Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

MATHEMATICAL CENTRE TRACTS 100

PROCEEDINGS
BICENTENNIAL CONGRESS
WISKUNDIG GENOOTSCHAP

PART I

Edited by

P.C. BAAYEN

D. VAN DULST

J. OOSTERHOFF

MATHEMATISCH CENTRUM AMSTERDAM 1979

ISBN 90 6196 168 8

P R E F A C E

Each year the WISKUNDIG GENOOTSCHAP, founded in Amsterdam in 1778, organizes a mathematical congress to stimulate research in mathematics and provide opportunities for personal contacts among Dutch mathematicians. On the occasion of the bicentennial celebration of the WISKUNDIG GENOOTSCHAP the 14th Dutch Mathematical Congress, to be held in Amsterdam in 1978, was to have a special festive character.

To achieve this goal it was decided by the organizing committee to have the major fields of mathematics currently practised in the Netherlands represented by invited speakers. As a consequence of this decision much greater emphasis was given to invited papers than is customary at Dutch Mathematical Congresses. Moreover, the third BROUWER Memorial Lecture, to be delivered by A. BOREL, was also incorporated in the congress. Together with the festive address of H. FREUDENTHAL, president of the WISKUNDIG GENOOTSCHAP at this time, these lectures were judged to be of sufficient interest to be published in the form of Proceedings.

In approaching speakers the program committee had to take into account that some prominent Dutch mathematicians were already overburdened by the organization of the bicentennial celebration and that others had recently acted as invited speakers at one of the last two Dutch Mathematical Congresses. Nevertheless, the invited papers represent a comprehensive cross section of contemporary pure and applied mathematics. The committee is pleased to note the important contributions of several speakers from abroad; without their papers a broad coverage of many domains of mathematics would have been far more difficult.

At the bicentennial congress the participants received a preliminary edition of these Proceedings. That edition was incomplete, e.g. the Presidential address and the BROUWER Memorial Lecture were missing (for obvious reasons). Also, some of the lecturers received their invitation rather late, and could not meet the dead-line for the preliminary publication. Some of the papers absent in the preliminary publication of these Proceedings are also published in the "NIEUW ARCHIEF VOOR WISKUNDE", the journal of the WISKUNDIG GENOOTSCHAP.

One very much appreciated lecture remains unwritten: B.L. van der WAERDEN's plenary talk on "Algebraic geometry in Holland between 1918 and 1928; reminiscences from my student's days". It was understood from the outset that this lecture would not be represented by a paper; the editors are happy, *however, that the lecturer himself does find a (pictorial) representation in these Proceedings.*

The publication of these Proceedings would not have been possible without the invaluable assistance of Mrs. S.J. Kuipers-Hoekstra. In this computer age we have become used to external memories: Mrs. Kuipers, in addition to being a living sample of this concept, also embodies the concept of an external conscience. Always cheerful, she managed to keep both authors and organizers at their tasks.

A tremendous lot of work for the bicentennial celebration of the WISKUNDIG GENOOTSCHAP was done by the Reproduction and Publishing service of the Mathematical Centre (apart from these Proceedings, several other publications had to be taken care of, all with the same dead-line). The editors thank Mr. D. Zwarst and Mrs. R.W.T. Riechelmann-Huis and all their staff for the work they did to get these Proceedings typed and printed, meticulously and well within the time schedule.

P.C. BAAYEN
D. van DULST
J. OOSTERHOFF

DUTCH MATHEMATICAL CONGRESSES

1965 / Enschede	1972 / Groningen
1966 / Heerlen/Valkenburg	1973 / Leiden
1967 / Nijmegen	1974 / Enschede
1968 / Eindhoven	1975 / Utrecht
1969 / Wageningen	1976 / Amsterdam
1970 / Delft	1977 / Rotterdam
1971 / Amsterdam	1978 / Amsterdam

BROUWER MEMORIAL LECTURES

1970	René THOM,	*Le degré Brouwerien en topologie différentielle moderne* Published in: Nieuw Archief voor Wiskunde (3) 19 (1971), 10-16.
1973	Abraham ROBINSON,	*Standard and non-standard number systems* Published in: Nieuw Archief voor Wiskunde (3) 21 (1973), 115-133.
1975	Armand BOREL,	*On the development of Lie group theory* Published in: these Proceedings, p.25 Also to be published in: Nieuw Archief voor Wiskunde (3) 27 (1979).

ORGANIZATION OF THE CONGRESS

The Bicentennial Congress of the WISKUNDIG GENOOTSCHAP was organized jointly by the Mathematical Institute of the Free University, the Mathematical Institute of the University of Amsterdam and the Mathematical Centre. These three Institutes together appointed the Organizing Committee whose members were: P.C. Baayen (Chairman), H. Bart, Tj. Blanksma, J.M. Buhrman, D. van Dulst, W.T. van Est, mrs. S.J. Kuipers-Hoekstra (secretary), J. Nuis, J. Oosterhoff, E. Slagt (Deputy Chairman), G.L. Wanrooij.
The congress was held in the main building of the Free University, at 29, 30 and 31 March, 1978.

At the occasion of the congress and the Bicentennial Celebration two expositions on the history and the cultural significance of mathematics were organized by a committee consisting of: F. van der Blij (Chairman), B.R.T. Arnold, F. Cohen, J. van de Craats, S.B. Engelsman, Mrs. M. Fournier, R. Frederik, K.W. Gnirrep, R.H. de Jong, P.H. Krijgsman, E. Thomas, G. Veeneman, Joh.H. Wansink.
One exposition was held in the same building as the Congress; a 100 page catalogue was published entitled "Tweehonderd jaar onvermoeide arbeid". Editors were R.H. de Jong, T.W.M. Jongmans and P.H. Krijgsman.
The other exposition, which was arranged around the 1694 edition of Abraham de Graaf's book "De Geheele Mathesis of Wiskonst", was aimed for a more general public and could be visited in Amsterdam, in the Library of the University of Amsterdam from 23 March to 26 April, in Utrecht, in the University Museum from 16 May to 21 June, and in Leiden, in the Museum Boerhaave from 10 July to 2 October, 1978.

In addition to the important support given by the two Amsterdam Universities and the Mathematical Centre, mention must be made of financial contributions from DSM Heerlen, Shell Nederland B.V., N.V. Philip's Gloeilampenfabrieken, I.B.M. Nederland N.V., the Actuarial Society of the Netherlands and Scheltema, Holkema, Vermeulen, booksellers.
Material contributions were presented by Control Data B.V. in Rijswijk.

