod_{i=1}^{n} \frac{\gamma_i/2}{\sinh \gamma_i/2} \right].$$

Nach Kodaira [9]) nennt man ein Geradenbündel F über V_n positiv, wenn seine Cohomologieklasse $f \in H^2(V_n, \boldsymbol{Z})$ als Fundamentalklasse einer kählerschen Metrik von V_n auftreten kann, (d.h. es existiert eine Kählermetrik $ds^2 = \sum g_{\alpha\bar{\beta}} dz^\alpha \overline{dz^\beta}$ mit der Eigenschaft, dass die geschlossene Form

$$\omega = \frac{i}{2} \sum g_{\alpha\bar{\beta}} dz^\alpha \wedge \overline{dz^\beta}$$

bei dem de Rhamschen Isomorphismus die Cohomologieklasse $f' \in H^2(V_n, \boldsymbol{R})$ repräsentiert, wobei f' das Bild von f bei dem natürlichen Homomorphismus $H^2(V_n, \boldsymbol{Z}) \to H^2(V_n, \boldsymbol{R})$ ist). Insbesondere ist F dann positiv, wenn F gleich einem Geradenbündel ist, das zu einem Hyperflächenschnitt von V_n bezüglich einer geeigneten Einbettung von V_n in einen projektiven Raum gehört. Es gilt der folgende

Satz (Kodaira [9])). *Wenn F^{-1} positiv, dann $H^i(V_n, F) = 0$ für $i \neq n$. Wenn $F \otimes K^{-1}$ positiv, dann $H^i(V_n, F) = 0$ für $i \neq 0$.*

(Die beiden Aussagen des vorstehenden Satzes gehen bei Verwendung von (7) ineinander über).

---

komplex-analytischen Geradenbündel den Divisorenklassen in natürlicher Weise eineindeutig entsprechen, siehe K. Kodaira and D. C. Spencer, Proc. Nat. Acad. Sci., U.S.A., **39**, 868—877 (1953).

[8]) In der Literatur wird oft auch die additive Schreibweise verwandt. Die Gruppe der komplex-analytischen Geradenbündel ist nichts anderes als die Cohomologiegruppe $H^1(V_n, \mathfrak{h})$. Eine ganzzahlige Potenz eines Geradenbündels ist stets im Sinne des Produktes von Geradenbündeln zu verstehen.

[9]) K. Kodaira, Proc. Nat. Acad. Sci., U.S.A., **39**, 1268—1273 (1953). — K. Kodaira, On Kähler varieties of restricted type, Proc. Nat. Acad. Sci., U.S.A., **40**, 313—316 (1954) und Ann. of Math., **60**, 28—48 (1954).

Wegen dieses Satzes von Kodaira und der Formel (15) hat man den *Satz*: *Wenn* $F \otimes K^{-1}$ *positiv ist, dann ist*

$$\dim H(V_n, F) = \varkappa_n \left[e^{f+\frac{1}{2}c_1} \prod_{i=1}^{n} \frac{\gamma_i/2}{\sinh \gamma_i/2} \right] \tag{15*}$$

$H^0(V_n, F)$ ist der komplexe Vektorraum aller globalen holomorphen Schnittflächen von F und ist, wenn man F durch einen Divisor D repräsentiert, isomorph zum komplexen Vektorraum aller derjenigen meromorphen Funktionen g von V_n, für die gD keine Polstellen hat.

Die Formel (11) und der Kodairasche Satz ergeben unmittelbar, dass für eine algebraische Mannigfaltigkeit V_n gilt:

$$1 + (-1)^n p_a = 1 + (-1)^n P_a = \chi(V_n)\ ^{10}). \tag{16}$$

Hierbei sind p_a, P_a die beiden klassischen arithmetischen Geschlechter; die Formel (16) wurde von Severi vermutet. Es wurde bereits erwähnt, dass $\chi(V_n)$ gleich dem Toddschen Geschlecht von V_n ist (siehe (14)).

2. *Vollständige Durchschnitte von Hyperflächen im komplexen projektiven Raum.*

2.1. Im komplexen projektiven Raum $\boldsymbol{P}_{n+r}$ von $n + r$ komplexen Dimensionen betrachte man r singularitätenfreie Hyperflächen $F^{(a_1)}, \ldots, F^{(a_r)}$ von den Graden $a_1, \ldots, a_r$. Es werde vorausgesetzt, dass diese Hyperflächen in „allgemeiner Lage" sind. (Das soll hier heissen, dass jeder Punkt $x \in F^{(a_1)} \cap \ldots \cap F^{(a_r)}$ in $\boldsymbol{P}_{n+r}$ eine Umgebung U besitzt, in der lokale Koordinaten $z_1, \ldots, z_{n+r}$ mit x als Zentrum definiert sind und in der $U \cap F^{(a_i)}$ durch $z_{n+i} = 0$ gegeben wird). Der Durchschnitt $F^{(a_1)} \cap \ldots \cap F^{(a_r)}$ ist dann eine kompakte komplexe Mannigfaltigkeit $V_n^{(a_1, \ldots, a_r)}$ von n komplexen Dimensionen, die singularitätenfrei in $\boldsymbol{P}_{n+r}$ eingebettet ist und damit eine algebraische Mannigfaltigkeit im Sinne von 1.3 ist. $V_n^{(a_1, \ldots, a_r)}$ wird „vollständiger Durchschnitt" der Hyperflächen $F^{(a_1)}, \ldots, F^{(a_r)}$ genannt. $V_n^{(a_1, \ldots, a_r)}$ hängt als differenzierbare Mannigfaltigkeit nur von $n, a_1, \ldots, a_r$ ab. Für festes $n, a_1, \ldots, a_r$ sind jedoch die möglichen algebraischen Mannigfaltigkeiten $V_n^{(a_1, \ldots, a_r)}$ i.a. nicht komplex-analytisch und auch nicht birational äquivalent, wie schon das Beispiel $n = 1$, $r = 1$, $a_1 = 3$ zeigt. $\boldsymbol{H}$ sei das zur Hyperebene von $\boldsymbol{P}_{n+r}$ gehörige Geradenbündel und $\tilde{\boldsymbol{H}}$ die Beschränkung von $\boldsymbol{H}$ auf $V_n^{(a_1, \ldots, a_r)}$. Ferner sei $h \in H^2(\boldsymbol{P}_{n+r}, \boldsymbol{Z})$ die Cohomologieklasse von $\boldsymbol{H}$ und $\tilde{h} \in H^2(V_n^{(a_1, \ldots, a_r)}, \boldsymbol{Z})$ die Cohomologieklasse von $\tilde{\boldsymbol{H}}$. Wir haben

---

[10]) K. Kodaira and D. C. Spencer, Proc. Nat. Acad. Sci., U.S.A., **39**, 641—649 (1953).

[11]) Wenn von *der* Chernschen Klasse gesprochen wird, ist die Summe $\sum_{i=0}^{\infty} c_r$ der Chernschen Klassen in den verschiedenen Dimensionen gemeint.

vor, das Polynom $\chi_y(V_n^{(a_1,\ldots,a_r)}, \tilde{\boldsymbol{H}}^k)$ zu berechnen [8]). Es wird sich herausstellen, dass dieses Polynom nur von $n, a_1, \ldots, a_r, k$ abhängt.

Satz: *Es sei* $V_n = V_n^{(a_1,\ldots,a_r)}$ *vollständiger Durchschnitt von Hyperflächen der Grade* $a_1, \ldots, a_r$ *im* $\boldsymbol{P}_{n+r}$. *Es gilt unter Einführung einer Unbestimmten* z

$$(1) \qquad \sum_{n=0}^{\infty} \chi_y(V_n, \tilde{\mathbf{H}}^k) z^{n+r} = \frac{(1+zy)^{k-1}}{(1-z)^{k+1}} \prod_{i=1}^{r} \frac{(1+zy)^{a_i} - (1-z)^{a_i}}{(1+zy)^{a_i} + y(1-z)^{a_i}}$$

Beweis: Die Chernsche Klasse [11]) von $\boldsymbol{P}_{n+r}$ ist $(1+h)^{n+r+1}$. Daraus folgt in bekannter Weise für die Chernsche Klasse von V_n

$$(2) \qquad c(V_n) = (1+\tilde{h})^{n+r+1}(1+a_1\tilde{h})^{-1} \ldots (1+a_r\tilde{h})^{-1},$$

wobei natürlich $(1+a_i\tilde{h})^{-1} = \sum_{j=0}^{n} (a_i h)^j$. Wir setzen

$$Q(x) = x(y+1)(1-e^{-x(y+1)})^{-1} - yx \text{ und } Q(x) = \frac{x}{R(x)}.$$

Dann ist nach dem Hauptsatz (1.3)

$$\chi_y(V_n, \tilde{\boldsymbol{H}}^k) = T_y(V_n, \tilde{\boldsymbol{H}}^k) =$$

$$\varkappa_n\Big[e^{(1+y)k\tilde{h}} \cdot \tilde{h}^{n+r+1} \cdot R(\tilde{h})^{-(n+r+1)} \cdot \prod_{i=1}^{r} \big((a_i\tilde{h})^{-1} R(a_i\tilde{h})\big)\Big]_{V_n} =$$

$$\varkappa_{n+r}\Big[e^{(1+y)kh} \cdot h^{n+r+1} \cdot R(h)^{-(n+r+1)} \cdot \prod_{n=1}^{r} R(a_i h)\Big]_{P_{n+r}}$$

Daher ist

$$\chi_y(V_n, \tilde{\boldsymbol{H}}^k) = \operatorname{resid}\,(e^{(1+y)kx} \cdot R(x)^{-(n+r+1)} \cdot \prod_{i=1}^{r} R(a_i x)),$$

wo resid $G(x)$ den Koeffizienten von x^{-1} in $G(x)$ bezeichnet. x ist eine Unbestimmte. Nach „Einführung einer neuen Variablen" $z = R(x)$ erhält man die zu beweisende Formel (1) wegen

$$e^{(1+y)x} = (1+zy)/(1-z), \quad R(ax) = \frac{(1+zy)^a - (1-z)^a}{(1+zy)^a + y(1-z)^a},$$

$$\frac{dz}{dx} = (1+zy)(1-z).$$

Bemerkung: Der vorstehende Satz wurde hier mit Hilfe des Hauptsatzes (1.3) bewiesen. Der Beweis des Hauptsatzes [12]) benutzt wesentlich die Thomsche Theorie [13]). Man kann den vorstehenden Satz auch ohne Thomsche Theorie

---

[12]) F. Hirzebruch, loc. cit. in [1]).

[13]) R. Thom, Comm. math. Helv. **28**, 17—86 (1954).

durch ein Induktionsverfahren mit Hilfe einer Formel von Kodaira-Spencer [14]) beweisen.

Ersetzt man in (1) y durch 0, dann geht (1) in die folgende Formel über

$$(1^*) \qquad \sum_{n=0}^{\infty} \chi(V_n, \tilde{H}^k) z^{n+r} = (1-z)^{-(k+1)} \prod_{i=1}^{r} (1-(1-z)^{a_i}).$$

Für $k = 0$ ist dies eine Formel für das arithmetische Geschlecht $\chi(V_n)$ der vollständigen Durchschnitte V_n, die sich in anderer Form bei Lefschetz findet (Ann. of Math. 17, 197—212 (1916)).

2.2. Die vollständigen Durchschnitte sind algebraische Mannigfaltigkeiten mit sehr speziellen Eigenschaften. Es gilt nämlich das folgende

Lemma: *Es sei* $V_n = V_n^{(a_1,\cdots,a_r)}$ *ein vollständiger Durchschnitt, und* $\tilde{h}$ *habe die in* 2.1 *angegebene Bedeutung. Dann gilt*:

1) V_n *ist für* $n \geqq 1$ *zusammenhängend und für* $n \geqq 2$ *einfach -zusammenhängend.*

2) *Es ist* $h^{p,q}(V_n) = \delta_{p,q}$ *für* $p + q \neq n$.

3) V_n *hat keine Torsion. Die Gruppe* $H^{2k}(V_n, \boldsymbol{Z})$ *ist für* $2k < n$ *unendlich-zyklisch und wird von* $\tilde{h}^k$ *erzeugt.*

Beweis: Dass V_n für $n \geqq 1$ zusammenhängend ist, ergibt sich zum Beispiel aus einem Satz von Bertini. Aus einem Satz von Chow [15]) folgt, dass V_n für $n \geqq 2$ einfach-zusammenhängend ist. — Da $h^{p,p} \geqq 1$ und $\sum_{p+q=k} h^{p,q} = b_k =$ k. Bettische Zahl von V_n, genügt es zum Beweis von (2) zu zeigen, dass b_k für $k < n$ gleich 1 oder gleich 0 ist, jenachdem k gerade oder ungerade ist. Nach dem Satz von Lefschetz [16]) über Hyperflächenschnitte ist aber für $k < n$ die k. Bettische Zahl von $V_{n+1}^{(a_1,\cdots,a_r,1)}$ gleich der von $V_n^{(a_1,\cdots,a_r)}$. Also ergibt sich die Behauptung über b_k durch Induktion über r. Für $r = 0$ ist sie sicherlich richtig (V_n = komplexer projektiver Raum). — Zum Beweis von 2) wurde der Lefschetzsche Satz nur für die Cohomologie mit reellen (oder komplexen) Zahlen als Koeffizienten benötigt. Der Lefschetzsche Satz gilt aber auch für die ganzzahlige Cohomologie [17]) und ergibt dann, wieder durch Induktion über r, die Behauptung 3).

Korollar: *Für einen vollständigen Durchschnitt* V_n *ist*

$$(3) \qquad \chi^p(V_n) = (-1)^{n-p} h^{p,n-p} + (-1)^p \text{ für } 2p \neq n,$$
$$\chi^m(V_n) = (-1)^m h^{m,m} \text{ für } 2m = n.$$

[14]) K. Kodaira and D. C. Spencer, Proc. Nat. Acad. Sci., U.S.A., **39**, 1273—1278 (1953), Formel (14).

[15]) W. L. Chow, Amer. J. Math., **74**, 726—736 (1952).

[16]) Vgl. zum Beispiel Kodaira-Spencer, loc. cit. in [14]).

[17]) Für einen neuen Beweis und für Anwendungen siehe die Arbeiten von Fáry (Cohomologie des variétés algébriques, Thèse, Collège de France, 1955).

2.3. Der Satz (2.1) gibt für $k = 0$ eine explizite Formel für die $\chi^p(V_n)$ und damit nach dem vorstehenden Korollar und nach 1.2 (9) auch für die $h^{p,q}(V_n)$. In dem folgenden Satz spezialisieren wir auf die vollständigen Durchschnitte von Quadriken.

Satz: *Es sei* $K_n^{(r)} = V_n^{(2,2,\ldots,2)}$ *vollständiger Durchschnitt von r Quadriken des* $\boldsymbol{P}_{n+r}$. *Dann ist*

$$(4) \qquad \sum_{n=0}^{\infty} \chi_y(K_n^{(r)}) z^n = (1 + yz)^{-1}(1 - z)^{-1}(2 + (y - 1)z)^r (1 + yz^2)^{-r}$$

Für $r = 1, 2, 3$ ergibt sich so

$V_n^{(2)}$, n ungerade : $h^{p,p} = 1$, alle anderen $h^{p,q} = 0$.

$V_n^{(2)}$, $n = 2m$: $h^{p,p} = 1$ für $p \neq m$, $h^{m,m} = 2$, alle anderen $h^{p,q} = 0$.

$V_n^{(2,2)}$, $n = 2m + 1$: $h^{p,p} = 1$, $h^{m,m+1} = h^{m+1,m} = m + 1$, alle anderen $h^{p,q} = 0$.

$V_n^{(2,2)}$, $n = 2m$: $h^{p,p} = 1$ für $p \neq m$, $h^{m,m} = 2m + 4$, alle anderen $h^{p,q} = 0$.

$V_n^{(2,2,2)}$, $n = 2m + 1$: $h^{p,p} = 1$, $h^{m,m+1} = h^{m+1,m} = (m + 1)(2m + 5)$, alle anderen $h^{p,q} = 0$.

$V_n^{(2,2,2)}$, $n = 2m$: $h^{p,p} = 1$ für $p \neq m$, $h^{m,m} = 3m^2 + 9m + 8$, $h^{m-1,m+1} = h^{m+1,m-1} = m(m + 1)/2$, alle anderen $h^{p,q} = 0$.

Bekanntlich ist für eine kählersche Mannigfaltigkeit V_n:

$$\chi_1(V_n) = \sum_{p=0}^{n} \chi^p(V_n) = \tau(V_n) = \text{Index } (V_n)\ ^{18}),$$

wo $\tau(V_n) = 0$ für ungerades n, und $\tau(V_{2m})$ gleich dem Überschuss der Anzahl der „Plus"-Zeichen über die Anzahl der „Minus"-Zeichen in der Diagonalform der durch $xx[V_{2m}](x \in H^{2m}(V_{2m}, \boldsymbol{R}))$ definierten quadratischen Form ist. Wir erhalten aus (4) für $y = 1$ die folgende Formel für die Indizes der vollständigen Durchschnitte von Quadriken

$$(4^*) \qquad \sum_{n=0}^{\infty} \tau(K_n^{(r)}) z^n = 2^r (1 - z^2)^{-1}(1 + z^2)^{-r}.$$

2.4. *Satz*: [19]) *Die Euler-Poincarésche Charakteristik von* $V_n = V_n^{(a_1,\ldots,a_r)}$ *werde mit* $E(V_n)$ *bezeichnet. Es ist*

$$(5) \qquad \sum_{n=0}^{\infty} E(V_n) z^n = (1 - z)^{-2} a_1 a_2 \ldots a_r \prod_{i=1}^{r} (1 + (a_i - 1)z)^{-1}$$

Beweis ergibt sich aus Satz (2.1) für $k = 0$ und $y = -1$.

Bemerkung: Da $c_n[V_n] = E(V_n)$, lässt sich der vorstehende Satz allein auf Grund der Formel (2) beweisen. Wenn die Euler-Poincarésche Charakteristik

[18]) W. V. D. Hodge, Proc. International Congress of Math., I, 182—191 (1950).

[19]) Diese Formel findet sich in anderer Form bei J. A. Todd, loc. cit. in [5]), p. 197.

des vollständigen Durchschnitts V_n bekannt ist, dann sind wegen Lemma (2.2) alle Bettischen Zahlen von V_n bekannt.

2.5. Für diesen Abschnitt sei V_n eine zusammenhängende algebraische Mannigfaltigkeit mit $H^2(V_n, \boldsymbol{Z}) = \boldsymbol{Z}$. Die vollständigen Durchschnitte sind für $n > 2$ solche algebraische Mannigfaltigkeiten [19a]).

Die Elemente $f \in H^2(V_n, \boldsymbol{Z})$ sind vom Typ (1,1) und treten deshalb alle als Cohomologieklassen von komplex-analytischen Geradenbündeln auf [20]). Die Gruppe $H^2(V_n, \boldsymbol{Z})$ hat zwei erzeugende Elemente. Dasjenige (eindeutig bestimmte) erzeugende Element, das als Cohomologieklasse eines positiven Geradenbündels auftritt (vgl. 1.4), werde mit g bezeichnet. Die erste Chernsche Klasse von V_n sei $c_1 = \mu g$. Bekanntlich ist $-c_1$ die Cohomologieklasse des kanonischen Geradenbündels K von V_n. Für ein komplex-analytisches Geradenbündel F mit der Cohomologieklasse f werde $f = \lambda(F)g$ gesetzt. $\lambda(F)$ ist eine wohlbestimmte ganze Zahl. Nach dem Satz von Kodaira (1.4) ist

$$(6) \qquad \begin{aligned} &H^i(V_n, F) = 0 \text{ für } i \neq n \text{ und } \lambda(F) < 0, \\ &H^i(V_n, F) = 0 \text{ für } i \neq 0 \text{ und } \lambda(F) + \mu > 0. \end{aligned}$$

Es sei zunächst $\mu > 0$.

Dann verschwindet $H^i(V_n, F)$ für alle i, wenn $-\mu < \lambda(F) < 0$.

Für $\lambda(F) \leqq -\mu$ ist nur $H^n(V_n, F)$ evtl. nicht null.

Für $\lambda(F) \geqq 0$ ist nur $H^0(V_n, F)$ evtl. nicht null.

Insbesondere verschwinden die g_i für $i \neq 0$. Also ist $\chi(V_n) = g_0 = 1$. (Wegen des Satzes von Kodaira (1.4) gilt allgemein: Wenn für eine algebraische Mannigfaltigkeit das Bündel K^{-1} positiv ist, dann verschwinden alle g_i für $i \neq 0$). Für eine algebraische Mannigfaltigkeit V_n mit $H^2(V_n, \boldsymbol{Z}) = \boldsymbol{Z}$ und $\mu > 0$ können alle Zahlen $\dim H^i(V_n, F)$ für jedes Geradenbündel F mit Hilfe der Formel 1.4 (15) berechnet werden.

Wenn $\mu \leqq 0$, dann bleibt es für $0 \leqq \lambda(F) \leqq -\mu$ unentschieden, ob mehr als eine Gruppe $H^i(V_n, F)$ nicht verschwindet.

2.6. Wir verwenden die Bezeichnungen des vorigen Abschnitts. Für einen vollständigen Durchschnitt $V_n = V_n^{(a_1, \ldots, a_r)}$ mit $n > 2$ ist $H^2(V_n, \boldsymbol{Z}) = \boldsymbol{Z}$. Es ist $g = \tilde{h}$ (vgl. 2.1 und Lemma (2.2)). Da $g_1(V_n) = 0$, verschwindet die Picardsche Mannigfaltigkeit von V_n und jedes Geradenbündel F ist durch $\lambda(F)$ in eindeutiger Weise bestimmt. Es ist $\mu = n + r + 1 - (a_1 + \ldots + a_r)$. Die g_i verschwinden für $0 < i < n$. (Es ist also $\chi(V_n) = 1 + (-1)^n g_n$ und daher

---

[19a]) Es werde hier hervorgehoben, dass jede *kählersche* Mannigfaltigkeit V_n mit $H^2(V_n, \boldsymbol{Z}) \simeq \boldsymbol{Z}$ algebraisch ist. V_n ist nämlich offenbar eine Hodge-Mannigfaltigkeit (Kähler variety of restricted type) und deshalb algebraisch (Kodaira, loc. cit. in [9])).

[20]) Satz von Lefschetz-Hodge. Vgl. Kodaira-Spencer, loc. cit. in [7]).

$p_a = P_a = g_n \geqq 0$.) Nach 2.5 und nach 2.1 (Formel (1*)) lassen sich für $\mu \geqq 0$ die Dimensionen sämtlicher Cohomologiegruppen $H^i(V_n, F)$, F beliebiges komplex-analytisches Geradenbündel über V_n, berechnen.

Beispiele:

1) Für den komplexen projektiven Raum $\boldsymbol{P}_n$ ($r = 0$ in 2.1 (1*)) ergibt sich

$$\chi(\boldsymbol{P}_n, \boldsymbol{H}^k) = \binom{n+k}{k} \text{ also}$$

dim $H^0(\boldsymbol{P}_n, \boldsymbol{H}^k) = \binom{n+k}{k}$ für $k \geqq 0$, dim $H^n(\boldsymbol{P}_n, \boldsymbol{H}^k) = \binom{-k-1}{n}$ für $-k \geqq n+1$

dabei ist $\boldsymbol{H}$ das zur Hyperebene von $\boldsymbol{P}_n$ gehörige Geradenbündel. Alle anderen Gruppen $H^i(\boldsymbol{P}_n, \boldsymbol{H}^k)$ verschwinden.

2) Für die in $\boldsymbol{P}_{n+1}$ eingebettete Quadrik $V_n^{(2)}$ ergibt sich

$$\chi(V_n^{(2)}, \tilde{\boldsymbol{H}}^k) = \binom{n+k}{n} + \binom{n+k-1}{n}. \text{ Es ist}$$

$$\dim H^0(V_n^{(2)}, \tilde{\boldsymbol{H}}^k) = \chi(V_n^{(2)}, \tilde{\boldsymbol{H}}^k) \text{ für } k \geqq 0 \text{ und}$$

$$\dim H^n(V_n^{(2)}, \tilde{\boldsymbol{H}}^k) = (-1)^n \chi(V_n^{(2)}, \tilde{\boldsymbol{H}}^k) \text{ für } -k \geqq n,$$

dabei ist $\tilde{\boldsymbol{H}}$ das zum Hyperebenenschnitt von $V_n^{(2)}$ gehörige Geradenbündel. Alle anderen Gruppen $H^i(V_n^{(2)}, \tilde{\boldsymbol{H}}^k)$ verschwinden.

3. *Riemann-Rochscher Satz und Pontrjaginsche Klassen.*

3.1. Für eine algebraische Mannigfaltigkeit V_n und ein komplex-analytisches Geradenbündel F über V_n wird $\chi(V_n, F)$ durch die Formel 1.4 (15) gegeben. $\dfrac{\gamma_i/2}{\sinh \gamma_i/2}$ ist eine Potenzreihe in γ_i^2. Die r-te elementar-symmetrische Funktion der γ_i^2 (aufgefasst als Polynom in den Chernschen Klassen c_i; vgl. 1.1) ist gleich der Pontrjagischen Klasse p_r von V_n, ($p_r \in H^{4r}(V_n, \boldsymbol{Z})$). Die Klassen p_r hängen nur von V_n als differenzierbarer Mannigfaltigkeit ab; sie hängen nicht von der Orientierung von V_n ab; sie bleiben bei allen differenzierbaren Homöomorphismen von V_n auf sich fest. Die Formel für $\chi(V_n, F)$ hat also die folgende Gestalt:

$$\chi(V_n, F) = \varkappa_n[A e^{f+\frac{1}{2}c_1}]_{V_n}, \tag{1}$$

wo f die Cohomologieklasse von F, c_1 die erste Chernsche Klasse von V_n und A ein Polynom in den Pontrjaginschen Klassen von V_n ist, dessen 0-dimensionales Glied gleich 1 ist.

3.2. Wir wenden die Bemerkungen von 3.1 speziell auf algebraische Mannigfaltigkeiten V_n mit $H^2(V_n, \boldsymbol{Z}) \simeq \boldsymbol{Z}$ an [19a]). Wir benutzen die Bezeichnungen von 2.5. Die Zahl $\chi(V_n, F)$ hängt nur von $\lambda(F)$ ab. Für $\lambda(F) = k$ setzen wir $\chi(V_n, F) = \chi(V_n, k)$, d.h.

$$\chi(V_n, k) = \varkappa_n[A e^{\left(k+\frac{\mu}{2}\right)g}]_{V_n}. \tag{2}$$

Es sei V'_n eine algebraische Mannigfaltigkeit, die differenzierbar und orientierungstreu homöomorph zu V_n ist. „Orientierungstreu" ist in Bezug auf die natürlichen Orientierungen von V_n und V'_n gemeint. Die Cohomologie von V_n werde mit der von V'_n, dem Homoömorphismus entsprechend, identifiziert. Alles, was sich auf V'_n bezieht, werde mit ' versehen. Um $\chi(V'_n, k)$ zu erhalten, muss man auf der rechten Seite von (2) μ durch μ' und g durch g' ersetzen. A bleibt unverändert. Natürlich ist $g = \pm g'$. Wegen der Orientierungstreue ist $g^n = g'^n$, woraus folgt, dass $g = + g'$ für ungerades n. Ferner ist $\mu \equiv \mu'$ (mod 2), da c_1 und c'_1 bei Reduktion mod 2 die Stiefel-Whitneysche Klasse $w_2 \in H^2(V_n, \boldsymbol{Z}_2)$ ergeben, welche topologisch invariant ist und bei Homöomorphismen von V_n auf sich festbleibt. Es folgt, dass *die Menge aller Zahlen $\chi(V_n, k)$ mit der Menge aller Zahlen $\chi(V'_n, k)$ übereinstimmt.* Insbesondere muss das arithmetische Geschlecht $\chi(V'_n)$ unter den $\chi(V_n, k)$ vorkommen.

Wenn n *ungerade* ist, $\chi(V'_n)$ bekannt ist und die Gleichung $\chi(V'_n) = \chi(V_n, k_0)$ für ein und nur ein k_0 zutrifft, dann ist μ' eindeutig bestimmt:

$$\left(k_0 + \frac{\mu}{2}\right) g = \frac{\mu'}{2} g' = \frac{\mu'}{2} g \text{ und } \mu' = 2k_0 + \mu.$$

Es sei n *gerade*. Dann gilt nach 1.3 (11): $\chi(V_n, k) = \chi(V_n, -\mu - k)$. Wenn $\chi(V'_n)$ bekannt ist und die Gleichungen $\chi(V'_n) = \chi(V_n, k_0) = \chi(V_n, -\mu - k_0)$ für ein und nur ein (nicht geordnetes) Zahlenpaar $(k_0, -\mu - k_0)$ zutreffen, dann ist $\mu' = \pm (2k_0 + \mu)$.

Beispiele:

1) $V_n = \boldsymbol{P}_n$ = komplexer projektiver Raum ($\mu = n + 1$). Die algebraische Mannigfaltigkeit V'_n sei differenzierbar und orientierungstreu homöomorph zu V_n. Da die bettischen Zahlen $b_i = b'_i$ gleich 1 oder 0 sind, jenachdem i gerade oder ungerade ist, und da

$$h^{p,p}(V'_n) \geqq 1 \text{ und } 2g'_{2p} + h^{p,p}(V'_n) \leqq b'_{2p} \text{ für } p > 0, \tag{3}$$

folgt $g'_i = 0$ für $i > 0$, also $\chi(V'_n) = 1$. — Die Zahl 1 kommt für *ungerades* n nur für $k = 0$ unter den Zahlen $\chi(\boldsymbol{P}_n, k)$ vor (vgl. 2.6, Beispiel 1)). Also ist $\mu' = \mu = n + 1 > 0$ für jede Mannigfaltigkeit V'_n. Also ist (siehe (2)) $\chi(\boldsymbol{P}_n, F) = \chi(V'_n, F')$, wenn die Cohomologieklassen von F und F' gleiche Vielfache von g bzw. g' sind, und nach 2.5 folgt dann: dim $H^0(\boldsymbol{P}_n, F) =$ dim $H^0(V'_n, F')$. Daraus hat Kodaira gefolgert, dass für ungerades n jede algebraische Mannigfaltigkeit V'_n, die differenzierbar homöomorph zu $\boldsymbol{P}_n$ ist [21]), komplex-analytisch

[21]) Wenn n ungerade ist und V'_n homöomorph zu $\mathbf{P}_n$ ist, dann ist V'_n auch orientierungstreu homöomorph zu $\mathbf{P}_n$, da $\mathbf{P}_n$ einen die Orientierung umkehrenden Homöomorphismus auf sich besitzt.

homöomorph zu $\boldsymbol{P}_n$ ist, (nicht veröffentlicht; vgl. Hirzebruch, loc. cit. in [6])). — Für *gerades* n ergibt sich $\mu' = \pm (n+1)$. Es ist anscheinend nicht bekannt, ob eine V'_n mit $\mu' = -(n+1)$ existiert.

2) $V_n = V_n^{(2)} = n$-dimensionale Quadrik (vgl. 2.3). Die algebraische Mannigfaltigkeit V'_n sei differenzierbar und orientierungstreu homöomorph zu $V_n^{(2)}$. Es ergibt sich wie im obigen Beispiel mit Hilfe von (3), dass $\chi(V'_n) = 1$. Unter Verwendung von 2.6, Beispiel 2), erhält man, dass $\mu' = \mu = n$ für ungerades n und $\mu' = \pm n$ für gerades n. Für ungerades n haben entsprechende Gruppen $H^0(V_n^{(2)}, F)$, $H^0(V'_n, F')$ wieder gleiche Dimensionen.

3.3. Es sei V_n eine algebraische Mannigfaltigkeit, deren Pontrjaginsche Klassen p_i für $i > 0$ verschwinden. Dann nimmt nach 3.1 die Formel für $\chi(V_n, F)$, F komplex-analytisches Geradenbündel über V_n, die folgende Gestalt an:

$$\chi(V_n, F) = \varkappa_n[e^{f+\frac{1}{2}c_1}] = (n!)^{-1}(f + \tfrac{1}{2}c_1)^n[V_n]. \tag{4}$$

f ist die Cohomologieklasse von F. Insbesondere gilt für das arithmetische Geschlecht:

$$\chi(V_n) = (n!)^{-1}2^{-n}c_1^n[V_n] \tag{4*}$$

Algebraische Mannigfaltigkeiten mit verschwindenden Pontrjaginschen Klassen sind zum Beispiel die homogenen Räume G/T, wo G eine kompakte einfache Liesche Gruppe und T ein maximaler Torus von G ist. Die auftretenden Zahlen $|\chi(V_n, F)|$ stimmen mit den Graden der irreduziblen Darstellungen von G überein. Die Stiefel-Whitneysche Klasse $w_2 \in H^2(G/T, \boldsymbol{Z}_2)$ ist gleich 0. Für Einzelheiten muss auf die in der Einleitung erwähnte gemeinsame Arbeit mit A. Borel hingewiesen werden. Wir bemerken hier nur, dass für eine algebraische Mannigfaltigkeit V_n mit $p_i = 0$ (für $i > 0$) und $w_2 = 0$ wegen Formel (4) folgendes gilt:

Für jedes Element $f \in H^2(V_n, \boldsymbol{Z})$, das vom Typ (1,1) ist, d.h. als Cohomologieklasse eines komplex-analytischen Geradenbündels auftritt [20]), hat man die Kongruenz

$$f^n \equiv 0 \pmod{n!} \tag{5}$$

Frage: Es sei V^{2n} eine kompakte orientierbare differenzierbare Mannigfaltigkeit der Dimension $2n$ mit $p_i = 0$ (für $i > 0$) und $w_2 = 0$. Gilt für jedes $f \in H^2(V^{2n}, \boldsymbol{Z})$ die Kongruenz (5)? (Vgl. Hirzebruch, loc. cit. in [6])).

