aequationes mathematicae

Volume 2, 1969

University of Waterloo, Ontario, Canada
Birkhauser Verlag, Basel, Switzerland

aequationes mathematicae

Annual subscription:
(3 fascicles per volume)
sFr. / DM. 40.—
US–$ 9.50. Single copy:
sFr. / DM 20.—
US–$ 4.90

aequationes mathematicae is published by the University of Waterloo and Birkhäuser Verlag. Price per volume (three issues) is 40 Swiss francs including postage and costs of mailing, payable at the office of the publisher, Birkhäuser Verlag in Basel, Switzerland or through your bookseller.

The journal will publish papers in pure and applied mathematics in general but, in particular, papers in functional equations, combinatorial and numerical analysis. The languages in which papers will generally be accepted are the international congress languages: English, French, German, Italian, and Russian.

Types of manuscripts that will be considered for publication include original research papers, reports of meetings and activities, problems and their solutions, expository papers and, as short communications, self-contained abstracts of papers which have been accepted for publication either by this or some other journal with referee system.

Manuscripts of papers should be submitted to the Editorial Office or to a member of the editorial board. Authors of papers should send the original and, preferably, two additional copies. Since it is the policy of the journal to publish in advance short communications of all results it accepts for publication, authors should include with their original submissions a self-explanatory abstract of their paper of length 25–100 lines.

PUBLISHER: Birkhäuser Verlag, 4000 Basel 10, Switzerland
EDITORIAL OFFICE: University of Waterloo, Faculty of Mathematics, Waterloo, Ontario, Canada

Contents Vol. 2 · 1969

Contents Vol. 2 · 1969

aequationes mathematicae

Vol. 2
No. 1
Pag. 1–136

Dedicated to
the 75th Birthday of
Alexander M. Ostrowski

University of Waterloo, Ontario, Canada
Birkhäuser Verlag, Basel, Switzerland

aequationes mathematicae

Annual subscription:
(3 fascicles per volume)
sFr. / DM 40.—
US–$ 9.50. Single copy:
sFr. / DM 20.—
US–$ 4.90

aequationes mathematicae is published by the University of Waterloo and Birkhäuser Verlag. Price per volume (three issues) is 40 Swiss francs including postage and costs of mailing, payable at the office of the publisher, Birkhäuser Verlag in Basel, Switzerland or through your bookseller.

The journal will publish papers in pure and applied mathematics in general but, in particular, papers in functional equations, combinatorial and numerical analysis. The languages in which papers will generally be accepted are the international congress languages: English, French, German, Italian, and Russian.

Types of manuscripts that will be considered for publication include original research papers, reports of meetings and activities, problems and their solutions, expository papers and short abstracts of papers which have been accepted for publication either by this or some other journal with referee system.

Manuscripts of papers should be submitted to the Editorial Office or to a member of the editorial board. Authors of papers should send the original and, preferably, two additional copies. Since it is the policy of the journal to publish in advance short communications of all results it accepts for publication, authors should include with their original submissions a self-explanatory abstract of their paper of length 25–100 lines.

PUBLISHER: Birkhäuser Verlag, 4000 Basel 10, Switzerland
EDITORIAL OFFICE: University of Waterloo, Faculty of Mathematics, Waterloo, Ontario, Canada

To A. M. Ostrowski

ALEXANDER M. OSTROWSKI

'To A. M. Ostrowski'

On the occasion of your 75th Birthday, mathematicians specialized in many directions, dedicate to you the papers collected in the present volume, as a symbol of their high regard for your contributions to mathematics and as a token of affection for you personally. Among these, you will particularly welcome the tributes from some of your distinguished pupils and collaborators.

It is proper at this time to refresh our memories of some of your outstanding work, but its breadth and depth make only a superficial sampling possible. However, we hope that this survey will encourage experts in some of the fields in which you have worked, to get acquainted with your accomplishments in others.

Your first published result caused general excitement in the mathematical community. HILBERT in his famous lecture to the International Congress of Mathematicians in Paris in 1900 stated a series of outstanding problems one of which was to show that

$$\zeta(z, s) = \sum n^{-s} z^n$$

does not satisfy any algebraic partial differential equation. This, and more, you solved in your thesis, by completely elementary methods.

In the theory of analytic functions, the subject of overconvergence, until your intervention, was merely a series of examples; after your work, the situation was completely clarified. Elsewhere in this domain, your fundamental contributions to the theory of the boundary correspondence and boundary distortion in conformal mapping, your investigations on the harmonic measure and harmonic majorants, uniqueness and factorization of analytic functions in the unit disk, your profound work on quasi-analytic functions, your researches on PICARD's and SCHOTTKY's theorems and on the Riemann ζ-function – all have a definitive place in the literature. Your introduction of the spherical metric in the complex plane, which you first used in your characterization of normal families, has been taken over into standard modern treatises on complex function theory.

In the field of algebra, we recall your essential advances in the theory of valuations which, before your research, was in a tentative state. Not only did you determine all possible valuations of the rational numbers, but you laid the foundations for the local treatment of many algebraic and arithmetic questions which are so topical at this time. Here we must also mention your work on the fundamental theorem of algebra, on Galois theory and on the theory of invariants.

You have enriched many parts of the theory of numbers, in particular, we note your work on WARING's problem, on diophantine approximation and on various problems in algebraic number theory.

In the theory of functions of real variables, we remember, as of particular interest to our 'Aequationes Mathematicae', your study of convex functions and of the Cauchy functional equation which has not been essentially improved during the last 40 years. Further, we call attention to your work on the topology of oriented line elements and your more recent research on positive operators.

Your longtime interest in computation – it was you who introduced the concept of a Horner – was intensified, to the benefit of all of us, by the advent of automatic computors. You have contributed to many areas of numerical analysis and your presence at various centers in the U.S.A. and elsewhere has had a profound influence on its development. You never hesitated to plan and carry out computational experiments, in conformal mapping and other areas. Your skill and scholarship in linear algebra has led to much progress, e.g., in the theory of norms and matrix inequalities, the analysis of methods of the solution of linear systems, the approximation to characteristic values (in particular, problems associated with the names of GERSCHGORIN and LYAPUNOV).

You have enriched the classical parts of the theory of equations and modernized much of it, including your recent designs of root finders. As early as 1937, you were deeply concerned with numerical stability and your papers on error estimates are models for the serious numerical analyst. No one who was not trained in the analytic theory of numbers could have completed – as you did, in a series of five papers – an investigation of one of the standard processes for determining characteristic values of matrices.

Apart from your major research papers and monographs, you have enriched our literature with your series 'Mathematische Miszellen' – these are not what your modest title implies; they are genuine masterpieces of mathematical ingenuity and competence which we read with as much pleasure as profit.

You have found time to issue your famous lectures and exercises on Differential and Integral Calculus so that those not fortunate enough to be able to attend your classes can benefit. English translations of these will soon further widen your influence on mathematical education.

Your Washington lectures on 'Solution of Equations and Systems of Equations' and your later researches in this basic field of numerical analysis have also been made available to a wider public in book form.

To this sketch of an impressive demonstration of mathematical versatility we would like to add our most cordial good wishes. In particular we hope that your wonderful creative powers will long remain with you in strength so that you can continue to enrich mathematics with beautiful new results just as you have been doing uninterruptedly from 1913 to the present.

THE EDITORS

Bibliographies

Bibliography of the Works of A. M. Ostrowski

1. *Zur Algebra der endlichen Felder* (Russian), S.-B. Phys. Math. Ges. Kiew *1913*, 1–37.
2. *Über einige Fragen der allgemeinen Körpertheorie*, J. Reine Angew Math. *143*, 255–284 (1913).
3. *Über die Existenz einer endlichen Basis bei gewissen Funktionensystemen*, Math. Ann. *78*, 94–119 (1918).
4. *Über einige Lösungen der Funktionalgleichung* $\phi(x)\cdot\phi(y) = \phi(xy)$, Acta Math. *41*, 271–284 (1918).
5. *Über sogenannte perfekte Körper*, J. Reine Angew. Math. *47*, 191–214 (1917).
6. *Neuer Beweis des Hölderschen Satzes, dass die Gammafunktion keiner algebraischen Differentialgleichung genügt*, Math. Ann. *79*, 286–288 (1918).
7. *Über eine neue Eigenschaft der Diskriminanten und Resultanten binärer Formen*, Math. Ann. *79*, 360–387 (1918).
8. *Über ein Analogon der Wronskischen Determinante bein Funktionen mehrerer Veränderlicher*, Math. Z. *4*, 223–230 (1919).
9. *Beweis der Irreduzibilität der Diskriminante einer algebraischen Form von mehreren Variablen*, Math. Z. *4*, 314–319 (1919).
10. *Über ganzwertige Polynome in algebraischen Zahlkörpern*, J. Reine Angew. Math. *149*, 117–124 (1919).
11. *Zur arithmetischen Theorie der algebraischen Grössen*, Nachr. Akad. Wiss. Göttingen Math-Phys. Kl. II *1919*, 279–298.
12. *Über die Existenz einer endlichen Basis bei Systemen von Potenzprodukten*, Math. Ann. *81*, 21–24 (1920).
13. *Zum ersten und vierten Gaussschen Beweis des Fundamentalsatzes der Algebra*, Nachr. Akad. Wiss. Göttingen Math.-Phys. Kl. II *1920*, Beiheft 50–58.
14. *Über die Reihe* $\sum_{n=0}^{\infty} q^{n^2} x^n$, Math. Ann. *82*, 64–67 (1921).
15. *Über Dirichletsche Reihen und algebraische Differentialgleichungen* (Inaugural-Dissertation), Math. Z. *8*, 241–298 (1920).
16. *Bemerkung zur Hardy-Littlewoodschen Lösung des Waringschen Problems*, Math. Z. *9*, 28–34 (1921).
17. (Mit E. Landau) *On the Diophantine Equation* $ay^2 + by + c = dx^n$, Proc. London Math. Soc. [2] *19*, 276–280 (1921).
18. *Über eine Eigenschaft gewisser Potenzreihen mit unendlich vielen verschwindenden Koeffizienten*, S.-B. Berlin Akad. Math.-Phys. Kl. *1921*, 557–565.
19. *Bemerkungen zur Theorie der Diophantischen Approximationen (I–III)*, Abh. Math. Sem. Univ. Hamburg *1*, 77–98 (1922).
20. *Auszug aus dem Brief von A. Ostrowski an L. Bieberbach*, Jber. Deutsch. Math.-Verein. *31*, 82–85 (1922).
21. *Notiz über einen Satz der Galoisschen Theorie*, Math. Z. *12*, 317–322 (1922).
22. *Über vollständige Gebiete gleichmässiger Konvergenz von Folgen analytischer Funktionen*, Abh. Math. Sem. Univ. Hamburg *1*, 327–350 (1922).
23. *Über ein algebraisches Übertragungsprinzip* (Habilitationsschrift), Abh. Math. Sem. Univ. Hamburg *1*, 281–326 (1922).
24. (Mit J. Schur) *Über eine fundamentale Eigenschaft der Invarianten einer binären Form*, Math. Z. *15*, 81–105 (1922).
25. *Über Singularitäten Taylorscher und Dirichletscher Reihen*, S.-B. Berlin Akad. Math.-Phys. Kl. *1923*, 39–44.

26. *Über die Bedeutung der Jensenschen Formel für die komplexe Funktionentheorie*, Acta Sci. Math. Szeged. *1*, 80–87 (1923).
27. Besprechung von H. Beck, *Koordinatengeometrie*, Gött. Gel. Anz. *1922*, 306–311.
28. *Über allgemeine Konvergenzsätze der komplexen Funktionentheorie*, Jber. Deutsch. Math.-Verein. *32*, 185–194 (1923).
29. *Über Potenzreihen, die überkonvergente Abschnittsfolgen besitzen*, S.-B. Berlin Akad. Math.-Phys. Kl. *1923*, 185–192.
30. *Über die Darstellung analytischer Funktionen durch Potenzreihen*, Jber. Deutsch. Math.-Verein. *32*, 286–295 (1923).
31. *Über eine neue Fragestellung in der algebraischen Invariantentheorie* (Habilitationsvortrag), Jber. Deutsch. Math.-Verein. *33*, 174–184 (1924).
32. *Mathematische Miszellen*, I: *Die Maxwellsche Erzeugung der Kugelfunktionen*, Jber. Deutsch. Math.-Verein. *33*, 245–251 (1924).
33. *Mathematische Miszellen*, II: *Über eine Reduktion der Integrabilitätsbedingungen für vollständige Differentiale*, Jber. Deutsch. Math.-Verein. *34*, 79–80 (1925).
34. *Zum Hölderschen Satz über $\Gamma(x)$*, Math. Ann. *94*, 248–251 (1925).
35. *Über Folgen analytischer Funktionen und einige Verschärfungen des Picardschen Satzes*, Math. Z. *24*, 215–258 (1926).
36. *Mathematische Miszellen*, III: *Über Nullstellen gewisser im Einheitskreise regulärer Funktionen und einige Sätze zur Konvergenz unendlicher Reihen*, Jber. Deutsch. Math.-Verein. *34*, 161–171 (1925).
37. *Über den Schottkyschen Satz und die Borelschen Ungleichungen*, S.-B. Berlin Akad. Math.-Phys. Kl. *1925*, 471–484.
38. *Mathematische Miszellen*, IV: *Bemerkung über Differenzierbarkeit unendlicher Funktionenfolgen*, Jber. Deutsch. Math.-Verein. *34*, 237–239 (1925).
39. *Mathematische Miszellen*, V: *Über eine Erweiterung des Irreduzibilitätsbegriffs und ihre Anwendung auf eine funktionentheoretisches Problem*, Jber. Deutsch. Math.-Verein. *35*, 91–96 (1926).
40. *On Representation of Analytical Functions by Power Series*, J. London Math. Soc. *1* (4), 251–263 (1926); Addendum *4* (1) (1929).
41. *On Hadamard's Test for Singular Points*, J. London Math. Soc. *1* (4), 236–239 (1926).
42. *Mathematische Miszellen*, VI: *Über einen Satz von Herrn Hadamard*, Jber. Deutsch. Math.-Verein. *35*, 179–182 (1926).
43. *Mathematische Miszellen*, VII: *Über Singularitäten gewisser mit Lücken behafteten Potenzreihen*, Jber. Deutsch. Math.-Verein. *35*, 269–280 (1926).
44. *Mathematische Miszellen*, VIII: *Funktionaldeterminanten und Abhängigkeit von Funktionen*, Jber. Deutsch. Math.-Verein. *36*, 129–134 (1927).
45. *Mathematische Miszellen*, IX: *Notiz zur Theorie der Diophantischen Approximationen*, Jber. Deutsch. Math.-Verein. *36*, 178–180 (1927).
46. *Mathematische Miszellen*, X: *Über das Poissonsche Integral und fast stetige Funktionen*, Jber. Deutsch. Math.-Verein. *36*, 349–353 (1927).
47. *Mathematische Miszellen*, XI: *Über den Lerchschen Satz*, Jber. Deutsch. Math.-Verein. *37*, 69–70 (1928).
48. *Mathematische Miszellen*, XII: *Bemerkungen zum Beweis des Budan-Fourierschen Satzes und Newton-Sylvesterschen Satzes*, Jber. Deutsch. Math.-Verein. *37*, 254–256 (1928).
49. *Mathematische Miszellen*, XIII: *Über Abhängigkeit linearer Systeme und Integrabilitätsbedingungen für Systeme linearer Differentialgleichungen in mehreren Variabeln*, Jber. Deutsch. Math.-Verein. *37*, 265–274 (1928).
50. J. A. Schouten, *Der Ricci-Kalkül*, Gött. Gel. Anz. *12*, 595–599 (1928).
51. *Über das Hadamardsche Singularitätenkriterium in der Theorie der Taylorschen und Dirichletschen Reihen*, S.-B. Berlin Math. Ges. *27*, 32–47 (1928).
52. *Mathematische Miszellen*, XIV: *Über die Funktionalgleichung der Exponentialfunktion und verwandte Funktionalgleichungen*, Jber. Deutsch. Math.-Verein. *38*, 54–62 (1929).

53. *Zur Theorie der konvexen Funktionen*, Comment. Math. Helv. *1*, 157–159 (1929).
54. *Mathematische Miszellen*, XV: *Zur konformen Abbildung einfach zusammenhängender Gebiete*, Jber. Deutsch. Math.-Verein. *38*, 168–182 (1929).
55. *Über Schwankungen analytischer Funktionen, die gegebene Werte nicht annehmen*, S.-B. Preuss. Akad. Wiss. Math.-Phys. Kl. *21*, 276–289 (1929).
56. *Sur quelques généralisations du produit d'Euler* $\prod_{\nu=0}^{\infty}(1+x^{2^\nu}) = \frac{1}{1-x}$, C. R. Acad. Sci. Paris *190*, 249–251 (1930).
57. *Über einige Verallgemeinerungen des Eulerschen Produktes* $\prod_{\nu=0}^{\infty}(1+x^{2^\nu}) = \frac{1}{1-x}$, Verh. Naturforsch. Ges. Basel *11*, 153–214 (1929).
58. *Zur Theorie der Überkonvergenz*, Math. Ann. *103*, 15–27 (1930).
59. *Mathematische Miszellen*, XVI: *Zur Theorie der linearen Diophantischen Approximationen*. Jber. Deutsch. Math.-Verein. *39*, 34–46 (1930).
60. *Über konforme Abbildung annährend kreisförmiger Gebiete*, Jber. Deutsch. Math.-Verein *39*, 78–81 (1930).
61. *Über quasianalytische Funktionen und Bestimmtheit asymptotischer Entwicklungen*, Acta Math. *53*, 181–266 (1929).
62. *Studien über den Schottkyschen Satz*, Rektoratsprogramm der Universität Basel für das Jahr 1931.
63. *Ganzzahligkeit in der Mathematik*, Tätigkeitsbericht der Math. Fachschaft an der Universität Heidelberg *1932*, 1–10.
64. *Asymptotische Abschätzungen der Argumentvariation einer Funktion, die die Werte 0 und 1 nicht annimmt*, Math. Z. *36*, 303–320 (1932).
65. *Asymptotische Abschätzung des absoluten Betrages einer Funktion, die die Werte 0 und 1 nicht annimmt*, Comment. Math. Helv. *5*, 55–87 (1933).
66. *Über den ersten und vierten Gauss'schen Beweis des Fundamental Satzes der Algebra*, in *C. F. Gauss Werke*, X (Springer-Verlag, Berlin 1933).
67. *Mathematische Miszellen*, XVII: *Notiz zu einem Satz von H. Bohr*, Jber. Deutsch. Math.-Verein. *42*, 161–165 (1932).
68. (Mit Th. Motzkin) *Über den Fundamentalsatz der Algebra*, S.-B. Berlin Akad. Math.-Phys. Kl. *1933*, 255–258.
69. *Über algebraische Funktionen von Dirichletschen Reihen*, Math. Z. *37*, 98–133 (1933).
70. *Über einen Satz von Leau und die analytische Fortsetzung einiger Klassen Taylorscher und Dirichletscher Reihen*, Math. Ann. *108*, 718–756 (1933).
71. *Mathematische Miszellen*, XVIII: *Notiz über den Wertvorrat der Riemannschen ζ-Funktion am Rande des kritischen Streifens*, Jber. Deutsch. Math.-Verein. *43*, 58–64 (1933).
72. *Über Nullstellen stetiger Funktionen zweier Variabeln*, J. Reine Angew. Math. *170*, 83–94 (1934).
73. *Sur les multiplicités des zéros des fonctions indéfiniment dérivables de deux variables*, Bull. Soc. Math. France *58* (2), 64–72 (1934).
73a. *Addition à la note: Sur les multiplicités des zéros des fonctions indéfiniment dérivables de deux variables*, Bull. Soc. Math. France *59* (2), 259–260 (1935).
74. *Bemerkung zur vorstehenden Note von Herrn S. Warschawski*, Math. Z. *38*, 684–686 (1934).
75. *Berührungsmasse, nullwinklige Kreisbogendreiecke und die Modulfigur*, Jber. Deutsch. Math.-Verein. *44*, 56–75 (1934).
76. *Untersuchungen zur arithmetischen Theorie der Körper (Die Theorie der Teilbarkeit in allgemeinen Körpern)*, Math. Z. *39*, 269–404 (1935).
77. *Über den Habitus der konformen Abbildung am Rande des Abbildungsbereiches*, Acta Math. *64*, 81–184 (1935).
78. *Bemerkungen über die Struktur von Ringen, die aus Polynomen in einer Variabel bestehen*, Acta Arith. *1*, 19–42 (1936).
79. *Beiträge zur Topologie der orientierten Linienelemente*, I, Compositio Math. *2*, 1–24 (1935).

80. *Beiträge zur Topologie der orientierten Linienelemente*, II–III, Compositio Math. *2*, 177–200 (1935).
81. *Sur la conversation des angles dans la transformation conforme d'un domaine au voisinage d'un point frontière*, C. R. Acad. Sci. Paris *202*, 727 (1936).
82. *Sur la transformation des plis dans la transformation conforme au voisinage d'un point frontière*, C. R. Acad. Sci. Paris *202*, 1135 (1936).
83. *Zur Randverzerrung bei konformer Abbildung* (Dem Gedächtnis von Leon Lichtenstein gewidmet), Prace Mat.-Fiz. *44*, 371–471 (1936).
84. *Sur une transformation de la série de Liouville-Neumann*, C. R. Acad. Sci. Paris *203*, 602 (1936).
85. *Konvergenzdiskussion und Fehlerabschätzung für die Newton'sche Methode bei Gleichungssystemen*, Comment. Math. Helv. *9*, 79–103 (1936–1937).
86. *Sur la détermination des bornes inférieures pour une classe des déterminants*, Bull. Soc. Math. France [2] *61*, 1–14 (1937).
87. *Über die Determinanten mit überwiegender Hauptdiagonale*, Comment. Math. Helv. *10*, 69–96 (1937).
88. *Über eine Modifikation des Newtonschen Näherungsverfahrens*, Inst. Math. Tbilissi *2*, 241–250 (1937).
89. *Sur quelques transformations de la série de Liouville-Neumann*, C. R. Acad. Sci. Paris *206*, 1345–1347 (1938).
90. *Über die Absolutabweichung einer differenzierbaren Funktion von ihrem Integralmittelwert*, Comment. Math. Helv. *10*, 226–227 (1938).
91. *Notiz über die Funktionaldeterminante von zwei Funktionen mit zwei gemeinsamen Nullstellen*, Comment. Math. Helv. *10*, 228–231 (1938).
92. *Über die Konvergenz und die Abrundungsfestigkeit des Newtonschen Verfahrens*, Rec. Math. Moscou *2* (44), 1073–1095 (1937).
93. *Über einen Fall der Konvergenz des Newtonschen Näherungsverfahrens*, Rec. Math. Moscou *3* (45), 253–258 (1938).
94. *Sur les modules des zéros des fonctions entières*, C. R. Acad. Sci. Paris *206*, 1541–1543 (1938).
95. *Sur l'approximation du déterminant de Fredholm par les déterminants des systèmes d'équations linéaires*, Ark. Mat. *26A*, 1–15 (1938).
96. *On a Theorem Concerning Identical Relations Between Matrices*, Quart. J. Math. Oxford Ser. *9*, 241–245 (1938).
97. *Sur la continuité relative des racines d'équations algébriques*, C. R. Acad. Sci. Paris *209*, 777–779 (1939).
98. *Note sur les produits des séries normales*, Bull. Soc. Roy. Sci. Liège *8*, 458–468 (1939).
99. *Sur la convergence et l'estimation des erreurs dans quelques procédés de résolution des équations numériques*, in *Grave-Festschrift* (Gosudarstv. Izdat. Tehn.-Teor. Lit., Moscow 1940), pp. 213–234.
100. *Recherches sur la méthode de Graeffe et les zéros des polynomes et des séries de Laurent*, Acta Math. *72*, 99–257 (1940).
101. *Mathematische Miszellen*, XIX, XX, XXI: *Zur integrallosen Bestimmung der Berührungstransformationen. Über eine Klasse von kanonischen Transformationen*, Verh. Naturforsch. Ges. Basel *52*, 35–48 (1941).
102. *Sur une classe de transformations différentielles dans l'espace à trois dimensions*, I, Comment. Math. Helv. *13*, 156–194 (1941).
103. *Sur une classe de transformations différentielles dans l'espace à trois dimensions*, II, Comment. Math. Helv. *14*, 23–60 (1942).
104. *On a Theorem by J. L. Walsh Concerning the Moduli of Roots of Algebraic Equations*, Bull. Amer. Math. Soc. *47*, 742–746 (1941).
105. *On the Definition of Contact Transformations*, Bull. Amer. Math. Soc. *47*, 760–763 (1941).
106. *Sur un théorème fondamental de la théorie des équations linéaires aux dérivées partielles*, Comment. Math. Helv. *15*, 217–221 (1943).
107. *Sur les transformations réversibles d'éléments de ligne*, Acta Math. *75*, 151–182 (1942).

108. *Note sur l'interversion des dérivations et les différentielles totales*, Comment. Math. Helv. *15*, 222–226 (1943).
109. *Sur les conditions de validité d'une classe de relations entre les expressions différentielles linéaires*, Comment. Math. Helv. *15*, 265–286 (1943).
110a. *Über algebraische Relationen zwischen unbestimmten Integralen*, Experientia *1*, 117–118 (1945).
110b. *Ein Unabhängigkeitssatz für irreduzible Integrale*, Experientia *1*, 195 (1945).
111. Besprechung von BAVINK: *Ergebnisse und Probleme der Naturwissenschaften*, Experientia *1*, 164–165 (1945).
112. *Sur les relations algébriques entre les intégrales indéfinies*, Acta Math. *78*, 315–318 (1946).
113. *Sur l'intégrabilité élémentaire de quelques classes d'expressions*, Comment. Math. Helv. *18*, 283–308 (1946).
114. *Sur l'inverse d'une transformation continue et biunivoque*, C. R. Acad. Sci. Paris *223*, 229–230 (1946).
115. *Nouvelle démonstration du théorème de Schoenflies pour les espaces à n-dimensions*, C. R. Acad. Sci. Paris *223*, 530–531 (1946).
116. *Sur la formule de Moivre-Laplace*, C. R. Acad. Sci. Paris *223*, 1090–1092 (1946).
117. *Sur le rayon de convergence de la série de Blasius*, C. R. Acad. Sci. Paris *227*, 580–582 (1948).
117a. *G. H. Hardy, Nachruf*, Experientia 5 (1949).
118. *On Some Generalizations of the Cauchy-Frullani Integral*, Proc. Nat. Acad. Sci. U.S.A. *35*, 612–616 (1949).
119. *Sur la variation de la matrice inverse d'une matrice donnée,* C. R. Acad. Sci. Paris *231*, 1019–1021 (1950).
120. *Un théorème d'existence pour les systèmes d'équations,* C. R. Acad. Sci. Paris *231*, 1114–1116 (1950).
121. *Note on Vincent's Theorem*, Ann. of Math. [2] *52*, 702–707 (1950).
122. *Generalization of a Theorem of Osgood to the Case of Continuous Approximation*, Proc. Amer. Math. Soc. *1*, 648–649 (1950).
123. *Un nouveau théorème d'existence pour les systèmes d'équations*, C. R. Acad. Sci. Paris *232*, 786–788 (1951).
124. *Sur une règle de Laguerre*, Ann. Mat. Pura Appl. [4] *31*, 65–68 (1950).
125. *Note on an Infinite Integral*, Duke Math. J. *18*, 355–359 (1951).
126. *La recherche des périodicités cachées. Analyse Harmonique* (Centre Nat. Rech. Sci., Paris 1949 [Colloques Internat. Centre Nat. Rech. Sci., No. 15]), pp. 93–95.
127. *Über das Nichtverschwinden einer Klasse von Determinanten und die Lokalisierung der charakteristischen Wurzeln von Matrizen*, Compositio Math. *9*, 209–226 (1951).
128. *Sur les conditions générales pour la régularité des matrices*, Rend. Mat. e Appl. (5) *10*, 156–168 (1951).
129. (With OLGA TAUSSKY) *On the Variation of the Determinant of a Positive Definite Matrix*, Nederl. Akad. Wetensch. Proc. [Ser. A. 54] = Indagationes Math. *13*, 383–385 (1951).
130. *Sur les matrices peu différentes d'une matrice triangulaire*, C. R. Acad. Sci. Paris *233*, 1558–1560 (1951).
131. *Note on Bounds for Determinants with Dominant Principal Diagonal*, Proc. Amer. Math. Soc. *3*, 26–30 (1952).
132. *Bounds for the Greatest Latent Root of a Positive Matrix*, J. London Math. Soc. *27*, 253–256 (1952).
133. *Sur quelques applications des fonctions convexes et concaves au sens de I. Schur*, J. Math. Pures Appl. (9) *31*, 253–292 (1952).
134. *On the Rounding off of Difference Tables for Linear Interpolation*, Math. Tables and Other Aids to Comput. *6*, 212–214 (1952).
135. *On the Convergence of Theodorsen's and Garrick's Method of Conformal Mapping*, in *Proc. Sympos. on Conformal Maps, 1949* (Nat. Bur. Standards, Washington, D.C. 1952 [Appl. Math. Ser., No. 18]), pp. 149–163.
136. *On a Discontinuous Analogue of Theodorsen's and Garrick's Method*, in *Proc. Sympos. on*

Conformal Maps, 1949 (Nat. Bur. Standards, Washington, D.C. 1952 [Appl. Math. Ser., No. 18]), pp. 165–174.

137. *Two Explicit Formulae for the Distribution Function of the Sums of n Uniformly Distributed Independent Variables*, Arch. Math. *3*, 451–459 (1952).
138. *On Over and Under Relaxation in the Theory of the Cyclic Single Step Iteration*, Math. Tables and Other Aids to Comput. *7*, 152–159 (1953).
139. *Simultaneous Systems of Equations*, in *Simultaneous Linear Equations and the Determination of Eigenvalues* (Nat. Bur. Standards, Washington, D.C. 1953 [Appl. Math. Ser., No. 29]), pp. 29–34.
140. *On Nearly Triangular Matrices*, J. Research Nat. Bur. Standards *52*, 319–345 (1954).
141. *Mathematische Miszellen*, XXII: *Über gewisse Ungleichungen zwischen monotonen Zahlenfolgen*, Jber. Deutsch. Math.-Verein *57*, 85–89 (1954).
142. *On Two Problems in Abstract Algebra Connected with Horner's Rule*, in *Studies in Mathematics and Mechanics Presented to Richard von Mises* (Academic Press, New York 1954), pp. 40–48.
143. *On the Linear Iteration Procedures for Symmetric Matrices*, Rend. Mat. e Appl. (5) *14*, 140–163 (1954).
144. *On the Spectrum of a One-Parametric Family of Matrices*, J. Reine Angew. Math. *193*, 143–160 (1954).
145. *Über die analytische Fortsetzung von Taylorschen und Dirichletschen Reihen*, Math. Ann. *129*, 1–43 (1955).
146. *Conformal Mapping of a Special Ellipse on the Unit Circle*, in *Experiments in the Computation of Conformal Maps* (Nat. Bur. Standards, Washington, D.C. 1955 [Appl. Math. Ser., No. 42]), pp. 1–2.
147. *Theodorsen's and Garrick's Method for Conformal Mapping of the Unit Circle into an Ellipse*, in *Experiments in the Computation of Conformal Maps* (Nat. Bur. Standards, Washington, D.C. 1955 [Appl. Math. Ser., No. 42]), pp. 3–5.
148. *Note on Bounds for Some Determinants*, Duke Math. J. *22*, 95–102 (1955).
149. *Über Evoluten und Evolventen ebener Kurven*, Arch. Math. *6*, 170–179 (1955).
150. *Un'applicazione dell'integrale di Stieltjes alla teoria elementare delle curve piane*, Atti. Accad. Naz. Lincei. Rend. Cl. Sci. Fis. Mat. Nat. (8) *18*, 373–375 (1955).
151. *Note on a Logarithm Algorithm*, Mathematical Tables and Other Aids to Comput. *9*, 65–68 (1955).
152. *Note sur les dérivées uniformes et les différentielles totales*, Comment. Math. Helv. *29*, 298–300 (1955).
153. *Über Normen von Matrizen*, Math. Z. *63*, 2–18 (1955).
154. *Note on a Theorem by A. Brauer*, Duke Math. J. *22*, 469–470 (1955).
155. *Sur les déterminants à diagonale dominante*, Bull. Soc. Math. Belg. *1954*, 46–51.
156. *Mathematische Miszellen*, XXIII: *Zur Differentialgeometrie der ebenen Parallelkurven*, Jber. Deutsch. Math.-Verein. *58*, 49–53 (1955).
157. *Determinanten mit überwiegender Hauptdiagonale und die absolute Konvergenz von linearen Iterationsprozessen*, Comment. Math. Helv. *30*, 175–210 (1955).
158. *Sur les critères de convergence et divergence dus à V. Ermakof* (à Trygve Nagell à l'occasion de son 60e anniversaire), Enseignement Math. (2) *1*, 224–257 (1956).
159. *Mathematische Miszellen*, XXIV: *Zur relativen Stetigkeit von Wurzeln algebraischer Gleichungen*, Jber. Deutsch. Math.-Verein *58*, 98–102 (1956).
160. *Über Verfahren von Steffensen und Householder zur Konvergenzverbesserung von Iterationen* (Mauro Picone zum 70. Geburtstag), Z. Angew. Math. Phys. *7*, 218–229 (1956).
161. *Über die Differenzierbarkeit von impliziten Funktionen*, Verh. Naturforsch. Ges. Basel *67*, 141–148 (1956).
162. *Zur Theorie der partiellen Differentialgleichungen erster Ordnung*, Math. Z. *66*, 70–87 (1956).
163. *Mathematische Miszellen*, XXV: *Über das Verhalten von Iterationsfolgen im Divergenzfall*, Jber. Deutsch. Math.-Verein. 59, 69–79 (1956).
164. *Über die Darstellung von symmetrischen Funktionen durch Potenzsummen*, Math. Ann. 132, 362–372 (1956).

165. *Über die Verbindbarkeit von Linien- und Krümmungselementen durch monoton gekrümmte Kurvenbogen*, Enseignement Math. [2] *2* (1956), 277–292.
166. *Le dévelopment de Taylor de la fonction inverse*, C. R. Acad. Sci. Paris *244*, 429–439 (1957).
167. *Les points d'attraction et de répulsion pour l'itération dans l'espace à n-dimensions*, C. R. Acad. Sci. Paris *244*, 288–289 (1957).
168. *Eine Verschärfung des Schubfächerprinzips in einem linearen Intervall*, Arch. Math. *8*, 1–10 (1957).
169. *Über näherungsweise Auflösung von Systemen homogener linearer Gleichungen*, Z. Angew. Math. Phys. *8*, 280–285 (1957).
170. *Mathematische Miszellen*, XXVI: *Zum Schubfächerprinzip in einem linearen Intervall*, Jber. Deutsch. Math.-Verein. *60*, 33–39 (1957).
171. *Mathematische Miszellen*, XXVII: *Über die Stetigkeit von charakteristischen Wurzeln in Abhängigkeit von den Matrizenelementen*, Jber. Deutsch. Math.-Verein. *60*, 40–42 (1957).
172. *Über die Evoluten von endlichen Ovalen*, J. Reine Angew. Math. *198*, 14–27 (1957).
173. *A Method of Speeding Up Iterations with Super-Linear Convergence*, J. Math. Mech. *7*, 117–120 (1958).
174. *On the Convergence of the Rayleigh Quotient Iteration for the Computation of the Characteristic Roots and Vectors*, I, Arch. Rational Mech. Anal. *1*, 233–241 (1958).
175. *Un critère d'univalence des transformations dans un* R^n, C. R. Acad. Sci. Paris *247*, 172–175 (1958).
176. *On the Bounds of a One-Parametric Family of Matrices*, J. Reine Angew. Math. *200*, 190–199 (1958).
177. *Wilhelm Süss*, Freiburg Univ.-Reden [N.F.] *28*, 1–16 (1958).
178. *On the Convergence of the Rayleigh Quotient Iteration for the Computation of the Characteristic Roots and Vectors*, II, Arch. Rational Mech. Anal. *1*, 423–428 (1958).
179. *On Trends and Problems in Numerical Approximation*, in *On Numerical Approximation, Proc. Sympos. Madison, Wisc., 1958* (The Univ. of Wisconsin Press, Madison, Wisc. 1959), pp. 1–10.
180. *On Gauss' Speeding Up Device in the Theory of Single Step Iteration*, Math. Tables and Other Aids to Comput. *12*, 116–132 (1958).
181. *Un nouveau critère d'univalence des transformations dans un* R^n, C. R. Acad. Sci. Paris *248*, 348–350 (1959).
182. *Über einige Sätze von Herrn M. Parodi*, Math. Nachr. *19*, 331–338 (1958).
183. *Über Eigenwerte von Produkten Hermitescher Matrizen*, Abh. Math. Sem. Univ. Hamburg *23*, 60–68 (1959).
184. *Three Theorems on Products of Power Series*, Compositio Math. *14*, 41–49 (1959).
185. *A Quantitative Formulation of Sylvester's Law of Inertia*, Proc. Nat. Acad. Sci. U.S.A. *45*, 740–744 (1959).
186. *On the Convergence of the Rayleigh Quotient Iteration for the Computation of the Characteristic Roots and Vectors*, III *(Generalized Rayleigh Quotient and Characteristic Roots with Linear Elementary Divisors)*, Arch. Rational Mech. Anal. *3*, 325–340 (1959).
187. *On the Convergence of the Rayleigh Quotient Iteration for the Computation of the Characteristic Roots and Vectors*, IV *(Generalized Rayleigh Quotient for Nonlinear Elementary Divisors)*, Arch. Rational Mech. Anal. *3*, 341–347 (1959).
188. *On the Convergence of the Rayleigh Quotient Iteration for the Computation of the Characteristic Roots and Vectors*, V *(Usual Rayleigh Quotient for Non-Hermitian Matrices and Linear Elementary Divisors)*, Arch. Rational Mech. Anal. *3*, 472–481 (1959).
189. *On the Convergence of the Rayleigh Quotient Iteration for the Computation of the Characteristic Roots and Vectors*, VI *(Usual Rayleigh Quotient for Nonlinear Elementary Divisors)*, Arch. Rational Mech. Anal. *4*, 153–165 (1959).
190. *Über Produkte Hermitescher Matrizen und Büschel Hermitescher Formen* (Dem Andenken an Leon Lichtenstein gewidmet), Math. Z. *72*, 1–15 (1959–1960).
191. *On Some Conditions for Non-Vanishing of Determinants*, Proc. Amer. Math. Soc. *12*, 268–273 (1961).

192. *A Regularity Condition for a Class of Partitioned Matrices*, Compositio Math. *15*, 23–27 (1961).
193. *On the Convergence of Gauss' Alternating Procedure in the Method of the Least Squares*, Ann. Mat. Pura Appl. (4) *48*, 229–236 (1959).
194. *Über geränderte Determinanten und bedingte Trägheitsindizes quadratischer Formen*, Monatsh. Math. *64*, 51–63 (1960).
195. *On Some Metrical Properties of Operator Matrices and Matrices Partitioned Into Blocks*, J. Math. Anal. Appl. *2*, 161–209 (1961).
196. *On the Eigenvectors Belonging to the Maximal Root of a Nonnegative Matrix*, Proc. Edinburgh Math. Soc. (2) *12*, 107–112 (1960–1961).
197. *Iterative Solution of Linear Systems of Functional Equations*, J. Math. Anal. Appl. *2*, 351–369, (1961).
198. *A Quantitative Formulation of Sylvester's Law of Inertia*, II, Proc. Nat. Acad. Sci. U.S.A. *46*, 859–862 (1960).
199. *On the Zeros of Bernoulli Polynomials of Even Order*, Enseignement Math. (2) *6*, 27–47 (1960).
200. *On an Inequality of J. Vincente Gonçalves*, Univ. Lisboa Revista Fac. Ci. [A] (2) *8*, 115–119 (1960).
201. (With H. SCHNEIDER) *Bounds for the Maximal Characteristic Root of a Non-Negative Matrix*, Duke Math. J. *27*, 547–553 (1960).
202. *Note on a Theorem by Hans Schneider*, J. London Math. Soc. *37*, 225–234 (1962).
203. *On Some Inequalities in the Theory of Matrices*, Scripta Math. *26*, 201–222 (1963).
204. *On Lancaster's Decomposition of a Matrix Differential Operator*, Arch. Rational Mech. Anal. *8*, 238–241 (1961).
205. (With H. SCHNEIDER) *Some Theorems on the Inertia of General Matrices*, J. Math. Anal. Appl. *4*, 72–84 (1962).
206. *On Positive Matrices*, Math. Ann. *150*, 276–284 (1963).
207. *Eine Vorzeichenregel in der Theorie der algebraischen Gleichungen*, von CARL RUNGE, herausgegeben von A. OSTROWSKI, Jber. Deutsch. Math.-Verein. *66*, 52–66 (1963).
208. *Il metodo del quoziente di Rayleigh* (C.I.M.E., Roma 1963), pp. 1–60.
209. *Sur l'analogue du théorème de Budan-Fourier pour les suites générales des polynomes*, J. Math. Pures Appl. (9) *43*, 49–58 (1964).
210. *On Runge's General Rule of Signs*, Ann. Acad. Sci. Fenn. [Ser. A] I *342*, 1–20 (1964).
211. *On Some Determinants with Combinatorial Numbers*, J. Reine Angew. Math. *216*, 25–30 (1964).
212. *On Approximations of Equations by Algebraic Equations*, J. Soc. Indust. Appl. Math. [Ser. B, Numer. Anal.] *1*, 104–130 (1964).
213. *On n-Dimensional Additive Moduli and Diophantine Approximations*, Acta Arith. *9*, 391–416 (1964).
214. *On Descartes Rule of Signs for Certain Polynomial Developments*, J. Math. Mech. *14*, 195–209 (1965).
215. *Positive Matrices and Functional Analysis*, in *Recent Advances in Matrix Theory* (Univ. Wisconsin Press, Madison, Wisc. 1964), pp. 81–101.
216. *On Ermakof's Convergence Criteria and Abel's Functional Equation*, Enseignement Math. (2) *11*, 103–122 (1965).
217. *Note sur les parties réelles et imaginaires des racines des polynômes*, J. Math. Pures Appl. (9) *44*, 327–329 (1965).
218. *Zur Entwicklung der numerischen Analysis*, Jber. Deutsch. Math.-Verein. *68*, 97–111 (1966).
219. *A Contribution to the Theory of the Fourier Integral Formula*, Math. Ann. *165*, 261–280 (1966).
220. *Sur une propriété des sommes des racines d'un polynome*, C. R. Acad. Sci. Paris [Sér. A–B] *263*, A46–A48 (1966).
221. *The Round-Off Stability of Iterations*, Z. Angew. Math. Mech. *47*, 77–81 (1967).
222. *General Existence Criteria for the Inverse of an Operator*, Amer. Math. Monthly *74*, 826–827 (1967).
223. *Contributions to the Theory of the Method of Steepest Descent*, Arch. Rational Mech. Anal. *26*, 257–280 (1967).

I. *Vorlesungen über Differential- und Integralrechnung*, Erster Band: *Funktionen einer Variablen* (Birkhäuser Verlag, Basel 1945).
2-te, neubearbeitete Aufl. (Birkhäuser Verlag, Basel – Stuttgart 1960 [Mathematische Reihe, Bd. 4, Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften]).
Lições de Calculo Differencial e Integral, I Volume: *Funções de uma Variável* (Fundaçao Calouste Gulbenkian, Lisboa 1967). (Translation into Portuguese.)
Differential and Integral Calculus with Problems, Hints for Solutions and Solutions, I (Scott, Foresman and Company, Glenview, Ill., USA 1968). (Translation into English.)

II. *Vorlesungen über Differential- und Integralrechnung*, Zweiter Band: *Differentialrechnung auf dem Gebiete mehrerer Variablen* (Birkhäuser Verlag, Basel 1951).
2-te, neubearbeitete Aufl. (Birkhäuser Verlag, Basel – Stuttgart 1961 [Mathematische Reihe, Bd. 5, Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften]).

III. *Vorlesungen über Differential- und Integralrechnung*, Dritter Band: *Integralrechnung auf dem Gebiete mehrerer Variablen* (Birkhäuser Verlag, Basel – Stuttgart 1954).
2-te, neubearbeitete Aufl. (Birkhäuser Verlag, Basel – Stuttgart 1967 [Mathematische Reihe, Bd. 7, Lehrbücher und Monographien aus dem Gebiete der exakten Wissenschaften]).

IV. *Solution of Equations and Systems of Equations* (Academic Press, New York – London 1960 [Pure and Applied Mathematics, Vol. IX]).
2-nd edition (Academic Press, New York – London 1966).

Research Papers

Solution générale de l'équation fonctionnelle $f[x+yf(x)]=f(x)f(y)$.

STANISŁAW WOŁODŹKO (Cracovie)

Dédié à A. M. Ostrowski à l'occasion de son 75e anniversaire

Introduction

Dans ce travail nous donnons la solution générale de l'équation fonctionnelle

$$f[x+yf(x)]=f(x)f(y) \tag{1}$$

ainsi que certaines conclusions qui résultent de cette solution.

L'équation (1) fut obtenue par S. GOŁĄB[1]) en cherchant les sous-groupes à trois paramètres du groupe centro-affine sur un plan

$$x' = ax + by$$
$$y' = cx + dy,$$

où a, b, c, d désignent les nombres réels tels que

$$W = \begin{vmatrix} a, & b \\ c, & d \end{vmatrix} \neq 0, \quad \text{et}$$

$$d = d(a, b, c). \tag{2}$$

En effet, en composant les transformations

$$\begin{aligned} x' &= ax + by \\ y' &= cx + d(a, b, c)\, y \end{aligned} \quad \text{ainsi que} \quad \begin{aligned} x'' &= a'x' + b'y' \\ y'' &= c'x' + d(a', b', c')\, y' \end{aligned}$$

nous obtenons

$$x'' = (aa' + cb')\, x + [a'b + b'd(a, b, c)]\, y$$
$$y'' = [ac' + cd(a', b', c')]\, x + [c'b + d(a, b, c)\, d(a', b', c')]\, y.$$

Comme la condition (2) doit être remplie, on aura

$$d[aa' + cb', a'b + b'd(a, b, c), ac' + cd(a', b', c')] = cb' + d(a, b, c)\, d(a', b', c').$$

[1]) Indépendamment de S. GOŁĄB une équation pareille fut obtenue par J. ACZÉL en partant de certains problèmes de la théorie des objets géométriques [1].

Reçu le 26 juli, 1967

De là, en posant

$$a=a'=1' \qquad b=b'=0, \qquad c'=x, \qquad c=y$$

nous obtenons

$$d[1,0,x+y\,d(1,0,x)]=d(1,0,x)\,d(1,0,y),$$

donc, en posant $f(x)=d(1,0,x)$ nous obtenons l'équation (1).

Dans le travail [2] S. Gołąb et A. Schinzel ont donné une vaste classe des solutions en particulier toutes les solutions régulières de l'équation (1) dans l'ensemble des fonctions réelles de la variable réelle. Les résultats les plus importants de ce travail contiennent les propositions données plus bas 1–5. En outre on a montré dans le travail [2] l'existence des solutions non mesurables de l'équation (1) en donnant certaines propriétés des solutions mesurables de ladite équation. Les solutions mesurables générales de l'équation (1) furent aussi étudiées par C. Gh. Popa dans le travail [3]. Les résultats de ces deux travaux concernant les solutions mesurables sont donnés dans la proposition 5. Nous allons maintenant citer les propositions 1–5 (avec quelques changements insignifiants).

Proposition 1. Les seules solutions continues de l'équation (1) sont les fonctions suivantes:

$$\left.\begin{aligned}
&\text{(a)} \quad f(x)\equiv 0,\\
&\text{(b)} \quad f(x)=1+ax,\\
&\text{(c)} \quad f(x)=\begin{cases} 0 & \text{pour } x\leqslant -\dfrac{1}{a}\\[2mm] 1+ax & \text{pour } x>-\dfrac{1}{a} \end{cases} \quad (a>0),\\
&\text{(d)} \quad f(x)=\begin{cases} 1+ax & \text{pour } x<\dfrac{1}{a}\\[2mm] 0 & \text{pour } x\geqslant \dfrac{1}{a} \end{cases} \quad (a>0),
\end{aligned}\right\} \qquad (3)$$

où a est une constante réelle arbitraire.

Parmi les solutions de l'équation (1) les auteurs du travail [2] distinguent les solutions dites triviales et les solutions micropériodiques. Nous dirons que la fonction f est une solution triviale de l'équation (1) si l'ensemble des valeurs de cette fonction est contenu dans l'ensemble $\{-1,0,1\}$. Nous dirons que f est une solution micropériodique de l'équation (1) si 0 est le point d'accumulation de ses périodes.

Proposition 2. La fonction f est une solution non-micropériodique et non triviale de (1) seulement s'il existe un nombre $a\neq 0$ et un groupe multiplicatif G qui ne contient seulement ± 1, mais aussi d'autres nombres, et G est tel que

$$f(x)=\begin{cases} 1+ax, & \text{quand } (1+ax)\in G,\\ 0 & \text{quand } (1+ax)\in G. \end{cases} \qquad (4)$$

Proposition 3. Si la solution non triviale de l'équation (1) n'est pas de la forme (3) (b), (c) ou (d), dans ce cas cette solution est discontinue en tout point d'une certaine demi-droite ou bien elle est de la forme (4) où le groupe G se compose de nombres a^n ou $\pm a^n$, a étant un nombre réel différent de zéro.

Proposition 4. L'équation (1) possède des solutions micropériodiques non triviales.

Proposition 5. Si la fonction f est une solution mésurable de l'équation (1), dans ce cas f est soit continu (de la forme (3)), soit égal presque partout à zéro.

Ce travail se compose de trois parties. Dans le § 1 nous donnerons un moyen d'obtenir les solutions de l'équation (1) appelé par nous «construction K». Nous montrerons ensuite que toute solution de l'équation (1) qui n'est pas identiquement égale à zéro peut être obtenue au moyen de cette construction (proposition 6). Le problème sera résolu dans l'espace linéaire arbitraire sur un corps quelconque. Dans la démonstration de la proposition 6 nous utiliserons l'axiome du choix. Dans le § 2 nous montrerons comment on peut obtenir les propositions 1–4 de la proposition 6. En même temps, la proposition 1 sera présentée en faisant des suppositions affaiblies (remarque 5). Dans le § 3 nous donnerons les solutions continues de l'équation (1) dans l'ensemble des fonctions complexes d'une variable complexe.

§ 1. Soit E l'espace linéaire sur un corps donné Φ. Les éléments de l'ensemble E seront désignés par les lettres latines, et les éléments de Φ par les lettres grecques. Pour les opérations dans l'ensemble E les symboles additifs, et dans l'ensemble Φ les symboles additifs-multiplicatifs seront utilisés. Pour une opération extérieure nous adopterons la convention suivante:

$$\lambda a = a \lambda .$$

Supposons que la fonction f soit l'application de l'ensemble E dans l'ensemble Φ, ce que nous noterons

$$f: E \to \Phi .$$

Nous passons maintenant à un moyen d'obtenir les solutions de l'équation (1).

Construction K

1. Nous choisissons un sous-groupe multiplicatif arbitraire Z' du groupe multiplicatif $\Phi' = \Phi \setminus \{0\}$.

2. Nous déterminons dans l'ensemble Z' une fonction arbitraire w dont les valeurs appartiennent à l'ensemble E, c'est-à-dire

$$w: Z' \to E .$$

3. Nous déterminons la fonction

$$a(\lambda, \mu) = w(\lambda \mu) - w(\lambda) - \lambda w(\mu) \quad \text{pour} \quad \lambda, \mu \in Z' . \tag{5}$$

4. Nous désignons par A' un sur-ensemble arbitraire de l'ensemble des valeurs de la fonction $a(\lambda, \mu)$, c'est-à-dire

$$a(Z' \times Z') \subset A'.$$

5. Nous désignerons par A'' l'ensemble de tous les produits λa où $a \in A'$ et $\lambda \in Z'$ (c'est-à-dire $A'' = Z' \cdot A'$).

6. L'ensemble A'' génère un groupe additif; nous le désignerons par A. Ainsi l'ensemble A se compose de toutes les combinaisons linéaires $k_1 a_1 + \cdots + k_n a_n$ où $a_1, \ldots, a_n \in A''$ et $k_1, \ldots, k_n$ sont des nombres entiers arbitraires, n étant également un entier quelconque.

7. Choisissons un sous-groupe multiplicatif arbitraire Z du groupe multiplicatif Z' tel que pour tout $\lambda \in Z$ et $\lambda \neq 1$ (voir la remarque 1)

$$w(\lambda) \bar{\in} A. \tag{6}$$

8. Déterminons la fonction (voir le lemme 3)

$$f(x) \overset{\text{df}}{=} \begin{cases} \lambda & \text{s'il existe } \lambda \in Z \text{ tel que } [x - w(\lambda)] \in A, \text{ si un tel } \lambda \text{ existe} \\ 0 & \text{pour les autres } x. \end{cases} \tag{7}$$

Remarque 1. Pour tout groupe Z' de l'ensemble A' et pour toute fonction w il existe un groupe Z remplissant la condition (6), p. ex. $Z = \{1\}$.

Nous allons maintenant démontrer cinq lemmes concernant les propriétés des ensembles et des fonctions qui apparaissent dans la construction K.

Lemme 1. *$a \in A$ pour un $a \in F$ si $\lambda a \in A$ pour un $\lambda \in Z$ et si $a \in A$ alors $\lambda a \in A$ pour tout $\lambda \in Z'$.*

Démonstration. 1. Soit $a \in A$. Il résulte de la définition même du groupe A qu'il existe un nombre entier positif n, des nombres entiers $k_1, \ldots, k_n$ et des éléments $a_1, \ldots, a_n \in A''$ tels que

$$a = k_1 a_1 + \cdots + k_n a_n. \tag{8}$$

D'autre part, il s'ensuit de la définition de l'ensemble A'' qu'il existe des $a_i' \in A'$ et des $\lambda_i' \in Z'$ tels que

$$a_i = a_i' \lambda_i' \quad (i = 1, \ldots, n). \tag{9}$$

De (8) et (9) nous obtenons

$$a = k_1 a_1' \lambda_1' + \cdots + k_n a_n' \lambda_n'.$$

In en résulte que

$$\lambda a = k_1 a_1' \lambda \lambda_1' + \cdots + k_n a_n' \lambda \lambda_n' \quad \text{pour tout } \lambda \in Z'. \tag{10}$$

Le groupe Z' étant un groupe multiplicatif, on a de (10) et de la définition des ensembles A'' et A

$$\lambda a \in A.$$

2. Soit maintenant $\lambda a \in A$ pour un $\lambda \in Z'$ et $a \in A$. De la partie précédente de la démonstration on obtient

$$[\lambda^{-1}(\lambda a)] \in A .$$

Ainsi $a \in A$.

LEMME 2. *Si* λ, $\mu \in Z'$, *on a* $[w(\lambda\mu) - w(\lambda) - \lambda w(\mu)] \in A$.

Démonstration. La proposition s'ensuit directement de la définition des ensembles A', A'' et de A.

LEMME 3. *La règle* (7) *définit une fonction.*

Démonstration. Il suffit de montrer que pour tout $x \in E$ il existe au plus un $\lambda \in Z$ pour lequel

$$[x - w(\lambda)] \in A .$$

Pour cela il suffit de montrer qui si λ, $\mu \in Z$ et $\lambda \neq \mu$, on aura pour chaque $x \in E$

$$[x - w(\lambda)] \overline{\in} A \quad \text{ou} \quad [x - w(\mu)] \overline{\in} A . \tag{11}$$

A étant un groupe additif, (11) a lieu si

$$[w(\lambda) - w(\mu)] \overline{\in} A \quad \text{pour} \quad \lambda \neq \mu \quad \text{et} \quad \lambda, \mu \in Z . \tag{12}$$

Pour démontrer (12) remarquons que du lemme 2, en posant μ et λ/μ au lieu de λ et μ, respectivement, on obtient

$$\left[w(\lambda) - w(\mu) - \mu w\left(\frac{\lambda}{\mu}\right) \right] \in A \quad \text{pour} \quad \lambda, \mu \in Z . \tag{13}$$

D'autre part, pour $\lambda \neq \mu$ et λ, $\mu \in Z$ on obtient à cause de (6) $w(\lambda/\mu) \overline{\in} A$, donc aussi (lemme 1) $\mu w(\lambda/\mu) \overline{\in} A$. De là et de (13) on arrive à (12).

LEMME 4. $f^{-1}(\{1\}) = A$.

Démonstration. Du lemme 2, en posant $\lambda = 1$, nous obtenons

$$- w(1) \in A .$$

De là, puisque A est un groupe additif, on a

$$w(1) \in A .$$

Il s'ensuit, que $x \in A$, si $[x - w(1)] \in A$ et dans ce cas seulement. De là et de (7) résulte la proposition à démontrer.

LEMME 5. *L'ensemble A est un ensemble de périodes de la fonction f.*

Démonstration. Il suffit de montrer que $a \in A$, si et seulement si pour tout $x \in E$

$$f(x + a) = f(x) . \tag{14}$$

1. Supposons que

$$a \in A. \tag{15}$$

Dans ce cas pour un $x \in E$ un des cas suivants doit subsister:

a) il existe $\lambda \in Z$ tel que

$$[x - w(\lambda)] \in A, \tag{16}$$

b) pour tout $\lambda \in Z$

$$[x - w(\lambda)] \bar{\in} A. \tag{17}$$

Dans le cas a) nous avons

$$f(x) = \lambda. \tag{18}$$

D'autre part, on conclue de (15) et de (16) que

$$[x + a - w(\lambda)] \in A.$$

Alors:

$$f(x + a) = \lambda.$$

De là et de (18) on obtient (14).
Dans le cas b) nous obtenons

$$f(x) = 0. \tag{19}$$

Mais de (15) et de (17) nous avons

$$[x + a - w(\lambda)] \bar{\in} A$$

pour tout $\lambda \in Z$. Il en résulte que

$$f(x + a) = 0.$$

De là et de (19) nous avons (14).

2. Supposons maintenant que pour tout $x \in E$

$$f(x + a) = f(x).$$

De là, en posant $x=0$, nous obtenons

$$f(a) = f(0). \tag{20}$$

Puisque $0 \in A$, il s'ensuit du lemme 4 que

$$(0) = 1.$$

De là, de (20) et du lemme 4 résulte que $a \in A$.

Nous démontrerons maintenant le

THÉORÈME 1. *La fonction f, non identiquement nulle, est une solution de l'équation* (1) *si et seulement s'il existe des ensembles Z', A' et Z et une fonction w qui remplie les conditions de la construction K telle que la fonction f soit la forme* (7).

Autrement dit, l'ensemble des solutions de l'équation (1) *(laissant à côté la solution identiquement nulle) est identique avec l'ensemble des fonctions qui peuvent être obtenues par la construction K.*

Démonstration de la suffisance. Nous supposons que pour la fonction f il existe des ensembles Z', A', Z et une fonction w tel que $f(x)$ soit de la forme (7). Nous montrerons que la fonction f est la solution de l'équation (1).

Pour les x, y arbitrairement déterminés de l'ensemble E trois éventualités peuvent avoir lieu:

a) il existe $\lambda, \mu \in Z$ tel que $[x-w(\lambda)] \in A$ et $[y-w(\mu)] \in A$,

b) $[x-w(\lambda)] \in A$ pour un certain $\lambda \in Z$ et $[y-w(\mu)] \overline{\in} A$ pour tout $\mu \in Z$,

c) $[x-w(\lambda)] \overline{\in} A$ pour chaque $\lambda \in Z$, y étant un élément quelconque de l'ensemble E.

Ad a). Puisqu'il existe des constantes $a, b \in A$ telles que

$$x = w(\lambda) + a \qquad \text{et} \qquad y = w(\mu) + b$$

on a alors

$$f[x + yf(x)] = f\{w(\lambda) + a + [w(\mu) + b] f[w(\lambda) + a]\}. \tag{21}$$

D'autre part, il résulte des lemmes 1 et 5 que a, b et $bf[w(\lambda)+a]$ sont des périodes de la fonction f. Il s'ensuit que

$$f\{w(\lambda) + a + [w(\mu) + b] f[w(\lambda) + a]\} = f\{w(\lambda) + w(\mu) f[w(\lambda)]\}. \tag{22}$$

De la définition de la fonction f nous avons

$$f[w(\lambda)] = \lambda \qquad \text{pour} \qquad \lambda \in Z. \tag{23}$$

Nous concluons du lemme 2 que

$$f[w(\lambda) + \lambda w(\mu)] = f[w(\lambda \mu)], \tag{24}$$

et de (23) et (24) nous obtenons

$$\{w(\lambda) + w(\mu) f[w(\lambda)]\} = \lambda \mu = f[w(\lambda)] f[w(\mu)] = f(x) f(y). \tag{25}$$

L'énonce du théorème résulte de (21), (22) et (25).

Ad b). Comme dans le cas a) nous constatons que

$$x = w(\lambda) + a \tag{26}$$

où $a \in A$. D'autre part, il s'ensuit de (7) que $f(y) = 0$.

Il suffit donc de montrer que

$$f[x + yf(x)] = 0.$$

Admettons, pour la démonstration par le contraire, que

$$f[x + yf(x)] = \mu \neq 0.$$

De là et de (26) s'ensuit que

$$[w(\lambda) + \lambda y - w(\mu)] \in A. \tag{27}$$

D'autre part on obtient du lemme 2

$$\left[w(\mu)-w(\lambda)-\lambda w\left(\frac{\mu}{\lambda}\right)\right]\in A. \tag{28}$$

A étant un groupe additif, nous obtenons de (27) et (28)

$$[w(\lambda)+\lambda y-w(\mu)]+\left[w(\mu)-w(\lambda)-\lambda w\left(\frac{\mu}{\lambda}\right)\right]=\lambda\left[y-w\left(\frac{\mu}{\lambda}\right)\right]\in A.$$

De là et du lemme 1 nous avons

$$\left[y-w\left(\frac{\mu}{\lambda}\right)\right]\in A,$$

ce qui impossible puisque

$$f(y)=0\neq\frac{\mu}{\lambda}.$$

Ad c). Dans ce cas $f(x)=0$. Donc aussi

$$f[x+yf(x)]=f(x)=0,$$

d'où résulte la thèse.

Démonstration de la nécessité. Supposons que la fonction f soit une solution de l'équation (1) qui n'est pas identiquement égale à zéro. Nous montrerons qu'il existe des ensembles Z', A' et Z ainsi que la fonction w tel que f à la forme (7). Nous citerons auparavant sept lemmes (lemmes 6–12). Puisque ces lemmes sont démontrés dans le travail [2], nous citons ces lemmes sans les démontrer.

LEMME 6. $f(0)=1$.

LEMME 7. *Si $f(x)=f(x')\neq 0$, la différence $x-x'$ est une période de la fonction f.* Supposons que

$$A_1 \overset{\text{df}}{=} f^{-1}(\{1\}).$$

LEMME 8. *L'ensemble A_1 est l'ensemble des périodes de la fonction f.*

LEMME 9. *L'ensemble A_1 est un sous-groupe additif du groupe additif E.*
Désignons par F l'ensemble des valeurs de la fonction f.

LEMME 10. *L'ensemble $F\setminus\{0\}$ est un sous-groupe multiplicatif du groupe multiplicatif $\Phi'(\overset{\text{df}}{=}\Phi\setminus\{0\})$.*

LEMME 11. $a\in A_1$, pour un $a\in E$ si et seulement si $\lambda a\in A_1$ pour tout $\lambda\in F\setminus\{0\}$.

LEMME 12. *Si $f(x)=\lambda\neq 1$, on a $f[x(1-\lambda)^{-1}]=0$.*
Il résulte de ce lemme que si la fonction f est une solution de l'équation (1) qui

n'est pas identiquement égale à l'unité, dans ce cas cette fonction possède les points où elle est égale à zéro. Nous allons utiliser ce fait dans les paragraphes suivants.

Soit

$$A_\lambda \overset{\mathrm{df}}{=} f^{-1}(\{\lambda\}) \quad \text{pour} \quad \lambda \in F \setminus \{0\}.$$

Il résulte des lemmes 7, 8 et 9 que l'ensemble A_λ est une couche du groupe additif E relativement au sous-groupe A_1. Nous choisirons un point dans chaque ensemble A_λ et nous le désignerons par x_λ. Il est évident que pour $\lambda \neq 1$

$$x_\lambda \bar{\in} A_1 . \tag{29}$$

Il s'ensuit de la définition des ensembles A_λ et des points x_λ que

$$f(x) = \begin{cases} \lambda & \text{s'il existe } \lambda \in F \setminus \{0\} \text{ tel que } [x - x_\lambda] \in A_1, \\ 0 & \text{pour les autres } x. \end{cases} \tag{30}$$

On voit maintenant de quelle manière ou peut obtenir la fonction f de la construction K. Il suffit pour cela de supposer que

$$Z' = Z = F \setminus \{0\}, \tag{31}$$

$$A' = A_1 \quad \text{et} \tag{32}$$

$$w(\lambda) = x_\lambda . \tag{33}$$

En effet, de la structure de l'équation (1), de la définition du point x_λ et des lemmes 7 et 8 il s'ensuit que

$$[x_{\lambda\mu} - x_\lambda - \lambda x_\mu] \in A_1 \quad \text{pour} \quad \lambda, \mu \in F \setminus \{0\}.$$

De là, ainsi que de (5) et (33), nous avons

$$a(\lambda, \mu) \in A_1 \quad \text{pour} \quad \lambda, \mu \in F \setminus \{0\}.$$

Nous en déduisons que l'ensemble A' défini par la relation (32) est un sur-ensemble de l'ensemble des valeurs de la fonction a. D'autre part, nous obtenons du lemme 11 que

$$A'' = A_1$$

et puisque A_1 est un groupe additif, on a

$$A = A_1 .$$

Le fait que (6) a lieu résulte de (29) et (33). De ces conclusions et de (30) nous avons (7).

REMARQUE 2. L'ensemble des valeurs différentes de zéro de la fonction f déterminée par la relation (7) est l'ensemble Z. On peut se poser la question, dans la construction K pourquoi nous sommes partis d'un certain sur-ensemble Z' de l'ensemble Z et non de l'ensemble des valeurs de la fonction à construire, comme le suggère la seconde partie de la démonstration du theorème 6 (v. (31)). Le procédé exposé dans la construction K est motivé parce que l'ensemble Z' ne peut pas remplir les conditions (6) (p. ex. dans le cas où $A=E$) ce que nous oxigeons de l'ensemble des valeurs de la fonction f.

REMARQUE 3. La question peut être posée si l'ensemble A' (introduit dans la construction K comme un sur-ensemble de l'ensemble des valeurs de la fonction a) peut être défini comme l'ensemble des valeurs de la fonction a. Nous allons donner un exemple duquel il résulte que la construction K ainsi modifiée ne fournira pas toutes les solutions de l'équation (1).

Considérons un groupe de KLEIN de quatrième ordre $(E, +)$. En désignant les éléments de l'ensemble E par 0, 1, 2, 3, nous obtiendrons la table d'addition suivante

+	0	1	2	3
0	0	1	2	3
1	1	0	3	2
2	2	3	0	1
3	3	2	1	0

Supposons que $(\Phi, +, .)$ désigne un corps à deux éléments. Si nous désignons les éléments de l'ensemble Φ par 0 et 1, nous obtiendrons les tables de règles suivantes:

+	0	1
0	0	1
1	1	0

.	0	1
0	0	0
1	0	1

L'opération extérieure du type $\Phi \times E \to E$ peut être définie au moyen de la table

.	0	1	2	3
0	0	0	0	0
1	0	1	2	3

Il est facile de voir que l'ensemble E est un espace linéaire sur un corps variable Φ.

Nous allons maintenant appliquer la construction K à la résolution de l'équation (1) dans l'espace linéaire E. Nous obtenons successivement

$$Z' = \{1\}$$

$$w(1) = k, \quad \text{où} \quad k \in E,$$

$$a(1, 1) = w(1) - w(1) - 1 \cdot w(1) = -w(1) = -k = k.$$

En considérant A' comme l'ensemble des valeurs de la fonction a nous avons

$$A' = \{k\}. \qquad \text{Alors} \qquad A'' = \{k\}.$$

De là et de la définition du groupe A, nous avons

$$A = \{0, k\}.$$

Puisque nous ne pouvons pas admettre que $Z=1$, nous obtenons quatre solutions de l'équation (1)

$$f_k(x) = \begin{cases} 1 & \text{pour} \quad x \in A = \{0, k\}, \\ 0 & \text{pour} \quad x \in E \setminus A \end{cases}$$

pour $k=0, 1, 2, 3$. Dans l'ensemble des fonctions obtenues ne se trouve pas la solution

$$f(x) \equiv 1.$$

Nous l'obtiendrons de la construction K en posant p. ex. $A'=E$.

REMARQUE 4. L'analogie entre les lemmes 1–5 et les lemmes 6–11 n'est qu'apparante puisque les lemmes 1–5 sont formulés en supposant que la fonction f est obtenue de la construction K, tandis que les lemmes 6–11 – en supposant que f est une solution de l'équation (1).

§ **2**. Dans le paragraphe nous nous occuperons des solutions réelles d'une variable réelle de l'équation (1). Nous démontrerons d'abord le lemme suivant.

LEMME 13. *La solution unique de l'équation*

$$w(\lambda\mu) = w(\lambda) + \lambda w(\mu), \quad \textit{pour} \quad \lambda, \mu \in G, \tag{34}$$

où G est un sous-groupe multiplicatif arbitraire du groupe des nombres réels, est la fonction

$$w(\lambda) = (1 - \lambda)\, c,$$

où c est une constante réelle arbitraire.

Démonstration. Si $G=\{1\}$, la proposition est évidente. Soit maintenant $G \neq \{1\}$. On a de (34)

$$w(\mu\lambda) = w(\mu) + \mu w(\lambda).$$

Il en résulte que

$$w(\lambda) + \lambda w(\mu) = w(\mu) + \mu w(\lambda).$$

Il s'ensuit que pour $\lambda \neq 1$ et $\mu \neq 1$ on a

$$\frac{w(\lambda)}{1-\lambda} = \frac{w(\mu)}{1-\mu}.$$

Donc

$$w(\lambda) = (1 - \lambda)\, c \quad \text{pour} \quad \lambda \neq 1, \tag{35}$$

c étant une certaine constante réelle.

D'autre part, en posant $\lambda=1$, et de (34) nous obtenons

$$w(1)=0.$$

Il en résulte que (35) a lieu pour tout $\lambda \in G$. Nous constatons au moyen d'une vérification directe, que (35) est une solution de l'équation (34) pour chaque c réel.

Nous allons maintenant démontrer, en nous basant sur le théorème 6, les propositions 1–4 citées dans l'introduction.

Démonstration de la proposition 1. Si la fonction f est une solution continue de l'équation (1), l'ensemble F des valeurs de cette fonction est un ensemble connexe. En excluant la solution égale identiquement à zéro, nous savons que l'ensemble $F \setminus \{0\}$ est un groupe multiplicatif. De là et du lemme 12 nous concluons qu'un des trois cas suivants doit subsister:

$$\begin{aligned} &\text{a)} \quad F=\{1\}, \\ &\text{b)} \quad F=(0,+\infty), \\ &\text{c)} \quad F=(-\infty,+\infty). \end{aligned}$$

Dans le cas a) nous avons $f(x)\equiv 1$. Dans le cas b) et c) nous pouvons remarque qu'il résulte du lemme 1 que le groupe A ne peut être composé que du nombre 0. Il s'ensuit que

$$A''=A'=\{0\}.$$

Alors

$$a(\lambda,\mu)\equiv 0 \quad \text{pour} \quad \lambda,\mu \in F\setminus\{0\}.$$

De là et du lemme 13 résulte que

$$w(\lambda)=(1-\lambda)\,c, \tag{36}$$

où c est une constante réelle.

D'autre part, puisque $A=\{0\}$, on a de (7) et de (36) que

$$f(x)=\begin{cases} \lambda & \text{pour } x=(1-\lambda)\,c \quad \text{et } \lambda \in F\setminus\{0\}, \\ 0 & \text{pour les autres } x. \end{cases}$$

De là nous obtenons dans le cas b) les solutions (3) (c), (d) et dans le cas c) la solution (3) (b).

Remarque 5. On sait que si un sous-groupe multiplicatif d'un groupe de nombres réels contient au moins un intervalle non dégéneré, dans ce cas ce sous-groupe est un groupe de nombres réels positifs, ou bien c' est un groupe de nombres réels différents de zéro. De là et de la démonstration de la proposition 1 (donnée plus haut) s'ensuit que les seules solutions de l'équation (1) dont l'ensemble des valeurs n'est pas un ensemble – frontière sont les solutions de la forme (3) (c), (d) et (3) (b) pour $a\neq 0$. Il en résulte que l'énoncé de la proposition 1 sera vrai si l'on suppose que seulement l'ensemble des valeurs de la fonction f – solution de l'équation (1) – est un ensemble connexe. En particulier les seules solutions de l'équation (1) qui ont la propriété Darboux sont les solutions du type (3).

Nous démontrerons maintenant le

LEMME 14. Si la fonction f est une solution non triviale de l'équation (1) et si l'ensemble $f^{-1}(\{1\})$ possède des éléments différents de zéro, dans ce cas la fonction f est une fonction micropériodique.

Démonstration. Il résulte de la supposition faite qu'il y a des nombres a et λ tels que $a \neq 0$ et $a \in A$, $|\lambda| \neq 1$ et $\lambda \in Z$. De là et du lemme 1 nous obtenons que $a\lambda^n \in A$ pour $n = 0, \pm 1, \ldots$. Puisque $\lim_{n \to +\infty} a\lambda^n = 0$ ou $\lim_{n \to -\infty} a\lambda^n = 0$, la fonction f doit être micropériodique.

Démonstration de la proposition 2

1. *Nécessité.* Supposons que la fonction non micropériodique f soit une solution non triviale de l'équation (1). De là et du lemme 14 résulterait que $A = \{0\}$. Alors, comme dans la démonstration de la proposition 1, il s'ensuit que

$$w(\lambda) = (1 - \lambda)\, c\,,$$

où c est une constante réelle; de (6) et de la supposition faite on aurait $c \neq 0$. Il suffirait donc de prendre $a = -(1/c)$ pour de (7) obtenir l'énoncé.

2. *Suffisance.* Supposons maintenant qu'il existe un nombre $a \neq 0$ et un groupe multiplicatif G contenant l'élément λ tel que $|\lambda| \neq 1$ et tel que (4) soit valable. Puisque la fonction $1 + ax$ épuise l'ensemble de nombres réels, il s'ensuit de (4) que le groupe G est l'ensemble des valeurs de f différents de zéro. En admettant maintenant dans la construction K que

$$Z' = Z = G\,, \qquad w(\lambda) = -(1 - \lambda)\frac{1}{a} \quad \text{et} \quad A' = \{0\}\,,$$

nous obtenons que la fonction f est de la forme (7), c'est-à-dire elle est une solution de l'équation (1). f est une solution non triviale (l'ensemble des valeurs de la fonction est le groupe G) et non micropériodique puisque elle n'est pas périodique ($A = \{0\}$).

Démonstration de la proposition 3. Considérons deux cas

$$1) \qquad A \neq \{0\}\,,$$
$$2) \qquad A = \{0\}\,.$$

Ad 1). De la supposition que la fonction f est une solution non triviale de l'equation (1) on conclue que cette fonction est micropériodique et comme elle n'est pas constante, elle doit être discontinue sur toute la droite.

Ad 2). Puisque nous excluons les solutions continues, donc (v. remarque 5) l'ensemble des valeurs de la fonction f est un ensemble-frontière. Nous obtenons de là (v. [2] p. 123) que trois cas sont possibles:

a) L'ensemble $F\backslash\{0\}$ se compose de nombres de la forme a^n ou $\pm a^n$ où a est un certain nombre réel différent de zéro et $n=0, \pm 1, \ldots$.

b) L'ensemble $F\backslash\{0\}$ est un ensemble dense dans l'intervalle $(0, +\infty)$.

c) L'ensemble $F\backslash\{0\}$ est dense dans l'intervalle $(-\infty, +\infty)$.

D'autre part, (en tenant compte de la proposition 2) la fonction f est de la forme (4). Alors, dans le cas a), nous obtenons la thèse; dans le cas b) nous avons la discontinuité de la fonction f sur la demidroite et dans le cas c) – la discontinuité sur toute la droite. D'où résulte la proposition.

Démonstration de la proposition 4. L'exemple d'une solution micropériodique non triviale de l'équation (1) qui était donné dans le travail [2] peut être obtenu de la construction K en procédant comme dans la seconde partie de la démonstration du théorème 6. Nous allons donner encore un exemple qui est en même temps l'illustration dont on peut se servir de notre construction pour obtenir des solutions de l'équation (1). Soit

$$Z' = Z = \{\lambda : \underset{k \in N}{\exists} \underset{s \in N}{\exists} (\lambda = 3^k 5^s)\},$$

où N désigne l'ensemble de nombres entiers et

$$w(3^k 5^s) = (1 - 3^k 5^s)\sqrt{2} + w'(3^k 5^s) \qquad \text{pour} \qquad k, s \in N,$$

où

$$w'(3^k 5^s) = \begin{cases} 0 & \text{pour les } s \text{ pairs, } k\text{-entiers,} \\ \frac{1}{2} & \text{pour les } s \text{ impairs, } k\text{-entiers.} \end{cases}$$

On aura dans ce cas

$$a(3^k 5^s, 3^r 5^t) = \begin{cases} 0 & \text{pour les } t \text{ pairs, } k, s, r\text{-entiers,} \\ -\frac{1}{2}(1 + 3^k 5^s) & \text{pour les } s, t \text{ impairs, } k\text{-entiers,} \\ \frac{1}{2}(1 + 3^k 5^s) & \text{pour les } s \text{ pairs, } t \text{ impairs, } k, r\text{-entiers.} \end{cases}$$

En posant maintenant

$$A' = A'' = A = \left\{x : \underset{a, r, h \in N}{\exists} \left(x = \frac{a}{3^r 5^h}\right)\right\},$$

nous obtenons la solution cherchée. Le fait que cette solution est micropériodique résulte du lemme 14.

§ 3. En terminant, nous allons considérer le problème des solutions continues de l'équation (1) dans l'ensemble des fonctions complexes d'une variable complexe. Nous montrerons comment on peut obtenir ces solutions au moyen de la construction K.

Supposons que f désigne la solution continue de l'équation (1)

$$f : C \to C,$$

où C est l'ensemble de nombres complexes. Pour éviter dans la suite la considération des cas triviaux, nous supposons que la fonction f ne soit pas identiquement égale à zéro ni à l'unité. Nous introdisons maintenant les désignations suivantes:

$$F = f(C).$$
$$R = \{r: \underset{\lambda \in F\setminus\{0\}}{\exists} (r = |\lambda|)\}.$$
$$\Omega = \{\omega: \underset{\lambda \in F\setminus\{0\}}{\exists} (\omega = \operatorname{Arg} \lambda)\}.$$
$$A = f^{-1}(\{1\}).$$
$$\operatorname{Re} A = \{x: \underset{a \in A}{\exists} (x = \operatorname{Re} a)\}.$$
$$\operatorname{Im} A = \{x: \underset{a \in A}{\exists} (x = \operatorname{Im} a)\}.$$

Le fonction f étant continue dans un ensemble connexe, l'ensemble F est également connexe. Il s'ensuit que

$$\text{l'ensemble } R \text{ est connexe.} \tag{37}$$

Dans le cas contraire il existerait $r_0 \bar{\in} R$ tel que dans deux intervalles $(0, r_0)$ et $(r_0, +\infty)$ il y aurait des éléments de l'ensemble R. Il s'ensuit qu'il y a des éléments de l'ensemble F qui appartiennent à l'intérieur du cercle limité par la circonférence $|z| = r_0$ et aussi des éléments qui sont à l'extérieur de ce cercle. Par contre aucun point du cercle $|z| = r_0$ ne fait partie de l'ensemble F. La conclusion obtenue est en contradiction avec la connexité de l'ensemble F.

D'autre part nous savons (lemme 10) que l'ensemble $F\setminus\{0\}$ est un sous-groupe du groupe multiplicatif des nombres complexes. Nous allons montrer que

$$\text{l'ensemble } R \text{ est un sous-groupe du groupe multiplicatif des nombres réels positifs.} \tag{38}$$

Choisissons dans ce but deux éléments arbitraires $r_1, r_2 \in R$. De là et de la définition des ensembles R et Ω résulte qu'il existe $w_1, w_2 \in \Omega$ tels que $\lambda_1 = r_1 e^{i\omega_1}, \lambda_2 = r_2 e^{i\omega_2}$ et aussi $\lambda_1, \lambda_2 \in F\setminus\{0\}$. L'ensemble $F\setminus\{0\}$ étant un groupe multiplicatif, on a aussi $\lambda_1, \lambda_2^{-1} \in F\setminus\{0\}$. D'autre part $\lambda_1, \lambda_2^{-1} = r_1 r_2^{-1} e^{i(\omega_1 - \omega_2)}$. De là et de la définition de l'ensemble R nous avons $r_1 r_2^{-1} \in R$, ce qui termine la démonstration.

Remarquons encore (ce qui n'est pas nécessaire pour les considérations ultérieures) que l'on peut pareillement montrer que l'ensemble Ω est un sous-groupe du groupe additif modulo 2π des nombres réels.

De (37) et (38) nous obtenons deux éventualités: $R = \{1\}$ ou $R = (0, +\infty)$. Nous montrerons maintenant que le cas $R = \{1\}$ doit être exclu. En effet, dans ce cas de la définition même de l'ensemble R résulterait que les valeurs de la fonction f différentes de zéro appartiennent a là circonférence $|z| = 1$. D'autre part la fonction f n'est pas égale identiquement à 1, donc aussi (lemme 12) $0 \in F$ ce qui est impossible puisque l'ensemble F est connexe. Finalement donc

$$R=(0,+\infty). \tag{39}$$

Rappelons ensuite (lemme 9), que l'ensemble A est sous-groupe du groupe additif des nombres complexes. De là, en raisonnant comme dans la démonstration de la conclusion (38), nous obtenons que les ensembles $\operatorname{Re} A$ et $\operatorname{Im} A$ sont des sous-groupes du groupe additif des nombres réels. De là, de (39) et du lemme 11 nous obtenons quatre cas possibles:

a) $\operatorname{Re} A=\operatorname{Im} A=\{0\}$ c'est-à-dire $A=\{0\}$,

b) $\operatorname{Re} A=(-\infty,+\infty)$, $\operatorname{Im} A=\{0\}$ c'est-à-dire $A=(-\infty,+\infty)$,

c) $\operatorname{Re} A=\{0\}$, $\operatorname{Im} A=(-\infty,+\infty)$, c'est-à-dire $A=\{z: \operatorname{Re} z=0\}$,

d) $\operatorname{Re} A=(-\infty,+\infty)$, $\operatorname{Im} A=(-\infty,+\infty)$, c'est-à-dire $A=C$.

La dernière possibilité doit être exclue parce que dans ce cas la seule solution de l'équation (1) serait la fonction $f(x)\equiv 1$. Evisageons donc les autres cas.

Ad a). Remarquons tout d'abord que dans ce cas nous avons

$$\Omega=[0,2\pi), \tag{40}$$

puisque dans le cas contraire existerait $w_0 \in [0,2\pi)$ tel que $w_0 \bar{\in} \Omega$. Par suite, de la définition de l'ensemble Ω et de (39) s'ensuit que

$$f(r e^{i\omega_0})=0$$

pour tout $r>0$, ce qui est impossible puisque la fonction f est continue et $f(0)=1$.

De (39), (40) et du lemme 12 résulte que $F=C$.

D'autre part, puisque dans ce cas nous avons $A=\{0\}$, donc de la construction K s'ensuit que

$$a(\lambda,\mu)\equiv 0 \quad \text{pour} \quad \lambda,\mu \in C' \overset{\text{df}}{=} C\backslash\{0\}.$$

Nous obtenons en conséquence

$$w(\lambda\mu)=w(\lambda)+\lambda w(\mu) \quad \text{pour} \quad \lambda,\mu \in C'$$

et

$$w: C' \to C.$$

De là, en raisonnant comme dans la démonstration du lemme 13, on a

$$w(\lambda)=(1-\lambda)c,$$

où c est une constante complexe arbitraire, mais (à cause de la condition (6)) différente de zéro. Par suite nous obtenons de la construction K

$$f(x) = \begin{cases} \lambda & \text{pour} \quad x = (1 - \lambda)\, c, \quad \text{et} \quad \lambda \in C \backslash \{0\}, \\ 0 & \text{pour} \quad x = c. \end{cases} \tag{41}$$

Ad b). Dans ce cas nous concluons du lemme 1, que

$$\Omega = \{0\}.$$

Il s'ensuit que

$$F = [0, +\infty).$$

D'autre part nous avons de la construction K

$$w(\lambda\mu) = w(\lambda) + \lambda\, w(\mu) + a(\lambda, \mu) \quad \text{pour} \quad \lambda, \mu \in F' \overset{\text{df}}{=} F \backslash \{0\}, \tag{42}$$

où

$$w: F' \to C,$$

et

$$a: F' \times F' \to (-\infty, +\infty).$$

De (42) nous obtenons

$$\operatorname{Im}[w(\lambda\mu)] = \operatorname{Im}[w(\lambda)] + \lambda \operatorname{Im}[w(\mu)].$$

De là et du lemme 13 nous avons

$$\operatorname{Im}[w(\lambda)] = (1 - \lambda)\, c,$$

où c est une constante arbitraire réelle différente de zéro; par suite

$$w(\lambda) = (1 - \lambda)\, ci + b(\lambda),$$

où

$$b: F' \to (-\infty, +\infty).$$

Remarquons encore que la condition

$$[x - w(\lambda)] \in A$$

est équivalente à la condition

$$\operatorname{Im} x = (1 - \lambda)\, c.$$

De là et de (7) nous tirons la conclusion, que

$$f(x) = \begin{cases} \lambda & \text{s'il existe } \lambda \in F \backslash \{0\} \text{ tel que } \operatorname{Im} x = (1 - \lambda)\, c, \\ 0 & \text{dans le cas contraire.} \end{cases} \tag{43}$$

Ad c). En raisonnant comme dans le cas b), nous abtenons

$$\Omega = \{0\}$$

et ensuite

$$w(\lambda\mu) = w(\lambda) + \lambda\, w(\mu) + a(\lambda, \mu) \quad \text{pour} \quad \lambda, \mu \in F', \tag{44}$$

où, de nouveau,

$$w: F' \to C.$$

Mais maintenant

$$a: F' \times F' \to A = \{z: \operatorname{Re} z = 0\}.$$

De (44) nous concluons que

$$\operatorname{Re}[w(\lambda\mu)] = \operatorname{Re}[w(\lambda)] + \lambda \operatorname{Re}[w(\mu)].$$

De là et du lemme 13 nous avons

$$\operatorname{Re}[w(\lambda)] = (1-\lambda)\, c$$

où c est une constante réelle arbitraire differente de zéro. Donc

$$w(\lambda) = (1-\lambda)\, c + b(\lambda)\, i;$$

b est une fonction arbitraire telle que

$$b: F' \to (-\infty, +\infty).$$

Remarquons encore que les conditions $[x - w(\lambda)] \in A$ et $\operatorname{Re} x = (1-\lambda)\, c$ sont équivalentes. De là et de la construction K nous obtenons

$$f(x) = \begin{cases} \lambda & \text{s'il existe } \lambda \in F\backslash\{0\} \text{ tel que } \operatorname{Re} x = (1-\lambda)\, c, \\ 0 & \text{dans le cas contraire.} \end{cases} \tag{45}$$

En résumant, nous obtenons de (41), (43) et (45)

Théorème 7. *Les seules solutions continues de l'équation* (1) *dans l'ensemble des fonctions complexes de la variable complexe sont les fonctions suivantes*:

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

(b) $f(x) = 1 + cx$, *où c est une constante arbitraire complexe,*

(c) $$f(x) = \begin{cases} 1 + c \cdot \operatorname{Im} x, & si\quad 1 + c \cdot \operatorname{Im} x > 0, \\ 0 & si\quad 1 + c \cdot \operatorname{Im} x \leqslant 0. \end{cases}$$

où c est une constante rèelle arbitraire

(d) $$f(x) = \begin{cases} 1 + c \cdot \operatorname{Re} x, & si\quad 1 + c \operatorname{Re} x > 0, \\ 0 & si\quad 1 + c \operatorname{Re} x \leqslant 0. \end{cases}$$

où c est une constante réelle arbitraire.

TRAVAUX CITÉS

[1] Aczél, J., *Beiträge zur Theorie der geometrischen Objekte*, III–IV, Acta Math. Acad. Sci. Hungar. *8*, 19–52 (1957).

[2] Gołąb, S. et Schinzel, A., *Sur l'équation fonctionelle* $f[x+yf(x)]=f(x)\,f(y)$, Publ. Math. Debrecen *6*, 113–125 (1959).

[3] Popa, C. G., *Sur l'équation fonctionnelle* $f[x+yf(x)]=f(x)f(y)$, Ann. Polon. Math. *17*, 193–198 (1965).

Wyzsza Szkola Ped.

Analytic Solutions of the Equation $\varphi(z) = h(z, \varphi[f(z)])$ with Right Side Contracting

by W. SMAJDOR (Katowice, Poland)

Dedicated to A.M. Ostrowski on the occasion of his 75th birthday

§1. In the present paper we are concerned with the functional equation

$$\varphi(z) = h(z, \varphi[f(z)]), \tag{1}$$

where f and h denote known functions of the type:

$$f: C^n \to C^n;$$
$$h: C^n \times C^m \to C^m$$

and φ of the type:

$$\varphi: C^n \to C^m$$

denotes the required function.

In the previous paper [9] (cf. also [8]) we have proved (under suitable assumptions regarding the functions h and f) the existence and uniqueness of local analytic solutions of equation (1) in the case where the characteristic roots of the matrix $\left[\frac{\partial f_i(z)}{\partial z_j}\bigg|_{z=0}\right]_{i,j\leqslant n}$ have absolute values smaller than 1. In the present paper we weaken this restriction allowing those roots to have absolute values not greater than 1, but assuming instead that the characteristic roots of the matrix $\left[\frac{\partial h^k(z,w)}{\partial w^l}\bigg|_{(z,w)=(0,0)}\right]_{k,l\leqslant m}$ have absolute values smaller than one.

In spaces C^n, C^m and $C^{m\times n}$ we define norms of the elements $z=(z_1,\dots,z_n)$, $w=(w^1,\dots,w^m)$ and $w^{(1)}=[w_i^k]_{k\leqslant m,\, i\leqslant n}$ by formulae

$$\|z\|_n = \left(\sum_{i=1}^{n}|z_i|^2\right)^{1/2}; \quad \|w\|_m = \left(\sum_{k=1}^{m}|w^k|^2\right)^{1/2}; \quad \|w^{(1)}\|_{m\times n} = \left(\sum_{k=1}^{m}\sum_{i=1}^{n}|w_i^k|^2\right)^{1/2}.$$

For a complex matrix $U=[u_l^k]_{k,l\leqslant m}$ we define the norm $\|U\|$ as

$$\|U\| = \sup_{\|w\|_m=1}\|Uw\|_m = \sup_{\|w\|_m=1}\left\{\sum_{k=1}^{m}\left|\sum_{l=1}^{m}u_l^k w^l\right|^2\right\}^{1/2}.$$

We assume the following hypotheses regarding the function h:

(I) The function $h(z,w)=h^1(z_1,\dots,z_n,w^1,\dots,w^m),\dots,h^m(z_1,\dots,z_n,w^1,\dots,w^m))$ is analytic for $\|z\|_n\leqslant r_0$ and $\|w\|_m\leqslant R_0$, and $h(0,0)=0$.

(II)
$$\|[a_{0,l}^k]_{k,l\leqslant m}\| < 1, \quad \text{where} \quad a_{0,l}^k = \frac{\partial h^k(z,w)}{\partial w^l}\bigg|_{(z,w)=(0,0)}, \quad k,l=1,\dots,m.$$

Received July 27, 1967

It follows by virtue of hypothesis (II) and from the continuity of the functions $\frac{\partial h^k(z,w)}{\partial w^l}$, $k, l=1,\dots,m$, that there exist numbers $\theta<1$, $r_1\leqslant r_0$ and $R\leqslant R_0$ such that

$$\left\|\left[\frac{\partial h^k(z,w)}{\partial w^l}\right]_{k,l\leqslant m}\right\|<\theta, \tag{2}$$

for $\|z\|_n\leqslant r_1$ and $\|w\|_m\leqslant R$.

First we shall investigate the existence of a solution of equation (1) for $f(z)=Sz$, i.e. for the equation

$$\varphi(z)=h[z,\varphi(Sz)], \tag{3}$$

where

$$S=\begin{bmatrix} s_1 & 0 & \dots & 0\\ 0 & s_2 & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & s_n\end{bmatrix}. \tag{4}$$

In the sequel we shall assume that

(III) $\qquad |s_i|\leq 1, \quad i=1,\dots,n.$

Let us define the functions $H_i^k(z,w,w^{(1)})$, $k=1,\dots,m$; $i=1,\dots,n$, as:

$$H_i^k(z,w,w^{(1)})=\frac{\partial h^k(z,w)}{\partial z_i}+\sum_{l=1}^{m}\frac{\partial h^k(z,w)}{\partial w^l}w_i^l s_i,$$

and let

$$H(z,w,w^{(1)})=[H_i^k(z,w,w^{(1)})]_{k\leqslant m,\,i\leqslant n}.$$

LEMMA 1. *Suppose that hypotheses* (I)–(III) *are fulfilled. Then for arbitrarily fixed point* $\mathring{w}^{(1)}\in C^{m\times n}$ *and for arbitrarily fixed positive number* ϱ *there exists a constant* L, *independent of* z, *such that the inequality*

$$\|H(z,\hat{w},\hat{w}^{(1)})-H(z,\hat{\hat{w}},\hat{\hat{w}}^{(1)})\|_{m\times n}\leqslant L\,\|\hat{w}-\hat{\hat{w}}\|_m+\theta\,\|\hat{w}^{(1)}-\hat{\hat{w}}^{(1)}\|_{m\times n}$$

holds for $\|z\|_n\leqslant r_1$; $\|\hat{w}\|_m\leqslant R$, $\|\hat{\hat{w}}\|_m\leqslant R$, $\hat{w}^{(1)}$, $\hat{\hat{w}}^{(1)}\in\bar{K}(\mathring{w}^{(1)},\varrho)$, *where*

$$K(\mathring{w}^{(1)},\varrho)=\{w^{(1)}\in C^{m\times n}\colon \|w^{(1)}-\mathring{w}^{(1)}\|_{m\times n}<\varrho\}.$$

Proof. Since the function $H(z,w,w^{(1)})$ is analytic, for arbitrarily fixed point $\mathring{w}^{(1)}\in C^{m\times n}$ and for arbitrarily fixed positive number ϱ there exists a constant L, independent of z and $\hat{\hat{w}}^{(1)}$, such that inequality

$$\|H(z,\hat{w},\hat{\hat{w}}^{(1)})-H(z,\hat{\hat{w}},\hat{\hat{w}}^{(1)})\|_{m\times n}\leqslant L\,\|\hat{w}-\hat{\hat{w}}\|_m \tag{5}$$

holds for $\|z\|_n\leqslant r_1$; $\|\hat{w}\|_m\leqslant R$, $\|\hat{\hat{w}}\|_m\leqslant R$ and $\hat{\hat{w}}^{(1)}\in\bar{K}(\mathring{w}^{(1)},\varrho)$. In virtue of (III) we have the inequality

$$\|H(z, \hat{w}, \hat{w}^{(1)}) - H(z, \hat{w}, \hat{\hat{w}}^{(1)})\|_{m\times n}^2 = \sum_{k=1}^{m}\sum_{i=1}^{n}\left|\sum_{l=1}^{m}\frac{\partial h^k(z, \hat{w})}{\partial w^l} s_i(\hat{w}_i^l - \hat{\hat{w}}_i^l)\right|^2$$

$$\leqslant \left(\sum_{i=1}^{n}\sum_{s=1}^{m}|\hat{w}_i^s - \hat{\hat{w}}_i^s|^2\right)\left\{\sum_{k=1}^{m}\left|\sum_{l=1}^{m}\frac{\partial h^k(z, \hat{w})}{\partial w^l}\frac{\hat{w}_i^l - \hat{\hat{w}}_i^l}{\left(\sum_{s=1}^{m}|\hat{w}_i^s - \hat{\hat{w}}_i^s|^2\right)^{1/2}}\right|^2\right\}.$$

By (2) each $\{\ \}$ term is less than θ^2, and by the definition of the norm of elements $w^{(1)}$ we have

$$\|H(z, \hat{w}, \hat{w}^{(1)}) - H(z, \hat{w}, \hat{\hat{w}}^{(1)})\|_{m\times n}^2 < \theta^2 \|\hat{w}^{(1)} - \hat{\hat{w}}^{(1)}\|_{m\times n}^2 \tag{6}$$

for $\|z\|_n \leqslant r_1$; $\|\hat{w}\|_m \leqslant R$, $\|\hat{\hat{w}}\|_m \leqslant R$ and $\hat{w}^{(1)}, \hat{\hat{w}}^{(1)} \in \bar{K}(\mathring{w}^{(1)}, \varrho)$. Finally we get by (5) and (6) the inequality of the lemma.

§2. Let $\varphi(z)$ be an analytic solution of equation (3) in a neighbourhood of $z=0$ fulfilling $\varphi(0)=0$. Let us put

$$c^{(1)} = [c_i^k]_{k\leqslant m,\, i\leqslant n} = \left[\frac{\partial\varphi^k(z)}{\partial z_i}\bigg|_{z=0}\right]_{k\leqslant m,\, i\leqslant n}, \qquad \varphi'(z) = \left[\frac{\partial\varphi^k(z)}{\partial z_i}\right]_{k\leqslant m,\, i\leqslant n}.$$

Differentiating both sides of relation (3) and then setting $z=0$ we get

$$c^{(1)} = \left[a_{i,0}^k + \sum_{l=1}^{m} a_{0,l}^k c_i^l s_i\right]_{k\leqslant m,\, i\leqslant n}, \tag{7}$$

where $a_{i,0}^k = \dfrac{\partial h^k(z, w)}{\partial z_i}\bigg|_{(z,w)=(0,0)}$.

Formula (7) shows that the value of the derivative of an analytic solution of equation (3) at the origin cannot be quite arbitrary. It must fulfil the system of linear equations (7).

We shall prove the following theorem.

THEOREM 1. *Let hypotheses* (I)–(III) *be fulfilled. Then for every matrix* $c^{(1)}$ *whose elements fulfil system* (7), *there exists exactly one solution* $\varphi(z)$ *of equation* (3) *analytic in a neighbourhood of* $z=0$, *and such that*

$$\varphi(0) = 0, \quad \varphi'(0) = c^{(1)}. \tag{8}$$

Proof. Let K be an arbitrarily chosen positive number. $H(z, 0, c^{(1)})$ is an analytic function in a neighbourhood of $z=0$. Thus there exists a number $r_2 \leqslant r_1$ such that the inequality

$$\|H(z, 0, c^{(1)}) - H(0, 0, c^{(1)})\|_{m\times n} \leqslant \tfrac{1}{2}(1-\theta)K \tag{9}$$

holds for $\|z\|_n \leqslant r_2$.

Now we choose a positive number r in such a manner that the following inequalities hold:

$$r < r_2, \tag{10}$$

$$r(K+\|c^{(1)}\|_{m\times n}) < \min\left(R, \frac{1}{2L}(1-\theta)K\right), \tag{11}$$

$$L\,r+\theta < \vartheta, \tag{12}$$

where

$$\theta < \vartheta < 1.$$

We define the function space $\mathscr{R}$ as the set of the functions $\varphi(z)$ which fulfil the following conditions:

(i) $\varphi(z)$ is analytic for $\|z\|_n<r$ and $\varphi'(z)$ is continuous for $\|z\|_n\leqslant r$;

(ii) $\varphi(0)=0$ and $\varphi'(0)=c^{(1)}$;

(iii) $\|\varphi'(z)-c^{(1)}\|_{m\times n}\leqslant K$ for $\|z\|_n\leqslant r$.

Next we define a metric ϱ in the set $\mathscr{R}$ putting

$$\varrho(\varphi_1,\varphi_2)=\sup_{\|z\|_n\leqslant r}\|\varphi_1'(z)-\varphi_2'(z)\|_{m\times n}, \qquad \varphi_1,\varphi_2\in\mathscr{R}. \tag{13}$$

By the mean-value theorem (cf. [1], ch. 8, § 5) we have for $\varphi_1, \varphi_2\in\mathscr{R}$

$$\|\varphi_1(z)-\varphi_2(z)\|_m\leqslant r\sup_{\|z\|_n\leqslant r}\|\varphi_1'(z)-\varphi_2'(z)\|_{m\times n}.$$

Hence and from the definition of metric (13) we infer

$$\sup_{\|z\|_n\leqslant r}\|\varphi_1(z)-\varphi_2(z)\|_m\leqslant r\varrho(\varphi_1,\varphi_2). \tag{14}$$

It follows from (14) and from the Weierstrass theorem (cf. [2], ch. 1, §4) that the space $\mathscr{R}$ with metric (13) is complete.

Now we consider the transformation $\psi=\Phi[\varphi]$ defined by the formula

$$\psi(z)=h[z,\varphi(Sz)]. \tag{15}$$

We shall show that Φ is a map from $\mathscr{R}$ into $\mathscr{R}$ and that it is a contraction map. Hence, in view of Banach's fixed-point theorem it follows that Φ has a unique invariant point, i.e. that there exists exactly one function φ which is analytic for $\|z\|_n<r$ and satisfies equation (3) and conditions (8).

If $\varphi\in\mathscr{R}$, then by (iii) and by the mean-value theorem we have

$$\|\varphi(z)\|_m\leqslant r\sup_{\|z\|_n\leqslant r}\|\varphi'(z)\|_{m\times n}\leqslant r(K+\|c^{(1)}\|_{m\times n}), \tag{16}$$

whence we get in view of (11)

$$\|\varphi(z)\|_m<R \quad \text{for} \quad \|z\|_n\leqslant r.$$

Thus the function $\psi(z)$ defined by (15) is analytic for $\|z\|_n<r$ and $\psi'(z)$ is continuous for $\|z\|_n \leqslant r$ We get according to the definition of H

$$\psi'(z) = H(z, \varphi(Sz), \varphi'(Sz)),$$

and on the other hand, since the elements of matrix $c^{(1)}$ fulfil system (7)

$$H(0, 0, c^{(1)}) = c^{(1)}.$$

Consequently $\psi(0)=0$ and $\psi'(0)=c^{(1)}$ Thus ψ fulfils conditions (i) and (ii).

It follows from (9), (10) and from Lemma 1 that

$$\begin{aligned}\|\psi'(z) - c^{(1)}\|_{m\times n} &\leqslant \|H(z, \varphi(Sz), \varphi'(Sz)) - H(z, 0, c^{(1)})\|_{m\times n} + \\ &\quad + \|H(z, 0, c^{(1)}) - H(0, 0, c^{(1)})\|_{m\times n} \\ &\leqslant L\,\|\varphi(Sz)\|_m + \theta\,\|\varphi'(Sz) - c^{(1)}\|_{m\times n} + \tfrac{1}{2}(1-\theta)\,K.\end{aligned}$$

By (16) and (iii) for φ we obtain hence

$$\|\psi'(z) - c^{(1)}\|_{m\times n} \leqslant L\,[r(K + \|c^{(1)}\|_{m\times n})] + \theta K + \tfrac{1}{2}(1-\theta)\,K.$$

This together with (11) yields

$$\|\psi'(z) - c^{(1)}\|_{m\times n} < K,$$

i.e. $\psi(z)$ fulfils condition (iii). Thus $\psi\in\mathscr{R}$.

Let us write $\psi_1=\Phi[\varphi_1]$, $\psi_2=\Phi[\varphi_2]$, $\varphi_1, \varphi_2\in\mathscr{R}$. By Lemma 1 we obtain the inequality

$$\begin{aligned}\|\psi_1'(z) - \psi_2'(z)\|_{m\times n} &= \|H(z, \varphi_1(Sz), \varphi_1'(Sz)) - H(z, \varphi_2(Sz), \varphi_2'(Sz))\|_{m\times n} \\ &\leqslant L\,\|\varphi_1(Sz) - \varphi_2(Sz)\|_m + \theta\,\|\varphi_1'(Sz) - \varphi_2'(Sz)\|_{m\times n}.\end{aligned}$$

Making use of (13) and (14) we get

$$\varrho(\psi_1, \psi_2) \leqslant (Lr + \theta)\,\varrho(\varphi_1, \varphi_2),$$

and finally by (12)

$$\varrho(\psi_1, \psi_2) < \vartheta\,\varrho(\varphi_1, \varphi_2).$$

Consequently Φ is a contraction map and the proof of the theorem has been completed.

§3. Let us denote by λ_k, $k=1,\dots,m$, the characteristic roots of the matrix $[a^k_{0,l}]_{k,l\leqslant m}$. We can replace hypothesis (II) by the following hypothesis

(IV) $$\lambda_0 = \max_{k=1,\dots,m} |\lambda_k| < 1.$$

We shall prove the following

LEMMA 2. *Under hypotheses* (III) *and* (IV) *the system of linear equations* (7) *possesses exactly one solution.*

Proof. For fixed i ($i=1,\dots,n$) system (7) has the following form

$$\begin{aligned}
-a_{i,0}^1 &= (a_{0,1}^1 s_i - 1)\, c_i^1 + a_{0,2}^1 c_i^2 s_i + \cdots + a_{0,m}^1 c_i^m s_i\\
-a_{i,0}^2 &= a_{0,1}^2 c_i^1 s_i + (a_{0,2}^2 s_i - 1)\, c_i^2 + \cdots + a_{0,m}^2 c_i^m s_i\\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\
-a_{i,0}^m &= a_{0,1}^m c_i^1 s_i + a_{0,2}^m c_i^2 s_i + \cdots + (a_{0,m}^m s_i - 1)\, c_i^m
\end{aligned}$$

Suppose that the determinant of this system is equal to zero. Then 1 is a characteristic root of the matrix $[s_i a_{0,l}^k]_{k,l\le m}$. Consequently $1/s_i$ ($s_i\ne 0$) is the characteristic root of the matrix $1/s_i[s_i a_{0,l}^k]_{k,l\le m}=[a_{0,l}^k]_{k,l\le m}$. Since $1/|s_i|\ge 1$, $1/s_i$ cannot be a characteristic root of the matrix $[a_{0,l}^k]_{k,l\le m}$. Thus the determinant of the above system cannot vanish and system (7) has exactly one solution.

THEOREM 2. *Let hypotheses* (I), (III) and (IV) *be fulfilled, and let the elements of the matrix* $c^{(1)}$ *fulfil system* (7)[1]. *Then there exists exactly one solution* $\varphi(z)$ *of equation* (3) *analytic in a neighbourhood of* $z=0$ *and such that* $\varphi(0)=0$.

Proof. It is known (cf. [5]) that for every $\varepsilon>0$ there exists a non-singular matrix $T=[t_l^k]_{k,l\le m}$ such that $\|T^{-1}[a_{0,l}^k]_{k,l\le m}T\|\le\lambda_0+\varepsilon$. We choose ε so small that $\lambda_0+\varepsilon<1$. The function $g(w)=Tw$ is analytic and maps the space C^m onto itself. We have for $\hat{h}(z,w)=g^{-1}\{h[z,g(w)]\}$ defined in the set $\bar{K}(0,r_0)\times g^{-1}[\bar{K}(0,R_0)]$:

$$\left[\frac{\partial\hat{h}^k(z,w)}{\partial w^l}\right]_{k,l\le m}=T^{-1}\left[\frac{\partial h^k(z,w)}{\partial w^l}\right]_{k,l\le m}T.$$

If $\varphi(z)$ fulfils equation (3) then $\psi(z)=g^{-1}[\varphi(z)]$ fulfils equation

$$\psi(z)=\hat{h}[z,\psi(Sz)] \tag{17}$$

and conversely. Thus, the fact that equation (3) has a solution is equivalent to the fact that equation (17) has a solution. The function $\hat{h}(z,w)$ fulfils hypotheses (I) and (II). Let

$$d^{(1)}=[d_i^k]_{k\le m,\,i\le n}=T^{-1}c^{(1)}.$$

If elements of the matrix $c^{(1)}$ fulfil system (7), then elements of the matrix $d^{(1)}$ fulfil the system

$$d^{(1)}=\left[\hat{a}_{i,0}^k+\sum_{l=1}^m\hat{a}_{0,l}^k d_i^l\right]_{k\le m,\,i\le n},$$

where

$$\hat{a}_{i,0}^k=\left.\frac{\partial\hat{h}^k(z,w)}{\partial z^i}\right|_{(z,w)=(0,0)},\quad \hat{a}_{0,l}^k=\left.\frac{\partial\hat{h}^k(z,w)}{\partial w^l}\right|_{(z,w)=(0,0)}.$$

Thus, all hypotheses of Theorem 1 are fulfilled. Consequently equation (17) possesses

[1]) It follows from Lemma 2 that there exists exactly one $c^{(1)}$.

exactly one analytic solution defined in a neighbourhood of $z=0$ and such that $\psi(0)=0$ and $\psi'(0)=d^{(1)}$. Let $\varphi(z)=g[\psi(z)]$. The function φ so defined is the unique analytic solution of equation (3) in a neighbourhood of $z=0$ fulfilling the condition $\varphi(0)=0$.

§4. Let us admit the new hypotheses:

(V) The function $f(z)$ of the type

$$f: C^n \to C^n$$

is analytic in a neighbourhood of $z=0, f(0)=0$.

In the sequel s_i will denote the characteristic roots of the matrix $\left[\frac{\partial f_i}{\partial z_j}\right]_{z=0}{}_{i,j\leq n}$ and S will be matrix (4).

(VI) There exists an analytic solution $\gamma(z)$ in a neighbourhood of $z=0$ of the equation

$$\gamma[f(z)] = S\gamma(z)$$

such that $\gamma(0)=0$ and $\gamma'(0)$ is a non-singular matrix.

In the case $n=1$, if $|s|<1$, hypothesis (VI) is always fulfilled (cf. [3]). In virtue of C. L. SIEGEL's theorem (cf. [6]) this hypothesis is also fulfilled for almost all s such that $|s|=1$. (For $n>1$, cf. [4]).

THEOREM 3. *Suppose that hypotheses* (I) *and* (III)–(VI) *are fulfilled. Then there exists exactly one analytic solution* $\varphi(z)$ *of equation* (1) *such that* $\varphi(0)=0$.

Proof. It follows from hypothesis (VI) that there exists a function γ which is analytic and invertible in a neighbourhood of $z=0$ and such that

$$S\,z = \gamma\{f[\gamma^{-1}(z)]\}. \tag{18}$$

If $\varphi(z)$ satisfies equation (1) then $\psi(z)=\varphi[\gamma^{-1}(z)]$ satisfies the equation

$$\psi(z) = \hat{h}(z, \psi\{\gamma[f(\gamma^{-1}(z))]\}), \tag{19}$$

where $\hat{h}(z, w)=h[\gamma^{-1}(z), w]$, and conversely. Thus functional equations (1) and (19) are equivalent. In view of (18) the latter can be written in the form

$$\psi(z) = \hat{h}[z, \psi(S\,z)]. \tag{20}$$

Since the hypotheses of Theorem 2 are fulfilled, there exists exactly one analytic solution $\psi(z)$ of equation (20) such that $\psi(0)=0$. The function $\varphi(z)=\psi[\gamma(z)]$ is the unique analytic solution of equation (1) fulfilling the condition $\varphi(0)=0$. This completes the proof.

§5. In this section we shall formulate a theorem on the existence of a solution of the linear equation

$$\varphi[f(z)] = g(z)\,\varphi(z) + F(z) \tag{21}$$

in the case $n=m=1$.

THEOREM 4. *Let the functions f, g and F be analytic in a neighbourhood of $z=0$ and let $f(0)=F(0)=0$, $|g(0)|\neq 1$ and $f'(0)=s$ with $|s|=1$. There exists a set A of linear Lebesgue measure 2π, contained in the circle $|z|=1$ such that if $s\in A$, then equation* (21) *possesses exactly one analytic solution $\varphi(z)$ in a neighbourhood of $z=0$ such that $\varphi(0)=0$.*

Proof. Let us assume that $|g(0)|>1$ and let

$$h(z, w) = \frac{w - F(z)}{g(z)}.$$

The function h so defined is analytic in a neighbourhood of $(z, w)=(0, 0)$, $h(0, 0)=0$ and $|h'_w(0, 0)|<1$. In virtue of SIEGEL's theorem (cf. [6]) there exists a set A of linear Lebesgue measure 2π contained in the circle $|z|=1$ such that if $s\in A$, then the Schröder equation

$$\gamma[f(z)] = s\gamma(z)$$

has exactly one analytic solution in a neighbourhood of $z=0$ such that $\gamma(0)=0$ and $\gamma'(0)=1$. Thus on account of Theorem 3 the equation

$$\frac{\varphi[f(z)] - F(z)}{g(z)} = \varphi(z)$$

has exactly one analytic solution such that $\varphi(0)=0$. This equation is equivalent to (21).

If $|g(0)|<1$, then putting $h(z, w)=g[f^{-1}(z)]w+F[f^{-1}(z)]$ and replacing $f(z)$ by $f^{-1}(z)$, the previous argument applies. Thus, the proof is finished.

It follows from the last theorem and paper [7] that the non-homogeneous linear equation (21) possesses a local analytic solution for almost all s such that $|s|=1$, where $|g(0)|<1$, or $|g(0)|>1$ or $g(0)=s^k$, $k=0, 1, 2, \ldots$. The case, where $|s|=1$, $|g(0)|=1$ and $g(0)\neq s^k$, (k an integer) is still left undecided.

REFERENCES

[1] DIEUDONNÉ, J., *Foundations of Modern Ananlysis* (Academic Press, New York–London 1960).

[2] HERVÉ, M., *Several Complex Variables*, Local Theory (Tata Inst. of Fundamental Research, Bombay; Oxford University Press, London 1963).

[3] KOENIGS, G., *Recherches sur les intégrales de certaines équations fonctionelles*, Ann. Sci. Ecole Norm. Sup. (*3*)*1*, Supplement, 3–41 (1884).

[4] MONTEL, P., *Lecons sur les recurrences et leurs applications*, Chapter 4 (Gauthier-Villars, 1957).

[5] OSTROWSKI, A. M., *Solution of Equations and Systems of Equations* (Academic Press, New York–London 1960).
[6] SIEGEL, C.L., *Iteration of analytic functions*, Ann. of Math. *(2)43*, 607–612 (1942).
[7] SMAJDOR, A. and SMAJDOR, W., *On the Existence and Uniqueness of Analytic Solutions of a Linear Functional Equation*, Math. Z. *98*, 235–242 (1967).
[8] SMAJDOR, W., *The Existence and Uniqueness of Analytic Solutions of the Functional Equation* $\phi(z-h(z, \phi[f(z)])$, Ann. Polon. Math. *19*, 37–45 (1967).
[9] SMAJDOR, W., *Local Analytic Solutions of the Functional Equation* $\phi(z)=h(z, \phi[f(z)])$ *in Multi-Dimensional Spaces*, Aequationes Math. 1, 20–36 (1968).

Math. Inst. of the Polish Acad. Sce.

An $0(h^2)$ Method for a Non-Smooth Boundary Value Problem

PHILIPPE G. CIARLET (Gien, France)

Dedicated to A.M. Ostrowski on the occasion of his 75th birthday

§ 1. Introduction

Consider the linear boundary value problem

$$u'' = p(x)\,u + q(x), \qquad 0 < x < 1, \tag{1.1}$$

$$u(0) = \alpha, \qquad u(1) = \beta. \tag{1.2}$$

Given a uniform partition of the interval [0, 1], of mesh size h, the standard three-point difference scheme as applied to (1.1)–(1.2) is known to yield an $0(h^2)$ accuracy at the mesh points. The argument, going back to GERSCHGORIN [3], assumes that $p(x)\geqslant 0$ on [0, 1] and that the solution $\phi(x)$ is a C^4-function on [0, 1]. This result can be easily extended to the case where $\phi(x)$ is a C^3-function, and $\phi'''(x)$ satisfies a Lipschitz condition on [0, 1]. However, the convergence will be only $0(h)$ (Resp. $o(1)$), if $\phi(x)$ is only a C^3-function (Resp. a C^2-function) (for details, see for example [8, Chapter 6]). The proofs generally use discrete Maximum Principle arguments, as is described in [9]. This is why a condition such as $p(x)\geqslant 0$ is required. However, this condition can easily be relaxed to $p(x)\geqslant\gamma > -8$, for some constant $\gamma\leqslant 0$ (cf. [5]). Likewise, a quasilinearity can be introduced in the right-hand side of (1.1) (cf. [2, 5]): the two-point boundary value problem

$$u'' = f(x, u), \qquad 0 < x < 1, \tag{1.3}$$

$$u(0) = \alpha, \qquad u(1) = \beta, \tag{1.4}$$

has a unique solution $\phi(x)$, whenever $\partial f/\partial u\,(x, u)$ exists and is bounded below by a constant $\gamma > -\pi^2$, for $0\leqslant x\leqslant 1$, and all real u. Again, if $\phi(x)$ is a C^4-function, the three-point difference scheme will yield an $0(h^2)$ accuracy. For such problems, a Maximum Principle is still valid when $\gamma > -8$ (if $-8\geqslant\gamma > -\pi^2$, the proofs are more refined and they use Energy-type arguments). Anyway, the error analysis for the three-point difference scheme is subject to the same limitations as in the linear case, depending upon the smoothness of $\phi(x)$.

In this paper, we consider a boundary value problem such as (1.3)–(1.4), and we assume the following:

$$f(x, u)\in C^0([0, 1]\times R), \tag{1.5}$$

Received September 28, 1967

there exists a constant γ such that

$$\left.\begin{aligned} &-\pi^2 < \gamma \leqslant \frac{f(x, u_1) - f(x, u_2)}{u_1 - u_2}, \\ &\qquad \text{for all} \quad 0 \leqslant x \leqslant 1, \quad -\infty < u_1, \quad u_2 < +\infty, \quad u_1 \neq u_2, \end{aligned}\right\} \tag{1.6}$$

for every positive number c, there exists a finite number $M(c)$ such that

$$\left.\begin{aligned} &\frac{f(x, u_1) - f(x, u_2)}{u_1 - u_2} \leqslant M(c), \\ &\qquad \text{for all} \quad 0 \leqslant x \leqslant 1, \quad |u_1| \leqslant c, \quad |u_2| \leqslant c, \quad u_1 \neq u_2. \end{aligned}\right\} \tag{1.7}$$

The boundary value problem (1.3)–(1.4), subject to the hypotheses of (1.5)–(1.6)–(1.7) has a unique solution $\phi(x)$ (cf. [2]). However, the solution might be *only a C^2-function.* In that case, all that has been shown by matrix methods is that the three-point difference scheme will yield a $o(1)$ error (i.e. uniform convergence as $h \to 0$). We will show in Section 2 (Theorem 1) how a variational method using piecewise linear approximations yields an $0(h^2)$ (Resp. $0(h^{3/2})$) error when $\gamma > -8$ (Resp. $-8 \geqslant \gamma > -\pi^2$) and finally a numerical example will be examined in Section 3.

The results we present here are part of [1].

§ 2. Proof of the Main Result (Theorem 1)

Denote by (P) the boundary-value problem of (1.3)–(1.4) subject to the hypotheses of (1.5)–(1.6)–(1.7). We also assume, without loss of generality, that the constants α, β of (1.4) are both set to be zero and that the constant γ of (1.6) is chosen once and for all to be *nonpositive.*

Let S denote the linear space of all real functions $w(x)$ satisfying $w(0) = w(1) = 0$, such that $w(x)$ is absolutely continuous in [0, 1] and $w'(x) \in L^2[0, 1]$, that is, S coincides with the well-known Sobolev space $W_0^{1,2}$ [0, 1]. Solving (P) is then equivalent to minimizing the following functional

$$F[w] = \int_0^1 \left\{ \tfrac{1}{2} [w'(x)]^2 + \int_0^{w(x)} f(x, \eta)\, d\eta \right\} dx \tag{2.1}$$

over the space S, and the discrete problem (P_N) consists in minimizing the functional $F[w]$ of (2.1) over appropriate finite dimensional subspaces S_N of S (for details, we refer the reader to [2]).

In this paper, we study the effect of minimizing the functional $F[w]$ over subspaces of piecewise linear functions. In so doing, we consider the special case $m = 1$ of the *smooth Hermite spaces* $H_0^{(m)}(\pi)$, defined for any positive integer m, which have been already considered in [2].

More precisely, given any positive integer N, let $\pi:0=x_0<\cdots<x_{N+1}=1$ denote a uniform partition of the unit interval with joints $x_i=ih$, $h=1/(N+1)$, $0\leqslant i\leqslant N+1$. Then the subspace $H_0^{(1)}(\pi)$ denotes the collection of all continuous functions $w(x)$ defined on $[0, 1]$ which are linear on each subinterval $[x_i, x_{i+1}]$, $0\leqslant i\leqslant N$, and which vanish at the end points. It is readily seen that the vector space $H_0^{(1)}(\pi)$ is a subspace of dimension N of S and, for brevity, it will be denoted by S_N in the sequel. A convenient basis in S_N can be constructed as follows: let $w_i(x)$, $1\leqslant i\leqslant N$, be the (unique) element in S_N which satisfies $w_i(x_j)=\delta_{ij}$, $1\leqslant j\leqslant N$. Then, the collection $\{w_i(x), 1\leqslant i\leqslant N\}$ constitutes a basis in S_N: any function $w(x)\in S_N$ can be written as

$$w(x)=\sum_{i=1}^{N} u_i w_i(x). \tag{2.2}$$

In [2], it was proved that the discrete problem (P_N) has one and only one solution $\hat{w}(x)=\sum_{i=1}^{N}\hat{u}_i w_i(x)$ in any finite dimensional subspace S_N of S, the unknowns $\hat{u}_i$, $1\leqslant i\leqslant N$, satisfying identically the following nonlinear system of equations:

$$\sum_{j=1}^{N}\left\{\int_0^1 w_i'(x)\,w_j'(x)\,dx\right\}\hat{u}_j+\int_0^1 w_i(x)f\left(x,\sum_{j=1}^{N}\hat{u}_j w_j(x)\right)dx=0,\qquad 1\leqslant i\leqslant N, \tag{2.3}$$

which simply represent the M equations $\partial F/\partial u_i[\sum_{j=1}^{N}\hat{u}_j w_j(x)]=0$, $1\leqslant i\leqslant N$. Such a system can be easily solved by standard numerical techniques (cf. [7]), and we again refer to [2] for details involved in its numerical solution. The error analysis developed in [2] for the subspaces $H_0^{(m)}(\pi)$ in general yielded an error estimate of the form

$$\|\phi-\hat{w}_N\|_\infty=\operatorname*{Max}_{0\leqslant x\leqslant 1}|\phi(x)-\hat{w}_N(x)|=0(h),$$

for the special case $m=1$.

We now present the following improvement, which had been conjectured in [10]:

THEOREM 1. With the assumptions of (1.5)–(1.6)–(1.7), we have

$$\|\phi-\hat{w}_N\|_\infty=0(h^{3/2}). \tag{2.4}$$

If the constant γ of (1.6) satisfies moreover

$$\gamma>-8, \tag{2.5}$$

then we have

$$\|\phi-\hat{w}_N\|_\infty=0(h^2),\qquad \textit{for sufficiently small } h. \tag{2.6}$$

The proof of Theorem 1 is rather lengthy and involves several intermediate steps.

First, given the solution $\phi(x)$, we define its S_N-interpolate $\tilde{w}(x)$ to be the unique function in S_N with the property:

$$\tilde{w}(x_i) = \phi(x_i), \quad 1 \leqslant i \leqslant N.$$

Also, let us note the following formulas which will be useful later on:

$$\int_0^1 w_i'(x)\, w_j'(x)\, dx = \left\{ \begin{array}{ll} \dfrac{2}{h}, & i = j, \\ -\dfrac{1}{h}, & |i-j| = 1, \\ 0, & |i-j| \geqslant 2, \quad 1 \leqslant i, j \leqslant N, \end{array} \right\} \tag{2.7}$$

$$\int_0^1 w_i(x)\, w_j(x)\, dx = \left\{ \begin{array}{ll} \dfrac{2h}{3}, & i = j, \\ \dfrac{h}{6}, & |i-j| = 1, \\ 0, & |i-j| \geqslant 2, \quad 1 \leqslant i, j \leqslant N. \end{array} \right\} \tag{2.8}$$

The central idea in the proof of Theorem 1 will be to compare the two functions $\tilde{w}(x)$ and $\hat{w}(x)$. Since $\hat{w}(x) = \sum_{i=1}^{N} \hat{u}_i w_i(x)$ satisfies the N equations (2.3), we are led to let

$$k_i = \sum_{j=1}^{N} \left\{ \int_0^1 w_i'(x)\, w_j'(x)\, dx \right\} \tilde{u}_j - \int_0^1 w_i(x) f(x, \sum_{j=1}^{N} \tilde{u}_j w_j(x))\, dx, \quad 1 \leqslant i \leqslant N, \tag{2.9}$$

the $\tilde{u}_j$'s being the coefficients of the expansion of $\tilde{w}(x)$ over the basis $\{w_i(x), 1 \leqslant i \leqslant N\}$, and we will compute upper bounds for the k_i's (Inequalities (2.13)).

An integration by parts gives

$$\int_0^1 w_i'(x)\, \phi'(x)\, dx + \int_0^1 w_i(x) f(x, \phi(x))\, dx = 0, \quad 1 \leqslant i \leqslant N. \tag{2.10}$$

Hence, by subtracting (2.10) from (2.9), we obtain

$$\left. \begin{array}{l} k_i = \displaystyle\int_0^1 w_i'(x)\, \{\tilde{w}'(x) - \phi'(x)\}\, dx \\ \qquad + \displaystyle\int_0^1 p(x)\, w_i(x)\, \{\tilde{w}(x) - \phi(x)\}\, dx, \quad 1 \leqslant i \leqslant N, \end{array} \right\} \tag{2.11}$$

where the function $p(x)$ is defined as follows:

$$p(x)=\begin{cases}\dfrac{f(x,\tilde{w}(x))-f(x,\phi(x))}{\tilde{w}(x)-\phi(x)}, & 0\leqslant x\leqslant 1, \quad \tilde{w}(x)\neq\phi(x),\\ 0, & 0\leqslant x\leqslant 1, \quad \tilde{w}(x)=\phi(x).\end{cases}$$

In [2], it was proved that an a priori bound for $\|\phi\|_\infty$ (valid also for $\|\hat{w}\|_\infty$, a fact that we will use in proving inequality (2.15)) can be computed. Since $\|\tilde{w}\|_\infty\leqslant\|\phi\|_\infty$, it follows from (1.6) and (1.7) that there exists a constant P independent of N such that

$$-\pi^2<\gamma\leqslant p(x)\leqslant P, \quad 0\leqslant x\leqslant 1. \tag{2.12}$$

Next, an integration by parts on each interval $[x_i, x_{i+1}]$, $0\leqslant i\leqslant N$, shows that $\int_0^1 w_i'(x)\{\tilde{w}'(x)-\phi'(x)\}\,dx=0$, $1\leqslant i\leqslant N$, by definition of the S_N-interpolate $\tilde{w}(x)$. This fact combined with the inequalities of (2.12) shows that $|k_i|\leqslant C\int_0^1 w_i(x)|\tilde{w}(x)-\phi(x)|\,dx$, $1\leqslant i\leqslant N$, where $C=\operatorname{Max}\{|P|,|\gamma|\}$. Finally, using a standard result in interpolation theory (cf. [2]), it follows that

$$|\tilde{w}(x)-\phi(x)|\leqslant\frac{M_2}{2}(x-x_i)(x_{i+1}-x), \quad x_i\leqslant x\leqslant x_{i+1}, \quad 0\leqslant i\leqslant N,$$

where $M_2=\|\phi''\|_\infty$. Thus from the definition of the basis functions $w_i(x)$, we obtain the estimates

$$|k_i|\leqslant Kh^3, \quad 1\leqslant i\leqslant N, \tag{2.13}$$

where $K=CM_2/12$.

Likewise, subtracting equations (2.3) from equations (2.9), and letting $v_i=\tilde{u}_i-\hat{u}_i$, $1\leqslant i\leqslant N$, we obtain

$$\left.\begin{aligned}&\sum_{j=1}^N\left\{\int_0^1 w_i'(x)\,w_j'(x)\,dx\right\}v_j\\ &\qquad+\sum_{j=1}^N\left\{\int_0^1 q(x)\,w_i(x)\,w_j(x)\,dx\right\}v_j=k_i, \quad 1\leqslant i\leqslant N,\end{aligned}\right\} \tag{2.14}$$

where again (as in (2.12)), there exists an explicitly computable bound Q independent of N such that the function $q(x)$ satisfies

$$-\pi^2<\gamma\leqslant q(x)\leqslant Q, \quad 0\leqslant x\leqslant 1. \tag{2.15}$$

In other words, the vector $\mathbf{v}=\tilde{\mathbf{u}}-\hat{\mathbf{u}}$ satisfies the matrix equation

$$A\mathbf{v}=\mathbf{k}, \tag{2.16}$$

where

$$A = (a_{ij}), \tag{2.17}$$

and

$$a_{ij} = \int_0^1 \{w_i'(x)\, w_j'(x) + q(x)\, w_i(x)\, w_j(x)\}\, dx, \quad 1 \leqslant i, \quad j \leqslant N. \tag{2.18}$$

The matrix A of (2.17) is clearly real and symmetric. By using the Rayleigh-Ritz Inequality $\pi^2 \int_0^1 [w(x)]^2\, dx \leqslant \int_0^1 [w'(x)]^2\, dx$, valid for any function $w(x) \in S$, it then follows that the matrix A is positive definite (incidentally, this is the reason why we imposed that the constant γ of (1.6) be bounded below away from $-\pi^2$).

At this point, we remark that if $f(x, u)$ is a function of x only, we obtain $\mathbf{k} = \mathbf{o}$. Since the matrix A is positive definite, we conclude that $\mathbf{v} = \mathbf{o}$. In other words, $\hat{u}_i = \tilde{u}_i = \phi(x_i)$, $1 \leqslant i \leqslant N$, which proves that the ordinates at the mesh points are given exactly, a fact that has been already observed in [2] and [6].

Going back to the matrix equation of (2.16), it follows from a well-known property of symmetric and positive definite matrices that

$$\|\mathbf{v}\|_2 \leqslant \frac{\|\mathbf{k}\|_2}{\lambda^1}, \tag{2.19}$$

where $\|\mathbf{u}\|_2 = \{\sum_{i=1}^N |u_i|^2\}^{1/2}$ is the classical Euclidean vector norm, and λ_1 is the smallest eigenvalue of A. Since we are interested in finding an upper bound for $\|\mathbf{v}\|_2$, we look for a lower bound on λ_1, for which we will use the characterization

$$\lambda_1 = \operatorname*{Inf}_{\sum_{i=1}^N y_i^2 = 1} \{\sum_{i,j=1}^N y_i a_{ij} y_j\}.$$

From the Rayleigh-Ritz inequality mentioned above, it follows that

$$\sum_{i,j=1}^N y_i a_{ij} y_j = \int_0^1 \left\{ \left[\sum_{j=1}^N y_j w_j'(x)\right]^2 + q(x) \left[\sum_{j=1}^N y_j w_j(x)\right]^2 \right\} dx$$

$$\geqslant \left(1 + \gamma/\pi^2\right) \int_0^1 \left[\sum_{j=1}^N y_j w_j'(x)\right]^2 dx.$$

Using formulas (2.7), we recognize that the expression

$$\frac{1}{h} \int_0^1 \left[\sum_{j=1}^N y_j w_j'(x)\right]^2 dx$$

is the quadratic form associated with the familiar matrix

$$A_0 = \frac{1}{h^2}\begin{vmatrix} 2 & -1 & & & \bigcirc \\ -1 & 2 & -1 & & \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \cdot \\ & & \cdot & \cdot & \cdot \\ \bigcirc & & -1 & 2 & -1 \\ & & & -1 & 2 \end{vmatrix}, \tag{2.20}$$

whose smallest eigenvalue $4/h^2 \sin^2 \pi h/2$ is bounded below by 4 for all h (cf. [8]). Thus,

$$\int_0^1 \left[\sum_{j=1}^{N} y_j w_j'(x)\right]^2 dx \geqslant 4h \sum_{j=1}^{N} y_j^2,$$

and hence $\lambda_1 \geqslant \varrho h$, where $\varrho = 4(1+\gamma/\pi^2)$ is independent of h.

Next, the estimates of (2.13) yield that

$$\|\mathbf{k}\|_2 = \left\{\sum_{i=1}^{N} |k_i|^2\right\}^{1/2} \leqslant K(N h^6)^{1/2} \leqslant K h^{5/2}, \quad \text{since} \quad N(h+1) = 1.$$

Thus, going back to the inequality of (2.19), we deduce that

$$\|\mathbf{v}\|_2 = \left\{\sum_{i=1}^{N} |v_i|^2\right\}^{1/2} \leqslant \frac{K}{\varrho} h^{3/2}. \tag{2.21}$$

As a consequence, we also have

$$\max_{1 \leqslant i \leqslant N} |v_i| \leqslant \frac{K}{\varrho} h^{3/2}, \tag{2.22}$$

and from the definition of the vector $\mathbf{v} = \tilde{\mathbf{u}} - \hat{\mathbf{u}}$, it finally follows that

$$\max_{1 \leqslant i \leqslant N} |\phi(x_i) - \hat{w}(x_i)| = 0(h^{3/2}). \tag{2.23}$$

It then follows that the error is $0(h^{3/2})$ *uniformly* on [0, 1], for let $x \in (x_i, x_{i+1})$ for some $0 \leqslant i \leqslant N$ (if $x = x_i$ or x_{i+1}, apply (2.23) directly). We have:

$$|\phi(x) - \hat{w}(x)| \leqslant |\phi(x) - \tilde{w}(x)| + |\tilde{w}(x) - \hat{w}(x)|,$$

and we have seen in deriving inequalities (2.13) that

$$|\phi(x) - \tilde{w}(x)| \leqslant \frac{M_2}{8} h^2, \quad \text{where} \quad M_2 = \|\phi''\|_\infty.$$

On the other hand, since both $\hat{w}(x)$ and $\tilde{w}(x)$ are linear in $[x_i, x_{i+1}]$, it follows that

$$|\tilde{w}(x) - \hat{w}(x)| \leqslant \mathrm{Max}\{|\phi(x_i) - \hat{w}(x_i)|, |\phi(x_{i+1}) - \hat{w}(x_{i+1})|\} = 0(h^{3/2}),$$

by using (2.23), which proves our assertion.

Hence, the first part (Equation (2.4)) of Theorem 1 is proved.'

We proceed now to show that with the further assumption $\gamma > -8$ of (2.5), the error is actually $0(h^2)$. Our method of proof will be based upon proving (Lemma 1) a Maximum Principle satisfied by the matrix A of (2.17). We need first recall a few definitions and results from matrix theory:

Let $A=(a_{ij})$ and $B=(b_{ij})$ be two $N \times N$ real matrices. Then $A \geqslant B$ if and only if $a_{ij} \geqslant b_{ij}$ for all $1 \leqslant i, j \leqslant N$.

Given any N-vector $\mathbf{v}=\{v_1, v_2, \ldots, v_N\}$, we define the norm $\|\mathbf{v}\|_\infty = \mathrm{Max}_{1 \leqslant i \leqslant n} |v_i|$. Given any $N \times N$ matrix $A=(a_{ij})$, the corresponding matrix norm $\|A\|_\infty = \mathrm{Sup}_{\|\mathbf{x}\|_\infty = 1} \|A\mathbf{x}\|_\infty$ is then $\|A\|_\infty = \mathrm{Max}_{1 \leqslant i \leqslant N} \sum_{j=1}^{N} |a_{ij}|$ (cf. [8, p. 15]).

A real and symmetric $N \times N$ matrix $A=(a_{ij})$ with $a_{ij} \leqslant 0$ for all $i \neq j$ is a *Stieltjes matrix* if and only if A is positive definite, or equivalently if its inverse is a non-negative matrix (i.e. $A^{-1} \geqslant 0$). For details, see [8, p. 85–87].

We now prove

LEMMA 1. *Let A be the matrix of* (2.17)–(2.18) *and let A_0 be the matrix of* (2.20). *Then,* for h sufficiently small, *we can decompose the matrix $h^{-1}A$ as*

$$h^{-1}A = A_1 + A_2, \tag{2.24}$$

where A_1 is a Stieltjes matrix which dominates A_0 (i.e. $A_1 \geqslant A_0$).

Moreover the following estimates hold

$$\|A_1^{-1}\|_\infty \leqslant \tfrac{1}{8}, \tag{2.25}$$

and

$$\|A_2\|_\infty \leqslant -\gamma < 8. \tag{2.26}$$

Proof: Write $q(x)$ as $q(x)=q^+(x)+q^-(x)$, where $q^+(x)=\mathrm{Max}\{q(x), 0\} \geqslant 0$, and $0 \geqslant q^-(x) = \mathrm{Min}\{q(x), 0\} \geqslant \gamma > -8$. Accordingly, we write the coefficients a_{ij} (Equations (2.18)) of the matrix A as

$$a_{ij} = \overset{+}{a}_{ij} + \overset{-}{a}_{ij}, \quad \text{where}$$

$$\overset{+}{a}_{ij} = \int_0^1 \{w_i'(x)\, w_j'(x) + q^+(x)\, w_i(x)\, w_j(x)\}\, dx, \quad \text{and}$$

$$\overset{-}{a}_{ij} = \int_0^1 q^-(x)\, w_i(x)\, w_j(x)\, dx.$$

Hence we can write $h^{-1}A = A_1 + A_2$ (Equation (2.24)) by letting $A_1 = (h^{-1}a_{ij}^{+})$ and $A_2 = (h^{-1}a_{ij}^{-})$. Using formulas (2.7)–(2.8), the coefficients of those two matrices can be written as

$$(A_1)_{ij} = \frac{1}{h^2}\begin{cases} 2 + \frac{2h^2}{3} q_{ii}^{+}, & i = j, \\ -1 + \frac{h^2}{6} q_{ij}^{+}, & |i-j| = 1, \\ 0, & |i-j| \geqslant 2, \quad 1 \leqslant i, j \leqslant N, \end{cases} \tag{2.27}$$

and

$$(A_2)_{ij} = \begin{cases} \frac{2}{3} q_{ii}^{-}, & i = j, \\ \frac{1}{6} q_{ij}^{-}, & |i-j| = 1, \\ 0, & |i-j| \geqslant 2, \quad 1 \leqslant i, j \leqslant N, \end{cases} \tag{2.28}$$

where (Q being the constant of the inequality of (2.15))

$$0 \leqslant q_{ij}^{+} \leqslant Q, \quad q_{ij}^{+} = q_{ji}^{+}, \quad 1 \leqslant i, j \leqslant N, \tag{2.29}$$

and

$$-8 < \gamma \leqslant q_{ij}^{-} \leqslant 0, \quad q_{ij}^{-} = q_{ij}^{-}, \quad 1 \leqslant i, j \leqslant N. \tag{2.30}$$

Clearly, $\|A_2\|_\infty \leqslant -\gamma < 8$, proving (2.26), and $A_1 \geqslant A_0$.

We now prove that $A_1^{-1} \geqslant 0$; first, A_1 is a real matrix with $(A_1)_{ii} > 0$, $1 \leqslant i \leqslant N$, and $(A_1)_{ij} \leqslant 0$, $1 \leqslant i, j \leqslant N$, $i \neq j$, provided h is small enough. Next A_1 is irreducibly diagonally dominant (cf. [8, p. 23]) for sufficiently small h. This, coupled with the above mentioned sign pattern implies that $A_1^{-1} \geqslant 0$ (even > 0), proving that A_1 is a Stieltjes matrix (cf. [8, p. 85]). From the fact that $A_1 \geqslant A_0$, we next deduce that $A_0^{-1} \geqslant A_1^{-1} \geqslant 0$ (cf. [8, p. 87]), and also that $\|A_1^{-1}\|_\infty \leqslant \|A_0^{-1}\|_\infty \leqslant \frac{1}{8}$ (cf. [8, p. 171]), which achieves the proof of Lemma 1.

We finally need an upper bound for $\|A^{-1}\|_\infty$. To obtain it, we use a classical procedure (cf. [4, p. 55]): by Lemma 1, we know that $\|A_1^{-1}A_2\|_\infty \leqslant -\gamma/8 < 1$. We write A^{-1} as $A^{-1} = h^{-1}(1 + A_1^{-1}A_2)^{-1}A_1^{-1}$, and we use the estimate

$$\|(1 + A_1^{-1}A_2)^{-1}\|_\infty \leqslant \frac{1}{1 - \|A_1^{-1}A_2\|_\infty} \leqslant \frac{8}{8 + \gamma},$$

which gives

$$\|A^{-1}\|_\infty \leqslant \frac{1}{h(8 + \gamma)}. \tag{2.31}$$

Returning to the matrix equation $A\mathbf{v} = \mathbf{k}$ of (2.16), combined with the estimates of (2.13), we finally obtain

$$\|\mathbf{v}\|_\infty = \max_{1 \leqslant i \leqslant N} |v_i| \leqslant \|A^{-1}\|_\infty \|\mathbf{k}\|_\infty \leqslant \frac{K}{8 + \gamma} h^2. \tag{2.32}$$

Hence, $\max_{1 \leqslant i \leqslant N} |\phi(x_i) - \hat{w}(x_i)| = 0(h^2)$, and the error is thus $0(h^2)$ uniformly on [0, 1], which achieves the proof of Theorem 1.

§ 3. Numerical Results

Consider the two-point boundary-value problem

$$u'' = u + q(x), \quad 0 < x < 1, \tag{3.1}$$

$$u(0) = u(1) = 0, \tag{3.2}$$

where

$$q(x) = \begin{cases} -12 - 5x + 6x^2, & 0 < x \leqslant \frac{1}{2}, \\ -12 - 5x + 6x^2 + 15(2x-1)^{1/2} - (2x-1)^{5/2}, & \frac{1}{2} \leqslant x < 1. \end{cases} \tag{3.3}$$

This problem is clearly a special case of a boundary-value problem of type (1.3)–(1.4) and the assumptions of (1.5)–(1.6) and (1.7) hold. Hence it has a unique solution $\phi(x)$ which can be seen to be given explicitly by:

$$\phi(x) = \begin{cases} -6x^2 + 5x, & 0 \leqslant x \leqslant \frac{1}{2}, \\ -6x^2 + 5x + (2x-1)^{5/2}, & \frac{1}{2} \leqslant x \leqslant 1. \end{cases} \tag{3.4}$$

The function $\phi(x)$ is in the class $C^2[0, 1]$. However, $\phi(x)$ has no further smoothness properties since $\lim_{\xi \to 1/2+} \phi'''(\xi) = +\infty$.

We have solved this problem via the three-point finite difference method (Table 1) and via the variational method (Table 2) with the same mesh sizes. The finite difference technique shows an oscillatory phenomenon coupled with a slow convergence while the variational method shows a faster convergence and an $0(h^2)$ error as shown by the last columns of the tables where we have computed the product $(N+1)^2 \cdot \varepsilon(N)$, where

$$\varepsilon(N) = \begin{cases} \max_{1 \leqslant i \leqslant N} |\phi(x_i) - u_i|, & \text{(Table 1)}, \\ \max_{1 \leqslant i \leqslant N} |\phi(x_i) - \hat{w}(x_i)|, & \text{(Table 2)}. \end{cases}$$

Table 1

Finite Differences

N	$\varepsilon(N)$	$(N+1)^2 \cdot \varepsilon(N)$
6	$1.08 \cdot 10^{-2}$	$5.30 \cdot 10^{-1}$
8	$7.49 \cdot 10^{-3}$	$6.07 \cdot 10^{-1}$
11	$2.80 \cdot 10^{-2}$	$4.03 \cdot 10^{-1}$
16	$3.03 \cdot 10^{-3}$	$8.76 \cdot 10^{-1}$
23	$9.60 \cdot 10^{-3}$	$5.53 \cdot 10^{0}$
32	$1.20 \cdot 10^{-3}$	$1.31 \cdot 10^{0}$

Table 2
Variational Method

N	$\varepsilon(N)$	$(N+1)^2 \cdot \varepsilon(N)$
6	$1.61 \cdot 10^{-3}$	$7.89 \cdot 10^{-2}$
8	$9.72 \cdot 10^{-4}$	$7.87 \cdot 10^{-2}$
11	$5.48 \cdot 10^{-4}$	$7.89 \cdot 10^{-2}$
16	$2.74 \cdot 10^{-4}$	$7.91 \cdot 10^{-2}$
23	$1.37 \cdot 10^{-4}$	$7.89 \cdot 10^{-2}$
32	$7.27 \cdot 10^{-5}$	$7.92 \cdot 10^{-2}$

One of the difficulties with the variational method is the computation of the second integral in equations (2.3). However in the numerical example given above, this integration can be carried out in closed form. For the general case, we refer the reader to a forthcoming doctoral dissertation by R. J. HERBOLD (Case Western Reserve University).

Acknowledgement

The author wishes to acknowledge with very deep gratitude his thesis advisor, Professor RICHARD S. VARGA who generously supported this research as a part of his doctoral dissertation, and to express his sincere thanks to Mr. MICHEL FREMOND of the Laboratoire Central des Ponts & Chaussées, Paris, France, for obtaining the above numerical results on the C.A.E. 510.

REFERENCES

[1] CIARLET, P. G., *Variational Methods for Nonlinear Boundary Value Problems*, Doctoral Thesis (Case Institute of Technology, 1966).

[2] CIARLET, P. G., SCHULTZ, M. H., and VARGA, R. S., *Numerical Methods of High-Order Accuracy for Nonlinear Boundary Value Problems*, I: *One Dimensional Problems*, Numer. Math. *9*, 394–430 (1967).

[3] GERSCHGORIN, S., *Fehlerabschätzung für das Differenzenverfahren zur Lösung partieller Differentialgleichungen*, Z. Angew. Math. Mech. *10*, 373–382 (1930).

[4] HOUSEHOLDER, A. S., *The Theory of Matrices in Numerical Analysis* (Blaisdell, New York 1964).

[5] LEES, M., *Lecture Notes on Two-Point Boundary Value Problems* (Case Institute of Technology, 1965).

[6] ROSE, M. E., *Finite Difference Schemes for Differential Equations*, Math. Comp. *18*, 179–195 (1964).

[7] SCHECHTER, S., *Iteration Methods for Nonlinear Problems*, Trans. Amer. Math. Soc. *104*, 179–189 (1962).

[8] VARGA, R. S., *Matrix Iterative Analysis* (Prentice-Hall, Englewood Cliffs, N.J. 1962).

[9] VARGA, R. S., *On a Discrete Maximum Principle*, SIAM J. Numer. Anal. *3*, 355–359 (1966).

[10] VARGA, R. S., *Hermite Interpolation-Type Ritz Methods for Two-Point Boundary Value Problems*. in *Partial Differential Equations*, Edited by J. H. BRAMBLE (Academic Press, New York 1966), pp. 365–373.

Case Institute of Technology, Cleveland, U.S.A., and
38, rue Louis Blanc, 45 – Gien, France.

Involutory Functions and Even Functions

H. SCHWERDTFEGER (Montreal Que. Canada)

Dedicated to A. M. Ostrowski on his 75th birthday

Introduction

In an earlier note [1] J. ACZEL has discussed relations between involutory functions, i.e. solutions of the functional equation

$$f(f(x)) = x \tag{0.1}$$

and even functions: $\varphi(-\xi)=\varphi(\xi)$, based on the remark that an involutory function $y=f(x)$ coincides with its own inverse and that therefore its graph in a cartesian x, y-coordinate system is symmetric with respect to the bisectrice $y=x$. Taking this line as the η-axis of a ξ, η-system obtained from the x, y-system by a $-45°$ rotation about the origin the connection between an involutory function and the corresponding even function is realized analytically by

$$\xi = \tfrac{1}{2}\sqrt{2}(x - f(x)), \quad \varphi(\xi) = \tfrac{1}{2}\sqrt{2}(x + f(x)).$$

We shall study involutory functions strictly in intervals of uniformity. The graphical representation then suggests that an involutory $f(x)$ defined for a non-negative x-interval can be defined in the corresponding non-positive interval so that it becomes an odd function: $f(-x)=-f(x)$. If such an $f(x)$ is continuous and has the positive fixed point $x_0=f(x_0)$, then also $-x_0$ is a fixed point.