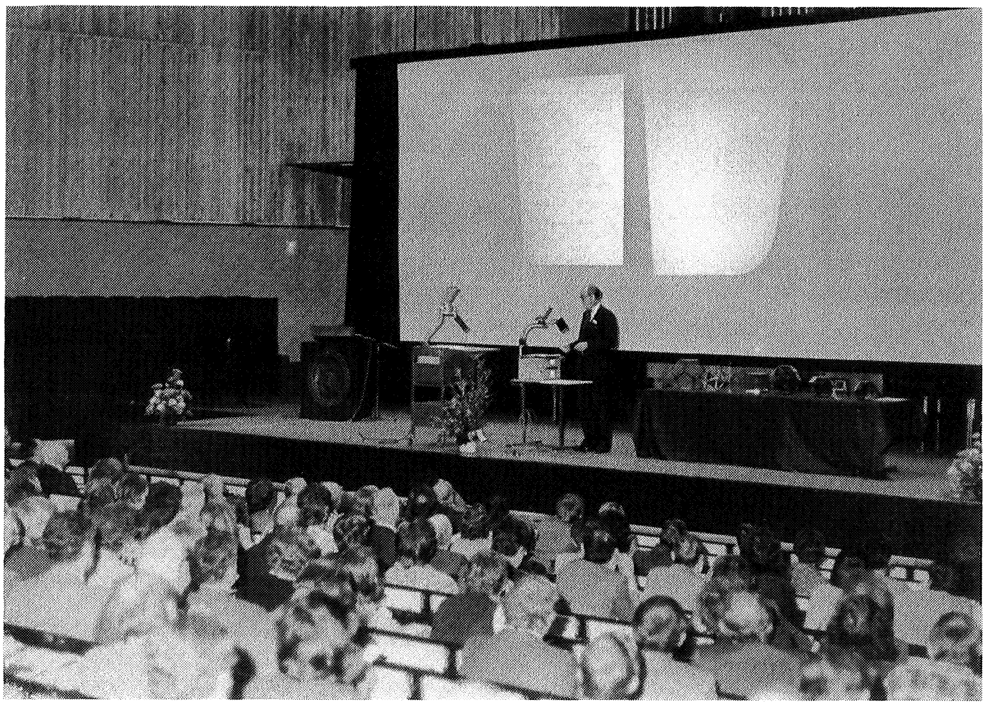
For the two expositions personal as well as material assistance was presented by the Museum Boerhaave (Leiden), the Mathematical Institute of the University of Utrecht, the University Museum of Utrecht, the University Library of Amsterdam, the Free University of Amsterdam and the I.O.W.O. (the institute for development of mathematical education).

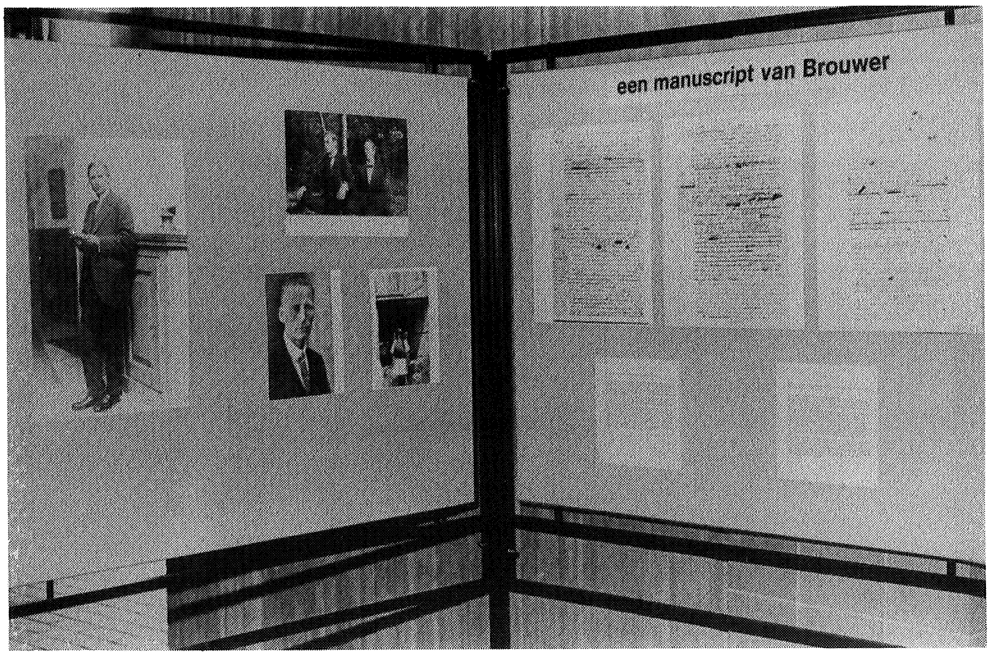
Some impressions from the Bicentennial congress

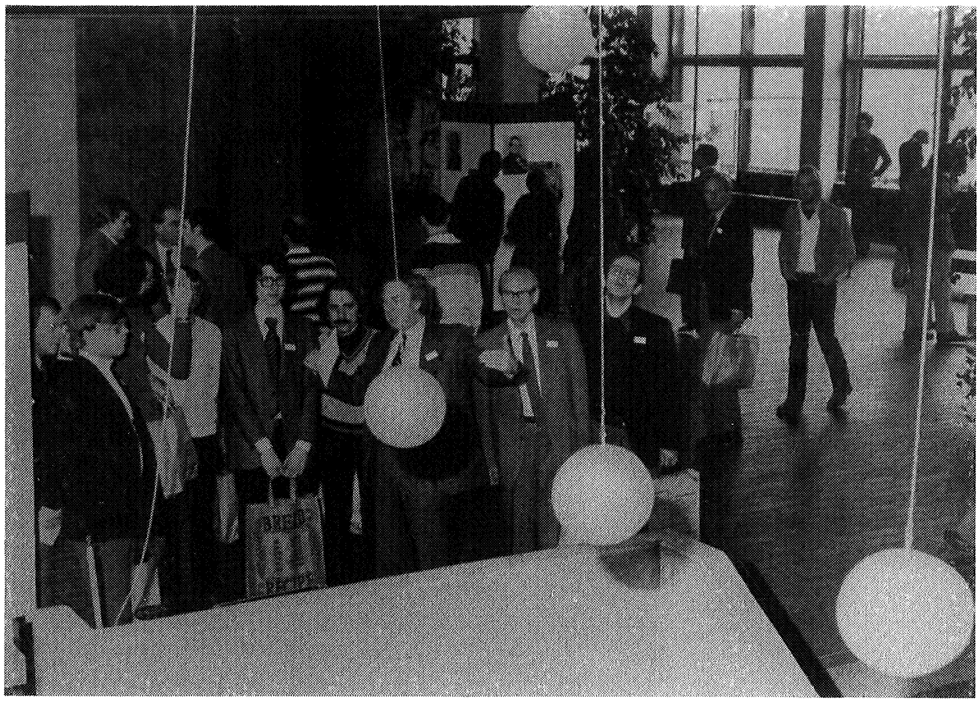
of the

Wiskundig Genootschap
"Een onvermoeide arbeid komt alles te boven"

1778 1978




een manuscript van Brouwer



LIST OF SHORT COMMUNICATIONS

JONGEN, H.Th.) TWILT, F.)	*Op zoek naar meerdere locale minima*
KAAS, R.) BUHRMAN, J.M.)	*De mediaan van de binomiale verdeling*
KINDT, M.	*Autowegen*
KLEVER, C.A.M.	*Gegeneraliseerde Hall vlakken*
KOSTENSE, P.J.	*De ideale rand van Martin en polaire verzamelingen*
KRUYSWIJK, D.	*Over getallen van de vorm n.n.!*
LAAN, C.G. van der	*Parameterschatting*
LAGERWERF, G.	*Wiskunde Gebruiken*
LANGE, J. de	*De Grand Canyon instap voor de behandeling van functies van twee variabelen*
LEEUWEN, J. van	*Een nuttig lemma voor contextvrij geprogrammeerde grammatica's*
LUNE, J. van der	*Convexe approximatie van integralen*
MAAS, W.A.K.	*Enkele aspecten van ongewenste oscillaties bij spanningstransformatoren*
MEESTER, F.R.M.	*Vaardigheden van een wiskundedocent(e), het gepland en ad hoc kunnen reageren*
MEULEN, S.G. van der) VELDHORST, M.)	*Datastructuren voor lineaire ruimten*
MOLENAAR, W.	*Bayesiaanse oplossingen van het m-groep regressie-probleem*
MULDER, H.M.	*Karakterisering van de n-kubus*
NAUWELAERTS, E.	*Localisatie van algebra's met polynoomidentiteiten*
OYSTAEYEN, F. van	*Birational ring extensions*
RIELE, H.J.J. te	*Berekeningen met betrekking tot het vermoeden van Mertens*
SCHIPPERS, H.	*Numerieke aanpak van de tijdsafhankelijke stroming t.g.v. een roterende, oneindige schijf*
SIMONS, J.L.	*Interactief blokken construeren bij de vakkendifferentiatie binnen het Algemeen Voortgezet Onderwijs*
SIMONS, J.L.	*Pang-Pang-Ki (PPK), een eenvoudig bordspel*
SCHRIJVER, A.	*Bewijs van Lovász' vermoeden over fractionele overdekkingen*
SCHURING, H.N.	*Toetsen toen en thans*
TEMME, N.M.	*Over de aymptotische ontwikkeling van de incomplete gammafunctie*
TIJS, S.H.	*Approximatie van spelen en programmeringsproblemen met deelspelen en deelprogramma's*
TURK, J.W.M.	*Stelsels multiplicatief afhankelijke getallen*