4. *Über virtuelle arithmetische Geschlechter.*

4.1. Es sei V_n eine kompakte komplexe Mannigfaltigkeit. Eine (nichtgeordnete) Menge von r komplex-analytischen Geradenbündeln $F_1, \ldots, F_r$ von V_n heisst virtuelle Untermannigfaltigkeit der Dimension $n - r$ von V_n. Es soll das virtuelle arithmetische Geschlecht $\chi(F_1, \ldots, F_r)_{V_n}$ dieser virtuellen Untermannigfaltigkeit definiert werden.

Definition:

$$\chi(F_1, \ldots, F_r)_{V_n} = \sum_{j=0}^{r} (-1)^j \sum_{i_1 < \ldots < i_j} \chi(V_n, F_{i_1}^{-1} \otimes \ldots \otimes F_{i_j}^{-1}).$$

Die innere Summe ist über alle möglichen $\binom{r}{j}$ Kombinationen zu erstrecken. Zum Beispiel ist

$r = 0$, $\chi(\ )_{V_n} = \chi(V_n)$

$r = 1$, $\chi(F_1)_{V_n} = \chi(V_n) - \chi(V_n, F_1^{-1})$

$r = 2$, $\chi(F_1, F_2)_{V_n} = \chi(V_n) - \chi(V_n, F_1^{-1}) - \chi(V_n, F_2^{-1}) + \chi(V_n, F_1^{-1} \otimes F_2^{-1})$

Das virtuelle arithmetische Geschlecht hat die folgenden Eigenschaften, die seine Definition rechtfertigen.

(1) $\chi(F_1, \ldots, F_r)_{V_n}$ hängt nicht von der Reihenfolge der F_i ab.

(2) Wenn F_1 ein Geradenbündel ist, das zu einem singularitätenfreien Divisor S gehört, dann ist

a) $\chi(F_1)_{V_n} = \chi(S)$ b) $\chi(F_1, \ldots, F_r)_{V_n} = \chi((F_2)_S, \ldots, (F_r)_S)_S$

Der Divisor S heisst dabei singularitätenfrei, wenn er lokal durch $z = 0$ gegeben werden kann. S ist also eine kompakte komplexe Mannigfaltigkeit der Dimension $n - 1$. Mit $(F_i)_S$ wird die Beschränkung von F_i auf S bezeichnet.

(3) Es gilt für $r \geqq 1$ die folgende „Funktionalgleichung".

$$\chi(A \otimes B, F_2, \ldots, F_r)_{V_n}$$
$$\chi(A, F_2, \ldots, F_r)_{V_n} + \chi(B, F_2, \ldots, F_r)_{V_n} - \chi(A, B, F_2, \ldots, F_r)_{V_n}$$

(F_i, A, B sind komplex-analytische Geradenbündel über V_n).

Aus (2) folgt: Wenn $F_1, \ldots, F_r$ ($1 \leqq r \leqq n$) zu singularitätenfreien Divisoren $S_1, \ldots, S_r$ gehören und wenn die Hyperflächen $S_1, \ldots, S_r$ von V_n sich „allgemein" schneiden, dann ist $\chi(F_1, \ldots, F_r)_{V_n}$ gleich dem arithmetischen Geschlecht der komplexen Mannigfaltigkeit $S_1 \cap S_2 \cap \ldots \cap S_r$, welche die Dimension $n - r$ hat. Die Formel (1*) von 2.1 gibt für $k = 0$ für den komplexen projektiven Raum die Werte aller möglichen virtuellen arithmetischen Geschlechter, wenn man für die a_i auch negative ganze Zahlen zulässt.

4.2. Für eine algebraische Mannigfaltigkeit V_n und komplex-analytische Geradenbündel $F_1, \ldots, F_r$ folgt aus der Definition (4.1) und aus 1.4 (15), dass

$$\chi(F_1, \ldots, F_r)_{V_n} = \varkappa_n \left[\prod_{i=1}^{r} (1 - e^{-f_i}) e^{c_1/2} \prod_{j=1}^{n} \frac{\gamma_j/2}{\sinh \gamma_j/2} \right], \tag{4}$$

wobei $f_i \in H^2(V_n, \mathbf{Z})$ die Cohomologieklasse von F_i ist. Da $1 - e^{-f_i}$ durch f_i teilbar ist, enthält der Ausdruck in $[\ldots]$ für $r > n$ kein Glied der topologischen Dimension $2n$, und es ergibt sich der

Satz: *Das virtuelle arithmetische Geschlecht* $\chi(F_1, \ldots, F_r)_{V_n}$ *verschwindet für* $r > n$, *wenn* V_n *eine algebraische Mannigfaltigkeit ist.*

Problem: *Ist der vorstehende Satz für beliebige kompakte komplexe Mannigfaltigkeiten* V_n *richtig?* Einfachster Fall ($n = 2$, $F_1 = F_2 = F_3 = F^{-1}$):

Ist die folgende Gleichung für eine kompakte komplexe V_2 und ein komplex-analytisches Geradenbündel F über V_2 immer richtig?

$$\chi(V_2) - 3\chi(V_2, F) + 3\chi(V_2, F^2) - \chi(V_2, F^3) = 0.$$

4.3. Aus der obigen Formel 4 und aus 1.4 (15) kann man sofort erhalten:

Satz: *Es sei* V_n *algebraisch und* F *ein komplex-analytisches Geradenbündel. Dann ist*

$$(5) \quad \chi(V_n, F) = \chi(V_n) + \chi(F)_{V_n} + \chi(F, F)_{V_n} + \ldots + \underset{n \text{ mal}}{\chi(F, \ldots, F)_{V_n}}$$

Beweis: Im Cohomologiering von V_n gilt die Gleichung

$$e^f = (1 - (1 - e^{-f}))^{-1} = \sum_{j=0}^{n} (1 - e^{-f})^j.$$

Der vorstehende Satz wurde im wesentlichen von Severi [22]) vermutet.

4.4. Der Satz (4.2) kann natürlich auch direkt bewiesen werden. In der Tat, nach einem bekannten Satz kann F_1 in der Form $S \otimes T^{-1}$ geschrieben werden, wo S, T Geradenbündel sind, die zu singularitätenfreien Divisoren gehören. Es folgt mit Hilfe der Funktionalgleichung 4.1 (3)

$$(6) \quad \chi(S, F_2, \ldots, F_r)_{V_n} = \chi(F_1, F_2, \ldots, F_r)_{V_n} + \chi(T, F_2, \ldots, F_r)_{V_n} - \chi(T, F_1, \ldots, F_r)_{V_n}$$

Der Satz (4.2) ergibt sich aus (6) und aus 4.1 (2) durch Induktion über n.

Auch die Severische Formel (5) kann direkt bewiesen. Auf Grund der Definition des virtuellen arithmetischen Geschlechtes ist nämlich die folgende Gleichung eine rein formale Identität

$$(5^*) \quad \chi(V_n, F) = \chi(V_n) + \chi(F)_{V_n} + \chi(F, F)_{V_n} + \ldots + \underset{n \text{ mal}}{\chi(F, \ldots, F)_{V_n}} + \underset{n+1 \text{ mal}}{\chi(F, \ldots, F)_{V_n}} - \chi(F^{-1}, \underbrace{F, \ldots, F}_{n+1 \text{ mal}})$$

Aus Satz (4.2) folgt jetzt (5).

4.5. Für eine algebraische Mannigfaltigkeit V_n und ein komplex-analytisches Geradenbündel F über V_n hängt $\chi(V_n, F)$ nach 1.4 (15) nur von der reellen Cohomologieklasse von F ab, welche sich ergibt, wenn man auf die ganzzahlige Cohomologieklasse f von F ($f \in H^2(V_n, \mathbf{Z})$ den natürlichen Homo-

[22]) F. Severi, Centre Belge Rech. math., Colloque Géom. algebrique, 9—55 (1950).

morphismus $H^2(V_n, \boldsymbol{Z}) \to H^2(V_n, \boldsymbol{R})$ anwendet. Kodaira und Spencer haben oft die folgende *Frage* gestellt:

Ist $\chi(V_n, F)$ für eine beliebige kompakte komplexe Mannigfaltigkeit nur von der reellen Cohomologieklasse von F abhängig?

Serre hat bemerkt, dass man die vorstehende Frage für eine gegebene kompakte komplexe Mannigfaltigkeit V_n und jedes komplex-analytische Geradenbündel F über V_n zu bejahen hat, wenn für beliebige komplex-analytische Geradenbündel A, B über V_n gezeigt werden kann, dass $\chi(V_n, A \otimes B^k)$, k durchlaufe die ganzen Zahlen, ein Polynom in k ist.

Satz: Wenn für eine kompakte komplexe V_n das virtuelle Geschlecht $\chi(F_1, \ldots, F_{n+1})_{V_n}$ immer verschwindet, dann ist für komplex-analytische Geradenbündel A, B über V_n immer $\chi(V_n, A \otimes B^k)$ ein Polynom in k.

Beweis: Wir beweisen die äquivalente Aussage, dass $\chi(A \otimes B^k)_{V_n}$ ein Polynom in k ist. Man fasse $\chi(A \otimes B^k)_{V_n}$ als eine Zahlenfolge auf. k durchlaufe alle ganzen Zahlen. Durch wiederholte Anwendung der Funktionalgleichung 4.1 (3) kann man die iterierten Differenzenfolgen bilden. Die Zahlen der $(n+1)$ Differenzenfolge sind Summen von Summanden der From $\pm \chi(F_1, \ldots, F_{n+1})_{V_n}$ die nach Voraussetzung verschwinden. Also ist $\chi(A \otimes B^k)_{V_n}$ ein Polynom n. Grades.

Der vorstehende Satz zeigt, dass die oben gestellte Frage mit dem Problem von 4.2 eng zusammenhängt.

Für weitere Einzelheiten der Theorie des virtuellen arithmetischen Geschlechts und für Definition und Eigenschaften des virtuellen χ_y-Geschlechtes, das für den Beweis des Hauptsatzes (1.3) wichtig ist, muss auf die in Vorbereitung befindliche Arbeit verwiesen werden (siehe Einleitung). Das virtuelle χ_y-Geschlecht geht für $y = 0$ in das virtuelle arithmetische Geschlecht über. — Es sollten in diesem § nur einige Bemerkungen und Beispiele gegeben werden. Bei einem systematischen Aufbau der Überlegungen, die schliesslich zu dem Hauptsatz (1.3) führen, hat die Theorie des virtuellen χ_y-Geschlechtes mit am Anfang zu stehen. Schliesslich werde noch bemerkt, dass das hier definierte virtuelle arithmetische Geschlecht für algebraische Mannigfaltigkeiten mit dem klassischen virtuellen arithmetischen Geschlecht übereinstimmt, da es durch seine Eigenschaften 4.1 (1)—(3), die auch von dem klassischen virtuellen Geschlecht erfüllt werden, für algebraische Mannigfaltigkeiten in eindeutiger Weise festgelegt ist. Zu beachten ist, dass das klassische arithmetische Geschlecht mit χ in der Beziehung $\chi(V_n) = 1 + (-1)^n p_a(V_n)$ steht [23]). Wenn etwas weiter oben von Übereinstimmung gesprochen wurde, dann ist natürlich gemeint, dass diese Beziehung auch im virtuellen Fall gilt, d.h. also:

---

[23]) Wir setzen hier voraus, dass V_n zusammenhängend ist.

$(-1)^{n-r+1} + (-1)^{n-r}\chi(F_1, \ldots, F_r)_{V_n} = \mathfrak{p}_a(F_1, \ldots, F_r)_{V_n} = P_a(F_1, \ldots, F_r)_{V_n}$

Die Funktionalgleichung 4.1 (3) ist für das klassische Geschlecht entsprechend abzuändern.

SCHÜTZENSTR. 19, BEI SPITZLEY,
WITTEN/RUHR, DEUTSCHLAND.

SOME RESULTS IN THE TRANSCENDENTAL THEORY OF ALGEBRAIC VARIETIES.

KUNIHIKO KODAIRA

1. *Introduction.* Recently the theory of analytic sheaves, developed mainly by H. Cartan [1], has been successfully applied to algebraic geometry by J.-P. Serre [2]. This theory is independent of potential theory. On the other hand it is possible to develope a theory of analytic sheaves on compact complex manifolds by a potential theoretic method. The present note is a survey of some results in this direction and their applications to algebraic geometry.

2. *Complex line bundles* [3]. Let M be a compact complex manifold of complex dimension n. By a *complex line bundle* F over M we shall mean an analytic fibre bundle whose fibre is a complex line (i.e. the complex number field) $\boldsymbol{C}$ and whose structure group is the multiplicative group $\boldsymbol{C}^*$ of complex numbers acting on $\boldsymbol{C}$. The bundle F may be described as follows: Let $\{U_j\}$ be a sufficiently fine finite covering of V and let π be the canonical projection of F onto V. Then the inverse image $\pi^{-1}(U_j)$ has a product structure: $\pi^{-1}(U_j) = U_j \times \boldsymbol{C}$, and $(z, \zeta_j) \in U_j \times \boldsymbol{C}$ is identical with $(z, \zeta_k) \in U_k \times \boldsymbol{C}$ if and only if $\zeta_j = f_{jk}(z) \cdot \zeta_k$, where $f_{jk}(z)$ is a non-vanishing holomorphic function defined in $U_j \cap U_k$. Under these circumstances we say that the bundle F is defined by the system $\{f_{jk}\}$ of the *transition functions* f_{jk}, and we call ζ_j the *fibre coordinate* of the point (z, ζ_j) on F over U_j. We identify two complex line bundles which are analytically equivalent. Moreover, for any pair of bundles F, G determined, respectively, by $\{f_{jk}\}$, $\{g_{jk}\}$, we define $F \pm G$ to be the bundle determined by $\{f_{jk} g_{jk}^{\pm 1}\}$. Then *the set* $\mathfrak{F} = \{F\}$ *of all complex line bundles* F *over* M *forms an additive group.*

Given a divisor D on M defined in each U_j by a local meromorphic function $f_j(z; D)$, we denote by $[D]$ the complex line bundle determined by the system $\{f_{jk}(z; D)\}$ of functions $f_{jk}(z; D) = f_j(z; D)/f_k(z; D)$. It is obvious that $[D]$ coincides with $[D']$ if and only if D is linearly equivalent to D'. Thus *each* $[D]$

---

[1]) Séminaire H. Cartan, 1951—1952.

[2]) Moreover Serre has generalized the theory of sheaves to the case of abstract algebraic varieties. See his paper in these Proceedings.

[3]) K. Kodaira and D. C. Spencer: Groups of complex line bundles over compact Kähler varieties, Proc. Nat. Acad. Sciences, **39** (1953), pp. 868—872.

may be identified with the divisor class of D and therefore *the divisor class group can be regarded as a subgroup of* $\mathfrak{F}$.

By the characteristic class $c(F)$ of F we shall mean the characteristic class of the principal bundle associated with F. $c(F)$ is an elemeht of the second cohomology group $H^2(M, \boldsymbol{Z})$ of M with coefficients in the integers $\boldsymbol{Z}$. For any element c of $H^2(M, \boldsymbol{Z})$ we denote by c_R the element of $H^2(M, \boldsymbol{R})$ corresponding to c under the canonical homomorphism $H^2(M, \boldsymbol{Z}) \to H^2(M, \boldsymbol{R})$, where $\boldsymbol{R}$ denotes the reals. In view of de Rham's theorem, c_R can be regarded as a class of closed real 2-forms on M. By a (p, q)-*form* we shall mean a $(p + q)$-form of type (p, q).

Theorem 1 [4]. *The homomorphism* $F \to c(F)$ *maps* $\mathfrak{F}$ *onto the subgroup* $H^2_{(1,1)}(M, \boldsymbol{Z})$ *of* $H^2(M, \boldsymbol{Z})$ *consisting of all* c *such that* c_R *contains a closed real* $(1, 1)$-*form.*

3. *Cohomologies with coefficients in some analytic sheaves.* Let Ω^p be the sheaf over M of germs of holomorphic p-forms and let $H^q(M, \Omega^p)$ be the q-th cohomology group of M with coefficients in Ω^p. Then, as was shown by Dolbeault [5], we have

$$H^q(M, \Omega^p) \simeq H^{p,q}, \tag{1}$$

provided that M *is a Kähler manifold*, where $H^{p,q}$ denotes the linear space of harmonic forms of type (p, q) on M. This result can be generalized to the case of sheaves of germs of holomorphic p-forms with coefficients in a complex line bundle.

By a form φ with coefficients in a complex line bundle F over M we shall mean a system $\{\varphi_j\}$ of forms defined respectively in U_j such that

$$\varphi_j = f_{jk} \cdot \varphi_k \quad \text{in } U_j \cap U_k,$$

where $\{f_{jk}\}$ is the system of transition functions defining the bundle F. For such a form $\varphi = \{\varphi_j\}$, we define $\bar{\partial}\varphi = \{(\bar{\partial}\varphi)_j\}$ and $\mathfrak{d}\varphi = \{(\mathfrak{d}\varphi)_j\}$ by

$$(\bar{\partial}\varphi)_j = \bar{\partial}\varphi_j, \qquad (\mathfrak{d}\varphi)_j = - * a_j \partial (a_j^{-1} * \varphi_j),$$

where $*\varphi_j$ denotes the dual form of φ_j with respect to a preassigned Hermitian metric $ds^2 = 2\sum g_{\alpha\bar{\beta}} dz^\alpha dz^{\bar{\beta}}$ on M and where a_j are real positive functions of class C^∞ defined respectively in U_j satisfying $a_j = |f_{jk}|^2 a_k$ in $U_j \cap U_k$. Moreover we introduce the complex Laplace-Betrami operator

$$\square_F = \bar{\partial}\mathfrak{d} + \mathfrak{d}\bar{\partial}$$

---

[4]) Kodaira and Spencer, loc. cit., p. 870.

[5]) P. Dolbeault: Sur la cohomologie des variété analytiques complexes, Compt. Rend. (Paris), **236** (1953), pp. 175—177.

and denote by $H^{p,q}(F)$ the linear space consisting of all forms $\varphi = \{\varphi_j\}$ satisfying $\square_F \varphi = 0$.

Now let $\Omega^p(F)$ be the sheaf over M of germs of holomorphic p-forms with coefficients in F and let $H^q(M, \Omega^p(F))$ be the q-th cohomology group of M with coefficients in $\Omega^p(F)$.

Theorem 2 [6]. *We have the isomorphism*

$$H^q(M, \Omega^p(F)) \simeq H^{p,q}(F). \tag{2}$$

This result implies that *each cohomolog group* $H^q(M, \Omega^p(F))$ *is a finite dimensional* $\boldsymbol{C}$*-modul* [7]. As one readily infers, the mapping

$$\varphi = \{\varphi_j\} \to \varphi^\dagger = \{\varphi_j^\dagger\}, \qquad \varphi_j^\dagger = a_j^{-1} * \bar{\varphi}_j$$

maps $H^{p,q}(F)$ isomorphically onto $H^{n-p,n-q}(-F)$. Hence we obtain *the isomorphism* [8]

$$H^q(M, \Omega^p(F)) \simeq H^{n-q}(M, \Omega^{n-p}(-F)). \tag{3}$$

We write for simplicity's sake $\Omega(F)$ for $\Omega^0(F)$. By the *canonical bundle* K over M will be meant the complex line bundle of n-forms over M. Since $\Omega^n(-F) \simeq \Omega(K-F)$, we have, as a special case of (3), the isomorphism

$$H^q(M, \Omega(F)) = H^{n-q}(M, \Omega(K-F)). \tag{4}$$

Incidentally, *in case* M *is a Kähler manifold*, we infer from (1) the *important formula*

$$\dim H^q(M, \Omega^p) = \dim H^p(M, \Omega^q). \tag{5}$$

We say that a real $(1,1)$*-form*

$$\gamma = i \sum \gamma_{\alpha\bar{\beta}} dz^\alpha \wedge dz^\beta$$

on M *is positive and write* $\gamma > 0$ *if the Hermitian matrix* $(\gamma_{\alpha\bar{\beta}})$ *is positive definite at each point on* M. Moreover, *we say that a complex line bundle* F *is positive if its characteristic class* $c_R(F)$ *contains a closed real* $(1,1)$*-form* $\gamma > 0$. Now we can prove by a differential-geometric method due to Bochner [9] the following

---

[6]) K. Kodaira: On cohomology groups of compact analytic varieties with coefficients in some analytic faisceaux, Proc. Nat. Acad. Sciences, **39** (1953), pp. 865—868.

[7]) This result reduces to a special case of a theorem of H. Cartan and J. P. Serre: Un théorème de finitude concernant les variété analytiques compactes, Compt. Rend. (Paris), **237** (1953), p. 128.

[8]) This result reduces to a special case of Serre's duality theorem. See J. P. Serre: Un théorème de dualité, Comment. Math. Helv., **29** (1955), pp. 9—26.

[9]) S. Bochner: Vector fields and Ricci curvature, Bull. Amer. Math. Soc. **52** (1946), pp. 776—797; Curvatures and Betti numbers, I and II, Annals of Math., **49** (1948), pp. 379—390, **50** (1949), pp. 77—93.

Theorem 3[10]. *The cohomology group* $H^q(M, \Omega(F))$ *vanishes for* $q \geqq 1$ *if the bundle* $F - K$ *is positive.*

4. *An intrinsic characterization of algebraic varieties* [11]. By a *Hodge metric* on a compact complex manifold M we shall mean a Kähler metric $ds^2 = 2\sum g_{\alpha\bar{\beta}} dz^\alpha d\bar{z}^\beta$ such that the associated exterior form $\omega = i\sum g_{\alpha\bar{\beta}} dz^\alpha \wedge d\bar{z}^\beta$ belongs to a cohomology class c_R, $c \in H^2(M, \boldsymbol{Z})$.

Given a complex line bundle F over M, we associate with F a „meromorphic" mapping Φ_F of M into a projective space $\mathfrak{S}$ in the following manner: Take a base $\{\varphi_0, \varphi_1, \ldots, \varphi_\lambda, \ldots, \varphi_r\}$ of the linear space $H^0(M, \Omega(F))$ consisting of all holomorphic sections of F and denote by $\varphi_{\lambda j}(z)$ the fibre coordinate of $\varphi(z)$ over U_j. Then, considering $(\varphi_{0j}(z), \varphi_{1j}(z), \ldots, \varphi_{rj}(z))$ as the homogeneous coordinates of a point in the projective space $\mathfrak{S}$, we have

$$(\varphi_{0j}(z), \ldots, \varphi_{rj}(z)) = (\varphi_{0k}(z), \ldots, \varphi_{rk}(z)) \qquad \text{for } z \in U_j \cap U_k.$$

Hence we can define the mapping Φ_F of M into $\mathfrak{S}$ by

$$z \to \Phi_F(z) = (\varphi_{0j}(z), \varphi_{1j}(z), \ldots, \varphi_{rj}(z)).$$

Now we can prove, with the help of Theorem 3, the following

Theorem 4. *Let* M *be a compact complex manifold carrying a Hodge metric. Then there exists on* M *a real* $(1, 1)$*-form* β *such that, for any complex line bundle* F *over* M *whose characteristic class* $c_R(F)$ *contains a closed real* $(1, 1)$*-form* $\gamma > \beta$, *the mapping* $z \to \Phi_F(z)$ *is a bi-regular mapping of* M *into a projective space.*

It is well known that every non-singular algebraic manifold imbedded in a projective space carries a Hodge metric. Assume conversely M to be a compact complex manifold carrying a Hodge metric with the associated form $\omega = i\sum g_{\alpha\bar{\beta}} dz^\alpha \wedge d\bar{z}^\beta$. Since $c_R \ni \omega$ implies that c belongs to $H^2_{(1,1)}(M, \boldsymbol{Z})$, it follows from Theorem 1 that there exists a complex line bundle F_1 over M with $c_R(F_1) = c_R \ni \omega$. Letting $F = hF_1$, where h is a sufficiently large positive integer, we have $c_R(F) \ni h\omega > \beta$, and therefore, by the above Theorem 4, Φ_F is a bi-regular mapping of M into a projective space. Thus we obtain the following

Theorem 5. *A compact complex manifold* M *is (bi-regularly equivalent to) a non-singular algebraic manifold imbedded in a projective space if and only if* M *can carry a Hodge metric.*

5. *Picard varieties* [12]. Let V be a non-singular *algebraic* variety of dimension n imbedded in a projective space. It follows from Theorem 4 that *every*

---

[10]) K. Kodaira, On a differential-geometric method in the theory of analytic stacks, Proc. Nat. Acad. Sciences, **39** (1953), pp. 1268—1273.

[11]) K. Kodaira, On Kähler varieties of restricted type. Annals of Math., **60** (1954), pp. 28—48.

complex line bundle F over V is determined by a divisor $D : F = [D]$. Consequently *the group $\mathfrak{F}$ of line bundles over V coincides with the divisor class group $\mathfrak{D}$ of V.* Since $c([D])$ is the dual of the homology class of the $(2n-2)$-cycle D on V, the kernel $\mathfrak{P}$ of the homomorphism $[D] \to c([D])$ of $\mathfrak{D}$ into $H^2(V, \boldsymbol{Z})$ consists of all divisor classes $[D]$ of divisors $D \sim 0$, where $\sim$ denotes the homology with coefficients in $\boldsymbol{Z}$. It can be shown that

$$\mathfrak{P} \simeq \frac{H^1(V, \Omega)}{j^* H^1(V, \boldsymbol{Z})} \tag{6}$$

where j^* is the homomorphism induced by the inclusion map $j : \boldsymbol{Z} \to \Omega$. Combining this with the isomorphism $H^1(V, \Omega) \simeq H^{0,1}$, we obtain

$$\mathfrak{P} \simeq H^{0,1}/\mathfrak{J} \tag{7}$$

where $\mathfrak{J}$ is the image in $H^{0,1}$ of $j^* H^1(V, \boldsymbol{Z})$. By identifying $\mathfrak{P}$ with the *complex torus* $H^{0,1}/\mathfrak{J}$, we introduce on $\mathfrak{P}$ a complex analytic structure. $\mathfrak{P} = H^{0,1}/\mathfrak{J}$ is called the *Picard variety* attached to V. With the help of (5) we infer from (6) that $\dim \mathfrak{P}$ *is equal to the number* $g_1 = \dim H^0(V, \Omega^1)$ *of linearly independent simple differentials of the first kind on* V:

$$\dim \mathfrak{P} = g_1. \tag{8}$$

Now it follows from Theorem 1 that

$$\mathfrak{D}/\mathfrak{P} \simeq H^2_{(1,1)}(V, \boldsymbol{Z}). \tag{9}$$

This leads immediately to the following criterion for algebraic $(2n-2)$-cycles due to Lefschetz and Hodge [13]: *An integral $(2n-2)$-cycle C on an algebraic manifold V of dimension n is homologous to a divisor if and only if the harmonic part HC of C is of type* (1,1).

6. *Characteristic linear systems of a complete continuous systems.* For simplicity's sake we write $\Gamma(F)$ for $H^0(V, \Omega(F))$. We associate with any member $\varphi(z) = (z, \varphi_j(z))$ of $\Gamma(F)$ the *effective divisor* $D = (\varphi)$ defined in each U_j by the holomorphic function $\varphi_j(z)$. Obviously φ is determined uniquely up to a multiplicative constant by its divisor $D = (\varphi)$. Moreover, in case $F = [D]$, the mapping $\varphi \to (\varphi)$ maps $\Gamma(F)$ onto the complete linear system $|D|$. We have therefore

$$\dim |D| + 1 = \dim \Gamma(F), \quad (F = [D]). \tag{10}$$

---

[12]) K. Kodaira and D. C. Spencer: Divisor class groups on algebraic varieties, Proc. Nat. Acad. Sciences, **39** (1953), pp. 872—877.

[13]) W. V. D. Hodge: The theory and applications of harmonic integrals, Cambridge University Press (1941), pp. 214—216.

Let S be a non-singular prime divisor on V. For an arbitrary line bundle F over V we denote by F_S the *restriction* of F to S and by r_S the *restriction map* of $\Omega(F)$ onto $\Omega(F_S)$. Moreover we define the *deficiency* def (F/S) of F over S by

$$\text{def}(F/S) = \dim \Gamma(F_S) - \dim r_S^* \Gamma(F),$$

where r_S^* denotes the homomorphism of $\Gamma(F)$ into $\Gamma(F_S)$ induced by r_S. In case $F = [D]$, def$([D]/S)$ *is equal to the deficiency of the linear system on* S *cut out by* $| D |$. In particular def$([S]/S)$ *is the deficiency of the characteristic system of the complete linear sytem* $| S |$. By the *irregularity* q of the algebraic manifold V we shall mean the maximum of def$([S]/S)$ for all non-singular prime divisor S on V:

$$q = \max_S \text{def } ([S]/S).$$

Now let $E = [S]$. Then we have the exact sequence

$$0 \to \Omega \to \Omega(E) \overset{r_S}{\to} \Omega(E_S) \to 0.$$

From the corresponding exact cohomology sequence

$$\ldots \to \Gamma(E) \overset{r_S^*}{\to} \Gamma(E_S) \to H^1(V, \Omega) \to H^1\big(V, \Omega(E)\big) \to \ldots$$

we infer with the help of Theorem 3 that def $([S]/S) \leqq \dim H^1(V, \Omega)$ and that the equality def $([S]/S) = \dim H^1(V, \Omega)$ holds if the bundle $[S] - K$ is positive. Combined with (5), this proves that *the irregularity* q *of an algebraic manifold* V *is equal to the number* g_1 *of linearly independent simple differentials of the first kind on* V *and that*

$$\text{def } ([S]/S) = q \ (= g_1) \tag{11}$$

if the bundle $[S] - K$ *is positive* [14].

It is easy to show that there exists on $V \times \mathfrak{P}$ a complex line bundle Ξ such that for each point $P \in \mathfrak{P}$, the restriction $\Xi_{V \times P}$ of Ξ to the subvariety $V \times P$ coincides with the line bundle P, where we consider $\Xi_{V \times P}$ as a line bundle over V in an obvious manner. Let $\mathscr{E}$ be a sufficiently positive line bundle over $\mathfrak{P}$. $\mathscr{E}$ or $E = [S]$ induces a complex line bundle over $V \times \mathfrak{P}$ in a canonical manner which will be denoted by the same symbol $\mathscr{E}$ or E. Now, *assuming that* $E - K$ *is positive*, we form the sum $F = E + \Xi + \mathscr{E}$ on $V \times \mathfrak{P}$. Then, using Theorem 3, we can prove that

$$r^*_{V \times P} \Gamma(F) = \Gamma(E + P), \tag{12}$$

---

[14]) K. Kodaira and D. C. Spencer: On a theorem of Lefschetz and the lemma of Enriques-Severi-Zariski, Proc. Nat. Acad. Siences, **39** (1953), pp. 1273—1278.

where $r_{V\times P}$ is the restriction map to $V \times P$. It follows from (12) that $d = \dim \Gamma(E + P)$ is independent of P and that $\mathscr{B} = \bigcup_{P \in \mathfrak{P}} \Gamma(E + P)$ is an analytic fibre bundle over $\mathfrak{P}$ whose fibre is the d-dimensional complex vector space $\boldsymbol{C}_d$ and whose structure group is the general linear group $GL(d, \boldsymbol{C})$. Replacing each fibre $\boldsymbol{C}_d$ of $\mathscr{B}$ by the corresponding projective space $\mathfrak{S}_{d-1}$ of dimension $d - 1$, we obtain from $\mathscr{B}$ an analytic fibre bundle Λ over $\mathfrak{P}$ whose fibre is the projective space $\mathfrak{S}_{d-1}$. By using Theorem 5 we infer readily that Λ is a non-singular algebraic variety imbedded in a projective space. Obviously we have the canonical map $\varphi \to \lambda = \lambda(\varphi)$ of $\mathscr{B}$ onto Λ.

Let $\mathscr{C}$ be the set of all effective divisors $D \sim S$ on V. Then we see that the map $\lambda \to D = D_\lambda$ defined by

$$\lambda = \lambda(\varphi) \leftarrow \varphi \to D_\lambda = (\varphi)$$

is a one-to-one algebraic correspondence between Λ and $\mathscr{C}$. Thus $\mathscr{C}$ is a complete continuous system in the sense of Italian geometers. Now it can be shown by using (7) and (8) that *the dimension of the characteristic system of $\mathscr{C}$ on S is equal to*

$$\dim \Lambda - 1 = \dim |S| + g_1 - 1.$$

Combining this with (11) we conclude that the characteristic system of $\mathscr{C}$ on S is complete. Thus we obtain the following:

Theorem 6. *Let S be a non-singular prime divisor on a non-singular algebraic manifold V such that the bundle $[S] - K$ is positive. Then the complete continuous system $\mathscr{C}$ containing S is uniquely determined and consists of all effective divisors $D \sim S$ on V. $\mathscr{C}$ can be parametrized by a non-singular algebraic manifold Λ which is an analytic fibre bundle over the Picard variety $\mathfrak{P}$ attached to V whose fibre is a projective space. Thus $\mathscr{C}$ consists of ∞^q linear systems, q being the irregularity of V. Moreover the characteristic linear system of $\mathscr{C}$ on S is complete.*

PROPRIÉTÉS ARITHMÉTIQUES DE CERTAINES FAMILLES DE COURBES ALGÉBRIQUES

A. NÉRON

I. *Introduction.*

Rappelons l'une des nombreuses formes de l'énoncé du théorème d'irréductibilité bien connu de Hilbert [4]:

„Soit u une application rationnelle d'une variété algébrique V sur une variété linéaire U de même dimension et supposons qu'il existe un même corps algébrique fini k sur lequel V, U et u soient définies. Pour un point M générique de u sur k, supposons que les composants de $u^{-1}(M')$ soient irrationnels sur $k(M)$. Alors il existe une infinité de spécialisations M' rationnelles sur k de M telles que $u^{-1}(M')$ soit défini et que ses composants soient irrationnels sur k".

Diverses généralisations de ce théorème ont été obtenues depuis Hilbert. En particulier, les travaux de W. Franz [3], Eichler [2], Inaba [5] et, plus récemment, F. K. Schmidt et G. Preuss [8] ont mis en évidence l'existence de différentes catégories remarquables très étendues de corps k pour lesquels le théorème reste valable.