There are other relations between involutory and even functions which apparently have not been observed so far. These relations are functional equations which, although they cannot be solved in full generality, may serve to define involutory functions in many special cases. The derivation of a functional equation will be based on a geometrical construction involving the graphs of the given functions. Only such solutions will be considered which are significant with respect to the geometrical interpretation.

In the following all constants, variables and functions will be supposed to be real, the functions defined in suitable intervals such that the occurring functional compositions are meaningful. Properties of smoothness and differentiability will be assumed in accordance with the requirements of the described analytical operations.

§ 1. Let an even function $g(x)\geqslant 0$ be given and let C be its graph in a cartesian coordinate system. Let F be a family of straight lines $l=l\,[x]$ with x as parameter, symmetric with respect to the y-axis so that if a line l passes through the point $(0, a)$,

Received September 28, 1967 and, in revised form, December 14, 1967

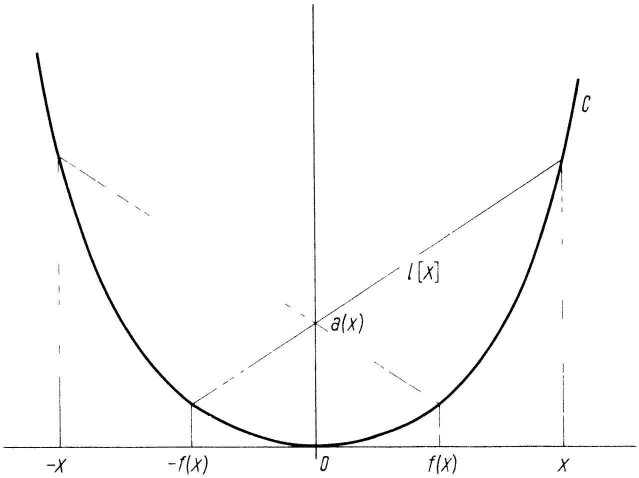


Figure 1
Definition of $f(x)$.

$a = a(x)$, with the slope $p = p(x)$, then through this point passes exactly one further line $l^* = l[-x] \in F$ with the slope $p^* = -p = p(-x)$. Fig. 1 illustrates the situation in the special case that $g(x)$ is convex and $g(0) = 0$. It is clear that $a(x)$ must also be an even function. Moreover we shall assume that there is an $x_0 > 0$ such that $a(x_0) = g(x_0)$. The figure suggests that for some interval I of positive values x there is an odd involutory function $f(x)$ such that if a line of F passes through the point $(x, g(x))$ of C the line meets C a second time[1]) at the point $(-f(x), g(f(x)))$. Then the line l^* passing through $(f(x), g(f(x)))$ and $(0, a(x))$ meets C a second time at the point

$$((-f(f(x)), g(f(\ (x)))) = (-x, g(x))$$

so that indeed $f(f(x)) = x$ for $x \in I$. For reasons of symmetry this relation is also satisfied for the corresponding negative x and so if $f(x) > 0$ for $x > 0$ we have $f(-x) = -f(x)$.

Since the lines through $(x, g(x))$, $(-f(x), g(f(x)))$ and through $(f(x), g(f(x)))$, $(-x, g(x))$ intersect the y-axis at the same point $(0, a(x))$ it follows that

$$a(f(x)) = a(x). \tag{1.1}$$

Moreover we have to assume $a(x) > 0$.

Between the three functions $g(x)$, $a(x)$ and $f(x)$ there exists a functional equation, readily derived from Fig. 1. Since

$$p(x) = \frac{g(x) - g(f(x))}{x + f(x)} = \frac{g(x) - a(x)}{x} = \frac{a(x) - g(f(x))}{f(x)} \tag{1.2}$$

---

1) There are situations in which this is not the case; cf. § 2.

we have

$$(g(x) - a(x))f(x) + (g(f(x)) - a(x))\, x = 0\,. \tag{1.3}$$

Either of the three functions: $g(x)$, $a(x)$, even and positive, and $f(x)$, odd, involutory and positive for $x>0$, may be considered as the unknown, the two others being prescribed in accordance with the geometrical interpretation. We observe that (1.3) always has the solution $f(x)= -x$, which, however, is geometrically insignificant and therefore to be ignored.[2]) Given $g(x)$ and $f(x)\neq -x$ the function $a(x)$ is readily found:

$$a(x) = \frac{g(x)f(x) + g(f(x))\, x}{f(x) + x};$$

it is even, positive and satisfies (1.1).

If x_0 is a fixed point of $f(x)$, then $a(x_0)=g(x_0)$ and conversely, if $g(x)$ is monotonic for $x\in I$.

§ **2.** Now let the two even functions $g(x)$ and $a(x)$ be given subject to the conditions indicated in § 1, so that $f(x)$ is the unknown odd involutory function, to be determined as a solution of (1.3) with the condition (1.1). From the geometrical interpretation it is evident that $f(x)$ might exist in several different intervals and that some of these may cover each other wholly or partially. We shall restrict ourselves to a local study of $f(x)$ which means that we take into consideration only a curve (or an arc of a curve) C and a line family F for which $f(x)$ exists in one certain interval I of positive values x. It will be seen that this is so if $g(x)$ has certain properties as shown in Fig. 2–5 and the only turning point, a maximum or a minimum of $a(x)$, lies on C. Assuming $g(x)$ and $a(x)$ sufficiently smooth it appears that every more general situation is made up of a finite number of these special cases some of which will be discussed.

First we shall explain the geometrical meaning of (1.3). Since the slope $p(x)$ is defined by $g(x)$ and $a(x)$ (cf. (1.2)), the line family F itself is defined by these two functions. Indeed the line $l\,[x]$ has the equation $\eta=p(x)\,\xi+a(x)$; the intersections of $l\,[x]$ with C are therefore defined by the roots of the equation

$$p(x)\,\xi + a(x) = g(\xi)\,. \tag{2.1}$$

One of its roots is $\xi=x$. Assuming $x>0$ we are interested in a second, negative, solution $\xi= -f(x)$. Thus (2.1) is essentially the same as equation (1.3).

Let us consider the case of a convex $g(x)$, increasing for $x>0$ as shown in Fig. 2 or Fig. 2a. The function $a(x)$ is also given by its graph K in the same coordinate system as C. Let x_0 be the point where the two curves meet: $p(x_0)=0, f(x_0)=x_0$.

[2]) This solution corresponds to the case that F is the family of tangents to C. It is, however, not covered by our initial assumptions.

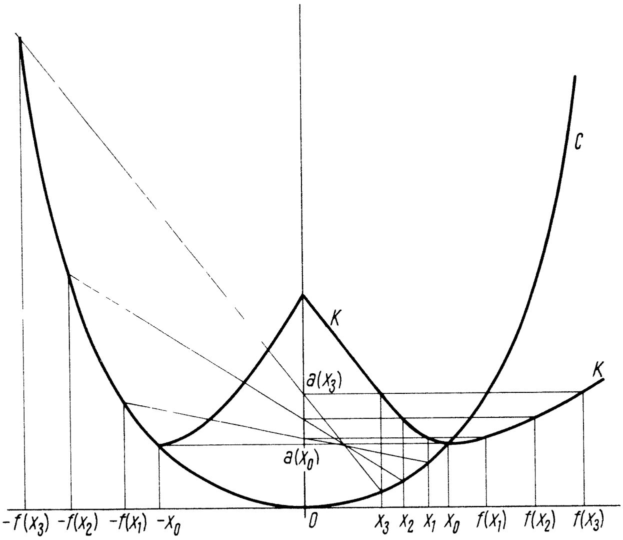


Figure 2

$g(x)$ convex, increasing for $x > 0$, $a(x)$ has minimum on C.

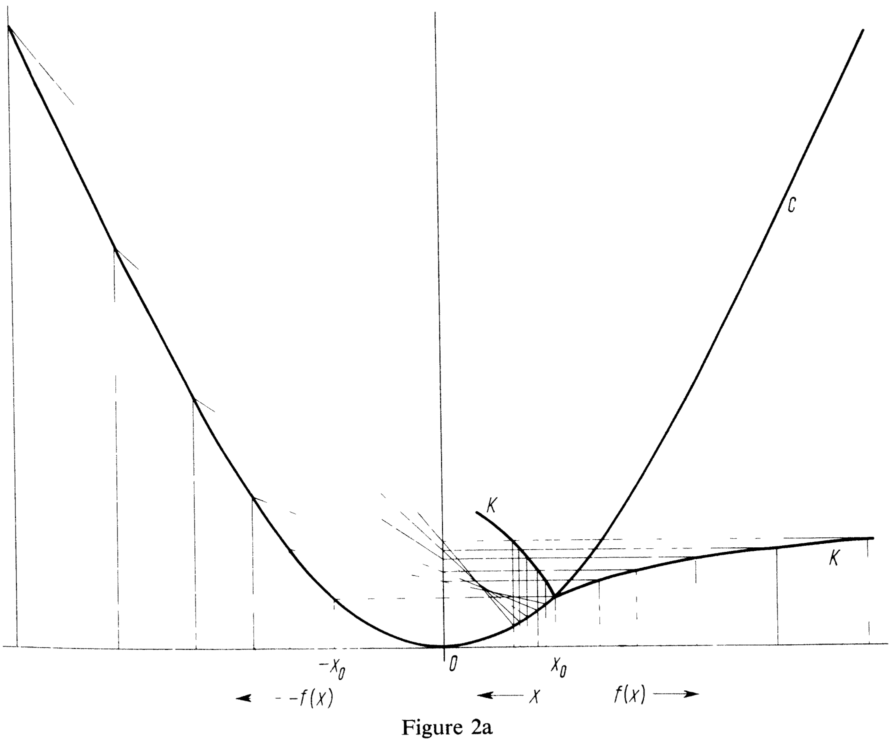


Figure 2a

$g(x)$ convex, increasing if $x > 0$, K has cusp on C.

If for $x<x_0$ we have $a(x)>a(x_0)$ so that $g(x)<a(x)$, it follows that $f(x)>x_0$ and because of (1.1) the function $a(x)$ is automatically defined for $x>x_0$ if its values are given for $x<x_0$; this is also evident from the construction indicated in the figures.

Assuming $g(x)$ and $a(x)$ continuous in a neighbourhood of x_0 we shall show that there also $f(x)$ will be continuous. Indeed the line through $(0, a(x))$ and $(x, g(x))$ performs a continuous motion; this means that $p(x)$ and $a(x)$ are continuous. Thus the intersection of $l\,[x]$ and C, i.e. the point $(-f(x), g(f(x)))$, moves continuously on C; hence $f(x)$ is continuous in an interval containing x_0, in particular at x_0; thus

$$f(x)\to x_0 \quad \text{as} \quad x\to x_0 .$$

Now we consider the horizontal line through $(x, a(x))$. It passes through the point $(f(x), a(x))$. As $x\to x_0$ these two approach the one point $(x_0, g(x_0))$ and the horizontal line approaches the line $l\,[x_0]$ which therefore has with the curve K only the point $(x_0, g(x_0))=(x_0, a(x_0))$ in common; hence if $a'(x)$ exists in the interval around x_0, then $a'(x_0)=0$. This also follows from the fact that $a(x)$ has a minimum for $x=x_0$. Indeed $a(x)>x_0$ and $y=f(x)>x_0$ as $x<x_0$; it follows that $a(y)=a(x_0)>x_0$. In Fig. 2a is shown a situation where $a'(x_0)$ does not exist.

Another special situation is illustrated in Fig. 3. It is again assumed that apart from the convex $g(x)$ (increasing if $x>0$) the even function $a(x)$ is prescribed for positive $x\leqslant x_0$. The figure then indicates – as in the preceding cases – how the curve K can be completed. In the present case K turns out to be a closed curve.

The interval I of uniformity of $f(x)$ is bounded left-hand by a point $x_1<x_0$ for which $f'(x_1)=0$. Since by (1.1)

$$a'(f(x))f'(x) = a'(x) \tag{2.2}$$

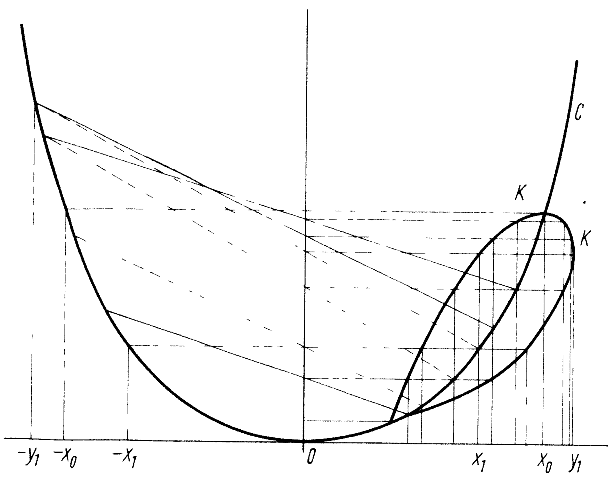


Figure 3
$g(x)$ convex, increasing for $x>0$, $a(x)$ has maximum on C.

we conclude that $a'(f(x_1))$ must be infinite; therefore the right-hand bound $y_1 > x_0$ of I is to be found as the solution of the equation $a'(y)^{-1}=0$ and so x_1 appears as the positive root $(<x_0)$ of $a(x)=a(y_1)$.

Two further special cases are represented in Fig. 4–5. They can be similarly discussed.

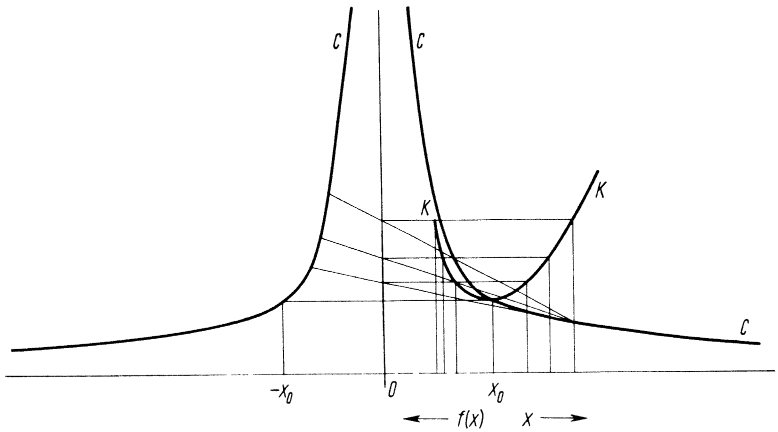


Figure 4
$g'(x)<0, g''(x)>0$ if $x>0$, $a(x)$ has minimum on C.

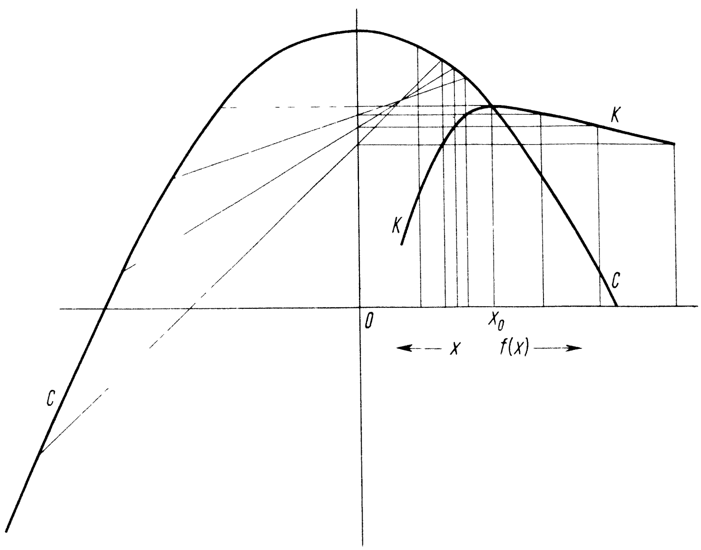


Figure 5
$g'(x)<0, g''(x)<0$ for $x>0$, $a(x)$ has maximum on C.

REMARK. The completion of K from a given part of K as required in these discussions is a geometrical construction which can be carried out for a curve K with respect to a given curve C, if C and K are in suitable mutual position. The construction is analytically realised by an element transformation (in the sense of LIE)

$$X = F(x, y, p), \quad Y = G(x, y, p), \quad \mathscr{P} = H(x, y, p)$$

so that if $p = dy/dx$ then $\mathscr{P} = dY/dX$. Indeed if $y = a(x)$ then $dy/dx = a'(x)$ and

$$X = f(x), \quad Y = a(f(x)) = a(x) = y, \quad \frac{dY}{dX} = \frac{a'(x)}{f'(x)} = a'(f(x)) = a'(X) \quad \text{by (2.2)}.$$

Thus the element transformation is a contact transformation. Over the uniformity interval of $f(x)$ it coincides with its own inverse. Since by (0.1) $f'(X)f'(x) = 1$, we have

$$x = f(X), \quad y = Y, \quad \frac{dy}{dx} = f'(x) \cdot \frac{dY}{dX} = \frac{a'(X)}{f'(X)}.$$

EXAMPLES: In the case $a(x) = g(x)$ the lines $l[x]$ are all horizontal; therefore $f(x) = x$.

Further let us assume that $a(x) = a$ is a positive constant. This function satisfies all the conditions, incl. (1.1). The fixed point x_0 of $f(x)$ is defined by $g(x_0) = a$. Here are some simple examples:

(a) Let $g(x) = x^2$; then (1.3) is a quadratic equation in $f(x)$:

$$xf^2 + (x^2 - a)f = ax$$

and its only significant solution is $f(x) = a/x$.

(b) $g(x) = |x|$. It is geometrically evident that a solution does not exist in the interval $-\frac{1}{2}a \leqslant x \leqslant \frac{1}{2}a$. The functional equation (1.3) becomes

$$(|x| - a)f + (|f| - a)x = 0$$

and

$$f(x) = \begin{cases} \dfrac{ax}{2x - a} & \text{if} \quad x > \frac{1}{2}a \\[2ex] \dfrac{-ax}{2x + a} & \text{if} \quad x < -\frac{1}{2}a. \end{cases}$$

(c) If $g(x) = |\sin x|$ $(-\frac{1}{2}\pi < x < \frac{1}{2}\pi)$ and $a > 1$ the geometrical problem of § 1 has no solution.

Finally a numerical example with a variable $a(x)$, cf. Fig. 6. Let again $g(x) = |x|$; moreover $a(x) = 2x_0 - x$ for $0 < x \leqslant x_0$. Then the line $l[x]$ has the equation

$$\eta = 2x_0 - x + 2\frac{x - x_0}{x}\xi.$$

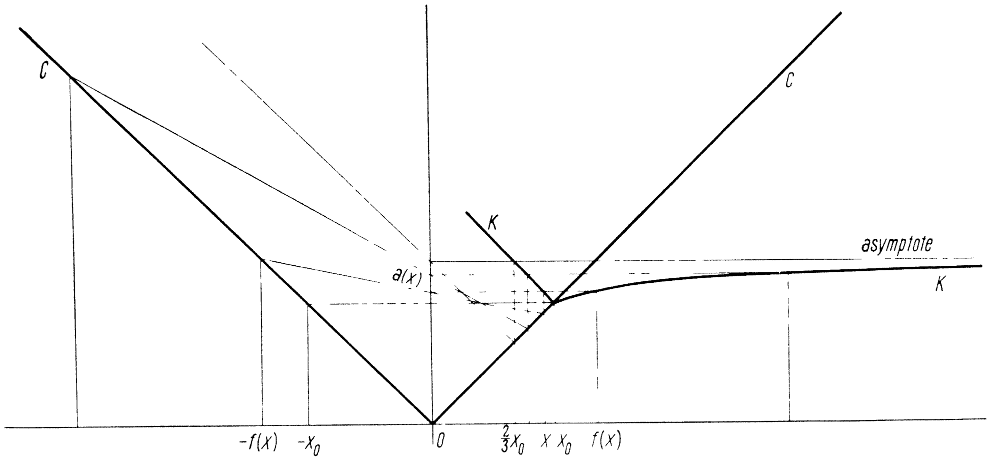


Figure 6
$g(x) = |x|$, $a(x) = 2x_0 - x$ $(x < x_0)$.

The (negative) abscissa of the (second) intersection of $l[x]$ with the graph C of $g(x)$ defines $-f(x)$. So we find

$$X = f(x) = x \frac{2x_0 - x}{3x - 2x_0} \geqslant x_0 \quad \text{for} \quad \tfrac{2}{3}x_0 < x \leqslant x_0$$

and therefore for all $X > x_0$:

$$x = f(X) = -\tfrac{1}{2}(3X - 2x_0) + \tfrac{1}{2}\sqrt{8X^2 + (X - 2x_0)^2}\,.$$

The (positive) interval I in which $f(x)$ is defined is thus seen to be $\frac{2}{3}x_0 < x < \infty$.

§ 3. Given an odd involutory function $f(x) > 0$ for $x > 0$ with the positive fixed point x_0 and a positive even function $a(x)$ satisfying (1.1) which has its maximum or minimum at x_0. We ask how far the equation (1.3) then determines a certain non-negative even function $g(x)$. We write (1.3) in the form

$$f(x)\,g(x) + x g(f(x)) = a(x)\,(f(x) + x)\,. \tag{3.1}$$

Evidently the function $a(x)$ is a solution of this equation. Thus if $h(x)$ is a suitable solution of the corresponding homogeneous equation

$$f(x)\,h(x) + x\,h(f(x)) = 0 \tag{3.2}$$

we have

$$g(x) = a(x) + h(x)\,. \tag{3.3}$$

Any solution of (3.2) can be written in the form $h(x) = x p(x)$ where $p(x)$ is an arbitrary function satisfying the conditions

$$p(f(x)) = -p(x) = p(-x).$$

Thus certainly $g(x)$ is an even function. If we interpret $p(x)$ as the slope of the line $l[x]$ we find that (3.3) coincides with one of the relations (1.2). Only at the fixed points $\pm x_0$ of $f(x)$ are the values of $g(x)$ defined: $g(x_0)=a(x_0)$, i.e. $p(x_0)=0$. The values of $p(x)$, and therefore those of $g(x)$, can be prescribed arbitrarily for positive $x<x_0$ after which $g(x)$ is completely defined.

§ **4.** There are various other ways of describing a symmetric line family F analytically with respect to a given even function $g(x)$ in order to determine a positive odd involutory function $f(x)$. For instance F may be given as the system of the tangents to the graph Γ of an even function $\varphi(X)>0$. Let us choose $g(x)$ again convex increasing and a suitably situated Γ (cf. Fig. 7). The line l touching Γ at the point

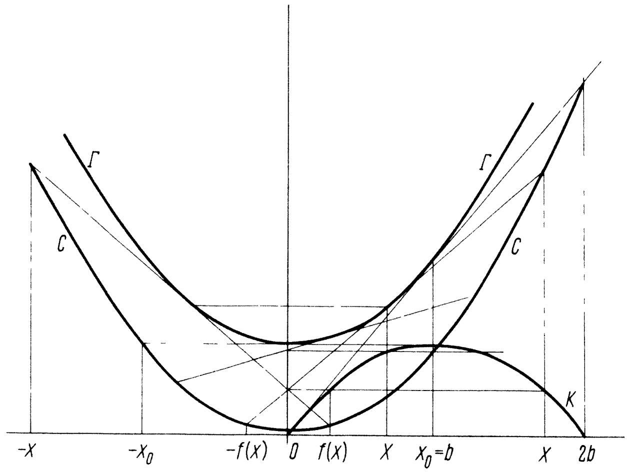


Figure 7
Construction of K for given C and Γ.

$(X, \varphi(X))$ has the equation $\eta=\varphi(X)+\varphi'(X)(\xi-X)$; hence we have the following relations between x and X:

$$p(x) = \varphi'(X) = \frac{g(x)-\varphi(X)}{x-X}, \quad a(x) = \varphi(X) - \varphi'(X)\,X. \tag{4.1}$$

With regard to (1.2) (second expression for $p(x)$) it is clear that $\varphi(0)=g(x_0)=a(x_0)$; thus by (4.1) $\varphi'(0)=0$ if $\varphi'(0)$ exists.

For suitable constellation of the curves C and Γ the equation (4.1) will have a solution $X=F(x)$. Since $p(x)$ is an odd function while $g(x)$ and $\varphi(X)$ are even, it is readily seen that $F(x)$ must be odd.

Using the first expression for $p(x)$ in (1.2) we obtain the functional equation

$$\varphi'(F(x))\,(x+f(x)) = g(x) - g(f(x)). \tag{4.2}$$

As an example let $g(x)=x^2$ and $\varphi(X)=X^2+b^2$ $(b>0)$. A simple calculation and reference to the geometrical situation (Fig. 7) shows that

$$X = F(x) = \begin{cases} x-b & \text{if} \quad x>0 \\ x+b & \text{if} \quad x<0\,. \end{cases}$$

Hence by (4.2)

$$f(x) = \begin{cases} -x+2b & \text{if} \quad 0<x<2b \\ -x-2b & \text{if} \quad 0>x>-2b\,. \end{cases}$$

From the second relation (4.1) we find $a(x)= -x(x-2b)$. Thus in this example the three curves C, K, Γ are congruent parabolae. The data fixing the mutual situation of the three curves are shown in Fig. 7. In particular it may be mentioned that the tangent to K at (0,0) is also tangent to Γ at $(b, 2b^2)$ and meets C at (0,0) and $(2b, 4b^2)$.

§ **5.** Further we consider what might be called the archimedean condition: Let the area enclosed by the graph C of the even function $g(x)$ represent a perpendicular section through a cylindrical homogeneous ship (swimmer) oscillating (rolling) around its position of equilibrium under the influence of gravity about a (variable) axis, perpendicular to the x, y-plane. With respect to this section of the ship the (plane) surface of the water marks the lines $l \in F$. According to Archimedes' principle of hydrostatics the area, enclosed by the curve C and any of the lines l has constant measure μ. This implies that

$$(x+f(x))\,(g(x)+g(f(x))) - 2\int_{-f(x)}^{x} g(t)\,dt = 2\mu\,. \tag{5.1}$$

Putting $f(x)=y$ and observing that $\int_0^x g(t)\,dt = G(x)$ is an odd function this equation can be written in the form

$$F(x, y) \equiv (x+y)\,(g(x)+g(y)) - 2(G(x)+G(y)) = 2\mu\,. \tag{5.2}$$

The function $F(x, y)$ is symmetric in x and y and the involutory function $f(x)$ is obtained by solving the equation (5.2) with respect to y. Because of the symmetry also $x=f(y)$. The relation between involutory and symmetric functions has also been pointed out in [1].

In the present case it is now possible to determine the curve Γ of sect. 4, i.e. the envelope of the family of lines l:

$$(g(x) - g(y))\,\xi - (x + y)\,\eta + g(x)\,y + x\,g(y) = 0$$

where x is the family parameter and $y = f(x)$. Using the standard procedure and the differential equation

$$y' = \frac{g(y) - g(x) - (x + y)\,g'(x)}{g(y) - g(x) + (x + y)\,g'(y)} \tag{5.3}$$

derived from (5.2) by differentiation, we obtain by a simple calculation

$$X = \tfrac{1}{2}(x - f(x)), \qquad Y = \tfrac{1}{2}(g(x) + g(f(x))) \tag{5.4}$$

as parameter representation of the curve Γ. Elimination of the parameter x would lead to the function $Y = \varphi(X)$.

It may be pointed out that the formulae (5.4) represent the second theorem of Dupin in the theory of swimming bodies; for the usual proof cf. [2].

In the special case $g(x) = x^2$ we find $F(x, y) = \frac{1}{3}(x + y)^3$ and therefore

$$y = f(x) = -x + \alpha, \qquad \alpha^3 = 6\mu.$$

For the curve Γ we have $Y = \varphi(X) = X^2 + \frac{1}{4}a^2$. Thus we come back to the formula with which we started in the example of § 4.

§ 6. Other suitable conditions on the line family, with or without immediate geometrical interpretation, can readily be formulated, although their explicit evaluation may be hard or even impossible. A condition with a geometrical, if not immediate mechanical, interpretation is the following: Let the lines l with respect to the curve C be fixed so that the 'submerged arc' of C, i.e. the arc under the line l, has constant length γ. Thus putting

$$S(x) = \int_0^x \sqrt{1 + g'(t)^2}\, dt$$

and noting that this is again an odd function, one has

$$S(x) + S(f(x)) = \gamma. \tag{6.1}$$

By differentiating (6.1) with respect to x we obtain the differential equation

$$y' = -\frac{\sqrt{1 + g'(x)^2}}{\sqrt{1 + g'(y)^2}}. \tag{6.2}$$

Its solution $y = f(x, c)$ contains an arbitrary constant c for which the proper value can be found by using (6.1).

The differential condition (6.2) of (5.3) can easily be generalized into

$$y' = H(x, y, g(x), g(y), g'(x), g'(y)) \tag{6.3}$$

where the function H of the six variables has to satisfy the relation

$$H(x, y; g(x), g(y); g'(x), g'(y)) \cdot H(y, x; g(y), g(x); g'(y), g'(x)) = 1$$

to make sure that the solution $y = f(x)$ of (6.3) can be involutory. Indeed from (0.1) follows $f'(x) f'(y) = 1$.

§ 7. In section 3 we have dealt with the problem of finding $g(x)$ for a given $f(x)$; this led to the linear equation (3.1). The equation (6.1) is of a similar type if we consider $S(x)$ as unknown function. Functional equations of this kind have been studied in [3] without the restriction that $f(x)$ be involutory.

Part of this paper was subject of a brief report at the Conference on Functional Equations, Waterloo, April 1967. In its present (revised) form the paper owes much to the constructive criticism of a referee.

REFERENCES

[1] Aczél, J., *A Remark on Involutory Functions*, Amer. Math. Monthly *55*, 638–639 (1948).
[2] de la Vallée Poussin, C.J., *Leçons de Mécanique Analytique*, Tome II (Gauthier Villars, Paris 1925), p. 277.
[3] Kordylewski, J. and Kuczma, M., *On Some Linear Functional Equations*, Ann. Polon. Math. *9*, 119–136 (1960–1961).

McGill University

A Grammar of Functions[1])

B. SCHWEIZER (Amherst, Mass., U.S.A.) and A. SKLAR (Chicago, Illinois, U.S.A.)

Dedicated to Professor A. Ostrowski on the occasion of his 75th birthday

Contents

PART I

Introduction

Since the end of the seventeenth century, the study of functions has been a focal point in the development of mathematics. To date, the major effort has gone into the study of special functions and classes of functions, usually defined on domains with a special structure. With the outstanding exception of group theory, and an early attempt by BABBAGE and HERSCHEL [3; 16], the study of functions from a purely algebraic point of view is a recent development. And this development has been uneven: while the theory of systems of one-place functions (both in themselves and in connection with other types of algebraic systems) can be regarded as well-established (see, e.g., [5; 6; 20; 23; 30; 32; 33]), the corresponding theory of multiplace and vector-valued functions, though the subject of much recent work (e.g., [10; 24; 26; 31]), is less advanced.

---

1) This work was supported in part by NSF Grants GP-7067 and GP-6303.

Received October 13, 1967

In this paper we begin the development of a comprehensive theory of multiplace vector-valued functions over arbitrary sets. Our approach is both an outgrowth from and a drastic modification of an earlier attempt [34]. Its basis is the consideration of the entire set $\mathscr{F}_\infty(S)$ of functions from subsets of Cartesian powers of a set S into other such subsets (see Section 2). We introduce two unary operations, L and R, and three binary operations into $\mathscr{F}_\infty(S)$. The first binary operation, which we call *serial composition* and denote by σ, coincides with ordinary function composition whenever the latter yields non-empty results. Like ordinary composition, serial composition is associative; in fact, as we show in Section 4, the system $(\mathscr{F}_\infty(S), \sigma, L, R)$ is a function semigroup (see [32; 33] and Section 1). We regard this result, which originally came as a complete surprise to us, as highly significant: On the one hand, it testifies to the naturalness of the operation σ; and on the other, it shows that the theory of function semigroups, developed for the abstract study of one-place functions, is of importance for the theory of multiplace functions, and thus for such topics as general Ω-algebras (see [7]).

In Section 5, we introduce the second binary operation, called *parallel composition* and denoted by π. Like σ, π is associative (though there is no identity element for it), and there are (conditional) interassociative and distributive laws connecting σ and π. The functions which we, following MENGER [23], call *selectors*, are studied in Section 6: these are the functions $J_{mn}(1 \leqq m \leqq n)$ such that $J_{mn}(x_1, \ldots, x_n) = x_m$ for all $(x_1, \ldots, x_n) \in S^n$. We show that, under σ and π, the single selector J_{22} generates not only all the selectors, but also all the *multiselectors* (vector-valued counterparts of selectors) as well. In Section 7 we define the third associative binary operation, denoted by β, in terms of σ, π, and multiselectors. The use of β permits us to simplify many expressions and proofs.

The four remaining sections are devoted to applications. In Section 8, we show that all the usual 'compositions' and 'modes of generation' of multiplace functions can be subsumed under σ, π, and β, plus the use of multiselectors. Linearity and generalizations of linearity are taken up in Section 9. It is shown that, wherever linearity is meaningful, the set of linear transformations is closed under each of the operations σ, π, and β. Among other things, this permits us to introduce three everywhere-defined associative operations on the set of all matrices over a ring, one of these operations being an extension of the ordinary matrix product. A brief indication of the fact that these considerations extend to multilinear transformations is also given. In Section 10, after discussing constants and place-fixing, we give a 'variable-free' definition of 'essential variables' and prove a theorem equivalent to the left-representation theorem for semigroups.

Finally, since we feel that the theory developed in this paper is the natural setting for the study of functional equations over a set, we very briefly touch on such equations in Section 11. The discussion ends with an application to the definition of recursive functions: We show that the entire set of primitive (resp., general) recursive functions can be characterized in terms of 4 (resp., 5) functions, instead of the usual countably infinite set of initial functions.

As this summary indicates, in this paper our entire approach to the subject is concrete rather than abstract. That is, we do not define our system as a set of abstract elements subject to certain postulates, but as a collection of concrete functions over a given underlying set. Still, axiomatization remains one of our aims*); and, looking ahead to this, we have consistently tried to make at least the statements of our theorems fit either a concrete or an abstract setting.

The original version (outlined in [34]) of cur present system was also concrete and dealt with essentially the same set of functions. But in most other respects the two versions differ greatly. In [34], a 2-place function, for instance, was defined as a 1-place function whose *values* are ordinary 1-place functions. This device, which was borrowed from combinatory logic (cf. [8], pp. 83–84) has been discarded, and all functions in the present system are on the same footing. In [34], a countably infinite set of binary operations was introduced. These have been superseded by the two operations σ and π: All the old operations are definable in terms of the new ones (see Section 8). Lastly, an equivalence relation employed in [34] has been dropped.

An announcement of the results of this paper appeared in [35]. Since then we have made a few changes, chiefly in notation. The only ones of significance are: (a) the use of σ instead of simple juxtaposition to denote serial composition; (b) the use of π to denote parallel composition, so that we write $F\pi G$ instead of (F, G); (*c*) the use of I_n instead of J_n to denote the identity function on S^n. Also the operation β did not appear in [35]. (For future use it may prove convenient to have several different notations for serial and parallel composition, in which case the notations of [35] can be revived.)

Because of its length, this paper is divided into two parts. Part I is in this issue and contains the Introduction and Sections 1–5; Part II will appear in a later issue and contain Sections 6–11; both parts include the complete Table of Contents and Bibliography.

Lemmas, theorems, and formal definitions are each numbered consecutively throughout the paper and referred to by number: Lemma 6, Theorem 11, Definition 5. Corollaries are referred to by the theorems to which they belong; thus 'Corollary 7' indicates the single corollary of Theorem 7, while 'Corollary 2.7' denotes Corollary 7 of Theorem 2. In the body of proofs, such references are abbreviated: thus, 'L6' 'T11', 'D5', 'C7', 'C2.7' for the examples cited. Displayed formulas are numbered consecutively within each section. They are referred to by number alone in the same section, and by section number and display number in a different section.

The influence of K. Menger is evident throughout this paper. We have also

*) *Note added in proof.* This aim has been achieved to the extent that we can characterize systems isomorphic to closed subsystems of $(\mathscr{F}_\infty(S), \sigma, \pi, L, R)$.

profited greatly from discussions with numerous colleagues, notably J. ACZÉL, V. D. BELOUSOV, M. A. MCKIERNAN, B. M. SCHEIN and H. I. WHITLOCK. We have gained much from our participation in the various conferences on functional equations held in Oberwolfach and in Waterloo. Finally, we thank the referees for their careful reading of the manuscript and for their many helpful comments and suggestions.

1. Functions and Function Semigroups

In this paper we take the notion of a (concrete) *function*, as well as the related notions of *domain, range, argument, value*, etc. for granted. The domain of a function f will be denoted by $\mathrm{Dom} f$; the range by $\mathrm{Ran} f$; and the value of f for $x \in \mathrm{Dom} f$ by fx. It is possible for either $\mathrm{Dom} f$ or $\mathrm{Ran} f$ to be empty; if either one is, then the other is also, and f is the *empty function* $\emptyset$. The usual way of defining a function f is first to identify $\mathrm{Dom} f$, and then to specify fx for every $x \in \mathrm{Dom} f$. Thus, given any set S we can define a function j_S, the *identity function* on S, by the specifications:

$$\mathrm{Dom}\, j_S = S, \tag{1}$$

$$j_S x = x \quad \text{for all} \quad x \in S. \tag{2}$$

With any function f we associate two related functions, Lf and Rf, as follows:

$$Lf = j_{\mathrm{Ran} f}, \qquad Rf = j_{\mathrm{Dom} f}. \tag{3}$$

If $f = Rf$ then $f = Lf$, and conversely; in this case we call f a *sub-identity*. The empty function $\emptyset$ is a subidentity.

Given any two functions f and g, we define a third function $f \circ g$ by:

$$\mathrm{Dom}(f \circ g) = \{x \mid x \in \mathrm{Dom}\, g,\ g x \in \mathrm{Dom} f\}, \tag{4}$$

$$(f \circ g)\, x = f(g x) \quad \text{for all} \quad x \in \mathrm{Dom}(f \circ g). \tag{5}$$

We call $f \circ g$ the *ordinary* (or *natural*) *composite* of f and g (in that order). Ordinary composition is an inherently associative operation: $(f \circ g) \circ h = f \circ (g \circ h)$ for any three functions f, g, h.

There is also an intrinsic partial ordering among functions: we say that f is a *restriction of* g, or g is an *extension* of f, and write $f \subseteq g$, if:

$$\mathrm{Dom} f \subseteq \mathrm{Dom}\, g, \tag{6}$$

$$f x = g x \quad \text{for all} \quad x \in \mathrm{Dom} f. \tag{7}$$

Composition and restriction are connected by the fact that $f \subseteq g$ if and only if $f = g \circ Rf$. Moreover, if $Rf \subseteq g$ and g is a subidentity, then $f \circ g = f$; similarly, if g is a subidentity and $Lf \subseteq g$, then $g \circ f = f$. It follows in particular that $Lf \circ f = f = f \circ Rf$ for any function f.

A set of functions closed under $\circ$, L, and R is a (concrete) *function system*. A function system with an identity element, i.e., an element j such that $j \circ f = f = f \circ j$ for all elements f of the system, is a *function semigroup.* (Since natural composition is associative, every set of functions closed under composition is automatically a semigroup. But not all semigroups of functions are function semigroups, or even function systems.)

We denote the set of all functions whose domains and ranges are subsets of a set S by $\mathscr{F}(S)$. For any non-empty set S, the set $\mathscr{F}(S)$ is always closed under $\circ$, L, and R. Therefore $(\mathscr{F}(S), \circ, L, R)$ is a function system; and since it has an identity element – j_S – it is a function semigroup. We call it the *primary function semigroup on S.*

For every function f there is a function g such that

$$f \circ g = Lf = Rg \tag{8}$$

(this assertion is in fact equivalent to the Axiom of Choice [32]). If g is related to f in this manner, we say that g is a *right-subinverse* of f and write $g[RS]f$. A function system in which every element has a right-subinverse in the system is said to have the *right-subinverse property.* Evidently, every primary function semigroup has the right-subinverse property.

An *abstract* function system (cf. [33]) is a quadruple $(\mathscr{S}, \circ, L, R)$ consisting of a non-empty set $\mathscr{S}$, a binary operation $\circ$ and two unary operations, L and R, on $\mathscr{S}$, subject to the following axioms:

AXIOM 1: The pair $(\mathscr{S}, \circ)$ is a semigroup.

AXIOM 2: For all elements f of $\mathscr{S}$:
(a) $LRf=Rf$, $RLf=Lf$,
(b) $Lf \circ f=f=f \circ Rf$.

AXIOM 3: For all pairs of elements f, g of $\mathscr{S}$:
(a) $L(f \circ g)=L(f \circ Lg)$, $R(f \circ g)=R(Rf \circ g)$,
(b) $Lf \circ Rg=Rg \circ Lf$,
(c) $Rf \circ g=g \circ R(f \circ g)$.

In [33], we showed that every (abstract) function system with the right-subinverse property is isomorphic to a subsystem of some primary function semigroup: and in [32], we gave a complete axiomatic characterization of primary function semigroups.*)

2. Functions over S

Throughout the rest of this paper, S will denote a fixed non-empty set. For any positive integer n, S^n will denote the n-fold Cartesian power of S. As usual we identify S^1 with S. Elements of S^n, i.e., ordered n-tuples of elements of S, will be written as

*) *Note added in proof.* For another characterization see: GLUSKIN, L. M., *Ideals of Semigroups of Transformations*, Mat. Sb. (N.S.) *47* (89), 111–130.

strings, e.g., $x_1 x_2 \dots x_n$. If $m \leqq n$, then the string $x_1 \dots x_m$ is an *initial segment* of the string $x_1 \dots x_n$, and $x_1 \dots x_n$ is an *extension* of $x_1 \dots x_m$. Note that every string is both an initial segment and an extension of itself.

If S' is any non-empty set of strings of elements of S, then $E(S')$ is the set of all extensions of all the strings of S'. If S' is the empty set, then $E(S')=S'$.

In particular, $E(S^n)$ is the union $\bigcup_{i=n}^{\infty} S^i$. The set $E(S)$ is thus the union of all the sets S^i, i.e., $E(S)$ is the set of all strings. The elements of $E(S)$ form a semigroup – the free semigroup on S – under *concatenation* [6]. Thus we can write, indifferently, $(x_1 \dots x_m) \dots x_n$ or $x_1 \dots x_n$.

DEFINITION 1. For any two positive integers d and r, the set of all functions whose domains are non-empty subsets of S^d and whose ranges are subsets of S^r will be denoted by $\mathscr{F}_{rd}(S)$. The union of all the sets $\mathscr{F}_{rd}(S)$, together with the empty function $\emptyset$, will be denoted by $\mathscr{F}_{\infty}(S)$. An element of $\mathscr{F}_{\infty}(S)$ will be called a function *over* S.

Note that the set $\mathscr{F}_{\infty}(S)$ is a *proper* subset of $\mathscr{F}(E(S))$, and that $\mathscr{F}_{11}(S) \cup \{\emptyset\} = \mathscr{F}(S)$. Functions over S will usually be written with Latin capitals, e.g., F, G, H; in particular, the identity function on S^n, which belongs to $\mathscr{F}_{nn}(S)$, will be denoted by I_n. Any non-empty function over S belongs to a unique $\mathscr{F}_{rd}(S)$; accordingly, we can make the following:

DEFINITION 2. If F is a non-empty function over S, then the *degree*[2] of F is the unique positive integer δF, and the *rank* of F is the unique positive integer ϱF, such that $F \in \mathscr{F}_{\varrho F, \delta F}(S)$. The *index* of F, denoted by ιF, is the difference $\delta F - \varrho F$.

A function of rank 1 will be called a *multiplace function* (over S); and a multiplace function of degree n will be called an *n-place function* (over S). The set of all multiplace functions of degree at least n, together with the empty function $\emptyset$, i.e., the union $\bigcup_{d=n}^{\infty} \mathscr{F}_{1d}(S) \cup \{\emptyset\}$, will be denoted by $\mathscr{M}_n(S)$. Thus $\mathscr{M}_1(S)$ comprises $\emptyset$ and all multiplace functions over S.

If $F \in \mathscr{F}_{\infty}(S)$ is non-empty and $\operatorname{Dom} F$ consists of all of $S^{\delta F}$, then we call F a *transformation* (over S). The set of transformations of degree d and rank r will be denoted by $\mathscr{T}_{rd}(S)$; and the set of all transformations over S will be denoted by $\mathscr{T}_{\infty}(S)$. Note that $\mathscr{T}_{rd}(S)$ is a proper subset of $\mathscr{F}_{rd}(S)$, and that $\mathscr{T}_{\infty}(S)$ is a proper subset of $\mathscr{F}_{\infty}(S)$.

The empty function $\emptyset$ has no intrinsic degree, rank, or index, but we adopt the convention that any suitable combination of these parameters may be attributed to $\emptyset$ as convenience dictates. Another convenient practice we shall follow is the sup-

[2]) Note that many authors (e.g., JÓNSSON and TARSKI [18]) use the term 'rank' to denote what we call 'degree'.

pression, whenever no ambiguity results, of explicit reference to S, and consequent abbreviation of $E(S)$, $\mathscr{F}_{rd}(S)$, $\mathscr{F}_{\infty}(S)$, $\mathscr{M}_n(S)$, $\mathscr{T}_{\infty}(S)$, to E, $\mathscr{F}_{rd}$, $\mathscr{F}_{\infty}$, $\mathscr{M}_n$, $\mathscr{T}_{\infty}$, etc.

It is readily verified that $\mathscr{F}_{\infty}$ is closed under R and L; in fact we have $RI_n = LI_n = I_n$, and $RF \subseteq I_{\delta F}$, $LF \subseteq I_{\varrho F}$ for any non-empty function F over S. Since $\mathscr{F}_{\infty}$ is closed under composition, it follows that $(\mathscr{F}_{\infty}, \circ, L, R)$ is a function system, in fact a subsystem of the primary function semigroup $(\mathscr{F}(E), \circ, L, R)$. However, from our point of view this system is defective. It has no identity element, hence is not a sub-function semigroup of $(\mathscr{F}(E), \circ, L, R)$. Moreover, the composite $F \circ G$ of two functions in $\mathscr{F}_{\infty}$ is non-empty *only* if $\delta F = \varrho G$. These defects can be remedied by modifying the binary operation of natural composition; the next two sections are devoted to this modification.

3. The Set $\mathscr{E}_{\infty}(S)$

DEFINITION 3. For any function F in $\mathscr{F}_{\infty}$, eF is the function in $\mathscr{F}(E)$ specified by the following conditions:

$$\mathrm{Dom}(eF) = E(\mathrm{Dom}\, F); \tag{1}$$

$$(eF)\, x_1 \cdots x_n = (F x_1 \cdots x_{\delta F}) \cdots x_n \quad \text{for any} \quad x_1 \cdots x_n \in \mathrm{Dom}(eF). \tag{2}$$

The set of all functions eF, i.e., the range of the function e, will be denoted by $\mathscr{E}_{\infty}(S)$ (briefly, $\mathscr{E}_{\infty}$).

Note that $e\emptyset = \emptyset$, while eF is a *proper* extension of F whenever F is non-empty.

LEMMA 1. *The function e of Definition 3 is one-to-one from* $\mathscr{F}_{\infty}$ *onto* $\mathscr{E}_{\infty}$.

Proof. Let F and G be distinct non-empty functions over S. If $\mathrm{Dom}\, F \neq \mathrm{Dom}\, G$, then $\mathrm{Dom}(eF) \neq \mathrm{Dom}(eG)$, whence $eF \neq eG$. If $\mathrm{Dom}\, F = \mathrm{Dom}\, G$, then again $eF \neq eG$, since eF and eG extend F and G, respectively, and the same function cannot be an extension of two distinct functions with the same domain. It is obvious that $eF = \emptyset$ if and only if $F = \emptyset$, and the proof of the lemma is complete.

Since e is one-to-one, it has an inverse e^{-1} which maps $\mathscr{E}_{\infty}$ onto $\mathscr{F}_{\infty}$. The notions of degree, rank and index can therefore be extended to elements of $\mathscr{E}_{\infty}$ by setting $\delta f = \delta(e^{-1} f)$, $\varrho f = \varrho(e^{-1} f)$, $\iota f = \iota(e^{-1} f)$, for any non-empty $f \in \mathscr{E}_{\infty}$.

To save parentheses, we shall frequently write $\hat{F}$, $\hat{G}$ instead of eF, eG, etc. Note that $\delta \hat{F} = \delta F$, $\varrho \hat{F} = \varrho F$ and $\iota \hat{F} = \iota F$.

LEMMA 2. *If F is a non-empty function over S, and $x_1 \dots x_n$ is a string in* $\mathrm{Dom}\, \hat{F}$, *then $\hat{F} x_1 \dots x_n$ is a string of length $n - \iota \hat{F} = n - \iota F$. If $y_1 \dots y_m$ is a string in* $\mathrm{Ran}\, \hat{F}$, *then $y_1 \dots y_m$ is the image under $\hat{F}$ of a string of length $m + \iota \hat{F} = m + \iota F$.*

Proof. By definition, if $x_1 \dots x_n \in \mathrm{Dom}\, \hat{F}$, then $n \geqq \delta F$ and $\hat{F} x_1 \dots x_n = (F x_1 \dots x_{\delta F}) \dots x_n$. But $\hat{F} x_1 \dots x_{\delta F}$ is a string of length ϱF, whence $\hat{F} x_1 \dots x_n$ is a string of length $n - \delta F + \varrho F = n - \iota F$. If $y_1 \dots y_m \in \mathrm{Ran}\, \hat{F}$, then $y_1 \dots y_m = \hat{F} x_1 \dots x_n$ for some string $x_1 \dots x_n \in \mathrm{Dom}\, \hat{F}$. Hence $m = n - \iota F$, whence $n = m + \iota F$.

Lemma 2 can be used to prove the following result, a special case of which yields a useful representation for the function e^{-1}.

LEMMA 3. *If F is a non-empty function over S and $n \geqq \delta F$, then $\hat{F} \circ I_n = I_{n-\iota F} \circ \hat{F}$ is a function in $\mathscr{F}_{n-\iota F, n}$. Similarly, for $m \geqq \varrho F$, $I_m \circ \hat{F} = \hat{F} \circ I_{m+\iota F}$ is a function in $\mathscr{F}_{m, m+\iota F}$. In particular,*

$$\hat{F} \circ I_{\delta F} = I_{\varrho F} \circ \hat{F} = I_{\varrho F} \circ \hat{F} \circ I_{\delta F} = F = e^{-1} \hat{F}. \tag{3}$$

Proof. It is easily seen that $\mathrm{Dom}(\hat{F} \circ I_n)$ is the set of all strings of length n in $\mathrm{Dom}\,\hat{F}$; this set is non-empty if $n \geqq \delta F$. By L2, it follows that $\mathrm{Ran}\,(\hat{F} \circ I_n)$ is the set of all strings of length $n - \iota F$ in $\mathrm{Ran}\,\hat{F}$. This set in turn is equal to $\mathrm{Ran}(I_{n-\iota F} \circ \hat{F})$. Applying L2 again, we find that $\mathrm{Dom}(I_{n-\iota F} \circ \hat{F}) = \mathrm{Dom}(\hat{F} \circ I_n)$. Hence $\hat{F} \circ I_n$ and $I_{n-\iota F} \circ \hat{F}$ are equal; and since $\mathrm{Dom}(\hat{F} \circ I_n) \subseteq S^n$, $\mathrm{Ran}(\hat{F} \circ I_n) \subseteq S^{n-\iota F}$, it follows that $\hat{F} \circ I_n \in \mathscr{F}_{n-\iota F, n}$. Similar arguments yield the second statement in the lemma. In particular, $\hat{F} \circ I_{\delta F} = I_{\varrho F} \circ \hat{F}$ is the restriction of $\hat{F}$ to the set of all strings of length δF in $\mathrm{Dom}\,\hat{F}$. But this latter set is precisely $\mathrm{Dom}\,F$. Hence $\hat{F} \circ I_{\delta F} = I_{\varrho F} \circ \hat{F}$ is the restriction of $\hat{F}$ to $\mathrm{Dom}\,F$, which is F. Finally, since $F = I_{\varrho F} \circ F$, we have $F = I_{\varrho F} \circ (\hat{F} \circ I_{\delta F}) = I_{\varrho F} \circ \hat{F} \circ I_{\delta F}$. This proves the lemma.

LEMMA 4. *If F is a non-empty function over S, and m, n are integers such that $\delta F \leqq m \leqq n$, then*

$$\hat{F} x_1 \ldots x_n = (\hat{F} x_1 \ldots x_m) \ldots x_n \quad \textit{for all} \quad x_1 \ldots x_n \in \mathrm{Dom}\,\hat{F}. \tag{4}$$

Proof. By (2), we have

$$\hat{F} x_1 \ldots x_n = (F x_1 \ldots x_{\delta F}) \ldots x_n = ((\hat{F} x_1 \ldots x_{\delta F}) \ldots x_m) \ldots x_n = (\hat{F} x_1 \ldots x_m) \ldots x_n,$$

since concatenation of strings in associative.

LEMMA 5. *The set $\mathscr{E}_\infty$ is closed under natural composition.*

Proof. Let $\hat{F}$ and $\hat{G}$ be in $\mathscr{E}_\infty$. If $\hat{F} \circ \hat{G}$ is empty, then $\hat{F} \circ \hat{G}$ is automatically in $\mathscr{E}_\infty$. Therefore we can suppose $\hat{F} \circ \hat{G}$ non-empty. Set $d = \max(\delta F + \iota G, \delta G)$, and $r = d - \iota F - \iota G = \max(\varrho F, \varrho G - \iota F)$. Then $d \geqq \delta G$, $d - \iota G = r + \iota F = \max(\delta F, \varrho G) \geqq \delta F$ and $r \geqq \varrho F$. Define:

$$H = I_r \circ \hat{F} \circ \hat{G} \circ I_d. \tag{5}$$

By L3, $I_r \circ \hat{F} \in \mathscr{F}_{r, r+\iota F}$ and $\hat{G} \circ I_d \in \mathscr{F}_{d-\iota G, d}$. Since $r + \iota F = d - \iota G$, it follows that H, if non-empty, is in $\mathscr{F}_{rd}$. We next show that $\mathrm{Dom}(\hat{F} \circ \hat{G}) = E(\mathrm{Dom}\,H)$. To this end, let $x_1 \ldots x_n \in \mathrm{Dom}(\hat{F} \circ \hat{G})$. Then $x_1 \ldots x_n \in \mathrm{Dom}\,\hat{G}$ and $\hat{G} x_1 \ldots x_n \in \mathrm{Dom}\,\hat{F}$. Hence $n \geqq \delta G$ and $n - \iota G \geqq \delta F$, whence $n \geqq d$. Since $d \geqq \delta G$, we have $x_1 \ldots x_{\delta G} \in \mathrm{Dom}\,G$, $x_1 \ldots x_d \in \mathrm{Dom}(\hat{G} \circ I_d)$, and $(\hat{G} \circ I_d)\, x_1 \ldots x_d = \hat{G} x_1 \ldots x_d$. From the fact that $\hat{G} x_1 \ldots x_d$ is an initial segment (of length $d - \iota G = r + \iota F$) of the string $\hat{G} x_1 \ldots x_n \in \mathrm{Dom}\,\hat{F}$, it follows that $(\hat{G} \circ I_d)\, x_1 \ldots x_d \in \mathrm{Dom}(I_r \circ \hat{F})$. Thus $x_1 \ldots x_d \in \mathrm{Dom}\,H$, which implies that $\mathrm{Dom}(\hat{F} \circ \hat{G}) \subseteq E(\mathrm{Dom}\,H)$.

Conversely, let $x_1 \dots x_n$ be an extension of $x_1 \dots x_d \in \operatorname{Dom} H$. Then, since $x_1 \dots x_d \in \operatorname{Dom}(\hat{G} \circ I_d) \subseteq \operatorname{Dom} \hat{G}$, we have $x_1 \dots x_n \in \operatorname{Dom} \hat{G}$. Similarly, since $\hat{G} x_1 \dots x_n$ is an extension of $(\hat{G} \circ I_d)\, x_1 \dots x_d \in \operatorname{Dom}(I_r \circ \hat{F}) \subseteq \operatorname{Dom} \hat{F}$, we have $\hat{G} x_1 \dots x_n \in \operatorname{Dom} \hat{F}$. Thus $x_1 \dots x_n \in \operatorname{Dom}(\hat{F} \circ \hat{G})$, whence $E(\operatorname{Dom} H) \subseteq \operatorname{Dom}(\hat{F} \circ \hat{G})$. Therefore $\operatorname{Dom}(\hat{F} \circ \hat{G}) = E(\operatorname{Dom} H) = \operatorname{Dom} \hat{H}$.

Finally, using L4, for any $x_1 \dots x_n \in \operatorname{Dom}(\hat{F} \circ \hat{G}) = \operatorname{Dom} \hat{H}$ we have

$$\begin{aligned}
\hat{H} x_1 \dots x_n &= (H x_1 \dots x_d) \dots x_n \\
&= ((I_r \circ \hat{F} \circ \hat{G} \circ I_d)\, x_1 \dots x_d) \dots x_n \\
&= ((I_r \circ \hat{F})((\hat{G} \circ I_d)\, x_1 \dots x_d)) \dots x_n \\
&= (\hat{F}(\hat{G} x_1 \dots x_d)) \dots x_n \\
&= \hat{F}((\hat{G} x_1 \dots x_d) \dots x_n) \\
&= \hat{F}(\hat{G} x_1 \dots x_n) \\
&= (\hat{F} \circ \hat{G})(x_1 \dots x_n).
\end{aligned}$$

Thus $\hat{F} \circ \hat{G} = \hat{H}$ and the proof is complete.

It is important to note that, unlike the composite of two functions F, G in $\mathscr{F}_\infty$, which is non-empty *only* if $\delta F = \varrho G$, the composite of two functions $\hat{F}, \hat{G}$ in $\mathscr{E}_\infty$ can be non-empty when $\delta F \neq \varrho G$. For example, if F is a transformation, then regardless of the relation between δF and ϱG, $\hat{F} \circ \hat{G}$ is non-empty whenever $\hat{G}$ is non-empty.

Straightforward computation yields:

LEMMA 6. *If F, G are in $\mathscr{F}_\infty$ and if $d = \max(\delta F + \iota G, \delta G)$, $r = \max(\varrho F, \varrho G - \iota F)$, then*

$$d = \delta F + \iota G \Leftrightarrow r = \varrho F \Leftrightarrow \delta F \geqq \varrho G\,; \tag{6}$$

$$d = \delta G \Leftrightarrow r = \varrho G - \iota F \Leftrightarrow \delta F \leqq \varrho G\,; \tag{7}$$

and, in each case, $d - r = \iota F + \iota G$.

LEMMA 7. *If $\hat{F}$ and $\hat{G}$ in $\mathscr{E}_\infty$ are such that $\hat{F} \circ \hat{G}$ is non-empty, then*

$$\delta(\hat{F} \circ \hat{G}) = \max(\delta F + \iota G, \delta G) \tag{8}$$

$$\varrho(\hat{F} \circ \hat{G}) = \max(\varrho F, \varrho G - \iota F) \tag{9}$$

$$\iota(\hat{F} \circ \hat{G}) = \iota F + \iota G. \tag{10}$$

Proof. Let H, d and r be as defined in the proof of L5. Then $\hat{H} = \hat{F} \circ \hat{G}$ whence (8) and (9) follow from $\delta \hat{H} = d = \delta H$ and $\varrho \hat{H} = \varrho H = r$; and (10) now follows from (8) and (9) and L6.

The preceding lemmas can be combined to yield the following useful result:

LEMMA 8. *Let F and G be non-empty functions over S. Then*

$$eF\circ eG=\begin{cases} e(F\circ eG), & \text{if} \quad \delta F \geqslant \varrho G, \\ e(eF\circ G), & \text{if} \quad \delta F \leqslant \varrho G, \\ e(F\ \ G), & \text{if} \quad \delta F = \varrho G. \end{cases} \tag{11}$$

Proof. If $eF\circ eG$ is non-empty, then we consider 3 cases: $\delta F \geqq \varrho G$, $\delta F \leqq \varrho G$, $\delta F = \varrho G$. If $\delta F \geqq \varrho G$, then $\varrho(eF\circ eG)=\varrho F$ by virtue of L6 and L7. Hence, using (3), we obtain

$$eF\circ eG = ee^{-1}(eF\circ eG) = e(I_{\varrho F}\circ eF\circ eG) = e(F\circ eG).$$

If $\delta F \leqq \varrho G$, then $\delta(eF\circ eG)=\delta G$. Therefore we have:

$$eF\circ eG = ee^{-1}(eF\circ eG) = e(eF\circ eG\circ I_{\delta G}) = e(eF\circ G).$$

If $\delta F = \varrho G$, then both $\delta(eF\circ eG)=\delta G$ and $\varrho(eF\circ eG)=\varrho F$. Hence,

$$eF\circ eG = e(I_{\varrho F}\circ eF\circ eG\circ I_{\delta F}) = e(F\circ G).$$

Finally, if $eF\circ eG$ is empty, then (11) is trivial. This completes the proof.

LEMMA 9. *An element $F\in\mathscr{F}_\infty$ is a subidentity if and only if $eF=\hat{F}$ is a subidentity.*

Proof. If $F\in\mathscr{F}_\infty$ is a subidentity, then F is the identity function on Dom F. By D3, it is immediate that eF is the identity function on $E(\operatorname{Dom} F)=\operatorname{Dom}(eF)$, whence eF is a subidentity. Conversely, if $eF\in\mathscr{E}_\infty$ is a subidentity, then $F=I_{\varrho F}\circ eF$ is the composite of two subidentities, and hence is itself a subidentity.

LEMMA 10. *The set $\mathscr{E}_\infty$ is closed under the operations R and L. In fact, for any $F\in\mathscr{F}_\infty$, we have:*

$$R(eF)=e(RF), \qquad L(eF)=e(LF). \tag{12}$$

Proof. Since RF is a subidentity, $e(RF)$ is a subidentity, specifically the identity function on $E(\operatorname{Dom}(RF))$. But by D4 and the definition of R, $E(\operatorname{Dom}(RF))=E(\operatorname{Dom} F)=\operatorname{Dom}(eF)$. Hence $e(RF)=R(eF)$. Similarly, $e(LF)$ is the identity function on $E(\operatorname{Dom}(LF))=E(\operatorname{Ran} F)$, so that it only remains to show that $E(\operatorname{Ran} F)=\operatorname{Ran}(eF)$. This is immediate if F is empty. If F is non-empty, let $y_1\ldots y_m$ be a string in $\operatorname{Ran}(eF)$. Then there is a string $x_1\ldots x_n$ in $\operatorname{Dom}(eF)$ such that $y_1\ldots y_m=(eF)\,x_1\ldots x_n$. But $(eF)\,x_1\ldots x_n=(Fx_1\ldots x_{\delta F})\ldots x_n$, and $Fx_1\ldots x_{\delta F}\in\operatorname{Ran} F$, whence $y_1\ldots y_m\in E(\operatorname{Ran} F)$. Conversely, if $y_1\ldots y_m\in E(\operatorname{Ran} F)$, then $m\geqq\varrho F$ and there is a string $x_1\ldots x_{\delta F}$ in Dom F such that $y_1\ldots y_{\varrho F}=Fx_1\ldots x_{\delta F}$. Hence $y_1\ldots y_{\varrho F}\ldots y_m=y_1\ldots y_m=(Fx_1\ldots x_{\delta F})\ldots y_m=(eF)\,x_1\ldots x_{\delta F}\ldots y_m$, so that $y_1\ldots y_m\in\operatorname{Ran}(eF)$. Hence $\operatorname{Ran}(eF)=E(\operatorname{Ran} F)$ for all $F\in\mathscr{F}_\infty$, and the proof of the lemma is complete.

As an immediate consequence of Lemmas 5 and 10, we have:

LEMMA 11. *The system $(\mathscr{E}_\infty,\circ,L,R)$ is a function semigroup with identity element $\hat{I}_1=eI_1$, and null element $\emptyset$.*

Proof. Since $\mathscr{E}_\infty$ is closed under $\frown$, R, and L, the quadruple $(\mathscr{E}_\infty, \frown, L, R)$ is a function system, a sub-system of the primary function semigroup $(\mathscr{F}(E), \frown, L, R)$. It only remains to show that $\hat{I}_1$ is an identity element for the system. This is immediate, for $\hat{I}_1$ is the identity function on $E(\text{Dom}\, I_1) = E(S)$, whence $\hat{I}_1 = j_{E(S)}$. But $j_{E(S)}$ is the identity element for $(\mathscr{F}(E), \frown, L, R)$ so *a fortiori* $\hat{I}_1$ is the identity element for the subsystem.

In any function system $(\mathscr{S}, \frown, L, R)$ the relation $\subseteq$ defined via:

$$f \subseteq g \quad \text{if and only if} \quad f = g \frown Rf \tag{13}$$

is a partial order on $\mathscr{S}$ (cf. Definition 3 and Theorem 5 in [33]). As remarked in Section 1, in concrete function semigroups or systems, in which the binary operation is natural composition, the relation $\subseteq$ coincides with restriction. In particular, the set $\mathscr{E}_\infty$ is partially ordered by restriction. In addition we have:

LEMMA 12. *The set of subidentities $\{\hat{I}_n\}$ is linearly ordered; in fact, $\hat{I}_m \subseteq \hat{I}_n$ if and only if $n \leqq m$.*

Proof. Since $\hat{I}_n$ is the identity function on $E(\text{Dom}\, I_n) = E(S^n) = \bigcup_{r=n}^{\infty} S^r$, it follows that $\text{Dom}\, \hat{I}_m \subseteq \text{Dom}\, \hat{I}_n$ if and only if $n \leqq m$.

Before going on to the next section, it is instructive to compare the mapping e with a mapping employed by A. S. DAVIS in [10] for similar purposes. DAVIS has no special symbol for his mapping: we shall denote it, *ad hoc*, by $\bar{e}$. The domain of e is $\mathscr{F}_\infty$. The domain of $\bar{e}$ on the other hand, is the union of $\mathscr{T}_\infty$ and $E(S)$. We shall confine our attention to the part common to $\text{Dom}\, e$ and $\text{Dom}\, \bar{e}$, which is $\mathscr{T}_\infty$. For any transformation F, $\text{Dom}(eF) = E(S^{\delta F}) = \bigcup_{n=\delta F}^{\infty} S^n$, while $\text{Dom}(\bar{e}F) = E(S) = \bigcup_{n=1}^{\infty} S^n$. Hence $\text{Dom}(eF) \subseteq \text{Dom}(\bar{e}F)$, with equality holding if and only if $\delta F = 1$. Now consider a string $x_1 \ldots x_n \in S^n$. If $n < \delta F$, $x_1 \ldots x_n \notin \text{Dom}(eF)$, but $x_1 \ldots x_n \in \text{Dom}(\bar{e}F)$, and $(\bar{e}F)x_1 \ldots x_n = Fz_1 \ldots z_{\delta F}$, where the string $z_1 \ldots z_{\delta F}$ is related to the string $x_1 \ldots x_n$ as follows: For $1 \leqq m \leqq \delta F$, $z_m = x_r$, where r is the least positive integer congruent to m modulo n. If $n \geqq \delta F$, then $x_1 \ldots x_n$ is in both $\text{Dom}(eF)$ and $\text{Dom}(\bar{e}F)$; but $(eF)\, x_1 \ldots x_n = (Fx_1 \ldots x_{\delta F}) \ldots x_n$, while $(\bar{e}F)\, x_1 \ldots x_n = Fx_1 \ldots x_{\delta F}$. Thus eF and $\bar{e}F$ agree *only* on $S^{\delta F}$; and $\text{Ran}(\bar{e}F) = \text{Ran}\, F$, while $\text{Ran}(eF)$ is a proper extension of Ran. F. It is readily seen from this that the sets $\text{Ran}\, e$ and $\text{Ran}\, \bar{e}$ are completely disjoint. About the only property the two sets have in common is that both of them are closed under natural composition. Finally, e is one-to-one, while it is easy to give examples that show that $\bar{e}$ is not.

4. Serial Composition

Using the machinery developed in the preceding section, we can now introduce our modification of natural composition in $\mathscr{F}_\infty$.

DEFINITION 4. Let F and G be functions over S. The *serial composite*, $F\sigma G$, of F and G is the function over S specified by:

$$F\sigma G = e^{-1}(eF \circ eG). \tag{1}$$

It follows at once from Lemma 8 that:

$$F\sigma G = \begin{cases} F \circ \hat{G}, & \text{if } \delta F \geqq \varrho G, \\ \hat{F} \circ G, & \text{if } \delta F \leqq \varrho G, \\ F \circ G, & \text{if } \delta F = \varrho G. \end{cases} \tag{2}$$

The expressions in (2) are easily modified so that $F\sigma G$ is always expressed as the natural composite of functions in $\mathscr{F}_\infty$. This results in:

LEMMA 13. *Let F and G be functions in $\mathscr{F}_\infty$. If either F or G is empty, then $F\sigma G=\emptyset$. Otherwise we have:*

$$F\sigma G = \begin{cases} F \circ (I_{\delta F} \circ \hat{G}), & \text{if } \delta F \geqq \varrho G, \\ (\hat{F} \circ I_{\varrho G}) \circ G, & \text{if } \delta F \leqq \varrho G, \\ F \circ G, & \text{if } \delta F = \varrho G; \end{cases} \tag{3}$$

and $I_{\delta F}\circ\hat{G}\in\mathscr{F}_{\delta F, \delta F+\iota G}$, $\hat{F}\circ I_{\varrho G}\in\mathscr{F}_{\varrho G-\iota F, \varrho G}$.

Proof. If either $F=\emptyset$ or $G=\emptyset$, then $eF \circ eG=\emptyset$, whence $F\sigma G=\emptyset$ by (1). If F and G are both non-empty and $\delta F\geqq\varrho G$, then

$$F\sigma G = F \circ \hat{G} = (F\circ I_{\delta F})\circ\hat{G} = F\circ(I_{\delta F}\circ\hat{G})$$

and, by L3, $I_{\delta F}\circ\hat{G}\in\mathscr{F}_{\delta F, \delta F+\iota G}$. This proves the first part of (3); similar arguments establish the rest.

In our announcement [35], the serial composite of F and G was defined in a manner equivalent to the following:

If either $F=\emptyset$ or $G=\emptyset$, then $F\sigma G=\emptyset$.

If F and G are non-empty then set

$$\mathrm{Dom}(F\sigma G) = \{x_1 \dots x_d | x_1 \dots x_{\delta G} \in \mathrm{Dom}\, G,\ y_1 \dots y_{\delta F} \in \mathrm{Dom}\, F\},$$

where $d=\max(\delta F+\iota G, \delta G)$, $x_1 \dots x_{\delta G}$ is the initial segment of length δG of $x_1 \dots x_d$, and $y_1 \dots y_{\delta F}$ is the initial segment of length δF of the string $(Gx_1 \dots x_{\delta G})\dots x_d$. If $\mathrm{Dom}(F\sigma G)$ is empty then $F\sigma G=\emptyset$. Otherwise, $F\sigma G$ is the function on $\mathrm{Dom}(F\sigma G)$ defined by:

$$(F\sigma G)\,x_1 \dots x_d = (F y_1 \dots y_{\delta F}) \dots y_{d-\iota G}, \tag{4}$$

for any $x_1 \dots x_d \in \mathrm{Dom}(F\sigma G)$, where $y_1 \dots y_{d-\iota G}=(Gx_1 \dots x_{\delta G})\dots x_d$.

Lemma 13 can be regarded as a more 'algebraic' version of the above definition. Like that definition, it serves to bring out the basic idea behind serial composition: We can look upon functions over S as devices that process strings of fixed lengths to produce other strings of fixed lengths. If $\delta F = \varrho G$, we can use the 'outputs' from G as 'inputs' to F. If δF does not match ϱG, we cannot do this directly; but we can, and do, proceed by modifying F or G. If $\delta F > \varrho G$, then we modify G; the resulting function $I_{\delta F} \circ \hat{G}$ turns out strings long enough to be accepted by F. If $\delta F < \varrho G$, then we modify F instead; the resulting function $\hat{F} \circ I_{\varrho G}$ now accepts strings from G. In both cases, the modification is the minimum necessary to enable our functions to be successfully connected in series: hence the name 'serial composition'.

After this metaphorical interlude, we turn to the basic properties of serial composition. The most fundamental of these is associativity. In fact, we have the following:

THEOREM 1. *The system* $(\mathscr{F}_\infty, \sigma, L, R)$ *is a function semigroup with identity element* I_1 *and null element* $\emptyset$.

Proof. We shall show that e^{-1} is an isomorphism from $(\mathscr{E}_\infty, \circ, L, R)$ onto $(\mathscr{F}_\infty, \sigma, L, R)$. Let $\hat{F}$, $\hat{G}$ be arbitrary elements of $\mathscr{E}_\infty$. By (1), we have

$$e^{-1}(\hat{F} \circ \hat{G}) = e^{-1}(e\,e^{-1}\hat{F} \circ e\,e^{-1}\hat{G}) = (e^{-1}\hat{F})\sigma(e^{-1}\hat{G}),$$

whence e^{-1}, being one-to-one and onto, is an isomorphism from the semi-group $(\mathscr{E}_\infty, \circ)$ onto the groupoid $(\mathscr{F}_\infty, \sigma)$. Since any homomorphic image of a semigroup with identity and zero is again a semigroup with identity and zero, $(\mathscr{F}_\infty, \sigma)$ is a semigroup with identity $e^{-1}\hat{I}_1 = I_1$ and null element $e^{-1}\emptyset = \emptyset$. Next, using L10, we have

$$e^{-1}(L\hat{F}) = e^{-1}(L(e\,e^{-1}\hat{F})) = e^{-1}(eL(e^{-1}\hat{F})) = L(e^{-1}\hat{F}),$$
$$e^{-1}(R\hat{F}) = e^{-1}(R(e\,e^{-1}\hat{F})) = e^{-1}(eR(e^{-1}\hat{F})) = R(e^{-1}\hat{F}),$$

so that e^{-1} preserves the unary operations L and R. Hence $(\mathscr{F}_\infty, \sigma, L, R)$ is isomorphic to the function semigroup $(\mathscr{E}_\infty, \circ, L, R)$, and therefore is itself a function semigroup.

Associativity of serial composition permits us to define positive integral powers of a function over S in the usual way: for any $F \in \mathscr{F}_\infty$, we set $F^1 = F$, and $F^{n+1} = F\sigma F^n$ for all $n \geqq 1$. If $F^2 = F$ then F is *idempotent*.

When Lemma 7 is restated in terms of serial composition we immediately have:

THEOREM 2. *Let F and G be functions over S. If $F\sigma G$ is non-empty, then*

$$\delta(F\sigma G) = \max(\delta F + \iota G, \delta G), \tag{5}$$

$$\varrho(F\sigma G) = \max(\varrho F, \varrho G - \iota F), \tag{6}$$

$$\iota(F\sigma G) = \iota F + \iota G. \tag{7}$$

Note that $\delta(F\sigma G)$ is independent of ϱF and that $\varrho(F\sigma G)$ is independent of δG. Note also that $\delta F = \varrho G$ if and only if both $\delta(F\sigma G) = \delta G$ and $\varrho(F\sigma G) = \varrho F$.

By induction or appropriate specialization, we arrive at the following corollaries to Theorem 2:

COROLLARY 1. *Let $F_1, F_2, \ldots, F_n$ be functions over S. If the serial composite $F_1 \sigma F_2 \sigma \ldots \sigma F_n$ is non-empty, then*

$$\left.\begin{aligned}\delta(F_1 \sigma F_2 \sigma \ldots \sigma F_n) = \max(\delta F_1 - \varrho F_2 + \delta F_2 - \varrho F_3 + \cdots + \delta F_n, \\ \delta F_2 - \varrho F_3 + \cdots + \delta F_n, \ldots, \delta F_n),\end{aligned}\right\} \tag{8}$$

$$\left.\begin{aligned}\varrho(F_1 \sigma F_2 \sigma \ldots \sigma F_n) = \max(\varrho F_1, \varrho F_1 - \delta F_1 + \varrho F_2, \ldots, \\ \varrho F_1 - \delta F_1 + \varrho F_2 - \delta F_2 + \cdots + \varrho F_n),\end{aligned}\right\} \tag{9}$$

$$\iota(F_1 \sigma F_2 \sigma \ldots \sigma F_n) = \iota F_1 + \iota F_2 + \cdots + \iota F_n. \tag{10}$$

COROLLARY 2. *If F^n is non-empty, then*

$$\delta F^n = \begin{cases} \delta F + (n-1)\,\iota F, & \text{if} \quad \iota F \geqq 0; \\ \delta F, & \text{if} \quad \iota F \leqq 0. \end{cases} \tag{11}$$

$$\varrho F^n = \begin{cases} \varrho F, & \text{if} \quad \iota F \geqq 0; \\ \varrho F - (n-1)\,\iota F, & \text{if} \quad \iota F \leqq 0. \end{cases} \tag{12}$$

$$\iota F^n = n(\iota F). \tag{13}$$

COROLLARY 3. *If F is idempotent then $\delta F^n = \varrho F^n = \delta F = \varrho F$ and $\iota F^n = \iota F = 0$, for all $n \geqq 1$.*

COROLLARY 4. *Let F be a one-place function. If $F \sigma G$ is non-empty, then $\delta(F\sigma G) = \delta G$, $\varrho(F\sigma G) = \varrho G$, $\iota(F\sigma G) = \iota G$. If $G \sigma F$ is non-empty then $\delta(G\sigma F) = \delta G$, $\varrho(G\sigma F) = \varrho G$, $\iota(G\sigma F) = \iota G$.*

COROLLARY 5. *If F is an m-place and G an n-place function, then $F\sigma G$ is empty or is an $(m+n-1)$-place function. Thus each of the sets $\mathcal{M}_n$ of multiplace functions of degree at least n is closed under serial composition.*

COROLLARY 6. *Let F_m be a p_m-place function for $m = 1, 2, \ldots, n$. Then the serial composite $F_1 \sigma F_2 \sigma \ldots \sigma F_n$ is either empty or is a $(p_1 + p_2 + \cdots + p_n - n + 1)$-place function.*

COROLLARY 7. *Let each of the n functions $F_1, F_2, \ldots, F_n$ be a p-place function. Then the serial composite $F_1 \sigma F_2 \sigma \ldots \sigma F_n$ is either empty or an $[n(p-1)+1]$-place function. In particular, if $p = 2$ then $F_1 \sigma F_2 \sigma \ldots \sigma F_n$ is either empty or an $(n+1)$-place function.*

Since $(\mathscr{E}_\infty, \circ, L, R)$ is partially ordered by restriction, the definition:

$$F \subseteq G \text{ in } \mathscr{F}_\infty \Leftrightarrow eF \subseteq eG \text{ in } \mathscr{E}_\infty, \tag{14}$$

induces a partial order in $(\mathscr{F}_\infty, \sigma, L, R)$. Now, by (3.13),

$$eF \subseteq eG \Leftrightarrow eF = eG \circ R(eF).$$

Using D4 and L13, we have

$$eF = eG \circ R(eF) = eG \circ e(RF) = e(G \sigma RF).$$

It follows that:

$$F \subseteq G \text{ in } \mathscr{F}_\infty \Leftrightarrow F = G\sigma RF, \tag{15}$$

i.e., the partial order induced in $(\mathscr{F}_\infty, \sigma, L, R)$ by (14) coincides with the standard function system partial order. However, in $(\mathscr{F}_\infty, \sigma, L, R)$ this standard partial order no longer coincides with restriction. The precise connection between the two different partial orders is brought out by the following discussion.

THEOREM 3. *Let $F, G \in \mathscr{F}_\infty$ be such that F is non empty and $F \subseteq G$, i.e., $F = G\sigma RF$. Then G is non-empty, $\iota F = \iota G$, and $\delta F - \delta G = \varrho F - \varrho G \geqq 0$.*

Proof. Since $G = \emptyset$ implies $G\sigma RF = \emptyset$ for any F, it follows that $G\sigma RF = F \neq \emptyset$ implies $G \neq \emptyset$. Next, since $\iota(RF) = 0$, upon using T2 we obtain:

$$\delta F = \delta(G\sigma RF) = \max(\delta G + \iota(RF), \delta(RF)) = \max(\delta G, \delta F) \geqq \delta G,$$

which implies $\delta F - \delta G \geqq 0$, and

$$\iota F = \iota(G\sigma RF) = \iota G + \iota(RF) = \iota G,$$

from which $\delta F - \delta G = \varrho F - \varrho G$ follows at once.

LEMMA 14. *If F and G are functions over S, then $F \subseteq G$ if and only if either: $F = \emptyset$; or:*

(i) *F and G are both non-empty,*

(ii) *All strings in* Dom *F are extensions of strings in* Dom *G,*

(iii) *$Fx_1 \dots x_{\delta F} = (Gx_1 \dots x_{\delta G}) \dots x_{\delta F}$ for all $x_1 \dots x_{\delta F} \in \mathrm{Dom}\, F$.*

Proof. If $F = \emptyset$, then $RF = \emptyset$, whence for any $G \in \mathscr{F}_\infty$ we have $G\sigma RF = G\sigma\emptyset = \emptyset = F$, i.e., $F \subseteq G$. Conversely, if $F \subseteq G$ for all $G \in \mathscr{F}_\infty$, then in particular $F \subseteq \emptyset$. Hence $F = \emptyset\sigma RF = \emptyset$.

If F is non-empty and $F \subseteq G$, then by T3, G is non-empty and $\delta F \geqq \delta G$. Now let $x_1 \dots x_{\delta F}$ be any string in $\mathrm{Dom}\, F = \mathrm{Dom}(G\sigma RF)$. By (2), $G\sigma RF = \hat{G} \circ RF$, whence it follows that $(RF)\, x_1 \dots x_{\delta F} = x_1 \dots x_{\delta F}$ is in Dom $\hat{G}$. Hence $x_1 \dots x_{\delta F}$ is an extension of a string in Dom G. Using (2) again, we have:

$$\begin{aligned} Fx_1 \dots x_{\delta F} &= (G\sigma RF)\, x_1 \dots x_{\delta F} \\ &= (\hat{G} \circ RF)\, x_1 \dots x_{\delta F} \\ &= \hat{G}((RF)\, x_1 \dots x_{\delta F}) \\ &= \hat{G}x_1 \dots x_{\delta F} = (Gx_1 \dots x_{\delta G}) \dots x_{\delta F}. \end{aligned}$$

This yields (i), (ii), and (iii). Finally, given (i), (ii), and (iii), we can reverse the above arguments to obtain $F = G\sigma RF$, whence $F \subseteq G$. This proves the lemma.

Combining Lemma 14 and the definition of restriction immediately yields:

LEMMA 15. *The partial order defined in $\mathscr{F}_\infty$ via* (15) *is strictly finer than the partial order by restriction, in the sense that:*

(a) *If F is a restriction of G, then $F \subseteq G$;*

(b) *If $F \subseteq G$ and $F \neq \emptyset$, then F is a restriction of G if and only if $\delta F = \delta G$. In particular, if G is any non-empty function and $n > \delta G$, then $G \sigma I_n \subseteq G$, but $G \sigma I_n$ is not a restriction of G.*

Proof. In view of L14 and the definition of restriction, it is clear that only the last statement in (b) requires explicit proof. As for that, we note that, since I_n is a subidentity, $G \sigma I_n \subseteq G$ holds automatically. Since I_n is a transformation and G is non-empty, $G \sigma I_n$ is non-empty. Finally, since $\delta(G \sigma I_n) = n > \delta G$, $G \sigma I_n$ cannot be a restriction of G. Indeed, $\mathrm{Dom}(G \sigma I_n)$ and $\mathrm{Dom}\, G$ are disjoint.

THEOREM 4. *The set of subidentities in $\mathscr{F}_\infty$ forms a lattice under the partial ordering $\subseteq$. The lattice has the maximum element I_1 and minimum element $\emptyset$. If F and G are subidentities, then $F \cap G = F \sigma G = G \sigma F$; and if F and G are non-empty, then $\delta(F \cap G) = \varrho(F \cap G) = \max(\delta F, \delta G)$, $\delta(F \cup G) = \varrho(F \cup G) = \min(\delta F, \delta G)$.*

Proof. The subidentities in any function semigroup form a meet semilattice, with $F \cap G = F \sigma G$ (see T6 of [32]). If F and G are non-empty, then, since $\iota F = \iota G = 0$, we have $\delta(F \sigma G) = \varrho(F \sigma G) = \max(\delta F, \delta G)$. Turning to unions, let F and G be non-empty subidentities (if either F or G is empty, then the existence of $F \cup G$ is trivial), and set $m = \min(\delta F, \delta G)$. Let $(\mathrm{Dom}\, F)_m$, $(\mathrm{Dom}\, G)_m$ be sets the consisting of the initial segments of length m of all strings in $\mathrm{Dom} F$, $\mathrm{Dom} G$, respectively; and let $S_m = (\mathrm{Dom}\, F)_m \cup (\mathrm{Dom}\, G)_m$. Since S_m is a subset of S^m, the function j_{S_m} is a subidentity in $\mathscr{F}_{mm}$. It is readily verified that j_{S_m} is the required union, $F \cup G$, whence we have $\delta(F \cup G) = \varrho(F \cup G) = m = \min(\delta F, \delta G)$. Finally, the inclusions $\emptyset \subseteq F \subseteq I_1$ for any subidentity F are trivial.

The lattice of subidentities of $\mathscr{F}_\infty$ is even complete, as an easy extension of the above proof shows.

Since e^{-1} is order-preserving, we can apply Lemma 12 to obtain:

COROLLARY 1. *The functions I_n form a chain within the lattice of subidentities of $\mathscr{F}_\infty$. Specifically, $I_m \subseteq I_n$ if and only if $m \geqq n$, whence $I_m \cap I_n = I_{\max(m,n)}$ and $I_m \cup I_n = I_{\min(m,m)}$.*

Combining this corollary, Lemma 15, Theorem 7 of [33], and the fact that $RF \subseteq I_{\delta F}$, $LF \subseteq I_{\varrho F}$ for any non-empty F, we obtain:

COROLLARY 2. *If $F \in \mathscr{F}_\infty$ is non-empty, then $F \sigma I_n = F$ if and only if $n \leqq \delta F$, and $I_m \sigma F = F$ if and only if $m \leqq \varrho F$.*

We can use Theorem 4 to justify a convenient definition of zero$^{\text{th}}$ powers of functions over S, as follows:

$$F^0 = R F \cup L F \qquad \text{for any non-empty } F \in \mathscr{F}_\infty . \tag{16}$$

Thus $\delta F^0 = \varrho F^0 = \min(\delta F, \varrho F)$, $\iota F^0 = 0$.

We regard F^0 as generated by F under serial composition just as we regard F^n for $n \geqq 1$ as so generated. This convention will be adhered to without further comment throughout the rest of this paper.

THEOREM 5. *Let $F \in \mathscr{F}_\infty$ be non-empty, and let m and n be positive integers. Then*

$$I_m \sigma F = F \sigma I_n \tag{17}$$

if and only if either:

(i) $m + \delta F = n + \varrho F$; *or*

(ii) $m \leqq \varrho F$ *and* $n \leqq \delta F$.

Moreover, the 2 sides of (17) *are both equal to F if and only if case* (ii) *holds. Cases* (i) *and* (ii) *are not mutually exclusive.*]

Proof. If (ii) holds, then (17) follows immediately from C4.2, and both sides of (17) are equal to F. If (ii) does not hold, but (i) does, then we have $m > \varrho F$, $n > \delta F$ and $m + \iota F = n$. In this case we use (2), L3, and (2) again to obtain:

$$I_m \sigma F = I_m \circ \hat{F} = \hat{F} \circ I_n = F \sigma I_n .$$

Conversely, if (17) holds and the 2 sides of (17) are both equal to F, then (ii) holds by virtue of C4.2. If (17) holds with neither side equal to F, then by C4.2 we have $m > \varrho F$ and $n > \delta F$. Hence we have:

$$\begin{aligned} m + \delta F &= \max(m, \varrho F) + \delta F \\ &= \max(m, \varrho F) + \iota F + \varrho F \\ &= \max(m + \iota F, \delta F) + \varrho F \\ &= \delta(I_m \sigma F) + \varrho F \\ &= \delta(F \sigma I_n) + \varrho F \\ &= \max(\delta F, n) + \varrho F \\ &= n + \varrho F . \end{aligned}$$

Thus (i) holds, and the proof of the theorem is complete.

Application of standard function-system identities now yields the following corollaries:

COROLLARY 1. *If* (17) *holds, then* $R(I_m \sigma F) = RF \sigma I_n$ *and* $L(F \sigma I_n) = I_m \sigma LF$.

COROLLARY 2. *If* $\delta F + \delta G \geqq \varrho G + 1$ *and* $RF = I_{\delta F}$, *then* $R(F \sigma G) = RG \sigma I_{\delta F + \iota G}$. *If* $\varrho F + \varrho G \geqq \delta F + 1$ *and* $LG = I_{\varrho G}$, *then* $L(F \sigma G) = I_{\varrho G - \iota F} \sigma LF$.

COROLLARY 3. *If* $RF = I_{\delta F}$ *and* $RG = I_{\delta G}$, *then* $R(F \sigma G) = I_{\delta(F \sigma G)}$; *i.e.,* $\mathscr{T}_\infty$ *is closed under serial composition. If* $LF = I_{\varrho F}$ *and* $LG = I_{\varrho G}$, *then* $L(F \sigma G) = I_{\varrho(F \sigma G)}$.

COROLLARY 4. *If* $RF = I_{\delta F}$, *then* $RF^n = I_{\delta F^n}$ *for all* $n \geqq 1$; *if* $LF = I_{\varrho F}$, *then* $LF^n = I_{\varrho F^n}$ *for all* $n \geqq 1$.

Note that Corollaries 3 and 4 extend standard results on domains and ranges of composites of 'full' functions and 'onto' functions.

THEOREM 6. *The function semigroup* $(\mathscr{F}_\infty, \sigma, L, R)$ *has the right-subinverse property.*

Proof. Every primary function semigroup has the right-subinverse property. Every function $F \in \mathscr{F}_\infty$ is in the primary function semigroup $(\mathscr{F}(E), \circ, L, R)$. Therefore for every $F \in \mathscr{F}_\infty$ there is a function $g \in \mathscr{F}(E)$ such that $\operatorname{Dom} g = \operatorname{Ran} F$, $\operatorname{Ran} g \subseteq \operatorname{Dom} F$, $F \circ g = LF = Rg$. If $F = \emptyset$, then $g = \emptyset$. If F is non-empty, then $\operatorname{Dom} g \subseteq S^{\varrho F}$, $\operatorname{Ran} g \subseteq S^{\delta F}$. Hence g is a function in $\mathscr{F}_\infty$, and $\delta g = \varrho F$, $\varrho g = \delta F$. By L13, $F \sigma g = F \circ g$. Hence $F \sigma g = LF = Rg$, so that g is a right-subinverse of F in $(\mathscr{F}_\infty, \sigma, L, R)$.

5. Parallel Composition

DEFINITION 5. Let F and G be functions over S. If either F or G is empty, then the *parallel composite*, $F \pi G$, of F and G is $\emptyset$. If F and G are non-empty, let $m = \max(\delta F, \delta G)$. Then $F \pi G$ is the function over S specified via:

$$\operatorname{Dom}(F \pi G) = \{x_1 \dots x_m | x_1 \dots x_{\delta F} \in \operatorname{Dom} F,\ x_1 \dots x_{\delta G} \in \operatorname{Dom} G\}, \tag{1}$$

$$(F \pi G)\, x_1 \dots x_m = F x_1 \dots x_{\delta F}\, G x_1 \dots x_{\delta G} \quad \text{for all} \quad x_1 \dots x_m \in \operatorname{Dom}(F \pi G). \tag{2}$$

THEOREM 7. *If F and G are functions in* $\mathscr{F}_\infty$ *and* $F \pi G$ *is non-empty, then:*

$$\delta(F \pi G) = \max(\delta F, \delta G), \tag{3}$$

$$\varrho(F \pi G) = \varrho F + \varrho G, \tag{4}$$

$$\iota(F \pi G) = \iota F + \iota G - \min(\delta F, \delta G). \tag{5}$$

Proof. Immediate from D5.

COROLLARY. *If* $F \in \mathscr{F}_\infty$, then $F \pi F = F$ *if and only if* $F = \emptyset$. *Thus there are no non-trivial idempotents in* $\mathscr{F}_\infty$ *under parallel composition. In particular, there is no identity element for parallel composition.*

THEOREM 8. *For any F and G in* $\mathscr{F}_\infty$ *we have:*

$$R(F \pi G) = R F \sigma R G. \tag{6}$$

Proof. From (1) it follows that $\operatorname{Dom}(F \pi G)$ is the set of all strings is S^m whose initial segments of length δF (resp., δG) belong to $\operatorname{Dom} F$ (resp., $\operatorname{Dom} G$). Thus,

$$\begin{aligned}\operatorname{Dom}(F \pi G) &= \operatorname{Dom}(F \sigma I_m) \cap \operatorname{Dom}(G \sigma I_m) \\ &= \operatorname{Dom}[R(F \sigma I_m) \sigma R(G \sigma I_m)].\end{aligned}$$

Hence, upon using well-known identities from the theory of function semigroups and the fact that $\delta(R F \sigma R G) = m$, we obtain:

$$\begin{aligned} R(F\,\pi\,G) &= R(F\,\sigma\,I_m)\,\sigma\,R(G\,\sigma\,I_m) \\ &= R\,F\,\sigma\,I_m\,\sigma\,R\,G\,\sigma\,I_m \\ &= (R\,F\,\sigma\,R\,G)\,\sigma\,I_m \\ &= R\,F\,\sigma\,R\,G\,. \end{aligned}$$

COROLLARY. *If* $RF = I_{\delta F}$ *and* $RG = I_{\delta G}$, *then* $R(F\pi G) = I_{\delta(F\pi G)}$. *Hence* $\mathcal{T}_\infty$ *is closed under parallel composition.*

Proof. By (6), $R(F\pi G) = I_{\delta F}\sigma I_{\delta G}$, which in turn is equal to $I_{\max(\delta F,\,\delta G)} = I_{\delta(F\pi G)}$ by virtue of T4 and C4.1.

THEOREM 9. *Parallel composition is associative.*

Proof. By T8, we have:

$$\begin{aligned} R[(F\,\pi\,G)\,\pi\,H] &= R(F\,\pi\,G)\sigma\,R\,H = R\,F\,\sigma\,R\,G\,\sigma\,R\,H \\ &= R\,F\,\sigma\,R(G\,\pi\,H) = R[F\,\pi(G\,\pi\,H)]\,, \end{aligned}$$

whence $\mathrm{Dom}[(F\pi G)\pi H] = \mathrm{Dom}[F\pi(G\pi H)]$. Hence $(F\pi G)\pi H = \emptyset$ if and only if $F\pi(G\pi H) = \emptyset$. If $(F\pi G)\pi H$ and $F\pi(G\pi H)$ are non-empty, they have a common domain and a common degree $d = \max(\delta F, \delta G, \delta H)$. Let $x_1 \dots x_d$ be any string in the common domain. Then we have:

$$\begin{aligned} [(F\,\pi\,G)\,\pi\,H]\,x_1 \dots x_d &= (F\,\pi\,G)\,x_1 \dots x_{\delta(F\pi G)}\,H\,x_1 \dots x_{\delta H} \\ &= F\,x_1 \dots x_{\delta F}\,G\,x_1 \dots x_{\delta G}\,H\,x_1 \dots x_{\delta H} \\ &= F\,x_1 \dots x_{\delta F}\,(G\,\pi\,H)\,x_1 \dots x_{\delta(G\pi H)} \\ &= [F\,\pi(G\,\pi\,H)]\,x_1 \dots x_d\,. \end{aligned}$$

Combining Theorems 7, 8, and 9 yields:

$$R(F_1\,\pi\,F_2\,\pi \dots \pi\,F_n) = R\,F_1\,\sigma\,R\,F_2\,\sigma \dots \sigma\,R\,F_n\,, \tag{7}$$

$$\delta(F_1\,\pi\,F_2\,\pi \dots \pi\,F_n) = \max(\delta F_1, \delta F_2, \dots, \delta F_n)\,, \tag{8}$$

$$\varrho(F_1\,\pi\,F_2\,\pi \dots \pi\,F_n) = \varrho\,F_1 + \varrho\,F_2 + \dots + \varrho\,F_n\,. \tag{9}$$

Parallel composition also has a weak cancellative property which may be formulated as follows:

THEOREM 10. *If* $RF \cup RG \subseteq RH$ *and either* $F\pi H = G\pi H$ *or* $H\pi F = H\pi G$, *then* $F = G$. *If* $F\pi F = G\pi G$, *then* $F = G$.

Proof. If $RF \cup RG \subseteq RH$, then $RF \subseteq RH$ and $RG \subseteq RH$. Hence $\mathrm{R F \sigma R H} = RH\sigma RF = RF$ and $RG\sigma RH = RH\sigma RG = RG$. So if either $F\pi H = G\pi H$ or $H\pi F = H\pi G$, then $RF = RG$. Hence $F = \emptyset$ if and only if $G = \emptyset$. If F and G are non-empty, let

$x_1 \dots x_{\delta F}$ be any string in their common domain. In case $F\pi H = G\pi H$ then $x_1 \dots x_{\delta F}$ is in $\mathrm{Dom}(F\pi H) = \mathrm{Dom}(G\pi H)$, and we have:

$$\begin{aligned} F x_1 \dots x_{\delta F} H x_1 \dots x_{\delta H} &= (F \pi H) x_1 \dots x_{\delta F} \\ &= (G \pi H) x_1 \dots x_{\delta F} \\ &= G x_1 \dots x_{\delta F} H x_1 \dots x_{\delta H} . \end{aligned}$$

Since two strings are equal only when identical, it follows that $Fx_1 \dots x_{\delta F} = Gx_1 \dots x_{\delta F}$ for any $x_1 \dots x_{\delta F} \in \mathrm{Dom}\, F = \mathrm{Dom}\, G$, whence $F = G$. Similarly, $F = G$ in case $H\pi F = H\pi G$. A similar (and simpler) argument establishes the second part of the theorem.

Formula (6) has exhibited one connection between parallel and series composition. The remainder of this section will be devoted to establishing other such connections, beginning with:

THEOREM 11. *If F is non-empty and H is a subidentity such that $\delta H = \delta F$ and $RF \subseteq H$, then for any G:*

$$F \pi G = F \sigma (H \pi G) = (F \sigma H) \pi G . \tag{10}$$

Proof. Let $m = \max(\delta F, \delta G)$. Since $\varrho(H\pi G) = \varrho H + \varrho G = \delta H + \varrho G = \delta F + \varrho G > \delta F$, it follows from T2 that $\delta(F\sigma(H\pi G)) = \delta(H\pi G) = \max(\delta H, \delta G) = m = \delta(F\pi G)$. Next, for any $x_1 \dots x_m \in \mathrm{Dom}(H\pi G)$, we have:

$$(H \pi G) x_1 \dots x_m = x_1 \dots x_{\delta F} G x_1 \dots x_{\delta G} . \tag{11}$$

Thus,

$$\begin{aligned} \mathrm{Dom}(F \sigma (H \pi G)) &= \{x_1 \dots x_m | x_1 \dots x_m \in \mathrm{Dom}(H \pi G),\ x_1 \dots x_{\delta F} \in \mathrm{Dom}\, F\} \\ &= \{x_1 \dots x_m | x_1 \dots x_{\delta H} \in \mathrm{Dom}\, H,\ x_1 \dots x_{\delta G} \in \mathrm{Dom}\, G, \\ &\qquad x_1 \dots x_{\delta F} \in \mathrm{Dom}\, F\} . \end{aligned}$$

But since $RF \subseteq H$ and $\delta H = \delta F$, $\mathrm{Dom}\, F \subseteq \mathrm{Dom}\, H$, so that the last set above is just $\mathrm{Dom}(F\pi G)$. Next, upon using (11) we obtain:

$$F \sigma (H \pi G) x_1 \dots x_m = F x_1 \dots x_{\delta F} G x_1 \dots x_{\delta G} = (F \pi G) x_1 \dots x_m ,$$

which proves that $F\pi G = F\sigma(H\pi G)$. Finally, since $F\sigma H = F$, we have $(F\sigma H)\pi G = F\pi G$.

COROLLARY 1. *For any non-empty F,*

$$F \pi G = F \sigma (I_{\delta F} \pi G) = F \sigma (R F \pi G) . \tag{12}$$

Using the associativity of parallel composition, a straightforward induction yields:

COROLLARY 2. *If all the functions* $G_1, \dots, G_{n-1}$ *are non-empty and if* $H_1, \dots, H_{n-1}$ *are subidentities such that* $\delta H_p = \delta G_p$ *and* $RG_p \subseteq H_p$ *for* $p = 1, \dots, n-1$, *then for any* G:

$$\left.\begin{aligned} G_1 \pi G_2 \pi \dots \pi G_{n-1} \pi G &= G_1 \sigma (H_1 \pi G_2) \sigma [(H_1 \sigma H_2) \pi G_3] \sigma \dots \\ &\qquad \sigma [(H_1 \sigma \dots \sigma H_{n-1}) \pi G] \\ &= G_1 \sigma (I_{m_1} \pi G_2) \sigma \dots \sigma (I_{m_{n-1}} \pi G), \end{aligned}\right\} \tag{13}$$

where $m_p = \max(\delta G_1, \dots, \delta G_p)$ *for* $p = 1, \dots, n-1$.

A relation valid in any function system takes the form $L(F \sigma G) \subseteq LF$ in $\mathscr{F}_\infty$. An application of this to (10) yields:

COROLLARY 3. $L(F \pi G) \subseteq LF$.

THEOREM 12. *If* $F \sigma G$ *is non-empty, and* $\delta F \leqq \varrho G$, *then*

$$F \sigma (G \pi H) = (F \sigma G) \pi H. \tag{14}$$

Proof. If $\delta F \leqq \varrho G$, then $\delta(F \sigma G) = \delta G$ by T2. Applying (12) twice, we obtain:

$$\begin{aligned} F \sigma (G \pi H) &= F \sigma [G \sigma (I_{\delta G} \pi H)] \\ &= (F \sigma G) \sigma (I_{\delta(F \sigma G)} \pi H) \\ &= (F \sigma G) \pi H. \end{aligned}$$

COROLLARY. *If* $\delta F \leqq \varrho G_1$, *then*

$$F \sigma (G_1 \pi G_2 \pi \dots \pi G_n) = (F \sigma G_1) \pi G_2 \pi \dots \pi G_n.$$

We shall refer to the property expressed in Theorem 12 as the (conditional) *interassociativity*[3] of series and parallel composition.

LEMMA 16. *For any three functions* F, G, H *in* $\mathscr{F}_\infty$, *we have:*

$$R((F \pi G) \sigma H) = R((F \sigma H) \pi (G \sigma H)). \tag{16}$$

Proof. Using (6) and several identities from the theory of function semigroups, we obtain:

$$\begin{aligned} R((F \pi G) \sigma H) &= R(R(F \pi G) \sigma H) \\ &= R(RF \sigma RG \sigma H) \\ &= R(RF \sigma H \sigma R(G \sigma H)) \\ &= R(F \sigma H \sigma R(G \sigma H)) \\ &= R(R(F \sigma H) \sigma R(G \sigma H)) \\ &= R(F \sigma H) \sigma R(G \sigma H) \\ &= R((F \sigma H) \pi (G \sigma H)). \end{aligned}$$

---

[3]) In view of the conditional nature of (14), we prefer the relatively neutral term 'interassociativity' to the more common term 'modularity', with or without an adjective.

LEMMA 17. *If either* $(F\pi G)\sigma H$ *or* $(F\sigma H)\pi(G\sigma H)$ *is non-empty, then* $\varrho[(F\pi G)\sigma H] = \varrho[(F\sigma H)\pi(G\sigma H)]$ *if and only if* $\varrho H \leqq \min(\delta F, \delta G)$.

Proof. Using T2 and T7, a simple computation yields:

$$\varrho((F\pi G)\sigma H) = \varrho F + \varrho G + \max(0, \varrho H - \max(\delta F, \delta G)),$$
$$\varrho((F\sigma H)\pi(G\sigma H)) = \varrho F + \varrho G + \max(0, \varrho H - \max(\delta F, \delta G)) + \max(0, \varrho H - \min(\delta F, \delta G)).$$

The two numbers differ by $\max(0, \varrho H - \min(\delta F, \delta G))$, which is 0 if and only if $\varrho H \leqq \min(\delta F, \delta G)$.

THEOREM 13. *If F, G, and H are functions in* $\mathscr{F}_\infty$, *then the right distributive relation*

$$(F\pi G)\sigma H = (F\sigma H)\pi(G\sigma H) \tag{17}$$

holds if and only if: either both sides of (17) *are empty, or* $\varrho H \leqq \min(\delta F, \delta G)$.

Proof. By L16, $\mathrm{Dom}((F\pi G)\sigma H) = \mathrm{Dom}((F\sigma H)\pi(G\sigma H))$ in all cases. Consequently, one side of (17) is empty if and only if the other side is, in which case (17) is trivial. If both sides are non-empty, then by L17, $\varrho H \leqq \min(\delta F, \delta G)$ is a necessary condition for (17) to hold.

To show that the condition is also sufficient, we first note that when $\varrho H \leqq \min(\delta F, \delta G)$, the common value, d say, of $\delta((F\pi G)\sigma H)$ and $\delta((F\sigma H)\pi(G\sigma H))$ is equal to $\iota H + \delta(F\pi G)$, and that $\delta(F\sigma H) = \delta F + \iota H$, $\delta(G\sigma H) = \delta G + \delta H$. Then for any $x_1 \ldots x_d$ in the common domain of $(F\pi G)\sigma H$ and $(F\sigma H)\pi(G\sigma H)$, we have:

$$\begin{aligned}((F\pi G)\sigma H)\, x_1 \ldots x_d &= (F\pi G)\,((H x_1 \ldots x_{\delta H}) \ldots x_d)\\ &= F[(H x_1 \ldots x_{\delta H}) \ldots x_{\delta F + \iota H}]\, G[(H x_1 \ldots x_{\delta H}) \ldots x_{\delta G + \iota H}]\\ &= (F\sigma H)\, x_1 \ldots x_{\delta F + \iota H} (G\sigma H)\, x_1 \ldots x_{\delta G + \iota H}\\ &= [(F\sigma H)\pi(G\sigma H)]\, x_1 \ldots x_d .\end{aligned}$$

COROLLARY. *If* $G_1, G_2, \ldots, G_n$ and H *are functions over S, then*

$$(G_1 \pi G_2 \pi \ldots \pi G_n)\sigma H = (G_1 \sigma H)\pi(G_2\sigma H)\pi \ldots \pi(G_n \sigma H) \tag{18}$$

holds if and only if: either both sides of (18) *are empty, or* $\varrho H \leqq \min(\delta G_1, \delta G_2, \ldots, \delta G_n)$.

REFERENCES

[1] ABEL, N. H., *Untersuchungen der Functionen zweier unabhängig veränderlichen Grössen x und y wie f(x, y), welche die Eigenschaft haben, dass f(z, f(x, y)) eine symmetrische Function von x, y, und z ist*, J. Reine Angew. Math. 1, 11–15 (1826).

[2] ACZÉL, J., *Lectures on Functional Equations and Their Applications* (Academic Press 1966).

[3] BABBAGE, C., *An Essay Towards the Calculus of Functions*, Phil. Trans. Royal Soc. 389–423 (1815); 179–256 (1816).

[4] BELOUSOV, V. D., *Medial Systems of Quasigroups*, Usp. Mat. Nauk 14 (89), 213–214 (1959).
[5] BERMAN, G. and R. J. SILVERMAN, *Embedding of Algebraic Systems*, Pacific J. Math. *10*, 777–786 (1960).
[6] CLIFFORD, A. H. and G. B. PRESTON, *The Algebraic Theory of Semigroups*, I. Amer. Math. Soc. Math. Surveys No. 7 (1961).
[7] COHN, P. M., *Universal Algebra* (Harper & Row 1965).
[8] CURRY, H. B. and R. FEYS, *Combinatory Logic* (North-Holland Publishing Co. 1958).
[9] DAVIES, R. O., *Two Theorems on Essential Variables*, J. London Math. Soc. 41, 333–335 (1966).
[10] DAVIS, A. S., *An Axiomatization of the Algebra of Transformations over a Set*, Math. Annalen 164, 372–377 (1966).
[11] DICKER, R. M., *The Substitutive Law*, Proc. London Math. Soc. *13*, 493–510 (1963).
[12] EVANS, T., *Abstract Mean Values*, Duke Math. J. *30*, 331–347 (1963).
[13] GLUSKIN, L. M., *Positional Operatives*, Dokl. Akad. Nauk SSSR *157*, 767–770 (1964). Mat. Sb. (N.S.) *68* (110), 444–472 (1965).
[14] GRAHAM, R. L., *On n-Valued Functionally Complete Truth-Functions*, J. Symb. Logic *32*, 190–195 (1967).
[15] HARTMANIS, J. and R. E. STEARNS, *Algebraic Structure Theory of Sequential Machines* (Prentice Hall 1966).
[16] HERSCHEL, J. F. W., *Consideration of Various Points of Analysis*, Phil. Trans. Royal Soc. (1814).
[17] HOSSZÚ, M., *On the Explicit Form of n-Group Operations*, Publ. Math. Debrecen *10*, 88–92 (1963).
[18] JÓNSSON, B. and A. TARSKI, *Boolean Algebras with Operators*, Part I, Amer. J. Math. *73*, 891–939 (1951).
[19] KUCZMA, M., *Functional Equations in a Single Variable*, Monografie Matematyczne, Vol. 46 (Warsaw 1968).
[20] LJAPIN, E. S., *Semigroups* (Amer. Math. Soc., 1963 [Translations of Mathematical Monographs, Vol. 3]).
[21] MAL'CEV, A. I., *Foundations of Linear Algebra* (W. H. Freeman 1963).
[22] MENGER, K., *Calculus, a Modern Approach* (Ginn 1955).
[23] MENGER, K., *An Axiomatic Theory of Functions and Fluents. The Axiomatic Method*, L. HENKIN *et al.*, Eds (North-Holland 1959), p. 454–473.
[24] MENGER, K., *On Substitutive Algebra and its Syntax*, Zeits. Math. Logik Grundl. Math. *10*, 81–104 (1964).
[25] MENGER, K. and H. I. WHITLOCK, *Two Theorems on the Generation of Systems of Functions*, Fund. Math. *58*, 229–240 (1966).
[26] NÖBAUER, W., *Über die Darstellung von universellen Algebren durch Funktionenalgebren*, Publ. Math. Debrecen *10*, 151–154 (1963).
[27] POST, E. L., *Introduction to a General Theory of Elementary Propositions*, Amer. J. Math. *43*, 163–185 (1921).
[28] POST, E. L., *The Two-Valued Iterative Systems of Mathematical Logic*, Annals of Math. Studies No. 5 (Princeton 1941).
[29] ROBINSON, J., *General Recursive Functions*, Proc. Amer. Math. Soc. *1*, 703–718 (1950).
[30] SCHEIN, B. M., *Relation Algebras*, Proceedings of the Interuniversity Scientific Symposium on General Algebra (Tartu 1966), p. 130–168.
[31] SCHEIN, B. M., *The Theory of Semigroups as a Theory of Superposition of Multiplace Functions*, ibid., p. 169–190.
[32] SCHWEIZER, B. and A. SKLAR, *The Algebra of Functions*, I, Math. Annalen *139*, 366–382 (1960); II, ibid. *143*, 440–447 (1961); III, ibid. *161*, 171–196 (1965).
[33] SCHWEIZER, B. and A. SKLAR, *Function Systems*, ibid. *172*, 1–16 (1967).
[34] SCHWEIZER, B. and A. SKLAR, *A Mapping Algebra with Infinitely Many Operations*, Colloq. Math. *9*, 33–38 (1962).
[35] SCHWEIZER, B. and A. SKLAR, *The Algebra of Multiplace Vector-Valued Functions*, Bull. Amer. Math. Soc. *73*, 510–515 (1967).

[36] SIERPIŃSKI, W., *Algèbre des ensembles*, Monografie Matematyczne, Vol. 23 (Warsaw 1957).
[37] SWIFT, J. D., *Algebraic Properties of N-Valued Propositional Calculi*, Amer. Math. Monthly *59*, 612–620 (1952).
[38] WEBB, D. L., *Generation of n-Valued Logic by One Binary Operation*, Proc. Nat. Acad. Sci. *21*, 252–254 (1935).
[39] WHEELER, R. F., *Complete Connectives for the 3-Valued Propositional Calculus*, Proc. London Math. Soc. (3) 16, 167–191 (1966).
[40] WHITLOCK, H. I., *A Composition Algebra for Multiplace Functions*, Math. Annalen 157, 167–178 (1964).
[41] ZUPNIK, D., *Polyadic Semigroups*, Publ. Math. Debrecen *14*, 273–279 (1967).

The University of Massachusetts
Illinois Institute of Technology

The Numerical Range of a Continuous Mapping of a Normed Space

F. F. Bonsall, B. E. Cain, and Hans Schneider[1]) (Edinburgh and Madison)

(Dedicated to A. Ostrowski on the occasion of his 75th birthday)

1. Introduction

Let X denote a normed linear space over the real or complex field, X' the dual space of X, $G(X)$ the set of all continuous mappings of $S(X)$ into X, where $S(X)$ is the unit sphere of X, i.e. the set of all $x \in X$ such that $\|x\| = 1$. Given $x \in S(X)$, let $D(x) = \{f \in X' : f(x) = \|f\| = 1\}$; and, given $T \in G(X)$, let $V(T, x) = \{f[T(x)] : f \in D(x)\}$. The *numerical range* $V(T)$ of a continuous mapping $T \in G(X)$ is defined by

$$V(T) = \bigcup \{V(T, x) : x \in S(X)\}.$$

In the special case when X is a Hilbert space and $D(x)$ can be identified with $\{x\}$, the numerical range of a linear operator has a long history [7]. Under the name 'field of values', the concept has been extended by F. L. Bauer [1] to linear operators on all finite dimensional normed linear spaces. The numerical range of a linear operator on a semi-inner-product space has been studied by G. Lumer [4]. A normed linear space X has, in general, many semi-inner-products that correspond to the norm of X. The choice of one of these semi-inner-products corresponds to the choice of a mapping $x \to f_x$ of $S(X)$ into X' such that $f_x \in D(x)$ for each x. Then the numerical range $W(T)$ for this semi-inner-product is given by

$$W(T) = \{f_x(Tx) : x \in S(X)\}.$$

Thus if T is a continuous linear operator, $V(T)$ is the union of all the numerical ranges $W(T)$ in the sense of Lumer.

When T is a continuous linear operator it is classical that $V(T)$ is a convex set if X is a Hilbert space [7], but an example is given in [5] of a linear operator on a two dimensional normed linear space for which $V(T)$ is not convex. Our main result is that $V(T)$ is connected for every normed linear space X and every $T \in G(X)$ (unless both X is the real numbers and $T(-\alpha) \neq -T(\alpha)$ where $S(X) = \{\alpha, -\alpha\}$). We give two proofs of this result and include an example of a (continuous) linear operator on a real or complex two dimensional semi-inner-product space for which $W(T)$ is not connected.

---

[1]) This research was supported in part by NSF Grant GP-7073.

Received November 13, 1967

Our first proof of the connectedness of $V(T)$ depends on the upper semi-continuity of the set-valued mapping $x \to D(x)$ with respect to the weak* topology on X'. We show that this mapping is also upper semi-continuous for the norm topology on X' when $X=(c_0)$, the sup norm space of sequences converging to 0, (and also of course when X has finite dimension) but not when $X=(c)$, the sup norm space of convergent sequences, or for certain other spaces. As an alternative to upper semi-continuity we also work with a topology on the set of subsets of X', which, although it is probably less familiar than upper semi-continuity, permits us to use the class of continuous functions and their widely studied properties in place of the class of upper semi-continuous mappings. Our second proof shows that when X is not the real numbers $P=\{(x,f)\in X\times X' : x\in S(X), f\in D(x)\}$ is connected in certain topologies, and this may be of interest in itself.

2. Connectedness of the numerical range

Let $\mathscr{P}(U)$ denote the set of all subsets of the set U. If E is a topological space, let $\{\mathscr{P}(U): U\subseteq E;\ U \text{ is open}\}$ be a basis for the *τ-topology* on $\mathscr{P}(E)$. Adjectives used with reference to the τ-topology will bear the prefix 'τ-', e.g. τ-open. A mapping $x\to A(x)$ of a topological space F into the set of subsets of a topological linear space E is *upper semi-continuous (usc)* on F if and only if for every $x\in F$ and every neighbourhood U of 0 in E there exists a neighbourhood V of x such that for all $y\in V$, $A(y)\subseteq A(x)+U$ (cf. [6], pages 35–36). There are other definitions of upper semi-continuity currently in use (cf. [2]). In fact, what we are calling τ-continuous is sometimes called upper semi-continuous.

LEMMA 1. *Let F be a topological space and let E be a topological linear space. If the mapping $x\to A(x)$ is τ-continuous, then it is usc. If for every x in F, $A(x)$ is a compact subset of E, then the function A is τ-continuous if and only if it is usc.*

Proof: Assume A is τ-continuous. If $x\in F$ and U is an open neighbourhood of 0 in E, then, since $\mathscr{P}(A(x)+U)$ is a τ-open subset of $\mathscr{P}(E)$, it follows that $A^{-1}[\mathscr{P}(A(x)+U)]=\{y\in F: A(y)\subseteq A(x)+U\}$ is an open subset of F. Hence $A^{-1}[\mathscr{P}(A(x)+U)]$ is a V whose existence is required in the definition of usc, and A is usc.

Assume that A is usc. Let $\mathscr{P}(U)$, with U an open subset of E, be a basic τ-open set. If $x\in A^{-1}(\mathscr{P}(U))$ then $A(x)\subseteq U$. The compactness hypothesis now provides a neighbourhood of 0 in E, denoted G, such that $A(x)+G\subseteq U$ (see [3] pages 35 and 36 for the details). Since A is usc there is a neighbourhood V of x such that for each $y\in V$, $A(y)\subseteq A(x)+G\subseteq U$. That is, $A(y)\in\mathscr{P}(U)$. Thus $x\in V\subseteq A^{-1}[\mathscr{P}(U)]$. So $A^{-1}[\mathscr{P}(U)]$ is open and A is τ-continuous.

An application of the Hahn-Banach Theorem shows that for each $x\in S(X)$ we have $D(x)\neq\emptyset$. In the weak* topology, $D(x)$ is a closed subset of the (solid) unit ball in X' and hence is compact (cf. [3] page 155). Since $D(x)$ is convex, $D(x)$ is connected in any topology which makes X' a topological linear space, because in any such topology $\alpha \to \alpha f + (1-\alpha) g$, $0 \leqslant \alpha \leqslant 1$ is a continuous function.

LEMMA 2. *Let $S(X)$ have the norm topology and let X' have the weak* topology. Then the mapping $x \to D(x)$ is τ-continuous and usc.*

Proof: Since $D(x)$ is compact in the weak* topology, Lemma 1 shows that it is sufficient to prove that $x \to D(x)$ is usc. Suppose that the mapping is not usc. Then there exist $x \in S(X)$ and a weak* neighbourhood U of 0 in X' such that for every positive integer n there exists $y_n \in S(X)$ and $f_n \in D(y_n)$ satisfying $\|y_n - x\| < 1/n$ and $f_n \notin D(x) + U$. Since $\|f_n\| = 1$, there exists a weak* cluster point g of $\{f_n\}$ with $\|g\| \leqslant 1$. Then

$$|g(x) - 1| \leqslant |g(x) - f_n(x)| + |f_n(x) - f_n(y_n)| \leqslant |g(x) - f_n(x)| + \|x - y_n\| .$$

Since $\|x - y_n\| < 1/n$ and g is a weak* cluster point of $\{f_n\}$, the righthand side can be made arbitrarily small by correctly choosing n, and so $g(x) = 1$. Therefore $g \in D(x)$. However, since g is a weak* cluster point of $\{f_n\}$ and U is a weak* neighbourhood of 0, we have $f_n \in g + U \subseteq D(x) + U$ for some n, which is contradictory.

LEMMA 3. *Let $T \in G(X)$, and let the scalar field have its usual topology. Then the mapping $x \to V(T, x)$ is a τ-continuous and usc mapping of $S(X)$ with the norm topology into the set of subsets of the scalar field.*

Proof: Observe that $V(T, x)$, for $x \in S(x)$ and $T \in G(X)$, is compact, because it is the image of the weak* compact set $D(x)$ under the weak* continuous mapping $f \to f(T(x))$. Therefore, by Lemma 1, it suffices to prove that $x \to V(T, x)$ is usc.

Let $x \in S(X)$ and $\varepsilon > 0$, and let $U = \{g \in X' : |g(T(x))| < \varepsilon/2\}$. Then U is a weak* neighbourhood of 0, and so, by Lemma 2 and the continuity of T, we may choose $\delta > 0$ such that for every $y \in S(X)$ with $\|x - y\| < \delta$ it follows that $\|T(x) - T(y)\| < \varepsilon/2$ and $D(y) \subseteq D(x) + U$. So if $y \in S(X)$, $\|x - y\| < \delta$, and $f \in D(y)$ then $f = g + u$ for some $g \in D(x)$ and $u \in U$. Since $g(T(x)) \in V(T, x)$ the distance from $f(T(y))$ to $V(T, x)$ is at most $|f(T(y)) - g(T(x))|$ and $|f(T(y)) - g(T(x))| \leqslant |f(T(y)) - f(T(x))| + |u(Tx)| < \|T(y) - T(x)\| + \varepsilon/2 < \varepsilon$. But $f(T(y))$ was an arbitrary point of $V(T, y)$, and so $V(T, y) \subseteq V(T, x) + \{t \in \text{scalar field} : |t| < \varepsilon\}$. Thus $x \to V(T, x)$ is usc.

The two proofs of the connectedness of the numerical range which we give use the connectedness of $S(X)$. Since $S(X)$ is disconnected only when X is $\mathbf{R}$, the real numbers, that case is treated separately. In fact when $X = \mathbf{R}$, $V(T) = \{1/\alpha\, T(\alpha)\} \cup \{-1/\alpha\, T(-\alpha)\}$ where $S(X) = \{\alpha, -\alpha\}$. This gives:

PROPOSITION. If $X = \mathbf{R}$, $V(T)$ is connected if and only if $T(-\alpha) = -T(\alpha)$, where $S(X) = \{\alpha, -\alpha\}$. In particular, if T is linear, $V(T)$ is connected.

Both proofs also use the following fact: If $\{L_x\}$ is a family of connected subsets of some topological space and if $G_1 \cup G_2 = \bigcup L_x$ is a decomposition of $\bigcup L_x$ into two non-empty disjoint sets, G_1 and G_2, open in the relative topology, then for each x the entire set L_x lies in either G_1 or G_2.

THEOREM 1. *Let $T \in G(X)$. If $X \neq \mathbf{R}$ then $\mathbf{V}(T)$ is connected.*

Proof: Suppose $X \neq \mathbf{R}$ and $V(T)$ is disconnected. Then $V(T) \subseteq H_1 \cup H_2$, where H_1 and H_2 are open sets giving a decomposition $V(T) = G_1 \cup G_2$ of $V(T)$ into disjoint, non-empty, relatively open sets $G_i = V(T) \cap H_i$ for $i=1$ and 2. By Lemma 3, the mapping $x \to V(T, x)$ is τ-continuous, and therefore, for $i=1$ and 2, the inverse image U_i of the τ-open set $\mathscr{P}(H_i)$ under $x \to V(T, x)$ is an open subset of $S(X)$, the domain of the mapping. For $x \in S(X)$, the set $V(T, x)$ is connected, being the image of the set $D(x)$, which is connected in the norm topology, under the norm continuous mapping $f \to f(Tx)$. Hence, by the sentence preceding the theorem, $V(T, x) \subset G_1 \subseteq H_1$ or $V(T, x) \subseteq G_2 \subseteq H_2$. Thus $x \in U_1$ or $x \in U_2$, but not both. We deduce that $S(X) = U_1 \cup U_2$, where $U_1 \cap U_2 = \emptyset$. But this is impossible, since $S(X)$ is connected, and the theorem is proved.

Here we begin our second proof of the connectedness of $V(T)$. Let P denote $\{(x, f) \in X \times X' : x \in S(X), f \in D(x)\}$. We shall first prove the following general theorem:

THEOREM 2. *Let $X \neq \mathbf{R}$, and let X have the norm topology. Let X' have a topology satisfying*

(a) *X' is a topological linear space,*

(b) *$x \to D(x)$ is usc for all $x \in S(X)$,*

(c) *$D(x)$ is compact, for all $x \in S(X)$.*

Then P is connected as a subset of $X \times X'$ with the product topology.

We shall then deduce:

COROLLARY. Let $X \neq \mathbf{R}$, and let $X \times X'$ be topologized by the product of the norm topology and the weak* topology. Then P is connected, as a subset of $X \times X'$.

Finally we shall show that this implies Theorem 1.

Proof of Theorem 2: We shall first show that every sequence $\{f_i\}$ in X' which is eventually in every neighbourhood of $D(x)$ has a limit point g in $D(x)$. For suppose this is false, then $D(x)$ has an open covering by sets each containing only finitely many f_i's. Since $D(x)$ is compact it has a finite covering by such sets. But the union of the sets in this finite covering is a neighbourhood of $D(x)$ containing only finitely many f_i's. Since this is a contradiction, the f_i's must have a limit point g in $D(x)$.

Let $\pi: P \to S(X)$ denote the projection mapping $(x, f) \to x$. Then π is a closed mapping. For suppose K is a closed subset of P and x is a limit point of $\pi(K)$. Then there exists $\{(x_i, f_i)\} \subset K$ such that $x = \lim x_i = \lim \pi(x_i, f_i)$. Since $t \to D(t)$ is usc, $\{D(x_i)\}$ is eventually within each neighbourhood of $D(x)$. Since $f_i \in D(x_i)$, $\{f_i\}$ is eventually in each neighbourhood of $D(x)$ and hence has a limit point g in $D(x)$. Thus $(x, g) \in \overline{\{(x_i, f_i)\}} \subseteq K$, whence $x \in \pi(K)$, and so $\pi(K)$ is closed.

Suppose that $P=G_1\cup G_2$ where the G_i's are non-empty, disjoint, open and closed subsets of P. For each $x\in S(X)$ the set $\pi^{-1}(x)$ is connected because it is homeomorphic to $D(x)$. The set $\pi^{-1}(x)$ must thus be a subset of either G_1 or G_2. It follows that $\pi(G_1)\cap\pi(G_2)=\emptyset$. Since $\pi(G_1)$ and $\pi(G_2)$ are closed and cover $S(X)$, this contradicts the connectedness of $S(X)$. Hence no such G_i's can exist and therefore P is connected.

The Corollary now follows immediately from Theorem 2, in view of Lemma 2 and the compactness of $D(x)$ in the weak* topology.

Let $X\times X'$ have the product topology formed from the norm topology of X and the weak* topology of X'. If X is infinite dimensional, it can be shown that the mapping $(x,f)\to f(x)$ is not continuous on $X\times X'$, because it is unbounded on every open subset of $X\times X'$. However:

LEMMA 4. *Let F be a norm-bounded subset of X'. Let $X\times X'$ have the product topology formed from the norm topology of X and the weak* topology of X'. Then the mapping $(x,f)\to f(x)$ defined on $X\times X'$ is a continuous mapping of the set $X\times F$ with the relative topology.*

Proof: Suppose that F is contained in a ball of radius r centered at the origin of X'. If (x_i,f_i) is a net in $X\times F$,

$$\begin{aligned}|f_i(x_i)-f(x)| &\leqslant |f_i(x-x_i)|+|f_i(x)-f(x)|\\ &\leqslant r\,\|x-x_i\|+|f_i(x)-f(x)|.\end{aligned}$$

Thus $f_i(x_i)$ will converge to $f(x)$ if both $x_i\to x$ in the norm topology and $f_i\to f$ in the weak* topology, i.e. if $(x_i,f_i)\to(x,f)$ in $X\times X'$.

Second Proof of Theorem 1. Let $X\times X'$ be topologized as it was in the Corollary to Theorem 2 and Lemma 4. We view the mapping $(x,f)\to\ [T(x)]$ defined on P as the composition of the continuous functions $(x,f)\to(T(x),f)\to f[T(x)]$. (Lemma 4 shows the continuity at the final step.) Since by the Corollary to Theorem 2, P is connected, it follows that $V(T)$ must be connected, as it is the image of P under a continuous function.

However, the numerical range $W(T)$ need not be connected:

EXAMPLE. Let X be $\mathbf{R}^2$ or $\mathbf{C}^2$ with the norm given for each $x=(\xi_1,\xi_2)\in X$ by $\|x\|=\max(|\xi_1|,|\xi_2|)$. Given $a=(\alpha_1,\alpha_2)\in S(X)$, and $x=(\xi_1,\xi_2)\in X$, let $f_a(x)$, T be defined by

$$f_a(x)=\begin{cases}\bar{\alpha}_1\xi_1 & \text{if } |\alpha_1|=1,\\ \bar{\alpha}_2\xi_2 & \text{if } |\alpha_1|<1.\end{cases}\qquad Tx=(\xi_1,0).$$

Then $f_a\in D(a)$, and T is a continuous linear operator. Also

$$f_a(Ta)=\begin{cases}1 & \text{if } |\alpha_1|=1\\ 0 & \text{if } |\alpha_1|<1.\end{cases}$$

Therefore the numerical range $W(T)$ in the sense of Lumer for the semi-inner-product space corresponding to the mapping $a \to f_a$ is the set with exactly two elements, 1 and 0.

3. Upper semi-continuity and τ-continuity of the mapping $x \to D(x)$ with the norm topology on $\mathbf{X}'$.

If X has finite dimension, the norm topology coincides with the weak* topology on X', and so, by Lemma 2, the mapping $x \to D(x)$ is upper semi-continuous and τ-continuous with respect to the norm topology on X'. We show that the mapping $x \to D(x)$ is also upper semi-continuous and τ-continuous in this sense when $X=(c_0)$, but not for certain other spaces X including the space (c).

Let $\mathbf{F}$ denote either the real or the complex field, and $\mathbf{P}$ the set of all positive integers. We denote by (m), as usual, the Banach space of all bounded mappings of $\mathbf{P}$ into $\mathbf{F}$ with the *sup norm*

$$\|x\|_\infty = \sup\{|x(n)| : n \in \mathbf{P}\} \qquad (x \in (m)),$$

and by (c) and (c_0) the subspaces of (m) consisting of all sequences that converge and converge to zero respectively. Also, as usual, we denote by (l_1) the Banach space of all mappings x of $\mathbf{P}$ into $\mathbf{F}$ such that $\|x\|_1 = \sum_{n=1}^{\infty} |x(n)| < \infty$, normed by $\|\cdot\|_1$.

THEOREM 3. *Let $X=(c_0)$. Then the mapping $x \to D(x)$ is a usc and τ-continuous mapping of $S(X)$ into subsets of X' with respect to the norm topologies in X and X'.*

Proof: To each element f of X' corresponds a sequence $\{\lambda_k\}$ of elements of $\mathbf{F}$ such that $\sum_{k=1}^{\infty} |\lambda_k| = \|f\|$ and

$$f(x) = \sum_{k=1}^{\infty} \lambda_k x(k) \qquad (x \in X).$$

Given $x \in S(X)$, let $E_x = \{k \in \mathbf{P} : |x(k)| = 1\}$. Then E_x is a non-empty finite set.

Let $a \in S(X)$ and $\varepsilon > 0$. Since the set $\{k \in \mathbf{P} : |a(k)| \geq 1/2\}$ is finite

$$\sup\{|a(k)| : k \in \mathbf{P} \backslash E_a\} = 1 - \eta$$

with $\eta > 0$. Choose δ with $0 < \delta < \min(\varepsilon, \eta)$, let $b \in S(X)$ with $\|b-a\|_\infty < \delta$, and let $f \in D(b)$. The sequence $\{\lambda_k\}$ corresponding to f satisfies

$$\sum_{k=1}^{\infty} |\lambda_k| = \sum_{k=1}^{\infty} \lambda_k b(k) = 1.$$

Therefore $\lambda_k = 0$ $(k \in \mathbf{P} \backslash E_b)$ and $\lambda_k = |\lambda_k| \overline{b(k)}$ for all k. Let $\mu_k = |\lambda_k| \overline{a(k)}$ $(k \in \mathbf{P})$, and let g be the element of X' corresponding to the sequence $\{\mu_k\}$. Since $\|b-a\|_\infty < \eta$, we have

$E_b \subseteq E_a$, and therefore $|a(k)|=1$ whenever $\lambda_k \neq 0$. Therefore

$$\|g\| = \sum_{k=1}^{\infty} |\mu_k| = \sum_{k=1}^{\infty} |\lambda_k| = 1\,,$$

$$g(a) = \sum_{k=1}^{\infty} \mu_k\, a(k) = \sum_{k=1}^{\infty} |\lambda_k| = 1\,.$$

Thus $g \in D(a)$. Also

$$\|f-g\| = \sum_{k=1}^{\infty} |\lambda_k - \mu_k|$$
$$= \sum_{k=1}^{\infty} |\lambda_k|\, |\overline{b(k)} - \overline{a(k)}| \leqslant \|b-a\|_\infty < \varepsilon\,.$$

Since f is an arbitrary element of $D(b)$, $g \in D(a)$, and $f=g+(f-g)$, this proves that $D(b) \subseteq D(a)+\{h \in X' : \|h\| < \varepsilon\}$. Thus $x \to D(x)$ is usc with respect to the norm topologies in X and X'. Furthermore $x \to D(x)$ will be τ-continuous if each $D(x)$ is compact. For each $x \in S(X)$, $D(x)$ is a subset of $\{f \in S(X') : \text{support}\,(f) \subseteq E_x\}$ which is homeomorphic to the compact set $S(\mathbf{F}^k)$, where k is the order of E_x and $\mathbf{F}^k$ has the norm $\|(t_1, \ldots, t_k)\| = \sum |t_i|$. Therefore since $D(x)$ is closed in the norm topology, it is compact.

That P is connected when $X=(c_0)$ and X' has the norm topology is a special case of Theorem 2.

THEOREM 4. *Let X be a linear subspace of (m) such that $(c) \subseteq X$. Then the mapping $x \to D(x)$ is not upper semi-continuous with respect to the norm topologies in X and X' (and therefore not τ-continuous).*

Proof: Given $n \in P$, let e_n, a, b_n denote the elements of X defined by

$$e_n(k) = \begin{cases} 1 & \text{if} \quad k=n \\ 0 & \text{if} \quad k \neq n \end{cases}, \qquad a(k) = 1 - \frac{1}{k}\ (k \in \mathbf{P}), \qquad b_n(k) = \begin{cases} a(k) & \text{if} \quad k<n \\ 1 & \text{if} \quad k \geqslant n \end{cases}.$$

Let f_n be the element of X' defined by $f_n(x)=x(n)$ $(x \in X)$. Given $g \in D(a)$, we have

$$g(e_n) = 0 (n \in \mathbf{P});$$

for we have $\|a+\xi e_n\|_\infty = 1$ whenever $|\xi| \leqslant 1/n$, and so

$$1 \geqslant |g(a+\xi e_n)| = |1+\xi g(e_n)| \qquad (|\xi| \leqslant 1/n),$$

which is impossible unless $g(e_n)=0$. Therefore, for all $g \in D(a)$,

$$\|f_n - g\| \geqslant |(f_n - g)(e_n)| = 1 \qquad (n \in \mathbf{P}).$$

However $f_n \in D(b_n)$ and $\|b_n - a\|_\infty = 1/n$. Thus b_n tends to a, but $D(b_n) \nsubseteq D(a)+U_1$, where $U_1 = \{f \in X' : \|f\| < 1\}$, and the result follows.

THEOREM 5. *Let $X=(l_1)$. Then the mapping $x \to D(x)$ is not upper semi-continuous with respect to the norm topologies in X and X' (and therefore not τ-continuous).*

Proof: Given $n \in \mathbf{P}$, let a, b_n be the elements of $S(X)$ defined by

$$a(k) = 1/2^k \quad (k \in \mathbf{P}), \qquad b_n(k) = \begin{cases} 2^n/(2^n - 1)2^k & (k \leqslant n) \\ 0 & (k > n). \end{cases}$$

Then

$$\|a - b_n\|_1 = \frac{1}{2^n - 1} \sum_{k=1}^{n} \frac{1}{2^k} + \sum_{k=n+1}^{\infty} \frac{1}{2^k} = \frac{1}{2^{n-1}}.$$

Let f_n be the functional defined by

$$f_n(x) = \sum_{k=1}^{n} x(k) \qquad (x \in X).$$

and let $g \in D(a)$. Since $a(k) > 0 \, (k \in \mathbf{P})$, we have

$$g(x) = \sum_{k=1}^{\infty} x(k) \qquad (x \in X).$$

Also $f_n \in D(b_n)$ and $\| f_n - g\| = 1$ $(n \in \mathbf{P})$. Thus the mapping $x \to D(x)$ is not upper semi-continuous with respect to the norm topologies in X and X'.

Note added in proof: Our lemma 2 is known, cf. Theorem 4.3, D. F. CUDIA: *The Geometry of Banach spaces. Smoothness.* Trans. Amer. Math. Soc. *110*, 284–314 (1964).

REFERENCES

[1] BAUER, F. L., *On the Field of Values Subordinate to a Norm*, Numer. Math. *4*, 103–111 (1962).
[2] BERGE, C., *Topological Spaces* (The MacMillan Company, New York 1963).
[3] KELLEY, J. L., NAMIOKA, I., and co-authors, *Linear Topological Spaces* (D. van Nostrand Company, New York 1963).
[4] LUMER, G., *Semi-Inner-Product Spaces*, Trans. Amer. Math. Soc. *100*, 29–43 (1961).
[5] NIRSCHL, N. and SCHNEIDER, H., *The Bauer Fields of Values of a Matrix*, Numer. Math. *6*, 355–365 (1964).
[6] RICKART, C. E., *General Theory of Banach Algebras* (D. van Nostrand Company, New York 1960).
[7] STONE, M. H., *Linear Transformations in Hilbert Space and Their Applications to Analysis* (Amer. Math. Soc., New York 1932 [Col. Publ., Nr. 15]).

University of Edinburgh
University of Wisconsin

Eigenvectors Obtained from the Adjoint Matrix

D. Ž. Djoković[1]) (Waterloo, Ont. Canada)

Dedicated to A. M. Ostrowski on his 75th birthday

1. Introduction

Let A be an n-square complex matrix. By A_k we denote the k-th column of A and similarly for other matrices. Let $\lambda_1, \lambda_2, \ldots$ be the distinct eigenvalues of A,

$$f(\lambda) = (\lambda - \lambda_1)^{r_1} (\lambda - \lambda_2)^{r_2} \ldots$$

the characteristic polynomial of A, and $B(\lambda)$ the adjoint matrix of $\lambda I - A$. Then

$$(\lambda I - A) \, B(\lambda) = f(\lambda) \, I. \tag{1}$$

Putting $\lambda = \lambda_1$ in (1) we get

$$(\lambda_1 I - A) \, B(\lambda_1) = 0.$$

We infer that each non-zero column of $B(\lambda_1)$ is an eigenvector of A belonging to λ_1. This simple remark appears in Gantmacher [2], p. 85–86. Let us suppose that some column, say the first column, of $B(\lambda_1)$ is zero. Let $(\lambda - \lambda_1)^{k_1}$ be the highest power of $\lambda - \lambda_1$ which divides the first column of $B(\lambda)$. In addition suppose that $k_1 < r_1$. By k_1 times differentiation of (1) with respect to λ and putting $\lambda = \lambda_1$ we conclude that the first column of $B^{(k_1)}(\lambda_1)$ is an eigenvector of A belonging to λ_1.

Hence, each column of $B(\lambda)$ which is not divisible by $(\lambda - \lambda_1)^{r_1}$ gives rise to an eigenvector of A belonging to λ_1. We can go further. Indeed, by differentiating $k_1, k_1 + 1, \ldots, r_1 - 1$ times the first columns in (1) we get

$$\begin{aligned}
&(\lambda_1 I - A) \, B_1^{(k_1)}(\lambda_1) = 0, \\
&(\lambda_1 I - A) \, B_1^{(k_1+1)}(\lambda_1) + (k_1 + 1) \, B_1^{(k_1)}(\lambda_1) = 0, \\
&\quad \vdots \\
&(\lambda_1 I - A) \, B_1^{(r_1-1)}(\lambda_1) + (r_1 - 1) \, B_1^{(r_1-2)}(\lambda_1) = 0.
\end{aligned}$$

Let x_i be the first column of $B^{(k_1+i-1)}(\lambda_1)$ divided by $(k_1 + i - 1)!$. Then

$$\begin{aligned}
A x_1 &= \lambda_1 x_1, \\
A x_2 &= \lambda_1 x_2 + x_1, \\
&\;\vdots \\
A x_{r_1-k_1+1} &= \lambda_1 x_{r_1-k_1+1} + x_{r_1-k_1}.
\end{aligned}$$

1) This research was supported in part by NRC Grant No. A-2972.

Received November 14, 1967

The vector x_1 is an eigenvector of A belonging to λ_1. The vectors $x_2, \ldots, x_{r_1-k_1+1}$ are generalized eigenvectors (Hauptvektoren) in the sense of [4], section 19.4.

This procedure for generating eigenvectors and generalized eigenvectors is discussed in [1] in the case of real matrices and real eigenvalues. The discussion is valid also for complex matrices. In [1] it was conjectured that by this procedure we obtain a system of eigenvectors and generalized eigenvectors of A belonging to λ_1 which span the λ_1-primary component of A. We recall (cf. [3], p. 132) that the λ_1-primary component of A may be defined as the subspace of vectors x such that

$$(\lambda_1 I - A)^{r_1} x = 0.$$

The purpose of this paper is to prove that this conjecture is true.

2. Invariant Factors of the Adjoint Matrix

We consider $\lambda I - A$ and $B(\lambda)$ as λ-matrices.

THEOREM 1. *Let* $\varphi_1(\lambda) \mid \varphi_2(\lambda) \mid \cdots \mid \varphi_n(\lambda)$ *be invariant factors of* $\lambda I - A$. *Then the invariant factors of its adjoint matrix* $B(\lambda)$ *are the polynomials*

$$\varphi_k^*(\lambda) = f(\lambda)/\varphi_{n-k+1}(\lambda), \qquad k = 1, \ldots, n.$$

Proof. Let $D_k(\lambda)$ resp. $D_k^*(\lambda)$ be the greatest common divisor of all k-rowed minors of $\lambda I - A$ resp. $B(\lambda)$. We note that $D_k(\lambda)$ and $D_k^*(\lambda)$ are monic polynomials in λ, by definition.

It is well known (cf. [2], p. 21) that between the minors of $\lambda I - A$ and $B(\lambda)$ there exists the relationship

$$\left| B(\lambda)\begin{pmatrix} i_1 \cdots i_k \\ j_1 \cdots j_k \end{pmatrix} \right| = (-1)^{\sum_{p=1}^{k}(i_p+j_p)} \left| (\lambda I - A)\begin{pmatrix} j'_1 \cdots j'_{n-k} \\ i'_1 \cdots i'_{n-k} \end{pmatrix} \right| f(\lambda)^{k-1}. \tag{2}$$

On the left hand side we have the minor of $B(\lambda)$ whose elements lie in the intersection of the rows $i_1, \ldots, i_k$ and the columns $j_1, \ldots, j_k$. The similar term on the right hand side denotes the corresponding minor of $\lambda I - A$. The indices $i'_1, \ldots, i'_{n-k}$ are complementary to $i_1, \ldots, i_k$, and so are $j'_1, \ldots, j'_{n-k}$ to $j_1, \ldots, j_k$.

From (2) it follows that

$$D_k^*(\lambda) = f(\lambda)^{k-1} D_{n-k}(\lambda), \qquad k = 1, \ldots, n.$$

Using the relationship between g.c.d. of minors and invariant factors we get

$$\begin{aligned} \varphi_k^*(\lambda) &= D_k^*(\lambda)/D_{k-1}^*(\lambda) \\ &= f(\lambda)^{k-1} D_{n-k}(\lambda)/f(\lambda)^{k-2} D_{n-k+1}(\lambda) \\ &= f(\lambda)/\varphi_{n-k+1}(\lambda), \qquad k = 2, \ldots, n. \end{aligned}$$

For $k=1$ we have

$$\varphi_1^*(\lambda) = D_1^*(\lambda) = D_{n-1}(\lambda) = f(\lambda)/\varphi_n(\lambda).$$

This proves our theorem.

Note that $f(\lambda)=\varphi_1(\lambda)\ldots\varphi_n(\lambda)$ and so

$$\varphi_k^*(\lambda)=\prod_{\substack{i=1\\ i\neq n-k+1}}^{n}\varphi_i(\lambda). \tag{3}$$

3. Generation of Eigenvectors

We shall say that a square λ-matrix is an elementary matrix if its determinant is a non-zero constant. The inverse of an elementary matrix is also elementary. There exist elementary matrices $E(\lambda)$ and $F(\lambda)$ (cf. [2], p. 141) such that

$$B(\lambda)=E(\lambda)\,S(\lambda)\,F(\lambda),$$

where $S(\lambda)$ has canonical diagonal form, i.e.,

$$S(\lambda)=\operatorname{diag}(\varphi_1^*(\lambda),\ldots,\varphi_n^*(\lambda)).$$

Let us put $C(\lambda)=E(\lambda)\,S(\lambda)$ and $G(\lambda)=F(\lambda)^{-1}$. Then $B(\lambda)=C(\lambda)\,F(\lambda)$ and from (1) we get

$$(\lambda I-A)\,C(\lambda)=f(\lambda)\,G(\lambda). \tag{4}$$

Let $(\lambda-\lambda_1)^{s_1}$ be the highest power of $\lambda-\lambda_1$ dividing $\varphi_1^*(\lambda)$. Then $(\lambda-\lambda_1)^{s_1}$ is also the highest power of $\lambda-\lambda_1$ dividing the first column of $C(\lambda)$. Indeed, since $E(\lambda)$ is elementary, its first column is not divisible by $\lambda-\lambda_1$. It remains to note that the first column of $C(\lambda)$ is equal to the first column of $E(\lambda)$ multiplied by $\varphi_1^*(\lambda)$. Similar statements hold for other columns of $C(\lambda)$. Let t be the number of elementary divisors of A which are powers of $\lambda-\lambda_1$. From (3) we conclude that $\varphi_k^*(\lambda)$ is divisible by $(\lambda-\lambda_1)^{r_1}$ or not according to whether $k>t$ or $k\leqslant t$.

We recall that $C_1(\lambda)$ denotes the first column of $C(\lambda)$. From (4), by differentiation in turn $s_1, s_1+1,\ldots, r_1-1$ times and putting $\lambda=\lambda_1$ we get

$$\begin{aligned}
&(\lambda_1 I-A)\,C_1^{(s_1)}(\lambda_1)=0,\\
&(\lambda_1 I-A)\,C_1^{(s_1+1)}(\lambda_1)+(s_1+1)\,C_1^{(s_1)}(\lambda_1)=0,\\
&\quad\vdots\\
&(\lambda_1 I-A)\,C_1^{(r_1-1)}(\lambda_1)+(r_1-1)\,C_1^{(r_1-2)}(\lambda_1)=0.
\end{aligned}$$

In this way we obtain an eigenvector of A belonging to λ_1 and generalized eigenvectors associated with it. From (3) again we see that r_1-s_1 is the greatest exponent u such that $(\lambda-\lambda_1)^u$ is an elementary divisor of A. Therefore the vectors

$$C_1^{(s_1)}(\lambda_1),\ldots,C_1^{(r_1-1)}(\lambda_1)$$

span a cyclic subspace corresponding to the elementary divisor $(\lambda-\lambda_1)^{r_1-s_1}$.

We have similar results concerning the second, third, ..., t-th column of $C(\lambda)$. The eigenvectors of A corresponding to λ_1 which were obtained from different columns of $C(\lambda)$ are linearly independent. Let $s_2, s_3, \ldots, s_t$ have similar meaning as s_1. Then these eigenvectors are

$$C_1^{(s_1)}(\lambda_1), \ldots, C_t^{(s_t)}(\lambda_1). \tag{5}$$

Since

$$C_k^{(s_k)}(\lambda_1) = E(\lambda_1)\, S_k^{(s_k)}(\lambda_1), \qquad k = 1, \ldots, t$$

we see that these eigenvectors are proportional to the corresponding columns of $E(\lambda_1)$. But $E(\lambda_1)$ is nonsingular and this implies that the vectors (5) are linearly independent.

Note that the last $n-t$ columns of the matrices

$$C(\lambda_1), C'(\lambda_1), \ldots, C^{(r_1-1)}(\lambda_1) \tag{6}$$

are zero. Therefore we have the following result:

All columns of the r_1 matrices (6) span the λ_1-primary component of A.