VEL, M. van de	*Interne karakterizering van ANR's*
VERSCHOREN, A.	*Enige aspecten der niet-commutatieve algebraische meetkunde*
VERWER, J.G.	*Split-methoden voor de numerieke integratie van parabolische vergelijkingen geformuleerd via de methode der lijnen*
VITÁNYI, P.M.B.	*Simulatie in lineaire tijd van een meerkopsbandeenheid met kop-tot-kop herzetinstructies door eenkopsbandeenheden*
VOLGENANT, A.	*Non-optimale handelsreizigertrajecten bij symmetrische problemen*
VOERHOEVE, M.	*De vergelijking* $1^k + \ldots + x^k + R(x) = y^z$
VRIES, J. de	*Compactificatie van G-ruimten*
VRIEZE, O.J.	*Perturbatietheorie voor spelen in normale vorm en stochastische spelen*
WAALL, R.W. van der	*De conductor nader bekeken*
WAL, J. van der	*Stoptijden en policy iteration voor niet verdisconteerde Markov beslissingsproblemen*
WATTEL, E.	*Souslin dendrons*
WIEGERINK, L.	*Wisbrug 200*
WOLKENFELT, P.H.M.	*Stabiliteit van numerieke methoden voor het oplossen van Volterra integro-differentiaal vergelijkingen*
WOUWE, J.M. van	*Go-ruimten met een* Σ*-net*

CONTENTS

PRESIDENTIAL ADDRESS

BROUWER MEMORIAL LECTURE

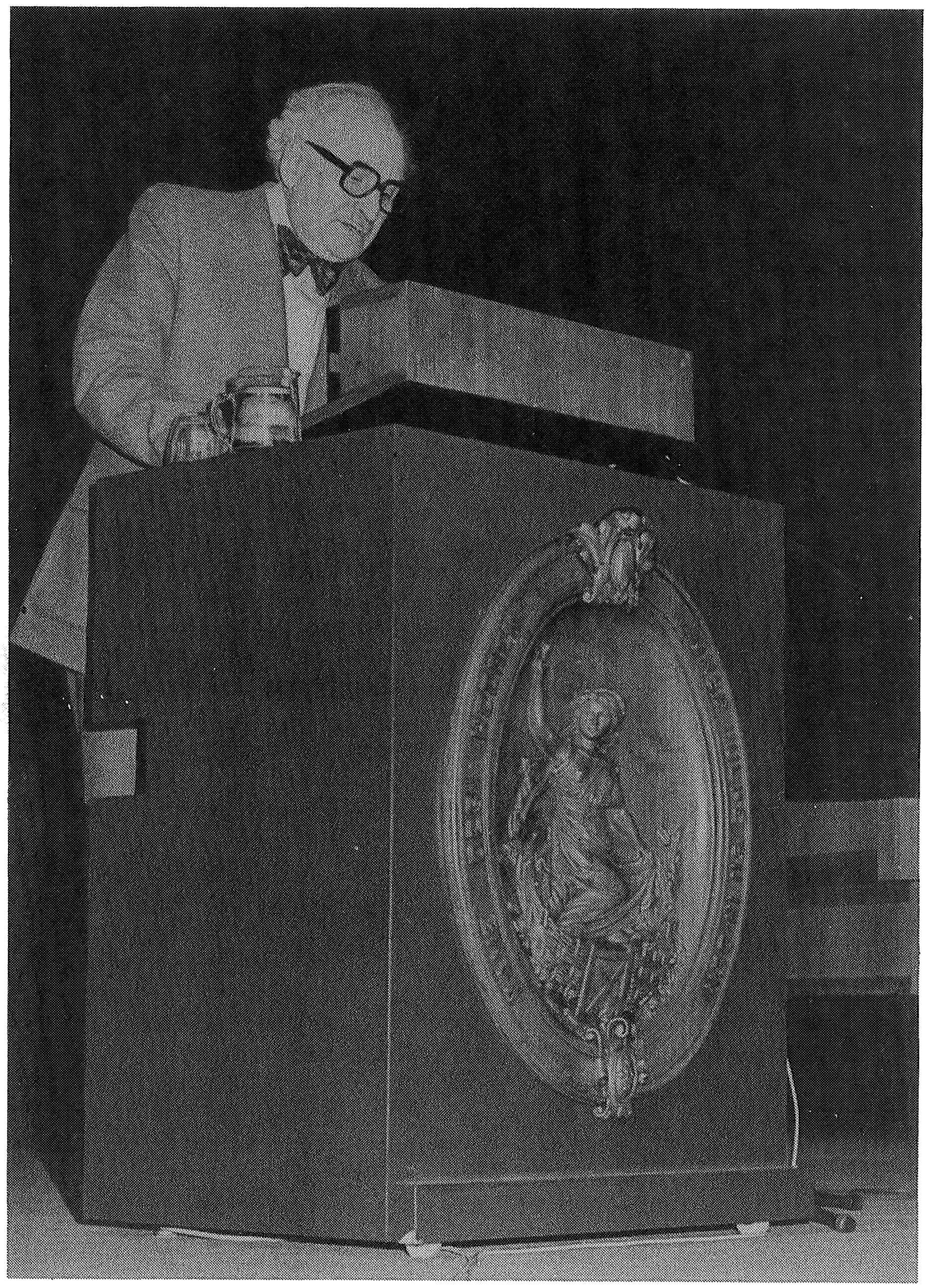
PLENARY ADDRESSES

INVITED LECTURES

CONTENTS OF PART II

NOTE. The following two invited lectures were presented at the congress, but no manuscript has been received for publication in these Proceedings:

SCOTT, D.S.	*What does the intuitionistic continuum look like*
SINGH VARMA, H.O.	*Geometric miscellenea*



TELKENS VIJFTIG JAREN

H. Freudenthal

TELKENS VIJFTIG JAREN

1778 - twee eeuwen geleden - dood van Rousseau, dood van Voltaire - vooravond van de patriottentijd [1] - ancien régime. Wisten de geregeerden van toen wel dat ze onder een ancien régime leefden? Konden zij bevroeden dat elf jaar later de Bastille bestormd zou worden, Condorcet [2], de secrétaire perpétuel van de Académie des Sciences, Lavoisier [3], grondlegger van de scheikunde, dat zij, de voortrekkers, in de gevangenis en op de guillotine hun leven zouden laten? Wat was er dan wel het gesprek van de dag? Eise Eisinga's [4] planetarium [5], de trots van Franeker en Friesland? Welneen, het functionneerde al, maar het was nog niet af. Dat een Wiskundig Genootschap werd opgericht onder de zinspreuk "Een onvermoeide arbeid komt alles te boven"? Wel, er werden in die tijd om de haverklap en onder velerlei zinspreuken genootschappen opgericht [6] - althans in de Zeven Provinciën.