Le théorème d'irréductibilité est, par ailleurs, étroitement rattaché à certains problèmes relatifs aux points rationnels sur les courbes algébriques. En particulier, on peut l'utiliser pour la recherche de courbes de rang élevé sur un corps algébrique donné, en l'appliquant au cas où U est la variété de paramètres d'une famille de courbes convenablement choisie; j'ai obtenu, par ce procédé, des courbes de genre 1 et de rang 10 et, pour tout entier $g \geqq 2$, des courbes de genre g et de rang $\geqq 3g + 6$ sur le corps des nombres rationnels [6].

Dans le but d'améliorer ces valeurs, je me suis demandé s'il existait un théorème analogue au théorème de Hilbert dans lequel U soit remplacée par une variété autre qu'une variété linéaire (ou unicursale) et possédant suffisamment de points rationnels sur k. J'ai obtenu une réponse affirmative à cette question dans le cas où U est une courbe de genre 1 (ou, plus généralement, une variété abélienne) et déduit de là l'existence sur le corps des nombres rationnels de courbes dont le rang atteint des valeurs supérieures d'une unité aux précédentes, c'est-à-dire 11 pour $g = 1$ et $3g + 7$ pour $g \geqq 2$.

2. *Un résultat préliminaire.*

Lemme. Soient E et F deux espaces numériques réels (resp. complexes) et soient G et H des sous-groupes de type fini de E et F respectivement. Soit f une fonction analytique réelle (resp. complexe) définie sur une partie compacte $\overline{E}$ de E, à valeurs dans F et qui n'induit une fonction polynomiale sur aucune droite de E recontrant $\overline{E}$. A tout $a \in G$ associons sa décomposition $a = \sum_i p_i a_i$ suivant une base $\{a_i\}$ $(i = 1, \ldots r)$ de G et à tout $b \in H$ associons sa décomposition $b = \sum_j q_j b_j$ suivant une base $\{b_j\}$ $(j = 1, \ldots s)$ de H. Soit ϱ un nombre réel positif. Pour tout entier N soient $\varphi(N)$ le nombre des points de $G \cap \overline{E}$ tels que:

$$|p_i| \leqq N \text{ pour tout } i \tag{1}$$

et $\psi(N)$ le nombre de ces mêmes points pour lesquels de plus $b = f(a)$ appartient à H et satisfait à

$$|q_j| \leqq N^{\varrho} \text{ pour tout } j \tag{2}$$

Alors il existe un nombre réel α $(0 < \alpha < 1)$ tel qu'on ait:

$$\psi(N) \leqq \varphi(N)^{\alpha}$$

dès que N est assez grand.

On peut se borner au cas réel, se ramener par projection au cas où F est une droite, supposer que G n'est pas contenu dans un sous-espace de E distinct de E et qu'on a $r > n$, en désignant par n la dimension de E.

Pour tout entier N, notons $\mathscr{G}(N)$ l'ensemble des points de G qui satisfont à (1) et par $\mathscr{H}(N)$ le sous-ensemble de $\mathscr{G}(N)$ formé des points a tels que $b = f(a)$ appartienne à H et satisfasse à (2). Une étude élémentaire de l'ensemble $\mathscr{G}(N)$ permet de montrer qu'il existe des constantes réelles positives N_0, β_0 et γ telles qu'à tout entier $N \geqq N_0$ et à tout nombre réel positif $\beta \leqq \beta_0$ on puisse associer un point $a(N, \beta)$ de G possédant les propriétés suivantes:

(I) L'ensemble $\mathscr{G}^*(N, \beta)$ des multiples entiers m. $a(N, \beta)$ de ce point tels que

$$m \leqq N^{\beta} \tag{3}$$

est contenu dans $\mathscr{G}(N)$

(II) On a

$$\| a(N, \beta) \| \leqq N^{-\gamma} \tag{4}$$

où $\| \ \|$ désigne la norme définie par la plus grande des valeurs absolues des coordonnées dans E

(III) On peut recouvrir $\mathscr{G}(N) \cap \overline{E}$ par des ensembles deux à deux sans point commun déduits de $\mathscr{G}^*(N, \beta)$ par des translations de manière que le nombre des points de la réunion de ces ensembles soit $\leqq 2\varphi(N)$.

Supposons la proposition fausse. Alors, quel que soit α, on peut trouver une suite infinie croissante d'entiers $N_1, N_2 \ldots N_k \ldots$ telle qu'on ait, pour tout k,

$$\psi(N_k) > \varphi(N_k)^{\alpha}$$

Prenons $\beta < \inf\left(\frac{r-n}{n},\ \beta_0\right)$ et $\alpha > 1 - \frac{\beta}{r-n}$.

D'après (III), il existe, pour tout k, un sous-ensemble $\mathscr{H}_k$ de $\mathscr{H}(N_k)$ déduit d'un sous-ensemble de $\mathscr{G}^*(N_k, \beta)$ par une translation définie par un élément d_k de $\overline{E}$ et dont le nombre des éléments est $\geqq \varkappa N_k^\delta$ où $\varkappa$ désigne une constante positive et où l'on pose $\delta = \beta - (r-n)(1-\alpha)$.

On peut supposer que f admet un développement en série convergent dans tout $\overline{E}$. A toute droite D de E rencontrant $\overline{E}$ et à toute représentation paramétrique de D de la forme $a = cu + d$, avec $c \in \overline{E}$, $||c|| = 1$ et $d \in \overline{E}$, associons la fonction

$$f_D(u) = \sum_{\lambda=0}^{\infty} A_\lambda u^\lambda$$

induite par f sur D. Par hypothèse, cette fonction n'est jamais un polynome. En utilisant un recouvrement fini convenable de l'ensemble des points $c \times d$ dans $E \times E$, on voit qu'on peut trouver une constante réelle positive ξ et un nombre fini d'ensembles $\mathscr{E}_h = \{l_{\nu,h}\}$ $(\nu = 1, \ldots s)$ de s entiers positifs satisfaisant aux conditions suivantes:

(I') pour tout couple $\{h, \nu\}$, on a

$$l_{\nu,h} - l_{\nu-1,h} > \frac{s\varrho}{\gamma} \tag{5}$$

(II') A toute droite D rencontrant E on peut faire correspondre un h tel que, pour la fonction f_D correspondante, on ait

$$A_{l_{\nu,h}} \geqq \xi \text{ pour tout } \nu \tag{6}$$

Posons $\bar{l} = \sup_{\nu,h}(l_{\nu,h})$. Appliquons ce qui précède à la droite D_k qui contient $\mathscr{H}_k$, en prenant $d = d_k$. Posons $f_{D_k}(u) = \sum_{\lambda=0}^{\infty} A_{\lambda,k} u^\lambda$. Soit $\mathscr{E}_{h_k}$ l'ensemble d'entiers correspondant. Nous écrirons pour simplifier l_ν au lieu de l_{ν,h_k}.

Pour tout $a \in \mathscr{H}_k$ et pour tout $l < l_1$ on a une relation de la forme

$$\Sigma_j q_j b_j - Q_l(m) \leqq \eta N_k^{\beta-\gamma} \tag{7}$$

où η est une constante indépendante des entiers l et k, et où Q_l est le polynome défini par $Q_l(m) = \sum_{\lambda=0}^{l-1} B_{\lambda,k} m^\lambda$, avec

$$B_{\lambda,k} = A_{\lambda,k} || a(N_k, \beta) ||^\lambda \tag{8}$$

Supposons k assez grand pour que $N_k^\delta \geqq \bar{l} + s$ et considérons, pour tout ν et pour tout ensemble $\{a_\mu\}$ $(\mu = 1, \ldots l'_\nu)$ de $l'_\nu = l_\nu + s$ points de $\mathscr{H}_k$, la matrice carrée d'ordre l'_ν dont les éléments de la μ ième ligne sont les entiers

$$q_j(a_\mu) \quad (j = 1, \ldots s) \quad \text{et} \quad [m(a_\mu)]^\lambda \quad (\lambda = 0, 1, \ldots l_\nu - 1)$$

Si le déterminant de cette matrice est $\neq 0$ pour un choix convenable de l'entier ν et de l'ensemble $\{a_\mu\}$, on peut trouver un ensemble d'entiers $\{t_\mu\}$ tels que $\sum_\mu t_\mu q_j(a_\mu) = 0$ pour tout j et que $\sum_\mu t_\mu [m(a_\mu)]^\lambda$ soit $\neq 0$ pour $\lambda = l_{\nu-1}$ et nul pour toute autre valeur $< l_\nu$. En combinant linéairement les relations (7) correspondant aux différents a_μ pour $t = t_\nu$, on obtient

$$B_{l_{\nu-1}} | \sum_\mu t_\mu (m(a_\mu))^{l_{\nu-1}}| \leqq \eta N_k^{l_\nu(\beta-\gamma)} \sum_\mu | t_\mu |$$

Or on peut supposer les t_μ choisis tels que

$$| t_\mu | \leqq \sup_{j,\mu} | q_j (a_\mu)|^s \sup_\mu | m(a_\mu)|^{\frac{1}{2} l_\nu (l_\nu - 1)}$$

pour tout μ et tout ν. Compte tenu des relations (2), (3), (4), (6) et (7), on obtient

$$N_k^{\zeta_\nu} \leqq \omega$$

où ω est une constante réelle positive et où

$$\zeta_\nu = \gamma(l_\nu - l_{\nu-1}) - s\varrho - \tfrac{1}{2}\beta l_\nu (l_\nu + 1)$$

Ceci est contradictoire pour k assez grand si on a pris β assez petit pour que les ζ_ν soient tous strictement positifs.

Si le déterminant de la matrice précédente est toujours nul, il existe pour tout ν une relation

$$L(q_1 \ldots q_s) = R_\nu(m),$$

où L est une forme linéaire et R_ν un polynome de degré effectif t_ν, satisfaite par tout point de $\mathscr{H}_k$. Par élimination des q_j, on en tire une relation algébrique $R(m) = 0$ satisfaite par tout $a \in \mathscr{H}_k$ Ceci est contradictoire avec le fait que le nombre des points de $\mathscr{H}_k$ tend vers l'infini avec k.

3. *Une variante du théorème d'irréductibilité.*

Théorème 1. *Soient A et B deux variétés abéliennes, u un homomorphisme de A sur B, et W une sous-variété non abélienne de A de même dimension n que B telle que $u(W)$ ait pour support B. Supposons que A, B, W et u soient définis sur un même corps algébrique fini k et que B possède une infinité de points rationnels sur k, non contenus dans une sous-variété abélienne de B. Pour tout nombre réel h soit $\varphi(h)$ le nombre des points M de B tels que l'intersection $u^{-1}(M)$. W soit définie et dont la hauteur* [1] *est inférieure à h et soit $\psi(h)$ le nombre de ces mêmes points pour*

[1]) Rappelons qu'on appelle hauteur d'un point algébrique M de l'espace projectif P_n, dont les coordonnées x ($= 0, 1, \ldots n$) sont des entiers d'un corps algébrique fini k l'expression $h(M)$ définie par

$$(h(M))^d = (N\omega)^{-1} \Pi_\sigma (\sup_i |x_\sigma^i|)$$

ou d désigne le degré de k, ω le *p.g.c.d.* des x_i et $N\omega$ sa norme relativement à k et où le produit Π est étendu à tous les isomorphismes de k dans le corps des nombres complexes. Cette notion, due à Weil [10] est équivalente à celle de complexité, introduite auparavant par Northcott [7].

lesquels de plus l'un des composants de $u^{-1}(M)$. *W est rationnel sur k. Alors, quel que soit le nombre réel positif* ε, *on a*

$$\psi(h) < \varepsilon\varphi(h)$$

dès que h est assez grand.

On sait que toute variété abélienne est en correspondance analytique biunivoque avec un tore muni d'une structure complexe. On sait, d'autre part, d'après Weil [9], que les points rationnels sur k d'une variété abélienne définie sur k forment un sous-groupe de type fini de cette variété. Soient donc $T = E/G$ et $U = F/H$ les tores respectivement associés à A et B, où l'on désigne par E et F des espaces numériques complexes et par G et H des sous-groupes convenables de ces espaces. Les images dans T et U des groupes des points rationnels sur k de A et B respectivement peuvent être représentés par des sous-groupes de type fini G' et H' de E et F.

Soit B_0 le sous-ensemble algébrique de B comprenant d'une part les points M en lesquels $u^{-1}(M)$. W n'est pas définie, ou admet des composants multiples, d'autre part les images par u des sous-variétés abéliennes de A contenues dans W (on peut, par récurrence sur n, exclure le cas où il existerait une infinité de telles sous-variétés, nécessairement déduites de l'une d'elles par des translations, dont la réunion recouvre W).

Le tore T (resp. U) est en correspondance biunivoque naturelle avec le parallélotope T' (resp. U') construit avec les éléments d'une base arbitraire de G(resp. H). Soient W', B'_0, u' les images respectives de W, B_0, u dans T', U' et $T' \times U'$. Soit γ un nombre réel > 0 et soit T'_0 l'intersection de T' avec la réunion des cubes de côté γ centrés aux différents points de B'_0.

On peut appliquer le lemme du § 2 aux espaces E et F, à leurs sous-groupes respectifs $G + G'$ et $H + H'$, au sous-ensemble $\overline{E} = T' - T'_0$ de E et à chacune des fonctions qui font correspondre à tout $a \in \overline{E}$ l'un des composants de $u'^{-1}(a) \cap W'$ On prend ensuite γ suffisamment petit et l'on tient compte des propriétés de la notion de hauteur (voir [6], Chap. III).

Théorème 2. *Soient V une courbe de genre* $\geqq 2$, *B une courbe de genre* 1, *et v une fonction rationnelle sur V non constante, à valeurs dans B. Supposons que V, B et v soient définis sur un même corps algébrique fini k. A tout K associons le nombre* $\varphi(h)$ *des points M de B rationnels sur k et de hauteur* $\leqq h$ *et le nombre* $\psi(h)$ *de ces mêmes points pour lesquels de plus l'un des composants de* $v^{-1}(M)$ *est rationnel sur k. Alors, quel que soit le nombre réel positif* ε, *on a*

$$\psi(h) < \varepsilon\varphi(h)$$

dès que h est assez grand (2).

Dans cet énoncé, la condition „V est de genre $\geqq 2$" est équivalente à la suivante: „*pour M générique, tout composant P de $v^{-1}(M)$ est tel que $k(P)$ soit une extension ramifiée de $k(M)$*".

On peut supposer B et V sans point multiple et identifier V avec son image canonique dans sa jacobienne J. On applique le théorème 1 en prenant pour A le produit $J \times B$, pour u la projection sur B et pour W le graphe de v dans $V \times B$.

Le théorème 2 s'étend enfin au cas où l'on remplace B par une variété abélienne de dimension $\geqq 2$ et V par une variété non birationnellement équivalente à une variété abélienne, compte tenu de résultats annoncés par Chow relatifs à la „variété d'Albanese" ou „variété de Picard duale" d'une variété algébrique.

4. *Propriétés arithmétiques des familles de courbes algébriques paramétrées par une variété abélienne.*

Soit B une variété abélienne définie sur un corps algébrique fini k et ayant une infinité de points rationnels sur k. Soient M un point générique de B sur k et C une courbe définie sur $K = k(M)$. Considérons le groupe de Mordell-Weil γ attaché à C et K et, pour toute spécialisation $M \to M'$ rationnelle sur k, considérons la spécialisation $C \to C'$ correspondante. Cherchons, en appliquant la même méthode que dans ([6], Chap. IV) à obtenir une infinité de points M' tels que le groupe γ' attaché à C' et k contienne un sous-groupe isomorphe à γ. Considérons sur la jacobienne J de C (qu'on peut supposer définie sur K, ainsi que la fonction canonique correspondante) une base $\{a_i\}$ $(i = 1, \ldots r)$ de γ, et à tout M' associons la spécialisation correspondante $\{a'_i\}(i = 1, \ldots r)$. Soient s un entier arbitraire et $\{u_h\}$ un système de représentants des classes du groupe fini $\gamma/s\gamma$ et, pour tout h, considérons les points $v_{\nu,h}$ de J qui satisfont à $sv_{\nu,h} = u_h$ et qui sont irrationnels sur K. Alors pour que le groupe engendré par les a'_i soit isomorphe à γ, il suffit de prendre M' tel que les spécialisations $v'_{\nu,h}$ des $v_{\nu,h}$ soient irrationnelles sur k.

Le théorème 2 permet d'affirmer qu'il existe bien une infinité de telles spécialisations, à condition de s'assurer que les extensions $K(v_{\nu,h})$ de K sont toutes ramifiées.

---

2) Ce résultat est naturellement contenu dans la conjecture de Mordell: „Toute courbe de genre $\geqq 2$ définie sur un corps algébrique fini k ne contient qu'un nombre fini de points rationnels sur k". Rappelons que C. Chabauty a démontré cette conjecture pour toute courbe dont le rang réduit est inférieur au genre [1].

5. *Existence de cubiques de rang* $\geqq 11$ *sur le corps des nombres rationnels.*

Dans tout ce qui suit, k désigne le corps des nombres rationnels. Soient dans un plan 3 droites rationnelles $D_i (i = 1, 2, 3)$ concourantes en un point O et sur chacune de ces droites un point rationnel P_i distinct de O. On peut trouver une cubique C_1 de genre 1 tangente, pour tout i, à D_i en P_i, ne passant pas par O et ayant une infinité de points rationnels. Considérons sur C_1 les tangentiels Q_i des points $P_i (i = 1, 2, 3)$ et adjoignons leur 5 autres points de C_1 de manière que les 8 points Q_j obtenus $(j = 1, \ldots 8)$ soient tels que 3 quelconques d'entre eux ne soient pas alignés ni 6 sur une même conique. Soit C une courbe générique du pinceau de cubiques π ayant pour points de base les Q_j et soient, pour $i = 1, 2, 3$, M_i et M_i' les points de $C \cap D_i$ autres que O et P_i. Alors, si, les Q_j ne satisfont pas à certaines conditions algébriques en nombre fini, le lieu B de $M = M_1 \times M_2 \times M_3$ dans le produit $D_1 \times D_2 \times D_3$ est une courbe de genre 1, qui possède un point rationnel M_0 (associé à la courbe C_0 de π qui passe par O). L'application du théorème d'irréductibilité (sous la forme de Hilbert) montre qu'on peut choisir les données de façon que M_0 ne soit pas d'ordre fini dans le groupe des points rationnels de B et que les 11 points Q_j et M_i soient indépendants dans le groupe des points de C rationnels sur $K = k(M)$. Enfin, la considération des spécialisations de M associées aux courbes de genre zéro de π permet de montrer que les extensions de K notées plus haut $K(v_{\nu,h})$ sont toutes ramifiées. Le théorème 2 entraîne l'existence d'une infinité de cubiques de rang réduit $\geqq 11$ sur k.

6. *Existence, pour* $g \geqq 2$, *de courbes de genre* g *et de rang* $\geqq 3g + 7$ *sur le corps des nombres rationnels.*

Soient encore trois droites D_i dans un plan concourantes en un point O et sur chacune d'elles un point rationnel P_i distinct de O. Soient de plus g points rationnels $S_k (k = 1, \ldots g)$ non situés sur les droites D_i. On peut montrer que, pour un choix convenable de ces points:

1. Il existe une courbe unicursale C_1 définie sur k, de degré $g + 2$, ayant O comme point multiple d'ordre g, les S_k comme points doubles et, pour tout i, tangente à D_i en P_i.

2. Il existe une courbe C_0 définie sur k, de genre g, de degré $g + 2$, ayant O comme point multiple d'ordre g, les S_k comme points simples, telle que $C_0 \cap C_1$ contienne en tout $2g + 4$ points Q_j distincts, tous rationnels, autres que O, les P_i et les S_k et que, pour tout i, $C_0 \cap D_i$ contienne en tout 2 points distincts M_{i0}, M_{i0}' tous deux rationnels, autres que O et P_i.

3. Désignant par π le pinceau déterminé par C_0 et C_1, par C une courbe générique de π et, pour tout i, par M_i et M_i' les points de $C \cap D_i$ autres que O et P_i, le lieu B de $M = M_1 \times M_2 \times M_3$ dans le produit $D_1 \times D_2 \times D_3$ est

une courbe de genre 1 possédant un point rationnel M_0 (Associé à C_0) qui est d'ordre infini dans le groupe des points rationnels de B.

4. On peut construire avec les points $M_i (i = 1, 2, 3)$ Q_j $(j = 1, \ldots 2g + 4)$ et $S_k (k = 1, \ldots g)$ un sous-groupe libre de rang $3g + 7$ du groupe γ attaché à C et à $K = k(M)$.

L'application du théorème 2 entraîne alors l'existence d'une infinité de courbes de π qui sont de genre g et de rang réduit $\geqq 3g + 7$ sur k.

BIBLIOGRAPHIE

[1] C. CHABAUTY, Sur les points rationnels des courbes algébriques de genre supérieur à l'unité (C. R. Acad. des Sc. de Paris. 1941, p. 882).

[2] M. EICHLER, Zum Hilbertschen Irreduzibilitätsatz (Math. Ann. Berlin, Bd. 116, 1939, p. 742).

[3] W. FRANZ, Untersuchungen zum Hilbertschen Irreduzibilitätsatz (Math. Z. Berlin, Bd 33, 1931, p. 275).

[4] D. HILBERT, Über die Irreduzibilität ganzen rationaler Funktionen mit ganzzahligen Koeffizienten. (Journal für die reine und angew. Math. Bd. 110, 1892 p., 104).

[5] E. INABA, Über den Hilbertschen Irreduzibilitätsatz (Japan J. Math. Bd. 19, (1944, p. 1).

[6] A. NÉRON, Problèmes arithmétiques et géométriques rattachés à la notion de rang d'une courbe algébrique dans un corps (Bulletin dc la Soc. Math. de France, 1952).

[7] D. G. NORTHCOTT, An inequality in the theory of arithmetic on algebraic varieties. (Proc. Cambridge philosophical soc. t. 45, 1949, p. 502).

[8] G. PREUSS et F. K. SCHMIDT, Über den Hilbertschen Irreduzibilitätsatz (Math. Nachr. Bd. 4, 1951, p. 348).

[9] A. WEIL, L'arithmétique sur les courbes algébriques (Acta mathematica, t. 52, 1929).

[10] A. WEIL, Arithmetic on algebraic varieties. (Ann. of Math. vol 53, No. 3, 1951).

SPECIALIZATION METHODS IN ALGEBRAIC GEOMETRY

D. G. NORTHCOTT.

Modern algebra has been applied with considerable success to problems which lie at the foundations of algebraic geometry, but it is none the less true that much of the language of abstract algebra is not well suited to such applications. For example, the concept of a *homomorphism* does not lend itself to the kind of pictorial thinking which is important in geometry. Fortunately, in this instance, the notion of a *specialization*, which we owe to van der Waerden, provides a substitute which is free from this objection. The ways in which this fruitful idea has been employed are very numerous and yet, it seems to me, one possible line of development has largely been neglected.

Let us recall the customary definition. Suppose that k is a field which is embedded in some large extension field Ω and let
$(x_1, x_2, \ldots, x_r)$ and $(\alpha_1, \alpha_2, \ldots, \alpha_r)$ be two sets each consisting of r elements taken from Ω. Then $(\alpha_1, \alpha_2, \ldots, \alpha_r)$ is said to be a specialization of $(x_1, x_2, \ldots, x_r)$ with respect to k if the two conditions

(i) $\varphi(X_1, X_2, \ldots, X_r)$ is a polynomial with coefficients in k;

(ii) $\varphi(x) = 0$;

together imply that $\varphi(\alpha) = 0$. When we have this situation we shall write $[x] \to [a]$ (*w. r. t.* k). Clearly $[x] \to [a]$ (*w. r. t.* k) is equivalent to each of the following two assertions:

(*a*) *there is a homomorphism of* $k[x]$ *on to* $k[\alpha]$ *in which* x_i *is mapped on to* α_i *and every element of* k *is left fixed*;

(*b*) *the variety, whose generic point over* k *is* (x), *contains the point* (α).

Suppose now that we wish to examine this question. *Given a specialization* $[x] \to [\alpha]$ (*w. r. t.* k) *and elements* $y_1, y_2, \ldots, y_s$ *in* Ω, *what elements* $\beta_1, \beta_2, \ldots, \beta_s$ *will have the property that*
$[x, y] \to [\alpha, \beta]$ (*w. r. t.* k)? This, it will be recognised, is a fundamental problem in the theory of irreducible correspondences. In different language, the problem is to find an extension of the homomorphism $k[x] \to k[\alpha]$ which will operate on the elements (y). Now the homomorphism $k[x] \to [\alpha]$ has as its kernel a prime ideal $\mathfrak{p}$. Let Q be the ring of quotients of $k[x]$ with respect to $\mathfrak{p}$ then a typical element of Q is of the form $f(x)/g(x)$ where $g(\alpha) \neq 0$. By mapping $f(x)/g(x)$ into $f(\alpha)/g(\alpha)$ we extend $k[x] \to k[\alpha]$ to a homomorphism of Q. The ring Q is,

of course, a local domain and if $\mathfrak{m}$ is its maximal ideal our homomorphism is (effectively) the natural mapping of Q on to its residue field $Q/\mathfrak{m} = K$ (say). We now see that we have

$$[x, y] \to [\alpha, \beta] \quad (w.\ r.\ t.\ k)$$

if and only if there is a homomorphism

$$Q\,[y] \to K\,[\beta]$$

in which y_i is mapped on to β_i and Q is mapped canonically on to its residue field. Such a homomorphism exists, however, if (and only if) the two conditions

(i) $\varphi\,(Y_1, Y_2, \ldots, Y_s)$ is a polynomial with coefficients in Q;

(ii) $\varphi\,(y) = 0$

together imply that $\bar{\varphi}(\beta) = 0$. Here $\bar{\varphi}(Y)$ is the polynomial obtained from $\varphi(Y)$ by reading it modulo $\mathfrak{m}$. A natural extension of our previous notation suggests itself at this point. We shall indicate this situation, when it arises, by writing

$$[y] \to [\beta] \quad (w.\ r.\ t.\ Q).$$

Specializations with respect to local a domain. The notation, which has just been introduced, virtually defines a specialization with respect to a local domain and we have been led to make this definition by simple considerations concerning a basic problem in the theory of correspondences. We shall come back to this fact later, but, at this point, let us examine, for a moment, the definition itself. It is worth noting, because it simplifies matters, that it will work for abstract local domains as well as for those which arise geometrically. When Q is an abstract local domain, the elements $y_1, y_2, \ldots, y_s$ must be taken from a field containing Q and $\beta_1, \beta_2, \ldots, \beta_s$ must come from an extension field of $K = Q/\mathfrak{m}$. No other comment is required to make precise the meaning of $[y] \to [\beta]$ $(w.\ r.\ t.\ Q)$ in this general case. At the other extreme, Q may be a field because a field is a very special kind of local domain. There is, however, no danger of confusion because $[y] \to [\beta]$ $(w.\ r.\ t.\ Q)$, interpreted according to the new definition, has the same significance as in the theory of van der Waerden.

Fortunately our progress consists of more than the production of an extra definition. As soon as one begins to think in terms of specializations with respect to a local domain, certain lines of attack stand out which were not at all obvious previously. Suppose, for instance, that the elements (y) belong to the quotient field of Q as will always happen if we apply this method to the study of a *birational* correspondence. Let $\overline{Q}$ be the completion of Q and to avoid complications let us suppose that $\overline{Q}$ is analytically irreducible, i.e. that $\overline{Q}$

is a domain. The elements (y) are now in the quotient fields of both Q and $\overline{Q}$ and, moreover, Q and $\overline{Q}$ have the same residue field. We can therefore compare the specializations of (y) with respect to Q, with the specializations of (y) with respect to $\overline{Q}$. The situation which one finds is gratifyingly simple because it turns out that the possible specializations are exactly the same in the two cases. It follows that we may, in certain circumstances, restrict our attention to *complete* local domains and this can be a great advantage. As an illustration I should like to give the chief parts of a proof of an important theorem on birational correspondences. This theorem is due to Zariski and it is usually referred to as the „Main Theorem."

The Main Theorem. Let T be a birational correspondence between two varieties V and V' and suppose that T has no fundamental elements on V'. By the transform $T\ [W]$ of W is meant the (generally reducible) subvariety of V' which satisfies:

(a) Each component of $T\ [W]$ corresponds to W under T;

(b) Every irreducible subvariety of V', which corresponds to W, is contained in $T\ [W]$.

Since there are no fundamental elements on V', it follows that if W' is a component of $T\ [W]$ then $\dim W \leqq \dim W'$. This is an elementary result. The Main Theorem, or rather a slight variant of it, asserts the following: *If V is analytically irreducible at W, and if there is a component W' of $T\ [W]$ with the property that* $\dim W = \dim W'$ *then W' is the only subvariety of V' which corresponds to W.*

Our previous remarks show that this theorem can be translated into one concerning the specializations, with respect to an analytically irreducible local domain Q, of a set of elements $(y_1, y_2, \ldots, y_s)$ belonging to its quotient field. More precisely, the corresponding result is this:

Suppose that $[y] \to [\beta]$ (w. r. t. Q) and that the β_i are algebraic over K. Suppose further that the specialization is such that if $[y] \to [\gamma]$ (w. r. t. Q) and $[\gamma] \to [\beta]$ (w. r. t. K) then necessarily $\dim_K(\gamma) = \dim_K(\beta)$. *In these circumstances, the only specializations of (y) w. r. t. Q are the conjugates of (β) over K.*

The proof of this is sufficiently short to be given in full.

Proof. As has already been observed we may suppose, without loss of generality, that Q is complete. Let $R = Q\ [y_1, y_2, \ldots, y_s]$ and let $\mathfrak{p}$ be the kernel of the homomorphism $R = Q\ [y] \to K\ [\beta]$, then our second hypothesis shows that $\mathfrak{p}$ is a minimal prime ideal of $R\mathfrak{m}$. Let Q' be the ring of quotients of R w. r. t. $\mathfrak{p}$ and let $\mathfrak{m}'$ be its maximal ideal. Then $Q'\mathfrak{m}$ is $\mathfrak{m}'$-primary and $Q'/\mathfrak{m}' = K(\beta)$ is algebraic over $Q/\mathfrak{m} = K$ consequently, since Q is complete, Q' is a finite Q-module.

Let $\mathfrak{P}$ be a prime ideal of $Q\ [y_1, y_2, \ldots, y_s]$ which contracts to $\mathfrak{m}$ in Q so

that $\mathfrak{P} \cap Q = \mathfrak{m}$. The assertion that the specializations of (y) with respect to Q are the conjugates of (β) over K is equivalent to saying that $\mathfrak{P}$ and $\mathfrak{p}$ must be the same. Now Q' and so, *a fortiori*, $Q[y]$ is an integral extension of Q consequently, since $\mathfrak{P} \cap Q = \mathfrak{m}$ is a maximal ideal, $\mathfrak{P}$ is a maximal ideal. Since Q' is an integral extension of Q it is also an integral extension of $Q[y]$, accordingly there is a maximal ideal of Q', which contracts to $\mathfrak{P}$. But $\mathfrak{m}'$ is the only maximal ideal of Q', hence $\mathfrak{P} = Q[y] \cap \mathfrak{m}' = R \cap \mathfrak{m}' = \mathfrak{p}$ by the properties of rings of quotients. The proof is now complete.

There is another application of these generalised specializations which concerns a key property of cones of tangents, but otherwise, so far as I know, their possibilities remain unexplored.

SHEFFIELD, ENGLAND.

GROUP VARIETIES AND DIFFERENTIAL FORMS

MAXWELL ROSENLICHT

1. The Albanese variety A of a given algebraic variety V is an abelian variety having the following characteristic properties:

1) there exists a rational map $\varphi : V \to A$ such that $\varphi(V)$ generates A.

2) if A' is an abelian variety and $\varphi' : V \to A'$ is a rational map, then $\varphi' = \psi\varphi$, where ψ is a rational map of A into A'.

If V is a nonsingular variety in a complex projective space, A and φ can be constructed as follows: Let $(\omega_1, \ldots, \omega_q)$ be a basis for the simple differentials of first kind on V. Then A is complex analytically homeomorphic to the multitorus which is the additive group of the space of q complex variables modulo the subgroup consisting of the lattice of periods of $(\omega_1, \ldots, \omega_q)$. To define φ, for any point $p \in V$ let $\varphi(p)$ be the point of A corresponding to the point $\left(\int_\gamma \omega_1, \ldots, \int_\gamma \omega_q\right)$ of the space of q complex variables, where γ is any path on V from p_0 to p, p_0 being some fixed point of V. The knowledge of A and φ conversely gives us $\omega_1, \ldots, \omega_q$, for these are merely the images under φ^{-1} of a basis for the invariant differential forms on A.

The most important case of the above is that in which V is a nonsingular curve; in this case A is its jacobian variety.

In abstract algebraic geometry periods of differentials have no meaning, but Weil has shown how many classical arguments involving differentials of the first kind can be made valid generally by replacing integration by maps of varieties into abelian varieties. Using this method, for example, one immediately gets a valid general proof of the theorem that the group of birational maps of a curve of genus greater than one onto itself is finite. In what follows we indicate how abstract analogues of certain results concerning differential forms with singularities (in particular simple differentials of second kind) can be given in terms of noncomplete commutative algebraic group varieties. In any characteristic, a standard Lie theory argument shows that a group variety of dimension n has exactly n linearly independent left invariant differential forms of degree one. These invariant differentials enable one to restate the results in a more "classical" language, but in the case of characteristic $p \neq 0$ the restatements in terms of differentials give very incomplete descriptions of what happens.

2. If G is a group variety and H a normal algebraic subgroup of G then G/H can be considered a group variety in a natural way and when this is done the usual homomorphism theorems hold. We say that G is an extension of H by G/H. If H is solvable (i.e. if there exists a normal chain of algebraic subgroups going down from H to $\{e\}$ such that the quotient of two successive members is either the multiplicative or the additive group in one variable) one can show that there exist cross sections for G/H in G; in particular G is then birationally equivalent to $H \times G/H$. The following is the main theorem on the structure of commutative group varieties.

Any commutative group variety is an extension of a solvable group variety by an abelian variety. A solvable commutative group variety is biregularly isomorphic to the direct product of a certain number of multiplicative groups in one variable by a group that is biregularly equivalent to an affine space; in characteristic zero this affine group is biregularly isomorphic to a product of additive groups in one variable, but this need not be so in characteristic $p \neq 0$.