Now we use the relation between $B(\lambda)$ and $C(\lambda)$:

$$C(\lambda) = B(\lambda)\, G(\lambda).$$

By differentiation we find

$$C^{(k)}(\lambda_1) = \sum_{i=0}^{k} \binom{k}{i} B^{(i)}(\lambda_1)\, G^{(k-i)}(\lambda_1)$$

$(k=0, 1, \ldots, r_1-1)$. It follows that all columns of the matrices (6) are linear combinations of the columns of matrices

$$B(\lambda_1), B'(\lambda_1), \ldots, B^{(r_1-1)}(\lambda_1). \tag{7}$$

We know already that the non-zero columns of (7) are the eigenvectors and generalized eigenvectors of A belonging to λ_1.

Hence we have proved

THEOREM 2. *The columns of the r_1 matrices* (7), *where $B(\lambda)$ is the adjoint matrix of $\lambda I - A$, span the λ_1-primary component of A.*

REFERENCES

[1] BRAND, L., *Matrices with Specified Eigenvalues and Associated Eigenvectors, Proper or Generalized*, Amer. Math. Monthly *74*, 640–648 (1967).
[2] GANTMACHER, F. R., *The Theory of Matrices*, Vol. I (New York 1960).
[3] JACOBSON, N., *Lectures in Abstract Algebra*, Vol. II (New York 1953).
[4] ZURMÜHL, R., *Matrizen und ihre technische Anwendungen* (Berlin 1961).

University of Waterloo

On Rota's Problem Concerning Partitions

R. Mullin (Waterloo, Ont. Canada)

Dedicated to Professor A. M. Ostrowski on the occasion of his 75th birthday

1. Introduction

Let P_n denote the set of partitions of a finite set of cardinality n. Two members of P_n are comparable (or dependent) if one is a refinement of the other; otherwise they are incomparable (or independent). In anticipation of later notation, we denote the set of partitions of P_n with exactly k parts by $L_k(P_n)$ or simply L_k, if there is no risk of ambiguity. Further if $\pi \in L_k$, we write $r(\pi)=k$. It is well known ([1], p. 91) that $|L_k(P_n)|=S(n, k)$, where $S(n, k)$ is a Stirling number of the second kind. We denote by $K(n)$ the largest integer for which $S(n, k)$ is a maximum. Clearly, since all partitions of $L_{K(n)}$ are independent, a largest set of independent elements (or largest antichain) in P_n has cardinality $B(n) \geqslant S(n, K(n))$. We show that $B(n)=S(n, K(n))$ for all positive integers n if and only if the maximum number of independent elements in $L_{K(n)+1} \cup L_{K(n)} \cup L_{K(n)-1}$ is $S(n, K(n))$.

2. Generating Sets

Since our proof will rely on induction, we require an investigation of the relation between independent sets of partitions on sets of cardinalities $n+1$ and n. Let I be an independent set of partitions on a set S_{n+1} of cardinality $n+1$ for $n \geqslant 1$. If some element x of S_{n+1} is removed from each member of I, a list of partitions of a set S_n is obtained. The elements of this list form a set which we call a generating set $G=G(I)$ for I. Members of I are obtained from some member g of $G(I)$ either by: (1) the addition of x to some part of g; or (2) by adding a new (singleton) part $\{x\}$ to g. Let us enlarge the concept of generating set to include any collection G^* of partitions of S_n by adding operation (3), we ignore any partitions of G^* we choose in the creation of an independent set I in S_{n+1}. Such a set will be called an extension of G.

Clearly if we use a generating set G^* to generate an independent set I, $G(I) \subseteq G^*$. We shall refer to $G(I)$ as the basic generating set for I relative to x. Let G^* be any generating set of partitions of $S(n)$. Let us endeavour to create as large an independent set I from G^* as possible.

Let us note that if operation (2) is performed on a partition π to obtain π' than neither operations (1) nor (3) may be applied to π, since any resultant of operation (1) on π would give an element comparable to π'.

Received November 15, 1967 and, in revised form, February 1, 1968

Further, if θ is finer than π and operation (2) is applied to π to obtain π', then operation (1) cannot be applied to θ since the resultant would be comparable with π'. Any of the above operations applied to incomparable elements leaves the resultant incomparable.

We further note that if $r(\pi)=t$, then (1) can be used to obtain up to t independent members of I. However, if x is adjoined to a part p of a partition π, to obtain π' and θ is finer than π, then x may not belong to any part of θ which is a subset of p, since the resultant would be comparable to π'.

Nevertheless, adding x to any or all parts of a set of incomparable elements of G^* produces incomparable partitions of S_{n+1}.

Now any extension I of G^* partitions G^* into three classes, $C_1=C_1(I)$, $C_2=C_2(I)$, and $C_3=C_3(I)$, those subject to operations (1), (2), and (3), respectively. Let us denote the number of elements of I obtained from a partition π under operation (1) by $S(\pi)$. Then we obtain the equation

$$|I| = \sum_{\pi\in C_1} S(\pi) + |C_2|. \tag{3.1}$$

Evidently C_2 is an independent set. Let p be a part of a partition π in a generating set G^*. We say that p is flagged (relative to I) if x is added to p in π in the construction of I. In this case, $\pi\in C_1$. Thus we may view $C_1\cup C_3$ as a sub-order of P in which certain parts of the partitions are distinguished (flagged). The flagging obeys the rule that if π is a refinement of π' and p is flagged in π', no subset of p is flagged in π. Conversely any sub-order with parts distinguished in accordance with this rule may be considered as a flagging of $C_1\cup C_3$ for some generating set G^* relative to some extension I. Such a flagging on a sub-order of P_n will be called admissable. If C is an admissably flagged partial order, we write

$$\psi(C) = \sum_{\pi\in C} S(\pi),$$

for any generating set G^* containing C, where, of course, $S(\pi)$ is the number of flagged parts of π. Thus equation (3.1) may be written as

$$|I| = \psi(C) + |C_2|. \tag{3.2}$$

3. Convergent (partial) orders

A finite (partial) order P with 0 element is said to be graded if for any element p of P, every maximal chain from 0 to p has the same length. Let $h(p)$ be the length of such a chain. Then $h(p)$ is called the height of p or rank of p. Given a ranked order P and a positive integer m we define $L_m(P)$, by

$$L_m(P) = \{p : p\in P, h(p) = m\}.$$

If there is no ambiguity as to the order in question, we abbreviate $L_m(P)$ to L_m.

If two elements, x and y, of P are comparable, we write xRy. If X is a subset of P, we define $M_m(P, X)$ or $(M_m(X))$ by

$$M_m(P, X) = \{p : p \in P,\ r(p) = m,\ pRx \text{ for some } x \in X\}.$$

We say that P is ascendingly convergent to L_m if $r<m$, and $W \subseteq L_r$ implies $|W| \leqslant |M_{r+1}(W)|$.

Similarly P is descendingly convergent to L_m if $r>m$, and $W \subseteq L_r$ implies $|W| \leqslant |M_{r-1}(W)|$. P is convergent to L_m if it is both ascendingly and descendingly convergent to L_m.

LEMMA 1. *If P is convergent to L_m, then $|L_m|$ is the greatest cardinality of a set of mutually incomparable elements of P.*

Proof: Let X be a set of mutually incomparable elements of P. Let $h = \min_{x \in X} \{h(x)\}$. If $h<m$, let $X' = \{x : x \in Y;\ h(x) = h\}$. Replacing X' by $M_{h+1}(X')$ gives a set X^* of mutually incomparable elements with $|X^*| \geqslant |X|$.

Repeating the operation we construct a set Y of mutually incomparable elements such that

(1) $h(y) \geqslant m$, for each $y \in Y$;

(2) $|Y| \geqslant |X|$.

By using descending convergence, we eventually form a set Z of mutually incomparable elements such that

(1) $h(z) = m$, for each $z \in Z$;

(2) $|Z| \geqslant |X|$.

Hence if X is a set of mutually incomparable elements, $|X| \leqslant |L_m|$. Since distinct elements of the same rank are incomparable, the lemma follows.

4. Some Properties of Stirling numbers

Let $\mathscr{S}(n)$ be the list of Stirling numbers $S(n, k)$ of the second kind, for any non-negative integer n and all integers k. (It is well-known that for $n \geqslant 1$, $S(n, k)$ is non-zero if and only if $1 \leqslant k \leqslant n$, and that $S(0, k) = \delta_{0,k}$. We shall say that $\mathscr{S}(n)$ is of type I if $n=0$, or if $S(n,1) < S(n,2) < \cdots < S(n, K(n)-1) < S(n, K(n))$ and $S(n, K(n)) > S(n, K(n)+1) > \cdots > S(n, n)$; and $\mathscr{S}(n)$ is of type II if $S(n,1) < S(n,2) < \cdots < S(n, K(n)-1)$, $S(n, K(n)-1) = S(n, K(n))$, and $S(n, K(n)) > S(n, K(n)+1) > \cdots > S(n, n)$.

It is well-known for $n \geqslant 1$, that $S(n, k) = S(n-1, k-1) + kS(n-1, k)$ (cf. [1], p. 33) and that $S(n, k) = \sum_{j=0}^{n-1} \binom{n-1}{j} S(j, k-1)$, (cf. [1], p. 43). We establish further properties below.

LEMMA 2. *For* $n \geqslant 1$, $\mathscr{S}(n)$ *is of type* I *or type* II, *and* $K(n)=K(n-1)+\varepsilon(n)$, *where* $\varepsilon(n)$ *is either* 0 *or* 1, *with* $\varepsilon(n)=1$ *if* $\mathscr{S}(n)$ *is of type* II.

Proof. (Induction). The lemma is trivially true for $n=1$. Let us assume that the lemma is valid for $1 \leqslant n \leqslant m$. We examine $\mathscr{S}(m+1)$, noting that the hypotheses imply that $K(m) \geqslant K(n)$ for $m \geqslant n \geqslant 1$. There are two cases to consider.

CASE 1. $\mathscr{S}(m)$ is of type II. Now

$$S(m+1, k) - S(m+1, k-1) = S(m, k-1) - S(m, k-2) + k[S(m, k) - S(m, k-1)] + S(m, k-1),$$

which is strictly positive, for $1 \leqslant k \leqslant K(m)$ by hypothesis, and the fact that $S(m, -1) = S(m, 0) = 0$ for $m \geqslant 1$. Further,

$$S(m+1, k) - S(m+1, k-1) = \sum_{j=0}^{m} \binom{m}{j} [S(j, k-1) - S(j, k-2)],$$

which is negative for $K(m)+1 \leqslant k \leqslant m+1$, since by hypothesis, $K(m-1)=K(m)-1$ if $\mathscr{S}(m)$ is of type II, hence $S(j, k-1)-S(j, k-2)$ is non-negative for $j \leqslant m$ and $k \geqslant K(m)+1$. Thus if $\mathscr{S}(m)$ is of type II, $\mathscr{S}(m+1)$ is of type I, and $\varepsilon(m+1)=0$.

CASE 2. $\mathscr{S}(m)$ is of type I.

Again we have

$$S(m+1, k) - S(m+1, k-1) = S(m, k-1) - S(m, k-2) + k[S(m, k) - S(m, k-1)] + S(m, k-1),$$

which is strictly positive for $1 \leqslant k \leqslant K(m)$. Also

$$S(m+1, k) - S(m+1, k-1) = \sum_{j=0}^{m} \binom{m}{j} [S(j, k-1) - S(j, k-2)],$$

which is negative for $K(m)+2 \leqslant k \leqslant m+1$, by hypothesis.

Let us examine $\Delta = S(m+1, K(m)+1) - S(m+1, K(m))$. $\mathscr{S}(m+1)$ is of type II if and only if $\Delta = 0$, in which case $K(m+1)=K(m)+1$, that is $\varepsilon(m+1)=1$. On the other hand if $\Delta \neq 0$, $\mathscr{S}(m+1)$ is of type I, and $\varepsilon(m+1)=0$ if Δ is negative, $\varepsilon(m+1)=1$ if Δ is positive. This establishes the lemma.

5. Equivalent propositions

Let us consider the order defined on P_n under the relation $\pi \leqslant \pi'$ if π' is a refinement of π. Clearly the order $(P_n, \leqslant)$ (which we normally abbreviate to P_n) is graded, and $h(\pi)=r(\pi)$. We prove the equivalence of the following propositions.

(i) P_n is convergent to $L_{K(n)}$, for $n=1, 2, 3, \ldots$.

(ii) The maximum number of independent partitions in P_n is $S(n, K(n))$, $n=1, 2, 3, \ldots$.

(iii) The maximum number of elements in $L_{K(n)+1}(P_n) \cup L_{K(n)}(P_n) \cup L_{K(n)-1}(P_n)$ is $S(n, K(n))$, for $n=1, 2, 3, \ldots$.

By Lemma 1, (i) implies (ii), and trivially (ii) implies (iii). We now prove

THEOREM 1. *Proposition* (iii) *implies proposition* (i).

Proof: (Induction). Proposition (i) is trivially true for $n=1, 2$. Let us assume that (i) is true for $n \leqslant m$. We consider P_{m+1} as being partitions of the linearly ordered set $1<2<3<\cdots<m<x$. It is now convenient to form a new representation of the elements of P_{m+1}. We proceed as follows:

Consider the sets

$$\begin{aligned} T_0 &= \{1, 2, 3, \ldots, m\} \\ T_1 &= \{\bar{1}, 2, 3, \ldots, m\} \\ T_2 &= \{1, \bar{2}, 3, \ldots, m\} \\ &\cdot \quad \cdot \; \cdot \; \cdot \; \ldots \; \cdot \\ &\cdot \quad \cdot \; \cdot \; \cdot \; \ldots \; \cdot \\ &\cdot \quad \cdot \; \cdot \; \cdot \; \ldots \; \cdot \\ T_m &= \{1, 2, 3, \ldots, \bar{m}\}. \end{aligned}$$

Let π be a partition of T_j, and p_j be the part of π which contains $\bar{j}$. We say that π is admissable if for every l in p_j, $l \geqslant j$. Also any partition of T_0 is admissable.

Let $Q_m(j)$ represent the set of admissable partitions of T_j. Let $Q_m = \bigcup_{j=0}^{m} Q_m(j)$.

Now consider the mapping from P_{m+1} into Q_m defined as follows. Let p be that part of $\pi(\in P_{m+1})$ which contains x. If $|p|=1$ delete x from π. If $|p|>1$, let j be the least element of p. In π, replace j by $\bar{j}$ and delete x. The resultant member of Q_m we denote by $\phi(\pi)$. Clearly ϕ is a bijection from P_{m+1} to Q_m.

We now show that P_{m+1} is ascendingly convergent to $L_{K(m+1)} = L_{K(m+1)}(P_{m+1})$. Let X be any subset of L_r, where $r<K(m+1)$. Hence $r \leqslant K(m)$. Define $\phi(X)$ by

$$\phi(X) = \{\phi(x) : x \in X\}.$$

Now form sets $R(X,j) = \phi(X) \cap Q_m(j)$.

Let $|R(X,j)| = g_j$. Thus

$$\sum_{j=0}^{m} g_j = |X|.$$

CASE 1. There are two cases to consider; $r<K(m)$, and $r=K(m)$. If $r<K(m)$, consider $P^{(j)} = (P_m, \leqslant)$ defined on T_j, $j=0, 1, 2, \ldots, m$. Since each member of $R(X,j)$

has rank r, for $1 \leqslant j \leqslant m$, we may form sets $S(j)=M_{r+1}(P^{(j)}, R(X,j))$. By the inductive hypothesis, if $h_j=|S(j)|$, $h_j \geqslant g_j$.

Similarly we form $S(0)=M_{r+1}(P^{(0)}, R(X,0))$ of cardinality h_0. Again $h_0 \geqslant g_0$. Each member of $S(j)$ is admissable, $j=0,1,2,\ldots,m$, since a refinement of an admissable partition is admissable. Further if $\pi \in S(j)$, $\phi^{-1}(\pi) \in M_{r+1}(X)$. Hence

$$|M_{r+1}(X)| \geqslant \sum_{j=0}^{m} h_j \geqslant \sum_{j=0}^{m} g_j = |X|, \quad \text{as required}.$$

CASE 2. If $r=K(m)$, then $K(m)<K(m+1)$, that is $\varepsilon(m+1)=1$ since otherwise $r=K(m+1)$, contrary to hypothesis. Thus $r+1=K(m+1)$, hence the maximum number of independent points in $L_r(P_{m+1}) \cup L_{r+1}(P_{m+1})$ is $|L_{r+1}(P_{m+1})|$. Let us assume that there is a subset X of $L_r(P_{m+1})$ such that $|M_{r+1}(X)|<|X|$.

Consider the set $Y=X \cup (L_{r+1}-M_{r+1}(X))$.

Now $|Y|>|L_{r+1}|$, and the members of Y are independent, by definition of $M_{r+1}(X)$, which is a contradiction. Hence for any $X \subseteq L_r$, $|M_{r+1}(X)| \geqslant |X|$. Hence P_{m+1} is ascendingly convergent to $L_{K(m+1)}$.

We now show that P_{m+1} is descendingly convergent to $L_{K(m+1)}$ subject to the validity of proposition (iii).

Let us assume that the theorem is valid for $n \leqslant m$. Suppose $X \subseteq L_r(P_{m+1})$, for $r>K(m+1)$. There are again two cases to consider.

CASE 3. $r-1>K(m)$.

Let x be a fixed element of P_{m+1}. Now $G=L_r(P_m) \cup L_{r-1}(P_m) \cup L_{r-2}(P_m)$ will serve as a generating set (with respect to x) for any set of incomparable elements of $L_r(P_{m+1}) \cup L_{r-1}(P_{m+1})$. Clearly any element of $L_{r-2}(P_m)$ must occur in $C_2(I)$ or $C_3(I)$. Also no member of $L_r(P_m)$ can occur in $C_2(I)$.

Suppose p is any part of cardinality $y(\geqslant 1)$ which occurs in some of the partitions of P_m. If p is flagged α times in $L_r(P_m)$ and β times in $L_{r+1}(P_m)$, p contributes $\alpha+\beta$ to $\psi(G)$.

Let $N_r(p)$ and $N_{r+1}(p)$ be the sets of partitions in which p is flagged in L_r and L_{r+1} respectively.

Removing part p from these partitions produces α members of $L_{r-1}(P_{m-|p|})$ and β members of $L_r(P_{m-|p|})$. These partitions are independent and since $r-1>K(m) \geqslant K(m-|p|)$, $P_{m-|p|}$ is descendingly convergent from L_r to L_{r-1}, and $\alpha+\beta$ is greatest when all members of L_{r-1} are flagged, (and no member of L_r is flagged). Thus $\psi(G)$ is largest when $C_1(I)=L_{r-1}(P_m)$.

Also by the inductive hypothesis, P_m is descendingly convergent to $L_{K(m)}$, hence since $r-2 \leqslant K(m)$, the largest independent set in $L_{r-1} \cup L_{r-2}$ is L_{r-2}. (We assume $m>2$.) Thus $|C_2|$ is a maximum when $C_2=L_{r-2}$. Thus I is maximized when $C_1=L_{r-1}$ and $C_2=L_{r-2}$. But in this instance $I=L_{r-1}(P_{m+1})$. But this implies that if $X \subseteq L_r(P_{m+1})$, then $|X| \leqslant |M_{r-1}(X)|$; as required.

CASE 4. If $r-1=K(m)$, then $\varepsilon(m+1)=0$. In this case we use the fact that the greatest number of elements in $L_{K(m+1)} \cup L_{K(m+1)+1}$ is $|L_{K(m+1)}|$ to show that if $X \subseteq L_r$, then $|X| \leqslant |M_{r-1}(X)|$, by arguments similar to those of case 2. This concludes the proof of the theorem.

There is an interesting graph-theoretic formulation of maximum value of $\psi(S)$ for $S=L_{K(m+1)+1}(P_m) \cup L_{K(m+1)}(P_m)$ (which is of interest when $\varepsilon(m+1)=0$). Consider the set of points

$$U = \{(\pi, p): \pi \in L_{K(m+1)+1}(P_m);\ p \text{ is a part of } \pi\}$$

and

$$V = \{(\pi, p): \pi \in L_{K(m+1)}(P_m);\ p \text{ is a part of } \pi\}\ .$$

Form the graph G as follows: (π, p) in U is joined to (θ, q) in V if and only if π is a refinement of θ and p is a subset of q. A set of vertices in G is independent if no pair is joined G. It is evident from the definition of $\psi(S)$ that max $\psi(S)$ is the cardinal number of the largest independent set in G.

The author wishes to express his gratitude to the referee for his comments, criticism, and suggestions which were most useful in the preparation of the present revised form of this paper.

REFERENCES

[1] RIORDAN, J., *An Introduction to Combinatorial Analysis* (John Wiley and Sons, New York, N.Y. 1958).

University of Waterloo

Some Properties of the Jordan Operator

EINAR HILLE (Eugene, Oregon)

To Professor Alexander Ostrowski on his 75th birthday

There is an extensive theory of Jordan algebras for which see, e.g., H. BRAUN and M. KOECHER [1]. This theory indicates the lively interest taken by the algebraist in the Jordan operator. In this brief note we aim to show that Jordan operators are also of some interest to the functional analyst.

1. Let $\mathfrak{B}$ be a non-commutative Banach algebra with unit element e and let $a \in \mathfrak{B}$ but not to the center of $\mathfrak{B}$. Define four linear bounded operators L_a, R_a, J_a and S_a on $\mathfrak{B}$ to $\mathfrak{B}$ by

$$L_a(x) = a\,x, \quad R_a(x) = x\,a, \quad J_a(x) = \tfrac{1}{2}(a\,x + x\,a), \quad S_a(x) = a\,x\,a \tag{1}$$

so that

$$J_a = \tfrac{1}{2}(L_a + R_a), \quad S_a = L_a R_a = R_a L_a. \tag{2}$$

These operators are elements of the algebra $\mathfrak{E}(\mathfrak{B})$ of linear bounded operators on $\mathfrak{B}$ to $\mathfrak{B}$. Here L_a is the image of the element a in the left regular representation of $\mathfrak{B}$ while R_a is the image in the right regular representation. J_a is the Jordan operator defined by a. On the other hand, S_a is the quadratic representation of $\mathfrak{B}$ as defined at a. For a general non-associative algebra the quadratic representation is given by

$$P_a = L_a(L_a + R_a) - L_{a^2} = R_a(L_a + R_a) - R_{a^2}$$

which in the associative case reduces to S_a.

We shall be concerned with the spectral properties of these operators. Spectra will be denoted by the letter 'σ', point spectra by the symbol $P\sigma$'. It is well known that the spectra of the operators L_a and R_a with respect to the algebra $\mathfrak{E}(\mathfrak{B})$ are identical and coincide with the spectrum of a with respect to $\mathfrak{B}$.

Since L_a and R_a commute the Gelfand representation theorem may be used to get a grip on the spectra of J_a and S_a. Their spectra with respect to $\mathfrak{E}(\mathfrak{B})$ are the same as the spectra with respect to the commutative algebra $\mathfrak{B}_0$ obtained by taking the second commutant of the algebra generated by L_a, R_a and the identity operator I. This gives

THEOREM 1. *The spectrum of J_a with respect to $\mathfrak{E}(\mathfrak{B})$ is contained in the mid-point set*

$$S_1 \equiv [\tfrac{1}{2}(\alpha + \beta) \mid \alpha \in \sigma(a), \beta \in \sigma(a)] \tag{3}$$

Received November 22, 1967 and, in revised form, February 3, 1968

while the spectrum of S_a is contained in the product set

$$S_2 \equiv [\alpha\beta \mid \alpha \in \sigma(a), \beta \in \sigma(a)]. \tag{4}$$

This result is well known to the algebraists in the finite dimensional case (See [1], Chapter VIII, Theorem 1.3.). In this case necessary and sufficient conditions are known in order that a particular combination $\frac{1}{2}(\alpha+\beta)$ or $\alpha\beta$ shall be a characteristic value. The following result is closely related to this criterion.

THEOREM 2. *Let $\alpha \in P\sigma[L_a]$, $\beta \in P\sigma[R_a]$ and suppose that $L_a(u)=\alpha u$, $R_a(v)=\beta v$, $u \neq 0$, $v \neq 0$. Then either the mapping $x \to uxv$ annihilates all x in $\mathfrak{B}$ or $\frac{1}{2}(\alpha+\beta) \in P\sigma[J_a]$* and $\alpha\beta \in P\sigma[S_a]$.

This follows from

$$J_a(uxv) = \tfrac{1}{2}(auxv + uxva) = \tfrac{1}{2}(\alpha+\beta)\,uxv, \tag{5}$$

$$S_a(uxv) = auxva = \alpha\beta uxv. \tag{6}$$

There is one case in which we can be sure that the second alternative holds. N. JACOBSON [4, p. 196] has introduced the notion of a prime ring. Such a ring has the important property that if $uxv=0$ for all x in the ring, then either $u=0$ or $v=0$. Finite matrix algebras and the operator algebras $\mathfrak{E}(\mathfrak{X})$ of linear bounded transformations on a Banach space into itself are examples of prime rings. This concept is eminently suitable for our problem and leads to

THEOREM 3. *If $\mathfrak{B}$ is a prime ring and the assumptions of Theorem 2 are valid, then $\frac{1}{2}(\alpha+\beta) \in P\sigma[J_a]$, $\alpha\beta \in P\sigma[S_a]$.*

COROLLARY. If $\mathfrak{B} = \mathfrak{M}_n$, the algebra of n by n matrices over the complex field, then $\sigma(J_a) = P\sigma(J_a) = S_1$ and $\sigma(S_a) = P\sigma(S_a) = S_2$.

For $\mathfrak{M}_n$ is a prime ring as well as a B-algebra under the usual definition of the algebraic operations and a suitable norm. Further the spectral values of a matrix are isolated and belong to the point spectrum of the corresponding L and R operators. All this is of course well known to the algebraists.

The following result belongs to the same range of ideas.

THEOREM 4. *If $\alpha \in P\sigma(L_a) \cap P\sigma(R_a)$ and if $au = ua = \alpha u$ where u is an idempotent $\neq 0$, then $\alpha \in P\sigma(J_a)$ and $\alpha^2 \in P\sigma(S_a)$.*

For

$$J_a(u) = \tfrac{1}{2}(au + ua) = \alpha u, \quad S_a(u) = aua = \alpha^2 u. \tag{7}$$

The following special case of Theorem 1 is of some independent interest.

THEOREM 5. *A necessary condition for the equation*

$$ax + xa = 0 \tag{8}$$

to have a non-trivial solution x is that a is either singular or has two spectral values of sum 0. *The condition is sufficient if $\mathfrak{B}$ is a prime ring.*

2. The resolvents of J_a and S_a may be found by a construction given by Yu. L. Daletsky [2] in 1953. Let us cover the spectrum of a by a finite number of closed ε-disks. Let Σ_ε be the union of these disks and form the mid-point set Δ_ε^1 and product set Δ_ε^2.

$$\Delta_\varepsilon^1 = [\tfrac{1}{2}(\alpha + \beta)],\ \Delta_\varepsilon^2 = (\alpha \beta),\ \alpha \in \Sigma_\varepsilon,\ \beta \in \Sigma_\varepsilon\,. \tag{9}$$

Let Λ_ε^1 be the complement of Δ_ε^1 and Λ_ε^2 the complement of Δ_ε^2. We denote the resolvent of a by $R(\lambda, a)$ with similar notation for other resolvents.

Theorem 6. *For any $\lambda \in \Lambda_\varepsilon^1$ the solution v of the equation*

$$\lambda y - \tfrac{1}{2}(a y + y a) = x \tag{10}$$

is given by

$$y = R(\lambda, J_a)[x] = \frac{1}{(2\pi i)^2} \int_{\Gamma_\varepsilon} \int_{\Gamma_\varepsilon} \frac{R(\alpha, a)\, \chi\, R(\beta, a)}{\lambda - \frac{1}{2}(\alpha + \beta)}\, d\alpha\, d\beta \tag{11}$$

where $\Gamma_\varepsilon = \partial \Delta_\varepsilon$. Similarly the solution z of the equation

$$\lambda z - a z a = x \tag{12}$$

for $\lambda \in \Lambda_\varepsilon^2$ is given by

$$z = R(\lambda, S_a)[x] = \frac{1}{(2\pi i)^2} \int_{\Gamma_\iota} \int_{\Gamma_\iota} \frac{R(\alpha, a)\, x\, R(\beta, a)}{\lambda - \alpha \beta}\, d\alpha\, d\beta\,. \tag{13}$$

We shall sketch the argument for the first case. The second is handled in a similar manner. Denoting the integrand in (11) by Q we can write $(\lambda I - J_a)[y]$ as

$$\int\!\!\int [\lambda - \tfrac{1}{2}(\alpha + \beta)]\, Q\, d\alpha\, d\beta + \int\!\!\int \tfrac{1}{2}(\alpha e - a)\, Q\, d\alpha\, d\beta + \int\!\!\int \tfrac{1}{2} Q (\beta e - a)\, d\alpha\, d\beta\,.$$

The first integral reduces to x, the second and the third are both 0 as is seen with the aid of the identities

$$(\alpha e - a)\, R(\alpha, a) = e\,, \qquad R(\beta, a)(\beta e - a) = e$$

which are valid on the boundary. This shows that

$$(\lambda I - J_a)\, R(\lambda, J_a)[x] = x$$

for $\lambda \in \Lambda_\varepsilon^1$. In a similar manner one verifies that

$$R(\lambda, J_a)[\lambda x - J_a(x)] = x\,.$$

In the matrix case $\mathscr{R}(\alpha, \mathscr{A})$ and $\mathscr{R}(\beta, \mathscr{A})$ are rational functions of α and β, respectively. Thus if we substitute their well known expansions in partial fractions in (11) and (13), the evaluation of the double integrals is reduced to the evaluation of numerical integrals which act as coefficients for certain matrix products. In the case of $\mathscr{R}(\lambda, J_{\mathscr{A}})$ the numerical integrals are of the form

$$I_{jk\mu\nu}(\lambda) = \frac{1}{(2\pi i)^2} \int_{\gamma_j} \int_{\gamma_k} (\alpha - \lambda_j)^{-\mu} (\beta - \lambda_k)^{-\nu} [\lambda - \tfrac{1}{2}(\alpha + \beta)]^{-1} \, d\alpha \, d\beta . \tag{14}$$

Here λ_j is the jth characteristic value of the matrix $\mathscr{A}$, $1 \leqq j, k \leqq n$. Further μ is a positive integer $\leqq m_j$, the multiplicity of λ_j as a pole of $\mathscr{R}(\alpha, \mathscr{A})$. The integral is multiplied by the factor $(\mathscr{Q}_j)^{\mu-1} \mathscr{X} (\mathscr{Q}_k)^{\nu-1}$ where $(\mathscr{Q}_j)^0 = \mathscr{P}_j$ with $\mathscr{P}_j$ the idempotent and $\mathscr{Q}_j$ the nilpotent associated with the characteristic value. Further, γ_j is an ε-circle centered at λ_j. A simple calculation gives

$$I_{jk\mu\nu}(\lambda) = (\tfrac{1}{2})^{\mu+\nu-2} \frac{(\mu + \nu - 2)!}{(\mu - 1)!\,(\nu - 1)!} [\lambda - \tfrac{1}{2}(\lambda_j + \lambda_k)]^{-(\mu+\nu-1)} . \tag{15}$$

This shows that $\mathscr{R}(\lambda, J_{\mathscr{A}})$ is also a rational function of λ and its poles are located at all the points $\frac{1}{2}(\lambda_j + \lambda_k)$. A particular pole γ will usually come from several terms in the expansion and the multiplicity of the pole is $\leqq \max(m_j + m_k - 1)$ where j and k run through those integers for which

$$\gamma = \tfrac{1}{2}(\lambda_j + \lambda_k) .$$

Actually the maximum is reached since we can choose a matrix $\mathscr{X}_0$ so that a particular

$$(\mathscr{Q}_j)^{m_j - 1} \mathscr{X}_0 (\mathscr{Q}_k)^{m_k - 1} \equiv \mathscr{X} \neq 0$$

while this $\mathscr{X}$ annihilates all other terms of the same degree in $(\lambda - \gamma)^{-1}$.

We get similar results for the operator $S_{\mathscr{A}}$. Here

$$\mathscr{R}(\lambda, S_{\mathscr{A}}) [\mathscr{X}] = \sum \frac{(\mu + \nu - 2)!}{(\mu - 1)!\,(\nu - 1)!} \lambda_k^{\mu-1} \lambda_j^{\nu-1} (\lambda - \lambda_j \lambda_k)^{-(\mu+\nu-1)} (\mathscr{Q}_j)^{\mu-1} \mathscr{X} (\mathscr{Q}_k)^{\nu-1} \tag{16}$$

The summation extends over the spectral values λ_j of $\mathscr{A}$ and $\mu \leqq m_j$, $\nu \leqq m_k$. The $\mathscr{P}_j$ and $\mathscr{Q}_j$ are as previously defined. All the products $\lambda_j \lambda_k$ are exhibited as poles of the resolvent and the multiplicity of a particular pole $\gamma = \lambda_j \mu_k$ is again $\leqq \max(m_j + m_k - 1)$ where j and k run through the integers for which $\gamma = \lambda_j \lambda_k$. Again the maximum is reached since we can choose $\mathscr{X}$ as a 'buffer' so as to keep any particular term in the expansion while annihilating the rest.

These results for $\mathfrak{M}_n$ extend, at least locally, to arbitrary B-algebras by virtue of a theorem which is essentially due to S. R. Foguel [3].

THEOREM 7. *Let γ be an isolated point of the set S_1 and suppose that the equation*

$$\gamma = \tfrac{1}{2}(\alpha + \beta) \tag{17}$$

has only a finite number of solutions (α, β) where α and β are in $\sigma(a)$, say

$$(\alpha_1, \beta_1), (\alpha_2, \beta_2), \ldots, (\alpha_p, \beta_p). \tag{18}$$

Suppose that for each j the numbers α_j and β_j are poles of $R(\lambda, a)$ of order m_j and n_j, respectively. Then γ is a pole of $R(\lambda, J_a)$ of order $\leqq \max(m_j + n_j - 1)$ and if $\mathfrak{B}$ is a prime ring the maximum is reached. If in this announcement S_1 is replaced by S_2 and $\frac{1}{2}(\alpha+\beta)$ by $\alpha\beta$, we obtain the corresponding result for $R(\lambda, S_a)$.

A proof can be based upon formulas (11) and (13) and the double decomposition of the spectrum suggested by (18). In the first case we get a number of integrals of type (14) which are holomorphic save for a single pole at $\lambda = \frac{1}{2}(\alpha_j + \beta_k)$ of order $\leqq m_j + n_k - 1$. In addition there are integrals arising from the complementary spectral sets all of which are holomorphic at the points under consideration. Only the integrals corresponding to $k = j$ give poles at γ. A similar argument applies to $R(\lambda, S_a)$.

3. In the preceding discussion we have obtained precise results only for the case in which the constituent singularities belong to the point spectrum of L_a and R_a. If they are isolated they are then poles of $R(\lambda, a)$. If $\mathfrak{B}$ is an operator algebra, the resolvents $R(\lambda, J_a)$ and $R(\lambda, S_a)$ may have other singularities than poles. The following simple example presents a case with a singular line.

We take $\mathfrak{B} = \mathfrak{E}[\mathfrak{C}]$ where $\mathfrak{C} = C[0, 1]$ so that $\mathfrak{B}$ is the algebra of linear bounded transformations on $C[0, 1]$ into itself. Take the two operators U and V defined by

$$U[f](t) = tf(t), \; V[g](t) = \int_0^t g(s)\, ds. \tag{19}$$

We consider J_U and aim to evaluate $R(\lambda, J_U)\, V$ operating on f in $C[0, 1]$. Here

$$R(\alpha, U)[f](t) = \frac{f(t)}{\alpha - t}, \tag{20}$$

so that $\sigma[U]$ is the interval $[0, 1]$ of the real axis. This interval is also the set S_1 for the operator J_U and the set S_2 for S_U. To prove this it is enough to exhibit an operator T in $\mathfrak{E}[\mathfrak{C}]$ such that $R(\lambda, J_U)\, T$ is not an element of $\mathfrak{E}[\mathfrak{C}]$ for any λ in S_1 and $R(\lambda, S_U)\, T$ is not an element of $\mathfrak{E}[\mathfrak{C}]$ for any λ in $S_2 = S_1$. Here we can take $T = V$ since

$$R(\lambda, J_U)\, V[f](t) = \frac{1}{(2\pi i)^2} \int_\Gamma \int_\Gamma \frac{1}{\alpha - t} \int_0^t \frac{f(s)\, ds}{\beta - s} \frac{d\alpha\, d\beta}{\lambda - \frac{1}{2}(\alpha + \beta)}. \tag{21}$$

Here Γ surrounds S_1 and a simple calculation gives

$$R(\lambda, J_U)\, V[f](t) = \int_0^t \frac{f(s)\, ds}{\lambda - \frac{1}{2}(s+t)}. \tag{22}$$

Similarly we get

$$R(\lambda, S_U)\, V[f](t) = \int_0^t \frac{f(s)\, ds}{\lambda - s\,t}. \tag{23}$$

Both expressions define holomorphic functions in $C[0, 1]$ for λ not in $S_1 = S_2$. For λ on the interval $[0, 1]$ the integrals normally do not exist at least not as elements of $C[0, 1]$. Thus every point of this interval belongs to the spectrum of J_U as well as the spectrum of S_U.

The operator V is quasi-nilpotent and so are the corresponding operators J_V and S_V. In this case $S_1 = S_2$ reduces to a single point, $\lambda = 0$, which is an isolated essential singularity of the resolvents.

Finally it is a pleasure to acknowledge help from a referee who supplied the reference to [1] and thus led me to revise the paper and sharpen the results.

REFERENCES

[1] Braun, H. and Koecher, M., *Jordan-Algebren* (Springer-Verlag, 1966 [Grundlehren d. Math. Wiss., No. 128]).
[2] Daletsky, Yu. L., *On the Asymptotic Solution of a Vector Differential Equation* (Russ.), Doklady Akad. Nauk SSSR *92*, 881–884 (1953).
[3] Foguel, S. R., Sums and Products of Commuting Spectral Operators, *Ark. Mat. 3*, 449–461 (1957).
[4] Jacobson, N., *Structure of Rings* (Amer. Math. Soc., Providence R.I. 1956 [Coll. Publications No. *37*]).

University of Oregon

Problems and Solutions

This section publishes problems and solutions believed to be new and interesting. Problems are designated by **P1, P2**, ..., solutions by **P1S1, P1S2**, ..., and remarks by **P1R1, P1R2**, ... Correspondence regarding this section should be sent to the Problems Editor. In case several similar solutions are received, the solutions may be edited with credits given the individual contributors.

The following problems are formulated Vol. 1, No. 3, of this journal in the report 'Fifth Annual Meeting on Functional Equations, Waterloo, Ontario', as Problems and Remarks 1, 3, 4, 5, 6, 21, 7, 8, 12, 13, 14, 16, 19 respectively: **P20** W. EICHHORN; **P21, P22** J. ACZEL; **P23** P. FISCHER; **P24, P32** G. SZEKERES; **P25** R. D. LUCE; **P26** G. TARGONSKI; **P27** E. VINCZE; **P28** J. H. B. KEMPERMAN; **P29** M. A. MCKIERNAN; **P30** V. D. BELOUSOV, S. L. SEGAL, G. SZEKERES; **P31** B. SCHWEIZER.

P28S1 – J. A. BAKER and S. L. SEGAL independently (this issue, p. 114)

The existence of additive functions $\phi: R \to R$ and sequences (a_k) such that $\sum a_k \phi(x^k)$ converges for all real x has been shown independently by the above authors by using derivations over the algebraic field. There still remain some unsolved questions concerning this problem, in particular, **P34** below.

P28R1 – S. L. SEGAL

In his solution **P28S1**, S. L. SEGAL has included the following elementary proof of the existence of derivations of the field of real numbers over the field of algebraic numbers, using well known results for finitely generated extension fields and Zorn's lemma. While not a new result (for example N. BOURBAKI, V. 5 § 9) his proof may be of interest relative to **P28**, and is therefore included here.

Given a field k, a derivation D in k and a finitely generated extension field G of k, the construction of a derivation D' in G such that D' restricted to k is D is considered in detail (in [1], pp. 11–14).

Quoting freely from [1], let A be the real algebraic numbers and R the real numbers. Define $D(x)=0$ for $x \in A$ and let δ be real and transcendental over A. Then there exists a derivation D' in $A(\delta)$ with $D'(\delta)=1$, and $D'(x)=0$ for $x \in A$. Consider the family of all fields containing $A(\delta)$ which are contained in R and which have a derivation agreeing with D' on $A(\delta)$. This family is partially ordered by algebraic extension and applying Zorn's Lemma has a maximal member, call it M.

We claim $M=R$. For if $M\neq R$, then there exists a simple extension field $M(\xi)$ of M such that

$$M \subsetneq M(\xi) \subseteq R$$

and a derivation in M which does not extend to $M(\xi)$. But this means that ξ is inseparably algebraic over M, contradicting the fact that M is of characteristic 0 (since A is of characteristic 0) and hence perfect. Hence $M=R$ and a function $L: R\to R$ with the desired properties exists.

REFERENCES

[1] WEIL, A., *Foundations of Algebraic Geometry*, American Mathematical Society Colloquium Publications, (New York, 1947).

P30S1 – S. L. SEGAL (this issue, p. 116)

The most general solution of the matrix functional equation system

$$F(X+Y)=F(X)+F(Y)$$
$$F\{[F(X)-X][F(Y)-Y]\}=\{0\}$$

which is bounded on some set of positive n^2-dimensional measure is found.

The author also poses questions concerning i) the existence of Hamel-type solutions, (see below **P30S2**) and ii) solutions over (finite) fields.

P30S2 – Z. MOSZNER (this issue, p. 118)

The author constructs a non-bounded solution of the system of equations using a Hamel basis.

P33 – B. SCHWEIZER

Let $C[0, 1]$ be the space of continuous functions on $[0, 1]$ into $[0, 1]$, under the usual sup norm. Let $\mathscr{F}=C[0, 1]\times C[0, 1]$ be endowed with one of the usual product norms. Let A be the set of all pairs (f, g) in $\mathscr{F}$ which commute under composition. A, a closed subset of a complete metric space, is thus itself a complete metric space. Let B be the set of all pairs of functions in $\mathscr{F}$ which have a common fixed point. W. M. BOYCE and J. P. HUNEKE have shown that $A\wedge B$ is a proper subset of A. I conjecture that $A\wedge B$ is of the first category, and possibly even nowhere dense, in A.

In the report of the 5th Annual Meeting on Functional Equations (Vol. 1, No. 3) B. SCHWEIZER conjectured on the nature of the space of functions $f: [0, 1]\to[0, 1]$ which commute and have a common fixed point (Isbell's conjecture). Since this original formulation, the results of Z. MOSZNER **P33S1** and M. KUCZMA **P33S2** below have led to the above reformulation of the problem.

P33S1 – Z. MOSZNER (this issue p. 120)

In the notation of **P33**, it is shown that B is nowhere dense in $\mathscr{F}$.

P33S2 – M. KUCZMA (this issue p. 123)

It is shown that also A is nowhere dense in $\mathscr{F}$.

P34 – S. L. SEGAL

Essentially we have shown (**P28S1**) that there exist Hamel-functions $L(x)$ such that $M(x) \overset{\text{Def}}{=} e^{-x} L(e^x)$ is also a non-linear additive function. Problem: describe $L(x)$ in terms of a Hamel basis.

Alternatively, given a derivation L, satisfying $L(x+y) = L(x) + L(y)$ and $L(xy) = xL(y) + yL(x)$ for x, y real: describe L in terms of a Hamel basis for the reals.

P23S1 – Z. MOSZNER (to appear in Aequationes Mathematicae)

Let ϕ be defined for $\chi > 0$ and satisfy (1) $\phi(\chi) \geqslant 0$, (2) $\phi(\lambda\chi) = \lambda\phi(\chi)$ for rational $\lambda > 0$, (3) $\phi(\chi + y) \leqslant \phi(\chi) + \phi(y)$. Problem **P23** (P. FISCHER) seeks a characterization of such ϕ, and conjectures that $\phi(\chi) = \|f(\chi)\|$ where $f: \boldsymbol{R} \to H$, H is a strictly convex space, and f is additive. Z. MOSZNER constructs a family of functions ϕ satisfying (1), (2) and (3) for which FISCHER's conjecture does not hold, using a Hamel basis.

P28S1

On a Problem of Kemperman Concerning Hamel Functions

J. A. BAKER, and, independently, S. L. SEGAL

At the Fifth annual meeting on Functional Equations held in Waterloo, Ontario, April, 1967, J. H. B. KEMPERMAN posed the problem:

Determine whether or not there exists a sequence $\{a_k\}$ of real constants ($a_k \neq 0$ for infinitely many k) and a corresponding non-linear additive function φ on the reals such that

$$\sum_{k=1}^{\infty} a_k \varphi(x^k) \tag{1}$$

is convergent for each real number x.

A real-valued function of a real variable φ will be called a *derivation* in case

$$\varphi(x+y) = \varphi(x) + \varphi(y) \quad \text{and}$$
$$\varphi(xy) = x\varphi(y) + y\varphi(x)$$

for all real x and y. If φ is a derivation it is easy to prove, by induction, that

$$\varphi(x^k) = k x^{k-1} \varphi(x) \tag{2}$$

for every real x and every positive integer k. It is also not difficult to show that a derivation must vanish at every algebraic number. Thus the only linear (continuous) derivation is the trivial derivation $\varphi \equiv 0$. The existence of many non-trivial derivations follows from [2], pages 120–131, and also [3]. We remark that a non-linear additive function is nowhere continuous, non-measurable, and has a graph which is dense in the plane, (see [1], pages 31–36).

Let $\{a_k\}$ be a sequence of real numbers such that $\sum_{k=0}^{\infty} a_k x^k$ converges for every real number x. Denote the sum of this series by $f(x)$ and let φ be a non-trivial derivation. Then

$$\varphi(x) f'(x) = \varphi(x) \sum_{k=1}^{\infty} k a_k x^{k-1} = \sum_{k=1}^{\infty} a_k (k x^{k-1} \varphi(x))$$

for every real x so that, by (2), $\sum_{k=1}^{\infty} a_k \varphi(x^k)$ converges absolutely to $\varphi(x) f'(x)$ for every real x.

J. A. BAKER notes further that if φ_1 and φ_2 are derivations and if we define

$$\varphi(x) \stackrel{\text{Def}}{=} \varphi_1(\varphi_2(x))$$

for every real x, then φ is additive and

$$\varphi(x^k) = k(k-1) x^{k-2} \varphi_1(x) \varphi_2(x) + k x^{k-1} \varphi_1(\varphi_2(x))$$

for each real x and every positive integer k. If $\{a_k\}$ and f are as above, then clearly $\sum_{k=1}^{\infty} a_k \varphi(x^k)$ converges absolutely to

$$f''(x)\,\varphi_1(x)\,\varphi_2(x) + f'(x)\,\varphi_1(\varphi_2(x))$$

for every real x. In general, if φ is the composition of a finite number of derivations then it is not difficult to show that $\sum_{k=1}^{\infty} a_k \varphi(x^k)$ converges absolutely for every real x with $\{a_k\}$ chosen as above.

S. L. Segal notes: we know that there exists a Hamel-function $L(x)$ such that $M(x) \overset{\text{Def}}{=} e^{-x} L(e^x)$ is also a non-linear additive function. Problem: describe $L(x)$ in terms of a Hamel basis.

The interesting question *Given a non-linear additive function* $\varphi(x)$, *for what sequences* $\{a_k\}$ *does* (1) *converge for all real* x remains open and is no doubt difficult.

S. L. Segal wishes to thank Charles Watts and Newcomb Greenleaf for conversation concerning derivations.

REFERENCES

[1] Aczél, J. *Lectures on Functional Equations and Their Applications* (Academic Press, New York and London 1966).
[2] Zariski, O. and Samuel, P., *Commutative Algebra*, Vol. 1. (D. Van Nostrand, Princeton 1958).
[3] Bourbaki, N., *Algèbre*, Vol. 5 (Hermann et Cie, Paris 1959), § 9.

University of Waterloo,
University of Rochester

P30S1

On a System of Matrix Functional Equations of Belousov

S. L. SEGAL

At the International Symposium in Functional Equations held in April, 1967 in Waterloo, Ontario, V. D. BELOUSOV asked for the solution of the system of matrix functional equations

$$F(X+Y)=F(X)+F(Y) \tag{1}$$

$$F([F(X)-X]\,[F(Y)-Y])=\{\{0\}\}. \tag{2}$$

where the domain and range of F are assumed to be the set of all $n \times n$ matrices over the real numbers for some fixed n, and $\{\{0\}\}$ indicates the $n \times n$ matrix all of whose entries are 0.

Throughout this note $\{\{C_{ij}\}\}$ indicates the matrix with entries C_{ij}; matrices without reference to their entries are denoted by upper-case Latin letters; lower case being reserved for the entries; $X=\{\{x_{ij}\}\}$, $Y=\{\{y_{ij}\}\}$, I is the $n \times n$ identity matrix.

In this note we prove the

THEOREM: *The most general solution of the system* (1), (2) *(with the matrices defined over the real numbers) which is bounded on an n^2-dimensional set of positive (Lebesgue n^2-dimensional) measure is*

$$F(X)=\{\{x_{ij}(1-e_{ij})\}\}$$

where either $e_{ij}=1$ or $e_{ip}e_{pj}=0$ for all p, $1 \leqslant p \leqslant n$.

REMARKS: Two interesting questions which remain are:

(i) What can be said about the existence of Hamel-type solutions of the system (1), (2)? (see **P30S2**).

(ii) What happens if the matrices are defined over a finite field?

PROOF: The equation (1) is easily reduced to the 1-dimensional case, and it then follows from well-known results that the most general solution of (1) which is bounded above on an n^2-dimensional subset of positive measure is

$$F(X)=\{\{f_{ij}x_{ij}\}\} \quad \text{where} \quad X=\{\{x_{ij}\}\} \tag{3}$$

(e.g. [1], p. 347–349, and references therein).

Defining $e_{ij}=1-f_{ij}$ and substituting (3) in (2) gives

$$F(\{\{-e_{ij}x_{ij}\}\}\,\{\{-e_{ij}y_{ij}\}\})=\{\{0\}\},$$

and so by (3) we get

$$(1 - e_{ij}) \sum_{k=1}^{n} e_{ik} x_{ik} e_{kj} y_{kj} = 0 \tag{4}$$

presumed to hold for all $x_{ij}, y_{ij}, 1 \leqslant i, j \leqslant n$.

Let $U_{\alpha,\beta}$ be defined as the matrix with a 1 in the α^{th} row and β^{th} column and otherwise having zeros for entries. For each i, j, and each p, $1 \leqslant p \leqslant n$, taking

$$\left.\begin{array}{l} X = U_{i,p}, \quad Y = U_{p,j} \quad \text{gives} \\ (1 - e_{ij})\, e_{ip} e_{pj} = 0, \quad 1 \leqslant p \leqslant n \end{array}\right\} \tag{5}$$

for each i, j; whence the theorem follows since it is easily checked that the condition is also sufficient.

Additional Remarks: (i) If $F(X) = \{\{x_{ij}(1 - e_{ij})\}\}$ can be realized by either just a pre- or just a post-multiplication; i.e. if $F(X)$ has either of the forms DX or XD, then $D^2 = D$.

In fact, suppose $X - F(X) = AXB$; then (2) becomes

$$A^2 X B A Y B^2 = A X B A Y B$$

for all matrices X, Y. Letting $X = Y = (BA)^{\dagger}$, where $(BA)^{\dagger}$ denotes the Moore-Penrose generalized inverse of BA (cf [2], [3]), we get

$$A^2 (BA)^{\dagger} B^2 = A (BA)^{\dagger} B$$

or,

$$B A^2 (BA)^{\dagger} B^2 A = BA. \tag{6}$$

If B or $A = I$, then (6) becomes $A^2 = A$ or $B^2 = B$ respectively, and clearly if $K^2 = K$ then also $(I - K)^2 = I - K$. (ii) A similar computation shows that if $X - F(X) = AXB$ and A and B are both non-singular then $B = A^{-1}$.

REFERENCES

[1] Aczél, J., *Lectures on Functional Equations and their Applications* (Academic Press, 1966).
[2] Penrose, R., *A Generalized Inverse for Matrices*, Proc. Camb. Phil. Soc. *51*, 406–413 (1955).
[3] Rado, R., *Note on Generalized Inverses of Matrices*, Proc. Camb. Phil. Soc. *52*, 600–601 (1956).

University of Rochester

P30S2

Réponse à un problème au sujet des fonctions additives

ZENON MOSZNER

Monsieurs V. D. BELOUSOV, S. L. SEGAL et G. SZEKERES ont formulé la conjecture (voir [1]) que le système des équations fonctionnelles

$$f(X+Y)=f(X)+f(Y) \tag{1}$$

$$f(f(X)\,f(Y))=f(f(X)\,Y+X\,f(Y)-X\,Y), \tag{2}$$

où f est une fonction de l'ensemble M^M où M est l'ensemble des $n\times n$ matrices des nombres reels, possède seulement des solutions de la forme

$$f(X)=\big(x_{ij}(1-e_{ij})\big) \tag{3}$$

où $X=(x_{ij})$ et pour chaque i et j:

$$e_{ij}=1 \quad \text{ou} \quad e_{ip}e_{pj}=0 \quad \text{pour chaque} \quad p=1,2,\dots,n.$$

M. SEGAL a démontré dans la note [2] que cette conjecture est exacte sous la supposition complementaire que la fonction f est borné sur un ensemble n^2-dimensionnel de mesure lebesguienne n^2-dimensionnelle positive. Il a posé dans la même note la question: qu'est qu'on peut dire au sujet de l'existence des solutions du type de Hamel pour les équations (1) et (2).

On peut démontrer facilement qu'il existe des solutions de cette sorte, donc qu'il existe des solutions des équations (1) et (2) qui ne sont pas de la forme (3).

En effet pour la fonction additive l'équation (2) est équivalente à l'équation

$$f((f(X)-X)\,(f(Y)-Y))=0. \tag{4}$$

Si nous posons $n=1$, la rélation (3) prend la forme

$$f(X)=(0) \quad \text{ou} \quad f(X)=X. \tag{5}$$

Si nous donc posons

$$f^*(w_0+w_1\,b_1+\cdots+w_m\,b_m)=w_1\,b_1+\cdots+w_m\,b_m$$

où $b_1, b_2, \dots, b_m$ sont les éléments différent pas rationnels et en outre arbitraires de base de Hamel $\mathscr{B}$ de l'ensemble des nombres reels, $w_0, w_1, \dots, w_m$ sont rationnels et arbitraires et $1\in\mathscr{B}$, dans ce cas la fonction f^* est additive, $f^*(x)=0$ pour x rationnel, $f^*(x)-x$ est un nombre rationnel pour chaque x et de là la fonction $f(x)=(f^*(x_{11}))$ remplit l'équation (4), n'étant en même temps de la forme (5).

TRAVAUX CITÉS

[1] Compte rendu de «Fifth Annual Meeting on Functional Equations» en Aequationes Mathematicae, P30, Vol. I.
[2] SEGAL, S. L., *On a System of Matrix Functional Equations of Belousov*, en Aequationes Mathematicae, P30S1, cette issue, p. 116.

Wyzsza Szkola Ped.

P33S1

A propos d'une conjecture de M. B. Schweizer

ZENON MOSZNER

M. B. SCHWEIZER a formulé, pendant la cinquième conférence annuèlle des équations fonctionnelles en l'Université de Waterloo [1], l'hypothèse suivante. Soit $C[0, 1]$ l'espace des fonctions continues, qui transforment l'intervalle $[0, 1]$ dans lui-même, muni de norme

$$\|f\| \overset{\text{df}}{=} \sup_{x\in[0,1]} |f(x)|.$$

Désignons par $\mathscr{K}$ l'ensemble des paires (f, g) des fonctions de l'espace $C[0, 1]$ telles que

$$f(g(x)) = g(f(x)) \quad \text{pour} \quad x\in[0, 1],$$

et qui ont le même point fixe, donc pour lesquelles il existe un point x_0 de $[0, 1]$ tel que[1])

$$f(x_0) = x_0 = g(x_0). \tag{1}$$

On conjecture que $\mathscr{K}$ est de $\text{I}^{\text{ère}}$ catégorie dans $C[0, 1]\times C[0, 1]$.

L'auteur de cette hypothèse ne dit pas quelle est la norme ou en général la métrique dans $C[0, 1]\times C[0, 1]$, mais on peut deviner qu'on prend une norme simple pour le produit cartesien, c'est-à-dire la norme

$$\|(f, g)\| = \|f\| + \|g\| \quad \text{ou} \quad \|(f, g)\| = \sqrt{\|f\|^2 + \|g\|^2} \quad \text{ou enfin} \quad \|(f, g)\| = \max(\|f\|, \|g\|).$$

Si cela a lieu, alors la conjecture ci-haut est exacte, puisque nous allons démontrer un peu plus: *l'ensemble B des paires de l'espace* $C[0, 1]\times C[0, 1]$ *qui remplissent la condition* (1) *est non-dense.*

---

[1]) Le symbole $\mathscr{K}$ correspond ici a $A \wedge B$ dans le problème **P33** et dans la note **P33S2**.

Remarquons d'abbord que l'ensemble B est fermé, donc il suffit de montrer qu'il est frontière.

Puisque la condition (x) est équivalente à la condition

$$\{x: f(x) - x = 0\} \cap \{x: g(x) - x = 0\} \neq \phi,$$

donc pour démontrer que B est un ensemble frontière, il suffit de démontrer le lemme suivant:

LEMME. *Soit F l'ensemble des fonctions f continues sur l'intervalle* $[0, 1]$ *et remplissants sur* $[0, 1]$ *la condition*

$$-x \leqslant f(x) \leqslant 1 - x. \tag{2}$$

Pour chaque $\varepsilon > 0$ et pour chaque paire (f, g) de $F \times F$ il existe une paire (f_1, g_1) de $F \times F$ telle que

$$\|(f, g) - (f_1, g_1)\| \leqslant \varepsilon$$

et

$$\{x: f_1(x) = 0\} \cap \{x: g_1(x) = 0\} = \phi.$$

Démonstration du lemme. Soit ε arbitraire, positif, et soit (f, g) une paire de $F \times F$. Nous allons démontrer d'abord qu'il existe une fonction f_1 de F telle que

$$\|f - f_1\| \leqslant \frac{\varepsilon}{2}$$

et l'ensemble

$$Z_1 = \{x: f_1(x) = 0\}$$

est non-dense.

L'ensemble

$$Z_2 = \{x: f(x) = 0\}$$

est fermé, donc chaque composante est un intervalle fermé. Si chaque composante se reduit à un point, l'ensemble Z_2 est non-dense et il suffit de poser $f_1 = f$. Dans le cas contraire sur chaque composante $[a, b]$ de Z_2 qui ne se reduit pas à un point (donc $a < b$) prenons pour f_1 la fonction continue qui remplit les conditions suivantes

$$f_1(a) = f_1(b) = 0, \quad 0 \leqslant f_1(x) \leqslant \min\left(b - x, \frac{\varepsilon}{2}\right),$$

et

$$f_1(x) \neq 0 \quad \text{sur} \quad (a, b).$$

Si nous posons $f_1 = f$ sur les autres points de l'intervalle $[0, 1]$ nous recevons la fonction qui remplit les conditions exigés.

Passons à présent à la définition de la fonction g_1. L'ensemble

$$E^* = \left\{x : |g(x)| < \frac{\varepsilon}{4}\right\}$$

est ouvert dans [0, 1], alors il est la somme d'une suite finie ou non des intervalles $\Delta_1, \Delta_2, \Delta_3 \ldots$ overts dans [0, 1] et disjoints. Prenons un intervalle Δ_ν et considérons les cas suivants – la fonction g prend sur les deux éxtremités a_ν et b_ν $(a_\nu < b_\nu)$ de Δ_ν

a) les valeurs différentes de zéro et du signe contraire, ou
b) les valeurs différentes de zéro et du même signé, ou
c) $g(a_\nu)=0$ et $g(b_\nu)\neq 0$, ou
d) $g(b_\nu)=0$ et $g(a_\nu)\neq 0$, ou enfin
e) $g(a_\nu)=g(b_\nu)=0$.

Dans le cas a) soit x_0 le point de l'intervalle Δ_ν qui n'appartient pas à Z_1 et posons g_1 sur Δ_ν comme la fonction continue, prenant la valeur zéro seulement dans le point x_0 et remplissant les conditions

$$g_1(a_\nu) = g(a_\nu), \quad g_1(b_\nu) = g(b_\nu), \quad -x \leqslant g_1(x) \leqslant 1 - x$$

et

$$|g_1(x)| < \frac{\varepsilon}{4}.$$

Dans le cas b) posons

$$g_1(x) = g(a_\nu) \quad \text{sur} \quad \Delta_\nu.$$

Si le cas c) a lieu, alors $a_\nu = 0$. Supposons que x_0 a la même signification que plus haut et soit g_1 une fonction continue qui remplit les conditions

$$g_1(b_\nu) = g(b_\nu), \quad -x \leqslant g_1(x) \leqslant 1 - x, \quad |g_1(x)| < \frac{\varepsilon}{4}$$

et si $g(b_\nu)>0$ on a $g_1(x)\neq 0$ sur Δ_ν, par contre dans le cas $g(b_\nu)<0$ a lieu $g_1(x)=0$ sur Δ_ν seulement pour $x=x_0$. Dans le cas d) (qui a lieu seulement pour $b_\nu=1$) nous définissons la fonction g_1 analogiquement comme dans le cas c).

Enfin dans le cas e) (qui a lieu si et seulement si $\Delta_\nu=[0, 1]$) nous prenons pour g_1 une fonction continue qui remplit les conditions

$$|g_1(x)| < \frac{\varepsilon}{4}, \quad -x \leqslant g_1(x) \leqslant 1 - x,$$

et $g_1(x)=0$ seulement pour $x=x_0$, où x_0 a la même signification comme plus haut.

En posant $g_1(x)=g(x)$ sur l'ensemble $[0, 1]\setminus E^*$, nous recevons la fonction qui avec la fonction f_1 satisfait aux conditions du lemme.

TRAVAUX CITÉS

[1] Compte rendu de «Fifth Annual Meeting on Functional Equations», Conjecture 10, Aequationes Mathematicae *1* (1968).

Wyzsza Szkola Ped.

P33S2

On a Conjecture of B. Schweizer

MAREK KUCZMA

At the 5th Meeting on Functional Equations B. SCHWEIZER has expressed a conjecture [5] which may be formulated as follows.

Let $C=C[0, 1]$ be the space of continuous functions on the interval $[0, 1]$ with values in $[0, 1]$, under the usual sup norm, and put $\mathscr{F}=C\times C$. Let $A\subset\mathscr{F}$ be the set of those $(f, g)\in\mathscr{F}$ that commute: $f\circ g=g\circ f$, and let $B\subset\mathscr{F}$ be the set of those $(f, g)\in\mathscr{F}$ that have a common fixed point: $f(x_0)=g(x_0)=x_0$, $x_0\in[0, 1]$. The conjecture is that the set $A\cap B$ is of the first category in $\mathscr{F}$.

Now, the truth of B. SCHWEIZER's conjecture is striking if one realizes how rare a feature is the permutability among pairs of continuous functions! In fact, we are going to show that *the set A is nowhere dense in $\mathscr{F}$*. The same applies, of course, to the set $A\cap B\subset A$.

A is evidently closed and thus it is enough to show that it has no inner points. Let $(f, g)\in A$ and suppose that one of the functions, say g, is not the identity. Then there exists an $x_0\in[0, 1]$ such that $y_0\overset{\mathrm{df}}{=}g(x_0)\neq x_0$. Given an $\varepsilon>0$, one can find a function $f_1\in C$ with the following properties

$$f_1(x_0)=f(x_0), \quad f_1(y_0)\neq f(y_0), \quad \|f-f_1\|<\varepsilon.$$

Now, the pair (f_1, g) is as close to (f, g) as one likes, provided ε has been chosen small enough, but $(f_1, g)\notin A$, since

$$f_1(g(x_0))=f_1(y_0)\neq f(y_0)=f(g(x_0))=g(f(x_0))=g(f_1(x_0)).$$

This shows that A has no inner points (adding one point consisting of two identity functions cannot change the situation).

As Z. MOSZNER [4] has recently proved, the set B also is nowhere dense in $\mathscr{F}$. However, his proof is much longer than that given above.

It was a long-standing conjecture [3] that $A \subset B$, i.e. $A \cap B = A$. This has been recently disproved by W. M. Boyce [1] and J. P. Huneke [2]. The latter result was the origin of B. Schweizer's conjecture. But in view of the present considerations as well as of those in [4] the problem of whether the set $A \cap B$ is in some respect small (e.g., whether it is of the first category) in A seems more adequate. (Similiarly, one can investigate the smallness of $A \cap B$ in B). A problem to this effect has indeed been raised by B. Schweizer (cf. **P33**).

Jagello University, Katowice, Poland
and
Mathematical Institute of the Polish Academy of Sciences

REFERENCES

[1] Boyce, W. M., *Commuting Functions with no Common Fixed Point*, Notices Amer. Math. Soc. *14*, 280 (1967).
[2] Huneke, J. P., *Two Counterexamples to a Conjecture on Commuting Functions of the Closed Unit Interval*, Notices Amer. Math. Soc. *14*, 284 (1967).
[3] Isbell, J. R., *Commuting Mappings of Trees*. Research problem 7. Bull. Amer. Math. Soc. *63*, 419 (1957).
[4] Moszner, Z., *A propos d'une conjecture de M. B. Schweizer*, Aequationes Math. **P33S1**.
[5] Schweizer, B., Conjecture 10, Fifth Annual Meeting on Functional Equations, Aequationes Math. *1* (1968).

Short Communications

This section consists of self-contained 25–100 line short communications which are prepublications of results, the details of which are to be published in either *aequationes mathematicae* or other journals with referee systems or equivalent systems of pre-reviewing papers. *Aequationes mathematicae* will endeavour to publish these short communications in the shortest possible time after the underlying papers have been accepted for publication. Unless indicated otherwise details of results described in this section will appear in subsequent issues of *aequationes mathematicae*.

Solution générale de l'équation fonctionnelle $f[x+yf(x)]=f(x)f(y)$*)

S. Wołodźko

Dans son travail l'auteur donne la construction fournissant les solutions de l'équation fonctionnelle $f[x+yf(x)]=f(x)f(y)$ dans l'espace linéaire E sur un corps variable Φ. Cette construction est la suivante.

Supposons que la fonction $f(x)$ soit l'application de l'ensemble E dans l'ensemble Φ, ce que nous noterons $f\colon E\to\Phi$. Nous passons maintenant à un moyen d'obtenir les solutions de l'équation en question.

1. Nous choisissons un sous-groupe multiplicatif arbitraire Z' du groupe multiplicatif $\Phi'=\Phi\setminus\{0\}$.
2. Nous déterminons dans l'ensemble Z' une fonction arbitraire $w(\lambda)$ dont les valeurs appartiennent à l'ensemble E, c'est-à-dire $w\colon Z'\to E$.
3. Nous déterminons la fonction $a(\lambda, \mu)=w(\lambda\mu)-w(\lambda)-\lambda w(\mu)$ pour $\lambda, \mu\in Z'$.
4. Nous désignons par A' un sur-ensemble arbitraire de l'ensemble des valeurs de la fonction $a(\lambda, \mu)$, c'est-à-dire $a(Z'\times Z')\subset A'$.
5. Nous désignerons par A'' l'ensemble de tous les produits λa, où $a\in A'$ et $\lambda\in Z'$.
6. L'ensemble A'' génère un groupe additif; nous le désignerons par A.
7. Choisissons un sous-groupe multiplicatif arbitraire Z du groupe multiplicatif Z' tel que pour tout $\lambda\in Z$ et $\lambda\neq 1$, $w(\lambda)\notin A$.
8. Déterminons la fonction

$$f(x)\overset{\mathrm{df}}{=}\begin{cases}\lambda & \text{s'il existe } \lambda\in Z \text{ tel que } [x-w(\lambda)]\in A\,,\\ 0 & \text{pour les autres } x \text{ restants}.\end{cases}$$

L'auteur applique cette construction pour les démonstrations des théorèms 1–4 du travail cité de S. Gołąb et A. Schinzel [1] ainsi que pour trouver les solutions

*) Cf. pp. 12–29.

continues de l'équation en question dans l'ensemble des fonctions complexes de la variable complexe. Le résultat est contenu dans le théorèm suivant.

PROPOSITION. Les seules solutions continues de l'équation en question dans l'ensemble des fonctions complexes de la variable complexe sont les fonctions suivantes.

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

$$f(x) = 1 + c \cdot x, \quad \text{où } c \text{ est une constante arbitraire complexe},$$

$$f(x) = \begin{cases} 1 + c \cdot \operatorname{Im} x, & \text{si} \quad 1 + c \cdot \operatorname{Im} x > 0 \\ 0 & \text{si} \quad 1 + c \cdot \operatorname{Im} x \leqslant 0, \end{cases} \quad (c \text{ – constante réelle}),$$

$$f(x) = \begin{cases} 1 + c \cdot \operatorname{Re} x, & \text{si} \quad 1 + c \cdot \operatorname{Re} x > 0 \\ 0 & \text{si} \quad 1 + c \cdot \operatorname{Re} x \leqslant 0. \end{cases} \quad (c \text{ – constante réelle}),$$

A Grammar of Functions*)

B. SCHWEIZER and A. SKLAR

For a given non-empty set S and positive integers d, r, let $\mathscr{F}_{rd}$ denote the set of all functions whose domains are (non-empty) subsets of S^d and whose ranges are subsets of S^r. If $F \in \mathscr{F}_{rd}$, then we call d the *degree* of F (written δF) and r the *rank* of F (written ϱF); the *index* of F (written ιF) is the integer $\delta F - \varrho F$. The union of all the sets $\mathscr{F}_{rd}$, together with the empty function $\emptyset$ is denoted by $\mathscr{F}_\infty$, and an element of $\mathscr{F}_\infty$ is called a *function over S*.

Into the set $\mathscr{F}_\infty$ we introduce 2 unary operations L, R, and 3 binary operations σ, π, β. For any function F over S, LF is the identity function on the range of F, and RF is the identity function on the domain of F. The binary operation σ, called *serial composition*, coincides with ordinary composition $\circ$ of (partial) transformations whenever the latter operation yields non-empty results; but $F\sigma G$ can be non-empty even when $\delta F \neq \varrho G$, in which case $F \circ G$ is always empty. Like ordinary composition, serial composition is associative; in fact the system $(\mathscr{F}_\infty, \sigma, L, R)$ is a *function semigroup* (cf. B. SCHWEIZER and A. SKLAR, *The Algebra of Functions*, Math. Annalen *139*, 366–382 (1960), *143*, 440–447 (1961), *161*, 171–196 (1965); *Function Systems*, Ibid. *172*, 1–16 (1967). In this function semigroup we have $\varrho RF = \delta RF = \delta F$, $\delta LF = \varrho LF = \varrho F$ whenever $F \neq \emptyset$, and $\delta(F\sigma G) = \max(\delta F + \iota G, \delta G)$, $\varrho(F\sigma G) = \max(\varrho F, \varrho G - \iota F)$, $\iota(F\sigma G) = \iota F + \iota G$, whenever $F\sigma G \neq \emptyset$.

The other binary operations, π (called *parallel composition*) and β, are also associative, and we have $\delta(F\pi G) = \max(\delta F, \delta G)$, $\varrho(F\pi G) = \varrho F + \varrho G$, $\delta(F\beta G) = \delta F + \delta G$, $\varrho(F\beta G) = \varrho F + \varrho G$, $\iota(F\beta G) = \iota F + \iota G$ whenever $F\pi G$ (resp., $F\beta G$) is non-empty. There are many relations, e.g., conditional distributive laws, connecting

*) Cf. pp. 62–85.

the binary operations, and some of these relations involve special classes of functions, called *selectors* and *multiselectors*, over S.

The structure thus developed has many applications. We show, for example, that all the usual 'compositions' and 'modes of generation' of multiplace functions can be subsumed under σ, π, and β, plus the use of multiselectors. We also indicate applications to questions of linearity and multilinearity, to constants, place-fixing, 'essential variables', and representation theorems, and to functional equations. We conclude with some remarks on the definition of recursive functions.

Remarks on the Square Norm *)

M. Hosszú

In a linear space G the function $f(x)=F(x, x)$ has the usual property of the square norm

$$f(x+y)+f(x-y)=2f(x)+2f(y), \quad x, y \in G \tag{1}$$

if $(x, y) \to F(x, y)$ is a real or complex valued bilinear function (cf. [2]). J. Aczél [1] has given the solution f of the functional equation (1) by means of a function F additive in both variables assuming that G is an abelian group and the range is another abelian group A in which every equation $4a=b$ has a unique solution $a \in A$ for arbitrarily fixed $a, b \in A$. If we omit this last solvability condition and the commutativity of $(G, +)$ then, assuming only $f(0)=0$, the following results can be established:

I) The functions

$$(x, y) \to F_1(x, y)=f(x+y)-f(x)-f(y), \; (x, y) \to F_2(x, y)=f(x+y)-f(x-y)$$

have the properties $F_2(x, x)=2F_1(x, x)=4f(x)$ and

$$\left.\begin{array}{l} F(x, y)=F(y, x), \quad F(x, 0)=0, \quad F(x+y, x+y)=F(y+x, y+x), \\ \qquad F(x+y, x-y)=F(y+x, x-y), \\ \qquad F(x+y, z)+F(x-y, z)=2F(x, z), \quad x, y, z \in G \end{array}\right\} \tag{2}$$

for every solution f of (1).

II) The function $x \to f(x)=F(x, x)$ satisfies $f(0)=0$ and

$$2f(x+y)+2f(x-y)=4f(x)+4f(y), \quad x, y \in G \tag{1'}$$

for every F having properties (2).

III) The function F_1 defined above is additive in both variables iff $f(x+y+z)=f(y+x+z)$ holds identically.

*) Received December 4, 1967

REFERENCES

[1] Aczél, J., *The General Solution of Two Functional Equations by Reduction to Functions Additive in Two Variables and with the Aid of Hamel Bases*, Glasnik Mat.-Fiz. i Astr. *20*, 65–73 (1965).

[2] Yosida, K., *Functional Analysis* (Berlin–Heidelberg–New York 1966), esp. pp. 39–40.

Ein allgemeiner Vierscheitelsatz für ebene Jordankurven*)

O. Haupt

Der sogenannte Vierscheitelsatz geht zurück auf S. Mukhopadhyaya (1909) und A. Kneser (1912); dabei wird unter einem Scheitel der betrachteten Kurve C (in der euklidischen Ebene) jeder Punkt $s \in C$ vertanden derart, dass beliebig kleine Umgebungen von s auf C (minsestens) vier Punkte enthalten, die auf einem Kreis liegen. (Diese Definition ist – jedenfalls für stetig gekrümmte Kurven C – gleichwertig mit der klassischen Definition des Scheitels als eines Punktes extremer Krümmung.) Während bei Kneser beliebige stetig gekrümmte Kurven C zugelassen sind, beschränkt sich Mukhopadhyaya auf (stetig gekrümmte) Ovale, verfügt aber andererseits über eine allgemeinere Beweismethode. Diese benutzt nämlich im wesentlichen nur die folgende Eigenschaft (K) der Kreise: Ein Kreis ist durch beliebige 3 seiner Punkte eindeutig bestimmt und ändert sich stetig mit solchen Punkten; überdies liefert sie Vierscheitelsätze nicht nur für Ovale, sondern für alle (Jordan-)Kurven C, welche – abgesehen von geeigneten Differenzierbarkeitseigenschaften (D) – der Bedingung (N) genügen: Auf der orientierten Kurve C besitzen, bei geeignet orientiertem Kreis K die Punkte von $C \cap K$ die gleiche Reihenfolge wie auf K. Demnach liefert die Methode von Mukhopadhyaya 4-Scheitelsätze für Kurven C, welche die Bedingungen (D) und (N) erfüllen, auch dann, wenn das System der Kreise ersetzt wird durch ein System $\mathfrak{k}$ von Kurven, welches die Eigenschaft (K) besitzt und im übrigen beliebig ist; man bemerkt dazu, dass sich der oben an erster Stelle erklärte Begriff des Scheitels wörtlich auf den Fall der Systeme $\mathfrak{k}$ überträgt. Die Methode von Mukhopadhyaya ist also durchaus topologischer Natur. Da bei Kneser die Bedingung (N) nicht benötigt wird, erhebt sich erstens die Frage nach der Entbehrlichkeit von (N) auch bei Mukhopadhyaya, und zweitens – entsprechend dem topologischen Charakter der Betrachtungen – die Frage nach einer topologisch formulierbaren Bedingung (D). Dass ohne ein solches (D) im allgemeinen nur zwei Scheitel vorhanden sind, zeigen Beispiele (2-Scheitelsatz). In der vorliegenden Arbeit werden beide Frage bejahend beantwortet durch den Nachweis, dass für Jordankurven C und beliebige Systeme $\mathfrak{k}$ der Vierscheitelsatz gilt, falls in der Umgebung eines jeden isolierten Scheitels eine gewisse topologische Beziehung (D) zwischen C und den Kurven $K \in \mathfrak{k}$ erfüllt ist.

Die vorliegende erste Mitteilung gibt eine Zusammenstellung der erforderlichen Definitionen und Hilfssätze sowie den ersten Teil des Beweises, der im wesentlichen

*) Eingegangen am 8.12.1967

auf kombinatorischen Betrachtungen beruht. – Eine zweite, später erscheinende Mitteilung bringt den noch fehlenden Beweis eines Existenzsatzes für Scheitel mit Hilfe des Grundgedankens von MUKHOPADHYAYA.

Non Negative Definite Solutions of Certain Differential and Functional Equations*)

E. LUKACS

Problems in probability theory or mathematical statistics are often formulated in terms of random variables or distribution functions since this is the natural way in which they occur. However, it is often desirable, and also possible, to reformulate these problems in such a way that they involve only the Fourier-Stieltjes transforms of distribution functions, called in probabilistic terminology characteristic functions.

The original problem takes then the form of a differential equation or of a functional equation which the characteristic function (or functions) must satisfy. However, it is not enough to solve this equation since the solution of the equation will in general contain also functions which are not characteristic functions and which do not lead to a solution of the original probabilistic or statistical problem. It is known (BOCHNER's theorem) that a function $f(t)$ is a characteristic function if, and only if, it is non-negative definite and if $f(0)=1$. The main difficulty in problems of this kind is often the determination of those solutions of a differential or functional equation which are non-negative definite.

The paper gives a survey of studies of these problems. Section 2 deals with certain functional equations while section 3 treats questions concerning the non-negative definite solutions of certain differential equations.

Sur le reste de certaines formules de quadrature**)

TIBERIU POPOVICIU

Le reste $R[f]$ de la formule de quadrature

$$\int_a^b f(x)\,dx = \sum_{\alpha=1}^{n} A_\alpha f(x_\alpha) + R[f] \tag{1}$$

où les noeuds x_α de l'axe réel sont distincts et les A_α sont des constantes réelles données, est de la forme

$$R[f] = A[\xi_1, \xi_2, \ldots, \xi_{m+2}; f] + B[\xi'_1, \xi'_2, \ldots, \xi'_{m+2}; f] \tag{2}$$

*) Received December 12, 1967
**) Reçu le 12.12.1967

La fonction f est supposée continue, m est le degré d'exactitude de la formule (1), les points ξ_α d'une part et les points ξ'_α d'autre part, sont distincts mais dépendent en général de la fonction f. Les constantes A, B sont indépendantes de la fonction f. Enfin $[y_1, y_2, \ldots, y_r; f]$ désigne la différence divisée, d'ordre $r-1$, de la fonction f sur les noeuds $y_1, y_2, \ldots, y_r$.

Lorsque dans (2) on peut prendre $B=0$ le reste $R[f]$ (ou la formule de quadrature (1)) est dit *de la forme simple.* Cette dernière notion est en étroite liaison avec la théorie des fonctions convexes d'ordre supérieur [1].

Dans le présent travail on montre comment, sous des hypothèses bien précisées, dans le cas où le reste n'est pas de la forme simple, on peut rétablir une sorte de simplicité par une généralisation convenable de la notion de différence divisée et de la convexité correspondante.

[1] Popoviciu, Tiberiu, Mathematica *1* (24), 95–142 1959.

Endomorphismenringe in der Galoisschen Theorie*)

Wolfgang Krull

Es sei N ein beliebiger Körper. Jedem $c \in N$ werde der durch „$c'a = c \cdot a$ für alle $a \in N$“ festgelegte Endomorphismus c' der additiven Gruppe N zugeordnet. Dann gilt: *Allgemeinster Unabhängigkeitssatz*: Sind $A_1, \ldots A_n$ paarweise verschiedene Automorphismen des Körpers (und damit Endomorphismen der additiven Gruppe) N, so folgt aus „$(\sum_{i=1}^{n} c'_i A_i)\,\alpha = 0$ für alle $\alpha \in N$“ stets $c'_1 = \cdots = c'_n = 0$. – Der Beweis arbeitet mit der Vandermondeschen Determinante; er ist konstruktiv und völlig elementar. – Es sei jetzt N separabel und normal über K mit der Galoisgruppe $G = \{A_1, \ldots, A_n\}$. Für einen beliebigen Körper L zwischen N und K werde $L' = \{c' \mid c \in L\}$ gesetzt. Nach Artin betrachtet man schon in der elementaren Galoisschen Theorie den Ring K'_G aller Endomorphismen $\{\sum_{i=1}^{n} a'_i A_i \mid a_i \in K\}$, der zum Gruppenring von G über K isomorph ist. Angesichts des allgemeinsten Unabhängigkeitssatzes erscheint es angebracht, neben K'_G gleich auch den Ring N'_G aller Endomorphismen $\{\sum_{i=1}^{n} c'_i A_i \mid c' \in N\}$ einzuführen. N'_G ist ein Vektorraum der Dimension n^2 über K', aber *nicht* zum Gruppenring von G über N isomorph, weil i.a. $A_i c' = (A_i c)' A_i \neq c' A_i$. Definiert man für den Körper L das Rechtsideal $R'_{G,L}$ von N'_G durch $R'_{G,L} = \{\phi \mid \phi c \in L$ für alle $c \in N\}$, so gelten die Sätze: Es ist $R'_{G,L}$ gleich der von dem Komplexprodukt $L' \cdot R'_{G,K}$ erzeugten additiven

*) Eingegangen am 13.12.1967

Gruppe. $R'_{G,K}$ ist ein *minimales* Rechtsideal von N'_G und ein Vektorraum der Dimension n über K'. Man hat $R'_{G,K}=\{\sum_{i=1}^{n}(A_i c)' A_i \mid c\in N\}$. – Das letzte Resultat zeigt, daß der schon durch seine Minimaleigenschaft ausgezeichnete Ring $R'_{G,K}$ eng mit der symmetrischen Bilinearform verknüpft ist, die man erhält, wenn man nach dem Vorbild von Dedekind die Menge $N\times N$ der Paare $\langle c,d\rangle$ $(c, d\in N)$ durch die Vorschrift „$(c,d)=\mathrm{Spur}\, c\cdot d$" bilinear in K abbildet. Die frühe Einführung von $R'_{G,K}$ dürfte also für den Aufbau der Galoisschen Theorie ebenso nützlich sein wie die von K'_G.

Functional Equations in Vector Spaces, Composition Algebras, and Systems of Partial Differential Equations*)

WOLFGANG EICHHORN

Let X be a vector space over a commutative field F of characteristic $\neq 2$ and let x be an arbitrary element in X. The following problem is considered: Find a linear mapping L of X into the vector space $\mathrm{Hom}(X,X)$ of the linear transformations of X with the property that there exists another such mapping, say M, such that the product of the two transformations $M(x)$ and $L(x)$ satisfies the functional equation

$$M(x)\,L(x)=\mu(x)\,I \qquad \begin{cases} I \text{ identity map } X\to X, \\ \mu\colon X\to F,\ \mathrm{char}\,F\neq 2,\ \mu\not\equiv 0. \end{cases} \tag{1}$$

The assumption that L and M are linear says that the following further functional equations hold:

$$\left.\begin{aligned} &L(\alpha x+\beta y)=\alpha L(x)+\beta L(y), \qquad M(\alpha x+\beta y)=\alpha M(x)+\beta M(y), \\ &\mu(\alpha x)=\alpha^2\mu(x) \qquad (\alpha,\beta\in F;\ x,y\in X). \end{aligned}\right\} \tag{1'}$$

One sees that μ is a quadratic form.

Two applications of the above problem shall be mentioned:

a) (1) is of interest in the theory of algebras: Every solution $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x)$ of (1) leads to the definition of an algebra $\bar{\mathscr{A}}$ over the vector space X by defining

$$x\,y := \bar{L}(x)\,y \tag{2}$$

as multiplication in X. $\bar{L}(x)$ is then called *left regular representation* of the element $x\in\bar{\mathscr{A}}$. Perhaps then new classes of non-associative algebras can be found.

b) In case $F=\mathbf{R}$, $\mu(x)$ non-degenerate, the above problem can be considered as

*) Received December 15, 1967

the following algebraic problem arising in the theory of systems of partial differential equations: Find systems

$$\left(\sum_{i=1}^{n} \gamma_k^{ij} \frac{\partial}{\partial \xi_i}\right)\begin{pmatrix}\varphi_1\\ \vdots \\ \varphi_n\end{pmatrix} = 0 \quad (\gamma_k^{ij} \in \mathbf{R};\, i, j, k = 1, 2, \dots n), \quad \text{short:} \quad L\left(\frac{\partial}{\partial x}\right) v = 0, \tag{3}$$

such that there exists another operator of the same kind, say $M(\partial/\partial x)$, with the following property:

$$\left.\begin{aligned} &M\left(\frac{\partial}{\partial x}\right) L\left(\frac{\partial}{\partial x}\right) v = \left(\alpha \frac{\partial^2}{\partial \xi_1^2} + \beta \frac{\partial^2}{\partial \xi_1 \partial \xi_2} + \dots + \omega \frac{\partial^2}{\partial \xi_n^2}\right) v = 0 \\ &\quad (\alpha, \beta, \dots, \omega \in \mathbf{R}), \quad \text{short:} \quad M\left(\frac{\partial}{\partial x}\right) L\left(\frac{\partial}{\partial x}\right) v = \mu\left(\frac{\partial}{\partial x}\right) v = 0. \end{aligned}\right\} \tag{4}$$

This problem is of interest in connection with

(i) generalization of function theory (especially if the partial differential equation (4) is the Laplace equation)

(ii) Dirac's 'linearization' of the wave equation

$$\left(\frac{\partial^2}{\partial \xi_1^2} - \frac{\partial^2}{\partial \xi_2^2} - \frac{\partial^2}{\partial \xi_3^2} - \frac{\partial^2}{\partial \xi_4^2}\right) \varphi(\xi_1, \xi_2, \xi_3, \xi_4) = 0. \tag{5}$$

Let $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x)$ be a solution of (1), let P, Q, R be non-singular linear transformations, let $\lambda \in F$, $\neq 0$. Then it is obvious that

$$\hat{L}(x) := P\, \bar{L}(Qx)\, R, \qquad \hat{M}(x) := \lambda R^{-1} \bar{M}(Qx) P^{-1}, \qquad \hat{\mu}(x) := \lambda \bar{\mu}(Qx) \tag{6}$$

is a solution of (1), too. In consequence of this remark we call two solutions $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x)$ and $\hat{L}(x)$, $\hat{M}(x)$, $\hat{\mu}(x)$ of (1) *isotopic* if there exist non-singular P, Q, R, and $\lambda \neq 0$ such that (6) holds. Note that the relation 'isotopy' is an equivalence relation and that the following definition is due to Albert: Two algebras $\bar{\mathscr{A}}$, $\hat{\mathscr{A}}$ with left regular representations $\bar{L}(x)$, $\hat{L}(x)$ are called *isotopic* if the first identity in (6) is valid.

THEOREM: *Let $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x)$ be a solution of* (1) *such that the algebra $\bar{\mathscr{A}}$ defined by $\bar{L}(x)$ is isotopic to an algebra $\mathscr{A}$ with identity element e. Then $\mathscr{A}$ is an alternative quadratic algebra. If, on the other hand, $\mathscr{A}$ is an alternative quadratic algebra and $L(x)$ its left regular representation, then one can easily find an $M(x)$ and a $\mu(x)$ such that $L(x)$, $M(x)$, $\mu(x)$ is a solution of* (1).

Consequences of the Theorem: A) The problem of determining all solutions $L(x)$, $M(x)$, $\mu(x)$ of (1) with the property that the algebras defined by $L(x)$ are isotopic to algebras with e is equivalent to the problem of determining all alternative quadratic algebras.

All alternative quadratic non-degenerate algebras, the so-called *composition algebras* $\mathscr{C}$, are known: They only exist in dimensions $n=1, 2, 4, 8$, and one has only the following examples:

$n=1$: $\mathscr{C}=Fe$

$n=2$: $\mathscr{C}$ is an extension field of F of degree 2 or isomorphic to $F \oplus F$

$n=4$: $\mathscr{C}$ is a quaternion algebra over F

$n=8$: $\mathscr{C}$ is a Cayley algebra over F.

B) Assume in the above Theorem that $\bar{\mu}$ is non-degenerate. Then $\mathscr{A}$ is one of the above composition algebras.

Since every division algebra is isotopic to an algebra with identity element e one has the following corollary of B):

C) If an algebra defined by a solution of (1) is a division algebra, then it is isotopic to one of the above composition algebras.

From C) follows

D) Every elliptic system of partial differential equations (3) with the property (4) can be transformed into a system

$$\left(\frac{\partial}{\partial \xi_1} e_1 + \frac{\partial}{\partial \xi_2} e_2 + \cdots + \frac{\partial}{\partial \xi_n} e_n\right)(\varphi_1 e_1 + \varphi_2 e_2 + \cdots + \varphi_n e_n) = 0,$$

where the $e_1, e_2, \ldots, e_n$ form the usual basis of the complex numbers, quaternions, or Cayley numbers.

Further results: E) If the dimension of X equals $n=p2^q$ (p odd, $q=0, 1, 2, \ldots$) then there does not exist any solution of (1) such that the rank of μ is greater than $2q+2$.

The statement E) contains in particular:

F) If the dimension n of X is finite and if μ in (1) is non-degenerate then solutions of (1) exist for $n=1, 2, 4, 8$ only. In other words: Systems of partial differential equations (3) with the property (4), μ non-degenerate, exist for $n=1, 2, 4, 8$ only.

G) In dimensions 2 and 4 there exist examples of solutions $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x)$ of (1), $\bar{\mu}$ non-degenerate, such that the algebras defined by $\bar{L}(x)$ are *not* isotopic to a composition algebra.

Zur Begründung der Theorie der automorphen Funktionen in mehreren Variablen*)

M. Eichler

Under a number of assumptions the following theorems on automorphic forms in n variables are proved:

A. There exist $n+1$ independent automorphic forms.

B1. The number of linearly independent automorphic forms of a given degree h is $O(h^n)$.

*) Received January 2, 1968

B2. The quotients of automorphic forms of equal degrees form a finite algebraic function field of n variables.

C1. Let $y_0(z), \ldots, y_n(z)$ be independent automorphic forms of equal degrees. Denote the domain of the variables by $\mathfrak{Z}$, the underlying group by Γ. The $y_\nu(z)$ map the quotient $\mathfrak{Z}/\Gamma$ on an open point set of a projective variety $\mathfrak{M}$ of dimension n. The closure of this set in $\mathfrak{M}$ is equal to $\mathfrak{M}$.

C2. The graded ring of automorphic form is generated by a finite set.

D. All Siegel modular forms can be generated by such whose Fourier expansions have rational integral coefficients.

The assumptions are easily verified in the case of the 'classical' modular groups. The group Γ is a subgroup of a Lie group of analytic mappings of $\mathfrak{Z}$ onto itself. Some assumptions concern the behaviour of the Fourier expansions in the cusps. The compactification (C1) hinges on the assumption, that a specialization to automorphic functions in one variable is possible. For the proof of D theta constants are used.

Canonical Decompositions, Stable Functions, and Fractional Iterates*)

A. SKLAR

Following B. SCHWEIZER and the author [*Function systems*, Math. Annalen *172*, 1–16 (1967)], functions are treated as elements of axiomatically defined 'function systems'. Each such system has a binary operation $\circ$ corresponding to composition, and 2 unary operations L, R corresponding to the projections of a function onto the identity functions on its range and domain, respectively.

It is shown that all functions f in an appropriate function system admit a 'canonical decomposition' into the composite $f_1 \circ f_2$ of an 'invertible' function f_1 and a 'strongly idempotent' function f_2 with $Rf_1 = Lf_2$. It is noteworthy that the 'concrete' analog of this abstract result is equivalent to the Axiom of Choice.

Next, canonical decompositions $f = f_1 \circ f_2$ such that $f^n = f_1^n \circ f_2$ for all iterates f^n of f are taken up. The 'separable' functions that admit such decompositions are characterized, and an important subclass of separable functions, the 'stable' functions, are characterized in turn. The 'concrete' analogs of these characterizations are the following: the function f is separable if the restriction of f to its range is an invertible function; f is stable if it is separable and its range is a subset of its domain. Thus the (real) sine function is stable; the (real) cosine function, on the other hand, is not even separable, but its 2nd iterate (the cosine of the cosine) is stable.

These notions are then applied to the study of the functional equation $f^n = g$, where g is a given, and f an unknown function, and $n > 1$ is a given integer. This 'fractional

*) Received January 8, 1968

iterate' equation has generally been studied only under the assumption that both f and g are invertible. Such restrictions can now be partially removed: thus we can, for example, characterize all stable fractional iterates of stable functions. Other applications of the results of this paper will appear later.

Some Graphical Properties of Matrices with Non-Negative Entries*)

A. L. DULMAGE and N. S. MENDELSOHN

A *bipartite graph* K has two vertex sets X and Y and a set of edges (which may be null) each of which is a pair of vertices (x, y), $x \in X$, $y \in Y$. A *transversal* of K is a subgraph G every vertex of which has exactly one edge. The *order* of a transversal is the number of its edges. If A is an $m \times n$ matrix of non-negative elements, *the bipartite graph K of A* (denoted by K_A) is the bipartite graph for which the vertex sets X and Y are the sets of rows and columns of A. The pair (x, y) is an edge of K_A if and only if the element in row x and column y of A is greater than zero. An $m \times n$ *sub-permutation matrix* of rank r is an $m \times n$ matrix of $mn-r$ zeros and r ones with at most one non-zero in any row or column. The bipartite graph of a sub-permutation matrix of rank r is a transversal of order r, together with $m+n-2r$ free vertices.

Let X and Y be arbitrary sets and let $f(x, y)$, $x \in X$, $y \in Y$, be a function with domain $X \times Y$ and range some subset of the non-negative reals. The bipartite graph K of the function f (denoted by K_f) has vertex sets X and Y and (x, y) is an edge of K if and only if $f(x, y) > 0$. A *sub-permutation function* g of rank r on $X \times Y$ is a function such that $g(x, y) = 0$ for all pairs (x, y), $x \times X$, $y \times Y$ with the exception of a set of pairs $(x_1, y_1), (x_2, y_2), \ldots (x_r, y_r)$, $x_i \neq x_j$, $y_i \neq y_j$ for $i \neq j$ for which $g(x, y) = 1$.

For a non-negative real valued function f with domain $X \times Y$ we define the bipartite graph K_f. The vertex sets of K_f are X and Y and the pair (x, y), $x \in X$ and $y \in Y$ is an edge if and only if $f(x, y) > 0$.

We say that a transversal G of a bipartite graph K *catches* a vertex if this vertex is a vertex of G, and that G catches a set U of vertices if every vertex of U is a vertex of G.

If W is a finite set, $|W|$ denotes the order of W, that is, the number of elements in W.

The bipartite graph $X \times Y$ is the graph with vertex sets X and Y such that every pair (x, y), $x \in X$, $y \in Y$ is an edge of the graph. If K is a bipartite graph with vertex set X and Y, and if $U \subset X$ and $V \subset Y$ then $(U \times V) \cap K$ denotes the bipartite graph with vertex sets U and V every edge of which is an edge (u, v) of K such that $u \in U$, $v \in V$.

If K is a bipartite graph with vertex sets X and Y then K has a unique subgraph which has the same edges as K and is such that every vertex of the subgraph has at least one edge. This subgraph is called the *essential* subgraph of K and is denoted by K'.

*) Received January 10, 1968

Reference will be made to irreducible and semi-irreducible graphs and to the canonical decomposition of a bipartite graph as defined in (2), (5) and (6).

Consider an $m \times n$ matrix $A\,(n \geqslant m)$ of non-negative reals with sum S and maximum row or column sum M. Let R_i be the sum of the elements in the ith row and C_j the sum of the elements in the jth column. Thus $M = \mathrm{Max}(R_i, C_j)\; i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$, and $S = \sum_{i=1}^{m} R_i = \sum_{j=1}^{n} C_j$. This $m \times n$ matrix can be augmented to an $n \times n$ matrix B of non-negative reals with the same S and M, by adding $n-m$ rows of zeros. The matrix B has all the row and column sums of A and in addition has $R_i = 0$ for $i = m+1, \ldots, n$. Let $\sigma = |S/M|$. As shown in (4), the matrix B may be extended to a matrix C with row and column sums A/σ by the addition of $n-\sigma$ rows and $n-\sigma$ columns provided $n > \sigma$. Let $B = (b_{ij})$. The matrix $C = (c_{ij})$ is defined as follows:

$$c_{ij} = b_{ij} \quad \text{for} \quad i \leqslant n, j \leqslant n,$$

$$c_{ij} = \frac{\dfrac{S}{\sigma} - R_i}{n - \sigma} \quad \text{for} \quad i \leqslant n, n+1 \leqslant j \leqslant 2n - \sigma,$$

$$c_{ij} = \frac{\dfrac{S}{\sigma} - C_j}{n - \sigma} \quad \text{for} \quad n+1 \leqslant i \leqslant 2n - \sigma, j \leqslant n,$$

$$c_{ij} = 0 \quad \text{for} \quad n+1 \leqslant i \leqslant 2n - \sigma, n+1 \leqslant j \leqslant 2n - \sigma.$$

Since C has equal row and column sums it is expressible in the form $\sum k_i P_i$ where the k_i are >0 and the P_i are permutation matrices (of order $2n-\sigma$). Thus the bipartite graph K_C of the matrix C has the property that for every edge e of K_C there exists a transversal G of order $2n-\sigma$ of which e is an edge. Let K_A be the bipartite graph of A. The transversal G has exactly $2(n-\sigma)$ edges which are not edges of K_A and hence has exactly $2n-\sigma-2(n-\sigma)=\sigma$ edges which are edges of K_A.

Now let A be a matrix of non-negative integers with sum S and maximum row or column sum M and let $\sigma = |S/M|$ (the greatest integer in S/M). It is shown in this paper that A is expressible as a sum of subpermutation matrices of rank σ and a subpermutation matrix of rank $q, 0 \leqslant q < \sigma$. Further, let K be a bipartite graph with finite vertex sets X and Y. Let S be the total number of edges and let M be the maximum number of edges which have the same vertex. It is shown also that K has a transversal G of order $\sigma = |S/M|$ with vertex sets U and V such that every vertex of X which has M edges is an element of U and every vertex of Y which has M edges is an element of V.

Let X and Y be finite or countably infinite sets and let f be a function with domain $X \times Y$ and range some subset of the non-negative reals such that the bipartite graph

K_f has a maximal transversal, and $S=\sum_{x\in X}\sum_{y\in Y} f(x,y)$ is finite. We conjecture that f is expressible as a possibly infinite sum $\sum c_i g_i$ where the c_i are positive reals and the g_i are sub-permutation functions on $X\times Y$ of rank r, if and only if $1\leqslant r\leqslant\sigma$.

On a 'Cube Functional Equation'*)

H. Haruki

In a previous paper in this journal we studied the following functional equation which we called a 'square functional equation':

$$\left.\begin{aligned} f(x+t,y+t)+f(x-t,y+t)&+f(x-t,y-t) \\ &+f(x+t,y-t)=4f(x,y), \end{aligned}\right\} \tag{1}$$

where x, y, t are real variables and $f(x,y)$ is a real-valued function of two real variables x, y in the whole xy-plane.

In this paper we consider the following functional equation:

$$\left.\begin{aligned} f(x+t,y+t,z+t)&+f(x+t,y+t,z-t) \\ &+f(x+t,y-t,z+t)+f(x+t,y-t,z-t) \\ &+f(x-t,y+t,z+t)+f(x-t,y+t,z-t) \\ &+f(x-t,y-t,z+t)+f(x-t,y-t,z-t) \\ &=8f(x,y,z), \end{aligned}\right\} \tag{2}$$

where x, y, t are real variables and $f(x,y,z)$ is a real-valued function of three real variables x, y, z in the whole xyz-space.

By considering the geometric meaning of (2), we call (2) a 'cube functional equation'.

We prove the following theorem:

Theorem: *If $f(x,y,z)$ is a real-valued continuous function of three real variables x,y,z in the whole xyz-space and satisfies* (2) *in the whole xyz-space, then and only then*

$$f(x,y,z)=\sum_{0\leqq i,j,k\leqq 5} c_{ijk}\frac{\partial^{i+j+k}}{\partial x^i\partial y^j\partial z^k}P(x,y,z),$$

where c_{ijk} $(0\leqq i,j,k\leqq 5)$ are real constants and $P(x,y,z)=xyz(y^2-z^2)(z^2-x^2)(x^2-y^2)$.

In order to prove this theorem, we use distributions and two lemmas.

*) Received January 22, 1968

Contents Vol. 2 · 1

Contents Vol. 2 · 1 (in continuation)

aequationes mathematicae

Vol. 2
No. 2/3
Pag. 137–408
1969

Dedicated to
the 75th Birthday of
Alexander M. Ostrowski
September 25th 1968

University of Waterloo, Ontario, Canada
Birkhäuser Verlag, Basel, Switzerland

aequationes mathematicae

Annual subscription:
(3 fascicles per volume)
sFr. / DM. 40.—
US-$ 9.50. Single copy:
sFr. / DM 20.—
US-$ 4.90

aequationes mathematicae is published by the University of Waterloo and Birkhäuser Verlag. Price per volume (three issues) is 40 Swiss francs including postage and costs of mailing, payable at the office of the publisher, Birkhäuser Verlag in Basel, Switzerland or through your bookseller.

The journal will publish papers in pure and applied mathematics in general but, in particular, papers in functional equations, combinatorial and numerical analysis. The languages in which papers will generally be accepted are the international congress languages: English, French, German, Italian, and Russian.

Types of manuscripts that will be considered for publication include original research papers, reports of meetings and activities, problems and their solutions, expository papers and, as short communications, self-contained abstracts of papers which have been accepted for publication either by this or some other journal with referee system.

Manuscripts of papers should be submitted to the Editorial Office or to a member of the editorial board. Authors of papers should send the original and, preferably, two additional copies. Since it is the policy of the journal to publish in advance short communications of all results it accepts for publication, authors should include with their original submissions a self-explanatory abstract of their paper of length 25–100 lines.

PUBLISHER: Birkhäuser Verlag, 4000 Basel 10, Switzerland © 1968
EDITORIAL OFFICE: University of Waterloo, Faculty of Mathematics, Waterloo, Ontario, Canada

Expository Papers

Non-Negative Definite Solutions of Certain Differential and Functional Equations

EUGENE LUKACS (Washington, D.C.)[1]

Dedicated to A. M. Ostrowski on the Occasion of his 75th Birthday

1. Introduction

Problems in probability theory or mathematical statistics are frequently stated in terms of random variables or distribution functions. Since there exists a one-to-one correspondence between distribution functions and their Fourier-Stieltjes transforms (called in probabilistic terminology characteristic functions) it is often possible to reformulate problems in probability theory or mathematical statistics so that they involve only characteristic functions. The original problem takes then the form of a differential equation or of a functional equation which the characteristic function (or functions) must satisfy. However, it is not enough to solve this equation since the solutions of the equation will in general contain also functions which are not characteristic functions. It is known (BOCHNER's theorem) that a function f is a characteristic function if, and only if, it is non-negative definite[2]) and $f(0)=1$.

The main difficulty in the above mentioned problems is often the task of determining those solutions of a differential or of a functional equation which are non-negative definite.

No systematic study of this question seems to be available at present. The purpose of this paper is to give a survey of some of the isolated results which are scattered in the literature. It is hoped that this exposition might stimulate further research in this area.

In section 2 we discuss certain functional equations, section 3 deals with questions concerning differential equations. We indicate also – as far as possible – the probabilistic motivation which led to the study of these equations.

[1]) Research supported by the Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force under AFOSR grant AF-AFOSR-437-65.

[2]) A complex valued function f of a real variable t is said to be non-negative definite if it is continuous and if for any positive integer N and any real $t_1, t_2, \ldots, t_N$ and any complex $\xi_1, \xi_2, \ldots, \xi_N$ the sum $\sum_{j=1}^{N} \sum_{k=1}^{N} f(t_j - t_k)\xi_j \bar{\xi}_k$ is real and non-negative.

Received December 12, 1967

2. Functional Equations

The first equation which we discuss is of great importance in probability theory. It occurs in connection with the study of the limit distributions of normalized sums of independently and identically distributed random variables. This study leads to the following problem: Determine the characteristic functions (non-negative definite functions) f which have the property that to every $b_1>0$, $b_2>0$ there corresponds a $b>0$ and a real γ such that the functional equation

$$f(b_1 t)\, f(b_2 t) = f(b t)\, e^{i\gamma t} \tag{2.1}$$

is satisfied for all real t. It can be shown that all non-negative definite solutions have the form

$$f(t) = \exp\left\{i a t - c|t|^{\alpha}\left[1 + i\beta \frac{t}{|t|}\, \omega(|t|, \alpha)\right]\right\} \tag{2.2}$$

where

$$\omega(|t|, \alpha) = \begin{cases} \operatorname{tg} \dfrac{\pi\alpha}{2} & \text{if } \alpha \neq 1 \\ \dfrac{2}{\pi} \log|t| & \text{if } \alpha = 1 \end{cases}$$

while a, c, β and α are real parameters such that

$$c \geqslant 0, \quad |\beta| \leqslant 1, \quad 0 < \alpha \leqslant 2. \tag{2.3}$$

If the conditions (2.3) are not satisfied then (2.2) is still a solution but is not non-negative definite. The solution is obtained by studying functional equations of the form

$$g(a u) = g(a_1 u) + g(a_2 u).$$

The characteristic functions (2.2) are called stable characteristic functions.

The next problem which we consider deals with two non-negative definite functions. The problem is to determine all characteristic functions f which satisfy the relation

$$f(t) = f(c t)\, f_c(t) \tag{2.4}$$

where $0<c<1$ while f_c is a characteristic function which depends on the parameter c. The non-negative definite solution of (2.4) are given by

$$\left.\begin{aligned} f(t) = \exp\Bigg\{ i t a - \frac{\sigma^2 t^2}{2} &+ \int_{-\infty}^{-0} \left(e^{itu} - 1 - \frac{itu}{1+u^2}\right) dM(u) \\ &+ \int_{+0}^{\infty} \left(e^{itu} - 1 - \frac{itu}{1+u^2}\right) dN(u) \Bigg\} \end{aligned}\right\} \tag{2.5}$$

where a is real, $\sigma^2 \geqslant 0$ while M and N satisfy the following conditions:

(i) M [resp. N] is non-decreasing in $(-\infty, 0)$ [resp. $(0, \infty)$] and $M(-\infty)= =N(+\infty)=0$,

(ii) $\int_{-\varepsilon}^{0} u^2 dM(u)$ and $\int_{0}^{\varepsilon} u^2 dN(u)$ are finite for all $\varepsilon>0$,

(iii) M and N have left and right derivatives and $u \to uM'(u)$ [resp. $uN'(u)$] is non-increasing for $u<0$ [resp. $u>0$]. Here $M'(u)$ and $N'(u)$ are either right or left derivatives and may be different at different points.

The characteristic functions satisfying (2.4) are called self-decomposable characteristic functions (or functions of the L-class); their distributions are the limit distributions of normalized sums of independently (but not necessarily identically) distributed random variables.

Equations (2.1) and (2.4) were first studied by P. LÉVY; for the proofs of the statements we refer the reader to GNEDENKO–KOLMOGOROV [1].

D. A. RAIKOV [14] showed that the normal distribution is characterized by the fact that all of its factors belong to the same type.[3]) He did this by investigating the functional equation

$$f(t) = e^{iat} f(c_1 t) f(c_2 t) \tag{2.6}$$

where c_1 and c_2 are positive real numbers while a is some real number. He succeeded in proving that the only non-negative definite solution of (2.4) which is not identically zero is the characteristic function of the normal distribution, that is the function

$$t \to f(t) = \exp[i\alpha t - \sigma^2 t^2/2]. \tag{2.7}$$

In the examples which we discussed until now it was possible to use probabilistic methods and to determine the non-negative definite solutions of the functional equation (2.6) [resp. (2.1) or (2.4)] without investigating all solutions. The situation is different in the problems which we discuss next. In these cases it will be necessary to determine first all solutions and then to select those which are non-negative definite.

A number of authors investigated the stochastic independence of linear forms. Final results were obtained almost simultaneously and independently by V. P. SKITOVICH and G. DARMOIS. They studied the following equation involving n complex valued functions of real variables:

$$\prod_{j=1}^{n} f_j(a_j u + b_j v) = \prod_{j=1}^{n} f_j(a_j u) f_j(b_j v) \tag{2.8}$$

---

3) In the theory of factorizations of characteristic functions one considers only factors which are themselves non-degenerate characteristic functions (that is which do not have the form $\exp(iat)$ with a real). We say that two characteristic functions f_1 and f_2 belong to the same type if $f_1(t) = e^{itb} f_2(at)$ with $a>0$ and b real.

where $a_1, \ldots, a_n; b_1, \ldots, b_n$ are real and such that $a_j b_j \neq 0$ for $j=1, 2, \ldots, n$. The above mentioned authors showed that the solution of this equation has the form

$$f_j(u) = \exp[P_j(u))] \quad (j = 1, 2, \ldots, n) \tag{2.8a}$$

where the $P_j(u)$ are polynomials of degree not exceeding n. (For details of the proof see E. LUKACS–R. G. LAHA [8]). The solutions (2.8a) are however not necessarily non-negative definite functions. A theorem, commonly called Marcinkiewicz' theorem (see J. MARCINKIEWICZ [13]) states that a function of the form $t \to e^{P(t)}$ where P is a polynomial can be non-negative definite only in case the degree of P does not exceed 2. Using well known properties of characteristic functions it is then possible to show that all $f_j(u)$ are characteristic functions of normal distributions possibly having different parameters, that is

$$f_j(u) = \exp[i u \alpha_j - \sigma_j^2 u^2/2] \quad (j = 1, 2, \ldots, n).$$

J. MARCINKIEWICZ derived the above mentioned theorem as a tool needed for his investigation of identically distributed linear forms. In this connection he studied the functional equation

$$\prod_j f(a_j t) = \prod_j f(b_j t) \tag{2.9}$$

where the product can be either finite or infinite. If it is infinite it is assumed to be uniformly convergent in every finite interval. The a_j and b_j are real numbers such that the sequence $\{|a_j|\}$ is not a permutation of the sequence $\{|b_j|\}$. MARCINKIEWICZ also assumed that f has derivatives of all orders. Using these assumptions he could show that necessarily $f(t) = \exp[P(t)]$ where P is a polynomial. It follows then, as in the previous case, that the only non-negative definite solution of (2.9) is the characteristic function (2.7).

The case of a finite product (2.9)

$$\prod_{j=1}^{r} f(a_j t) = \prod_{j=1}^{r} f(b_j t) \tag{2.9a}$$

is of particular interest and is the object of a study of Yu. V. LINNIK [6]. He showed (§ 55 of [6]) that (2.9a) admits non-negative definite solutions different from (2.7) if the assumption that f has derivatives of all orders is dropped. Yu. V. LINNIK presents a deep and exhaustive study of the equation (2.9a), his results are based on the investigation of the analytic properties of the entire function

$$z \to G(z) = \sum_{j=1}^{r} (|a_j|^z - |b_j|^z).$$

The equation

$$\prod_j f(a_j t) = f(t) \tag{2.10}$$

was studied by E. LUKACS–R. G. LAHA [2]. Here no differentiability assumption is needed, only the inequality

$$\infty > \Sigma a_j^2 \geqslant 1$$

had to be assumed. The solution of (2.10) is then again (2.7). This work was motivated by studying a finite or infinite linear form in independently and identically distributed random variables and assuming that the distribution function of the linear form is identical with the common distribution of the random variables occurring in the linear form.

We conclude this section by listing some equations which occur in connection with the characterization of certain stochastic processes. These equations differ in two respects from those treated in the first part of this section: (1) They involve the logarithms of characteristic functions instead of the characteristic functions themselves, (2) they contain integrals.

In the following we denote by f a characteristic function and by φ its logarithm. Let A and B be two possible real numbers, $A < B$, and suppose that a and b are two continuous functions in $[A, B]$ such that $\max_{A \leqslant t \leqslant B} |a(t)| \neq \max_{A \leqslant t \leqslant B} |b(t)|$. The equation

$$\int_A^B \varphi[u\,a(t)]\,dt = \int_A^B \varphi[u\,b(t)]\,dt \tag{2.11}$$

occurs in connection with a characterization of the Wiener process. It can be shown that φ is a polynomial so that it follows from the theorem of MARCINKIEWICZ that $f = \exp\varphi$ has the form (2.7). A number of similar equations or systems of such equations were studied by R. G. LAHA and E. LUKACS ([3], [4], [5]) and by E. LUKACS ([9], [10]).

3. Differential Equations

The study of non-negative definite solutions of differential equations was originally motivated by certain characterization problems. We describe briefly one example. It is desired to determine a non-negative definite function f which has the property that $\varphi = \log f$ satisfies the differential equation

$$\varphi^{(p+r)}(t) - i^r \varphi^{(p)}(t) = i^{p+r} C \tag{3.1}$$

with initial conditions

$$\varphi^{(s)}(0) = i^s \kappa_s \quad [s = 0, 1, \ldots, (p + r - 1)]\,. \tag{3.1a}$$

Here $\varphi^{(s)}(t) = (d^s/dt^s)\,\varphi(t)$ while the κ_s are certain real parameters (called cumulants) of the distribution whose characteristic function is f. Equation (3.1) occurs in connection with the characterization of the Poisson distribution by a regression property. Equation (3.1) can be solved easily, however, it is rather complicated to determine the non-negative definite functions $f=\exp\varphi$ obtainable from the solutions φ of (3.1). One gets (see [11] and [12])

$$f(t)=\exp\{\lambda_1(e^{it}-1)+\delta_r\lambda_2(e^{-it}-1)+c_1\,it-\varepsilon_p c_2 t^2/2\}$$

where $\lambda_1\geqslant 0$, $\lambda_2\geqslant 0$, $c_2\geqslant 0$, c_1 real and

$$\delta_r=\begin{cases}0 & \text{if } r \text{ is an odd integer}\\ 1 & \text{if } r \text{ is an even integer}\end{cases}\quad\text{while}\quad \varepsilon_p=\begin{cases}0 & \text{if } p=1\\ 1 & \text{if } p>1\end{cases}\quad (p \text{ integer}).$$

A number of similar differential equations were studied by various authors who worked on characterization problems. If one surveys these works one notices the absence (and also the desirability) of a general theory. Such a theory would be useful even if it does not yield the non-negative definite solutions of differential equations but succeeds only to give analytical properties of the non-negative definite solutions. We conclude this section by describing the first successful attempt in this direction. For details of this work we refer to papers by Yu. V. LINNIK [7] and A. A. ZINGER–Yu. V. LINNIK [15].

We consider an ordinary differential equation with real constant coefficients $A_{j_1,\ldots,j_n}$

$$\sum A_{j_1\ldots j_n} i^{-(j_1+\cdots+j_n)} f^{(j_1)}(t)\ldots f^{(j_n)}(t)=c[f(t)]^n. \tag{3.2}$$

The summation runs over all non-negative integers $j_1,\ldots,j_n$ such that

$$j_1+\cdots+j_n\leqslant p. \tag{3.3}$$

Here p is an integer such that at least one of the coefficients $A_{j_1\ldots j_n}$ with $j_1+\cdots+j_n=p$ does not vanish. Let m be the order of the differential equation (3.2). We introduce the polynomial

$$A(x_1,\ldots,x_n)=\frac{1}{n!}\sum{}^*\sum A_{j_1\ldots j_n} x_{k_1}^{j_1}\ldots x_{k_n}^{j_n}. \tag{3.4}$$

The first summation $\sum^*$ is taken over all permutations $(k_1,\ldots,k_n)$ of the first n integers while the second summation runs over all $(j_1,\ldots,j_n)$ satisfying the inequality (3.3). We say that the differential equation (3.2) is positive-definite if $A(x_1,\ldots,x_n)$ is a non-negative polynomial. A. A. ZINGER and Yu. V. LINNIK obtained the following results

(A) All non-negative definite functions f which satisfy in a neighborhood of the origin a positive definite differential equation have derivatives of all orders at the origin.

(B) All non-negative definite functions which satisfy in a neighborhood of the origin a positive definite differential equation (3.2) of order $m\geqslant n-1$ are entire functions.

A. A. Zinger–Yu. V. Linnik [15] imposed also further conditions on the differential equation (3.2) which assured that the non-negative definite solutions have necessarily the form (2.7), that is they obtained a characterization of the normal distribution. However, these conditions are so restrictive that many known characterizations of the Normal population are excluded. Several known characterizations of the Gamma distribution lead to differential equations similar to (3.1) which are not positive definite. It would be interesting to obtain results similar to (A) and (B) for equations of this type, of course it might be necessary to replace the condition that the equation be positive definite by a different assumption. The study of differential equations whose non-negative definite solutions are rational functions would also be valuable.

REFERENCES

[1] Gnedenko, B. V. and Kolmogorov, A. N., *Limit Distributions for Sums of Independent Random Variables* (Addison-Wesley, Cambridge, Mass. 1954).

[2] Laha, R. G. and Lukacs, E., *On Linear Forms Whose Distribution is Identical with that of a Monomial*, Pacific J. Math. *15*, 207–214, (1965).

[3] Laha, R. G. and Lukacs, E., *On a Characterization of the Wiener Process*, in *Trans. of the Second Prague Conference* (1960), pp. 307–312.

[4] Laha, R. G. and Lukacs, E., *On Identically Distributed Stochastic Integrals*, in *Trans. of the Third Prague Conference* (1964), pp. 467–474.

[5] Laha, R. G. and Lukacs, E., *On Linear Forms and Stochastic Integrals*, in *Proc. 35th Session of the International Statistical Institute, Belgrade*, Vol. 2 (1965), pp. 828–840.

[6] Linnik, Yu. V., *Linear Forms and Statistical Criteria*, I, II (Russ.), Ukrain Math. Ž. *5*, 207–243, 247–290 (1953), – *Selected Transl. in Math. Stat. and Probab.*, Amer. Math. Soc. *3*, 1–90 (1962).

[7] Linnik, Yu. V., *On Polynomical Statistics in Connection with the Analytical Theory of Differential Equations* (Russ.), Vestnik Leningrad Univ. *11*, 35–48 (1956). = *Selected Transl. in Math. Stat. and Probab.*, Amer. Math. Soc. *1* 191–206 (1961).

[8] Lukacs, E. and Laha, R. G., *Applications of Characteristic Functions* (Charles Griffin, London 1964).

[9] Lukacs, E., *Some Results in the Theory of the Wiener Integral*, in *Trans. of the Fourth Prague Conference* (1967), pp. 29–43.

[10] Lukacs, E., *Une caractérisation des processus stables et symétriques*, C. R. Acad. Sci. Paris *264*, 959–960 (1967).

[11] Lukacs, E. *On the Characterization of a Family of Populations Which Includes Poisson Populations*, Ann. Univ. Sci. Budapest. Eötvös Sect. Math. *3–4*, 159–175 (1960–1961).

[12] Lukacs, E., *Some Extensions of a Theorem of Marcinkiewicz*, Pacific J. Math. *8*, 487–501 (1958).

[13] Marcinkiewicz, J. *Sur une propriété de la loi de Gauss* Math. Z. *44*, 612–618 (1939), also in *Collected Papers* (P.W.N., Warszawa 1964), pp. 463–469.

[14] Raikov, D. A. *On the Decomposition of Gauss and Poisson Laws*, Izv. Akad. Nauk SSSR Ser. Mat. *2*, 91–124 (1938).

[15] Zinger, A. A. and Linnik, Yu. V., *On a Class of Differential Equations with Applications to Certain Questions of Regression Theory* (Russ.), Vestnik Leningrad Univ. *12*, 121–130 (1957). = *Selected Transl. in Math. Stat. and Probab.* Amer. Math. Soc. *3*, 181–190 (1962).

The Catholic University of America

Research Papers

Über ein Funktionalgleichungssystem der Informationstheorie

Zoltán Daróczy (Debrecen, Hungary)

Herrn Professor Alexander M. Ostrowski zum 75. Geburtstag gewidmet

Einleitung

In den Arbeiten [5], [6] und [7] von J. Kampé de Fériet und B. Forte kann man eine axiomatische Grundlegung des Informationsbegriffes finden. In diesen Untersuchungen wird die Information ohne Benützung des Wahrscheinlichkeitbegriffes erklärt. In der Arbeit [7] spielt ein Funktionalgleichungssystem eine Rolle, das auch in sich selbst interessant ist. Dieses System wurde von C. Baiocchi in der Arbeit [3] gelöst. Das Hauptergebnis von [3] ist das folgende:

Es bezeichne R die Menge der reellen Zahlen und es sei $f: R \to R$ eine Funktion, für die die Bedingungen

$$f[x+f(y)]+f(y)=f[y+f(x)]+f(x) \quad (x, y \in R), \tag{1}$$

$$f(-x)=f(x)+x \quad (x \in R), \tag{2}$$

$$\lim_{x \to -\infty} f(x)=0 \tag{3}$$

erfüllt sind. Ist f stetig, so hat $f(x)$ die Darstellung

$$f(x)=-\tfrac{1}{2}\{|x|+x\}, \tag{4}$$

oder

$$f(x)=-\frac{1}{A}\ln(1+e^{Ax}), \tag{5}$$

wobei $A>0$ ein konstanter Wert ist.

Die Grundidee des Beweises ist die folgende: Man kann zeigen, dass f im Falle $f(x) \neq -\frac{1}{2}\{|x|+x\}$ eine stetige Ableitung der Ordnung zwei besitzt, und aus dieser Behauptung folgt eine Differentialgleichung für die unbekannte f. Die Lösungen dieser Differentialgleichung sind die Funktionen (5).

In dieser Arbeit wollen wir alle stetigen Lösungen des Funktionalgleichungssystems (1) und (2) ohne der Bedingung (3) angeben. Unsere Methode ist von der Beweisidee von C. Baiocchi grundsätzlich verschieden. Wir werden einen für die

Eingegangen den 2.1.68

Informationstheorie auch sonst wichtigen Satz von P. ERDÖS ([4], vgl. auch [8], [9], [2] und [1]) verwenden.

§ 1. Hilfssätze

Wir beginnen mit folgender

DEFINITION 1. $\mathfrak{M}$ ist die Menge aller Funktionen $f: R \to R$, für die die Funktionalgleichungen

$$f[x+f(y)]+f(y)=f[y+f(x)]+f(x) \quad (x, y \in R) \tag{1}$$

$$f(-x)=f(x)+x \quad (x \in R) \tag{2}$$

erfüllt sind.

HILFSSATZ 1. *Ist $f \in \mathfrak{M}$, so ist die durch*

$$x \circ y = y - f(x-y) \quad (x, y \in R) \tag{6}$$

definierte Operation kommutativ und assoziativ.

Beweis. Aus (2) folgt

$$x \circ y = y - f(x-y) = y - f(y-x) - (y-x) = x - f(y-x) = y \circ x,$$

damit haben wir die Kommutativität bewiesen. Von den Gleichungen (1), (2) und von der Kommutativität folgt

$$\begin{aligned}(x \circ y) \circ z &= z - f(x \circ y - z) = z - f[y - z - f(x-y)] \\ &= z - f[z - y + f(x-y)] - [z - y + f(x-y)] = y - f[x - y + f(z-y)] \\ &\quad - f(z-y) = z \circ y - f(x - z \circ y) = x \circ (z \circ y) = x \circ (y \circ z).\end{aligned}$$

Damit ist der Beweis des Hilfssatzes vollendet.

DEFINITION 2. Es sei $f \in \mathfrak{M}$. Die zur Funktion f gehörende Folge $\{\varphi_n\}$ $(n=1, 2, \ldots)$ definieren wir folgendermassen:

$$\varphi_1 = 0, \quad \varphi_2 = -f(\varphi_1), \quad \varphi_3 = -f(\varphi_2), \ldots, \varphi_{n+1} = -f(\varphi_n), \ldots \tag{7}$$

HILFSSATZ 2. *Ist $\{\varphi_n\}$ die zu einer Funktion $f \in \mathfrak{M}$ gehörige Folge, so gilt die Funktionalgleichung*

$$\varphi_{n+m} = \varphi_n \circ \varphi_m \quad (n, m = 1, 2, \ldots). \tag{8}$$

Beweis. Wegen (6) gilt die Gleichung

$$\varphi_1 \circ \varphi_1 = \varphi_1 - f(\varphi_1 - \varphi_1) = -f(0) = \varphi_2,$$

d.h., es gilt die Identität

$$\varphi_1 \circ \varphi_n = \varphi_{1+n} \tag{9}$$

im Falle $n=1$. Setzen wir jetzt voraus, dass die Gleichung (9) für eine $n \geqq 1$ gilt; so erhalten wir wegen des Hilfssatzes 1

$$\varphi_1 \circ \varphi_{n+1} = \varphi_{n+1} \circ \varphi_1 = \varphi_1 - f(\varphi_{n+1} - \varphi_1) = -f(\varphi_{n+1}) = \varphi_{n+2},$$

damit haben wir die Identität (9) für alle n bewiesen. Es sei jetzt n fix, dann bedeutet (9), dass

$$\varphi_k \circ \varphi_n = \varphi_{k+n} \tag{10}$$

im Falle $k=1$ gilt. Aus (10) und (9) folgt falls (10) für ein $k \geqslant 1$ gilt,

$$\varphi_{k+1} \circ \varphi_n = (\varphi_1 \circ \varphi_k) \circ \varphi_n = \varphi_1 \circ (\varphi_k \circ \varphi_n) = \varphi_1 \circ \varphi_{k+n} = \varphi_{k+n+1},$$

und damit ist der Beweis vollendet.

HILFSSATZ 3. *Ist $\{\varphi_n\}$ die zu einer $f \in \mathfrak{M}$ gehörige Folge, so gilt die Funktionalgleichung*

$$\varphi_{nm} = \varphi_n + \varphi_m \quad (n, m = 1, 2, \ldots). \tag{11}$$

Beweis. Nach dem Hilssatz 2 gilt

$$\varphi_{n+m} = \varphi_n \circ \varphi_m = \varphi_m - f(\varphi_n - \varphi_m). \tag{12}$$

Setzen wir jetzt $n=m$ in (12), so erhalten wir

$$\varphi_{2n} = \varphi_n - f(0) = \varphi_n + \varphi_2,$$

d.h., (11) gilt im Falle $m=2$. Nehmen wir an, dass $\varphi_{kn} = \varphi_k + \varphi_n$ gilt, so erhalten wir aus dem Hilfssatz 2

$$\varphi_{(k+1)n} = \varphi_{kn+n} = \varphi_{kn} \circ \varphi_n = \varphi_n - f(\varphi_{kn} - \varphi_n) = \varphi_n - f(\varphi_k) = \varphi_n + \varphi_{k+1},$$

damit haben wir den Hilfssatz 3 bewiesen.

§ 2. Über die stetigen Lösungen von (1)

In diesem § beweisen wir einen Hilfssatz über die stetigen Lösungen der Funktionalgleichung (1).

HILFSSATZ 4. *Ist $f(x)$ eine stetige Lösung der Funktionalgleichung* (1), *so ist $f(x) \geqq 0$ oder $f(x) \leqq 0$ für alle $x \in R$.*

Beweis. Nehmen wir an, dass die Behauptung des Hilfssatzes falsch ist. Dann existieren zwei Zahlen $x_1 \neq x_2$, so dass die Ungleichungen $f(x_1) < 0$ und $f(x_2) > 0$ gelten. Dann gibt es – wegen der Stetigkeit von f – zwischen x_1 und x_2 einen Punkt x_0 derart, dass $f(x_0)$ verschwindet. Setzen wir in (1) $x = x_0$, so ergibt sich

$$f[x_0 + f(y)] = 0 \tag{13}$$

für alle $y \in R$. Wegen der Stetigkeit von f nimmt $f(y)$ alle Werte aus dem Intervall $[f(x_1), f(x_2)]$ an, d.h. aus (13) folgt

$$f(x) = 0 \quad \text{falls} \quad x \in [x_0 + f(x_1), x_0 + f(x_2)]. \tag{14}$$

Mit vollständiger Induktion werden wir zeigen, dass die Gleichung

$$f(x)=0 \quad \text{falls} \quad x\in[x_0+nf(x_1),\, x_0+nf(x_2)] \tag{15}$$

für alle natürlichen Zahlen n gilt. Für diesen Zweck setzen wir voraus, dass (15) für eine fixgehaltene n erfüllt ist; so erhalten wir aus (13)

$$f[x+f(y)]=0 \quad \text{falls} \quad x\in[x_0+nf(x_1),\, x_0+nf(x_2)]$$

für alle $y\in R$. Wir wissen schon, dass $f(y)$ alle Werte aus $[f(x_1), f(x_2)]$ annimt, daher gilt

$$f(x)=0 \quad \text{falls} \quad x\in[x_0+(n+1)\,f(x_1),\, x_0+(n+1)\,f(x_2)].$$

Im Falle $n\to\infty$ gilt $[x_0+nf(x_1)]\to\infty$ (wegen der Ungleichung $f(x_1)<0$) und $[x_0+nf(x_2)]\to+\infty$ (wegen der Ungleichung $f(x_2)>0$), d.h., aus (15) folgt

$$f(x)=0$$

für alle $x\in R$. Dies ist aber ein Wiederspruch, d.h. $f(x)\geqq 0$ oder $f(x)\leqq 0$ für alle $x\in R$ ist.

§ 3. Das Hauptergebnis

In diesem § wollen wir das folgende Hauptergebnis beweisen.

SATZ. *Die allgemeinsten stetigen Lösungen des Funktionalgleichungssystems* (1) *und* (2) *sind die Funktionen*

$$f(x)=-\frac{1}{A}\ln\left(1+e^{Ax}\right), \tag{16}$$

und

$$f(x)=\pm\tfrac{1}{2}\{|x|\mp x\}, \tag{17}$$

wobei $A\neq 0$ ein konstanter Wert ist. (Die Lösungen (17) kann man aus (16) mit dem Grenzübergang $A\to\pm\infty$ erhalten).

Beweis. Nach dem Hilfssatz 4. ist $f(x)\geqq 0$ oder $f(x)\leqq 0$ für alle $x\in R$. *Erstens sei* $f(x)\geqq 0$. Dann gilt wegen der Gleichung (2)

$$f(x)+x=f(-x)\geqq 0 \quad \text{für} \quad x\in R,$$

d.h., für alle $x\in R$ gilt die Ungleichung

$$-f(x)\leqq x. \tag{18}$$

Aus (18) und aus der Definition der Folge $\{\varphi_n\}$ folgt

$$\varphi_n\geqq -f(\varphi_n)=\varphi_{n+1} \quad (n=1,2,\ldots) \tag{19}$$

d.h., die Folge $\{\varphi_n\}$ ist monoton abnehmend.

Nach dem bekannten Satz von P. ERDÖS [4] (Siehe noch [8], [9]) gilt die Darstellung

$$\varphi_n = a \ln n \qquad (n = 1, 2, \ldots), \tag{20}$$

wenn die Folge $\{\varphi_n\}$ die Eigenschaften (11) und (19) hat. Wegen der Ungleichung (19) ist der konstante Wert $a \leqq 0$. Jetzt wollen wir zwei Fälle unterscheiden.

1) *Es sei* $f(0) \neq 0$. Dann ist $\varphi_2 = -f(0) \neq 0$, d.h. in (20) gilt $a = (1/A) < 0$. Mit der Berücksichtigung der Gleichungen (12) und (20) erhalten wir $f((1/A) \ln n - (1/A) \ln m) = (1/A) \ln m - (1/A) \ln(n+m)$ für alle natürliche Zahlen n und m. Aus dieser Gleichung folgt

$$f\left(\frac{1}{A} \ln \frac{n}{m}\right) = -\frac{1}{A} \ln\left(1 + \frac{n}{m}\right),$$

d.h., es gilt

$$f\left(\frac{1}{A} \ln r\right) = -\frac{1}{A} \ln(1 + r)$$

für alle positiven rationalen Zahlen $r = n/m$. Aus der Stetigkeit von $f(x)$ folgt dann

$$f\left(\frac{1}{A} \ln t\right) = -\frac{1}{A} \ln(1 + t)$$

für alle $t > 0$. Es sei jetzt $x = (1/A) \ln t (x \in (-\infty, \infty)$ falls $t \in (0, \infty))$, so gilt

$$f(x) = -\frac{1}{A} \ln(1 + e^{Ax})$$

für alle $x \in R$. Damit haben wir die Lösungen (16) mit $A < 0$ erhalten.

2) *Es sei* $f(0) = 0$. Dann ist, wegen (7), $\varphi_n = 0$ $(n = 1, 2, \ldots)$. Aus der Ungleichung (18) folgt

$$\lim_{x \to -\infty} f(x) = \infty,$$

d.h., f nimmt alle Werte aus dem Intervall $[0, \infty)$ an. Setzen wir $y = 0$ in (1), so erhalten wir wegen $f(0) = 0$

$$f[f(x)] = 0,$$

woraus

$$f(t) = 0 \qquad \text{falls} \qquad t = f(x) \in [0, \infty) \tag{21}$$

folgt. Aus der Gleichung (2) erhalten wir

$$f(t) = -t \qquad \text{falls} \qquad t < 0. \tag{22}$$

Aus (21) und (22) folgt die Lösung

$$f(x) = \tfrac{1}{2}\{|x| - x\}$$

für alle $x \in R$. Diese ist eine der Lösungen (17).

Zweitens sei $f(x) \leqq 0$. In diesem Falle erhalten wir analogerweise die Lösungen (16) mit $A>0$ und die zweite der Lösungen (17). Damit haben wir den Satz restlos bewiesen.

Unser Satz enthält den von BAIOCCHI [3] als Korollar (unter den stetigen Lösungen (16) und (17) von (1) und (2), wird (3) nur durch (4) und (5) erfüllt).

§ 4. Probleme

PROBLEM 1. Im § 1. haben wir keine Stetigkeitsvoraussetzung über die Funktionen $f \in \mathfrak{M}$ ausgenützt. Dies führt auf das Problem der allgemeinen Lösung des Funktionalgleichungssystems (1) und (2).

PROBLEM 2. Im § 2. haben wir ein Ergebnis über die stetigen Lösungen der Funktionalgleichung (1) bewiesen. Nun fragen wir: Welche Funktionen sind die allgemeinen stetigen Lösungen der Funktionalgleichung (1)? Unsere Vermutung ist folgende: Die alle stetigen Lösungen der Funktionalgleichung (1) sind $f(x)=g(x+a)$ oder $f(x)\equiv a$, wobei $g(x)$ eine der Funktionen (16) oder (17) und $a \in R$ ein konstanter Wert ist.

LITERATURVERZEICHNIS

[1] ACZÉL, J., *Lectures on Functional Equations and Their Applications* (Academic Press, New York–London 1966).

[2] ACZÉL, J., und DARÓCZY, Z., *Charakterisierung der Entropien positiver Ordnung und der Shannonschen Entropie*, Acta Math. Acad. Sci. Hungar. *14*, 95–121 (1963).

[3] BAIOCCHI, C., *Su un sistema di equazioni funzionali connesso alla teoria dell'informazione*, Boll. Un. Mat. Ital. *22*, 236–246 (1967).

[4] ERDÖS, P., *On the Distribution Function of Additive Functions*, Ann. of Math. *47*, 1–20 (1946).

[5] KAMPÉ DE FÉRIET, J. und FORTE, B., *Information et probabilité*, I, C. R. Acad. Sci. Paris *265 A*, 110–114 (1967).

[6] KAMPÉ DE FÉRIET, J. und FORTE, B., *Information et probabilité*, II, C. R. Acad. Sci. Paris *265 A*, 142–146 (1967).

[7] KAMPÉ DE FÉRIET, J. et FORTE, B., *Information et probabilité*, III, C. R. Acad. Sci. Paris *265 A*, 350–353 (1967).

[8] RÉNYI, A., *On a Theorem of P. Erdös and Its Application in Information Theory*, Mathematica Cluj *1*, 341–344 (1959).

[9] RÉNYI, A., *On Measures of Entropy and Information*, in *Proc. Fourth Berkeley Symposium on Statistics and Probability 1960*, Vol. 1 (Berkeley, Calif. 1961), pp. 547–561.

L. Kossuth University, Debrecen

Some Graphical Properties of Matrices with Non-Negative Entries

A. L. DULMAGE and N. S. MENDELSOHN (Winnipeg, Manitoba, Canada)

To Alexander M. Ostrowski On His Seventy-Fifth Birthday

1. Introduction

The parameters S and M were introduced in (1), (3) and (4) in connection with the stochastic rank and the term rank of a matrix. The results of this paper involve these same parameters.

Two of these results are the following. Let A be a matrix of non-negative integers with sum S and maximum row or column sum M and let $\sigma=[S/M]$ (the greatest integer in S/M). It is shown in theorem 1 that A is expressible as a sum of sub-permutation matrices of rank σ and a sub-permutation matrix of rank $q, 0\leqslant q<\sigma$. Further, let K be a bipartite graph with finite vertex sets X and Y. Let S be the total number of edges and let M be the maximum number of edges which have the same vertex. It is shown as a consequence of theorem 3 that K has a transversal G of order $\sigma=[S/M]$ with vertex sets U and V such that every vertex of X which has M edges is an element of U and every vertex of Y which has M edges is an element of V.

Theorem 1 may be proved as a corollary of theorem 3 but the connection of theorem 1 with the doubly stochastic extension of a matrix is interesting. It is on this connection that the proof of theorem 1 in section 3 is based.

2. Definitions and Notation

A *bipartite graph* K has two vertex sets X and Y and a set of edges (which may be null) each of which is a pair of vertices $(x, y), x\in X, y\in Y$. A *transversal* of K is a subgraph G every vertex of which has exactly one edge. The *order* of a transversal is the number of its edges. If A is an $m\times n$ matrix of non-negative elements, *the bipartite graph K of A* (denoted by K_A) is the bipartite graph for which the vertex sets X and Y are the sets of rows and columns of A. The pair (x, y) is an edge of K_A if and only if the element in row x and column y of A is greater than zero. An $m\times n$ *sub-permutation matrix* of rank r is an $m\times n$ matrix of $mn-r$ zeros and r ones with at most one non-zero in any row or column. The bipartite graph of a sub-permutation matrix of rank r is a transversal of order r, together with $m+n-2r$ free vertices.

Let X and Y be arbitrary sets and let $f(x, y), x\in X, y\in Y$, be a function with

Received January 10, 1968

domain $X \times Y$ and range some subset of the non-negative reals. The bipartite graph K of the function f (denoted by K_f) has vertex sets X and Y and (x, y) is an edge of K if and only if $f(x, y) > 0$. A *sub-permutation function* g of rank r on $X \times Y$ is a function such that $g(x, y) = 0$ for all pairs (x, y), $x \in X$, $y \in Y$ with the exception of a set of r pairs

$$(x_1, y_1), (x_2, y_2), \dots (x_r, y_r), x_i \neq x_j, y_i \neq y_j \quad \text{for} \quad i \neq j$$

for which $g(x, y) = 1$.

We say that a transversal G of a bipartite graph K *catches* a vertex if this vertex is a vertex of G, and that G catches a set U of vertices if every vertex of U is a vertex of G.

If W is a finite set, $|W|$ denotes the order of W, that is, the number of elements in W.

The bipartite graph $X \times Y$ is the graph with vertex sets X and Y such that every pair (x, y), $x \in X$, $y \in Y$, is an edge of the graph. If K is a bipartite graph with vertex set X and Y, and if $U \subset X$ and $V \subset Y$ then $(U \times V) \cap K$ denotes the bipartite graph with vertex sets U and V every edge of which is an edge (u, v) of K such that $u \in U$, $v \in V$.

If K is a bipartite graph with vertex sets X and Y, then K has a unique subgraph which has the same edges as K and is such that every vertex of the subgraph has at least one edge. This subgraph is called the *essential* subgraph of K and is denoted by K'.

Reference will be made to irreducible and semi-irreducible graphs and to the canonical decomposition of a bipartite graph as defined in (2), (5) and (6).

3. The Decomposition of a Matrix of Non-Negative Integers

Consider an $m \times n$ matrix A $(n \geqslant m)$ of non-negative reals. Let R_i be the sum of the elements in the ith row and C_j the sum of the elements in the jth column. Thus

$$M = \max(R_i, C_j) \qquad i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n,$$

and

$$S = \sum_{i=1}^{m} R_i = \sum_{j=1}^{n} C_j.$$

This $m \times n$ matrix can be augmented to an $n \times n$ matrix B of non-negative reals with the same S and M, by adding $n - m$ rows of zeros. The matrix B has all the row and column sums of A and in addition has $R_i = 0$ for $i = m+1, \dots, n$. Let $\sigma = [S/M]$. As shown in (4), the matrix B may be extended to a matrix C with row and column sums S/σ by the addition of $n - \sigma$ rows and $n - \sigma$ columns provided $n > \sigma$. Such an extension is called a $(n - \sigma, S/\sigma)$ doubly stochastic extension. Let $B = (b_{ij})$. The matrix $C = (c_{ij})$ is defined as follows:

$$c_{ij} = b_{ij} \quad \text{for} \quad i \leqslant n, j \leqslant n,$$

$$c_{ij} = \frac{\frac{S}{\sigma} - R_i}{n - \sigma} \quad \text{for} \quad i \leqslant n, n + 1 \leqslant j \leqslant 2n - \sigma,$$

$$c_{ij} = \frac{\frac{S}{\sigma} - C_j}{n - \sigma} \quad \text{for} \quad n + 1 \leqslant i \leqslant 2n - \sigma, j \leqslant n,$$

$$c_{ij} = 0 \quad \text{for} \quad n + 1 \leqslant i \leqslant 2n - \sigma, n + 1 \leqslant j \leqslant 2n - \sigma.$$

Since C has equal row and column sums it is expressible in the form $\sum k_i P_i$ where the k_i are >0 and the P_i are permutation matrices (of order $2n-\sigma$). Thus the bipartite graph K_C of the matrix C has the property that for every edge e of K_C there exists a transversal G of order $2n-\sigma$ of which e is an edge. Let K_A be the bipartite graph of A. The transversal G has exactly $2(n-\sigma)$ edges which are not edges of K_A and hence has exactly $2n-\sigma-2(n-\sigma)=\sigma$ edges which are edges of K_A.

We are now in a position to prove our first theorem.

THEOREM 1. *Let A be an $m \times n$ matrix of non-negative integers with sum S and maximum row or column sum M and let $\sigma = [S/M]$. Let t be an integer, $1 \leqslant t \leqslant \sigma$, and let p and q be the unique integers such that $p \geqslant 0$, $0 \leqslant q < t$ and $S = (M+p)\,t + q$. Then A is expressible as a sum of $(M+p)$ $m \times n$ sub-permutation matrices of rank t and* (if $q > 0$) *one $m \times n$ sub-permutation matrix of rank q.*

Proof: Since $t \leqslant \sigma \leqslant S/M$, we have $S \geqslant Mt$ and hence the integers p and q always exist.

There are two cases to consider.

CASE 1: $p=q=0$. Since $S=Mt$, we have $t=\sigma$. Let G be a transversal of K_C of order $2n-\sigma$ and let H be the intersection of G and K_A. Let X and Y be the vertex sets of K_A and let I and J be the maximum subsets of X and Y respectively such that $R_x = M$, $x \in I$, and $C_y = M$, $y \in J$. Let U and V be the vertex sets of H. Since $S/\sigma - M = 0$, no edge of K_C joins a vertex of I or J to any vertex not in K_A and we have $I \subset U$ and $J \subset V$. In other words, the transversal H catches the vertex sets I and J. Let P_1 be the $m \times m$ sub-permutation matrix of which H is the bipartite graph. The rank of P_1 is σ. Let $A - P_1 = A_1$. A_1 is an $m \times n$ matrix of non-negative integers of sum $S_1 = S - \sigma$ with maximum row or column sum $M_1 = M - 1$. Also

$$\frac{S_1}{M_1} = \frac{S - \frac{S}{M}}{M - 1} = \frac{S}{M}.$$

Thus

$$\sigma_1 = \frac{S_1}{M_1} = \sigma .$$

Let C_1 be the matrix with $n-2\sigma_1$ rows and columns which is the $(n-\sigma_1, M_1)$ doubly stochastic extension (4) of A_1. Let G_1 be a transversal of K_{C_1} of order $2n-\sigma_1$, let H_1 be the intersection of G_1 and K_{A_1} and let P_2 be the $m\times n$ sub-permutation matrix of which H_1 is the bipartite graph. Let $A_1 - P_2 = A_2$. The process is repeated until $A_M = 0$. Thus $A = P_1 + P_2 + \cdots + P_M$. A is the sum of M $m\times n$ sub-permutation matrices of rank t.

CASE 2: At least one of p and q is positive. Since $S/M = t + (pt+q)/M$, we have $t < S/M$.

If S/M is an integer, we have $t<\sigma$. Let G_1 be a transversal of K_C of order $2n-\sigma$ and let $H_1 = G_1 \cap K_A$. Let I and J be defined as above. Since H_1 catches I and J, there exists at least one edge $e_1 = (x_1, y_1)$ of H_1 such that $x_1 \in I$ or $y_1 \in J$. Let L_1 be the transversal of order $\sigma - 1$ which results from removing e_1 from H_1.

If S/M is not an integer, we have $t \leqslant \sigma$ and $S/\sigma > M$. Since $S/\sigma - M$ is positive, K_C has at least one edge $e_2 = (x_2, y_2)$ such that $x_2 \in I$ and $y_2 \notin K_A$ or $x_2 \notin K_A$ and $y_2 \in J$. Let G_2 be a transversal of K_C of order $2n-\sigma$ of which e_2 is an edge and let $L_2 = G_2 \cap K_A$. The transversal L_2 has order σ and I or J has at least one vertex which is not a vertex of L_2.

If $p>0$ let N_1 be a transversal of order t which is a subgraph of L_1 or L_2 (according as S/M is or is not an integer), let Q_1 be the sub-permutation matrix of which N_1 is the graph and let $A - Q_1 = A_1$. For the matrix A_1, we have $S_1 = S - t$ and $M_1 = M$ so that $S_1 = (M_1 + p - 1)\, t + q$. If $p > 1$, we have $t < S_1/M_1$ and the process may be repeated. Eventually we find $A - Q_1 - Q_2 - \cdots - Q_p = A_p$. For the matrix A_p we have $S_p = S - pt$ and $M_p = M$ so that $S_p = Mt + q$.

If $q>0$ we repeat the process except that we let N be a transversal of order q which is a subgraph of L_1 or L_2 and let Q be the sub-permutation matrix of which N is the bipartite graph. The matrix $A_{p+1} = A - Q_1 - Q_2 - \cdots - Q_p - Q$ has $S_{p+1} = S - pt - q$ and $M_{p+1} = M$ so that $S_{p+1} = Mt$.

The matrix A_{p+1} is a matrix of the type discussed in case 1, and hence is expressible as $A_{p+1} = P_1 + P_2 + \cdots + P_M$ where the P_i are sub permutation matrices of rank t.

Thus

$$A = \sum_{i=1}^{m} P_i + \sum_{i=1}^{p} Q_i + Q$$

as required.

COROLLARY 1: *Let K be a bipartite graph with S edges and let M be the maximum number of edges at any vertex. Let $\sigma = [S/M]$. Let t be an integer, $1 \leqslant t \leqslant \sigma$, and let p and*

q be the unique integers such that $p \geqslant 0$, $0 \leqslant q < t$ and $S = (M+p)\,t + q$. Then K can be expressed as the union of $M+p$ transversals of order t and (if $q > 0$) *a transversal of order q. No two of these transversals have an edge in common.*

If S/M is an integer it has been shown in proving theorem 1 that there exists a transversal of order σ which catches all the rows and columns of maximum sum. (The proof given is valid also for a matrix of non-negative reals.) However, if S/M is not an integer, the following example shows that it is not possible to show the existence of such a catching transversal simply by first reducing S to $S_1 = \sigma M$, keeping M fixed. Let A be the 3×3 matrix with $a_{31} = 0$, $a_{32} = 0$, and the other elements equal to 1. We have $M = 3$, $S = 7$, $\sigma = 2$. In whatever way we reduce S to $S_1 = 6$ keeping $M_1 = M = 3$, we must reduce the sum for at least one of row 1, row 2, and column 3. For the matrix A_1 the ratio S_1/M_1 is an integer and, by the above result, there exists a catching transversal of order 2. This transversal catches the rows and columns of A_1 with sum M_1, but may not catch all the rows and columns of A with sum M. However the catching transversal of order σ does exist when S/M is not integral, as we shall see on proving theorem 3 of the next section.