Misschien dat een tijdgenoot als belangrijkste feit genoteerd zou hebben de aankomst uit Wenen in Parijs van Mesmer [7] op het hoogtepunt van zijn roem. Of wist U niet wie Mesmer was? Hij heeft zijn naam gegeven aan wat men Mesmerisme of dierlijk magnetisme ging noemen. Hij was, meen ik, de eerste in een reeks, die zich nog steeds voortzet, in een lange stoet aan de zelfkant van de wetenschap.

1778 - wat was er toen in de wiskunde aan de hand? In 1778 verscheen de eerste publicatie van iemand, die er straks vermaardheid door zou verwerven als schepper van een nieuwe wiskunde: Karl Friedrich - niet Gauss, die toen één jaar oud was, maar - Hindenburg [8], stichter en grootmeester van de combinatorische school. Nooit van gehoord? Terecht! Sic transit gloria mundi - vooral als die roem onverdiend was. Niet alleen achteraf bezien, maar van meet af aan: gewichtigdoenerij met trivialiteiten - en toch vermaard. Dat kan ook in de wiskunde.

1778 - nog iets van belang? Johann Karl Schultze's [9] tafelwerk in twee delen [10] - een verschijnsel karakteristiek voor dat tijdperk, dat afgesloten zou worden, enkele jaren later, met de tafels van Vega [11], een naam die U vermoedelijk wel iets zegt. Het was het tijdperk niet alleen van de genootschappen, maar ook van de tafelwerken. Om dit te staven: de wederwaardigheden van de Oostenrijker A.F. Felkel [12], uitvinder van een factorenmachine, een gewone rekenmachine, een leesmachine, een taalmachine, een meetinstrument. In 1777 werd zijn factorentafel, tot 144000 gevorderd, uitgegeven en voortzetting tot 20 000 000 aangekondigd; van het tot 2 000 000 voltooide manuscript werd het deel tot 408 000 op staatskosten gedrukt, maar bij gebrek aan afzet tijdens de Turkse oorlog (1788) tot infanteriepatronenpapier verwerkt - wederom: sic transit gloria mundi.

In 1778 deelde de secrétaire perpétuel van de Académie des Sciences, Condorcet, mede [13] dat drie jaar eerder

> *L'Académie a pris... la résolution de ne plus examiner aucune solution du Problème de la duplication du cube, de la trisection de l'angle ou de la quadrature du cercle, ni aucune machine annoncée comme un mouvement perpétuel*

en hij lichtte dit besluit toe met uitspraken, die prematuur en voor die tijd nauwelijks verantwoord waren: Immers dat de derdegraadsvergelijkingen van de eerste twee problemen niet met vierkantswortels konden worden opgelost, kon in die tijd zeer waarschijnlijk worden geacht, maar een bewezen feit was het allerminst; evenmin dat voor de kwadratuur van het cirkelsegment transcendente functies onvermijdelijk waren; het is waar dat Condorcet t.a.v. π en het perpetuum mobile [14] nog slagen om de arm hield.

Daar maakten de mensen zich toen dus druk om. Misschien mist U in Condorcet's lijst het parallellenpostulaat en de algebraïsche vergelijkingen van hogere dan de vierde graad. Voor echte belangstelling in meetkunde was het nog een halve eeuw te vroeg. Maar wat grote en wijdse problemen aangaat, dan lijkt het, als men de literatuur van toen doorbladert, alsof het er drie waren: alle algebraïsche vergelijkingen oplossen, alle integralen uitrekenen, alle differentiaalvergelijkingen integreren. En dat laatste dan als formele bezigheid - nog niemand dacht er aan, differentiaalvergelijkingen in het kader te plaatsen van begin- of randwaardeproblemen.

De algemene algebraïsche vergelijkingen oplossen - zelfs Bézout [15] kwam er niet uit. Maar zijn inspanningen werden anderszins beloond- in 1779 zou zijn hoofdwerk [16] verschijnen, de eliminatie-theorie, als het ware het beginpunt van de algebraïsche meetkunde. Hoewel, de grote doorbraak, die beslissend zou blijken, was enkele jaren eerder in stilte geschied: Lagrange's Réflexions [17], die Ruffini, Abel en Galois zouden inspireren: het inzicht in de opbouw van resolventen door middel van wat we nu de symmetrische groep noemen, en het resignerend besef dat de tot en met de vierde graad succesrijk gebleken methoden vanaf de vijfde graad zouden falen.

Wat deed Laplace [18] rond 1778? Onnoemelijk veel. Vooral werkte hij aan zijn Système du Monde, dat ik altijd nog eens zou willen bestuderen. Enkele bladzijden eruit, net nog in december 1778 gefiatteerd, - de voortplanting van golven in kanalen [19] - trokken mijn aandacht, toen ik me met Cauchy bezig hield, want ze waren de eerste aanzet tot diens indrukwekkende jeugdwerk over golven [20].

Hoewel druk bezig met kansrekening - vooral met het exploiteren van Bayes' principe- is Laplace in 1778 nog niet aan de normale verdeling toe, zoals ze al aan de Moivre [21] bekend was, en evenmin aan de goede foutenwet. Wel is in 1778 de oorsprong van de Laplace-transformatie [22] te zoeken, of veeleer het idee om Lagrange's formele opvatting van differentie en differentiatie als operator analytisch te rechtvaardigen door voortbrengende functies, waaruit later de voortbrengende integralen zouden ontstaan [23].

De differentiaalvergelijking van Laplace - ook hiervoor is het nog te voreg [24]. Maar het idee van de potentiaal is er al, 1777, dankzij Lagrange [25], zelfs van continu verdeelde massa's.

Bézout, Lagrange, Laplace - ik zou er nog enkele namen aan kunnen toevoegen. Eén heeft er in elk geval tot hier toe ontbroken: Euler [26], oud en blind, maar ongeremd productief als een jongeling. De twee verhandelingen waarin hij de kwadratische reciprociteitswet formuleerde, vermoedelijk in 1772 geschreven, werden in 1783 gepubliceerd [27], en als ik het gemiddelde neem van deze twee jaartallen, kom ik - hoe kan het beter? - zowat in 1778 terecht. Jammer voor Legendre [28] die de wet in 1785 opnieuw moest ontdekken. Pas Gauss zou hem echt bewijzen. [29]

Zonder een rekenkundige truc kan ik in Eulers productie iets dichter bij 1778 aanwijzen en dan nog iets dat als beginpunt de aandacht van de historicus trekt. Het is met 1777 gedateerd [30] en als ik er aan toevoeg dat Lagrange in 1779 hetzelfde onderwerp [31] behandelde, kom ik weer gemiddeld in 1778 terecht. Het is het eerste optreden van begrip en term "afbeelding" - latijn "repraesentatio" - algemeen afbeeldingen van het boloppervlak op het vlak, en dan gespecialiseerd tot conforme afbeeldingen, die ook complex geschreven worden - een idee dat in ander verband, zoals ook de Cauchy-Riemann differentiaalvergelijkingen, teruggaat tot d'Alembert [32]).

* * *

"Jubileum" komt van het hebreeuwse joveljaar, dat met de periode 50 werd gevierd. Wiskundig Genootschap jubileerde dus voor 't eerst in 1828. Wat valt er over 1828 te vertellen? De aardappelziekte, de uitvinding van de feuilletonroman? Neen, ik bedoel in de wiskunde.

In 1828 verschenen van de hand van Alphonse Quetelet [33)] de "Instructions populaires sur le calcul des probabilités" [34)], een klein, en zoals de titel zegt, populair boekje. Het zou niet de moeite waard zijn ervan te reppen, als de jonge auteur niet in 't zelfde jaar, nog net in de decembervergadering van de Brusselse Akademie [35)] de aanzet had gegeven tot het levenswerk dat hem beroemd zou maken- "De eerste ontmoeting tussen de wiskunde en de sociale wetenschappen" heb ik het elders [36)] genoemd.