The most important special examples of such commutative group varieties which are not trivial direct products are the so-called generalized jacobian varieties. These can be constructed from curves with singularities in a way analogous to that used to get the ordinary jacobian variety of a nonsingular curve. If J is the generalized jacobian corresponding to the curve V, there exists a natural rational map φ of V into J having properties similar to the universal mapping property of φ when V is nonsingular. Generalized jacobians are important largely because they can be used to obtain information about arbitrary commutative group varieties and because they are known rather explicitly; in fact the generalized jacobian of a curve V is an extension of a certain solvable group variety, whose structure depends only on the nature of the singularities of V, by the ordinary jacobian variety of V.

3. In the case of a function field of one variable of characteristic zero a differential is of the second kind at a given place if its residue at that place is zero. The differential can then be approximated as closely as one wishes at the given place by exact differentials. If the characteristic is $p \neq 0$, trouble is caused by differentials of the form dx/x^{np+1} which all have zero residue when $x = 0$ but cannot be closely approximated by exact differentials. To guarantee the finite dimensionality of the vector space of differentials that are everywhere of the second kind modulo exact differentials one has to exclude these bad differentials and hence one defines a differential to be of second kind if at each place it can be expressed as the sum of an exact differential and a differential that has no pole at the place. The following fact gives a rationale for this definition.

If $\varphi : V \to J$ is the canonical map of a curve into its generalized jacobian, then the inverse images on V of the invariant differential forms on J are all of the second kind if and only if the maximal solvable subgroup of J is the direct product of additive groups in one variable.

In the classical case of a nonsingular algebraic variety in a complex projective space the main theorem on simple differentials of second kind says that the vector space of differentials of second kind modulo exact differentials has dimension equal to twice the irregularity. Consideration of the canonical map into the Albanese variety shows that the theorem can be considered as one that concerns only abelian varieties and as such deals with the possible extensions of direct products of additive groups in one variable by a given abelian variety. From this point of view the main theorem is:

Let A be an abelian variety. Let G be a commutative group variety which is an extension of a product of additive groups in one variable by A. Then if $\dim G > 2 \dim A$, *G is the direct product of an additive group in one variable by a group of the same type as G. Furthermore such a G of dimension* $2 \dim A$ *can be found which does not admit this direct product decomposition.*

A somewhat simpler statement of this theorem can be given in terms of factor sets. First we remark that it is easy to show that any group extension of a commutative group variety by an abelian variety is again commutative. It follows that commutative group extensions of the solvable commutative group variety H by the abelian variety A correspond to functions $f : A \times A \to H$ satisfying the relation

$$f(p_1 + p_2, p_3) + f(p_1, p_2) = f(p_1, p_2 + p_3) + f(p_2, p_3),$$

where a function $f(p_1, p_2)$ is to be regarded as trivial if it is of the form

$$f(p_1, p_2) = g(p_1) + g(p_2) - g(p_1 + p_2).$$

If H is taken to be the additive group in one variable, then the nontrivial commutative group extensions of H by A can be made into a vector space over the constant field in the obvious way, and the previous theorem then says that this vector space has the same dimension as A. The proof rests on the use of generalized jacobians. On the other hand, the following result is almost trivial.

Let the group variety G be an extension of H, the direct product of a number of additive groups in one variable, by the abelian variety A. Then in characteristic $p \neq 0$ (but not necessarily in characteristic 0) there exists a rational homomorphism of G onto $H \times A$.

This last theorem explains the apparent anomaly of the following, which itself is a good example of the inadequacy of a purely formal approach toward differentials in characteristic $p \neq 0$.

The dimension of the vector space of differentials of second kind modulo exact differentials on a curve of genus g is $2g$ or g, depending on whether the ground field has characteristic 0 or $p \neq 0$.

4. A related but somewhat different geometric theory for differentials of second kind valid for any characteristic can be given in terms of fiber vundles. Let V be a nonsingular variety. Then a principal fiber bundle over V having as group the additive group in one variable can be defined by a set of open sets (in the Zariski topology of V) covering V, say $U_1, \ldots, U_s$, and rational functions $f_1, \ldots, f_s$ on V such that $f_i - f_j$ is finite on $U_i \cap U_j$ for all $i, j = 1, \ldots, s$. We identify this with the bundle that comes from replacing $\{U_i\}$ by a refinement and using the same $\{f_i\}$ (with repetitions), or by using the same $\{U_i\}$ and replacing $\{f_i\}$ by $\{g_i\}$ such that $f_i - g_i$ is finite on U_i for each i, or by using the same $\{U_i\}$ and adding any fixed function F to each f_i. Such fiber bundles form a vector space over the field of constants; in characteristic zero this vector space is shown to be isomorphic to the space of differentials of second kind modulo differentials of first kind and modulo exact differentials by associating with the bundle got from $\{U_i\}$ and $\{f_i\}$ a closed differential having the same singularities on each U_i as df_i. Hence (in characteristic zero) the dimension of this space is the irregularity of V (= dimension of the Picard or Albanese variety of V). In the case of arbitrary characteristic this same result can be shown directly for curves and we can use the previously mentioned results to get it for abelian varieties. It also seems possible that this is enough to prove it generally.

The consideration of principal fiber bundles having as group the multiplicative group in one variable, or of group extensions of this group by a given abelian variety, leads to the theory of the Picard variety but does not seem to add anything new to that theory.

5. A large number of problems along these lines suggest themselves. For example, one can try to find a similar theory for extensions of more general solvable group varieties by abelian varieties. It would also be interesting (even if rather poor in applications) to have more information on differentials of the purely formal type, i.e. things of the form $\Sigma f_i\, dx_i$. E.g. it is still not known in characteristic $p \neq 0$, except in special cases, whether a differential without singularities is necessarily closed. Finally there is the very fundamental problem of finding some sort of simple geometric theory, valid for any characteristic, for differential forms of higher degree. Little is known about this last problem but group varieties do not seem to give the answer.

NORTHWESTERN UNIVERSITY

GEOMETRY UPON AN ALGEBRAIC VARIETY

BENIAMINO SEGRE

I. Algebraic geometry — that is to say, the branch of geometry which deals with the properties of entities represented by algebraic equations — has in recent years developed in two distinct directions, which in a sense are opposed to one another. One of these directions is called *abstract* in as much as it is concerned with algebraic equations defined over commutative fields subject only to slight restrictions; here the means employed are purely algebraic, including in particular ideal theory and valuation theory. The other direction may properly be called *geometrical*; this usually deals with algebraic equations in the complex domain, and from time to time appeals to ideas and methods of analytic and projective geometry, topology, the theories of analytic functions and of differential forms.

The dualism between these two disciplines has close relationship and affinity with that which, three centuries ago, arose between *l'esprit géométrique* of Descartes and *l'esprit de finesse* of Pascal, and which, in the past century, on the one hand divided the geometers into analysts of the school of Plücker and synthesists of the school of Steiner and, on the other, the algebraists into purists à la Dedekind and arithmetizers à la Kronecker. However, this dualism, instead of proving harmful to geometry, offers undoubted advantages when the two lines of development, with their respective merits and possibilities, are regarded not as contrasting but as complementary.

We cannot fail to recognise in the abstract method and its technique a peculiar elegance, an impeccable logical coherence, and to appreciate the importance of the results so far obtained by it, particularly in the study of the foundations of geometry and the difficult questions concerning the singularities of algebraic varieties. But equally we cannot fail to recognise that the geometrical approach, with its greater concreteness, lends itself better to the formulation and initial study of new concepts and problems; and that it presents an incomparable wealth and colour of its own, due to the interweaving of many diverse strands, to the subtle and perspicuous play of geometrical intuition, and to the possibility of readily constructing examples and investigating special cases. We may also point out that, in the geometrical discipline, corresponding to a more definite notion of algebraic variety, there is a much wider range of subjects and a far greater number of orientations and contacts with other

important branches of mathematics, which have found, and are finding, therein inspiration and extensions beyond the purely algebraic field.

In the present address I shall leave aside altogether the abstract method, partly for lack of time and partly because exposisions of the subject have recently been given by Hodge {46} [1]), Zariski [1] and Groebner {35}; for similar reasons I shall not speak of the relations between algebraic geometry and number theory, for which I may refer to the discourse by André Weil {100} at the 1950 Congress. Even so, the panorama, when thus delimited and restricted to the contributions of the last few years, would still be too vast, and the terrain too impenetrable, if I attempted to include significant but relatively particular results such as the recent researches of Conforto [2]), Weil {101—103}, Igusa {50—51}, Chow {19, 20}, Andreotti {3, 6} and Barsotti {13} on Abelian varieties; those of Severi {90, 94} and Conforto {87} on quasi-Abelian varieties; those of Roth {74—77} on elliptic, hyperelliptic and pseudo-Abelian varieties; the studies of Roth {67—73, 78}, Baldassarri {10—12} and Segre {86} on questions of unirationality and birationality of algebraic varieties, with special regard to threefolds; the researches of Chisini {18} and his school on the theory of braids and branch curves of multiple planes; the results of Andreotti {4, 5, 7—9} and of Conforto and Gherardelli {21} on the torsion of algebraic varieties, on the problem of uniformisation and the classification of the irregular surfaces; those of Godeaux {27—34} on multiple surfaces, of Gaeta {22—24} on families of space curves, of Scott {79, 80} and Todd {97} on correspondences between algebraic surfaces; and, finally, the investigation, by Severi {91, 92}, Néron {64, 65} and Segre {81, 83}, of various arithmetic questions which arise in algebraic geometry.

I propose instead to discuss, though only summarily, certain recent developments in the geometry on an algebraic variety, which (I hope) will be fairly clear even to non-specialists; and, at the same time, to glance at some of the still unsolved problems in the subject.

II. Let M be an algebraic variety of dimension $d \geqq 1$, defined over the complex field; and suppose that M is irreducible and non-singular (i.e. free from multiple points). By "geometry on M" we mean the study of *relative invariant* properties, that is, those properties which are unchanged by *regular* birational transformations of M; among these, special importance attaches to those which are *absolutely invariant*, that is, are unaltered by *all* possible bira-

---

[1]) The numbers in curly brackets refer to the bibliography at the end; those in square brackets are from the monograph {84} by B. Segre, which we do not reproduce here.

[2]) Fabio Conforto, whose premature death we all deeply regret, was to have addressed the present Congress on the subject of singular Abelian varieties. For a detailed account of his work see B. Segre {87}.

tional transformations of M (including those with exceptional elements). In this study the first salient feature that presents itself is the aggregate of the algebraic subvarieties of M; any two of these — which we suppose to be pure and of dimensions a, b — define an intersection, also algebraic, each component of which has dimension $\geqq a + b - d$. In general, but not always, the dimension of each component is exactly $a + b - d$: we then say that the two varieties present the regular case.

These notions can be extended to subvarieties which are impure, effective or virtual, namely to the aggregates

$$\sum \alpha_i A_i \tag{1}$$

formed by a finite number of subvarieties A_i of M, each taken with an assigned multiplicity α_i, where α_i is an integer (positive, zero or negative). The aggregates (1) may be associated in an obvious manner by addition and subtraction, thereby determining an Abelian group of infinite base. In this group we can define intrinsically, and in various ways, a subgroup of varieties which are to be considered as equivalent to zero, in such a manner that for the corresponding lateral classes — called *classes* (or *systems*) *of equivalence* — there subsists a multiplication, commutative and associative, geometrically related to the intersection in the regular case; this gives rise to an *equivalence ring* on M.

The useful types of equivalence, introduced by Severi [4—7] 20 years ago, are essentially four in number; in non-increasing order of restriction these are: *rational, algebraic, topological* and *enumerative* equivalence. In the case of two *hypersurfaces* (varieties of dimension $d - 1$) of M, the first reduces to the classical linear equivalence, that is, the relation which intercedes between two hypersurfaces of constant level for the same rational function of a variable point of M. In the case $d > 1$, having fixed the type of equivalence, we can similarly consider an infinity of equivalence rings, one for each of the irreducible subvarieties of M, and hence a multiplication law which (in addition to the classes determined by the factors of the product) depends on the particular subvariety of M on which the product is taken. We may complete the picture by introducing other operations of a less obvious character; it will suffice to say that these correspond geometrically to the search for a residual intersection in the non-regular case.

III. If A is a hypersurface of M, the successive powers

$$A, -A^2, \ldots, (-1)^d A^{d-1}, (-1)^{d+1} A^d,$$

effected in the rational equivalence ring on M and taken with alternating signs, form a *covariant succession of immersion of A in M*, consisting of pure subvarieties of A having the respective dimensions

$$k, k - 1, \ldots, 1, 0, \tag{2}$$

where $k = d - 1$. In the case of a variety A of dimension $k < d - 1$, it is still possible to obtain an analogous succession by proceeding as follows (Cf. Segre {84, 85}).

Suppose for simplicity that A is irreducible, non-singular and of dimension k, where $0 \leqq k \leqq d - 2$. Then at each point P of A the varieties M and A have (projective, complex) tangent spaces $[d]$, $[k]$ of respective dimensions d and k, and such that $[d]$ contains $[k]$. The spaces $[k + 1]$ of $[d]$ passing through $[k]$ can be mapped by "points" of a space of dimension $d - k - 1$, there being one such "point" for each mode of tending to P from a point of M not belonging to A. By substituting, in a suitable manner, these ∞^{d-k-1} abstract points for P and repeating the operation for each of the ∞^k points of A, we may show that M can be transformed into a new algebraic variety M', defined by M and A to within a regular birational transformation, which we say is deduced from M by the *dilatation of base A* (Cf. Segre [6]). The varieties M and M' are thus related by a birational correspondence, Θ say, which has no fundamental points on M' and which admits A as locus of fundamental points on M; that is, Θ dilates A into a hypersurface A' of M', which is thus a fibred variety of base A, the fibres being the ∞^k spaces $[d - k - 1]$ which are transforms of the points P of A.

If we now consider the powers

$$A'^{d-k},\ A'^{d-k+1},\ \ldots,\ A'^{d-1},\ A'^{d}, \tag{3}$$

effected in the rational equivalence ring on M', we see that they form a succession of pure subvarieties of A', with the respective dimensions (2). Now take the varieties (3) in order with the signs

$$(-1)^{d-k-1},\ (-1)^{d-k},\ \ldots,\ (-1)^{d},\ (-1)^{d+1},$$

and transform the varieties so obtained by the inverse Θ^{-1} of Θ; we thereby obtain a succession of pure varieties of A, which we denote by

$$A_{M,0},\ A_{M,1},\ \ldots,\ A_{M,k-1},\ A_{M,k} \tag{4}$$

with respective dimensions (2), and such that the first coincides with A. These are naturally defined to within an equivalence; we say that they form the *covariant succession of immersion of A in M*.

A detailed analysis, into which I cannot enter here, shows that the diverse geometrical problems of intersection, regular or irregular, and various other questions in which varieties possessing multiple points can also occur, can all be solved by making use of covariant successions of immersion. Moreover, the latter are closely linked with the so-called *canonical varieties*, as I now propose to indicate.

An algebraic variety M defines (to within a regular birational transformation) the variety $N = M \times M$ of its ordered point-pairs, called the product of

M by itself; this is an algebraic variety of dimension $2d$. M can then be identified with the so-called diagonal variety of N, i.e. the locus of the points which map in N the coincident point-pairs of M. It is thus permissible to consider the covariant succession of immersion of M in N; this is a relative invariant of M; and, by simple algebraic operations within the rational equivalence ring of M, we can deduce from it the *canonical succession*

$$M_0^*,\ M_1^*,\ \ldots,\ M_{d-1}^*,\ M_d^*, \tag{5}$$

consisting of pure varieties of dimensions $d, d-1, \ldots, 1, 0$ (the first of which coincides with M) such that the corresponding classes of equivalence are *invariants* of M.

In a similar manner we define the canonical succession

$$A_0^*,\ A_1^*,\ \ldots,\ A_{k-1}^*,\ A_k^* \tag{6}$$

on any pure k-dimensional subvariety A (possibly virtual) of M; and we may prove that the covariant varieties of immersion (4) are related to the canonical varieties by the formulae

$$A_i^* = \sum_{j=0}^{i} (-1)^j A_{M,j}\, M_{i-j}^* \quad (i = 0, 1, \ldots, k). \tag{7}$$

These equations can be solved for $A_{M,j}$ and hence, in any given problem, we can replace the covariant varieties of immersion by the canonical varieties. It follows that, in order to go deeply into the geometry upon an algebraic variety, it is essential to study the operation of *canonisation*, i.e. the association of the succession (6) with any subvariety A of M.

To begin with, we shall describe some of the results already established in this connection, omitting (for lack of time) to speak of the relations between the canonical varieties and the Jacobian varieties of linear systems of hypersurfaces. We shall then indicate some of the important questions which still remain unanswered.

In the case where A is a hypersurface of M, (7) reduces simply to

$$A_i^* = \sum_{j=0}^{i} A^{j+1} M_{i-j}^* \quad (i = 0, 1, \ldots, d-1).$$

An analogous result holds for varieties which are *complete intersections*; thus, for example, if $P = (AB)_M$ is a regular and simple intersection of two hypersurfaces A, B of M, we have

$$P_i^* = P \sum_{j=0}^{i} (A^j + A^{j-1}B + \ldots + AB^{j-1} + B^j) M_{i-j}^* \quad (i = 0, 1, \ldots, d-2).$$

Also, if A denotes any non-singular variety of M, the successive powers of A in the equivalence ring of M are expressible algebraically in terms of the

canonical varieties of A and M. Further, a knowledge of the latter enables us in many cases to determine the canonical varieties of the varieties A', M' which are deduced from A, M by applying a dilatation Θ of base A. Thus, denoting by $[P]'$ an arbitrary point in the space $[d-k-1]$ of A' which corresponds to any point P of A by Θ, we have the results (Segre {85}):

$$A'^{*}_{a-1} = (-1)^{d-k-1}[(d-k)A^{*}_{k}]',$$
$$M'^{*}_{d} = \Theta(M^{*}_{d}) + (-1)^{d-k}[(d-k-1)A^{*}_{k}]',$$

where the meaning of each symbol in square brackets is deducible by addition.

More generally, it would be exceedingly important to determine the transformation law for the canonical varieties (5) of M, under any non-regular birational transformation. Among other things, this would enable us to study the algebraic combinations of the canonical varieties which are *absolutely* invariant for such transformations.

Among the algebraic combinations of (5), those which reduce to point-sets have a particular interest. These are obtained in the following manner. Denoting by $i_1, i_2, \ldots, i_l$ any partition of the number d, we see that the product $M^{*}_{i_1} M^{*}_{i_2} \ldots M^{*}_{i_l}$ defines a set of points on M (determined to within an equivalence); we represent this by the symbol $(i_1 i_2 \ldots i_l)_M$, and the number of points in the set by $[i_1, i_2 \ldots i_l]_M$. If with each partition $(i) = i_1, i_2, \ldots, i_l$ of d we associate any integer whatever, say $\lambda_{i_1 i_2 \ldots, i_l}$, we may define on M an invariant series of equivalence, consisting of sets equivalent to

$$\sum_{(i)} \lambda_{i_1 i_2 \ldots i_l} (i_1 i_2 \ldots i_l)_M. \tag{8}$$

The number of points in this set, namely

$$\sum_{(i)} \lambda_{i_1 i_2 \ldots i_l} [i_1 i_2 \ldots i_l]_M, \tag{9}$$

is a numerical invariant of M, depending on M in such a way as to reduce to zero if — and, for a generic M, only if — the numbers λ are all zero.

Presumably the only invariant point-sets of M are equivalent to a set of type (8); also any invariant variety on a generic M may presumably be expressed algebraically in terms of the canonical varieties. This question has not yet been studied; it presents serious difficulties because, as we shall soon see, for $d \geqq 2$ the situation with regard to the invariant numbers is quite different.

V. On M, the effective hypersurfaces K equivalent to a canonical hypersurface M^{*}_{1} form a complete linear system $|K|$, possibly empty, whose dimension, augmented by unity, we denote by $g_d = g_d(M)$, writing $g_d = 0$ when there is no effective K. The number g_d, called the *geometric genus* of M, is an absolute invariant of M, equal to the number of linearly independent everywhere finite

d-ple integrals on M. Similar invariants, for $d \geqq 2$, are the numbers $g_i = g_i(M)$ $(i = 1, 2, \ldots, d-1)$ which give the numbers of analogous i-ple integrals on M.

For a generic M of dimension $d \geqq 2$, none of these invariants can be expressed in the form (9), while — as we shall see — the alternating sum

$$g_d - g_{d-1} + g_{d-1} - \ldots + (-1)^{d-1} g_1 \tag{10}$$

can be so expressed. For a given M, the calculation of g_i $(i < d)$ in general presents difficulties which at the moment are insuperable.

Something, however, is known about the geometric genus g_d, owing to the fact that we can construct the canonical system $|K|$ algebraically by referring to a model of M in space $[d+1]$, endowed with the so-called ordinary singularities; in fact it is well known that $|K|$ is cut on such a model — of order n, say — residually to the double hypersurface, by primals of order $n-d-2$. We may thus attempt to determine the dimension of $|K|$, and hence g_d, by using the classical *postulation* (or *characteristic*) *formula*: such a formula is, however, valid only for primals of sufficiently high order [3]), and thus we cannot affirm that it is applicable for primals of order $n-d-2$. It is therefore very curious that in this case the formula should lead to an *invariant* of M, which moreover, for $d \geqq 2$, is in general distinct from g_d; this is called the *arithmetic genus*. The theory of the arithmetic genus was initiated, in the case $d = 2$, during the second half of the last century by Cayley, Zeuthen and Noether, and then completed by Italian and French geometers. The extension to varieties of higher dimension was first made by Severi [2] in a fundamental memoir of 1909, where two definitions for the arithmetic genus are proposed, which may be presented in the following manner.

Let A be any hypersurface (effective or virtual) of M, and consider two arbitrary hypersurfaces C, C' of M, subject only to suitable conditions of generality (such as those satisfied by prime sections of M). Let $\delta_e(A)$ denote the effective dimension $(\geqq -1)$ of the complete linear system $|A|$ determined by A on M; then, by the postulation formula, we know that, if x and x' are sufficiently large integers, the expression

$$\delta_e(A + xC + x'C')$$

must be a polynomial of order d in x and x', with integral coefficients. The constant term in this polynomial coincides with the constant terms in the polynomials (of orders d in x and x' respectively) expressing

$$\delta_e(A + xC), \qquad \delta_e(A + x'C')$$

for sufficiently large values of x and x'. The equality of these two constant

---

[3]) For the limits of validity of this formula, see Gaeta {25, 26} and Severi {93}.

terms shows that they furnish an integer, $\delta_v(A)$ say, depending only on A and M, which we call the *virtual dimension* of $|A|$.

The two arithmetic genera $p_a = p_a(M)$ and $P_a = P_a(M)$ are then obtained by assuming

$$p_a = (-1)^d \delta_v(0), \qquad P_a = \delta_v(K) + 1 - (-1)^d,$$

where 0 denotes the zero of linear equivalence on M. Thus p_a and P_a are relative invariants of M [4]; and it is a very remarkable fact that they are always *equal*. This important property has been the subject of investigation by Severi [2, 9], Albanese {1}, Todd [3], Zariski {104}, Kodaira {53}, and Kodaira and Spencer {61}. It is to the last two authors that the first general and complete demonstration of this result is due; they have obtained it by appealing to the recent theory (of Leray and H. Cartan) of analytic stacks. The same authors have also established the equality between p_a and the alternating sum (10), previously conjectured by Severi [2], thereby demonstrating the invariance of the arithmetic genus [5]. Similar questions have been considered by Kodaira in a very recent and important work {60}, where other notable results are obtained, including the *Riemann-Roch theorem*, which expresses the effective dimension $\delta_e(A)$ as a function of certain invariants of A and M, on the hypothesis that the system $|A|$ is sufficiently ample (so as, among other things, to contain partially the canonical system $|K|$) [6]. Kodaira also proves the theorem concerning *the completeness of the (linear) characteristic system* defined on the generic A by the complete continuous system which totally contains A [7].

VI. The theory of the arithmetic genus can be established on a quite different basis, as Todd [3] has shown, in a notable work of 1937. In this work Todd, inspired by an idea of Severi [2], seeks to establish the invariance of P_a by showing that the expression

$$t(M) = P_a(M) + (-1)^d, \tag{11}$$

by an appropriate choice of the coefficients λ in the *rational* field, can be written in the form (9): that is to say, that there exists a relation of the form

$$\mu(d) \cdot t(M) = \sum_{(i)} \mu_{i_1 i_2 \ldots i_l}(d) \cdot [i_1 i_2 \ldots i_l]_M, \tag{12}$$

[4]) Cf. Severi [2, 9], Albanese {2}.

[5]) For the extension of this theory to analytic varieties, cf. Spencer {95}, Kodaira {57, 58}, and Kodaira and Spencer {62}. A tentative demonstration of the equality of P_a and the sum (10) is to be found in Kodaira {54}; see also Kodaira {59}. Concerning the relations between the postulation formula and the arithmetic genus, see Muhly and Zariski {63}, Groebner {36}.

[6]) Besides the classical cases $d = 1, 2$, the Riemann-Roch theorem, under more or less restrictive conditions, had previously been established for threefolds by Severi [2], Segre [3], Kodaira {53}. For any value of d, see Hirzebruch {43}.

[7]) For the case of threefolds, cf. Segre [3].

where the coefficients μ and $\mu_{(i)}$ are relatively prime integers, the first of which is positive, and dependent only on the dimension d of M.

It cannot, however, be said that this result has been then completely established, in as much as it has been obtained by appealing to a certain postulate[8]). Nevertheless, Todd must be given the credit for calculating the coefficients in (12) in the cases $d \leqq 6$, and for verifying that, for a variety M of dimension $d \leqq 6$, which is a product

$$M = A \times B \tag{13}$$

of two varieties A, B, we have the formula

$$t(M) = t(A) \,.\, t(B). \tag{14}$$

An analogous formula for the arithmetic genus p_a of the product of any number of given varieties has been actually obtained by Gaeta [1] in 1952.

Recently Hirzebruch {40} has shown how the coefficients in (12) can be calculated explicitly for any value of d, on the hypothesis that (14) holds for the variety (13). A functional equation for $t(M)$ is thus obtained, which admits one and only one solution reducing to $(-1)^d$ for any projective space of dimension d; this solution is found by using the *multiplicative sequences*, due to Hirzebruch, in relation with the development in power series of $-x(e^{-x}-1)^{-1}$. In this way, with any algebraic variety M there is associated a rational number $T(M) = (-1)^d t(M)$, given by (12), called the *Todd genus*: but it is not thereby clear whether such a number always satisfies (11), nor even whether it is necessarily an integer. In this connection Hirzebruch {41, 42}, overcoming considerable difficulties, has in the first place shown that the product of $t(M)$ by 2^{d-1} is an integer; the same author then proceeds to obtain (11), which implies the integral character of $t(M)$, using for this purpose the theory of analytic stacks, as appears from the preliminary note {43}.

The above developments could be invested with a more geometrical character by the following procedure. With regard to (12), we can introduce the *invariant series of equivalence* $|\eta(M)|$, defined by

$$\eta(M) = \sum_{(i)} \mu_{i_1 i_2 \ldots i_l}(d) \,.\, (i_1 i_2 \ldots i_l)_M. \tag{15}$$

We should then have to show how to arrive at a direct definition of the sets (15), and to prove that the number $[\eta(M)]$ of points in such a set is equal to

$$[\eta(M)] = \mu(d) \,.\, [P_a(M) + (-1)^d]. \tag{16}$$

By virtue of (12), this would be tantamount to establishing (11), giving at the same time a geometrico-functional interpretation of the result.

---

[8]) In Kodaira {59}, with the use of harmonic integrals, the validity of this result is verified for varieties which are complete regular intersections of primals in any given space.

Now, in point of fact, this has already been accomplished in the particular cases $d = 2, 3$ (a circumstance which Hirzebruch seems to have overlooked); and it should also be feasible for $d > 3$. In the case $d = 2$, the sets (15) belong to the *Enriques series* (cf. Enriques [1] and Campedelli {16}):

$$\eta(M) = (2)_M + (1, 1)_M,$$

and (16) is none other than the classical relation of Noether {66}:

$$(I + 4) + (p^{(1)} - 1) = 12(p_a + 1).$$

For $d = 3$, (15) is given quite simply by

$$\eta(M) = (1, 2)_M,$$

and (16) reduces to

$$[\eta(M)] = 24(P_a - 1),$$

both results being in a paper by Segre [1] (formulae (82), (60)) dating from 1934. We may add that the authors in question have given simple constructions for the sets $\eta(M)$ and $2\eta(M)$ in the respective cases $d = 2, 3$, by considering a net of hypersurfaces on M. It would be interesting to obtain an analogous construction for $d > 3$; lacking this, we could endeavour to give a direct definition of the sets (15) by proceeding inductively in the following way.

In the first place, using some of the results described in § IV, we can easily show that all the coefficients μ in (15) must be taken to be zero if we require the result

$$\eta(M + N) = \eta(M) + \eta(N)$$

to hold for every pair M, N of hypersurfaces of a generic $(d + 1)$-dimensional variety V. We should then have to prove that, if P denotes the variety $(MN)_V$, and if $\eta(P)$ has already been defined inductively, the coefficients $\mu_{(i)}$ in (15) can be determined (and, by what has been said, uniquely) so that, for every choice of V, M, N, we have

$$\eta(M + N) = \eta(M) + \eta(N) + \nu(d) \cdot \eta(P),$$

where $\nu(d)$ is a positive integer which is prime to the $\mu_{(i)}$'s. Since, as we know,

$$P_a(M + N) = P_a(M) + P_a(N) + P_a(P),$$

this implies that $\mu(d)$ must satisfy the relation

$$\mu(d) = \nu(d) \cdot \mu(d - 1). \tag{17}$$

From this method of introducing the series of equivalence $| \eta |$, we could deduce geometrical forms for the relations of Severi [2] and Albanese [1] [9]), between the arithmetic genera of the various powers $K^0 = M, K^1 = K, K^2, \ldots,$

[9]) Relations equivalent to Albanese's were afterwards obtained independently by Todd and Maxwell [1]. For the equivalence between the two sets of relations see Todd {98}.

K^d of a canonical hypersurface K on M. Thus, with regard to the Severi relation, we have the formula

$$[(-1)^d - 1]\eta(M) = \sum_{i=1}^{d} (-1)^i \frac{\mu(d)}{\mu(d-i)} \eta(K^i), \tag{18}$$

which has already been established for the first few values of d[10]), and which it would be interesting to prove in general.

VII. The preceding results have already led to extensions in various directions, and readily suggest others. To begin with, there is no difficulty in extending the concept of dilatation to compact *complex* varieties: this has already been achieved by Hopf {49} by means of the so-called σ-process, in the special case where the base of the dilatation reduces to a point; of this, interesting applications have been made by Hopf, Behnke and Stein {14}, and Hirzebruch {39}. As regards the extension to the general case we have so far only a few results of local character (Segre {82}) and considerations of homology on dilated varieties (Guggenheimer {38}). It would be well worth while to carry these studies further, specially in relation to the covariant varieties of immersion, the canonical varieties and intersection problems in the non-regular case, as a result of which — in view of recent work by Hodge [1], Chern {17}, Vesentini {99} and Guggenheimer {37} — it would be possible to introduce Chern's characteristic classes by a new and simpler method.

It is known that there exist compact *complex* varieties which are not equivalent to algebraic varieties (Cf. for example Calabi and Eckermann {15}), that an important category intermediate between these two types is given by the *Kähler varieties* (Cf. Hodge {48}), and that — with the use of harmonic integrals — various properties of algebraic varieties can be carried over to the last-named. This has been brilliantly achieved by Kodaira {52, 55} and by Spencer {95}, as regards the Riemann-Roch theorem and the arithmetic genus. We may further point out that almost all the researches on the Todd genus have been conducted by Hirzebruch for varieties even more general than complex varieties (the so-called *almost complex* manifolds). This author {40} has also characterised the Todd genus by means of its invariance under Hopf's σ-process, and — for varieties of dimensions which are multiples of 4 — has exhibited a relation between the genus and the inertia index of certain quadratic forms {41, 42, 44}. A comparison with a known result of Hodge {45, p. 224} leads to the presumption that there exists, in the algebraic case, a relation (which it would be worth investigating) between these ideas and the notion of geometric genus.

---

[10]) For $d = 3$ this is to be found in Segre [1] (formula (58)), observing that $\mu(0) = 1$, $\mu(1) = 2$, $\mu(2) = 12$, $\mu(3) = 24$ (Cf. formulae (50), (54) of Segre [1]). By virtue of (17), one could probably deduce from (18) that $\nu(d) = 2$ for all *odd* values of d.

Here too the algebraic case is revealed as a source of fruitful suggestions; and, moreover, it is not to be supposed that apparent extensions always have a greater significance. Thus it has been recently shown by Kodaira {56} that certain Kähler varieties, which Hodge {47} has called Kähler manifolds of *restricted type*, are always regularly equivalent to non-singular algebraic varieties. In addition, Kodaira {60} has found a notable case in which certain varieties which are "abstract" in the sense of Weil can be realised as algebraic varieties in the usual sense.

Even if we limit ourselves to the study of algebraic varieties, there is no lack of interesting questions which remain to be examined; among them are the specially important problems which present themselves when the series and systems of equivalence are considered in the *effective*, instead of the virtual, field. And it may be noted that the only adequate weapons of attack which we possess for such a purpose to-day are algebro-geometric in character.

In this order of ideas, we may first mention the problems concerning the dimensions of series and systems of equivalence; about these, absolutely nothing is known save in the classic case of linear systems of hypersurfaces and for an attempt made by Severi {89} to extend the Riemann-Roch theorem to series of equivalence on a surface.

Another notable type of question is that which links the construction of the canonical varieties with the use of differential forms of the first species, in the manner originally indicated by Severi [3], Todd {96} and Eger [2, 3]; another question treats of series and systems of equivalence from the topological point of view and in relation to the theory of correspondences (Cf. Severi {88}, [6]).

It would also be worth while to investigate the operation of canonisation in connection with the various compositions within the equivalence rings on a given algebraic variety, with particular regard to the iteration of this operation.