4. The Existence of a Transversal which Catches the Rows and Columns which Have Sums Close to M

We begin with a theorem on the existence of certain transversals of a graph of which one of the vertex sets may be countably infinite.

THEOREM 2. *Let X be a finite set of order n and Y a finite or countably infinite set. Let f be a function with domain $X \times Y$ and range some subset of the non-negative reals such that $R_x = \sum_{y \in Y} f(x, y)$ is finite for all $x \in X$ and let $\sum_{x \in X} f(x, y) = C_y$. Let $S = \sum_{x \in X} R_x = \sum_{y \in Y} C_y$ and let $M = \operatorname{Max}(R_x, C_y)$, $x \in X$, $y \in Y$. (Since S is finite, there exists $y_0 \in Y$ such that $\operatorname{Sup}_{y \in Y}(C_y) = C_{y_0}$). Let Z be the subset of Y consisting of all y such that $C_y > n/(n+1)\,M$ and let the order of Z be v. Let K_f be the bipartite graph of the function f. Then either $Z = \emptyset$ or the order of Z is less than or equal to n and the subgraph $(X \times Z) \cap K_f$ of K_f has a transversal of order v.*

COROLLARY. *If in addition to the assumptions of 2, we assume that K_f has a transversal of order n, then K_f has a transversal G of order n which catches the vertex set Z.*

Proof: Let $S^* = \sum_{y \in Z} C_y$. We have

$$S^* = \sum_{y \in Z} C_y > v \frac{n}{n+1} M \geqslant \frac{v}{n+1} S.$$

Since $S^* \leqslant S$ we have $v/(n+1) < 1$ so that $v \leqslant n$. (If $C_y > 0$ for at least $n+2$ distinct y's, the condition $C_y > n/(n+1)\,M$ may be replaced by $C_y \geqslant n/(n+1)\,M$). If $(X\times Z)\cap K_f$ has no transversal of order v then by (2) theorem 2 the exterior dimension of this graph is less than v. Let (A, B), $A\subset X$, $B\subset Z$, be a minimal exterior pair [(2), (5)] for $(X\times Z)\cap K_t$. We have

$$v > |A| + |B| = v - k,\ k > 0.$$

Also,

$$(v-k)\,M \geqslant S^* > v\,\frac{n}{n+1}\,M \geqslant v\,\frac{v}{v+1}\,M$$

so that

$$v - k > \frac{v^2}{v+1},$$

which gives a contradiction.

Proof of Corollary. Let P and Q be the vertex sets of a transversal H of a bipartite graph K of finite exterior dimension. If the order of H is less than the maximal order ϱ, then the first algorithm in (6) replaces H by a transversal G which is maximal and catches the vertex sets P and Q of H. Now let H be the transversal of order v of $(X\times Z)\cap K_f$. If $v=n$ this transversal is satisfactory since it catches Z. If $v<n$, then the algorithm in (6) provides us with the required transversal G.

Alternatively, theorem 1 of (1) may be stated as follows. Let K be a bipartite graph with vertex sets X and Y and let H and L be any two transversals of K. (These transversals need not be maximal, in fact, the graph need not have a maximal transversal). Let U be the first vertex set of H and let V be the second vertex set of L. Thus H is a maximal transversal of $(U\times Y)\cap K$ and L is a maximal transversal of $(X\times V)\cap K$. Then there exists a transversal G the edges of which are edges of H or L, which catches the vertex sets U and V. The corollary follows at once from this theorem. We are now in a position to prove the main theorem.

THEOREM 3. *Let X and Y be finite or countably infinite sets and let f be a function with domain $X\times Y$ and range some subset of the non-negative reals such that the bipartite graph K_f has a maximal transversal, and $S=\sum_{x\in X}\sum_{y\in Y} f(x,y)$ is finite. Let ϱ be the order of the maximal transversal. Let*

$$R_x = \sum_{y\in Y} f(x,y)$$

and

$$C_y = \sum_{x\in X} f(x,y)$$

and let

$$M = \max_{x\in X,\, y\in Y} (R_x, C_y).$$

Let $\sigma=[S/M]$ and let

$$\lambda = \max\left[\frac{S}{M(\sigma+1)},\ \frac{\varrho}{\varrho+1}\right].$$

Let I be the subset of X consisting of all x such that $R_x > \lambda M$ and let J be the subset of Y consisting of all y such that $C_y > \lambda H$. Then corresponding to every integer t, $\sigma \leqslant t \leqslant \rho$, there exists a transversal G_t of order t which catches the vertex sets I and J.

Proof: The proof is in two parts. First we construct a transversal G_ϱ which catches I and J. Then, corresponding to every transversal G_t, $t > \sigma$, which catches I and J, we construct a transversal G_{t-1} which catches I and J.

Consider any maximal transversal G and the canonical decomposition of K_f. Let X and Y be partitioned as $X = X_1 + X_2 + X_3$ and $Y = Y_1 + Y_2 + Y_3$, so that $(X_1 \times Y_1) \cap K_t$ is the horizontal tail of the decomposition, $(X_2 \times Y_2) \cap K_t$ contains the irreducible subgraphs of the core, and $(X_3 \times Y_3) \cap K_t$ is the vertical tail as indicated in figure 1. (The transversal G is indicated schematically by the x's).

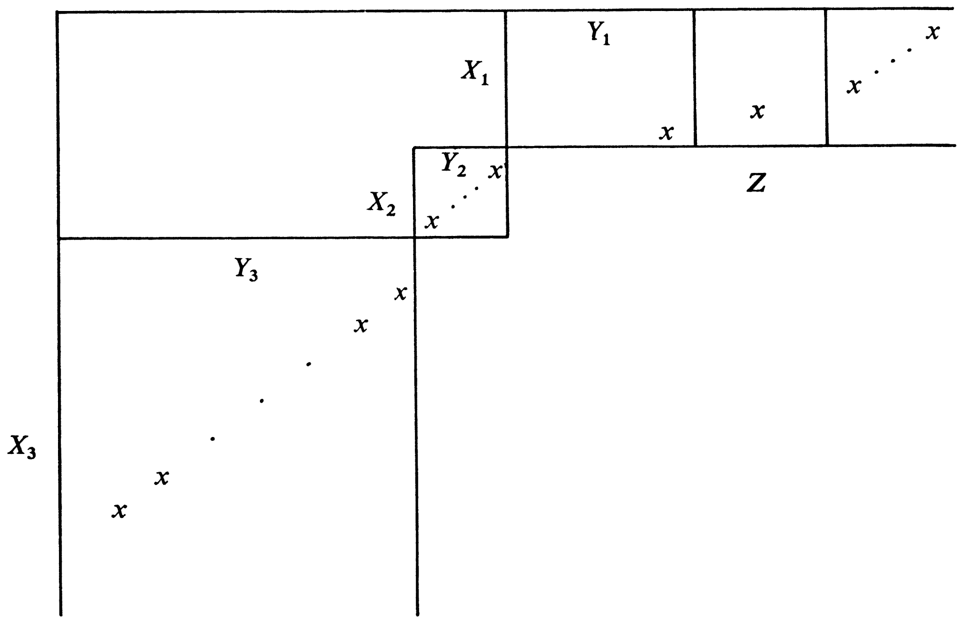


Figure 1.

G catches all the vertices of the sets $I \cap (X_1 + X_2)$ and $J \cap (Y_2 + Y_3)$.
Let

$$|X_1| = n,$$

let

$$M' = \max\left(\sum_{y \in Y_1} f(x, y), \sum_{x \in X_1} f(x, y)\right), \quad x \in X_1, \quad y \in Y_1$$

and let Z be the subset of Y_1 consisting of all y such that $C_y > n/(n+1)\ M'$. By theorem 2 there exists a transversal G' of $(X_1 \times Y_1) \cap K_t$ of order n which catches all the vertices of Z. Since

$$\frac{n}{n+1} M' \leqslant \frac{n}{n+1} M \leqslant \frac{\varrho}{\varrho+1} M \leqslant \lambda M ,$$

we see that G' catches all the vertices of $J \cap Y_1$. Similarly there exists a maximal transversal G'' of $(X_3 \times Y_3) \cap K$ which catches all the vertices of $I \cap X_3$. Let $H = (X_2 \times Y_2) \cap G$. The transversal which is the union of G', H and G'' is the transversal G_ϱ which catches I and J.

Now let G_t, $t > \sigma$, be a transversal of order t which catches I and J. If G_t has an edge e which catches no vertex of I or J, G_{t-1} is found by deleting e. In other cases, the transversal G_t may be decomposed into three disjoint subgraphs as shown in figure 2. Let G_t^1 consist of those edges of G_t which catch a vertex of I but do not catch

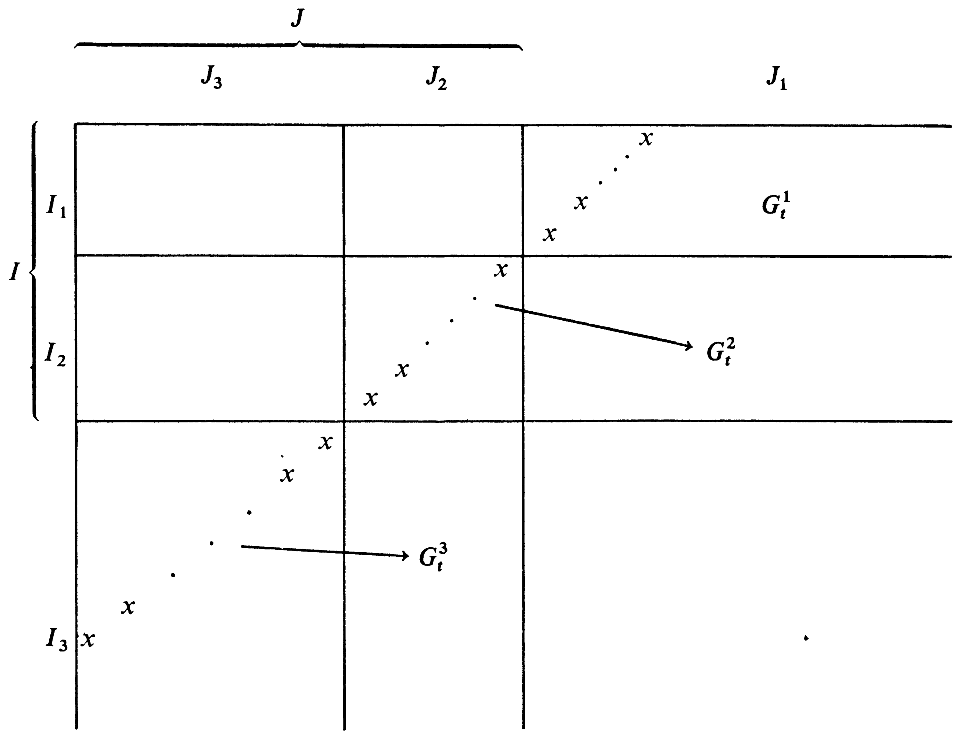


Figure 2.

a vertex of J, Let G_t^2 consist of those edges of G_t which catch a vertex of I and a vertex of J, and let G_t^3 consist of those edges of G_t which catch a vertex of J but do not catch a vertex of I. Thus $G_t = G_t^1 \cup G_t^2 \cup G_t^3$. We let the vertex sets of G_t^i be I_i, J_i, $i=1, 2, 3$. We have

$$|I_1| > 0 \quad \text{for otherwise} \quad |J| = t$$

and

$$\sum_{y \in J} C_y > t\lambda M \geqslant \frac{(\sigma+1)\,S}{M(\sigma+1)} M = S.$$

Similarly, $|J_3| > 0$.

We now construct a finite sequence of subgraphs $L'_1, L'_2, \ldots, L'_w$ of K_f. The subgraph $L_1 = (I_1 \times J) \cap K_f$ has edges, for otherwise,

$$S \geqslant \sum_{x \in I_1} R_x + \sum_{y \in J} C_y > t\lambda M \geqslant S.$$

Let L'_1 be the essential subgraph of L_1 and let A_1, B_1 be its vertex sets. If $B_1 \cap J_3$ is not null the sequence terminates at L'_1 but if $B_1 \subset J_2$ we construct L'_2 as follows. Let C_1 be the first vertex set of that subgraph of G_t of which B_1 is the second vertex set. The subgraph $[I_1 \times (J - B_1)] \cap K_f$ has no edges. The subgraph $L_2 = [C_1 \times (J - B_1)] \cap K_f$ has edges for otherwise we have

$$S \geqslant \sum_{x \in I_1 \cup C_1} R_x + \sum_{y \in J - B_1} C_y > t\lambda M \geqslant S.$$

Let L'_2 be the essential subgraph of L_2 and let A_2, B_2 be its vertex sets. If $B_2 \cap J_3$ is not null the sequence terminates at L'_2 but if $B_2 \subset J_2$ we construct L'_3. Continuing, if L'_k has been constructed with vertex sets A_k, B_k and if $B_k \cap J_3$ is not null the sequence terminates at L'_k but if $B_k \subset J_2$ we construct L'_{k+1} as follows. Let C_k be the first vertex set of that subgraph of G_ϱ of which B_k is the second vertex set. The subgraph

$$\left[\left(\bigcup_{i=1}^{k-1} C_i \cup I_1\right) \times \left(J - \bigcup_{i=1}^{k} B_i\right)\right] \cap K_f$$

has no edges. The subgraph

$$L_k = \left[C_k \times \left(J - \bigcup_{i=1}^{k} B_i\right)\right] \cap K_f$$

has edges, for otherwise we have

$$S \geqslant \sum_{k} R_x + \sum C_y > t\lambda M \geqslant S; \qquad x \in \bigcup_{i=1}^{k} (C_i \cup I_1) \quad y \in J - \bigcup_{i=1}^{k} B_i$$

Let L'_{k+1} be the essential subgraph of L_{k+1}. Since the sets B_i are disjoint and non null, there must exist w such that $B_w \cap J_3 \neq 0$. Thus the sequence terminates at L'_w with $w \leqslant |J_2| + 1$. Let x_1 be a vertex of the set $B_w \cap J_3$ and let (x_1, y_0) be an edge of $L'_w \cap (I \times J_3)$. There exists a unique edge (x, y) of G_t^3 which has the vertex y_0. Let this edge be (x_0, y_0). We form a chain of $2w+1$ edges of which the first and second edges are (x_0, y_0) and (x_1, y_0). Let (x_1, y_1) be the unique edge (x, y) of G_t^2 which has $x=x_1$. We have $y_1 \in A_{w-1}$ and there exists at least one edge (x_2, y_1) of L'_{w-1} and a unique edge (x_2, y_2) of G_t. The chain terminates at an edge (x_w, y_w) of G'_t. Thus the first edge of this chain belongs to G_t^3, the third, fifth, ... $(2w-1)^{th}$ belong to G_t^2 and the $(2w+1)^{th}$ to G_t^1 (if $w=1$ there are no edges of G_t^2). The second, fourth, ...$2w^{th}$ edges belong to $L'_w, L'_{w-1}, \ldots, L'_1$ respectively. Now delete the edges (x_0, y_0), (x_1, y_1), ..., (x_w, y_w) of G_t and add the edges (x_0, y_1), (x_1, y_2), ... (x_{w-1}, y_w). The resulting transversal is G_{t-1}, a transversal of order $t-1$ which catches I and J.

COROLLARY 2. *Let I and J be defined as in theorem 3 with the assumptions of theorem 3. If S/M is not integral, K_f has a transversal of order $\sigma+1$ which catches I and J.*

Proof: We have $\varrho \geqslant S/M > \sigma$, so that $\varrho \geqslant \sigma+1$.

THEOREM 4. *Let a function f, R_x, C_y, S, M and σ be defined as in theorem 2. If $R_x = M$ for $x \in X$, then K_f has a transversal of order n.*

Proof: We have $S = \sum_{x \in X} R_x = nM$.

Thus

$$\varrho \geqslant \frac{S}{M} = n .$$

In theorem 3, if we are content to catch the rows and columns with sum equal to M we need not assume that K_f has a maximal transversal. This remark is made explicit in the next theorem.

THEOREM 5. *Let X and Y be finite or countably infinite sets and let f, S, R_x, C_y, M and G be defined as in theorem 3 with S finite. Let I be the subset of X consisting of all x such that $R_x = M$ and let J be the subset of Y consisting of all Y such that $C_y = M$. Then corresponding to every integer $t \geqslant \sigma$, if K_f has no maximal transversal and to every integer t, $\sigma \leqslant t \leqslant \varrho$ if t has a maximal transversal, there exists a transversal G_t of order t which catches the vertex sets I and J.*

Proof: We may assume K_f has no maximal transversal since otherwise the proof follows by theorem 3. Since S is finite, the sets I and J are finite. By theorem 4 the subgraph $(I \times Y) \cap K_f$ has a transversal H of order $|I|$ and the subgraph $(X \times J) \cap K_f$ has a transversal L of order $|J|$. By theorem 1 of (1) as described in the alternative proof of the corollary to theorem 2 there exists a transversal G_q of order q every edge of which is an edge of H or an edge of L such that G_q catches I and J. Since K_f has no maximal transversal, every transversal can be extended, that is, for every transversal G_t of order t there exists an edge e disjoint from G_t. The union of G_t and the graph with this single edge is a transversal G_{t+1} of order $t+1$. Thus if $q \leqslant \sigma$ the proof is complete. If $\sigma < q$ the proof is complete except for $\sigma \leqslant t < q$. In proving theorem 3 it was shown that if G_t catches I and J and if the order t of G_t is $> \sigma$, there exists a transversal G_{t-1} of order $t-1$ which catches I and J and a similar proof is valid here. This completes the proof of theorem 5.

There is an immediate corollary.

COROLLARY. *Let X and Y be finite or countably infinite sets and let I and J be defined as in theorem 5. If S/M is integral and if K_f has an edge (x_1, y_1) such that x_1 is not in I and y_1 is not in J then K_f has a transversal G of order $\sigma - 1$ which catches I and J.*

Proof: Let f' be the function on $X \times Y$ such that $f' = f$ for $x \neq x_1, y \neq y_1$ and $f'(x_1, y_1) = 0$. For f', we have $S' = S - f(x_1, y_1)$, $M' = M$ and $\sigma' = [S'/M'] = \sigma - 1$. $K_{f'}$ is a proper subgraph of K_f. By theorem 5, $K_{f'}$ has a transversal of order σ' $(= \sigma - 1)$ which catches I and J.

5. Algorithms for Finding the Transversals

There are algorithms for finding the various transversals. The transversal of theorem 2 of $(X \times Z) \cap K_f$ may be found using algorithm 1 of (6). The transversal G of the corollary to theorem 2 may be found as indicated in the proof using algorithm 1 of (6) beginning with the transversal of $(X \times Z) \cap K_f$.

The transversal G_t of theorem 3 may be found as follows. First find a maximal transversal using algorithm 1 of (6) and then find the canonical decomposition using algorithm 2 of (6). Then use theorem 2 to find the transversals G' and G''. Thus G_ϱ is found. Finally in proving theorem 3 we have given an algorithm for finding a transversal G_{t-1} from G_t when $t > \sigma$.

The transversal in theorem 4 is maximal and can be found using algorithm 1 of (6).

In theorem 5 the transversals H and L may be found using algorithm 1 of (6) and the transversal G_q using the algorithm described in theorem 1 (1). Then if $q < t$, G_t may be found from G_q using algorithm 1 of (6) and if $\sigma \leqslant t < q$, G_t may be found using the algorithm described in theorem 3.

The decomposition of a matrix A of non-negative integers can be carried out as described in theorem 1 or alternatively as follows. For any t, $1 \leqslant t \leqslant \sigma$, let $S = (M+p)t + q$ as in theorem 1. If p or q is positive, then $t < S/M$ and if S/M is integral, $t < \sigma \leqslant \varrho$ and if S/M is not integral $t \leqslant \sigma < \varrho$. Thus $t < \varrho$ if p or q is positive. Now find a transversal G_ϱ of order ϱ which catches I and J as described in theorem 5. Let e be an edge of G_ϱ which catches a vertex of I or J. The transversal found by deleting e from G_ϱ can replace the transversal L_1 or L_2 of theorem 1. Eventually we have A_{p+1} with $S_{p+1} = Mt$ and $\sigma_{p+1} = t$. Then we can find a transversal G_t which catches the sets I and J of A_{p+1}. This transversal can be used to find P_1 and the sub-permutation matrices $P_2, \ldots, P_M$ are found in a similar way.

6. A Conjecture Concerning the Convex Polyhedral Cone of Sub-Permutation Functions

An equivalent statement of the theorem in (3) is the following. An $m \times n$ matrix A of non-negative reals with sum S and maximum row or column sum M is expressible as a finite sum $\sum c_i P_i$ where the c_i are positive reals and the P_i are sub-permutation matrices of rank $r$if and only if $1 \leqslant r \leqslant \sigma$. Also, if $A = \sum c_i P_i$, then $\sum c_i = M$ if and only if $r = S/M$, otherwise $\sum c_i > M$. The theorem is proved using a doubly stochastic extension of A.

Now let f be a function satisfying the assumptions of theorem 3. We conjecture that f is expressible as a possibly infinite sum $\sum c_i g_i$ where the c_i are positive reals and the g_i are sub-permutation functions on $X \times Y$ of rank r, if and only if $1 \leqslant r \leqslant \sigma$. The necessity is immediate. The conjecture concerns the sufficiency. Since any sub-permutation function g of rank $t > 1$ is expressible as a finite sum $\sum c_i g_i$ involving positive c_i and sub-permutation functions g_i of rank $t-1$, it is sufficient to consider the case $r = \sigma$.

If S/M is not integral, then by the corollary to theorem 5, K_f has a transversal G of order $\sigma + 1$ which catches I and J. Let $e = (x_0, y_0)$ be an edge of G such that R_{x_0} or C_{y_0} is equal to M. Let G_1 be the transversal obtained from G by deleting the edge e. Let g_1 be the sub-permutation function of which G_1 is the graph. Let c_1 be the minimum of $S/\sigma - M$ and the values of $f(x, y)$ for (x, y) an edge of G_1. Consider $f_1 = f - c_1 g_1$. Either K_f has at least one edge that is not an edge of K_{f_1}, or $S_1/M_1 = (S - c_1 \sigma)/M = \sigma$ so that S_1/M_1 is integral. It is not clear however that in repeating this process, we eventually get S/M integral.

If S/M is integral, consider a transversal G of order σ which catches I and J and let c be the minimum of $M - \lambda M$ (λ defined as in theorem 3) and the values of $f(x, y)$ for (x, y) an edge of G. Let g be the sub-permutation function of which G is the graph and let $f_1 = f - cg$. Either K_f has at least one edge that is not an edge of K_{f_1} or $M_1 = \lambda M$. We have, of course,

$$\frac{S_1}{M_1} = \frac{S - c\sigma}{M - c} = \frac{S}{M}.$$

Since S/M is integral,

$$\frac{S}{M(\sigma + 1)} = \frac{\sigma}{\sigma + 1},$$

and since $\varrho \geqslant \sigma$ we have

$$\lambda = \frac{\varrho}{\varrho + 1}$$

Since K_{f_1} is a subgraph of K_f, the order ϱ_1 of the maximal transversal of K_{f_1} is less than or equal to ϱ, and hence

$$\lambda_1 = \frac{\varrho_1}{\varrho_1 + 1} \leqslant \frac{\varrho}{\varrho + 1} = \lambda.$$

Thus, if the process is repeated, either an edge is removed from K_f or M is reduced by a factor which is less than or equal to λ.

Thus if f is a function satisfying the conditions of theorem 3, then for every positive integer n there exist non-negative integers p and q $(p+q=n)$ and n positive reals c_i and n sub-permutation functions g_i of order σ such that the function

$$f_n = f - \sum_{i=1}^{n} c_i g_i$$

is a function with range some subset of the non-negative reals which has the following properties.

$$\sigma_n = \frac{S_n}{M_n} = \sigma = \frac{S}{M}. \tag{1}$$

K_{f_n} is a subgraph of K_f, and K_f has at least p edges that are not edges of K_{f_n}. (2)

$$M_n \leqslant \lambda^q M. \tag{3}$$

In conclusion note that if f were defined as in theorem 5, then for S/M integral no matter how small the value of C, we might have $M_1 > M - C$ so that $S_1/M_1 < S/M$

REFERENCES

[1] DULMAGE, A. L. and MENDELSOHN, N. S., *Some Generalizations of the Problem of Distinct Representatives*, Canad. J. Math. *10*, 230–241 (1958).

[2] DULMAGE, A. L. and MENDELSOHN, N. S., *Coverings of Bipartite Graphs*, Canad. J. Math. *10*, 517–534 (1958).

[3] DULMAGE, A. L. and MENDELSOHN, N. S., *The Convex Hull of Sub-Permutation Matrices*, Proc. Amer. Math. Soc. *9*, 253–254 (1958).

[4] DULMAGE, A. L. and MENDELSOHN, N. S., *The Term and Stochastic Ranks of a Matrix*, Canad. J. Math. *11*, 269–279 (1959).

[5] DULMAGE, A. L. and MENDELSOHN, N. S., *A Structure Theory of Bipartite Graphs of Finite Exterior Dimension*, Trans. Roy. Soc. Can. [Series 3] *LIII*, Sect. III, 1–13.

[6] DULMAGE, A. L. and MENDELSOHN, N. S., *Two Algorithms for Bipartite Graphs*, J. Soc. Indust. Appl. Math. *11*, 183–194 (1963).

University of Manitoba

Some Results on Roots of Unity, with an Application to a Diophantine Problem

MORRIS NEWMAN (Washington, D.C.)

Dedicated to Prof. A. M. Ostrowski on the occasion of his 75th birthday

Let R be the rational field, and let $R_k = R(\zeta_k)$ be the field of the kth roots of unity, where k is any positive integer and

$$\zeta_k = \exp(2\pi i/k).$$

Let m, n, be arbitrary positive integers, and assume (for definiteness) that

$$m \geqslant n. \tag{1}$$

Put

$$(m, n) = \delta, \quad [m, n] = \Delta, \tag{2}$$

so that $mn = \delta\Delta$. It is known that if $f(k)$ is any multiplicative arithmetic function, then

$$f(m) f(n) = f(\delta) f(\Delta). \tag{3}$$

In particular, (3) holds for $f(k) = \varphi(k)$, the Euler function.

If the fields F and G are each normal finite algebraic extensions of R, then so is $\{F, G\}$ (the smallest subfield of the complex numbers containing both F and G), and

$$(\{F, G\} : F) = (G : F \cap G). \tag{4}$$

(Here if the field A is a finite algebraic extension of the field B, then $(A:B)$ denotes the degree of the extension).

We shall prove

LEMMA 1. *The fields R_m, R_n, satisfy*

$$R_m \cap R_n = R_\delta, \tag{5}$$

$$\{R_m, R_n\} = R_\Delta. \tag{6}$$

LEMMA 2. *If*

$$R(\zeta_m + \zeta_m^{-1}) = R(\zeta_n + \zeta_n^{-1}), \tag{7}$$

then $m = n$ or $m = 2n$.

These lemmas will be applied to the diophantine equation

$$\sin \pi x \sin \pi y = \tfrac{1}{4} \quad (x, y, \text{rational}). \tag{8}$$

Received November 28, 1967

We shall determine all solutions of (8), thus settling a problem posed by R. GRAHAM (see [2]). In fact we shall determine all rationals x and y such that $\sin\pi x \sin\pi y$ is rational.

The method used to prove Lemma 2 can be applied to show that if ζ is a primitive kth root of unity belonging to R_n, then $k=n$ or $k=2n$.

A proof of (5) for the case when m and n are distinct primes has been given by E. JOHNSON in [1].

Proof of Lemma 1. Since $m|\,\Delta$, $n|\,\Delta$, we have $\zeta_m\in R_\Delta$, $\zeta_n\in R_\Delta$, and hence

$$\{R_m, R_n\} \subset R_\Delta .$$

Also $\zeta_m=\zeta_\Delta^{\Delta/m}=\zeta_\Delta^{n/\delta}$, $\zeta_n=\zeta_\Delta^{\Delta/n}=\zeta_\Delta^{m/\delta}$. Since $(m/\delta, n/\delta)=1$, this implies that $\zeta_\Delta\in\{R_m, R_n\}$, and hence that

$$R_\Delta \subset \{R_m, R_n\} .$$

It follows that $\{R_m, R_n\}=R_\Delta$, and so (6) is proved.

Since $\delta|m$, $\delta|n$, we have that $\zeta_\delta\in R_m$, $\zeta_\delta\in R_n$, and hence that

$$R_\delta \subset R_m \cap R_n .$$

Now (4) and (6) imply that

$$(R_\Delta : R_m) = (R_n : R_m \cap R_n),$$

from which we find, by (3), that

$$(R_m \cap R_n : R) = \frac{\varphi(m)\,\varphi(n)}{\varphi(\Delta)} = \varphi(\delta).$$

But $(R_\delta : R)=\varphi(\delta)$, and $R_\delta\subset R_m\cap R_n$. It follows that $R_\delta=R_m\cap R_n$, and so (5) is proved. Thus we have completed the proof of the lemma.

Proof of Lemma 2. Since $\zeta_m+\zeta_m^{-1}$ belongs to both R_m and R_n, it must also belong to their intersection, which is R_δ, by Lemma 1. Since $\zeta_m+\zeta_m^{-1}$ is real, it must in fact belong to $R(\zeta_\delta+\zeta_\delta^{-1})$. It follows that

$$\deg(\zeta_m+\zeta_m^{-1}) \leqslant \deg(\zeta_\delta+\zeta_\delta^{-1}),$$

and similarly that

$$\deg(\zeta_n+\zeta_n^{-1}) \leqslant \deg(\zeta_\delta+\zeta_\delta^{-1}).$$

Hence $\varphi(m)\leqslant\varphi(\delta)$, $\varphi(n)\leqslant\varphi(\delta)$. Now use the fact that $\varphi(ab)\geqslant\varphi(a)\,\varphi(b)$ for all positive integers a, b, to get

$$\varphi(\delta) \geqslant \varphi(m) = \varphi\left(\delta\cdot\frac{m}{\delta}\right) \geqslant \varphi(\delta)\,\varphi\left(\frac{m}{\delta}\right),$$

$$\varphi(\delta) \geqslant \varphi(n) = \varphi\left(\delta\cdot\frac{n}{\delta}\right) \geqslant \varphi(\delta)\,\varphi\left(\frac{n}{\delta}\right).$$

These inequalities imply that

$$\varphi\binom{m}{\delta} \leqslant 1, \quad \varphi\binom{n}{\delta} \leqslant 1;$$

thus m/δ and n/δ can each be only 1 or 2. Since $m \geqslant n$, the possibilities for $m/n = (m/\delta)/(n/\delta)$ are therefore only 1 or 2. This completes the proof of the lemma.

We now go on to the diophantine equation (8). Since $\sin \pi x = \sin \pi (1-x)$, there is no loss of generality in assuming that $0 < x, y \leqslant \frac{1}{2}$. We shall prove

THEOREM 1. *All solutions of the diophantine equation* (8), *normalized so that* $0 < x \leqslant y \leqslant \frac{1}{2}$, *are given by*

$$\begin{array}{c|c|cc} x & \frac{1}{6} & \frac{1}{10} & \frac{1}{12} \\ \hline y & \frac{1}{6} & \frac{3}{10} & \frac{5}{12} \end{array} \tag{9}$$

Proof of Theorem 1. Put

$$x = \frac{a}{m}, \quad (a, m) = 1, \quad 1 \leqslant a \leqslant \frac{m}{2},$$

$$y = \frac{b}{n}, \quad (b, n) = 1, \quad 1 \leqslant b \leqslant \frac{n}{2}.$$

We can assume that $m \geqslant n$. Then (8) becomes

$$(\zeta_{2m}^{a} - \zeta_{2m}^{-a})(\zeta_{2n}^{b} - \zeta_{2n}^{-b}) = -1. \tag{10}$$

Squaring (10), we have that

$$(\zeta_m^a + \zeta_m^{-a} - 2)(\zeta_n^b + \zeta_n^{-b} - 2) = 1. \tag{11}$$

This equation implies that

$$R(\zeta_m^a + \zeta_m^{-a}) = R(\zeta_n^b + \zeta_n^{-b}).$$

Since these fields are normal and $(a, m) = 1$, $(b, n) = 1$, we get that

$$R(\zeta_m + \zeta_m^{-1}) = R(\zeta_m^a + \zeta_m^{-a}),$$

$$R(\zeta_n + \zeta_n^{-1}) = R(\zeta_n^b + \zeta_n^{-b}).$$

Hence

$$R(\zeta_m + \zeta_m^{-1}) = R(\zeta_n + \zeta_n^{-1}).$$

Then Lemma 2 implies that $m = n$ or $m = 2n$; thus

$$\zeta_n = \zeta_m \quad \text{or} \quad \zeta_m^2,$$

and so must belong to R_m.

Determine a' so that $aa' \equiv 1 \bmod m$, and apply the automorphism of R_m

$$\zeta_m \to \zeta_m^{a'},$$

to (11). We get

$$(\zeta_m+\zeta_m^{-1}-2)(\zeta_n^{a'b}+\zeta_n^{-a'b}-2)=1.$$

which implies that

$$\tfrac{1}{4}=\left|\sin\frac{\pi}{m}\right|\cdot\left|\sin\frac{a'b\pi}{n}\right|$$

$$\leqslant\frac{\pi}{m}\cdot 1,$$

so that $m\leqslant 4\pi$. Hence

$$m\leqslant 12. \tag{12}$$

Inequality (12) reduces the solution of (8) to the examination of finitely many cases. We can save ourselves some work by observing that (10) implies that $1-\zeta_m$ is a unit, which can only happen if m is divisible by at least 2 distinct primes. Thus m can have only the values 6, 10, 12. The proof of the theorem is now concluded by a direct examination of all possible cases.

If we notice that $4\sin\pi x\sin\pi y$ is an algebraic integer, we can proceed as above to prove

THEOREM 2. *All rational values of x and y such that* $\sin\pi x\sin\pi y$ *is a positive rational, normalized so that* $0<x\leqslant y\leqslant\frac{1}{2}$, *are given by*

x	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{1}{10}$	$\frac{1}{12}$
y	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	$\frac{1}{6}$	$\frac{3}{10}$	$\frac{5}{12}$

. (13)

REFERENCES

[1] JOHNSON, E. C., Skew-Hadamard abelian group difference sets. *J. Algebra 4*, (1966), 388–402.
[2] *Proc. Conference on Number Theory*, held at the University of Colorado, Boulder, Colorado, 1963.

National Bureau of Standards

Two-Point Boundary-Value Problems and Iteration

Richard Bellman (Los Angeles)

Dedicated to A. Ostrowski on his 75th Birthday

§ 1. Introduction

The connection between the theory of iteration and analytic differential equations subject to initial value problems is classical; cf. [1], [2], where many references may be found. Let us indicate here an analogous connection between iteration of a more complex type and two-point boundary-value problems. The theory of invariant imbedding, [3], [4], applied to two-point boundary-value problems yields the interesting functional equations which provide the linkage.

Consider, to illustrate the general results, the scalar differential equations

$$\left.\begin{aligned} u' &= g(u, v), \quad & u(a) &= c_2, \\ v' &= h(u, v), \quad & v(0) &= c_1, \end{aligned}\right\} \tag{1}$$

where t, the independent variable, ranges over $[0, a]$.

These equations may be associated with a transport process in a one-dimensional rod.

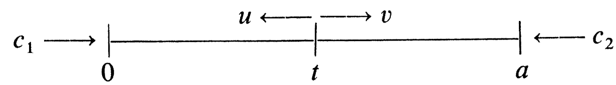


Figure 1

Here the function $u(t)$ denotes the intensity of left-hand flux at an internal point and $v(t)$, that of the right-hand flux. The quantities c_1 and c_2 represent the intensities of left-hand and right-hand incident fluxes.

§ 2. Invariant Imbedding

We consider the equation in (1.1) as a member of a family of problem characterized by the parameters, c_1, c_2, and a. We take $a \geqslant 0$, and initially $-\infty < c_1, c_2 < \infty$. In general, the two-point boundary-value problem will not possess a unique solution for unrestricted values of c_1, c_2, and a. Under simple assumptions on g and h, it is easy to establish the existence of a solution for small a. The functional equation used below is designed *inter alia* to circumvent some analytic and computational difficulties associated with (1.1) for parameter values which escape a direct approach. For a conventional approach, see [5].

Received November 29, 1967.

Introduce the functions

$$\left.\begin{aligned} R(c_1, c_2, a) &= \text{the value of } u(0), \text{ the 'reflected flux'}, \\ T(c_1, c_2, a) &= \text{the value of } v(a), \text{ the 'transmitted flux'}. \end{aligned}\right\} \qquad (1)$$

For the following discussion, we suppose that these functions exist and are uniquely defined for $a \geqslant 0$ and all c_1 and c_2. The physically interesting case is $c_1, c_2 \geqslant 0$. The required existence and uniqueness can often be demonstrated by non-constructive means in some direct fashion by means of a variational principle.

In the isotropic case $R(c_1, c_2, a) = T(c_2, c_1, a)$. In general this symmetry does not hold.

§ 3. Functional Equations

Examining Figure 1, and focussing on the interval $[0, t]$, we obtain the two equations

$$\left.\begin{aligned} R(c_1, c_2, a) &= R(c_1, u(t), t), \\ v(t) &= T(c_1, u(t), t). \end{aligned}\right\} \qquad (1)$$

Similarly, a consideration of the interval $[t, a]$ yields

$$\left.\begin{aligned} T(c_1, c_2, a) &= T(v(t), c_2, a-t), \\ u(t) &= R(v(t), c_2, a-t). \end{aligned}\right\} \qquad (2)$$

Note that $u(t)$ and $v(t)$ depend on a. We have suppressed the dependence.

Limiting versions of these equations as $t \to 0$ and $t \to a$ yield associated partial differential equations, [6]. From the second equation in (1) and (2) we can solve for the 'internal fluxes', u and v, in terms of the reflection and transmission functions R and T. Using these results in the first equations in (1) and (2), we obtain the desired functional equations. For the case where g and h are linear in u and v, see [7].

§ 4. Iteration

Let us introduce the basic functions

$$\left.\begin{aligned} r(c_1, c_2) &= R(c_1, c_2, 1), \\ t(c_1, c_2) &= T(c_1, c_2, 1). \end{aligned}\right\} \qquad (1)$$

From (5.1) and (3.2), we have, for $a = 2$, $t = 1$,

$$\left.\begin{aligned} &\text{(a)} \quad R(c_1, c_2, 2) = r(c_1, u(1)), \\ &\text{(b)} \qquad\qquad v(1) = t(c_1, u(1)), \\ &\text{(c)} \quad T(c_1, c_2, 2) = t(v(1), c_2), \\ &\text{(d)} \qquad\qquad u(1) = r(v(1), c_2). \end{aligned}\right\} \qquad (2)$$

From (b) and (d), we obtain relations of the form

$$\left.\begin{aligned} u(1) &= w_1(c_1, c_2), \\ v(1) &= z_1(c_1, c_2). \end{aligned}\right\} \tag{3}$$

Using (a) and (c), we have

$$\left.\begin{aligned} R(c_1, c_2, 2) &= r(c_1, w_1(c_1, c_2)), \\ T(c_1, c_2, 2) &= t(z_1(c_1, c_2), c_2), \end{aligned}\right\} \tag{4}$$

the desired iterative relation. Consider next the case where $a = N$, an integer. Using $t = 1$, we have from (3.1) and (3.2),

$$\left.\begin{aligned} v(1) &= t(c_1, u(1)), \\ u(1) &= R(v(1), c_2, N-1). \end{aligned}\right\} \tag{5}$$

From these two relations, we can determine $u(1)$ and $v(1)$,

$$\left.\begin{aligned} u(1) &= w_{N-1}(c_1, c_2), \\ v(1) &= z_{N-1}(c_1, c_2). \end{aligned}\right\} \tag{6}$$

Finally, we obtain a relation of the form

$$\left.\begin{aligned} R(c_1, c_2, N) &= r(c_1, w_{N-1}(c_1, c_2)), \\ T(c_1, c_2, N) &= t(z_{N-1}(c_1, c_2), c_2). \end{aligned}\right\} \tag{7}$$

Thus, starting with the functions $r(c_1, c_2)$, $t(c_1, c_2)$ for the unit length, we can generate the reflection and transmission functions for an arbitrary integral length.

§ 5. Constant Right-Hand Incident Flux

The results simplify to a great degree if we assume that the right-hand flux is constant, say $c_2 = 0$. Let

$$f_N(c_1) = \text{the value of } u(0) \text{ for } a = N. \tag{1}$$

Then, referring to Figure 1, we obtain the simpler functional equations

$$\left.\begin{aligned} f_N(c_1) &= r(c_1, u(1)), \\ u(1) &= f_{N-1}(v(1)), \\ v(1) &= t(c_1, u(1)). \end{aligned}\right\} \tag{2}$$

This provides the desired relation between f_N and f_{N-1}, upon eliminating $u(1)$ and $v(1)$ from the three equations.

§ 6. The Analytic Case

A most important case is that where g and h are analytic about $u=v=0$ with the forms

$$\left.\begin{aligned} g(u, v) &= a_1 u + a_2 v + \cdots, \\ h(u, v) &= b_1 u + b_2 v + \cdots. \end{aligned}\right\} \quad (1)$$

In this case, we can set

$$\left.\begin{aligned} f_N(c_1) &= r_{1N} c_1 + r_{2N} c_1^2 + \cdots, \\ u(1) &= u_1 c_1 + u_2 c_1^2 + \cdots, \\ v(1) &= v_1 c_1 + v_2 c_1^2 + \cdots. \end{aligned}\right\} \quad (2)$$

where the coefficients are independent of c_1 and use (5.2) to determine the coefficients r_{1N}, r_{2N} recurrently in terms of the u_i and v_i. These latter are determined by means of the differential equation. The expansions will be valid for $|c_1|$ small under various assumptions concerning a_1, a_2, b_1, and b_2.

The case where $N=\infty$ is particularly simple since (5.2) reduces to

$$\left.\begin{aligned} f(c_1) &= r(c_1, u(1)), \\ u(1) &= f(v(1)), \\ v(1) &= t(c_1, u(1)), \end{aligned}\right\} \quad (3)$$

setting $f(c_1)=f_\infty(c_1)$.

Similar results hold for the functions $R(c_1, c_2, N)$ and $T(c_1, c_2, N)$ and for the multidimensional case.

Subsequently, we will discuss the asymptotic behavior of the foregoing functions as $N\to\infty$ and the use of these methods for numerical solution.

REFERENCES

[1] Bellman, R., *Perturbation Techniques in Mathematics, Physics, and Engineering* (Holt, Rinehart & Winston, New York 1964).

[2] Harris, T.E., *The Theory of Branching Processes* (Springer-Verlag, Berlin 1963).

[3] Bellman, R., Kalaba, R., and Prestrud, M., *Invariant Imbedding and Radiative Transfer in Slabs of Finite Thickness* (American Elsevier Publ. Co., New York 1963).

[4] Bellman, R., Kagiwada, H., Kalaba, R., and Prestrud, M., *Invariant Imbedding and Time-dependent Processes.* (American Elsevier Publ. Co., New York 1964).

[5] Bellman, R. and Kalaba, R., *Quasilinearization and Nonlinear Boundary Value Problems* (American Elsevier Publ. Co., New York 1965).

[6] Bellman, R., Kalaba, R., and Wing, G.M., *Invariant Imbedding and the Reduction of Two-point Boundary Value Problems to Initial-Value Problems,* Proc. Nat. Acad. Sci. USA *46*, 1646–1649 (1960).

[7] Bellman, R., *Scattering Processes and Invariant Imbedding* (to appear).

University of Southern California

An Application of Minimal Solutions of Three-Term Recurrences to Coulomb Wave Functions

WALTER GAUTSCHI (La Fayette, Indiana)[1])
Dedicated to Professor ALEXANDER OSTROWSKI on his 75th birthday

1. In a recent article [3] we dealt with various recurrence algorithms for computing minimal solutions of linear second-order difference equations–solutions, that is, which grow more slowly than any other linearly independent solution. Such solutions (if they exist) are uniquely determined up to a multiplicative constant. The value of this constant may be determined by specifying *one* initial value, or, more generally, by specifying the value of an infinite series in this solution. In the latter case, it is possible to obtain the respective solution without reference to any initial data.

Among several examples we considered in particular the regular Coulomb wave functions $F_L(\eta,\varrho)$ (see [1] for notations). If we let

$$f_L = \frac{2^L L!}{(2L)!\,C_L(\eta)} F_L(\eta,\varrho), \qquad C_L(\eta) = \frac{2^L e^{-\pi\eta/2}|\Gamma(L+1+i\eta)|}{(2L+1)!}, \tag{1.1}$$

then f_L is a minimal solution of

$$\frac{L[(L+1)^2+\eta^2]}{(L+1)(2L+3)} y_{L+1} - \left[\eta + \frac{L(L+1)}{\varrho}\right] y_L + \frac{L(L+1)}{2L-1} y_{L-1} = 0 \qquad (L=1,2,3,\ldots), \tag{1.2}$$

and we have the following infinite series relation [3],

$$\sum_{L=0}^{\infty} \lambda_L f_L = \varrho\, e^{\omega\varrho}, \qquad \lambda_L = i^L P_L^{(i\eta,-i\eta)}(-i\omega). \tag{1.3}$$

Here, $P_n^{(\alpha,\beta)}(x)$ denotes the Jacobi polynomial of degree n. The parameters η,ϱ,ω are assumed to be real, with $\varrho>0$, $\omega\geqslant 0$. Provided ω is chosen appropriately, the algorithms mentioned above lead to effective schemes of computing f_L over an extended range of the parameters η, ϱ, and for as many values of L as are desired [3, § 7], [4]. An advantage of this approach is the absence of any need to compute $F_0(\eta,\varrho)$, which is known to be tedious, calling for a variety of methods in different regions of the parameters [2].

As one proceeds to large values of η, however, the generation of the coefficients λ_L becomes subject to serious loss of accuracy due to cancellation errors. The reason

[1]) Computer Sciences Department, Purdue University, Lafayette, Indiana, and Argonne National Laboratory, Argonne, Illinois. This work was performed in part under the auspices of the United States Atomic Energy Commission.

Received November 29, 1967.

for this can be traced to a peculiar phenomenon associated with the recurrence relation for the Jacobi polynomials of purely imaginary parameters and variable. The phenomenon, already observed in [3], but left unexplained there, is briefly described in section 2. In section 3 we further elucidate this phenomenon and, at the same time, provide a simple scheme to eliminate the cancellation problem which it causes. The algorithm that so results proves to be effective for an almost unlimited region of the parameters η, ϱ, and L. The only factor restricting its use on a digital computer appears to be the possible occurrence of 'overflow' when $|\eta|$ is very large. These matters will be discussed in section 4.

2. From the well-known recurrence relation for Jacobi polynomials, one finds that λ_L satisfies

$$\lambda_{L+1} = \frac{2L+1}{L+1}\,\omega\,\lambda_L + \frac{L^2+\eta^2}{L(L+1)}\,\lambda_{L-1} \qquad (L=1,2,3,\ldots), \tag{2.1}$$

$$\lambda_0 = 1, \qquad \lambda_1 = \omega - \eta. \tag{2.2}$$

(In particular, all λ_L are real.) It is readily seen, that (2.1) possesses a minimal solution, whenever $\omega \neq 0$, which we denote by λ'_L, assuming $\lambda'_0 = 1$. The desired solution λ_L is known to be nonminimal [3]. It would appear, therefore, that (2.1) and (2.2) lend themselves conveniently for the accurate generation of λ_L. This is indeed the case as long as η is not too large. As $\eta \to \infty$, it was observed, however, that λ_L 'approaches' the minimal solution λ'_L in the sense that $\lambda_1 - \lambda'_1 \to 0$. Therefore, the initial values of λ_L, as η becomes large, will ultimately be indistinguishable (in finite arithmetic) from those of λ'_L, even though for large L the two solutions behave quite differently. In fact, $\lambda_L \to \infty$ as $L \to \infty$, while $\lambda'_L \to 0$ as $L \to \infty$. It is clear, therefore, that the solution λ_L cannot be determined accurately from initial values, when η is large, unless one resorts to multiple-precision arithmetic.

If it were possible to compute

$$\varepsilon = \lambda_1 - \lambda'_1 \tag{2.3}$$

accurately, then the following device can be used [3].

Let λ''_L be the solution of (2.1) defined by

$$\lambda''_0 = -\lambda'_1, \qquad \lambda''_1 = 1. \tag{2.4}$$

Then

$$\lambda_L = \lambda'_L + \frac{\varepsilon}{1+\lambda'^2_1}\,(\lambda''_L + \lambda'_1\,\lambda'_L), \tag{2.5}$$

which shows that for small ε the solution λ_L initially follows closely λ'_L until the dominance of λ''_L outweighs the smallness of ε. All terms in (2.5) can be computed accurately: λ'_L by the algorithms mentioned at the beginning of section 1, ε by assumption, and λ''_L by straightforward application of (2.1) and (2.4). It may be noted, in this respect, that relative errors δ_0, δ_1 in the initial values λ''_0, λ''_1 give rise to comparable relative errors in λ''_L, when L is large, namely errors approximately equal to $\delta_0 \lambda_1'^2/(1+\lambda_1'^2)$ and $\delta_1/(1+\lambda_1'^2)$, respectively. This indicates that the solution λ''_L is computationally well defined.

3. We proceed now to derive an explicit expression for ε defined in (2.3). We may assume $\eta > 0$, in which case $\omega > 0$ [cf. (4.2) below].

First of all we note that a minimal solution of (2.1) is given by

$$\lambda_L^{min} = i^{L+1} Q_L^{(i\eta, -i\eta)}(-i\omega),$$

where $Q_n^{(\alpha,\beta)}(x)$ denotes the Jacobi function of the second kind. This follows from the asymptotic formula[2]

$$(x-1)^\alpha (x+1)^\beta Q_n^{(\alpha,\beta)}(x) \sim n^{-1/2} [x - (x^2-1)^{1/2}]^{n+1} \phi(x) \quad (n \to \infty), \tag{3.1}$$

in which x is a real or complex number outside the segment $[-1, 1]$, α and β are real or complex with $\mathrm{Re}\,\alpha > -1$, $\mathrm{Re}\,\beta > -1$, and $\phi(x) \neq 0$ is regular outside of $[-1, 1]$ and independent of n. It is understood in (3.1) that one takes that branch of $x - (x^2-1)^{1/2}$ for which $|x - (x^2-1)^{1/2}| < 1$. Letting $x = -i\omega$, $\alpha = i\eta$, $\beta = -i\eta$, one readily obtains

$$\lambda_{L+1}^{min} / \lambda_L^{min} \sim \omega - (\omega^2+1)^{1/2} \quad (L \to \infty).$$

On the other hand, it is known [3, p. 66] that

$$\lambda_{L+1}/\lambda_L \sim \omega + (\omega^2+1)^{1/2} \quad (L \to \infty),$$

showing that λ_L^{min} is indeed minimal. It follows, therefore, that

$$\lambda'_L = \frac{\lambda_L^{min}}{\lambda_0^{min}} = i^L \frac{Q_L^{(i\eta,-i\eta)}(-i\omega)}{Q_0^{(i\eta,-i\eta)}(-i\omega)}.$$

In order to evaluate λ_0^{min}, we make use of [5, p. 75]

$$Q_0^{(\alpha,\beta)}(x) = 2^{\alpha+\beta} \frac{\Gamma(\alpha+1)\,\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)} (x-1)^{-\alpha-1} (x+1)^{-\beta} \times F\left(\alpha+1, 1; \alpha+\beta+2; \frac{2}{1-x}\right),$$

where $F(a, b; c; x)$ denotes the hypergeometric function. Assuming $\alpha + \beta = 0$ (as is the

---

[2]) See [5, p. 223], where the result is obtained for real $\alpha > -1$, $\beta > -1$. The derivation by the method of steepest descent, however, is valid also when α and β are complex, with $\mathrm{Re}\,\alpha > -1$, $\mathrm{Re}\,\beta > -1$.

case in our context), we have

$$F\left(\alpha+1,1;2;\frac{2}{1-x}\right)=\sum_{\nu=0}^{\infty}\frac{(\alpha+1)(\alpha+2)\cdots(\alpha+\nu)}{(\nu+1)!}\left(\frac{2}{1-x}\right)^{\nu}$$
$$=\sum_{\nu=0}^{\infty}\frac{(-1)^{\nu}}{\nu+1}\binom{-\alpha-1}{\nu}\left(\frac{2}{1-x}\right)^{\nu},$$

which, on applying

$$\sum_{\nu=0}^{\infty}(-1)^{\nu}\binom{-\alpha-1}{\nu}\frac{z^{\nu}}{\nu+1}=\frac{1}{z}\int_0^z(1-t)^{-\alpha-1}\,dt=\frac{1}{\alpha z}\left[(1-z)^{-\alpha}-1\right],$$

becomes

$$F\left(\alpha+1,1;2;\frac{2}{1-x}\right)=\frac{1-x}{2\alpha}\left[\left(1-\frac{2}{1-x}\right)^{-\alpha}-1\right].$$

Therefore,

$$Q_0^{(\alpha,-\alpha)}(x)=\Gamma(\alpha+1)\Gamma(-\alpha+1)(x-1)^{-\alpha-1}(x+1)^{\alpha}\frac{1-x}{2\alpha}\left[\left(1-\frac{2}{1-x}\right)^{-\alpha}-1\right]$$
$$=-\frac{\Gamma(\alpha+1)\Gamma(-\alpha+1)}{2\alpha}\left[1-(x+1)^{\alpha}(x-1)^{-\alpha}\right],$$

provided that

$$-\pi<\arg(x-1)+\arg\left(1-\frac{2}{1-x}\right)\leqslant\pi.$$

Letting $x=-i\omega$ (which satisfies this condition), and $\alpha=i\eta$, we obtain

$$\lambda_0^{min}=iQ_0^{(i\eta,-i\eta)}(-i\omega)=-\frac{\Gamma(1+i\eta)\Gamma(1-i\eta)}{2\eta}\left[1-(1-i\omega)^{i\eta}(-1-i\omega)^{-i\eta}\right].$$

By an elementary computation one finds that

$$(1-i\omega)^{i\eta}(-1-i\omega)^{-i\eta}=e^{-2\eta\phi},$$

where

$$\phi=\operatorname{arc\,tan}\frac{1}{\omega}.$$

Since, furthermore, $\Gamma(1+i\eta)\Gamma(1-i\eta)=\pi\eta/\sinh(\pi\eta)$, we finally obtain

$$\lambda_0^{min}=-\frac{\pi}{2\sinh(\pi\eta)}\left(1-e^{-2\eta\phi}\right). \tag{3.2}$$

To determine λ_1^{min}, we use [5, p. 80]

$$Q_1^{(\alpha,\beta)}(x) = \tfrac{1}{2}[(\alpha+\beta+2)\,x + \alpha - \beta]\,Q_0^{(\alpha,\beta)}(x) - 2^{\alpha+\beta-1}(\alpha+\beta+2)\frac{\Gamma(\alpha+1)\,\Gamma(\beta+1)}{\Gamma(\alpha+\beta+2)}(x-1)^{-\alpha}(x+1)^{-\beta},$$

and find

$$\lambda_1^{min} = -\,Q_1^{(i\eta,\,-i\eta)}(-\,i\omega) = (\omega-\eta)\,\lambda_0^{min} + \frac{\pi\eta}{\sinh(\pi\eta)}\,e^{-2\eta\phi}. \tag{3.3}$$

Combining (3.2) and (3.3), we get

$$\lambda_1' = \frac{\lambda_1^{min}}{\lambda_0^{min}} = \omega - \eta - \frac{2\eta}{e^{2\eta\phi}-1}. \tag{3.4}$$

Therefore, in view of (2.2), (2.3), the desired expression for ε is

$$\varepsilon = \lambda_1 - \lambda_1' = \frac{2\eta}{e^{2\eta\phi}-1}, \quad \phi = \arctan\frac{1}{\omega}. \tag{3.5}$$

This result both explains the phenomenon described in section 2, and provides a simple formula to compute ε, and thus λ_L by means of (2.5).

4. The values of the exponentials in (1.1), (1.3), and (3.5), for large $|\eta|$, may become so large (or so small) as to exceed the range of permissible floating point numbers on a particular computer. If this range is given by $[10^{-R}, 10^{R}]$, such 'overflow' will occur in any of the following three cases,

$$e^{\pi|\eta|/2} > 10^R, \quad e^{\omega\varrho} > 10^R, \quad e^{2|\eta|\phi} > 10^R. \tag{4.1}$$

By definition of ω [3, p. 65], we have

$$\omega\varrho = \begin{cases} \pi\eta \quad (\tau \geqslant 1), \\ \eta\,[\pi - 2\arccos\sqrt{\tau} + 2\sqrt{\tau(1-\tau)}] \quad (0<\tau<1), \\ 0 \quad (\tau<0), \end{cases} \tag{4.2}$$

where $\tau = \varrho/2\eta$. Since $0 \leqslant \pi - 2\arccos\sqrt{\tau} + 2\sqrt{\tau(1-\tau)} \leqslant \pi$ for $0 \leqslant \tau \leqslant 1$, it follows that $0 \leqslant \omega\varrho \leqslant \pi|\eta|$. Moreover, $2|\eta|\phi = 2|\eta|\arctan(1/\omega) \leqslant \pi|\eta|$. Therefore, none of the cases in (4.1) will arise if $|\eta|$ is restricted to satisfy

$$e^{\pi|\eta|} \leqslant 10^R, \quad \text{i.e.} \quad |\eta| \leqslant \frac{R\ln 10}{\pi} = (.7329\ldots)\,R. \tag{4.3}$$

On the CDC 3600, e.g., one has $R = 308$, so that on this computer the restriction (4.3) amounts to $|\eta| \leqslant 225.7\ldots$.

Another place where overflow may occu·· is in the generation of the quantities λ_L'' by (2.1), (2.4), when $\eta > 0$. As $L \to \infty$, one finds

$$\lambda_L'' \sim \frac{1 + \lambda_1'^2}{2\eta} (e^{2\eta\phi} - 1)\, \lambda_L,$$

and in view of the known asymptotic behavior of λ_L (cf. [3, p. 66]), and (3.4),

$$\lambda_L'' \simeq \frac{1 + (\omega - \eta)^2}{2\eta} (2\pi L)^{-1/2} (1 + \omega^2)^{-1/4} e^{\eta\phi} [\omega + \sqrt{\omega^2 + 1}]^{L+1/2},$$

having assumed $\exp(2\eta\phi) \gg 1$. Roughly, then, $\lambda_L'' \simeq \exp(\eta\phi)\,[\omega + \sqrt{\omega^2 + 1}]^L$, and to avoid overflow we should have

$$e^{\eta\phi} [\omega + \sqrt{\omega^2 + 1}]^L \leqslant 10^R.$$

Letting ν denote the largest value of L for which λ_L'' is required (an estimate for ν may be found in [3, p. 69]) the last inequality is satisfied if

$$\eta \arctan\left(\frac{1}{\omega}\right) + \nu \ln(\omega + \sqrt{\omega^2 + 1}) \leqslant R \ln 10. \tag{4.4}$$

We note that ν depends not only on the parameters η, ϱ, and L, but also on the desired accuracy for $F_L(\eta, \varrho)$. If six significant digits are required, for example, it was found that (4.4) holds true for $0 \leqslant \eta \leqslant 100$, $0.1 \leqslant \varrho \leqslant 200$, $0 \leqslant L \leqslant 100$, if $R = 308$ as before.

The region in which our recurrence algorithm is applicable (using standard floating point artithmetic) is thus delineated by the two inequalities (4.3) and (4.4).

REFERENCES

[1] Abramowitz, M., *Coulomb Wave Functions*, in: *Handbook of Mathematical Functions* (edited by M. Abramowitz and I. A. Stegun), NBS Applied Mathematics Series, No. 55 (Washington, D.C. 1964), Chapt. 14.

[2] Fröberg, C.E., *Numerical Treatment of Coulomb Wave Functions*, Rev. Modern Phys. *27*, 399–411 (1955).

[3] Gautschi, W., *Computational Aspects of Three-Term Recurrence Relations*, SIAM Rev. *9*, 24–82 (1967).

[4] Gautschi, W., *Algoritm 292, Regular Couloumb wave functions*. Comm. ACM *9*, 793–795 (1966).

[5] Szegő, G., *Orthogonal Polynomials*, Revised ed., AMS Colloquium Publ., Vol. 23, (New York 1959.)

Purdue University

On the Distribution of Prime Divisors

P. ERDŐS (Budapest, Hungary)

Dedicated to the 75th birthday of Professor A. OSTROWSKI

Denote by $V(n)$ the number of distinct prime factors of n. A well known theorem of HARDY and RAMANUJAN [8] states that for almost all n $V(n) = (1+o(1)) \log_2 n$ and a special case of a result of KAC and myself [3] states that the density of integers n satisfying

$$V(n) > \log_2 n + c(\log_2 n)^{1/2}$$

is $1/\sqrt{2\pi} \int_c^\infty c^{-x^2/2} dx$ (almost all n means for all neglecting a sequence of density 0, $\log_k n$ denotes the k-fold iterated logarithm).

Denote by $v(n; u, v)$ the number of prime factors p of n satisfying $u<p<v$. Let $u=u(x)$, $v=v(x)$, and assume that $\log_2 v - \log_2 u \to \infty$, TURÁN [11] proved that for all but $o(x)$ integers $n<x$ $v(n; u, v) = (1+o(1)) (\log_2 v - \log_2 u)$. We now investigate the case when u and v depend on n. If the dependence is regular Turán's method carries through without too much difficulty. In the general case, somewhat unexpectedly, $\log_2 v - \log_2 u \to \infty$ is not sufficient for $v(n; u, v) = (1+o(1)) (\log_2 v - \log_2 u)$, but in fact we show that this holds uniformly in u and v under rather mild conditions.

In fact we prove

THEOREM 1. *Assume* $(\log_2 v - \log_2 u)/\log_3 n \to \infty$. *Then we have for almost all n uniformly in u and v*

$$V(n; u, v) = (1 + o(1)) (\log_2 v - \log_2 u).$$

Theorem 1 is best possible. In fact we have

THEOREM 2. *There are two continuous functions* $f_1(c)$ *and* $f_2(c)$, $f_1(0)=\infty$, $f_1(\infty)=1$, $f_2(c)$ *is strictly decreasing for* $0<c<\infty$; $f_2(c)=0$ for $0\leqslant c\leqslant 1$, $f_2(\infty)=1$, $f_2(c)$ *is strictly increasing in* $1<c<\infty$, *satisfying for almost all n and for every* $c>0$

$$\max V(n; u, v) = (1 + o(1)) f_1(c) (\log_2 v - \log_2 u)$$

and

$$\min V(n; u, v) = (1 + o(1)) f_2(c) (\log_2 v - \log_2 u)$$

where the max *and* min *is taken with n fixed over the values* $1\leqslant u<v\leqslant n$ *satisfying*

$$\log_2 v - \log_2 u > c \log_3 n$$

We will prove Theorem 1 in full detail, the proof of Theorem 2 is similar but more complicated and will be omitted (see [4]).

Theorems 1 and 2 can be generalised for a large class of additive functions but

Received December 1, 1967.

we will not discuss this here. I only formulate the probabilistic theorem which corresponds to theorems 1 and 2 in case of Rademacher functions $r_k(x)$.

THEOREM 3. *For all x neglecting a set of measure 0 we have*

$$\lim \frac{1}{v-u} \sum_{k=u}^{v} r_k(x) = 0$$

uniformly in u and v if $(v-u)/\log u \to \infty$.

Further

$$\lim \frac{1}{v-u} \sum_{k=u}^{v} r_k(x) = f_1(c)$$

where $u \to \infty$ and $v > u + c \log u$. $f(c) = 1$ for $0 \leqslant c \leqslant 1$, $f(\infty) = 0$.

$f(c)$ is strictly decreasing in $1 < c < \infty$.

The proof of Theorem 3 follows from simple independence arguments and will not be given here. Theorem 3 could be generalised for other independent functions but I have not investigated how far this generalisation will go.

Let $p_1 < \cdots < p_{v(n)}$ be the distinct prime factors of n. In a previous paper [5] I stated (it can be proved by the methods of probabilistic number theory [5]), that roughly speaking the order of magnitude of the i-th prime factor of n is $\exp \exp i$ ($\exp z = c^z$). More precisely for every $\varepsilon > 0$ and $\eta > 0$ there is an i_0 so that for all but εx integers $n \leqslant x$ we have for all $i_0 < i \leqslant v(n)$

$$(1-\eta)\, i < \log\log p_i < (1+\eta)\, i .$$

In fact in [5] a sharper result is stated.

By the same method I can prove that for every $\varepsilon > 0$ and $\eta > 0$ there is an i_0 so that for all but εx integers $n \leqslant x$ we have for all $i_0 < i \leqslant v(n)$

$$(1-\eta)\, i < \log\log n - \log\log p_i < (1+\eta)\, i .$$

I now state without proof a few results about prime factors of integers which can be obtained by standard methods of probabilistic number theory (see [9]).

We have for almost all integers n

$$\sum_{\substack{p_i \mid n \\ p_i > \exp \exp i}} \frac{1}{i} = \left(\tfrac{1}{2} + o(1)\right) \log_3 n . \tag{1}$$

In fact more generally we have for almost all n

$$\sum{}' \frac{1}{i} = (1 + o(1)) \int_c^{\infty} e^{-x^2/2}\, dx \log_3 n \tag{2}$$

where the dash indicates that the summation is extended over the $p_i|n$, $p_i >$ exp exp $(i + c\,i^{1/2})$.

(1) and (2) follows from a result of CHUNG and myself [2] (which is a generalisation of a result of PAUL LEVY [10]) and BRUNS method (similarly as in [3]).

On the other hand it is not true that for almost all n

$$\sum_{\substack{p_i|n\\ p_i > \exp\exp i}} 1 = (\tfrac{1}{2} + o(1)) \log_2 n\,. \tag{3}$$

Instead of (3) the following result holds: Let $j_1 < \cdots$ be a sequence of integers. Put $\sum_{j_r < y} 1 = A(y)$ and assume $A(y) \to \infty$, $A(2y)/A(y) \to 1$ as $y \to \infty$. Then for almost all integers

$$\sum_{\substack{p_{j_r}|n\\ p_{j_r} > \exp\exp j_r}} 1 = (\tfrac{1}{2} + o(1))\, A(\log_2 n)$$

From the arc sine law [6] and BRUNS method we obtain the following result. The density of integers n for which

$$\frac{1}{\log_2 n} \sum_{\substack{p_i|n\\ p_i < \exp\exp i}} 1 < \alpha$$

holds, equals $2/\pi$ arc sin $\alpha^{1/2}$. This shows that (3) is not true.

Denote by $f(\alpha)$ the density of integers the largest prime factor of which is $< n^{\alpha}$. It is easy to see that $f(\alpha)$ exists and is a continuous strictly increasing function of α, $f(0) = 0, f(1) = 1$. DICKMAN, DE BRUIJN, BUCHSTAB [1] and others obtained more or less explicit formulas for $f(\alpha)$. Denote by $f_i(\alpha)$ the density of the integers n for which $p_{v(n)-i} < n^{\alpha}$. It is easy to see by methods similar to those used for $f(\alpha) = f_0(\alpha)$ that $f_i(\alpha)$ is a strictly increasing continuous function of α, $f_0(\alpha)\, f_i(0) = 0$, $f_i(1/i+1) = 1$. As far as I know $f_i(\alpha)$ has never been computed explicitely. It follows by the methods of probabilistic number theory [3] that the density of integers n for which

$$p_i > \exp\exp(i + c\,i^{1/2})$$

and

$$p_{V(n)-j} > \exp \log n / e^{j - c\,j^{1/2}}$$

approaches as $i \to \infty, j \to \infty$ $1/\sqrt{2\pi} \int_c^{\infty} e^{-x^2/2} dx$.

Denote by $\alpha(i, k)$ the density of the integers the i-th prime factor of which is p_k. Clearly $\alpha(i, k)$ exists for every i and k (because the sequence of numbers the i-th prime factor of which equals p_k can be obtained by set theoretic operations from a finite number of arithmetic progressions), and is positive for $k \geq i$. It might be of interest to determine max $\alpha(i, k)$, or at least to obtain an asymptotic formula for it. I only succeeded in obtaining here some rather crude results.

Now we prove Theorem 1. Because of the slow growth of the iterated logarithms it clearly will suffice to prove the following.

THEOREM 1'. *To every $\varepsilon>0$ there is an A so that for every $x>x_1(\varepsilon, A)$ the number of integers $n<x$ for which for every $(u<v\leqslant x)$*

$$\log_2 v - \log_2 u > A \log_3 x \tag{4}$$

we have

$$(1-\varepsilon)(\log_2 v - \log_2 u) < V(n; u, v) < (1+\varepsilon)(\log_2 v - \log_2 u)$$

is $x+o(x)$.

To prove Theorem 1' we only have to show that for $x>x_0(\varepsilon, A)$ the number of integers $n<x$ for which there are values $u<v\leqslant x$ satisfying (4) and for which

$$V(n; u, v) < (1-\varepsilon)(\log_2 v - \log_2 u) \tag{5}$$

or

$$V(n; u, v) > (1+\varepsilon)(\log_2 v - \log_2 u) \tag{6}$$

is $o(x)$.

Put $w_i = \exp\exp i$. First we prove

LEMMA 1. *The number of integers $n\leqslant x$ for which*

$$\max_i V(n; w_i, w_{i+1}) > \log_3 x \tag{7}$$

is $o(x)$.

The number of integers $n\leqslant x$ for which $V(n; w_i, w_{i+1})>\log_3 x$ holds is clearly at most $\sum' [x/a_j]$ where in $\sum'$ the summation is extended over the integers a_j which are the product of $t=[\log_3 x]$ distinct primes p, $w_i<p<w_{i+1}$. Now clearly by the well known theorem of Mertens $\sum_{p<y} 1/p = \log_2 y + c + o(1)$ we have uniformly in i

$$\sum{}' \left[\frac{x}{a_j}\right] < x\Bigg(\sum_{w_i<p<w_{i+1}} \frac{1}{p}\Bigg)^t \Bigg/ t = x\,\frac{1+o(1)}{t!} = o\left(\frac{x}{\log_2 x}\right). \tag{8}$$

In (7) there are only $\log_2 x$ choices of i, thus Lemma 1 follows immediately from (8).

From Lemma 1 we easily deduce that to prove Theorem 1' it suffices to consider in (5) and (6) only those u's and v's for which

$$u = w_i, \quad v = w_j, \quad 1 \leqslant i < j \leqslant \log_2 x, \quad j - i > A \log_3 x, \tag{9}$$

LEMMA 2. *For all but $o(x)$ integers $n\leqslant x$*

$$V(n; x^{1/\log_2 x}, x) = (1 + o(1)) \log_3 x .$$

Lemma 2 follows immediately by the method of TURÁN [11].

From Lemma 2 it easily follows that in (9) we can assume $j \leqslant \log_2 x - \log_3 x$ or $w_j \leqslant \exp(\log x/\log_2 x)$.

Denote ($p^\alpha \| n$ means that $p^\alpha | n$ but $p^{\alpha+1} \nmid n$)

$$A(n) = \prod{}' p^\alpha$$

where the dash indicates $p^\alpha \| n, p \leqslant \exp(\log x/\log_2 x)$.

LEMMA 3. *For all but $o(x)$ integers $n \leqslant x$ we have $A(n) < x^{1/2}$.*

Lemma 3 is well known, but for the sake of completeness we give the simple proof. We evidently have by the theorem of MERTENS ($L = \exp(\log x/\log_2 x)$)

$$\prod_{n=1}^{x} A(n) < \prod_{p<L} p^{\sum_{k=1}^{\infty} x/p^k} = \exp\left(x \sum_{p<L} \frac{\log p}{p-1}\right) < \exp(2x \log x/\log_2 x). \tag{10}$$

From (10) we obtain that the number of integers $n \leqslant x$ for which $A(n) \geqslant x^{1/2}$ is less than $4x/\log_2 x$, which proves the Lemma.

LEMMA 4. *Let i and j satisfy* (9). *Then the number of integers $S(i,j)$ satisfying $A(n) < x^{1/2}$ which satisfy* (5) *or* (6) *is $o(x/(\log_2 x)^2)$ uniformly in i and j.*

Before we prove Lemma 4 we deduce Theorem 1′ from our four Lemmas. We already deduced from Lemmas 1 and 2 that to prove theorem 1′ it suffices to consider the values i and j satisfying (9) and from Lemma 3 we obtain that we only have to consider the $n \leqslant x$ with $A(n) < x^{1/2}$. There are fewer than $(\log_2 x)^2$ values of i and j satisfying (9), hence from Lemma 4 the number of integers which satisfy (5) and (6) for some i and j satisfying (9) is $o(x)$ as stated.

Thus to complete the proof of Theorem 1, we only have to prove Lemma 4. Let $i < j$ satisfy (9) and denote

$$A_{i,j}(n) = \prod p^\alpha, \quad \text{where} \quad p^\alpha \| n, \quad w_i < p < w_j.$$

LEMMA 5. *Let $t \leqslant x^{1/2}$. The number of integers $f_{i,j}(x, t)$ for which $n \leqslant x$ and $A_{i,j}(n) = t$ is less than $c_1 x/t \exp(j-i)$.*

$f_{i,j}(x, t)$ clearly equals the number of integers $m \leqslant x/t$ which have no prime factor p, satisfying $w_i < p < w_j$. It immediately follows from BRUNS method [7] (here we use $A(n) < x^{1/2}$)) and the well known theorem of MERTENS $\prod_{w_i<p<w_j} (1 - 1/p) < c_2/\exp(j-i)$ that $f_{i,j}(x, t)$ is less than

$$c_3 \frac{x}{t} \prod_{w_i<p<w_j} \left(1 - \frac{1}{p}\right) < c_1 x/t \exp(j-i),$$

which proves Lemma 5.

Denote now by $a_1 < \cdots$ the integers not exceeding $x^{1/2}$ which are composed entirely from the primes $w_i<p<w_j$ and for which

$$V(a_r) < (1-\varepsilon)(j-i) \tag{11}$$

or

$$V(a_r) > (1+\varepsilon)(j-i). \tag{12}$$

We evidently have

$$\sum \frac{1}{a_r} < \sum_{u<(1-\varepsilon)(j-1)} \left(\sum \frac{1}{P}\right)^u \Big/ u! + \sum_{u>(1+\varepsilon)(j-1)} \left(\sum \frac{1}{P}\right)^u \Big/ u! \tag{13}$$

where P runs through all the powers of the primes p, satisfying $w_i<p<w_j$. By the theorem of Mertens we obtain by a simple calculation (using $j-i>A \log_3 x$)

$$\left.\begin{aligned}\sum \frac{1}{a_r} &< \sum_{u<(1-\varepsilon)(j-i)} \frac{(j-i+c_4)^u}{u!} + \sum_{u>(1+\varepsilon)(j-i)} \frac{(j-i+c_4)^u}{u!} \\ &< \exp(j-i+c_4)(1-c_5\varepsilon)^{-c_6\varepsilon(j-i)} = o\left(\exp(j-i)/(\log_2 x)^2\right)\end{aligned}\right\} \tag{14}$$

for sufficiently large $A=A_0(\varepsilon)$. (14) states a well known property of the exponential series, namely in the expansion $e^z=\sum_{k=0}^{\infty} z^k/k!$, the main contribution comes from the terms $k=(1+o(1))z$.

We evidently have from Lemma 5 and (14)

$$S(i,j) = \sum_r f_{i,j}(x,a_r) < \sum_r \frac{c_1 x}{a_r \exp(j-i)} = o\left(\frac{x}{(\log_2 x)^2}\right),$$

which proves Lemma 4 and thus completes the proof of Theorems 1′ and 1.

REFERENCES

[1] DE BRUIJN, N.G., *On the number of positive integers* $\leqslant x$ *and free of prime factors and* $\geqslant y$, *Nederl. Akad. Wetensch. Indag. Math. 13*, 50–60 (1951).

[2] CHUNG, K.L., and ERDŐS, P., *Probability Limit Theorems Assuming Only the First Moment*, Mem. Amer. Math. Soc. No. 6, *1951.*

[3] ERDŐS, P. and KAC, M., *The Gaussian Law of Errors in the Theory of Additive Number Theoretic Functions*, Amer. J. Math. *62*, 738–742 (1940).

[4] ERDŐS, P., *Some Remarks on Prime Factors of Integers*, Canad. J. Math. *11*, 161–167 (1959), and *On some Properties of Prime Factors of Integers*, Nagoya Math. J. *27*, 617–623 (1966).

[5] ERDŐS, P., *On the distribution function of additive functions*, Ann. of Math. *47*, 1–20 (1946).

[6] See e.g. ERDŐS, P., KAC, M., *On the number of positive sums of independent random variables*, *Bull. Amer. Math. Soc. 53*, 1011–1020 (1947).

[7] See e.g. HALBERSTAM, H., ROTH, K.F., *Sequences*, Vol. I (Clarendon Press, Oxford 1966), Chapt. IV.

[8] HARDY, G.H. and RAMANUJAN, S., *The normal number of prime factors of n*, Quart. J. Math. *48*, 76–92 (1917).

[9] KUBYLYUS, I. P., *Probabilistic methods in number theory*, Amer. Math. Soc. Transl. [2] *19*, 47–85 (1962).
[10] LEVY, P., *Théorie de l'addition des variables aléatoires* (Gauthier-Villars, Paris 1937).
[11] TURÁN, P., *On a theorem of Hardy and Ramanujan*, J. London Math. Soc. *9*, 274–276 (1934).

Mathematical Research Institute
Hungarian Academy of Sciences

On a Certain Limitation of Eigenvalues of Matrices

PAUL TURÁN (Budapest, Hungary)

To the 75th birthday of A. M. Ostrowski

1. Let **A** be an $n\times n$ matrix with arbitrary complex entries and $\lambda_1, \lambda_2, \dots, \lambda_n$ its eigenvalues so that

$$|\lambda_1| \geqq |\lambda_2| \geqq \cdots \geqq |\lambda_n|. \tag{1.1}$$

The limitations of all or of some eigenvalues with various quantities directly connected with the elements of **A** form a vast literature in which OSTROWSKI is one of the leading contributors. These investigations are connected with applications in the technique (determination of 'zones of danger') and economics (certain economical indices) which give to it a special interest and flavor. In what follows I shall deal with a limitation which – contrary to the largest part of the pertaining literature – does not work with the absolute values of the entries. I assert the following:

THEOREM. *The absolute smallest eigenvalue λ_n of the matrix* **A** *satisfies the inequality*

$$|\lambda_n| \leqq \max_{\nu=1,2,\dots,n} |\mathrm{trace}\,\mathbf{A}^\nu|^{1/\nu}$$

The inequality is best-possible and all cases where equality occur can be described.

The proof of this theorem will be given in **2.**, **3.** and **4.**; **5.** will contain some further remarks and problems.

2. The theorem will be a quick consequence of the following

LEMMA. *If $z_1, z_2, \dots, z_n$ are complex numbers with*

$$\min_j |z_j| \geqq 1 \tag{2.1}$$

then

$$\max_{\nu=1,2,\dots,n} |z_1^\nu + \cdots + z_n^\nu| \geqq 1.$$

Equality occurs if and only if the $(z_1, z_2, \dots, z_n)$-system consists of n vertices of a regular $(n+1)$-gon inscribed into the unit-circle.

This lemma occurred in the German edition of my book[1]) *Eine neue Methode in der Analysis und deren Anwendungen* (1953), p. 25. as Satz I, but without giving all extremal systems. This was the only general theorem on power sums which I did not

[1]) A completely rewritten and enlarged English version will appear in the Interscience Tracts Series.

Received December 1, 1967

apply in some different subject in the rest of the book. Now even this theorem has an application.

For the proof of this lemma let

$$z^n + a_1 z^{n-1} + \cdots + a_n = 0 \tag{2.2}$$

be the equation with the zeros $z_1, z_2, \ldots, z_n$. Then the Newton-Girard formulae give with the notation

$$s_\nu = z_1^\nu + z_2^\nu + \cdots + z_n^\nu$$

the relations

$$\left.\begin{array}{rl} s_1 + a_1 & = 0 \\ s_2 + a_1 s_1 + 2a_2 & = 0 \\ \vdots & \vdots \\ s_n + a_1 s_{n-1} + \cdots + n a_n & = 0 \end{array}\right\} \tag{2.3}$$

Since

$$|a_n| = |z_1 z_2 \ldots z_n| \geqq 1\,, \tag{2.4}$$

we get

$$\max_{\nu=1,\ldots,n} |a_\nu| \geqq 1\,. \tag{2.5}$$

Let the index j be defined (not necessarily uniquely) by

$$\max_{\nu=1,\ldots,n} |a_\nu| = |a_j| \geqq 1\,. \tag{2.6}$$

Then from the jth equation of (2.3), putting

$$\max_{\nu=1,\ldots,n} |s_\nu| = M\,, \tag{2.7}$$

we get

$$j\,|a_j| \leqq M(1 + |a_1| + \cdots + |a_{j-1}|) \leqq M(1 + (j-1)\,|a_j|) \leqq Mj\,|a_j| \tag{2.8}$$

which proves the first part of the lemma.

3. Concerning the sign of equality in the lemma, (2.8) and (2.4) give at once that

$$|a_1| = |a_2| = \cdots = |a_n| = 1 \tag{3.1}$$

and hence from (2.4) and (2.1)

$$|z_1| = |z_2| = \cdots = |z_n| = 1\,. \tag{3.2}$$

We have, of course,

$$\max_{\nu=1,\ldots,n} |s_\nu| = 1\,. \tag{3.3}$$

The first equation in (2.3) gives

$$|s_1| = |a_1| = 1\,;$$

further, replacing the $(z_1, z_2,\ldots, z_n)$ extremal-system by $(z_1 e^{-i\alpha}, z_2 e^{-i\alpha},\ldots, z_n e^{-i\alpha})$ if necessary, it follows that

$$s_1 = -1, \quad a_1 = 1. \tag{3.4}$$

Suppose we have already proved for $1 \leqq \mu < n$ and such a 'normed' extremal-system that

$$\left.\begin{aligned} s_1 = s_2 = \cdots = s_\mu = -1 \\ a_1 = a_2 = \cdots = a_\mu = 1. \end{aligned}\right\} \tag{3.5}$$

From the $(\mu+1)$th equation of (2.3) we get, using (3.5),

$$s_{\mu+1} = \mu - (\mu+1)\, a_{\mu+1}. \tag{3.6}$$

This and (3.3) give

$$1 \geqq |s_{\mu+1}| = |(\mu+1)\, a_{\mu+1} - \mu| \geqq (\mu+1)\,|a_{\mu+1}| - \mu \tag{3.7}$$

which owing to (3.1) equals 1.

Hence in (3.7) the equality sign holds everywhere, i.e.

$$a_{\mu+1} = 1,$$

and, from (3.6), also

$$s_{\mu+1} = -1$$

holds. But this means that the normed extremal-system satisfies the equation

$$z^n + z^{n-1} + z^{n-2} + \cdots + z + 1 = 0.$$

4. Our theorem follows at once from the lemma. We apply it with

$$z_j = \frac{\lambda_j}{\lambda_n}, \quad j = 1, 2, \ldots, n. \tag{4.1}$$

This gives the existence of an integer $1 \leqq \nu_0 \leqq n$ such that

$$\left|\sum_{j=1}^{n} \lambda_j^{\nu_0}\right| \geqq |\lambda_n|^{\nu_0}$$

i.e.

$$|\lambda_n| \leqq \left|\sum_{j=1}^{n} \lambda_j^{\nu_0}\right|^{1/\nu_0} \leqq \max_{\nu=1,\ldots,n} \left|\sum_{j=1}^{n} \lambda_j^{\nu}\right|^{1/\nu} \tag{4.2}$$

Since the eigenvalues of $\mathbf{A}^\nu$ are the λ_j^ν numbers we get

$$\left|\sum_{j=1}^{n} \lambda_j^{\nu}\right| = |\operatorname{trace} \mathbf{A}^\nu|$$

which with (4.2) completes the proof of our theorem.

Equality in the theorem can occur obviously when the numbers in (4.1) are $(n+1)$th roots of unity. This occurs then and only then when the eigenvalues of $\mathbf{A}$ lie on the vertices of a regular $(n+1)$-gon with its center in the origin.

5. It is natural to ask whether or not analogous theorems hold for the absolutely largest eigenvalue of $\mathbf{A}$. In this connection we can assert two results.

All eigenvalues of $\mathbf{A}$ *lie on the disk*

$$|z| \leqq \max_{\nu=1,2,\ldots,2n-1} |\operatorname{trace} \mathbf{A}^\nu|^{1/\nu}. \tag{5.1}$$

This inequality is best possible.

The proof follows as in **4.** but the lemma must be replaced by the following theorem of Cassels [1] (which was found and proved by Ostrowski [2] too).

If

$$\max_{j=1,\ldots n} |z_j| = 1 \tag{5.2}$$

then

$$\max_{\nu=1,2,\ldots,2n-1} |s_\nu| \geqq 1. \tag{5.3}$$

The case of equality occurs e.g. for the matrix

$$a_{11} = 1,$$
$$a_{ik} = 0, \quad \text{otherwise}.$$

But here the equality can be attained in immensely many cases; e.g. as Dancs [3] showed there are $(z_1^*, z_2^*, \ldots, z_n^*)$-systems of cardinality of continuum such that they are essentially different,

$$\max_{j=1,\ldots n} |z_j^*| = 1$$

and even

$$|z_1^{*\nu} + \cdots + z_n^{*\nu}| \leqq 1$$

holds for $\nu = 1, 2, \ldots, 3n-4$.

The second result runs as follows:

All eigenvalues of $\mathbf{A}$ *lie on the disk*

$$|z| < 3 \max_{\nu=1,\ldots,n} |\operatorname{trace} \mathbf{A}^\nu|^{1/\nu}. \tag{5.4}$$

The proof follows again as in **4.** but the lemma must be replaced by the following deep refinement of Satz II of my book due to Atkinson[2]).

[2]) In his paper [4] he proved (5.5) only with $\frac{1}{6}$; the sharper inequality (5.5) I know only from a manuscript he kindly showed me.

From (5.2) we have the inequality

$$\max_{\nu=1,2,\ldots,n} |s_\nu| > \tfrac{1}{3}. \tag{5.5}$$

In (5.4) the constant is not best possible. The best possible constant depends of course on the following extremal problem.

For the $(z_1, z_2, \ldots, z_n)$-systems with (5.2) what is the minimum M_n of

$$\max_{\nu=1,2,\ldots,n} |z_1^\nu + z_2^\nu + \cdots + z_n^\nu|$$

and what are the extremal systems?

The problem is easy for $n=2$ and gives

$$M_2 = \frac{1}{\sqrt{2}}(\sqrt{5}-1) \sim 0{,}8740320. \tag{5.6}$$

For $n=3$ J. Lawrinowicz characterized the extremal systems and found[3]) after a difficult analysis

$$M_3 \sim 0{,}8247830. \tag{5.7}$$

(5.6)–(5.7) would indicate that M_n is monotonically *decreasing*, contrary to the conjecture

$$M_n = 1 - \varepsilon_n, \qquad \lim_{n\to\infty} \varepsilon_n = 0,$$

but so far no numerical $\Theta < 1$ is known such that for *infinitely many* n's for suitable $(z_1', z_2', \ldots, z_n')$-systems with $\max\limits_{j=1,\ldots,n} |z_j'| = 1$ the inequality

$$\max_{\nu=1,2,\ldots,n} |z_1'^\nu + z_2'^\nu + \cdots + z_n'^\nu| \leqq \Theta \tag{5.8}$$

holds. The best result in this direction is due to J. Komlós, A. Sárközy and E. Szemerédi [6] who proved (5.8) replacing Θ by

$$1 - \frac{1-\varepsilon}{2}\frac{\log n}{n}$$

for arbitrarily small $\varepsilon > 0$.

REFERENCES

[1] Cassels, J. W. S., *On the Sums of Powers of Complex Numbers*, Acta Math. Acad. Sci. Hungar. *7*, 283–290 (1956).

[2] Ostrowski, A., *Sur une propriété des sommes des racines d'un polynome*, C. R. Acad. Sci. Paris *263*, 46–48 (1966).

[3]) In a forthcoming paper. The characterisation can be found in his paper [5].

[3] DANCS, I., *On Power-Sums of Complex Numbers* (Hungarian), Mat. Lapok *13*, 108–114 (1962).
[4] ATKINSON, F. V., *On Sums of Powers of Complex Numbers*, Acta Math. Acad. Sci. Hungar. *12*, 185–188 (1961).
[5] LAWRINOWICZ, J., *Calculation of a Minimum Maximorum of Complex Numbers*, Bull. Soc. Sci. Lettres Łodz *11*, 1–9 (1960).
[6] KOMLÓS, J., SÁRKÖZY, A. and SZEMERÉDI, E., *On Power-Sums of Complex Numbers* (Hungarian), Mat. Lapok *15*, 337–347 (1964).

Mathematical Institute of the Academy of Sciences

A Remark on the Square Norm

M. Hosszú (Miskolc, Hungary)

Dedicated to Prof. A. M. Ostrowski on his 75th Birthday

Let F be a real or complex valued bilinear function defined for the elements x, y of a linear space G. It is known [1] that the function

$$x \to f(x) = F(x, x)$$

has the usual property of the square norm:

$$f(x+y) + f(x-y) = 2f(x) + 2f(y), \qquad x, y \in G. \quad {}^{1}) \tag{1}$$

Here we investigate the purely algebraic nature of this property assuming only the following:

α) $(G, +)$ is a group (not necessarily abelian);

β) $f\colon G \to A$ is a mapping of G into A and $(A, +)$ is an abelian group;

γ) $f(0)=0$ holds for the unit element of G resp. A.

Assuming α)–γ), we can observe that the functional equation (1) implies the following:

$$f(-y) = \ (y), \qquad y \in G; \tag{2}$$

$$f(x+y) = f(y+x), \qquad x, y \in G; \tag{3}$$

$$f(2x) = 4f(x), \qquad x \in G; \tag{4}$$

$$f(x+y+x-y) = f(2x), \qquad x, y \in G. \tag{5}$$

In fact, (2) follows from (1) by putting $x=0$. (3) can be seen by changing x and y in (1) since (2) implies $f(y-x)=f(x-y)$. We obtain (4) by substituting $y=x$ in (1). Finally, using (1), (3) and (4), we get

$$\begin{aligned} f(x+y+x-y) &= 2f(x+y) + 2f(x-y) - f[x+y-(x-y)] = 4f(x) + 4f(y) \\ -f(x+y+y-x) &= 4f(x) + 4f(y) - f(-x+x+2y) \\ &= 4f(x) + 4f(y) - f(2y) = f(2x) \end{aligned}$$

and this verifies (5).

---

[1]) In [2] there is defined the notion of pre-Hilbert space as a linear space admitting a solution f of the functional equation (1). In [1] the solutions f of (1) are given in the form $f(x) = F(x, x)$, where F is a function additive in both variables, assuming that G is an abelian group and the range of f is an other abelian group A in which equations of the form $4a = b$ have unique solutions $a \in A$. Here we omit this solvability condition and the commutativity of $(G, +)$.

Received December 4, 1967

Let us define the functions

$$(x, y) \to F_1(x, y) = f(x+y) - f(x) - f(y) = f(x) + f(y) - f(x-y),$$
$$(x, y) \to F_2(x, y) = f(x+y) - f(x-y) = 2F_1(x, y),$$

where f satisfies (1). By (2)–(5) we can verify that both F_1 and F_2 satisfy the following relations:

$$F_2(x, x) = 2F_1(x, x) = 4f(x); \tag{6}$$

$$F(x, y) = F(y, x); \tag{7}$$

$$F(x, 0) = 0; \tag{8}$$

$$F(x+y, x+y) = F(y+x, y+x); \tag{9}$$

$$F(x+y, x-y) = F(y+x, x-y); \tag{10}$$

$$F(x+y, z) + F(x-y, z) = 2F(x, z) \tag{11}$$

for every $x, y, z \in G$.

In fact, (6)–(9) are obvious. (10) follows immediately since

$$f(x+y+x-y) = f(y+x+x-y) = f(2x)$$

is true in accordance with (5) and (3). Finally, (11) can be seen by

$$F_1(x+y, z) + F_1(x-y, z) = f(x+y+z) - f(x+y) - f(z) + f(x-y+z)$$
$$- f(x-y) - f(z) = f(x+y+z) + f(x-y+z) - f(x+y) - f(x-y) - 2f(z)$$
$$= f(z+x+y) + f(z+x-y) - 2f(x) - 2f(y) - 2f(z) = 2f(z+x) + 2f(y)$$
$$- 2f(x) - 2f(y) - 2f(z) = 2f(x+z) - 2f(x) - 2f(z) = 2F_1(x, z),$$

and

$$F_2(x+y, z) + F_2(x-y, z) = 2F_1(x+y, z) + 2F_1(x-y, z)$$
$$= 4F_1(x, z) = 2F_2(x, z).$$

Conversely, if we have a function F with properties (7)–(11), then

$$f_1(x) = F(x, x)$$

has the property γ and satisfies the functional equation

$$2f(x+y) + 2f(x-y) = 4f(x) + 4f(y), \qquad x, y \in G. \tag{1'}$$

In fact, γ) holds obviously. In order to verify (1′), let us consider

$$2F(x, x) = F(x+y, x) + F(x-y, x),$$
$$2F(y, y) = F(y+x, y) + F(y-x, y),$$
$$2F(x+y, x) = F(x+y, x+y) + F(x+y, x-y),$$

$$2F(x-y,x)=F(x-y,x+y)+F(x-y,x-y),$$
$$2F(y+x,y)=F(y+x,y+x)+F(y+x,y-x),$$
$$2F(y-x,y)=F(y-x,y+x)+F(y-x,y-x)$$

which are true by (11) and (7). But (11), (8) and (7) imply

$$F(-y,z)=-F(y,z)=F(y,-z),\qquad y,z\in G, \tag{12}$$

consequently we have (1′)

$$4F(x,x)+4F(y,y)=2F(x+y,x+y)+2F(x-y,x-y)$$

since

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

and

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

One can observe that the assumption

δ) $2a\neq 0 \quad \text{for } a\neq 0 \quad \text{in } A$

implies the equivalence of (1) and (1′). Moreover, α)–β) and δ) implies that every solution f of (1) satisfies γ) also. This can be seen by putting $x=y=0$ in (1)[2]).

Finally, remark that (11)–(12) is equivalent with the additivity:

$$F(x+y,z)=F(x,z)+F(y,z),\qquad x,y,z\in G \tag{13}$$

if and only if

$$F(x+y,z)=F(y+x,z) \tag{14}$$

holds for every x, y and z in G (not only for $z=x-y$, as in (10)). In fact, by changing x and y in (11), we have

$$F(y+x,z)+F(y-x,z)=2F(y,z)$$

hence

$$F(x+y,z)+F(y+x,z)=2F(x,z)+2F(y,z).$$

Thus for the equivalence of (11)–(12) and (13) it is nccessary and sufficient that equation (14) hold. Specifically, if we have

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

where

$$f(x+y)=f(y+x),$$

---

[2]) There is no loss of generality in assuming γ) since $g(x)=f(x)-f(0)$ satisfies both (1) and γ) for every solution f of (1) (because then $2f(0)=0$).

then the condition (14) is equivalent to

$$f(x + y + z) = f(y + x + z), \qquad x, y, z \in G.$$

Summing up the following theorems are proved:

THEOREM 1. *The functions*

$$(x, y) \to F_1(x, y) = f(x + y) - f(x) - f(y),$$
$$(x, y) \to F_2(x, y) = f(x + y) - f(x - y)$$

have the properties (6)–(11) *for every solution* f *of the functional equations* (1) *under assumptions* α)–γ).

THEOREM 2. *The function*

$$x \to f(x) = F(x, x) \qquad \textit{satisfies } (1') \textit{ and } \gamma) \textit{ for every } F$$

having properties (7)–(11).

THEOREM 3. *Provided* α)–γ) *hold for a solution* f *of* (1), *the function* F_1 *defined above is additive in both variables iff* $f(x+y+z)=f(y+x+z)$ *holds identically.*

REFERENCES

[1] ACZÉL, J., *The General Solution of Two Functional Equations by Reduction to Functions Additive in Two Variables and with the Aid of Hamel Bases,* Glasnik Mat.-Fiz. Astronom. Društvo Mat. Fiz. Hrvatske [2] *20*, 65–73 (1965).

[2] YOSIDA, K., Functional Analysis (Springer-Verlag, Berlin–Heidelberg–New York 1966), esp. pp. 39–40.

Stability of General Systems of Linear Equations

VICTOR PEREYRA (Caracas, Venezuela)

To Professor Alexander Ostrowski, on his 75th birthday

1. Introduction

A common problem in many applications is to find minimal least squares solutions to systems of simultaneous linear equations. By this we mean:

Given the $m \times n$ matrix A and the m-vector b, find an n-vector x which makes $\|Ax-b\|_2$ minimum, and that has minimum L_2 norm among all those vectors which have the same property.

It is well known that the solution to this problem is unique and given by

$$x = A^+ b \tag{1.1}$$

where A^+ is the Moore-Penrose pseudoinverse of A [8].

When the rank of A is well determined the computation of x as given by (1.1) does not constitute a very difficult problem and several working methods are available [2, 4, 5, 7, 10].

When the rank of A is not well determined then the problem is ill-posed and most methods fail to define an 'appropriate' solution. The term appropriate here is not casual since the violent ill-conditioning arising from almost linearly dependent columns of A makes it extremely difficult to define an x as in (1.1). This stems from the fact that small changes in the matrix A cause very large changes in the computed solution $\hat{x}$. See also [3, Chapt. 15] for an interesting discussion in this respect.

DEFINITION. Let A be an $m \times n$ matrix, b an m-vector, ε, δ two positive numbers. We shall say that A is *ε, δ-stable with respect to b* if for all matrix perturbations ΔA, such that

$$\frac{\|\Delta A\|}{\|A\|} \leqslant \delta$$

the minimal least squares solutions $x = A^+ b$, $\hat{x} = x + h = (A + \Delta A)^+ b$ are such that

$$\frac{\|h\|}{\|x\|} \leqslant \varepsilon .$$

A stronger concept than that of ε, δ-stability is the following:

Given the real non-negative number δ_0, and the real non-negative function $\varepsilon(\delta)$ defined in $(0, \delta_0)$, we shall say that A is *strongly ε-stable at b* if the function $(.)^+ b$ is

Received December 4, 1967 and, in revised form, November 12, 1968

continuous at A and for any δ, $0<\delta<\delta_0$

$$\|(A + \Delta A)^+ b - A^+ b\| \leqslant \varepsilon(\delta)\, \|A^+ b\|$$

provided that $\|\Delta A\| \leqslant \|A\|\, \delta$.[1])

Some properties of strong stability are:

(a) If A is strongly ε-stable then it is strongly ε'-stable for any ε' such that

$$\varepsilon(\delta) \leqslant \varepsilon'(\delta), \quad (0 < \delta < \delta_0).$$

(b) Because of the continuity there does exist an $\varepsilon(\delta)$ (in fact infinitely many) such that A is ε-stable and $\varepsilon(\delta)=o(1)$ as $\delta\to 0$.

(c) If A is strongly $\tilde{\varepsilon}$-stable then it is ε, δ-stable for any ε such that $\tilde{\varepsilon}(\delta)\leqslant\varepsilon$.

Our intention is two-fold. On one side we shall find quantitative (and computable) a posteriori upper bounds for the relative variation

$$\frac{\|x - \hat{x}\|}{\|x\|}$$

stemming from perturbations on the matrix A and the right hand side b. On the other side, we shall use these bounds in order to solve the following problem:

PROBLEM I. 'Given a general $m\times n$ matrix A, an m vector b, and ε, δ two positive numbers, find B, a set of linearly independent columns of A, such that $\tilde{A}=BB^+A$ is ε, δ-stable with respect to b, *and* $\|(I-BB^+)\,A\|=\|A-\tilde{A}\|$ is minimum'.

This is another way of saying that, since the problem of finding minimal least squares solutions might be ill-conditioned, the user should supply the additional parameters ε, δ in order to make the problem well posed.

Observe that if B contains all the linearly independent columns of A then $\tilde{A}\equiv A$ and this will mean that A itself is ε, δ-stable and no modification of the given matrix is necessary in order to satisfy the requirements of the problem. Since the two conditions involved in Problem I are opposite, i.e. less columns will probably make $\tilde{A}$ stable but will make the representation of A worse, it is clear that in this way we are trying to modify automatically the basic model A in order to make it stable, while simultaneously trying to introduce the minimum violence compatible with this purpose.

The method used in the analysis is non other than Wilkinson's 'backward error analysis' (see [13]).

The kinds of perturbations we allow are those that preserve the rank of A. In the course of the work we find a bound for

$$\frac{\|(E + \Delta E)^+ - E^+\|}{\|E^+\|}$$

---

[1]) The stability concepts introduced here are, of course, relative to the problem we are solving.

where E is a full rank matrix. Stewart [12] produces an L_2 bound of this type for a general matrix A which does not seem to be comparable with ours. Ben Israel [1] obtains bounds for

$$\frac{\|x - \hat{x}\|}{\|x\|}$$

in the case in which the equation $Ax = b$ is solvable (i.e. $b \in \text{range}(A)$), and the perturbations ΔA are restricted not only in size but also to belong to a subspace which depends upon the matrix A. In this case he is able to obtain bounds which parallel the classical ones for nonsingular square matrices (cf. [3, 13]). Bjorck [15] and Golub and Wilkinson [14] also obtain bounds for the full rank case.

2. Notation and Definitions

The vector and matrix norms considered in this paper are any compatible, symmetric norms. That is, norms satisfying

$$\|Ax\| \leqslant \|A\| \, \|x\|$$

for any matrix-vector pair for which the operation makes sense, and also with the property $\|A\| = \|A^*\|$.[2])

With regard to this last property we mention that if $\|.\|$ is any matrix norm then $|||A||| = \max(\|A\|, \|A^*\|)$ is a symmetric norm. In any case the use of this type of norm is not essential to obtain the results we are interested in, but makes the bounds and intermediate computations simpler.

Some of the definitions below will be norm dependent. Unless it becomes necessary to clarify the context we will use but one symbol for the defined objects and we will presume that a fixed norm is used throughout any given process.

DEFINITIONS. For a general $m \times n$ matrix $A(m \geqslant n)$ we define:

(a) A^+ the *Moore-Penrose generalized inverse* or *pseudoinverse* of A; A^+ is the unique $n \times m$ matrix which satisfies: (i) $AA^+A = A$, (ii) $A^+AA^+ = A^+$, (iii) $(AA^+)^* = AA^+$, (iv) $(A^+A)^* = A^+A$ [8].

(b) $\mathscr{R}(A) = \{y : y = Ax, x \in R_n\} \subset R_m$ the *range subspace* of A;

(c) $pk(A) = \|A^+\| \, \|A\| = pk(A^*)$ the *pseudocondition number* of A;

(d) $P_A = AA^+$ the *orthogonal projection on the range subspace* of A;

(e) $\widehat{Ab}$ the angle between a vector b and $\mathscr{R}(A)$;

(f) We shall say that ΔA is a *perturbation* on A if the dimensions of the matrix ΔA are equal to those of A and iff $\text{rank}(A + \Delta A) = \text{rank}(A)$.

---

[2]) For any matrix A, A^* is the conjugate transpose or adjoint of A.

For an arbitrary matrix A and a perturbation ΔA we define the following functionals:

(g) $\omega(A)=pk(A)\dfrac{\|\Delta A\|}{\|A\|}$ the *amplified relative perturbation*;

(h) $\eta(A)=1+\omega(A)$;

(i) $\beta(A)=\omega(A)(\eta(A)+pk(A))$ the *regularity factor*;

(j) If $\dfrac{\|\Delta A\|}{\|A\|}\leqslant\delta$

$$\alpha(A)=\delta pk(A)[1+pk(A)(\delta+1)].$$

We recall some useful well known properties:

(k) A is a full column rank matrix iff A^*A is nonsingular. Furthermore we have in this case that

$$A^+=(A^*A)^{-1}A^*;$$

(l) A is a full row rank matrix iff AA^* is nonsingular. Also

$$A^+=A^*(AA^*)^{-1};$$

(m) If A is an arbitrary matrix and B is a subset of linearly independent columns of A that expand $\mathscr{R}(A)$ (i.e. B has full column rank$=$rank(A)) then

$$A=BC$$

where $C=B^+A$ has full row rank.

Proof. Assume that B is formed with the first columns of A, and call $\bar{B}$ the matrix of the remaining columns; thus in block form we have $A=(B,\bar{B})$. Since the columns of $\bar{B}$ are linearly dependent on the columns of B we can write $\bar{B}=B\tilde{B}$ for some matrix $\tilde{B}$. Thus

$$BB^+A=(BB^+B,BB^+B\tilde{B})=(B,B\tilde{B})=(B,\bar{B})=A.$$

Also $C=B^+A=(I,\bar{B})$ since $B^+B=I$. Thus rank$(C)=$rank$(A)=$rank$(B)=$number of rows of C.

(n) $(A^+)^*=(A^*)^+$.

3. Perturbation of the Right Hand Side

We shall now study the effect that perturbations on the right hand side b have on the minimal least squares solution.

LEMMA 3.1. *Let A be an $m\times n$ matrix, b and k m-vectors. Let $x=A^+b$. If $A\neq 0$ and b is not orthogonal to $\mathscr{R}(A)$ then the h that gives the minimal least squares solution* *to* $A(x+h)=b+k$ *satisfies*

$$\frac{\|h\|}{\|x\|} \leqslant pk(A) \frac{\|k\|}{\|P_A b\|}. \tag{3.1}$$

Proof. From its definition we obtain

$$h = A^+ k,$$

and therefore

$$\|h\| \leqslant \|A^+\| \, \|k\|. \tag{3.2}$$

On the other hand

$$Ax = AA^+ b = P_A b$$

and

$$\|P_A b\| \leqslant \|A\| \, \|x\|$$

which gives

$$\frac{\|P_A b\|}{\|A\|} \leqslant \|x\|. \tag{3.3}$$

Combining (3.2) and (3.3) we get

$$\frac{\|h\|}{\|x\|} \leqslant \|A\| \, \|A^+\| \frac{\|k\|}{\|P_A b\|},$$

which is (3.1).

Of course (3.2) is valid even if b is orthogonal to $\mathscr{R}(A)$, i.e. $AA^+b=0$.

4. Continuity of $(.)^+b$. Full Rank Case

In this paragraph we shall prove that, for full rank matrices, $(.)^+b$ is continuous, and moreover that those matrices are strongly ε-stable for suitable ε's (and δ_0) that we shall exhibit.

First of all we produce a representation for the pseudoinverse of a perturbed, full rank matrix.

LEMMA 4.1. *Let* E *be a full rank matrix and* ΔE *a perturbation on it. Let*

$$\Omega = E^+ \{\Delta E + E^{*+} \Delta E^* (E + \Delta E)\}, \tag{4.1}$$

$$\Lambda = \{\Delta E + (E + \Delta E) \, \Delta E^* E^{*+}\} \, E^+. \tag{4.2}$$

If E *is full column rank then* $(I+\Omega)$ *is nonsingular and*

$$(E + \Delta E)^+ = (I + \Omega)^{-1} E^+ (I + E^{*+} \Delta E^*). \tag{4.3}$$

If E *is full row rank then* $(I+\Lambda)$ *is nonsingular and*

$$(E + \Delta E)^+ = (I + \Delta E^* E^{*+}) \, E^+ (I + \Lambda)^{-1}. \tag{4.4}$$

Proof. Let E then be full column rank and let ΔE be a perturbation on E. Thus $(E+\Delta E)$ has full column rank (cf. § 2, (f)). Then $(E+\Delta E)^*(E+\Delta E)$ is nonsingular (cf. § 2, (k)), and we can write

$$\begin{aligned}(E + \Delta E)^* (E + \Delta E) &= E^*E + E^*\Delta E + \Delta E^* (E + \Delta E) \\ &= (E^*E) [I + E^+\Delta E + (E^*E)^{-1}\Delta E^* (E + \Delta E)] = (E^*E) (I + \Omega), \qquad (4.5)\end{aligned}$$

where we used the identity $E^+E^{*+} = (E^*E)^{-1}E^*E(E^*E)^{-1} = (E^*E)^{-1}$.

Therefore $I+\Omega$ is nonsingular and

$$[(E + \Delta E)^* (E + \Delta E)]^{-1} = (I + \Omega)^{-1} (E^*E)^{-1}.$$

Finally

$$(E + \Delta E)^+ = (I + \Omega)^{-1} (E^*E)^{-1} (E^* + \Delta E^*) = (I + \Omega)^{-1} E^+ (I + E^{*+} \Delta E^*)$$

which is (4.3).

Let E be a full row rank matrix and ΔE a perturbation on E. If we apply the first part of the Theorem to the full column rank matrix E^* and the corresponding perturbation ΔE^*, then by taking adjoints in the final result (4.4) follows. It is enough to observe that

$$\Lambda(E) = [\Omega(E^*)]^*. \qquad (4.6)$$

LEMMA 4.2. *Let E be a full rank matrix and ΔE an arbitrary matrix of the same dimensions as E. If*

$$\beta(E) < 1 \qquad (4.7)$$

then ΔE is a perturbation on E.

Proof. First of all, let E be a full column rank matrix. From (4.5) we deduce that $(E+\Delta E)$ has full column rank iff $(I+\Omega)$ is nonsingular. But $(I+\Omega)$ is nonsingular if $\|\Omega\| < 1$.

We shall show now that $\|\Omega\| \leqslant \beta(E)$ (thus the name of regularity factor we gave to $\beta(E)$).

$$\begin{aligned}\|\Omega\| &= \|E^+ \{\Delta E + E^{*+}\Delta E^* (E + \Delta E)\}\| \\ &\leqslant \|E^+\| \, \|E\| \left\{\frac{\|\Delta E\|}{\|E\|} + \|E^{*+}\| \, \|E^*\| \frac{\|\Delta E^*\|}{\|E^*\|} \left(1 + \frac{\|\Delta E\|}{\|E\|}\right)\right\} \\ &= \omega(E) (\eta(E) + pk(E)) = \beta(E).\end{aligned}$$

If E is of full row rank then a similar argument shows that for $(E+\Delta E)$ to be nonsingular it is sufficient that $\|\Lambda\| < 1$, and that also in this case $\|\Lambda\| \leqslant \beta(E)$. Therefore (4.7) implies, in both the full column and the full row rank cases, that $(E+\Delta E)$ is nonsingular and consequently that ΔE is a perturbation.

LEMMA 4.3. *Let E, ΔE, and $\beta(E)$ be as in Lemma 4.2. Then*

$$\frac{\|(E+\Delta E)^+ b - E^+ b\|}{\|E^+ b\|} \leqslant \frac{1}{1-\beta(E)}\left(\beta(E) + \omega(E)\frac{\|E^+\|\ \|b\|}{\|E^+ b\|}\right). \tag{4.8}$$

Proof. Again, we have to consider the full column and full row rank cases separately. Let E be a full column rank matrix. According to (4.3) and using the Neumann series for $(I+\Omega)^{-1}$ we get

$$(E+\Delta E)^+ b - E^+ b = \sum_{j=1}^{\infty} (-\Omega)^j E^+ b + \sum_{j=0}^{\infty} (-\Omega)^j E^+ E^{*+} \Delta E^* b$$

thus

$$\frac{\|(E+\Delta E)^+ b - E^+ b\|}{\|E^+ b\|} \leqslant \frac{1}{1-\beta(E)}\left\{\beta(E) + \|E^+\|\ pk(E)\frac{\|\Delta E\|}{\|E\|}\ \frac{\|b\|}{\|E^+ b\|}\right\}.$$

A similar argument using the representation (4.4) gives the result for the full row rank case.

COROLLARY 4.3. *Under the hypotheses of Lemma 4.3 we have that*

$$\frac{\|(E+\Delta E)^+ - E^+\|}{\|E^+\|} \leqslant \frac{\beta(E)+\omega(E)}{1-\beta(E)}.$$

Now we can prove the main result of this Section.

THEOREM 4.4. *Let E be an $m \times n$ full rank matrix and b an m-vector non orthogonal to $\mathscr{R}(E)$. Let*

$$\delta_0(E) = \frac{1}{2\,pk(E)}\{\sqrt{(1+pk(E))^2+4} - (1+pk(E))\}, \tag{4.9}$$

$$\alpha(E) = pk(E)\,\delta\{1 + pk(E)(\delta+1)\}. \tag{4.10}$$

For any $\delta < \delta_0$ let ΔE be any $m \times n$ matrix satisfying

$$\frac{\|\Delta E\|}{\|E\|} \leqslant \delta,$$

and let

$$\varepsilon(\delta) = \frac{1}{1-\alpha}\{\alpha + \delta pk(E)^2 \cos^{-1}(\widehat{Eb})\}, \quad (0 < \delta < \delta_0).\ [3] \tag{4.11}$$

Then E is strongly ε-stable at b.

Proof. From (i) in §2 and (4.9) we see that for any $\delta < \delta_0$ and ΔE as specified, the corresponding $\beta(E) < 1$. Therefore Lemma 4.3 is applicable. Since

[3] Here $\cos^{-1}(Eb)$ is short hand for $\frac{\|b\|}{\|EE^+b\|}$. If the norm used is the Euclidean norm then this coincides with the standard trigonometrical notion.

$\|EE^+b\| \leqslant \|E\| \, \|E^+b\|$, then
$$\frac{\|EE^+b\|}{\|E\|} \leqslant \|E^+b\|$$

and from (4.8) we get that
$$\frac{\|(E+\Delta E)^+ b - E^+ b\|}{\|E^+ b\|} \leqslant \frac{1}{1-\beta(E)} \left\{ \beta(E) + pk(E)^2 \frac{\|\Delta E\|}{\|E\|} \cos^{-1}(\widehat{Eb}) \right\}, \tag{4.12}$$

or replacing
$$\frac{\|\Delta E\|}{\|E\|} \quad \text{by} \quad \delta$$
$$\frac{\|(E+\Delta E)^+ b - E^+ b\|}{\|E^+ b\|} \leqslant \frac{1}{1-\alpha} \{\alpha + \delta pk(E)^2 \cos^{-1}(\widehat{Eb})\} = \varepsilon(\delta)$$

as we wanted to prove.

Observe that in the full row rank case $\cos^{-1}(\widehat{Eb}) = 1$.

5. The General Case

Let A be an arbitrary $m \times n$ matrix and ΔA a perturbation on it. Let B be a full rank matrix containing *all* the independent columns of A. Let $C = B^+ A$; from (*m*) §2 we know that $A = BC$. Let ΔB be a matrix formed from ΔA by taking the same columns as those taken from A to form B. Let
$$\max \left(\frac{\|\Delta B\|}{\|B\|}, \frac{\|\Delta A\|}{\|A\|} \right) \leqslant \delta .$$

Let $\beta(B) < 1$, so ΔB is a perturbation on B. Let ΔC be defined by means of the identity
$$C + \Delta C = (B + \Delta B)^+ (A + \Delta A). \tag{5.1}$$

(Observe that from the hypotheses it follows that $A + \Delta A = (B + \Delta B)(C + \Delta C)$.)

Now we want to find a bound for the relative perturbation
$$\frac{\|\Delta C\|}{\|C\|} .$$

LEMMA 5.1. *Let A, B, C, ΔA, ΔB, ΔC, δ be as above. If $\alpha(B) < 1$ then*
$$\frac{\|\Delta C\|}{\|C\|} \leqslant \frac{2\alpha(B)}{1-\alpha(B)} \equiv \gamma . \tag{5.2}$$

Proof. $\Delta C = (B + \Delta B)^+ (A + \Delta A) - C =$
$$[(I+\Omega)^{-1} - I]\, C + (I+\Omega)^{-1} B^+ [B^{*+} \Delta B^* (A + \Delta A) + \Delta A] .$$

Thus

$$\frac{\|\Delta C\|}{\|C\|} \leqslant \frac{1}{1-\beta(B)}\left\{\beta(B) + \frac{\|B^+\| \, \|A\|}{\|C\|}\,(pk(B)\,\delta(1+\delta)+\delta)\right\}.$$

Since $\beta(B) \leqslant \alpha(B)$ (5.2) follows.

We shall assume now that δ is so restricted as to ensure that $\beta(C) < 1$. Thus Lemma 4.2 tells us that $C + \Delta C$ has full row rank. As we saw in Theorem 4.4 this will be satisfied if

$$\frac{\|\Delta C\|}{\|C\|} \leqslant \gamma < \frac{1}{2\,pk(C)}\{\sqrt{[1+pk(C)]^2+4} - [1+pk(C)]\} \equiv \delta' \tag{5.3}$$

We want to use the decomposition $A = BC$ (and $A + \Delta A = (B + \Delta B)(C + \Delta C)$) in order to apply our results on the full rank case to this general problem. To do that we proceed in two stages. Firstly we set the following least squares problem:

$$(B + \Delta B)(y + k) = b \tag{5.4}$$

where as usual $y = B^+ b$.

Applying Theorem 4.4 to B, ΔB we see that

$$\frac{\|k\|}{\|y\|} \leqslant \frac{1}{1-\alpha(B)}\{\alpha(B) + \delta pk(B)^2 \cos^{-1}(\widehat{Bb})\}. \tag{5.5}$$

Now we can set out a second least squares problem:

$$(C + \Delta C)(x + h) = y + k$$

where $x = C^+ B^+ b$, and this will give us the desired h. First we get

$$h = (C + \Delta C)^+ (y + k) - C^+ B^+ b$$

and therefore

$$h = [(C + \Delta C)^+ - C^+]\, B^+ b + (C + \Delta C)^+ k.$$

Taking norms and dividing by $\|x\|$ we get

$$\frac{\|h\|}{\|x\|} \leqslant \frac{\|[(C+\Delta C)^+ - C^+]\, B^+ b\|}{\|C^+ B^+ b\|} + \|C\| \, \|(C+\Delta C)^+\| \, \frac{\|k\|}{\|y\|}, \tag{5.6}$$

where we have used the inequality

$$\frac{\|y\|}{\|C\|} \leqslant \|x\|,$$

which is easily obtained from $Cx = B^+ b = y$. From (5.6) we now can prove the following result.

THEOREM 5.2. *Let A be an arbitrary $m \times n$ matrix, b an m-vector, B a maximal set of linearly independent columns of A, $C = B^+A$. Let ΔA be an $m \times n$ matrix perturbation associated with A, ΔB the one corresponding to B as deduced from ΔA, and $\Delta C = (B + \Delta B)^+(A + \Delta A) - C$. If*

$$\max\left(\frac{\|\Delta A\|}{\|A\|}, \frac{\|\Delta B\|}{\|B\|}\right) \leqslant \delta$$

is so restricted as to make $\alpha(B) < 1$, and (5.3) *is valid then A is ε-stable for*

$$\varepsilon(\delta) = \frac{1}{1 - \alpha(C)}\left\{\gamma pk(C)^2 + \alpha(C) + \frac{pk(C)\,\eta(C)}{1 - \alpha(B)}\left[\alpha(B) + \delta pk(B)^2 \cos^{-1}(\widehat{Bb})\right]\right\}. \tag{5.7}$$

Proof. The ε-stability of A, according to the def. in § 1 will be proved if for any perturbation ΔA on A, with norm less than $\delta \cdot \|A\|$, it follows that the h defined by $x + h = (A + \Delta A)^+ b$ $(x = A^+ b)$ satisfies $\|h\| \leqslant \varepsilon(\delta)\,\|x\|$.

If in Theorem 4.4 we replace E by C and b by B^+b and then put the bound so obtained in (5.6) we get

$$\frac{\|h\|}{\|x\|} \leqslant \frac{1}{1 - \alpha(C)}(\alpha(C) + \gamma pk(C)^2) + \|C\|\;\|(C + \Delta C)^+\|\;\frac{\|k\|}{\|y\|}.$$

Using (5.5) and the representation for $(C + \Delta C)^+$ in this last expression we get

$$\frac{\|h\|}{\|x\|} \leqslant \frac{1}{1 - \alpha(C)}(\alpha(C) + \gamma pk(C)^2) + \frac{\|C\|\;\|C^+\|}{1 - \alpha(C)}\,\eta(C)$$
$$\times \frac{1}{1 - \alpha(B)}\{\alpha(B) + \delta pk(B)^2 \cos^{-1}(\widehat{Bb})\} = \varepsilon(\delta),$$

as desired.

Observe that the whole expression is $0(\delta)$. The terms that are linear in δ and contain a factor of the form $pk(B)^2$ will be dominant in the case of ill conditioning. It is clear from its definition that C is well conditioned (all its singular values are 0 or 1), thus $pk(C)$ is not a troublesome quantity.

6. Applications

(a) *The Computation of Minimal Least Squares Solutions of Systems with Undetermined Rank*

We shall consider now how to apply the results of § 5 to the solution of Problem I (cf. § 1).

In [10] (see also [7, Chap. 2.4, §2a.1, pp. 369–372]) a method due to Rosen [11] for computing the pseudoinverse of a matrix was implemented as an Algol 60 program. Certain modifications introduced into the original procedure were implicitly directed toward the solution of Problem I. The difficulty there was that the user had to provide some parameters which were, in principle, unrelated to the problem he wanted to solve, i.e. to find minimal least squares solutions. Well conditioned problems, as usual, did not create troubles. Our claim is that ill-posed problems of the form (1.1) are to be modified *before* any attempt is made to solve them. How to modify them depends, of course, on the user needs. What we propose in setting Problem I is in fact an automatic way of modifying ill designed models in order to eliminate highly correlated variables. The aim is to obtain answers which will behave smoothly with respect to the data, and to do that with minimum violence to the original model. The algorithm we propose now for solving Problem I can be implemented by introducing minor modifications to the computer program in [10].

Algorithm. A matrix B is constructed by considering the columns of A one at a time. Recalling the statement of Problem I, let ε, δ be the stability parameters. We shall admit a column of A in the base B only if $\tilde{A}=BB^+A$ is ε, δ-stable. To implement this check, let us assume that B_q is a basis for which the corresponding $\tilde{A}_q$ is ε, δ-stable, and let u be a column of A, not in B_q, which is under examination. Let also $r_q=\|A-\tilde{A}_q\|$. We shall provisionally form $B_{q+1}=(B_q, u)$, $C_{q+1}=B^+_{q+1}A$, C^+_{q+1}, and test $\tilde{A}_{q+1}=$ $=B_{q+1}C_{q+1}$ for ε, δ-stability according to (5.7). Of course it is assumed that u is not *exactly* linearly dependent on the columns of B_q (in fact the limitation is overflow of the computer in the computation of B^+_{q+1}), otherwise it is rejected a priori without further questioning.

If $\tilde{A}_{q+1}$ is ε, δ-stable then we compute r_{q+1} and check if $r_{q+1}<r_q$; if so then u is accepted and B_{q+1} replaces B_q, otherwise u is rejected. If $\tilde{A}_{q+1}$ does not satisfy the condition (5.7) then u is also rejected.

In all cases, the next step is to examine the following column in A and so on, until all the columns have been exhausted and a basis B satisfying the requirements of Problem I have been found.

Observe that the choice of a particular basis B depends upon the ordering of the columns of A (or rather, on the order in which they are taken for examination), but that in any case only one cycle is necessary since a column which has been rejected when compared with a certain set of columns B' will certainly be rejected if compared with any set $B\supset B'$.

Let us call u a column of A that makes $\cos^{-1}(\widehat{bu})$ minimum. Assume for simplicity that $\|u\|=1$. Then,

$$u^+ = \frac{\|u^*\|}{\|u\|^2} = 1 ,$$

and $pk(u)=1$. Thus, a sufficient condition for Problem I to have a solution is that

$$\delta\cdot\left[\frac{2+\cos^{-1}(\widehat{bu})+\delta}{1-2\delta-\delta^2}\right]\leqslant\varepsilon$$

where $\delta<\sqrt{2}-1$; this is so because under those conditions $B\equiv u$ is ε, δ-stable (see Theorem 4.4).

(b) *Estimation of the Error Caused by a Model Modification.*

There are applications in which the dimensions of the matrix A are very large. In some cases the phenomenon under study allows one to modify A in such a way that the new model does not differ much from the original one but it has a simpler structure. For instance, in crystallographic computations it is known that in certain cases the normal equations of the linearized model are approximately block diagonal near the solution. What is more important, the size of the blocks is small compared to that of the original problem, where 200 parameters is not an unusual occurrence (cf. [9, specially paper 17, pp. 170–187 by R. A. Sparks] for more details).[4])

It is important in these cases to be able to estimate the effect of this modification on the computed least squares solution, and also the effect on the statistical quantities related to the elements of the inverse of the matrix of normal equations $A^T A$.

For this application the $m\times n$ matrix A will have full column rank and we shall then use the results of §4. Let then A and $\tilde{A}$ be $m\times n$ full column rank matrices, $\Delta A=A-\tilde{A}$ be the perturbation, and b the right hand side vector. Let $x=A^+b$, $\tilde{x}=\tilde{A}^+b$. We want to estimate

$$\frac{\|h\|}{\|\tilde{x}\|}=\frac{\|x-\tilde{x}\|}{\|\tilde{x}\|}$$

in terms of

$$\frac{\|\Delta A\|}{\|\tilde{A}\|}\leqslant\delta,\quad pk(\tilde{A}),$$

and the other data of the problem. But this is exactly what we shall obtain from (4.12) in Theorem 4.4 if we put $\tilde{A}$ instead of E and ΔA instead of ΔE. Recalling the definition of $\alpha(\tilde{A})=pk(\tilde{A})\,\delta\,[1+pk(\tilde{A})\,(\delta+1)]$, and that $\beta(\tilde{A})\leqslant\alpha(\tilde{A})$ we can write this result in the following way

$$\frac{\|x-\tilde{x}\|}{\|\tilde{x}\|}\leqslant\frac{\delta pk(\tilde{A})}{1-\alpha(\tilde{A})}\{1+pk(\tilde{A})\,[\delta+\cos^{-1}(\widehat{\tilde{A}b})+1]\}\tag{6.1}$$

provided that $\alpha(\tilde{A})<1$.

From the decomposition (4.5) it follows that

$$\frac{\|(A^*A)^{-1}-(\tilde{A}^*\tilde{A})^{-1}\|}{\|(\tilde{A}^*\tilde{A})^{-1}\|}\leqslant\frac{\alpha(\tilde{A})}{1-\alpha(\tilde{A})}.\tag{6.2}$$

The inequality (6.2) is nothing but a bound for the relative variation of the variance-covariance matrix for the derived parameters (cf. [6, Ch. 4]), due to the perturbation ΔA.

---

[4]) I am indebted to Professor L. Becka, Chemistry Dpt. of the U. Central de Venezuela, for bringing up this application and sharing with me his ample experience on the problem.

REFERENCES

[1] Ben Israel, A., *On Error Bounds for Generalized Inverses*, SIAM J. Numer. Anal. *3*, 585–592 (1966).

[2] Cho, C. Y., Cline, R. E. *et al.*, *Talks on Generalized Inverses and Solutions of Large, Approximately Singular Linear Systems*, MRC Tech. Rep. 644, April 1966, The University of Wisconsin, Madison.

[3] Forsythe, G., and Moler, C., *Computer Solution of Linear Algebraic Equations* (Prentice Hall, New Jersey 1967).

[4] Golub, G., *Least Squares, Singular Values and Matrix Approximation*, and P. Businger, *An Algol Program for Computing the Singular Value Decomposition*, Stanford University Tech. Rep. CS73, July 1967.

[5] Golub, G., and Kahan, W., *Calculating the Singular Values and Pseudoinverse of a Matrix*, J. SIAM Numer. Anal. *2*, 205–224 (1965).

[6] Hamilton, W. C., *Statistics in Physical Sciences* (Ronald Press Co., New York 1964).

[7] Korganoff, A., and Pavel-Parvu, M., *Elements de Théorie des Matrices Carrées et Rectangles en Analyse Numérique* (Dunod, Paris 1967).

[8] Penrose, R., *A Generalized Inverse for Matrices*, Proc. Camb. Philos. Soc. *51*, 406–413 (1955).

[9] Pepinsky, R., Robertson, J. M. and Speakman, J. C. (Editors), *Computing Methods and the Phase Problem in X-Ray Crystal Analysis* (Pergamon Press, Oxford 1961).

[10] Pereyra, V., and Rosen, J. B., *Computation of the Pseudoinverse of a Matrix of Unknown Rank*, Stanford University Tech. Rep. CS13, Sept. 1964, CR *6* 7948.

[11] Rosen, J. B., *Minimum and Basic Solutions to Singular Linear Systems*, J. SIAM *12*, 156–162 (1964).

[12] Stewart, G. W., *On the Continuity of the Generalized Inverse*, Notices AMS, Abstract 650–54, p. 932 (November 1967). To appear in SIAM J. Numer. Anal.

[13] Wilkinson, J. H., *Rounding Errors in Algebraic Processes* (Prentice Hall, New Jersey 1963).

[14] Golub, G. H., and Wilkinson, J. H., *Note on the Iterative Refinement of Least Squares Solution*, Numer. Math. *9*, 139–148 (1966).

[15] Björck, A., *Solving Linear Least Squares Problems by Gram–Schmidt Orthogonalization*, BIT *7*, 1–21 (1967).

Universidad central de Venezuela

A Less Formal Approach to Kaluza–Klein Formalism

M. A. McKiernan (Waterloo, Ont., Canada)

Dedicated to Professor A. Ostrowski on the occasion of his 75th birthday

Abstract

The 'action' integrals (a) $\lambda(\tau_1)=\int_{\tau_0}^{\tau_1}\sqrt{g_{ij}\dot{y}^i\dot{y}^j}\,d\tau$ and (b) $\lambda(\tau_1)=\int_{\tau_0}^{\tau_1}\{\sqrt{h_{ij}\dot{x}^i\dot{x}^j}-B_i\dot{y}^i\}\,d\tau$, corresponding respectively to gravitational and gravitational-electromagnetic phenomena, are shown to be related under continuous groups of null translations. This relation motivates a modified Kaluza–Klein formalism for which the classical cylindrical metric preserving transformations (c) $y^5=x^5+f^5(x^j)$, $y^i=f^i(x^j)$ for $i=1,2,3,4$ are replaced by (d) $y^5=x^5$, $y^i=f^i(x^j,x^5)$. *The cylindrical metric of* V^5 *is nevertheless preserved under* (d), since it is assumed that V^5 admits a metric of the form $(\dot{y}^5)^2-g_{ij}(y^k)\,\dot{y}^i\dot{y}^j$ (corresponding to (a)) and that (d) defines a continuous group of null translations in the V^4 metric defined by g_{ij} when x^5 is considered the group parameter. Application of (d) leads to the cylindrical metric $(\dot{x}^5+B_i\dot{x}^i)^2-h_{ij}\dot{x}^i\dot{x}^j$ corresponding to (b). The resulting electromagnetic fields $F_{ij}=B_{i,j}-B_{j,i}$ are then related to the curvatures of the V^4 corresponding to g_{ij} and h_{ij}; in particular it is shown that $B_i B_j \underset{g}{R}^{ij}=-\frac{1}{4}F_{ij}F^{ij}$ and $F^{ij}_{\cdot;j}=B_j\underset{g}{R}^{ij}$. When $\underset{g}{R}_{ij}=0$ it is shown that F_{ij} is a null electromagnetic field which is generally non-trivial. Some physical and geometric interpretations of the mathematical results are also presented.

1. Introduction[1]

The 'action' integral

$$\lambda(\tau_1)=\int_{\tau_0}^{\tau_1}\sqrt{g_{ij}(y)\,\dot{y}^i\dot{y}^j}\,d\tau+\lambda(\tau_0),\quad i=1,2,3,4, \tag{1}$$

of general relativity is parameter invariant; when the parameter is proper time, $g_{ij}\dot{y}^i\dot{y}^j=1$, whence $\lambda(\tau_1)=\tau_1+k$. Hence λ in (1) is proper time for any choice of parameter τ.

Equation (1) may be written in the form

$$\dot{\lambda}^2=g(y;\dot{y},\dot{y}),\qquad \dot{\lambda}^2=g_{ij}(y^k)\,\dot{y}^i\dot{y}^j, \tag{2}$$

where $g(y;\xi,\eta)$ is a bilinear form in ξ,η. In the case of special relativity where $g_{kj}=-\delta_{kj}$ $(y^4=i\,c\,t)$, the transformations $y=\phi(x)$ which leave the metric tensor invariant are the Lorentz transformations; equivalently, a map $\phi\colon x\to y$ with inverse

[1]) For the first half of the paper we give both the index notation and its translation into the modern idiom. However, since the later half is more closely related to classical applied mathematics, we will use only index notation leaving its translation to the interested reader.

Received December 4, 1967

map $\phi^{-1}: y \to x$ is a Lorentz transformation if and only if the Lie difference of the metric tensor $-\delta_{ij}$, with respect to this map, vanishes. The y frame of reference (or, preferably, the frame $\bar{y}$ 'carried along' by ϕ) is then attributed to an observer in 'uniform motion' with respect to an observer attached to the x frame of reference.

This concept readily generalizes mathematically for the case of an arbitrary metric tensor g_{ij}; a map $\phi: x \to y$ will be called a *general Lorentz transformation with respect to* g_{ij} if the Lie difference of g_{ij} with respect to ϕ vanishes. If the x frame is thought of as that of an observer A whose world line is C_A, the y (or $\bar{y}$) frame may be interpreted as the reference frame of an observer B whose world line is C_B where $\phi: C_A \to C_B$. Physical assertions based solely on the metric must be the same for observers A and B since ϕ is metric preserving; in particular, if C_A is a geodesic, so also is C_B. Two such observers will be called *comparable*, but unlike the special relativistic case this in no way implies any concept of 'uniform motion' of A relative to B or vice versa. Indeed, the analysis indicates that a comparable observer may well appear to be accelerating.

In the metric of (1) or (2) consider a one parameter family of general Lorentz transformations, σ the parameter,

$$y = \phi_\sigma(x),\ x = \phi_\sigma^{-1}(y); \qquad y^i = \phi_\sigma^i(x^j),\ x^i = \overset{-1}{\phi}{}_\sigma^i(y^j), \tag{3}$$

which form a (local) group of transformations so that

$$\left.\begin{aligned} &\mathring{\phi}_\sigma(\varphi_\sigma^{-1}(y)) = B(y); && \mathring{\varphi}_\sigma^{-1}\{\phi_\sigma(x)\} = \bar{B}(x), \\ &\frac{dy^i}{d\sigma} = B^i(y^j); && \frac{dx^i}{d\sigma} = \bar{B}^i(x^j), \end{aligned}\right\} \tag{4}$$

where $\mathring{\varphi}_\sigma$ denotes differentiation of ϕ_σ with respect to σ. (We use ϕ_σ^{-1} rather than $\phi_{-\sigma}$ for convenience.) Under the map ϕ_σ let $\hat{\varphi}_\sigma(x)$ denote the induced map for the tangent spaces; that is, ξ in the tangent space at x is mapped to $\hat{\varphi}_\sigma(x)\,\xi$ in the tangent space at $\phi_\sigma(x)$; hence $\hat{\varphi}_\sigma(x)$ corresponds to the matrix formed from the partial derivatives $\partial y^i/\partial x^j$. Since ϕ_σ is a general Lorentz transformation for every σ, clearly (1) and (2) are invariant under ϕ_σ for each fixed σ. More precisely $g\{\phi_\sigma(x);\ \hat{\varphi}_\sigma(x)\,\xi,\ \hat{\varphi}_\sigma(x)\,\eta\} \equiv g\{\phi_0(x);\ \hat{\varphi}_0(x)\,\xi,\ \hat{\varphi}_0(x)\,\eta\}$ and *it is assumed that* ϕ_0 *is the identity map.*

Let $y^i(\tau)$ be the world line of a particle, and hence $\lambda(\tau)$ given by (1) or (2) is the proper time for this particle. *We wish to study the form of* (1) *or* (2) *when referred to a general Lorentz frame x, given that this frame of reference varies with the proper time of the particle studied*; hence the parameter σ in (3) is to be related to the proper time λ of the particle.

If this relation is assumed to be of the form $\sigma = \lambda$, and if $y = \phi_\lambda(x)$ is assumed to be not only a general Lorentz transformation group but in fact a group of null translations in the $g_{ij}(y)$ metric, then (1) and (2) are transformed respectively into the forms (details in section 2)

$$\lambda(\tau_1) = \int_{\tau_0}^{\tau_1} \{\sqrt{h_{ij}(x)\,\dot{x}^i\,\dot{x}^j} - B_i(x)\,\dot{x}^i\}\,d\tau + \lambda(\tau_0), \tag{1*}$$

and

$$(\dot{\lambda} + B_i(x)\,\dot{x}^i)^2 = h_{ij}(x)\,\dot{x}^i\,\dot{x}^j, \quad \text{where} \quad h_{ij} = g_{ij} + B_i\,B_j. \tag{2*}$$

But (1*) is the action integral for a charged particle in the electromagnetic field corresponding to $F_{ij}=B_{i,j}-B_{j,i}$. It is further shown that the transition from (1) to (1*) under null translations may be interpreted within the Kaluza–Klein formalism. (See for example [1], [2], [3]).

We briefly outline this formalism for completeness. Consider a V^5 (coordinates y^α for $\alpha=1, \ldots, 5$ ($i, j=1, \ldots, 4$)) which is cylindrical in that the metric ds is given by

$$\left(\frac{ds}{d\tau}\right)^2 = (\dot{y}^5)^2 + \gamma_{5i}(y^k)\,\dot{y}^i\,\dot{y}^5 + \gamma_{ij}(y^k)\,\dot{y}^i\,\dot{y}^j,$$

and the $\gamma_{\alpha\beta}$ are independent of y^5. The y^α frames are the 'preferred' frames, and hence lead to the group of transformations of the form

$$y^5 = x^5 + f^5(x^i), \qquad y^i = f^i(x^j) \tag{*}$$

which preserve the cylindrical metric of the V^5. It has been shown that Maxwell's equations can be derived from a variational principle applied to the scalar density $\sqrt{-g}\,P$ where $g_{ij}=\gamma_{ij}-\gamma_{5i}\,\gamma_{5j}$ and P is the scalar curvature in V^5. Several objections have been raised against this theory. In [1] for example, W. Pauli points out that 'there is however no justification for the particular choice of the five dimensional curvature scalar P as integrand of the action integral, from the standpoint of the restricted group (*) of the cylindrical metric'. Since P is invariant under all coordinate transformations of V^5, it would seem that P lacks particular significance relative to the restricted transformations (*). In [2], M. A. Tonnelat states 'on critiquait, en particulier, le rôle artificiel de la 5^e^ coordonnée que manifeste la nécessité d'une condition cylindrique'; and further 'toute assimilation des $\gamma_{\alpha\beta}$ avec le champ unifié ... ne permettra pas une transformation covariante des composantes-gravitation en compasantes-électromagnétisme et reciproquement' since under (*), the γ_{ij} never contribute to the γ_{5i}.

It will be seen that continuous groups of general Lorentz transformations, in fact, null translations, lead not only to a relation between (1) and (1*) but to a modification of the Kaluza–Klein formalism in which the above quoted critiques are no longer applicable. Specifically, not only is it assumed that V^5 is cylindrical, but also that it admits a cylindrical coordinate system in which $\gamma_{55}=1$ and $\gamma_{5i}=0$. Hence the curvature scalar P depends only on γ_{ij} and coincides with the curvature scalar R

based on the V^4 corresponding to the metric tensor γ_{ij}. The transformations (*) are replaced by the transformations in V^5 which leave y^5, γ_{55} and γ_{ij} invariant:

$$(**)\qquad y^5 = x^5, \quad y^i = \phi^i(x^j, x^5),$$

where $\phi^i(x^j, x^5)$ is a general Lorentz transformation relative to γ_{ij} for every x^5, and where $\partial y^i/\partial x^5$ defines a null vector. Since γ_{55} and γ_{ij} are assumed invariant under (**),

(i) the curvature scalar P in V^5 depends on, and solely on, the only invariants of $\gamma_{\alpha\beta}$ under (**) and coincides with the R of the V^4;
(ii) while $\gamma_{5i}=0$ in one cylindrical frame, γ_{5i} is not invariant under (**), thereby implying a transformation of gravitational components into electromagnetic components and conversely;
(iii) y^5 is identified with the previous λ, the proper time of the particle under consideration, and is therefore not simply an artificial parameter.

We note further that the acceptance of $y^5=\lambda$ as time like in no way contradicts the usual hypothesis of the Kaluza formalism which asserts that the signature of $\gamma_{\alpha\beta}$ is $(---+-)$ (where now y^4 and y^5 are assumed real). Briefly (if not trivially) explained, this results from the position of λ and y on different sides of equation (2).

2. Transition to Action Integral of Charged Particles

We now assume the parameter σ of the previous section to be identified with the proper time λ of the particle. Hence $y=\phi_\lambda(x)$ implies

$$\dot y = \mathring{\phi}_\lambda(x)\,\dot\lambda + \hat{\phi}_\lambda(x)\,\dot x, \qquad \dot y^i = \frac{dy^i}{d\lambda}\dot\lambda + \frac{\partial y^i}{\partial x^j}\dot x^j, \tag{5}$$

along the world line of the particle, and (2) becomes

$$\dot\lambda^2 = g\{\phi_\lambda(x);\mathring{\phi}_\lambda(x),\mathring{\phi}_\lambda(x)\}\,\dot\lambda^2 + 2g\{\phi_\lambda(x);\mathring{\phi}_\lambda(x),\hat{\phi}_\lambda(x)\,\dot x\}\,\dot\lambda + g\{\phi_\lambda(x);\hat{\phi}_\lambda(x)\,\dot x,\hat{\phi}_\lambda(x)\,\dot x\},$$

or alternatively

$$\dot\lambda^2 = g_{rs}\frac{\partial y^r}{\partial\lambda}\frac{\partial y^s}{\partial\lambda}\dot\lambda^2 + 2g_{rs}\frac{\partial y^s}{\partial\lambda}\frac{\partial y^r}{\partial x^i}\dot x^i\,\dot\lambda + g_{rs}\frac{\partial y^r}{\partial x^i}\frac{\partial y^s}{\partial x^j}\dot x^i\,\dot x^j. \tag{6}$$

Since ϕ_λ is a general Lorentz transformation group, the last term of (6) is independent of λ, and is in fact simply $g(x;\dot x,\dot x)$. Conversely, if ϕ_λ reduces to the identity map for some λ_0, and if the last term in (6) is independent of λ, then ϕ_λ is clearly a general Lorentz transformation. We wish to show that if ϕ_λ is a Lorentz transformation group, then the other terms in (6) are also independent of λ.

The following lemmas are well known in the case that ϕ_λ forms a group of transformations. However, both for completeness and for applications when ϕ_λ does not form a group, we present the proofs (which differ little from the classical proofs).

LEMMA 1: *Let $y=\phi_\lambda(x)$, with inverse $x=\phi_\lambda^{-1}(y)$, be a one parameter family of transformations of class C^1 in x, y and λ and define*

$$B(y,\lambda) \overset{\text{df}}{=} \mathring{\phi}_\lambda\{\phi_\lambda^{-1}(y)\}, \quad \bar{B}(x,\lambda) \overset{\text{df}}{=} \mathring{\phi}_\lambda^{-1}\{\phi_\lambda(x)\}. \tag{7}$$

Then

$$\mathring{\phi}_\lambda(x) = -\hat{\phi}_\lambda(x)\,\bar{B}(x,\lambda) \quad \mathring{\phi}_\lambda^{-1}(y) = -\hat{\phi}_\lambda^{-1}(y)\,B(y,\lambda). \tag{8}$$

Proof: Since ϕ_λ and ϕ_λ^{-1} are inverse transformations for each λ,

$$\phi_\lambda\{\phi_\lambda^{-1}(y)\} \equiv y, \quad \phi_\lambda^{-1}\{\phi_\lambda(x)\} \equiv x,$$

and differentiation with respect to λ yields

$$\mathring{\phi}_\lambda\{\phi_\lambda^{-1}(y)\} \equiv -\hat{\phi}_\lambda\{\phi_\lambda^{-1}(y)\}\,\mathring{\phi}_\lambda^{-1}(y),$$
$$\mathring{\phi}_\lambda^{-1}\{\phi_\lambda(x)\} \equiv -\hat{\phi}_\lambda^{-1}\{\phi_\lambda(x)\}\,\mathring{\phi}_\lambda(x),$$

which in view of the definitions of $B(y,\lambda)$ and $\bar{B}(x,\lambda)$ yields

$$\hat{\phi}_\lambda\{\phi_\lambda^{-1}(y)\}\,\mathring{\phi}_\lambda^{-1}(y) = -B(y,\lambda),$$
$$\hat{\phi}_\lambda^{-1}\{\phi_\lambda(x)\}\,\mathring{\phi}_\lambda(x) = -\bar{B}(x,\lambda).$$

Since $\hat{\varphi}_\lambda\{\phi_\lambda^{-1}(y)\}$ is inverse to $\hat{\varphi}_\lambda^{-1}(y)$, (similarly for $\hat{\varphi}_\lambda(x)$) the lemma follows.

LEMMA 2: *Let $x=x(\tau)$ define a curve in terms of the parameter τ, and let $\lambda=\lambda(\tau)$ also depend on the parameter τ. Then the image curve $y(\tau)$ under the transformation ϕ_λ of Lemma 1 has as tangent vector $\dot{y}$ given by*

$$\dot{y} = \hat{\phi}_\lambda(x)\,[\dot{x} - \bar{B}(x,\lambda)\,\dot{\lambda}]. \tag{9}$$

Proof: Equation (9) is simply (5) in view of (8).

LEMMA 3: *If the transformations ϕ_λ of Lemma 1 are such that $g\{\phi_\lambda(x);\hat{\varphi}_\lambda(x)\,\xi, \hat{\varphi}_\lambda(x)\,\eta\}$ is independent of λ, then the vector field $\bar{B}(x,\lambda)$, for each λ, is a Killing vector field in the metric $\mathring{g}(x;\xi,\eta)$ defined by*

$$\mathring{g}(x;\xi,\eta) \overset{\text{def}}{=} g\{\phi_{\lambda_0}(x);\hat{\phi}_{\lambda_0}(x)\,\xi, \hat{\phi}_{\lambda_0}(x)\,\eta\} \quad \textit{for fixed} \quad \lambda_0.$$

Alternatively

$$\overset{*}{\bar{B}}_{i;j}(x,\lambda) + \overset{*}{\bar{B}}_{j;i}(x,\lambda) \equiv 0,$$

where ';' denotes covariant differentiation (for each fixed λ) relative to the induced metric tensor

$$\mathring{g}_{ij}(x) \overset{\text{def}}{=} g_{rs}(\phi_{\lambda_0}(x))\,\frac{\partial\phi_{\lambda_0}^r(x)}{\partial x^i}\,\frac{\partial\phi_{\lambda_0}^s(x)}{\partial x^j},$$

and $\overset{*}{\bar{B}}_i = \mathring{g}_{ij}\, \bar{B}^j$. *If there exists a* λ_0 *such that* ϕ_{λ_0} *is the identity mapping, then* $\overset{*}{g}$ *coincides with* g.

Proof: Since $g_{rs}(\phi_\lambda(x))\, \partial_i\, \phi^r_\lambda(x)\, \partial_j\, \phi^s_\lambda(x)$ is assumed independent of λ (where ϕ^r_λ is the r^{th} component of the map ϕ_λ and where ∂_i denotes the differentiation with respect to x^i), the derivative of this expression with respect to λ contains three terms whose sum vanishes. If each term is treated individually one obtains, in view of (8), (the argument x is temporarily suppressed).

$$g_{rs}(\phi_\lambda)\, \partial_i\, \mathring{\phi}^r_\lambda\, \partial_j\, \phi^s_\lambda = -\, g_{rs}(\phi_\lambda)\, \partial_k\, \phi^r_\lambda\, \partial_j\, \phi^s_\lambda\, \partial_i\, \bar{B}^k - g_{rs}(\phi_\lambda)\, \partial_i\, \partial_k\, \phi^r_\lambda\, \partial_j\, \phi^s_\lambda \bar{B}^k\,,$$

$$g_{rs}(\phi_\lambda)\, \partial_i\, \phi^r_\lambda\, \partial_j\, \mathring{\phi}^s_\lambda = -\, g_{rs}(\phi_\lambda)\, \partial_i\, \phi^r_\lambda\, \partial_k\, \phi^s_\lambda\, \partial_j\, \bar{B}^k - g_{rs}(\phi_\lambda)\, \partial_i\, \phi^r_\lambda\, \partial_j\, \partial_k\, \phi^s_\lambda \bar{B}^k\,,$$

and

$$\frac{\partial g_{rs}(\phi_\lambda)}{\partial y^p}\, \mathring{\phi}^p_\lambda\, \partial_i\, \phi^r_\lambda\, \partial_j\, \phi^s_\lambda = -\, \frac{\partial g_{rs}(\phi_\lambda)}{\partial y^p}\, \partial_k\, \phi^p_\lambda\, \partial_i\, \phi^r_\lambda\, \partial_j\, \phi^s_\lambda \bar{B}^k.$$

Summation on the left must yield zero, while the summation on the right may be evaluated for $\lambda = \lambda_0$ to give

$$0 = -\, \mathring{g}_{kj}(x)\, \partial_i\, \bar{B}^k(x, \lambda) - \mathring{g}_{ik}(x)\, \partial_j\, \bar{B}^k(x, \lambda) - \partial_k\, \mathring{g}_{ij}(x)\, \bar{B}^k(x, \lambda),$$

from which the lemma follows.