1828, de geestelijke geboortedatum van Quetelet de statisticus, de - ondanks voorlopers - eerste statisticus. In 1835 zou zijn grote werk verschijnen, Sur l'homme et le développement de ses facultés - essai d'une physique sociale [37)]. Het werk maakte furore. Waarom? Als catalogus van cijfers, als portret van de gemiddelde mens, l'homme moyen? Neen, wegens een passage, volgende op een uittreksel uit de Franse criminaalstatistiek [38)]

> *Cette constance avec laquelle les mêmes crimes se produisent annuellement dans le même ordre et attirent les mêmes peines dans les mêmes proportions, est un des faits les plus curieux que nous apprennent des statistiques des tribunaux; je me suis particulièrement attaché à la mettre en évidence dans mes différents écrits; je n'ai cessé de répéter chaque année:* Il est un budget qu'on paie avec une régularité effrayante, c'est celui des prisons, des bagnes et des échafauds; c'est celui-là surtout qu'il faudrait s'attacher à réduire; *et chaque année les nombres sont venus confirmer mes prévisions, à tel point, que j'aurais pu dire, peut-être avec plus d'exactitude: Il est un tribut que l'homme acquitte avec plus de régularité que celui qu'il doit à la nature ou au trésor de l'Etat, c'est celui qu'il paie au crime! Triste condition de l'espèce humaine! Nous pouvons énumérer d'avance combien d'individus souilleront leurs mains du sang de leurs semblables, combient seront faussaires, combien empoisonneurs, à peu près comme on peut énumérer d'avance les naissances et les décès qui doivent avoir lieu.*
>
> *La société renferme en elle les germes de tous les crimes qui vont se commettre, en même temps que les facilités nécessaires à leur développement. C'est elle, en quelque sorte, qui prépare ces crimes, et le coupable n'est que l'instrument qui les exécute...*
>
> *Cette observation, qui peut paraître décourageante au premier abord, devient consolante au contraire quand on l'examine de près, puisqu'elle montre la possibilité d'améliorer les hommes, en modifiant*

> *leurs institutions, leurs habitudes, l'état de leurs lumières, et en général, tout ce qui influe sur leur manière d'être. Elle ne nous présente, au fond, que l'extension d'une loi déjà bien connue de tous les philosophes... c'est ce que tant que les mêmes causes subsistent, on doit attendre le retour des mêmes effets...*

Inderdaad had Quetelet deze uitspraak zowat jaar aan jaar herhaald, nadat hij hem voor 't eerst - ongeveer net zo - in december 1828 had gedaan, en hij zou hem blijven herhalen tot de oren van zijn tijdgenoten ervan zouden tuiten. De standvastigheid van het aantal misdaden, volgens soorten opgesplitst - waarop was deze uitspraak gebaseerd toen Quetelet hem voor 't eerst in 1828 deed? Op drie jaren statistiek - het was veeleer een voorspelling, aangedurfd door iemand die de wet van de grote getallen op zijn duimpje kende. Quetelet zou een der grote 19e eeuwers worden, opgehemeld en verguisd. Van de emoties, die hij zou opwekken, was hijzelf het eerste slachtoffer - een geëngageerd wiskundige, zou men heden zeggen. Maar er waren er ook spoedig die roken wat voor vlees ze met Quetelet in hun kuip hadden: een vinnige aanval van ene den Tex [39] al in 1829:

> *Ik ben verre van de toepassing van getalberekeningen, daar waar die gepast is, af te keuren. Ik mag lijden, dat men berekent, welke waarschijnlijkheid er is, dat uit een zeker getal nummers een bepaald nummer worde getrokken, hoeveel waarschijnlijker het is, dat uit 10 witte en 2 zwarte boonen er eene witte dan eene zwarte zal getrokken worden. Ik erken de nuttigheid van het berekenen van waarschijnlijkheden voor alle soorten van assurantiën, maar om te zeggen, dat onze meeste kennis slechts op waarschijnlijkheden berust, en dat men daarom overal die waarschijnlijkheid moet berekenen;*
> *... om aldus door getallen uit te drukken de waarschijnlijkheid, dat een regterlijk vonnis regtvaardig of onregtvaardig, dat een afgelegd getuigenis waarheid of leugen bevat, zoo dit alles, ik erken het, op het voetspoor van anderen, door denzelfden Hoogleraar Quetelet geschied is; neen, ik kan en mag mijne vrees, van op die wijze al de waarde van pligten en zedeleer te zien verloren gaan, niet verbergen. Men moest nog maar gaan berekenen, welke waarschijnlijkheid er is, dat een pas geboren kind een dief, moordenaar enz. zal worden, om dan, wanneer hij dit werkelijk geworden moge zijn, zich met de force irrésistible des nombres te verdedigen...*

Wel, nog in 't zelfde jaar heeft Quetelet de "penchant au crime" ingevoerd - misschien is den Tex het geweest die hem op dit idee heeft gebracht - en uitgerekend: 0,00003621 - fatsoenlijken en onfatsoenlijken uiteraard over één kam geschoren. Hè, met zo'n beetje "penchant au crime" bega je toch geen misdaad en voor een oppassend iemand is zelfs dit te veel!

Maar Quetelet had ook zijn aanhangers en leerlingen - de merkwaardigste: Florence Nightingale, "The Lady with the Lamp" volgens de romantische traditie, maar daarnaast een zeer reële "tante", die de dood ging bestrijden in de lazaretten van de Krim, met sterftestatistieken en met Quetelet's sociale fysica als tweede deel van Gods openbaring.

Zoveel kostbare tijd aan Quetelet besteed! Valt er echt niets anders over 1828 te vertellen? Ja, een heleboel - alleen het is zo'n beetje triviaal.

Op 12 augustus 1828 schreef Legendre in de voorrede tot het eerste supplement van zijn Traité des fonctions elliptiques [40]:

> *... à peine mon ouvrage avait-il vu le jour ... que j'appris avec autant d'étonnement que de satisfaction, que deux jeunes géomètres MM. Jacobi de Koenigsberg et Abel de Christiania, avaient réussi à perfectionner considérablement la théorie des fonctions elliptiques dans les points les plus élevés.*

Stoort het U dat ik na me breedvoerig met Quetelet bezig te hebben gehouden, dit onder de trivialiteiten rangschik? Wel, er zal geen wiskundige in de zaal zitten die de namen van Abel [41] en Jacobi [42] niet kent. Maar hoevelen onder U zegt de naam van Quetelet iets?

Zijn elliptische functies dan zo belangrijk? Ze waren het, als oefenterrein, als startpunt voor complexe functietheorie in het groot, voor Riemann oppervlakken, voor abelse integralen, thetafuncties, algebraïsche meetkunde.

1828, een annus mirabilis in de wiskunde! In en rond 1828: het theorema van Abel [43], het onmogelijkheidsbewijs voor de 5e graadsvergelijking [44], de oplossing van vergelijkingen met commutatieve Galoisgroep, waar het woord "abels" voor "commutatief" vandaan komt [45], het eerste diepgaande voorbeeld van epsilontiek in Dirichlets bewijs van de Fourierreeks [46], Moebius' Barycentrische Calcuul [47], Plücker's homogene coördinaten [48], Sturm's ketens [49], Cauchy midden in de uitwerking van de elasticiteitstheorie [50] - en, last not least, de niet-euclidische meetkunde van Bolyai [51] en Lobačevski [52] - voorwaar een annus mirabilis.