Finally, we should not neglect the problems of classification, concerning which very little is known for varieties of three or more dimensions. In particular there is Severi's conjecture — confirmed by him [6, chap. VI] solely in the case of surfaces — that the Abelian varieties of rank unity are the only algebraic varieties for which the various canonical systems of lower dimension all reduce to the zero of rational equivalence.

All these questions are undoubtedly important and difficult, even if none of them has been mentioned by von Neumann in his memorable discourse on "Unsolved problems in mathematics". To the former, as to the latter, it will be the task of the future to provide an answer!

BIBLIOGRAPHY

[1] G. ALBANESE, Sul genere aritmetico delle varietà algebriche a quattro dimensioni, Rend. Acc. Lincei, (5) 33 $(1924)_1$, 179—182.

[2] G. ALBANESE, Invarianza del genere P_a di una varietà algebrica a quattro dimensioni, Rend. Acc. Lincei, (5) 33 $(1924)_1$, 210—214.

[3] A. ANDREOTTI, Sopra le involuzioni appartenenti ad una varietà di Picard, Acta Pontif. Ac. Sc., 14 (1950), 107—115.

[4] A. ANDREOTTI, Sopra la varietà di Picard d'una superficie algebrica e sulla classificazione delle superficie irregolari, Rend. Acc. Lincei, (8) 10 (1951), 380—385.

[5] A. ANDREOTTI, Sopra il problema dell'uniformizzazione per alcune classi di superficie algebriche, Mem. Acc. dei XL, (4) 2 (1951), 111—127.

[6] A. ANDREOTTI, Sopra le varietà di Picard di una superficie algebrica, Mem. Acc. dei XL, (4) 2 (1951), 129—137.

[7] A. ANDREOTTI, Recherches sur les surfaces irrégulières, Mém. Ac. Roy. Belgique, 27 (1952), fasc. 4, 56 pp.

[8] A. ANDREOTTI, Recherches sur les surfaces algébriques irrégulières, Mém. Ac. Roy. Belgique, 27 (1952), fasc. 7, 36 pp.

[9] A. ANDREOTTI, Sopra il gruppo della torsione unidimensionale delle varietà algebriche, Rend. Semin. Mat. Torino, 12 (1952—3), 239—264.

[10] M. BALDASSARRI, Su un criterio di riduzione per un sistema algebrico di varietà, Rend. Semin. Mat. Padova, 19 (1950).

[11] M. BALDASSARRI, Sulle V_3 contenenti un sistema lineare triplamente infinito di superficie razionali, Rend. Sem. Mat. Padova, 20 (1951), 135—152.

[12] M. BALDASSARRI, Una condizione per l'esistenza di unisecanti, Rend. Acc. Lincei, (8) 12 (1952), 390—397.

[13] I. BARSOTTI, A note on Abelian varieties, Rend. Circ. Mat. Palermo, (2) 2 (1953), 236—257.

[14] H. BEHNKE—K. STEIN, Modification komplexer Mannigfaltigkeiten und Riemannscher Gebiete, Math. Ann., 124 (1951), 1—16.

[15] E. CALABI—B. ECKMANN, A class of compact, complex manifolds which are not algebraic, Ann. of Math., 58 (1953), 494—500.

[16] L. CAMPEDELLI, Intorno ad alcune serie invarianti di gruppi di punti sopra una superficie, Rend. Acc. Lincei, (6) 17 (1933), 37—40.

[17] S. CHERN, On the characteristic classes of complex sphere bundles and algebraic varieties, Amer. Journ., 75 (1953), 565—597.

[18] O. CHISINI, Sulla costruzione a priori delle trecce caratteristiche, Ann. di Mat., (4) 33 (1952), 353—366.

[19] W. L. CHOW, On Picard varieties, Amer. J. Math., 74 (1952), 895—909.

[20] W. L. CHOW, The Jacobian variety of an algebraic curve, Amer. J. Math., 76 (1954), 453—476.

[21] F. CONFORTO—F. GHERARDELLI, Classificazione delle superficie ellittiche con un fascio ellittico di curve di genere tre, Ann. di Mat., (4) 33 (1952), 273—351.

[22] F. GAETA, Sulle curve sghembe algebriche di residuale finito, Ann. di Mat., (4) 27 (1948), 177—241.

[23] F. GAETA, Nuove ricerche sulle curve sghembe algebriche di residuale finito e sui gruppi di punti del piano, Ann. di Mat., (4) 31 (1950), 1—64.

[24] F. GAETA, Caratterizzazione delle curve origini di una catena di resti minimali, I e II, Rend. Acc. Lincei, (8) 12 (1952), 270—273, 387—389.

[25] F. GAETA, Sur la limite inférieure l_0 des valeurs de l pour la validité de la postulation régulière d'une variété algébrique, Comptes Rend. de l'Ac. des Sc., 234 (1952), 1121—1123.

[26] F. GAETA, Détermination de la chaîne syzygétique des idéaux matriciels parfaits et son application à la postulation de leurs variétés algébriques associées, Comptes Rendus de l'Ac. des Sc., 234 (1952), 1833—1835.

[27] L. GODEAUX, Osservazioni sui punti uniti delle involuzioni cicliche appartenenti ad una superficie algebrica, Rend. di Mat., (5) 9 (1950), 70—73.

[28] L. GODEAUX, Sur quelques surfaces algébriques représentant des involutions cycliques, Bull. Ac. Roy. Belgique, I—IV, (5) 37 (1951), 819—825, 826—835, 938—949, 1106—1119.

[29] L. GODEAUX, Recherches sur les points de diramation des surfaces multiples, Bull. Ac. Roy. Belgique, (5) 37 (1951), 111—120.

[30] L. GODEAUX, Sur la singularité d'une surface multiple en un point de diramation, Bull. Ac. Roy. Belgique, (5) 38 (1952), 50—70.

[31] L. GODEAUX, Mémoire sur les surfaces multiples, Mém. Ac. Roy. Belgique, (2) 27 (1952), n. 1626, 80 pp.

[32] L. GODEAUX, Sur quelques points de diramation de seconde espèce et de troisième catégorie d'une surface multiple, I et II, Bull. Ac. Roy. Belgique, (5) 38 (1952), 755—765, 898—907.

[33] L. GODEAUX, Les singularités des points de diramation isolés des surfaces multiples. Coll. Géom. alg. (Liège 1952), 225—241.

[34] L. GODEAUX, Sur l'ordre d'une involution cyclique appartenant à une surface algébrique, Bull. Soc. Roy. Liège, 22 (1953), 77—84.

[35] W. GRÖBNER, La théorie des idéaux et la Géométrie algébrique, Coll. de Géom. alg. (Liège, 1952), 129—144.

[36] W. GRÖBNER, Über das arithmetische Geschlecht einer algebraischen Mannigkeit, Arch. Math., 3 (1952), 351—359.

[37] H. GUGGENHEIMER, Interpretazione topologica dei covarianti di Beniamino Segre, Rend. Acc. Lincei, (8) 16 (1954), 331—334.

[38] H. GUGGENHEIMER, Omologia delle dilatazioni, Rend. Acc. Lincei, (8) 17 (1954).

[39] F. HIRZEBRUCH, Über vierdimensionale Riemannsche Flächen mehrdeutiger analytische Funktionen von zwei komplexen Veränderlichen, Math. Ann., 126 (1953), 1—22.

[40] F. HIRZEBRUCH, Arithmetic genera and the theorem of Riemann-Roch for algebraic varieties, The Inst. Adv. Study mim. Notes, 11 pp.

[41] F. HIRZEBRUCH, On Steenrod reduced powers in oriented manifolds, The Inst. Adv. Study mim. Notes, 11 pp.

[42] F. HIRZEBRUCH, The index of an oriented manifold and the Todd genus for an almost complex manifold, The Inst. Adv. Study mim. Notes, 43 pp.

[43] F. HIRZEBRUCH, Arithmetic genera and the theorem of Riemann-Roch for algebraic varieties, The Inst. Adv. Study mim. Notes, 11 pp.

[44] F. HIRZEBRUCH, On Steenrod's reduced powers, the index of inertia, and the Todd genus, Proc. Nat. Ac. Sc., 39 (1953), 951—956.

[45] W. V. D. HODGE, The theory and application of harmonic integrals, Cambridge, The University Press, 1941.

[46] W. V. D. HODGE, Some recent development in the theory of algebraic varieties, Journ. London Math. Soc., 25 (1950), 148—157.

[47] W. V. D. HODGE, A special type of Kähler manifold, Proc. London Math. Soc., (3) 1 (1951), 104—117.

[48] W. V. D. HODGE, Structure problems for complex manifolds, Rend. di Mat., (5) 11 (1952), 101—110.

[49] H. HOPF, Über komplex-analytische Mannigfaltigkeiten, Rend. di Mat., (5) 10 (1951), 169—182.

[50] J. IGUSA, Some remarks on the theory of Picard varieties, Journ. Math. Soc. Japan, 3 (1951), 245—248.

[51] J. IGUSA, On the Picard varieties attached to algebraic varieties, Amer. Journ. Math., 74 (1952), 1—22.

[52] K. KODAIRA, The theorem of Riemann-Roch on compact analytic surfaces, Amer. Journ. Math., 73 (1951), 813—875.

[53] K. KODAIRA, The theorem of Riemann-Roch for adjoint systems on 3-dimensional algebraic varieties, Ann. of Math., (2) 56 (1952), 298—342.

[54] K. KODAIRA, Arithmetic genera of algebraic varieties, Proc. Nat. Ac. Sc., 38 (1952), 527—533.

[55] K. KODAIRA, On the theorem of Riemann-Roch for adjoint systems on Kählerian varieties, Proc. Nat. Ac. Sc., 38 (1952), 522—527.

[56] K. KODAIRA, On Kähler varieties of restricted type, Princeton Univ. mim. Notes, 27pp.

[57] K. KODAIRA, On cohomology groups of compact analytic varieties with coefficients in some analytic faisceaux, Princeton Univ. mim. Notes, 7 pp.

[58] K. KODAIRA, On a differential-geometric method in the theory of analytic stacks, Princeton mim. Notes, 11 pp.

[59] K. KODAIRA, The theory of harmonic integrals and their applications to algebraic geometry, Princeton Univ. mim. Notes (1953).

[60] K. KODAIRA, Some results in the transcendental theory of algebraic varieties, Ann. of Math., 59 (1954), 86—134.

[61] K. KODAIRA—D. C. SPENCER, On arithmetic genera of algebraic varieties, Princeton Univ. mim. Notes, 17 pp.

[62] K. KODAIRA—D. C. SPENCER, On a theorem of Lefschetz and the lemma of Enriques-Severi-Zariski, Princeton Univ. mim. Notes, 12 pp.

[63] H. T. MUHLY—O. ZARISKI, Hilbert's characteristic function and the arithmetic genus of an algebraic variety, Trans. Amer. Math. Soc., 69 (1950), 78—88.

[64] A. NÉRON, La théorie de la base pour les diviseurs sur les variétés algebriques, Coll. Géom. alg. Liège (1952), 119—126.

[65] A. NÉRON, Problèmes arithmétiques et géométriques rattachés à la notion de rang d'une courbe algébrique dans un corps, Bull. Soc. Math. France, 80 (1952), 101—166.

[66] M. NOETHER, Zur Theorie des eindeutigen Entsprechens algebraischer Gebilde, Math. Ann., 2 (1870).

[67] L. ROTH, Sulle V_3 algebriche su cui l'aggiunzione si estingue, Rend, Acc. Lincei, (8) 9 (1950), 246—250.

[68] L. ROTH, Sulle V_3 algebriche che possiedono un sistema anticanonico, Atti 4° Congr. U.M.I. (1951), II, 434—439.

[69] L. ROTH, Algebraic threefolds, Rend, di Math., (5) 10 (1951), 297—376.

[70] L. ROTH, On a class of unirational varieties, Proc. Cambridge Phil Soc., 47 (1951), 496—503.

[71] L. ROTH, Some threefolds on which adjunction terminates, Proc. Cambridge Phil. Soc., 48 (1952), 233—242.

[72] L. ROTH, Sulle V_3 algebriche generate da congruenze di curve, Rend. Acc. Lincei, (8) 12 (1952), 66—70.

[73] L. ROTH, On threefolds of linear genus unity, Ann. di Mat., (4) (1953), 247—276.

[74] L. ROTH, Hyperelliptic threefolds, Proc. Cambridge Phil. Soc., 49 (1953), 397—409.

[75] L. ROTH, On elliptic threefolds, Rend. Circ. Mat. Palermo, (2) 2 (1953), 141—158.

[76] L. ROTH, On threefolds which contain conguences of elliptic curves, Rend. di Mat., (5) 12 (1953), 387—425.

[77] L. ROTH, Pseudo-Abelian varieties, Proc. Cambridge Phil. Soc., 50 (1954).

[78] L. ROTH, Sull'estensione di un teorema di Castelnuovo-Humbert, Rend. Acc. Lincei, (8) 15 (1953), 376—380.

[79] D. B. SCOTT, On the fundamental theorem for point-point correspondences with valency on an algebraic surface, Acta Pontif. Ac. Sc., 14 (1950), 57—61.

[80] D. B. SCOTT, The united curve of a point-curve correspondence on an algebraic surface, and some related topological characters of the surface, Proc. London Math. Soc., (2) 51 (1950), 308—324.

[81] B. SEGRE, Arithmetical questions on algebraic varieties, London, The Athlone Press, 1951.

[82] B. SEGRE, Dilatazioni e comportamenti associati nel campo analitico, Rend. Circ. Mat. Palermo, (2) 1 (1952), 373—379.

[83] B. SEGRE, Sui corpi risolventi delle equazioni algebriche, Rend. Acc. Lincei, (8) 13 (1952), 335—340.

[84] B. SEGRE, Nuovi metodi e risultati nella geometria sulle varietà algebriche, Ann. di Mat., (4) 35 (1953), 1—128.

[85] B. SEGRE, Dilatazioni e varietà canoniche sulle varietà algebriche, Ann. di Mat., (4) 37 (1954), 139—155.

[86] B. SEGRE, Intorno ad alcune generalizzazioni di un teorema di M. Noether, Rend. di Math., (5) 14 (1954), 75—84.

[87] B. SEGRE, L'opera scientifica di Fabio Conforto, Rend. di Mat., (5) 14 (1954), 48—74.

[88] F. SEVERI, I sistemi d'equivalenza sulle varietà algebriche e le loro applicazioni, Atti 1° Congr. U.M.I., Firenze 1937, 58—68.

[89] F. SEVERI, Ulteriori sviluppi della teoria delle serie d'equivalenza sulle superficie algebriche, Comm. Pontif. Ac. Sc., 6 (1942), 977—1029.

[90] F. SEVERI, Funzioni quasi abeliane, Pontif. Ac. Sc. Scripta varia, 4 (1947), 327 p.

[91] F. SEVERI, Legami tra certe proprietà aritmetiche delle superficie e la teoria della base, Rend. di Mat., (5) 9 (1950), 59—69.

[92] F. SEVERI, Ulteriori complementi alla teoria della base, Rend. Circ. Mat. Palermo, (2) 1 (1952), 71—87.

[93] F. SEVERI, Un'osservazione sul limite d'applicabilità della formula di postulazione per una varietà algebrica, Atti Acc. Ligure di Sc. e Lettere, 9 (1952), 1—6.

[94] F. SEVERI, Sulla caratterizzazione dei corpi di funzioni quasi abeliane, Convegno di Geom. diff. (1953).

[95] D. C. SPENCER, Cohomology and the Riemann-Roch theorem, Princeton Univ. mim. Notes, 19 pp.

[96] J. A. TODD, Note on the canonical series of a V_d, Proc. London Math. Soc., (2) 43 (1937), 139—147.

[97] J. A. TODD, On the overlap of an algebraic surface, Journ. London Math. Soc., 26 (1951), 73—74.

[98] J. A. TODD, On the invariants of the canonical system of a V_d, Proc. Cambridge Phil. Soc., 49 (1953), 410—412.

[99] E. VESENTINI, Classi caratteristiche e varietà covarianti d'immersione, Rend. Acc. Lincei, (8) 16 (1954), 199—204.

[100] A. WEIL, Number-theory and algebraic geometry, Proc. Internat. Congress of Math. 1950, vol. II (Amer. Math. Soc., 1952), 90—100.

[101] A. WEIL, Variétés abéliennes et courbes algébriques (Actual. Sc. et Ind., 1064), Paris, Hermann, 1948.

[102] A. WEIL, Variétés abéliennes, Colloque Internat. Paris (1950), 125—127.

[103] A. WEIL, On Picard varieties, Amer. Journ. Math., 74 (1952), 865—894.

[104] O. ZARISKI, Complete linear systems on normal varieties and a generalization of a lemma of Enriques-Severi, Ann. of Math. (2) 55 (1952), 552—592.

THE TRIANGLE AS A GEOMETRIC VARIABLE

J. G. SEMPLE

This address appeared in "Mathematika" (1954), 80—88.

COHOMOLOGIE ET GÉOMÉTRIE ALGÉBRIQUE

JEAN-PIERRE SERRE

De nombreux problèmes de géométrie algébrique classique peuvent être formulés et étudiés de la façon la plus commode au moyen de la théorie des faisceaux: c'est ce que montrent clairement les travaux récents de Kodaira-Spencer (cf. [3], [4], ainsi que d'autres notes publiées en 1953 aux Proc. Nat. Acad. Sci. USA) et de Hirzebruch [2]. Il était naturel d'essayer d'étendre ces méthodes à la géométrie algébrique „abstraite", sur un corps de caractéristique quelconque; dans ce qui suit, je me propose de résumer rapidement les principaux résultats que j'ai obtenus dans cette direction.

1. Propriétés générales des faisceaux algébriques cohérents sur une variété projective.

Dans toute la suite, le corps de base k sera un corps commutatif, algébriquement clos, de caractéristique quelconque. Dans l'espace projectif $\boldsymbol{P}_r(k)$, de dimension r sur k, nous choisirons une fois pour toutes un système de coordonnées homogènes $t_0, \ldots, t_r$.

Soit X une *sous-variété* de $\boldsymbol{P}_r(k)$, c'est-à-dire l'ensemble des zéros communs à une famille de polynômes homogènes en $t_0, \ldots, t_r$. Une sous-variété de X sera appelée un sous-ensemble fermé; X se trouve ainsi muni d'une topologie, la *topologie de Zariski*, qui en fait un espace quasi-compact (le théorème de Borel-Lebesgue est valable). La notion de *faisceau* sur X se définit, comme d'ordinaire, par la donnée d'une famille de groupes abéliens $\mathscr{F}_x$, $x \in X$, et d'une topologie sur l'ensemble $\mathscr{F}$, somme des $\mathscr{F}_x$; la projection canonique $\pi : \mathscr{F} \to X$ doit être un homéomorphisme local, et l'application $(f, g) \to f - g$ doit être continue là où elle est définie (cf. [6], n°. 1).

Soit $\mathscr{F}(X)$ le faisceau des germes de fonctions sur X, à valeurs dans k. Si x est un point de X, soit S_x l'ensemble des fractions rationnelles en $t_0, \ldots, t_r$ qui peuvent s'écrire $f = P/Q$, où P et Q sont des polynômes homogènes de même degré, et $Q(x) \neq 0$; S_x n'est autre que l'anneau local de x sur $\boldsymbol{P}_r(k)$. L'opération de restriction à X est un homomorphisme $\varepsilon_x : S_x \to \mathscr{F}(X)_x$ dont nous désignerons l'image par $\mathcal{O}_x$; l'anneau $\mathcal{O}_x$ est *l'anneau local de x sur X*; lorsque x parcourt X, les $\mathcal{O}_x$ forment un sous-faisceau du faisceau $\mathscr{F}(X)$, que nous désignerons par $\mathcal{O}$ (ou par $\mathcal{O}_X$ lorsque nous voudrons préciser la variété X); le faisceau $\mathcal{O}$ est appelé le *faisceau des anneaux locaux* de X.

Un faisceau $\mathscr{F}$ sur X est appelé un *faisceau algébrique* si c'est un fais ceau de $\mathscr{O}$-modules, c'est-à-dire si chaque $\mathscr{F}_x$ est muni d'une structure de $\mathscr{O}_x$-module unitaire, variant continûment avec x. Désignons par $\mathscr{O}^p$ (p entier $\geqq 0$) la somme directe de p faisceaux isomorphes à $\mathscr{O}$; un faisceau algébrique $\mathscr{F}$ est dit *cohérent* si l'on peut recouvrir X par des ouverts U tels que, au-dessus de chacun d'eux, il existe une suite exacte de faisceaux:

$\mathscr{O}^p \xrightarrow{\varphi} \mathscr{O}^q \xrightarrow{\psi} \mathscr{F} \to 0$ (p et q étant des entiers convenables),

où φ et ψ soient des homomorphismes $\mathscr{O}$-linéaires définis au-dessus de U. Les faisceaux algébriques cohérents jouissent des mêmes propriétés formelles que les faisceaux analytiques cohérents de la théorie de Cartan-Oka (voir [6], Chap. I, § 2 et Chap. II, § 2).

Les *groupes de cohomologie* $H^q(X, \mathscr{F})$ de l'espace X à valeurs dans un faisceau $\mathscr{F}$ se définissent par le procédé de Čech. Plus précisément, soit $\mathfrak{U} = \{U_i\}_{i \in I}$ un recouvrement ouvert de X; une q-cochaîne de $\mathfrak{U}$ à valeurs dans $\mathscr{F}$ est, par définition, un système $f_{i_0 \ldots i_q}$, où chaque $f_{i_0 \ldots i_q}$ est une section de $\mathscr{F}$ au-dessus de $U_{i_0} \cap \ldots \cap U_{i_q}$; on pose

$$(df)_{i_0 \ldots i_{q+1}} = \sum_{j=0}^{j=q+1} (-1)^j f_{i_0 \ldots \hat{i}_j \ldots i_{q+1}}.$$

Les q-cochaînes de $\mathfrak{U}$ ($q = 0, 1, \ldots$), ainsi que l'opérateur d, constituent un complexe $C(\mathfrak{U}, \mathscr{F})$, dépendant du recouvrement $\mathfrak{U}$. On définit alors $H^q(X, \mathscr{F})$ comme la limite inductive des groupes de cohomologie des complexes $C(\mathfrak{U}, \mathscr{F})$. Les groupes $H^q(X, \mathscr{F})$ jouissent des propriétés habituelles des groupes de cohomologie; en particulier, $H^0(X, \mathscr{F})$ est canoniquement isomorphe au groupe $\Gamma(X, \mathscr{F})$ des sections de $\mathscr{F}$ au-dessus de X. A toute suite exacte de faisceaux

$$0 \to \mathscr{A} \to \mathscr{B} \to \mathscr{C} \to 0,$$

où $\mathscr{A}$ est un faisceau algébrique cohérent, est attachée une *suite exacte de cohomologie* ([6], n°. 47):

$$\ldots \to H^q(X, \mathscr{B}) \to H^q(X, \mathscr{C}) \to H^{q+1}(X, \mathscr{A}) \to H^{q+1}(X, \mathscr{B}) \to \ldots$$

Lorsque $\mathscr{F}$ est un faisceau algébrique cohérent sur X, les groupes de cohomologie $H^q(X, \mathscr{F})$ possèdent des propriétés particulières importantes. On a tout d'abord ([6], n°. 66):

Théorème 1. *Les groupes $H^q(X, \mathscr{F})$ sont des espaces vectoriels de dimension finie sur k, nuls pour $q >$ dim X.*

Avant d'énoncer le théorème 2, introduisons une notation. Soit U_i l'ensemble des points $x \in X$ où $t_i \neq 0$; les U_i, $0 \leqq i \leqq r$, forment un recouvrement ouvert de X. Si $\mathscr{F}$ est un faisceau algébrique cohérent sur X, soit $\mathscr{F}_i$ la restriction de $\mathscr{F}$ à U_i; n étant un entier quelconque, la multiplication par $(t_j/t_i)^n$ est un isomorphisme

$$\theta_{ij}(n) : \mathscr{F}_i \to \mathscr{F}_j.$$

défini au-dessus de $U_i \cap U_j$; comme l'on a $\theta_{ij}(n) \circ \theta_{jk}(n) = \theta_{ik}(n)$ au-dessus de $U_i \cap U_j \cap U_k$, on peut définir un faisceau $\mathscr{F}(n)$ à partir des $\mathscr{F}_i$ par recollement au moyen des isomorphismes $\theta_{ij}(n)$. Au-dessus de U_i, $\mathscr{F}$ et $\mathscr{F}(n)$ sont isomorphes, ce qui montre que $\mathscr{F}(n)$ est un faisceau algébrique cohérent. Le théorème suivant ([6], n°. 66) indique quelles sont les propriétés de $\mathscr{F}(n)$ lorsque n tend vers $+\infty$:

Théorème 2. *Pour n assez grand, on a:*

a) $H^0(X, \mathscr{F}(n))$ *engendre le $\mathscr{O}_x$-module $\mathscr{F}(n)_x$ quel que soit $x \in X$.*

b) $H^q(X, \mathscr{F}(n)) = 0$ *pour tout $q > 0$.*

On peut également étudier $H^q(X, \mathscr{F}(n))$ pour n tendant vers $-\infty$. On obtient ([6], n°. 74):

Théorème 3. *Soit q un entier $\geqq 0$. Pour que $H^q(X, \mathscr{F}(-n))$ soit nul pour n assez grand, il faut et il suffit que $\operatorname{Ext}_{S_x}^{r-q}(\mathscr{F}_x, S_x)$ soit nul pour tout $x \in X$.*

Dans l'énoncé ci-dessus, $\mathscr{F}_x$ est considéré comme un S_x-module, au moyen de l'homomorphisme $\varepsilon_x : S_x \to \mathscr{O}_x$ défini plus haut; les Ext sont relatifs à l'anneau S_x (pour leur définition, voir [1]).

2. Le théorème de dualité.

Nous supposerons à partir de maintenant que X est une variété *sans singularités*, irréductible, et de dimension m.

Si p est un entier $\geqq 0$, nous noterons $W^{(p)}$ l'espace fibré des p-covecteurs tangents à X; c'est un espace fibré algébrique, à fibre vectorielle, et de base X (pour la définition de ces espaces, voir [7], ainsi que [5], n°. 4 et [6], n°. 41). Si V est un espace fibré algébrique à fibre vectorielle quelconque, nous noterons $\mathscr{S}(V)$ le faisceau des germes de sections régulières de V; nous désignerons par V^* l'espace fibré *dual* de V, et par $\tilde{V}$ l'espace fibré $V^* \otimes W^{(m)}$. Le faisceau $\mathscr{S}(W^{(p)})$ n'est autre que le faisceau Ω^p des germes de formes différentielles de degré p; le faisceau $\mathscr{S}(\tilde{V})$ est canoniquement isomorphe à $\operatorname{Hom}(\mathscr{S}(V), \Omega^m)$.

Lemme. $H^m(X, \Omega^m)$ *est un espace vectoriel de dimension* 1 *sur k.*

Lorsque X est une courbe ($m = 1$), ce résultat est une conséquence classique du théorème des résidus. A partir de là, on raisonne par récurrence sur m. Si C désigne le diviseur découpé sur X par un polynôme homogène de degré n, suffisamment „général", on définit (cf. [4]) une suite exacte de faisceaux:

$$0 \to \Omega^m \to \Omega^m(n) \to \Omega_C^{m-1} \to 0,$$

où Ω_C^{m-1} désigne le faisceau des germes de formes différentielles de degré $m-1$ sur la variété C. Pour n assez grand, le théorème 2 montre que $H^q(X, \Omega^m(n)) = 0$ si $q \neq 0$; la suite exacte de cohomologie montre alors que $H^m(X, \Omega^m)$ est isomorphe à $H^{m-1}(C, \Omega_C^{m-1})$, d'où le résultat, compte tenu de l'hypothèse de récurrence.

Soit maintenant V un espace fibré algébrique à fibre vectorielle, de base X. Puisque $\mathscr{S}(\tilde{V})$ est isomorphe à Hom $(\mathscr{S}(V), \Omega^m)$, on a un homomorphisme canonique:

$$\mathscr{S}(V) \otimes \mathscr{S}(\tilde{V}) \to \Omega^m;$$

cet homomorphisme donne naissance à un cup-produit qui est une application bilinéaire de $H^q(X, \mathscr{S}(V)) \times H^{m-q}(X, \mathscr{S}(\tilde{V}))$ dans $H^m(X, \Omega^m)$. D'après le lemme précédent, $H^m(X, \Omega^m)$ est une espace vectoriel de dimension 1 sur k; on obtient donc ainsi une forme bilinéaire sur $H^q(X, \mathscr{S}(V)) \times H^{m-q}(X, \mathscr{S}(\tilde{V}))$, définie à la multiplication près par un scalaire.

Théorème 4. *La forme bilinéaire définie ci-dessus met en dualité les espaces vectoriels* $H^q(X, \mathscr{S}(V))$ *et* $H^{m-q}(X, \mathscr{S}(\tilde{V}))$.

Ce théorème est l'analogue, dans le cas abstrait, du „théorème de dualité" de [5]. On le démontre par récurrence sur $m = \dim X$; pour $m = 1$, il résulte facilement de la dualité entre différentielles et classes de répartitions; le passage de $m - 1$ à m se fait au moyen de suites exactes analogues à celle utilisée dans la démonstration du lemme ci-dessus; les théorèmes 2 et 3 y jouent un rôle essentiel.

Un cas particulier important est celui où V est l'espace fibré associé à un diviseur D de X (cf. [7], ainsi que [5], n°. 16). Dans ce cas, $\mathscr{S}(V)$ est isomorphe au faisceau $\mathscr{L}(D)$ défini de la manière suivante: un élément de $\mathscr{L}(D)_x$ est une fonction rationnelle f sur X, dont le diviseur (f) vérifie l'inégalité $(f) \geqq -D$ au voisinage de x. On a alors $\mathscr{S}(\tilde{V}) = \mathscr{L}(K - D)$, K désignant un diviseur de la classe canonique de X (cf. [8]), et le théorème 4 prend la forme suivante:

Corollaire. *Les espaces vectoriels* $H^q(X, \mathscr{L}(D))$ *et* $H^{m-q}(X, \mathscr{L}(K - D))$ *sont en dualité.*

3. Caractéristiques d'Euler-Poincaré et formule de Riemann-Roch.

Si $\mathscr{F}$ est un faisceau algébrique cohérent sur X, nous poserons:

$$h^q(X, \mathscr{F}) = \dim_k H^q(X, \mathscr{F}) \quad \text{et} \quad \chi(X, \mathscr{F}) = \sum_{q=0}^{q=m} (-1)^q h^q(X, \mathscr{F}).$$

On montre facilement ([6], n°. 80) que $\chi(X, \mathscr{F}(n))$ est un polynôme en n, de degré $\leqq m$. D'après le théorème 2, $\chi(X, \mathscr{F}(n)) = h^0(X, \mathscr{F}(n))$ pour n assez grand; appliquant ceci au faisceau $\mathscr{F} = \mathscr{O}$, on voit que $\chi(X, \mathscr{O}(n))$ est égal, pour *tout* n, à la fonction caractéristique de Hilbert de la variété X (voir [8], § 10). En particulier, $\chi(X, \mathscr{O})$ est égal au terme constant de la fonction caractéristique, d'où ([6], n°. 80):

Théorème 5. $\chi(X, \mathscr{O})$ *est égal au genre arithmétique de* X.

(Nous appelons genre arithmétique la quantité notée $1 + (-1)^m p_a(X)$ dans [8]).

A partir de maintenant, nous écrirons $\chi(X)$ au lieu de $\chi(X, \mathcal{O})$.

Si H est une sous-variété de X, sans singularités et de dimension $m-1$, on a une suite exacte de faisceaux:

$$0 \to \mathcal{L}(-H) \to \mathcal{O} \to \mathcal{O}_H \to 0.$$

La suite exacte de cohomologie associée à cette suite exacte de faisceaux montre que $\chi(X, \mathcal{O}) = \chi(H, \mathcal{O}_H) + \chi(X, \mathcal{L}(-H))$, autrement dit:

$$\chi(H) = \chi(X) - \chi(X, \mathcal{L}(-H)).$$

Considérons alors un diviseur D quelconque, et soit $\chi_X(D)$ son *genre arithmétique virtuel* ([8], § 11). Si E est une section hyperplane de X, on voit aisément que $\chi_X(D+nE)$ est un polynôme en n; il en est de même de $\chi(X) - \chi(X, \mathcal{L}(-D-nE))$; de plus, la formule ci-dessus montre que ces deux expressions sont égales pour n assez grand. Elles le sont donc pour tout n, ce qui donne:

Théorème 6. *Pour tout diviseur D, on a $\chi_X(D) = \chi(X) - \chi(X, \mathcal{L}(-D))$.*

En remplaçant D par $-D$, on peut écrire le théorème précédent sous la forme:

Formule de Riemann-Roch. $\chi(X, \mathcal{L}(D)) = \chi(X) - \chi_X(-D)$.

Appliquons cette formule au cas $m=2$. On a

$$h^0(X, \mathcal{L}(D)) = \dim_k \Gamma(X, \mathcal{L}(D)) = l(D),$$

et $h^2(X, \mathcal{L}(D)) = h^0(X, \mathcal{L}(K-D)) = l(K-D)$, d'après le théorème de dualité. On obtient donc:

$$l(D) - h^1(X, \mathcal{L}(D)) + l(K-D) = \chi(X) - \chi_X(-D).$$

On retrouve donc bien *l'inégalité de Riemann-Roch* ([8], § 13):

$$l(D) + l(K-D) \geqq \chi(X) - \chi_X(-D),$$

et l'on voit en outre que $h^1(X, \mathcal{L}(D))$ n'est pas autre chose que la superabondance de D.

Remarque. D'après le théorème de dualité, on a:

$$\chi(X, \mathcal{L}(K-D)) = (-1)^m \chi(X, \mathcal{L}(D)).$$

En particulier, $\chi(X, \mathcal{L}(K)) = (-1)^m \chi(X)$, ce qui, joint au théorème 6, donne:

$$\chi_X(-K) = \begin{cases} 2\chi(X) & \text{si } m \text{ est impair} \\ 0 & \text{si } m \text{ est pair.} \end{cases}$$

Avec les notations de [8], § 13, ceci s'écrit $P_a(X) = p_a(X)$, conformément à une conjecture de Severi.