If the family of transformations $y = \phi_\lambda(x)$ forms a group, then ϕ_{λ_0} may be chosen as the identity map and $\mathring{g} = g$; also the tangent vector field $\bar{B}(x, \lambda)$ is then independent of λ and denoted simply by $\bar{B}(x)$.

THEOREM 1: *Consider the action integral*

$$\lambda(\tau_1) = \int_{\tau_0}^{\tau_1} \sqrt{g_{ij}(y)\, \dot{y}^i\, \dot{y}^j}\, d\tau + \lambda(\tau_0), \tag{1}$$

and a continuous group of Lorentz transformations $y = \phi_\sigma(x)$ *in which the group parameter* σ *varies with the proper time* λ *of the particle; assume that* σ *and* λ *are related by*

$$\dot{\sigma} = K\, \dot{\lambda}, \quad K \text{ a real constant}.$$

If the tangent vector field $B^i(y)$, *defined by*

$$B(y) = \mathring{\phi}_\sigma(\phi_\sigma^{-1}(y)),$$

is light-like (that is, $g_{ij}\, B^i\, B^j = 0$) *then* (1) *is given by*

$$\lambda(\tau_1) = \int_{\tau_0}^{\tau_1} \{\sqrt{h_{ij}(x)\, \dot{x}^i\, \dot{x}^j} - K\, \bar{B}_i(x)\, \dot{x}^i\}\, d\tau + \lambda(\tau_0) \tag{10}$$

in the x frame of reference, where

$$\bar{B}(x) = \mathring{\varphi}_\sigma^{-1}(\phi_\sigma(x)), \qquad \bar{B}_i = g_{ij}\,\bar{B}^j,$$

and

$$h_{ij}(x) = g_{ij}(x) + K^2\,\bar{B}_i(x)\,\bar{B}_j(x).$$

Further, the covariant components of $\bar{B}(x)$ *are also given by* $\bar{B}_i = h_{ij}\,\bar{B}^j$; *the* $\bar{B}^i$ *satisfy* $h_{ij}\,\bar{B}^i\,\bar{B}^j = 0$.

Proof: $\lambda(\tau_1)$ as given by (1) satisfies the differential equation

$$\dot{\lambda}^2 = g_{ij}(y)\,\dot{y}^i\,\dot{y}^j, \qquad \dot{\lambda}^2 = g(y;\dot{y},\dot{y}), \tag{2}$$

and under the transformation $y = \phi_\lambda(x)$, the tangent vector $\dot{y}$ becomes

$$\dot{y}^i = \frac{\partial y^i}{\partial x^k}\{\dot{x}^k - K\,\bar{B}^k(x)\,\dot{\lambda}\}, \qquad y = \mathring{\varphi}_\sigma(x)\,\{\dot{x} - K\,\bar{B}(x)\,\dot{\lambda}\}, \tag{11}$$

by lemma 2. Since $y = \phi_\sigma(x)$ is a Lorentz transformation for every σ,

$$\left.\begin{aligned} g_{ij}(\phi_\sigma(x))\,\frac{\partial y^i}{\partial x^r}\frac{\partial y^j}{\partial x^s} &= g_{rs}(x), \\ g\{\phi_\sigma(x);\,\mathring{\varphi}_\sigma(x)\,\xi,\,\mathring{\varphi}_\sigma(x)\,\eta\} &= g(x;\xi,\eta), \end{aligned}\right\} \tag{12}$$

and substitution of (11) in (2) yields

$$\left.\begin{aligned} \dot{\lambda}^2 &= g_{rs}(x)\,\dot{x}^r\,\dot{x}^s - 2\,K\,g_{rs}(x)\,\dot{x}^r\,\bar{B}^s(x)\,\dot{\lambda} + K^2\,g_{rs}(x)\,\bar{B}^r(x)\,\bar{B}^s(x)\,\dot{\lambda}^2, \\ \dot{\lambda}^2 &= g(x;\dot{x},\dot{x}) - 2\,K\,g(x;\dot{x},\bar{B}(x))\,\dot{\lambda} + K^2\,g(x;\bar{B}(x),\bar{B}(x))\,\dot{\lambda}^2. \end{aligned}\right\} \tag{13}$$

We wish to show that the last term in (13) vanishes; that is, that $\bar{B}(x)$ is also a light-like vector in the g metric. But this is trivial since by lemma 1, we have $B(y) = -\mathring{\varphi}_\sigma(\phi_\sigma(y))\,\mathring{\varphi}_\lambda^{-1}(y)$, and with $y = \phi_\sigma(x)$,

$$B(\phi_\sigma(x)) = -\,\mathring{\varphi}_\sigma(x)\,\bar{B}(x). \tag{14}$$

But by (12),

$$\begin{aligned} g(x;\bar{B}(x),\bar{B}(x)) &= g\{\phi_\sigma(x);\,\mathring{\varphi}_\sigma(x)\,\bar{B}(x),\,\mathring{\varphi}_\sigma(x)\,\bar{B}(x)\} \\ &= g(y;B(y),B(y)) = 0. \end{aligned}$$

Hence (13) may be solved for $\dot{\lambda}$ by 'completing the square'; that is,

$$\begin{aligned} \{\dot{\lambda} + K\,g_{rs}(x)\,\bar{B}^s(x)\,\dot{x}^r\}^2 &= \{g_{rs}(x) + K^2\,\bar{B}_r(x)\,\bar{B}_s(x)\}\,\dot{x}^r\,\dot{x}^s, \\ \{\dot{\lambda} + K\,g(x;\bar{B}(x),\dot{x})\}^2 &= g(x;\dot{x},\dot{x}) + K^2\,g(x;\bar{B}(x),\dot{x})\,g(x;\bar{B}(x),\dot{x}), \end{aligned}$$

from which (10) follows. The last assertions of the theorem are trivial since $\bar{B}_i\,\bar{B}^i = 0$. q.e.d.

Since $\bar{B}_i$ and h_{ij} in (10) are related to the g_{ij} of (1) and (2), the question arises as to what relations must hold between h_{ij} and $\bar{B}_i$ in order that (10) be derivable from an action principle of the form (1) as stated in theorem 1. Clearly $\bar{B}_i$ must be a null field relative to h_{ij} and further $\bar{B}^i$ must be tangential to the trajectories of a general Lorentz transformation group relative to the metric $g_{ij}=h_{ij}-K^2\,\bar{B}_i\,\bar{B}_j$.

In the form (10), the four vector potential $\bar{B}^i$ gives rise to the electro-magnetic field tensor F_{ij} defined by

$$F_{ij}=\frac{\partial\bar{B}_i}{\partial x^j}-\frac{\partial\bar{B}_j}{\partial x^i}=\bar{B}_{i,j}-\bar{B}_{j,i},$$

$$F(x;\xi,\eta)=h(x;D_\eta\,\bar{B},\xi)-h(x;D_\xi\,\bar{B},\eta),$$

where ',' refers to covariant differentiation in the h_{ij} metric; alternatively, D refers to the Riemannian connection in the space with metric h.

Since F_{ij} is unchanged by the addition of a gradient to $\bar{B}_i$ (a gauge transformation) the assertion that $\bar{B}^i$ should be a null field places essentially no restriction on F_{ij}; for if f is a scalar function satisfying the first order partial differential equation

$$h^{ij}\,\bar{B}_i\,\bar{B}_j+2\,h^{ij}\,\bar{B}_i\,\partial_j f+h^{ij}\,\partial_i f\,\partial_j f=0,$$

then $\bar{\bar{B}}_i=\bar{B}_i+\partial_i f$ satisfies $h_{ij}\,\bar{\bar{B}}^i\,\bar{\bar{B}}^j=0$ while the extremals of (10) are invariant under gauge transformations.

THEOREM 2: *Given an action integral of the form*

$$\lambda(\tau_1)=\int_{\tau_0}^{\tau_1}\{\sqrt{h_{ij}(x^k)\,\dot{x}^i\,\dot{x}^j}-K\,\bar{B}_i(x^k)\,\dot{x}^i\}\,d\tau+\lambda(\tau_0) \tag{10}$$

in which the gauge is chosen such that

$$h_{ij}(x^k)\,\bar{B}^i(x^k)\,\bar{B}^j(x^k)=0,\quad \text{where}\quad \bar{B}^i=h^{ij}\,\bar{B}_j. \tag{15}$$

Assume that $\bar{B}_i$ defines a motion in the metric $h_{ij}-\bar{B}_i\,\bar{B}_j$; that is,

$$\underset{B_i}{\mathscr{L}}(h_{ij}-K^2\,\bar{B}_i\,\bar{B}_j)=0,\quad (\underset{B}{\mathscr{L}}\ \textit{denotes the Lie derivative}); \tag{16}$$

then the system of differential equations

$$\frac{dx^i}{d\sigma}=\bar{B}^i(x^k) \tag{17}$$

defines a local group of transformations

$$x^i=x^i(\sigma;y^1,\ldots,y^n),\qquad x=\psi_\sigma(y)$$

where

$$\overset{*}{\psi}_\sigma\{\psi_\sigma^{-1}(x)\}=\bar{B}(x),\qquad \overset{*}{\psi}{}_\sigma^{-1}\{\psi_\sigma(y)\}\overset{\text{def}}{=}B(y),$$

under which (10), *with the assumption that* $\dot{\sigma} = K\,\dot{\lambda}$, *is transformed into*

$$\lambda(\tau_1) = \int_{\tau_0}^{\tau_1} \sqrt{g_{ij}(y)\,\dot{y}^i\,\dot{y}^j}\;d\tau + \lambda(\tau_0) \tag{1}$$

where $g_{ij}(y) = h_{ij}(y) - K^2\,\bar{B}_i(y)\,\bar{B}_j(y)$.

Proof: The proof consists essentially in reversing the steps of the proof of theorem 1. However, since the theorem is stated in the index notation, a proof in index free notation requires some preliminary definitions and observations:

$$\left.\begin{aligned} h(x;\xi,\eta) &\overset{\text{df}}{=} h_{ij}(x)\,\xi^i\,\eta^j, \qquad A(x;\xi) \overset{\text{df}}{=} h(x;\bar{B}(x),\xi),\\ g(x;\xi,\eta) &\overset{\text{df}}{=} h(x;\xi,\eta) - A(x;\xi)\,A(x;\eta), \end{aligned}\right\} \tag{18}$$

whence (15) implies

$$A(x;\bar{B}(x)) = h(x;\bar{B}(x),\bar{B}(x)) = g(x;\bar{B}(x),\bar{B}(x)) = 0\,. \tag{19}$$

It is well known [4] that (17) defines a local group of transformations $x = \psi_\sigma(y)$, given that the initial conditions are chosen such that ψ_σ includes the identity mapping, say for $\sigma = 0$. Then (16) implies that ψ_σ defines a motion in the g metric and hence (noting that the identity for $\sigma = 0$ implies $\mathring{g} = g$ in lemma 3),

$$g\{\psi_\sigma(y);\,\hat{\psi}_\sigma(y)\,\xi,\,\hat{\psi}_\sigma(y)\,\eta\} \equiv g(y;\xi,\eta)\,. \tag{20}$$

The arguments in the derivation of lemmas 1 and 2 apply to the transformations $x = \psi_\sigma(y)$, and hence

$$\dot{x} = \hat{\psi}_\sigma(y)\,\{\dot{y} - K\,B(y)\,\dot{\lambda}\}\,, \tag{21}$$

$$\bar{B}(\psi_\sigma(y)) = -\,\hat{\psi}_\sigma(y)\,B(y)\,. \tag{22}$$

Substitution of $x = \psi_\sigma(y)$ in (19) yields, in view of (20) and (22),

$$A\{\psi_\sigma(y);\,\bar{B}(\psi_\sigma(y))\} = g\{y;\,B(y),\,B(y)\} = 0\,,$$

and, by definitions (18), again using (22) and (20),

$$\left.\begin{aligned} A\{\psi_\sigma(y);\,\hat{\psi}_\sigma(y)\,\xi\} &= h\{\psi_\sigma(y);\,\bar{B}(\psi_\sigma(y)),\,\hat{\psi}_\sigma(y)\,\xi\}\\ &= -\,g\{\psi_\sigma(y);\,\hat{\psi}_\sigma(y)\,B(y),\,\hat{\psi}_\sigma(y)\,\xi\}\\ &= -\,g\{y;\,B(y),\,\xi\}\,. \end{aligned}\right\} \tag{23}$$

The theorem may now be proved by differentiating (10) with respect to the upper limit, obtaining in the index free notation the differential equation

$$\dot{\lambda}^2 + 2\,K\,A(x;\dot{x})\,\dot{\lambda} = g(x;\dot{x},\dot{x})\,.$$

The substitution $x = \psi_\sigma(y)$ yields, in view of (21), (20) and (24),

$$\dot{\lambda}^2 - 2\,K\,g\{y;\,B(y),\,\dot{y} - K\,B(y)\,\dot{\lambda}\}\,\dot{\lambda} = g\{y;\,\dot{y} - K\,B(y)\,\dot{\lambda},\,\dot{y} - K\,B(y)\,\dot{\lambda}\},$$

which, in view of (23) becomes simply

$$\dot{\lambda}^2 = g(y;\,\dot{y},\,\dot{y}),$$

thereby proving the theorem.

Equation (16) states that $\bar{B}(x)$ defines a *motion* [4] in the g metric and hence $\bar{B}_i$ must satisfy Killing's equation

$$\bar{B}_{i,j} + \bar{B}_{j,i} = 0 \tag{25}$$

where ',' denotes covariant differentiation relative to g_{ij}. Equation (15) implies also (as stated in (19))

$$g_{ij}\,\bar{B}^i\,\bar{B}^j = 0, \tag{19}$$

from which follows $\bar{B}^i\,\bar{B}_{i,k} = 0$ and hence, with (25), $\bar{B}_{i,k}\,\bar{B}^k = 0$. Then clearly

$$\bar{B}^i_{,k}\,\bar{B}^k = 0,$$

and the curves satisfying (17) are then null geodesics in the metric g_{ij} since by (17)

$$\frac{\partial \dot{x}^i}{\partial\sigma} = \bar{B}^i_{,k}(x)\,\bar{B}^k(x) = 0, \quad \text{where} \quad \dot{x}^i = \frac{dx^i}{d\sigma},$$

while (19) becomes $g_{ij}\,\dot{x}^i\,\dot{x}^j = 0$. This result simply asserts, in the special case here considered, the well known fact [4] that a Killing vector field of constant magnitude is tangential to a *translation*, since a translation is a metric preserving family of transformations whose trajectories are geodesics (briefly, a geodesic motion).

LEMMA 4: *The vector field $\bar{B}^i(x)$ in theorems* 1 *and* 2 *defines a null translation in the metric of g_{ij}.*

It may be of interest to note that the Killing equation (25) relative to the g_{ij} metric may be written relative to the h_{ij} metric in the form

$$\bar{B}_{i|j} + \bar{B}_{j|i} = \bar{B}^r[\bar{B}_j\,F_{ri} + \bar{B}_i\,F_{rj}]$$

where '|' refers to covariant differentiation relative to h_{ij}. We omit the proof since this result is not relevant to the purpose of this paper, namely: the motivation for and formulation of a modified Kaluza–Klein formalism.

3. The Modified Kaluza-Formalism

Equation (2) may be written

$$\dot{s}^2 = \dot{\lambda}^2 - g_{ij}(y^k)\,\dot{y}^i\,\dot{y}^j, \quad \text{when} \quad \dot{s} = 0,$$

while equation (13) may be written

$$\dot{s}^2 = \dot{\lambda}^2 + 2\,K\,\bar{B}_i(x)\,\dot{x}^i\,\dot{\lambda} - g_{ij}(x)\,\dot{x}^i\,\dot{x}^j\,, \qquad \dot{s} = 0\,,$$

where

$$g_{ij} = h_{ij} - K^2\,\bar{B}_i\,\bar{B}_j\,, \qquad h_{ij}\,\bar{B}^i\,\bar{B}^j = 0\,, \qquad \bar{B}^i = h^{ij}\,\bar{B}_j = g^{ij}\,\bar{B}_j\,.$$

Hence theorems 1 and 2 can be re-interpreted as follows:

Corollary 1: *Let V^5 be a five dimensional space in which there exists a* (cylindrical) *coordinate frame y^α* (Greek indices run from 1 to 5, Latin indices from 1 to 4) *such that the metric is given by* s *where*

$$\dot{s}^2 = -\,\gamma_{ij}(y^k)\,\dot{y}^i\,\dot{y}^j + (\dot{y}^5)^2\;; \tag{26}$$

that is, $\gamma_{55}=1$, $\gamma_{5i}=0$, γ_{ij} is independent of y^5. If a transformation $y^\alpha = y^\alpha(x^\beta)$ leaves y^5, γ_{55}, and the quadratic form $\gamma_{ij}\,\xi^i\,\eta^j$ invariant, then the transformation is of the form

$$y^5 = x^5\,, \qquad y^i = \phi^i_{x^5}(x^j)\,, \tag{27}$$

where $\phi_{x^5}(x)$ is a family of null translations (x^5 the parameter) *in the space V^4 with metric tensor γ_{ij}. If $\gamma_{5i}(x^k)$ denotes the* (covariant) *components of the tangent vector field to this translation, then in the x coordinate frame* (26) *has the form*

$$\dot{s}^2 = -\,\gamma_{ij}(x^k)\,\dot{x}^i\,\dot{x}^j + 2\gamma_{5i}(x^k)\,\dot{x}^i\,\dot{x}^5 + (\dot{x}^5)^2 \tag{28}$$

and

$$\gamma^{ij}\,\gamma_{5i}\,\gamma_{5j} = 0\,.$$

Proof: Since $y^\beta = y^\beta(x^\alpha)$ is to leave the form $\gamma_{ij}\,\xi^i\,\eta^j$ invariant, and y^5 invariant, it follows that the transformation may be written in the form (27), where $y^i = \phi^i_{x^5}(x^j)$ is a general Lorentz transformation in the space V^4 with metric γ_{ij}. The form (28) is obtained by precisely the same argument used in theorem 1 simply by setting $x^5 = y^5 = \lambda = \sigma$, $g_{ij} = \gamma_{ij}$, and $\bar{B}_i = \gamma_{5i}$ (in effect we are treating only the case $K=1$).

Corollary 2: *Let V^5 be a five dimensional space with metric satisfying equation* (28). *Assume further that γ_{5i} in* (28) *are the* (covariant) *components of the tangential vector field to a null translation in the V^4 with metric tensor γ_{ij}. Then the null translations $y^i = \phi^i_{x^5}(x^j)$ where*

$$\frac{dy^i}{dx^5} = \gamma^{ij}(x^k)\,\gamma_{j5}(x^k)\,,$$

together with $y^5 = x^5$, define a transformation in V^5 under which (28) *reduces to* (26).

Proof: Since (28) may be written in the form

$$\dot{s}^2 = (\dot{x}^5 + \gamma_{5i}\,\dot{x}^i)^2 - (\gamma_{ij} + \gamma_{5i}\,\gamma_{5j})\,\dot{x}^i\,\dot{x}^j\,,$$

the identification $x^5=\lambda$, $\gamma_{5i}=\bar{B}_i$, $\gamma_{ij}=g_{ij}$ yields

$$\dot{s}^2 = (\dot{\lambda} + \bar{B}_i\,\dot{x}^i)^2 - h_{ij}\,\dot{x}^i\,\dot{x}^j,$$

where $h_{ij}=g_{ij}+\bar{B}_i\,\bar{B}_j$; this clearly corresponds to (10). The assertion that γ_{5i} defines the tangent vector field to a null translation in the $\gamma_{ij}=g_{ij}=h_{ij}-\bar{B}_i\,\bar{B}_j$ metric is equivalent to (15) and (16). Hence the corollary is simply an alternate formulation of theorem 2, when $K=1$.

The space V^5 and the coordinate transformations discussed in (27) differ from the Kaluza space and corresponding transformations in that

(a) the vanishing of γ_{5i} is assumed in (26) for some (cylindrical) frame,

(b) not only is the cylindrical nature of the $\gamma_{\alpha\beta}$ assumed preserved, with $y^5=x^5$, but also the invariance of both γ_{55} *and* the form $\gamma_{ij}\,\xi^i\,\eta^j$ is here assumed.

The curvature tensor $P^{\alpha}_{\beta\gamma\delta}$ in V^5, in view of (26), depends solely on $\gamma_{55}=1$ and γ_{ij} along with its partial derivatives, and it is precisely γ_{55} and γ_{ij} which are invariant under our transformations. Relative to the transformations considered in §1, §2, or in Corollaries 1, 2 the selection of the curvature scalar P as integrand of the action integral is now amply justified; *it depends on, and solely on, the only invariants under the transformations considered, and coincides with the choice of R in the gravitational theory represented in* (26).

The fifth dimension $y^5=x^5$ is here considered invariant under the transformations, while $y^5=x^5+\phi^5(x^i)$ is the usual equation in the Kaluza–Klein formalism. In view of the physical interpretation of y^5 as the proper time of the particle considered, the invariance of proper time (which is simply the 'action' *both* in (1) and (10)) seems also amply justified.

4. Fields Derivable by Null Translations

For the purposes of this section, it is sufficient to consider $\sigma=\lambda$ in section 2. Hence we write simply $y=\phi_\lambda(x)$.

THEOREM 3: *Let the action integral*

$$\lambda = \int \sqrt{-g_{ij}\,\dot{y}^i\,\dot{y}^j}\,d\tau$$

be transformed, under a null translation corresponding to the null vector field $\bar{B}^i(x)$ in the g_{ij} metric, into the action integral

$$\lambda = \int \sqrt{-h_{ij}\,\dot{x}^i\,\dot{x}^j} - \bar{B}_i\,\dot{x}^i\,d\tau$$

where $h_{ij}=g_{ij}+\bar{B}_i\,\bar{B}_j$. Let $\underset{g}{R}(y)$ and $\underset{h}{R}(x)$ be the scalar curvature tensors formed from the g and h metrics respectively. Let $F_{ij}\overset{\text{def}}{=}\bar{B}_{i,j}-\bar{B}_{j,i}$. Then the determinant of F_{ij}

vanishes $(\det(F_{ij})=0)$ *and*

$$\underset{g}{R} = \underset{h}{R} - \tfrac{1}{4}F_{ij}\,F^{ij} \quad \text{where} \quad F^{ij} = g^{ir}\,g^{js}\,F_{rs} = h^{ir}\,h^{js}\,F_{rs}. \tag{29}$$

Note: The '$-$' sign in (29) results from the choice $\gamma_{55}=1$ which, as indicated in the introduction, does not conflict with the assertion that $x^5=\lambda$ is time like.

Proof: Although (29) may be derived by tedious explicit expansion, it is simpler to refer to the results of the Kaluza–Klein formalism. The first action integral of the theorem corresponds to the cylindrical metric

$$(\dot{y}^5)^2 + g_{ij}(y^k)\,\dot{y}^i\,\dot{y}^j$$

in a five dimensional space V^5, where again Latin indices assume the values $1,\dots,4$ and Greek indices $1,\dots,5$. The curvature scalar P in this V^5 clearly depends only on the $g_{ij}(y^k)$; in fact $P=R(g_{ij}(y))$ where $R(g_{ij}(y))$ denotes the scalar curvature in a four dimensional space V^4 with metric coefficients $g_{ij}(y^k)$. The form of the second action integral of the theorem implies that V^5 also has the cylindrical metric

$$\{\dot{x}^5 + \bar{B}_i(x^k)\,\dot{x}^i\}^2 + h_{ij}(x^k)\,\dot{x}^i\,\dot{x}^j,$$

with reference to the x^α coordinate system; alternatively, V^5 has the metric

$$(\dot{x}^5)^2 + 2\,\bar{B}_i(x^k)\,\dot{x}^i\,\dot{x}^5 + g_{ij}(x^k)\,\dot{x}^i\,\dot{x}^j.$$

But the Kaluza–Klein formalism states ([1], page 230) that in the x^α frame, the curvature scalar P of V^5 is given by

$$P(x^\alpha) = R(h_{ij}(x^k)) - \tfrac{1}{4}F_{ij}(x^k)\,F^{ij}(x^k) \quad \text{where} \quad F_{ij} = \partial_j\,B_i - \partial_i\,B_j,$$

where again $R(h_{ij}(x^k))$ is the curvature scalar in a V^4 with metric $h_{ij}(x^k)\,\dot{x}^i\,\dot{x}^j$. Since $y^i=\phi^i_{x^5}(x^j)$ is a Lorentz transformation (for every x^5) in the g metric, it follows that $R(g_{ij}(y))$ becomes $R(g_{ij}(x))$ under this Lorentz transformation, and (29) follows.

THEOREM 4:[2]) *Let ',' and '|' denote respectively covariant differentiation in the* g_{ij} *and the* h_{ij} *metrics. Then* $h^{ir}\,F_{rj}=g^{ir}\,F_{rj}$; *that is, either* h^{ij} *or* g^{ij} *may be used to raise the indices of* F_{ij}. *If* $\underset{g}{R}_{ijkr}$ *and* $\underset{h}{R}_{ijkr}$ *denote respectively the curvature tensors corresponding to the metrics* g_{ij} *and* h_{ij}, *then*

$$\left.\begin{array}{ll}
\text{(a)}\ \ F_{kj,i} = 2\,\bar{B}_r\,\underset{g}{R}{}^{r}_{\cdot ijk} & \text{(b)}\ \ F_{kj|i} = 2\,\bar{B}_r\,\underset{h}{R}{}^{r}_{\cdot ijk} \\
\text{(c)}\ \ F^{ij}{}_{,j} = 2\,\bar{B}_j\,\underset{g}{R}{}^{ij} & \text{(d)}\ \ F^{ij}{}_{|j} = 2\,\bar{B}_j\,\underset{h}{R}{}^{ij} \\
\text{(e)}\ \ \tfrac{1}{4}F_{ij}\,F^{ij} = -\,\bar{B}^i\,\bar{B}^j\,\underset{g}{R}_{ij} = -\,\bar{B}^i\,\bar{B}^j\,\underset{h}{R}_{ij}. &
\end{array}\right\} \tag{30}$$

---

[2]) Formulas (30) a, b, c, d are due to Professor D. LOVELOCK (Bristol); the author wishes to thank him for his considerable assistance.

Proof: Since $\bar{B}^i \bar{B}_{i,j} = -\bar{B}^i \bar{B}_{j,i} = 0$ it follows that $\bar{B}^i F_{ij} = 0$. It is readily verified that $h_{ij} = g_{ij} + \bar{B}_i \bar{B}_j$ implies $h^{ij} = g^{ij} - \bar{B}^i \bar{B}^j$, whence

$$h^{ij} F_{jk} = g^{ij} F_{jk} - \bar{B}^i \bar{B}^j F_{jk} = g^{ij} F_{jk}.$$

The assertions in (30) may be proved as follows: since $F_{ij} = \bar{B}_{i,j} - \bar{B}_{j,i} = 2\,\bar{B}_{i,j}$ we see that

$$F_{ij,k} = 2\,\bar{B}_{i,jk} = 2(\bar{B}_{i,kj} + \bar{B}_r \underset{g}{R}{}^r_{\cdot ijk}) = F_{ik,j} + 2\,\bar{B}_r \underset{g}{R}{}^r_{\cdot ijk},$$

and (30-a) follows from the identity $F_{ij,k} + F_{ki,j} + F_{jk,i} \equiv 0$. Similarly for (30-b). Formulas (30-c) and (30-d) are immediate consequences of the preceding formulas. To prove (30-e), note that $g^{ij} \bar{B}_i \bar{B}_j = 0$ implies $g^{ij} \bar{B}_i \bar{B}_{j,s} = 0$ and hence

$$\begin{aligned} g^{ij} \bar{B}_{i,r} \bar{B}_{j,s} &= -g^{ij} \bar{B}_i \bar{B}_{j,rs} = g^{ij} \bar{B}_i \bar{B}_{r,js} \\ &= \bar{B}^j \{\bar{B}_{r,sj} - \bar{B}^k \underset{g}{R}_{krsj}\}. \end{aligned}$$

The relation $\bar{B}^r{}_{,r} = 0$ (in view of (25)) thus implies

$$g^{rs} g^{ij} \bar{B}_{i,r} \bar{B}_{j,s} = -\bar{B}^j \bar{B}^k \underset{g}{R}_{kj};$$

on the other hand,

$$F_{ij} F^{ij} = g^{ir} g^{js} (\bar{B}_{i,j} - \bar{B}_{j,i})(\bar{B}_{r,s} - \bar{B}_{s,r}) = 4\,g^{ir} g^{js} \bar{B}_{i,j} \bar{B}_{r,s}$$

which proves (30-e), since the above proof is equally valid with ',' replaced by '|' and $\underset{g}{R}_{ijkr}$ replaced by $\underset{h}{R}_{ijkr}$. q.e.d.

If F_{ij} is related explicitly to the electric and magnetic fields $\vec{E}$ and $\vec{H}$ respectively by the expression (for example [5] page 58 and 63)

$$F_{ij} = \begin{pmatrix} 0 & H_3 & -H_2 & -i\,E_1 \\ -H_3 & 0 & H_1 & -i\,E_2 \\ H_2 & -H_1 & 0 & -i\,E_3 \\ i\,E_1 & i\,E_2 & i\,E_3 & 0 \end{pmatrix},$$

then the field invariants $F_{ij} F^{ij}$ and $\det(F_{ij})$ are given in terms of $\vec{E}$ and $\vec{H}$ by

$$F_{ij} F^{ij} = 2(|\vec{H}|^2 - |\vec{E}|^2), \qquad \det(F_{ij}) = -(\vec{E}\cdot\vec{H})^2.$$

In terms of $\vec{E}$ and $\vec{H}$, theorem 4 implies the

COROLLARY 3: *Under the hypothesis of theorem* 3, $\vec{E}\cdot\vec{H}$ *always vanishes, while* $|\vec{H}|^2 - |\vec{E}|^2$ *will also vanish if* $\underset{g}{R}_{ij} = 0$. *The current* $F^{ij}{}_{|j}$ *vanishes if* $\underset{h}{R}_{ij} = 0$.

(It can be shown that $F^{ij}{}_{|j} = F^{ij}{}_{,j}$ and hence the current vanishes if either $\underset{g}{R}_{ij} = 0$ or $\underset{h}{R}_{ij} = 0$.) In view of the identical vanishing of the field invariant $\vec{E}\cdot\vec{H}$, it is of interest to investigate more fully the vanishing of the invariant $|\vec{H}|^2 - |\vec{E}|^2$; the vanishing of the latter would imply that only null electromagnetic fields are derivable by the above methods. As stated in lemma 4, the vector fields $\bar{B}^i(x)$ define null translations, that is, a motion (in the metric of g_{ij}) for which every point follows a null geodesic. It is known [4] that any motion in a V^n is characterized as follows: there exists a co-ordinate frame x^i for V^n in which the components of the metric tensor g_{ij} are independent of one co-ordinate, say x^1; the curves of parameter x^1 are then the trajectories of this motion. In a V^4, suppose g_{ij} independent of x^1, and let the indices $\alpha, \beta, \gamma \ldots = 2, 3, 4$ while $i, j, k \ldots = 1, 2, 3, 4$. Then the metric in V^4 is given by the quadratic form

$$g_{11}(x^\gamma)(\dot{x}^1)^2 + 2g_{1\beta}(x^\gamma)\,\dot{x}^1\,\dot{x}^\beta + g_{\alpha\beta}(x^\gamma)\,\dot{x}^\alpha\,\dot{x}^\beta .$$

If B^i is tangential to this motion then its components must be $[B^1, 0, 0, 0]$, and hence $B_i(x^j) = g_{i1}(x^j)\,B^1(x^j)$; that B^i corresponds to a motion implies $B_{i,j} + B_{j,i} = 0$; that is, $\partial_j(g_{i1}\,B^1) + \partial_i(g_{j1}\,B^1) = 2\,[ij, 1]\,B^1$. Since g_{ij} is independent of x^1 it follows that $2\,[ij, 1] = \partial_j\,g_{i1} + \partial_i\,g_{j1}$ and hence

$$g_{i1}\,\frac{\partial B^1}{\partial x^j} + g_{j1}\,\frac{\partial B^1}{\partial x^i} = 0 .$$

If we set $i = j = 1$, it follows that B^1 is independent of x^1; but then with $i = 1$ follows $\partial B^1/\partial x^j = 0$, and hence *if B^i is tangent to the motion along x^1, then $B^1 = k$, a constant, and $B^\alpha = 0$, in this coordinate frame.*

Since

$$B_i = k\,g_{1i} ,$$

in order that F_{ij} should not vanish for this motion one requires only that g_{1i} should not be a gradient, and hence

$$F_{ij} = k\left(\frac{\partial g_{1i}}{\partial x^j} - \frac{\partial g_{1j}}{\partial x^i}\right).$$

Finally, *in order that the motion be a geodesic motion*, viz. $g_{ij}\,B^i\,B^j$ *constant, it is further simply required that g_{11} be constant in this frame.*

It is clear from the above that the motion cannot be a null geodesic motion unless $k = 0$. However it is also known [4] that if B^i and $\bar{B}^i$ generate translations (geodesic motions), then $c_1\,B^i + c_2\,\bar{B}^i$ also generates a translation, for arbitrary constants c_1 and c_2. Hence, *if along with the field B^i: $[k, 0, 0, 0]$, a time translation corresponding to $\bar{B}^i$: $[0, 0, 0, \bar{k}]$ can be found, then $\bar{\bar{B}}^i$: $[c_1\,k, 0, 0, c_2\,\bar{k}]$ also generates a motion, and*

$$g_{ij}\,\bar{\bar{B}}^i\,\bar{\bar{B}}^j = c_1^2\,g_{11}\,k^2 + 2\,c_1\,c_2\,g_{14}\,k\,\bar{k} + c_2^2\,g_{44}\,\bar{k}^2$$

can vanish since c_2 and/or $\bar{k}$ can be imaginary in our $V^4(x^4 = i\,c\,t)$. Hence, for the existence of a null geodesic motion in the V^4 of relativity it suffices to show the existence of a time translation and at least one space translation. The existence of a time translation is often considered as the criterion for a stationary gravitational field [6].

That $|\vec{H}|^2 - |\vec{E}|^2$ does not vanish in general may be seen as follows: consider a metric of the form

$$(\dot{x}^1)^2 + 2\, g_{1\alpha}\, \dot{x}^1\, \dot{x}^\alpha + g_{\alpha\beta}\, \dot{x}^\alpha\, \dot{x}^\beta + 2\, g_{\alpha 4}\, \dot{x}^\alpha\, \dot{x}^4 + (\dot{x}^4)^2$$

where α, $\beta = 2$, 3 and the g_{ij} are independent of x^1 and $x^4 = i\,c\,t$. Let B^i be given by $[k, 0, 0, i\,k]$, and therefore B^i corresponds to a motion for which

$$g_{ij}\, B^i\, B^j = k^2 + (i\, k)^2 = 0\,,$$

hence a *null translation*. Then B_i is given by

$$k\,[1,\, g_{21} + i\, g_{24},\, g_{31} + i\, g_{34},\, i]$$

in which all terms are independent of x^1 and x^4. It is readily verified that

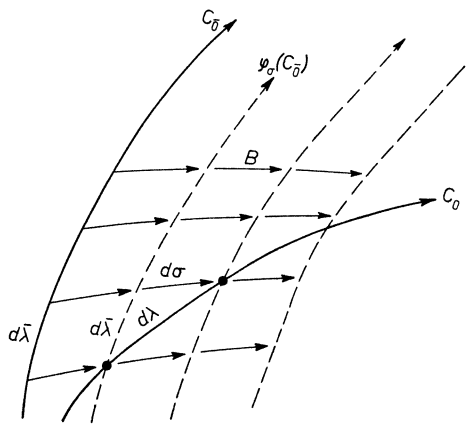
$$F_{23} = -\,F_{32} = \left(\frac{\partial g_{21}}{\partial x^3} - \frac{\partial g_{31}}{\partial x^2}\right) + i\left(\frac{\partial g_{24}}{\partial x^3} - \frac{\partial g_{34}}{\partial x^2}\right)$$

while all other F_{ij} vanish. Clearly F_{23}, and hence also $F_{ij}\, F^{ij}$, is not zero in general (given $\det(g^{\alpha\beta}) \neq 0$).

5. Geometric and Physical Interpretation of $y = \phi_\sigma(x)$

Let 0 and $\bar{0}$ be two observers with C_0: $y = y(\lambda)$ the world line of observer 0 in terms of his proper time λ, and $C_{\bar{0}}$: $x = x(\bar{\lambda})$ that of $\bar{0}$ in terms of his proper time $\bar{\lambda}$. Thus

$$g\left\{y;\, \frac{dy}{d\lambda},\, \frac{dy}{d\lambda}\right\} = 1 \quad \text{and} \quad g\left\{x;\, \frac{dx}{d\bar{\lambda}},\, \frac{dx}{d\bar{\lambda}}\right\} = 1\,. \tag{31}$$



Let B be a null Killing vector field corresponding to a continuous group of null translations $\xi = \phi_\sigma(\eta)$; the trajectories of the translations are then the world lines of light signals, and *we assume that these signals join* $C_{\bar{0}}$ *to* C_0. Hence a signal received by 0 from the clock of $\bar{0}$ has travelled along a trajectory of the group ϕ_σ. By assumption therefore, each value of $\bar{\lambda}$ determines a unique null geodesic joining $x(\bar{\lambda})$ to a unique event $y(\lambda)$ on C_0. Alternatively, *every emission time* $\bar{\lambda}$ *along* $C_{\bar{0}}$ *determines* a unique group parameter $\sigma = \sigma(\bar{\lambda})$ and *a unique reception time* $\lambda = \lambda(\bar{\lambda})$ *along* C_0 such that

$$y\{\lambda(\bar{\lambda})\} = \varphi_{\sigma(\bar{\lambda})}\{x(\bar{\lambda})\}. \tag{32}$$

The relativistic Doppler shift is then given by the reciprocal of $d\lambda/d\bar{\lambda}$. This Doppler shift is readily calculated by differentiation of (32) with respect to $\bar{\lambda}$; one thereby obtains, as in (9),

$$\frac{dy}{d\lambda}\cdot\frac{d\lambda}{d\bar{\lambda}} = \varphi_\sigma(x)\left\{\frac{dx}{d\bar{\lambda}} - \bar{B}(x)\frac{d\sigma}{d\bar{\lambda}}\right\}$$

where $\bar{B}$ is the null Killing vector field corresponding to the inverse transformation $\eta = \phi_\sigma^{-1}(\xi)$. Since ϕ_σ is a general Lorentz transformation for every σ, it follows that

$$g\left\{y;\frac{dy}{d\lambda},\frac{dy}{d\lambda}\right\}\cdot\left(\frac{d\lambda}{d\bar{\lambda}}\right)^2 = g\left\{x;\frac{dx}{d\bar{\lambda}} - \bar{B}\frac{d\sigma}{d\bar{\lambda}},\frac{dx}{d\bar{\lambda}} - \bar{B}\frac{d\sigma}{d\bar{\lambda}}\right\}$$

which, in view of (31), implies

$$\left(\frac{d\lambda}{d\bar{\lambda}}\right)^2 = 1 - 2\,g\left\{x;\frac{dx}{d\bar{\lambda}},\bar{B}\right\}\frac{d\sigma}{d\bar{\lambda}}. \tag{33}$$

In terms of an arbitrary parameter τ, (33) may be rewritten in the form

$$\dot{\lambda}^2 = \dot{\bar{\lambda}}^2 - 2\,g\{x;\dot{x},\bar{B}\}\,\dot{\sigma}. \tag{34}$$

To illustrate that (33) is indeed a formula for the Doppler shift, consider the case of special relativity; for simplicity, we consider only x and t. The world lines $C_{\bar{0}}$ and C_0 are given in terms of their respective proper times by

$$C_0:\begin{cases} y(\lambda) = \dfrac{\beta\,\lambda}{\sqrt{1-\beta^2}} \\ t(\lambda) = \dfrac{\lambda}{c\sqrt{1-\beta^2}} \end{cases} \quad\text{and}\quad C_{\bar{0}}:\begin{cases} x(\bar{\lambda}) = \dfrac{\bar{\beta}\,\bar{\lambda}}{\sqrt{1-\bar{\beta}^2}} \\ \bar{t}(\bar{\lambda}) = \dfrac{\bar{\lambda}}{c\sqrt{1-\bar{\beta}^2}} \end{cases}$$

where $c\,\beta = dy/dt$ and $c\,\bar{\beta} = dx/d\bar{t}$. The group of null translations $\xi = \phi_\sigma(\eta)$ is given by

$$\varphi_0:\begin{cases} \xi = \eta + (\alpha\,c)\,\sigma \\ \bar{\tau} = \tau + \alpha\,\sigma \end{cases} \quad \text{for arbitrary constant} \quad \alpha,$$

where B is the vector $[\alpha\, c, \alpha]$ and $\bar{B}$ the vector $[-\alpha\, c, \; -\alpha]$. Finally, g_{ij} is the diagonal matrix $\{-1, c^2\}$. Equations (32) then become

$$\frac{\bar{\beta}\,\bar{\lambda}}{\sqrt{1-\bar{\beta}^2}} + (\alpha\, c)\,\sigma = \frac{\beta\,\lambda}{\sqrt{1-\beta^2}}$$

$$\frac{\bar{\lambda}}{c\sqrt{1-\bar{\beta}^2}} + \alpha\,\sigma = \frac{\lambda}{c\sqrt{1-\beta^2}}$$

which may be solved for σ and λ in terms of $\bar{\lambda}$ to obtain

$$\sigma = -\frac{\beta - \bar{\beta}}{\alpha\, c(1-\beta)\sqrt{1-\bar{\beta}^2}}\,\bar{\lambda} \quad \text{and} \quad \lambda = \sqrt{\frac{(1+\beta)(1-\bar{\beta})}{(1-\beta)(1+\bar{\beta})}}\,\bar{\lambda}. \tag{35}$$

When the source $\bar{0}$ is a rest, $\bar{\beta}=0$ and the second of equations (35) reduces to the well-known expression $\bar{\lambda}=(1-\beta)^{1/2}/(1+\beta)^{1/2}\,\lambda$ for the Doppler shift due to longitudinal motion.

It remains to verify that (33) is also the relativistic Doppler shift. In view of the previously given expressions for $\bar{B}$, $x(\bar{\lambda})$, g_{ij}, and (35), equation (33) becomes

$$\begin{aligned}\left(\frac{d\lambda}{d\bar{\lambda}}\right)^2 &= 1 - 2\left\{-\frac{\bar{\beta}}{\sqrt{1-\bar{\beta}^2}}(-\alpha\, c) + c^2\,\frac{1}{c\sqrt{1-\bar{\beta}^2}}(-\alpha)\right\}\cdot\frac{\beta-\bar{\beta}}{\alpha\, c(1-\beta)\sqrt{1-\bar{\beta}^2}}\\ &= 1 + 2\,\frac{(1-\bar{\beta})(\beta-\bar{\beta})}{(1-\bar{\beta}^2)(1-\beta)}.\end{aligned}$$

It is readily verified that the above equation agrees with (35), viz.

$$\left(\frac{d\lambda}{d\bar{\lambda}}\right)^2 = \frac{(1-\bar{\beta})(1+\beta)}{(1+\bar{\beta})(1-\beta)}.$$

Equation (34) is also of interest since, by (31),

$$\dot{\bar{\lambda}}^2 = g(x;\dot{x},\dot{x}) \tag{36}$$

and (34) may be written in the form

$$\dot{\lambda}^2 = g(x;\dot{x},\dot{x}) - 2\,g(x;\dot{x},\bar{B})\,\dot{\sigma}. \tag{37}$$

While (36) expresses the proper time $\bar{\lambda}$ of the particle whose world line is $x(\tau)$, (37) *expresses this same proper time in terms of the observer's clock.* In the cases previously considered, it was assumed that $\dot{\sigma}=K\,\dot{\lambda}$ for some constant K; this assumption in conjunction with (37) leads to the 'action' integral

$$\lambda = \int \{\sqrt{h_{ij}\,\dot{x}^i\,\dot{x}^j} - \bar{B}_i\,\dot{x}^i\}\,d\tau \tag{1*}$$

while (36) implies

$$\bar{\lambda} = \int \sqrt{g_{ij}\, \dot{x}^i \dot{x}^j}\, d\tau . \tag{1}$$

Hence (1) and (1*) both express the proper time $\bar{\lambda}$ of the particle; however, (1*) expresses this proper time in terms of $\lambda(\bar{\lambda})$, the reception time at which the observer 0 'sees' $\bar{\lambda}$ on the particle's clock.

6. The Case $\vec{E}\cdot\vec{H}\neq 0$

In the previous sections it was assumed that for $\lambda=0$ the transformation $y=\phi_\lambda(x)$ reduced to the identity map. The Kaluza–Klein formalism then implies that ϕ_λ represents a null translation, and that the field invariant $\vec{E}\cdot\vec{H}$ vanishes. While the preceding analysis indicates several possibilities for generalization to fields for which $\vec{E}\cdot\vec{H}\neq 0$, (for example, $B_i B^i$ non-constant, or B_i tangent to more general families of null geodesics) it may be fruitful to consider the analysis below, which lies entirely within the present framework.

Given the form

$$\dot{\lambda}^2 - g(y; \dot{y}, \dot{y}), \tag{36}$$

and the transformation $y=\phi_\lambda(x)$, $x=\phi_\lambda^{-1}(y)$, one may again define the vector fields $B(y)$, $\bar{B}(x)$, by [3]

$$B(y) = \mathring{\phi}_\lambda(\phi_\lambda^{-1}(y)), \quad \bar{B}(x) = \mathring{\phi}_\lambda^{-1}(\phi_\lambda(x)).$$

As previously shown, it follows that

$$\dot{y} = \hat{\varphi}_\lambda(x)\cdot\{\dot{x} - \bar{B}(x)\,\dot{\lambda}\},$$

and substitution into (36) yields

$$\dot{\lambda}^2 - g[\phi_\lambda(x); \hat{\varphi}_\lambda(x)\{\dot{x} - \bar{B}\,\dot{\lambda}\}, \hat{\varphi}_\lambda(x)\{\dot{x} - \bar{B}\,\dot{\lambda}\}]. \tag{37}$$

The cylindrical axiom of the Kaluza–Klein formalism demands that $g\{\phi_\lambda(x); \hat{\varphi}_\lambda(x)\,\xi, \hat{\varphi}_\lambda(x)\,\eta\}$ be independent of λ, and hence λ may be chosen arbitrarily, in particular $\lambda=0$, which suggests the definition

$$\overset{*}{g}(x; \xi, \eta) \overset{\text{def}}{=} g\{\phi_0(x); \hat{\varphi}_0(x)\,\xi, \hat{\varphi}_0(x)\,\eta\}.$$

Hence $\phi_\lambda(x)$ must be a motion as before, but a motion in the induced $\overset{*}{g}$ metric, as indicated in lemma 3. If $\overset{*}{g}=g$ then the mapping $\phi_0(x)$ is an isometry in the space V^4 with metric g. In general, one can only say that

$$\overset{*}{\bar{B}}_{i;j} + \overset{*}{\bar{B}}_{j;i} = 0 \quad \text{where} \quad \overset{*}{\bar{B}}_i = \overset{*}{g}_{ij}\,\bar{B}^j \quad \text{and ";"}$$

[3]) Note that this is also a restriction on ϕ_λ since generally one would have $B(y, \lambda)$ and $\bar{B}(x, \lambda)$.

denotes covariant differentiation in the induced $\overset{*}{g}$ metric. Hence (37) reduces to

$$\dot{\lambda}^2 - \overset{*}{g}\{x; \dot{x} - \bar{B}\,\dot{\lambda}, \dot{x} - \bar{B}\,\dot{\lambda}\},$$

that is,

$$[1 - \overset{*}{g}(x; \bar{B}, \bar{B})]\,\dot{\lambda}^2 + 2\,\overset{*}{g}(x; \bar{B}, \dot{x})\,\dot{\lambda} - \overset{*}{g}(x; \dot{x}, \dot{x}).$$

The Kaluza–Klein formalism then demands that $\bar{B}$ be a null vector field in the induced metric $\overset{*}{g}$. In component notation, $\bar{B}^i(x)$ must satisfy

$$\overset{*}{\bar{B}}_{i;\,j} + \overset{*}{\bar{B}}_{j;\,i} = 0 \qquad \overset{*}{g}_{i\,j}\,\bar{B}^i\,\bar{B}^j = 0,$$

and hence $\bar{B}^i$ defines a null translation in the induced $\overset{*}{g}$ metric; clearly $\bar{B}^i$ does not necessarily define a null translation in the g metric unless of course $\overset{*}{g}$ is the result of a translation in the g metric.

It is clear that

$$\overset{*}{F}_{i\,j} = \overset{*}{\bar{B}}_{i;\,j} - \overset{*}{\bar{B}}_{j;\,i} = \overset{*}{\bar{B}}_{i,\,j} - \overset{*}{\bar{B}}_{j,\,i}$$

again satisfies the condition $\det(\overset{*}{F}_{i\,j}) = 0$, but the tensor $F_{i\,j}$, given by

$$F_{i\,j} = \bar{B}_{i,\,j} - \bar{B}_{j,\,i}$$

does not necessarily satisfy $\det(F_{i\,j}) = 0$, that is, it is still possible that $\vec{E}\cdot\vec{H}$ does not vanish. It is hoped that a more detailed analysis can be presented in the near future.

REFERENCES

[1] PAULI, W., *Theory of Relativity* (Pergamon Press, N.Y. 1958).
[2] TONNELAT, M. A., *Les théories unitaires de l'électromagnétisme et de la gravitation* (Gauthier-Villars, Paris 1965).
[3] BERGMAN, P., *Introduction to the Theory of Relativity* (Prentice-Hall, N.Y. 1965).
[4] EISENHART, L. P., *Continuous Groups of Transformations* (Dover Publications, N.Y. 1961).
[5] LANDAU, L. D. and LIFSHITZ, E. M., *The Classical Theory of Fields* (Pergamon Press, London, Paris, 1959).
[6] SYNGE, J. L., *Relativity: General Theory* (North-Holland Publishing Co., Amsterdam 1960).

University of Waterloo

Über die Koebesche Konstante $\frac{1}{4}$

KARL SZILÁRD, Budapest

Herrn Professor Alexander Ostrowski zum 75-sten Geburtstag

Es sei die analytische Funktion $w=f(z)$ im Einkeitskreise $|z|<1$ definiert, $f(0)=0$ und $|f'(0)|=1$. Für den Fall, dass die Abbildung $z\to w=f(z)$ schlicht ist, wurde bekanntlich von L. BIEBERBACH im Jahre 1916 bewiesen, dass der grösstmögliche Radius R der Kreisscheibe $|w|<R$ die in allen Mengen $\mathfrak{B}$ der Bildpunkte $w=f(z)$ (d.h. für alle schlichten Funktionen mit der obigen Normierung) enthalten ist, genau den Wert $\frac{1}{4}$ besitzt. Dass ein solcher grösstmöglicher Radius existiert (dessen Wert somit eine absolute Konstante sein muss) wurde von P. KOEBE schon im Jahre 1907 bewiesen. Aus allgemeinen Sätzen von W. K. HAYMAN [1] folgt, dass die Forderung der Schlichtheit durch eine schwächere ersetzt werden kann („schwache Schlichtheit"); es bleibt richtig, dass $R=\frac{1}{4}$.

Hier soll gezeigt werden, dass erstens die Behauptung $R=\frac{1}{4}$ nicht nur für die „schwach schlichten" normierten, sondern auch für eine weitere Klasse von nicht notwendigerweise schlichten Abbildungen gültig bleibt, und zweitens, dass für eine gegebene normierte Abbildung $z\to w=f(z)$ der Wert dieses grösstmöglichen Radius von unten abgeschätzt werden kann, wenn in gewissem (später zu präzisierenden) Sinne das Verhalten der Bildpunktmenge $\mathfrak{B}$ in bezug auf die zum Punkte $w=0$ (wir wollen diesen Punkt durch O bezeichnen) nächsten Randpunkte von $\mathfrak{B}$ bekannt ist. In welchem Sinne dieses Verhalten in bezug auf die nächsten Randpunkte gemeint ist, wird aus dem nächstfolgenden Hilfssatze klar. Dieser Hilfssatz bezieht sich auf einen speziellen Fall. Der Übergang zum allgemeineren Falle geschieht durch Benutzung eines Satzes von G. PÓLYA und G. SZEGÖ [2] über das Nichtabnehmen des inneren Radius eines Gebietes in bezug auf einen gewissen Punkt, wenn man das Gebiet einer Symmetrisierung (nach PÓLYA, oder nach STEINER) unterwirft (siehe auch [3], Seite 82).

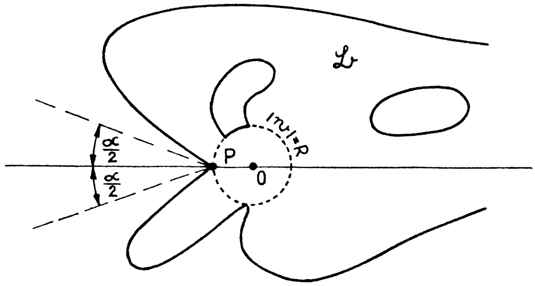
HILFSSATZ. *Die analytische Funktion $w=f(z)$ sei im Einheitskreise $|z|<1$ definiert, $f(0)=0$ und $|f'(0)|=1$. Die Menge $\mathfrak{B}$ der Werte w, die von der Funktion $f(z)$ mindestens einmal angenommen werden, soll folgende Eigenschaft haben: Einer derjenigen Randpunkte von $\mathfrak{B}$* (der Randpunkt *P*) *die einen minimalen Abstand R vom Punkte $w=0$* (vom Punkte *O*) *haben, ist die Spitze eines Winkels von der Grösse α $(0\leqq\alpha\leqq\pi)$ dessen Halbierende die Gerade OP ist derart, dass dieser Winkel* (in seinem Inneren und auch auf den Schenkeln) *keinen Punkt der Menge $\mathfrak{B}$ enthält* (s. Fig. 1).

Eingegangen am 28.12.67

Wir behaupten dann:

$$R \geqq \frac{1}{4 - \dfrac{2\alpha}{\pi}},$$

wobei diese Abschätzung die bestmögliche ist.



w - Ebene

Figur 1

Beweis: Wir können ohne Einschränkung der Allgemeinheit annehmen, dass der Punkt P auf der negativen Hälfte der reellen Achse der w-Ebene liegt, d.h., dass die Abscisse von P gleich $(-R)$ ist. (Denn wäre dies nicht der Fall, so würden wir statt $f(z)$ eine Funktion $f_1(z) = e^{i\omega} f(z)$ mit einem passend gewählten ω-Wert betrachten, die den angeführten Voraussetzungen, insbesondere mit Beibehaltung des Wertes von R genügt.) Wir fassen nun das Gebiet $\mathfrak{G}$ der w-Ebene, das aus allen Punkten w mit

$$-\pi + \frac{\alpha}{2} < \arg(w + R) < \pi - \frac{\alpha}{2}$$

besteht, ins Auge und konstatieren, dass $\mathfrak{B}$ ein Teilgebiet von $\mathfrak{G}$ ist. Durch die Funktion

$$z_1(w) = \frac{(w+R)^{\frac{1}{2-(\alpha/\pi)}} - R^{\frac{1}{2-(\alpha/\pi)}}}{(w+R)^{\frac{1}{2-(\alpha/\pi)}} + R^{\frac{1}{2-(\alpha/\pi)}}}$$

wird das Gebiet $\mathfrak{G}$ auf das Innere des Einheitskreises $|z_1| < 1$ abgebildet, wobei der Punkt $w = 0$ in den Punkt $z_1 = 0$ übergeführt wird. Setzt man anstatt w in $z_1(w)$ die Funktionswerte $w = f(z)$ ein, so bekommt man eine analytische Funktion

$$z_1[f(z)] = F(z)$$

durch welche der Einheitskreis $|z| < 1$ auf ein Teilgebiet des Einheitskreises $|z_1| < 1$

abgebildet wird, wobei $F(0)=0$ ist. Nach dem Schwarz'schen Lemma ist dann

$$\left|\frac{dF(z)}{dz}\right|_{z=0} \leqq 1 .$$

Nach der Kettenregel können wir schreiben:

$$\left|\frac{dF(z)}{dz}\right|_{z=0} = \left|\frac{dz_1}{dw}\right|_{w=0} \cdot \left|\frac{df}{dz}\right|_{z=0} = \left|\frac{dz_1}{dw}\right|_{w=0} \leqq 1,$$

da der zweite Faktor laut Voraussetzung gleich Eins ist. Nun ist

$$\left(\frac{dz_1}{dw}\right)_{w=0} = \frac{2\, R^{\frac{1}{2-(\alpha/\pi)}} \cdot R^{\frac{-1+(\alpha/\pi)}{2-(\alpha/\pi)}}}{4 \cdot R^{\frac{2}{2-(\alpha/\pi)}}\left(2-\frac{\alpha}{\pi}\right)} = \frac{1}{R\left(4-\frac{2\alpha}{\pi}\right)}.$$

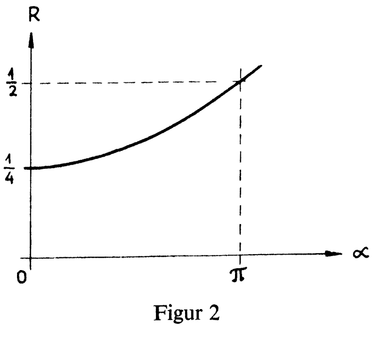
Somit erhalten wir:

$$\frac{1}{R\left(4-\frac{2\alpha}{\pi}\right)} \leqq 1, \quad \text{oder} \quad R \geqq \frac{1}{4-\frac{2\alpha}{\pi}}$$

und das ist die erste Behauptung des Hilfssatzes. Die zweite Behauptung ist, dass diese Abschätzung nicht verschärft werden kann. Das leuchtet aber ein, wenn man bedenkt, dass das Gebiet $\mathfrak{G}$ eines der möglichen Gebiete $\mathfrak{B}$ ist und für dieses Gebiet ist $F(z)=z$ und somit $F'(z)\equiv 1$, woraus folgt, dass die vorige Ungleichung in diesem Falle in die Gleichung

$$R = \frac{1}{4-\frac{2\alpha}{\pi}}$$

übergeht. Damit ist der ausgesprochene Hilfssatz bewiesen. Die Abhängigkeit des Radius R vom Winkel α ist auf der Figur 2. dargestellt, insbesondere sei vermerkt, dass für $\alpha=0$, $R=\frac{1}{4}$ und für $\alpha=\pi$, $R=\frac{1}{2}$ ist.



Figur 2

Nun wollen wir zeigen, dass der Radius R des grössten Kreises um den Punkt $w=0$, dessen innere Punkte sämtlich Bildpunkte sind, mit Hilfe einer Eigenschaft von $\mathfrak{B}$ in bezug auf die Menge der zum Punkte $w=0$ nächsten Randpunkte P abgeschätzt werden kann. Die zu $w=0$ nächsten Randpunkte liegen auf der Kreislinie $|w|=R$, sie können dort eine endliche, oder unendliche Menge bilden. Jedenfalls bewirken sie, dass, falls wir die Punktmenge $\mathfrak{B}$ nach der Methode von PÓLYA symmetrisieren, wobei wir die Symmetrisierung in bezug auf die positive Hälfte der reellen Achse der w-Ebene vornehmen, die symmetrisierte Punktmenge $\mathfrak{B}^*$ einen zu $w=0$ nächsten Randpunkt auf der negativen Hälfte der reellen Achse mit der Abseisse $w=-R$ haben wird. Dies bleibt auch richtig, wenn die Halbgerade, wonach wir symmetrisiert haben, die positive Hälfte der reellen Achse enthält, aus einem Punkte zwischen $w=0$ und $w=-R$ ausgeht, und der Punkt $w=-R$ für $\mathfrak{B}$ ein nächster Randpunkt war. Nun ist die Eigenschaft die wir der Punktmenge $\mathfrak{B}$ (die ein Gebiet ist) auferlegen wollen: *Nach der angeführten Symmetrisierung soll das symmetrisierte Gebiet $\mathfrak{B}^*$ einfach zusammenhängend sein.* Da der unendlich ferne Punkt $w=\infty$ weder zu $\mathfrak{B}$, noch zu $\mathfrak{B}^*$ gehört und $w=-R$ ein Randpunkt ist, so kann der Rand von $\mathfrak{B}^*$ nicht nur aus einem Punkte bestehen, d.h. der Rand von $\mathfrak{B}^*$ ist ein Kontinuum, das den Punkt $w=-R$ enthält und da er in bezug auf die reelle Achse auch symmetrisch ist, kann $\mathfrak{B}^*$ keinen Punkt der reellen Achse enthalten, dessen Abscisse kleiner als $(-R)$ ist (dies sieht man ein, wenn man bedenkt, dass $w=\infty$ nicht zu $\mathfrak{B}^*$ gehört). $\mathfrak{B}^*$ ist gewiss einfach zusammenhängend, wenn $\mathfrak{B}$ einfach zusammenhängend war (s. [3] Seite 71) und $\mathfrak{B}$ ist gewiss einfach zusammenhängend, wenn die Abbildung $z \to w=f(z)$ schlicht ist, doch $\mathfrak{B}^*$ kann auch für nicht schlichte Abbildungen einfach zusammenhängend sein, zum beispiel wenn die Abbildung $z \to w=f(z)$ „schwach schlicht" („weakly univalent" s. [1] Seite 157) ist und die Symmetrisierung nach der positiven Hälfte der reellen Achse vorgenommen wurde, $\mathfrak{B}^*$ kann aber für weitere nicht schlichte Abbildungen einfach zusammenhängend sein, wie man auch an Hand von Beispielen erkennen kann. Wir wollen jetzt den zu beweisenden Satz formulieren.

Die analytische Funktion $w=f(z)=a_1z+a_2z^2+\cdots$ sei für $|z|<1$ definiert, $|a_1|=1$. Die Menge $\mathfrak{B}$ der w-Werte die von $f(z)$ mindestens einmal angenommen werden, soll folgende Eigenschaft haben: Es soll ein Halbstrahl durch den Punkt $w=0$ und seine Verlängerung durch einen zu $w=0$ nächsten Randpunkt gehen so, dass die nach diesem Halbstrahl Pólya-symmetrisierte Punktmenge $\mathfrak{B}^*$ der Menge $\mathfrak{B}$ als einfach zusammenhängendes Gebiet ausfällt, oder es soll $\mathfrak{B}^*$ als einfach zusammenhängendes Gebiet ausfallen, wenn man nach einem aus $w=0$ ausgehenden Halbstrahl symmetrisiert. Dann behaupten wir: Der Radius R des grössten Kreises mit dem Mittelpunkt $w=0$, dessen Punkte sämtlich Bildpunkte $w=f(z)$ sind, ist $\geqq \frac{1}{4}$. Wenn es ausserdem möglich ist einen Winkel von der Grösse α anzugeben, dessen Spitze mit einem zu $w=0$ nächsten Randpunkte von $\mathfrak{B}^*$ zusammenfällt so, dass kein Punkt von $\mathfrak{B}^*$ in diesem Winkelraume liegt, so ist

$$R \geqq \frac{1}{4 - \frac{2\alpha}{\pi}}, \quad (0 \leqq \alpha \leqq \pi).$$

Beweis: Der Einfachkeit halber (und ohne die Allgemeinheit einzuschränken) nehmen wir an, dass der Halbstrahl, nach dem die Pólya-Symmetrisierung vorgenommen wurde, von einem Punkte Q der negativen Hälfte der reellen Achse zwischen $w = -R$ und $w = 0$ der w-Ebene (oder aus dem Punkte $w = 0$) ausgeht und die positive Hälfte der reellen Achse enthält. Auch können wir annehmen, dass $w = -R$ ein nächster Randpunkt von $\mathfrak{B}$ war. Das symmetrisierte Gebiet $\mathfrak{B}^*$ enthält sämtliche Punkte der Kreisfläche $|w| < R$ und mindestens einen Randpunkt P mit der Abcisse $(-R)$ d.h. diese Kreisfläche ist auch für $\mathfrak{B}^*$ die grösstmögliche von der betrachteten Art. Da das Gebiet $\mathfrak{B}^*$ laut Voraussetzung einfach zusammenhängend ist und sein Rand (wie vorhin schon erwähnt) nicht nur aus einem Punkte besteht, in bezug auf die reelle Achse symmetrisch ist und den unendlich fernen Punkt $w = \infty$ nicht enthält, so kann es keinen Punkt der reellen Achse mit einer Abscisse $\leqq -R$ enthalten.

Betrachten wir jetzt eine in dem Kreise $|z| < 1$ definierte analytische Funktion $f^*(z) = a_1^* z + a_2^* z^2 + \cdots$ die den Einheitskreis auf das Gebiet $\mathfrak{B}^*$ *schlicht* abbildet. Es ist $f(0) = f^*(0) = 0$. Laut des anfangs erwähnten Satzes von Pólya und Szegö können wir behaupten:

$$|a_1^*| \geqq 1.$$

Der Satz von Pólya und Szegö besagt hier nämlich, dass wenn r_0 der innere Radius des Gebietes $\mathfrak{B}$ in bezug auf $w = 0$ ist und die Symmetrisierung mit Hilfe einer Halbgeraden vorgenommen wurde, die durch den Punkt $w = 0$ geht, und r_0^* der innere Radius des symmetrisierten Gebietes $\mathfrak{B}^*$ in bezug auf $w = 0$ ist, so haben wir: $r_0^* \geqq r_0$. Weiterhin gilt: $|a_1| \leqq r_0$ und $|a_1^*| = r_0^*$ (letzteres wegen der Schlichtheit der Abbildung durch $f^*(z)$; siehe [3], Seite 80), so dass wir schreiben können:

$$1 = |a_1| \leqq r_0 \leqq r_0^* = |a_1^*|.$$

Nun genügt die Funktion

$$\frac{f^*(z)}{a_1^*} = z + \frac{a_2^*}{a_1^*} z^2 + \cdots$$

gewiss den Voraussetzungen des Hilfssatzes mit $\alpha = 0$, nur müssen wir statt R die Grösse $R/|a_1^*|$ schreiben und somit können wir behaupten:

$$\frac{R}{|a_1^*|} \geqq \tfrac{1}{4} \quad \text{und umso mehr} \quad R \geqq \tfrac{1}{4}.$$

Wenn aber ausserdem das Gebiet $\mathfrak{B}^*$ ein solches ist, dass es einem Randpunkt w mit $|w| = R$ besitzt, der Spitze eines Winkels von der Grösse α ist, der keinen Punkt von

$\mathfrak{B}^*$ enthält, so wird der Randpunkt $w=-R$ dieselbe Eigenschaft haben, wobei die reelle Achse der Winkelhalbierende dieses Winkels ist (dies folgt aus der Symmetrie von $\mathfrak{B}^*$ in bezug auf die reelle Achse und aus den Eigenschaften der Symmetrisation nach PÓLYA). In diesem Falle können wir wiederum den Hilfssatz auf die Funktion

$$\frac{f^*(z)}{a_1^*}$$

anwenden und folgern:

$$\frac{R}{|a_1^*|} \geqq \frac{1}{4-\frac{2\alpha}{\pi}}$$

und umso mehr:

$$R \geqq \frac{1}{4-\frac{2\alpha}{\pi}}$$

wie behauptet würde.

Es ist nicht schwer, Beispiele von nicht schlichten (auch nicht „schwach schlichten") Funktionen zu konstruieren, die den Voraussetzungen des eben bewiesenen Satzes genügen.

Die Behauptung: $R \geqq \frac{1}{4}$ kann für den Fall, dass das Pólya-symmetrisierte, einfach zusammenhängende Gebiet $\mathfrak{B}^*$ von $\mathfrak{B}$ durch eine Symmetrisierung in bezug auf die positive Hälfte der reellen Achse der w-Ebene erhalten wurde, auch aus dem Satze von W. K. Hayman bewiesen werden, dass für alle im Kreise $|z|<1$ definierten analytischen Funktionen $w=f(z)$ mit $f(0)=0$ und $f'(0)=1$ es eine Kreislinie $|w|=r$ mit $r \geqq \frac{1}{4}$ existiert die in $\mathfrak{B}$ enthalten ist (den extremalen Fall ausgenommen, s. [1], Satz IV). Da nämlich die Kreislinie $|w|=r$ nach der Symmetrisierung zu $\mathfrak{B}^*$ gehören wird und $\mathfrak{B}^*$ einfach zusammenhängend ist, so gehören alle w mit $|w| \leqq r$ zu $\mathfrak{B}^*$. Das kann aber nur sein, wenn alle w-Werte $|w| \leqq r$ zu $\mathfrak{B}$ gehörten und der Beweis wird wie vorhin vollendet.

LITERATURHINWEISE

[1] HAYMAN, W. K., *Some Applications on the Transfinite Diameter to the Theory of Functions*, J. Analyse Math. *1*, 155–179 (1951).

[2] PÓLYA, G. and SZEGÖ, G., *Inequalities on the Capacity of a Condenser*, Amer. J. Math. *67*, 1–31 (1945).

[3] HAYMAN, W. K., *Multivalent Functions* (Cambridge 1958 [Cambridge Tracts in Mathematics and Mathematical Physics, No. *48*]).

Mathematisches Institut der Ungarischen Akademie der Wissenschaften

Das Minimum von $D/f_{11}\, f_{22}\, \ldots\, f_{55}$ für reduzierte positive quinäre quadratische Formen

B. L. VAN DER WAERDEN (Zürich)

Alexander Ostrowski gewidmet

§ 1. Problemstellung

Für reduzierte positive quadratische Formen in 5 Veränderlichen

$$f(x) = \sum_{1}^{5} f_{ii} x_i^2 + \sum_{i<k} f_{ik} x_i x_k \tag{1}$$

gilt nach Minkowski die „fundamentale Ungleichung"

$$D \geqq c f_{11} f_{22} f_{33} f_{44} f_{55} \tag{2}$$

wo D die Diskriminante ist[1]). Es fragt sich, was der beste Wert von c, also die untere Grenze des Quotienten

$$Q = \frac{D}{f_{11} \cdots f_{55}} \tag{3}$$

ist. Wir werden beweisen: $\operatorname{Min} Q = \frac{1}{8}$.

§ 2. Vorbereitungen

Wenn die Form f reduziert ist, so bedeutet das, dass die Basisvektoren $\mathbf{e}_1, \ldots, \mathbf{e}_5$ des Gitters so gewählt sind:

$\mathbf{e}_1$ ist ein kürzester Gittervektor $\neq 0$.

$\mathbf{e}_2$ ist ein kürzester Gittervektor, der mit $\mathbf{e}_1$ zusammen ein primitives System bildet, u.s.w. bis:

$\mathbf{e}_5$ ist ein kürzester Gittervektor, der mit $\mathbf{e}_1, \ldots, \mathbf{e}_4$ zusammen das ganze Gitter aufspannt.

Jeder Gittervektor $\mathbf{x}$ kann ganzzahlig durch die Basisvektoren ausgedrückt werden:

$$\mathbf{x} = \mathbf{e}_1 x_1 + \cdots + \mathbf{e}_5 x_5 .$$

Die *Norm* des Vektors $\mathbf{x}$ ist der Wert der Form $f(x)$:

$$N(\mathbf{x}) = f(x) .$$

1) Siehe etwa B. L. van der WAERDEN, *Reduktionstheorie der positiven quadratischen Formen*, Acta mathematica *96*, 265 (1956). Diese Arbeit wird im folgenden mit R zitiert.

Received December 6, 1967

Die *sukzessiven Minima* der Form und die zugehörigen Minimumvektoren $\mathbf{m}_1, \ldots, \mathbf{m}_5$ werden nach R so definiert:

$\mathbf{m}_1$ ist ein kürzester Gittervektor $\neq 0$.

$\mathbf{m}_2$ ist ein kürzester von $\mathbf{m}_1$ linear unabhängiger Gittervektor, u.s.w. bis:

$\mathbf{m}_5$ ist ein kürzester von $\mathbf{m}_1, \ldots, \mathbf{m}_4$ linear unabhängiger Gittervektor. Die Normen der Minimumvektoren sind die sukzessiven Minima:

$$N_i = N(\mathbf{m}_i) \quad (i = 1, \ldots, 5).$$

Nach R, § 7 kann man immer $\mathbf{m}_1, \ldots, \mathbf{m}_4$ gleich $\mathbf{e}_1, \ldots, \mathbf{e}_4$ wählen. Dann ist also

$$\mathbf{m}_i = \mathbf{e}_i, \quad N_i = N(\mathbf{e}_i) = f_{ii} \quad (i = 1, \ldots, 4).$$

Für $\mathbf{m}_5$ gibt es zwei Möglichkeiten:

1) Entweder das fünfte Minimum N_5 ist gleich $f_{55} = N(\mathbf{e}_5)$; dann kann man $\mathbf{m}_5 = \mathbf{e}_5$ wählen und man hat

$$Q = \frac{D}{f_{11} \cdots f_{55}} = \frac{D}{N_1 \ldots N_5} \geqq \tfrac{1}{8} \tag{4}$$

wobei der Wert $\frac{1}{8}$ zur dichtesten Kugelpackung in 5 Dimensionen gehört.

2) Oder N_5 ist kleiner als $f_{55} = N(\mathbf{e}_5)$. Setzt man nun

$$\mathbf{m}_5 = \mathbf{e}_1 t_1 + \cdots + \mathbf{e}_5 t_5, \tag{5}$$

so kann man t_5 als positiv annehmen (sonst würde man $\mathbf{m}_5$ durch $-\mathbf{m}_5$ ersetzen). Wäre $t_5 = 1$, so würde $\mathbf{m}_5$ mit $\mathbf{e}_1, \ldots, \mathbf{e}_4$ das ganze Gitter aufspannen und man könnte $\mathbf{e}_5$ durch den kürzeren Vektor $\mathbf{m}_5$ ersetzen, was nicht geht. Also ist t_5 mindestens 2.

Die Gittervektoren $\mathbf{x}$ mit $x_5 = 0$ bilden ein Gitter in dem von $\mathbf{e}_1, \ldots, \mathbf{e}_4$ aufgespannten vierdimensionalen Teilraum R_4. Die mit $x_5 = 1$ nennen wir „Vektoren der ersten Schicht", die mit $x_5 = 2$ „Vektoren der zweiten Schicht", u.s.w. Die Schichten liegen in äquidistanten parallelen Hyperebenen $x_5 = 0, x_5 = 1$, etc. Der Abstand zwischen je zwei aufeinanderfolgenden Hyperebenen sei $h_5 = h$. Die Diskriminante der Form, die aus $f(x)$ entsteht, indem man $x_5 = 0$ setzt, sei D_4. Dann ist nach R

$$D = D_4 \cdot h^2.$$

Wäre $t_5 > 2$, d.h. wäre $\mathbf{m}_5$ ein Vektor der dritten oder einer noch höheren Schicht, so wäre

$$N_5 = N(\mathbf{m}_5) \geqq 9 h^2, \tag{6}$$

also

$$D = D_4 \cdot h^2 \leqq f_{11} f_{22} f_{33} f_{44} \cdot \tfrac{1}{9} N_5 = \tfrac{1}{9} N_1 N_2 N_3 N_4 N_5,$$

was zur Ungleichung von Minkowski

$$D \geqq \tfrac{1}{8} N_1 N_2 N_3 N_4 N_5$$

im Widerspruch steht. Somit kommt im Falle 2) nur $t_5=2$ in Frage. Statt (5) kann man also im Falle 2) schreiben

$$\mathbf{m}_5 = \mathbf{e}_1 t_1 + \mathbf{e}_2 t_2 + \mathbf{e}_3 t_3 + \mathbf{e}_4 t_4 + \mathbf{e}_5 \cdot 2 . \tag{7}$$

Diejenigen Gittervektoren

$$\mathbf{x} = \mathbf{e}_1 x_1 + \cdots + \mathbf{e}_4 x_4 + \mathbf{e}_5 x_5$$

bei denen x_5 gerade ist, lassen sich durch

$$\mathbf{e}_1 = \mathbf{m}_1, \ldots, \mathbf{e}_4 = \mathbf{m}_4 \text{ und } \mathbf{m}_5$$

ausdrücken. Sie bilden im Gitter G ein Teilgitter H vom Index 2, das von den Minimalvektoren $\mathbf{m}_1, \ldots, \mathbf{m}_5$ aufgespannt wird. Die Gittervektoren $\mathbf{x}$ mit ungeradem x_5 bilden eine Nebenklasse H' von H in G.

In der Wahl der Minimumvektoren $\mathbf{m}_1, \ldots, \mathbf{m}_5$ steckt eine Willkür, aber ihre Normen $N_1, \ldots, N_5$ sind nach R, § 5 eindeutig bestimmt. Auch das Teilgitter H ist eindeutig bestimmt; es wird nämlich von allen Gittervektoren mit Normen $\leqq N_5$ aufgespannt. Die Vektoren $\mathbf{x}'$ von H' haben im Falle 2) Normen $>N_5$ und das Minimum dieser Normen ist $f_{55}=N(\mathbf{e}_5)$. Somit ist f_{55} eindeutig durch das Gitter bestimmt, unabhängig von der Wahl der Basisvektoren $\mathbf{e}_1, \ldots, \mathbf{e}_4$. Diese Eindeutigkeit von f_{55} gilt in beiden Fällen 1) und 2); denn im Falle 1) ist ja $f_{55}=N_5$.

In folgenden werden wir immer $N_1=f_{11}=1$ annehmen. Das ist eine ganz unwesentliche Einschränkung, denn man kann die Form $f(x)$ immer durch N_1 dividieren.

§ 3. Reduktion auf den Fall $N_1 = N_2 = \cdots = N_5 = 1$

Satz 1. *Bei der Bestimmung der unteren Grenze der Quotienten Q kann man sich auf solche Formen beschränken, deren sukzessive Minima N_i alle gleich* 1 *sind.*

Satz 1 wird bewiesen sein, sobald gezeigt ist, dass es zu jeder Form f eine Form f^* mit $Q^* \leqq Q$ gibt, deren sukzessive Minima alle Eins sind.

Allgemein gilt

$$N_1 \leqq N_2 \leqq \cdots \leqq N_5 . \tag{8}$$

Wir nehmen nun an, dass $N_1, \ldots, N_k$ alle gleich 1 sind, aber N_{k+1} grösser als 1:

$$1 = N_1 = \cdots = N_k < N_{k+1} . \tag{9}$$

Wir führen (wie in R, § 2) rechtwinklige Koordinaten $\xi_1, \ldots, \xi_5$ ein, wobei jedes ξ_i nur von $x_1, \ldots, x_i$ abhängt. Nun kann man das Gitter in ein anderes Gitter im gleichen metrischen Raum transformieren, durch die affine Transformation

$$\left.\begin{aligned} \xi'_1 &= \xi_1 \\ &\cdots \\ \xi'_k &= \xi_k \\ \xi'_{k+1} &= \beta\,\xi_{k+1} \\ &\cdots \\ \xi'_5 &= \beta\,\xi_5 \end{aligned}\right\} \tag{10}$$

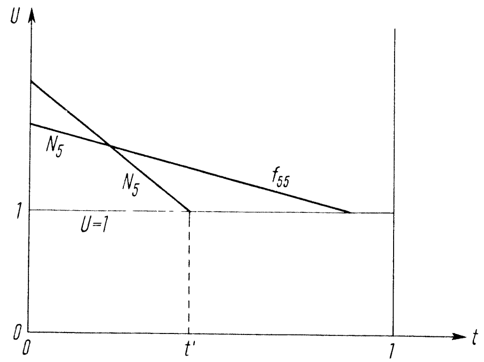
die jedem Punkt P mit Koordinaten ξ_i einen Punkt P' mit Koordinaten ξ'_i zuordnet.

Wir setzen $\beta^2 = 1 - t$ und studieren das Verhalten der Minima $N_1, \ldots, N_5$, wenn t stetig von 0 bis nahe bei 1 anwächst. Für $t=0$ haben $N_1, \ldots, N_5$ die ursprünglichen Werte. Wenn t wächst, bleiben $N_1, \ldots, N_k$ zunächst konstant, aber N_{k+1} nimmt stetig ab und wird für einen gewissen t-Wert gleich N_k. Um das einzusehen, untersuchen wir, wie die Norm eines Gittervektors $\mathbf{x}$ sich durch die Transformation T ändert. Wird $\mathbf{x}$ durch die Transformation T in $\mathbf{x}'$ transformiert, so gilt

$$\left.\begin{aligned} N(\mathbf{x}') &= \xi_1'^2 + \cdots + \xi_5'^2 \\ &= \xi_1^2 + \cdots + \xi_k^2 + \beta^2\,\xi_{k+1}^2 + \cdots + \beta^2\,\xi_5^2 \\ &= \xi_1^2 + \cdots + \xi_k^2 + (1 - t)\,(\xi_{k+1}^2 + \cdots + \xi_5^2). \end{aligned}\right\} \tag{11}$$

Die Norm $N(\mathbf{x}')$ ist also eine lineare Funktion von t. Sie ist konstant, wenn $\mathbf{x}$ in dem von $\mathbf{e}_1, \ldots, \mathbf{e}_k$ aufgespannten Teilraum R_k liegt, weil dann $\xi_{k+1}, \ldots, \xi_5$ alle Null sind. Liegt aber $\mathbf{x}$ nicht in R_k, so ist $N(\mathbf{x}')$ eine abnehmende lineare Funktion von t.

Die Gittervektoren in R_k haben konstante Normen, die von der Transformation T nicht beeinflusst werden. Die Normen aller anderen Gittervektoren sind wegen (9) für $t=0$ grösser als Eins. Nimmt t zu, so nehmen die Normen $u=N(\mathbf{x}')$ linear ab. Stellen wir sie als Funktionen von t graphisch dar, so erhalten wir in der (t, u)-Ebene



Figur 1
Die Normen der Gittervektoren als Funktionen von t im Falle $k=4$.

lauter Geraden mit negativer Steigung. N_{k+1} ist das Minimum der Normen der nicht in R_k liegenden Gittervektoren. Solange keine der abnehmenden Geraden in Fig. 1 die Gerade $u=1$ schneidet, bleiben $N_1, \ldots, N_k$ gleich 1 und N_{k+1} bleibt grösser als Eins. Wir wollen zeigen, dass es einen t-Wert zwischen 0 und 1 gibt, für den N_{k+1} gleich 1 wird.

Nach R, § 2 gilt

$$D = h_1^2 h_2^2 h_3^2 h_4^2 h_5^2 , \qquad (12)$$

wobei die Höhen $h_1, \ldots, h_k$ im Raum R_k liegen und $h_{k+1}, \ldots, h_5$ senkrecht dazu. Bei der Transformation T bleiben $h_1^2, \ldots, h_k^2$ ungeändert, während $h_{k+1}^2, \ldots, h_5^2$ mit $1-t$ multipliziert werden. Also wird D mit $(1-t)^{5-k}$ multipliziert:

$$D' = (1-t)^{5-k} D . \qquad (13)$$

Nun ist aber nach MINKOWSKI

$$D \geqq \tfrac{1}{8} N_1 N_2 \ldots N_5 > \tfrac{1}{8}$$

also $8D > 1$. Bei der Transformation T werden $N_{k+1}, \ldots, N_5$ verkleinert. Solange sie grösser als 1 bleiben, bleibt die Diskriminante grösser als $\frac{1}{8}$:

$$D' = (1-t)^{5-k} D > \tfrac{1}{8} .$$

Daraus folgt $(1-t)^{5-k} > (8D)^{-1}$.

Setzen wir

$$(8D)^{-1} = (1-\tau)^{5-k}$$

so folgt $1-t > 1-\tau$, also $t < \tau$. Spätestens an der Stelle $t=\tau$ muss die stückweise lineare Funktion N_{k+1} also den Wert 1 annehmen. Der Wert von t, für den $N_{k+1}=1$ wird, sei t'. Man erhält dann für $t=t'$ durch die Transformation T ein neues Gitter mit

$$1 = N_1' = \cdots = N_k' = N_{k+1}' .$$

Wir zeigen nun, dass der Quotient Q sich bei der Transformation T nicht vergrössern kann.

Am einfachsten ist der Beweis im Falle $k=4$. In diesem Falle bleibt der von $\mathbf{e}_1, \ldots, \mathbf{e}_4$ aufgespannte lineare Raum R_4 bei der Transformation T ungeändert. Die Minimalvektoren $\mathbf{m}_1, \ldots, \mathbf{m}_4$ liegen in diesem Raum und bleiben ebenfalls ungeändert. Ihre Normen $N_1, \ldots, N_4$ sind vor und nach der Transformation T gleich Eins. Alle Abstände senkrecht zu diesem Raum werden bei der Transformation mit β multipliziert. Die Abstände zwischen den Schichten der Fig. 1 werden ebenfalls mit β multipliziert; die Reihenfolge der Schichten bleibt ungeändert. Betrachten wir nun den Quotienten

$$Q = \frac{D}{f_{11} \cdots f_{55}} = \frac{D}{N_1 \ldots N_4 \cdot f_{55}} = \frac{D}{f_{55}} ,$$

so wird der Zähler D bei der Transformation T mit $\beta^2=(1-t)$ multipliziert. Der Nenner ist die kleinste Norm eines Vektors der ersten Schicht. Ist $\mathbf{x}'$ ein Vektor, der nach der Transformation T die kleinste Norm hat, so hat man nach (11)

$$f'_{55}=N(\mathbf{x}')\geqq(1-t)\cdot N(\mathbf{x})\geqq(1-t)f_{55}.$$

Daraus folgt

$$Q'=\frac{D'}{f'_{55}}\leqq\frac{(1-t)\,D}{(1-t)f_{55}}=\frac{D}{f_{55}}=Q.$$

Damit ist der Fall $k=4$ erledigt.

Ist $k<4$, so kann eine Schwierigkeit auftreten, weil die Minimumvektoren $\mathbf{m}_1,\dots,\mathbf{m}_4$ nach der Transformation nicht mehr Minimumvektoren zu sein brauchen und der von ihnen aufgespannte Raum R_4 nach der Transformation vielleicht nicht mehr der von den Minimumvektoren aufgespannte Raum ist.

Es gibt insgesamt nur endlich viele Vektoren, die vor oder nach der Transformation T als Minimumvektoren $\mathbf{m}_i$ oder Basisvektoren e_i in Betracht kommen; denn es gibt nur endlich viele Gittervektoren, die nach der Transformation Normen $\leqq f_{55}$ haben. Alle anderen Gittervektoren haben nach der Transformation und daher um so mehr vor der Transformation Normen grösser als f_{55} und daher auch grösser als N_5.

Diesen endlich vielen Gittervektoren entsprechen endlich viele Geraden in der (t, u)-Ebene.

Sie haben nur endlich viele Schnittpunkte, deren Abzissen die Strecke von 0 bis t' in endlich viele Teilstrecken teilen. Auf jeder dieser Teilstrecken bleiben die Räume $R_1,\dots,R_4$ konstant. Die Minimumvektoren $\mathbf{m}_1,\dots,\mathbf{m}_4$ und $\mathbf{e}_5$ ändern sich auf jeder Teilstrecke stetig und ihre Normen sind lineare Funktionen von t.

Auf der ersten Teilstrecke kann die Funktion

$$Q=\frac{D}{f_{11}\cdots f_{55}}=\frac{D}{N_1\dots N_4\cdot f_{55}} \tag{14}$$

nur verkleinert, nicht vergrössert werden. Der Zähler in (14) wird nämlich bei der Transformation T mit $(1-t)^{n-k}$ multipliziert und der Nenner mit dem gleichen oder einem grösseren Faktor. Also ist der Wert Q_1 am Ende der ersten Teilstrecke höchstens gleich dem ursprünglichen Wert Q_0.

Am Ende der ersten Teilstrecke ist der Wert Q_1 von Q eindeutig bestimmt, unabhängig von der Wahl der Minimumvektoren $\mathbf{m}_1,\dots,\mathbf{m}_4$, denn die Faktoren N_i und f_{55} sind eindeutig bestimmt. Also kann man für die zweite Teilstrecke genau so weiter schliessen. So erhält man schliesslich am Ende der letzten Teilstrecke eine reduzierte Form f^* mit

$$Q^*\leqq Q \quad\text{und}\quad 1=N_1^*=\cdots=N_k^*=N_{k+1}^*.$$

Sind nun alle N_i^* gleich Eins, so sind wir fertig. Sind aber nur k' von den Grössen N_i^* gleich Eins $(k<k'<5)$, so kann man den gleichen Beweis mit k' statt k noch einmal führen. Nach höchstens vier Schritten kommt man zum Ziel:

$$N_1 = \cdots = N_5 = 1 . \tag{15}$$

Damit ist Satz 1 bewiesen.

§ 4. Der konvexe Bereich K

Wenn (15) erfüllt ist, so gilt auch

$$f_{11} = f_{22} = f_{33} = f_{44} = 1 \tag{16}$$

und (siehe R, § 7)

$$f_{55} \leqq \tfrac{5}{4} . \tag{17}$$

Da weiter für reduzierte Formen

$$|f_{ik}| \leqq f_{ii} \quad \text{für} \quad i<k \tag{18}$$

gilt, so sind alle f_{ik} absolut beschränkt.

Die Reduktionsbedingungen

$$f(s_1, \ldots, s_5) \geqq f_{kk} \quad \text{für} \quad (s_k, \ldots, s_5) = 1 \tag{19}$$

definieren einen konvexen Bereich B im 15-dimensionalen Raum R_{15} der Koeffizienten f_{ik}. Zu den Bedingungen (19) gehören die folgenden:

$$\left.\begin{aligned} N(\mathbf{m}_2) &= f_{22} \geqq f_{11} \\ N(\mathbf{m}_3) &= f_{33} \geqq f_{11} \\ N(\mathbf{m}_4) &= f_{44} \geqq f_{11} \\ N(\mathbf{m}_5) &= f(t_1, t_2, t_3, t_4, t_5) \geqq f_{11} . \end{aligned}\right\} \tag{20}$$

Ist nun

$$N_1 = N_2 = N_3 = N_4 = N_5 , \tag{21}$$

so gilt in den vier Ungleichungen (20) das Gleichheitszeichen. Durch die vier Gleichungen, die man so erhält, ist für jeden Satz Koeffizienten $t_1, \ldots, t_5$ ein linearer elfdimensionaler Teilraum R_{11} in R_{15} definiert. Durch die Normierung $f_{11} = 1$ wird R_{11} auf einen R_{10} reduziert.

Die Räume R_{10} und R_{11} hängen noch von den ganzzahligen Koeffizienten $t_1, \ldots, t_5$ ab. Im Falle 1) des § 2 haben $t_1, \ldots, t_5$ die Werte 0, 0, 0, 0, 1. Im Falle 2) ist $t_5 = 2$; die übrigen t sind, wie wir noch sehen werden, ungerade Zahlen[2]). Jedenfalls sind nur endlich viele Wertsysteme $t_1, \ldots, t_5$ möglich (R, § 10), also gibt es auch nur endlich

---

[2]) Minkowski hat gefunden, dass $t_1, \ldots, t_4$ alle gleich ± 1 sein müssen (*Gesammelte Werke* I, S. 154). Jedoch werden wir dieses Ergebnis nicht brauchen.

viele Räume R_{10}, in denen wirklich reduzierte Formen mit $f_{55}>1$ liegen. In jedem Raum R_{10} definieren die Reduktionsbedingungen (19) einen beschränkten konvexen Bereich K.

§ 5. Das Minimum von Q in K

SATZ 2. *Das Minimum von Q in K wird in einer Ecke des Polyeders K angenommen.*

Beweis: Die Funktion

$$Q=\frac{D}{f_{11}\cdots f_{55}}=\frac{D}{f_{55}}$$

ist stetig, hat also in K ein Minimum c. Dann hat $D-cf_{55}$ das Minimum Null. Die Funktion D ist aber nach MINKOWSKI konkav (R, § 14), und f_{55} ist linear, also ist $D-cf_{55}$ ebenfalls konkav. Eine konkave Funktion nimmt ihr Minimum in einer Ecke des konvexen Körpers K an. In dieser Ecke ist also $D-cf_{55}=0$, also $D/f_{55}=c$. Damit ist Satz 2 bewiesen.

Bei der Anwendung des Satzes 2 muss man zwischen den Fällen 1) und 2) unterscheiden. Im Falle 1) ist $f_{55}=N_5=1$. Diese Gleichung gilt im ganzen Bereich K, den wir in diesem Fall K_1 nennen. Das Minimum von $Q=D/f_{55}$ in K_1 ist gleich dem Minimum von D, also gleich $\frac{1}{8}$.

Der Fall 2) ist durch die Ungleichung $f_{55}>1$ gekennzeichnet. Diese Ungleichung gilt in den inneren Punkten des konvexen Bereiches, den wir in diesen Fall K_2 nennen. In gewissen Randpunkten von K_2 kann sehr wohl $f_{55}=1$ sein. Diese Randpunkte gehören dan sowohl zu K_2 als zu K_1. In solchen Randpunkten ist natürlich $Q\geqq\frac{1}{8}$. Wir haben zu untersuchen, ob es andere Ecken von K_2 geben kann, in denen $Q<\frac{1}{8}$ ist. Die Antwort wird sein: Nein. Es gibt zwar ein lokales Minimum von Q in einer Ecke E, aber dieses lokale Minimum ist $\frac{1}{5}$.

§ 6. Das raumzentrierte Gitter im Falle 2)

Für die folgenden Rechnungen wollen wir nicht $\mathbf{e}_1,\ldots,\mathbf{e}_5$, sondern $\mathbf{m}_1,\ldots,\mathbf{m}_5$ als Basisvektoren wählen. Sie spannen allerdings nicht das ganze Gitter G, sondern nur das Teilgitter H vom Index 2 auf. Die Basisvektoren $\mathbf{e}_1,\ldots,\mathbf{e}_4$ sind gleich $\mathbf{m}_1,\ldots,\mathbf{m}_4$. Um $\mathbf{e}_5$ durch $\mathbf{m}_1,\ldots,\mathbf{m}_5$ auszudrücken, muss man (7) nach $\mathbf{e}_5$ auflösen. Das gibt

$$\mathbf{e}_5=\mathbf{m}_1c_1+\cdots+\mathbf{m}_4c_4+\mathbf{m}_5\cdot\tfrac{1}{2}, \tag{22}$$

wobei $c_1,\ldots,c_4$ ganze oder halbganze Zahlen sind.

Sind alle c_i ganze Zahlen plus $\frac{1}{2}$, so haben wir ein *raumzentriertes Gitter*. Wenn aber einige c_i ganz sind, so haben wir ein *seitenzentriertes Gitter*. Wir zeigen, dass der letztere Fall nicht vorkommen kann.

Es sei etwa c_1 eine ganze Zahl. Dann ligt der Vektor

$$\mathbf{e}_5' = \mathbf{e}_5 - \mathbf{m}_1 c_1 = \mathbf{m}_2 c_2 + \mathbf{m}_3 c_3 + \mathbf{m}_4 c_4 + \mathbf{m}_5 \cdot \tfrac{1}{2}$$

in dem von $\mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4, \mathbf{m}_5$ aufgespannten vierdimensionalen Raum R_4', und zwar in der ersten Schicht, da der letzte Koeffizient $\frac{1}{2}$ ist. Die vier sukzessiven Minima des Gitters in R_4' sind alle 1, weil $\mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4$ und $\mathbf{m}_5$ die Norm 1 haben. Nach R, § 7 liegt dann aber in der ersten Schicht ein Vektor $\mathbf{e}''$ kleinster Norm, dessen Norm gleich dem vierten Minimum, also gleich Eins ist. Dieser Vektor wäre kürzer als $\mathbf{e}_5$, was nicht geht, da $\mathbf{e}_5$ ein kürzester Vektor der ersten Schicht des ganzen Gitters sein sollte.

Es bleibt also nur den Fall des raumzentrierten Gitters. Die c_i sind in diesem Fall ganze Zahlen plus $\frac{1}{2}$. Die Gittervektoren sind

$$\mathbf{z} = \mathbf{m}_1 z_1 + \cdots + \mathbf{m}_5 z_5 , \tag{23}$$

wobei die Koeffizienten $z_1, \ldots, z_5$ entweder alle ganz (dann liegt z im Teilgitter H) oder alle halbganz sind (dann liegt z in der Nebenklasse H').

Die Norm des Vektors $\mathbf{z}$ ist eine quadratische Form in den Koeffizienten $z_1, \ldots, z_5$, die wir $g(z)$ nennen:

$$N(\mathbf{z}) = \sum g_{ii} z_i^2 + \sum g_{ik} z_i z_k = g(z), \tag{24}$$

wobei die zweite Summe nur über die Paare i, k mit $i<k$ zu erstrecken ist.

Lässt man in (24) für die z_i nur ganzzahlige Werte zu, so erhält man aus (24) die Normen der Vektoren des Teilgitters H. Die Form g ist also die zu diesem Teilgitter gehörige quadratische Form.

Die Koeffizienten f_{ik} der Form f sind einfache lineare Funktionen der Koeffizienten g_{ik} der Form g (und umgekehrt). Insbesondere ist

$$\left.\begin{aligned} f_{55} = N(\mathbf{e}_5) &= g(c_1, c_2, c_3, c_4, \tfrac{1}{2}) \\ &= \sum g_{ii} c_i^2 + \sum g_{ik} c_i c_k \quad \text{mit} \quad c_5 = \tfrac{1}{2}, \end{aligned}\right\} \tag{25}$$

wobei die c_i ganze Zahlen plus $\frac{1}{2}$ sind.

§ 7. Die Reduktionsbedingungen

Im folgenden betrachten wir ausschliesslich solche Formen g, welche die Bedingungen

$$N_1 = N_2 = \cdots = N_5 = 1$$

oder

$$g_{11} = g_{22} = \cdots = g_{55} = 1 \tag{26}$$

erfüllen.

Die Reduktionsbedingungen für die Form f lassen sich sehr einfach durch die

Koeffizienten der Form g ausdrücken. Sie lauten, wenn (26) berücksichtigt wird:

$$g(z_1, \ldots, z_5) \geqq 1 \quad \text{für} \quad (z_1, \ldots, z_5) \neq (0, \ldots, 0) \tag{27}$$

$$g(z_1, \ldots, z_5) \geqq f_{55} \quad \text{für} \quad z_5 = \tfrac{1}{2}, \tag{28}$$

wobei f_{55} durch (25) gegeben ist. Das Problem ist, unter diesen Bedingungen das Minimum von $Q = D/f_{55}$ zu bestimmen.

Zu den Nebenbedingungen dieses Minimumproblems gehört auch noch die Gleichung (25):

$$f_{55} = g(c_1, c_2, c_3, c_4, \tfrac{1}{2}).$$

Diese ist lästig, da die halbganzen Zahlen c_1, c_2, c_3, c_4 nicht bekannt sind. Wir ersetzen daher f_{55} lieber durch eine neue Unbekannte v. Genauer stellen wir das Problem so: Das Minimum von $Q = D/v$ ist zu bestimmen unter den Nebenbedingungen

$$g_{11} = g_{22} = \cdots = g_{55} = 1 \tag{26}$$

$$g(z_1, \ldots, z_5) \geqq 1 \quad \text{für} \quad (z_1, \ldots, z_5) \neq (0, \ldots, 0) \tag{27}$$

$$g(z_1, \ldots, z_5) \geqq v \quad \text{für} \quad z_5 = \tfrac{1}{2} \tag{29}$$

$$v \geqq 1, \tag{30}$$

wobei die z_i in (27) und (29) entweder alle ganz oder alle ganz plus $\frac{1}{2}$ sein sollen.

Die Diskriminante D des Gitters G ist ein Viertel der Diskriminante des Teilgitters H oder der Form g, und diese hängt nur von den g_{ik} ab. Hält man also die g_{ik} fest und vergrössert v so weit als möglich, ohne (29) zu verletzen, so wird $Q = D/v$ kleiner. Das Minimum von Q wird also nur dann erreicht, wenn in einer von den Ungleichungen (29) das Gleichheitszeichen gilt, also wenn für gewisse halbganze $c_1, \ldots, c_4$

$$g(c_1, c_2, c_3, c_4, \tfrac{1}{2}) = v$$

gilt. Das bedeutet: An der Stelle des Minimums von Q ist v genau gleich dem f_{55} der Gleichung (25). Um die Gleichung (25) brauchen wir uns jetzt nicht mehr zu kümmern: sie ist beim Minimum automatisch erfüllt.

Wir haben jetzt ein Minimumproblem in dem 11-dimensionalen Raum R' der Unbekannten g_{ik} $(i<k)$ und v zu lösen. Die Einschränkungen (27), (29) und (30) definieren in diesem Raum R' einen beschränkten Bereich B'. Das Minimum von Q in B' kann nur in einer Ecke des Polyeders B' angenommen werden.

§ 8. Die vergrösserte Bereich B'

Um das Problem zu vereinfachen, lassen wir von den Ungleichungen (27) und (29) zunächts einige ausser Betracht. Nur die folgenden sollen beibehalten werden:

$$v \geqq 1 \tag{30}$$

$$g(z_1, \ldots, z_5) \geqq v \quad \text{mit} \quad z_i = \pm \tfrac{1}{2}. \tag{31}$$

Diese 33 Ungleichungen definieren einen konvexen Körper K', der B' umfasst. Wir werden gleich sehen, dass K' beschränkt ist. In K' hat $Q = D/v$ ein Minimum; es wird notwendigerweise in einer Ecke angenommen. Wir durchmustern nun alle Ecken von K' mit folgendem Ergebnis. In einer einzigen Ecke E ist $v = f_{55} > 1$; diese Ecke erfüllt sämtliche Ungleichungen (27), (29) und ergibt ein lokales Minimum $Q = \frac{1}{5}$. In allen anderen Ecken ist $v = f_{55} = 1$. Die zugehörigen Formen sind sämtlich reduziert; folglich gilt für sie $D \geqq \frac{1}{8}$ und daher wegen $v = 1$ auch $Q \geqq \frac{1}{8}$. Also ist das Minimum von Q im Bereich K' mindestens $\frac{1}{8}$. Um so mehr gilt das im Teilbereich B'.

Um den eben skizzierten Gedankengang auszuführen, ziehen wir zunächst einige Folgerungen aus den Ungleichungen (31). Eine von ihnen lautet

$$g(\tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}, \tfrac{1}{2}) \geqq v, \tag{32}$$

oder mit 4 multipliziert

$$g(1, 1, 1, 1, 1) \geqq 4v \geqq 4. \tag{33}$$

Genau so beweist man allgemein

$$g(\pm 1, \pm 1, \pm 1, \pm 1, \pm 1) \geqq 4. \tag{34}$$

Schreibt man (33) ganz aus und bringt die 5 Glieder $g_{ii} = 1$ auf die rechte Seite, so erhält man

$$\sum g_{ik} \geqq -1, \tag{35}$$

wobei wie immer über alle Paare ik mit $i > k$ zu summieren ist. Genau so ergibt (34) eine Ungleichung von der Form

$$\sum \pm g_{ik} \geqq -1. \tag{36}$$

Die Ungleichung (35), die aus (33) entstanden ist, bezeichnen wir als Ungleichung $+++++$. Ersetzt man etwa die letzten beiden $+$ durch $-$, so erhält man die Ungleichung $+++--$, etc. Insgesamt gibt es 32 Ungleichungen (36), entsprechend den 32 Folgen von je 5 Vorzeichen. Von diesen 32 Ungleichungen sind aber nur 16 verschieden, weil die Unkehrung aller 5 Vorzeichen jeweils dieselbe Ungleichung ergibt.

Addiert man die zwei Ungleichungen (36), deren Vorzeichenreihe mit $++++$ anfängt, und dividiert durch 2, so erhält man

$$g_{12} + g_{13} + g_{14} + g_{23} + g_{24} + g_{34} \geqq -1 \tag{37}$$

oder

$$g(1, 1, 1, 1, 0) \geqq 3. \tag{38}$$

Genau so beweist man für jede Folge von 4 Vorzeichen

$$g(\pm 1, \pm 1, \pm 1, \pm 1, 0) \geqq 3, \tag{39}$$

wobei die Null auch an einer anderen Stellen stehen darf, z.B.

$$g(1, 0, 1, -1, 1) \geqq 3.$$

Addiert man die vier Ungleichungen (36), deren Vorzeichenreihe mit $+++$ anfangen, und dividiert durch 4, so erhält man

$$g(1, 1, 1, 0, 0) \geqq 2 \tag{40}$$

und ebenso allgemein

$$g(\pm 1, \pm 1, \pm 1, 0, 0) \geqq 2, \tag{41}$$

wobei die Nullen auch an anderen Stellen stehen dürfen.

Addiert man die acht Ungleichungen (36), deren Vorzeichen mit $++$ anfangen, und dividiert durch 8, so erhält man $g_{12} \geqq -1$ oder

$$g(1, 1, 0, 0, 0) \geqq 1. \tag{42}$$

und ebenso

$$g(\pm 1, \pm 1, 0, 0, 0) \geqq 1, \tag{43}$$

wobei die Nullen auch an anderen Stellen stehen dürfen.

Schliesslich hat man wegen (26)

$$g(1, 0, 0, 0, 0) = 1, \tag{44}$$

wobei die Eins an beliebiger Stelle stehen darf.

§ 9. Schaleneinteilung

Wir teilen die Gittervektoren mit Ausnahme des Nullvektors in „Schalen" ein. Zur ersten Schale rechnen wir die Vektoren $\mathbf{z}$, deren Koordinaten z_i alle $\pm\frac{1}{2}$ sind. Die n-te Schale besteht aus den Vektoren mit

$$\text{Max.}\,|z_i| = \frac{n}{2}.$$

Die Ungleichungen (31) besagen, dass die Vektoren der ersten Schale Normen $\geqq v$ haben. Die daraus folgenden Ungleichungen (34), (39), (41), (43) und die Gleichung (44) besagen, dass die Vektoren der zweiten Normen $\geqq 1$ haben, wobei das Minimum 1 wirklich angenommen wird. Wir wollen nun zeigen, dass die Vektoren aller höheren Schalen ebenfalls Normen $\geqq 1$ haben. Daraus folgt insbesondere, dass die zum Teilgitter H gehörige Form $g(x)$ reduziert und daher positiv ist.

Gesetzt, es gäbe eine Form g_1, die zwar die Ungleichungen (31) mit $v \geqq 1$, aber nicht alle Ungleichungen (27) erfüllen würde. Dann würde mindestens eine Ungleichung (27) nicht erfüllt sein, d.h. es gäbe einen Vektor $\mathbf{z}$ in der dritten oder einer höheren Schale mit Norm $N(\mathbf{z})<1$. Wir nehmen nun eine Form g_0, die sämtliche Ungleichungen erfüllt, z.B.

$$g_0(z) = z_1^2 + z_2^2 + z_3^2 + z_4^2 + z_5^2, \tag{45}$$

und wir verbinden die Pukte g_0 und g_i im 10-dimensionalen Raum der Koeffizienten g_{ik} $(i<k)$ durch eine Strecke

$$g = g_0 + t(g_1 - g_0).$$

Es muss einen grössten Wert t geben, für welchen die Form g gerade noch alle Ungleichungen (27) erfüllt. Nennen wir diesen Wert t^*. Für einen etwas grösseren Wert $t^*+\varepsilon$ ist dann eine der Ungleichungen

$$g(z_1, \ldots, z_5) \geqq 1$$

verletzt, d.h. es gilt $g(z_1, \ldots, z_1)<1$. Für den Wert t^* ist aber die Ungleichung $g(z)\geqq 1$ noch erfüllt, also gilt für $t=t^*$ das Gleichheitszeichen

$$g(z_1, \ldots, z_5) = 1, \tag{46}$$

wobei $\mathbf{z}$ ein Vektor aus der dritten oder einer höheren Schale ist. Einer der Beträge $|z_i|$ muss also mindestens $1\frac{1}{2}$ sein. Indem wir, wenn nötig, die Minimumvektoren $m_1, \ldots, m_5$ umnumerieren, können wir $|z_5|\geqq\frac{3}{2}$ erreichen. Ist z_5 negativ, so ersetzen wir den Vektor $\mathbf{z}$ durch $-\mathbf{z}$. Dann wird also $z_5\geqq\frac{3}{2}$, d.h. der Vektor $\mathbf{z}$ liegt in der dritten oder einer höheren Schicht.

Nun gilt aber, wie wir früher gesehen haben – siehe (6) – für alle Vektoren der dritten und der höheren Schichten

$$N(\mathbf{z}) \geqq 9\,h^2.$$

Wenn $\mathbf{z}$ die Norm 1 haben soll, so ist $9\,h^2\leqq 1$, also

$$D = D_4 h^2 \leqq f_{11} f_{22} f_{33} f_{44} \cdot \tfrac{1}{9} = \tfrac{1}{9} < \tfrac{1}{8},$$

was unmöglich ist. Also kann ein Vektor der dritten oder einer höheren Schicht unmöglich die Norm 1 haben, d.h. (46) ist unmöglich.

Damit ist beweisen:

SATZ 3. *Alle Formen g, die die Ungleichungen* (31) *mit $v\geqq 1$ erfüllen, sind reduziert und daher positiv.*

§ 10. Das Minimum von Q

Im Bereich der positiven Formen ist D eine konkave Funktion. Daher gilt der Beweis des Satzes 2 auch für den Bereich K', der durch die Ungleichungen (30) und (31) definiert ist, und wir erhalten

Satz 4. *Das Minimum von Q in K' wird in einer Ecke des Polyeders K' angenommen.*

In (31) kann man sich wegen

$$g(-z_1, \ldots, -z_5) = g(z_1, \ldots, z_5)$$

auf den Fall $z_5 = +\frac{1}{2}$ beschränken. Man hat also nur 16 verschiedene Ungleichungen (31) und eine Ungleichung (30).

Eine Ecke des Polyeders K' wird durch 11 unabhängige Gleichungen definiert, die man erhält, indem man in 11 von den 17 Ungleichungen = statt $\geqq$ schreibt. Nun ergeben sich 2 Möglichkeiten, die unseren früheren Fällen 1) und 2) entsprechen, nämlich:

A. Man schreibt in (30) das Gleichheitszeichen. Dann ist $v = f_{55} = 1$ und man erhält den früheren Fall 1).

B. Man hat in (30) das Zeichen $>$. In diesem Fall muss man aus den Gleichungen

$$g(z_1, \ldots, z_5) = v \tag{47}$$

elf unabhängige auswählen.

Nun haben aber alle Gleichungen (47) eine gemeinsame Lösung, nämlich die Form g_0 – siehe (45) – mit

$$g_{ii} = 1, \quad g_{ik} = 0 \quad (i < k), \quad v = \tfrac{5}{4}. \tag{48}$$

Wählt man also aus den Gleichungen (47) irgend elf unabhängig aus so erhält man als einzige Lösung die Form (48). Das zugehörige Gitter G ist kubisch raumzentriert. Die Diskriminante des Gitters ist $\frac{1}{4}$, also ist der Quotient $Q = D/v$ gleich $\frac{1}{5}$. Dieser Wert ist, wie man leicht sieht, ein lokales Minimum.

Die Form g_0 erfüllt offensichtlich alle Bedingungen (27), (29) und (30). Die zum Gitter G gehörige Form f ist also reduziert. Aber auch im Falle A sind die Bedingungen (27) und (29) erfüllt. Die Bedingungen (27) folgen nämlich, wie wir in § 9 gesehen haben, aus (30) und (31). Die Bedingungen (29) aber folgen aus (27), denn im Falle A ist $v = 1$.

Wir sehen also, dass sämtliche Ecken des konvexen Bereiches K' die Ungleichungen (27) und (29) erfüllen, also zum Bereich B' gehören. Dann aber muss der ganze Bereich K' in B' enthalten sein. Da umgekehrt B' in K' enthalten ist, so folgt

$$B' = K'. \tag{49}$$

Geht man wieder vom elfdimensionalen Raum R' auf R_{10} zurück, indem man $v = f_{55}$ setzt, so erhält man den

Satz 5. *Alle Reduktionsbedingungen* (27) *und* (29) *folgen aus den 17 Bedingungen*

$$g(\pm\tfrac{1}{2}, \pm\tfrac{1}{2}, \pm\tfrac{1}{2}, \pm\tfrac{1}{2}, \tfrac{1}{2}) \geqq f_{55} \tag{50}$$

$$f_{55} \geqq 1, \tag{51}$$

und den 5 Gleichungen

$$g_{11} = g_{22} = \cdots = g_{55} = 1\,. \tag{26}$$

Im Falle A ist $Q=\frac{1}{8}$. Im Falle B ist $Q=\frac{1}{5}$. Also ist Q im Bereich $B'=K'$ überall $\geqq \frac{1}{8}$. Daraus folgt:

SATZ 6. *Das Minimum von Q ist $\frac{1}{8}$. Die dichteste Kugelpackung erreicht das Minimum.*

Mathematisches Institut
der Universität Zürich

Ein allgemeiner Vierscheitelsatz für ebene Jordankurven

I. Mitteilung. Vorbereitende Betrachtungen. Erster Teil des Beweises

OTTO HAUPT (Erlangen)

Herrn Alexander Ostrowski zum 75. Geburtstag mit herzlichen Wünschen

Einleitung

1. In seiner bekannten Arbeit hat A. KNESER [1][1]) den Vierscheitelsatz für stetig gekrümmte Jordankurven C (in der euklidischen Ebene E) bewiesen (Anderer Beweis bei H. KNESER [1]); unter einem Scheitel von C wird dabei jede Maximal- und Minimalstelle der Krümmung verstanden.

2. Nachstehend soll der Knesersche Satz zu einer Aussage verallgemeinert werden, welche der Topologie in E angehört (vgl. die Formulierung in § 3). Es wird nämlich *Erstens* das System $\mathfrak{K}$ der Kreise ersetzt durch irgend ein System $\mathfrak{k}$ von Kurven, deren jede – kurz gesagt (vgl. im Folgenden Nr. 3.1.) – durch $k=k(\mathfrak{k})=3$ ihrer Punkte eindeutig bestimmt und stetige Funktion solcher 3 Punkte ist. Und *Zweitens* tritt an Stelle der Forderung der Stetigkeit der Krümmung eine Forderung, die sich nur auf das Verhalten der Jordankurven C bezüglich $\mathfrak{k}$ in der Umgebung eines jeden isolierten Scheitels bezieht; dabei ist als Scheitel von C bezüglich $\mathfrak{k}$, kurz $\mathfrak{k}$-Scheitel, bezeichnet jeder Punkt $s \in C$ derart, dass in jeder Umgebung von s auf C mindestens 4 Punkte von C auf einem $K \in \mathfrak{k}$ liegen. – Die Existenz von Systemen $\mathfrak{k}$ (gemäss Erstens), die vom System der Kreise verschieden sind, ist bekannt (vgl. z.B. BENZ [1]). – Ferner sind die an einen $\mathfrak{k}$-Scheitel gemäss Zweitens gestellten Anforderungen widerspruchsfrei; sie sind nämlich insbesondere im Falle stetiger Krümmung für die Scheitel bezüglich der Kreise erfüllt (und sogar schon im Falle der Eindeutigkeit des freien Schmiegkreises in einem $\mathfrak{K}$-Scheitel.) Der Satz von KNESER ist mithin als Spezialfall in unserer topologischen Verallgemeinerung enthalten (vgl. H [1]).

BEMERKUNG. Neben topologischen Begriffsbildungen werden zur Vereinfachung einiger Formulierungen Begriffe aus der direkten Infintesimalgeometrie benutzt, nämlich z.B. der der $\mathfrak{k}$-Paratingente als Verallgemeinerung des freien Schmiegkreises.

3. In der vorliegenden Arbeit wird angenommen, dass $C \cap K$ endlich ist für jedes $K \in \mathfrak{k}$. Auf den Fall, dass es $K \in \mathfrak{k}$ mit unendlichem $C \cap K$ gibt, wird später zurückzu-

---

[1]) Ziffern in eckiger Klammer im Text verweisen auf das Literaturverzeichnis am Ende dieser Note.

Eingegangen am 8.12.1967

kommen sein. – Eine Sonderrolle spielt der Fall, dass $C \cap K$ maximal vier Punkte enthält für jedes $K \in \mathfrak{k}$ (vgl. Nr. 3.3.); dieser Fall ist schon früher erledigt (vgl. H.–K. [1], Nr. 4.1.4.1.ff.). Der Gedankengang des Beweises für die übrigen Fälle gliedert sich so: Zunächst werden (§§ 1 und 2) Bemerkungen zusammengestellt, welche die gegenseitige Lage zweier Kurven (in der Ebene) betreffen (und die rein topologischer Natur sind). Im § 3 werden die Vierscheitelsätze im Bezug auf Systeme $\mathfrak{k}$ formuliert und es wird der Beweis zurückgeführt (vgl. Nr. 3.4.3. und 3.4.3.1.) auf den eines Existenzsatzes (genauer auf den eines Kontraktionssatzes), dessen Beweis an anderer Stelle gegeben werden soll.

4. Modifiziert man die in Ziff. 2 oben, Zweitens, angedeuteten Forderungen an die $\mathfrak{k}$-Scheitel ein wenig, so erweist sich jeder isolierte $\mathfrak{k}$-Scheitel als vom Punktordnungswert 4 und als gemeinsamer Endpunkt zweier Teilbogen von C je vom Punktordnungswert 3; auch ergibt sich, im Falle C nur endlich viele $\mathfrak{k}$-Scheitel besitzt, dass die $\mathfrak{k}$-Scheitel von je gleicher Signatur sich gegenseitig trennen, das also gleichviel $\mathfrak{k}$-Scheitel positiver und negativer Signatur vorhanden sind, mindestens also je zwei. Dies soll in einer späteren Note gezeigt werden. Übrigens liefert die ebene nicht-euklidische (hyperbolische) Geometrie ebenfalls einen 4-Scheitelsatz, welcher dem klassischen in der euklidischen (oder vielmehr in der konformen) Ebene entspricht, worauf später zurückzukommen sein wird.

Herzlich zu danken habe ich auch diesmal wieder Herrn Hermann Künneth für vielfach Kritik und Verbesserungsvorschläge (vgl. insbesondere die Nr. 2.2. und 2.3.).

Anmerkung. (I) Ohne zusätzliche Forderungen an dic k-Scheitel (etwa im Sinne von Ziff. 2., Zweitens) gilt nur noch ein *Zweischeitelsatz*, wofür schon im Falle der Kreise gewisse Ovale (mit stetiger Tangente und) mit zwei Unstetigkeiten der Krümmung Beispiele liefern (vgl. allgemein H.–K. [1]. Nr. 4.5.). – (II) Unbeantwortet bleibe hier die Frage, ob der allgemeine Vierscheitelsatz eine Ausnahmestellung einnimmt in dem Sinne, dass bei Systemen $\mathfrak{k}$ mit $\mathfrak{k}(k)=2t+1$, $t \geqslant 2$, kein ähnlich allgemeiner $(k+1)$-Scheitelsatz gilt. Man möchte vermuten, dass der Fall $k(\mathfrak{k})=3$ tatsächlich eine Ausnahme darstellt.

§ 1. Bezeichnungen. Vorbemerkungen

1.1. Der sogenannte Grundbereich $G=\overline{G}$, in welchem sich die Betrachtungen abspielen, sei eine abgeschlossene Kreisscheibe in der euklidischen Ebene E (oder ein topologisches Bild von ihr). Der Rand $G-\underline{G}$ von G sei mit G_g bezeichnet.

Weiter sei $\mathfrak{q}$ das System bestehend aus allen (einfachen, abgeschlossenen) in G

enthaltenen Bogen, die bis auf ihre Endpunkte in $\underline{G}$ liegen, sowie aus allen (in G enthaltenen, einfachen, geschlossenen) Kurven (Jordankurven) Q mit leerem oder einpunktigen $Q \cap G_g$. Jeder Bogen aus $\mathfrak{q}$ lässt sich zu einer Jordankurve (in G) ergänzen, indem man einen der (beiden) durch die Endpunkte des Bogens begrenzten Teilbogen von G_g hinzufügt. Demgemäss kann (und – soweit es bequem ist – soll) jedes $Q \in \mathfrak{q}$ als Jordankurve vorausgesetzt werden derart, dass $Q \cap G_g$ leer oder einpunktig oder ein Bogen ist. Wir *verabreden* dabei: Es soll $x \in Q$ immer bedeuten, dass $x \in Q - Q \cap G_g$.

Für jedes $Q \in \mathfrak{q}$ seien die beiden Komponenten von $G - Q$ als die Seiten $Q(\alpha)$ von Q bezeichnet, wobei $\alpha = +$ oder $\alpha = -$ ist.

Später (vgl. Nr. 3.1.ff.) wird statt $\mathfrak{q}$ ein Teilsystem $\mathfrak{k}$ von $\mathfrak{q}$ betrachtet, dessen Elemente K als Ordnungscharakteristiken (kurz OCh) bezeichnet werden, und die, unter anderem, die Eigenschaft besitzen, durch je drei ihrer Punkte eindeutig bestimmt zu sein (Vgl. Nr. 3.1.1.).

1.1.1. Ist $Q \in \mathfrak{q}$ *orientiert*, so auch jeder Teilbogen B von Q, der (ev. bis auf seine Endpunkte) in $\underline{G}$ liegt. Es sei $y_1, \ldots, y_t \in B$, $t \geq 3$; liegt dann $y_{\tau-1}$ auf B vor y_τ für jedes $\tau = 2, \ldots, t$, so sagt man, die $y_1, \ldots, y_t$ seien in dieser Reihenfolge entsprechend der Orientierung von Q bzw. B angeordnet oder die Reihenfolge im t-tupel $\mathfrak{t} = (y_1, \ldots, y_t)$ entspreche der Orientierung von B (Betr. die Definition von „vor" bzw. „hinter" vgl. H.-K. [1]. Nr. 1.1.).

Jede Teilfolge $y_{\tau_1}, \ldots, y_{\tau_r}$, $3 \leq r \leq t-1$, eines solchen $\mathfrak{t}$ mit $\tau_{\varrho-1} < \tau_\varrho$, $\varrho = 2, \ldots, r$, heisse (in dieser Reihenfolge) Q-konsekutiv (in $\mathfrak{t}$); sie heissen hingegen (in dieser Reihenfolge) direkt *Q-konsekutiv* (in $\mathfrak{t}$), wenn sogar $\tau_\varrho = \varrho$ ist für alle $\varrho = 1, \ldots, r$, wenn also „zwischen" den y_{τ_ϱ} keine, von ihnen verschiedene unter den y_τ aus $\mathfrak{t}$ liegen.

1.1.2. Es seien $C, Q \in \mathfrak{q}$. Dabei wird meist $C \subset \underline{G}$ sein, sodass C und $C \cap Q$ fremd zu G_g ist.

ANNAHME. Für die jeweils betrachteten $Q \in \mathfrak{q}$ soll stets $C \cap Q$ *endlich* und alle Punkte von $C \cap Q$ sollen *Schnittpunkte*[2]) von C mit Q sein.

1.1.3. Die Mächtigkeit einer Menge $M \subset G$ werde mit POW(M) und als Punktordnungswert von M bezeichnet. Ist POW$(M \cap Q)$ für jedes $Q \in \mathfrak{q}$ endlich bzw. maximal gleich m, so bezeichnet man M als von höchstens endlichem POW$(M; \mathfrak{q})$ bzw. man setzt POW$(M; \mathfrak{q}) = m$. Weil $C \cap Q$ für $C \subset \underline{G}$ endlich sein und nur Schnittpunkte enthalten soll (vgl. Nr. 1.1.2.), ist POW$(C \cap Q) \equiv 0$ (mod 2).

1.1.4. Es sei C *orientiert*. Sind $x_1, \ldots, x_n \in C \cap Q$, so heisse das n-tupel $\mathfrak{x} = (x_1, \ldots, x_n)$ enthalten in $C \cap Q$, in Zeichen $\mathfrak{x} \subset C \cap Q$. Ferner heisse $\mathfrak{x}$ normal bzw. fast-

[2]) Für Bezeichnungen, die im Folgenden nicht weiter erklärt sind, sei auf H.-K. [1] verwiesen.

normal in $C \cap Q$, wenn *bei passender Orientierung von Q* die $x_1, \ldots, x_n$ (in dieser Reihenfolge) sowohl direkt C- als direkt Q-konsekutiv bzw. zwar direkt C-konsekutiv aber (wenigstens) Q-konsekutiv sind in $C \cap Q$ d.h. in dem t-tupel $\mathfrak{t} = (y_1, \ldots, y_t)$ der auf C entsprechend der Orientierung von C geordneten Schnittpunkte y_τ von $C \cap Q$.

Zwei entsprechend der Orientierung von C angeordnete t'- bzw. t''-tupel $\mathfrak{t}' = (x'_1, \ldots, x'_{t'})$ bzw. $\mathfrak{t}'' = (x''_1, \ldots, x''_{t''})$ auf C heissen quasi-fremd, wenn die kleinsten, alle x'_τ bzw. alle x''_τ enthaltenden Teilbogen C' bzw. C'' von C fremd sind bis auf höchstens Endpunkte, also

$$C' \cap C'' \subset \{x'_1\} \cup \{x'_{t'}\} \cup \{x''_1\} \cup \{x''_{t''}\}\,; \qquad 3 \leqslant t', 3 \leqslant t''.$$

1.1.5. Es sei $R \in \mathfrak{q}$ und $a, b \in R$ mit $a \neq b$. Dann bezeichne $R(a|b)$ einen durch die beiden (End-)Punkte a, b begrenzten, abgeschlossenen Teilbogen B von R. Der grösste offene in $R(a|b)$ enthaltene Bogen, also $B - \{a\} - \{b\}$ sei mit $\underset{\sim}{R}(a|b) = \underset{\sim}{B}$ bezeichnet.

1.1.6. Es seien $C, Q \in \mathfrak{q}$ mit $C \subset \underline{G}$ und beide orientiert; ferner sei $\mathfrak{t} = (x_1, \ldots, x_n) \subset C \cap Q$ mit in dieser Reihenfolge direkt C-konsekutiven Schnittpunkten $x_1, \ldots, x_n$, $n = 2k \geqslant 4$. Es sei V bzw. H eine vordere bzw. hintere Umgebung von x_1 bzw. von x_n auf Q. Man sagt, es besitze $\mathfrak{t}$ die Signatur α mit $\alpha = \pm 1$, in Zeichen $\alpha = \mathrm{sign}(\mathfrak{t})$, wenn für hinreichend kleine V, H gilt $\underset{\sim}{V} \cup \underset{\sim}{H} \subset C(\alpha)$.

1.1.7. Ist $R \in \mathfrak{q}$ orientiert, so heisse $B = R(a|b)$ im Sinne von R orientiert, wenn B eine auf R hintere Umgebung von a und (daher) eine auf R vordere Umgebung von b enthält; a wird dann auch als *Anfangspunkt* des orientierten Bogens B bezeichnet. Ist $z_\nu \in \underset{\sim}{B}$, $\nu = 1, \ldots, n$; $n \geqslant 1$, und sind die $a, z_1, \ldots, z_n, b$ in dieser Reihenfolge R- bzw. B-konsekutiv, so deutet man dies auch an durch die Schreibweise $R(a|z_1|\ldots|z_n|b)$ statt $R(a|b)$, womit dann $R(a|b)$ eindeutig bestimmt ist.

1.2. VOR. (1) *Es seien $C, Q \in \mathfrak{q}$ mit $C \subset \underline{G}$ und es sei Q sowie C orientiert.* – (2) *Ist $s', s'' \in C \cap Q$ mit $s' \neq s''$, so sei $C' = C(s'|s'')$ bzw. $Q' = Q(s'|s'')$, also $C' \subset C$, $Q' \subset Q$.* – (3) *Es sei $H' \subset Q'$ bzw. $V'' \subset Q'$ eine hintere Umgebung von s' auf Q bzw. eine vordere Umgebung von s'' auf Q.*

BEH. (I) *Hinreichend kleine $\underset{\sim}{H}'$ und $\underset{\sim}{V}''$ liegen auf der gleichen bzw. auf verschiedenen Seiten von C, wenn und nur wenn* $\mathrm{POW}(C' \cap Q) \equiv 0$ *bzw.* $\equiv 1 \pmod 2$ *ist. Speziell ist also $\underset{\sim}{H}' \cup \underset{\sim}{V}'' \subset C(\alpha)$ bei passendem, α wenn s', s'' direkt C-konsekutiv in $C \cap Q$ sind.* (II) *Jedes fast-normale* 4-*tupel aus $C \cap Q$ besitzt eine bestimmte Signatur.* – (III) *Sind $x, x', x'' \in C \cap Q$ direkt C-konsekutiv in $C \cap Q$ und ist $Q'' = Q(x|x'') \subset Q$ mit $x' \notin Q''$, so gilt* $\mathrm{POW}(C \cap \underset{\sim}{Q}'') \equiv 1 \pmod 2$*; insbesondere ist also $C \cap \underset{\sim}{Q}'' \neq \emptyset$.*

Bew. Betr. Beh. (I). Die Beh. ergibt sich daraus, dass $C \cap Q$ endlich ist und nur

Schnittpunkte enthält und dass ferner C und auch (vgl. Nr. 1.1.) Q Jordankurven sind, also $\mathrm{POW}(C \cap Q) \equiv 0$ ist. – *Beh.* (II) folgt aus Beh. (1). – *Betr. Beh.* (III). Gemäss Beh. (I) liegen Umgebungen von x' und x'' auf $\underset{\sim}{Q}''$ auf verschiedenen Seiten von C.

1.3. VOR. *Es seien C, $Q \in \mathfrak{q}$ mit $C \subset \underline{G}$; und es sei C orientiert. Es seien s', $s'' \in C \cap Q$ mit $s' \neq s''$ direkt C-konsekutiv. Ausserdem sei einer der offenen, von s' und s'' begrenzten Teilbogen von Q fremd zu C; dieser Teilbogen von Q sei mit $\underset{\sim}{T} = \underset{\sim}{Q}(s' \mid s'')$ bezeichnet.*

BEH. (1) *Ist* $\mathrm{POW}(C \cap Q) \geqslant 4$, *so gehören s' und s'' zu einem fast-normalen* 4-*tupel* $\mathfrak{t}' = (q', s', s'', q'') \subset C \cap Q$*; dabei liegt also q' auf C vor s', ferner s' vor s'' und s'' vor q'' (und q'' vor q') $(q' \neq q'')$.* – (2) *Aus* $\underset{\sim}{T} \subset C(\alpha)$ *folgt* $\mathrm{sign}\,(\mathfrak{t}') = \alpha$.

Bew. Wegen $\mathrm{POW}(C \cap Q) \geqslant 4$ und weil C Kurve mit $C \subset \underline{G}$ ist, existieren q', $q'' \in C \cap Q$ derart, dass q', s', s'', q'' in dieser Reihenfolge direkt C-konsekutiv sind (in $C \cap Q$). Zu zeigen ist: Es sind q', s', s'', q'' in eben dieser Reihenfolge zugleich (mindestens) Q-konsekutiv, bei passend orientiertem Q. In der Tat: Wegen $C \cap \underset{\sim}{T} = \emptyset$ ist $\underset{\sim}{T} \subset C(\alpha)$ für ein gewisses α und es sind s', s'' direkt Q-konsekutiv bei einer jetzt festzuhaltenden, Orientierung von Q, bei der $\underset{\sim}{H}' \subset C(\alpha)$ gilt für eine hinreichend kleine hintere Umgebung H' von s' auf Q. Weil q', s' direkt C-konsekutiv sind, gilt $\underset{\sim}{V}' \subset C(\alpha)$ für eine auf Q vordere Umgebung V' von q' (Nr. 1.2.). Wir schliessen jetzt indirekt, nehmen also an, es seien q', s', s'', q'' nicht Q-konsekutiv. Weil s', s'' direkt Q-konsekutiv sind, sind dann q'', s', s'', q' Q-konsekutiv (in dieser Reihenfolge) auf $C \cap Q$. Es sei $T' = Q(s'' \mid q') \subset Q$ mit $q'' \notin T'$, ferner sei $C' = C(q' \mid s') \subset C$ mit $s'' \notin C'$. Es ist dann $J = C' \cup T \cup T'$ eine Jordankurve, durch die $\underline{G}$ in zwei Gebiete G', G'' zerlegt wird. Es gilt etwa $\underset{\sim}{C}(s' \mid s'') \subset G'$, wobei $\underset{\sim}{C}(s' \mid s'')$ fremd zu q', q'' ist. Wegen $\underset{\sim}{T} \subset C(\alpha)$ und $\underset{\sim}{V}' \subset C(\alpha)$ liegt auch eine Umgebung von q' auf $\underset{\sim}{T}'' = Q - T - T'$ in G'. Andererseits ist $\underset{\sim}{C}(s' \mid s'' \mid q'') \cap J = \{s''\}$, also $q'' \in G''$ und daher $\underset{\sim}{T}'' \cap J \neq \emptyset$. Widerspruch.

1.4. VOR. *Es seien C, $Q \in \mathfrak{q}$, mit $C \subset \underline{G}$ und orientiertem C. Ferner seien x', s', s'', $x'' \in C \cap Q$ direkt C-konsekutiv in $C \cap Q$ (in dieser Reihenfolge).*

BEH. *Folgende beiden Aussagen sind gleichwertig:* (1) *Ein von s' und s'' begrenzter Teilbogen T von Q ist fremd zu x' und x''.* – (2) *Das* 4-*tupel* $\mathfrak{t} = (x', s', s'', x'')$ *ist fast normal (in $C \cap Q$).*

ZUSATZ. Liegt $\underset{\sim}{T}$ in der Nähe von s' (und von s'') in $C(\alpha)$, so ist $\mathrm{sign}(\mathfrak{t}) = \alpha$.

Bew. Aus (1) *folgt* (2). Da x', s', s'', x'' direkt C-konsekutiv sind, ist nut zu zeigen, dass x', s', s'', x'' Q-konsekutiv sind, wenn Q so orientiert ist, dass eine hintere Umgebung von s' auf Q in T liegt; dabei ist $T = Q(s' \mid s'') \subset Q$ fremd zu x' und x''. Daher sind auf Q nur die Reihenfolgen (1) x', s', s'' oder (2) x'', s', s'' und (1′) s', s'', x'' oder (2′) s', s'', x' denkbar; dabei ist z.B. (1) dahin zu verstehen, dass der Bogen $Q(x' \mid s' \mid s'')$ fremd zu x'' ist und entsprechend für (2), (1′), (2′). Es genügt, (2) als unmöglich nachzuweisen, da man für (2′) entsprechend schliesst. Im Fall (2) gilt aber: Es sei $C''=C(s'|x'')\subset C$ und $Q''=Q(x''|s')\subset Q$ der von s' und x'' begrenzte, zu x' fremde Teilbogen von C bzw. von Q. Wegen $\mathrm{POW}(C''\cap Q)=3$ liegen (gemäss Nr. 1.2.) Umgebungen S' von s' und X'' von x'' auf $\underset{\sim}{Q}''$ auf verschiedenen Seiten von C, etwa $S'\subset C(-\alpha)$ und $X''\subset C(\alpha)$, Analog wie beim Beweise in Nr. 1.3. schliesst man (unter Bezugnahme auf die Jordankurve $J=C''\cup Q''$) auf einen Widerspruch. Es ist also nur (1) und (1′) möglich. – *Aus* (2) *folgt* (1) gemäss der Definition des fast normalen 4-tupels.

1.5. VOR. (1) *Es seien* $C, Q\in\mathfrak{q}$ *mit* $C\subset\underline{G}$ *und mit* $\mathrm{POW}(C\cap Q)\geqslant 6$. – (2) *Es seien* $x_1', x_1, x_2, x_3, x_3''\subset C\cap Q$ *in dieser Reihenfolge C-direkt konsekutiv und* x_1, x_2, x_3 *überdies direkt Q-konsekutiv in* $C\cap Q$.

BEH. (I) *Die 4-tupel* $\mathfrak{t}'=(x_1', x_1, x_2, x_3)\subset C\cap Q$ *und* $\mathfrak{t}''=(x_1, x_2, x_3, x_3'')\subset C\cap Q$ *sind beide fast-normal. – Es ist* $\mathrm{sign}(\mathfrak{t}')\neq\mathrm{sign}(\mathfrak{t}'')$.

Bew. Die Beh. (I) folgt aus Nr. 1.3., weil gemäss Vor. (2) $C\cap\underset{\sim}{Q}(x_1|x_2)=C\cap\underset{\sim}{Q}(x_2|x_3)=\emptyset$, wenn $x_3\notin Q(x_1|x_2)$ und $x_1\notin Q(x_2|x_3)$. Da ferner Umgebungen von x_1 auf $\underset{\sim}{Q}(x_1|x_2)$ und von x_2 auf $\underset{\sim}{Q}(x_2|x_3)$ auf verschiedenen Seiten von C liegen, folgt auch die Beh. (II).

1.5.1. VOR. (1) *Es seien alle Punkte von* $C\cap K$ *Schnittpunkte; ferner seien* 1, 2, $3, 4\in C\cap K$ *direkt C-konsekutiv, wobei zur Abkürzung i statt* x_i *geschrieben ist,* $i=1,\ldots,4$. – (2) *Es sei* $K(1|2)\cap K(3|4)=\emptyset$ *und* $\underset{\sim}{K}(1|2)\cup\underset{\sim}{K}(3|4)\subset C(\alpha)$. – (3) *Es sei* $\mathrm{POW}(C\cap K)\geqslant 6$.

BEH. (I) *Es ist* (1, 2, 3, 4) *ein fast normales* 4-*tupel der Signatur* $-\alpha$. – (II) *Es gibt* $x, y\in C\cap K$ *derart, dass* x, 1, 2, 3, 4, y *in dieser Reihenfolge direkt C-konsekutiv und dass* $(x, 1, 2, 3)$ *sowie* $(2, 3, 4, y)$ *fast normale* 4-*tupel der Signatur* α *sind.*

Bew. Betr. Beh. (I). Indirekt. Andernfalls sind 1, 2, 4, 3 K-konsekutiv (gemäss Vor. (2)). Es sei $T'=K(2|4)$ bzw. $T''=K(3|1)$ der zu 3 bzw. zu 4 fremde Teilbogen von K. Dann ist $C(2|3|4)\cup T'$ eine Jordankurve J derart, dass eine Umgebung von 3 auf $\underset{\sim}{T}''$ und 1 auf verschiedenen Seiten von J liegen, sodass $T'\cap T''\neq\emptyset$ wäre, während doch K einfache Kurve ist. – *Betr. Beh.* (II). Gemäss Vor. (3) sowie (1), ferner weil $C\subset\underline{G}$ Kurve ist, existieren Schnittpunkte x, y derart, dass die x, 1, 2, 3, 4, y direkt C-konsekutiv sind. Gemäss Vor. (2) folgt der Rest der Beh. (II) aus Nr. 1.4.

§ 2. Existenz von Paaren fast normaler quasi-fremder 4-tupel verschiedener Signatur

2.0. Satz. VOR. *Es sei C eine orientierte Kurve mit* $C\subset\underline{G}$. *Ferner sei* $\mathrm{POW}(C;\mathfrak{q})$ *höchstens endlich, aber* $\mathrm{POW}(C;\mathfrak{q})\geqslant 6$.

BEH. *Für jedes* $Q \in \mathfrak{q}$ *mit* $\mathrm{POW}(C \cap Q) \geqslant 6$ *enthält* $C \cap Q$ *(mindestens) vier fast normale* 4-*tupel* $\mathfrak{p}', \mathfrak{p}'', \mathfrak{n}', \mathfrak{n}''$ *von folgender Art: Es sind* $\mathfrak{p}'$ *und* $\mathfrak{p}''$ *quasi-fremd, ebenso* $\mathfrak{n}'$ *und* $\mathfrak{n}''$. *Ausserdem ist* $\mathrm{sign}(\mathfrak{p}') = \mathrm{sign}(\mathfrak{p}'')$ *und* $\mathrm{sign}(\mathfrak{n}') = \mathrm{sign}(\mathfrak{n}'')$, *aber* $\mathrm{sign}(\mathfrak{p}') = -\mathrm{sign}(\mathfrak{n}')$. (Wie immer sollen alle Punkte von $C \cap Q$ Schnittpunkte sein [Vgl. Nr. 1.1.2.]). *Eine Ausnahme machen lediglich* (vgl. Nr. 2.2.) *gewisse Q im Falle* $\mathrm{POW}(C \cap Q) = 6$ (welche erst später (Nr. 3.4.5.) einer einschlägigen näheren Untersuchung unterzogen werden).

Der *Beweis* dieses Satzes wird in den Nr. 2.1.–2.3. geführt. Der Gedankengang ist dabei folgender: Zuerst wird der Fall $\mathrm{POW}(C \cap Q) = 6$ behandelt und anschliessend der Fall $\mathrm{POW}(C \cap Q) = 8$, von welchem aus vollständige Induktion von $\mathrm{POW}(C \cap Q) = 2n \geqslant 8$ auf $\mathrm{POW}(C \cap Q) = 2(n+1)$ möglich ist. Induktion von $2n = 6$ aus ist wegen der im Satz erwähnten Ausnahmen nicht durchführbar.

2.1. VERABREDUNG. Es sei $C \cap Q = \bigcup_{\mu=1}^{m} \{x_\mu\}$; dabei ist $m = 2n$, da alle x_μ Schnittpunkte sein sollen. Wir *schreiben* μ statt x_μ, $\mu = 1, \ldots, m$, und ν statt $m + \nu$ für $\nu = 1, 2, \ldots$. Die $1, 2, \ldots, m$ sollen *in dieser Reihenfolge direkt C-konsekutiv* sein in $C \cap Q$; es soll also μ *vor* $\mu + 1$ auf C liegen. Dagegen seien die μ in der Reihenfolge $1', 2', \ldots, m'$ *direkt Q-konsekutiv* in $C \cap Q$ bei einer Orientierung von Q. Jede derart in irgend einem $C \cap Q$ bei einer Orientierung von Q mögliche (direkt Q-konsekutive) Reihenfolge $1', 2', \ldots, m'$ werde als eine zulässige und abgekürzt mit $((\mu'))$ bezeichnet. O.B, d.A. werden wir uns auf den Fall $1 = 1'$ beschränken.

2.1.1. VORBEMERKUNG. *Sind* μ', $\nu' \in ((\mu'))$ *bei zulässigem* $((\mu'))$ *und ist* $\nu' = \mu' + 2k$ $k \geqslant 1$, *so sind* μ' *und* ν' *in* $((\mu'))$ *durch eine ungerade Anzahl von Elementen aus* $((\mu'))$ *getrennt. Speziell können also* μ', ν' *oder* ν', μ' *nicht direkt Q-konsekutiv in* $C \cap Q$ *sein.*

Bew. Liegt z.B. ν' auf Q hinter μ' und ist $C' = C(\mu' \mid \mu' + 1 \mid \ldots \mid \nu') \subset C$, so liegen wegen $\nu' \equiv \mu'$ (mod 2) eine vordere bzw. eine hintere Umgebung von μ' bzw. von ν' auf $\underset{\sim}{C}'$ auf verschiedenen Seiten von Q. Die Beh. folgt jetzt aus Nr. 1.2.

2.1.2. VOR. *Es seien* $\mu, \zeta, \nu, \omega \in C \cap Q$ *in dieser Reihenfolge C-konsekutiv, also* $\zeta = \mu + m'$, $\nu = \mu + r$, $\omega = \mu + r + s$ *mit* $1 \leqslant m' \leqslant r - 1$, $2 \leqslant r$, $1 \leqslant s$; *dabei sei* $C \subset \underline{G}$.

BEH. *Weder* μ, ν, ω, ζ *noch* μ, ν, ζ, ω *sind, je in dieser Reihenfolge,* direkt *Q-konsekutiv (in* $C \cap Q$*).*

Bew. Indirekt. (I) Es seien μ, ν, ω, ζ direkt Q-konsekutiv. Wir betrachten $C' = C - C(\mu \mid \zeta \mid \nu \mid \omega)$ und $Q' = Q(\mu \mid \nu \mid \omega) \subset Q$. Dann ist $J = C' \cup Q'$ eine Jordankurve, auf deren einer Seite ζ liegt, auf deren anderer Seite aber eine Umgebung von ω auf $Q'' = Q - Q'$. Daher ist $C' \cap Q'' \neq \emptyset$; also sind ω, ζ nicht direkt Q-konsekutiv. – (II) Es seien μ, ν, ζ, ω direkt Q-konsekutiv. Es sei $C' = C(\mu \mid \zeta) \subset C$ fremd zu ν und $Q' = Q(\mu \mid \nu \mid \zeta) \subset Q$. Dann ist $J = C' \cup Q'$ eine Jordankurve, auf deren einer Seite

ω liegt, auf deren anderer Seite aber eine Umgebung von ζ auf $Q''=Q-Q'$ liegt, sodass $\underset{\sim}{C}'\cap\underset{\sim}{Q}''\neq\emptyset$. Widerspruch.

2.2. Der Fall POW$(C\cap Q)=6$.

Es sei also $C\cap Q=\{1,2,3,4,5,6\}$, wobei die 1, 2, 3, 4, 5, 6 in dieser Reihenfolge direkt C-konsekutiv seien (sowie sämtlich Schnittpunkte in $C\cap Q$). Gemäss Nr. 2.1.1. kann in den zulässigen $((\mu'))$ (höchstens) $2'=2$ oder $2'=4$ oder $2'=6$ sein; da $2'=6$ durch Umorientierung von C auf den Fall $2'=2$ zurückführbar ist, kann (und soll) $2'=6$ ausser Betracht bleiben. Ebenfalls gemäss Nr. 2.1.1. können ferner in den zulässigen $((\mu'))$ neben 12 (höchstens) (I) die Paare 34 und 56 sowie (II) die Paare 36 und 45; weiter neben 14 (höchstens) (III) die Paare 23 und 56 sowie (IV) die Paare 25 und 36 auftreten; dabei sind ausser 34, 56 usw. auch jeweils die „Transponierten“ 43, 65 usw. in Betracht zu ziehen. Gemäss Nr. 2.1.1. bzw. 2.1.2. ist Fall (IV) ausgeschlossen, weil 2536, 3625 bzw. 2563, 6325 nicht möglich sind. Überprüft man entsprechend die Fälle (I)–(III) und beachtet, dass mit einem $((\mu'))$ auch alle topologische äquivalenten der Beh. des Satzes in Nr. 2.0. genügen, so folgt:

Es ist die Beh. des Satzes in Nr. 2.0. *nur nachzuprüfen für die Fälle* (I') $((\mu'))=1,2,3,4,5,6)$ und (II') $((\mu'))=(1,2,3,6,5,4)$.

Der Beweis dieser Behauptung lässt sich, ebenso wie für den Fall POW$(C;\,\mathfrak{k})\geqslant 8$ (Nr. 2.3.) nach Herrn KÜNNETH anschaulich auch so führen: Man betrachte der einfacheren Formulierung halber statt C und seinem Innern eine Kreisperipherie $\mathfrak{P}$ und die von $\mathfrak{P}$ begrenzte offene Kreisscheibe $\mathfrak{S}$. Sind die Punkte 1, 2, 3, 4, 5, 6 zyklisch auf $\mathfrak{P}$ angeordnet, so sind die in I.–III. oben angegebenen drei Paare jeweils dadurch gekennzeichnet, dass die drei Verbindungssehnen je der beiden Punkte eines Paares fremd sind (diese Sehnen entsprechen je einem durch die beiden Punkte begrenzten, zu C fremden offenen Teilbogen von Q). Die den Fällen I.–III. entsprechenden „Sehnenfiguren“ seien als Figuren I.–III. bezeichnet; sie sind offensichtlich die einzig möglichen. Man sieht, dass die Figuren II. und III. topologisch äquivalent sind, sodass man sich auf I. und II. beschränken kann. Diese sind nun noch zu *ergänzen* durch paarweise fremde, einfache ausserhalb $\mathfrak{S}$ verlaufende Bogen, deren Endpunkte *verschiedenen* Paaren angehören; jeder der Punkte 1,…, 6 ist Endpunkt genau eines dieser Bogen. (Dabei scheiden jetzt die durch Nr. 2.1.1. und 2.1.2. gekennzeichneten Fälle aus). Die übrigbleibenden Fälle erweisen sich unmittelbar als in zwei Klassen topologisch äquivalenter zerfallend; w.z.z.w.