Niets omtrent Gauss? Wel degelijk - de differentiaalmeetkunde [53] van de oppervlakken, met de kromming van Gauss, geodetische lijnen, curvatura integra en theorema egregium. Aan wie moet ik de prijs van dat jaar toekennen? Ik denk aan iemand die je anders zou vergeten: George Green [54], geen wiskundige, maar een natuurkundige. Vereist dit een toelichting? Ik dacht van neen. De naam "Green" spreekt voor zichzelf.

* * *

En dan met een kangoeroe-sprong van 50 jaar naar 1878. Ik pak een naam die U allen kent, omdat elke catalogus van tweedehandse wiskunde ermee begint: E.A. Abbott, auteur van "Flatland, a romance of many dimensions, by a square" [55]. U kent zeker ook het idee van het boekje: een flatlander in drie dimensies te laten denken en bijgevolg ons in vier dimensies. "Vier dimensies" is lang een geliefkoosd populair mathematisch onderwerp geweest, mystieke wiskunde als het ware, zoals vele boeken getuigen [56].

De belangstelling voor meer dimensies, gekromde ruimten en niet-euclidische meetkunde bij filosofen en geletterde leken was gewekt door Helmholtz [57] en aangewakkerd door zijn interpreten. Dat Gauss deze geheimen in zijn graf had mee willen nemen, kwam er nog bij om de nieuwsgierigheid te prikkelen. Er heerste nogal verwarring over de verschillende onderwerpen onderling en ook met de projectieve meetkunde, waar - hoe zot - parallellen elkaar snijden. De mensen waren er veelal tegen. Maar sommigen waren er gelukkig mee, vooral met de vierde dimensie.

Een ervan, in 1878, die de vierde dimensie te pas kwam, was de Leipzigse hoogleraar Zöllner [58], die na aanvankelijk verdienstelijk astrofysisch werk al wat steken had laten vallen toen hij in de ban raakte van William Crookes en tenslotte van het amerikaanse medium Henry Slade, die een tournée door Europa maakte.

Felix Klein [59] bekent dat hij het geweest is, die zonder er erg in te hebben, Zöllner op het spoor van de vierde dimensie heeft gezet, toen hij hem vertelde, dat een in onze ruimte verknoopte gesloten kromme in vier dimensies vanzelf ophoudt een knoop te zijn [60]. Zöllner zag er een middel in om experimenteel te toetsen of spiritistische verschijnselen via de vierde dimensie plaats vinden. Hij stelde Slade de proef voor, die er met het gebruikelijke "we will try it" op reageerde, en ziedaar [61]:

> *Dieser Versuch ist mir nun mit Hilfe des amerikanischen Mediums Hr. Henry Slade zu Leipzig am 17. Dezember 1877 Vormittags 11 Uhr innerhalb einer Zeit von wenigen Minuten gelungen. Die nach der Natur gezeichnete Abbildung des mehr als 1 Millimeter starken und festen Bindfadens mit den 4 Knoten .... ist auf Tafel IV dargestellt .... Während ich die Schürzung nur eines Knotens gewünscht hatte, waren nach wenigen Minuten die auf Taf. IV abgebildeten vier Knoten in dem Bindfaden.*

Jammer dat Alexander de Grote geen Slade tot zijn beschikking had, aldus Zöllner, hij had de Gordiaanse knoop dan niet met zijn zwaard behoeven door te hakken [62]. Want wat is er niet allemaal via de vierde dimensie mogelijk! Spirits schrijven in diverse talen in een gesloten lei, magnetiseren eenzijdig een breinaald, laten tafels verdwijnen en ineens van 't plafond vallen. Alleen een proef om twee uit verschillende houtsoorten gedraaide ringen in elkaar te schakelen lukt niet en evenmin de omzetting via de vierde dimensie van rechtsdraaiend wijnsteenzuur in linksdraaiend druivenzuur.

Felix Klein vertelt nog dat Slade later ontmaskerd werd - ik heb het niet na kunnen trekken. Ook constateert hij hier het begin van de

grosse Mystifikation..., die bald in Verbindung mit Hypnotismus, Suggestion, religiösem Sektierertum, populärer Naturphilosophie usw. lange Zeit die Köpfe beherrschte.

Waarvoor de wiskunde toch niet allemaal goed - of slecht - is. Nog in mijn jeugd was "die vierde dimensie" zowat synoniem met wat toen spiritisme heette. Dat duurde - meen ik - tot de vierde dimensie nieuw emplooi vond in de populariteit opeisende relativiteitstheorie. Maar in 1878 was Einstein pas -1 jaar oud, hetzelfde jaar trouwens, waarin Michelson [63)] het instrument van zijn befaamde proef afmaakte.

Het lijkt een ironie dat het jaar, waarin de vierde dimensie voor 't eerst te algemenen nutte werd aangewend, tevens het jaar was waarin de dimensie als zodanig wiskundig werd afgeschaft - althans voorlopig. In 1878 publiceerde Georg Cantor zijn eerste grootse ontdekking [64)]: de gelijkmachtigheid van continua van alle "dimensies" - ook het begrip machtigheid duikt hier voor 't eerst op. Toen kwam dus het aloude begrip "dimensie" op de tocht te staan - om het met een modewoord te zeggen - en bleef zelfs nog langer op de tocht dan de lichtsnelheid na Michelson, te weten tot Brouwer de topologische invariantie van de dimensie bewees [65)].

Wat was er anders nog aan de hand in en rond 1878? Dedekinds 11^{e} supplement bij Dirichlet [66)] groeide naar zijn definitieve vorm toe. Voor Kronecker heb ik voor dat jaar een artikel opgescharreld, dat nu niet karakteristiek Kronecker is, hoewel het over de Kroneckerse karakteristiek handelt [67)]. Felix Klein werkte met veel verbeeldingskracht en noeste vlijt aan zijn automorfe functies, nog onbewust van het feit dat binnen drie à vier jaar een vuurpijl de lucht in zou schieten, Poincaré [68)] genaamd, die Klein's gloeiwormpjes zou doen verbleken. Lie [69)] zat middenin het scheppingswerk van zijn [70)]

neue Theorie, die ich die Theorie der Transformationsgruppen nenne.

Sylvester [71)] publiceerde in 1878 18 papers, maar Cayley [72)] overtrof hem met een 24-tal. Als ik echter niet tel, maar ook weeg, moet ik de prijs van het jaar toekennen aan een jong, veelbelovend wiskundige: Frobenius [73)], met het idee de compositie van bilineaire vormen als product van lineaire afbeeldingen te interpreteren, met de karakterisering van het kwaternionenlichaam als nuldelervrije algebra, met de bilineaire covariant van Pfaffse vormen - de opmaat van de lineaire groepentheorie.

Nog iets. Maar dan van 1879, Schubert's Kalkül der abzählenden Geometrie [74] - een boeiend fantastisch algoritme, na een eeuw ondanks Van der Waerden [75], Severi [76] en Gerretsen [77] nog niet als zodanig gerechtvaardigd - ik bedoel als naieve methode, dus niet alleen wat de uitkomsten betreft. In 1942 heb ik, in de ban van het Schubert-Kalkül, een prijsvraag van Wiskundig Genootschap oplossende, beloofd de lacune te vullen. Belofte maakt schuld - met dit haast 100 jaar slapende werk wilde ik ook mijn eigen geweten wakker schudden.

* * *

Weer een halve eeuw verder: 1928. Wie ik dan allereerst ontmoet: mijzelf temidden van een kring van toekomstige en gevestigde wis- en natuurkundigen. Na drie keer uit een historisch perspectief te hebben gekeken, laat ik de vierde keer een ooggetuigenis afleggen.