4. Questions non résolues.

Nous venons d'étendre au cas abstrait quelques uns des résultats connus

dans le cas classique. Mais il y en a d'autres dont l'extension paraît plus difficile. Citons notamment:

a) Soit $h^{p,q} = \dim_k H^q(X, \Omega^p)$. A-t-on $h^{p,q} = h^{q,p}$? La dimension de la variété de Picard de X est-elle égale à $h^{1,0}$?

(Signalons que le théorème de dualité entraîne l'égalité de $h^{p,q}$ et de $h^{m-p,\, m-q}$).

b) Si V est un espace fibré algébrique, à fibre vectorielle, de base X, $\chi(X, \mathscr{S}(V))$ est-il égal à un polynôme en les classes canoniques de V et de la structure tangente à X (cf. [2])?

On peut également se demander si les $B_n = \sum_{p+q=n} h^{p,q}$ coïncident avec les „nombres de Betti'' qui interviennent dans les conjectures de Weil relatives à la fonction zêta de X (la variété X étant supposée définie sur un corps fini).[1])

Bibliographie

[1] H. Cartan and S. Eilenberg. Homological Algebra. Princeton Math. Ser., n°. 19.

[2] F. Hirzebruch. Arithmetic genera and the theorem of Riemann-Roch for algebraic varieties. Proc. Nat. Acad. Sci. USA, **40**, 1954, p. 110—114.

[3] K. Kodaira and D. C. Spencer. On arithmetic genera of algebraic varieties. Proc. Nat. Acad. Sci. USA, **39**, 1953, p. 641—649.

[4] K. Kodaira and D. C. Spencer. On a theorem of Lefschetz and the lemma of Enriques-Severi-Zariski. Proc. Nat. Acad. Sci. USA, **39**, 1953, p. 1273—1278.

[5] J.-P. Serre. Un théorème de dualité. Comment. Math. Helv., **29**, 1955, p. 9—26.

[6] J.-P. Serre. Faisceaux algébriques cohérents. Ann. of Math. **61**, 1955, p. 197—278.

[7] A. Weil. Fibre-spaces in algebraic geometry (Notes by A. Wallace). Chicago Univ., 1952.

[8] O. Zariski. Complete linear systems on normal varieties and a generalization of a lemma of Enriques-Severi. Ann. of Math., **55**, 1952, p. 552—592.

---

[1]) J. Igusa vient de résoudre négativement deux des questions posées ci-dessus: il a construit une variété X avec $h^{0,1} = h^{1,0} = 2$, alors que la dimension de la variété de Picard de X est 1 et que le premier nombre de Betti de X (au sens de Weil) est 2.

Cf. J. Igusa. *On some problems in abstract algebraic geometry*. Proc. Nat. Acad. Sci. USA, **41**, 1955.

FONCTIONS ET VARIÉTÉS QUASI-ABÉLIENNES

FRANCESCO SEVERI *)

Il s'agit-là des fonctions analytiques, méromorphes de π variables indépendantes avec $\mu \leqq 2\pi$ périodes [1]). La théorie est commencée en 1947. (Voir mon mémoire „Funzioni quasi abeliane" un volume dans „Scripta varia" Ac. Pont. 1947).

Si $\mu = 2\pi$ (et alors nous écrirons, suivant l'usage, p au lieu de π) on tombe sur les fonctions abéliennes. Dans ce cas les composantes des périodes forment une *matrice de* Riemann, caractérisée par deux conditions, une qualitative et l'autre quantitative. A propos d'une telle matrice il faut rappeler les recherches remarquables de Gaetano Scorza. Le théorème d'existence relatif assure l'existence d'un corps de fonctions abéliennes, attaché à une matrice de Riemann quelconque ω, donnée. Le même corps est défini par toute matrice *équivalente* à ω, dans un sens bien précis [2]).

Parmi les matrices équivalentes à ω il y a des matrices *normales*, c'est-à-dire de type $| A\Omega |$, ou A est une matrice carré (d'ordre p) à éléments tous nuls sauf ceux de la diagonale principale qui sont $\frac{2\pi i}{d_1}, \ldots, \frac{2\pi i}{d_p}$; $d_1, \ldots, d_p$ étant des entiers ($d_1 = 1$, d_s divisible par d_{s-1}), appellés *diviseurs* du corps; et Ω est une matrice carré symétrique satisfaisante à une seule condition qualitative [3]).

Avant de nous borner aux fonctions quasi abéliennes ($\mu < 2\pi$) ajoutons quelques mots sur la *variété de Picard* V_p, attachée à un corps de fonctions abéliennes de p variables [4]).

---

*) Conférence tenue le 6 septembre 1954 au „Symposion of Algebraic Geometry" (Amsterdam).

[1]) L'étude des fonctions avec $\mu > 2\pi$ périodes se reduise à l'étude des fonctions quasi abéliennes par un convenable changement de variables.

[2]) Deux matrices A, A^* s'appellent *équivalentes* s'il existe une matrice complèxe non dégénérée β et une matrice unimodulaire à éléments entiers B, telle que $A^* = \beta AB$. Cette relation est réflexive, symétrique et transitive.

[3]) Pour les développements plus recents de la théorie des fonctions abéliennes je renvois aux belles leçons: *Funzioni abeliane e matrici di Riemann* di Fabio Conforto (Corsi dell' Istituto Nazionale di Alta Matematica, 1942).

[4]) La dénomination de *variete de Picard* a été introduite dans des travaux de Enriques-Severi et de Castelnuovo. En partant d'une variété algébrique irréductible quelconque W d'irrégularité superficielle $p > 0$ on obtient, d'après Castelnuovo (1905), une *première*

V_p est une variété abélienne, c'est-à-dire que les coordonnées de ses points sont des fonctions du corps et en outre elle est de *rang* 1, c'est-à-dire qu'à chaque point de V_p correspond un seul point dans un des parallèletopes des périodes. Nous nous référons à un modèle non singulier, ne contenant pas des variétés exceptionnelles; ce qu'il est possible d'obtenir.

Toute variété abélienne (de rang quelconque) se représente avec une involution sur V_p.

La V_p fur recontrée pour la première fois par Picard dans ses recherches classiques (commencées pour $p = 2$ en 1889) sur les variétés algébriques admettant un groupe continu, abélien, transitif, ∞^p, d'automorphismes birationnels. Nous envisageons la variété surtout de ce point devue, de sorte à comprendre les deux types de variétés dont nous avons parlé tout à l'heure dans la note [4]).

Des transformations infinitésimales du groupe Picard déduit aisement l'existence sur V_p de p intégrales algébriques simples. Il lui échappa que ces intégrales ne sont pas nécessairement de 1re espèce, comme il le croyait. Le malentendu fut corrigé par Painlevé en 1903.

J'avais dans une petite Note de 1907 (Ist. Veneto) remarqué que pour caractériser la surface de Picard au moyen des invariants géométriques, l'hypothèse de l'existence du groupe continu ∞^2, transitif, n'était pas suffisante, mais qu'il fallait ajouter l'hypothèse de la transitivité *absolue* du groupe, c'est-à-dire que sur la surface, dépourvue de courbes exceptionnelles, n'existait aucune courbe ou point invariant. Au point de vue analytique cela signifie que si la transitivité n'est pas absolue, les deux intégrales simples de la surface ne peuvent pas être tous les deux de 1re espèce.

Eh! bien, la même conclusion est valable pour une dimension quelconque. C'est-à-dire que si une variété algébrique V_π (non singulière) possède un groupe continu abélien, ∞^π, Γ, d'automorphismes birationnels, ou bien Γ est *absolument* transitif et alors V_π est une variété de Picard d'irregularité superficielle

---

V_p de Picard. Elle est l'image birationnelle de la totalité ∞^p des systèmes linéaires de hypersurfaces d'un système irréductible complet de hypersurfaces de W. Une *seconde* V_p (abélienne de rang 1 suivant la términologie qui va suivre), s'obtient d'après le conférencier (F. Severi, *Un teorema d'imversione per l'integrali semplici di I^a specie appartenenti ad una superficie algebrica*, Atti del R. Istituto Veneto di Scienze, Lettere ed Arti, 1913) moyennant la matrice de Riemann des intégrales simples de première espèce attachées à W. Le conférencier a remarqué pour la première fois (en 1913) que les deux variétés peuvent être distinctes. Un lien fort élegant entre les deux variétés a été établi par M. Andreotti (Lincei, 1951: Ac. Royale de la Belgique, 1952). Dans des travaux récemment publiés dans l'Am. Journal la seconde variété est appellée *variété de Albanese*, tandis que cet ancien et vaillant disciple du conférencier l'a rencontrée seulement presque vingt après la publication de la Note de Severi (Istituto Veneto).

$\pi = p$, de sorte qu'il y a sur la variété p intégrales simples de 1re espèce; ou bien Γ est *généralement* transitif et ils existent sur V_π de sous variétés non exceptionnelles invariantes et alors $\pi > p$, p étant l'irregularité superficielle de V_π et les π intégrales simples existant sur V_π en vertu des transformations infinitésimales du groupe, ne sont pas tous de première espèce. On tombe ainsi sur les variétés et sur les fonctions quasi abéliennes.

La recherche de Painlevé ne touche pas aux groupes continus attachés aux fonctions envisagées; mais ce lien est dans la nature même du problème, comme nous le verrons de suite.

En raison de l'importance de la recherche de Painlevé publiée dans le volume des Acta mathematica pour la célébration du centenaire d'Abel, j'ajouterai d'autres indications. La recherche est développée jusqu'au fond pour $\pi = 2$, avec le but principal de démontrer le théorème algébrique d'addition pour les fonction de plusieurs variables, thèoréme que Weierstrass donnait dans ses leçons, mais dont on ne connaissait pas la démonstration avant Painlevé. Dans le cadre de ces fonctions se presentaient soit les fonctions abéliennes, soit leurs dégénérescences, qui sont précisement les fonctions quasi abéliennes.

Les résultats de Painlevé combinés avec ceux de l'Auteur amènent à la conclusion que les fonctions quasi abéliennes (en particulier abéliennes) épuissent l'ensemble des fonctions continues de π variables, à dérivées premières continues, qui admettent un théorème d'addition algébrique. (Elles résultent en conséquence analytiques méromorphes, périodiques).

Nous appellons *variété quasi abélienne de Picard* V_π une variété algébrique (dont nous considérons un modèle non singulier) possedant un groupe continu ∞^π, abélien, Γ, généralement transitif d'automorphismes birationnels. Elle joue dans la théorie des fonctions quasi abéliennes le même role qui est joué par la variété de Picard dans la théorie des fonctions abéliennes.

L'inversion sumultanée du système linéaire Σ des intégrales simples $u_1, u_2, \ldots, u_\pi$, existant sur V_π en vertu de Γ, donne l'uniformisation de V_π au moyen de fonctions quasi abéliennes d'un corps, dont nous indiquérons dans la suite la construction. La V_π est de rang 1 par rapport aux périodes de ce corps de fonctions, mais il y a des exceptions à la biunivocité.

Le système Σ est *complet* par rapport à Γ; c'est-à-dire qu'il renferme toute intégrale simple de V_π invariant pour Γ.

Au système Σ appartient toute intégrale de 1re espèce $u_1, u_2, \ldots, u_p$ de V_π, ou p $(0 \leqq p \leqq \pi)$ est l'irregularité superficielle de V_π. Soit Σ_0 le système linéaire des intégrales de 1re espèce. Etant $p < \pi$, dans Σ il existe des intégrales de 2me et 3me espèce, formant dans leur ensemble un système linéaire subordonné Σ'. Il renferme un nombre maximum δ_1 $(\geqq 0)$ d'intégrales *essentiellement de* 3me *espece*, $u_{p+1}, \ldots, u_{p+\delta_1}$, formant un système linéaire Σ_1. En disant

„essentiellement" de 3me espèce nous entendons exprimer que toute combinaison linéaire à coéfficients constants, non tous nuls, d'intégrales de Σ_1, est de 3me espèce. Analoguement par rapport aux intégrales de 2de espèce, pour lesquelles nous aurons un nombre maximum δ_2 ($\geqq 0$) d'intégrales essentiellement de 2de espèce $u_{p+\delta_1+1}, \ldots, u_{p+\delta_1+\delta^2}$, formant un système linéaire Σ_2. Le système Σ' a la dimension $\delta_1 + \delta_2 - 1$ et Σ est le système minimum renfermant $\Sigma_0, \Sigma_1, \Sigma_2$. C'est-à-dire que $\pi = p + \delta_1 + \delta_2$. Les systèmes Σ_1, Σ_2 peuvent être choisis arbitrairement en Σ', sous la condition de leur définition.

On a ainsi trois *entiers caractéristiques* p, δ_1, δ_2 de la variété V_π associée au groupe Γ, c'est-à-dire à la matrice des périodes des intégrales $u_1, u_2, \ldots, u_\pi$.

Nous verrons tout à l'heure qu'étant donné π et étant choisie la decomposition $\pi = p + \delta_1 + \delta_2$ de π, ou p, δ_1, δ_2 sont des entiers non négatifs arbitraires (sous la condition $\pi = p + \delta_1 + \delta_2$), il existe toujours des variétés V_π possédantes les entiers caractéristiques choisis.

Nous verrons aussi que le groupe Γ n'est pas déterminé univoquement par V_π, lorsque $\pi > p$ (tandis qu'il est parfaitement déterminé si $\pi = p$).

Une des grandes difficultés dans l'étude de V_π ($p < \pi$) est que cette variété renferme un nombre infini de variétés exceptionnelles. C'est le même qui se présente (suivant un théorème classique de Castelnuovo-Enriques) pour les surfaces appartenant aux classes des reglées et des surfaces rationnelles. Il n'existent pas pour ces surfaces des modèles dépourvus de courbes exceptionnelles, tandis qu'ils existent pour toute autre surface. C'est le même pour V_π, comme nous allons le voir.

On démontre aisément que l'ensemble de points, I, invariant pour Γ, est constitué des pôles et des points logarithmiques des intégrales de Σ'. Mais ces points ne sont pas les seuls points singuliers pour ces intégrales. Il y a aussi pour chacune d'elles des points d'indétermination qui remplissent des variétés algébriques pures à $\pi - 2$ dimensions. Soit M leur ensemble. Alors $I + M$ est une variété algébrique pure H, $\infty^{\pi-1}$, qui renferme tout point singuliers, les points d'indéterminations se présentant aussi comme points d'accumulations des pôles et des points logarithmiques.

Eh! bien, chaque composante de M est exceptionnelle vis-à-vis des transformations de Γ. Soit $E_{\pi-2}$ une telle composante. Alors une transformation (non identique) de Γ change $E_{\pi-2}$ dans une variété $E'_{\pi-1}$, remplie par une involution $\infty^{\pi-2}$ de courbes rationnelles γ, chacune desquelles correspond à un point de $E_{\pi-2}$. En changeant la transformation, la variété $E'_{\pi-1}$ décrit sur V_π un système irréductible $\infty^{\pi-1}$, de variétés analogues et les γ attachées aux variétés $E'_{\pi-1}$ de ce système constituent une involution $\infty^{\pi-1}$, J, qui compose le système susdit. Par le point générique de V_π passe une et une seule courbe γ.

Chaque involution des courbes γ est invariante pour Γ. On a ainsi autant

d'involutions analogues que des variétés d'indétermination; mais en realité il existe sur V_π une infinité de telles involutions. Une seule d'elles est suffisante pour nos déductions ultérieures.

J'ai en effet démontré récemment (Convegno di geom. diff., Venezia, 1953) *que sur toute courbe rationnelle γ d'une involution J est possible de fixer un point sans introduire des irrationalités arithmétiques* (par rapport aux paramètres de γ en J).

Cela complète dans un point fondamental la théorie que j'avais développée en 1947, parce qu'on peut à présent éliminer l'ennuyante hypothèse de travail, que j'avais appelé „hypothèse L": c'est-à-dire l'hypothèse de l'existence d'un modèle (non singulier) de V_π tel qu'aucun composante $\infty^{\pi-1}$ de la variété invariante H soit exceptionnelle.

Nous pouvons maintenant énoncer les théorèmes fondamentaux de la théorie des fonctions quasi abéliennes:

Théorème de caractérisation. Pour qu'une V_π possédant un groupe continu du type Γ ait l'irregularité superficielle $p < \pi$ (c'est-à-dire qu'elle soit quasi abélienne) il faut et il suffit que Γ renferme un sous groupe rationnel ∞^δ, ou $\delta < \pi$ (et alors $p = \pi - \delta$). La V_π est en conséquence le produit directe d'une variété de Picard V_p et d'un espace linéaire S_δ; et viceversa.

Si $p = 0$ (et par suite $\delta = \pi$) *la V_π est rationnelle.* Tout cela explique à priori l'existence sur V_π d'un nombre infini de variétés exceptionnelles et de groupes de type Γ.

Pour établir les théorèmes d'existence, qui vont suivre, il faut préablement déterminer les expressions des intégrales simples de 2de et 3me espèce d'une V_p de Picard en fonction des intégrales de 1re espèce; ce qui donne une extension remarquable des formules classiques de la théorie des fonctions elliptiques et des fonctions abéliennes *speciales*. Avec ces moyens on arrive plus aisément aux théorèmes suivantes:

I Théorème d'existence. Il existe toujours un corps de fonctions quasi abéliennes dont la matrice des périodes est la même de celle des périodes des p intégrales simples de première espèce d'une V'_p de Picard et de δ_1 intégrales simples de 3^{me} espece et de δ_2 intégrales simples de 2^{me} espèce, génériquement choisies sur V'_p.

En disant „génériquement choisies" nous entendons que les variétés logarithmiques des intégrales de 3me espèce soient *simplement liées*[5]) et que les δ_2 intégrales de 2de espèce soient *essentiellement de 2de espèce*.

[5]) Dans mes premiers travaux sur la base j'ai demontré que les variétés logarithmiques (nécessairement pures et de dimension $\pi - 1$ si la variété ambiante est ∞^π) d'une intégrale de 3me espèce, sont algébriquement liées. On doit supposer pour notre théorème qu'il n'y ait qu'un de ces liens indépendants.

Après une substitution linéaire à determinant non nul sur les intégrales conçues comme des variables inépendantes, la matrice formée par leur $2p + \delta_1$ périodes (cycliques et polaires) se réduit à la *forme normale* suivante

$$(T) \qquad \begin{vmatrix} A & \Omega & 0 \\ 0 & \Omega_1 & B \\ 0 & \Omega_2 & 0 \end{vmatrix},$$

ou $| A\Omega |$ est la matrice de Riemann attachée à V'_p; $| 0\Omega_1 |$, $| 0\Omega_2 |$ sont les matrices respectives des périodes cycliques des intégrales de 3^{me} et de 2^{de} espèce; et B, matrice carré d'ordre δ_1, est la matrice des périodes polaires des intégrales de 3^{me} espèce.

Une ultérieure substitution linéaire (à déterminant non nul) sur les dernières intégrales réduit égal à zéro tout élément de B, sauf ceux de la diagonale principale qui deviennent égaux à $2\pi i$.

II Théorème d'existence. Sur une V_π quasi abélienne d'irrégularité superficielle $p < \pi$, il existe un nombre infini de groupes continus du type Γ associés, suivant le théorème I, au mème corps de fonctions quasi abéliennes, et ces groupes dépendent du choix d'une transformation crémonienne arbitraire dans un espace linéaire.

Ajoutons que parmi les groupes Γ sur une même V_π il y en a qui satisfont à l'hypothèse L, et d'autres qui en échappent.

Mais le plus significatif des théorèmes d'existence est le suivant.

III Théorème d'existence. Etant donné une matrice du type (T), pour qu'il existe un correspondant corps de fonctions quasi abéliennes il faut et il suffit que la matrice $| A\Omega |$ soit riemannienne. Les autres matrices partielles $| \Omega_1 |$, $| \Omega_2 |$ peuvent etre choisies arbitrairement.

Résultat d'une simplicité inattendue vis-à-vis des difficultés qu'on a du surmonter pour l'atteindre.

Pour ce qui concerne les conséquences de ces théorèmes je renvoi à mon mémoire du 1947.

Je me borne a rappeler (parce qu'il est nécessaire pour la suite) que la détermination du nombre des modules d'un corps K de fonctions quasi abéliennes conduit à envisager un quatrième entier caractéristique du corps, qui est (lui même) comme les caractères p, δ_1, δ_2, invariant vis-à-vis des transformations birationnelles: c'est le rang ϱ de la matrice carré $| B |$.

Les applications à la géométrie algébrique sont nombreuses et très intéressantes. Je me suis borné (dans le mémoire de 1947) seulement à quelques essais initiaux en traitant complètement le cas particulier $\pi = 2$.

Le role de la surface de Kummer dans la théorie des surfaces hyperelliptiques est ici joué par la surface de Plücker (surface de complex d'un complex

quadratique général). Elle est rationnelle, de 4me ordre, avec une droite double ordinaire et 8 points doubles ordinaries hors de la droite.

On trouvera plusieurs problèmes qui interessent soit l'analyse soit la géométrie, dans mon mémoire de 1947 et particulièrement au n. 57.

Parmi les problèmes non resolus il y a le problème de déterminer les conditions auxquelles doit satisfaire une matrice ayant π horizontales et $\mu < 2\pi$ verticales pour qu'elle puisse être réduite au type (T) et elle soit pourtant la matrice des périodes d'un corps de fonctions quasi abéliennes.

Je termine en résumant les remarquables contributions à la théorie données dans ces derniers temps par Fabio Conforto. Je saisis l'occasion pour envoyer encore une fois mes souvenirs à la mémoire de l'ami perdu.

Le but principal de Conforto était d'arriver à une théorie arithmétique des fonctions quasi abéliennes analogues jusqu'au possible à celle des fonctions abéliennes.

Mais il y a là des différences profondes et des difficultée absolument nouvelles à surmonter.

La raison principale n'est pas seulement qu'à un corps de fonctions quasi abéliennes réponde un nombre infini de groupes Γ, mais aussie qu'à une matrice (T) peuvent correspondre plusieurs corps de fonctions quasi abéliennes; tandis que dans le cas abélien il y a correspondence biunivoque entre le corps et le groupe Γ, entre la matrice des périodes et le corps. C'est cette correspondence que dans ce dernier cas permet de substituer à l'étude des propriétés fonctionnelles l'étude des propriétés arithmétiques de la matrice de Riemann, comme l'a fait Scorza. Dans le cas quasi-abélien le détachement du fonctionnel du coté arithmétique n'est pas possible.

Conforto posa à la base de ses recherches la théorie des correspondances univoques sur une V_π quasi abélienne de Picard répresenté par des congruences du type

$$u' \equiv \Lambda u + \lambda \pmod{\omega} \tag{1}$$

étant u la matrice à π horizontales et à 1 verticale dont les éléments sont les intégrales $u_1, u_2, \ldots, u_\pi$, c'est-à-dire les variables du corps de fonctions quasi abéliennes envisagé sur V_π; ω la matrice des périodes, Λ une matrice carré d'ordre π et λ une matrice à π horizontales et à une verticale.

Pour que la congruence ait un sense il faut et il suffit que ω satisfait à une *relation de Hurwitz-Conforto* (comme on a proposé de la nommer), c'est-à-dire

$$\omega I = \Lambda \omega$$

ou I est une matrice carré convenable d'ordre μ (μ nombre des périodes, c'est-à-dire de verticales de ω) à *éléments entiers*.

Les rapports entre les matrice A et I donnent lieu à des faits imprevus en comparaison avec le cas abélien. En outre tandis que dans le dernier cas la (1) donne toujours une correspondance *algébrique*, dans le cas quasi-abélien ($\mu < 2\pi$) la correspondence, tout en restant analytique, peut être *transcendante*.

Ces propriétés sont contenues, avec bien d'autres, dans plusieurs notes et mémoires de Conforto, publiées entre 1948 et 1954. Le dernier (postume, dans les Annali di Matematica) contient la détermination complète des correspondances (1); un autre mémoire de 1949 donne toutes les rélations de Hurwitz-Conforto. Ici joue un rôle fondamentale le caractère ϱ, dont j'ai parlé plus haut (rang de la matrice B). Les matrices quasi abéliennes qui présentent l'analogie plus étroite avec les matrices de Riemann (et que comprennent en particulier ces dernières) sont celles pour lesquelles $\varrho = \delta_2$.

Cette théorie arithmétique-fonctionnelle, suivant le programme de Conforto, est bien loin d'être achevée.

Plusieurs élèves de l'Istituto di Alta Matematica en Rome travaillent maintenant suivant les directives laissées par Conforto. P.ex. un de ces élèves, M. Benedicty a étudié les propriétés des fonctions quasi abéliennes que j'avais nommées *speciales* et les conditions pour que une matrice du type (T) soit équivalente à une matrice du même type, avec des entiers caractéristique et des diviseurs différents. Il vient maintenant d'envisager le problème de déterminer ces entiers et ces diviseurs pour toute matrice du type (T) équivalente à une matrice donnée de tel type.

ROME, ISTITUTO NAZIONALE DI ALTA MATEMATICA.

PROBLÈMES RÉSOLUS ET PROBLÈMES NOUVEAUX DANS LA THÉORIE DES SYSTÈMES D'ÉQUIVALENCE

FRANCESCO SEVERI *)

J'ai publié le premier Mémoire sur cette théorie en 1932, après plus de trente ans de reflexions.

Dès le commencement de mon activité scientifique j'avais acquéri la persuasion que les instruments classiques (systèmes linéaires) n'étaient plus suffisants pour développer la géométrie des transformations birationnelles sur les variétés algébriques; mais qu'il fallait envisager sur ces variétés les systèmes algébriques non linéaires de sousvariétés et en outre des systèmes convenables, encore inconnus, faisant jouer non seulement les fonctions rationnelles isolées (comme dans les systèmes linéaires) mais aussi bien les ensembles de plusieurs fonctions rationnelles, dont la considération simultanée aurait conduit certainement à des invariants nouveaux.

Je suis arrivé ainsi, dans un premier temps, il y a un démisiècle, à l'équivalence algébrique et aux problèmes de la base qui s'y rattachent; et dans un second temps à l'équivalence rationnelle.

Comme ici je parlerai presque exclusivement d'équivalence rationnelle, je dirai tout court „équivalence", en ajoutant une qualification seulement si j'envisagerai d'autres espèces d'équivalence.

Je rappelle vitement les définitions fondamentales [1]).

Soit M_r une variété irréductible (dont nous envisageons toujours un modèle non singulier); et $|A^1|$, $|A^2|$, . . ., $|A^{r-h}|$ soient $r-h$ systèmes linéaires de hypersurfaces sur M_r, génériques dans le sens qu'on va fixer.

Nous supposons que la variété V_h commune ad $r-h$ hypersurfaces prises

---

*) Conférence tenue le 10 septembre 1954 au „Symposion of Algebraic Geometry" (Amsterdam).

[1]) Voir mes leçons sur les *Serie, sistemi d'equivalenza e corrispondenze algebriche sulle varietà algebriche* récueillies par MM. Conforto et Martinelli (Roma, Edizioni Cremonese, 1942), ainsi que le rapport de F. Conforto, *Lo stato attuale della teoria dei sistemi d'equivalenza e della teoria delle corrispondenze algebriche tra varietà* (Atti del Convegno matematico presso l'Istituto Nazionale di Alta Matematica, in Roma, 1942). Voir aussi la mise à point plus récente de l'Auteur dans "Rendiconti Lincei" (16 avril et 11 mai 1955; notice ajoutée en corrigéant les épreuves).

génériquement une pour chacun de ces systèmes, soit pure et ait exactement la dimension h et non plus grande.

Alors l'ensemble des variétés V_h, complété, s'il est nécessaire, par les variétés à h dimensions (*specialisations*), chacune desquelles est d'accumulation pour les V_h susdites, résulte algébrique (de plus rationnel). C'est ce que nous nommons un *système intersection ou système élémentaire d'espèce h* (une *série* pour $h = 0$).

Va sans dire que chacune des specialisations auxquelles nous avons fait allusion est de dimension h, même s'elle est contenue dans l'intersection de $r - h$ hypersurfaces particulières, des systèmes donnés, se coupant dans une variété pure ou impure, avec componentes de dimension $> h$ [2]).

Le symbole d'intersection $A^1A^2 \ldots A^{r-h}$ représente ainsi toujours une variété effective pure de dimension h, qui est indéterminée (toutefois d'une manière précise) seulement pour des A particulières dans les systèmes linéaires donnés.

Il est bon d'ajouter explicitement que l'intersection effective $A^1 \ldots A^{r-h}$ doit être envisagée en attachant à toute componente irréductible sa multiplicité d'intersection, suivant des précisations bien connues [3]).

On peut remarquer (mais la chose ici n'est pas essentiel!) que si le système élémentaire a la dimension > 1, la partie variable de sa variété générique est irréductible ou bien elle a seulement des composantes simples (à multiplicité d'intersection égale à 1). Cela découle de l'application répétée d'un théorème classique de Bertini [4]).

L'ensemble des systèmes élémentaires fait déjà jouer dans notre géométrie les groupes de $r - h$ fonctions rationnelles du point de M_r.

Mais la considération des systèmes élémentaires n'est pas suffisante pour créer la théorie que nous songeons, car l'ensemble des systèmes élémentaires d'espèce donnée, sur M_r, ne constituant pas un groupe (abélien) par rapport à

---

[2]) Ces spécialisations, obtenues ainsi comme des limites, constituent l'équivalence fonctionnelle de la composante de dimension anormale, dans le sens envisagé dans mes leçons citées (voir la pag. 17).

[3]) Voir p. ex. Severi, *Le diverse concezioni di varietà nella geometria algebrica* (Memorie della Accademia Nazionale dei XL, vol. II, Roma, 1951).

[4]) Une autre propriété des systèmes élémentaires qui ne trouve pas application dans la suite est que chacun système élémentaire possède un certain caractère involutoire. On peut même donner au symbole $A^1A^2 \ldots A^{r-h}$ une signification plus ample, à équivalence *linèaire* près des hypersurfaces de M_r. Cette extension [pag. 73 de mes leçons citées] n'est pas nécessaire dans cette simplification de l'exposition originaire. Toutefois elle se trouvera aussi développée dans une Note que l'Auteur publiera prochainement (dans le volume en honneur de Lefschetz) sur la valeur du symbole d'intersection virtuelle des variétés algébriques par rapport soit à l'équivalence algébrique soit à celle rationnelle.

la somme, ne pourrait pas conduire à des résultats vraiment intéressants. En effet, la considération des systèmes linéaires sur une hypersurface est réellement fructueuse en tant qu'ils forment un groupe abélien par rapport à la somme.

J'avais remarqué, dans mes leçons citées (pagg. 37 et 80) que la somme de deux systèmes élémentaires n'est toujours un système élémentaire, en donnant l'exemple bien simple des droites de l'espace ordinaire. Le système des droites est élémentaire; mais non plus le système des m-ple de droites (m donné $\geqq 2$).

De cela naissait la nécessité d'envisager des systèmes que tout en faisant jouer simultanément plusieur fonctions rationnelles, donnaient lieu à des groupes abéliens. J'abouti en 1934 à la définition des systèmes cherchés, de la manière suivante, qui apparait aujourd'hui la plus simple et naturelle [5]).

Appellons-nous pour brevité *différences élémentaires* les différences de couples de variétés pures de dimension h, appartenant à un même système élémentaire.

Deux variétés (virtuelle ou effectives) V_h, $\overline{V}_h$ s'appellent *rationnellement équivalentes* (ou simplement équivalentes) et nous écrivons

$$V_h \equiv \overline{V}_h,$$

lorsque la variété virtuelle $V_h - \overline{V}_h$ est une combinaison linéaire (à coéfficients entiers, positifs, négatifs ou nuls) de différences élémentaires. Lorsqu'on doit rappeler la dimension de la variété ambiant on parle d'équivalence *r-dimensionnelle* des variétés ∞^h ou *d'espèce h*.

De la remarque faite que les systèmes élémentaires ne forment pas un ensemble fermé par rapport à la somme, naisse le problème de savoir si pour des M_r particulières, cet ensemble peut être fermé. M. Morin a repondu à cette démande pour $r = 2$ (1941) en démontrant que si sur une surface la somme de deux systèmes élémentaires est *toujours* un système élémentaire la surface appartient à la classe des surfaces réglées (rationnelles ou irrationnelles). Le problème est ouvert pour $r > 2$.

Pendant la construction de la théorie j'avais fait plusieurs fois allusion au point de vue groupale des définitions données. M. Todd heureusement développa à son tour plus tard ce point de vue, et il posa ainsi les diverses espèces d'équivalence sur un même plan général.

Voici la conception de M. Todd. L'ensemble des variétés virtuelles (pures) de dimension donnée h sur M_r, constitue, par rapport à l'addition, un groupe abélien G. Choisissons un sous groupe quelconque H de G comme classe zéro et

---

[5]) Une idée des difficultés, pour ainsi dire psychologiques, qui s'opposaient au débouchement naturel de la recherche est donnée par la pénétrante récension de A. Comessatti dans le Bollettino dell' Unione Matematica italiana, 1942."

disons que V_h, $\overline{V}_h$ sont équivalentes (par rapport à H) lorsque leur différence est un élément de H.

Dans l'*équivalence rationnelle* le groupe H est le groupe (minimum) renfermant toutes les différences élémentaires et leurs combinaisons linéaires à coéfficients entiers; dans l'*équivalence algébrique* les éléments de H sont les différences des couples de variétés de chaque système irréductible de variétés V_h; etc.

Va sans dire que toute équivalence jouit des propriétés réflexive, symétrique et transitive et que si H est invariant elle est invariante vis-à-vis des transformations birationnelles, comme dans les deux types d'équivalence envisagés.

La nécessité de considérer l'équivalence rationnelle dans le domaine virtuel, amène que dans cette théorie on doit rénoncer à la notion usuelle de système d'équivalence *complet*.

En effet la totalité T des variétés équivalentes à une variété donnée n'est pas ni algébrique ni dimensionnelle. Cela n'est pas surprenant parce que le même se passe pour l'ensemble des hypersurfaces *virtuelles linéairement* équivalentes à une hypersurface donnée. Dans le champs effectif par contre, ce dernier ensemble possède une dimension déterminée et il est linéaire, ce qu'est basilaire dans la théorie de l'équivalence linéaire et ce qui donne origine au théorème de Riemann-Roch.