Für die dem *Fall* II'. entsprechende Figur lässt sich jetzt die Beh. des Satzes in Nr. 2.0. unmittelbar verifizieren. Und zwar ergibt sich bei passender Wahl von $C(+)$ und $C(-)$:

(II') $\mathfrak{p}'=(1,2,3,4)$, $\mathfrak{p}''=(4,5,6,1)$; $\mathfrak{n}'=(6,1,2,3)$, $\mathfrak{n}''=(3,4,5,6)$.

Betr. *Fall* I'. vgl. Nr. 3.4.5.

2.3. Der Fall $\mathrm{POW}(C \cap Q) = 8$.

Entsprechend wie in Nr. 2.2. stellt man mit Hilfe der Nr. 2.1.1. und 2.1.2. fest, dass – von topologisch äquivalenten abgesehen – nur Sehnenfiguren in Betracht kommen, die den folgenden Tripeln von Paaren (1, 2) usw. entsprechen: I. 12, 34, 56, 78. – II. 12, 34, 58, 67. – III. 14, 23, 58, 67. Und als zugehörige $((\mu'))$ ergeben sich: Zu I. (1, 2, 3, 4, 5, 6, 7, 8). – Zu II. (a) (1, 2, 3, 4, 5, 8, 7 6) sowie (1, 2, 3, 4, 7, 6, 5, 8); und (b) (1, 2, 7, 6, 3, 4, 5, 8). – Zu III. (a) (1, 2, 3, 8, 7, 4, 5, 6) sowie (1, 2, 5, 6, 7, 4, 3, 8); und (b) (1, 2, 3, 8, 5, 6, 7, 4) und (1, 2, 7, 4, 5, 6, 3, 8).

Dabei erweisen sich als topologisch äquivalent noch je die beiden 8-tupel in II.(a), in III.(a) und in III.(b).

Zu dem 8-tupel in I. und in II.(b), ferner je zu dem in II.(a), in III.(a) und in III.(b) an erster Stelle stehenden 8-tupel $((\mu'))$ gehören dann als $\mathfrak{p}', \ldots, \mathfrak{n}''$:

I. : $\mathfrak{p}' = (1, 2, 3, 4)$, $\mathfrak{p}'' = (5, 6, 7, 8)$; $\mathfrak{n}' = (2, 3, 4, 5)$, $\mathfrak{n}'' = (6, 7, 8, 1)$.
II.(a): $\mathfrak{p}' = (1, 2, 3, 4)$, $\mathfrak{p}'' = (6, 7, 8, 1)$; $\mathfrak{n}' = (2, 3, 4, 5)$, $\mathfrak{n}'' = (5, 6, 7, 8)$.
II.(b): $\mathfrak{p}' = (3, 4, 5, 6)$, $\mathfrak{p}'' = (7, 8, 1, 2)$; $\mathfrak{n}' = (2, 3, 4, 5)$, $\mathfrak{n}'' = (5, 6, 7, 8)$.
III.(a): $\mathfrak{p}' = (1, 2, 3, 4)$, $\mathfrak{p}'' = (6, 7, 8, 1)$; $\mathfrak{n}' = (8, 1, 2, 3)$, $\mathfrak{n}'' = (4, 5, 6, 7)$.
III.(b): $\mathfrak{p}' = (1, 2, 3, 4)$, $\mathfrak{p}'' = (5, 6, 7, 8)$; $\mathfrak{n}' = (4, 5, 6, 7)$, $\mathfrak{n}'' = (8, 1, 2, 3)$.

Damit ist für den Fall $\mathrm{POW}(C \cap Q) = 8$ die Beh. in Nr. 2.0. bewiesen.

2.4. Der Fall $\mathrm{POW}(C \cap Q) > 8$.

Die Beh. des Satzes in Nr. 2.0. *ist richtig allgemein für jedes* $m = 2n$, $n \geq 4$.

Bew. durch Induktion von $n = 4$ aus. Die Beh. sei für ein $n \geq 4$ schon bewiesen. Es sei $m = 2(n+1)$ und $\mathrm{POW}(C \cap Q) = m$, wobei $C \cap Q = \{1, 2, \ldots, m\}$ mit (in dieser Reihenfolge) C-direkt konsekutiven $1, 2, \ldots, m$. Wir unterscheiden zwei Fälle:

FALL (A). Die $1, 2, \ldots, m$ sind in dieser Reihenfolge auch direkt Q-konsekutiv bei geeignet orientiertem Q.

FALL (B). Es liegt Fall (A) nicht vor.

2.4.1. BETR. FALL (A). Hier ist die Existenz von $\mathfrak{p}'$, $\mathfrak{p}''$, $\mathfrak{n}'$, $\mathfrak{n}''$ für $n = 4$ in Nr. 2.3., Fall (I), bewiesen. Für $n \geq 5$ kann man unmittelbar so schliessen: Jedes 4-tupel $\mathfrak{t}_\nu = (\nu, \nu+1, \nu+2, \nu+3)$ ist fast normal für alle ν mit $1 \leq \nu \leq 2n-3$, wobei $2n-3 \geq 7$, ferner $\mathfrak{t}_1$ und $\mathfrak{t}_5$ bzw. $\mathfrak{t}_2$ und $\mathfrak{t}_6$ quasi-fremd und bzw. von gleicher Signatur, also $\mathrm{sign}(\mathfrak{t}_1) = \mathrm{sign}(\mathfrak{t}_5)$ und $\mathrm{sign}(\mathfrak{t}_2) = \mathrm{sign}(\mathfrak{t}_6)$, aber $\mathrm{sign}(\mathfrak{t}_1) = -\mathrm{sign}(\mathfrak{t}_2)$. – Damit ist die Beh. in Nr. 2.0. für den Fall (A) und für alle $m = 2n \geq 8$ bewiesen.

2.4.2. BETR. FALL (B). Hier existieren bekanntlich (vgl. H.–K. [1], Nr. 4.5.2., Satz 2 und Bew). $x' = r$, $x'' = r + s$ mit $s = 1 + 2q$, $1 \leq q$, von folgender Art: Es sind x'', x' in dieser Reihenfolge direkt Q-konsekutiv in $C \cap Q$, d.h. es ist der zu $r+1$ fremde Teilbogen $\underset{\sim}{X} = \underset{\sim}{Q}(x'' \mid x')$ von $\underset{\sim}{Q}$ fremd zu C, sodass $\underset{\sim}{X} \subset C(\alpha)$ bei passendem α.

Ferner sind die zu r fremden Teilbogen $\underset{\sim}{X}_\mu=\underset{\sim}{Q}(r+\mu|r+\mu+1)$ von Q fremd zu C und es gilt $\underset{\sim}{X}_\mu\subset C(\alpha)$, $\mu=1, 3,\ldots, 1+2(q-1)$.

Zur Vereinfachung unterscheiden wir noch:

FALL (B′): Es ist $q=1$. – FALL (B″) Es ist $q\geqslant 2$.

2.4.2.1. BETR. FALL (B′).

2.4.2.1.I. Setzen wir $y'=r+1, y''=r+2$ und $Y=X_1=Q(y'\,|\,y'')$, so sind x', y', y'', x'' in dieser Reihenfolge direkt C-konsekutiv und Q-konsekutiv, ferner ist $\underset{\sim}{X}\cup\underset{\sim}{Y}\subset C(\alpha)$. Es ist also $\mathfrak{t}=(x', y', y'', x'')$ ein fast normales 4-tupel der Signatur α. Wir ersetzen nun Q durch eine Kurve $Q'\in\mathfrak{q}$ mit $\mathrm{POW}(C\cap Q')=m-2=2n$ und derart, dass

$$C\cap Q=(C\cap Q')\cup\{y'\}\cup\{y''\}$$

mit zu Q' fremden y', y'' und dass Q' mit Q bis auf eine beliebig kleine Umgebung U von $Y=\bar{Y}$ auf Q übereinstimmt (Eine Konstruktion für ein solches Q' wird weiter unten angegeben; vgl. Nr. 2.4.2.1.III.).

Wegen $\mathrm{POW}(C\cap Q')=2n\geqslant 8$ existieren nach Induktionsannahme fast normale 4-tupel $\mathfrak{p}'$, $\mathfrak{p}''$, $\mathfrak{n}'$, $\mathfrak{n}''$ in $C\cap Q'$ (vgl. Nr. 2.0., Satz, Beh.).

Um im Folgenden die Abhängigkeit dieser $\mathfrak{p}',\ldots,\mathfrak{n}''$ von Q bzw. von Q' und von α deutlich hervortreten zu lassen setzen wir

$$\mathfrak{p}'=\mathfrak{t}'(+1;Q'),\quad \mathfrak{p}''=\mathfrak{t}''(+1;Q')\quad\text{und}\quad \mathfrak{n}'=\mathfrak{t}'(-1;Q'),\quad \mathfrak{n}''=\mathfrak{t}''(-1;Q')$$

und entsprechend $\mathfrak{t}'(+1;Q)$ statt $\mathfrak{t}'(+1;Q')$ usw. für Q statt Q'.

2.4.2.1.II. Wir haben jetzt zu zeigen: Beim Übergang von Q' zu Q bleibt das einzelne $\mathfrak{t}'(\beta;Q')$, $\mathfrak{t}''(\beta;Q')$; $\beta=\pm 1$, etwa $\mathfrak{t}'(\beta;Q')$ entweder *Erstens* erhalten oder es wird *Zweitens* $\mathfrak{t}'(\beta;Q')$ ersetzt durch ein $\mathfrak{t}'(\beta;Q)$, welches quasi-fremd ist zu einem in $C\cap Q$ enthaltenen $\mathfrak{t}''(\beta;Q)$.

Bew. betr. Erstens. Bleibt $\mathfrak{t}'(\beta;Q')$ beim Übergang zu Q erhalten, so ist $\mathfrak{t}'(\beta;Q')$ selbst ein $\mathfrak{t}'(\beta;Q)$. Dieses $\mathfrak{t}'(\beta;Q)$ ist nun quasifremd zu einem $\mathfrak{t}''(\beta;Q)$, falls ein $\mathfrak{t}''(\beta;Q')$ existiert, das ebenfalls beim Übergang von Q' zu Q erhalten bleibt. Sonst ist aber $\mathfrak{t}'(\beta;Q)$ quasi-fremd zu einem nicht in $C\cap Q'$ enthaltenen $\mathfrak{t}''(\beta;Q)$, wie sich aus dem Beweis für Zweitens (vgl., unten) ergibt.

Bew. Betr. Zweitens. Ein $\mathfrak{t}'(\beta;Q')$, welches nicht beim Übergang von Q' zu Q erhalten bleibt, ist jedenfalls ein solches, dem gleichzeitig x' *und* x'' angehören. Gehört nämlich höchstens eines der x', x'' zu $\mathfrak{t}'(\beta;Q')$, so ist dieses $\mathfrak{t}'(\beta;Q')$ quasifremd zu $\mathfrak{t}'(\alpha;Q)=(x', y', y'', x'')$ und es ist daher $\mathfrak{t}'(\beta;Q')$ ein $\mathfrak{t}'(\beta;Q)$, bleibt also erhalten, im Widerspruch zur Annahme. Daher hat man, weil x', x'' direkt C-konsekutiv *in* $C\cap Q'$ sind, entweder *Fall* (a) $\mathfrak{t}'(\beta;Q')=\mathfrak{t}'=(u', x', x'', u'')$ oder, wenn Fall (a) nicht vorliegt, *Fall* (b) $\mathfrak{t}'=(v', v'', x', x'')$ oder *Fall* (c) $\mathfrak{t}'=(x', x'', v', v'')$.

Betr. Zweitens, Fall (a). Es ist $\mathfrak{t}=(x', y', y'', x'')$ fast normal in $C\cap Q$ mit $\mathrm{sign}(\mathfrak{t})=$ $=\alpha$, also $\mathfrak{t}$ ein $\mathfrak{t}''(\alpha;Q)$. Existiert ein in $C\cap Q'$ fast normales $\mathfrak{t}'=\mathfrak{t}'(\beta;Q')=(u', x',$

x'', u''), so ist auch $\operatorname{sign}(\mathfrak{t}') = \operatorname{sign}(\mathfrak{t})$, weil $\underset{\sim}{X} \cup \underset{\sim}{Y} \subset C(\alpha)$. Ferner wird $\mathfrak{t}$ von $\mathfrak{t}'$ umfasst, d.h. es ist $C(x'|y'|y''|x'') \subset C(u'|x'|x''|u'')$. Nach Induktionsannahme ist aber $\mathfrak{t}'$ quasi-fremd zu einem $\mathfrak{t}''(\alpha; Q')$, welches, gemäss der Konstruktion von Q', zugleich ein $\mathfrak{t}''(\alpha; Q)$ sein muss. Es enthält also $C \cap Q$ jedenfalls quasi-fremde $\mathfrak{t}'(\alpha; Q)$ und $\mathfrak{t}''(\alpha; Q)$. Daneben könnte in $C \cap Q'$ ein $\mathfrak{t}'(-\alpha; Q')$ enthalten sein, welches beim Übergang von Q' zu Q nicht erhalten bleibt; ein solches müsste aber dann zum Fall (b) oder (c) gehören, welchen Fällen wir uns jetzt zuwenden.

Betr. Zweitens. Fall (b) *und* (c). Existiert, gemäss Fall (b), in $C \cap Q'$ ein fast normales $\mathfrak{t}' = \mathfrak{t}'(\beta; Q') = (v', v'', x', x'')$, so sind, bei entsprechender Orientierung von C und Q, die x'', x', v'', v' in $C \cap Q'$ direkt C-konsekutiv sowie Q'-konsekutiv (in dieser Reihenfolge). Wegen $\underset{\sim}{X} \subset C(\alpha)$ gilt $\operatorname{sign}(\mathfrak{t}') = \beta = -\alpha$. Zufolge der Konstruktion von Q' und weil $\mathfrak{t}$ fast normal in $C \cap Q$ ist, sind Q-konsekutiv die y', y'', x'', x', v'', v' in dieser Reihenfolge, also die y', x', v'', v' sowohl Q-konsekutiv als direkt C-konsekutiv (in $C \cap Q$). Daher ist $\mathfrak{t}'' = (y', x', v'', v') \subset C \cap Q$ fast normal und es ist $\operatorname{sign}(\mathfrak{t}'') = -\operatorname{sign}(\mathfrak{t}) = -\alpha$; denn der zu v' und y' fremde Teilbogen $\underset{\sim}{Q}(v''|x') \subset Q$ liegt in der Nähe von v'' und von x' in $C(-\alpha)$, weil $\underset{\sim}{X} \subset C(\alpha)$. Andererseits ist sign $(\mathfrak{t}') = \operatorname{sign}(\mathfrak{t}'')$ mit $\mathfrak{t}' = \mathfrak{t}'(-\alpha; Q')$. Beim Übergang von Q' zu Q kann also $\mathfrak{t}'$ durch $\mathfrak{t}''$ ersetzt werden, wenn noch gezeigt ist: $\mathfrak{t}''$ ist quasi-fremd zu einem $\mathfrak{t}^* = \mathfrak{t}(-\alpha; Q)$. Zufolge Induktionsannahme existiert aber in $C \cap Q'$ ein zu $\mathfrak{t}'$ quasi-fremdes fast normales $\mathfrak{t}^*$ derart, dass $\mathfrak{t}'$ und $\mathfrak{t}^*$ höchstens v' oder (und) x'' gemeinsam haben und gleiche Signatur besitzen; wegen $C(y'|x'|v''|v') \subset C(x''|x'|v''|v')$ ist also auch $\mathfrak{t}''$ quasi-fremd zu $\mathfrak{t}^*$ und es bleibt ausserdem $\mathfrak{t}^*$ beim Übergang von Q' zu Q erhalten. Somit enthält auch $C \cap Q$ (mindestens) zwei quasi-fremde, fast normale 4-tupel der Signatur $-\alpha$. Entsprechend schliesst man im Fall (c).

2.4.2.1.III. KONSTRUKTION VON Q' (vgl. Nr. 2.4.2.1., Ziff. I.). Es sei Q so orientiert, dass x'' auf X vor x', dass also eine hintere Umgebung von x'' auf Q in X liegt. Dann liegt y' auf Y vor y'', weil $\mathfrak{t} = (x', y', y'', x'')$ fast normal in $C \cap Q$ ist. Es gibt nun (beliebig kleine) vordere bzw. hintere Umgebungen V' von y' bzw. H'' von y'' auf Q derart, dass $\underset{\sim}{V}' \cup \underset{\sim}{H}'' \subset C(-\alpha)$ ist. Weiter gibt es (beliebig kleine) Umgebungen W in E von $\underset{\sim}{C}(y'|y'') = \underset{\sim}{C}' \subset C - C \cap Q$ derart, dass (bei hinreichend kleinem V', H'') auch $V' \cup H'' \subset W \cap Q \subset U$ sowie dass $C \cap Q \cap W = \{y'\} \cup \{y''\}$ ist. Ist nun $V' = Q(\bar{y}'|y')$, $H'' = Q(y''|\bar{y}'')$, so gibt es einfache Bogen $F \subset W \cap C(-\alpha)$ mit Endpunkten $\bar{y}'$, $\bar{y}''$ und mit $\underset{\sim}{F} \cap Q = \emptyset = F \cap C$. Und dann ist $(Q - Y - V' - H'') \cup F$ eine Kurve $Q' \in \mathfrak{q}$ der geforderten Beschaffenheit.

2.4.2.2. BETR. FALL (B″). Es existieren also $x', y', y'', z'\, z'', \ldots, x'' \in C \cap Q$, die in dieser Reihenfolge direkt C-konsekutiv sowie Q-konsekutiv sind (vgl. Nr. 2.4.2.) Dabei ist $\underset{\sim}{X} \subset C(\alpha) \cap Q$ und ebenso $\underset{\sim}{Y} = \underset{\sim}{Q}(y'|y'') \subset C(\alpha) \cap Q$ und $\underset{\sim}{Z} = \underset{\sim}{Q}(z'|z'') \subset C(\alpha) \cap$

$\cap Q$ mit $\underset{\sim}{X}\cap C=\underset{\sim}{Y}\cap C=\underset{\sim}{Z}\cap C=\emptyset$. Mithin ist $(x', y', y'', z')=\mathfrak{t}$ fast normal in $C\cap Q$ mit $\operatorname{sign}(\mathfrak{t})=\alpha$. Konstruiert man also entsprechend wie in Nr. 2.4.2.1. III. ein Q', indem man $Y\cup V'\cup H''$ aus Q wegschneidet, so tritt hier z' an Stelle des x'' in Nr. 2.4.2.1. III. und man schliesst dann wie oben.

§ 3. Ein Vierscheitelsatz für ebene Jordankurven

Als Hilfmittel bei dem folgenden Beweise eines Vierscheitelsatzes dienen die in §§ 1 und 2 zusammengestellten Bemerkungen.

3.1. Voraussetzungen. Definitionen. Wir betrachten im Grundbereich G ein Teilsystem $\mathfrak{k}$ von $\mathfrak{q}$ (vgl. Nr. 1.1.) mit folgenden Eigenschaften:

3.1.1. Jedes $K\in\mathfrak{k}$ ist durch beliebige 3 seiner Punkte $x_i\in K-K\cap G_g$, $i=1, 2, 3$, eindeutig bestimmt und stetige Funktion dieser x_i (vgl. H.–K. [1], Nr. 1.1.1. und Nr. 1.1.3.); ist $x_i\in K$, $i=1, 2, 3$, so schreiben wir für K auch *ausführlicher* $K(x_1, x_2, x_3)$. Dabei bedeutet $x\in K$ immer soviel wie $x\in K-K\cap G_g$ (vgl. Nr. 1.1.). Die Stetigkeit von $K=K(x_1, x_2, x_3)$ in den x_i besagt unter anderem: Ist $K=K(x_1, x_2, x_3)\in\mathfrak{k}$, also mit $x_\iota\neq x_j$ für $i\neq j$, ist ferner $x_\iota=\lim_n x_{\iota n}$, $i=1, 2, 3$, so existiert $K_n=K(x_{in}, x_{\iota n}, x_{in})$ $\in\mathfrak{k}$ für schliesslich alle n und es gilt $K=\lim_n K_n$.

3.1.1.1. Definition. Als Punktordnungswert $\mathrm{POW}(x; C; \mathfrak{k})$ *eines* Punktes $x\in C$ wird erklärt das Minimum (soweit es existiert) der Punktordnungswerte aller Umgebungen U von x auf C (Existiert dieses Minimum nicht, ist also $\mathrm{POW}(U; \mathfrak{k})$ für alle hinreichend kleinen U endlich, aber nicht beschränkt oder unendlich, so spricht man von endlichem bzw. unendlichem $\mathrm{POW}(x; C; \mathfrak{k})$).

3.1.2. Es sei $C\subset \underline{G}$ eine Jordankurve, von welcher gefordert wird: (I) (1) es ist $\mathrm{POW}(C; \mathfrak{k})$ höchstens endlich. – (2) Für jedes $x\in C$ mit $\mathrm{POW}(x; C; \mathfrak{k})<4$ ist $\mathrm{POW}(x; C; \mathfrak{k})=3$. – (II) (1) Es seien x_ι, $x_{in}\in C$ mit $x_i\neq x_j$ für $i\neq j$; $i, j=1, 2, 3; n=1, 2, \ldots$. Weiter sei $x_\iota=\lim_n x_{in}$ und es soll $K_n=K(x_{1n}, x_{2n}, x_{3n})\in\mathfrak{k}$ existieren, also mit $x_{in}\in K_n$; $n=1, 2, \ldots$. Unter diesen Annahmen soll $L=lim_n K_n$ existieren und es soll L eine OCh sein, also $L=K(x_1, x_2, x_3)\in\mathfrak{k}$. – (2) Weiter soll gelten: Für jeden Punkt $x\in C$ ist jede ($\mathfrak{k}$-)*Paratingente* $P(x)$ eine *mehrpunktige Ordnungscharakteristik* (vgl. H.–K. [1], Nr. 4.2.3.); dabei ist eine $\mathfrak{k}$-Paratingente $P(x)$ an C im Punkt x erklärt als $P(x)=\lim_n K_n$ mit $K_n\in\mathfrak{k}$, wobei für die OCh gilt: Es gibt eine gegen x konvergierende Folge von Umgebungen U_n von x auf C, $n=1, 2, \ldots$, und Punkte $x_{in}\in U_n\cap K_n$ mit $x_{in}\neq x_{jn}$ für $i\neq j$; $i, j=1, 2, 3$ (also $K_n=K(x_{1n}, x_{2n}, x_{3n})$ und $x=lim_n x_{in}$). Das System der $\mathfrak{k}$-Paratingenten $P(x)$ an C im Punkt x sei mit $\mathfrak{p}(x)$ bezeichnet. – Liegen alle x_{in}, $i=1, 2, 3$; $n=1, 2, \ldots$, in vorderen bzw. alle in hinteren Umgebungen von x auf C, so wird der $\lim_n K_n$ als eine *vordere* bzw. als eine *hintere* ($\mathfrak{k}$-)*Paratingente*

$P_v(x)$ bzw. $P_h(x)$ an C in x bezeichnet und das System dieser $P_v(x)$ bzw. $P_h(x)$ mit $\mathfrak{p}_v(x)$ bzw. mit $\mathfrak{p}_h(x)$.

Anmerkung. (a) Gemäss der Forderung (II) (1) folgt aus der Mehrpunktigkeit von $P(x)$, und allgemeiner der eines $L=\lim K_n$, schon $P(x)\in\mathfrak{k}$ bzw. $L\in\mathfrak{k}$. In der Tat: Als Limes von Kontinuen ist L ein (mehrpunktiges) Kontinuum, enthält also jedenfalls 3 verschiedene Punkte x_i, $i=1, 2, 3$. Wegen $L=\lim K_n$ existieren $x_{in}\in K_n$ mit $x_i=\lim x_{in}$, $i=1, 2, 3$; $n=1, 2, \ldots$. Gemäss (II) (1) folgt daraus $L\in\mathfrak{k}$. – (b) Die Forderungen (I) (2) und (II) (1) zusammen mit denen aus Nr. 3.1.1. bezüglich $\mathfrak{k}$ ziehen nach sich: *Durch beliebige* 3 *verschiedene Punkte* $y_i\in C$ *geht (genau) eine* OCh $K=K(y_1, y_2, y_3)$. *Ferner ist* $\mathrm{POW}(C;\mathfrak{k})\geqslant 4$. In der Tat: Wegen (I) (2) ist $\mathrm{POW}(C;\mathfrak{k})\geqslant 3$; weil aber C Kurve (in $\underline{G}$) ist, folgt $\mathrm{POW}(C;\mathfrak{k})\geqslant 4$ (vgl. Nr. 3.1.1.). Es seien nun $y_i\in C$ beliebig, wobei y_i auf dem orientierten C vor y_{i+1} liegt, $i=1, 2, 3$. Ist $C'=C(y_1\,|\,y_2\,|\,y_3)\subset C$, so wählen wir $x\in\underset{\sim}{C}''=C-C'$. Gemäss (I) (2) gibt es in $\underset{\sim}{C}''$ drei Punkte z_i, $i=1, 2, 3$, die auf einem $K=K(z_1, z_2, z_3)\in\mathfrak{k}$ liegen. O. B. d. A. wird dabei z_i als vor z_{i+1} auf C gelegen angenommen. Gemäss Nr. 3.1.1. existiert eine hintere Umgebung H von z_3 derart, dass $K(z_1, z_2, z_3')\in\mathfrak{k}$ für jedes $z_3'\in\bar{H}$ existiert. Vermöge analytischer Induktion (vgl. Lorenzen [1]) kann man nun auf eine hintere Umgebung des hinteren Endpunktes von H übergehen bzw. dieses Verfahren bis zur Erreichung von y_3 fortsetzen. Es existiert also $K(z_1, z_2, y_3)\in\mathfrak{k}$. Entsprechend führt man z_2 in y_2 bei festen z_1, y_3 über usw.

3.2. Definition. Ein Punkt $s\in C$ heisse ($\mathfrak{k}$-)Scheitel von C, wenn $\mathrm{POW}(s;C;\mathfrak{k})\geqslant 4$ ist. In jeder Umgebung von s auf C gibt es also (mindestens) 4 Punkte, die auf dem gleichen $K\in\mathfrak{k}$ liegen.

3.3. Der Fall $\mathrm{POW}(C;k)=4$.

3.3.1. Es seien x', $x''\in C$ vorgegeben mit $x'\neq x''$.

Definitionen. Mit $\mathfrak{k}(x')$ bzw. $\mathfrak{k}(x', x'')$ sei das System aller $K\in\mathfrak{k}$ bezeichnet, für die $x'\in K$ bzw. x', $x''\in K$ gilt. Es sei weiter $y\in C$ mit $y\neq x'$ bzw. mit $x'\neq y\neq x''$. Man bezeichnet dann y als einen $\mathfrak{k}(x')$- bzw. $\mathfrak{k}(x', x'')$-Scheitel, wenn beliebig kleine Umgebungen von y auf C (mindestens) 3 bzw. (mindestens) 2 Punkte enthalten, die auf dem gleichen $K\in\mathfrak{k}(x')$ bzw. $K\in\mathfrak{k}(x', x'')$ liegen.

3.3.1.1. Forderung. Es sei y_n', $y_n''\in C$ mit $y_n'\neq y_n''$ und $K_n'\in\mathfrak{k}(x', x'')$ mit y_n', $y_n''\in K_n'$, ferner existierte $y=\lim y_n'=\lim y_n''$ mit $x'\neq y\neq x''$ sowie $L=\lim K_n$. *Dann soll L eine* OCh *sein* ($L\in\mathfrak{k}(x', x'')$) Man bezeichnet ein solches L als eine $\mathfrak{k}(x', x'')$-(Scheitel)-Paratingente *an* C (in $\mathfrak{k}(x', x'')$-Scheitel y), in *Zeichen*: $L=P(y;x', x'')$.

Entsprechend sei y_n', y_n'', $y_n\in C$ mit $y_n\neq y_n''\neq y_n\neq y_n'$ und y_n', y_n'', $y_n\in K_n\in\mathfrak{k}(x')$; es existiere $y=\lim y_n'=\lim y_n''=\lim y_n$ mit $y\neq x'$ sowie $L=\lim K_n$. Dann soll wieder

$L \in \mathfrak{k}(x')$ sein. Und man bezeichnet L als eine $\mathfrak{k}(x')$-(Scheitel-)Paratingente $P(y; x')$ in y an C.

3.3.2. VIERSCHEITELSATZ FÜR DEN FALL POW$(C; \mathfrak{k}) = 4$.

VOR. (1) *Es sollen die in* Nr. 3.1.1. *und* 3.1.2. *angegebenen Annahmen erfüllt sein.* – (2)- *Bei beliebigen* $x' \in C$ *bzw.* x', $x'' \in C$ *existiere in jedem (von* x' *bzw. von den* x', x'' *verschiedenen)* $\mathfrak{k}(x')$- bzw. $\mathfrak{k}(x', x'')$-*Scheitel* y genau *eine* $\mathfrak{k}(x')$ – *bzw. genau eine* $\mathfrak{k}(x', x'')$-*Paratingente* $P(y)$, *die selbst* OCh *(und mehrpunktig) ist. Ebenso soll es in jedem* $\mathfrak{k}$-*Scheitel von* C *nur eine einzige* $\mathfrak{k}$-*Paratingente* $P(y)$ *an* C *geben.*

BEH. *Wenn* POW$(C; \mathfrak{k}) = 4$ *ist, besitzt* C genau *vier* $\mathfrak{k}$-*Scheitel.*

ZUSATZ. Jeder $\mathfrak{k}$-Scheitel von C ist signiert und stark-symmetrisch (im Sinne von Nr. 3.4.1.). Sind ferner s_0, s_1, s_2, s_3 die vier $\mathfrak{k}$-Scheitel von C in der einer Orientierung von C entsprechenden Reihenfolge, so ist $\text{sign}(s_i) = (-1)_i \text{sign}(s_0)$, $i = 1, 2, 3$, (Vgl. auch H [1]).

ANMERKUNG. Dieser Vierscheitelsatz für POW$(C; \mathfrak{k}) = 4$ ist schärfer als der in Nr. 3.4.2 für POW$(C; \mathfrak{k}) \geqslant 6$ angegebene insofern, als der erstere die Existenz nicht nur von *mindestens* sondern von *genau* vier $\mathfrak{k}$-Scheitel behauptet.

Bew. (I) Zunächst ist jedes 4-tupel (x_1, x_2, x_3, x_4) aus Schnittpunkten von C mit einem $K' \in \mathfrak{k}$ normal, d.h. die x_1, x_2, x_3, x_4 sind, in dieser Reihenfolge, sowohl direkt C-konsekutiv als (direkt) K'-konsekutiv. Sind nämlich – bei orientiertem C – die x_i direkt C-konsekutiv, so gibt es eine Orientierung von K', bei welcher x_1 und x_2 direkt K'-konsekutiv sind, wie aus POW$(C; \mathfrak{k}) = 4$ folgt. Und daraus ergibt sich weiter, dass dann x_1, x_2, x_3, x_4 direkt K'-konsekutiv sind. – (II) Zufolge der Vor (1) und (2) zusammen mit der Normalität von C bezüglich $\mathfrak{k}$ (siehe Ziff. (I)) ist H.–K. [1], Nr. 4.1.4.3.1., Satz, für den Fall $k = 3$ anwendbar.

Bew. des Zusatzes. Gemäss der Beh. des Satzes ist jeder $\mathfrak{k}$-Scheitel von C isoliert. Gemäss Ziff. (I) des Bew. des Satzes besitzt jeder $\mathfrak{k}$-Scheitel von C eine $\mathfrak{k}$-normale Umgebung auf C; und gemäss Vor. (2) existiert in jedem $\mathfrak{k}$-Scheitel genau eine $\mathfrak{k}$-Paratingente. Daher folgt der Zusatz aus H [1], Nr. 3.2., Satz, sowie Nr. 3.3., Satz 1., Beh. (2), für $k = 3$.

3.4. Der Fall höchstens endlichen POW$(C; k) > 6$.

3.4.1. DEFINITIONEN.

(a) Ein $\mathfrak{k}$-Scheitel s von C heisse signiert, wenn eine Seite $C(\alpha)$ von C und eine Umgebung W von s auf C existiert derart, dass jedes, in W enthaltene fast normale 4-tupel $\mathfrak{t} = (x_1, x_2, x_3, x_4) \subset W \cap K$ die Signatur α besitzt und dass es fast normale 4-tupel in beliebiger Nähe von s gibt. Es heise dann α die Signatur des $\mathfrak{k}$-Scheitels s, in Zeichen $\alpha = \text{sign}(s)$.

(b) Ein $\mathfrak{k}$-Scheitel s von C heisse stark symmetrisch, wenn folgendes gilt: Es gibt eine Umgebung W von s auf C sowie in W enthaltene vordere und hintere Umgebungen V bzw. H von s auf C mit folgender Eigenschaft: Für jedes $K \in \mathfrak{k}$ mit $\mathrm{POW}(\underset{\sim}{V} \cap K) = 3$ gilt $\mathrm{POW}(W \cap K) \geqslant 4$; und entsprechend folgt $\mathrm{POW}(W \cap K) \geqslant 4$ aus $\mathrm{POW}(\underset{\sim}{H} \cap K) = 3$.

Anmerkung. Beispiele signierter und zugleich stark symmetrischer $\mathfrak{k}$-Scheitel liefern die isolierten Scheitel im Sinne der klassischen Differentialgeometrie (vgl. H [1], Nr. 4.4.); dabei ist also $\mathfrak{k}$ das System der Kreise und Kreisbogen in einer Kreisscheibe G und $C \subset \underline{G}$ besitzt stetige (endliche) Krümmung.

3.4.2. Es gilt nun der folgende

Vierscheitelsatz für den Fall $\mathrm{POW}(C; \mathfrak{k}) \geqslant 6$.

Vor. (1) *Es seien die in* Nr. 3.1. *angegebenen Forderungen bezüglich* $\mathfrak{k}$ *und* C *erfüllt; insbesondere sei also* $\mathrm{POW}(C; \mathfrak{k})$ *höchstens endlich.* – (2) *Es sei* $\mathrm{POW}(C; \mathfrak{k}) \geqslant 6$. *Und jeder isolierte* $\mathfrak{k}$*-Scheitel sei signiert und stark symmetrisch.*

Beh. *Es besitzt C mindestens vier* $\mathfrak{k}$*-Scheitel.*

Anmerkung. (1) Gemäss Vor. (1) enthält insbesondere kein $C \cap K$ einen Teilbogen von K. – (2) Der Fall, dass $\mathrm{POW}(C; \mathfrak{k})$ unendlich ist, bleibe späterer Erörterung vorbehalten. Es bleibe hier auch dahin gestellt, inwieweit die Höchstens-Endlichkeit von $\mathrm{POW}(C; \mathfrak{k})$ im Falle endlich vieler $\mathfrak{k}$-Scheitel Folge der Vor. (2) ist.

3.4.3. Beweis des Vierscheitelsatzes für $\mathrm{POW}(C; \mathfrak{k}) \geqslant 6$.

(I) Besitzt C unendlich viele $\mathfrak{k}$-Scheitel, so ist die Beh. richtig. Daher wird im Folgender jeder Scheitel als isoliert (sowie als signiert und als stark symmetrisch) angenommen.

(II) Abkürzungen. Es sei „α-Scheitel" oder „α-*Sch.*" Abkürzung für „$\mathfrak{k}$-Scheitel der Signatur α", ferner „4-f. n." oder nur „f. n." für „fast normales 4-tupel" und „α f. n." für „fast normales 4-tupel der Signatur α", schliesslich „*Q-kons.*" bzw. „*dir. Q-kons.*" für „Q-konsekutiv" bzw. „direkt Q-konsekutiv". – Ist $\mathfrak{t} = (1, 2, 3, 4)$ ein f. n., so sei $C(\mathfrak{t}) = C(1 | 2 | 3 | 4) \subset C$ gesetzt. Ist weiter auch $\mathfrak{t}' = (1', 2', 3', 4')$ ein f. n. und ist $C(\mathfrak{t}') \subset C(\mathfrak{t})$, so kann $\mathfrak{t}' \subset \mathfrak{t}$ geschrieben werden.

Es sei der Bogen B bzw. die Kurve Q gegeben als topologisches Bild eines Intervalls J bzw. einer Kreisperipherie P. Ferner sei T ein Teilbogen von B oder ein echter Teilbogen von Q. Man setzt dann $|T|$ gleich der Länge des Urbildes von T in J bzw. in $\mathfrak{P}$. Es gilt: $|T| = |\underset{\sim}{T}|$. Für hinreichend kleines $|T|$ ist der Durchmesser von T beliebig klein; und umgekehrt. Für Teilbogen T', T'' von B bzw. von Q mit $\underset{\sim}{T'} \cap \underset{\sim}{T''} = \emptyset$ gilt $|T'| + |T''| = |T' \cup T''|$.

(III) *Ein Teilbogen* $T=\bar{T}$ *von* C *enthält einen* α*-Scheitel genau dann, wenn eine Folge von* α *f.n.* $\mathfrak{t}_n$, $n=1, 2, \ldots$, *existiert* (*mit* $\operatorname{sign}(\mathfrak{t}_n)=\alpha$ *und*) *mit* $T\cap C(\mathfrak{t}_n)\neq\emptyset$ *für alle* n *sowie mit* $\lim|C(\mathfrak{t}_n)|=0$.

In der Tat: Ist $s\in T$ ein α-Sch., so existieren Folgen derartiger $\mathfrak{t}_n$. – Umgekehrt: Aus $T\cap C(\mathfrak{t}_n)\neq\emptyset$ folgt wegen der Kompaktheit von $D(\mathfrak{t}_n)=T\cap C(\mathfrak{t}_n)\subset C$, dass eine Teilfolge $(\mathfrak{t}'_m)_m$ von $(\mathfrak{t}_n)_n$ existiert, für welche $\{D(\mathfrak{t}'_m);\ m=1, 2, \ldots\}$ einen Häufungspunkt h besitzt. Wegen $\lim|C(\mathfrak{t}'_m)|=0$ liegen schliesslich alle $C(\mathfrak{t}'_m)$ in beliebiger Nähe von h. Daher ist h ein $\mathfrak{k}$-Scheitel; und da h signiert sein soll, ist h sogar ein α-Sch.

Anmerkung. Beim Beweis ist nur benutzt, dass die Scheitel signiert (und isoliert) sein sollen, nicht aber ihre starke Symmetrie.

3.4.3.1. Existenzsatz. Ist $\mathfrak{t}_0$ ein fast normales 4-tupel mit $\operatorname{sign}(\mathfrak{t}_0)=\alpha$, so existieren Folgen von fast normales 4-tupeln $\mathfrak{t}_n$, $n=1, 2, \ldots$, der folgenden Art: Die Folge beginnt mit $\mathfrak{t}_0$; es ist $\operatorname{sign}(\mathfrak{t}_n)=\operatorname{sign}(\mathfrak{t}_0)$ für jedes n; die kleinsten, $\mathfrak{t}_n$ enthaltenden Teilbogen $C(\mathfrak{t}_n)$ von C ziehen sich auf einen Punkt $s\in\underset{\sim}{C}(\mathfrak{t}_0)$ zusammen (d.h. dass schliesslich alle $C(\mathfrak{t}_n)$ in beliebig kleinen Umgebungen von s auf C enthalten, also die $|C(\mathfrak{t}_n)|$ beliebig klein sind).

Der Beweis dieses Existenzsatzes wird in einer zweiten Mitteilung erbracht.

1. Anmerkung. Bei dem Beweis des Existenzsatzes wird benutzt, dass die α-Scheitel sämtlich stark symmetrisch sein sollen.

2. Anmerkung. Gemäss H.–K. [1], Nr. 4.5.4. ff., lässt sich aus jedem f.n. $\mathfrak{t}_0$ durch einen Kontraktionsprozess eine Folge von f.n. $\mathfrak{t}_n$ gewinnen, die gegen einen $\mathfrak{k}$-Scheitel konvergieren. Bei dem dortigen Kontraktionsprozess ist aber nicht garantiert, dass $\operatorname{sign}(\mathfrak{t}_0)=\operatorname{sign}(\mathfrak{t}_n)$ ist für alle n. Es ist also über die Signatur des Scheitels, der durch den Kontraktionsprozess gewonnen wird, nichts bekannt. Daher wird zum Beweise des Existenzsatzes ein modifizierter Kontraktionsprozess entwickelt.

3.4.4. Aus dem Existenzsatz (Nr. 3.4.3.1.) sowie aus Nr. 3.4.3., Ziff. (III), folgt unmittelbar:

Ist $\mathfrak{t}_0=(1, 2, 3, 4)$ *ein fast normales* 4*-tupel in* $C\cap K$, *so gibt es ein* $T_0=\bar{T}_0\subset\underset{\sim}{C}(\mathfrak{t}_0)==\underset{\sim}{C}(1|2|3|4)$ *derart, dass in* T_0 *(mindestens) ein* $\mathfrak{k}$*-Scheitel* s *mit* $\operatorname{sign}(s)=\operatorname{sign}(\mathfrak{t}_0)$ *enthalten ist.*

3.4.5. Erledigung des Falles $\operatorname{POW}(C;\mathfrak{k})=6$.

(A) *Im Fall* II'. der Nr. 2.2. liefern die beiden f.n. quasi-fremden 4-tupel der Signatur α (gemäss Nr. 3.4.4.) zwei verschiedene Scheitel $s'(\alpha)$, $s''(\alpha)$ je der Signatur α; entsprechend für die beiden quasi-fremden f.n. 4-tupel der Signatur $-\alpha$. Da aber alle Scheitel signiert sein sollen, sind Scheitel verschiedener Signatur verschieden. Daraus folgt die Existenz von mindestens 4 Scheiteln usw.

(B) Im Fall I'. der Nr. 2.2. kann man so schliessen: Es sei $(1, 2, 3, 4, 5, 6) \subset C \cap K$ ein diesem Fall zugehöriges 6-tupel (von Schnittpunkten). Dann ist $\mathfrak{t}' = (1, 2, 3, 4)$ und $\mathfrak{t}'' = (2, 3, 4, 5)$ f.n. mit $\operatorname{sign}(\mathfrak{t}') = -\operatorname{sign}(\mathfrak{t}'')$. Es existieren also Scheitel $s(\mathfrak{t}')$, $s(\mathfrak{t}'')$ mit $\operatorname{sign}(s(\mathfrak{t}')) = -\operatorname{sign}(s(\mathfrak{t}''))$ und mit $s(\mathfrak{t}') \subset \underset{\sim}{C}(1|2|3|4)$, $s(\mathfrak{t}'') \subset \underset{\sim}{C}(2|3|4|5)$. Wir unterscheiden jetzt die Fälle:

FALL (I′1) Es ist $s(\mathfrak{t}') \in \underset{\sim}{C}(1|2|3)$. – FALL (I′2) Es ist $s(\mathfrak{t}') \in C(3|4)$, wobei $C(3|4)$ durch $\underset{\sim}{C}(1|2|3) \cap C(3|4) = \emptyset$ eindeutig bestimmt ist.

Betr. Fall (I′1). Da $\mathfrak{r}' = (3, 4, 5, 6)$ ebenfalls f.n. ist mit $\operatorname{sign}(\mathfrak{r}') = \operatorname{sign}(s(\mathfrak{t}'))$, existiert ein Scheitel $s(\mathfrak{r}') \in \underset{\sim}{C}(3|4|5|6)$ mit $\operatorname{sign}(s(\mathfrak{r}')) = \operatorname{sign}(s(\mathfrak{t}'))$. Und wegen $\underset{\sim}{C}(1|2|3) \cap C(3|4|5|6) = \emptyset$ ist $s(\mathfrak{r}') \neq s(\mathfrak{t}')$. Es existieren also zwei verschiedene Scheitel der gleichen Signatur wie $s(\mathfrak{t}')$.

Betr. Fall (I′2). Da hier $\mathfrak{r}' = (5, 6, 1, 2)$ f.n. ist mit $\operatorname{sign}(\mathfrak{r}') = \operatorname{sign}(\mathfrak{t}')$, existiert $s(\mathfrak{r}') \subset \underset{\sim}{C}(5|6|1|2)$ mit $\operatorname{sign}(s(\mathfrak{r}')) = \operatorname{sign}(s(\mathfrak{t}'))$. Wegen $C(3|4) \cap \underset{\sim}{C}(5|6|1|2) = \emptyset$ ist wieder $s(\mathfrak{r}') \neq s(\mathfrak{t}')$.

Zyklische Vertauschung von i mit $i+1$, wobei 2 statt $6+1$ zu setzen ist, zeigt, dass man für $s(\mathfrak{t}'')$ mit $\operatorname{sign}(\mathfrak{t}'') = -\operatorname{sign}(\mathfrak{t}')$ (vgl. oben) ebenso auf die Existenz eines $s(\mathfrak{r}'')$ mit $\operatorname{sign}(s(\mathfrak{r}'')) = \operatorname{sign}(\mathfrak{t}'') = -\operatorname{sign}(s(\mathfrak{t}'))$ und mit $s(\mathfrak{r}'') \neq s(\mathfrak{t}'')$ schliessen kann. Damit ist der Vierscheitelsatz für den Fall $\operatorname{POW}(C; \mathfrak{k}) = 6$ bewiesen.

3.4.6. Erledigung der Fälle höchstens endlichen $\operatorname{POW}(C; \mathfrak{k}) \geqslant 8$.

Gemäss Nr. 2.3. und 2.4. existieren hier stets zwei Paare je quasi-fremder, f.n. 4-tupel je von gleicher Signatur, während die Signaturen von 4-tupel, die verschiedenen Paaren zugehören, verschieden sind. Wie in Nr. 3.4.5., (A), ergibt sich die Richtigkeit des Vierscheitelsatzes mit Hilfe des Existenzsatzes (Nr. 3.4.3.1.).

LITERATURHINWEISE

Benz, W., [1] *Über Möbiusebenen*, Jber. Deutsch. Math.-Verein. *63*, 1–27 (1966).

Haupt, O., und Künneth, H. [1] *Geometrische Ordnungen* (Springer-Verlag, Berlin–Heidelberg–New York 1967). Zitiert mit H.-K. [1].

Haupt, O., [1] *Bemerkungen zum Kneserschen Vierscheitelsatz*, Abh. Math. Sem. Univ. Hamburg *31*, 218–238 (1967). Zitiert mit H [1].

Kneser, A., [1] *Bemerkungen über die Anzahl der Extreme der Krümmung auf geschlossenen Kurven und über verwandte Fragen in einer nichteuklidischen Geometrie*, in *Festschrift z. 70. Geburtstag von H. Weber* (Leipzig–Berlin 1912), pp. 170–180.

Kneser, H., [1] *Neuer Beweis des Vierscheitelsatzes*, Christian Huygens *2* (1922–1923), 315–318.

Lorenzen, P., [1] *Differential und Integral* (Frankfurt (Main) 1965).

Universität Erlangen–Nürnberg

Sur le reste de certaines formules de quadrature

TIBERIU POPOVICIU à Cluj (Roumanie)

Dédié à M. A. Ostrowski à l'occasion de son 75ième anniversaire

1. On peut exprimer de plusieurs manières le reste $R[f]$ de la formule de quadrature

$$\int_a^b f(x)\,dx = \sum_{\alpha=1}^{n} A_\alpha f(x_\alpha) + R[f] \tag{1}$$

où x_α $(\alpha=1, 2,\dots, n)$ sont des points distincts de l'axe réel et A_α, $\alpha=1, 2,\dots, n$ des constantes réelles données, indépendantes de la fonction f. Nous supposerons que a et b, où $a<b$, soient finis, que toutes les fonctions considérées soient réelles, et nous désignerons par I un intervalle contenant les points a, b, x_α, $(\alpha=1, 2,\dots, n)$.

Soit m le degré d'exactitude du reste $R[f]$ (ou de la formule de quadrature (1)), donc le nombre, bien déterminé par les conditions: $R[1]=R[x]=\cdots=R[x^m]=0$, $R[x^{m+1}]\neq 0$. Si $R[1]\neq 0$ nous prenons $m=-1$ et si $R[1]=0$ nous avons $0\leqq m\leqq 2n-1$.

Si f est une fonction continue sur l'intervalle I nous avons [1]

$$R[f] = A[\xi_1, \xi_2,\dots, \xi_{m+2}; f] + B[\xi'_1, \xi'_2,\dots, \xi'_{m+2}; f]. \tag{2}$$

Les points ξ_α, d'une part, et les points ξ'_α, d'autre part, sont distincts mais dépendent en général de la fonction f. Les constantes A, B sont indépendantes de la fonction f. Nous avons $A+B=R[x^{m+1}]$ et $[y_1, y_2,\dots, y_r; f]$ désigne la différence divisée, d'ordre $r-1$, de la fonction f sur les points (ou noeuds) $y_1, y_2,\dots, y_r$. Si $m\geqq 0$ les points ξ_α, ξ'_α, peuvent être choisis à l'intérieur de l'intervalle I et si de plus la fonction f a une $(m+1)$-ième dérivée $f^{(m+1)}$ à l'intérieur de I, on a

$$R[f] = A\frac{f^{(m+1)}(\xi)}{(m+1)!} + B\frac{f^{(m+1)}(\xi')}{(m+1)!} \tag{3}$$

ξ, ξ', étant deux points de l'intérieur de l'intervalle I et A, B étant, d'ailleurs, les mêmes constantes que dans (2).

Si on peut prendre $B=0$ dans (2) ou dans (3), on dit que le reste $R[f]$ est *de la forme simple*. Nous avons donné autrefois des conditions nécessaires et suffisantes pour qu'il en soit ainsi [1], mais le reste n'est pas toujours de la forme simple.

Nous allons montrer que même si le reste n'est pas de la forme simple, dans le sens précédent, nous pouvons, dans certains cas, introduire des différences divisées

Reçu le 12.12.67

modifiées de manière que le reste devient de la forme simple par rapport à ces nouvelles différences divisées. La définition de la différence divisée qui interviendra ici résultera de ce qui suit.

2. Soit toujours m le degré d'exactitude de $R[f]$ et désignons par $k(0 \leqq k \leqq n)$ le nombre des points x_α compris dans l'intervalle ouvert (a, b). Si $k > 0$ nous pouvons supposer que les points $x_1, x_2, \ldots, x_k$, sont dans (a, b) et si $k < n$ que les points x_{k+1}, $x_{k+2}, \ldots, x_n$, sont en dehors de (a, b). Désignons par c, d $(c \leqq a < b \leqq d)$, les extrémités de l'intervalle I et considérons les polynomes

$$P_\alpha = (x - c)^\alpha - \frac{R[(x - c)^\alpha]}{R[(x - c)^{m+1}]}(x - c)^{m+1} \qquad (\alpha = 0, 1, \ldots, n + k)$$

qui sont bien définis puisque $R[(x-c)^{m+1}] = R[x^{m+1}] \neq 0$ et qui vérifient les égalités

$$R[P_\alpha] = 0, (\alpha = 0, 1, \ldots, n + k). \tag{4}$$

Il est facile de voir qu'on doit avoir $m \leqq n + k - 1$. En effet si nous prenons le polynome

$$l = (x - x_1)^2 (x - x_2)^2 \ldots (x - x_k)^2 (x - x_{k+1})(x - x_{k+2}) \ldots (x - x_n)$$

(dont la forme est facile à écrire si $k = 0$ ou $k = n$), nous avons

$$R[l] = \int_a^b l(x)\, dx \neq 0. \tag{5}$$

Désignons maintenant par $D(y_1, y_2, \ldots, y_{n+k+1}; f)$ le déterminant des valeurs des fonctions $P_0, P_1, \ldots, P_m, P_{m+2}, P_{m+3}, \ldots, P_{n+k}, f$ (si $m = -1$ des fonctions $P_1, P_2, \ldots$, P_{n+k}, f) sur les points $y_1, y_2, \ldots, y_{n+k+1}$, et par $D^*(y_1, y_2, \ldots, y_{n+k})$ le mineur correspondant à l'élément $f(y_{n+k+1})$ de ce déterminant, donc le déterminant des valeurs des fonctions $P_0, P_1, \ldots, P_m, P_{m+2}, P_{m+3}, \ldots, P_{n+k}$ sur les points $y_1, y_2, \ldots, y_{n+k}$. Nous avons $D(y_1, y_2, \ldots, y_{n+k+1}; x^{m+1}) = (-1)^{n+k-m-1} V(y_1, y_2, \ldots, y_{n+k+1})$ où $V(y_1, y_2, \ldots$, $y_{n+k+1})$ est le déterminant de Vandermonde des nombres $y_1, y_2, \ldots, y_{n+k+1}$.

La nouvelle différence divisée que nous introduisons est définie par la formule

$$[y_1, y_2, \ldots, y_{n+k+1}; f]^* = \frac{D(y_1, y_2, \ldots, y_{n+k+1}; f)}{D(y_1, y_2, \ldots, y_{n+k+1}; x^{m+1})}. \tag{6}$$

Dans cette définition nous avons supposé que les points y_α sont distincts. Mais on peut étendre la définition aussi au cas des points y_α non pas nécessairement distincts par un passage à la limite dans la formule (6). Ceci revient a maintenir la formule de définition (6) en modifiant convenablement le déterminant D (et aussi les déterminants

D^*, V) tel que nous l'avons expliqué antérieurement [1]. Ceci exige l'existence d'un certain nombre de dérivées de la fonction f. De cette manière la différence divisée (6) est définie quels que soient les points y_α distincts ou non.

3. Avec les notations précédentes considérons la fonction continue $\varphi(x)$ qui en dehors des points x_α est égale à la combinaison linéaire $l(x)[x_1, x_1, x_2, x_2, \ldots, x_k, x_k, x_{k+1}, x_{k+2}, \ldots, x_n, x; f]^*$ des fonctions $P_0, P_1 \ldots, P_m, P_{m+1}, P_{m+2}, \ldots, P_{n+k}, f$. On a alors $\varphi(x_\alpha)=0$ $(\alpha=1, 2, \ldots, n)$ donc aussi

$$R[\varphi] = \int_a^b \varphi(x)\, dx$$

et compte tenant de (4), (5), nous avons

$$R[f] = \frac{R[x^{m+1}]}{R[l]} \int_a^b \varphi(x)\, dx .$$

En remarquant que la fonction l ne change pas de signe sur (a, b), il en résulte que

$$R[f] = R[x^{m+1}]\,[\xi_1, \xi_2, \ldots, \xi_{n+k+1}; f]^* \tag{7}$$

où les ξ_α sont des points distincts de l'intérieur de I (dépendant en général de la fonction f), *à condition qu'on puisse appliquer les théorèmes de la moyenne aux différences divisées* (6). Je réfère le lecteur pour ces théorèmes à un travail précédent [1]. Cette condition est sûrement vérifiée si:

(H). *Le déterminant* $D^*(y_1, y_2, \ldots, y_{n+k})$ *est* $\neq 0$, *quels que soient les points* $y_1, y_2, \ldots, y_{n+k}$ *distincts ou non.*

Dans la démonstration précédente de la formule (7) on a supposé que la fonction continue f ait une dérivée (au moins sur les points $x_1, x_2, \ldots, x_k$). On peut voir facilement que le résultat obtenu est valable sous la seule hypothèse de continuité sur l'intervalle fermé $[a, b]$ de la fonction f.

Si nous désignons par $W(g_1, g_2, \ldots, g_s)$ le wronskien des fonctions $g_1, g_2, \ldots, g_s$, si nous supposons de plus que la fonction f ait une $(n+k)$-ième dérivée sur l'intérieur de I et si la condition (H) est vérifiée, nous avons (le wronskien a ses lignes et ses colonnes dans l'ordre habituel)

$$R[f] = \frac{(-1)^{n+k-m-1}}{1!\,2!\ldots(n+k)!} R[x^{m+1}]\{W(P_0, P_1, \ldots, P_m, P_{m+2}, P_{m+3}, \ldots, P_{n+k}, f)\}_{x=\xi}$$

ξ étant un point intérieur de I. Remarquons que le second membre de cette formule est de la forme

$$\sum_{\alpha=0}^{n+k-m-1} Q_\alpha(\xi) f^{(m+1+\alpha)}(\xi)$$

où Q_α est un polynome de degré α indépendant de la fonction f $(\alpha=0, 1, \ldots, n+k-m-1)$.

4. Si $p_0=1$ et p_α $(\alpha=0, 1, \ldots, n+k)$ sont les fonctions symétriques fondamentales des nombres $y_\alpha - c$, $(\alpha=1, 2, \ldots, n+k)$, nous avons

$$D^*(y_1, y_2, \ldots, y_{n+k}) =$$
$$= V(y_1, y_2, \ldots, y_{n+k}) \sum_{\alpha=1}^{n+k-m-1} (-1)^\alpha \frac{R[(x-c)^{m+1+\alpha}]}{R[(x-c)^{m+1}]} P_{n+k-m-1-\alpha}$$

Il en résulte que la condition (H) est vérifiée si les nombres

$$\sum_{\alpha=0}^{n+k-m-1} (-1)^\alpha \frac{R[(x-c)^{m+1+\alpha}]}{R[(x-c)^{m+1}]} \binom{s}{n+k-m-1-\alpha} (d-c)^{n+k-m-1-\alpha}$$
$$(s = 0, 1, \ldots, n+k) \qquad (8)$$

sont tous positifs ou tous négatifs.

Exemple. Considérons la formule de quadrature

$$\int_0^2 f(x)\, dx = \frac{2f(0) + 6f(1) + 2f(2)}{5} + R[f].$$

Dans ce cas nous avons $k=m=1$ et nous pouvons prendre $c=0$, $d=2$. Le reste n'est pas de la forme simple habituelle (on a $R[f]=\frac{7}{375}f''(\xi)-\frac{32}{375}f''(\xi')$ si la dérivée seconde existe). Dans ce cas les nombres (8) sont tous de même signe et différents de 0, donc la condition (H) est vérifiée. Si la fonction f est continue sur [0, 2] et a une dérivée quatrième sur (0, 2), nous avons

$$R[f] = -\tfrac{1}{60}[4f''(\xi) - 4(\xi-1)f'''(\xi) + (2\xi^2 - 4\xi + 3)f^{\text{IV}}(\xi)], \quad 0 < \xi < 2.$$

5. On peut facilement étendre les considérations précédentes aux formule de quadrature où le second membre contient aussi linéairement certaines des valeurs sur les points x_α des dérivées succesives de la fonction f et aussi à des formules d'approximation linéaire correspondant à des fonctionnelles linéaires plus générales.

BIBLIOGRAPHIE

[1] Popoviciu, T., *Sur le reste dans certaines formules linéaires d'approximation de l'analyse*, Mathematica, Cluj *1* (24), 95–142 (1959).

Institutul de Calcul

Endomorphismenringe in der Galoisschen Theorie

WOLFGANG KRULL (Bonn)

Herrn Alexander Ostrowski zum 75. Geburtstag am 25.9.1968 gewidmet.

Es sei N ein separabler Normaloberkörper des Körpers K vom endlichen Grade $[N:K]=n$. Seine Galoisgruppe sei $G=\{A_1,\dots A_n\}$. Mit $K(G)$ bezeichnen wir den Gruppenring von G über K, also den Ring aller Linearformen $\sum_{i=1}^{n} a_i A_i$ $(a_i \in K)$ mit den Multiplikationsregeln $(a_i A_i)(a_k A_k)=(a_i a_k)(A_i A_k)$. Folgt man bei der Entwicklung der Galoisschen Theorie für N dem Vorbild von ARTIN, so wird man zunächst N als *Vektorraum* über K betrachten und nicht nur die A_i als Endomorphismen von N interpretieren, sondern auch jedem $a \in K$ den Endomorphismus a' zuordnen, der durch die Vorschrift '$a'\alpha = a\cdot\alpha$ für alle $\alpha \in N$' definiert ist. Vermeidet man, was doch wohl empfehlenswert ist, die Identifizierung von a und a', so bildet die Menge $\{a' \mid a \in K\}$ einen zu K kanonisch isomorphen Körper K', und es ist der Ring K'_G aller Endomorphismen $\sum_{i=1}^{n} a'_i A_i$ zum Gruppenring $K(G)$ isomorph aufgrund des *Unabhängigkeitssatzes*:

Aus $(\sum_{i=1}^{n} a'_i A_i)\alpha = 0$ *für alle* $\alpha \in N$ *folgt stets* $a_1=\dots=a_n=0$, *die* A_i *sind also über* K' *linear unabhängig.*

Angesichts der grundsätzlichen Bedeutung des Unabhängigkeitssatzes dürfte der Hinweis nicht überflüssig sein, daß dieser Satz in einer viel allgemeineren Fassung als sie für den Spezialfall der Galoisschen Theorie nötig ist, völlig elementar bewiesen werden kann:

ALLGEMEINER UNABHÄNGIGKEITSSATZ: *Es sei* N *ein ganz beliebiger Körper; jedem* $\eta \in N$ *werde der durch '*$\eta'\alpha = \eta\cdot\alpha$ *für alle* $\alpha \in N$*' festgelegte Endomorphismus* η' *der additiven Gruppe* N *zugeordnet; sind dann* $A_1,\dots A_n$ *irgendwelche paarweise verschiedene Automorphismen des Körpers* N *(und damit natürlich erst recht Endomorphismen der additiven Gruppe* N*), so folgt aus* $(\sum_{i=1}^{n} \eta'_i A_i)\alpha = 0$ *für alle* $\alpha \in N$ *stets* $\eta_1=\dots=\eta_n=0$.

Beim Beweis werde zunächst angenommen, es existiere ein $\alpha \in N$, derart daß $A_i\alpha \neq A_k\alpha$ für $i\neq k$ mit $i,k=1,\dots n$. Dann ist die Vandermonde'sche Determinante $|A_i\alpha^k|$ $(i=1,\dots n;\ k=0,\dots n-1)$ von 0 verschieden. Sind also die Elemente $\eta_1,\dots\eta_n$ aus N nicht alle 0, so ist auch $\sum_{i=1}^{n} \eta_i(A_i\alpha^k) = \sum_{i=1}^{n} (\eta'_i A_i)\alpha^k \neq 0$ für mindestens eine der Elemente $\alpha^0, \alpha, \alpha^2, \dots \alpha^{n-1}$.

Eingegangen am 13.12.1967

Um den allgemeinsten Fall zu erfassen, erweitere man N durch Adjunktion von abzählbar unendlich vielen Unbestimmten $u_1, u_2, \dots$ zum Körper N_u und erweitere jeden Automorphismus A von N stillschweigend zu einem Automorphismus A von N_u durch die Vorschrift '$Au_i = u_i$ für jedes i'. Sind nun die Automorphismen $A_1, \dots A_n$ von N paarweise verschieden, gibt es also für $i \neq k$ stets ein $\alpha_{ik} \in N$, für das $A_i \alpha_{ik} \neq A_k \alpha_{ik}$, so existiert für hinreichend großes M in N_u eine Linearform $\varphi(u) = \sum_{i=1}^{M} \alpha_i u_i$ mit $\alpha_i \in N$, derart daß $A_i \varphi(u) \neq A_k \varphi(u)$ für $i \neq k$ und es wird $\varphi_k(u) = \varphi(u)^k$ eine homogene Form k-ten Grades in $u_1, \dots u_M$ über N ($k = 0, 1, 2, \dots$; $\varphi_0(u) = 1$, $\varphi_1(u) = \varphi(u)$). Nach dem ersten Teil des Beweises weiß man nun: Sind die Elemente $\eta_1, \dots \eta_n$ aus N nicht alle geleich 0, so gibt es ein k_0 mit $0 \leqslant k_0 \leqslant n-1$, derart daß $\sum_{i=1}^{n} \eta_i \cdot A_i \varphi_{k_0}(u) \neq 0$, und daraus folgt weiter: Für mindestens einen Koeffizienten α von $\varphi_{k_0}(u)$ wird $\sum_{i=1}^{n} \eta_i (A_i \alpha) = \sum_{i=1}^{n} (\eta'_i A_i) \alpha \neq 0$.

Man beachte: Der Beweis ist konstruktiv. Er zeigt: Der allgemeine Unabhängigkeitssatz folgt einfach aus der Bemerkung, daß zwar stets $A\alpha^k = (A\alpha)^k$, aber – von trivialen Ausnahmefällen abgesehen – niemals $\eta' \alpha^k = (\eta' \alpha)^k$ ist. Es hat nun offenbar auch in der gewöhnlichen Galoisschen Theorie keinen Sinn, mit einem Spezialfall zu beginnen, indem man nur die Endomorphismen a' ($a \in K$) und nicht beliebige Endomorphismen η' ($\eta \in N$) zuläßt. D.h. aber: Man wird, vom Unabhängigkeitssatz ausgehend, nicht zuerst auf den Endomorphismenring K'_G des K-Moduls N sondern gleich auf den Endomorphismenring $N'_G = \{ \sum_{i=1}^{n} \eta'_i A_i \mid \eta_i \in N \}$ der Abelschen Gruppe N geführt (Bezeichnungen wie zu Beginn der Note). Allerdings hat N'_G eine wesentlich kompliziertere Struktur als K'_G. Vor allem ist N'_G nicht mehr zum Gruppenring $N(G)$ isomorph, weil in N'_G nicht die Multiplikationsregel $(\eta'_i A_i)(\eta'_k A_k) = (\eta'_i \eta'_k)(A_i A_k)$, sondern die Multiplikationsregel $(\eta'_i A_i)(\eta'_k A_k) = (\eta'_i (A_i \eta_k)')(A_i A_k)$ gilt. Es fragt sich daher: Besitzt der Ring N'_G überhaupt leicht zu erfassende, an sich interessante Eigenschaften, die es rechtfertigen, sich schon in der elementaren Galoisschen Theorie näher mit ihm zu beschäftigen und nicht gleich zu dem einfacheren Unterring K'_G überzugehen? Es soll gezeigt werden, daß diese Frage zu bejahen ist und daß vor allem das Studium von N'_G auf einen neuen Ring $R'_{G,K}$ (einfacher Natur) führt, der zum mindesten gleichberechtigt neben K'_G tritt.

Bekanntlich gibt es zu den über K' (und über N') linear unabhängigen Endomorphismen $A_1, \dots A_n$ stets ein $\alpha \in N$, derart daß $A_1 \alpha, \dots A_n \alpha$ über K linear unabhängig sind (Satz von der Existenz der Normalbasis). J. Neukirch hat sich nun die Frage vorgelegt, ob ein entsprechender Satz auch dann gilt, wenn $B_1, \dots B_n$ beliebige über N' linear unabhängige Endomorphismen aus N'_G sind, und er fand die folgende negative Antwort (briefliche Mitteilung):

Wählt man α so, daß $A_1\alpha, \dots A_n\alpha$ über K linear unabhängig sind, und setzt man

$$P_i = \sum_{k=1}^{n} (A_k A_i \alpha)' A_k \qquad (i = 1, \dots n), \tag{1}$$

so sind die P_i zwar linear unabhängig über N', aber es gilt $P_i\gamma \in K$ $(i=1,\dots n)$ für alle $\gamma \in N$. $P_1\gamma, P_2\gamma, \dots P_n\gamma$ sind also stets linear abhängig über K.

Die lineare Unabhängigkeit der P_i ergibt sich in bekannter Weise aus der Tatsache, daß die Determinante $|A_k A_i \alpha|$ $(k, i=1,\dots n)$ wegen der Wahl von α von 0 verschieden ist. Andererseits hat man

$$\left(\sum_{k=1}^{n} (A_k A_i \alpha)' A_k\right) \beta = \sum_{k=1}^{n} ((A_k A_i \alpha)' A_k) \beta$$
$$= \sum_{k=1}^{n} (A_k A_i \alpha) \cdot A_k \beta = \sum_{k=1}^{n} A_k((A_i \alpha) \beta) = \mathrm{Spur}((A_i \alpha) \beta) \in K$$

für jedes $\beta \in N$. – Das Neukirchsche Ergebnis gestattet nun leicht die Lösung einer anderen Aufgabe, die sich eigentlich im Rahmen der Galoisschen Theorie sofort aufdrängt, sobald man einmal den Ring N'_G gebildet hat. Ist L ein beliebiger Körper zwischen K und N, so liegt es auf der Hand, die Menge $R'_{G,L}$ aller der $\phi \in N'_G$ zu betrachten, die der Bedingung '$\phi\alpha \in L$ *für alle* $\alpha \in N$' genügen. Die $R'_{G,L}$ sind offenbar *Rechtsideale* in N'_G, aber, vom Trivialfall $L=N$ abgesehen, *keine zweiseitigen Ideale.* Denn für jedes $\phi \neq 0$ aus N'_G bildet wegen des Unabhängigkeitssatzes schon die Menge $\{\eta'\phi \mid \eta \in N\}$ ganz N *auf* sich selbst ab. Wir denken uns nun die P_i im Sinne Neukirchs durch Auszeichnung eines bestimmten α fest gewählt, und behaupten:

Es ist $R'_{G,L} = L'P_1 + \cdots + L'P_n$ für jeden Körper L zwischen K und N. Dabei ist natürlich $L'P_1 + \cdots + L'P_n = \{\sum_{i=1}^{n} \lambda'_i P_i \mid \lambda_i \in L\}$.

Beweis: Die $R'_{G,L}$ sind offenbar alles Linksvektorräume über K'.

Das gleiche gilt für die $L'P_1 + \cdots + L'P_n$. Außerdem ist nach der Grundeigenschaft der P_i stets $R'_{G,L} \supseteq L'P_1 + \cdots + L'P_n$. Es genügt daher, zu zeigen, daß stets $R'_{G,L}$ und $L'P_1 + \cdots + L'P_n$ über K' die gleiche Dimension haben. Die Dimension von $L'P_1 + \cdots + L'P_n$ ist gleich $[L:K] \cdot n$ wegen der linearen Unabhängigkeit der P_i, und die von N'_G speziell ist gleich n^2. Man braucht daher nur für jedes L' einen Vektorraum $S'_{G,L}$ anzugeben, der über K' die Dimension $n \cdot (n - [L:K])$ besitzt, und bei dem $N'_G = S'_{G,L} + R'_{G,L}$, aber $\{0\} = S'_{G,L} \cap R'_{G,L}$ wird, woraus sich nach der linearen Algebra sofort für $R'_{G,L}$ die gewünschte Dimension $[L:K] \cdot n$ ergibt. – Es sei nun etwa, indem wir im Augenblick zur Ersparung der $'$ stets K mit K' und L mit L' identifizieren, $[L:K]=r$, $[N:L]=s$ und $1=\gamma_1, \gamma_2, \dots \gamma_r$ bzw. $1=\delta_1, \delta_2, \dots \delta_s$ eine Basis von L über K bzw. N über L. Dann hat $LP_1 + \cdots + LP_n$ über K die Basis $\{\gamma_i P_k \mid i=1,\dots r; k=1,\dots n\}$, und es wird infolgedessen sicher $N'_G = S'_{G,L} + R'_{G,L}$, wenn wir unter $S'_{G,L}$ den Vektorraum über K mit der Basis $\{\delta_l \gamma_i P_k \mid l=2,\dots s;\ i=1,\dots r;\ k=1,\dots n\}$ verstehen, dessen

Dimension den gewünschten Wert $n(n-r)=n(n-[L:K])$ besitzt. Es sei nun $\phi=\sum_{l=2}^{s}\sum_{i=1}^{r}\sum_{k=1}^{n} a_{lik}\delta_l\gamma_i P_k$ beliebig aus $S'_{G,L}$ und $\eta\in N$. Dann hat man $\phi\eta=\sum_{l=2}^{s}\sum_{i=1}^{r}\delta_l\gamma_i\times\sum_{k=1}^{n} a_{lik}(P_k\eta)$, und nach Wahl der δ_l, γ_i, sowie wegen $P_k\eta\in K$ $(k=1,\ldots n)$ ist $\phi\eta\in L$ gleichwertig mit $\sum_{k=1}^{n} a_{lik}(P_k\eta)=\sum_{k=1}^{n}(a_{lik}P_k)\eta=0$ für $l=2,\ldots s$; $i=1,\ldots r$. Es führt also die Annahme "$\phi\eta\in L$ für alle η" wegen der linearen Unabhängigkeit der P_k und des Unabhängigkeitssatzes auf $a_{lik}=0$ $(l=2,\ldots s;\ i=1,\ldots r;\ k=1,\ldots n)$, d.h. $\phi=0$.

Das Hauptinteresse beim Studium der Rechtsideale von N'_G richtet sich naturgemäß jetzt auf $R'_{G,K}$. Indem wir die zuletzt weggelassenen $'$ wieder schreiben, erhalten wir die folgende *basisfreie Charakterisierung* von $R'_{G,K}$:

Es ist $R'_{G,K}=\{\sum_{k=1}^{n}(A_k\eta)'A_k \mid \eta\in N\}$.

In der Tat, wegen $R'_{G,K}=K'P_1+\cdots+K'P_n$ haben wir für ein beliebiges $P\in R'_{G,K}$ eine Darstellung:

$$P=\sum_{i=1}^{n} a'_i P_i=\sum_{i=1}^{n} a'_i\sum_{k=1}^{n}(A_k A_i\alpha)'\,A_k$$
$$=\sum_{k=1}^{n}\left[\sum_{i=1}^{n} a'_i(A_k A_i\alpha)'\,A_k\right]=\sum_{k=1}^{n}\left[A_k\sum_{i=1}^{n} a_i(A_i\alpha)\right]'A_k$$

Da nun die Elemente $A_i\alpha$ eine Normalbasis von N über K darstellen, wird $N==\{\sum_{i=1}^{n} a_i A_i\alpha \mid a_i\in K\}$, d.h., es ergibt sich $P=\sum_{k=1}^{n}(A_k\eta)'A_k$, wobei η den Körper N durchläuft, wenn man für P alle Elemente aus $R'_{G,K}$ zuläßt. – Die invariante Charakterisierung von $R'_{G,K}$ läßt die zunächst nur begrifflich erschlossene Rechtsidealeigenschaft von $R'_{G,K}$ unmittelbar erkennen; denn man hat:

$$\left[\sum_{k=1}^{n}(A_k\eta)'\,A_k\right]A_l=\sum_{k=1}^{n}\,[A_k A_l(A_l^{-1}\eta)]'\,A_k A_l=\sum_{k=1}^{n}(A_k\vartheta)'\,A_k \quad \text{mit} \quad \vartheta=A_l^{-1}\eta,$$

und weiter:

$$\left[\sum_{k=1}^{n}(A_k\eta)'\,A_k\right]\vartheta'=\sum_{k=1}^{n}(A_k\eta)'\,(A_k\vartheta')=\sum_{k=1}^{n}((A_k\eta)'\,(A_k\vartheta)')\,A_k=\sum_{k=1}^{n}(A_k(\eta\vartheta))'\,A_k\,.$$

Darüber hinaus ergibt sich:

$R'_{G,K}$ ist ein minimales Rechtsideal in N'_G.

Es sei nämlich $P=\sum_{k=1}^{n}(A_k\eta)'A_k\neq 0$ beliebig aus $R'_{G,K}$ und ϑ beliebig aus N. Dann haben wir $P\vartheta'=\sum_{k=1}^{n}(A_k(\eta\vartheta))'A_k$, und da $\eta\vartheta$ mit η ganz N durchläuft, wird $R'_{G,K}$ das durch P erzeugte Rechtsideal. – Wichtiger aber ist, daß das Rechtsideal $R'_{G,K}$, auf das wir, ausgehend von N'_G durch eine naturgegebene Überlegung zwangsweise kamen, auch auf einem ganz anderen Wege gewonnen werden kann. Betrachtet man den Körper N über K vom Standpunkt der linearen Algebra, so kann man von der *Spurenbildung* als einer fundamentalen linearen Operation ausgehen, und man erinnert sich dann sofort daran, daß in der Diskriminantentheorie (letzten Endes nach DEDEKIND) die symmetrische *Bilinearform* (η, ϑ) über K eine hervorragende Rolle spielt, die jedes Paar $\langle \eta, \vartheta \rangle$ aus $N \times N$ durch die Vorschrift $(\eta, \vartheta) = \mathrm{Spur}(\eta \cdot \vartheta) = \sum_{k=1}^{n} A_k(\eta \cdot \vartheta)$ bilinear in K abbildet. Hält man nun η in der Bilinearform (η, ϑ) fest, während ϑ ganz N durchläuft, so entsteht gerade die lineare Abbildung $P = \sum_{k=1}^{n} (A_k \eta)' A_k$ von N in K; die Abbildungen P bilden wegen der Bilinearität von (η, ϑ) einen Vektorraum über K', – und dieser Vektorraum ist gerade das minimale Rechtsideal $R'_{G,K}$. – Damit haben wir das Endergebnis: Will man dem Normalkörper N mit Hilfe der Gruppe G einen Vektorraum über K' zuordnen, so hat man nicht *eine*, sondern zwei ausgezeichnete Möglichkeiten. Denn es tritt neben den als Endomorphismenring von N in sich gedeuteten Gruppenring K'_G der Vektorraum $R'_{G,K}$, der aus Endomorphismen von N in K besteht und natürlich (als Rechtsideal in N'_G) gleichfalls Ringeigenschaft besitzt. Der fundamentale Zusammenhang von $R'_{G,K}$ mit der Spurenbildung, ganz abgesehen von unseren idealtheoretischen Betrachtungen, weist darauf hin, daß man in Zukunft der Bedeutung von $R'_{G,K}$ für die Galoissche Theorie größere Beachtung schenken sollte.

Mathematisches Institut der Universität Bonn

Quasi-Residual Designs

J. F. Lawless, R. C. Mullin (Waterloo, Ont., Canada),
R. G. Stanton (Toronto, Ont., Canada)

Dedicated to A. Ostrowski on His 75th Birthday

1. Introduction

We may partially specify a balanced incomplete block design by naming its parameter set (v, b, r, k, λ); such a design is an arrangement of v varieties into b blocks such that each block contains k distinct varieties, each variety occurs in r blocks, each pair of varieties occurs in λ blocks. It is well-known (see, for example, Bose [3] or Ryser [8]) that

$$b\,k = r\,v, \tag{1}$$

$$\lambda(v-1) = r(k-1). \tag{2}$$

Also [see Fisher (5)], one has the inequality

$$b \geqslant v. \tag{3}$$

If a given block is repeated α times (that is, there is a set of $\alpha+1$ identical blocks), then Stanton and Sprott [11] have stengthened the Fisher inequality to give

$$b \geqslant (\alpha+1)\,v - (\alpha-1). \tag{4}$$

In particular, it is implicit from (4) that, if $b<2v$, then there are no repeated blocks.

If $b=v$ (and $r=k$), we speak of a symmetrical balanced incomplete block design or a v-k-λ system (see Mann [7], Ryser [8], Stanton [9]). From such a design, one can always form a residual design by deleting a single block and the varieties in it; this residual design has parameter set

$$(V, B, R, K, \Lambda),$$

where $V=v-k$, $B=v-1$, $R=k$, $K=k-\lambda$, $\Lambda=\lambda$.

We shall speak of a quasi-residual design as one which has the parameters of a residual design. It was first observed by Bhattacharya [2] that there exist quasi-residual designs which are not residual designs; his example is readily available in Hall [6]. It is the purpose of this paper to discuss quasi-residual designs, using the results of Stanton-Sprott [11] and Stanton-Mullin [10] to specify structural properties of such designs.

Received December 13, 1967

2. Block Intersections in Quasi-Residual Designs

Let B_1 be a fixed block, and let a_i denote the frequency with which the integer i appears among the block-intersection numbers (that is, the number of elements in $B_1 \cap B_j$). Then the numbers a_i satisfy the following relations (we use the form given in (11)).

$$\sum_{\varepsilon=0}^{K} a_i = B - 1 \tag{5}$$

$$\sum_{i=0}^{K} i\, a_i = K(R-1) \tag{6}$$

$$\sum_{i=0}^{K} i^2 a_i = K(K\Lambda - K - \Lambda + R) \tag{7}$$

In a quasi-residual design, (7) simplifies to

$$\sum_{i=0}^{K} i^2 a_i = K^2 \Lambda . \tag{8}$$

We note that the original symmetric design has parameters

$$(v, v, k, k, \lambda),$$

where $v = 1 + k(k-1)/\lambda$. Hence

$$V = (k^2 - \lambda k - k + \lambda)/\lambda, \quad B = k(k-1)/\lambda,$$
$$R = k, \quad K = k - \lambda, \quad \Lambda = \lambda .$$

This parameter set can be simplified to

$$V = K(K + \lambda - 1)/\lambda, \quad B = (K+\lambda)(K+\lambda-1)/\lambda .$$
$$R = K + \lambda, K, \lambda,$$

and equations (5), (6), and (8) become

$$a_0 + a_1 + a_2 + \cdots + a_K = (K^2 + \lambda^2 + 2K\lambda - K - 2\lambda)/\lambda, \tag{9}$$

$$a_1 + 2a_2 + \cdots + K a_K = K(K + \lambda - 1), \tag{10}$$

$$a_1 + 4a_2 + \cdots + K^2 a_K = K^2 \lambda . \tag{11}$$

Multiply these equations by $\lambda(\lambda-1)$, $-(2\lambda-1)$, 1, respectively, and add. (These multipliers are chosen so as to eliminate the terms in a_λ and $a_{\lambda-1}$). Then a little algebra produces

$$\sum_{i=0}^{K} [i - \lambda]\,[i - (\lambda - 1)]\, a_i = \lambda(\lambda-1)(\lambda-2). \tag{12}$$

Equation (12) is a fundamental result from which a number of other results follow.

3. Quasi-Residual Designs with Extremal Intersections

From equation (12), we see that $a_i=0$ if

$$[i-\lambda][i-(\lambda-1)] > \lambda(\lambda-1)(\lambda-2). \tag{13}$$

This quadratic in i can be solved and the result embodied in

THEOREM 1. *In a quasi-residual design, $a_i=0$ if*

$$i > \tfrac{1}{2}[2\lambda-1+\sqrt{1+4\lambda(\lambda-1)(\lambda-2)}].$$

It is interesting to consider those extremal designs in which i takes its largest possible value. This will occur if

$$1+4\lambda(\lambda-1)(\lambda-2)=p^2, \tag{14}$$

where p is a positive integer.

The first few solutions of this equation are $\lambda=1, 2, 3$, and 7. In these cases the extremal designs with a_i as large as possible have the following forms.

$$\lambda=1, \quad a_i=0 \quad \text{for} \quad i>1 \quad \text{(see Theorem 1)}. \tag{1}$$

$$\lambda=2, \quad a_i=0 \quad \text{for} \quad i>2 \quad \text{(see Theorem 2)}. \tag{2}$$

$$\lambda=3, \quad a_5=1 \quad \text{(see Theorem 3, Case II)}. \tag{3}$$

$$\lambda=7, \quad a_{21}=1, \quad a_7=(K^2-36K)/7-45, \quad a_6=7K+49. \tag{4}$$

It can be shown that no other solutions of Equation (14) are possible.

4. The Cases $\lambda=1$ and $\lambda=2$

If $\lambda=1$, equation (12) becomes

$$\sum_{i=0}^{K}(i-1)(i)\,a_i=0. \tag{15}$$

It follows, since $a_i \geqslant 0$, that $a_i=0$ for $i>1$. We thus have

THEOREM 2. *In a quasi-residual design with $\lambda=1$, we must have $V=K^2$, $B=K^2+K$, $R=K+1$, with $a_0=K-1$, $a_1=K^2$.*

This property, namely, that every block has the same intersection properties with the other blocks, carries over to $\lambda=2$. For in that case equation (12) becomes

$$\sum_{i=0}^{K}(i-2)(i-1)\,a_i=0. \tag{16}$$

Again since $a_i \geqslant 0$, we see that $a_0 = 0$, $a_i = 0$ for $i > 2$; hence we obtain

THEOREM 3. *In a quasi-residual design with* $\lambda = 2$, *we have* $V = K(K+1)/2$, $B = (K+2)(K+1)/2$, $R = K+2$, *with* $a_1 = 2K$, $a_2 = (K^2 - K)/2$.

The results of Theorem 3 can be employed to verify the correctness of the discussion given by BURAU [3] of the design (16, 16, 6, 6, 2), as opposed to the larger number of non-isomorphic systems claimed by ATIQULLAH [1]. Of course, Theorems 2 and 3 follow, in a different manner, from the Hall-Connor result (see [6]) that, for $\lambda = 1$ and $\lambda = 2$, every quasi-residual design is a residual design.

4. The Case $\lambda = 3$

The case $\lambda = 3$ is especially interesting since, for the first time, it is possible to have $a_i > 0$ for $i > \lambda$, in virtue of the Bhattacharya example mentioned in Section 1. Equation (12) becomes

$$\sum_{i=0}^{K} (i-3)(i-2)\, a_i = 6. \tag{17}$$

Since $(i-3)(i-2) > 6$ for $i > 5$, we see that $a_i = 0$ for $i > 5$; hence (15) reduces to

$$3(a_0 + a_5) + (a_1 + a_4) = 3. \tag{18}$$

This Diophantine equation has six solutions as described in

THEOREM 4. *In a quasi-residual design with* $\lambda = 3$, *we have* $V = K(K+2)/3$, $B = (K+3)(K+2)/3$, $R = K+3$. *The numbers* a_i *are given by one of the following cases* ($a_i = 0$ *for* $i > 5$). *In particular,* $a_K = 0$ *in all cases, that is, it is not possible to have a repeated block in a quasi-residual design with* $\lambda = 3$. *Note that the value of* a_i *depends both on the design and on the fixed block* B_1 *under consideration.*

Table 1

Case	a_0	a_1	a_2	a_3	a_4	a_5
I	1	0	$3K$	$(K^2-4K)/3$	0	0
II	0	0	$3K+5$	$(K^2-4K)/3-5$	0	1
III	0	0	$3K+6$	$(K^2-4K)/3-8$	3	0
IV	0	1	$3K+3$	$(K^2-4K)/3-5$	2	0
V	0	2	$3K$	$(K^2-4K)/3-2$	1	0
VI	0	3	$3K-3$	$(K^2-4K)/3+1$	0	0

Two main questions are concerned with the existence of designs of the various types, as well as the possibility of extending them to symmetric designs (obviously not possible for those cases with $a_i > 0$ for some $i > 3$).

We now prove

THEOREM 5. *In a quasi-residual design with* $\lambda=3$, *if* $a_5=1$ (for some block B_1), *then* $K=15$.

Proof. If a block of type II exists, then we have

$$a_5=1, \quad a_2=3K+5, \quad a_3=\frac{K^2-4K}{3}-5, \quad a_0=a_1=a_4=0. \tag{1}$$

Consider then two blocks B_1 and B_2, where $|B_1\cap B_2|=5$.

Then, from the results (1) it is clear that, for the remaining $B-2$ blocks B_j, either $|B_i\cap B_j|=2$ or $|B_i\cap B_j|=3$ for $i=1, 2$, and $J=3,\ldots, B$.

Designate the blocks B_1 and B_2 as

$$\begin{aligned} &B_1\colon \theta_1\,\theta_2\,\theta_3\,\theta_4\,\theta_5\,a_1\,a_2\cdots a_{K-5} \quad (a_i\neq b_j)\quad i,j=1,2,\ldots,K-5\\ &B_2\colon \theta_1\,\theta_2\,\theta_3\,\theta_4\,\theta_5\,b_1\,b_2\cdots b_{K-5}\end{aligned}$$

Consider now the remaining blocks $B_3,\ldots, B_B$. Let us split these blocks into 3 basic types.

Type 1: B_j contains *two* varieties from *each* of B_j and B_2.

Type 2: B_j contains *two* varieties from *one* of B_1 or B_2, and *three* varieties from the other.

Type 3: B_j contains *three* varieties from *each* of B_1 and B_2.

Now, we form various sub-types, depending on the number of θ_i's which appear in the blocks $\{B_j\}$.

We observe that the following types are possible:

Type 1: (2 varieties from each of B_1 and B_2).
- *11: One* of the θ_i's occurs.
- *12: Two* of the θ_i's occur.
- *13: None* of the θ_i's occurs.

Type 2: (2 varieties from one of B_1 or B_2, three from the other).
- *21: One* of the θ_i's occurs.
- *22: Two* of the θ_i's occur.
- *23: None* of the θ_i's occurs.

Type 3: (3 varieties from each of B_1 and B_2).
- *31: One* of the θ_i's occurs.
- *32: Two* of the θ_i's occur.
- *33: Three* of the θ_i's occur.
- *34: None* of the θ_i's occurs.

The occurrences of θ_i's, a_i's and b_j's in each of these types are then as follows:

Type	*Type*	*Type*
11: $\theta_i a_j b_k \ldots$	*21:* $\theta_i a_j a_k b_l \ldots$ *or* $\theta_i a_j b_k b_l \ldots$	*31:* $\theta_i a_j a_k b_l b_m \ldots$
12: $\theta_i \theta_j \ldots$	*22:* $\theta_i \theta_j a_k \ldots$ *or* $\theta_i \theta_j b_k \ldots$	*32:* $\theta_i \theta_j a_k b_l \ldots$
13: $a_i a_j b_k b_l \ldots$	*23:* $a_i a_j a_k b_l b_m \ldots$ *or* $a_i a_j b_k b_l b_m \ldots$	*33:* $\theta_i \theta_j \theta_k \ldots$
		34: $a_i a_j a_k b_l b_m b_n \ldots$

Consider now the varieties $\theta_1, \ldots, \theta_5, a_1, \ldots, a_{K-5}, b_1, \ldots, b_{K-5}$, which occur in B_1 and B_2, and consider the types of *pairs* which these varieties can give. We may form the following table:

Type of Pair	*Number of occurrences of this type of pair in* $B_3, \ldots, B_B$
(i) $a_i b_j$	$n_1 = 3(K-5)^2$
(ii) $a_i a_j$ or $b_k b_l$	$n_2 = 2(K-5)(K-6)$
(iii) $\theta_i a_j$ or $\theta_i b_k$	$n_3 = 20(K-5)$
(iv) $\theta_i \theta_j$	$n_4 = 10$

This table is immediate, since B_1 and B_2 are

$$B_1 : \theta_1 \theta_2 \theta_3 \theta_4 \theta_5 a_1 a_2 \cdots a_{K-5}$$
$$B_2 : \theta_1 \theta_2 \theta_3 \theta_4 \theta_5 b_1 b_2 \cdots b_{K-5}$$

and each pair occurs exactly 3 times.

Let the number of blocks of type ij be x_{ij}. Then we have 10 unknowns x_{11}, X_{12}, x_{13}, x_{21}, x_{22}, x_{23}, x_{31}, x_{32}, x_{33}, x_{34}, and the following relations hold, observing our block types listed above.

$$x_{11} + 4x_{13} + 2x_{21} + 6x_{23} + 4x_{31} + x_{32} + 9x_{34} = n_1 \quad (1)$$
$$2x_{13} + x_{21} + 4x_{23} + 2x_{31} + 6x_{34} = n_2 \quad (2)$$
$$2x_{11} + 3x_{21} + 2x_{22} + 4x_{31} + 4x_{32} = n_3 \quad (3)$$
$$x_{12} + x_{22} + x_{32} + 3x_{33} = n_4 \quad (4)$$
$$x_{11} + 2x_{12} + x_{21} + 2x_{22} + x_{31} + 2x_{32} + 3x_{33} = 5(R-2) \quad (5)$$
$$2x_{11} + 4x_{13} + 3x_{21} + x_{22} + 5x_{23} + 4x_{31} + 2x_{32} + 6x_{34} = 2(R-1)(K-5) \quad (6)$$
$$x_{11} + x_{12} + x_{13} + x_{21} + x_{22} + x_{23} + x_{31} + x_{32} + x_{33} + x_{34} = B-2 \quad (7)$$

Here

$$n_1 = 3(K-5)^2 \qquad n_2 = 2(K-5)(K-6)$$
$$n_3 = 20(K-5) \qquad n_4 = 10$$
$$R = K+3 \qquad B = \frac{(K+2)(K+3)}{3}$$

We can not solve these equations uniquely, but let us solve for

$X_1=(x_{11}+x_{12}+x_{13})$, $X_2=(x_{21}+x_{22}+x_{23})$, and $X_3=(x_{31}+x_{32}+x_{33}+x_{34})$,

noting that X_i=number of blocks of type i ($i=1, 2, 3$). Then it follows, from Equations (1)–(7), that

$$X_1+X_2+X_3=B-2,$$
$$4X_1+5X_2+6X_3=10(R-2)+2(R-1)(K-5),$$
$$2X_1+2X_2=-2n_1+n_2-n_3-2n_4+10(R-2)+4(R-1)(K-5).$$

Hence

$$X_1=2(15-K),$$
$$X_2=10(K-5),$$
$$X_3=\frac{(K-15)(K-4)}{3}.$$

Evidently $K=15$ is the only integer for which all of these values are non-negative. This completes the proof of Theorem 5.

If such a design exists, it has parameters (85, 102, 18, 15, 3). Further, for such a design all remaining blocks $B_3,\dots, B_{102}$ are of type 2, if B_1 and B_2 are the type II blocks of Table 1. We may also deduce that $X_{21}=60$, $X_{22}=10$, and $X_{23}=30$. Although the existence of such a design is unknown, we note that the corresponding symmetrical design with $v=103$, $k=18$, $\lambda=3$ does not exist.

While complete results are not yet available, it is worth noting that Cases III, IV, and V can, like Case II, occur for only a finite number of values of K. This result requires considerable investigation, and will be proved in a later paper.

REFERENCES

[1] ATTIQULLAH, M., *On Configurations and Non-Isomorphisms of Some Incomplete Block Designs*, Sankhya *20*, 227–248 (1958).
[2] BHATTACHARYA, K. N., *On a New Symmetrical Balanced Incomplete Block Design*, Bull. Calcutta Math. Soc. *36*, 91–96 (1945).
[3] BOSE, R. C., *On the Construction of Balanced Incomplete Block Designs*, Ann. Eugenics *9*, 353–399 (1939).
[4] BURAU, W., *Über die zur Kummerkonfiguration analogen Schemata von 16 Punkten und 16 Blöcken und ihre Gruppen*, Abh. Math. Sem. Univ. Hamburg *26*, 129–144 (1963/64).
[5] FISHER, R. A., *An Examination of the Different Possible Solutions of a Problem in Incomplete Blocks*, Ann. Eugenics *10*, 52–75 (1940).
[6] HALL JR., M., *Block Designs*, Chapter 13, in *Applied Combinatorial Mathematics*, ed. E. F. BECKENBACH (John Wiley, New York 1964).
[7] MANN, H. B., *Analysis and Design of Experiments* (Dover, New York 1949).
[8] RYSER, H. J., *Combinatorial Mathematics*, Math. Assoc. Amer. (John Wiley, New York 1963).

[9] STANTON, R. G., *Types of Statistical Designs* (Mimeographed Notes), Canad. Math. Congress, Seminar, Saskatoon (1963).
[10] STANTON, R. G. and MULLIN, R. C., *Existence and Uniqueness of BIBDs*, J. Combinatorial Theory (to appear).
[11] STANTON, R. G. and SPROTT, D. A., *Block Intersections in Balanced Incomplete Block Designs*, Canad. Math. Bull. 7, 539–548 (1964).

University of Waterloo
York University

Some Remarks on a Functional Equation Characterising the Root

MAREK KUCZMA (Katowice, Poland)

Dedicated to Professor Alexander M. Ostrowski on the occasion of his 75th birthday

§ 1. At 4th Meeting on the Functional Equations (Oberwolfach, July 1966) Professor A. OSTROWSKI presented the following result [3].

THEOREM A. *If the functions $f(x)$ and $g(x)$ are rational and the sequence of iterates $f^n(x)$ is not periodic, then algebraic solutions $\varphi(x)$ of the functional equation*

$$\varphi[f(x)] = g(x)\,\varphi(x) \tag{1}$$

are k-th roots of rational functions.

The following theorem yields an interesting illustration of that result.

THEOREM 1. *The only algebraic solutions of the functional equation*

$$\varphi(x^{p+1}) = x\,\varphi(x), \tag{2}$$

where $p \geqslant 2$ is a fixed integer, are the functions

$$\varphi(x) = c\sqrt[p]{x}, \tag{3}$$

with an arbitrary (complex) constant c.

Proof. Let $\varphi(x)$ be an algebraic solution of (2). The function $\psi(x) = \varphi(x)/\sqrt[p]{x}$ is an algebraic solution of the functional equation

$$\psi(x^{p+1}) = \psi(x) \tag{4}$$

and by Theorem A $\psi(x) = \sqrt[k]{R(x)}$, where $R(x)$ is rational. $R(x)$ is also a solution of (4), and thus it must be constant, since it takes the same value on the infinite set of points x_n with $x_{n+1} = x_n^{p+1}$, $x_0 \neq 0$. Hence also $\psi(x)$ is constant, say $\psi(x) \equiv c$, and (3) follows.

§ 2. Functions (3) can also be characterized as the analytic solutions of equation (2). We shall appeal to a theorem due to Mrs. W. SMAJDOR [4]. Below we quote a special case of her theorem concerning the linear homogeneous equation (cf. also [2], Chapter VIII).

THEOREM B. *Let functions $f(x)$ and $g(x)$ be analytic in a neighbourhood of a point ξ, $f(\xi) = \xi$, $0 \leqslant |f'(\xi)| < 1$. If $g(\xi) = 1$ or $g(\xi) = [f'(\xi)]^k$, where k is a positive integer, then the functional equation*

$$\varphi(x) = g(x)\,\varphi[f(x)] \tag{5}$$

Received December 14, 1967.

has a unique one-parameter family of local analytic solutions in a neighbourhood of ξ. *In the other cases* $\varphi(x)\equiv 0$ *is the only analytic solution of equation* (5) *in a neighbourhood of* ξ.

In the particular case, where $g(\xi)=1$, the local analytic solutions of equation (5) in a neighbourhood of ξ are given by the formula

$$\varphi(x) = c\prod_{\nu=0}^{\infty} g[f^{\nu}(x)],$$

where $f^{\nu}(x)$ is the ν-th iterate of $f(x)$ and c is the parameter.

Let $\mathfrak{R}$ denote the complex plane cut along the negative real axis. (The negative real axis could be replaced by an arbitrary ray issuing from the origin and not passing through the point $x=1$.) We have the following

THEOREM 2. *Functions* (3), *where* $\sqrt[p]{x}$ *denotes the branch of the root-function in* $\mathfrak{R}$ *that takes the value* 1 *for* $x=1$, *are the only analytic solutions of equation* (2) *in* $\mathfrak{R}$.

Proof. We write equation (2) in the form

$$\varphi(x) = \sqrt[p+1]{x}\,\varphi(\sqrt[p+1]{x}), \qquad |x-1|<1,$$

and by Theorem B ($\xi=1$) we obtain functions (3) as the only analytic solutions of (2) in a neighbourhood of $x=1$. However, since an analytic function is uniquely determined by its values in an arbitrarily small region, (3) must hold everywhere in $\mathfrak{R}$.

Theorem 2 results also from the following, more general one.

THEOREM 3. *For every* c, *function* (3), *where* $\sqrt[p]{x}$ *denotes the branch of the root-function in* $\mathfrak{R}$ *that takes the value* 1 *for* $x=1$, *is the only solution of equation* (2) *in* $\mathfrak{R}$ *such that*

$$\lim_{x\to 1}\varphi(x) = c. \tag{6}$$

The proof of the above theorem is based on the following

LEMMA. *Let*

$$x_{n+1} = \sqrt[p+1]{x_n}, \qquad x_0\in\mathfrak{R}, \tag{7}$$

where $\sqrt[p+1]{x}$ *is the branch of the root-function in* $\mathfrak{R}$ *with* $\sqrt[p+1]{1}=1$. *Then* $x_n\in\mathfrak{R}$ *for every* n *and, if* $x_0\neq 0$, $\lim_{n\to\infty} x_n=1$.

Proof of the lemma. Let $\operatorname{Arg} x_n=\theta_n$, $-\pi<\theta_n<\pi$. Then $\theta_{n+1}=\theta_n/(p+1)$ and thus $x_n\in\mathfrak{R}$ is proved by induction. Moreover, obviously $\lim_{n\to\infty}\theta_n=0$, whereas $\lim_{n\to\infty}|x_n|=1$ whenever $x_0\neq 0$.

Proof of the theorem. Let $\varphi(x)$ be a solution of equation (2) in $\mathfrak{R}$ fulfilling (6).

Then the function $\psi(x)=\varphi(x)/\sqrt[p]{x}$ satisfies equation (4) for $x\in\mathfrak{K}$, $x\neq 0$, and fulfils the condition

$$\lim_{x\to 1}\psi(x)=c. \tag{8}$$

Taking an arbitrary $x_0\in\mathfrak{K}$, $x_0\neq 0$, and defining the sequence x_n by (7), we get by induction in view of (4) $\psi(x_n)=\psi(x_0)$, whence by (8) and by the Lemma $\psi(x_0)=c$. Hence (3) holds for all $x\in\mathfrak{K}$, $x\neq 0$. For $x=0$, (3) results directly from (2) on setting $x=0$.

Whe shall present also a variant of the above proof. This variant does not require a use of $\sqrt[p]{x}$, but unfortunately it still does require a use of $\sqrt[p+1]{x}$.

Let $\varphi(x)$ be a solution of equation (2) in $\mathfrak{K}$ fufilling (6). Put

$$\gamma_1(x)=\frac{1}{c}\varphi(x^p),\qquad \gamma_2(x)=\left(\frac{1}{c}\varphi(x)\right)^p.$$

(2) and (6) imply that the functions γ_1 and γ_2 satisfy the functional equation

$$\gamma(x^{p+1})=x^p\gamma(x) \tag{9}$$

and fulfil the condition

$$\lim_{x\to 1}\gamma(x)=1. \tag{10}$$

The same argument as previously leads to the conclusion that $\gamma(x)/x=1$, i.e. $\gamma(x)=x$. In other words, the function $c^{-1}\varphi(x)$ is inverse to x^p, i.e. $\varphi(x)=c\sqrt[p]{x}$.

The equality $\gamma(x)=x$ may also be obtained directly from (9) and (10). Taking $x_0\in\mathfrak{K}$ we obtain from (9) for sequence (7)

$$\gamma(x_0)=\left(\prod_{\nu=1}^{n}x_\nu^p\right)\gamma(x_n),$$

whence by (10) and by the Lemma

$$\gamma(x_0)=\prod_{\nu=1}^{\infty}x_\nu^p=x_0.$$

Theorem 3 remains valid also in the case of a real variable, and the proof is essentially the same:

THEOREM 4. *For every (real) c, function* (3), *where* $\sqrt[p]{x}$ *denotes the arithmetic root of* $x\geqslant 0$, *is the only solution of equation* (2) *in* $[0,\infty)$ *fulfilling condition* (6).

The special case $p=2$ of the above theorem was proved in [1].

§ 3. In Theorems 2 and 3 the point $\xi=1$ played a particular rôle. Similar results can be obtained if $\xi=1$ is replaced by

$$\xi_k=\cos\frac{2k\pi}{p}+i\sin\frac{2k\pi}{p},\qquad k=1,\dots,p-1.$$

(All these points as well as $\xi=0$ are the fixed points of the function $f(x)=x^{p+1}$.) However, the situation changes if we take $\xi=0$: equation (2) has no non-trivial analytic solution in a neighbourhood of the origin. Below we are going to prove a more general result.

We consider equation (1) and let

$$f(x)=\xi+(x-\xi)^{p+1}F(x),\qquad F(\xi)\neq 0,\tag{11}$$

$$g(x)=(x-\xi)^{q}G(x),\qquad G(\xi)\neq 0,\tag{12}$$

and we assume that

$$q\geqslant 1.\tag{13}$$

Setting $x=\xi$ in (1) we obtain $\varphi(\xi)=0$ and thus we may write

$$\varphi(x)=(x-\xi)^{r}\Phi(x),\qquad \Phi(\xi)\neq 0,\qquad r\geqslant 1,\tag{14}$$

or $\varphi(x)\equiv 0$. Leaving the latter case aside, we get by inserting (11), (12) and (14) into (1)

$$(x-\xi)^{r(p+1)}(F(x))^{r}\,\Phi[f(x)]=(x-\xi)^{q+r}G(x)\,\Phi(x).\tag{15}$$

Comparing the orders of the zeros of the left-hand side and the right-hand side of (15) at $x=\xi$ gives

$$rp=q,\tag{16}$$

which means that q must be a multiple of p. In this case we get from (15)

$$\Phi(x)=\frac{(F(x))^{r}}{G(x)}\,\Phi[f(x)],\tag{17}$$

i.e. an equation of form (5). By Theorem B equation (17) has non-trivial local analytic solutions if and only if $(F(\xi))^{r}=G(\xi)$ and these solutions are given by

$$\Phi(x)=c\prod_{\nu=0}^{\infty}\frac{(F[f^{\nu}(x)])^{r}}{G[f^{\nu}(x)]}.$$

Hence we get the following result.

THEOREM 5. *Let $f(x)$ and $g(x)$ be analytic in a neighbourhood of ξ and suppose that they have form* (11) *and* (12) *with* (13), *respectively. If p divides q and $(F(\xi))^{r}=G(\xi)$, where r is determined by* (16), *then equation* (1) *has a unique one-parameter family of analytic solutions in a neighbourhood of ξ. These solutions are given by the formula*

$$\varphi(x)=c(x-\xi)^{r}\prod_{\nu=0}^{\infty}\frac{(F[f^{\nu}(x)])^{r}}{G[f^{\nu}(x)]}.$$

In other cases $\varphi(x)\equiv 0$ is the only analytic solution of equation (1) *in a neighbourhood of ξ.*

In the case of equation (2), where $\xi=0$, we have $q=1$ whereas $p\geqslant 2$. Thus Theorem 5 implies the following

COROLLARY. *Equation* (2) *has no non-trivial local analytic solution in a neighbourhood of* $x=0$.

REFERENCES

[1] BUREK, J., and KUCZMA, M., *Einige Bemerkungen über monotone und konvexe Lösungen gewisser Funktionalgleichungen*, Math. Nachr. *36*, 121–134 (1968).
[2] KUCZMA, M., *Functional Equations in a Single Variable* (PWN, Warszawa 1968 [Monografie Matematyczne, Vol. 46]).
[3] OSTROWSKI, A., *Über algebraische Lösungen Φ der Funktionalgleichung $\Phi(\varphi(x))=g(x)\,\Phi(x)$, für rationale $g(x)$, denen rationale φ entsprechen*, in *Die vierte Tagung über Funktionalgleichungen, Oberwolfach, 10.–16. Juli 1966*, Aequationes Math. *1*, 135–149 (1968).
[4] SMAJDOR, W., *On the Existence and Uniqueness of Analytic Solutions of the Functional Equation* $\varphi(z)=h(z,\varphi[f(z)])$, Ann. Polon. Math. *19*, 37–45 (1967).

Mathematical Institute of the Polish Academy of Sciences and Jagiellonian University, Branch in Katowice, Poland

Funktionalgleichungen in Vektorräumen, Kompositionsalgebren und Systeme partieller Differentialgleichungen[1])

WOLFGANG EICHHORN (Würzburg, Deutschland)

Herrn Professor Dr. A. M. Ostrowski zum 75. Geburtstag gewidmet

Einleitung

In dieser Arbeit wird das folgende Problem behandelt: X sei ein Vektorraum über einem beliebigen kommutativen Körper K der Charakteristik ungleich 2, x sei ein beliebiges Element aus X, $\mathrm{Hom}(X, X)$ sei die Algebra der linearen Transformationen von X. Gesucht sind lineare Abbildungen L von X in die Algebra $\mathrm{Hom}(X, X)$, zu denen weitere solche Abbildungen, nennen wir sie M, existieren derart, dass für das Produkt der beiden Transformationen $M(x)$ und $L(x)$

$$M(x)\,L(x) = \mu(x)\,I \qquad \left\{\begin{array}{l} I \text{ die identische Abbildung } X \to X \\ \mu \text{ eine Abbildung } X \to K,\ \mu(x) \not\equiv 0 \\ \text{Charakteristik von } K \text{ ungleich } 2 \end{array}\right\} \tag{1}$$

gilt.

In der Sprechweise der Theorie der Funktionalgleichungen kann man dieses Problem auch so formulieren: Gesucht sind Abbildungen $L: X \to \mathrm{Hom}(X, X)$, $M: X \to \mathrm{Hom}(X, X)$ und $\mu: X \to K$, die sowohl der Funktionalgleichung (1) als auch den folgenden Funktionalgleichungen genügen:

$$L(\alpha x + \beta y) = \alpha L(x) + \beta L(y) \tag{2}$$
$$M(\alpha x + \beta y) = \alpha M(x) + \beta M(y) \tag{3}$$
für alle $x, y \in X$ und $\alpha, \beta \in K$.

Aus (1), (2) und (3) ergibt sich sofort die weitere Funktionalgleichung

$$\mu(\alpha x) = \alpha^2 \mu(x) \qquad \text{für alle } x \in X \text{ und } \alpha \in K$$

sowie, wenn man

$$\mu(x, y) = \mu(y, x) \quad \text{für} \quad \tfrac{1}{2}[\mu(x+y) - \mu(x) - \mu(y)] \tag{4}$$

[1]) Die sich auf das Problem der Existenz elliptischer „verallgemeinerter Cauchy-Riemannscher Differentialgleichungssysteme“ beziehenden Untersuchungen der vorliegenden Arbeit bildeten einen Abschnitt in der Habilitationsschrift (Würzburg 1965) des Verfassers, die bis auf diesen Abschnitt inzwischen publiziert ist: [10], [11]. Die vorliegende Arbeit wurde während eines Gastaufenthaltes des Verfassers an der Faculty of Mathematics der University of Waterloo (Kanada) im Herbstsemester 1967 abgeschlossen und zum Teil durch NRC Grant Nr. A-2972 gefördert.

Eingegangen am 15.12.1967, und in veränderter Form am 27.3.1968

schreibt,

$$\begin{aligned}\mu(\alpha x + \beta z, y) &= \alpha\mu(x, y) + \beta\mu(z, y)\\ \mu(x, \alpha y + \beta z) &= \alpha\mu(x, y) + \beta\mu(x, z)\end{aligned} \quad \text{für alle } x, y, z \in X \text{ und } \alpha, \beta \in K.$$

$\mu(x, y)$ ist also eine symmetrische Bilinearform, $\mu(x)=\mu(x, x)$ eine quadratische Form.

Es ist bemerkenswert, dass die obige Fragestellung als eine Verallgemeinerung des klassischen Problems der Komposition der quadratischen Formen betrachtet werden kann; siehe Abschnitt I. Diese Verallgemeinerung ist beispielsweise aus den im folgenden unter 1 und 2 genannten Gründen von Interesse:

1. Man kann $L(x)$ mit den Eigenschaften (1), (2) als die linksreguläre Darstellung des Elements x einer (nicht notwendig assoziativen oder kommutativen) Algebra $\mathscr{A}$ im Vektorraum X auffassen, das heisst in X eine bilineare Multiplikation $(x, y) \to xy$, also eine bilineare Abbildung von $X \times X$ in X, durch[2])

$$xy := L(x)\,y \quad \left\{\begin{array}{l} x, y \in X;\ L(x)\,y \text{ bedeutet Anwendung der} \\ \text{linearen Transformation } L(x) \text{ auf } y \in X \end{array}\right\} \tag{5}$$

erklären. Die so mit Hilfe der Lösungen des Funktionalgleichungssystems (1), (2), (3) definierte Klasse von Algebren enthält möglicherweise neben der bekannten Klasse der Kompositionsalgebren (und der zu diesen isotopen[3]) Algebren) bisher noch nicht untersuchte Algebren. Aus den Abschnitten IV und V geht hervor, dass das tatsächlich der Fall ist, und zwar selbst dann, wenn die quadratische Form μ *nicht-ausgeartet* ist, das heisst wenn aus $\mu(x, X)=0$ stets $x=0$ folgt; unter dieser Voraussetzung allerdings *nur* dann, wenn die über $L(x)$ vermöge (5) definierte Algebra zu keiner Algebra mit Einselement isotop ist, und selbst dann nur in denjenigen Dimensionen $n \geqslant 2$, in denen auch Kompositionsalgebren existieren: $n=2, 4, 8$.

Bei der Herleitung dieses Ergebnisses wie auch bei dem in Abschnitt IV behandelten Problem der Existenz von Lösungen des Funktionalgleichungssystems (1), (2), (3) in Abhängigkeit von der Dimension des Vektorraumes X und dem Rang der quadratischen Form μ spielen Mengen antikommutativer (n, n)-Matrizen eine wichtige Rolle. Solche Mengen sind seit HURWITZ [14], [15] im Zusammenhang mit dem Problem der Komposition der quadratischen Formen und seit der Aufstellung

---

[2]) Im folgenden wird stets $a := b$ geschrieben, wenn a durch b definiert ist.

[3]) Liegt zwei Algebren $\overline{\mathscr{A}}$ und $\hat{\mathscr{A}}$ derselbe Vektorraum X zugrunde und ist $L(x)$ die linksreguläre Darstellung des Elements x von $\overline{\mathscr{A}}$, $\hat{L}(x)$ die des Elements x von $\hat{\mathscr{A}}$, so heisst nach ALBERT [1], S. 696, $\overline{\mathscr{A}}$ *isotop zu* $\hat{\mathscr{A}}$, wenn nichtsinguläre lineare Transformationen P, Q, R existieren derart, dass $\hat{L}(x) = PL(Qx)R$ gilt. Man überlegt sich leicht, dass die Relation „Isotopie" eine Äquivalenzrelation ist und somit eine Klasseneinteilung der Algebren gleicher Dimension über demselben Körper K ermöglicht.

der Diracschen Theorie des Elektrons im Zusammenhang mit dieser Theorie immer wieder untersucht worden, allerdings nicht – wie in der vorliegenden Arbeit – über einem beliebigen Körper der Charakteristik ungleich 2, sondern stets nur über dem reellen oder komplexen Zahlkörper.

Von den bisher noch nicht erwähnten Abschnitten II und III befasst sich der eine mit Eigenschaften der Lösungen von (1), (2), (3) und der durch diese Lösungen definierten Algebren, der andere mit Beziehungen zwischen diesen Algebren und den Kompositionsalgebren.

2. Betrachtet man das Gleichungssystem (1), (2) und (3) unter der Voraussetzung endlicher Dimension n des Vektorraumes X für eine feste Basis in X, so kann man L und M in (1), (2), (3) als (n, n)-Matrizen auffassen, die lineare Funktionen der Komponenten $\xi_1, \ldots, \xi_n$ des Vektors

$$x = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix} \text{ aus } X \text{ sind: } \quad L(x) = \left(\sum_{i=1}^{n} \alpha_{ijk} \xi_i \right), \quad M(x) = \left(\sum_{i=1}^{n} \beta_{ijk} \xi_i \right) \tag{6}$$

$(\alpha_{ijk}, \beta_{ijk} \in K;\ i, j, k = 1, \ldots, n)$. Wählt man für K speziell den reellen Zahlkörper $\mathbf{R}$ und schreibt man

$$\frac{\partial}{\partial x} \quad \text{für} \quad \begin{pmatrix} \partial/\partial\xi_1 \\ \vdots \\ \partial/\partial\xi_n \end{pmatrix} \quad \text{und } v \text{ für den Vektor} \quad \begin{pmatrix} \varphi_1 \\ \vdots \\ \varphi_n \end{pmatrix},$$

dessen Komponenten $\varphi_1, \ldots, \varphi_n$ zweimal stetig partiell differenzierbare reellwertige Funktionen der n reellen Veränderlichen $\xi_1, \ldots, \xi_n$ seien, so ist

$$L\left(\frac{\partial}{\partial x}\right) v = 0 \tag{7}$$

ein System von n partiellen Differentialgleichungen erster Ordnung, dem eine einzelne (nicht notwendig elliptische) partielle Differentialgleichung zweiter Ordnung, nämlich

$$\mu\left(\frac{\partial}{\partial x}\right) v = M\left(\frac{\partial}{\partial x}\right) L\left(\frac{\partial}{\partial x}\right) v = 0, \tag{8}$$

in analoger Weise zugeordnet ist, wie die klassischen Cauchy-Riemannschen Differentialgleichungen mit der Laplaceschen Gleichung in zwei Veränderlichen verknüpft sind. Das in einem Teil der vorliegenden Arbeit behandelte Problem der Existenz von Systemen der Art (7) mit der Eigenschaft (8) bei nichtausgearteter quadratischer Form μ – sie seien im folgenden *verallgemeinerte Cauchy-Riemannsche Differentialgleichungssysteme*, kurz: *VCR-Systeme* genannt – ist nicht nur im Zusammenhang mit Fragen der Verallgemeinerung der Funktionentheorie von Interesse. Aus dem folgenden Grunde hat es auch Bedeutung für die theoretische Physik: Es schliesst das

Problem ein, ob neben den bekannten Fällen

α) Laplacesche Gleichung $$\frac{\partial^2\varphi}{\partial\xi_1^2}+\frac{\partial^2\varphi}{\partial\xi_2^2}=0$$

β) Wellengleichung $$\frac{\partial^2\varphi}{\partial\xi_1^2}-\frac{\partial^2\varphi}{\partial\xi_2^2}-\frac{\partial^2\varphi}{\partial\xi_3^2}-\frac{\partial^2\varphi}{\partial\xi_4^2}=0$$

weitere (in der Differentiationsordnung homogene) partielle Differentialgleichungen zweiter Ordnung mit konstanten Koeffizienten $\alpha, \beta, \ldots, \omega$

$$\alpha\frac{\partial^2\varphi}{\partial\xi_1^2}+\beta\frac{\partial^2\varphi}{\partial\xi_1\partial\xi_2}+\cdots+\omega\frac{\partial^2\varphi}{\partial\xi_n^2}=0,\quad \text{kurz: } \mu\left(\frac{\partial}{\partial x}\right)\varphi=0\quad (\mu \text{ nichtausgeartet}),$$

existieren, die ebenfalls „*linearisiert*“ [4]) werden können, das heisst in ein System von n Differentialgleichungen erster Ordnung der Art (7) mit der Eigenschaft (8) verwandelbar sind.

Für den Fall der Laplaceschen Gleichung in n Veränderlichen wurde dieses Problem von Olga Taussky-Todd [27] und von Stiefel [26] untersucht. Mit Zusammenhängen zwischen den Lösungen von Systemen der Art (7) und den Lösungen einzelner partieller Differentialgleichungen von im allgemeinen höherer als zweiter Ordnung befassen sich Teile der Arbeiten [8] und [9] des Verfassers.

Die vorliegende Arbeit hatte zunächst nur das Ziel, die genannten Untersuchungen von Olga Taussky-Todd und Stiefel in dem unter 2 beschriebenen Rahmen weiterzuführen. Die Anregung, die jetzt vorliegende allgemeine Problemstellung zu behandeln, verdanke ich Herrn Professor M. Koecher. Sie hat zur Folge, dass Ergebnisse über VCR-Systeme als Nebenprodukte von Resultaten über das Funktionalgleichungssystem (1), (2), (3) abfallen, so zum Beispiel eine Klassifizierung der elliptischen VCR-Systeme (siehe den Zusatz 2 zu Satz 3) oder die Aussage, dass VCR-Systeme – gleich ob elliptisch oder nicht – nur für $n=2, 4, 8$ existieren; siehe den Zusatz zu Satz 5. Auf das Problem einer Klassifizierung *aller* VCR-Systeme wird in einer anderen Arbeit eingegangen werden.

I. Die Fragestellung der vorliegenden Arbeit als Verallgemeinerung des klassischen Problems der Komposition der quadratischen Formen in n Veränderlichen

Es sei X wieder ein Vektorraum über einem Körper K der Charakteristik ungleich 2. Man sagt von einer nichtausgearteten quadratischen Form $\mu: X\to K$, sie *gestatte*

---

4) Ausdrucksweise im Lehrbuch [29], S. 1096, von Weizel. Ein im obigen Sinne der Wellengleichung β) zugeordnetes VCR-System wird in der theoretischen Physik mit Hilfe der sogenannten Paulischen Matrizen als einzelne Gleichung geschrieben und *Weylsche Gleichung des Neutrinos* genannt.

Komposition, wenn in X eine bilineare Multiplikation $(x, y) \rightarrow xy$ $(x, y, xy \in X)$ existiert derart, dass

$$\mu(xy) = \mu(x)\,\mu(y) \tag{9}$$

gilt. Bezeichnet man mit JACOBSON [16] jede Algebra mit Einselement e über einem Körper K der Charakteristik ungleich 2, in der eine Identität (9) mit nichtausgearteter quadratischer Form μ erfüllt ist, als *Kompositionsalgebra*, so gilt der folgende Satz (siehe etwa JACOBSON [16] oder SCHAFER [25], S. 73), der für den Fall K = reeller oder komplexer Zahlkörper, μ = Summe von reinen Quadraten, zuerst von HURWITZ [14], [15] bewiesen wurde:

Kompositionsalgebren $\mathscr{C}$ existieren nur in den Dimensionen $n = 1, 2, 4, 8$. In diesen sind der Reihe nach genau die folgenden Fälle möglich:

a) $\mathscr{C} = Ke$

b) *$\mathscr{C}$ ist Erweiterungskörper von K vom Grade 2 oder isomorph zu $K \oplus K$*

c) *$\mathscr{C}$ ist eine Quaternionenalgebra*[5]) *über K*

d) *$\mathscr{C}$ ist eine Cayley-Algebra*[5]) *über K.*

Es ist bemerkenswert, dass Kompositionsalgebren unendlicher Dimension nicht existieren. Es gibt aber auch keine unendlichdimensionale *Divisions*algebra *ohne Einselement* mit der Eigenschaft (9), μ nichtausgeartet, denn eine solche wäre zu einer unendlichdimensionalen Kompositionsalgebra isotop; vgl. KAPLANSKY [18], S. 957. Weiter ist bekannt – siehe KAPLANSKY [18], S. 957, oder JACOBSON [16], S. 56 und 57 – dass jede endlichdimensionale Algebra *ohne Einselement*, die Komposition gestattet, zu einer Kompositionsalgebra, also zu einer der Algebren a) bis d), isotop ist.

Im folgenden wird gezeigt: *Im Falle endlicher Dimension n des Vektorraumes X ist das Problem der Komposition der quadratischen Formen ein Spezialfall des in der Einleitung formulierten Problems der Lösung des Funktionalgleichungssystems* (1), (2), (3).

Beweis: Schreibt man – siehe (4) – $\mu(x, x)$ für $\mu(x)$ und – siehe (5) – $L(x)\,y$ für xy, so erhält die „Kompositionsgleichung" (9) die Gestalt

$$\mu(L(x)\,y, L(x)\,y) = \mu(x, x)\,\mu(y, y)$$

oder, wenn man mit $L^*(x)$ die zu $L(x)$ bezüglich der Bilinearform μ adjungierte lineare Transformation bezeichnet,

$$\mu(L^*(x)\,L(x)\,y, y) = \mu(x, x)\,\mu(y, y).$$

Diese Gleichung ist dann und nur dann erfüllt, wenn die Gleichung

$$L^*(x)\,L(x) = \mu(x, x)\,I \qquad (I \text{ die } (n, n)\text{-Einheitsmatrix})$$

---

[5]) Zur Definition siehe etwa BRAUN und KOECHER [4], S. 221.

gilt. Diese aber ist offensichtlich ein Spezialfall von (1), das heisst das Problem ihrer Lösung ein Spezialfall des im folgenden behandelten Problems der Lösung des Funktionalgleichungssystems (1), (2), (3).

II. Einige Eigenschaften der Lösungen des Gleichungssystems (1), (2), (3) und der durch diese Lösungen definierten Algebren

Durch eine leichte Rechnung bestätigt man

HILFSSATZ 1. *Ist* $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x)$ *eine Lösung des Funktionalgleichungssystems* (1), (2), (3), *sind* P, Q, R *beliebige nichtsinguläre lineare Transformationen und ist* $\lambda \neq 0$ *ein beliebiges Element aus* K, *so ist*

$$\hat{L}(x) := P\bar{L}(Qx)R, \quad \hat{M}(x) := \lambda R^{-1}\bar{M}(Qx)P^{-1}, \quad \hat{\mu}(x) := \lambda\bar{\mu}(Qx) \tag{10}$$

ebenfalls eine Lösung.

Bezeichnet man jede durch eine Lösung des Funktionalgleichungssystems (1), (2), (3) gemäss $xy := L(x)y$ definierte Algebra als *(*$ML = \mu I$*)-Algebra* und vergegenwärtigt man sich die in Fussnote 3 gegebene Definition des Begriffs Isotopie von Algebren, so besagt also Hilfssatz 1 nichts anderes als

HILFSSATZ 1'. *Jede zu einer (*$ML = \mu I$*)-Algebra isotope Algebra ist ebenfalls eine (*$ML = \mu I$*)-Algebra. Mit anderen Worten: Die Eigenschaft* $ML = \mu I$ *ist isotopie-invariant.*

Es sei darauf hingewiesen, dass dagegen die Eigenschaft einer Algebra, Komposition zu gestatten, nicht isotopie-invariant ist. Beispiel: Die zur Algebra **C** der komplexen Zahlen isotope Algebra mit der Basis e_1, e_2 und der Multiplikationstafel $e_1^2 = e_1$, $e_2^2 = -e_1$, $e_1e_2 = e_2e_1 = 2e_2$, das heisst mit der linksregulären Darstellung

$$L(\xi_1 e_1 + \xi_2 e_2) = \begin{pmatrix} \xi_1 & -\xi_2 \\ 2\xi_2 & 2\xi_1 \end{pmatrix}$$

ist zwar wie **C** eine kommutative ($ML = \mu I$)-Divisionsalgebra, erlaubt aber keine Komposition.

Im folgenden benutzen wir die von (4) herrührende Schreibweise $\mu(x, x)$ für $\mu(x)$. Weiter nennen wir aus naheliegenden Gründen zwei Lösungen $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x, x)$ und $\hat{L}(x)$, $\hat{M}(x)$, $\hat{\mu}(x, x)$ des Systems (1), (2), (3) *zueinander isotop*, wenn sie gemäss (10) ineinander transformierbar sind. Offenbar ist die Relation „Isotopie" eine Äquivalenzrelation.

HILFSSATZ 2. *Im Falle endlicher Dimension des Vektorraumes* X *ist das Problem der Lösung des Funktionalgleichungssystems* (1), (2), (3) *äquivalent zum Problem der*

Lösung der quadratischen Gleichung

$$L(x)^2 = 2\,\mu(x,a)\,L(x) - \mu(x,x)\,I \qquad \left\{\begin{array}{l} L\colon X \to \mathrm{Hom}(X,X), \text{ linear} \\ a \in X \text{ so, dass } \mu(a,a) \neq 0 \end{array}\right\} \tag{11}$$

in dem folgenden Sinne: Es sei $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x,x)$ *eine Lösung von* (1), (2), (3). *Dann existiert eine Lösung* $\hat{L}(x)$, $\hat{\mu}(x,x)$ *von* (11) *derart, dass*

$$\hat{L}(x), \quad \hat{M}(x) := 2\,\hat{\mu}(x,a)\,I - \hat{L}(x), \quad \hat{\mu}(x,x) \tag{12}$$

eine zu $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x,x)$ *isotope Lösung von* (1), (2), (3) *ist; umgekehrt gibt es zu jeder – auch unendlichdimensionalen – Lösung* $\hat{L}(x)$, $\hat{\mu}(x,x)$ *von* (11) *ein* $\hat{M}(x)$, *nämlich* $\hat{M}(x) := 2\hat{\mu}(x,a)\,I - \hat{L}(x)$, *derart, dass* $\hat{L}(x)$, $\hat{M}(x)$, $\hat{\mu}(x,x)$ *eine Lösung des Systems* (1), (2), (3) *ist.*

Beweis. Nach Voraussetzung gilt

$$\bar{M}(x)\,\bar{L}(x) = \bar{\mu}(x,x)\,I \quad \text{mit} \quad \bar{\mu}(a,a) \neq 0 \quad \text{für geeignetes } a \in X\,. \tag{13}$$

Dann ist $\bar{M}(a)/\bar{\mu}(a,a)$ gleich $\bar{L}(a)^{-1}$; im Falle unendlicher Dimension wäre das zwar linksinvers, aber nicht notwendig rechtsinvers zu $\bar{L}(a)$. Definiert man nun

$$\hat{L}(x) := \bar{L}(a)^{-1}\,\bar{L}(x), \qquad \hat{M}(x) := \frac{\bar{M}(x)\,\bar{L}(a)}{\bar{\mu}(a,a)}, \qquad \hat{\mu}(x,x) := \frac{\bar{\mu}(x,x)}{\bar{\mu}(a,a)},$$

so wird (13) zu

$$\hat{M}(x)\,\hat{L}(x) = \hat{\mu}(x,x)\,I \quad \text{mit} \quad \hat{L}(a) = \hat{M}(a) = I, \quad \hat{\mu}(a,a) = 1\,. \tag{14}$$

Offenbar sind die beiden Lösungen $\bar{L}$, $\bar{M}$, $\bar{\mu}$ und $\hat{L}$, $\hat{M}$, $\hat{\mu}$ von (1), (2), (3) zueinander isotop. Wir zeigen nun, dass (14)

$$\hat{M}(x) = 2\,\hat{\mu}(x,a)\,I - \hat{L}(x) \tag{15}$$

impliziert, woraus die erste Behauptung des Hilfssatzes folgt: Ersetzt man x in (14) durch $x+y$ und beachtet man, dass $\hat{L}(x)$, $\hat{M}(x)$, $\hat{\mu}(x,y)$ linear bzw. bilinear sind, so erhält man

$$\hat{M}(x)\,\hat{L}(y) + \hat{M}(y)\,\hat{L}(x) = 2\,\hat{\mu}(x,y)\,I$$

und hieraus für $y=a$ gerade (15). Die zweite Behauptung des Hilfssatzes bestätigt man durch eine leichte Rechnung.

In der Sprechweise der Theorie der Algebren kann man Hilfssatz 2 folgendermassen formulieren:

HILFSSATZ 2′. *Jede endlichdimensionale* $(ML=\mu I)$*-Algebra ist zu einer Algebra* $\mathscr{A}$ *isotop, deren linksreguläre Darstellung* $L(x)$ *für geeignetes* $a \in \mathscr{A}$ *einer quadratischen Gleichung der Art* (11) *genügt. Umgekehrt ist jede – auch unendlichdimensionale – Algebra mit einer solchen linksregulären Darstellung eine* $(ML=\mu I)$*-Algebra.*

Von nun an wird über den unseren Betrachtungen zugrunde liegenden Vektorraum X nicht mehr vorausgesetzt, dass er endlichdimensional ist. $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x,x)$ sei eine Lösung des Funktionalgleichungssystems (1), (2), (3) derart, dass die durch $\bar{L}(x)$ vermöge $xy:=\bar{L}(x)\,y$ definierte $(ML=\mu I)$-Algebra $\bar{\mathscr{A}}$ zu einer Algebra $\mathscr{A}$ mit Einselement e isotop ist. Für die linksreguläre Darstellung $L(x)$ von $\mathscr{A}$ folgt dann $L(e)=I$ und – weil $\mathscr{A}$ nach Hilfssatz 1′ ebenfalls eine $(ML=\mu I)$-Algebra ist – die Existenz eines $M(x)$ mit $M(e)=I$ derart, dass $M(x)\,L(x)=\mu(x,x)\,I$ gilt. Diese Gleichung führt wie am Ende des Beweises von Hilfssatz 2 auf $M(x)=2\mu(x,e)\,I-L(x)$ und damit auf die quadratische Gleichung

$$L(x)^2=2\mu(x,e)\,L(x)-\mu(x,x)\,I\,. \tag{16}$$

Diese besagt, dass für $x, y\in\mathscr{A}$ die Identität

$$x(xy)=2\mu(x,e)\,xy-\mu(x,x)\,y \tag{17}$$

gilt. Setzt man $y=e$, so wird hieraus

$$x^2=2\mu(x,e)\,x-\mu(x,x)\,e\,, \tag{18}$$

das heisst die Algebra $\mathscr{A}$ ist *quadratisch*. Multipliziert man diese Gleichung von rechts mit y, so stimmt ihre rechte Seite mit der von (17) überein. Es gilt also $x^2y=x(xy)$, das heisst die Algebra $\mathscr{A}$ ist *linksalternativ*. Nach ALBERT [2], S. 322 und 323, ist jede quadratische linksalternative Algebra auch *rechtsalternativ* $[yx^2=(yx)\,x]$ und damit *alternativ*. Jede alternative Algebra ist übrigens *flexibel*, das heisst es gilt $x(yx)=(xy)\,x$; zum Beweis ersetze man x durch $x+y$ in $x^2y=x(xy)$.

Ergänzend zu den obigen Betrachtungen überlegen wir uns noch folgendes:

1. Eine Algebra, die zu keiner Algebra mit Einselement e isotop ist, kann zu keiner quadratischen Algebra isotop sein; eine solche besitzt nämlich ex definitione ein Einselement.

2. Mit Hilfe der linksregulären Darstellung $L(x)$ des Elements x einer jeden alternativen quadratischen Algebra kann die folgende Lösung von (1), (2), (3) gewonnen werden:

$$L(x),\qquad M(x):=2\mu(x,e)\,I-L(x),\qquad \mu(x,x)\,. \tag{19}$$

Fasst man die obigen Ergebnisse zusammen, so gelten also die drei Behauptungen von

SATZ 1. *$\bar{L}$, $\bar{M}$, $\bar{\mu}$ sei eine Lösung des Funktionalgleichungssystems* (1), (2), (3), *$\bar{\mathscr{A}}$ sei die über $\bar{L}(x)$ vermöge $xy:=\bar{L}(x)\,y$ definierte Algebra.* 1. *Genau dann, wenn $\bar{\mathscr{A}}$ zu einer Algebra mit Einselement e isotop ist*[6]*, ist $\bar{\mathscr{A}}$ zu einer alternativen quadratischen*

[6]) Wie man dem Beweis von Satz 3 entnimmt, ist eine Algebra dann und nur dann zu einer Algebra mit e isotop, wenn sie zwei Elemente a, b besitzt derart, dass sowohl $L(a)$ als auch $R(b)$, die rechtsreguläre Darstellung des Elements b, nichtsingulär sind.

Algebra isotop, das heisst zu einer Algebra, in der die folgenden Identitäten gelten:

$$x^2 y = x(x y), \quad y x^2 = (y x) x, \quad x(y x) = (x y) x, \quad x^2 = 2\mu(x, e) x - \mu(x, x) e.$$

2. *Aus „$\bar{\mathscr{A}}$ isotop zu $\mathscr{A}$ mit e" folgt: „$\mathscr{A}$ alternativ und quadratisch".*

3. *Aus „$\mathscr{A}$ alternativ und quadratisch" folgt: Die linksreguläre Darstellung* $L(x)$ *des Elements x aus $\mathscr{A}$ gestattet die Konstruktion einer Lösung von* (1), (2), (3), *nämlich* (19).

Man hat also eine vollständige Übersicht über die Klassen isotoper Lösungen von (1), (2), (3) mit der im Satz genannten Eigenschaft, wenn man alle alternativen quadratischen Algebren der Charakteristik ungleich 2 kennt. Für den Fall *nichtausgearteter* quadratischer Form μ in (18) sind diese Algebren bekannt*): Es sind gerade die Kompositionalgebren, also die auf S. 291 aufgeführten Algebren a), b), c), d) in den Dimensionen 1, 2, 3, 4; vgl. etwa JACOBSON [16] oder SCHAFER [25], S. 73–75; BRAUN und KOECHER [4], S. 224, beschränken sich auf den Beweis dieser Behauptung unter der Voraussetzung endlicher Dimension.

III. Beziehungen zwischen den Kompositionsalgebren und den durch die Lösungen der Gleichung $ML = \mu I$ definierten Algebren

Wegen des eben erwähnten Ergebnisses über die nichtausgearteten alternativen quadratischen Algebren kann man die Behauptungen 2 und 3 von Satz 1 für den Fall nichtausgearteter quadratischer Form μ folgendermassen fassen:

SATZ 2. *Ist eine* $(ML = \mu I)$*-Algebra mit nichtausgearteter quadratischer Form* μ *zu einer Algebra $\mathscr{A}$ mit Einselement isotop, so ist $\mathscr{A}$ eine Kompositionsalgebra, also eine der auf S.* 291 *aufgeführten Algebren* a), b), c), d), *die nur in den Dimensionen* 1, 2, 4, 8 *existieren. Jede Kompositionsalgebra ist eine* $(ML = \mu I)$*-Algebra mit Einselement.*

Hieraus folgt eine

CHARAKTERISIERUNG DER KOMPOSITIONSALGEBREN: *Eine Algebra mit Einselement in einem Vektorraum X über einem Körper der Charakteristik ungleich* 2 *ist genau dann eine Kompositionsalgebra, wenn zu ihrer linksregulären Darstellung* $L(x)$ *eine weitere lineare Abbildung* $M: X \to Hom(X, X)$ *existiert derart, dass das Produkt* $M(x)\,L(x)$ *die identische Abbildung, multipliziert mit einer nichtausgearteten quadratischen Form, ergibt.*

Eine weitere Konsequenz von Satz 2 ist

SATZ 3. *Jede* $(ML = \mu I)$*-Divisionsalgebra ist zu einer Kompositionsalgebra isotop. Diese ist ebenfalls eine Divisionsalgebra.*

Die zweite Behauptung folgt aus der Tatsache, dass die Eigenschaft einer Algebra,

*) Zusatz bei der Korrektur: Auch über den ausgearteten Fall weiss man inzwischen Bescheid; vgl. die Dissertation (München 1968) von L. ZAGLER, einem Schüler von M. KOECHER.

Divisionsalgebra zu sein, isotopie-invariant ist. Zum Beweis der ersten braucht wegen Satz 2 nur gezeigt zu werden, dass jede $(ML=\mu I)$-Divisionsalgebra zu einer Algebra mit Einselement isotop ist und eine nichtausgeartete quadratische Form μ besitzt.

In der Tat: Jede Divisionsalgebra $\bar{\mathbf{D}}$ ist zu einer Divisionsalgebra mit Einselement isotop: Sei nämlich in $\bar{\mathbf{D}}$

$$x\,y = \bar{L}(x)\,y = \bar{R}(y)\,x \quad \begin{cases} \bar{L}(x),\ \bar{R}(y) \text{ die links- bzw. rechtsreguläre} \\ \text{Darstellung des Elements } x \text{ bzw. } y \text{ aus } \bar{\mathbf{D}} \end{cases}$$

und damit – wenn A und B zwei lineare Transformationen des Vektorraumes X bedeuten –

$$(A\,x)\,(B\,y) = \bar{L}(A\,x)\,B\,y = \bar{R}(B\,y)\,A\,x, \tag{20}$$

so definiere man eine zu $\bar{\mathbf{D}}$ isotope Algebra $\mathbf{D}$ durch die aus $\bar{L}(x)$ und $\bar{R}(x)$ vermöge

$$L(x) := \bar{L}(\bar{R}(b)^{-1}\,x)\,\bar{L}(a)^{-1} \qquad (a,\ b \neq 0 \text{ aus } \bar{\mathbf{D}}, \text{ fest})$$

hervorgehende linksreguläre Darstellung; dieser entspricht – siehe (20) – die rechtsreguläre Darstellung $R(x)=\bar{R}(\bar{L}(a)^{-1}x)\,\bar{R}(b)^{-1}$. In $\mathbf{D}$ ist $x=ab$ Einselement e, denn $L(ab)=R(ab)=I$. Für die Elemente x aus $\mathbf{D}$ gilt nach Satz 1 eine Identität der Art $x^2=2\mu(x,e)\,x-\mu(x,x)\,e$, das heisst $x[x-2\mu(x,e)\,e]=-\mu(x,x)\,e$. Der letzten Identität entnimmt man, dass die quadratische Form nichtausgeartet ist. Wäre sie nämlich ausgeartet, so existierte in $\mathbf{D}$ ein Element $c\neq\lambda e$ (λ beliebig aus K) derart, dass $\mu(c,c)=0$ gälte. Man hätte dann $c[c-2\mu(c,e)\,e]=0$, das heisst $\mathbf{D}$ enthielte Nullteiler im Widerspruch zur Voraussetzung, dass $\bar{\mathbf{D}}$ und damit $\mathbf{D}$ Divisionsalgebren sind. Damit ist Satz 3 vollständig bewiesen.

Da die einzigen reellen Kompositionsalgebren, die zugleich Divisionsalgebren sind, der reelle Zählkörper $\mathbf{R}$, der komplexe Zahlkörper $\mathbf{C}$, der Schiefkörper $\mathbf{Q}$ der Hamiltonschen Quaternionen und die nichtassoziative (aber alternative) Algebra $\mathbb{C}$ der klassischen „Cayleyschen Zahlen“ sind, gilt der folgende

Zusatz 1 zu Satz 3. *Jede reelle $(ML=\mu I)$-Divisionsalgebra ist zu einer der Algebren* $\mathbf{R}$, $\mathbf{C}$, $\mathbf{Q}$, $\mathbb{C}$ *isotop.*

Dieses Ergebnis kann man im Hinblick auf die in der Einleitung erwähnten VCR-Systeme auch so ausdrücken:

Zusatz 2 zu Satz 3. *Ist ein Differentialgleichungssystem der Art*

$$\left(\sum_{i=0}^{n-1} \alpha_{ijk}\frac{\partial}{\partial\xi_i}\right)\begin{pmatrix}\varphi_0\\ \vdots \\ \varphi_{n-1}\end{pmatrix} = 0, \quad \text{kurz: } \bar{L}\left(\frac{\partial}{\partial x}\right) v = 0 \quad (n \geqslant 2) \tag{21}$$

ein elliptisches VCR-System, das heisst gibt es einen weiteren Operator $\bar{M}(\partial/\partial x)$ von *der Art des Operators in* (21) *mit der Eigenschaft*

$$\bar{M}\left(\frac{\partial}{\partial x}\right)\bar{L}\left(\frac{\partial}{\partial x}\right)v=\bar{\mu}\left(\frac{\partial}{\partial x}\right)v=0 \quad (\bar{\mu}\ \textit{positiv oder negativ definit}),$$

so ist es stets möglich, nichtsinguläre (n, n)*-Matrizen* P, Q, R *zu finden derart, dass der durch Transformation mit diesen Matrizen aus* $\bar{L}$ *entstehende Operator*

$$L\left(\frac{\partial}{\partial x}\right):=P\bar{L}\left(Q\frac{\partial}{\partial x}\right)R \tag{22}$$

die linksreguläre Darstellung von

$$e_0\frac{\partial}{\partial\xi_0}+e_1\frac{\partial}{\partial\xi_1}+\cdots+e_{n-1}\frac{\partial}{\partial\xi_{n-1}} \qquad (n=2,4,8)$$

ist, wobei e_0 = *Einselement,* $e_1,\ldots,e_{n-1}$ *die übliche Basis genau einer der Algebren* **C**, **Q**, $\mathbb{C}$ *ist.*

Ein elliptisches VCR-System kann also im Sinne von (22) auf genau eines der folgenden VCR-Systeme transformiert werden:

1. Auf die klassischen Cauchy-Riemannschen Differentialgleichungen,
2. auf das den Fueterschen linksregulären Quaternionenfunktionen zugrunde liegende VCR-System (vgl. etwa Fueter [13], Moisil [21], Eichler [12]),
3. auf ein mit den Cayleyschen Zahlen zusammenhängendes VCR-System.

IV. Die Existenz von Lösungen der Gleichung $ML=\mu I$ in Abhängigkeit von Dim X und Rang μ

Im folgenden wird das Funktionalgleichungssystem (1), (2), (3) unter der Voraussetzung betrachtet, dass der zugrunde liegende Vektorraum X von endlicher Dimension n über dem Grundkörper K der Charakteristik ungleich 2 ist; von der quadratischen Form $\mu\not\equiv 0$ in (1) wird zunächst nicht angenommen, dass sie nichtausgeartet ist.

Es sei $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x)$ eine Lösung von (1), (2), (3). Die Komponenten des Vektors $x\in X$ bezüglich einer Basis von X werden im folgenden mit $\xi_0, \xi_1,\ldots,\xi_{n-1}$ bezeichnet. Nach einem bekannten Satz aus der Theorie der quadratischen Formen – siehe etwa van der Waerden [28], S. 167 und 168 – lässt sich $\bar{\mu}$ auf eine Summe von $r\leqslant n$ Quadraten transformieren; r heisst bekanntlich der *Rang* von $\bar{\mu}$. In der Sprechweise von S. 292 gibt es also eine zu $\bar{L}$, $\bar{M}$, $\bar{\mu}$ isotope Lösung der Art

$$L(x),\ M(x),\ \mu(x)=(\xi_0^2+\alpha_1\xi_1^2+\cdots+\alpha_{r-1}\xi_{r-1}^2) \qquad (\alpha_1,\ldots,\alpha_{r-1}\neq 0 \text{ aus } K);$$

$$M(x)\,L(x)=(\xi_0^2+\alpha_1\xi_1^2+\cdots+\alpha_{r-1}\xi_{r-1}^2)\,I \qquad \left\{\begin{matrix}\alpha_1,\ldots,\alpha_{r-1}\neq 0 \text{ aus } K\\ I \text{ die } (n,n)\text{-Einheitsmatrix}.\end{matrix}\right\} \tag{23}$$

Die (n, n)-Matrizen $L(x)$ und $M(x)$ hängen vom Vektor x linear ab, das heisst sie haben die Gestalt

$$L(x) = \sum_{\nu=0}^{n-1} A_\nu \xi_\nu, \quad M(x) = \sum_{\nu=0}^{n-1} B_\nu \xi_\nu \quad \left\{\begin{matrix} A_\nu, B_\nu \ (n, n)\text{-Matrizen} \\ \text{mit Elementen aus } K. \end{matrix}\right\} \tag{24}$$

Da im folgenden isotope Lösungen von (1), (2), (3) nicht als wesentlich verschieden angesehen werden, kann $A_0 = B_0 = I = (n, n)$-Einheitsmatrix vorausgesetzt werden. Dann folgt aber aus (23), (24), dass $B_\varrho = -A_\varrho$ $(\varrho = 1, \ldots, n-1)$ und also

$$A_\sigma^2 = -\alpha_\sigma I \qquad (\alpha_\sigma \neq 0 \text{ aus } K; \sigma = 1, \ldots, r-1) \tag{25}$$

$$A_\tau^2 = 0 \qquad (\tau = r, \ldots, n-1) \tag{26}$$

$$A_\lambda A_\mu + A_\mu A_\lambda = 0 \qquad (\lambda, \mu = 1, \ldots, n-1; \lambda \neq \mu) \tag{27}$$

sein muss. Matrizen A_λ, A_μ mit der Eigenschaft (27) nennt man *antikommutativ.*

Mengen antikommutativer (n, n)-Matrizen mit gewissen zusätzlichen Eigenschaften spielen seit HURWITZ [14], [15] in zahlreichen Publikationen eine wichtige Rolle, beispielsweise (in zeitlicher Reihenfolge) in Arbeiten von RADON [24], EDDINGTON [6], [7] (Fall $n=4$), NEWMAN [22], LITTLEWOOD [20], JORDAN–V. NEUMANN–WIGNER [17], BRAUER–WEYL [3], ECKMANN [5], STIEFEL [26], KESTELMAN [19], PUTTER [23]. Alle diese Arbeiten beschränken sich auf den reellen oder komplexen Fall, so dass der folgende Hilfssatz bewiesen werden muss:

HILFSSATZ 3. *Es sei K ein beliebiger Körper der Charakteristik ungleich* 2 *und n eine beliebige natürliche Zahl, die in der Gestalt $n = p2^q$ (p ungerade, $q = 0, 1, 2, \ldots$) dargestellt sei. Eine Menge antikommutativer (n, n)-Matrizen über K mit der Eigenschaft* (25) *enthält niemals mehr als $2q+1$ Matrizen. Es gibt Körper, in denen $2q+1$ solche Matrizen existieren.*[7])

Ist dieser Hilfssatz bewiesen, so sieht man unmittelbar ein:

SATZ 4. *Es sei $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x)$ eine Lösung des Funktionalgleichungssystems* (1), (2), (3) *für den Fall beliebiger endlicher Dimension $n = p2^q$ (p ungerade, $q = 0, 1, 2, \ldots$) von X. Der Rang der quadratischen Form $\bar{\mu}(x)$ sei r. Dann gilt $r \leqslant 2q+2$. Mit anderen Worten: Ist* $\operatorname{Dim} X = n = p2^q$, *so gibt es keine Lösung von* (1), (2), (3) *derart, dass der Rang von μ grösser als $2q+2$ ist. Die Schranke $2q+2$ für r ist minimal im folgenden Sinn: Für spezielle Grundkörper K ist $r = 2q+2$ möglich.*

Für nichtausgeartetes μ, das heisst für $r = n$, erhält man hieraus

SATZ 5. *Endlichdimensionale Lösungen des Funktionalgleichungssystems* (1), (2),

---

[7]) Seit HURWITZ [14], [15] ist bekannt, dass im Falle K = komplexer Zahlkörper stets $2q+1$, aber nie $2q+2$ antikommutative (n, n)-Matrizen mit der Eigenschaft (25), $\alpha_\sigma = \pm 1$, existieren.