1928 - de jojo-rage was even vlug verdwenen als hij verschenen was, maar de relativiteitstheorie-rage had een taaier bestaan. Dankzij Spengler's "Untergang des Abendlandes" was de taalschat met de woorden "functie" en "functioneel" verrijkt. In de meetkunde waren "Ricci-Kalkül" en Blaschke school "still going strong" met de nadruk op "still". Hardy en Littlewood hadden 16 publicaties in 1928, maar geen over analytische getallentheorie, die door Van der Corput, Schnirelman en Vinogradov belet werd de zachte dood te sterven, die sommigen onder ons haar toewensten. In felle kleuren bloeide nog de Duitse variëteit complexe functietheorie, waarvan heden als laatste nabloei het Bieberbach-vermoeden getuigt. O quae mutatio rerum sinds een Landau-titel als "Über die Blochsche Konstante und zwei verwandte Weltkonstanten" [78] de wereld kon doen wankelen. Weet je nog wel, oudje, die opwinding bij Landau's bewijs [79] in 6 regels van de irreducibiliteit van de vergelijking voor de cirkelverdeling en het nog eenvoudigere, al was het over een hele bladzij, van Issai Schur [80]. En Issai Schurs simpel bewijs [81] voor de stelling van Stickelberger dat de discriminant van een algebraïsch getallenlichaam congruent 0 of 1 mod 4 is? Of de nieuwe bewijzen voor de analyticiteit van oplossingen van elliptische differentiaalvergelijkingen? [82] Of, zekere dinsdag, Eberhard Hopf's bewerking dat de - later naar hem genoemde - singuliere integraalvergelijking geen oplossing had, en de volgende dinsdag zijn colloquiumvoordracht mèt de oplossing van dezelfde vergelijking? [83]. Of op hetzelfde Colloquium Von Neumann [84] over Gesellschaftsspiele [85] of over de verdeling van het boloppervlak in drie congruente deelverzamelingen, waarvan twee, geschikt verdraaid, ook al het boloppervlak opvullen? [86] Of even vreemd als indrukwekkend C.L. Siegel [87], die op slinkse manier werd overgehaald, de eed te breken, geen tijdschriftartikelen meer te publiceren.

Wat deden we zo in die tijd? De een dook diep in Hilberts Zahlbericht [88], de ander in Julia's iteraties [89], anderen in eindige groepen of hun voorstellingen, weer een ander, dankzij Karl Loewner [90] in de continue tegenhanger ervan, de continue groepen van Sophus Lie over Elie Cartan [91] tot Hermann Weyl [92] - Weyls grootse publicaties [93] op dit gebied waren nog kersvers, terwijl hijzelf met deze wetenschap al de quantummechanica was binnengestapt. Ook die maakte furore bij ons: Schrödinger als opvolger van Planck en Heisenberg in het Physikalische Kolloquium. Alleen wat Van der Waerden straks Moderne Algebra [94] zou noemen, deed het nog niet bij ons ondanks één fanatiekeling voor p-adiek - ik denk niet dat ik vóór een uitstapje in 1930 naar Artin in Hamburg ooit Steinitz [95], Emmy Noether [96] en Krull [97] gelezen heb. Grondslagen ja - dankzij Von Neumann. Hilbertruimte ja, dankzij Erhardt Schmidt [98] en Von Neumann.

En dan natuurlijk topologie. Maar niet in de stijl van G.T. Whyburn [99], Sierpinski [100], W.L. Ayres [101], de olympikers van 1928 met resp. 26, 18 en 12 publicaties. Ook niet in de stijl van een periode die in 1928 afgesloten werd met Karl Menger's boek Dimensionstheorie [102], maar veeleer langs de lijn die Heinz Hopf doortrok vanuit Brouwer over Erhardt Schmidt.

Ik zou nog een poos kunnen doorgaan met te vertellen wat je in 1928 las en wat je hoorde, op colleges, colloquia en in de wandeling. Over het meest opwindende uit die tijd heb ik nog gezwegen. Het gesprek van de dag - bij ons en buiten - was wat - na de politieke revolutie van 1918 - een revolutie in de wiskunde leek. Of was het maar - na de Kapp-Putsch van 1920 - een tot falen gedoemde putsch? Ik bedoel het intuïtionisme. Is het Hilbert geweest die het als een "putsch" had afgedaan? In elk geval voor wie er mee sympathiseerden, was "putschist" een geuzennaam [103]. In 1927 had Brouwer in Berlijn college gegeven; in 1928 hield hij zijn indrukwekkende Weense lezing [104] en verscheen Hilberts tweede Hamburgse lezing [105] met Hermann Weyls repliek [106].

Overdrijf ik als ik deze affaire het meest opwindende rond 1928 noem? Ik neem Hilbert zelf tot getuige, ik citeer uit zijn lezing [107]

> *Ich staune unter diesen Umständen darüber, dass ein Mathematiker an der strengen Gültigkeit des Tertium non datur zweifelt. Ich staune noch mehr darüber dass, wie es scheint, eine ganze Gemeinde von Mathematikern sich heute zusammengefunden hat, die das gleiche tut. Ich staune am meisten über die Tatsache, dass überhaupt auch im Kreise der Mathematiker die Suggestivkraft eines einzelnen temperamentvollen und geistreichen Mannes die unwahrscheinlichsten und exzentrischsten Wirkungen auszuüben vermag.*

En dan beklaagt Hilbert zich bitter over mensen die durven twijfelen aan zijn bewijs van de "Widerspruchslosigkeit" en aan zijn oplossing van het continuumprobleem, waarvoor immers alleen nog Bernays [108)] en Ackermann [109)] enkele hulpstellinkjes hoeven aan te dragen.

Nog drie jaar en dan verschijnt Gödels vermaarde artikel [110)] - de onbeslisbaarheidsstelling. De zeepbel van wat Von Neumann eens Hilberts slechte geweten heeft genoemd, wordt doorgeprikt. Brouwer zelf zweeg.

Et le combat cessa, faute de combattants.

* * *

En nu vandaag de dag, 1978. Of moet ik de tijdmachine laten doorrazen tot 2028? Het zou op hetzelfde neerkomen - de naaste toekomst of de verste. Het meest opzienbarende in het laatste verleden was ongetwijfeld de oplossing van het vierkleurenprobleem [111)], en nog wel met grootscheepse assistentie van de computer. Misschien even opzienbarend: de constructie van een onbreekbare code [112)] - elementaire getallentheorie, maar wederom geassisteerd door de computer. Wel, dit was al weer gisteren. Het heden telt alleen voorzover het drachtig is met de toekomst.

Ik heb met deze flitsen niet alleen op de toppen van de wiskunde gemikt. Hoe komt de wiskunde op anderen over - was een impliciete vraag: Logaritmentafels en cirkelkwadratuur, statistieken, vierde dimensie, grondslagenstrijd - was achtereenvolgens het antwoord. Hoe zou het antwoord heden luiden? Computers - supersnelle met reusachtige geheugens, en zakcomputers (zoals er zakagenda's zijn) om de eerbied voor de grote computers wat af te zwakken.

De Wiskunde is afgedaald uit ivoren torens naar het marktplein. Ik denk dat ik het niet te zeer mis heb, als ik beweer, dat productie en producentental, consumptie en consumentental en het aantal hunner die wiskunde onderwijzen en leren - dat elk van deze groepen in elke halve eeuw sinds de oprichting van Wiskundig Genootschap met een factor tien is gegroeid. Ten onzent komt thans de meerderheid van de jeugd tot 16 jaar met een soort wiskunde in aanraking. Wiskundigen en wiskunde penetreren diep in wetenschap, techniek en maatschappij. Getalsmatig bekeken lijken we spoedig aan verzadiging toe. Laten we ons niet zelf bedriegen. De factor 10 per halve eeuw, waarmee ik de expansie van de wiskunde heb willen beschrijven, is zeker niet toepasselijk op de groei van begrip en goed gebruik van de wiskunde. Integendeel, met de expansie van de wiskunde zouden wanbegrip en misbruik zelfs kunnen zijn toegenomen.