Si l'on se borne au domaine effectif, comme dans quelques problèmes il est utile ou nécessaire de le faire, on peut envisager les systèmes complets dans le sens usuel, même s'il s'agit de systèmes d'équivalence.

Précisément: on peut prendre à l'intérieur de T des *systemes d'équivalence* qui soient algébriques et irréductibles. Si un tel système est formé par des variétés effectives on dira qu'il est *complet* lorsqu'il n'est pas contenu en T dans un système analogue plus ample. La seule différence (seule mais très importante!) à ce sujet, entre équivalence linéaire et équivalence rationnelle, est que dans l'équivalence rationnelle on ne peut pas affirmer toujours l'unicité d'un système complet renferment une variété donnée, ce que se verifie toujours dans l'équivalence linéaire [6]).

Une dernière remarque. Un système d'équivalence (r-dimensionnelle)

---

[6]) Et même dans l'equivalence algébrique, toutefois si l'on rénonce à la condition que le système complet soit irréductible. En effet les variétés algébriques de M_r algébriquement équivalentes à une variété donnée, se distribuent en un nombre fini de systèmes algébriques irréductibles et le système réduetible ainsi obtenu est déterminé d'une façon unique par une quelconque de ses variétés. Voir Severi, *Il punto di vista gruppale nei vari tipi d'equivalenza sulle varietà algebriche* (Commentarii mathematici helvetici, 1948, p. 200).

d'espèce h ne rempli pas nécessairement la variété M_r toute entière: il peut remplir seulement une sous variété N_s ($s < r$). Le système même n'est pas alors nécessairement un système d'équivalence (s-dimensionnelle) à l'intérieur de N_s. (On donne aisement des exemples). S'il est un système d'équivalence même dans N_s nous disons que ses variétés sont *demifixes* par rapport à M_r.

On peut supprimer ou ajouter à chaque variété d'un système d'équivalence des variétés fixes ou démifixes: le système reste un système d'équivalence.

Nous sommes maintenant en mesure d'exposer les propriétés fondamentales des systèmes d'équivalence.

Les système linéaires peuvent être coupés sur M_r par les systèmes de hypersurfaces de l'espace linéaire S_d ou M_r est plongée. J'ai donné en 1937 une construction projective analogue des systèmes d'équivalence:

Un système de variétés effectives ou virtuelles à $d + h - r$ dimensions de S_d coupe sur M_r, eventuellement hors de variétés fixes ou démifixes, un systeme d'équivalence d'espece h. Réciproquement: tout système d'équivalence d'espece h sur M_r peut se couper ainsi.

La totalité des variétés effectives ∞^h, d'un ordre donné m, de S_d, est elle-même un système algébrique d'équivalence, dont les composantes sont des systèmes irréductibles complets, qu'on appelle *familles* de variétés (pures, effectives) d'ordre m.

Pour ce qui concerne les variétés effectives sur M_r on peut démander si un système irréductible de telles variétés équivalentes peut être coupé sur M_r par un système irréductible de variétés effectives à $d + h - r$ dimensions de S_d. Je l'ai démontré seulement pour les séries d'équivalence de groupes effectives sur une surface (1935, 1942). Je crois que le théorème soit vrai en général.

L'instrument géométrique que j'ai employé pour ces questions est une convenable représentation d'une V_h de S_d ($h < d - 1$), que j'ai nommée *représentation monoïdale graduale.*

Dans les représentations connues (celle classique monoïdale de Cayley-Halphen et celle de la figuration principale de Kronecker, que j'eu l'occasion d'éclaircir dans une utile discussion avec M. Perron) la variété est représentée ou bien comme intersection partielle de $d - h$ hypersurfaces se coupant ulterieurement dans une variété de structure connue ou bien comme intersection complète d'un nombre surabondant de hypersurfaces.

En envisageant le problème dans le domaine virtuel, au moyen de la représentation graduelle on aboutit à une solution de simplicité extrême:

Toute V_h (effective ou virtuelle) de S_d est intersection complète de $d - h$ hypersurfaces virtuelles.

Chaque V_h se présente ainsi en S_d comme une somme algébrique de variétés

effectives ∞^h, chacune desquelles est intersection complète de $d - h$ hypersurfaces.

De cela découle le corollaire très suggestif suivant:

Dans le domaine virtuel tout système d'équivalence sur M_r est un système élémentaire.

La proposition réciproque est presque évidente.

En particulier les familles de V_h effectives d'ordre donné m en S_d sont les *traces* dans le domaine effectif d'un système virtuel d'équivalence, qui est intersection complète de $d - h$ systèmes linéaires virtuels de hypersurfaces; et qui par conséquent est *rationnel.*

Les théorèmes ci-dessus, que j'avais donné en 1937, ont été obtenus à nouveau en 1951 par M. Samuel par une méthode mixte d'algèbre abstraite et de géométrie „italienne" (comme l'Auteur l'appelle!), dans des corps plus généraux du corps complex.

En vertu du théorème démontré, donnant une condition nécessaire et suffisante, on pourrait prendre le théorème même comme définition des systèmes d'équivalence; mais la simplicité de la définition n'evitérait pas les difficultés essentielles de la théorie.

Pour achever les lignes principales du tableau, il faut ajouter le théorème suivant (que j'ai donné en 1937 en liaison avec les théorèmes précédents).

Toute système d'équivalence sur M_r est renfermé dans un système rationnel de variétés de meme ordre et dimension. Réciproquement: Tout systeme rationnel (il suffit même qu'il soit unirationnel) est un systeme d'équivalence.

Il ne faut pas oublier qu'il s'agit de systèmes de variétés *virtuelles*, sans quoi on tomberai en des fautes ou malentendus.

P.ex. les familles effectives d'ordre et dimension donné de S_d appartiennent à un système rationnel virtuel d'équivalence, comme nous l'avons déjà remarqué, mais on ne peut pas conclure qu'elles sont des systèmes rationnels.

J'ai plusieurs raisons pour croire qu'en realité ces familles soient des systèmes du moins unirationnels; mais il faudrait le démontrer [7]).

La *rationalité* de notre type d'équivalence donne l'analogie plus étroite possible avec le *linéarité* de l'équivalence entre hypersurfaces. Qu'il soit impossible d'avoir une relation générale d'équivalence linéaire entre les V_h ayant $h < d - 1$, est déjà démontré du fait que dans un espace linéaire les plus simples familles de variétés, qui sont les familles d'espaces linéaires,

---

[7]) Dans la Note pour le volume Lefschetz, dont j'ai parlé dans 4), je démontre que tout système d'équivalence (rationnelle) de variétés effectives sur M_r jouit de la propriété que deux variétés quelconques sont toujours contenues dans un système rationnel ∞^1 de variétés du système ce que rend plus vraisemblable la propriété énoncée dans le texte.

sont des grassmanniennes, c'est-à-dire des variétés d'ordre invariantif relatif > 1 [8]).

Après les théorèmes basilaires nous pouvons parler des propriétés des systèmes d'équivalence invariants vis-à-vis des transformations birationnelles.

Au point de vue hystorique, il faut rappeler, en première ligne, la série $| S |$ sur une surface algébrique F. J'ai donné cette série dans le premier mémoire de 1932, avec une construction valable seulement sur les surfaces irrégulières et tout de suite, dans la même année, avec une construction valable sur toute surface.

La série $| S |$, d'ordre $I + 4$, I étant l'invariant de Zeuthen-Segre de la surface, donne l'interprétation fonctionnelle de cet invariant.

Une seconde série invariante, se présentant par elle-même, est la série caractéristique virtuel $| \Omega |$ du système canonique (impur) virtuel de F. Elle donne l'interprétation fonctionnelle du genre linéaire de la surface.

Les deux séries jouissent de l'invariance *relative*, c'est-à-dire qu'elles sont invariantes seulement vis-à-vis des transformations birationnelles sans exceptions.

Toutefois leur somme $| S + \Omega |$ est invariante par toute transformation birationnelle, même avec exceptions: est un invariante *absolu*. Elle est d'ordre 12 $(p_a + 1)$, ou p_a est le genre numératif de F; de sorte qu'elle donne l'interprétation fonctionnelle de p_a.

Ayant ainsi les interprétations fonctionnelles de tous les invariants numératifs de la surface, on peut suivre la même méthode de traduction fonctionnelle des passages numératifs, que j'ai employé pour obtenir la série $| S |$, en transformant un procédé numératif de mon maître Corrado Segre. On arrive alors, en supérant des difficultés plus ou moins sensibles, à interpréter fonctionnellement toute rélation numérative, ayant un sens sur une surface quelconque; ce qui grandit les horizons de la géométrie numérative.

Plusieurs interprétations de ce genre ont été obtenues par Enriques, Campedelli, B. Segre, Eger, Thullen.

Dans le mémoire de 1932 j'avais nommé *série canonique* la série $| S |$ de la surface. La dénomination est aujourd'hui généralement adoptée pour la série $| S |$ et analoguement pour ses extensions; toutefois en réalité $| S |$ n'a plus droits de $| \Omega |$ à s'appeller canonique.

Mais il faut ajouter que $| S |$ a des analogies très étroites avec la série canonique d'une courbe.

---

[8]) Deux variétés non singulières sont dites du *même ordre invariantif relatif* si elles peuvent être posées en correspondance birationnelle sans exceptions. Les variétés d' ordre invariantif relatif 1 sont dites *linéaires*. Voir mes leçons, *Serie, sistemi d'équivalenza* etc., n. 20.

P.ex. lorsqu'on envisage $|S|$ sur une surface irrégulière (comme je l'ai fait la première fois que j'ai considéré la série), la *série d'irrégularité* (formée elle-même de groupes S) se présente comme le lieux des groupes jacobiens des intégrales simples de 1re espèce de F, en analogie avec la série canonique d'une courbe, comme lieu des groupes jacobiens de ses intégrales abéliennes de 1re espèce.

Une seconde analogie se présente en relation au problème de la détermination des courbes et des surfaces irrationnelles possédant des groupes continus d'automorphismes birationnels.

Une courbe (elliptique) possédant un groupe continu abélien ∞^1, absolument transitif, d'automorphismes birationnels, est caractérisée par la propriété que sa série canonique est la série zéro de l'équivalence linéaire.

Analoguement une surface (hyperelliptique de Picard) possédant un groupe continu abélien ∞^2, absolument transitif, d'automorphismes birationnels, est caractérisée par la propriété que son système canonique est le système zéro de l'équivalence linéaire et sa série $|S|$ est la série zéro de son équivalence rationnelle (Severi).

La seule condition que la série $|S|$ soit la série zéro caractérise les surfaces elliptiques de Enriques possédant un groupe continu algébrique, abélien, ∞^1 d'automorphismes birationnels (Dantoni).

D'une troisième analogie nous dirons dans la suite.

Passons aux variétés à plusieurs dimensions.

La théorie invariantive des systèmes d'équivalence sur les variétés à trois dimensions a commencé avec un mémoire de M. B. Segre (1934), suivi après par d'autres mémoires du même auteur.

Sur une M_3 on trouve naturellement une double extension de la série $|S|$: c'est-à-dire un système d'équivalence constitué de courbes et une série d'équivalence. Cette dernière donne l'interprétation fonctionnelle de l'invariante de Zeuthen-Segre pour les M_3.

Pour une M_r de dimension quelconque, après quelques idées générales de De Franchis (1932, 1936, 1940), on a les recherches de M. Eger (1937) et surtout l'étude plus générale de M. Todd (1937). Dans cette étude on introduit les systèmes d'équivalence invariants (qu'on appelle deshormais „canoniques") de dimensions de 0 à $r-1$ (ceux de dimension $r-1$ est le système linéaire canonique, bien connu depuis Clebsch et Noether), qui donnent la complète extension de la série $|S|$ et des systèmes définis sur les M_3 par Segre. Todd envisage aussi (comme l'avait fait Segre pour $r=3$) les systèmes intersection des systèmes canoniques à deux à deux en toute manière possible. On a ainsi naturellement d'autres invariants d'ou on déduit l'interprétation fonctionnelle de plusieurs rélations numératives.

Ces systèmes jouissent seulement de la propriété d'invariance vis-à-vis des transformations birationnelles sans exceptions. Sur le problème difficile d'établir le comportement vis-à-vis des transformations birationnelles quelconques on a des recherches de Todd (1937 et 1938) et de B. Segre (1954) se rapportant essentiellement, ces dernières, à quelque cas des transformations appelées dilatations.

M. Todd a commencé aussi à envisager (1939) des systèmes d'équivalence *covariants* d'une variété V_h sur M_r $(h < r)$. Ils se rélient aux anciens *caractères d'immersion* que j'avais introduit en 1902 dans un mémoire sur les intersections des variétés algébriqus.

Mais d'autres contributions remarquables et riches de rapports topologiques se trouvent dans un récent mémoire de M. B. Segre (Annali di Matematica), qui sera suivi bientot par un second sous presse.

Puisque M. Segre même a déjà fait en propos une conférence ici, je puis me borner à un très court apercu qu'il ne pourrait pas manquer dans mon rapport.

En relation avec les sous variétés V d'une M_r, Segre étude les anneaux d'équivalence introduit par M. Todd. Les éléments de l'anneau (attaché à une V) qui possèdent un élément inverse, donnent les successions attachées à V, constituées par V et par des convenables sous variétés de V. On définit aussi une succession covariante d'immersion de V en M_r, qui étend la notion de succession de variétés caractéristiques d'une hypersurface et que l'on peut faire aussi dériver de celle-ci par une transformations birationnelle convenable, qui *dilate* V dans une hypersurface.

La *succession canonique* de V se déduit par des simples opérations algébriques à partir de la succession covariante d'immersion de la variété diagonale de $V \times V$ (qui est elle-même un modèle de V) sur le produit $V \times V$.

La méthode suivie par M. Segre pour obtenir les successions canoniques étend la rélation connue (Comessatti, Severi, Todd) entre la série $|S|$ d'une surface ou d'une variété M_r et le groupe caractéristique de la variété diagonale $\overline{M}_r$ de $M_r \times M_r$ sur le même produit. Sur $\overline{M}_r$ (qui est birationnellement équivalent à M_r) la série d'équivalence $(\overline{M}_r \times \overline{M}_r)$ n'est autre chose que la série $(-1)^r|S|$.

Le résultat est valable jusqu'à $r = 1$ (voir mon „Trattato di geometria algebrica", p. 269) si l'on prend pour série $|S|$ la série canonique. Cela donne la troisième analogie entre la série $|S|$ d'une surface et la série canonique d'une courbe, à laquelle nous avons fait allusion ci-dessus.

Dans la théorie des sytèmes canoniques il faut aussi rappeler un interessant mémoire de M. Hodge qui pose en lumière des liens entre ces systèmes et la théorie de la classe caractéristique de Stiefel-Whitney, étendue par Chern aux

variétés complexes avec une métrique hermitienne (les variétés algébriques en particulier).

Je rappelle enfin, encore une fois, à propos des systèmes canoniques, ma prévision qu'une variété de Picard ∞^p soit caractérisée par la propriété que tous ses systèmes canoniques de la dimension 0 à la dimension $p-1$ soient les variétés zéro de l'équivalence rationnelle respective. La réponse à cette question semble très importante, car on ne connait aucune série de caractère invariants dont les valeurs puissent caractériser une variété de Picard (tandis que cette série de caractères est connue pour une surface: $p_a = -1$, $p_g = P_4 = 1$).

Il reste aussi ouverte la question de caractériser les variétés sur lesquelles seulement une partie des systèmes canoniques se réduit à zéro. Elles possédent peut-être des groups continus de automorphisme birationnels.

Pour achever les grandes lignes de notre tableau nous devons encore parler des rapports entre les systèmes d'équivalence de variétés effectives ou virtuelles, la topologie et la théorie des intégrales des différentes dimensions attachées aux fonctions rationnelles sur M_r [9]).

Commencons par le point de vue topologiques. Il faut rappeler d'abord quelques définitions.

Soient $R(M_r)$, $R(V_h)$, ... les variétés riemanniennes des variétés algébriques M_r, V_h, ...; et soit Σ un système algébrique irréductible de V_h (effectives ou virtuelles) sur M_r.

Nous disons que Σ est un système *à circulation nulle ou pseudonulle* par rapport à une certain dimension (réelle) l, lorsque tout l-cycle de $R(M_r)$, appartenant à Σ, c'est-à-dire rempli de variété $R(V_h)$, est un cycle nul ou respectivement pseudonul (c'est-à-dire un diviseur de zéro).

Nous disons que Σ est à *circulation algébrique* par rapport à une certains dimension l (réelle et nécessairement paire) lorsque tout l-cycle de $R(M_r)$ appartenant à Σ, est un cycle algébrique.

Nous disons enfin que Σ est à *torsions nulles* lorsque tout cycle pseudonul de dimension quelconque appartenant à Σ est en conséquence nul.

J'ai demontré que *tout système d'équivalence est à torsions nulles et à circulations nulles par rapport aux dimensions impaires et à circulation algébrique par rapport aux dimensions paires.*

La réciproque est-elle vraie? J'ai pu démontrer seulement que *si un système de variétés est à circulation nulle ou pseudonulle par rapport aux dimensions impaires et à circulation algébrique par rapport aux dimensions paires, il est un système de pseudoéquivalence.*

---

[9]) On pourra consulter en propos deux Notes de l'Auteur dans les Comptes rendus, déc. 1955 (ajoutée en corrigéant les épreuves).

Nous appellons *système de pseudoéquivalence* tout système dont le multiple suivant un entier convenable $\lambda > 1$ est un système d'équivalence.

Dans le domaine de l'équivalence linéaire le phénomène de la pseudo-équivalence ne se présente pas, car on démontre aisément que si un système continu Σ de hypersurfaces ($h = r - 1$) jouit de la propriété ci-dessus, les hypersurfaces de Σ sont linéairement équivalentes.

Il est fort probable qu'en ajoutant l'hypothèse plus restreinte que le système de variétés soit à torsions nulles on puisse aboutir à la conclusion que le système est d'équivalence. La question est ouverte.

D'autre coté l'existence de systèmes de pseudoéquivalence est certainement liée aux torsions de M_r ou de la variété représentative du système Σ.

En effet j'ai demontré que si M_r, Σ *n'ont aucune torsion le théoreme réciproque du premier théoreme rappelé est vrai.*

L'application de ces théorèmes aux surfaces dont nos connaissances sont plus amples, conduit à des conséquences remarquables.

Ainsi on trouve que *la condition nécessaire et suffisante pour que les points d'une surface constituent une série de pseudoéquivalence est que la surface soit régulière* ($p_g = p_a$) *et de genre zéro* ($p_g = p_a = 0$).

Si la surface possède torsion, comme p.ex. dans le cas de la surface de 6^{me} ordre de Enriques, ayant le bigenre $P_2 = 1$, la série des points est de pseudo-équivalence et non d'équivalence. On a ainsi un exemple effectif de série de pseudoéquivalence.

Si la surface ayant $p_g = p_a = 0$ est sans torsion, comme dans le cas des surfaces rationnelles, la série de ses points est d'équivalence. Est-ce-qu'il y a des surfaces irrationnelles dont les points sont équivalents deux à deux? La question n'est pas résolue. J'ai pu seulement déterminer les caractères des surfaces irrationnelles eventuelles à points équivalents; mais je ne sais pas si elles existent.

Ces questions sont étroitement liées aux problèmes très difficiles de caractériser topologiquement les classes de variétés birationnellement équivalents. Je donnerai quelques exemples en propos à la prochaine occasion du centenaire de Riemann à Berlin.

Au point de vue topologique les surfaces à points équivalents sont caractérisées de la valeur zéro de la connexion linéaire et de la torsion et de la propriété de ne pas posséder que des cycles bidimensionnels algébriques.

Passons maintenant au point de vue transcendant. Je rappelerai d'abord que ma définition originaire de la série $| S |$, comme série d'irrégularité d'une surface, a été étendue par M. Todd (1937) aux variétés superficiellement irrégulières. Mais il est tout naturel de supposer qu'il existe des liens entre les systèmes canoniques d'une variété et ses intégrales multiples de 1re espèce. La

question générale est ouverte. On peut citer en propos un mémorie de M. Hodge (1937) contenant des résultats très suggestifs.

En référant mes résultats sur le sujet, je dois me borner aux théorèmes conclusifs. C'est peut-être peu, mais pour le moment il n'y a pas d'autre.

En premier lieu j'ai donné un théorème que, pour son analogie avec le théorème classique d'Abel sur les courbes, j'ai nommé *le théorème d'Abel pour les séries à circulation linéaire nulle ou pseudonulle* sur une surface:

La condition nécessaire et suffisante pour qu'une série irréductible de groupes de points sur une surface F, d'irrégularité q, soit à circulation nulle ou pseudonulle est que les q formes différentielles de 1re espèce attachées à F, donnent des sommes nulles sur les groupes de la série.

Si les groupes de la série sont effectifs, de sorte qu'on la peut aussi supposer complète par rapport à sa dimension et à la propriété des q formes différentielles, on démontre qu'elle est nécessairement algébrique et on obtient une sorte de *théorème de Riemann-Roch* donnant la dimension de la série, qui est univoquement déterminée par un quelconque des ses groupes.

Pour les séries à circulation algébrique sur une surface on a le théorème d'Abel suivant:

La condition nécessaire et suffisante pour qu'une série irréductibles de groupes de points sur une surface F soit de pseudoéquivalence est que les sommes des formes différentielles linéaires et quadratiques de 1re espèce attachées à F soient nulles.

Les séries complètes à circulation algébrique sont-elles algébriques? Je ne le sais pas, quoique je penche pour l'affirmative. J'ai donné le théorème de Riemann-Roch rélatif à ces séries.

Le théorème conclusif de ces recherches caractérise une série de pseudoéquivalence sur une surface par le fait que les sommes des formes différentielles de 1re espèce, linéaires et quadratiques, relatives aux groupes de la série, sont nulles.

On pourrait trouver aisément des conditions nécessaires liant les formes différentielles de 1re espèce et des differents degrés aux systèmes d'équivalence et de pseudoéquivalence sur une variété M_r, mais la chose deviendra vraiment importante seulement lorsqu'on aboutira à des conditions nécessaires et *suffisantes* caractérisantes ces systèmes.

Les instruments pour achever une théorie, qui sera une de plus artistiques de notre science, seront sans doute soit les intégrales harmoniques, dont la théorie a été crée avec tant de succès par M. Hodge, soit les formes tensorielles introduites heureusement par M. Kähler en 1932 et dont l'Auteur même a commencé, dans ces dernier temps, à développer systématiquement la théorie.

Je devrai maintenant parler des applications des systèmes d'équivalence à la théorie générale des correspondances algébriques entre variétés algébriques;

mais cela n'est pas possible ici. J'en parlerai peut-être prochainement au Colloque pour le centenaire Poincaré.

Je désire plûtot, en terminant, de répéter les paroles d'un jeune et fort capable géomètre francais: je parle de M. Samuel, car il formule une prévision, correspondante à mes vifs souhaits, d'une union de plus en plus intime entre notre esprit géométrique et l'algèbre abstraite.

„Toute une ligne de chercheurs, dépuis Dedekind et Weber jusqu'à C. Chevalley, A. Weil, O. Zariski et leurs élèves, a montré qu'il était possible d'appliquer à la géométrie algébrique les méthodes de l'algèbre abstraite. On s'est également apercu que maints résultats obtenus par „des éclairs d'intuition d'Italiens privilégiés" se pouvaient démontrer par les méthodes rigoureuses de l'algèbre; ainsi s'est tempérée peu à peu la méfiance éprouvée par les algébristes vis-à-vis des méthodes italiennes; et le temps n'est peut-être pas si éloigné où une méthode universelle de traduction permettra, sans changer les idées générales, de faire rentrer toute démonstration „italienne" dans le cadre rigoureux de l'algèbre. Ce temps verra probablement une floraison de mémoires, „d'esploitation" assez analogue à la superproduction contemporaine de hyperaxiomatisateur en mal de généralisation".

ROME, ISTITUTO NAZIONALE DI ALTA MATEMATICA.

THE INVARIANT THEORY OF LINEAR SETS ON AN ALGEBRAIC VARIETY

B. L. VAN DER WAERDEN

Introduction.

Enriques has developed a birational transformation theory for linear sets of curves on an algebraic surface. His main transformation rule is: The excess (i.e. the difference effective minus virtual multiplicity) of every fixed curve, point or neighbour point has to remain invariant under birational transformations. Thus, an invariant notion of "full linear set" (Vollschar, serie completa) is obtained. This theory presupposes the resolution of singularities and the theory of neighbour points.

In an earlier paper (Acta Salamanticensia I, 1947) I have generalized the theory to n dimensions, replacing the neighbour points by valuations of dimension $(n-1)$. This theory still presupposed the resolution of singularities.

In this paper a new foundation will be given, based upon invariant notions only, and not dependent upon the resolution of singularities. No particular model variety V will be used, only the field F of rational functions on the variety and the set U of all $(n-1)$-dimensional exponential valuations of this field. It is known that these valuations may be defined by $(n-1)$-dimensional non-singular subvarieties on suitable models, but these models play no part in the theory.

Leading idea.

According to Enriques' theory, every full linear set can be obtained as the difference of two sets without excess. A set without excess arises from a linear set of forms $\Sigma \lambda_k F_k$ by intersection with V. Fixed parts are left aside, and the virtual multiplicities are defined to be equal to the effective ones. The linear set remains the same, if all F_k are multiplied or divided by one form. Dividing by one of the F_k's, we obtain a *modulus of functions* $\Sigma \lambda_k f_k$, which determines the linear set on V. Now, the notion of a modulus of functions is invariant under birational tranformations. The effective multiplicities are defined by valuations, and the notion of a valuation is invariant too.

General theory.

We start with a field F, whose elements $f, g, \ldots$ will be called *functions*, and with a set U of Krull valuations $u, v, \ldots$ with values from an ordered

group Γ. The value of a function f in the valuation v will be denoted by fv. We have

$$(fg)v = fv + gv$$

$$(f+g)v \geqq \text{Min}\,(fv, gv)$$

In the application of the general theory to algebraic geometry the field F is finitely generated over a subfield K, the field of constants, and the valuations are supposed to give value zero to the constants. The valuations may be restricted to those of dimension $(n-1)$, where n is the degree of transcendency of F over K. In the general theory these restrictions are unnecessary: the theory is valid for every field F and every set of valuations U.

Next we consider *modules* M, generated by a finite number of functions f_k in F. The module M consists of all sums $\Sigma\, n_k f_k$ with integer coefficients n_k. If a field of constants K is given, the sums $\Sigma\, \lambda_k f_k$ may be included in the module or not.

The *effective value* φv of a module M in a valuation v is defined by

$$\varphi v = \varphi_M v = \text{Min}\, f_k v = \text{Min}\, fv \quad (f \text{ in } M).$$

φv may be regarded as a function of v with values in Γ. Such a function will be called a *distribution* (of effective values).

The *closure* $\overline{M}$ of a module M consists of all functions f such that

$$fv \geqq \varphi_M v \text{ for all } v.$$

The closure is a *closed* or *full module* in the following sense: it cannot be enlarged without lowering some of the effective values φv. There is a one-to-one correspondence between full modules M and their distributions φ.

The *full sum* $M + N$ of two full modules M and N is the closure of the module product MN. In calling $\overline{MN}$ a sum rather than a product, we follow the Italian geometers. A confusion with the Dedekind module sum (M, N) is not likely to arise. In an arithmetical theory, the full sum may be called full product.

The distribution φ_{M+N} corresponding to the full sum $M + N$ is the sum of the distributions φ_M and φ_N. Hence the full modules M form an abelian semi-group S, isomorphic to the semi-group of their distributions φ. It follows that the semi-group S can be imbedded in an abelian group G: the group of *virtual modules* $M - N$, corresponding to the distributions $\varphi_M - \varphi_N$.

In order to obtain Enriques' group of virtual linear sets, a factor group G/H must be formed, H being the group of modules (f) generated by single functions f.

Special cases.

1. If F is an algebraic number field and U the set of all non-archimedean valuations, G is the group of fractional ideals. The distribution φ_M corresponding to a closed module (or ideal)

$$M = P_1^{a_1} P_2^{a_2} \dots P_r^{a_r} \tag{1}$$

gives the value a_k to the P-adic valuations v_k defined by the prime ideals P_k in (1), and zero to all others. G/H is the group of ideal classes.

2. If F is a function field of one variable, K its field of constants, U the set of all discrete (say integer-valued) valuations giving value zero to all constants, and V a projective model of F without singularities, there is a one-to-one correspondence between valuations v and sets of K-conjugate points on V. Every distribution $\varphi = \varphi_M - \varphi_N$ gives effective values $\varphi v = 0$ to all valuations v with the exception of a finite number of v's, say v_k, to whom arbitrary integer values a_k are given. If these values are attached to the corresponding points of V, a K-rational *divisor* on V is obtained. Hence G is isomorphic to the group of K-rational divisors on V, and G/H to the group of divisor classes.

3. If F is a function field of n variables and U the set of all $(n-1)$-dimensional valuations giving value zero to all constants, the full modules $\overline{M}$ have finite dimensions, because on every projective model V the functions f_k have only a finite number of pole divisors, with limited multiplicities, and there is only a finite number of linearly independent functions having given pole divisors with limited multiplicities. If the difference $\overline{M} - \overline{N}$ is effective, i.e. if there are functions f besides zero, such that $f\overline{N}$ is contained in $\overline{M}$, these functions also form a K-module of finite dimension. An important and difficult problem is: to determine the closure $\overline{M}$ and its dimension for given M.

Final remark.

The general theory is nearly trivial, but the problems, to which it gives rise in special cases, seem to be very difficult.

ZÜRICH.

ON THE DEFINITION OF RATIONAL EQUIVALENCE OF CYCLES ON A VARIETY

Summary of a correspondence between F. SEVERI, P. SAMUEL and B. L. VAN DER WAERDEN, started by a discussion on foundations at the Symposium on Algebraic Geometry on September 10th, 1954.

B. L. VAN DER WAERDEN

In the discussion following Severi's lecture on Sep. 10th, general agreement was reached concerning the following points.

1. If $A_1, \ldots, A_s$ are cycles of dimension $h_1, \ldots, h_s$ on a variety V of dimension r without singular points, the intersection cycle $A_1 \ldots A_s$ is uniquely defined as a cycle of dimension

$$k = h_1 + \ldots + h_s - (s-1)r$$

provided the point-set intersection of the carriers of $A_1, \ldots, A_s$ has the normal dimension k. The definition of the intersection symbol $A_1 \ldots A_s$ has been given by Severi [1]). Equivalent definitions have been given by Chevalley [2]) and Weil.[3])

2. However, if the dimension of the point-set intersection is larger than k, the symbol $A_1 \ldots A_r$ is defined only as a member of a variety of algebraically equivalent cycles. Now, rational equivalence is stricter than algebraic equivalence. Hence, in setting up a definition of rational equivalence, $A_1 \ldots A_s$ can be treated as a uniquely defined cycle only in the case of normal dimension k of the point-set intersection.

At the end of the discussion, Samuel proposed the following definition of rational equivalence:

Let the symbol $A_1 \ldots A_s$ be defined only in the case of a normal dimension of the point-set intersection of the varieties carrying the cycles $A_1, \ldots, A_s$.

---

[1]) F. SEVERI, Ueber die Grundlagen der algebraischen Geometrie, Abh. math. Sem. Hamburg Univ. 9, p. 335 (1933). See also v. D. WAERDEN, Zur alg. Geom. 14, § 6, Math. Ann. 115, p. 635 (1938).

[2]) C. CHEVALLEY, Intersections of algebraic and algebroid varieties, Trans. Amer. Math. Soc. 57, p. 1 (1945).

[3]) A. WEIL, Foundations of algebraic geometry, New York 1946, Chapter VI and Appendix III. For the equivalence of the definitions see J. IGUSA, On the algebraic geometry of CHEVALLEY and WEIL, J. Math. Soc. Japan 1, p. 198 (1949) and P. SAMUEL, La notion de multiplicité en Algèbre et en Géométrie algébrique, Paris 1951, p. 57.

For divisors, i.e. for cycles of dimension $r - 1$ on M_r, let $A \sim 0$ mean linear equivalence of A to the zero divisor. Now consider for a fixed $s \leqq r$, all intersections of normal dimension $r - s$

$$H = A_1 \ldots A_s$$

where all factors A_j are divisors ~ 0. These intersections and their specializations generate, in the additive group G of cycles of dimension $r - s$, a subgroup H. Let two cycles be called L-equivalent, if they are congruent modulo H.

At the Symposium, everybody seemed to be satisfied with this definition. However, in a letter of September 24th 1954, Samuel himself made the following objection.

The definition of L-equivalence just given does not allow the definition of the intersection of two equivalence classes. For, if X and Y are cycles such that $X \cdot Y$ is defined and X equivalent to zero, it does not follow that $X \cdot Y$ is equivalent to zero. Hence the equivalence relation is too fine to be used for forming a ring of L-equivalence classes.

A correspondence between Samuel, Severi and myself led to the conclusion that the definition of L-equivalence is really too fine and ought to be replaced by one of Severi's earlier definitions.

According to the definition given by Severi on p. 216 of his paper "Il punto di vista gruppale nei vari tipi di equivalenza sulle varietà algebriche", Comm. Math. Helv. 21 (1948), the elements of an equivalence system are sums or differences of *complete* intersections $B_k = A_1 \ldots A_{r-k}$, where $A_1, \ldots, A_{r-k}$ are divisors variable in linear systems. Now if this definition is compared to the definition of the notion *sistema d'intersezione completa* as given in Severi's book [4]) it becomes clear that the word *complete* ought not to be taken in the litteral sense. This is also clear from the following example.

On a quadratic V_4 in S_5 every divisor A is a complete intersection of V_4 with a hypersurface in S_5, and hence every complete intersection B_k has even order. Thus, according to the definition just quoted, cycles of odd order (e.g. single points of V_4) cannot be members of equivalence systems.

It follows that the "complete intersection definition" must be modified by allowing at least the addition of a fixed cycle to the sums and differences mentioned before. This modified definition was given by Severi in his letter of February 10th 1955, and also in his paper "On the symbol of virtual intersection of algebraic varieties", of which Severi kindly sent me the manuscript. Let us call this definition I.