(3) *mit nichtausgeartetem* μ *existieren dann und nur dann, wenn der Vektorraum* X *die Dimension* $1, 2, 4$ *oder* 8 *hat.*

Beispiele von Lösungen des Gleichungssystems (1), (2), (3) in den genannten Dimensionen kann man mit Hilfe der linksregulären Darstellungen der Kompositionsalgebren konstruieren. Weitere Lösungen in den Dimensionen 2 und 4 werden in Abschnitt V bestimmt.

ZUSATZ ZU SATZ 5. *Verallgemeinerte Cauchy-Riemannsche Differentialgleichungssysteme gibt es nur für* $n=2, 4$ *und* 8.

Beweis von Hilfssatz 3. Es sei K^* ein algebraischer Erweiterungskörper von K, der die Quadratwurzeln sowohl des Elements -1 als auch der in (25) auftretenden (von Null verschiedenen) Elemente $-\alpha_1, \ldots, -\alpha_{r-1}$ aus K enthält. Wir zeigen, dass die Maximalzahl antikommutativer $n=p2^q$-reihiger quadratischer Matrizen C_σ über K^* mit der Eigenschaft $C_\sigma^2=-I$ gleich $2q+1$ ist. Damit ist dann der Hilfssatz bereits bewiesen, denn existierten über K *mehr als* $2q+1$ antikommutative Matrizen A_σ mit (25), so wäre $C_\sigma := A_\sigma/\sqrt{\alpha_\sigma}$ eine Menge von *mehr als* $2q+1$ antikommutativen Matrizen C_σ über K^* mit $C_\sigma^2=-I$.

Sei also $C_1, \ldots, C_{m(n)}$ eine maximale Menge antikommutativer (n, n)-Matrizen über K^*, deren Quadrat jeweils gleich $-I$ ist. Dieselbe Eigenschaft hat dann auch die Menge $TC_1T^{-1}, \ldots, TC_{m(n)}T^{-1}$ für jede nichtsinguläre (n, n)-Matrix T. Man kann also beispielsweise C_1 in der Normalform

$$C_1 = \begin{pmatrix} D_1 & & & 0 \\ & D_2 & & \\ & & \ddots & \\ 0 & & & D_k \end{pmatrix} \quad \text{mit „Begleitmatrices“} \quad D_\kappa = \begin{pmatrix} 0 & 0 & \ldots & 0 & \beta_{\kappa 1} \\ 1 & 0 & \ldots & 0 & \beta_{\kappa 2} \\ 0 & 1 & & \vdots & \vdots \\ \vdots & & \ddots & 0 & \vdots \\ 0 & \ldots & 0 & 1 & \beta_{\kappa\lambda_\kappa} \end{pmatrix} \tag{28}$$

($\beta_{\kappa 1}, \ldots, \beta_{\kappa\lambda_\kappa} \in K$; $\kappa=1, \ldots, k$) annehmen; siehe etwa VAN DER WAERDEN [28], S. 161 und 162. Aus $C_1^2=-I$ folgt, dass die D_κ in (28) höchstens zweireihig sein können, also

$$D_\kappa = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \quad \text{oder} = (i) \text{ oder} = (-i);$$

$i := \sqrt{-1}$ liegt nach Voraussetzung in K^*. Nun ist aber

$$\begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & i \\ 1 & -i \end{pmatrix}^{-1} = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix},$$

das heisst C_1 ist ähnlich einer Diagonalmatrix mit ausschliesslich $\pm i$ in der Diagonalen.

Hieraus kann man bereits schliessen, dass $m(n)=1$ für *ungerades* n ist. Ist nämlich $m(n)>1$, so folgt aus der Antikommutativität der Matrizen C_σ, dass n gerade sein muss: Es gilt ja $C_2 C_1 C_2^{-1} = -C_1$, das heisst die Eigenwerte $\pm i$ von C_1 gehen bei Vorzeichenänderung in sich über.

Sei nun *n gerade*. Dann kann also C_1 als Diagonalmatrix mit ausschliesslich $\pm i$ in der Diagonalen angenommen werden, ja sogar mit gleich vielen $+i$ wie $-i$; diese kann man durch eine Ähnlichkeitstransformation ordnen, das heisst C_1 darf in der Gestalt

$$C_1 = \begin{pmatrix} i\mathbf{1} & 0 \\ 0 & -i\mathbf{1} \end{pmatrix} \quad \left[\mathbf{1} \text{ die } \left(\frac{n}{2}, \frac{n}{2}\right)\text{-Einheitsmatrix}\right]$$

vorausgesetzt werden. Nun kann man weiterschliessen, wie das NEWMAN [22] bei der Berechnung der Maximalzahl antikommutativer *komplexer* Matrizen C_σ mit $C_\sigma^2 = -I$ durchgeführt hat: Wegen

$$C_1 C_\alpha + C_\alpha C_1 = 0 \text{ und } C_\alpha^2 = -I \qquad [\alpha = 2, \ldots, m(n)]$$

besitzen die C_α die Form

$$C_\alpha = \begin{pmatrix} 0 & \mathfrak{C}_\alpha \\ -\mathfrak{C}_\alpha^{-1} & 0 \end{pmatrix} \quad \left[\mathfrak{C}_\alpha \text{ nichtsinguläre } \left(\frac{n}{2}, \frac{n}{2}\right)\text{-Matrizen}\right].$$

Aus $C_\alpha C_\beta + C_\beta C_\alpha = 0$ ergibt sich $\mathfrak{C}_\alpha \mathfrak{C}_\beta^{-1} + \mathfrak{C}_\beta \mathfrak{C}_\alpha^{-1} = 0$ $(\alpha, \beta \geqslant 2;\ \alpha \neq \beta)$. Hieraus folgt

$$(\mathfrak{C}_\gamma \mathfrak{C}_2^{-1})^2 = -\mathbf{1}, \quad (\mathfrak{C}_\gamma \mathfrak{C}_2^{-1})(\mathfrak{C}_\delta \mathfrak{C}_2^{-1}) + (\mathfrak{C}_\delta \mathfrak{C}_2^{-1})(\mathfrak{C}_\gamma \mathfrak{C}_2^{-1}) = 0 \quad (\gamma, \delta \geqslant 3, \gamma \neq \delta),$$

das heisst die $(n/2, n/2)$-Matrizen $\mathfrak{C}\gamma \mathfrak{C}_2^{-1}$ $[\gamma = 3, \ldots, m(n)]$ bilden wie die Matrizen $C_\sigma[\sigma = 1, \ldots, m(n)]$ eine Menge von antikommutativen Matrizen, deren Quadrat jeweils die negative Einheitsmatrix liefert. Bezeichnet man die Maximalzahl solcher $(n/2, n/2)$-Matrizen mit $m(n/2)$, so gilt also $m(n/2) \geqslant m(n) - 2$.

Sei nun $\mathfrak{E}_\delta$ $[\delta = 1, \ldots, m(n/2)]$ eine maximale Menge von solchen $(n/2, n/2)$-Matrizen. Dann ist

$$\begin{pmatrix} i\mathbf{1} & 0 \\ 0 & -i\mathbf{1} \end{pmatrix}, \quad \begin{pmatrix} 0 & -\mathbf{1} \\ \mathbf{1} & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & \mathfrak{E}_\delta \\ \mathfrak{E}_\delta & 0 \end{pmatrix}$$

eine Menge von (n, n)-Matrizen, die ebenfalls die erwünschten Eigenschaften besitzen. Also gilt $m(n) \geqslant m(n/2) + 2$.

Zusammen mit der vorher erhaltenen Ungleichung ergibt sich $m(n) = m(n/2) + 2$. Beachtet man, dass, wie oben gezeigt wurde, $m(p) = 1$ für ungerades p gilt, so folgt hieraus $m(n) = m(p2^q) = 2q + 1$, was zu beweisen war.

V. Lösungen der Gleichung $ML = \mu I$ mit nichtausgeartetem μ, die nicht auf Kompositionsalgebren führen

SATZ 6. $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x)$ *sei eine Lösung des Funktionalgleichungssystems* (1), (2), (3), *die Dimension des zugrunde liegenden Vektorraumes X sei* 2, *die quadratische*

Form $\bar{\mu}$ sei nichtausgeartet. Dann lässt sich stets eine zu $\bar{L}(x)$, $\bar{M}(x)$, $\bar{\mu}(x)$ isotope Lösung

$$L(x) = \xi_0 I + \xi_1 A_1, \quad M(x) = \xi_0 I - \xi_1 A_1, \quad \mu(x) = \xi_0^2 + \alpha_1 \xi_1^2 \tag{29}$$

[I die (2, 2)-Einheitsmatrix, $\alpha_1 \neq 0$] angeben, die eine der beiden sich gegenseitig ausschliessenden Eigenschaften besitzt:

a) *$L(x)$ ist die linksreguläre Darstellung einer zweidimensionalen Kompositionsalgebra*

b) *A_1 in (29) ist die Einheitsmatrix I.*

Beweis. Nach dem vorigen Abschnitt ist eine Lösung von (1), (2), (3) mit den im Satz genannten Eigenschaften zu einer Lösung (29) mit $A_1^2 = -\alpha_1 I$ isotop. Enthält der zugrunde liegende Körper K kein Element, dessen Quadrat gleich $-\alpha_1$ ist, so kann A_1 durch eine Ähnlichkeitstransformation auf die Gestalt $\begin{pmatrix} 0 & -\alpha_1 \\ 1 & 0 \end{pmatrix}$ gebracht werden; siehe (28). Mit diesem A_1 ist $L(x)$ in (29) die linksreguläre Darstellung einer zweidimensionalen Kompositionsalgebra. Existiert dagegen eine Quadratwurzel aus $-\alpha_1$ in K, so kann – da zwischen isotopen Lösungen nicht unterschieden wird – $\mu(x)$ in (29) in der Gestalt $\mu(x) = \xi_0^2 - \xi_1^2$ geschrieben werden, und man hat $A_1^2 = I$. Eine (2, 2)-Matrix mit dieser Eigenschaft ist stets – siehe (28) – zu genau einer der folgenden ähnlich:

$$\left.\begin{array}{l} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} \\ \left[\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \text{ und } \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \text{ sind ähnlich zu } \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\right]. \end{array}\right\} \tag{30}$$

Mit der ersten dieser Matrizen für A_1 ist $L(x)$ in (29) wieder die linksreguläre Darstellung einer zweidimensionalen Kompositionsalgebra, mit den beiden übrigen Matrizen erhält man

$$L(x) = \begin{pmatrix} \xi_0 + \xi_1 & 0 \\ 0 & \xi_0 + \xi_1 \end{pmatrix} \quad \text{bzw.} \quad L(x) = \begin{pmatrix} \xi_0 - \xi_1 & 0 \\ 0 & \xi_0 - \xi_1 \end{pmatrix}, \tag{31}$$

das heisst die linksregulären Darstellungen der Elemente x zweier isotoper Algebren, deren rechtsreguläre Darstellungen

$$R(x) = \begin{pmatrix} \xi_0 & \xi_0 \\ \xi_1 & \xi_1 \end{pmatrix} \quad \text{bzw.} \quad R(x) = \begin{pmatrix} \xi_0 & -\xi_0 \\ \xi_1 & -\xi_1 \end{pmatrix}$$

singulär sind; nach Fussnote 6 auf S. 294 sind diese Algebren zu keiner Algebra mit Einselement und damit zu keiner Kompositionsalgebra isotop.

Wie man aus (31) abliest, kann Satz 6 auch folgendermassen formuliert werden:

Satz 6′. *Ist eine zweidimensionale ($ML = \mu I$)-Algebra mit nichtausgeartetem μ zu keiner Kompositionsalgebra isotop, so ist sie zu der durch die Basis e_0, e_1 mit $e_0^2 = e_0$,*

$e_0 e_1 = e_1$, $e_1 e_0 = e_0$, $e_1^2 = e_1$ definierten assoziativen, aber nichtkommutativen Algebra isotop.

Auch in der Dimension 4 existieren Lösungen $L(x)$, $M(x)$, $\mu(x)$ von (1), (2), (3) mit nichtausgeartetem μ derart, dass $L(x)$ *zu keiner linksregulären Darstellung des Elements x einer Kompositionsalgebra isotop ist*; das sei anhand des folgenden Beispiels gezeigt, in dem $\alpha_1 \neq 0$, $\alpha_2 \neq 0$, β, γ und δ beliebige Elemente aus dem Grundkörper K mit $-\beta^2 - \alpha_1 \gamma^2 - \alpha_2 \delta^2 =: \alpha_3 \neq 0$ sind:

$$\left.\begin{aligned} L(x) &= \begin{pmatrix} \xi_0 + \beta\xi_3 & -\alpha_1\xi_1 + \alpha_1\gamma\xi_3 & -\alpha_2\xi_2 + \alpha_2\delta\xi_3 & 0 \\ \xi_1 + \gamma\xi_3 & \xi_0 - \beta\xi_3 & 0 & \alpha_2\xi_2 - \alpha_2\delta\xi_3 \\ \xi_2 + \delta\xi_3 & 0 & \xi_0 - \beta\xi_3 & -\alpha_1\xi_1 + \alpha_1\gamma\xi_3 \\ 0 & -\xi_2 - \delta\xi_3 & \xi_1 + \gamma\xi_3 & \xi_0 + \beta\xi_3 \end{pmatrix}, \\ M(x) &= 2\xi_0 I - L(x), \qquad \mu(x) = \xi_0^2 + \alpha_1\xi_1^2 + \alpha_2\xi_2^2 + \alpha_3\xi_3^2. \end{aligned}\right\} \tag{32}$$

Zunächst bestätigt man durch Rechnung, dass $M(x)\,L(x) = \mu(x)\,I$ gilt. Ist (32) die linksreguläre Darstellung des Elements $x = \xi_0 e_0 + \xi_1 e_1 + \xi_2 e_2 + \xi_3 e_3$ einer Algebra, so erhält man für die Basiselemente e_0, e_1, e_2, e_3 die folgende Multiplikationstafel, die anfangs wie die der üblichen Basis einer Quaternionenalgebra aussieht:

$$\begin{aligned} &e_0 e_i = e_i \quad (i = 0, 1, 2, 3), && e_1 e_2 = -e_2 e_1 = e_3, \quad e_1^2 = -\alpha_1 e_0, \quad e_2^2 = -\alpha_2 e_0, \\ &e_j e_0 = e_j \quad (j = 0, 1, 2), && e_1 e_3 = -\alpha_1 e_2, \quad e_2 e_3 = \alpha_2 e_1, \\ &e_3 e_0 = \beta e_0 + \gamma e_1 + \delta e_2, && e_3 e_1 = \alpha_1 \gamma e_0 - \beta e_1 - \delta e_3, \\ &e_3 e_2 = \alpha_2 \delta e_0 - \beta e_2 + \gamma e_3, && e_3^2 = -\alpha_2 \delta e_1 + \alpha_1 \gamma e_2 + \beta e_3. \end{aligned}$$

Da nicht gleichzeitig $e_3^2 e_3 = e_3 e_3^2$ und $(e_1 + e_3)^2\,(e_1 + e_3) = (e_1 + e_3)\,(e_1 + e_3)^2$ sein kann, handelt es sich hier um eine $(ML = \mu I)$-Algebra, die *nicht einmal potenzassoziativ* ist. Zum Beweis, dass sie zu keiner Kompositionsalgebra isotop ist, berechne man die rechtsreguläre Darstellung des Elements x:

$$R(x) = \begin{pmatrix} \xi_0 & -\alpha_1\xi_1 & -\alpha_2\xi_2 & \beta\xi_0 + \alpha_1\gamma\xi_1 + \alpha_2\delta\xi_2 \\ \xi_1 & \xi_0 & \alpha_2\xi_3 & \gamma\xi_0 - \beta\xi_1 - \alpha_2\delta\xi_3 \\ \xi_2 & -\alpha_1\xi_3 & \xi_0 & \delta\xi_0 - \beta\xi_2 + \alpha_1\gamma\xi_3 \\ \xi_3 & \xi_2 & -\xi_1 & -\delta\xi_1 + \gamma\xi_2 + \beta\xi_3 \end{pmatrix}.$$

Man überzeugt sich leicht, dass deren Determinante identisch in $\xi_0, \xi_1, \xi_2, \xi_3, \alpha_1, \alpha_2, \beta, \gamma, \delta$ verschwindet, das heisst nach Fussnote 6 auf S. 294, dass die über (32) definierte Algebra zu keiner Algebra mit Einselement isotop ist. Also kann sie nach den Sätzen 1 und 2 zu keiner Kompositionsalgebra isotop sein.

LITERATURVERZEICHNIS

[1] ALBERT, A. A., *Non-Associative Algebras*, I: *Fundamental Concepts and Isotopy*, Ann. of Math. *43*, 685–707 (1942).

[2] ALBERT, A. A., *On Right Alternative Algebras*, Ann. of Math. *50*, 318–328 (1949).
[3] BRAUER, R. and WEYL, H., *Spinors in n Dimensions*, Amer. J. Math. *57*, 425–449 (1953).
[4] BRAUN, H. und KOECHER, M., *Jordan-Algebren* (Springer-Verlag, Berlin–Heidelberg–New York 1966 [Grundlehren der Mathematischen Wissenschaften, Bd. 128]).
[5] ECKMANN, B., *Gruppentheoretischer Beweis des Satzes von Hurwitz-Radon über die Komposition der quadratischen Formen*, Comment. Math. Helv. *15*, 358–366 (1942/43).
[6] EDDINGTON, A. S., *On Sets of Anticommuting Matrices*, J. London Math. Soc. *7*, 58–68 (1932).
[7] EDDINGTON, A. S., *On Sets of Anticommuting Matrices*, Part II: *The Factorization of E-Numbers*, J. London Math. Soc. *8*, 142–152 (1933).
[8] EICHHORN, W., *Funktionentheorie in Algebren über dem reellen Zahlkörper und ihre Anwendung auf partielle Differentialgleichungen* (Würzburg 1961).
[9] EICHHORN, W., *Verallgemeinerte Cauchy-Riemannsche und Laplacesche Differentialgleichungen*, Math. Z. *84*, 279–293 (1964).
[10] EICHHORN, W., *Über die multiplikative Cauchysche Funktionalgleichung für die Hamiltonschen Quaternionen und die Cayleyschen Zahlen*, J. Reine Angew. Math. *221*, 2–13 (1966).
[11] EICHHORN, W., *Über die multiplikativen Abbildungen endlichdimensionaler Algebren in kommutative Halbgruppen*, J. Reine Angew. Math. 231, 10–46 (1968).
[12] EICHLER, M., *Allgemeine Integration einiger partieller Differentialgleichungen der mathematischen Physik durch Quarternionenfunktionen*, Comment. Math. Helv. *12*, 212–224 (1939/40).
[13] FUETER, R., *Die Funktionentheorie der Differentialgleichungen $\Delta u = 0$ und $\Delta\Delta u = 0$ mit vier reellen Variablen*, Comment. Math. Helv. 7, 307–330 (1934/35).
[14] HURWITZ, A., *Über die Composition der quadratischen Formen von beliebig vielen Variablen*, Nachr. Ges. Wiss. Göttingen Math.-Phys. Kl. *1898*, 309–316.
[15] HURWITZ, A., *Über die Komposition der quadratischen Formen*, Math. Ann. *88*, 1–25 (1923).
[16] JACOBSON, N., *Composition Algebras and Their Automorphisms*, Rend. Circ. Mat. Palermo *7*, 55–80 (1958).
[17] JORDAN, P., VON NEUMANN, J., and WIGNER, E., *On an Algebraic Generalization of the Quantum Mechanical Formalism*, Ann. of Math. *35*, 29–64 (1934).
[18] KAPLANSKY, J., *Infinite-Dimensional Quadratic Forms Admitting Composition*, Proc. Amer. Math. Soc. *4*, 956–960 (1953).
[19] KESTELMAN, H., *Anticommuting Linear Transformations*, Canad. J. Math. *13*, 614–624 (1961).
[20] LITTLEWOOD, D. E., *Note on the Anticommuting Matrices of Eddington*, J. London Math. Soc. *9*, 41–50 (1934).
[21] MOISIL, G. C., *Sur les quaternions monogenes*, Bull. Sci. Math. *55*, 168–174 (1931).
[22] NEWMAN, M. H. A., *Note on an Algebraic Theorem of Eddington*, J. London Math. Soc. *7*, 93–99 (1932).
[23] PUTTER, J., *Maximal Sets of Anti-Commuting Skew-Symmetric Matrices*, J. London Math. Soc. *42*, 303–308 (1967).
[24] RADON, J., *Lineare Scharen orthogonaler Matrizen*, Abh. Math. Sem. Univ. Hamburg *1*, 1–14 (1922).
[25] SCHAFER, R. D., *An Introduction to Nonassociative Algebras* (Academic Press, New York–London 1966 [Pure and Applied Math. Vol. 22]).
[26] STIEFEL, E., *On Cauchy-Riemann Equations in Higher Dimensions*, J. Res. Nat. Bur. Standards [Sect. B] *48*, 395–398 (1952).
[27] TAUSSKY, O., *An Algebraic Property of Laplace's Differential Equation*, Quart. J. Math. Oxford Ser. *10*, 99–103 (1936).
[28] VAN DER WAERDEN, B. L., *Algebra*, 2. *Teil*, 4. Aufl. (Springer Verlag, Berlin–Göttingen–Heidelberg 1959).
[29] WEIZEL, W., *Lehrbuch der theoretischen Physik*, Band 2: *Struktur der Materie*, 2. verbesserte Aufl. (Springer Verlag, Berlin–Göttingen–Heidelberg 1958).

Universität Würzburg und University of Waterloo

Additive Inhaltsmasse im positiv gekrümmten Raum

W. MAIER und A. EFFENBERGER (Jena, East Germany)

Herrn Professor Dr. A.M. Ostrowski zum 75. Geburtstag gewidmet

Unter den Aufgaben, die im Rahmen der mehrdimensionalen Geometrie noch immer nicht in abschliessender Weise behandelt werden konnten, steht die Frage nach der Inhaltsmessung von Simplexen in Räumen fester Krümmung seit C. F. GAUSS und J. BOLYAI zur Diskussion. Durch mehrfache Lotkonstruktion entsteht im vieldimensionalen Raum positiver oder negativer Krümmung das durch L. SCHLÄFLI als Orthoschem bezeichnete besondere Simplex. Die Vorzugsstellung der Orthoscheme gegenüber der umfassenden Menge beliebiger Simplexe besteht nicht nur darin, dass die Anzahl der zur Konstruktion eines Orthoschems (OS) nötigen Bestimmungsstücke wesentlich kleiner ist als für beliebige Simplexe. Vielmehr bestehen zwischen den OS-Winkeln gruppentheoretische Zusammenhänge, wie sie Gauss am Pentagramma von NEPER genauer untersucht hat [5]. Andrerseits ist jedes OS im Fall negativer Krümmung ergänzungsgleich einer Linearkombination von asymptotischen Orthoschemen, und diese seit N. LOBATSCHEWSKI klassische Reduktionsmethode ist auf reelle Mannigfaltigkeiten positiver Krümmung nicht übertragbar. Deshalb erscheint es angebracht, solche Funktionalgleichungen ins Auge zu fassen, welche allgemeinen simplizialen Zerlegungen im Raum positiver Krümmung zugeordnet werden können. Berichten werden wir im anschliessenden Paragraphen eins über eine algebraische Methode, die man Herrn Dr. WEISSBACH verdankt, während in den folgenden Paragraphen geometrische Leitmotive von Herrn EFFENBERGER zu neuen Ergebnissen führen.

§ 1. Einteilung der Simplexe nach Typen

Der euklidische R_n sei mit kartesischen Koordinaten x_ν vermessen, so dass $\prod_{\nu=1}^{n} dx_\nu =: dX$ als Raumelement benutzt werden kann.

Die aus den $\{x_\nu\}$ gebildete einspaltige Matrix X erlaubt nach Multiplikation mit der gestürzten Matrix, die Kennzeichnung der $(n-1)$-dimensionalen Kugel

$$\bar{X}X = 1 \tag{1}$$

zu geben, auf welcher unsere Simplexe verwirklicht werden sollen. Der Übergang zu Polarkoordinaten erlaubt, den Durchschnitt G eines vom Nullpunkt ausstrahlenden Kegels K mit der Kugel (1) in der Gestalt

$$G = \frac{2^{1-n/2}}{\Gamma\left(\frac{n}{2}\right)} \int_K dX\, e^{-1/2\, \bar{X}X} \tag{2}$$

wiederzugeben [3].

Eingegangen am 16.12.1967

Mit

$$\det(c_i^k) \neq 0; \quad (c_i^n, \ldots, c_i^n) =: C_i, \quad C_i X =: p_i$$

können die n Ebenen $p_i = 0$ als Berandung von K gewählt werden; die aus (1) ausgeschnittenen Simplexe haben den Inhalt S_n, und aus (2) folgt dann

$$S_n = \frac{2^{1-n/2}}{\Gamma\binom{n}{2}} \int\limits_{0<p_i} dX\, e^{-1/2\, X X}\,. \tag{2'}$$

Die Anpassung des Integranden (2′) an die Integrationsgrenze $p_i = 0$ gelingt mit

$$\begin{pmatrix} p_1 \\ \vdots \\ p_n \end{pmatrix} =: P =: C X$$

in der Gestalt

$$S_n = \frac{2^{1-n/2}}{\Gamma\binom{n}{2}} \det(C^{-1}) \int\limits_{0<p_i} dP\, e^{-1/2\, P \bar{C}^{-1} C^{-1} P}\,. \tag{2''}$$

Die $\binom{n}{2}$ zwischen den Grenzebenen liegenden Keilwinkel (h, k) bestimmen sich aus

$$\cos(h, k) = \frac{C_h C_k}{\sqrt{C_h \bar{C}_k}\sqrt{C_k \bar{C}_h}}\,. \tag{3}$$

Durch Einführung der symmetrischen Matrix

$$\begin{pmatrix} 1 & -\cos(1,2) \ldots & -\cos(1,n) \\ -\cos(2,1) & 1 & \\ \vdots & \ddots & \\ -\cos(n,1) & & 1 \end{pmatrix} =: W \tag{3'}$$

aus den positiven Zahlen $C_i \bar{C}_i =: t_i$ als Elemente einer Diagonalmatrix T kann das System der Aussagen (3) zusammengefasst werden zur Matrizengleichung

$$W = T^{-1/2} C \bar{C} T^{1/2}\,. \tag{3''}$$

Der Fall eines regulären Simplexes mit Keilwinkeln $(h, k) =: \alpha$; $h, k = 1, \ldots, n$ wurde schon von H. Ruben [7] erledigt. Mit der Hilfsgrösse

$$\frac{1 + (1-n)\cos\alpha}{1 + \cos\alpha} =: h$$

vereinfacht sich (2″) zu

$$S_n = \frac{2^{1-n/2}}{\sqrt{h}\binom{n}{2}} \int\limits_{0<p_i} dP\, e^{-1/2\left[\sum_{\nu=1}^{n} p_\nu^2 - (h-1)/n h \left(\sum_{\nu=1}^{n} p_\nu\right)^2\right]}\,. \tag{4}$$

Das in der Fehlertheorie gebrauchte uneigentliche Integral

$$(2\pi)^{-1/2} \int_{-\infty}^{x} dt\, e^{-1/2\, t^2} = : f(x) \tag{4'}$$

erlaubt die Zurückführung von (4) auf eine einfache Integration

$$S_n = \frac{2^{1-n/2}}{\Gamma\left(\frac{n}{2}\right)} \cdot (2\pi)^{(n-1)/2} \sqrt{\frac{n}{h-1}} \int_{-\infty}^{+\infty} dx\, e^{(-x^2)/2} \left[f\left(x \sqrt{\frac{h-1}{n}} \right) \right]^n . \tag{4''}$$

Durch Orthogonaltransformation der quadratischen Form im Exponenten des Integranden (2″) gelangt Herr WEISSBACH zu einen Eigenwertproblem, dessen Lösung die Menge aller Simplexinhalte aufgliedert nach Typen verschiedener Schwierigkeit. Bezeichnet man die speziell von Herrn RUBEN bestimmten Simplexe als solche vom Typ 0, so zeigt die Betrachtung der Vielfachheit von Eigenwerten einer eingehenden Matrix bei kleinen Typenzahlen die Möglichkeit zur Reduktion des n-fachen Integrals (2″) entsprechend (4′, 4″); insbesondere erweist sich die Menge der Simplexe vom Typ 1 nicht abhängig von $\binom{n}{2}$ Parametern, sondern mit $3<n$ nur von $n<\binom{n}{2}$ geeignet gewählten Stücken. Vorbereitend wähle man eine n-reihige Orthogonalmatrix Y sowie eine positive Diagonalmatrix U, so dass $C=YU^{1/2}$ zur Umformung von (2′) in die Gestalt

$$S_n = \frac{2^{1-n/2}}{\Gamma\left(\frac{n}{2}\right)} \det(U^{-1}) \int_{0<p_i} dP\, e^{-1/2\, \bar{P} Y U^{-1} \bar{Y} P}$$

führt.

Mit der weiteren Festsetzung $Y(U-E)^{1/2}=:A$ und $(E+\bar{A}A)^{-1/2}\,\bar{A}P=:P^*$ erhalten wir

$$S_n = \frac{2^{1-n/2}}{\Gamma\left(\frac{n}{2}\right)} \det\{(E+\bar{A}A)^{-1/2}\} \int_{0<p_i} dP\, e^{-1/2\, (\bar{P}P - \bar{P}^* P^*)} . \tag{5}$$

Diejenigen Eigenwerte von $T^{1/2}\, W T^{1/2}$, welche den Wert Eins annehmen, bewirken eine Rangerniedrigung der Diagonalmatrix $U-E$; existieren nun $(n-m)$ solche Eigenwerte, so kann die quadratische Form $\bar{P}^* P^*$ reduziert werden auf die Summe von nur $m<n$ quadrierten linearen Formen. Dementsprechend erkläre man durch

$$\min_T \{\mathrm{Rg}(W - T^{-1})\} = : m \tag{6}$$

den Typ des betrachteten Simplexes; bleibt $m<n-1$, so vereinfacht sich auch der Integrationsprozess (5). Als nächst einfacher Fall neben den von Herrn RUBEN bearbeiteten regulären Simplexen erscheint so die Menge der Simplexe vom Typ $m=1$. Mit $g, h, k, l=1,\dots,n$, $g\neq h\neq k\neq l$ wird in diesem Fall

$$\begin{vmatrix}\cos(g,h) & \cos(g,l)\\ \cos(k,h) & \cos(k,l)\end{vmatrix}=0. \tag{6'}$$

Zur Lösung des Systems (6') führt eine Parameterfolge $\{a_h\}$, deren Elemente sich als reelle oder reinimaginäre Zahlen herausstellen, die im letzteren Fall einzuschränken sind durch die Forderung $-1<\sum_{v=1}^{n} a_v^2$. Für die gesuchten Keilwinkel gilt dann

$$-\cos(h,k)=\frac{a_h}{\sqrt{1+a_h^2}}\cdot\frac{a_k}{\sqrt{1+a_k^2}}.$$

Als Bezeichnung wählen wir deshalb für die Inhalte der Simplexe vom Typ 1

$$S_n(W)=:S_n^1(a_1,\dots,a_n).$$

Unter den von Herrn WEISSBACH gefundenen Funktionalgleichungen dieser besonderen Simplexklasse erwähnen wir als durchsichtiges Beispiel

$$S_n^1(a_1,\dots,a_n)=\sum_{v=1}^{n} S_n^1\left(\frac{a_1}{a_v},\dots,\frac{a_{v-1}}{a_v},\frac{-1}{a_v},\frac{a_{v-1}}{a_v},\dots,\frac{a_n}{a_v}\right). \tag{6'''}$$

Diese Identität kann betrachtet werden als ein der Algebra entnommenes Gegenstück von Aussagen für sphärische Inhalte zu den der Geometrie entnommenen Darstellungen eines hyperbolischen Simplexes durch asymptotische Simplexe einfacherer Art, wie sie erstmals LOBATSCHEWSKI gegeben hat.

Gegenüber den Hilfsmitteln der linearen Algebra sind im folgenden solche der kombinatorischen Topologie zu benutzen.

§ 2. Innere Geometrie konvexer sphärischer Simplexe

Auf der Hypersphäre vom Radius 1 im euklidischen R_n, im folgenden kurz Sphäre genannt, bestimmen n nicht in einer Hyperebene durch den Sphärenmittelpunkt liegende Punkte mit Nummern $1,\dots,n$ ein $(n-1)$-dimensionales nichtentartetes konvexes sphärisches *Simplex*

$$S_n=(1,\dots,n). \tag{1}$$

Das *Inhaltsmass* eines Simplexes werde mit demselben Symbol (1) wie das Simplex selbst bezeichnet.

Eine Teilmenge $\{v_1,\dots,v_j\}\subseteq\{1,\dots,n\}$ der Eckpunktmenge bestimmt ein $(j-1)$-

dimensionales Simplex $S_j=(v_1,\ldots, v_j)$, welches wir *Untersimplex* $S_j \subseteq S_n$ nennen wollen.

Wegen der Isomorphie zwischen der Menge der Untersimplexe von S_n und der Potenzmenge der Menge $\{1,\ldots, n\}$ bilden die Untersimplexe eines Simplex S_n einen Booleschen Verband, genannt *Untersimplexverband* von S_n. Die Menge der Untersimplexe S_m von S_n mit

$$S_k \subseteq S_m \subseteq S_j \tag{2}$$

nennen wir ein *Intervall* $[S_k, S_j]$ des Untersimplexverbandes und die Zahl $(j-k)$ die *Dimension* dieses Intervalls.

Die zweidimensionalen Intervalle $[S_k, S_{k+2}]$ haben 4 Elemente, etwa

$$\left.\begin{matrix} (v_1,\ldots, v_k) \\ (v_1,\ldots, v_k, v_{k+1}) \quad (v_1,\ldots, v_k, v_{k+2}) \\ (v_1,\ldots, v_k, v_{k+1}, v_{k+2}). \end{matrix}\right\} \tag{3}$$

Wir bezeichnen den euklidischen Einbettungsraum niedrigster Dimension des Simplexes $S_j=(\mu_1,\ldots, \mu_j)$ mit $R_j=R(\mu_1,\ldots, \mu_j)$ und ordnen jedem zweidimensionalen Intervall (3) den *Winkel*

$$0<(k; v_1,\ldots, v_{k+2})<\pi \tag{4}$$

zu, welcher zwischen den Räumen $R(v_1,\ldots, v_k, v_{k+1})$ und $R(v_1,\ldots, v_k, v_{k+2})$ an ihrem gemeinsamen Unterraum $R(v_1,\ldots, v_k)$ gemessen wird. Die Zahl k heisse die *Ordnung* des Winkels in $[S_0, S_n]$, vgl. [2]. Unter dem System der Winkel eines Intervalls verstehen wir die Winkel aller zweidimensionalen Teilintervalle.

Wir können jedes Intervall $[S_h, S_j]=[(v_1\ldots v_h), (v_1\ldots v_h\ldots v_j)]$ isomorph mit Erhaltung der Winkel bzw. dual isomorph mit Übergang zu den Supplementwinkeln auf einen Untersimplexverband abbilden. Die Bilder der Untersimplexe bezeichnen wir durch

$$(h; v_1,\ldots, v_m) \quad \text{bzw.} \quad (j; v_1,\ldots, v_m) \quad \text{mit} \quad h\leqslant m\leqslant j \tag{5}$$

und nennen h die Ordnung und j die duale Ordnung des Intervalls $[S_h, S_j]$. Die Abbildung auf den dual isomorphen Untersimplexverband $[(j; v_1\ldots v_j), (j; v_1\ldots v_h)]$ erfolgt durch Polarabbildung im Einbettungsraum $R(v_1\ldots v_j)$ und die Abbildung auf den isomorphen Untersimplexverband $[(h; v_1\ldots v_h), (h; v_1\ldots v_j)]$ durch nachfolgende Polarenabbildung im Einbettungsraum $R(j; v_1\ldots v_h)$ von $(j; v_1\ldots v_h)$.

Wir ordnen jedem Intervall $[(v_1\ldots v_h), (v_1\ldots v_h\ldots v_j)]$ als *Volumen* den Inhalt des Simplexes $(h; v_1\ldots v_j)$ und als *duales* Volumen den Inhalt des Simplexes $(j; v_1\ldots v_h)$ zu.

Die Volumina der zweidimensionalen Intervalle sind gerade ihre Winkel, die dualen Volumina sind die Supplemente der Winkel, d.h.

$$(k+2; v_1\ldots v_k)=\pi-(k; v_1\ldots v_{k+2}). \tag{6}$$

Die Volumina der Intervalle der Ordnung 0 stimmen mit den Volumina ihrer maximalen Elemente überein, weshalb wir die Ordnung 0 im Symbol für die Intervallvolumina weglassen wollen:

$$(0;\, \nu_1 \ldots \nu_m) = (\nu_1 \ldots \nu_m) \tag{7}$$

Nach üblichen Bezeichnungen sind die Intervallvolumina die Volumina von Auffangsimplexen [2] bei nichtverschwindender und von Untersimplexen bei verschwindender Ordnung. Wir verwenden einige Begriffe aus dem Einbettungsraum:

$$a_{k;\,\nu_1 \ldots \nu_{k+2}} := \sin^2\{\tfrac{1}{2}(k;\, \nu_1 \ldots \nu_{k+2})\} \tag{8}$$

ist das Quadrat der halben Sehne zwischen den Eckpunkten des Grosskreisbogens $(k;\, \nu_1 \ldots \nu_{k+2})$. Es folgt

$$\left.\begin{aligned} \sin(k;\, \nu_1 \ldots \nu_{k+2}) &= \sqrt{4\, a_{k;\,\nu_1 \ldots \nu_{k+2}}(1 - a_{k;\,\nu_1 \ldots \nu_{k+2}})} \\ \cos(k;\, \nu_1 \ldots \nu_{k+2}) &= 1 - 2\, a_{k;\,\nu_1 \ldots \nu_{k+2}} \end{aligned}\right\} \tag{9}$$

Wir betrachten ein Simplex

$$(k';\, \nu_1 \ldots \nu_{h'}), \tag{10}$$

dessen Untersimplexverband ein nichtleeres System von Winkeln besitzen soll, d.h. $|h'-k'| \geqslant 2$. Hiernach setzen wir

$$\left.\begin{aligned} k' = k;\quad h' - 2 = h \quad &\text{für} \quad k' \leqslant h' - 2 \\ h' = h;\quad k' - 2 = k \quad &\text{für} \quad h' \leqslant k' - 2\,. \end{aligned}\right\} \tag{11}$$

Die für Betrachtungen im Einbettungsraum bevorzugten Winkel der Ordnung 0 im Verband $[(k';\, \nu_1 \ldots \nu_{k'}), (k';\, \nu_1 \ldots \nu_{h'})]$, genannt *Kanten* des Simplexes $(k';\, \nu_1 \ldots \nu_{h'})$. sind

$$\left.\begin{aligned} &(k;\, \nu_1 \ldots \nu_k\, \mu_{k+1}\, \mu_{k+2}) \\ &\qquad \text{mit} \quad \{\mu_{k+1}, \mu_{k+2}\} \subseteq \{\nu_{k+1} \ldots \nu_{h+2}\} \quad \text{für} \quad k \leqslant h\,, \\ &\pi - (k;\, \nu_1 \ldots \nu_h\, \mu_{h+1} \ldots \mu_{k+2}) \\ &\qquad \text{mit} \quad \{\mu_{h+1} \ldots \mu_{k+2}\} = \{\nu_{h+1} \ldots \nu_{k+2}\} \quad \text{für} \quad h \leqslant k\,. \end{aligned}\right\} \tag{12}$$

Ferner finden wir die zugehörigen Quadrate halber Sehnen der Kanten

$$\left.\begin{aligned} a_{k;\,\nu_1 \ldots \nu_k\, \mu_{k+1}\, \mu_{k+2}} \quad &\text{für} \quad k \leqslant h \\ 1 - a_{k;\,\nu_1 \ldots \nu_h\, \mu_{h+1} \ldots \mu_{k+2}} \quad &\text{für} \quad h \leqslant k\,. \end{aligned}\right\} \tag{13}$$

Nehmen wir zu den $|h'-k'|$ Eckpunkten von $(k';\, \nu_1 \ldots \nu_{h'})$ noch den Mittelpunkt der Sphäre hinzu, finden wir zu diesen $(|h' - k'| + 1)$ Eckpunkten ein $|h' - k'| + 1$ dimensionales euklidisches Simplex, dessen Volumen wir etwas abweichend von [8] mit

$$\underset{|h' - k'|}{L}(k';\, \nu_1 \ldots \nu_{h'}) \tag{14}$$

bezeichnen. Zur Berechnung dieses Volumens benötigen wir die Determinante

$$D(k;h;\nu_1\dots\nu_n):=\begin{cases}(-1)^{h-k}\begin{vmatrix}0 & \cdots & \cdots & 1\\ \vdots & \ddots \; a_{k;\,\nu_1\dots\nu_k\,\mu_{k+1}\,\mu_{k+2}} & \vdots & \vdots\\ \vdots & \cdots & 0 & 1\\ 1 & \cdots & 1 & 2\end{vmatrix} & \text{mit } \{\mu_{k+1},\mu_{k+2}\}\subseteq\{\nu_{k+1}\dots\nu_{h+2}\} \text{ für } k\leqslant h\\[2em] (-1)^{k-h}\begin{vmatrix}0 & \cdots & \cdots & 1\\ \vdots & \ddots \; 1-a_{k;\,\nu_1\dots\nu_h\,\mu_{h+1}\dots\mu_{k+2}} & \vdots & \vdots\\ \vdots & \cdots & 0 & 1\\ 1 & \cdots & 1 & 2\end{vmatrix} & \text{mit } \{\mu_{h+1}\dots\mu_{k+2}\}=\{\nu_{h+1}\dots\nu_{k+2}\} \text{ für } h\leqslant k\end{cases} \tag{15}$$

Die Zeiger μ_i durchlaufen hier sämtliche angegebenen Werte. μ_{k+1} sei Zeilen- und μ_{k+2} Spaltenindex. Die Anordnung der Zeilen und Spalten erfolge nach dem Auftreten des Index in der Anordnung $\nu_{k+1},\dots,\nu_{h+2}$ für $k\leqslant h$ bzw. $\nu_{h+1},\dots,\nu_{k+2}$ für $h\leqslant k$. Für $k=h$ sind 2 Werte definiert, welche jedoch übereinstimmen. Mit (15) wird

$$L(k';\nu_1\dots\nu_{h'})=\frac{2^{1/2|h'-k'-1|}}{|h'-k'-1|!}\cdot\sqrt{D(k;h;\nu_1\dots\nu_n)}, \tag{16}$$

woraus die Bedingung

$$D(k;h;\nu_1\dots\nu_n)\geqslant 0 \quad \text{für alle} \quad k;h;\nu_1\dots\nu_n \tag{17}$$

folgt. Die hiermit im Einbettungsraum geometrisch gedeutete Determinante $D(k;h;\dots)$ wird auch in der nichteuklidischen Simplextheorie Bedeutung haben. Zunächst soll jedoch noch weiter definiert werden:

Werden im Symbol $D(k;h;\dots)$ je l der als Zeilen- bzw. Spaltenindizes auftretenden Punktnummern über- bzw. unterstrichen,

$$D(k;h;\dots\overline{\nu_{r_1}}\dots\underline{\nu_{s_1}}\dots\overline{\nu_{r_l}}\dots\underline{\nu_{s_l}}\dots), \tag{18}$$

so sollen in den Determinanten (15) die Zeilen mit überstrichener und die Spalten mit unterstrichener Nummer gestrichen und die Determinante dann mit $(-1)^l$ multipliziert werden.

Für $k=h$ können zwei entgegengesetzt gleiche Werte definiert sein, so dass wir hier immer festlegen müssen, ob die Determinante zum Fall $k\leqslant h$ oder zum Fall $h\leqslant k$ zu rechnen ist.

Als Spezialfall des Determinantensatzes von Sylvester finden wir die Gleichung

$$\left.\begin{aligned} & D(k;h;\ldots v_{j_1}\ldots v_{j_2}\ldots v_{j_3}\ldots v_{j_4}\ldots)\cdot D(k;h;\ldots \overline{v_{j_1}}\ldots \overline{v_{j_2}}\ldots \underline{v_{j_3}}\ldots \underline{v_{j_4}}\ldots) \\ & \quad = D(k;h;\ldots \overline{v_{j_1}}\ldots v_{j_2}\ldots \underline{v_{j_3}}\ldots v_{j_4}\ldots)\cdot D(k;h;\ldots v_{j_1}\ldots \overline{v_{j_2}}\ldots v_{j_3}\ldots \underline{v_{j_4}}\ldots) \\ & \quad - D(k;h;\ldots \overline{v_{j_1}}\ldots v_{j_2}\ldots v_{j_3}\ldots \underline{v_{j_4}}\ldots)\cdot D(k;h;\ldots v_{j_1}\ldots \overline{v_{j_2}}\ldots \underline{v_{j_3}}\ldots v_{j_4}\ldots) \\ & \quad \text{mit} \quad k+1 \leqslant \begin{Bmatrix} j_1 < j_2 \\ j_3 < j_4 \end{Bmatrix} \leqslant h+2 \quad \text{bzw.} \quad h+1 \leqslant \begin{Bmatrix} j_1 < j_2 \\ j_3 < j_4 \end{Bmatrix} \leqslant k+2, \end{aligned}\right\} \tag{19}$$

wo gleichzeitig in sämtlichen Determinanten noch weitere Streichungen vorgenommen werden können.

Wir formulieren einen in den Sehnenquadraten rationalen Ausdruck

$$J(k;h;v_1\ldots v_n) := \frac{D(k;h;\ldots v_{h+1}\,v_{h+2}\ldots)\cdot D(k;h;\ldots \underline{\overline{v_{h+1}}}\,\underline{\overline{v_{h+2}}}\ldots)}{D(k;h;\ldots \underline{\overline{v_{h+1}}}\,v_{h+2}\ldots)\cdot D(k;h;\ldots v_{h+1}\,\underline{\overline{v_{h+2}}}\ldots)} \tag{20}$$

oder nach (19)

$$1 - J(k;h;v_1\ldots v_n) = \frac{D^2(k;h;\ldots \overline{v_{h+1}}\,\underline{v_{h+2}}\ldots)}{D(k;h;\ldots \underline{\overline{v_{h+1}}}\,v_{h+2}\ldots)\cdot D(k;h;\ldots v_{h+1}\,\underline{\overline{v_{h+2}}}\ldots)}. \tag{21}$$

Für die in (22) und (23) eingehenden Wurzeln legen wir uns auf das positive Vorzeichen bei (20) und auf das Vorzeichen von $D(k;h;\ldots \overline{v_{h+1}}\,\underline{v_{h+2}}\ldots)$ bei (21) fest. Mit diesen Wurzelfixierungen ist

$$\sqrt{J(k;h;v_1\ldots v_n)} = \sin(h;v_1\ldots v_{h+2}) = \sin[\pi - (h;v_1\ldots v_{h+2})] \tag{22}$$

$$\left.\sqrt{1 - J(k;h;v_1\ldots v_n)} = \begin{cases} \cos(h;v_1\ldots v_{h+2}) & \text{für} \quad k \leqslant h \\ \cos[\pi - (h;v_1\ldots v_{h+2})] & \text{für} \quad h \leqslant k \end{cases} \quad \text{für alle} \quad k;h;v_1\ldots v_n. \right\} \tag{23}$$

Auf den Beweis durch vollständige Induktion in den beiden Dreiecksschemata $k \leqslant h$ und $h \leqslant k$ durch Anwendung des sphärischen cos-Satzes wollen wir hier verzichten.

Analog hierzu sind die beiden folgenden Aussagen: Der Wert der Funktion $J(k;h;v_1\ldots v_n)$ ist invariant gegenüber Änderungen von k und gegenüber Tauschmengen in den Punktmengen $\{v_1\ldots v_h\}$, $\{v_{h+1}, v_{h+2}\}$ bzw. $\{v_{h+3}\ldots v_n\}$. Die erste Aussage ist gehaltvoller, weshalb wir die Funktion $J(k;h;v_1\ldots v_n)$ als *Ordnungsinvariante* bezeichnen.

Die Invarianzeigenschaften von $J(k;h;v_1\ldots v_n)$ beschreiben vollständig die innere Geometrie, d.h. die zwischen den Winkeln bestehenden Zusammenhänge im sphärischen Simplex. Insbesondere können wir etwa vermöge (23) sämtliche Winkel eines Simplex aus denen einer bevorzugten Ordnung k berechnen, wonach im System der Winkel noch $\binom{n}{2}$ Freiheitsgrade bestehen.

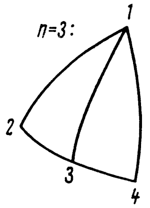
§ 3. Funktionalgleichungen zur Volumenadditivität

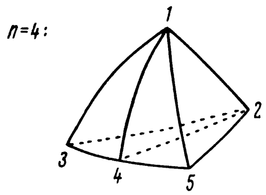
Wir verlassen nun die innere Geometrie der Simplexe und wenden uns dem Problem der Inhaltsmessung zu. Unter den kennzeichnenden Eigenschaften für das Simplexvolumen als Funktion eines Bestimmunssystems spielt die Volumenadditivität bei ergänzungsgeometrischen Konstruktionen, in denen nur Simplexe beteiligt sind, eine grosse Rolle. Wir betrachten hier nur eine spezielle Klasse ergänzungsgeometrischer Konstruktionen, welche sich auf den Spezialfall der Ergänzung zweier nichtentarteter Simplexe $S_n^{(1)}$ und $S_n^{(2)}$ zu einem Simplex $S_n^{(3)}$ zurückführen lässt. Es sei

$$\left.\begin{aligned} S_n^{(1)} &= (1, \ldots, n-2, n-1, n) \\ S_n^{(2)} &= (1, \ldots, n-2, n, n+1) \\ S_n^{(3)} &= (1, \ldots, n-2, n-1, n+1) \end{aligned}\right\} \tag{1}$$

also

$$(1, \ldots n-2, n-1, n) + (1, \ldots n-2, n, n+1) = (1, \ldots n-2, n-1, n+1) \tag{2}$$





Das Simplexvolumen ist $\binom{n}{2}$-parametrig. Bei Berücksichtigung der Bindungen zwischen den Winkeln der 3 Simplexe bleibt als analytischer Ausdruck der Volumenadditivität (2) eine $\left[\binom{n}{2}+1\right]$-stufige Funktionalgleichung: $\binom{n}{2}$ Parameter hatte etwa das Simplex $(1 \ldots n)$, und durch einen weiteren Parameter, welchen wir jedoch erst später einführen wollen, wird die Lage des Punktes $(n+1)$ fixiert.

Mit der Festlegung

$$\left.\begin{aligned} v_j &= n \\ v_l &= \begin{Bmatrix} n-1 \\ n+1 \end{Bmatrix} \text{ in Ausdrücken mit dem oberen Zeiger } \begin{Bmatrix} 1 \\ 2 \end{Bmatrix} \end{aligned}\right\} \tag{3}$$

können wir als Bedingung der Konstruktion (2) sofort einige Winkel des Simplexes $S_n^{(2)}$ durch Winkel von $S_n^{(1)}$ ausdrücken:

$$\left.\begin{aligned} &(h; v_1 \ldots v_{h+2})^{(2)} = \\ &\quad \begin{cases} (h; v_1 \ldots v_{h+2})^{(1)} & \text{für } j \leqslant h,\ l \notin \{h+1, h+2\} \text{ und für } h+2 < l \\ \pi - (h; v_1 \ldots v_{h+2})^{(1)} & \text{für } j \leqslant h,\ l \in \{h+1, h+2\} \end{cases} \end{aligned}\right\} \tag{4}$$

Weitere Kenntnisse über die Winkel von $S_n^{(2)}$ gewinnen wir durch die Berücksichtigung der inneren Geometrie der sphärischen Simplexe, d.h. indem wir die k-Invarianten (2.22) und (2.23) in (4) einsetzen.

Legen wir uns nun etwa auf die Quadrate halber Sehen der Kanten als Bestimmungssystem des Simplexvolumens fest, so finden wir nach längerer Rechnung das Gleichungssystem

$$\left.\begin{array}{l}(\lambda_1, \lambda_2)^{(2)} = (\lambda_1, \lambda_2)^{(1)} \quad \text{für} \quad \lambda_1, \lambda_2 \in \{1 \ldots n-2\} \\ \dfrac{D(0; h^*; n, \overline{n-1}; v_1 \ldots \underline{v_h}^* \ldots v_{n-2})}{D(0; h^*; n, \overline{n+1}; v_1 \ldots \underline{v_h}^* \ldots v_{n-2})} = -\sqrt{\dfrac{D(0; h; n, n-1, v_1 \ldots v_{n-2})}{D(0; h; n, n+1, v_1 \ldots v_{n-2})}} \\ \text{für alle} \quad h, h^* \quad \text{mit} \quad \begin{cases} 0 \leqslant h \leqslant n-2 \\ 1 \leqslant h^* \leqslant n-2 \end{cases} \\ \text{und für} \quad \{v_1 \ldots v_{n-2}\} = \{1 \ldots n-2\}. \end{array}\right\} \tag{5}$$

Die beiden Seiten sind hiernach invariant gegenüber Änderung von h^* bzw. h in den angegebenen Grenzen, wir sprechen von *h-Invarianten.* Als letzten Parameter der Ergänzung von $S_n^{(1)}$ und $S_n^{(2)}$ zu $S_n^{(3)}$ wählen wir die einfachere h-Invariante unter der Wurzel und legen diesen Parameter sogleich auf seinen einfachsten Wert 1 fest,

$$\frac{D(0; h; n, n-1, v_1 \ldots v_{n-2})}{D(0; h; n, n+1, v_1 \ldots v_{n-2})} = \frac{\sin^2(n-1, n)}{\sin^2(n, n+1)} := 1. \tag{6}$$

Hieraus und aus (5) mit $h^* = 1$ folgen die beiden Lösungen für das Bestimmungssystem von $S_n^{(2)}$:

$$\left.\begin{array}{l} a_{n,n+1} = 1 - a_{n-1,n} \\ a_{v,n+1} = 1 - a_{v,n-1} \end{array} \quad \text{für} \quad v \in \{1 \ldots n-2\} \right\} \tag{7}$$

($S_n^{(1)}$ und $S_n^{(2)}$ ergänzen sich zu einem sphärischen Zweieck);

$$\left.\begin{array}{l} a_{n,n+1} = a_{n-1,n} \\ a_{v,n+1} = -a_{v,n-1} + 2(a_{v,n} + a_{n-1,n} - 2a_{v,n} \cdot a_{n-1,n}) \quad \text{für} \quad v \in \{1 \ldots n-2\} \end{array}\right\} \tag{8}$$

($S_n^{(1)}$ und $S_n^{(2)}$ ergänzen sich mit Kantenverdopplung).

Die Ergänzungen zum Zweieck und die aus diesem Prototyp ableitbaren Funktionalgleichungen haben zur Charakterisierung des Simplexvolumens wenig Aussagekraft. Ihre Bedeutung könnte darin liegen, dass sich mit ihrer Hilfe bekannte Funktionalgleichungen geometrisch neu deuten lassen. So entsteht z.B. für $n=3$ die zweistufige arc cos-Funktionalgleichung in der symmetrischen Form

$$g\begin{pmatrix} x \\ y\,z \end{pmatrix} := \frac{2x^2}{(x+y)(x+z)} - 1 \tag{9}$$

$$H\left[g\begin{pmatrix} x \\ y\,z \end{pmatrix}\right] + H\left[g\begin{pmatrix} y \\ z\,x \end{pmatrix}\right] + H\left[g\begin{pmatrix} z \\ x\,y \end{pmatrix}\right] = 0 \quad \text{für} \quad H(s) = c \cdot \operatorname{arc\,cos} s. \tag{9'}$$

Die Stufe 3 ist nur scheinbar, da die 3 Argumente identisch die arc cos-Multiplikationsgleichung

$$r = s\,t - \sqrt{1-s^2}\cdot\sqrt{1-t^2} \tag{10}$$

befriedigen. Gewisse aus der klassischen Geometrie herrührende Ungleichungen in x, y und z weisen eine Unsymmetrie auf, die auf die bevorzugte Stellung eines der Summanden in (9′) als Inhaltsfunktion eines Zweiecks zurückzuführen ist.

Ähnlich wie (9′) lässt sich auch die dreistufige arc cos-Funktionalgleichung rein geometrisch begründen.

Fasst man die Ergänzung mit Kantenverdopplung als Prototyp einer erweiterten Klasse von ergänzungsgeometrischen Konstruktionen auf, so charakterisieren die entstehenden Funktionalgleichungen die Inhaltsfunktion des Simplexes stärker.

Wir ergänzen die Simplexe $(1,\dots,n-2,n+q-1,n+q)$ mit $q\in\{0,\dots,q_0\}$ zu einem Simplex $(1,\dots,n-2,n-1,n+q_0)$, also

$$\sum_{q=0}^{q_0}(1,\dots,n-2,n+q-1,n+q) = (1,\dots,n-2,n-1,n+q_0). \tag{11}$$

Die Festlegungen

$$\left.\begin{aligned}(n+q-1,n+q) &:= (n-1,n) \quad \text{für} \quad q\in\{1,\dots,q_0\} \quad \text{und}\\ (n-1,n+q_0) &= (q_0+1)\cdot(n-1,n) < \pi\end{aligned}\right\} \tag{12}$$

bewirken, dass je 2 Nebensimplexe mit Kantenverdopplung ergänzt werden und dass der Ergänzungssimplex $(1,\dots,n-2,n-1,n+q_0)$ aller Simplexe $(1,\dots,n-2,n+q-1,n+q)$ konvex ist.

Nach Einführung des Symbols

$$(1\dots n) =: S_n(q_{\kappa\lambda})_{\kappa,\lambda\in\{1\dots n\}} =: S_n\begin{pmatrix}a_{\kappa\lambda}\\ a_{\kappa,n-1}\\ a_{\kappa,n}\\ a_{n-1,n}\end{pmatrix}_{\kappa,\lambda\in\{1\dots n-2\}} \tag{13}$$

für die Inhaltsfunktion des Simplexes entsteht aus (11)

$$\sum_{q=0}^{q_0} S_n\begin{pmatrix}a_{\kappa\lambda}\\ a_{\kappa,n+q-1}\\ a_{\kappa,n+q}\\ a_{n+q-1,n+q}\end{pmatrix} = S_n\begin{pmatrix}a_{\kappa\lambda}\\ a_{\kappa,n-1}\\ a_{\kappa,n+q_0}\\ a_{n-1,n+q_0}\end{pmatrix}, \tag{14}$$

wo κ und λ jeweils die Menge $\{1\dots n-2\}$ durchlaufen.

Wir haben nun alle Argumente aus dem Bestimmungssystem von $(1\dots n)$ zu berechnen, wonach aus (14) eine $\binom{n}{2}$-stufige Funktionalgleichung entsteht. Hierzu verwenden wir die Tschebyscheffschen Polynome 1. und 2. Art, welche etwa durch die dreigliedrige Rekursion

$$f_{m+1}(x) - 2x f_m(x) + f_{m-1}(x) = 0 \tag{15}$$

mit den Anfangsbedingungen

$$\left.\begin{array}{lll} T_0(x) = 1 & \text{und} & P_0^*(x) = 1 \\ T_1(x) = x & & P_1^*(x) = 2x \end{array}\right\} \tag{16}$$

gekennzeichnet seien [6].

Für die Argumente in (14) finden wir zunächst nach (12)

$$a_{n+q-1,\,n+q} = a_{n-1,\,n} \tag{17}$$

und

$$a_{n-1,\,n+q_0} = \tfrac{1}{2}\left[1 - T_{q_0+1}(1 - 2a_{n-1,\,n})\right]. \tag{18}$$

Ersetzen wir ferner in (5) die Punktnummer $n+1$ durch $n+q$ und wenden (17) an, so entsteht als Verallgemeinerung von (8)

$$\left.\begin{array}{l} a_{\kappa,\,n+q} = \tfrac{1}{2}\left[1 + (1 - 2a_{\kappa,\,n-1})\cdot P_{q-1}^*(1 - 2a_{n-1,\,n}) - (1 - 2a_{\kappa,\,n}) \right. \\ \qquad \left. \times\, P_q^*(1 - 2a_{n-1,\,n})\right] \quad \text{mit} \quad \kappa \in \{1 \ldots n-2\} \quad \text{und} \quad q \in \{1 \ldots q_0\} \end{array}\right\}, \tag{19}$$

womit wir alle Argumente in (14) durch das Bestimmungssystem von $(1 \ldots n)$ ausgedrückt und damit eine $\binom{n}{2}$-stufige Funktionalgleichung erhalten haben. Wir beabsichtigen, $q_0 \to \infty$ gehen zu lassen und dabei die Kante $(n-1,\, n+q_0)$ festzuhalten. Wegen des Verhaltens der Tschebyscheffschen Polynome sind die Darstellungen (18) und (19) der Argumente von (14) ungeeignet für diesen Grenzübergang.

Da bei dem Prozess nur das Simplex $(1, \ldots, n-2, n-1, n+q_0)$ unberührt bleibt, während das Simplex $(1 \ldots n)$ wegen $(n-1, n) \to 0$ durch Verlust seiner Dimension entartet, sollen nach dem Grenzübergang die Winkel des Simplex $(1, \ldots, n-2, n-1, n+q)$ das Parametersystem bilden.

Während wir also die Argumente auf der rechten Seite von (14) stehenlassen, finden wir für die Argumente auf der linken Seite statt (19) aus (5) die Darstellung

$$\left.\begin{array}{l} a_{\kappa,\,n+q} = \tfrac{1}{2}\{1 - \cos(\kappa n)\cdot\cos[q\cdot(n-1, n)] \\ \qquad\qquad + \cos(1; n, \kappa, n+q_0)\cdot\sin(\kappa n)\cdot\sin[q\cdot(n-1, n)]\} \\ \qquad\qquad \text{mit} \quad \kappa \in \{1 \ldots n-2\} \quad \text{und} \quad q \in \{-1, 0, 1, \ldots, q_0\}. \end{array}\right\} \tag{20}$$

L. Schläfli [8] führte in die Inhaltsmessung den Kalkül der Volumenquotienten ein:

$$\frac{(1\ldots n)}{L(1\ldots n)} =: \phi_n(a_{\kappa\lambda})_{\kappa,\lambda\in\{1\ldots n\}} =: \phi_n\begin{pmatrix} a_{\kappa\lambda} \\ a_{\kappa,n-1} \\ a_{\kappa,n} \\ a_{n-1,n}\end{pmatrix}_{\kappa,\lambda\in\{1\ldots n-2\}} . \tag{21}$$

Wir rufen uns noch einmal die Tatsache ins Gedächtnis zurück, dass jedes Paar von Nebensimplexen durch Kantenverdopplung ergänzt wird, was für den entsprechenden Parameter die Festlegung (6) nach sich zieht. Nach (2.16) bedeutet dies, dass sich (21) auch für die Volumenquotienten relativ einfach formulieren lässt, da die Volumenquotienten der Simplexe $(1,\ldots,n-2, n+q-1, n+q)$ alle den gleichen Nanner haben. Es entsteht

$$\begin{aligned} \phi_n\begin{pmatrix} a_{\kappa\lambda} \\ a_{\kappa,n-1} \\ a_{\kappa,n+q_0} \\ a_{n-1,n+q_0}\end{pmatrix} &= \sqrt{\frac{D(0;n-2;n,n+q_0,1,\ldots,n-2)}{D(0;n-2;n-1,n+q_0,1,\ldots,n-2)}} \\ &\times \sqrt{\frac{D(0;n-2;n,n-1,1,\ldots,n-2)}{D(0;n-2;n,n+q_0,1,\ldots,n-2)}} \cdot \sum_{q=0}^{q_0} \phi_n\begin{pmatrix} a_{\kappa\lambda} \\ a_{\kappa,n+q-1} \\ a_{\kappa,n+q} \\ a_{n-1,n}\end{pmatrix} \end{aligned} \tag{22}$$

wo κ und λ überall die Menge $\{1\ldots n-2\}$ durchlaufen, mit Bestimmung $\{a_{k,\nu}\}$ aus (20). Die vorkommende h-Invariante ersetzen wir durch

$$\sqrt{\frac{D(0;n-2;n,n-1,1,\ldots,n-2)}{D(0;n-2;n,n+q_0,1,\ldots,n-2)}} = \frac{\sin(n,n-1)}{\sin(n,n+q_0)}. \tag{23}$$

Setzen wir Holomorphie des Volumenquotienten ϕ_n an der Stelle des Verschwindens eines Bestimmungsstückes $a_{n-1,n}$ in sämtlichen Veränderlichen voraus, so können wir bei Berücksichtigung eines Restgliedes in (22) $a_{n-1,n}=0$ setzen. Mit den weiteren Abschätzungen

$$\left.\begin{aligned} a_{n-1,n} &= O[(n-1,n)^2] \\ \sin(n-1,n) &= (n-1,n) + O[(n-1,n)^3] \end{aligned}\right\} \tag{24}$$

können wir den Grenzübergang $q_0 \to \infty$ durchführen.

Ersetzen wir beim Grenzübergang

$$\left.\begin{aligned} (n-1,n) &\to dt \\ \left.\begin{aligned} q\cdot(n-1,n) \\ (q-1)\cdot(n-1,n) \end{aligned}\right\} &\to t, \end{aligned}\right\} \tag{25}$$

so geht die Summe (22) über in ein Integral.

Die Punktnummer n kann nach dem Grenzübergang überall durch $n-1$ ersetzt werden und kommt dann nicht mehr vor, so dass wir den Punkt $n+q_0$ neu benennen können. Mit

$$g_\kappa(t) := \tfrac{1}{2}[1 - \cos(\kappa, n-1)\cdot\cos t + \cos(1; n-1, \kappa, n)\cdot\sin(\kappa, n-1)\cdot\sin t] \tag{26}$$

entsteht endgültig aus (22)

$$\left.\phi_n\begin{pmatrix} a_{\kappa\lambda} \\ a_{\kappa,n-1} \\ a_{\kappa,n} \\ a_{n-1,n}\end{pmatrix} = \frac{1}{\sin(n-1,n)}\cdot\int_0^{(n-1,n)} dt\cdot\phi_n\begin{pmatrix} a_{\kappa\lambda} \\ g_\kappa(t) \\ g_\kappa(t) \\ 0\end{pmatrix}\quad\right\} \tag{27}$$

wo κ und λ überall die Menge $\{1 \ldots n-2\}$ durchlaufen.

Als Folge der Volumenadditivität bei der ergänzungsgeometrischen Konstruktion (11) mit (12) kann also der Volumenquotient eines nichtentarteten Simplex $(1 \ldots n)$ als Integral über Volumenquotienten von dimensionsentarteten Simplexen mit einer verschwindenden Kante geschrieben werden, und die Bestimmung der verbleibenden Volumenquotienten reduziert sich auf eine lokale Aussage. Eine beschränkte weitere Zurückführung ist möglich durch Darstellung des Integranden in (27) als Integral über Volumenquotienten mit zwei punktfremden verschwindenden Kanten. Nach Wiederholung des Verfahrens ergibt sich ϕ_n als $\left[\frac{n}{2}\right]$-faches Integral über Volumenquotienten mit $\left[\frac{n}{2}\right]$ paarweise punktfremden verschwindenden Kanten, wo $[x]$ die grösste ganze Zahl nicht grösser als x bedeutet. (27) lässt sich verallgemeinern zu

$$\left.\begin{aligned}\phi_n\begin{pmatrix} a_{\kappa\lambda} \\ g_\kappa(t_1) \\ g_\kappa(t_2) \\ \sin^2\{\tfrac{1}{2}(t_2-t_1)\}\end{pmatrix} &= \frac{1}{\sin(t_2-t_1)}\cdot\int_{t_1}^{t_2} dt\cdot\phi_n\begin{pmatrix} a_{\kappa\lambda} \\ g_\kappa(t) \\ g_\kappa(t) \\ 0\end{pmatrix} \\ &\text{mit}\quad 0 < t_2 - t_1 < \pi,\end{aligned}\right\} \tag{28}$$

wo κ und λ jeweils die Menge $\{1 \ldots n-2\}$ durchlaufen.

Diese Darstellung löst nicht nur die von uns betrachteten Funktionalgleichungen (14) mit (17), (18), (19), sondern z.B. auch diejenige der Ergänzung zweier Simplexe zu einem dritten bei Verzicht auf die Parameterfestlegung (6): Die Volumenadditivität folgt aus der Additivität des Riemannschen Integrals in (28). Der Parameter (6), welcher gegenüber den restlichen Parametern eine Sonderstellung einnahm, geht also in die Lösung der Funktionalgleichungen nicht ein.

SCHRIFTTUM

[1] ACZÉL, J., *Vorlesungen über Funktionalgleichungen* (Berlin 1961), S. 20–315.
[2] BÖHM, J., *Coxeters Integrationsmethode in gekrümmten Räumen*, Math. Nachr. *21*, 179–214 (1964).
[3] KNESER, H., *Simplexinhalte i.d. nichteuklidischen Geometrie*, Deutsche Math., *1*, 337–340 (1936).
[4] LOBATSCHEWSKI, N. J., *Imaginäre Geometrie*, Übersetzung H. LIEBMANN, (Leipzig 1904), S. 82f.
[5] MAIER, W., *Nichteuklidische Volumina*, Sitzungsber. Sächs. Akad. Wiss., Math.-Nat. Klasse, *107*, H. 6, 3–19 (1967).
[6] NAAS, J. und SCHMID, H. L., *Mathematisches Wörterbuch*, II, 755.
[7] RUBEN, H., *Skew-Regular Simplices in Hyperspherical Space*, Acta Math., *103*, 1–23 (1960).
[8] SCHLÄFLI, L., *Ges. math. Abhandl.*, Bd. 3 (Basel 1956), S. 21–39.
[9] WEISSBACH, B., *Simplexinhalt in Räumen konstanter Krümmung*, Promotionsschrift T. H. Magdeburg (1966), S. 44–65.

A Special Class of Doubly Stochastic Matrices

ALAN J. HOFFMAN (New York)

Dedicated to A. M. Ostrowski on his 75th birthday

1. Introduction

We shall be dealing with real vectors in R^n and square real matrices of order n. Such a matrix $S=(s_{ij})$ is said to be doubly stochastic if $s_{ij}\geqslant 0$, and all row and column sums of S are 1. A permutation matrix P is a doubly stochastic matrix in which every entry is 0 or 1. A vector $x=(x_1,\dots,x_n)\in R^n$ is said to be in descending order if $x_1\geqslant\cdots\geqslant x_n$, and we write $x\in D(n)$. For any $x\in R^n$, x^D is the unique vector in $D(n)$ of the form Px, where P is a permutation matrix.

The best known facts about doubly stochastic matrices are the theorems of BIRKHOFF [2] and HARDY, LITTLEWOOD and POLYA [3], to be described below. The significance of these results can be seen in OSTROWSKI [6], and our interest arose in their use [1] in proving a theorem of OSTROWSKI and TAUSSKY [7]. An excellent survey of results and problems in the study of doubly stochastic matrices is given by MIRSKY in [5].

(1.1) BIRKHOFF'S THEOREM: *The set of doubly stochastic matrices is the convex hull of the permutation matrices.*

(1.2) HARDY, LITTLEWOOD AND POLYA'S THEOREM: If $x, y\in R^n$, then there exists a doubly stochastic matrix S such that $x=Sy$ if and only if

$$(x^D, u_k)\leqslant(y^D, u_k),\qquad k=1,\dots,n-1, \tag{1.3}$$

$$(x^D, u_n)=(y^D, u_n), \tag{1.4}$$

where $u_k(k=1,\dots,n)$ is the vector whose first k co-ordinates are 1, all other co-ordinates 0.

Most proofs of (1.2) assume $x, y\in D(n)$ as a preliminary maneuver. In this note we assume $x, y\in D(n)$ and focus on the question of finding a minimal convex subset $\mathscr{R}(n)$ of the set of symmetric doubly stochastic matrices of order n such that $y\in D(n)$, $R\in\mathscr{R}(n)$, implies $Ry\in D(n)$. To explain the results, we introduce some definitions and notations.

Let $N=\{1,\dots,n\}$, and let $\mathscr{P}(n)$ be the set of all partitions ϱ of N each of which has the property that each summand consists of consecutive integers. Thus, $\mathscr{P}(4)=\{(1,2,3,4),(12,3,4),(1,23,4),(123,4),(1,2,34),(12,34),(1,234),(1234)\}$. We clal

Received December 19, 1967 and, in revised form, February 1, 1968

$\mathscr{P}(n)$ the set of consecutive partitions of N and denote any particular consecutive partition ϱ of N by listing the subsets in order $\varrho=(\varrho_1,\ldots,\varrho_k)$. We use the symbol $|X|$ to denote the number of elements in the set X. For any consecutive partition ϱ in $\mathscr{P}(N)$ we define the symmetric doubly-stochastic matrix M^ϱ by the rule:

$$M^\varrho_{i,j}=\begin{Bmatrix} \dfrac{1}{|\varrho_t|} & \text{if} \quad i,j\in\varrho_t \\ 0 & \text{otherwise} \end{Bmatrix}. \tag{1.5}$$

We use the symbol $\mathscr{M}^{\mathscr{P}(n)}$ to denote the 2^{n-1} matrices of the form (1.5).

Next, we use the symbol $\mathscr{R}(n)$ to denote all symmetric doubly stochastic matrices $R=(r_{ij})$ of order n satisfying

$$r_{1n}\geqslant 0 \tag{1.6}$$

$$\sum_j r_{ij}=1 \quad \text{for all} \quad i; \tag{1.7}$$

$$r_{ij}=r_{ji} \quad \text{for all} \quad i,j; \tag{1.8}$$

$$r_{11}\geqslant r_{12}\geqslant\cdots\geqslant r_{1n}; \tag{1.9}$$

$$r_{nn}\geqslant r_{n-1,n}\geqslant\cdots\geqslant r_{1n}; \tag{1.10}$$

$$r_{ij}+r_{i-1,j+1}\geqslant r_{i-1,j}+r_{i,j+1} \quad \text{if} \quad 2\leqslant i\leqslant j\leqslant n-1. \tag{1.11}$$

Matrices satisfying (1.6), (1.8)–(1.11) are called 'uniformly tapered' in [9] and [10]. Our principal results, in analogy to (1.1) and (1.2) are

THEOREM 1. *$\mathscr{R}(n)$ is the convex hull of $\mathscr{M}^{\mathscr{P}(n)}$.*

THEOREM 2. *If $x, y\in D(n)$, then they satisfy* (1.3) *and* (1.4) *if and only if there exists an $R\in\mathscr{R}(n)$ such that $x=Ry$.*

We shall give three proofs of theorem 2. The first is in the spirit of the proof of (1.2) given by RADO in [8]; the second is the same as one given some years ago in an unpublished note [4]. The second proof has an obvious pictorial interpretation, and since theorem 2 includes (1.2), it is to our taste the most intuitive demonstration of (1.2). The third (and simplest) proof, by induction, is due to D. Ž. DJOKOVIĆ.

2. Proof of Theorem 1

It is clear that every matrix in $\mathscr{M}^{\mathscr{P}(n)}$ is in $\mathscr{R}(n)$. To prove theorem 1, we must show that every matrix $R\in\mathscr{R}(n)$ is a convex combination of matrices in $\mathscr{M}^{\mathscr{P}(n)}$.

Let R be given. We first show

$$\text{if} \quad i\leqslant j\leqslant n-1, \quad r_{ij}\geqslant r_{i,j+1}. \tag{2.1}$$

This is certainly true for $i=1$ by (1.9). Assume inductively that it has been shown for $i=k-1\leqslant n-2$, so we have

$$r_{k-1,j}\geqslant r_{k-1,j+1} \quad \text{for} \quad k\leqslant j.$$

By (1.11),

$$r_{kj}+r_{k-1,j+1}\geqslant r_{k-1,j}+r_{k,j+1}.$$

Adding the last two inequalities verifies (2.1) for $i=k$ and the induction is complete. Using (1.10) and (1.11) in a similar manner, we have

$$\text{if} \quad 2\leqslant i\leqslant j \qquad r_{ij}\geqslant r_{i-1,j}. \tag{2.2}$$

We also note that (2.1), (2.2) and (1.8) imply

$$\text{if} \quad n-1\geqslant i\geqslant j, \quad r_{ij}\geqslant r_{i+1,j}, \quad \text{and} \tag{2.3}$$

$$\text{if} \quad i\geqslant j\geqslant 2, \qquad r_{ij}\geqslant r_{i,j-1}. \tag{2.4}$$

Note that, using (1.6), the foregoing imply $r_{ij}\geqslant 0$ for all i,j.

Next, let i_1 be the largest index j such that $r_{1j}>0$. If $i_1<n$, there must be at least one index $j>i_1$ such that $r_{i_1+1,j}>r_{i_1,j}$. Otherwise, by (2.3), we have $r_{i_1+1,j}\leqslant r_{i_1,j}$ for all j. By (1.7), this means $r_{i_1+i,j}=r_{i_1,j}$ for all j, especially for $j=1$. By (1.8), $r_{1,i_1+1}=r_{i_1+1,1}=r_{i_1,1}=r_{1,i_1}$, contradicting the definition of i_1.

If $i_1<n$, let $i_2>i_1$ be the largest index j such that $r_{i_1+1,j}>r_{i_1,j}$. If $i_2<n$, let $i_3>i_2$ be the largest index j such that $r_{i_2+1,j}>r_{i_2,j}$, and so on. This produces an increasing sequence $1\leqslant i_1<i_2<\cdots<i_k<n$. Let $\varrho=(\varrho_1,\ldots,\varrho_k)$ be the consecutive partition in which

$$\varrho_1=\{1,\ldots,i_1\}, \quad \varrho_2=\{i_1+1,\ldots,i_2\},\ldots, \quad \varrho_k=\{i_k+1,\ldots n\}.$$

We contend that the matrix M^ϱ defined by (1.5) has the property that each of the inequalities (1.6), (1.9)–(1.11) satisfied as an equality by R is satisfied as an equality by M^ϱ. If $r_{1n}=0$, obviously $M^\varrho_{1n}=0$, which checks our assertion for (1.6). If $M^\varrho_{1j}>M^\varrho_{1,j+1}$, then $j=i_1$, so $r_{1j}>0$, $r_{1,j+1}=0$. Thus $r_{ij}=r_{1,j+1}$ implies $M^\varrho_{1j}=M^\varrho_{1,j+1}$, which disposes of (1.9). If $M^\varrho_{in}>M^\varrho_{i-1,n}$, then $i=i_k+1$, so $r_{in}>r_{i-1,n}$, and our contention is demonstrated for (1.10).

If $2\leqslant i\leqslant j\leqslant n-1$ and

$$M^\varrho_{ij}+M^\varrho_{i-1,j+1}>M^\varrho_{i-1,j}+M^\varrho_{i,j+1},$$

then there exists a $t<k$ such that $i=i_t+1, j=i_{t+1}$. But this means $r_{ij}>r_{i-1,j}$, $r_{i,j+1}=r_{i-1,j+1}$, so

$$r_{ij}+r_{i-1,j+1}>r_{i-1,j}+r_{i,j+1},$$

which demonstrates the contention for (1.11).

If $R=M^\varrho$ we are finished. If not, then it follows from what we have just proved that, for sufficiently small positive α, the matrix

$$\frac{1}{1-\alpha}(R-\alpha M^\varrho) \tag{2.5}$$

is also in $\mathscr{R}(n)$. Let $\alpha_1<1$ be the largest α such that the matrix given by (2.5) is in $\mathscr{R}(n)$. Let $R=R_0$, $M^\varrho=M^{\varrho^1}$

$$R_1=\frac{1}{1-\alpha_1}(R_0-\alpha_1 M^{\varrho^1});$$

then

$$R_0=(1-\alpha_1)R_1+\alpha_1 M^{\varrho^1}.$$

Note that R_1 satisfies every equality in (1.6), (1.9)–(1.11) that R_0 does, and there is at least one additional inequality in (1.6), (1.9)–(1.11) that is a strict inequality for R_0 but an equality for R_1.

Now treat R_1 in the same manner as R_0, and so on. We obtain a sequence of matrices R_t, M^{ϱ^t}

$$\left.\begin{aligned} R_0&=(1-\alpha_1)R_1+\alpha_1 M^{\varrho^1}, & 0<\alpha_1<1\\ R_1&=(1-\alpha_2)R_2+\alpha_2 M^{\varrho^2}, & 0<\alpha_2<1\\ R_{t-1}&=(1-\alpha_t)R_t+\alpha_t M^{\varrho^t}, & 0<\alpha_t<1, \end{aligned}\right\} \tag{2.6}$$

where the number of inequalities in (1.6), (1.9)–(1.11) satisfied as equalities in R_t is a strictly increasing function of t. Since there are only a finite number of such inequalities, we must finally encounter *at* such that $R_t\in\mathscr{M}^{\mathscr{P}(n)}$. Then (2.6) shows that $R=R_0$ is a convex combination of matrices in $\mathscr{M}^{\mathscr{P}(n)}$.

3. Proof of Theorem 2

Let $\mathscr{M}(y)$ be the convex hull of all vectors of the form $M^\varrho y$, $\varrho\in\mathscr{P}(n)$; our task is to show that (1.3) and (1.4) imply $x\in\mathscr{M}(y)$. Since $\mathscr{M}(y)$ is a closed convex set, it is sufficient by the hyperplane separation theorem to show that there is no vector $z\neq 0$ such that

$$(x,z)>\max_{\varrho\in\mathscr{P}(n)}(M^\varrho y,z). \tag{3.1}$$

We first show that there exists $M^\varrho z\in D(n)$ such that

$$(x,z)\leqslant(x,M^\varrho z). \tag{3.2}$$

Now it is well-known that

$$\alpha_1 \geqslant \cdots \geqslant \alpha_s, \quad b_1 \leqslant \cdots \leqslant b_s \quad \text{implies} \tag{3.3}$$
$$\sum \alpha_i b_i \leqslant \frac{1}{s} \sum b_i \sum \alpha_i .$$

If $z \in D(n)$, (3.2) follows from $\varrho = (\varrho_1, \dots, \varrho_n)$, $\varrho_t = \{t\}$.
If $z \notin D(n)$, define $\varrho = (\varrho_1, \dots, \varrho_k)$ by the rule

$$z_{|\varrho_1|} > z_{|\varrho_1|+1},$$
$$z_{|\varrho_1|+|\varrho_2|} > z_{|\varrho_1|+|\varrho_2|+1},$$
$$\dots$$
$$z_{|\varrho_1|+\cdots+|\varrho_k|} > z_{|\varrho_1|+\cdots+|\varrho_k|+1},$$

$z_i \leqslant z_{i+1}$ for all other values of $i \leqslant n-1$. Then

$$(x, z) \leqslant (x, M^\varrho z), \tag{3.4}$$

by (3.3). If $M^\varrho z \in D(n)$, we are finished. If not, treat $M^\varrho z$ as z was treated, leading to a partition σ and a matrix M^σ. Note that the partition ϱ is a refinement of the partition σ so that $M^\sigma M^\varrho = M^\sigma$, and from (3.4)

$$(x, z) \leqslant (x, M^\varrho z) \leqslant (x, M^\sigma M^\varrho z) = (x, M^\sigma z).$$

Continuing in this way, we come eventually to a partition τ such that

$$(x, z) \leqslant (x, M^\tau z), \quad M^\tau z \in D(n),$$

and (3.2) is established.

Let $w = M^\tau z$. It follows from (3.1) that

$$(x, w) \geqslant (x, z) > (M^\tau y, z) = (y, M^\tau z) = (y, w), \tag{3.5}$$

where $x, y, w \in D(n)$. At this point, the argument rejoins [8]. Since $w \in D(n)$, $w = \sum_{k=1}^{n} \alpha_k u_k$, where $\alpha_k \geqslant 0$, $k = 1, \dots, n-1$. But then (3.5) is incompatible with (1.3) and (1.4), so (3.1) is false and the theorem is proven.

We now offer a second proof of theorem 2. Let $y \in D(n)$, and let $K(y)$ be the set of all vectors x such that $x \in D(n)$ and (1.3) and (1.4) hold. It is sufficient to prove that the extreme points of the polyhedron $K(y)$ are precisely the vectors of the form $M^\varrho y$, $\varrho \in \mathscr{P}(n)$.

We first show that every $M^\varrho y$ is extreme. Let $\varrho = (\varrho_1, \dots, \varrho_k)$, and assume $x = M^\varrho y = -\frac{1}{2}(z + w)$, z, $w \in K(y)$. Since

$$x_1 + \cdots + x_{|\varrho_1|+\cdots+|\varrho_t|} = y_1 + \cdots + y_{|\varrho_1|+\cdots+|\varrho_t|}$$

for all t it follows that

$$z_1 + \cdots + z_{|\varrho_1| + \cdots + |\varrho_t|} = x_1 + \cdots + x_{|\varrho_1| + \cdots + |\varrho_t|} \tag{3.6}$$

for all t. This implies

$$\sum_{i \in \varrho_t} z_i = \sum_{i \in \varrho_t} x_i \quad \text{for all} \quad t. \tag{3.7}$$

But since each x_i in ϱ_t is the same, it follows from $z, w \in D(n)$ that for each i, $z_i = x_i = w_i$. Hence $x = M^{\varrho} y$ is extreme.

Suppose $x \in K(y)$ cannot be expressed in the form $M^{\varrho} y$ for any $\varrho \in \mathscr{P}(n)$. This means that there exists an index $i \leqslant n-1$ such that

$$x_i > x_{i+1}, \quad \text{and} \tag{3.8}$$

$$\sum_{j=1}^{i} x_j < \sum_{j=1}^{i} y_j. \tag{3.9}$$

Let

$$A = \{j \mid x_j = x_i\}, \qquad B = \{j \mid x_j = x_{i+1}\}.$$

Suppose

$$\sum_{t=1}^{j} x_t = \sum_{t=1}^{j} y_t \quad \text{for some} \quad j \in A. \tag{3.10}$$

Since $x_{j+1} = x_j$ and $y_{j+1} \leqslant y_j$, we would have (in view of (1.3))

$$\sum_{t=1}^{j+1} x_t = \sum_{t=1}^{j+1} y_t. \tag{3.11}$$

Continuing, we would contradict (3.9). Thus (3.10) is false and

$$\sum_{t=1}^{j} x_t < \sum_{t=1}^{j} y_t \quad \text{for all} \quad j \in A. \tag{3.11}$$

A similar argument shows that

$$\sum_{t=1}^{j} x_t < \sum_{t=1}^{j} y_t \quad \text{for all} \quad j \in B, \quad j \leqslant i + |B| - 1. \tag{3.12}$$

Let $\varepsilon > 0$ be chosen so that $|A|\,|B|\,\varepsilon$ is smaller than the difference between right and left hand sides of all inequalities (3.11) and (3.12) and also $x_{i-|A|} - x_{i-|A|+1}$ (if $i - |A| \geqslant 1$) and $x_{i+|B|} - x_{i+|B|+1}$ (if $i + |B| \leqslant n-1$). Define $z = (z_1, \ldots, z_n)$ and $w = (w_1, \ldots, w_n)$ by the rule

$$\begin{aligned} &z_j = x_j = w_j && \text{if} \quad j \notin A, \quad j \notin B, \\ &z_j = x_j + |B|\,\varepsilon, \quad w_j = x_j - |B|\,\varepsilon && \text{if} \quad j \in A, \\ &z_j = x_j - |A|\,\varepsilon, \quad w_j = x_j + |A|\,\varepsilon && \text{if} \quad j \in B, \end{aligned}$$

Then the preceding stipulations show $z, w \in K(y)$ and

$$x = \tfrac{1}{2}(z + w).$$

Thus x is not extreme.

The third proof is due to D. Ž. Djoković. If $n=1$, the theorem is obviously true, and we use induction.

Case 1. Let $(x, u_k) = (y, u_k)$ for some $k<n$. Let x' (x'') be the vector in R^k (R^{n-k}) whose components coincide with the first k (last $n-k$) components of x. Similarly we define y' and y''. The pairs x', y' and x'', y'' satisfy the conditions analogous to (1.3) and (1.4). By inductive hypothesis $x' = R'y'$, $x'' = R''y''$ where $R' \in \mathscr{R}(k)$, $R'' \in \mathscr{R}(n-k)$. We infer that $x = Ry$ where R is the direct sum $R = R' \oplus R''$. It is easy to see that $R \in \mathscr{R}(n)$. Indeed, we can represent R' and R'' as convex combinations

$$R' = \sum \alpha_\varrho M'^{\varrho}, \qquad R'' = \sum \beta_\sigma M''^{\sigma}$$

where $M'^{\varrho} \in \mathscr{M}^{\mathscr{P}(k)}$, $M''^{\sigma} \in \mathscr{M}^{\mathscr{P}(n-k)}$.

Then

$$R = \sum_{\varrho, \sigma} \alpha_\varrho \beta_\sigma (M'^{\varrho} \oplus M''^{\sigma}) \in \mathscr{R}(n).$$

This disposes with case 1.

Case 2. Let $(x, u_k) < (y, u_k)$ for all $k = 1, \ldots, n-1$. Let R_0 be the matrix whose all entries are $1/n$, and $z = R_0 y$. Then

$$x_1 \geqslant \frac{1}{n} \sum_{i=1}^{n} x_i = z_1 = \cdots = z_n.$$

Therefore there is a unique $\alpha \in (0, 1]$ such that the vector $v = \alpha z + (1-\alpha) y$ satisfies $(x, u_k) \leqslant (v, u_k)$ for all k, with equality for $k=n$ and at least one another k. By Case 1 there exists $R \in \mathscr{R}(n)$ such that

$$x = Rv = R(\alpha R_0 y + (1-\alpha) y) = [\alpha R_0 + (1-\alpha) R]\, y$$

which proves the theorem.

REFERENCES

[1] Bellman, R. and Hoffman, A., *On a Theorem of Ostrowski and Taussky*, Arch. Math. Basel *5*, 123–127 (1954).

[2] Birkhoff, G., *Tres observaciones sobre el algebra lineal*, Univ. Nac. Tucuman Rev. [Ser. A] *5*, 147–150 (1946).

[3] Hardy, G. H., Littlewood, J. E., and Polya, G., *Some Simple Inequalities Satisfied by Convex Functions*, Messenger of Math. *58*, 145–152 (1929); *Inequalities* (Cambridge 1934).

[4] Hoffman, A. J., *On an Inequality of Hardy, Littlewood and Polya*, Nat. Bur. Standards, Report No. 2977 (1953).

[5] MIRSKY, L., *Results and Problems in the Theory of Doubly-Stochastic Matrices*, Z. Wahrscheinlichkeitstheorie und verw. Gebiete *1*, 319–334 (1963).
[6] OSTROWSKI, A., *Sur quelques applications des fontions convexes et concaves au sens de I. Schur*, J. Math. Pures Appl. *31*, 253–292 (1952).
[7] OSTROWSKI, A. and TAUSSKY, O., *On the Variation of the Determinant of a Positive Definite Matrix*, Nederl. Akad. Wetensch. Proc. [Ser. A] *54*, 383 (1951).
[8] RADO, R., *An Inequality*, J. London Math. Soc. *27*, 1–6 (1952).
[9] GUILLEMIN, E. A., *On the Analysis and Synthesis of Single-Element-Kind Networks*, IRE Trans., CT-7, 303–312 (1960).
[10] WEINBERG, L., *A Survey of Linear Graphs: Fundamentals and Applications to Network Theory*, Matrix Tensor Quart. 81–89, 103–115 (1964).

Thomas J. Watson Research Center

The Minimum Value of a Definite Integral II

L. J. MORDELL (Cambridge, England)

Dedicated to A. Ostrowski on his 75th Birthday

In a recent[1]) note, I gave a very simple proof of Bowman's result (BOWMAN and GERARD, *Higher Calculus* (Camb. Univ. Press 1967), p. 327) that $1/(n+1)$ is the minimum value for real $a_1, \ldots, a_n$ of

$$\int_0^\infty e^{-x}(1 + a_1 x + \cdots + a_n x^n)^2 \, dx .$$

The method can be generalized to find the minimum value M for real a_i of some integrals of the form

$$\int_q^p f(x)\,(a_0 + a_1 x + \cdots + a_n x^n)^2 \, dx , \tag{1}$$

where $f(x)$ is such that the integrals $\int_q^p f(x)\, x^r dx \, (r \geqslant 0)$, exist and the coefficient a_l of the term $a_l x^l$ in the bracket is given as 1. We need not hereafter mention the limits p, q which are fixed for given $f(x)$.

A minimum actually exists and by the usual process, the a's are determined by the equations

$$\int f(x)\, x^r (a_0 + a_1 x + \cdots + a_n x^n) \, dx = 0 , \quad (0 \leqslant r \leqslant n, r \neq l). \tag{2}$$

On multiplying (2) by a_r, summing for r, and subtracting from (1), we have

$$M = \int f(x)\, x^l (a_0 + a_1 x + \cdots + a_n x^n) \, dx . \tag{3}$$

Suppose now that

$$\int f(x)\, x^s dx = g(s)/h(s), \tag{4}$$

where $g(s)$, $h(s)$ are polynomials in s. Then, from (2),

$$a_0 \frac{g(r)}{h(r)} + a_1 \frac{g(r+1)}{h(r+1)} + \cdots + a_n \frac{g(r+n)}{h(r+n)} = 0 , \quad (r = 0, 1, \ldots l-1, l+1, \ldots n) \tag{5}$$

---

[1]) *The Minimum Value of a Definite Integral,* Mathematical Gazette **52** (1968), 135–136.

Received December 19, 1967

Hence we have the identity in z

$$\left.\begin{aligned} a_0 \frac{g(z)}{h(z)} + a_1 \frac{g(z+1)}{h(z+1)} + \cdots + a_n \frac{g(z+n)}{h(z+n)} = \\ = \frac{z(z-1)\ldots(z-l+1)(z-l-1)\ldots(z-n)}{h(z)\,h(z+1)\ldots h(z+n)} \, k F(z), \end{aligned}\right\} \tag{6}$$

where $F(z)$ is a polynomial in z, and k is independent of z and is inserted since in the applications, $F(z)$ is easily determined except for a constant factor. Also, from (3) and (4), on putting $z=l$ in (6);

$$M = \frac{(-1)^{n-l}\, l!\,(n-l)!\, k F(l)}{h(l)\,h(l+1)\ldots h(l+n)}. \tag{7}$$

We now give some applications. Take first

$$f(x) = x^m,\ p = 1,\ q = 0,\ m > -1\,.$$

Now

$$\int_0^1 x^{m+r}\, dx = \frac{1}{m+r+1},$$

and so from (6),

$$\left.\begin{aligned} \frac{a_0}{m+z+1} + \frac{a_1}{m+z+2} + \cdots + \frac{a_n}{m+z+n+1} = \\ = \frac{k z(z-1)\ldots(z-l+1)(z-l-1)\ldots(z-n)}{(m+z+1)(m+z+2)\ldots(m+z+1+n)}. \end{aligned}\right\} \tag{8}$$

For clearly $F(z)=1$, since on multiplying out, both sides of (8) are polynomials in z of degree n. The coefficients $a_0, a_1, \ldots$ are given by splitting the right hand side into partial fractions and expressing the condition that $a_l=1$. Multiply by $m+z+l+1$ and put $z=-m-l-1$. Then

$$1 = \frac{k(-1)^{n-l}(m+l+1)\ldots(m+l+1+n)}{l!\,(n-l)!\,(m+2l+1)}.$$

From (7),

$$M = \frac{k(-1)^{n-l}\, l!\,(n-l)!}{(m+l+1)\ldots(m+l+1+n)},$$

and so

$$M = \frac{\{l!\,(n-l)!\}^2\,(m+2l+1)}{\{(m+l+1)\ldots(m+l+1+n)\}^2}. \tag{9}$$

The case $l=m=0$, which gives $M=1/(n+1)^2$, was given me by BOWMAN.

There was no difficulty here in expressing the condition that $a_l=1$. As $F(z)$ is easily found in the applications, the real problem is to find the coefficients $a_0, a_1, \dots a_l, \dots$ in the expansion of the right hand side of (6) given by the left hand side. This is quite simple for the case

$$f(x)=e^{-x}x^m,\ p=\infty,\ q=0,\ m>-1\,.$$

Since

$$\int_0^\infty e^{-x}x^{m+r}\,dx=\Gamma(m+r+1),$$

this case is not of the type (4), but the method applies. For (5) becomes

$$a_0\Gamma(m+r+1)+\cdots+a_n\Gamma(m+r+1+n)=0,$$

or, after dividing by $\Gamma(m+r+1)$,

$$a_0+a_1(m+r+1)+\cdots+a_n(m+r+1)\,..\,(m+r+n)=0\,.$$

Then we have the identity,

$$\left.\begin{aligned}&a_0+a_1(m+z+1)+\cdots+a_n(m+z+1)\dots(m+z+n)=\\&\qquad=kz(z-1)\dots(z-l+1)(z-l-1)\dots(z-n),\end{aligned}\right\}\qquad(10)$$

since both sides are polynomials of degree n (the factor z is omitted when $l=0$, and the factor $z-n$ when $l=n$).

Also we find

$$M=(-1)^{n-l}k\Gamma(m+l+1)\,l!(n-l)!\,.$$

In (10), change z into $z-m$. Then identically

$$\left.\begin{aligned}&a_0+a_1(z+1)+\cdots+a_n(z+1)\dots(z+n)=\\&=k(z-m)\dots(z-m-l+1)(z-m-l-1)\dots(z-m-n)=kG(z),\end{aligned}\right\}\qquad(11)$$

say. We show now that

$$l!\,a_l=k\left\{G(-1)-lG(-2)+\frac{l(l-1)}{2!}G(-3)\,..\right\}=k\{\Delta^l G(z)\}_{z=-1}\,,$$

where $\Delta G(z)=G(z)-G(z-1)$.

Put $z=-1$ in (11) and so $a_0=kG(-1)$. Change z into $z-1$ in (11), subtract and put $z=-2$. Then $a_1=k(G(-1)-G(-2))$. The value of a_l follows easily by successive differences. Hence, if $a_l=1$,

$$l!=k\{\Delta^l G(z)\}_{z=-1}\,,$$

and so

$$\frac{M}{l!}=\frac{(-1)^{n-l}\Gamma(m+l+1)\,l!(n-l)!}{\{\Delta^l G(z)\}_{z=-1}},$$

The result takes a simple form when $l=0$ or n. When $l=0$ put $z=-1$ in (11) and so

$$1 = k(-1)^n (2+m)\dots(1+n+m).$$

Also

$$M = k(-1)^n \Gamma(m+1)\, n!,$$

and so

$$M = \frac{n!\,\Gamma(m+1)}{(2+m)\dots(1+n+m)}.$$

When $m=0$, this is BOWMAN's result. When $l=n$, equate coefficients of z^n in (11). Then $1=k$; also $M=k\Gamma(m+n+1)\, n!$ and so

$$M = n!\,\Gamma(m+n+1).$$

Another illustration is given by

$$\int_0^1 x^{a-1}(1-x)^{b-1}(a_0 + a_1 x + \dots + a_n x^n)^2\, dx,$$

where $a>0$ and b is a positive integer. Then

$$\int_0^1 x^{a-1+r}(1-x)^{b-1}\, dx = \frac{\Gamma(a+r)\,\Gamma(b)}{\Gamma(a+r+b)} = \frac{\Gamma(b)}{(a+r)\dots(a+r+b-1)},$$

and so the denominator is a polynomial of degree b in r.
Hence (6) becomes

$$\frac{a_0\Gamma(b)\,\Gamma(a+z)}{\Gamma(a+z+b)} + \dots + \frac{a_n\Gamma(b)\,\Gamma(a+z+n)}{\Gamma(a+z+n+b)} =$$

$$= \frac{kz(z-1)\dots(z-l+1)(z-l-1)\dots(z-n)}{(a+z)\dots(a+z+b-1+n)},$$

since on multiplying out, we get two polynomials of degree n. There seems now no simple expression for a_l except when $l=0$ or n.
When $l=0$, multiply out by $a+z$ and put $z=-a$. Then, with $a_0=1$, we find

$$1 = (-1)^n k(a+1)\dots(a+n)/1.2\dots(b+n-1).$$

Also

$$M = \frac{(-1)^n k n!}{a(a+1)\dots(a+b+n-1)}.$$

Hence

$$M = \frac{n!(n+b-1)!}{(a+1)\ldots(a+n)\,a(a+1)\ldots(a+b+n-1)}.$$

When $l=n$, multiply out by $a+b+z-1+n$ and put $z=1-n-a-b$. Then

$$(-1)^{b-1} = \frac{k(-1)^n(a+b+n-1)\ldots(a+b+2n-2)}{(-1)^{b+n-1}(b+n-1)!}.$$

Also

$$M = \frac{kn!}{(a+n)\ldots(a+b+2n-1)},$$

and so

$$M = \frac{n!(n+b-1)!}{(a+b+n-1)\ldots(a+b+2n-2)(a+n)\ldots(a+b+2n-1)}.$$

St. John's College

On Asymptotically Regular Solutions of a Linear Functional Equation

R. R. COIFMAN and M. KUCZMA (Chicago, U.S.A. and Katowice, Poland)

Dedicated to Professor A. M. Ostrowski on the occasion of his 75th birthday

In the present paper we consider the linear functional equation

$$\varphi[f(x)] - \varphi(x) = h(x), \tag{1}$$

in which $\varphi(x)$ is the unknown function, and $f(x)$, $h(x)$ are given functions defined on a fixed open interval $(0, a)$, $0<a\leqslant\infty$. (All functions considered in the present paper are real-valued functions of a real variable.) Equation (1) belongs to the most important functional equations in a single variable and has been dealt with (as well as its particular cases) by many authors (cf. [4], in particular Chapters II and V).

We shall be concerned with continuous solutions $\varphi(x)$ of (1) such that the limit

$$\lim_{x\to 0+} \{\varphi(\lambda x) - \varphi(x)\} \tag{2}$$

exists for every $\lambda>0$. Such solutions will be called *asymptotically regular*. The object of the present paper is to give a necessary and sufficient condition for the existence of a continuous and asymptotically regular solution of equation (1) (it turns out that such a solution is unique up to an additive constant). This result is related to a result of the second author [3] concerning monotonic solutions of equation (1).

Asymptotically regular functions were investigated by J. KARAMATA [2]. We shall make use of the following lemma which expresses a basic property of convergence (2).

LEMMA 1 (cf. [2]). *Let $F(x)$ be a measurable function on $(0, a)$ such that the limit*

$$\lim_{x\to 0+} \{F(\lambda x) - F(x)\} = L(\lambda) \tag{3}$$

exists for every $\lambda>0$. Then there exists a constant c such that

$$L(\lambda) = c\log\lambda,$$

and convergence (3) is uniform in λ on every compact interval $[\lambda_1, \lambda_2]\subset(0, \infty)$.

COROLLARY. Under conditions of lemma 1, for every sequence $x_n>0$, $x_n\to 0$, and for every sequence $\lambda_n\to\lambda_0>0$, we have

$$\lim_{n\to\infty} \{F(\lambda_n x_n) - F(x_n)\} = c\log\lambda_0.$$

Received December 19, 1967 and, in revised form March 14, 1968

Concerning $f(x)$ we shall assume that it is continuous and strictly increasing in $(0, a)$ and fulfils the conditions

$$0 < f(x) < x \quad \text{for} \quad x \in (0, a), \tag{4}$$

$$\lim_{x \to 0+} f(x)/x = s, \qquad 0 < s < 1. \tag{5}$$

We shall say that $f(x)$ fulfilling (4) and (5) is *regular with respect to iteration* iff the discrete group of natural iterates $f^n(x)$ of $f(x)$:

$$f^0(x) = x, \quad f^{n+1}(x) = f[f^n(x)], \qquad n = 0, \pm 1, \pm 2, \dots,$$

can be imbedded in a continuous group $f^u(x)$, $u \in (-\infty, \infty)$:

$$f^{u+v}(x) = f^u[f^v(x)], \qquad u, v \in (-\infty, \infty), \quad f^1(x) = f(x),$$

such that

$$\lim_{x \to 0+} f^u(x)/x = s^u \quad \text{for} \quad u \in (-\infty, \infty)$$

(cf. e.g. [1], [5], [4; Chapter IX]). The following lemma gives a necessary and sufficient condition for $f(x)$ to be regular with respect to iteration.

LEMMA 2 (cf. [1], [5], [4; Chapter IX]). *A continuous and strictly increasing function $f(x)$ fulfilling* (4) *and* (5) *is regular with respect to iteration if and only if for some $x_0 \in (0, a)$ the limit*

$$\lim_{n \to \infty} f^n(x)/f^n(x_0) = R(x) \tag{6}$$

exists and $R(x)$ is continuous and strictly increasing in $(0, a)$.

The function $R(x)$ occurring in the above lemma satisfies the Schröder equation

$$R[\ (x)] = sR(x) \tag{7}$$

in $(0, a)$, and the regular iterates $f^u(x)$ of $f(x)$ are given by

$$f^u(x) = R^{-1}(s^u R(x)). \tag{8}$$

Moreover, there exists a constant $\eta \neq 0$ such that for every $\lambda > 0$

$$\lim_{x \to 0+} R(\lambda x)/R(x) = \lambda^\eta \tag{9}$$

(cf. [2; Theorem III]). In other words, the function $\log R(x)$ is asymptotically regular (it follows from (6), (7) and from the strict monotonicity of $R(x)$ that $R(x) > 0$ in $(0, a)$).

The result of the present paper is contained in the following

THEOREM. *Let $f(x)$ be continuous, strictly increasing and regular with respect to* *iteration on* $(0, a)$ *and let it fulfil conditions* (4) *and* (5). *Let* $h(x)$ *be continuous on* $(0, a)$ *and suppose that there exists the limit*

$$\lim_{x \to 0+} h(x) = \gamma. \tag{10}$$

Then equation (1) *has a continuous and asymptotically regular solution* $\varphi(x)$ in $(0, a)$ *if and only if the series*

$$\sum_{n=0}^{\infty} \{h[f^n(x)] - h[f^n(x_0)]\} \tag{11}$$

converges for some $x_0 \in (0, a)$ *uniformly in* x *on every compact subinterval of* $(0, a)$. *If this condition is fulfilled, then the solution is given by*

$$\varphi(x) = \varphi(x_0) + \frac{\gamma}{\log s} \log R(x) - \sum_{n=0}^{\infty} \{h[\ ^n(x)] - h[f^n(x_0)]\}, \tag{12}$$

where $R(x)$ *is given by* (6), *and thus the solution is determined uniquely by its value at* x_0[1]).

Proof. Suppose that $\varphi(x)$ is a continuous and asymptotically regular solution of equation (1) in $(0, a)$. Let us take an arbitrary sequence $x_n > 0$ tending to zero. In view of the corollary to lemma 1 (we take $F(x) = \varphi(x)$, $\lambda_n = f(x_n)/x_n$) we have

$$\lim_{n \to \infty} \{\varphi[f(x_n)] - \varphi(x_n)\} = c \log s,$$

since by (5) $\lim_{n \to \infty} f(x_n)/x_n = s$. This implies, the sequence x_n being arbitrary, that

$$\lim_{x \to 0+} \{\varphi[f(x)] - \varphi(x)\} = c \log s,$$

which gives by (1) and (10) $c \log s = \gamma$ and

$$c = \gamma/\log s. \tag{13}$$

Making use once more of the corollary to lemma 1, with the specialization

$$F(x) = \varphi(x), \quad x_n = \ ^n(x_0), \quad \lambda_n = f^n(x)/f^n(x_0),$$

we obtain in view of (6) and (13)

$$\lim_{n \to \infty} \{\varphi[f^n(x)] - \varphi[f^n(x_0)]\} = \frac{\gamma}{\log s} \log R(x). \tag{14}$$

---

1) It is readily seen that if limit (6) exists for a certain $x_0 \in (0, a)$, then it exists for every $x_0 \in (0, a)$ (cf. e.g. [4; Chapter VI]). Similarly, if series (11) converges for a certain $x_0 \in (0, a)$, then it converges for any x_0. Therefore there is no restriction of the generality in assuming that x_0 occurring in (6) and (11) is the same.

It results from lemma 1 that (14) holds uniformly in x on every compact subinterval of $(0, a)$.

On the other hand, we have from (1) by induction

$$\varphi[f^n(x)] = \varphi(x) + \sum_{i=0}^{n-1} h[f^i(x)], \qquad n = 1, 2, 3, \ldots, \tag{15}$$

whence

$$\varphi[f^n(x)] - \varphi[f^n(x_0)] = \varphi(x) - \varphi(x_0) + \sum_{i=0}^{n-1} \{h[f^i(x)] - h[f^i(x_0)]\}. \tag{16}$$

The uniform convergence of series (11) results from the uniform convergence in (14). Formula (12) is a consequence of (16) and (14) and the uniqueness statement follows from (12).

Now suppose that series (11) converges uniformly on every compact subinterval of $(0, a)$. Then we may define the function $\varphi(x)$ by (12) (where $\varphi(x_0)$ may be regarded as an arbitrary constant). $\varphi(x)$ is evidently continuous in $(0, a)$ and it follows from (7) and (10) that $\varphi(x)$ satisfies equation (1) in $(0, a)$. It remains to prove that $\varphi(x)$ is asymptotically regular.

Formula (16), resulting directly from the fact that $\varphi(x)$ satisfies equation (1) in $(0, a)$, shows now that $\varphi[f^n(x)] - \varphi[f^n(x_0)]$ converges as $n \to \infty$ uniformly on every compact subinterval of $(0, a)$. We have by (8)

$$f^n(x) = R^{-1}(s^n R(x))$$

and, since by (6) $R(x_0) = 1$,

$$f^n(x_0) = R^{-1}(s^n).$$

Consequently the sequence

$$\varphi[R^{-1}(s^n R(x))] - \varphi[R^{-1}(s^n)]$$

converges as $n \to \infty$ uniformly on every compact subinterval of $(0, a)$. Setting $R(x) = \lambda$, $\varphi[R^{-1}(x)] = F(x)$, we get the existence of the limit

$$\lim_{n \to \infty} \{F(\lambda s^n) - F(s^n)\}, \tag{17}$$

and the convergence is uniform in λ on every compact subinterval of $(0, \lambda^*)$, where $\lambda^* = \lim_{x \to a-} R(x)$. But we may write (17) as

$$\lim_{n \to \infty} \{[F((\lambda s^{-k}) s^{n+k}) - F(s^{n+k})] + [F(s^{n+k}) - F(s^n)]\}$$

with an arbitrary positive integer k, and this shows that the convergence in (17) is uniform in λ on every compact subinterval of $(0, \infty)$, since the limit

$$\lim_{n \to \infty} [F(s^{n+k}) - F(s^n)] = \lim_{n \to \infty} \{\varphi[f^{n+k}(x_0)] - \varphi[f^n(x_0)]\}$$
$$= \lim_{n - \infty} \sum_{i=n}^{n+k-1} h[f^i(x_0)] = k\gamma$$

(cf. (15)) obviously exists. Consequently there exists also limit (3), for every $\lambda > 0$, which shows that $F(x)$ is asymptotically regular.

Now let $y_n > 0$ be an arbitrary sequence tending to zero and let us take an arbitrary $\lambda > 0$. We put

$$x_n = R(y_n), \quad \lambda_n = R(\lambda y_n)/R(y_n).$$

Then $x_n \to 0+$ and by (9) $\lim_{n\to\infty} \lambda_n = \lambda^\eta > 0$. In virtue of the corollary to lemma 1 the limit

$$\lim_{n\to\infty} \{F(\lambda_n x_n) - F(x_n)\} = \lim_{n\to\infty} \{\varphi(\lambda y_n) - \varphi(y_n)\}$$

exists. Since the sequence $y_n \to 0+$ has been arbitrary, this shows that limit (2) exists for every $\lambda > 0$, i.e. $\varphi(x)$ is asymptotically regular. This completes the proof.

REFERENCES

[1] COIFMAN, R. R., *Sur l'unicité des solutions de l'équation d'Abel-Schröder et l'itération continue*, J. Austral. Math. Soc. *5*, 36–47 (1965).

[2] KARAMATA, J., *Sur un mode de croissance régulière des fonctions*, Mathematica, Cluj, *4*, 38–53 (1930).

[3] KUCZMA, M., *Sur une équation fonctionnelle*, Mathematica, Cluj, *3* (26), 79–87 (1961).

[4] KUCZMA, M., *Functional Equations in a Single Variable* (P.W.N., Warszawa 1968 [Monografie Matematyczne, 46)].

[5] LUNDBERG, A., *On Iterated Functions with Asymptotic Conditions at a Fixpoint*, Ark. Mat. *5*, 193–206 (1963).

University of Chicago, U.S.A., and
Jagello University, Katowice Poland.

Interpolation by Analytic Functions of Bounded Growth[1])

DAVID G. CANTOR (Los Angeles), D. L. HILLIKER (Irvine) and E. G. STRAUS
(Los Angeles, California, U.S.A.)

To A. Ostrowski on his 75th birthday

1. Introduction

In this paper we examine those sequences in a complex domain on which arbitrarily assigned values can be assumed by functions which are meromorphic or holomorphic in a certain domain and satisfy growth conditions. In §2 we discuss interpolation by meromorphic functions and obtain exact conditions for interpolation by functions of bounded characteristic in the unit disk and for interpolation by meromorphic functions of bounded order in the plane. In § 3 we consider interpolation by entire functions and obtain bounds on the interpolating function in terms of the sequence of points and the sequence of values. Finally, in § 4, we obtain a necessary and sufficient condition on a sequence so that all values whose growth rate is compatible with interpolation by entire functions of order $\leqslant \varrho$ can be so interpolated. The result is completely analogous to CARLESON's theorem ([1], p. 196) for interpolation by bounded functions in the unit disk. Since the functions of order $\leqslant \varrho$ do not form a Banach space the method of proof is different, though the arguments are simpler.

Throughout the paper we use ε for any arbitrarily small positive number and c for any sufficiently large constant, which may depend on the ε in the expression (if any) but not on any other variables.

2. Interpolation by Meromorphic Functions

In this section we wish to investigate the bounds on functions meromorphic in a domain D with arbitrarily prescribed values (w_n) at a sequence (z_n) of points in D. To be specific, let (z_n) be a sequence of distinct points of D with finite multiplicities (m_n). If the z_n have no limit points in D then there exists a function $P(z)$ analytic in D with zeros of multiplicity m_n at the points z_n. Let $D=\bigcup_r D_r$ where the D_r are bounded domains and $\bar{D}_{r_1}\subset D_{r_2}$ for $r_1<r_2$, where $\bar{D}_r$ is the closure of D_r. We associate with $P(z)$ a growth function

$$M(r)=\max\{\max_{z\in D_r}|P(z)|, 1\}.$$

[1]) The preparation of this paper was sponsored in part by NSF Grant GP-5497.

Received December 20, 1967

The main result of this section can now be stated as follows.

2.1. THEOREM. *Given any complex values* $(w_{n,\mu})$; $\mu=1,\ldots,m_n$; $n=1,2,\ldots$ *There exists a function* $f(z)$ *meromorphic in* D *so that*

$$f^{(\mu-1)}(z_n)=w_{n,\mu};\quad \mu=1,\ldots,m_n;\quad n=1,2,\ldots \tag{2.2}$$

and $f(z)=g(z)/h(z)$, *where* g, h *are analytic in* D *and*

$$|g(z)|\leqslant M(r)^2;\quad |h(z)|\leqslant M(r)\quad \text{for}\quad z\in D_r.$$

Proof. As usual we construct a function $A(z)$ meromorphic in D with poles of order $\leqslant m_n$ at z_n and principal part

$$Q(z,z_n)=\frac{u_{n1}}{z-z_n}+\frac{u_{n2}}{(z-z_n)^2}+\cdots+\frac{u_{n\,m_n}}{(z-z_n)^{m_n}}$$

where the $u_{n\mu}$ are determined so that

$$\frac{d^{\mu-1}}{dz^{\mu-1}}(P(z)\,Q(z,z_n))|_{z=z_n}=w_{n\mu};\quad \mu=1,\ldots,m_n. \tag{2.3}$$

In order to construct $A(z)$ we first construct a sequence (z_n^*) in D where $z_n^*\neq z_n$ but $|z_n^*-z_n|$ is sufficiently small to assure the convergence of

$$A(z)=\sum_{n=1}^{\infty}(Q(z,z_n)-Q(z,z_n^*)) \tag{2.4}$$

for all $z\in D$, $z\neq z_n$, $z\neq z_n^*$ and the convergence of

$$P^*(z)=P(z)\prod_{n=1}^{\infty}\left(\frac{z-z_n^*}{z-z_n}\right)^{m_n}. \tag{2.5}$$

We carry this construction out in the following two lemmas.

2.6. LEMMA. *There exists a sequence of neighborhoods* U_n of z_n *in* D *so that for* $z_n^*\in U_n$ *the product* $P^*(z)$ *in* (2.5) *exists and satisfies* $|P^*(z)|\leqslant 2\,M(r)$ *for* $z\in D_r$.

Proof. We define U_n by induction as $U_n=\{z\,\|z-z_n|<\delta_n\}$, so that

$$P_N^*(z)=P(z)\prod_{n=1}^{N}\left(\frac{z-z_n^*}{z-z_n}\right)^{m_n};\quad z_n^*\in U_n;\quad N=0,1,2,\ldots$$

satisfies $|P_N^*(z)|\leqslant(2-2^{-N})\,M(r)$ for $z\in D_r$. Now assuming that we have chosen δ_n for $n<N$, we can choose δ_N so small that $U_N\subset D$ and

$$\left|\frac{z-z_N^*}{z-z_N}\right|^{m_N}<\left(1+\frac{\delta_N}{|z-z_N|}\right)^{m_N}<\frac{2-2^{-N}}{2-2^{-N+1}} \tag{2.7}$$

for all z with $|z-z_N|\geqslant r_N$ where $r_N\leqslant 1$ is less than the distance of z_N to the boundary of D. For $|z-z_N|\leqslant r_N$, $z\in D$ we have

$$|P_N^*(z)-P_{N-1}^*(z)|\leqslant|P_{N-1}^*(z)|\left|\left(1+\frac{\delta_N}{|z-z_N|}\right)^{m_N}-1\right|<\delta_N 2^{m_N}\left|\frac{P_{N-1}^*(z)}{(z-z_N)^{m_N}}\right|<2^{-N} \tag{2.8}$$

by applying the maximum modulus principle to the analytic function $P_{N-1}^*(z)/(z-z_N)^{m_N}$ and choosing

$$\delta_N<2^{-N-m_N}r_N^{m_N}/\max_{|z-z_N|=r_N}|P_{N-1}^*(z)|.$$

Combining (2.7) and (2.8) we get

$$|P_N^*(z)|\leqslant\max\left\{\frac{2-2^{-N}}{2-2^{-N+1}}\cdot(2-2^{-N+1})\,M(r),(2-2^{-N+1})\,M(r)+2^{-N}\right\}$$
$$\leqslant(2-2^{-N})\,M(r)\quad\text{for all}\quad z\in D_r.$$

Thus (2.7) and (2.8) show that $P_N^*(z)$ converges uniformly to $P^*(z)$ in every D_r and that $|P^*(z)|\leqslant 2M(r)$ for $z\in D_r$ for all choices of z_n^* in U_n.

2.9. LEMMA. *We can choose the neighborhoods U_n of Lemma 2.6 so small that*

$$g(z)=P(z)\,P^*(z)\,A(z) \tag{2.10}$$

satisfies $|g(z)|<M(r)^2$ *for* $z\in D_r$ *whenever* $z_n^*\in U_n$ *for all* n.

Proof. As in the proof of Lemma 2.6 we restrict δ_n successively so that for all $z_n^*\in U_n$ we have

$$|Q(z,z_n)-Q(z,z_n^*)|<\frac{1}{2^{n+1}}\quad\text{for}\quad|z-z_n|\geqslant r_n \tag{2.11}$$

and

$$|P(z)\,P^*(z)|\,|Q(z,z_n)-Q(z,z_n^*)|<\frac{1}{2^{n+1}}\quad\text{for}\quad|z-z_n|\leqslant r_n. \tag{2.12}$$

Combining (2.11) and (2.12) we get

$$\left.\begin{aligned}|g(z)|&\leqslant|P(z)|\,|P^*(z)|\,\Sigma\,|Q(z,z_n)-Q(z,z_n^*)|\\&<M(r)\cdot 2M(r)\sum_{n=1}^{\infty}2^{-n-1}=M(r)^2\quad\text{for}\quad z\in D_r.\end{aligned}\right\} \tag{2.13}$$

Proof of Theorem 2.1 *resumed.* Now let $f(z)=P(z)\,A(z)=g(z)/h(z)$ where $g(z)=\frac{1}{2}P(z)\,P^*(z)\,A(z)$ and $h(z)=\frac{1}{2}P^*(z)$ as constructed in Lemmas 2.6 and 2.9.

2.14. COROLLARY. *A sequence (z_n) of distinct points in the unit disk with multiplicities (m_n) has the property that for any complex $(w_{n\mu})$; $\mu=1,\ldots,m_n$; $n=1,2,\ldots$*

we can interpolate by a meromorphic function of bounded characteristic (that is the ratio of two bounded analytic functions) if and only if

$$\Sigma m_n(1-|z_n|)<\infty \tag{2.15}$$

that is, if and only if the z_n are zeros of multiplicity m_n of a nonzero bounded analytic function (*see* [1], p. 64).

Proof. The necessity of (2.15) is immediate since we can choose $w_{11}=1$ and $w_{n\mu}=0$ for all other indices. Then the $(z_n)_2^\infty$ with multiplicities (m_n) are zeros of a nonzero function of bounded characteristic and hence zeros of a nonzero bounded function. The sufficiency follows immediately from Theorem 2.1 where $P(z)$ is the Blaschke product and $M(r)=1$.

2.16. Corollary. A sequence (z_n) of distinct complex numbers with multiplicities (m_n) has the property that for any complex $(w_{n\mu})$ we can interpolate by a meromorphic function of order $\leqslant \varrho$ if and only if

$$\Sigma' m_n|z_n|^{-\varrho-\varepsilon}<\infty \tag{2.17}$$

where Σ' is extended over all $z_n \neq 0$. In other words if and only if the z_n are the zeros with multiplicity m_n of an entire function of order $\leqslant \varrho$.

Proof. The necessity follows just as in Corollary 2.14. The sufficiency follows from Theorom 2.1 where $P(z)$ is the Hadamard product and $M(r)=c\exp(r^{\varrho+\varepsilon})$.

Theorem 2.1 also gives some information about interpolation by meromorphic functions of order ϱ and bounded type, but since we square the bound $M(r)$ the information is not as complete.

2.18. Corollary. *A sequence (z_n) of distinct complex numbers with multiplicities (m_n) has the property that for any complex $(w_{n\mu})$ we can interpolate by a meromorphic function of order and type $\leqslant(\varrho,\sigma)$ if there exists an entire function $P(z)$ with zeros of multiplicity m_n at z_n and order and type $\leqslant(\varrho,\sigma/2)$; and only if there exists such a function with order and type $\leqslant(\varrho,\sigma)$.*

3. Interpolation by Entire Functions

We again consider a sequence (z_n) of distinct complex numbers without limit points in the finite plane and with multiplicities (m_n). With this sequence we associate the following functions. First the counting functions

$$n(r)=\sum_{|z_k|\leqslant r} m_k \quad \text{and} \quad m(r)=\max_{|z_n|\leqslant r} m_n. \tag{3.1}$$

Next the Weierstrass product

$$P(z)=\prod_{n=1}^{\infty} E_{k_n}(z,z_n)^{m_n}, \tag{3.2}$$

where

$$E_{k_n}(z, z_n) = \begin{cases} z \quad \text{if} \quad z_n = 0 \\ \left(1 - \frac{z}{z_n}\right) \exp\left(\frac{z}{z_n} + \cdots + \frac{1}{k_n}\left(\frac{z}{z_n}\right)^{k_n}\right) \quad \text{if} \quad z_n \neq 0 \end{cases} \tag{3.3}$$

and the k_n are chosen so that $\Sigma' m_n |z_n|^{-k_n-1}/k_n < \infty$ where Σ' is extended over the nonzero z_n. Next the growth function

$$M(r) = \max\{\max_{|z| \leqslant r} |P(z)|, 1\}. \tag{3.4}$$

(If we choose

$$k_n < (1 + \varepsilon) \log n(\sqrt{2}\,|z_n|)/\log |z_n|$$

we get

$$M(r) < c \exp(n(2r)^{1+\varepsilon}) \tag{3.5}$$

so that we could replace $M(r)$ by the right side of (3.5) wherever it occurs in the following discussion.) Finally we construct a nondecreasing positive function $B(r)$ so that

$$\frac{1}{m_n!} |P^{(m_n)}(z_n)| \geqslant 1/B(|z_n|) \quad \text{for all} \quad n. \tag{3.6}$$

Now consider a sequence of values $(w_{n\mu})$; $\mu = 1, \ldots, m_n$; $n = 1, 2, \ldots$. With it we associate a nondecreasing function $C(r)$ so that

$$|w_{n\mu}| \leqslant (\mu - 1)!\, C(|z_n|); \qquad \mu = 1, 2, \ldots, m_n. \tag{3.7}$$

3.8. THEOREM. *There exists an entire function $f(z)$ so that*

$$^{(\mu-1)}(z_n) = w_{n\mu}, \qquad \mu = 1, \ldots, m_n; \quad n = 1, 2, \ldots$$

and

$$|f(z)| \leqslant c\{2n(2|z|)\, C(2|z|)\,(1 + B(2|z|)\, M(2|z|))^{m(2|z|)}\}^{(1+\varepsilon)\log_2 |z|}.$$

Proof. We wish to construct a Mittag-Leffler series with poles of order $\leqslant m_n$ at z_n whose principal parts are

$$Q(z, z_n) = \frac{u_{n1}}{z - z_n} + \frac{u_{n2}}{(z - z_n)^2} + \cdots + \frac{u_{n m_n}}{(z - z_n)^{m_n}}. \tag{3.9}$$

If we expand

$$P(z) = a_{n m_n}(z - z_n)^{m_n} + \cdots + a_{n\, 2m_n - 1}(z - z_n)^{2m_n - 1} + \cdots \tag{3.10}$$

then the interpolation conditions (2.3) become

$$\left.\begin{aligned} w_{n\,1} &= a_{n\,m_n} u_{n\,m_n} \\ w_{n\,2} &= a_{n\,m_n+1}\, u_{n\,m_n} + a_{n\,m_n} u_{n\,m_n-1} \\ \frac{w_{n\,3}}{2!} &= a_{n\,m_n+2}\, u_{n\,m_n} + a_{n\,m_n+1}\, u_{n\,m_n-1} + a_{n\,m_n} u_{n\,m_n-2} \\ &\cdots \\ \frac{w_{n\,m_n}}{(m_n-1)!} &= a_{n\,2\,m_n-1}\, u_{n\,m_n} + \cdots + a_{n\,m_n} u_{n\,1}\,. \end{aligned}\right\} \tag{3.11}$$

By the Cauchy inequality we have

$$|a_{n\,k}| \leqslant M(|z_n| + 1). \tag{3.12}$$

Using (3.6), (3.7) and (3.12) and solving (3.11) by successive substitutions we get

$$\left.\begin{aligned} |u_{n\,\mu}| &\leqslant B(|z_n|)\, C(|z_n|)\,(1 + B(|z_n|)\, M(|z_n| + 1))^{m_n-\mu} \\ &\leqslant B(|z_n|)\, C(|z_n|)\,(1 + B(|z_n|)\, M(|z_n| + 1))^{m(|z_n|)-\mu}. \end{aligned}\right\} \tag{3.13}$$

Thus the Mittag-Leffler series

$$A(z) = \Sigma\left(\left(\frac{z}{z_n}\right)^{l_{n\,1}} \frac{u_{n\,1}}{z - z_n} + \cdots + \left(\frac{z}{z_n}\right)^{l_{n\,m_n}} \frac{u_{n\,m_n}}{(z - z_n)^{m_n}}\right) \tag{3.14}$$

converges for $z \neq z_n$ if we choose $l_{n\,\mu} = 0$ if $z_n = 0$ and

$$\left.\begin{aligned} l_{n\,\mu} &= [\log_2^+ (|u_{n\,\mu}|\, n(2|z_n|))] + 1 \\ &\leqslant [\log_2^+ \{B(|z_n|)\, C(|z_n|)\, n(2|z_n|)\,(1 + B(|z_n|)\, M(|z_n| + 1))^{m(|z_n|)-1}\}] + 1 \\ &\underset{\text{def}}{=} L(|z_n|) \quad \text{for} \quad z_n \neq 0. \end{aligned}\right\} \tag{3.15}$$

Here, as usual $\log^+ x = \max\{0, \log x\}$.

In order to estimate $f(z) = P(z)\, A(z)$ we divide the sum on the right of (3.14) into three parts $A = \Sigma_1 + \Sigma_2 + \Sigma_3$ where Σ_1 is extended over all z_n with $|z_n| \leqslant 2|z| - 1$, $|z_n - z| \geqslant 1$; Σ_2 is extended over all z_n with $|z_n - z| < 1$ and Σ_3 is extended over all z_n with $|z_n| > 2|z| - 1$, $|z_n - z| \geqslant 1$.

Using (3.13) and (3.15) we get, with $\alpha = \min_{z_n \neq 0} |z_n|$,

$$|\Sigma_1| \leqslant n(2|z|)\,(|z|/\alpha)^{L(2|z|)}\, B(2|z|)\, C(2|z|)(1 + B(2|z|)\, M(2|z|))^{m(2|z|)-1}. \tag{3.16}$$

For large $|z|$ we can write $|z|/\alpha < 2^{(1+\varepsilon)\log_2 |z|}$ so that

$$\left.\begin{aligned} &(|z|/\alpha)^{L(2|z|)} \\ &\quad < \{2\, B(2|z|)\, C(2|z|)\, n(2|z|)\,(1 + B(2|z|)\, M(2|z|))^{m(2|z|)-1}\}^{(1+\varepsilon)\log_2 |z|}. \end{aligned}\right\} \tag{3.17}$$

It is clear that all the other terms on the right of (3.16) can be subsumed in the exponent on the right of (3.17) so that we get

$$\left.\begin{aligned} |P\,\Sigma_1| &\leqslant M(|z|)\, |\Sigma_1| \\ &< \{2\, n(2|z|)\, C(2|z|)\,(1 + B(2|z|)\, M(2z))^{m(2|z|)}\}^{(1+\varepsilon)\log_2 |z|} \end{aligned}\right\} \tag{3.18}$$

for all large $|z|$. To estimate Σ_3 we write $\Sigma_3=\Sigma_{31}+\Sigma_{32}+\cdots$ where Σ_{31} is extended over all z_n with $2|z|-1<|z_n|\leqslant 4|z|$, $|z_n-z|\geqslant 1$ and Σ_{3k} is extended over all z_n with $2^k|z|<|z_n|\leqslant 2^{k+1}|z|$ for all $k>1$. Then for $|z|\geqslant 1$, $k>1$

$$\left.\begin{aligned}|\Sigma_{3k}| &< n(2^{k+1}|z|)\max_{|z_n|>2^k|z|}\left(\frac{1}{2^k}\right)^{l_{n\mu}}\frac{|u_{n\mu}|}{|z-z_n|^\mu}\\ &\leqslant n(2^{k+1}|z|)\frac{1}{(2^k-1)|z|}\max_{|z_n|>2^k|z|}2^{-l_{n\mu}}|u_{n\mu}|\\ &\leqslant\frac{1}{2^k-1}\,n(2^{k+1}|z|)\frac{|u_{n\mu}|}{n(2^{k+1}|z|)|u_{n\mu}|}=\frac{1}{2^k-1}.\end{aligned}\right\}\tag{3.19}$$

It is equally easy to see that Σ_{31} is bounded and hence

$$\Sigma_3=\Sigma_{31}+\Sigma_{32}+\cdots\quad\text{is bounded}.\tag{3.20}$$

Finally we estimate for $|z|\geqslant 2$

$$\left.\begin{aligned}|P(z)\,\Sigma_2| &\leqslant n(|z|+1)\left(\frac{|z|}{|z|-1}\right)^{L(|z|+1)}\max_{|z_n|<|z|+1}|u_{n\mu}|\left|\frac{P(z)}{(z-z_n)^\mu}\right|\\ &\leqslant n(2|z|)2^{L(2|z|)}B(2|z|)C(2|z|)(1+B(2|z|)M(2|z|))^{m(2|z|)-1}\\ &\times\max_{\substack{|\zeta|=|z|+2\\|z_n|<|z|+1}}\left|\frac{P(\zeta)}{(\zeta-z_n)^\mu}\right|\\ &<2n^2(2|z|)B^2(2|z|)C^2(2|z|)\times\\ &\qquad\times(1+B(2|z|)M(2|z|))^{2m(2|z|)-2}\times M(2|z|).\end{aligned}\right\}\tag{3.21}$$

Comparing (3.18), (3.19) and (3.21) we see that the upper bound in (3.18) dominates the upper bounds for (3.19) and (3.21). Thus

$$|f(z)|=|P(z)\,A(z)|<\{2n(2|z|)\,C(2|z|)\,(1+B(2|z|)\,M(2|z|))^{m(2|z|)}\}^{(1+\varepsilon)\log_2|z|}$$

for all large z.

3.22. COROLLARY. *Assume that* $m(r)=0(r^\varepsilon)$ and

$$n(r)=0(r^{\varrho+\varepsilon})\tag{3.23}$$

so that the Hadamard product $P(z)$ *satisfies* $\operatorname{ord}P\leqslant\varrho$. *If in addition*

$$B(r)=0(e^{r^{\varrho+\varepsilon}}).\tag{3.24}$$

Then for every sequence of values, $(w_{n\mu})$, *with*

$$C(r)=0(e^{r^{\varrho+\varepsilon}}),\tag{3.25}$$

there exists an entire function $f(z)$ *satisfying* ord $f \leqslant \varrho$ *and*

$$f^{(\mu-1)}(z_n) = w_{n\mu}; \quad \mu = 1, \ldots, m_n; \quad n = 1, 2, \ldots.$$

Note that for any $(w_{n\mu})$ *for which interpolation by an entire function* $f(z)$ *of order* $\leqslant \varrho$ *is possible there must exist an upper bound function* $C(r)$ *which satisfies* (3.25).

4. Interpolation by Entire Functions of Bounded Order

4.1. Definition. A sequence (z_n) of distinct complex numbers with multiplicities (m_n) is an *order* ϱ *interpolation sequence* if; for every sequence $|w_{n\mu}|$ satisfying $|w_{n\mu}| \leqslant C(|z_n|)$; $\mu = 1, \cdots, m_n$; $n = 1, 2, \ldots$ where $C(r) = 0(\exp(\varrho + \varepsilon))$; there exists an entire $f(z)$ with ord $f \leqslant \varrho$ and

$$f^{(\mu-1)}(z_n) = w_{n\mu}; \quad \mu = 1, \ldots, m_n; \quad n = 1, 2, \ldots.$$

In this section we shall restrict attention to sequences with bounded multiplicities $m_n \leqslant m$ and therefore, without loss of generality we may assume $m_n = m$ for all n. Our main result is the following.

4.2. Theorem. *A sequence* (z_n) *of distinct complex numbers with multiplicity* m *is an order* ϱ *interpolation sequence if and only if*

$$\sum_{z_n \neq 0} |z_n|^{-\varrho-\varepsilon} < \infty \quad \textit{or, equivalently,} \quad n(r) = 0(r^{\varrho+\varepsilon}) \tag{4.3}$$

and the Hadamard product $P(z) = \prod_{n=1}^{\infty} E_{[\varrho]}(z, z_n)^m$, *where* $E_k(z, z_n)$ *is defined in* (3.3), *satisfies*

$$|P^{(m)}(z_n)| \geqslant 1/B(|z_n|) \quad \textit{where} \quad B(r) = 0(\exp(r^{\varrho+\varepsilon})). \tag{4.4}$$

Proof. The sufficiency is a consequence of Corollary 3.22.
To prove necessity we use the following.

4.5. Lemma. *Given a Banach space* F *of functions analytic in a domain* D *so that the functionals* $L_{\mu z_0} f = f^{(\mu)}(z_0)$ *are continuous for* $\mu = 0, 1, \ldots, m-1$ *and all* $z_0 \in D$. *Given a sequence* (z_n) *of distinct points in* D *and a Banach space* S *of sequences* $(w_{n\mu})$; $\mu = 1, \ldots, m$; $n = 1, 2, \ldots$. *Assume that for every* $(w_{n\mu}) \in S$ *there exists an* $f \in F$ *so that* $f^{(\mu-1)}(z_n) = w_{n\mu}$; $\mu = 1, \ldots, m$; $n = 1, 2, \ldots$. *Then there exists a constant* M *such that for every* $(w_{n\mu}) \in S$ *there exists an* $f^* \in F$ *with*

$$f^{*(\mu-1)}(z_n) = w_{n\mu} \quad \text{and} \quad \|f^*\|_F \leqslant M \|(w_{n\mu})\|_S. \tag{4.6}$$

Proof. We first construct the Banach space F_S consisting of those functions $f \in F$ for which $(f^{(\mu-1)}(z_n)) \in S$ and with norm

$$\|f\|_{F_S} = \|f\|_F + \|(f^{(\mu-1)}(z_n))\|_S. \tag{4.7}$$

This space contains a subspace

$$I = \{f \mid f \in F; f^{(\mu-1)}(z_n) = 0;\ \mu = 1, \dots, m;\ n = 1, 2, \dots\}$$

and we can construct the factor space $B = F_S/I$ with

$$\|f + I\|_B = \inf_{g \in I} \|f + g\|_{F_S}.$$

By hypothesis the mapping

$$f + I \to (f^{(\mu-1)}(z_n))$$

is a one-to-one continuous mapping of B onto S. Thus, by the closed graph theorem for Banach spaces it has a continuous inverse

$$(w_{n\mu}) \to f + I$$

with $f \in F, f^{(\mu-1)}(z_n) = w_{n\mu}$ and

$$\|f + I\|_B \leqslant M \, \|(w_{n\mu})\|_S. \tag{4.8}$$

If we choose $(w_{n\mu})$ so that $\|(w_{n\mu})\|_S = 1$, we can choose the representative f^* in $f + I$ so that

$$\|f^*\|_{F_S} = \|f^*\|_F + 1 \leqslant M + 1$$

or

$$\|f^*\|_F \leqslant M. \tag{4.9}$$

We now consider Banach spaces of entire functions

$$\left.\begin{aligned} F_\varepsilon &= \{f \mid |f(z)| = 0(\exp(|z|^{\varrho+\varepsilon})\} \\ \|f\|_{F\varepsilon} &= \|f\|_\varepsilon = \sup_z |f(z)| \exp(-|z|^{\varrho+\varepsilon}) \end{aligned}\right\} \tag{4.10}$$

and the Banach space of bounded sequences

$$\left.\begin{aligned} S &= \{(w_{n\mu}) \mid |w_{n\mu}| = 0(1)\} \\ \|w_{n\mu}\|_S &= \sup_{n,\mu} |w_{n\mu}|. \end{aligned}\right\} \tag{4.11}$$

Proof of Theorem 4.2 *resumed.* Now the functions of order $\leqslant \varrho$ are exactly the functions in $\bigcap_{\varepsilon > 0} F_\varepsilon$. Therefore by hypothesis and Lemma 4.5 there exists an $M_\varepsilon > 0$ so that for every $\mu \in \{0, 1, \dots, m-1\}$ and $i \in \{1, 2, \dots\}$ there exists an $f_{\varepsilon i \mu} \in F_\varepsilon$ with

$$f_{\varepsilon i v}^{(\mu)}(z_j) = \delta_{ij} \delta_{\mu v} \quad \text{and} \quad \|f_{\varepsilon i v}\|_\varepsilon \leqslant M_\varepsilon. \tag{4.12}$$

Let (z_n') be the sequence of zeros of $f_{\varepsilon i v}$. Every z_n except z_i occurs at least m times in (z_n'). Let

$$B_{ir}(z) = \prod_{|z_n'| < r} \frac{r(z - z_n')}{r^2 - \bar{z}_n' z}, \tag{4.13}$$

where r is chosen so that $|z_n'| \neq r$ for all n. Then $B_{ir}(z)$ is analytic in $|z|<r$ and $|B_{i\nu r}(z)| = 1$ for $|z|=r$. Thus $B_{i\nu r}(z)/f_{\varepsilon i \nu}(z)$ is analytic without zeros in $|z| \leqslant r$. If we set

$$B^*_{i\nu r}(z) = B_{i\nu r}(z)\left(\frac{r(z-z_i)}{r^2 - \bar{z}_i z}\right)^{-\nu}$$

$$f^*_{\varepsilon i \nu}(z) = f_{\varepsilon i \nu}(z)\left(\frac{r(z-z_i)}{r^2 - \bar{z}_i z}\right)^{-\nu}$$

then by the minimum modulus principle and (4.12)

$$\left.\begin{aligned}
|B^*_{i\nu r}(z_i)| &= |B^*_{i\nu r}(z_i)/f^{(\nu)}_{\varepsilon i \nu}(z_i)| \\
&= \left|B^*_{i\nu r}(z_i)/(\nu!\left(\frac{r}{r^2 - |z_i|^2}\right)^{\nu} f^*_{\varepsilon i \nu}(z_i)\right| \\
&= \frac{1}{\nu!}\left(\frac{r^2 - |z_i|^2}{r}\right)^{\nu} |B^*_{i\nu r}(z_i)/f^*_{\varepsilon i \nu}(z_i)| \\
&\geqslant \frac{1}{\nu!}\left(\frac{r^2 - |z_i|^2}{r}\right)^{\nu} \min_{|z|\leqslant r} |B_{i\nu r}(z)/f_{\varepsilon i \nu}(z)| \\
&\geqslant \frac{1}{\nu!}\left(\frac{r^2 - |z_i|^2}{r}\right)^{\nu} (\max_{|z|\leqslant r} |f_{\varepsilon i \nu}(z)|)^{-1} \\
&\geqslant \frac{1}{\nu!}\left(\frac{r^2 - |z_i|^2}{r}\right)^{\nu} \frac{1}{M_\varepsilon} \exp(-r^{\varrho+\varepsilon}).
\end{aligned}\right\} \tag{4.14}$$

If we choose $|z_i| < \frac{1}{2}r$, $r \geqslant 2$ and set $M = m!\, M_\varepsilon$, we get

$$|B^*_{i\nu r}(z_i)| \geqslant \frac{1}{M} \exp(-r^{\varrho+\varepsilon}). \tag{4.15}$$

We now wish to show that (4.14) implies (4.4). But for $|z_i| < \frac{1}{2}r$ and $r \geqslant 2$ we have

$$\left.\begin{aligned}
&\frac{1}{M}\exp(-r^{\varrho+\varepsilon}) \leqslant |B^*_{i\nu r}(z_i)| \\
&\qquad = \prod_{\substack{|z_n'|<r \\ z_n' \neq z_i}} \left|\frac{r(z_i - z_n')}{r^2 - \bar{z}_n' z_i}\right| \leqslant \prod_{\substack{|z_n|<r \\ z_n \neq z_i}} \left|\frac{r(z_i - z_n)}{r^2 - \bar{z}_n z_i}\right|^m \\
&\qquad = \prod \left(\frac{|z_n|}{r}\right)^m \prod \left|\frac{1 - z_i/z_n}{1 - z_i \bar{z}_n/r^2}\right|^m \\
&\qquad \leqslant \prod (2^m |1 - z_i/z_n|^m) \leqslant 2^{n(r)} \prod_{\substack{|z_n|<r \\ z_n \neq z_i}} |1 - z_i/z_n|^m .
\end{aligned}\right\} \tag{4.16}$$

Using (4.3) we see that for every $\varepsilon > 0$ there exists a $c > 0$ independent of i and r so that

$$\prod_{\substack{|z_n| < r \\ z_n \neq z_i}} |1 - z_i/z_n|^m > c \exp(-r^{\varrho+\varepsilon}) \tag{4.17}$$

whenever $|z_i| < r/2$. Now

$$|P^{(m)}(z_i)| = m! \prod_{z_n \neq z_i} |E_{[\varrho]}(z_i, z_n)|^m \geqslant \prod\nolimits_1 \prod\nolimits_2 \tag{4.18}$$

where $\prod_1$ extends over all z_n with $|z_n| < r$, $z_n \neq z_i$ and $\prod_2$ extends over all z_n with $|z_n| > r$. Here r is chosen so that $|z_n| \neq r$ for all n and $2|z_i| < r < 3|z_i|$.

$$\left.\begin{aligned} \prod\nolimits_1 &\geqslant \prod_{\substack{|z_n| < r \\ z_n \neq z_i}} |1 - z_i/z_n|^m \exp\left(-m\Sigma\left(\left|\frac{z_i}{z_n}\right| + \cdots + \frac{1}{[\varrho]}\left|\frac{z_i}{z_n}\right|^{[\varrho]}\right)\right) \\ &\geqslant c \exp(-r^{\varrho+\varepsilon}) \exp\left(-m \sum_{|z_n| < r} \varrho \frac{r^{\varrho+\varepsilon}}{|z_n|^{\varrho+\varepsilon}}\right) \\ &\geqslant c \exp(-r^{\varrho+\varepsilon}). \end{aligned}\right\} \tag{4.19}$$

Also

$$\left.\begin{aligned} \prod\nolimits_2 &\geqslant \exp\left(-m \sum_{|z_n| > r > 2|z_i|} \left(\frac{1}{[\varrho]+1}\left|\frac{z_i}{z_n}\right|^{[\varrho]+1} + \cdots\right)\right) \\ &\geqslant \exp\left(-2m \sum_{|z_n| > r} \left|\frac{z_i}{z_n}\right|^{[\varrho]+1}\right) > \exp\left(-m\Sigma\left(\frac{r}{|z_n|}\right)^{\varrho+\varepsilon}\right) \\ &> c \exp(-r^{\varrho+\varepsilon}). \end{aligned}\right\} \tag{4.20}$$

If we multiply (4.19) and (4.20) and use the fact that $|z_i| > r/3$ we get (4.4).

The necessity of (4.3) is obvious if we choose, say, $w_{11} = 1$ and $w_{n\mu} = 0$ otherwise, so that $z_2, z_3, \ldots$ are the zeros of a nonzero entire function of order $\leqslant \varrho$.

It is clear that one could formulate similar theorems for functions analytic in a domain and for functions satisfying other growth conditions.

REFERENCE

[1] K. HOFFMAN, *Banach Spaces of Analytic Functions* (1962).

University of California, Los Angeles
University of California, Irvine
and
University of California, Los Angeles

Reports of Meetings

Die sechste Tagung über Funktionalgleichungen

Oberwolfach, 17–22, Juni 1968

Die sechste Tagung über Funktionalgleichungen fand vom 17. bis 22. Juni im Mathematischen Forschungsinstitut Oberwolfach statt. Sie stand, wie bisher, unter der Leitung der Professoren J. Aczél (Waterloo, Ont. – Bochum), O. Haupt (Erlangen – Nürnberg) und A. M. Ostrowski (Basel).

Leider konnte der dritte Tagungsleiter Herr Prof. Haupt wegen Erkrankung an der Tagung nicht teilnehmen.

Die Tagung wurde durch Herrn Prof. Ostrowski eröffnet. Anschliessend hat der Vertreter des Birkhäuser Verlags, Herr Einsele, das erste Heft der Aequationes Mathematicae den Teilnehmern überreicht.

Neben Lösung einzelner Funktionalgleichungen und allgemeinen Sätzen bezüglich Funktionalgleichungstypen im Reellen und Komplexen kamen Gleichungen in Algebren und abstrakten Räumen noch stärker zu Geltung als bei früheren Tagungen. Besonders möchten wir Anwendungen von verallgemeinerten Funktionen und Zusammenhänge mit Differentialgleichungen sowie ausführliche Behandlung von Funktionalungleichungen hervorheben. Erfreulicherweise haben zahlreiche Vorträge verschiedene Anwendungen aus der Geometrie, der Kontrolltheorie und der Informationstheorie eingehend behandelt.

In der an die Vorträge anschliessenden, oft sehr lebhaften Diskussion wurden verschiedene Fragestellungen erörtert und weiter geklärt. Die Problem- und Bemerkungssitzungen erwiesen sich auch diesmal als sehr anregend und nützlich und konnten sogar die knappen Zusammenfassungen einiger Vorträge aufnehmen, die nicht ausführlich abgehalten werden konnten.

Die Anzahl der Teilnehmer war die doppelte derjenigen bei den beiden letzten Oberwolfacher Funktionalgleichungstagungen. Insgesamt waren 36 wissenschaftliche Teilnehmer aus folgenden 10 Ländern anwesend: Australien (1), Deutschland (6), Italien (1), Jugoslawien (3), Kanada (7), Polen (1), Rumänien (6), Schweiz (1), Ungarn (4), Vereinigte Staaten (6).

Mit Bedauern wurde festgestellt, dass zahlreiche eingeladene Kollegen, insbesondere aus Polen und Ungarn, nicht kommen konnten.

Die Teilnehmer waren:

Aczél, J. (Waterloo, Ont., Kanada – Bochum, Deutschland)

Bajraktarević, M. (Sarajevo, Jugoslawien)

Baker, J. A. (Waterloo, Ont., Kanada)
Benz, W. (Bochum, Deutschland)
Cargo, D. P. (Amherst, Mass., USA)
Chen, Y. (Bochum, Deutschland)
Djordjević, R. Ž. (Niš, Jugoslawien)
Dragomir, A. (Timişoara, Rumänien)
Eichhorn, W. (Würzburg, Deutschland)
Fenyő, I. (Budapest, Ungarn)
Fischer, P. (Budapest, Ungarn)
Forte, B. (Pavia, Italien)
Haruki, H. (Waterloo, Ont., Kanada)
Howroyd, T. (Melbourne, Australien)
Ionescu, D. V. (Cluj, Rumänien)
Kannappan, Pl. (Waterloo, Ont., Kanada)
Kucharźewski, M. (Katowice, Polen)
Kurepa, S. (Zagreb, Jugoslawien)
Maier, W. (Jena, Deutschland)
McKiernan, M. A. (Waterloo, Ont., Kanada)
Melchior, U. (Bochum, Deutschland)
Mokanski, J. P. (Guelph, Ont., Kanada)
Muszély, G. (Budapest, Ungarn)
Oguztöreli, M. N. (Edmonton, Alta, Kanada)
Olariu, V. (Bucureşti, Rumänien)
Ostrowski, A. M. (Basel, Schweiz)
Popa, C. (Timişoara, Rumänien)
Radó, F. (Cluj, Rumänien)
Segal, S. (Rochester, N.Y., USA)
Sklar, A. (Chicago, Ill., USA)
Schmidt, H. (Würzburg, Deutschland)
Schweizer, B. (Amherst, Mass., USA)
Stamate, I. (Cluj, Rumänien)
Targonski, Gy. (Bronx, N.Y., USA)
Vincze, E. (Miskolc, Ungarn)
Zupnik, B. (Chicago, Ill., USA)

Auch die diesjährige Tagung war überaus fruchtbar und ungewöhnlich anregend. Sogar nach der offiziellen Sitzungszeit haben sich oft kleinere Gruppen zu ausserplanmässigen Informationsvorträgen und Diskussionen versammelt. Eine Einladung, die nächste Tagung im September 1969 in Kanada zu veranstalten, wurde dankend angenommen. Auf einhelligen Wunsch aller Teilnehmer wurde daher nach Rücksprache mit der Leitung des Forschungsinstituts Oberwolfach mit deren Zustimmung für den August 1970 die Veranstaltung einer Tagung über Funktionalgleichungen in Oberwolfach in Aussicht genommen.

Kurzfassungen der Vorträge sowie die Problemstellungen und Bemerkungen folgen (getrennt voneinander) in chronologischer Reihenfolge.

Vortragsauszüge

H. HARUKI: On a 'Cube Functional Equation'

In this lecture we solve the functional equation

$$\begin{aligned} f(x+t, y+t, z+t) + f(x+t, y+t, z-t) \\ + f(x+t, y-t, z+t) + f(x+t, y-t, z-t) \\ + f(x-t, y+t, z+t) + f(x-t, y+t, z-t) \\ + f(x-t, y-t, z+t) + f(x-t, y-t, z-t) \\ = 8f(x, y, z), \end{aligned}$$

where $f(x, y, z)$ is a real-valued continuous function of three real variables x, y, z in the whole xyz-space and x, y, z, t are real variables.

M. A. MCKIERNAN: Boundedness on a Set of Positive Measure and the Mean Value Property Characterizes Polynomials on a Space V^n

Let $\{y_i\}_{i\in I}$ be a finite collection of fixed vectors in an n-dimensional vector space $\mathscr{V}^n$ over the reals $\mathbf{R}$, where card $(I)=N\geqq n$. Let $\{\mu_i\}_{i\in I}$ be a set of real numbers such that $\sum_{i\in I}\mu_i=1$. We consider functions $f:\mathscr{V}^n\rightarrow\mathbf{R}$ which satisfy the functional equation

$$\sum_{i\in I}\mu_i f(x+ty_i)=f(x) \quad \text{for all} \quad x\in V^n,\ t\in\mathbf{R},\ t\geqslant 0. \tag{1}$$

THEOREM 1. *If f satisfies* (1) *and if* i) $\sum_{i\in J}\mu_i\neq 0$ *for all* $J\subseteq I$, ii) *the y_i span* $\mathscr{V}^n$, iii) *f is bounded on some set of positive measure in $\mathscr{V}^n$, then f is a polynomial on $\mathscr{V}^n$ of degree at most $N(N-1)/2$.*

This theorem is an immediate corollary of the more general

THEOREM 2. *Let $\mathscr{V}$ be a module over a commutative ring Q; let $A\subseteq Q$ be closed under addition; for $y\in\mathscr{V}$, let T_y denote the translation operator defined on all functions $f:\mathscr{V}\rightarrow\mathbf{R}$ by $(T_y f)(x)=f(x+y)$; let ω_t, for $t\in Q$, be the operator defined by*

$$\omega_t=\sum_{i\in I}\mu_i(T_{ty_i}-1)$$

where $\{\mu_i\}_{i\in I}\subset\mathbf{R}$, $\{y_i\}_{i\in I}\subset\mathscr{V}$, card $(I)=N$. *Let A be the commutative algebra* (over $\mathbf{R}$)

of all finite linear combinations of translation operators. If $\sum_{i\in I}\mu_i=1$, $\sum_{i\in J}\mu_i\neq 0$ for all $J\subset I$, then the ideal Ω_A generated in A by the operators $\{\omega_t\}_{t\in A}$ includes all finite difference operators of the form

$$(T_{ty_1}-1)^{k_1}(T_{ty_2}-1)^{k_2}\cdot\cdots\cdot(T_{ty_N}-1)^{k_N} \quad \textit{for} \quad k_1+k_2+\cdots+k_N>\frac{N(N-1)}{2}$$

for all $t\in A$.

W. EICHHORN: Funktionalgleichungen in Vektorräumen und Algebren

Solche Funktionalgleichungen (Urbild- und/oder Bildbereich der unbekannten Funktionen ein Vektorraum oder eine Algebra) sind unter anderem von Interesse

(i) als *definierende* Relationen für Klassen von Algebren (siehe *Problem* 10, weiter unten),

(ii) als Hilfsmittel zur Strukturuntersuchung *gegebener* Algebren,

(iii) bei der Lösung von Funktionalgleichungssystemen, sofern es möglich ist, diese als eine *einzelne* Gleichung in einer Algebra aufzufassen.

Zu jedem der drei Gesichtspunkte wurden Beispiele gegeben.

W. BENZ: Über die Funktionalgleichung $f(1+x)+f(1+f(x))=1$

Die folgende Resultate wurden vorgelegt: Es gibt genau einen injektiven Antiendomorphismus f der multiplikativen Gruppe L^* eines (nicht notwendig kommutativen) Körpers L, der für $x\neq 0$, -1 der Funktionalgleichung $f(1+x)+f(1+f(x))=1$ genügt, nämlich $f(x)=1/x$. Weiterhin: Gibt es einen injektiven Endomorphismus f der Gruppe L^*, der für $x\neq 0$, -1 der genannten Funktionalgleichung genügt, so ist L kommutativ. Auf zwei Anwendungen wurde hingewiesen: Enthält eine auf der projektiven Geraden über L minimal dreifach transitive Gruppe Γ alle Translationen und Dilatationen, so ist L kommutativ und $\Gamma\equiv \mathrm{PGL}(2, L)$. Gilt in der Kettengeometrie (k, L, Γ), k ein im Zentrum von L gelegener Unterkörper, das Winkelaxiom von Wilhelm Süss, so ist L kommutativ und $\Gamma\equiv \mathrm{PGL}(2, L)$.

C. POPA: Structures abstraites attachées à des équations fonctionnelles.

Dans le travail on considère l'équation fonctionnelle:

$$F(x, f_1(\alpha_1(x)), \ldots, f_n(\alpha_n(x)))=0$$

où F, α_i sont des applications connues et on cherche les applications f_i.

En associant à un ensemble une famille de relations d'équivalence définies sur cet ensemble, on obtient une structure que nous avons nommée ε-structure ou structure d'équivalence. En définissant convenablement les homomorphismes entre les ε-structures, on aboutit à la conclusion que chaque solution de l'équation fonctionnelle considérée est un tel homomorphisme.

D. P. CARGO: Function Semigroups and Functional Equations.

The relations L, R, D and H, first studied by I. A. Green, are completely characterized for a Function System J with the right subinverse property. These results are applied to (1) the generalized idempotency equation $g^{n+1}=g$ and to (2) Babbage's equation $g^n=m$ where m is a given subidentity.

THEOREM 1. *If J is the semigroup of all partial functions from a set X to itself, then the general solution f of* (2) *is uniquely a disjoint union $f=f_0 \cup f$, where* (a) f_0 *is a permutation on the range of f,* (b) $f^n(x)=x$ *for all* $x \in \operatorname{dom} f_0$. (c) $\operatorname{dom} f_0 \cap \operatorname{dom} f_1 = \phi$, (d) $\operatorname{dom} f_0 \cap \operatorname{ran} f_1 = \phi$ *and* (e) $f_1^n = \phi$.

THEOREM 2. *Let $n=1$ and let e be a fixed idempotent. Suppose all solutions in e/H to the equation $g^n=e$ are known. Then we can determine* (a) *all permutation solutions to* (2) *for any $m \in e/D$;* (b) *all solutions to* (2), *for any $m \in e/D$, provided the functions f_1 such that $f_1^n=\phi$ are known;* (c) *all solutions to* (1) *which lie in e/D.*

A. SKLAR: Lösungen der allgemeinen Konjugationsgleichungen

Es seien M_1, M_2 nicht-leere Mengen, und n eine positive ganze Zahl. Es sei G eine Funktion, die eine Teilmenge von M_1^n in M_1 abbildet, und H eine Funktion, die eine Teilmenge von M_2^n in M_2 abbildet. Wir nennen eine Funktionalgleichung der Gestalt:

$$F(G(x_1, \ldots, x_n)) = H(F(x_1), \ldots, F(x_n)),$$

in welcher die unbekannte Funktion F die Menge M_1 in M_2 abbildet, eine „allgemeine Konjugationsgleichung". Diese enthalten als Spezialfälle die am häufigsten untersuchten Funktionalgleichungen, z.B. die Abelsche und die Schrödersche Gleichungen für $n=1$, die Cauchy und verwandte Gleichungen für $n=2$. Wir können alle diese Gleichungen bearbeitet durch B. Schweizer und Verf. [The algebra of multiplace vector-valued functions, Bull. AMS *73*, 510–515, 1967; A grammar of functions, Aequat. Math., to appear] sowie dem Begriff der „kanonischen Zerlegung" einer Funktion [Verf., Canonical decompositions, stable functions and fractional iterates, Aequat. Math., to appear] mit Hilfe von algebraische Systeme gemeinsam behandeln. Speziell ist es möglich, eine nützliche algebraische Charakterisierung aller Lösungen einer Konjugationsgleichung zu geben.

D. ZUPNIK: Interassociativity

Two groupoids $(\mathcal{S}, F)$ and $(\mathcal{S}, G)$ are GF-associative if

$$F(G(x, y), z) = G(x, F(y, z)) \quad \text{for every} \quad x, y, z \in \mathcal{S}.$$

GF-associativity may be characterized by: Every right multiplication R_x of F commutes with every left multiplication λ_y of G.

THEOREM 1. *Let $(\mathcal{S}, F)$ and $(\mathcal{S}, G)$ be GF-associative. Then*

a) *if F has a right identity e, then*

$$G(x, y) = F(x, fy) \quad \textit{for every} \quad x, y \in \mathcal{S}$$

where $f = \mathcal{S}_e$ has domain γ and as range a subset of the right nucleus of F

$$(= \{x \mid FF_{yzx} = F_y F_{zx}\}).$$

b) *if F has a right identity e, and G has e as left identity, then $F = G$ is a semigroup.*

c) *if there exist $a \in \mathcal{S}$ such that the right multiplication R_a of F, and the left multiplication $\lambda_a \in G$ are both onto $\mathcal{S}$, then there exists a semigroup $(\mathcal{S}, H)$ with e as identity such that*

$$H(x, y) = F(x, f^* y) = G(g^* x, y)$$

where f^* is a function such that $R_a f^* = j$ and g^* is a function such that $\lambda_a g^* = j$. (By j, I we denote the identity function on $\mathcal{S}$).

It is shown, that it is necessary to assume an identity at least implicitly in order to relate GF-associativity to a semigroup. In this connection the following is shown:

THEOREM 2. *Let $(\mathcal{S}, F)$ be an aritrary groupoid with right identity. Then if the composition of L_a and L_b is also a left multiplication of $(\mathcal{S}, F)$, then*

$$L_a L_b = L_{Fab},$$

THEOREM 3. *Let $F = \{f\}$ bea set of functions with* Ran $f \subseteq$ Dom $f = \mathcal{S}$*, and such that there exists $a \in \mathcal{S}$ so that for every $x \in \gamma$ there exists at least one $f \in F$ for which $fa = x$. Then for every $x \in \mathcal{S}$ there exists at most one function t with* Ran $t \subseteq$ Dom $t = \mathcal{S}$ *for which $ta = x$, and which commutes with every $f \in F$. I.e. if*

$$t_i f = f t_i \quad \textit{for every} \quad f \in F \quad \textit{and}$$
$$t_1(a) = t_2(a), \quad \textit{then} \quad t_1 = t_2.$$

It is shown, that theorem 3 is the basis for the result of Ljapin on 'magnifying elements' of a semigroup.

B. SCHWEIZER: Semigroups on the space of probability distribution functions

Let Δ be the set of all functions F which are left-continuous and non-decreasing from **R** into [0, 1], i.e., probability distribution functions in the extended sense. For any F, G in Δ and any $h \geqq 0$, let A and B be the properties defined by: $A(F, G, h) \Leftrightarrow F(x-h)-h \leqq G(x)$ for $-1/h \leqq x \leqq h+1/h$ and $B(F, G, h) \Leftrightarrow G(x) \leqq F(x+h)+h$ for $-h-1/h \leqq x \leqq 1/h$; and let $\mathscr{L}(F, G)=\inf\{h \mid A(F, G, h) \text{ and } B(F, G, h)\}$. D. Sibley has shown that $\mathscr{L}$ is a metric on Δ, that the metric space $(\Delta, \mathscr{L})$ is compact and that, for any sequence $\{F_n\}$ in Δ, $\mathscr{L}(F_n, F) \to 0$ iff $\{F_n\}$ converges weakly to F. Continuous, associative operations (i.e., topological semigroups) on the space $(\Delta, \mathscr{L})$ are of interest in the study of triangle inequalities for probabilistic metric spaces and in other connections as well. Convolution is one such; and others may be defined with the aid of the known solutions to the topological semigroup problem on the unit interval. However, the general problem of finding all non-isomorphic topological semigroups (i.e., all non-equivalent solutions to the functional equation of associativity) on the space $(\Delta, \mathscr{L})$ is unsolved (**P 40**).

M. Bajraktarević: Sur les solutions générales de certaines équations fonctionnelles

En supposant que $\mathscr{C}$ soit un corps commutatif; X_i $(i=1, \ldots, n)$, $X=X_1 \times X_2 \times \cdots \times X_n$, Y – des ensembles donnés non vides dont les éléments sont désignés respectivement par x_i, $x=(x_1, \ldots, x_n)$, y; f_i $(i=0, 1, \ldots, n)$ des applications $f_0: X \to \mathscr{C}$, $f_i: X_i \times Y \to \mathscr{C}$ on donne les solutions générales des équations fonctionnelles

$$f_0(x) + \sum_{i=1}^{n} f_i(x_i, y) = 0,$$

$$f_0(x) + \prod_{i=1}^{n} f_i(x_i, y) = 0,$$

$$f_0(x) + \prod_{i=1}^{n} f_i(x_i, y) + \sum_{i=p+1}^{n} f_i(x_i, y) = 0,$$

$$f_0(x) + \prod_{i=1}^{n} f_i(x_i, y) \sum_{i=p+1}^{n} f_i(x_i, y) = 0,$$

$$f_0(x) + \prod_{i=1}^{p} f_i(x_i, y) + \prod_{i=p+1}^{n} f_i(x_i, y) = 0,$$

$$f_0(x) + \sum_{i=1}^{p} f_i(x_i, y) \sum_{i=p+1}^{n} f_i(x_i, y) = 0$$

et d'un certain nombre d'équations fonctionnelles se réduisant aux équations citées ci-dessus par des transformations simples.

M. Kucharźewski: Einige Ergebnisse über die Funktionalgleichungen mit Matrizenargumenten

Bezeichnen wir mit $\mathrm{GL}(n, R)$ bzw. $\overset{\frown}{\mathrm{GL}}(n, R)$ die Gruppe bzw. Halbgruppe aller quadratischen Matrizen der Ordnung n über dem reellen Zahlkörper R. Das folgende Funktionalgleichungssystem

$$F(B\,A) = F(B)\,F(A) \tag{1}$$

$$g(B\,A) = F(B)\,g(A) + g(B) \tag{2}$$

$$F:\quad \mathrm{GL}(n, R) \to \overset{\frown}{\mathrm{GL}}(m, R)$$

$$g:\quad \mathrm{GL}(n, R) \to R^m$$

wird betrachtet.

Insbesondere werden die Lösungen von (1), (2) im Falle $n=2$, $m=3$ bestimmt (Z. Karenska).

Ist F eine Lösung von (1), so stellt das Paar

$$F, \qquad g = (F - e)\cdot v \tag{3}$$

eine Lösung von (1), (2) dar, wenn e eine Einheitsmatrix der Ordnung m und v ein konstanter Vektor ist.

Es entsteht jetzt die Frage, wann (3) die einzige Lösung von (1), (2) ist. Jede der nachstehenden Bedingungen ist dafür hinreichend.

1° Es gibt ein $\varrho_0 \neq 0$, so dass die Matrix $F(\varrho_0 E) - e$ nicht singulär ist (M. Kucharźewski, M. Kuczma).

2° Es gibt ein $\varrho_0 \neq 0$, so dass die Matrix $F(E_1(\varrho_0)) - e$ nicht singulär ist (A. Zajtz).

3° Die Matrixfunktion $F(x)$ nicht reduzibel und $\mathrm{Det}\,F(x) \not\equiv 1$ ist.

E bedeutet die Einheitsmatrix der Ordnung n und $E_1(\varrho)$ entsteht, wenn man in der ersten Reihe und der ersten Spalte von E, ϱ einstetzt.

Aus diesen Bedingungen folgt, dass (3) die einzige Lösung von (1), (2) ist, wenn F die Form

$$F(X) = \varphi(\Delta)\, \underset{(1)}{X} \times \underset{(2)}{X} \ldots \underset{(p)}{X} \times \underset{(1)}{Y} \times \underset{\cdots}{\cdots} \times \underset{(q)}{Y} \quad (\Delta = \mathrm{Det}\, X)$$

hat $\big(Y = (X^{-1})^T$; das Zeichen „ד bedeutet das Kroneckersche Produkt der Matrizen$\big)$ (A. Zajtz).

B. Forte: On a System of Functional Equations in Information Theory

We start from the axiomatic definition of the measure of the amount of information without probabilities. Then we try to derive among others both Shannon's and Rényi's entropies for incomplete distributions as special measures of information. This leads to the following system of functional equations:

$$(\mathfrak{P}) \begin{cases} \theta(u, -x) = x + \theta(1-u, x) \quad \text{for} \quad \forall (u, x) \in B \\ \theta\left[1-u, x+\theta\left(\dfrac{1-v}{u}, y\right)\right] + \\ \qquad + \theta\left(\dfrac{1-v}{u}, y\right) = \theta\left[1-v, y+\theta\left(\dfrac{1-u}{v}, x\right)\right] + \theta\left(\dfrac{1-u}{v}, x\right) \\ \text{for} \quad \forall (u, v) \in \mathfrak{B}, \quad x \in R, \quad y \in R, \end{cases}$$

where R is the set of real numbers, and

$$B = \{(u, x): \quad u \in [0, 1], \quad x \in R\},$$
$$\mathfrak{B} = \{(u, v): \quad u \in (0, 1], \quad v \in (0, 1], \quad u + v \geqq 1\}.$$

One can prove the following two theorems:

THEOREM 1. *If:*

i) $\theta(u, x) \in C(B)$,

ii) $\theta(u, 0) = 0$ *(entropy idempotency),*

iii) $\theta(u, x)$ *is strictly monotonic (increasing) with respect to* x *for every* $u > 0$,

iv) $\theta(u, x)$ *satisfies* $(\mathfrak{P})$,

v) $\theta(0, x) = 0$

then either

$$\theta(u, x) = \frac{1}{1-\alpha} \log[1 - u + ue^{-(1-\alpha)x}] \tag{1}$$

or

$$\theta(u, x) = -ux. \tag{2}$$

Function (1) *defines Renyi's entropy, function* (2) *defines Shannon's entropy.*

THEOREM 2. *If:*

i) $\theta(u, x) = C^3(B)$,

ii) $\theta(u, x)$ *satisfies* $(\mathfrak{P})$,

then:

$$\theta(u, x) = \frac{1}{1-\alpha} \log[(1-u)^u + u^u e^{-(1-\alpha)x}], \quad \text{or} \tag{1}$$

$$\theta(u, x) = -ux, \quad \text{or} \tag{2}$$

$$\theta(u, x) = -\frac{1}{k} \log(1 + e^{kx}), \quad k \neq 0. \tag{3}$$

R. Ž. DJORDJEVIĆ: Remarques concernant une equation fonctionnelle

Considérons l'équation fonctionnelle suivante:

$$f(z+x, y, z) + f(y, y+z, x) + f(z, x, x+y) = 0, \tag{1}$$

ùo $x, y, z \in R$ R(l'ensemble des nombres réels) et $f: R^3 \to R$.

L'équation (1) est le cas particulier $g=h=f$, de l'équation

$$f(z+x, y, z) + g(y, y+z, x) + h(z, x, x+y) = 0.$$

Il n'est pas difficile de démontrer que cette équation a la solution générale dans la forme

$$f(u, v, w) = F(u-w, v, w), \quad g(u, v, w) = G(u, v-u, w),$$
$$h(u, v, w) = -F(v, w-v, u) - G(w-v, u, v),$$

où F et G sont des fonctions arbitraires de R.

Mes efforts, effectués en vue de résoudre l'équation (1), ne m'ont permis de déterminer aucune solution non triviale de l'équation (1).

Cependant, j'ai obtenu les résultats suivants:

THÉORÈME 1. *Si la fonction f satisfait à l'équationels fonctionnelle* (1), *nous avons.*

$$f(u, v, w) + f(v, w, u) + f(w, u, v) = 0.$$

THÉORÈME 2. *L'équation* (1) *n'est verifiée par aucun polynôme* ($\neq 0$) *de degré moins de cinq.*

THÉORÈME 3. *L'équation* (1) *n'a pas des solutions sous la forme d'un polynôme de degré n arbitraire, qui comprend tous les monômes de degré au plus n.*

Les résultats obtenus nous incitent à poser les trois hypothèses suivantes:

HYPOTHÈSE 1. *L'équation fonctionnelle* (1) *n'a pas de solution sous la forme d'un polynôme.*

HYPOTHÈSE 2. *L'équation fonctionnelle* (1) *n'a aucune fonction analytique comme solution non triviale.*

HYPOTHÈSE 3. *L'équation fonctionnelle* (1) *n'a pas de solutions non triviales de tout.*

W. MAIER: Additive Inhaltsmasse im positiv gekrümmten Raum

Die auf H. Kneser zurückgehende Integraldarstellung von Inhaltsmassen im R_{n-1} fester Krümmung wurde für reguläre Simplexe durch H. Ruben mit Fehlerintegralen als Integrandenfaktoren in Zusammenhang gebracht. Die Vielfachheit von Eigenwerten einer quadratischen Form benutzte B. Weissbach zur Klassifikation von Simplexen entsprechend dem Zerfall gewisser Integrale; Simplexe vom Typ 1 hängen nur von n unabhängigen Bestimmungsstücken ab, und genügen einer einfachen Funktionalgleichung. Die letzten Arbeiten von L. Schläfli zur vieldimensionalen Geometrie behandelten Quotienten von Inhaltsmassen aus Räumen verschiedener

Krümmung. Mit Begriffen der Verbandstheorie gelang es A. Effenberger, aus den Winkeln verschiedener Ordnung im Simplex gewisse Invarianten aufzubauen. Durch Konstruktionen der Ergänzungsgeometrie können damit Schläflis Quotienten im Fall fester Nenner zur Herleitung linearer Funktionalgleichungen für konvexe Simplexe benutzt werden.

F. RADÓ: Behandlung von Fragen über Kollineationen mit Funktionalgleichungsmethoden

Ist K ein Schiefkörper, $K^* = K \setminus \{0\}$, $\bar{K} = K \cup \{\infty\}$, so setzen wir $x + \infty = \infty + x = \infty$ für $x \in K$, $x \cdot \infty = \infty \cdot x = \infty$ für $x \in K^* \cup \{\infty\}$ ($\infty + \infty$, $0 \cdot \infty$ sind nicht erklärt). Die Abbildung $\bar{e}: \bar{K} \to \bar{K}$ heisst ein *Endomorphismus von* $\bar{K}$, falls $\bar{e}(x+y) = \bar{e}(x) + \bar{e}(y)$, $\bar{e}(xy) = \bar{e}(x)\,\bar{e}(y)$, jedesmal wenn die bzgl. Operationen erklärt sind. Die Einschränkung $e: M \to K$ von $\bar{e}$ auf $M = \bar{e}^{-1}(K)$ soll als ein *partieller Endomorphismus von* K bezeichnet werden. Man bezeichnet $M^* = e^{-1}(K^*)$. Es wird im Vortrag eine Kennzeichnung der partiellen Endomorphismen ohne Benutzung des Elements ∞ gegeben.

Es sei P die Punktmenge der projektiven Ebene über K und $\mathscr{C} \subset P$. Die Kollineation $\varphi: \mathscr{C} \to P$ wird als *ausgeartet* bezeichnet, wenn es Geraden g_1, g_2 mit $\varphi(\mathscr{C}) \subset g_1 \cup g_2$ gibt. Es gelten die folgenden Verallgemeinerungen eines Satzes von J. Aczél und W. Benz:

Ist die Kollineation $\varphi: \mathscr{C} \to P$ nicht ausgeartet und enthält $\mathscr{C}$ vier Geraden, so ist φ bis auf projektive Kollineationen als Faktoren durch

$$K(x_1, x_2, x_3) \to K(e(\varrho x_1), e(\varrho x_2), e(\varrho x_3))$$

gegeben, wo e ein partieller Endomorphismus von K ist und ϱ wie folgt bestimmt wird: $\varrho x_i \in M$, $i = 1, 2, 3$, $\exists j$, $\varrho x_j \in M^*$ (was immer möglich ist).

Wird darüber hinaus die Injektivität von φ in einem einzigen Punkt verlangt, die Menge $\mathscr{C}$ dagegen so verallgemeinert, dass sie nur noch 3 Geraden und einen Punkt zu enthalten braucht, so gilt statt der obigen Darstellung die folgende:

$$K(x_1, x_2, x_3) \to K(e(x_1), e(x_2), e(x_3)),$$

wobei e ein Endomorphismus von K ist.

M. N. OĞUZTÖRELI: A Class of Functional Equations in Optimal Control Theory

Let B be a Banach space of continuous functions $v = v(t)$ defined for $t \geqq 0$, and V a subset of B. Let the functions $u = u(t)$ $(\equiv u(t; v))$ be defined by

$$
\left.\begin{aligned}
u(t;v) = \phi(t) &+ \int_0^t K_1(t;\sigma)\,v(\sigma)\,d\sigma \\
&+ \frac{\lambda}{2}\int_0^t\int_0^t K_2(t;\sigma_1,\sigma_2)\,v(\sigma_1)\,v(\sigma_2)\,d\sigma_1\,d\sigma_2 + \cdots \\
&+ \frac{\lambda^n}{(n+1)!}\int_0^t\int_0^t\cdots\int_0^t K_{n+1}(t;\sigma_1,\dots,\sigma_{n+1})\,v(\sigma_1) \\
&\quad \dots v(\sigma_{n+1})\,d\sigma_1\,d\sigma_2\dots d\sigma_{n+1} + \cdots
\end{aligned}\right\} \tag{1}
$$

where λ is a real parameter, $\phi(t)$ is a given function continuously differentiable for $t \geqq 0$, $K_j(t;\sigma_1,\dots,\sigma_j)$ $(j=1,2,3,\dots)$ are continuously differentiable in t for $t \geqq 0$, continuous and symmetric in $\sigma_1,\dots,\sigma_j$ for $0 \leqq \sigma_i \leqq t$ $(i=1,2,\dots,j)$, and the functional power series in (1) is supposed to be uniformly convergent for $t \in I_0$, for each $v \in V$ and for $|\lambda| \leqq R$, R being a positive constant and $I_0 = [0, \gamma]$ a given compact interval.

Let $\mathscr{I}$ be the set of all intervals $I = [\alpha, \beta]$ contained in I_0. Let

$$F_1(u,v) = F(u(t), v(t), t \in I) \quad (I \in \mathscr{I}) \tag{2}$$

be a given non negative functional defined on the space $U \times V \times \mathscr{I}$, where U denotes the space of all possible $u = u(t; v)$. The functional $F_1(u, v)$ is supposed to be strongly differentiable in $u \in U$ and in $v \in V$ for each $I \in \mathscr{I}$; and if $I_1, I_2 \in \mathscr{I}$ and if $I_1 \cap I_2 \neq \phi$, or, I_1 and I_2 have only one point in common, then

$$F_{I_1 \cup I_2}(u,v) = F_{I_1}(u,v) + F_{I_2}(u,v). \tag{3}$$

Further, it is also supposed that $F_1(u, v)$ is continuous in I and if $I_\alpha = [\alpha, \beta] \in \mathscr{I}$, then

$$\frac{\partial}{\partial\alpha} F_{I_\alpha}(u,v) = \lim_{\beta\to\alpha} \frac{F_{I_\alpha}(u,v)}{\beta - \alpha} \tag{4}$$

exists and is strongly differentiable in u and v. A possible form for $F_I(u, v)$ is

$$F_I(u,v) = \int_\alpha^\beta Q(u(t), v(t))\,dt, \tag{5}$$

where $Q(u, v)$ is a non-negative, continuously differentiable function in u and v.

The optimization problem considered here consists of the finding of a $v^0 \in V$ for which the functional $F_{I_0}(u, v)$ assumes its minimum. This problem leads us to the following functional equation

$$\frac{\partial}{\partial \tau}\{\min_{v \in V} F_{I_\tau}(u, v)\} = -\min_{v \in V}\left\{\frac{\partial}{\partial \tau} F_{I_\tau}(u, v)\right\}, \tag{6}$$

where $\tau \in I_0$ and $I_\tau = [\tau, \gamma]$. Certain natural consequences of (6) are discussed.

The special case where $F_I(u, v)$ is defined by (5) has been also discussed and shown that extreme points of the functional $(\partial/\partial\tau) F_{I_\tau}(u, v)$ satisfy a Volterra equation of the first kind involving a functional power series similar to the right hand side of (1).

(This is an abstract of a paper under publication in the Rend. Acc. Lincei 1968.)

I. Fenyő: Über ein Problem von M. Hosszú

Es wird die Funktionalgleichung

$$f(x + y - xy) + f(xy) = f(x) + f(y)$$

im Bereich der Distributionen (im Sinne von L. Schwartz) mittels gewisser Distributionentransformationen gelöst.

V. Olariu: Sur les itérés d'un ordre quelquonque des quelques opérateurs différentielles

On part de l'équation fonctionnelle

$$f_\lambda(x) * f_\mu(x) = f_{\lambda+\mu}(x) \tag{1}$$

où λ et μ sont des nombres complexes et $f_\lambda(x)$ est une distribution (au sens de Schwartz) qui dépend analitiquement de λ, le signe $*$ indique la convolution des fonctions entre quelles il est posé. Si les distributoins $f_\lambda(x)$ ont des propriétés convenables, on peut définir, a l'aide de (1) toutes les soulutions élémentaires d'un opérateur differentiel, étroitement lié à $f_\lambda(x)$ et de ses itérés.

Ainsi, si $f_\lambda(x)$ pour $\lambda = 0$ est la distribution de Dirac, $f_\lambda(x)|_{\lambda=0} = \delta(x)$, (1) montre que $f_{-\lambda}(x)$ est une sorte de «inversion élémentaire» de $f_\lambda(x)$ et si $f_\lambda(x)|_{\lambda=-k} = L^k\,\delta$, $k = 1, 2, \ldots$, (1) donne aussi que $f_k(x)$ est la solution élémentaire de l'opérateur L^k. (L^k est l'itéré d'ordre k de L); il résulte que si on peut définir les itérés d'un ordre quelquonque L^λ de l'opérateur L, (1) donne les solutions élémentaires correspondents.

Exemples. Si

$$f_\lambda(x) = \frac{x_+^{\lambda-1}}{\Gamma(\lambda)} e^{ax}$$

$$x_+ = \begin{cases} x & \text{si } x > 0 \\ 0 & \text{si } x \leqq 0 \end{cases}$$

on $a f_0(x) = \delta(x)$ et $f_{-k} = L^k \delta(x)$;

$$Ly = \frac{dy}{dx} - ay$$

a étant un scalaire ou une matrice.

Si

$$f_\lambda(x, t) = \frac{t_+^{\lambda-1-n/2}}{\Gamma(\lambda)(4\pi)^{n/2}} e^{-|x|^2/4t}$$

$$(|x|^2 = x_1^2 + \cdots + x_n^2)$$

on a aussi $f_0(x, t) = \delta(x)$ $x = (x_1, \ldots, x_n)$ et

$$f_{-k}(x, t) = L^k, \quad Lu = \frac{\partial u}{\partial t} - \Delta_n u$$

$$\left(\Delta_n u = \frac{\partial^2 u}{\partial x_1^2} + \cdots + \frac{\partial^2 u}{\partial x_n^2}\right).$$

GY. TARGONSKI: Endomorphisms of Function Algebras and Schröder's Equation

The paper is part of a program to increase the use of functional analysis in the theory of functional equations of one variable.

ω is a mapping of an arbitrary set U into itself; Z is a linear algebra. A contravariant exponential functor assigns to u the function set Z^U and to ω the mapping Z^ω of Z^U into itself. Z^ω turns out to be a linear endomorphism of Z^U, of the form $Z^\omega f = f_{0\omega}$. This occurs in the Schröder equation and its generalization, the linear functional equation of the first order (the Kordylewski-Kuczma equation).

The 'reversal equation' algebra is this: in a given function algebra, every endomorphism generated by a contravariant exponential functor, in other words a right composition ('substitution'). As known, the answer is affirmative for the algebra of all continuous mappings of a compact Hausdorff space into the reals; but compactness of the domain is not necessary since the answer is also affirmative for the algebra of all complex polynomials. The answer is however negative for the algebra of all bounded functions on the unit interval; if E is a proper subset of the unit interval, ψ_E its characteristic function, and $h(x)$ a mapping of the unit interval into itself, then

$$(\Omega f)(x) = \psi_E(x) f[h(x)]$$

is an endomorphism which cannot be represented as a right composition. It can however be so represented if we restrict ourselves to the subalgebra of real functions vanishing at some fixed x_0.

Necessary and sufficient conditions for the 'reversal equation' to be answerable in the affirmative are not known (**P 41**).

J. A. BAKER: The Functional Equation $f(x+y)f(x-y)=f(x)^2-f(y)^2$

Using a result of W. H. Wilson, Bull. Amer. Math. Soc. *26*, 300–312 (1919) we prove the following:

THEOREM. *Let f be a complex-valued function defined on a real vector space X such that*

$$f(x+y)f(x-y)=f(x)^2-f(y)^2$$

for all $x, y \in X$.

I. f is continuous along rays (i.e. $\forall x \in X$ the mapping $r \to f(rx)$ is a continuous function of the real variable r), then either

(a) f is linear (i.e. $f(rx+sy)=rf(x)+sf(y)$ for all $x, y \in X$ and all real r and s) or

(b) $f(x)=c \sin L(x)$ $\forall x \in X$ where c is a complex constant and L is a complex-valued linear function defined on X.

II. X is a linear topological space and f is continuous then I obtains and in case (b) L is continuous.

III. If $X=R^n$ and f is measurable on some subset of positive measure then f is continuous.

Part II. generalizes a result of S. Kurepa, Ann. Pol. Math. *10*, 1–5, 1961, and Part III. generalizes results of S. Kurepa, Monatsh. d. Math. *64*, 321–329, 1960, and S. L. Segal, Amer. Math. Monthly *70*, 306–308, 1963.

A. OSTROWSKI: Über eine Klasse von Funktionalungleichungen

Setzt man $T(f, g)=\int_0^1 fg\,dx-\int_0^1 f\,dx \int_0^1 g\,dx$, so wurde die Grüss'sche Ungleichung $|T(f, g)| \leqq \frac{1}{4}\operatorname{Osz} f \operatorname{Osz} g$ von neuem bewiesen und im Zusammenhang mit einigen Tschebyscheffschen Un gleichungen auf die allgemeinsten linearen Mittelwertsbildungen übertragen. Von den weiteren Resultaten seien angeführt:

$$|T(f, g)| \leqq \tfrac{1}{8}\left(\int_0^1 f'^2\,dx\right)^{1/2}\left(\int_0^1 g'^2\,dx\right)^{1/2}; \tag{1}$$

$$\int_0^1\int_0^1 \left|\frac{f(x)-f(y)}{x-y}\right|^s dx\,dy \leqq (\lg 4)\int_0^1 |f'(x)|^s\,dx, \quad (s \geqq 1). \tag{2}$$

GY. MUSZÉLY: Über die stetigen Lösungen der Ungleichung

$$pf(p)+(1-p)f(1-p) \geqslant pf(q)+(1-p)f(1-q).$$

Die in dem Intervall (0, 1) stetigen Lösungen lassen sich in der Form

$$f(p) = (1-p)\cdot g(p-\tfrac{1}{2}) + \int_0^{p-1/2} g(t)\,dt + C$$

darstellen, wobei g eine in dem Intervall $(-\frac{1}{2}, \frac{1}{2})$ ungerade, stetige, wachsende Funktion und C eine beliebige Konstante sind.

P. Fischer: Sur l'inégalité $\sum_{i=1}^{n} p_i f(p_i) \geqslant \sum_{i=1}^{n} p_i f(q_i)$

On s'occupe de l'inégalité suivante

$$\sum_{i=1}^{n} p_i f(p_i) \geqslant \sum_{i=1}^{n} p_i f(q_i) \tag{$*$}$$

où $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$ et $p_i > 0$, $q_i > 0$ $(i = 1, 2, \ldots, n)$ et $n \geqslant 2$, mais fixé.

On démontre les résultats suivants

THÉORÈME 1. *Tous les solutions de* $(*)$ *sont monotones.*

THÉORÈME 2. *Tous les solutions de* $(*)$ *sont différentiables si* $n \geqslant 3$.

THÉORÈME 3. *On considère la décomposition de* f *dans la forme* $f = g_1 + g_2 + h$ *où* h *est une fonction des sauts,* g_1 *est une fonction absolument continue et* g_2 *est une fonction singulière continue, alors aussi* h_1, g_1 *et* g_2 *satisfont à* $(*)$.

On donne encore la solution générale de $(*)$ dans la domaine des fonctions de sauts et celle des fonctions absolument continues. Ces résultats sont généralisations de résultats de MM. Aczél et M. Pfanzagl (Metrika *11*, 91–105, 1966).

J. Aczél: Über Zusammenhänge zwischen Differential- und Funktionalgleichungen

Eine Differentialgleichung $y'' = \varphi(x, y, y')$ lässt mit y auch $x \to y(h(x))$ als Lösung zu, wenn $\varphi(h, y(h), y'(h))\, h'^2 + y'(h)\, h'' = \varphi(x, y(h), y', (h)\, h')$ ist. Dies geht in die von y unabhängige Gestalt $h'' = f(x, h, h')$ über, wenn $\varphi(x, y, y'h') = \varphi(h, y, y')\, h'^2 + y'f(x, h, h')$ d.h.

$$\varphi(x, y, zu) = \varphi(v, y, z)\, u^2 + zf(x, v, u) \tag{1}$$

ist.

Die Herren A. Moór und L. Pintér haben (1) durch Zurückführung auf

$$f(x, y, zu) = f(v, y, z)\, u^2 + zf(x, v, u) \tag{2}$$

und

$$\gamma(zu) = \gamma(z)\, u^2 + z\gamma(u) \tag{3}$$

und die Gleichung (3) unter Differenzierbarkeitsvoraussetzung gelöst (Publ. Math. *13*, 207–223, 1966).

Hier werden (3), (1) und (2) kürzer und ohne irgendwelche Regularitätsvoraussetzung (auch nicht Stetigkeit) gelöst.

PL. KANNAPPAN: A Functional Equation for the Cosine

It is well known that, the complex-valued, measurable solutions of D'Alembert's functional equation

$$f(x+y)+f(x-y)=2f(x)f(y), \quad \text{for} \quad x, y \text{ real}, \tag{A}$$

are $f(x)=\cos ax$, where a is *any* complex constant. Here, the functional equation

$$f(x+y+2A)+f(x-y+2A)=2f(x)f(y), \tag{B}$$

where f is a complex-valued, measurable function of a real variable and $A \neq 0$ is a real constant, is considered. It is shown that f is continuous and that apart from the trivial solutions ($f=0$, 1), the only functions which satisfy (B) are cosine functions $\cos ax$ and $-\cos bx$ where for the constants a and b only a *denumerable* set of (real) numbers is admissible. Equation (B) is similar to the equation

$$f(x-y+A)-f(x+y+A)=2f(x)f(y)$$

considered by E. B. Van Vleck (Ann. Math. (2) 13, 154, 1913), where f is assumed real and continuous and the general solution is $f(x)=\sin cx$, for a sequence $c=((4j+1)\pi)/2A$ ($j=0, 1, \ldots$).

T. D. HOWROYD: Some Uniqueness Theorems for Functional Equations

If $w=H(u, v, x, y)$ is continuous and implicitly defines v as a continuous function of w, u, x, y; F is continuous, strictly increasing (or decreasing) and satisfies a Lipschitz condition in its first place; ϕ has measurable majorant on a set of positive measure and satisfies

$$\psi(F(x, y))=H(\phi(x), \phi(y), x, y)$$

then ϕ is locally bounded. Further if $r\in[-\frac{1}{2}, 1]$, $s\in(0, \infty)$ and

$$|H(u, v, x, y)-H(U, V, x, y)| \leqq rs(|u-U|+|v-V|)$$
$$|H(u, u, x, x)-H(v, v, x, x)| \geqq s|u-v|,$$

then ϕ is uniquely determined by two initial conditions. These results have extensions to the case where x, y are complex varables or n-dimensional vectors.

E. VINCZE: Über eine Klasse der alternativen Funktionalgleichungen

Wir betrachten die Funktionalgleichung

$$F[f(x+y), f(x), f(y)] = 0, \tag{1}$$

wobei x, y die Elemente einer beliebigen Abelschen Gruppe Q sind, die gesuchte Funktion f die Gruppe Q in (oder auf) einen beliebigen Körper K der Charakteristik 0 abbildet, weiterhin $F(u, v, w)$ ein gegebenes Polynom von u, v, w ist. Es sei weiter

$$F_1[f(x+y), f(x), f(y)] = F_2[f(x+y), f(x), f(y)] \tag{1a}$$

eine beliebige „Umordnung" der Funktionalgleichung (1) in dem folgenden Sinne: Die Gleichung (1a) entsteht aus (1) durch endlich viele Additionen, Multiplikationen (und ihren Inversoperationen) der angewandten Funktionen $f(x+y)$, $G[f(x), f(y)]$, wobei $G(u, v)$ eine rationale Funktion von u, v ist, mit einverstanden auch die Fälle, wo $G(u, v) = G_1(u)$, $G(u, v) = G_2(v)$, $G(u, v) =$ konst. sind. Wir nennen die Funktionalgleichung

$$h\{F_1[f(x+y), f(x), f(y)]\} = h\{F_2[f(x+y), f(x), f(y)]\}$$

(oder ihre „Umordnung" in dem obigen Sinne) eine algebraisch alternative Funktionalgleichung von (1), wenn die rationale Funktion h keine eindeutige Inversfunktion hat. Unter anderem gilt der folgende

SATZ: *Jede algebraisch alternative Gleichung der beiden Funktionalgleichungen*

$$f(x+y) - f(x) - f(y) = 0, \tag{2}$$

$$g(x+y) - Ag(x)\,g(y) - Bg(x) - Bg(y) - C = 0, \quad B^2 = AC + B \tag{3}$$

besitzt nur dieselben nichtkonstanten Lösungen, wie (2) *bzw.* (3).

D. V. Ionescu: Sur l'équation fonctionnelle de M. Fréchet

On sait que pour la différence d'ordre n de la fonction f, on a la représentation $\Delta_h^n f(x) = \int_x^{x+nh} \varphi(s) f^n(s)\,ds$, lorsque $f \in C^n[x, x+nh]$, qui est un cas particulier de la représentation $[x_0, x_1, \ldots, x_n; f] = \int_{x_0}^{x_n} \varphi(s) f^n(s)\,ds$, lorsque $f \in C^n[x_0, x_n]$. Dans cette formule la fonction φ est positive sur (x_0, x_n) et l'on a $\int_{x_0}^{x_n} \varphi(s)\,ds = 1/n!$.

M. Fréchet a considéré la différence $\Delta_{h_1,\ldots,h_n}^n f(x)$ à n pas $h_1, \ldots, h_n$ et a démontré que les seules solutions continues de l'équation fonctionnelle $\Delta_{h_1,\ldots,h_n}^n f(x) = 0$, quelsque soient $h_1, \ldots, h_n$ sont les polynomes de degré au plus égal a $n-1$.

Dans cette communication nous donnons la représentation

$$\mathop{\Delta}_{h_1,\ldots,h_n}^{n} f(x) = \int_x^{x+h_1+\cdots+h_n} \psi(s) f^n(s)\,ds$$

lorsque $f \in C^n[x, x+h_1+\ldots+h_n]$ et sous certaines hypothèses sur les $h_1, \ldots, h_n$. Nous montrons que la fonction ψ s'obtient par un problème aux limites, qu'elle est positive sur l'intervalle $(x, x+h_1+\cdots+h_n)$ et que

$$\int_x^{x+h_1+\cdots+h_n} \psi(s)\,ds = h_1\, h_2 \ldots h_n .$$

De cette représentation on peut déduire des conclusions intéressantes sur certaines équations fonctionnelles.

I. STAMATE: Funktionalgleichungen der Polynome

Es werden die hauptsächlichsten Ergebnisse über die Funktionalgleichungen von M. Fréchet, Th. Anghelutza, Brouwer, A. Marchaud, T. Popoviciu, D. V. Ionescu, J. Herbrand etc, erwähnt.

Die Arbeit ist in folgende Kapitel eingeteilt: Funktionalgleichungen einer Veränderlichen, Funktionalgleichungen mit mehreren Veränderlichen, abstrakte Funktionen, als Limes definierte Polynome, Integro-Funktionalgleichungen.

In der Zusammenfassung werden die vom Verfasser bearbeiteten Integro-Funktional-gleichungen eingeführt, in denen die Gleichungen mehrere unbekannte Funktionen enthalten.

S. KUREPA: Relations Between Additive Functions

We consider additive functions f and g such that

$$f(A(x)) = P(x)\, g(B(x))$$

holds for all but at most a finite number of x's, with A and B quotients of polynomials and P a continuous function. Details will appear in this Journal.

Problemstellungen und Bemerkungen

1. Remarque sur la série

$$g(x) = \sum a_r (\varphi(x))^r$$

où $\varphi(x+y) = \varphi(x) + \varphi(y)$. E. Vincze a posé la question suivante: Comment est-ce possible caractériser les fonctions $g(x)$ qui peuvent être écrites sous la forme suivante:

$$g(x) = \sum_{r=0}^{\infty} a_r (\varphi(x))^r$$

où x est réel arbitraire, les a_r sont réels et où $\varphi(x+y) = \varphi(x) + \varphi(y)$.

On donne plusieurs théorèmes pour caractériser ces fonctions, parmi lesquels sont les suivants: Si φ est une fonction non-continue et si la fonction entière $x \to F(x) = \sum a_r x^r$ n'est pas identiquement égal à une constante, on a que la fonction g ne peut pas être mesurable dans aucun intervalle. On donne la généralisation suivante de ce théorème: Soit donnée une fonction réelle φ qui possède la propriété suivante: Si A est un ensemble de mesure positive, alors l'image $\varphi(A)$ est dense partout. Soit $x \to F(x) = \sum a_r x^r$ une telle fonction entière (a_r soit réel) pour laquelle $F'(x) \not\equiv 0$, alors on a que la fonction $x \to F(\varphi(x))$ ne peut être continue dans aucun point, de plus elle ne peut être mesurable dans aucun intervalle.

P. Fischer

2. Problem. O. E. Gheorghiu (Bul. Sti. Tehn. Timişoara *11*(*25*), 391–393, 1966) has mentioned the functional equation

$$F(x, y)\, F(z, t) = F(xz - yt,\, xt + yt + yz) \qquad (*)$$

stating only that $F(x, y) = (x^2 + xy + y^2)^a$ with arbitrary constant a is a solution. The equation implies

$$F(x, y) = \begin{cases} m(y)\, g(x/y) & y \neq 0 \\ m(x) & y = 0, \end{cases}$$

where m is a multiplicative function:

(i) $m(xy) = m(x)\, m(y)$

and

(ii) $g(u)\, g(v) = m(u + v + 1)\, g\left(\dfrac{uv - 1}{u + v + 1}\right) (u + v + 1 \neq 0)$

(iii) $g(u)\, g(-1 - u) = m(-1 - u - u^2)$.

How can the general solution of (*) be determined explicitly? How that of (i), (ii), (iii)? **(P 42)**

J. Aczél

3. Remark. A detailed report was given on the method of W. A. Luxemburg (Lectures on A. Robinson's theory of infinitesimals and infinitely large numbers. Second ed. 1964. p. 82) on the solution of the Cauchy equation (*) $f(x+y) = f(x) + f(y)$. With means of non-standard analysis the following was proved: if f is a solution of (*) on $\mathbf{R}$ and f is bounded on some interval in $\mathbf{R}$, then $f(x) = xf(1)$ for all $x \in \mathbf{R}$.

I. Fenyő

4. Remark. M. V. Zareckij's paper «Sur quelques équations fonctionnelles liées à l'équation de Cauchy» (Russian), Doklady Ak. Nauk Beloruss. SSR *11*, 487–491, 1967 states that

$$F(x+y, z) + F(x, y) = F(x, y+z) + F(y, z) \tag{1}$$

is necessary and sufficient in order that

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

should have a solution f. This statement is false, but it becomes true if we combine (1) with

$$F(x, y) = F(y, x), \tag{2}$$

because the general solution of (1) is

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

where f is arbitrary, T is an arbitrary biadditive ($T(x_1+x_2, y)=T(x_1, y)+T(x_2, y)$; $T(x, y_1+y_2)=T(x, y_1)+T(x, y_2)$) skew-symmetric ($T(y, x)=-T(x, y)$) function, while the *general solution of the system* (1), (2) *is*

$$F(x, y) = f(x+y) - f(x) - f(y). \tag{3}$$

See e.g. J. Aczél, Glasnik mat. fiz. astron. *20*, 65–73, 1965, where also the general solution of the functional equation

$$f(x+y) + f(x-y) = 2f(x) + 2f(y) \tag{4}$$

is given as $f(x)=\frac{1}{2}F(x, x)$ where F is an arbitrary biadditive, symmetric function. *The function F belonging to a given solution f* can be given by

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

The formal coincidence of (5) with (3) suggests more general research on such dualities as that which seems to exist between (4) and the system (1), (2).

J. Aczél

5. Remark to Problem 2, above. The general complex-valued (real-valued) solution of the functional equation

$$F(x, y)\, F(z, t) = F(xz - yt, xt + yz + yt),$$

assumed valid for all real x, y, z and t is

$$F(x, y) = \varphi\left(x + \frac{y}{2} + \frac{i\sqrt{3}}{2}\, y\right)$$

where φ is an arbitrary complex-valued (real-valued) function of a complex variable satisfying

$$\varphi(z\omega) = \varphi(z)\, \varphi(\omega)$$

for all complex z and ω.

The measurable (complex-valued) solutions are $F \equiv 0$, $F \equiv 1$, and

$$F(x, y) = \begin{cases} 0 & \text{if} \quad x = y = 0 \\ \left(x + \frac{y}{2} + \frac{i\sqrt{3}}{2} y\right)^n e^{c \ln (x^2 + xy + y^2)} & \text{if} \quad (x, y) \neq (0, 0) \end{cases}$$

where n is an arbitrary integer and c a complex constant.

The measurable real valued solutions are

$$F \equiv 0, F \equiv 1 \quad \text{and} \quad F(x, y) = \begin{cases} 0 & \text{if} \quad x = y = 0 \\ (x^2 + xy + y^2)^a & \text{if} \quad (x, y) \neq (0, 0) \end{cases}$$

where a is a real constant.

J. A. BAKER

6. Remark. The above remark gives answer also to the other question which I have posed under 2:

$$m(x) = \varphi(x) \quad \text{(the restriction of } \varphi \text{ to reals)}$$

$$g(x) = \varphi\left(x + \tfrac{1}{2} + \frac{i\sqrt{3}}{2}\right).$$

J. ACZÉL

7. Remarque. On considère l'équation suivante:

$$\|f(x + y)\| = \|f(x) + f(x)\| \qquad (*)$$

où cette relation est valide presque partout (au sens de mesure de Lebesgue de deux dimensions). On peut démontrer que pour tous f on peut trouver une fonction F qui a les propriétés suivantes

$$F(x + y) = F(x) + F(y)$$

et $F(x) = f(x)$ presque partout (au sens de measure de Lebesgue d'une dimension).

P. FISCHER

8. Remark to the above problem posed by P. Fischer: I believe that it would be desirable to establish the extension in the following sense (also for the real Cauchy equation): if the functional equation is supposed to be satisfied for all $(x, y) \in S$ then the extension should satisfy it everywhere and be equal to f on

$$S' = \{x \mid \exists y : (x, y) \in S\} \cup \{y \mid \exists x : (x, y) \in S\} \cup \{z = x + y \mid (x, y) \in S\}.$$

(P 43).

J. ACZÉL

9. Remark. Means in n-space.

The following theorem is proven. The case $n=1$ of the theorem was proven by Aczél [*On mean values*, Bull. Amer. Math. Soc. *54*, 392–400 (1948)]. R^n denotes the n-dimensional Euclidean vector space.

THEOREM. *Suppose that M is either an open n-cell or a closed n-cell and that there is a continuous function $(x, y)\to xy$ from $M\times M$ to M which is cancellative ($ax=ay$ or $xa=ya$ implies $x=y$) and medial ($xy\cdot uv=xu\cdot yv$). Suppose further that M contains an idempotent e ($e=ee$) which, in the case that M is a closed n-cell, is not in the bounding $(n-1)$-sphere. Then there exist commuting, invertible linear transformations L_1, L_2: $R^n\to R^n$ and a homeomorphism f of M onto a subset of R^n such that $f(xy)=L_1(f(x))+L_2(f(y))$ for all x, y in M. If, in addition to the above hypotheses, we have commutativity ($xy=yx$) and idempotency ($x=xx$) in M, then there exists a homeomorphism f of M onto a subset of R^n such that $f(xy)=(f(x)+f(y))/2$ for all x, y in M.*

K. SIGMON

10. Problem. V sei ein Vektorraum über einem Körper K, L: $V\to\mathrm{Hom}(V, V)$, eine Abbildung von V in die Algebra der linearen Transformationen von V. L sei linear:

$$L(\alpha x+\beta y)=\alpha L(x)+\beta L(y) \quad (\alpha, \beta\in K;\ x, y\in V).$$

Durch jede solche Abbildung L ist eine Algebra in V bestimmt, wenn man als Multiplikation $(x, y)\to xy\in V$ in V definiert:

$$xy := L(x)\,y \tag{1}$$

($=$Anwendung der linearen Transformation $L(x)$ auf den Vektor $y\in V$). Umgekehrt definiert jede Algebra in V eine solche Abbildung L.

Gesucht ist eine Funktionalgleichung (oder ein System von Funktionalgleichungen) für Abbildungen L mit den genannten Eigenschaften derart, dass die durch die Lösungen $L(x)$ vermöge (1) definierten Algebren eine *neue* Klasse nichtassoziativer Algebren bilden.

Ein Beispiel einer solchen Funktionalgleichung ist

$$M(x)\,L(x)=\mu(x)\,I, \begin{cases} M\colon V\to \mathrm{Hom}(V, V), & \text{linear} \\ \mu\colon V\to K \\ I \text{ die Identität } V\to V. \end{cases}$$

Diese Funktionalgleichung wurde behandelt in W. Eichhorn: Funktionalgleichungen in Vektorräumen, Kompositionsalgebren und Systeme partieller Differentialgleichungen. Aequationes Math., im Druck **(P 44)**.

W. EICHHORN

11. Remark. A result and a problem for Boolean algebras.

From the axiomatic definition of the information we have the following problem, formulated for Boolean algebras. Let A be a Boolean algebra and let G be an arbitrary abelian group (specially, the additive group of real numbers). We define the set A^{*2} by

$$A^{*2} = \{(x, y): x \in A, y \in A, x \cap y = 0\}.$$

Let $\varphi: A^{*2} \to G$ be a mapping, satisfying the conditions

$$\varphi(x, y) = \varphi(y, x) \quad [(x, y) \in A^{*2}] \tag{1}$$

and

$$\begin{gathered}\varphi(x \cup y, z) + \varphi(x, y) = \varphi(x, y \cup z) + \varphi(y, z) \\ [(x, y) \text{ and } (x \cup y, z) \in A^{*2}].\end{gathered} \tag{2}$$

It is trivial, that if $g: A \to G$ is an arbitrary mapping $\varphi: A^{*2} \to G$ is defined by

$$\varphi(x, y) = g(x) + g(y) - g(x \cup y) \qquad (x, y) \in A^{*2},$$

then $\varphi(x, y)$ is a solution of the system of functional equations (1) and (2).

We have proved the following

THEOREM. *Let A be a finite Boolean algebra and G an abelian group. If the mapping $\varphi: A^{*2} \to G$ satisfies the conditions* (1) *and* (2), *then there exists a mapping $g: A \to G$, such that*

$$\varphi(x, y) = g(x) + g(y) - g(x \cup y) \tag{3}$$

holds for all $(x, y) \in A^{*2}$.

For an infinite Boolean algebra we do not know whether the theorem holds or not. This is an open question (**P 45**).

Z. Daróczy

12. Remark. Functional equations on algebraic systems. The functional equation

$$f_0(x + y) + f_1(x - y) = \sum_{k=1}^{n} f_{2k}(x) f_{2k+1}(y) \tag{1}$$

containing several unknown functions $f_i: G \to R$ where $(G, +)$ is a group and $(R, +, \cdot)$ a ring, leads to a generalized additive type equation using identity

$$\begin{aligned} &f_0(x + y + z) + f_1(x + z + y) = \\ &\quad = \sum_{k=1}^{n} [f_{2k}(x + y) f_{2k+1}(z) - f_{2k}(x) f_{2k+1}(z - y) + f_{2k}(x + z) f_{2k+1}(-y)] \end{aligned}$$

and supposing

$$f_0(x + y + z) = f_0(x + z + y) \quad \text{or} \quad f_1(x + y + z) = f_1(x + z + y), \quad x, y, z \in G \tag{2}$$

The method can be applied also for more general equations as e.g.

$$f_0(x \circ y) + f_1(x * y) = F[f_1(x), \ldots, f_n(x); f_{n+1}(y), \ldots, f_{2n}(y)],$$

where F is given and $\circ$, $*$ are given binary operations satisfying certain generalized associative laws resp. suppositions similar to (2).

M. Hosszú

13. Remark. In connection with the discussion of derivations, I would like to call attention to the following result of W. Nöbauer [Funktionen auf kommutativen Ringen, Math. Ann. *147*, 166–175, 1962]: Let $(U, +, \circ, \cdot)$ be the tri-operational algebra of functions over an integral domain I; let R be the set of rational functions in U; and let D be a transformation from R into U satisfying the three conditions

$$D(f+g) = Df + Dg, \tag{1}$$

$$D(f \cdot g) = (Df) \cdot g + f \cdot (Dg), \tag{2}$$

$$D(f \circ g) = D(f \circ g) \cdot Dg. \tag{3}$$

If $|I|$ is finite, then D is trivial in the sense that $Df = 0$ for every f in R. If $|I|$ is infinite, then D is trivial or the ordinary (formal) derivative. This shows that adding the chain rule to the sum and product rules has important consequences.

B. Schweizer

14. Problem. It is known, that e.g. in the family of all power series in two real variables, convergent everywhere, the harmonic functions are exactly those satisfying the functional equation

$$u(x, y) = 2\mathrm{Re}\, u\left(\frac{x+iy}{2}, \frac{y-ix}{2}\right).$$

Is it possible to derive from this a characterisation entirely in terms of real variables? **(P 46).**

Gy. Targonski

15. Remarque. On donne la solution générale de l'équation fonctionnelle matricielle

$$\prod_{i=1}^{n} A_i(x_{i_1}, \ldots, x_{i_{p_i}}) = 0$$

sous l'hypotèse que les variables sont indépendantes et appartiennent à un ensemble arbitraire M; les éléments des matrices considerées sont des fonctions définies sur M avec des valeurs dans un corps commutatif K. Pour trouver la solution de cette équation on a utilisé des résultats de la théorie des matrices semi-inverses et inverses généralisées.

I. a. Pour l'équation matricielle $AX=B$ la condition de comptabilité est: $B=AA^{\perp}B$, et la solution est

$$X = A^{\perp}B + (E - A^{\perp}A)\,U\,.$$

b. On peut écrire la condition de compatibilité sous la forme: $B=AA^{\perp}U$ ou $(E-AA^{\perp})\,B=0$. Des résultats analogues pour l'équation $XA=B$.

II. Soit l'équation $A(x)\,B(y)\,C(z)=0$, qui est un cas particulier de l'équation considérée (les variables sont indépendantes) et le lemme établi par E. Arghiriade (Rev. de math. pure et appl. no. 9. 1967) est le suivant:

L'équation fonctionnelle matricielle $A(x,y)\,B(z)=0$ admet la solution génerale $A(x,y)=A_1(x,y)\,P$, $B(z)=QB_1(z)$, où A_1, B_1 sont des matrices arbitraires, P, Q étant des matrices constantes soumises a la condition: $PQ=0$.

Nous avons: $A(x)\,B(y)=M(x,y)\,P$

$$C(z) = QN(z) \tag{1}$$

avec $PQ=0$, d'où $A(x)\,B(y)\,(E-P^{\perp}P)=0$. Donc

$$A(x) = M_1\,P_1\,; \quad B(y)\,(E - P^{\perp}P) = Q_1N_1(y)\,, \tag{2}$$

avec $P_1Q_1=0$.

La condition de compatibilité étant $N_1(y)=V(y)\,(E-P^{\perp}P)$, nous avons

$$B(y) = Q_1V(y)\,(E - P^{\perp}P) + W(y)\,(P^{\perp}P)\,. \tag{3}$$

Les relations (1), (2), (3) ($P^{\perp}$ est la semi-inverse de P) donnent la solution générale dans le cas consideré. La téchnique utilisée est la même pour le cas générale.

En considérant l'équation de J. Aczél: $F(x,z)=G(x,y)\,H(y,z)$ pour des matrices rectangulaires de la mème manière (dans certaines conditions) nous avons obtenus quelques résultats, de quels nous nous ocuperons dans une autre note.

E. ARGHIRIADE – A. DRAGOMIR

16. Remark on the paper read by J. Aczél.

It appears that one can formulate an aspect of Aczél's lecture and of the paper by A. Moór and L. Pintér (Publ. Math. Debrecen *13*, 207–223, 1966), as follows. Given an algebra F of n times differentiable functions, D_n a not necessarily linear differential operator of order n, and S the solution set $\{f \mid D_n(f)=0\}$. Question: given an $f_0 \in F$, $D_n(f_0)=0$, does there exist an n-parameter family of substitution operators $H_{\bar{r}_n}$, such that $H_{\bar{r}_n}y_0$ is the solution set S, or possibly some interesting subset of S; possibly the choice of the 'generating element' f_0 is also relevant. If D_n is linear, a study of the commutator $[D_n, H_{\tilde{r}_n}]$ may be helpful. It is not impossible that early work by P. Appell (Acta Math. *15*, 281–315, 1891) may have some connections with this. **(P 47)**.

GY. TARGONSKI

17. Bemerkung zum Vortrag von Herrn Professor Fenyő.

Es handelt sich um die Funktionalgleichung

$$f(x)+f(y)=f(x+y-xy)+f(xy) \tag{1}$$

und wir nehmen an, dass f an den Stellen $x=\pm 1$ stetig ist. Mit $y=-x$ folgt aus (1)

$$g(x)\stackrel{\text{def}}{=}f(x)+f(-x)=f(x^2)+f(-x^2)=g(x^2)=\cdots=g(x^{2^k})=\cdots$$

Daraus ergibt sich mit $x=y^{2^{-k}}$ die Gleichung

$$g(y)=g(y^{2^{-k}}) \quad (\mathrm{R}=0,1,2,\ldots).$$

Wegen der Stetigkeit von f bei $x=\pm 1$ folgt auch

$$f(y)+f(-y)=g(y)=\lim_{k\to\infty} g(y^{2^{-k}})=g(1)=2b$$

dass heisst

$$\varphi(x)\stackrel{\text{def}}{=}f(x)-b=-[f(-x)-b]=-\varphi(-x) \tag{2}$$

ist eine ungerade Funktion. Damit geht die Gleichung (1) in

$$\varphi(x)+\varphi(y)=\varphi(x+y-xy)+\varphi(xy) \tag{3}$$

über, und dabei gilt auch

$$\varphi(x)+\varphi(-y)=\varphi(x-y+xy)+\varphi(-xy), \tag{4}$$

die mit $y\Rightarrow -y$ aus (3) folgt. Addieren wir die Gleichungen (3) und (4), so ergibt sich wegen (2)

$$2\varphi(x)=\varphi(x+y-xy)+\varphi(x-y+xy).$$

Dies ist aber die Jensensche Gleichung, was man mit den Substitutionen $x+y-xy=u$ und $x-y+xy=v$ leicht einsieht, das heisst wegen der Stetigkeit von $\varphi(x)$ (obgleich erst bei $x=\pm 1$) erhält man – wie bekannt ist – die Lösung $\varphi(x)=ax+c$; infolge (2) ist $c=0$. Wir haben also $f(x)=ax+b$ als Lösungen unter der oben erwähnten Annahme.

E. VINCZE

18. Problem. Was sind die Lösungen der Funktionalgleichung

$$f_0\left(\sum_{i=1}^{\infty} a_i\right)=\sum_{i=1}^{\infty} f_i(a_i)$$

für die reellen Funktionen $f_0, f_1, f_2, \ldots$ von reellen Veränderlichen, wobei $\sum a_i$ eine (nicht unbedingt absolut) konvergente Reihe mit der Nebenbedingung $|a_i|\geqq|a_{i+1}|>0$ $(i=1,2,\ldots)$ ist **(P 48)**.

E. VINCZE

19. Remark. The solution of Belousov's system of matrix functional equations

(i) $$F(X+Y)=F(X)+F(Y)$$

(ii) $$F[(F(X)-X)(F(Y)-Y)]=0$$

where variables and function values are $n\times n$ real matrices is given under the assumption of boundedness above on a set of positive Lebesgue n^2-dimensional measure.

S. L. SEGAL

20. Remark. Kemperman's question: To find an additive non measurable function Φ on the reals and non-zero constants a_k such that $\sum a_k\Phi(x^k)$ converges precisely when $\sum a_k x^k$ converges.

S. L. SEGAL

21. Remark. On the set of solutions of the translation equation. James Hamilton (Fordham University, Bronx, New-York) found the following results. Let $F(u, v)$ be a function defined for all reals, and a solution of the translation equation

$$F[F(u,v),w]=F[u,v+w]. \tag{1}$$

A one parameter family of functions of one variable is given by

$$F(c,t)=f_c(t), \quad (c=\text{const.})$$

the 'index set' Λ_F is defined as the set of values $F(c, 0)$, where c ranges over all reals. Three statements are given:

I. The function $F(c_1,t)-F(c_2,t)$ either vanishes everywhere, or nowhere, and then $\{f_c(t)\}$ is isomorphic to Λ_F.

The equivalence relation of 'shift relatedness' between $F(c_1, t)$ and $F(c_2, t)$ prevails if

$$\exists a\forall t\; F(c_1,t+a)=F(c_2,t). \tag{2}$$

This relation generates a decomposition of $\{f_c(t)\}$ (and that of the index set) into disjoint classes. Thus e.g. $F(u, v)=ue^v$ has three equivalence classes, $u>0$, $u=0$ and $u<0$. Let now Λ_α be the subset of Λ_F belonging to one equivalence class, and S_α the corresponding subset of $\{f_c(t)\}$. Then

II. The codomain of each function is identical with Λ_α, and conversely, the set of all functions in $\{f_c(t)\}$ with a given codomain forms an equivalence class, and the codomain is the corresponding subset of the index set Λ_F.

Hamilton concludes by studying two different solutions $F(u, v)$ and $G(u, v)$ of (1). Decomposing $\{f_c(t)\}$ and $\{g_c(t)\}=\{G(c,t)\}$ into equivalence classes he finds

III. For a class $S_\alpha \subset \{f_c(t)\}$ to be shift related to a class $V_\beta \subset \{g_c(t)\}$ it is necessary that $\Lambda_\alpha = \Lambda_\beta$. Here shift relatedness is to be understood in the generalized sense $\exists a \forall t\, F(c_1, t) = G(c_2, t+a)$.

GY. TARGONSKI

22. Bemerkung. Einige Bemerkungen über die Cauchy–Binetschen Funktionalgleichung.

Man betrachtet von dem Standpunkte der allgemeinen Lösung die Funktionalgleichungen

$$f_1(xy) = 2f_1(x)\,f_1(y) + C$$
$$f_1(xy) = f_1(x)\,f_1(y) + f_2(x)\,f_3(y) + C$$
$$f_1(xy) = 2f_2(x)\,f_3(y) + C$$

Diese Funktionalgleichungen sind von der Funktion $f_1(x) = f(x, 0, 0, 0)$ verifiziert, wo $f(x, y, z, t)$ die Lösung der Cauchy–Binetschen Funktionalgleichung

$$\begin{aligned} f(\alpha x + \beta y + \gamma z, \alpha_1 x + \beta_1 y + \gamma_1 z, \alpha t + \beta v + \gamma w, \alpha_1 t + \beta_1 v + \gamma_1 w) \\ = f(x, y, t, v)\, f(\alpha, \alpha_1, \beta, \beta_1) + f(x, z, t, w)\, f(\alpha, \alpha_1, \gamma, \gamma_1) \\ + f(y, z, v, w)\, f(\beta, \beta_1, \gamma, \gamma_1) \end{aligned}$$

Man stellt fest, dass die vier Funktionen $f_1(x) = f(x, 0, 0, 0)$, $f_2(x) = f(0, x, 0, 0)$, $f_3(x) = f(0, 0, x, 0)$, $f_4(x) = f(0, 0, 0, x)$ die Matrizengleichung

$$\begin{bmatrix} f_1(xy) & f_2(xy) \\ f_3(xy) & f_4(xy) \end{bmatrix} = \begin{bmatrix} f_1(x) & f_2(x) \\ f_3(x) & f_4(x) \end{bmatrix} \cdot \begin{bmatrix} f_1(y) & f_2(y) \\ f_3(y) & f_4(y) \end{bmatrix} + \begin{bmatrix} C & C \\ C & C \end{bmatrix}$$

erfüllen.

O. GHEORGHIU – B. CRSTICI

* * *

Dies war die erste Funktionalgleichungstagung seit der Einweihung des neuen Gästehauses, dessen moderne und elegante Einrichtungen allgemeinen Beifall fanden. Trotz des Personalmangels haben sich die Teilnehmer aufs freundlichste umsorgt gefühlt. Die Leiter der Tagung möchten die angenehme Pflicht erfüllen, zugleich im Namen aller Teilnehmer der Institutsleitung auf herzlichste für die Ermöglichung der Tagung und ihrer vorzüglichen Organisation zu danken.

Bericht zusammengestellt von I. FENYŐ (Rostock)

Problems and Solutions

This section publishes problems and solutions believed to be new and interesting. Problems are designated by **P1**, **P2**,..., solutions by **P1S1**, **P1S2**, ..., and remarks by **P1R1**, **P1R2**, ... Correspondence regarding this section should be sent to the Problems Editor, Prof. M. A. McKiernan, Faculty of Mathematics, University of Waterloo, Ont., Canada. In case several similar solutions are received, the solutions may be edited with credits given the individual contributors.

P2S1 – Marek Kuczma

Let D_n^r be the space of all functions

$$f = (f_1, \ldots, f_n): R^n \to R^n$$

which are of class C^r in R^n and have a positive Jacobian:

$$\det\left(\frac{\partial f_j}{\partial x_i}\right) > 0 \quad \text{in} \quad R^n.$$

Further, let Q_n^r be the set of all squares of functions from D_n^r:

$$Q_n^r = \{f : f = g^2,\ g \in D_n^r\},$$

where the squares are taken in the sense of the iteration: $g^2 = g(g)$. Then

(i) For any positive integers r, n, we have $D_n^r - Q_n^r \neq \emptyset$.

(ii) For $n = 1$ and for any positive integer r (including $r = +\infty$) every function $f \in D_1^r$ can be represented as a superposition of at most four functions belonging to the class Q_1^r.

This gives a partial answer to a problem by Z. Moszner (Aequationes Math. *1* (1968), 150, **P2**). The details are contained in two papers submitted to Annales Polonici Mathematici:

[1] M. Kuczma, Fractional iteration of differentiable functions.

[2] M. Kuczma, On squares of differentiable functions.

P20S1 – G. Cross (Aeq. Math. Vol. 1, No. 3, pp. 298–300)

Generalized homogeneity in the form $F\{hx, hy\} = \theta(h, x, y)\, F(x, y)$ is studied. A note from Z. Moszner suggests the addition of the forms $f(x, y) = 1/\theta(1/x, x, y)$ for $x = 0$, $f(0, y) = 1/\theta(1/y, 0, y)$ for $y = 0$, to the note by G. Cross.

P23S1 – Z. Moszner (this issue pp. 380)

As announced in Aeq. Math., Vol. 2, No. 1.

P35 – M. Kuczma

Let Δ denote a convex domain in $\mathscr{R}^n$, and consider functions from $\mathscr{R}^n$ to $\mathscr{R}$. Then a set $T \subseteq \Delta$ is said to be of class $\mathscr{A}$ if every convex function on Δ, which is bounded above on T, is also continuous on Δ; $T \in \mathscr{B}$ if every additive function on Δ, bounded above on T, is continuous on Δ; $T \in \mathscr{C}$ if every additive function on Δ, bounded from both sides on T, is continuous on Δ.

Find necessary and sufficient conditions for a set $T \subset \mathscr{R}^n$ (in terms of the nature of the set only) to belong to classes $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$.

P36 – M. Kuczma

Prove or disprove the equality $\mathscr{A} = \mathscr{B}$.

P37 – J. Aczel and P. Erdös

Does there exist a Hamel basis $\mathscr{H}$ such that if $a \in \mathscr{H}$, then also $1/a \in \mathscr{H}$?

Further, is there a Hamel basis such that if $a \in \mathscr{H}$, then all powers of a are in $\mathscr{H}$ (perhaps all positive and negative powers)?

P38 – P. Erdös

Let $f(x)$ be a solution of $f(x+y) = f(x) + f(y)$ and assume that for every x, $\text{Lim}_{k=\infty} f(x^k) = 0$. Does it follow that $f(x) = 0$?

Remark: J. A. Baker and S. L. Segal have shown **P28S1** that a non-measurable solution exists for which $\text{Lim}_{k=\infty} f(x^k) = 0$ for every $|x| < 1$.

P39 – F. Rothberger

Let $f(x)$ be a solution of $f(x+y) = f(x) + f(y)$ and assume that $f(x)$ and $f(1/x)$ always have the same sign. Does it follow that $f(x) = cx$?

P40 – B. Schweizer

Formulated on p. 354, this issue.

P41 – Gy. Targonski

Formulated on p. 361, this issue.

P42 – J. Aczel

Formulated on p. 367, this issue.

P42S1 – J. A. Baker

See Remarks 5 and 6, pp. 368–369, this issue.

P43 – J. Aczel

Formulated as Remark 8, p. 369, this issue.

P44 – W. Eichhorn

Formulated on p. 370, this issue.

P45 – Z. Daroczy

Formulated at the end of remark 11, p. 371, this issue.

P46 – Gy. Targonski

Formulated on p. 372, this issue.

P47 – Gy. Targonski

Formulated in Remark 16, p. 373, this issue.

P48 – E. Vincze

Formulated on p. 374, this issue.

P48S1 – J. A. Baker (this issue p. 387)

The paper characterizes sequences f_i, $i=0, 1, 2 \ldots$, of functions from a real Banach to a real normed linear space such that $f_0(\Sigma_{i=1}^{\infty} x_i) = \Sigma_{i=1}^{\infty} f_i(x_i)$ for every convergent $\Sigma_{i=1}^{\infty} x_i$ with $|x_i| \geqslant |x_{i+1}| > 0$.

P23S1

Sur une hypothèse au sujet des functions subadditives

ZENON MOSZNER (Cracovie)

M. F. Fischer a formulé, pendant la cinquième conférence annuélle des équations fonctionnelles a l'Université de Waterloo, le problème suivant (problème 5 dans [1]).

Considérons la fonction φ qui est définie pour $x>0$ et satisfait aux relations suivantes:

$$\varphi(x) \geqslant 0 \tag{1}$$

$$\varphi(\lambda x) = \lambda\varphi(x) \quad \text{pour tout nombre rationnel positif } \lambda, \tag{2}$$

$$\varphi(x+y) \leqslant \varphi(x) + \varphi(y). \tag{3}$$

Le problème est le suivant: comment peut on caractériser des fonctions φ?

M. Fischer a formulé l'assertion que pour tout φ on peut trouver une fonction f et un espace strictement convexe $\mathscr{H}$ tels que: $f(x)$ satisfait à l'équation de Cauchy pour x réel arbitraire, $f(x)\in\mathscr{H}$,

$$\varphi(x) = \|f(x)\| \quad \text{si} \quad x > 0. \tag{4}$$

Nous démontrerons que *cet assertion n'est pas exacte.*

Soient

a) b un nombre arbitraire, différent de zéro,

b) ψ une fonction non-negative, définie pour chaque $x \neq -b$ et

$$\text{convexe et croissante pour } x < -b, \tag{5}$$

$$\text{convexe et décroissante pour } x > -b, \tag{6}$$

c) $\mathscr{B}$ une base de Hamel, à la quelle appartient b.

Appelons un nombre x libre si le nombre b ne parait pas dans le développement du nombre x sur la base $\mathscr{B}$, et posons

$$\varphi(x) = \begin{cases} 0, & \text{si } x \text{ est un nombre libre et positif}, \\ \psi\left(\dfrac{x}{w} - b\right)|w|, & \text{si } 0 < x = wb + w_1b_1 + \cdots + w_nb_n, \\ & \text{où } w, w_1, \ldots, w_n \text{ sont des nombres rationnels}, \\ & w \neq 0 \text{ et } b, b_1, \ldots, b_n \text{ sont des nombres} \\ & \text{differents de base } \mathscr{B}. \end{cases} \tag{7}$$

On voit facilement que la fonction φ remplit les conditions (1) et (2). Nous allons démontrer qu'elle remplit aussi l'inégalité (3).

Received Mav 22 1968

Considérons les deux cas:

1) $x+y$ est un nombre libre, dans ce cas $\varphi(x+y)=0$, donc (3) a lieu évidemment,

2) $x+y$ n'est pas libre, dans ce cas considérons les deux sous-cas suivants:

(α) x n'est pas libre et y est libre, ou

(β) x et y ne sont pas des nombres libres.

Le cas: x libre et y n'est pas libre se ramène au cas (α) par le changement de désignations. Le cas: x et y sont libres ne peut pas être d'après 2).

Ad (α). Soit $x=wb+r$ où r est un nombre libre. Nous avons d'àprès (7):

$$\varphi(x+y)=\varphi(wb+r+y)=\psi\left(\frac{r+y}{w}\right)|w|,$$

$$\varphi(x)=\psi\left(\frac{r}{w}\right)|w|,\qquad \varphi(y)=0,$$

donc l'inégalité (3) a dans ce cas la forme suivante

$$\psi\left(\frac{r+y}{w}\right)\leqslant\psi\left(\frac{r}{w}\right). \tag{8}$$

Si $w>0$, donc d'après $x=wb+r>0$, nous avons

$$\frac{r}{w}>-b$$

et puisque $y>0$ alors

$$-b<\frac{r}{w}<\frac{r}{w}+\frac{y}{w}.$$

Il en résulte d'après (6) l'inégalité (8).

Si $w<0$, donc

$$\frac{r}{w}+\frac{y}{w}<\frac{r}{w}<-b$$

et d'après (5) nous avons aussi (8).

Ad (β). Soient $x=wb+r$ et $y=\bar{w}b+\bar{r}$ où r et $\bar{r}$ sont libres. Nous avons d'après 2): $w+\bar{w}\neq 0$ et

$$\varphi(x+y)=|w+\bar{w}|\,\psi\left(\frac{r+\bar{r}}{w+\bar{w}}\right),\quad \varphi(x)=|w|\,\psi\left(\frac{r}{w}\right),\quad \varphi(y)=|\bar{w}|\,\psi\left(\frac{\bar{r}}{\bar{w}}\right).$$

Nous devons démontrer que

$$|w+\bar{w}|\,\psi\left(\frac{r+\bar{r}}{w+\bar{w}}\right)\leqslant|w|\,\psi\left(\frac{r}{w}\right)+|\bar{w}|\,\psi\left(\frac{\bar{r}}{\bar{w}}\right). \tag{9}$$

Désignons $r/w=\alpha$, $\bar{r}/\bar{w}=\beta$, $w/(w+\bar{w})=p$. Dans ce cas

$$\frac{\bar{w}}{w+\bar{w}}=1-p$$

et l'inégalité (9) prend la forme:

$$\psi(p\alpha + (1-p)\beta) \leqslant |p|\,\psi(\alpha) + |1-p|\,\psi(\beta). \tag{10}$$

Les nombres x et y étant positifs, donc

$$w(b+\alpha) > 0 \quad \text{et} \quad \bar{w}(b+\beta) > 0$$

et de là nous avons les cas suivants:

I. $w>0$, $\alpha>-b$, $\bar{w}>0$, $\beta>-b$;
II. $w<0$, $\alpha<-b$, $\bar{w}<0$, $\beta<-b$;
III. $w>0$, $\alpha>-b$, $\bar{w}<0$, $\beta<-b$;
IV. $w<0$, $\alpha<-b$, $\bar{w}>0$, $\beta>-b$.

Dans les cas I et II nous avons

$$0 < p < 1 \quad \text{et} \quad 0 < 1-p < 1$$

et puisque la fonction ψ dans l'intervalle $(-\infty, -b)$ et dans l'intervalle $(-b, +\infty)$ est convexe d'après (5) et (6), donc l'inégalité (10) a lieu dans ces cas.

Le cas IV se ramène au cas III par le changement de désignations.

Dans le cas III nous avons $p>1$ si $w>-\bar{w}$ et $p<0$ si $w<-\bar{w}$. Si $p>1$ nous avons

$$-b < \alpha < p\alpha + (1-p)\beta,$$

donc d'après (6)

$$\psi(\alpha) \geqslant \psi(p\alpha + (1-p)\beta).$$

Mais dans ce cas

$$\psi(\alpha) \leqslant p\psi(\alpha) + (p-1)\psi(\beta),$$

donc l'inégalité (10) a lieu.

Si $p<0$ nous avons

$$p\alpha + (1-p)\beta < \beta,$$

donc d'après (5)

$$\psi(p\alpha + (1-p)\beta) \leqslant \psi(\beta) \leqslant -p\psi(\alpha) + (1-p)\psi(\beta),$$

alors l'inégalité (10) a lieu aussi dans ce cas.

Nous avons donc terminée la démonstration de l'inégalité (10).

Supposons à présent que pour la fonction φ, définie plus haut, l'égalité (4) ait lieu.

Pour x libre et positif nous avons $\varphi(x)=0$, donc $f(x)=\tilde{0}$, où $\tilde{0}$ désigne zéro de l'espace $\mathscr{H}$. Puisque f est une fonction additive, il en résulte que

$$f(y+x) = f(y)$$

pour chaque y et pour chaque x libre et positif.

D'où

$$\varphi(x+y) = \|f(y+x)\| = \|f(y)\| = \varphi(y)$$

pour chaque x libre et positif et chaque y positif.

Soit y un nombre positif qui n'est pas libre et posons $y = wb + r$ où r est un nombre libre. Nous avons pour chaque x libre et positif

$$\varphi(x+y) = \psi\left(\frac{r+x}{w}\right)|w| = \varphi(y) = \psi\left(\frac{r}{w}\right)|w|,$$

donc

$$\psi\left(\frac{r+x}{w}\right) = \psi\left(\frac{r}{w}\right)$$

pour chaque x positif et libre et chaque $y = w(b + r/w) > 0$.

Si $w > 0$, alors $r/w > -b$. Puisque y est en outre arbitraire, l'ensemble des nombres r/w est dense dans l'intervalle $(-b, +\infty)$ et puisque l'ensemble des nombres x/w est dense dans l'intervalle $(0, +\infty)$, donc la fonction ψ est constante (étant convexe) dans l'intervalle $(-b, +\infty)$. Par analogie nous pouvons démontrer que cette fonction est aussi constante dans l'intervalle $(-\infty, -b)$ (ce qui n'est pas essentielle dans la suite).

Il résulte de nos considérations que si nous prenons pour la fonction ψ une fonction qui remplit toutes les conditions dans b), n'étant pas en même temps constante dans un des intervalles $(-\infty, -b)$ ou $(-b, +\infty)$ et si nous définissons la fonction φ par la formule (7), la fonction φ montre que l'assertion de M .Fischer mène a une contradiction.

Le problème fondamentale de M. Fischer reste évidemment ouverte. Il se pose aussi un problème complémentaire suivant:

– comment peut on caractériser parmi les fonctions φ qui remplissent les conditions (1), (2) et (3) des fonctions qui remplissent l'affirmation de M. Fischer?

Pour répondre à ce problème remarquons d'abord que si (4) a lieu, alors

$$\varphi(|x|) = \|f(x)\| \quad \text{pour} \quad x \neq 0,$$

puisque la fonction f est paire. En posant $\varphi(0) = 0$ nous avons l'égalité (4) aussi pour $x = 0$. Il en résulte que

$$\varphi(|x_1 + x_2|) = \|f(x_1 + x_2)\| = \|f(x_1) + f(x_2)\| \leqslant \|f(x_1)\| + \|f(x_2)\| = \varphi(|x_1|) + \varphi(|x_2|),$$

donc

$$\varphi(|x_1 + x_2|) \leqslant \varphi(|x_1|) + \varphi(|x_2|) \tag{11}$$

pour chaque x_1 et x_2. *L'inégalité* (11) *est donc une condition nécessaire pour que* φ, *remplissant les conditions* (1), (2) *et* (3), *remplisse aussi l'assertion de M. Fischer.*

Supposons à présent que la condition (11) a lieu pour une fonction remplissant (1), (2) et (3). Dans ce cas l'ensemble

$$E_1 = \{x : \varphi(|x|) = 0\}$$

est un sous-espace linéaire de l'espace linéaire E des nombres réels sur le corps des nombres rationnels. Considérons un élément Q de l'espace quotient E/E_1. Nous avons

$$\varphi(|x|) = \text{const.} \quad \text{pour} \quad x \in Q. \tag{12}$$

En effet si $(x_1 - x_2) \in E_1$, donc

$$\varphi(|x_1|) = \varphi(|x_2 + (x_1 - x_2)|) \leqslant \varphi(|x_2|) + \varphi(|x_1 - x_2|) = \varphi(|x_2|),$$

donc $\varphi(|x_1|) \leqslant \varphi(|x_2|)$. En remplacant x_1 par x_2 et inversement nous avons (12).

On peut démontrer facilement qu'en posant

$$\|Q\| = \varphi(|x|) \quad \text{pour} \quad x \in Q \in E/E_1, \tag{13}$$

nous normons l'espace E/E_1. Ensuite, en prenant pour $f(x)$ la fonction

$$f(x) = Q \quad \text{pour} \quad x \in Q \quad \text{et} \quad Q \in E/E_1,$$

nous voyons que cette fonction remplit les conditions pour la fonction f dans l'assertion de M. Fischer ($\mathscr{H} = E/E_1$) et que (4) a lieu.

L'inégalité (11) *est donc aussi une condition suffisante pour qu'une fonction* φ *remplissant les conditions* (1), (2) *et* (3), *remplisse aussi l'affirmation de M. Fischer pour un espace* $\mathscr{H}$ *pas nécessairement strictement convexe.*

Remarquons enfin que l'espace E/E_1 considéré plus haut est strictement convexe (ou – dans d'autres termes – strictement normé) par la norme (13) si et seulement si pour x_1 et x_2 tels que $\varphi(|x_1|) \neq 0 \neq \varphi(|x_2|)$ on a *l'équivalence* suivante

$$\varphi(|x_1 + x_2|) = \varphi(|x_1|) + \varphi(|x_2|) \Leftrightarrow \bigvee_{\lambda \text{ rationel}} (\lambda > 0 \text{ et } \varphi(|x_2 - \lambda x_1|) = 0). \tag{14}$$

Il se pose aussi la question de savoir s'il existe parmi les fonctions de la forme (7), des fonctions qui remplissent l'assertion de M. Fischer?

Nous savons déjà que dans ce cas

$$\varphi(x) = \begin{cases} c_1 |w| & \text{pour} \quad w > 0, \\ c_2 |w| & \text{pour} \quad w < 0, \end{cases}$$

où w a la même signification que dans la formule (7). D'après (1) nous avons $c_1 \geqslant 0$ et $c_2 \geqslant 0$.

Remarquons que si x a la forme suivante

$$x = wb + w_1 b_1 + \cdots + w_n b_n$$

dans ce cas

$$|x| = x = wb + w_1 b_1 + \cdots + w_n b_n \quad \text{ou}$$
$$|x| = -x = (-w)\, b + (-w_1)\, b_1 + \cdots + (-w_n)\, b_n,$$

donc

$$|x| = (\operatorname{sgn} x)\, wb + (\operatorname{sgn} x)\, w_1 b_1 + \cdots + (\operatorname{sgn} x)\, w_n b_n,$$

où il n'y a d'aucune liaison entre $\operatorname{sgn} x$ et w.

L'inégalité (11) a donc la forme suivante

$$\alpha\, |\bar{w} + \bar{\bar{w}}| \leqslant \beta\, |\bar{w}| + \gamma\, |\bar{\bar{w}}|, \tag{15}$$

où $\bar{w}$ et $\bar{\bar{w}}$ sont les coefficients des nombres réels x_1 et x_2 arbitraires dans la base $\mathscr{B}$ et

a) $$\alpha = \begin{cases} c_1 & \text{si} \quad [\operatorname{sgn}(x_1 + x_2)]\, (\bar{w} + \bar{\bar{w}}) > 0, \\ c_2 & \text{si} \quad [\operatorname{sgn}(x_1 + x_2)]\, (\bar{w} + \bar{\bar{w}}) < 0, \end{cases}$$

b) $$\beta = \begin{cases} c_1 & \text{si} \quad (\operatorname{sgn} x_1)\, \bar{w} > 0, \\ c_2 & \text{si} \quad (\operatorname{sgn} x_1)\, \bar{w} < 0, \end{cases}$$

c) $$\gamma = \begin{cases} c_1 & \text{si} \quad (\operatorname{sgn} x_2)\, \bar{\bar{w}} > 0, \\ c_2 & \text{si} \quad (\operatorname{sgn} x_2)\, \bar{\bar{w}} < 0. \end{cases}$$

Admettons que $0 \leqslant c_1 < c_2$, $\bar{w} > 0$, $\bar{\bar{w}} > 0$, $\operatorname{sgn} x_1 > 0$, $\operatorname{sgn} x_2 < 0$, $\operatorname{sgn}(x_1 + x_2) < 0$. Dans ce cas l'inégalité (15) nous donne

$$c_2 (\bar{w} + \bar{\bar{w}}) \leqslant c_1 \bar{w} + c_2 \bar{\bar{w}},$$

donc $c_2 \leqslant c_1$. Nous recevons la même contradiction dans le cas où $c_2 < c_1$. Il résulte donc de nous considérations que $c_1 = c_2 = c \geqslant 0$.

Passons à présent à l'équivalence (14). Pour $c \neq 0$ (le cas $c = 0$ est banal) le membre droit de cette équivalence a lieu, d'après (7), dans ce cas et seulement dans ce cas s'il existe un nombre rationel λ positif pour lequel le nombre $|x_2 - \lambda x_1|$ est libre. Cette dernière condition a lieu si et seulement si les nombres x_1 et x_2 ont dans la base $\mathscr{B}$ les coefficients $\bar{\bar{w}}$ et $\bar{w}$ pour b de même signe. Le membre gauche de l'équivalence (14) a dans notre cas la forme

$$c\, |\bar{w} + \bar{\bar{w}}| = c\, |\bar{w}| + c\, |\bar{\bar{w}}|,$$

c'est-à-dire la forme

$$|\bar{w} + \bar{\bar{w}}| = |\bar{w}| + |\bar{\bar{w}}|,$$

ce qui a lieu aussi seulement dans le cas si $\bar{w}$ et $\bar{\bar{w}}$ ont le même signe. L'équivalence (14) a donc lieu dans notre cas.

Nous avons donc démontré que *parmi les fonctions de la forme* (7) *seulement les fonctions* $\varphi(x) = c\,|w|$ *où* $c \geqslant 0$ *remplissent l'assertion de M. Fischer.*

A propos des solutions (7) on peut démontrer aussi que

1) *chaque fonction* $\varphi(x)$ *qui remplit les conditions* (1), (2) *et* (3) *et pour laquelle il existe* $b \neq 0$ *tel que* $\varphi(x)=0$ *pour chaque* x *libre et positif, doit être de la forme* (7) *avec la fonction* ψ *remplissant* b),

2) *il existe des solutions* (1), (2) *et* (3) *qui ne sont pas de la forme* (7), par exemple $\varphi(x)=cx$ pour $c>0$.

TRAVAUX CITÉS

[1] *Fifth Annual Meeting on Functional Equations*, Aequationes Math. *1*, 275–305 (1968).

P48S1

On a Problem of E. Vincze

J. A. BAKER (Waterloo)

At the sixth annual meeting on Functional Equations held at Oberwolfach, June 1968, E. Vincze raised the following problem 'Determine all sequences $\{f_i\}_{i=0}^{\infty}$ of real-valued functions of a real variable satisfying

$$f_0\left(\sum_{i=1}^{\infty} x_i\right) = \sum_{i=1}^{\infty} f_i(x_i) \tag{1}$$

for every convergent series $\sum_{i=1}^{\infty} x_i$ such that $|x_k| \geqslant |x_{k+1}| > 0$ $k=1, 2, \ldots$'. Here we answer a slightly more general question. By a convex cone in a real vector space $\mathscr{X}$ we understand a subset $\mathscr{C}$ of $\mathscr{X}$ which is closed under addition and under multiplication by positive scalars ([2] page 135). Note that we do not insist on $0 \in \mathscr{C}$.

THEOREM *Let $\mathscr{X}$ be a real Banach space and let $\mathscr{C}$ be a convex cone in $\mathscr{X}$ with non-void interior. Let* $\mathbf{Y}$ *be a real normed linear space and for $i=0, 1, 2, \ldots,$ suppose $f_i : \mathscr{C} \to \mathbf{Y}$ such that*

$$f_0\left(\sum_{i=1}^{\infty} x_i\right) = \sum_{i=1}^{\infty} f_i(x_i) \tag{1}$$

for every convergent series $\sum_{i=1}^{\infty} x_i$ such that $x_k \in \mathscr{C}$, $\|x_k\| > \|x_{k+1}\| > 0$ for all $k=1, 2, \ldots$ and $\sum_{k=1}^{\infty} x_k \in \mathscr{C}$. Then there exists a unique continuous linear transformation $T : \mathscr{X} \to \mathbf{Y}$ and a sequence $\{\beta_i\}_{i=0}^{\infty} \subset \mathbf{Y}$ such that

$$f_k(x) = Tx + \beta_k \quad \text{for} \quad x \in \mathscr{C} \quad \text{and} \quad k = 0, 1, 2 \ldots \quad \text{and} \tag{2}$$

$$\sum_{k=1}^{\infty} \beta_k = \beta_0 \,. \tag{3}$$

The converse is trivial.

Proof. We may assume that $\mathscr{C}$ contains all its limit points except perhaps 0. If not, consider a subcone $\mathscr{C}'$ of $\mathscr{C}$ generated by some closed ball in $\mathscr{C}$. (i.e. If $\mathscr{B}$ is a closed ball in $\mathscr{C}$ let $\mathscr{C}' = \{tb \mid t > 0 \text{ and } b \in \mathscr{B}\}$.) Then $\mathscr{C}'$ satisfies all the conditions impossed on $\mathscr{C}$ and $\mathscr{C}'$ contains all its limit points except 0. Our proof would then be valid on $\mathscr{C}'$ and the uniqueness gives us the proof for $\mathscr{C}$.

For $x, y \in \mathscr{C}$ let $F(x, y) = f_0(x+y) - f_1(x)$. We will show that $F(x, y)$ is independent of x. In fact, if $x, x', y \in \mathscr{C}$, choose $\{x_k\}_{k=2}^{\infty} \subset \mathscr{C}$ such that

$$\min\{\|x\|, \|x'\|\} > \|x_2\| > \|x_3\| > \cdots > \|x_k\| > \|x_{k+1}\| > \cdots > 0$$

Received October 13, 1968

and

$$\sum_{k=2}^{\infty} x_k = y.$$

Then by (1) we have

$$F(x, y) = \sum_{k=2}^{\infty} f_k(x_k) = F(x', y).$$

Let

$$F(x, y) = g(y) \quad \text{for} \quad x, y \in \mathscr{C}.$$

Then

$$f_0(x + y) = f_1(x) + g(y) \quad \text{for every} \quad x, y \in \mathscr{C}.$$

Now $\mathscr{C}$ is a semigroup under $+$ and hence, (see [1] or [3]) there exists $A: \mathscr{C} \to \mathbf{Y}$ and $\beta_0, \beta_1, \gamma \in \mathbf{Y}$ such that

$$\begin{aligned} A(x + y) &= A(x) + A(y) \\ f_0(x) &= A(x) + \beta_0 \\ f_1(x) &= A(x) + \beta_1 \quad \text{and} \\ g(x) &= A(x) + \gamma \quad \text{for every} \quad x, y \in \mathscr{C}. \end{aligned} \tag{4}$$

By (1) and (4) we have

$$A\left(\sum_{i=1}^{\infty} x_i\right) + \beta_0 = A(x_1) + \beta_1 + \sum_{i=2}^{\infty} f_i(x_i)$$

or, since A is additive,

$$A\left(\sum_{i=2}^{\infty} x_i\right) = (\beta_1 - \beta_0) + \sum_{i=2}^{\infty} f_i(x_i)$$

for every $\sum_{i=2}^{\infty} x_i$ such that $\|x_k\| > \|x_{k+1}\| > 0$, $x_k \in \mathscr{C}$ for all $k = 2, 3, \dots$ and $\sum_{k=2}^{\infty} x_k \in \mathscr{C}$.

Repeating the above argument and using induction we find that for each $k = 0, 1, 2, \dots$ there exists $\beta_k \in \mathbf{Y}$ such that

$$f_k(x) = A(x) + \beta_k \quad \text{for every} \quad x \in \mathscr{C}. \tag{5}$$

Now we show that A has a unique continuous additive extension to $\mathscr{X}$. It is easy to show that $\mathscr{C} - \mathscr{C} = \{x - y \mid x, y \in \mathscr{C}\}$ is a subspace of $\mathscr{X}$. Also, since $\mathscr{C}$ has non-void interior, $\mathscr{C} - \mathscr{C}$ also has non-void interior. Hence $\mathscr{C} - \mathscr{C} = \mathscr{X}$. Let $T: \mathscr{X} \to \mathbf{Y}$ be defined by $T(x - y) = A(x) - A(y)$ for $x, y \in \mathscr{C}$. It is easily verified that T is well-defined, additive and that T extends A.

Suppose T is not continuous. Then T is not bounded on any of the bounded sets

$$S_n = \left\{x \in \mathscr{C} \,\middle|\, \frac{1}{2^{n+1}} < \|x\| < \frac{1}{2^n}\right\}$$

since these sets all have non-void interior. Thus there exists $x_k \in \mathscr{C}$ such that

$$\frac{1}{2^{k+1}} < \|x_k\| < \frac{1}{2^k}$$

and

$$|f_k(x_k)| = |T(x_k) + \beta_k| > 1$$

for each $k = 1, 2, \ldots$. But this contradicts (1) since $\sum_{k=1}^{\infty} x_k$ converges. Hence T is continuous. Therefore T is linear.

It follows from (1), (5) and the continuity of T that $\beta_0 = \sum_{k=1}^{\infty} \beta_k$.

It remains to prove uniqueness. Suppose $f_k(x) = T'(x) + \beta'_k$ for all $x \in \mathscr{C}$, $k = 0, 1, 2 \ldots$ where $T' : \mathscr{X} \to \mathbf{Y}$ and $\{\beta'_k\}_{k=0}^{\infty} \subset \mathbf{Y}$.

Then

$$T(x) + \beta_k = T'(x) + \beta'_k$$

for all x in some open set U, interior to $\mathscr{C}$. Thus the linear map $T - T'$ is constant on the open set U.

Hence

$$T - T' \equiv 0 \quad \text{and} \quad \beta_k = \beta'_k \quad \text{for all} \quad k = 0, 1, 2 \ldots.$$

REFERENCES

[1] J. Aczel, *On a Generalization of the Functional Equations of Pexider*, Publ. Inst. Math., Beograd *4* (18), 77–80 (1964).

[2] R. E. Edwards, *Functional Analysis* (Holt, Rinehart and Winston, New York 1965).

[3] E. Vincze, *Über eine Verallgemeinerung der Pexiderschen Funktionalgleichungen*, Studia Univ., Babes-Bolyai (1962).

Short Communications

This section consists of self-contained 25–100 line short communications which are prepublications of results, the details of which are to be published in either *aequationes mathematicae* or other journals with referee systems or equivalent systems of pre-reviewing papers. *Aequationes mathematicae* will endeavour to publish these short communications in the shortest possible time after the underlying papers have been accepted for publication. Unless indicated otherwise details of results described in this section will appear in subsequent issues of *aequationes mathematicae.*

A Less Formal Approach to Kaluza–Klein Formalism*)

M. A. McKiernan

In the space V^4 of relativity, coordinate y (or y^i) and metric $g(y; \dot{y}, \dot{y})$ (or g_{ij}), let $y = \phi_\sigma(x)$ be a one parameter group of null translations; that is, if

$$\mathring{\phi}_\sigma(x) \overset{\mathrm{df}}{=} \frac{d}{d\sigma}\phi_\sigma(x); \qquad B(y) \overset{\mathrm{df}}{=} \mathring{\phi}_\sigma(\phi_\sigma^{-1}(y)); \qquad \bar{B}(x) \overset{\mathrm{df}}{=} \mathring{\phi}_\sigma^{-1}(\phi_\sigma(x))$$

then we assume $g_{ij} B^i B^j = 0 = g_{ij} \bar{B}^i \bar{B}^j$ and $\underset{B}{\mathscr{L}} g_{ij} = 0$, where $\underset{B}{\mathscr{L}}$ denotes the Lie derivative. If σ is allowed to vary with the proper time λ of a particle, say $\dot{\sigma} = \dot{\lambda}$ where

$$\lambda = \int \sqrt{g_{ij}\, \dot{y}^i \dot{y}^j}\, d\tau, \tag{1}$$

then in the x frame of reference the action integral has the form

$$\lambda = \int \sqrt{h_{ij}(x)\, \dot{x}^i \dot{x}^j} - \bar{B}_i(x)\, \dot{x}^i\, d\tau \tag{2}$$

where $h_{ij} = g_{ij} + \bar{B}_i \bar{B}_j$, $h_{ij} \bar{B}^i \bar{B}^j = 0$, $\underset{B}{\mathscr{L}} h_{ij} = 0$, and $h_{ij} \bar{B}^j = g_{ij} \bar{B}^j$. Conversely, if (2) is given, and $\bar{B}_i$ is a null translation in the h metric, then $x = \phi_\lambda^{-1}(y)$ exists, transforming (2) into (1).

This analysis suggests a modified Kaluza-Klein formalism in a cylindrical V^5. The modified formalism is used to show that if $F_{ij} \overset{\mathrm{df}}{=} \bar{B}_{i,j} - \bar{B}_{j,i}$ is the electromagnetic field tensor, then the Ricci tensors $\underset{g}{R_{ij}}$ and $\underset{h}{R_{ij}}$ derived from the g and h metrics satisfy

*) Cf. pp. 207–226.

$$\underset{g}{R} = \underset{h}{R} - \tfrac{1}{4}(F_{ij} F^{ij}) \quad \text{and} \quad \bar{B}^i \bar{B}^j \underset{g}{R_{ij}} = -\tfrac{1}{4} F_{ij} F^{ij},$$

while det F_{ij} always vanishes. Hence if the original g metric is flat, this process leads to null fields, $\vec{E} \cdot \vec{H} = 0$, $\vec{E}^2 - \vec{H}^2 = 0$. It can also be shown that $F^{ij}_{/j} = 2 \bar{B}_j \underset{g}{R^{ij}}$, where '/' refers to covariant differentiation in either the g or h metrics, whence no current can be obtained from a flat space.

Stability of General Systems of Linear Equations*)

VICTOR PEREYRA

A common problem in many applications is that of finding minimal least squares solutions to systems of simultaneous linear equations. Given an $m \times n$ matrix A and an m-vector b a *minimal least squares solution* to $Ax = b$ is the unique n-vector x that minimizes $\|Ax - b\|_2$ and has minimum norm among all those vectors having such a property. It is well known that the solution to this problem is given by $x = A^+ b$, where A^+ is the Moore-Penrose generalized inverse of A.

When the rank of A is well determined this problem can be solved numerically with relative ease. When the rank of A is not well defined then the problem becomes ill-conditioned and most known methods encounter difficulties in providing the minimal least squares solution.

The purpose of this paper is to introduce a new problem, related to that of finding minimal least squares solutions, which will be well posed even in the case of rank indeterminacy.

PROBLEM I: 'Given A, b as above, ε, δ two positive numbers, find B, a subset of linearly independent columns of A, such that $\tilde{A} = BB^+ A$ has the property that for any rank preserving matrix perturbation ΔA with

$$\frac{\|\Delta A\|}{\|A\|} \leqslant \delta, \quad \frac{\|[(\tilde{A} + \Delta\tilde{A})^+ - A^+] b\|}{\|\tilde{A}^+ b\|} \leqq \varepsilon$$

holds. Among all the B's having this property the one making $\|(I - BB^+) A\| = \|A - \tilde{A}\|$ a minimum should be chosen'.

In order to be able to solve Problem I we develop a set of inequalities which permit us to estimate the effect that perturbations on the data have on the computed minimal least squares solutions.

*) Cf. pp. 194–206.

Beziehung der ebenen verallgemeinerten nichteuklidischen Geometrie zu gewissen Flächen im pseudominkowskischen Raum *)

O. VARGA (Budapest)

Bekannt ist folgendes: In der Fernebene eines pseudoeuklidischen Raumes ist die hyperbolische nichteuklidische Geometrie gültig. Ein Flächenmodell in diesem Raum liefert eine Schale desjenigen zweischaligen Hyperboloides, das konjugiert zu dem die Massbestimmung festlegenden einschaligen Hyperboloid ist. Genauer besagt das obige, dass das Bogenelement des zweischaligen Hyperboloides, gemessen in der Metrik des zu ihm konjugierten einschaligen Hyperboloides übereinstimmt mit dem Bogenelement derjenigen hyperbolischen Geometrie, die in der Fernebene durch die Schnittellipse des gemeinsamen Asymptotenkegels der betrachteten Hyperboloide induziert wird.

In vorliegender Arbeit wird eine Verallgemeinerung der obigen Fragestellung untersucht. Man gehe von einer zentralsymmetrischen konvexen Fläche eines affinen Raumes aus. Durch Einführung einer Fernebene gehe man zu einem projektiven Raum über. Die durch den Mittelpunkt der konvexen Fläche gehende, und diese in einer zentralsymmetrischen Kurve K schneidende Ebene, längs der es einen umschriebenen Zylinder gibt, soll durch eine Kollineation zur Fernebene gemacht werden. Durch Auszeichnung dieser Ebene erhält man schliesslich einen affinen Raum in der die ursprüngliche – ebenfalls der Kollineation unterworfene – Fläche gestaltlich einem zweischaligen Hyperboloid ähnelt. Aus dem Zylinder wurde dabei ein Kegel. Beide Flächen schneiden die neue Fernebene in der Kurve K. Diese Kurve induziert nach D. HILBERT eine verallgemeinerte (hyperbolische) nichteuklidische Geometrie. Schliesslich kann man eine solche Fläche einführen, die im Äusseren des Kegels liegt, denselben zum Asymptotenkegel besitzt und in Bezug auf den Ursprung sternförmig ist. Wir bezeichnen den Raum, in dem diese Fläche eine Massbestimmung festlegt, als pseudominkowskisch. Er ist die naturgemässe Verallgemeinerung des pseudoeuklidischen Raumes. Das Ergebnis der Arbeit besteht in der Ermittlung des Zusammenhanges des Bogenelementes das zu Hilbertschen Massbestimmung gehört, mit dem Bogenelement des verallgemeinerten zweischaligen Hyperboloides, gemessen in der pseudominkowskischen Metrik. Ist das Bogenelement auf der Fläche ds_M und das, durch Zentralprojektion in der Fernebene erhaltene in der Hilbertschen Massbestimmung gemessene ds_H, so stimmen beide Bogenelemente bis auf einen, durch eine eindeutige Funktion bestimmten Faktor überein. Im klassischen Falle ist diese Funktion identisch gleich eins.

*) Eingegangen am 3.1.1968.

Semigroup Actions and Dimension*)

DAVID STADTLANDER

Given a topological semigroup T and a Hausdorff space X [called the state-space], an act (Act) is a continuous function from $T \times X$ into (onto) X [whose values are denoted by juxtaposition] such that $s(tx)=(st)x$ for each $s, t \in T$ and $x \in X$. We say that T acts (Acts) on X. Groups, semigroups, transformation groups and automata furnish some of the many examples of acts.

In 'Thread Actions' [to appear, Duke Math. J.], the author gave a set-theoretic definition of a class $\mathscr{K}$ of pairs (X, A); X a compact metric space and A a closed subset of X. It was shown there that given a compact metric space X, a closed subset A of X and a metric thread semigroup T, $(X, A) \in \mathscr{K}$ iff T Acts on X with $OX = A$. In particular then, if $(X, A) \in \mathscr{K}$, A is a strong deformation retract of X.

Given a compact semigroup T Acting on a compact space X, we introduce an equivalence relation δ on X [$x \delta y$ iff $x \cup Tx = y \cup Ty$] which partitions the state-space according to the mobility of its points under mappings from the semigroup T. It is shown that if T is a normal semigroup [$tT = Tt$ for each $t \in T$], then to each $x \in X$, there corresponds an idempotent $e \in T$ such that $\langle H(e), \delta[x] \rangle$ is a topological transformation group transitive on $\delta[x]$. This result is applied to relate $\dim X$ and $\dim T$ for normal semigroups T and also to obtain the description of a state-space given in

THEOREM 2.12: *Let the metric continuum semigroup T with identity [not a group] Act on the compact metric space X. Then there is a space Y, an idempotent $e \in K(T)$ [the minimal ideal of T], and a monotone mapping p of X onto Y such that*

(1) $(Y, p(eX)) \in \mathscr{K}$ and

(2) $p^{-1}p(x)$ is the underlying space of an abelian topological group for each $x \in X$.

If X is non-degenerate and finite dimensional,

$$\dim(p^{-1}p(x)) < \dim(X/eX) \leqslant \dim X \quad \text{for each} \quad x \in X - eX.$$

Actions by one-dimensional semigroups are also studied. In particular we obtain the two results which follow.

COROLLARY 3.13: *Let the one-dimensional metric continuum semigroup T with identity* Act *on the compact metric space X. Letting p denote the natural map of X upon X/δ, $(X/\delta, p(K(T)X)) \in \mathscr{K}$.*

THEOREM 3.5: *Let the one-dimensional metric continuum semigroup T with identity act on the space X. Let $Y = Tx$ for some $x \in X$.*

---

*) Received Feburuary 22, 1968.

I. If $K(T)$ consists of left zeros and
   (a_1) T is locally connected or
   (b_1) The complement in T of the set of weak cutpoints is countable,
   Then $\dim Y \leqslant 1$.
II. If $K(T)$ is a group and
   (a_2) $T/K(T)$ is locally connected or
   (b_2) The complement in $T/K(T)$ of the set of weak cutpoints is countable,
   Then $\dim Y \leqslant 1$.
III. Otherwise if (a_2) or (b_2) holds, then
   $\dim Y \leqslant \max\{1, \dim[x \in Y \mid Tx = \{x\}]\}$.

Theorem 3.5 and its corollaries extend the dimension theorems given by L. W. Anderson and R. P. Hunter in 'Homomorphisms and dimension' [Math. Annalen *147* (1962), 248–268] and 'On one dimensional semigroups' [Math. Annalen *146* (1962), 383–396].

Divergence-Free Tensorial Concomitants *)

D. Lovelock

In an n-dimensional manifold with local coordinates $x^i (i=1,\dots,n)$ we consider the n^2 quantities $g_{ij}=g_{ij}(x^h)$ which are the components of a symmetric tensor of covariant valency two with nonvanishing determinant g. If L is a scalar density which is a function of g_{ij} and its first two derivatives i.e.

$$L = L(g_{ij}, g_{ij,k}, g_{ij,kh}) \tag{1}$$

then we may associate with it the Euler-Lagrange expression

$$E^{ij}(L) \equiv \frac{\partial L}{\partial g_{ij}} - \frac{\partial}{\partial x^h}\left(\frac{\partial L}{\partial g_{ij,h}}\right) + \frac{\partial^2}{\partial x^h \partial x^k}\left(\frac{\partial L}{\partial g_{ij,kh}}\right). \tag{2}$$

It has been shown by Rund that E^{ij} are the components of a symmetric tensor density which is divergence-free i.e.

$$E^{ij}{}_{|j} \equiv 0,$$

where the vertical bar denotes partial covariant differentiation. In general E^{ij} will be of fourth order in g_{ij}. In the theory of general relativity a particular scalar density of the type (1) is used, viz.

$$L = a\sqrt{g}\,R + b\sqrt{g}$$

where a and b are constants and R is the curvature scalar, and the corresponding

*) Received September 16, 1968.

Euler-Lagrange expression (2) is

$$E^{ij} = a\sqrt{g}(R^{ij} - \tfrac{1}{2}g^{ij}R) - \tfrac{1}{2}b\sqrt{g}\, g^{ij} \tag{3}$$

where R^{ij} is the Ricci tensor. However, this E^{ij} is of *second* order in g_{ij}. In a previous article the problem of characterising scalar densities (1) for which E^{ij} is of third and second order in g_{ij} was discussed and, for $n=2$, 3 and 4 all such scalar densities were exhibited.

The present article is concerned with tensor densities of contravariant valency two, the components A^{ij} of which are symmetric and satisfy the conditions:

(a) A^{ij} is a concomitant of g_{hk} and its first two derivatives, i.e.

$$A^{ij} = A^{ij}(g_{hk}, g_{hk,r}, g_{hk,rs});$$

(b) A^{ij} is divergence-free, i.e.

$$A^{ij}{}_{|j} \equiv 0 .$$

All such tensor densities are constructed and exhibited. It is found that the number m of independent tensor densities with these properties depends crucially on the dimension of the space, in fact

$$m = \begin{cases} n/2 & \text{if } n \text{ is even}, \\ (n+1)/2 & \text{if } n \text{ is odd}. \end{cases}$$

It is furthermore shown that every such A^{ij} is the Euler-Lagrange expression E^{ij} corresponding to a suitably chosen L of the type (1), viz.

$$L = c g_{ij} A^{ij},$$

where c is a constant. If $n=4$ the only tensor density with these properties is the one usually used in general relativity, viz. (3).

A Quasi-Monte Carlo Method for Computing Double and Other Multiple Integrals *)

S. K. Zaremba

Independently of each other, Hlawka [1] and Korobov [2] showed that integrals over multidimensional unit cubes could be efficiently computed by averaging integrand values over sets of points formed by the multiples, reduced modulo 1 in each coordinate, of suitable lattice points. This applies particularly to functions represented by absolutely convergent multiple Fourier series. Unfortunately, no method of finding good lattice points in more than two dimensions other than by tedious trial and error is known at present. Lattice points offered by Korobov, although described

*) Received March 5, 1968.

by him as "optimum", are by no means the best possible. On the other hand, the present author proved [3] that in two dimensions lattice points of the form of

$$\langle 1/u_n, u_{n-1}/u_n \rangle, \tag{*}$$

where u_n is the n-th Fibonacci number, are best in a well-defined sense. Computation methods using such lattice points appear to be more efficient than the classical methods based on iterated integration.

In order to admit the required multiple Fourier expansion, the integrand must be capable of being redefined outside the unit cube so as to produce a function which is periodic with a unit period in each coordinate, while remaining sufficiently smooth. There are ways of reducing general integrands to such functions. A method proposed by KOROBOV consists of introducing a change of variables as a result of which the new integrand vanishes together with its partial derivatives up to a required order on the boundary of the multidimensional cube. However, in two dimensions such a method is bound to produce inside the square inordinately large values of the mixed partial derivatives of the integrand, entailing a corresponding increase in the error of integration, and this effect would become rapidly accentuated with an increase in the number of dimensions. Instead, the present author suggests adding to the integrand, say $f(x, y)$, a polynomial in y with coefficients depending on x, and a polynominal in x with coefficients depending on y. If r is the degree of these polynomials, they can easily be so fitted as to ensure that the values of the modified function f^* and of its partial derivatives up to

$$\frac{\partial^{2r-2} f^*}{\partial x^{r-1}\, \partial y^{r-1}}$$

agree on opposite sides of the unit square, the contributions of these polynomials to the integral over the square vanishing identically. If the original function has a partial derivative

$$\frac{\partial^{2r} f}{\partial x^r\, \partial y^r}$$

of bounded variation in the sense of HARDY and KRAUSE, the error of integration is $0(\log u_n/u_n^{r+1})$ when the u_n distinct multiples modulo 1 of (*) are used. This method of converting the integrand can obviously be extended to an arbitrary number of dimensions; the amount of work involved grows, roughly speaking, only linearly with the dimension.

Pending the availability of really good lattice points in more than two dimensions, trial computations were carried out in two dimensions by the proposed method. For comparison the same integrals were also computed by classical methods, as well as by KOROBOV's method. The latter method, compared with that proposed by the present author, produced errors about 10 to 1,000,000 times greater for the same integrand with the same number of points.

REFERENCES

[1] HLAWKA, E., *Zur angenäherten Berechnung mehrfacher Integrale*, Monatsh. Math. *66*, 140–151 (1962).
[2] KOROBOV, N. M., *Teoretikochislovye metody v priblizhennom analize* [Numbertheoretical Methods in Numerical Analysis] (Fizmatgiz, Moscow 1963).
[3] ZAREMBA, S. K., *Good Lattice Points, Discrepancy and Numerical Integration*, Ann. Mat. Pura Appl. (iv) *73*, 293–318 (1966).

On Collineations on Three and on Four Lines in a Projective Plane *)

VÁCLAV HAVEL (Brno)

By a collineation on a point subset of a projective plane is meant a map $\sigma: C \to C$ sending any three collinear points of C onto collinear points.

In the sequel let P be a projective plane with ternary ring (S, τ) according to reference points O (origin), X (improper point of the first axis), Y (improper point of the second axis) and Z (unity point). The results are as follows: *1.* If σ is a collineation on $OX \cup OY \cup XY$ such that $O^\sigma = O$, $X^\sigma = X$, $X^{\sigma^{-1}} = \{X\}$; $Y^\sigma = Y$, $Y^{\sigma^{-1}} = \{Y\}$; $(OX)^\sigma \neq \{O, X\}$, $(OY)^\sigma \neq \{O, Y\}$, then there exist maps α, β, γ of S into S such that $O^\alpha = O^\beta = O^\gamma = O$, $O^{\alpha^{-1}} = O^{\beta^{-1}} = O^{\gamma^{-1}} = \{O\}$, $S^\alpha \neq \{O\}$, $S^\beta \neq \{O\}$, $S^\gamma \neq \{O\}$ and $\tau(a, b, c) = 0 \Rightarrow \tau(a^\alpha, b^\beta, c^\gamma) = 0$. *2.* The converse of *1.* holds true too. *3.* As a corollary to *1.*, if (S, τ) satisfies the linearity condition and the additive right inverse property, then various consequences are deduced. *4.* If (S, τ) satisfies the linearity condition and if ϱ is a collineation on $OX \cup OY \cup XY \cup YZ$ with the same assumptions as in *1.*, then $1^\alpha = 1$, $\beta = \gamma$ and β is an additive homomorphism. *5.* The converse of *4.* is also true. *6.* As a corollary to *4.*, if especially (S, τ) satisfies the additive right inverse property and if $Z^\sigma = Z$, then $\alpha = \beta = \gamma$ and α is an endomorphism of (S, τ).

A Sine Functional Equation **)

J. A. BAKER

Let f be a complex-valued function defined on the real vector space X and satisfying

$$f(x+y)f(x-y) = f(x)^2 - f(y)^2$$

*) Received April 23, 1968.
**) Received May 1, 1968.

for all x, $y \in X$. If f is continuous along rays then either (i) f is linear (additive and real homogeneous) or (ii) $f(x) = c \sin L(x)$ for all $x \in X$ where c is complex constant and L is linear. If X has a linear topology, f is continuous and (ii) holds, then L is continuous. If $X = R^n$ and f is measurable on some subset of positive Lebesque measure, then f is continuous. The real-valued case is also considered.

Endomorphismen von ebenen Viergeweben (Beitrag zu einem Problem von J. Aczél)*)

Václav Havel (Brno)

Es sei **P** eine projektive Ebene und **Q** ein Viergewebe in **P**, das sämtliche Punkte von **P** enthält. In **P** werden die Referenzpunkte geeignet gewählt, und der zugeordnete Ternärkörper T von **P** soll die Linearitätsbedingung erfüllen. Es werden drei Sälze abgeleitet, in denen die Endomorphismen von **Q** durch die Eigenschaften von T charakterisiert werden. Im Falle eines Viergewebes **Q** erster Art werden die Endomorphismen von **Q** (mit einem zusätzlichen Fixpunkt) durch gewisse additive Endomorphismen von T ausgedrückt. Im Falle eines Viergewebes **Q** zweiter Art werden die Endomorphismen von **Q** (wieder mit einem zusätzlichen Fixpunkt) einfach durch nichttriviale Endomorphismen von T ausgedrückt. Im Falle eines Viergewebes **Q** dritter Art und eines T mit Gruppenaddition werden die Endomorphismen von **Q** durch nichttriviale Endomorphismen von T ausgedrückt.

On the Fundamental Approximation Theorems of D. Jackson, S. Bernstein and Theorems of M. Zamansky and S. B. Stečkin **)

P. L. Butzer and K. Scherer

The direct theorems of D. Jackson and inverse theorems of S. Bernstein as well as their generalizations by A. Zygmund play a fundamental role in approximation of periodic functions by trigonometric polynomials. Of further importance are results by M. Zamansky and S. B. Stečkin.

It is first shown that the assertions of the theorems of Jackson, Bernstein, Zamansky and Stečkin are equivalent to another for polynomials of best approximation. More precisely, let $t_n(x)$ be a trig. polynomial of degree $\leqslant n$ and T_n the corresponding linear space. Let

$$E_n(f) = \inf_{t_n \in T_n} \|f - t_n\|_{C_{2\pi}} \quad (f \in C_{2\pi};\ n \in \mathbb{N}),$$

and $t_n^* = t_n^*(f; x) \in T_n$ be the polynomial of best approximation corresponding to

*) Received May 9, 1968.
**) Received May 13, 1968.

$f \in C_{2\pi}$. We write $f \in \mathrm{Lip}^* \alpha$ if

$$|f(x+h) + f(x-h) - 2f(x)| \leqslant M|h|^\alpha \quad \text{for all} \quad x.$$

THEOREM A. *Let $f \in C_{2\pi}$. The following assertions are equivalent for $0<k<r+\alpha<l$ and $0<\alpha<2$ $(r, k, l \in \mathbb{N})$:*

a) $E_n(f) = \|f - t_n^*(f)\|_{C_{2\pi}} = O(n^{-r-\alpha})$;
b) $f^{(k)} \in C_{2\pi}$ *and* $\|f^{(k)} - t_n^{*(k)}(f)\|_{C_{2\pi}} = O(n^{k-r-\alpha})$;
c) $\|t_n^{*(l)}(f)\|_{C_{2\pi}} = O(n^{l-r-\alpha})$;
d) $f^{(r)} \in \mathrm{Lip}^* \alpha$.

The implication a)⇒d) is the theorem of Bernstein, d)⇒a) is that of Jackson, a)⇒b) is that of Stečkin, a)⇒c) is that of Zamansky. The theorem also asserts that the converses of the Zamansky and Stečkin results are valid for $C_{2\pi}$-functions.

Secondly, the latter theorem is generalized to Banach spaces in the setting of the theory of intermediate spaces. Indeed, let X be a (B)-space, $P_0 = \{0\} \subset P_1 \subset P_2 \subset \cdots \subset P_n \subset \cdots$ a sequence of subspaces of X, and

$$E_n(f) = \inf_{p_n \in P_n} \|f - p_n\|_X \quad (f \in X;\ n \in \mathbb{N}).$$

Suppose that there exists an element $p_n^*(f) \in P_n$ of best approximation to $f \in X$. Let Y be an approximation space of X, thus $a(B)$-subspace such that $P_n \subset Y \subset X (n \in \mathbb{N})$ with continuous embedding. Generalizing the notions of the classical inequalities of Jackson and Bernstein it is further assumed that a Jackson and Bernstein type inequality of order σ, $\sigma \geqslant 0$, is satisfied with respect to Y:

$$E_n(f) \leqslant C n^{-\sigma} \|f\|_Y \quad (f \in Y;\ n \in \mathbb{N})$$
$$\|p_n\|_Y \leqslant D n^{\sigma} \|p_n\|_X \quad (p_n \in P_n;\ n \in \mathbb{N}),$$

C and D being positive constants.

THEOREM B. *Let Y_1 and Y_2 be two approximation spaces of X such that the preceding inequalities are satisfied with respect to Y_1 and Y_2 with orders σ_1 and σ_2, respectively. The following assertions are equivalent for $0 \leqslant \sigma_1 < \theta < \sigma_2$ and $1 \leqslant q \leqslant \infty$:*

a) $\{n^\theta E_n(f)\} \in l_*^q$, *i.e.* $\left\{\sum_{n=1}^{\infty} [n^\theta E_n(f)]^q \frac{1}{n}\right\}^{1/q} < \infty$;
b) $f \in Y_1$ *and* $\{n^{\theta-\sigma_1} \|f - p_n^*(f)\|_{Y_1}\} \in l_*^q$;
c) $\{n^{\theta-\sigma_2} \|p_n^*(f)\|_{Y_2}\} \in l_*^q$;
d) $f \in (X, Y_2)_{\theta/\sigma_2, q; K}$.

In case $q=\infty$, assertion a) is to be interpreted as $\sup_n [n^\theta E_n(f)]<\infty$; similarly for b) and c). (This case contains Theorem A.) Moreover, for $1\leqslant q\leqslant\infty$, $0<\theta<1$,

$$(X, Y)_{\theta, q; K} = \left\{f\in X: \left(\int_0^\infty [t^{-\theta}K(t,f)]^q \frac{dt}{t}\right)^{1/q} < \infty\right\},$$

where for $0<t<\infty, f\in X$ the function norm

$$K(t,f;X;Y) = \inf_{f=f_1+f_2,\ f_2\in Y} (\|f_1\|_X + t\,\|f_2\|_Y).$$

Note that the latter spaces are intermediate spaces of X and Y under the corresponding norm, thus (B)-spaces such that $Y\subset(X, Y)_{\theta, q; K}\subset X$. In case $q=\infty$, assertion d) reads $K(t,f;X;Y_2)=O(t^{\theta/\sigma_2})$ and in this sense it is a generalization of the Lipschitz spaces.

The theorems stated above are typical of those established. An essential fact is that all proofs rest upon the Jackson and Bernstein type inequalities and in all applications only these need be verified.

Thirdly, the counterpart of Theorem B is established for a general class of approximation processes on (B)-spaces. These are bounded, commutative linear operators on X to itself which approximate the identity and satisfy modified Jackson and Bernstein-type inequalities. In particular, 'polynomial' operators V_n are studied, i.e. those having the property $V_n(f)\in P_n$, $n\in\mathbb{N}, f\in X$.

New applications are possible to the spaces $C_{2\pi}$, $C(T_n)$; $L^p(T_n)$ (T_n being the n-dimensional torus) and $L^p(E_n)$, $1\leqslant p\leqslant\infty$, and to the singular integrals of Fejér, Jackson, Rogosinski-Bernstein, Riesz.

Nerves of Simplicial Complexes*)

BRANKO GRÜNBAUM

If a finite simplicial complex $\mathscr{K}$ is considered as covered by its family of facets (i.e., maximal closed simplices, principal simplices), one may define in the usual fashion the nerve $\mathscr{N}(\mathscr{K})$ of $\mathscr{K}$, which is also a simplicial complex. The following results are typical of those established on the relation between $\mathscr{K}$ and $\mathscr{N}(\mathscr{K})$.

THEOREM 1. *A simplical complex $\mathscr{M}$ is isomorphic to the complex $\mathscr{N}(\mathscr{K})$ for some simplicial complex $\mathscr{K}$ of dimension d if and only if each edge (i.e., 1-simplex) of $\mathscr{M}$ belongs to at most d facets of $\mathscr{M}$, and each vertex of $\mathscr{M}$ belongs to at most $d+1$ facets of $\mathscr{M}$.*

Let a simplicial complex $\mathscr{K}$ be called *taut* provided each of its vertices is the

*) Received June 14, 1968.

intersection of facets of $\mathscr{K}$. Triangulated manifolds or pseudomanifolds are taut, as are many other interesting classes of complexes.

THEOREM 2. *If $\mathscr{K}$ is a taut complex then $\mathscr{N}(\mathscr{K})$ is taut and $\mathscr{K}$ is isomorphic to $\mathscr{N}^2(\mathscr{K})=\mathscr{N}(\mathscr{N}(\mathscr{K}))$. Conversely, if $\mathscr{K}$ is isomorphic to $\mathscr{N}^2(\mathscr{K})$ then $\mathscr{K}$ is taut.*

Theorem 2 shows the existence of a duality between $\mathscr{K}$ and $\mathscr{N}(\mathscr{K})$ for taut complexes; it generalizes in a certain sense the well-known dualities for convex polytopes, and for combinatorial cell complexes.

THEOREM 3. *For every simplicial complex $\mathscr{K}$ there exists an n such that $\mathscr{N}^n(\mathscr{K})$ is a taut complex of the same homotopy type as $\mathscr{K}$.*

Discrete Variational Green's Function. I *)

PHILIPPE G. CIARLET

Given a real boundary value problem

$$Lu(x)=f(x), \quad x\in\Omega, \quad \text{and} \quad u(x)=0, \quad x\in\partial\Omega,$$

where Ω is a bounded open subset of R^n, and L is a second-order symmetric and positive definite linear elliptic operator, it is known that if the data are smooth enough, its solution ϕ is given, for any function f, by

$$\phi(x)=-\int_\Omega G(x,\xi)\,f(\xi)\,d\xi,$$

$G(x,\xi)$ being the associated *Green's function.*

One can approximate the solution ϕ of such a problem by using a variational approximation procedure: given a finite dimensional subspace S^N of the Sobolev space $H_0^1(\Omega)$, there exists a unique function $\phi^N\in S^N$ such that

$$a(\phi^N, w)=-(f, w), \quad \text{for all} \quad w\in S^N,$$

or equivalently such that

$$F[\phi^N]=\operatorname{Inf}\{F[w];\ w\in S^N\}, \quad \text{with} \quad F[w]\equiv a(w,w)+2(f,w),$$

where (u, v) denotes the inner product in $L^2(\Omega)$, and $a(u, v)$ denotes the extension of the continuous bilinear form $-(Lu, v)$ to the space $H_0^1(\Omega)$.

The object of this paper is first to show that, as in the continuous case, there exists a function $G^N(x,\xi)$, called the *discrete variational Green's function* (relative to a given subspace S^N), such that the unique variational approximation ϕ^N obtained

*) Received June 20, 1968.

as above is given by

$$\phi^N(x) = -\int_\Omega G^N(x, \xi) f(\xi)\, d\xi,$$

for any right-hand side f. Moreover, it is proved that $G^N(x, \xi)$ has a particularly simple expression. More precisely,

$$G^N(x, \xi) \equiv \sum_{p=1}^{N} \frac{1}{\lambda_p^N} \phi_p^N(x)\, \phi_p^N(\xi),$$

where λ_p^N (resp. ϕ_p^N), $1 \leqslant p \leqslant N$, are the approximate eigenvalues (resp. approximate orthonormalized eigenfunctions) of the operator L, obtained by applying the Rayleigh-Ritz procedure over the same subspace S^N.

Then, using a result of Aronszajn–Smith on reproducing kernels, a necessary and sufficient condition is given which insures that a *discrete variational maximum principle* holds in a given subspace S^N, i.e. $\phi^N \geqslant 0$ whenever $f \leqslant 0$, or equivalently $G^N(x, \xi) \geqslant 0$, which is thus the discrete counterpart of the maximum principle for the operator L.

The existence of discrete Green's functions as well as discrete maximum principles have been extensively discussed, in particular by Bramble and Hubbard, in the case where boundary value problems are approximated by finite differences schemes, and the central idea of the present paper is to show that similar properties can also be established in the case of variational approximation schemes.

Linear Operator Equations on a Partially Ordered Vector Space *)

Erich Bohl

This paper is concerned with the study of some properties of linear monotone operators and with a general method for obtaining error estimates for a solution of linear operator equations. The following notations will be used:

X denotes a real vector space endowed with a partial ordering relation $\leqslant$ derived from an Archimedean cone K which contains order units. $\mathring{K}$ will stand for the set of all order units of K. It is assumed that $X \neq \{\theta\}$ where θ denotes the zero element in X. Any order unit $e \in \mathring{K}$ defines two real functionals on X according to the formulae

$$q(x, e) = \operatorname{Min}\{\varrho : \varrho \in R, -x + \varrho e \geqslant \theta\}, \quad \|x\|_e = \operatorname{Max}(q(x, e), q(-x, e))$$

for all $x \in X$, R denotes the set of all real numbers. The functional $\| \ \|_e$ is a norm on X. If $e, e' \in \mathring{K}$ then the corresponding norms $\| \ \|_e$, $\| \ \|_{e'}$ are equivalent.

A linear operator B on X is called monotone, if $B(K) \subset K$. Such an operator is

*) Received July 20, 1968.

bounded with respect to every norm $\| \ \|_e$ and the operator norm $\|B\|_e$ is given by $\|B\|_e = \|Be\|_e$.

THEOREM 1: *For any linear monotone operator B on X the following statements are equivalent:*
(a) $z \in K$ *and* $(I-B)z \in \mathring{K}$; *(b)* $z \in \mathring{K}$ *and* $q(Bz, z) = \|Bz\|_z = \|B\|_z < 1$. *I denotes the identity operator.*

THEOREM 2: *For any linear monotone operator B on X the following statements are equivalent:*
(i) *there exists* $e \in \mathring{K}$ *such that* $\|B\|_e < 1$;
(ii) *there exists* $e \in \mathring{K}$ *such that* $\|B^k\|_e < 1$ *for some natural number* $k \geqslant 1$;
(iii) *for every* $e' \in \mathring{K}$ *and every iterative procedure* $y_{n+1} = By_n + e'$, $y_0 \in K$, *there exists a number* $k = k(e', y_0)$ *such that* $y_k \in \mathring{K}$ *and* $(I-B)y_k \in \mathring{K}$;
(iv) *there exists* y_0, $g \in K$ *such that for the iterative procedure* $y_{n+1} = By_n + g$, *there exists a number k with the properties that* $y_k \in \mathring{K}$ *and* $(I-B)y_k \in \mathring{K}$.

THEOREM 3: *Let* A_1, A_2 *be linear monotone operators on X. Let X be complete or* A_1, A_2 *be completely continuous with respect to some norm* $\| \ \|_e$. *Let the operator* $\hat{A} = A_1 + A_2$ *satisfy one of the conditions* (i)–(iv). *Then the following statements hold:*
(v) *there is precisely one solution* $\bar{x} \in X$ *of the equation* $x = (A_1 - A_2)x + b$, $b \in X$ *and the iterative procedure* $x_{n+1} = (A_1 - A_2)x_n + b$ *converges (with respect to some norm* $\| \ \|_e$*) for each initial approximation* $x_0 \in X$ *to* $\bar{x}$;
(vi) *let* $y_k \in \mathring{K}$ *be constructed by either* (iii) *or* (iv) $(B = \hat{A})$ *then we have the error estimate*

$$- \|x_0 - x_1\|_{(I-\hat{A})y_k} \hat{A}^n y_k \leqslant \bar{x} - x_n \leqslant \|x_0 - x_1\|_{(I-\hat{A})y_k} \hat{A}^n y_k \quad \text{for all} \quad n \geqslant 0 .$$

If for some n the element $\hat{A}^n y_k$ *is an order unit, the error estimate takes the form*

$$\|\bar{x} - x_n\|_{\hat{A}^n y_k} \leqslant \|x_0 - x_1\|_{(I-\hat{A})y_k} ;$$

(vii) *if* A_2 *is the zero operator statement* (v) *is equivalent to each of the statements listed in Theorem* 2.

There are various well known error estimates which turn out to be special cases of the estimate given in (vi). We are now going to point out only three of them:

a) $\|\bar{x} - x_n\|_y \leqslant \|(I - \hat{A})^{-1} \hat{A}^n\|_y \, \|x_0 - x_1\|_y$ for any $y \in \mathring{K}$.

b) Let (iv) be satisfied for $g = \theta$. The resulting estimate is known for linear systems in the real space R^N:

$$- \|x_0 - x_1\|_{(y_k - y_{k+1})} y_{k+1} \leqslant \bar{x} - x_1 \leqslant \|x_0 - x_1\|_{(y_k - y_{k+1})} y_{k+1} .$$

c) Let us consider a linear system $x = Ax + b$ in the real space R^N where $A = (a_{ik})$ is a real $(N\times N)$-matrix and $b\in R^N$. Introducing the real numbers $\alpha_i=\sum_{k=1}^{N}|a_{ik}|$ for $i=1,\ldots,N$ and assuming that $\alpha_i<1$ for $i=1,\ldots,N$ we obtain from (vi)

$$|\bar{t}_i - t_i^1| \leqslant \alpha_i \operatorname*{Max}_j \frac{|t_j^0 - t_j^1|}{1-\alpha_j} \quad \text{for} \quad i = 1, \ldots, N,$$

where $\bar{x}=(\bar{t}_i)$, $x_0=(t_i^0)$, $x_1=(t_i^1)=Ax_0+b$, and $\bar{x}$ is the unique solution of the system $x=Ax+b$.

Let us introduce the two real functionals defined on $\mathring{K}$

$$\lambda(x, B) = -q(-Bx, x), \quad \bar{\lambda}(x, B) = q(Bx, x),$$

where B is any linear operator on X.

A linear operator B on X is called strongly monotone if B is monotone and if for each $x\in K-\mathring{K}$, $x\neq\theta$, there exists a natural number $n=n(x)\geqslant 1$ such that $B^n x\in\mathring{K}$.

THEOREM 4: *Let B be a linear monotone operator on X and let there exist a linear strongly monotone completely continuous operator B′ on X which commutes with B. Then the following statements hold:*

(*a*) *there exists an order unit z and real numbers* $\mu_0>0$, λ_0 *such that* $Bz=\lambda_0 z$, $B'z=\mu_0 z$, $\|z\|_e=1$;

(*b*) *if* $x_{n+1}=B'x_n$, $x_0\in\mathring{K}$, *then* $x_n\in\mathring{K}$, *and hence* $x_n\neq\theta$ *for all* $n\geqslant 0$ *and* $\lim\lambda(x_n, B)=\lim\bar{\lambda}(x_n, B)=\lambda_0$;

(*c*) $\lambda(x_n,B)\leqslant\lambda(x_{n+1},B)\leqslant\lambda_0\leqslant\bar{\lambda}(x_{n+1},B)\leqslant\bar{\lambda}(x_n,B)$ *for* $n\geqslant 0$;

(*d*) $z_n=\|x_n\|_e^{-1}x_n$ *converges to z (with respect to* $\|\ \|_e$);

(*e*) $|\alpha|\leqslant|\lambda_0|=\|B\|_z\leqslant\|B\|_{e'}$ *for all real eigenvalues* α *of B and all* $e'\in\mathring{K}$. *Consequently,* $\lambda_0=0$ *if and only if B is the zero operator and*

$$\lambda_0 = \operatorname*{Min}_{e\in\mathring{K}} \|B\|_e = \operatorname*{Min}_{e\in\mathring{K}} q(Be, e)$$

THEOREM 5: *For a linear monotone completely continuous operator B on X the following statements are equivalent:*

(α) *B is strongly monotone and there exists an order unit e such that* $\|B\|_e<1$;

(β) *for each* $y_0\in K$, $y_0\neq\theta$ *the iterative procedure* $y_{n+1}=By_n$ *has the property that there exists a number* $k=k(y_0)$ *such that* $y_k-y_{k+1}\in\mathring{K}$.

On Homomorphisms of the General Linear Group *)

D. Ž. Djoković

Let $GL_n(K)$ and $SL_n(K)$ be the general and special linear group respectively over a division ring K. We denote by K^* the multiplicative group of K and by C the

*) Received August 22, 1968.

commutator subgroup of K^*. Let

$$\phi: \mathrm{GL}_n(K) \to \mathrm{GL}_m(K)$$

be a group homomorphism. Our main result is: If $m<n$ then $\mathrm{Ker}\,\phi \supset \mathrm{SL}_n(K)$. This result can be expressed also in the following way: If $m<n$ then there exists a group homomorphism

$$\psi: K^*/C \to \mathrm{GL}_m(K)$$

such that

$$\phi(X) = \psi(\det X) \quad \text{for all} \quad X \in \mathrm{GL}_n(K).$$

Here, det denotes Dieudonné's determinant which is a homomorphism from $\mathrm{GL}_n(K)$ onto K^*/C. This generalizes a recent result of M. Kucharczewski and A. Zajtz (Annales Polonici Math. *18* (1966), 205–225) which is a special case of our result for $K=R=$the field of real numbers. They have obtained an analogous result when $K=R$ and $m=n$.

Let $M_n(F)$ be the algebra of $n \times n$ matrices over a field F. We give also a very simple proof of the following fact:
If $f: M_n(F) \to F$ satisfies $f(XY)=f(X)f(Y)$ for all $X, Y \in M_n(F)$ then there exists $g: F \to F$ such that $f(X)=g(\det X)$ for all $X \in M_n(F)$ and g satisfies $g(ab)=g(a)\,g(b)$ for all $a, b \in F$.

Eine Verallgemeinerung des Begriffes der homogenen Produktionsfunktion *)

WOLFGANG EICHHORN

Die Klassen der homogenen und (allgemeiner) der homothetischen Produktionsfunktionen sind aus verschiedenen Gründen für viele Zwecke der Produktionstheorie zu eng. Deshalb wird die im Vergleich zur Funktionalgleichung $\Phi(\lambda x)=\lambda\alpha\,\Phi(x)$ der vom Grade $\alpha>0$ homogenen Produktionsfunktionen $\Phi: \mathbb{R}^n_+ \to \mathbb{R}_+$ ($\mathbb{R}_+$ die nichtnegativen reellen Zahlen) allgemeinere Funktionalgleichung

$$\Phi(\lambda x) = \varphi\left(\lambda, \frac{x}{|x|}\right)\Phi(x) \quad \begin{cases} \lambda > 0, \quad x \in \mathbb{R}^n_+, \quad x \neq 0, \\ |x| = \text{euklidische Norm von } x, \\ \varphi(\lambda, x/|x|) \in \mathbb{R}_+ \end{cases}$$

eingeführt und auf vom Standpunkt der Produktionstheorie interessante Lösungen hin untersucht. Es zeigt sich, daß eine Produktionsfunktion Φ dann und nur dann einer Gleichung dieser Art genügt, wenn φ die Form

$$\varphi\left(\lambda, \frac{x}{|x|}\right) = \lambda^{\psi\left(\frac{x}{|x|}\right)} \quad \left(\lambda > 0, \quad x \neq 0, \quad \psi\left(\frac{x}{|x|}\right) > 0\right)$$

*) Eingegangen am 5.11.1968. Einzelheiten erscheinen demnächst in: Unternehmensforschung.

hat. Nichtsdestoweniger ist die so definierte Klasse von Produktionsfunktionen sehr viel allgemeiner als die der homogenen Produktionsfunktionen; dies wird anhand von Beispielen gezeigt, die die Mängel der homogenen und der homothetischen Produktionsfunktionen nicht besitzen.

Universität Würzburg

Mehrproduktunternehmungen mit linearen Expansionswegen*)

WOLFGANG EICHHORN und WERNER OETTLI

Die Produktionsfunktion F ordne jedem Aufwandsvektor v (= Vektor der Quantitäten der aufgewendeten Produktionsfaktoren) die Menge $F(v)$ derjenigen Produktvektoren x (=Vektoren der Quantitäten der produzierten Güter) zu, die mit Hilfe von v produzierbar sind. F heißt *homogen vom Grade* $\alpha>0$, wenn für beliebiges $\lambda>0$ und v stets $F(\lambda v)=\lambda\alpha F(v)$ gilt. Ein Aufwandsvektor v^* wird *Minimalkostenkombination zur Erzeugung des Produktvektors* x^* genannt, wenn sich bei seinem Einsatz geringere Kosten ergeben als bei Einsatz irgendeines anderen Aufwandsvektors v mit der Eigenschaft $x^*\in F(v)$.

SATZ: *Ist F homogen vom Grade $\alpha>0$ und wird ein gegebener Produktvektor x^* mit Hilfe der Minimalkostenkombination v^* produziert, so ist – falls die Preise der Produktionsfaktoren vom Grade β homogene Funktionen des Aufwandsvektors v sind – λv^* Minimalkostenkombination zur Erzeugung des Produktvektors $\lambda\alpha x^*$. Es liegt dann also ein linearer Expansionsweg vor.*

Analoge Sätze werden für den Fall des Umsatzmaximums und des Gewinnmaximums bewiesen. Schließlich wird gezeigt, daß homogene Produktionsfunktionen neben linearen auch nichtlineare Expansionswege besitzen können, und daß andererseits die Existenz linearer Expansionswege auch noch für diejenigen Produktionsfunktionen gesichert ist, die der Funktionalgleichung

$$F(v) = \varphi(|v|)\,F\left(\frac{v}{|v|}\right) \quad \left\{\begin{array}{l} v \neq 0,\ |v| = \text{euklidische Norm des} \\ \text{Vektors } v,\ \varphi \text{ positivwertig, streng} \\ \text{monoton wachsend, } \varphi(1) = 1 \end{array}\right.$$

genügen.

Universität Würzburg und
IBM-Forschungslaboratorium Zürich

*) Eingegangen am 5.11.1968. Einzelheiten erscheinen demnächst in einem Sammelband „Operations Research-Verfahren", herausgegeben von R. Henn, H. P. Künzi und H. Schubert.

Contents Vol. 2 2/3 1969

Expository Papers

Research Papers

Reports of Meetings

Short Communications