---

[4]) F. SEVERI, Serie, sistemi d'equivalenza, e corrispondenze algebriche sulle varietà algebriche, Roma 1942 p. 35 and 75.

In order to show that Definition I leads to a ring of equivalence classes, the following lemma must be proved: If a fixed cycle C intersects the complete intersections $B_k = A_1 \ldots A_{r-k}$, where $A_1, \ldots, A_{r-k}$ are divisors variable in linear systems, the intersections $B_k C$ are contained in a system of equivalence. In his letter of April 23rd, 1955, Severi informs me that he is able to give a very simple geometrical proof of this lemma.

Another possibility, mentioned by Severi in his letter of Febr. 10th, 1955, would be, to form sums and differences of 3 types of cycles, viz.:

1) complete intersections $B_k = A_1 \ldots A_{r-k}$,

2) semifixed cycles $CA_1 \ldots A_{r-k-h}$, where C is a fixed cycle of dimension $n - h$ $(k < n - h < n)$,

3) fixed cycles C of dimension k.

If this definition II is adopted, the lemma is no longer necessary in order to prove that the equivalence classes form a ring. The lemma implies the equivalence of definitions I and II.

Personally, I prefer the definition given on p. 70 of Severi's book quoted before. This definition allows the exclusion of semifixed intersection systems from a complete intersection system or their inclusion at will ("se si vuole"). In the language of abstract algebra, this definition III may be formulated as follows.

Let V be an indivisible variety of dimension r, without singularities, defined over a field K. Let $|A_1|, \ldots, |A_{r-k}|$ be linear systems of divisors on V and let $A_1, A_2, \ldots$ be independent generic elements of these linear systems. To be more explicit, A_1 is defined as the intersection cycle of the form

$$F_\lambda = F_o + \lambda_1 F_1 + \ldots + \lambda_t F_t$$

with V minus the fixed part of this intersection cycle. Just so, A_2 is defined by a form G_μ, etc. The $\lambda, \mu, \ldots$ are independent indeterminates.

It is well-known that A_1 is irreducible over $K(\lambda)$. On A_1, the forms G_μ (or the cycles A_2) cut out a linear system. Suppose that the cycles of this linear system are not fixed as μ varies, but contain a variable part which may be denoted by $(A_1 \cdot A_2)_{var}$ or by $(V \cdot F_\lambda \cdot G_\mu)_{var}$. Since $(A_1 A_2)_{var}$ is irreducible over $K(\lambda, \mu)$, we may again form the intersection with the forms H_ν defining $|A_3|$, etc. In this way the variable part

$$B = (A_1 A_2 \ldots A_{r-k})_{var} = (V F_\lambda G_\mu \ldots)_{var}$$

is uniquely defined as a variety of dimension k, irreducible over $K(\lambda, \mu, \ldots)$, and also as a cycle on V, because the multiplicity of B as a part of the intersection $V F_\lambda G_\mu \ldots$ is uniquely defined. In the "classical case" (characteristic zero) this multiplicity is 1, by Bertini's theorem. In the non-classical case, the

field K and the forms $F_\lambda, \ldots$ defining $| A_1 |, \ldots$ may be chosen in such a way that the intersection multiplicity becomes $= 1$, but this is only a minor detail.

The variable part $B = (A_1 \ldots A_{r-k})_{var}$ may also be characterized by the property that it is variable over the whole variety V, as $\lambda, \mu, \ldots$ vary. This means that a generic point of B over $K(\lambda, \mu, \ldots)$ is a generic point of V over K. We express this by saying that B has *full variability over* V. We restrict ourselves to those cases, in which the complete intersection $A_1 \ldots A_{r-k}$ actually contains a component of the right dimension k and of full variability over V.

The coordinates of the cycle B are rational functions of the parameters $\lambda, \mu, \ldots$. Hence the generic cycle B, and all its specializations over K, form a rational variety of cycles. This variety will be called an *intersection system of full variability on* V.

If W is any subvariety of V of dimension $h \geqq k$, defined over K, we may also define intersection systems of full variability over W. Such a system will be called an *intersection system of restricted variability on* V. In the special case $h = k$, the system consists of the fixed cycle W only. Both kinds of intersection systems, of full and of restricted variability, will be called *elementary systems of species* k *on* V, *defined over the field* K.

The definition of an elementary system is somewhat more restricted than Severi's, because the omission of parts of smaller variability, which is left free by Severi, is made compulsory by our definition. This is done because, from Severi's wording, it is not quite clear what systems of semifixed intersection cycles may or may not be included. [5]) For the final notion "system of rational equivalence", the omission of systems of smaller variability is without importance, because such systems may be added afterwards. This was pointed out by Severi in his letter of Febr. 10th, 1955.

The final definition is as follows. Let $| B_1 |, | B_2 |, \ldots, | B_m |$ be elementary systems of species k on V, defined over the field K. Let B_1 be any element of $| B_1 |$, generic or not, and B_2 any element of $| B_2 |$, etc. The sum or difference

$$C = B_1 \pm B_2 \pm \ldots \pm B_m \tag{1}$$

and all its specializations over K, form an irreducible variety of cycles. Every subvariety of this variety, whether reducible or not, is called a *system of rational equivalence* on V.

---

[5]) Consider e.g. the intersection of a fixed straight line with a pencil of conics in the plane. From the wording of Severi's definition one might think that it is permissible to reject, for every conic, one of the two points of intersection "at will" and to retain the other. However, the resulting system would not be algebraic. Moreover, this interpretation is excluded by Severi's explanation on p. 72: "mentre le componenti semifisse descrivono sistemi d'equivalenza su varietà subordinate ad M_r."

Obviously, we could also have assumed $B_1, B_2, \ldots, B_m$ to be independent generic elements of the elementary systems $| B_j |$. The generic cycle C then defines a *rational* variety of cycles $|C|$, and systems of rational equivalence may be defined as subvarieties of such a rational variety $| C |$.

Two cycles C_1 and C_2 are called *rationally equivalent*, if they belong to one system of rational equivalence. Obviously, this relation is reflexive (for every single cycle C_1 forms, by itself, a system of rational equivalence) and symmetric. We now prove that it is transitive.

Let C_1 and C_2 be specializations of a cycle C of the form (1), and let C_2 and C_3 be specializations of a cycle C' of the same form (1). We may suppose that in the expressions

$$C = B_1 \pm B_2 \pm \ldots \pm B_m$$
$$C' = B_1' \pm B_2' \pm \ldots \pm B_m'$$

the terms B_i and B_j' are all formed with new indeterminates, algebraically independent with respect to the field of definition K_2 of the cycle C_2. Now among the specializations of the cycle

$$C + C' - C_2$$

we have

$$C_1 + C_2 - C_2 = C_1$$

as well as

$$C_2 + C_3 - C_2 = C_3$$

Hence C_1 is rationally equivalent to C_3.

A *rational equivalence class* or R-class may now be defined as the set of all cycles equivalent to a given cycle.

In order to define the intersection ring of rational equivalence classes, we have to show that, if C and D are any two cycles on V, of dimensions h and k, with $h + k \geqq r$, there exists a cycle C' rationally equivalent to D such that the intersection $C' \cdot D$ is defined as a cycle of dimension $h + k - r$. By Severi's methods, the construction of C' is easy.

The equivalence class of the intersection $C' \cdot D$ remains the same, if C' is replaced by an equivalent cycle C'', or if D is replaced by an equivalent D'; this follows immediately from the definition of rational equivalence.

In this way we obtain a satisfactory set of definitions, equivalent to the set given in Severi's book.

ZÜRICH.

ABSTRACT VERSUS CLASSICAL ALGEBRAIC GEOMETRY

ANDRÉ WEIL

The word "classical", in mathematics as well as in music, literature or most other branches of human endeavor, may be taken in a chronological sense; it then means anything which antedates whatever one chooses to consider as "modern", and may be used to describe remote antiquity or the achievements of yesteryear, according to the mood and the age of the speaker. Sometimes, too, it is purely laudatory and is applied to any piece of work which is thought to be of permanent value.

Here, however, while discussing algebraic geometry, I wish to use the words "classical" and "abstract" in a strictly technical sense which will be explained presently. Until not long ago algebraic geometers did their work exclusively with reference to the field of complex numbers; at the same time they worked on non-singular models, or at any rate their concern with multiple points was merely in order to try to push them out of the way by suitable birational transformations. Thus transcendental and topological tools of various kinds were available, and it was merely a matter of individual taste, personal inclination or expediency whether to use them or not on any given occasion. The most decisive progress ever made in the theory of algebraic curves was achieved by Riemann precisely by introducing such methods. Later authors took considerable pains to obtain the same results by other means. In so doing, they were motivated, at least in part, by the fact that Riemann had given no justification for Dirichlet's principle and that it took many years to find one. Similarly, the use of topological methods by Poincaré and Picard, not to mention some more recent writers, has often been such as to justify doubts about the validity of their proofs, while conversely it has happened that theorems which had merely been made plausible by so-called geometrical reasoning were first put beyond doubt by the transcendental theory.

Now we have progressed beyond that stage. Rigor has ceased to be thought of as a cumbersome style of formal dress that one has to wear on state occasions and discards with a sigh of relief as soon as one comes home. We do not ask any more whether a theorem has been rigorously proved but whether it has been proved. At the same time we have acquired the techniques whereby our predecessors' ideas and our own can be expanded into proofs as soon as they have reached the necessary degree of maturity; no matter whether such ideas are

based on topology or analysis, on algebra or geometry, there is little excuse left for presenting them in incomplete or unfinished form.

What, then, is the true scope of the various methods which we have learnt to handle in algebraic geometry? The answer is obvious enough. Let us call "classical" those methods which, by their very nature, depend upon the properties of the real and of the complex number-fields; such methods may be derived from topology, calculus, convergent series, partial differential equations or analytic function-theory. As examples, one may quote the use of the differential calculus in the proof of the Kronecker-Castelnuovo theorem, of theta-functions in the theory of elliptic curves and abelian varieties, of topology in the proof of the "principle of degeneracy". Let us call "abstract" those methods which, being basically algebraic, are essentially applicable to arbitrary ground-fields; this includes for instance the theory of differentials of the first, second and third kinds (but of course not that of their integrals) and the greater part of the "geometric" proofs of the Italian school. Thus it is plain that, in all cases where an abstract proof is available, it may be expected to yield more than any classical proof for the same result. No one could deny this unless he had made up his mind to ignore fields of non-zero characteristic and was prepared to maintain that a theorem in algebraic geometry which has been proved for the field of complex numbers can always be extended to any field of characteristic 0. There are indeed many cases where this is so; quite often, however, the extension can only be made to algebraically closed fields. As to denying any existence to algebraic geometry of non-zero characteristic, not merely would this, in view of recent developments, amount to denying motion; it would also deprive algebraic geometry of a rich and promising field of possible applications to number-theory, where one cannot do without reduction modulo p.

At present, abstract methods also possess the invaluable advantage of being inherently applicable to varieties with arbitrary singularities, while only the non-singular varieties fall within the scope of all but the most elementary of the classical methods known to us. For instance we have now a fully developed abstract theory of the so-called Picard and Albanese varieties attached to a given algebraic variety. The corresponding classical theory depends upon Hodge's existence theorem for simple integrals of the first kind. In order to apply the latter to a given variety, of course over the field of complex numbers, one has to transform it first into a non-singular variety; this is a famous problem for which no general solution is yet available. The former method, however, requires no other preliminary step than the normalization of the given variety, a very simple process of universal scope; once this has been done, it is not subject to any limitation whatsoever concerning the groundfield. The situation is similar for Severi's theorem of the base and its extension by Néron to the abstract case.

On the other hand, partly because of our habits of thought, partly for more substantial reasons, it is frequently much easier to prove a theorem in the classical case and by classical methods than to prove it abstractly. For instance the so-called principle of degeneracy is almost trivial in the classical case, for elementary topological reasons; its abstract proof by Zariski is an awe-inspiring achievement and requires the formidable apparatus of the abstract meromorphic functions. In the theory of abelian varieties, it is rather obvious that such a variety over complex numbers is a compact commutative Lie group and is therefore isomorphic to a torus of topological dimension $2n$ where n is the algebraic dimension of the variety. The number of elements of given order r in the group is then r^{2n}. This is still so in the abstract case provided r is prime to the characteristic; but there is no easy proof for it at present. Here we meet with a theorem whose abstract formulation requires an assumption involving the characteristic. If we reformulate it, however, by saying that division by r defines an extension of degree r^{2n} of the field of algebraic functions on the variety, then it remains always true, whatever the characteristic may be. This example is fairly typical. If a result which can be formulated in purely algebraic terms is known to be true in the classical case, it almost invariably happens that there is a corresponding result in the abstract theory; just what this may be is sometimes a matter for guesswork.

I do not mean to suggest that classical methods have no other purpose than occasionally to give easier proofs for abstract results under suitable additional assumptions. Algebraic varieties are objects of considerable interest to analysts and topologists; it is right and proper that they should study them for their own sake or as special cases of more general objects with no counterpart in algebraic geometry, for instance complex or quasi-complex manifolds. From the point of view of the algebraic geometer, however, it cannot be denied that the chief use of classical methods is to lend plausibility to results which have then to be attacked directly. Some examples of this will now be discussed in greater detail.

The first one deals with correspondences between curves. If C, C' are two curves, one says that a correspondence between them, i.e. a cycle on the surface $C \times C'$, is equivalent to 0 if it is linearly equivalent to an element of the group generated by all curves $P \times C'$ (where P is any point of C) and $C \times P'$ (where P' is any point of C'). One of Castelnuovo's most interesting theorems gives an "enumerative" criterion for equivalence; it attaches to every correspondence an integer $\delta(X) \geqq 0$, the so-called "equivalence defect", such that $\delta(X) = 0$ is necessary and sufficient for X to be equivalent to 0. Let X' be the correspondence between C' and C obtained by interchanging the two factors of the product $C \times C'$. The multiplication of correspondences being defined in a fairly obvious

manner, $X'X$ is then a correspondence between C and itself. If Z is any such correspondence, let $f(Z)$ be the number of its fixed points, i.e. its intersection-number with the diagonal of the product $C \times C$; let $d(Z)$, $d'(Z)$ be its degrees, i.e. its intersection-numbers with the curves $P \times C$, $C \times P$, respectively. Put

$$S(Z) = d(Z) + d'(Z) - f(Z).$$

It is easily seen that $S(Z)$ is the same for any two equivalent correspondences and that it has the formal properties of a trace on the ring of classes of correspondences. Castelnuovo's equivalence defect can then be expressed as $\delta(X) = S(X'X)$; and his theorem can be written as $S(X'X) \geqq 0$, with $S(X'X) = 0$ if and only if X is equivalent to 0. In this form, it may be regarded as the fundamental theorem on correspondences; for instance, the so-called Riemann hypothesis for function-fields follows from it almost immediately.

Castelnuovo's proof was "geometric"; in other words, it was such that its translation into abstract terms was essentially a routine matter once the necessary techniques had been created; in fact, all modern proofs are based upon the ideas introduced by him and supplemented by the later work of other Italian geometers, particularly Severi, on the same subject. It will now be shown how a rather simple proof can be given in the classical case by using transcendental and topological methods.

In the first place, any correspondence Z between two non-singular varieties V and W over complex numbers induces homomorphisms of the homology groups of V into those of W. If V, W and Z have the same dimension, these homomorphisms map the homology group of V for any dimension into that of W for the same dimension. If V is the same as W, the number of fixed points of Z (its intersection-number with the diagonal) is given by Lefschetz's formula as the alternating sum of the traces of the endomorphisms induced by Z on the homology groups of V. From this it follows immediately that in the case of a curve the integer $S(Z)$ defined above is the trace of the endomorphism induced by Z on the homology group H of C for dimension 1. If g is the genus of C, H is a free abelian group of rank $2g$; for a given choice of generators, an endomorphism of H is represented by an integral-valued square matrix of order $2g$.

Let X be a correspondence between two curves C, C'. Let H, H' be the homology groups of dimension 1 for C and C'; let $\gamma_1, \ldots, \gamma_{2g}$ be generators for H, and $\gamma'_1, \ldots, \gamma'_{2g'}$ generators for H'. Call $E = \| e_{\lambda\mu} \|$ the intersection-matrix for the γ_λ, which is skew-symmetric of determinant 1; call E' the similar matrix for the γ'_ϱ. By a well-known theorem in topology, the cycles $P \times C'$, $C \times P'$ and $\gamma_\lambda \times \gamma'_\varrho$ generate the homology group of dimension 2 for $C \times C'$, so that X must be homologous on $C \times C'$ to a linear combination

$$d \,.\, (P \times C') + d' \,.\, (C \times P') + \sum_{\lambda,\varrho} a_{\lambda\varrho} \,.\, (\gamma_\lambda \times \gamma'_\varrho).$$

Put $A = ||a_{\lambda\varrho}||$. It is easily seen that the matrices of the homomorphisms of H into H' and of H' into H induced respectively by X and by X' are

$$L = {}^t(EA) \quad , \quad L' = AE'$$

where ^{t} denotes the transpose of a matrix. This gives

$$L' = E^{-1} \,.\, {}^tL \,.\, E'.$$

Consider now on C the harmonic differentials, i.e. the real parts of the differentials of the first kind on C; the vector-space of such differentials is of (real) dimension $2g$. Take for this a basis consisting of forms ω_λ respectively homologous to the γ_λ in the sense of de Rham, i.e. such that

$$\int_{\gamma_\lambda} \omega_\mu = e_{\lambda\mu} \qquad (\lambda, \mu = 1, 2, \ldots, 2g).$$

By de Rham's theorems, E is then also the matrix of the integrals $\iint \omega_\lambda \wedge \omega_\mu$ taken on C.

The differential ζ_λ of the first kind with the real part ω_λ has an imaginary part which is also harmonic and can therefore be written as $\sum_\mu c_{\mu\lambda}\omega_\mu$. Put $J = ||c_{\lambda\mu}||$. From the fact that $i\zeta_\lambda$ is again of the first kind, it follows at once that $J^2 = -1$, where 1 denotes the unit matrix. We have $\zeta_\lambda \wedge \zeta_\mu = 0$; integrating this over C and expressing ζ_λ, ζ_μ in terms of the ω_λ, we find $E = {}^tJ \,.\, E \,.\, J$ or the equivalent relation ${}^t(EJ) = EJ$, expressing that EJ is a symmetric matrix. If ζ is any differential of the first kind, we have $i\zeta \wedge \bar{\zeta} \geqq 0$ everywhere, C being oriented in the usual manner. Integrating this over C, we find that the quadratic form with the matrix EJ is positive-definite. These statements on EJ are substantially identical with Riemann's bilinear relations and inequalities for the periods of the integrals of the first kind; nor does the proof just given differ in substance from Riemann's.

Now let again X be as above; define the forms ω'_ϱ, ζ'_ϱ and the matrix J' for C' just as ω_λ, ζ_λ, J have been defined for C. The differential form $\zeta_\lambda \wedge \zeta'_\varrho$ induces 0 on every component of X since such components are algebraic subvarieties of $C \times C'$. Therefore $\iint \zeta_\lambda \wedge \zeta'_\varrho$, taken on X, must be 0. Expressing X as above in terms of a homology basis on $C \times C'$ and expressing ζ_λ, ζ'_ϱ in terms of the ω_λ, ω'_ϱ, one finds, by taking the real and imaginary parts of the double integral, two equivalent relations, one of which is

$$(EA)E' = {}^tJ \,.\, (EA) \,.\, E'J'.$$

Take the transpose of this relation, remembering that ${}^t(EA) = L$ and that

$E'J'$ is symmetric; multiply to the left by E'^{-1} and to the right by $J^{-1} = -J$; we get

$$LJ = J'L.$$

This expresses the fact that X induces a linear mapping of the complex vector-space of differentials of the first kind on C into the corresponding space for C', or also a complex homomorphism of the jacobian variety of C into that of C'.

All this is well-known. The inequality $S(X'X) \geqq 0$ is now easy to prove. In fact, $S(X'X)$ is no other than the trace of the matrix $L'L$, which is the same as that of $M = J^{-1}L'LJ$. This may be written as

$$M = J^{-1} \cdot E^{-1} \cdot {}^tL \cdot E'LJ = (EJ)^{-1} \cdot {}^tL \cdot E'J'L.$$

As the quadratic form with the matrix EJ is positive-definite, it can be transformed into a sum of $2g$ squares by a suitable substitution U. This gives ${}^tU \cdot EJ \cdot U = 1$ and therefore $(EJ)^{-1} = U \cdot {}^tU$. The trace of M is the same as that of the matrix

$$N = U^{-1}MU = {}^t(LU) \cdot (E'J') \cdot (LU).$$

Put $E'J' = \| s_{\varrho\sigma} \|$, $LU = \| x_{\varrho\lambda} \|$; then the trace of N is

$$Tr(N) = \sum_{\lambda} \left(\sum_{\varrho, \sigma} s_{\varrho\sigma} x_{\varrho\lambda} x_{\sigma\lambda} \right).$$

Since $E'J'$ is positive-definite, it is clear that the right-hand side is $\geqq 0$ and that it is > 0 except when $LU = 0$ i.e. when $L = 0$. In order to complete the proof, it only remains to show that L cannot be 0 unless X is equivalent to 0; this is an easy consequence of Abel's theorem.

If I may be allowed a personal note here, this is precisely how I first persuaded myself of the truth of the abstract theorem even before I had perceived the connection between the trace $S(Z)$ and Castelnuovo's equivalence defect. No one with any experience in such matters will fail to acknowledge the cogency of such an argument, even though no proof can be based on it.

Is it possible to extend these results to higher dimensions? Many facts point to a generalization of the Riemann hypothesis which can be stated as follows. Let V be a variety over the field k with q elements. Then there is for each integer ν a correspondence I_ν between V and itself such that to each point of V with coordinates $x_1, \ldots, x_N$ there corresponds by I_ν the point with the coordinates $x_1^{q^\nu}, \ldots, x_N^{q^\nu}$. The fixed points for I_ν are precisely those which have their coordinates in the extension k_ν of degree ν of k. Let N_ν be the number of such points; if V is non-singular, this is the intersection-number of I_ν with the diagonal of $V \times V$.

Now in the classical case the numbers N_ν of fixed points for the successive powers of a given correspondence Z between a compact non-singular variety V

and itself is given by Lefschetz's formula as being equal to

$$N_\nu = \sum_{h=0}^{2n} (-1)^h \sum_{i=1}^{B_h} (\alpha_{hi})^\nu \qquad \text{(A)}$$

where n is the complex dimension of V, B_h its Betti number for the (topological) dimension h, and the α_{hi}, for $1 \leqq i \leqq B_h$, are the characteristic roots of the endomorphism induced by Z on the Betti group of V for the dimension h. This makes it plausible that in the case described above the number N_ν of fixed points of the correspondence I_ν is given by a formula of this type.

Not only has this been found to be so in all cases where the N_ν could actually be computed, but it turns out that the α_{hi} are of absolute value $q^{h/2}$ in all such cases. For $n = 1$ the latter fact is precisely the Riemann hypothesis; if true in general, it is therefore the generalization we have been looking for. Analogy suggests that it must depend upon some generalization of Castelnuovo's theorem or rather of the inequality $S(X'X) \geqq 0$; if so, then presumably this generalization might admit a comparatively easy proof in the classical case by means of Hodge's theory of harmonic differentials.

Before coming to the next example, let me recall the concept of numerical equivalence. Two cycles of the same dimension on a non-singular complete variety are said to be numerically equivalent if their intersection-numbers with every cycle of the complementary dimension are equal whenever they are both defined. In the classical case, two cycles which are homologous to each other are obviously equivalent in this sense; this implies at once that the group of equivalence classes of cycles of a given dimension is finitely generated. In the abstract case, Néron's theorem shows that this is so for divisors (cycles of dimension $n - 1$ on a variety of dimension n) and therefore also for cycles of dimension 1; to prove it for dimensions between 1 and $n - 1$ seems still beyond our reach at present.

Again in the classical case, more precise results are known under special assumptions. For instance, there are varieties whose homology groups are all generated by algebraic cycles; this implies that they vanish for the odd dimensions. If V and W are such varieties, all algebraic cycles on $V \times W$ must then be numerically equivalent to linear combinations of cycles of the form $X \times Y$, where X is a cycle on V and Y is a cycle on W. This must be so, in particular, if V and W are non-singular rational surfaces, since the homology groups of such surfaces are known to have the property in question. Making use of Néron's theorem, we thus get the following purely algebraic statement. Let S, S' be two non-singular rational surfaces; let the X_i be generators for the group of divisor-classes on S modulo algebraic equivalence; let the X'_j be the generators for the corresponding group on S'; then every cycle of dimension 2 on $S \times S'$ is

numerically equivalent to a linear combination of the cycles $P \times S'$, $S \times P'$ (where P is a point of S and P' a point of S') and $X_i \times X_j'$. It does not seem hopeless to try to find an abstract proof for this statement.

Let us for a moment assume it to be true. Applying it to the diagonal of $S \times S$, one deduces immediately from it the validity of Lefschetz's fixed point formula for S in the following form: if X is a correspondence of dimension 2 between S and itself, the number of its fixed points is $d(X) + d'(X) + S(X)$, where $d(X)$, $d'(X)$ are the degrees of X (its intersection-numbers with $P \times S$ and with $S \times P$) and $S(X)$ is the trace of the endomorphism induced by X on the group of divisor-classes on S modulo numerical equivalence. This can then be applied as above to a surface S defined over a finite field k of q elements and to the number N_ν of points of S with coordinates in the field k_ν with q^ν elements. One finds that N_ν is of the form

$$N_\nu = q^{2\nu} + q^\nu \sum_i (\varepsilon_i)^\nu + 1$$

where the ε_i are the characteristic roots for a certain linear substitution of finite order and are therefore roots of unity.

Under the same assumption, one can then verify in this case the following general conjecture. Let V be a non-singular complete variety of dimension n over an algebraic number-field K; for the sake of simplicity we assume that it is embedded in a projective space and write a set of equations for it as $F_\mu(X_0, X_1, \ldots, X_N) = 0$, where the F_μ are homogeneous polynomials with coefficients in the ring of integers of K. Let $B_0, B_1, \ldots, B_{2n}$ be the Betti numbers of V (with $B_0 = B_{2n} = 1$, since V is irreducible, and $B_h = B_{2n-h}$ by the duality theorem). Let $\mathfrak{P}$ be a prime ideal in K such that the equations $F_\mu = 0$, reduced modulo $\mathfrak{P}$, define a non-singular variety $V_\mathfrak{P}$ of dimension n over the residue field $K_\mathfrak{P}$ of K mod. $\mathfrak{P}$; it is not hard to show that all but a finite number of prime ideals in K have that property. Assuming the validity of a formula of type (A) for $V_\mathfrak{P}$, and assuming (as is the case in all examples which could be treated so far) that the integers B_h in it are no other than the Betti numbers of V, call $\alpha_{hi}(\mathfrak{P})$, for $0 \leqq h \leqq 2n$, $1 \leqq i \leqq B_h$, the numbers occurring in the right-hand side of the formula (A) for the variety $V_\mathfrak{P}$; as mentioned before, these numbers are of absolute value $q^{h/2}$ whenever they can be calculated. Put now

$$\Phi_h(s) = \prod_\mathfrak{P} \prod_i \left(1 - \alpha_{hi}(\mathfrak{P}) \cdot N\mathfrak{P}^{-s}\right)^{-1}.$$

Then our examples indicate that $\Phi_h(s)$ coincides (except for a finite number of factors) with the Euler product for a Dirichlet series which can be continued in the whole plane and satisfies a functional equation of the familiar type

$$\Psi(s) = \Psi(h + 1 - s)$$

where Ψ is the product of the Dirichlet series, of a gamma factor and of an exponential factor. It is tempting to surmise that this is always so, but I have little hope that a general proof may soon be found. For non-singular rational surfaces at any rate the results stated above would imply that $\Phi_2(s)$, except for a finite number of factors, is the same as a suitable L-function (in the sense of Artin) for a certain extension of K. For instance, for a non-singular cubic surface in the projective 3-space, one thus gets an L-function belonging to the extension of K determined by the 27 straight lines on the surface. The Galois group for this is known; it is a group of order $2^7 \cdot 3^4 \cdot 5$ and has a simple subgroup of index 2. In general, therefore, the function $\Phi_2(s)$ which we may expect to belong to a given cubic surface is essentially an L-function of a definitely non-abelian type. Here is a rather unexpected connection between number-theory and algebraic geometry.

TABLE OF CONTENTS

SECTION I: ALGEBRA AND THEORY OF NUMBERS

Stated addresses

Davenport, H., Simultaneous diophantine approximation 9
Erdös, P., On additive arithmetical functions and applications of probability to number theory 13
Hlawka, E., Das inhomogene Problem in der Geometrie der Zahlen 20
Jacobson, N., Some aspects of the theory of representations of Jordan algebras 28
Maass, H., Differentialgleichungen und automorphe Funktionen. 34
Néron, A., Propriétés arithmétiques de certaines familles de courbes algébriques 40,481
Northcott, D. G., Specialization methods in algebraic geometry. 41,489

SECTION II: ANALYSIS

Stated addresses

Behnke, H., Funktionentheorie auf komplexen Mannigfaltigkeiten. 45
Bureau, F. J., Les solutions élémentaires et le problème de Cauchy 58
Cartwright, M. L., Some aspects of the theory of non-linear vibrations 71
Cesari, L., Retraction, homotopy, integral 77
Chandrasekharan, K., On some problems in Fourier analysis 85
Erdélyi, A., Asymptotic solutions of differential equations with transition points 92
Hayman, W. K., The coefficients of schlicht and allied functions. 102
Hille, E., Some aspect of Cauchy's problem 109
Kodaira, K., Some results in the transcendental theory of algebraic varieties 117, 474
Myrberg, P. J., Über automorphe Funktionen 118
Pauc, Chr., Contributions à la théorie de la dérivation de fonctions d'ensemble . . 127
Ważewski, T., Sur une méthode topologique de l'examen de l'allure asymptotique des intégrales des équations différentielles 132
Zygmund, A., Hilbert Transforms in E^n 140

SECTION III: GEOMETRY AND TOPOLOGY

Stated addresses

Coxeter, H. S. M., Regular honeycombs in hyperbolic space 155
Eckmann, B., Zur Cohomologietheorie von Räumen und Gruppen 170
Freudenthal, H., La topologie dans les fondements de la géometrie. 178
Montgomery, D., Topological transformation groups 185
Serre, J. P., Cohomologie et géométrie algébrique 189, 515
Yano, K., On pseudo-Hermitian and pseudo-Kählerian manifolds 190

SECTION IV: PROBABILITY AND STATISTICS

Stated addresses

Blackwell, D., Controlled random walks 201, 336
Doob, J. L., Interrelations between Brownian motion and potential theory . . . 202
Fortet, R., Lois des grands nombres pour des éléments aléatoires généraux. 205, 360

SECTION V: MATHEMATICAL PHYSICS AND APPLIED MATHEMATICS

Stated addresses

Collatz, L., Fehlermaszprinzipien in der praktischen Analysis. 209
Fichera, G., Methods of functional linear analysis in mathematical physics . . . 216
Hestenes, M. R., Hilbert space methods in variational theory and numerical analysis 229
Kampé de Fériet, J., Problèmes mathématiques posés par la mécanique statistique de la turbulence 237

Rellich, F., Halbbeschränkte Differentialoperatoren höherer Ordnung 243
Stoker, J. J., Some recent progress in the theory of surface waves in water . . 251
Weinstein, A., Elliptic and hyperbolic axially symmetric problems 264

SECTION VI: LOGIC AND FOUNDATIONS

Stated addresses

Lorenzen, P., Die Fiktion der Überabzählbarkeit 273
Mostowski, A., Development and applications of the „projective" classification of sets of integers. 280
Barkley Rosser, J., The relative strength of Zermelo's set theory and Quine's new foundations 289

SECTION VII: PHILOSOPHY, HISTORY AND EDUCATION

Stated addresses

Daltry, C. T., Self-education by children in mathematics using Gestalt methods — i.e. learning-through-insight 297
Kurepa, G., The role of mathematics and mathematician at present time. . . . 305
Piene, K., School mathematics for universities and for life. 318

SYMPOSIUM A: STOCHASTIC PROCESSES

Dantzig, D. van, Opening speech 329
Fréchet, R. M., Allocution 332
Blackwell, D., Controlled random walks 201, 336
Blanc-Lapierre, A., Sur quelques applications de la théorie des fonctions aléatoires à l'optique 339
Doob, J. L., Present state and future prospects of stochastic process theory . . . 348
Fortet, R. M., Les fonctions aléatoires comme éléments aléatoires dans un espace de Banach 356
Fortet, R. M., Lois des grands nombres pour des éléments aléatoires généraux 205, 360
Hille, E., Perturbation methods in the study of Kolmogoroff's equations 365
Kendall, D. G., and Reuter, G. E. H., Some pathological Markov processes with a denumerable infinity of states and the associated semigroups of operators on l, 377
Lévy, P., Processus semi-markoviens. 416
Lindley, D. V. The estimation of velocity distributions from counts 427

SYMPOSIUM B: ALGEBRAIC GEOMETRY

Gröbner, W., Über die idealtheoretische Grundlegung der algebraischen Geometrie 447
Hirzebruch, F., Der Satz von Riemann-Roch in Faisceau-theoretischer Formulierung; einige Anwendungen und offene Fragen 457
Kodaira, K., Some results in the transcendental theory of algebraic varieties 117, 474
Néron, A., Propriété sarithmétiques de certaines familles de courbes algébriques 40, 481
Northcott, D. G., Specialization methods in algebraic geometry. 41, 489
Rosenlicht, M., Group varieties and differential forms 493
Segre, B., Geometry upon an algebraic variety 497
Semple, J. G., The triangle as a geometric variable 514
Serre, J. P., Cohomologie et géométrie algébrique 189, 515
Severi, F., Fonctions et variétés quasi-abéliennes 521
Severi, F., Problèmes résolus et problèmes nouveaux dans la théorie des systèmes d'équivalence. 529
Waerden, B. L. van der, The invariant theory of linear sets of an algebraic variety 542
Waerden, B. L. van der, On the definition of rational equivalence of cycles on a variety 545
Weil, A., Abstract versus classical algebraic geometry 550