Hoe komt de wiskunde op anderen af? U kunt er zich een idee van vormen als U de examenvraagstukken MAVO, HAVO, VWO ter hand neemt. Een carricatuураal beeld? Ja, want onderwijs is in de regel beter dan de examens die dienen om het te toetsen. Desniettemin, de kloof tussen echte wiskunde en dat soort wiskunde is enorm. Hebben wij, op 't hoogste niveau, iets misdaan of iets nagelaten in het kweken van begrip voor wat wiskunde wezenlijk is en wat je wezenlijk met wiskunde kunt doen?

Het is niet een kwestie van de hand in eigen boezem steken. Onze cultuur in de breedte toegankelijk maken, is een algemeen pedagogisch en sociaal probleem - alleen meen ik dat de bijdrage die de wiskundige hiertoe kan leveren, paradigmatisch zou zijn. Een taak voor de volgende halve eeuw.

Totzover het wanbegrip. En dan het misbruik van de wiskunde, dat trouwens veelal voortvloeit uit wanbegrip of dankzij wanbegrip welig kan tieren. Beweringen staven, hard maken met cijfers, is een goede zaak, althans als de cijfers deugen en er niet mee geknoeid wordt. Geen onderdeel van de wiskunde wordt zo critiekloos toegepast als de mathematische statistiek, die expres uitgevonden werd, om empirische gegevens critisch te bewerken. Van het hoge aanzien van de wiskunde profiteert de pseudo-wiskunde; wiskunde is thans een ritueel op tal van wetenschapsgebieden: wiskundige franje om de geloofwaardigheid en respectabiliteit van onderzoek te verhogen. Kunnen wij het ons als wiskundigen permitteren, ons er niets van aan te trekken? Een rhetorische vraag waar we in de volgende halve eeuw het antwoord op verschuldigd zijn. Laten we inmiddels afspreken: geen wetenschapsbeoefenaar draagt zo grote verantwoordelijkheid in wat hij doet en laat, voor onderwijs en voorlichting, als de wiskundige.

Op vijf tijdstippen in de historie, gescheiden door vier keer een halve eeuw ben ik mijn zoeklicht gaan richten. Leek het soms of ik de draak stak met de wiskunde of met wat pretendeerde wiskunde te zijn? Wel, als het een carnavalsoptocht was, dan is toch met een boetepreek geëindigd.

Als wiskundige mag je wel, moet je zelfs, de draak steken met het tientallige stelsel. Het getal 200 is geen wiskunde, maar feest. Wiskunde is altijd feest, dagelijks feest. En een heel groot feest, als zoveel wiskundigen bijeen zijn. Feesten zijn er om gevierd te worden. En dit feest vooral.

NOTEN

1) gerekend van 1780.
2) Antoine de Condorcet 1743-1794.
3) Antoine L. Lavoisier 1743-1794.
4) Eise J. Eisinga, 1744-1828.
5) Bouw begonnen 1774, reeds "gangbaar" in 1778, voltooid in 1781. Zie de beschrijving door J.H. van Swinden "Beschrijving van en konst-stuk verbeeldende een volleedig beweegbaar hemels-gestel uitgedagt en vervaardigd door Eise Eisenga, Franeker 1780.
6) Hollandsche Maatschappij der Wetenschappen 1752, Zeeuwsch Genootschap der Wetenschappen 1765, Bataafsch Genootschap der Proefondervindelijke Wijsbegeerte 1769, Teylers Tweede Genootschap 1778, Provinciaal Utrechtsch Genootschap van Kunsten en Wetenschappen 1773, Natuurkundig Gezelschap te Utrecht 1777, Felix Meritis 1777, Natuurkundig Gezelschap te Middelburg 1780, Doctrina et Amicitia 1788.
7) Franz of Anton Mesmer, 1733 of 1734 - 1815.
8) 1741-1808. Infinitinomii dignitatem indeterminarum leges ac formulae. Göttingen 1778.
9) 1749-1790.
10) Neue und erweiterte Sammlung logarithmischer, trigonometrischer und anderer Tafeln, Berlin 1778.
11) Georg Freiherr von Vega 1756-1802. Logarithmische trigonometrische und andere zum Gebrauch der Mathematik eingerichtete Tafeln und Formeln, Wien 1783.
12) 1750 - ? Gegevens volgens Allg. Deutsche Biogr.6, 612.
13) Histoire de l'Acad.R. année 1775, 61-66 (1778).
14) Zie omtrent perpetuum mobile: A.W.J.G. Ord-Hume, The history of an obsession.
15) Etienne Bezout 1730-1783.
16) Théorie générale des equations algébriques, Paris 1779.
17) J.L. Lagrange (1736-1813), Réflexions sur la résolution algébrique des équations, Nouv.Mémoires Acad.R. Berlin 1770 (1772), 134-215; 1771 (1773), 138-254 = Oeuvres III, 205-421.
18) Pierre Simon de Laplace 1749-1827.
19) Recherches sur plusieurs points du système du monde; suite, Mém.Acad.R. Paris 1776 (1779), speciaal XXXVIII = Oeuvres IX (1893), 283-310, speciaal 301-310.

NOTEN (2)

20) A.L. Cauchy (1789-1857), prijsvraag van 1816, verschenen 1827 = Oeuvres (2)1, 1-318.

21) A. de Moivre (1667-1754), The Doctrine of Chances, 2nd ed. 1738.

22) Preciezer: P.S. Laplace, Mémoires sur l'usage du calcul des différences partielles dans la théorie des suites. Mém.Acad.R. Paris 1777 (1780) = Oeuvres IX (1893) 313-335. Mémoire sur les suites. Mém.Acad.R.Paris 1779 (1782), 207-309 = Oeuvres X (1894), 1-89.

23) Zie H. Freudenthal, Operatoren - van Heaviside tot Mikusinski, Simon Stevin 33 (1959), 13-19.

24) Laplace, Théorie des attractions des sphéroides et de la figure des planètes. Mém.Acad.R.Paris 1782 (1785), 113-196, spec. 135 = Oeuvres X(1894), 341-419, spec. 362.

25) Lagrange, Remarques générales sur le mouvement de plusieurs corps qui s'attirent mutuellement en raison inverse des carrés des distances. Nouv.Mem.Acad.Berlin 1777 = Oeuvres IV (1869), 401-418, speciaal 402; Sur l'équation séculaire de la Lune. Mém.Acad.Paris,savants étrangers 7 (1773) = Oeuvres VI (1873), 335-399, speciaal 349-350.

26) 1707-1783.

27) Observationes circa divisionem quadratorum per numeros primos. Disquisitio accuratior circa residua ex divisione quadratorum altiorumque potestatum per numeros primos relicta. Opuscula analytica I, 64-84, 121-156 = Comm.Arith. (1849),I, 477-486, 487-506 = Opera (1) 3(1941), 497-512, 513-543. In feite heeft Euler de reciprociteitswet al in 1744-46 bezeten. Comm. Petropol. 14 (1744), 151-181 = Comm.Arith(1849), I,35-49 = Opera(1)2 (1917), 194-222.

28) A.M. Legendre 1752-1833.
Recherches d'analyse indéterminée. Memoires Acad.R. Paris 1785 (1788), 465-559.

29) C.F. Gauss, Disquisitiones arithmeticae 1801 = Werke I (1863), Speciaal: Sectio quarta.

30) De repraesentatione superficiei sphaerica super plano. Acta Petr. 1777, I, 107-132 (1778) = Opera (1) 28 (1955), 248-275.
Vertaling Ostwalds Klassiker 93.

31) Sur la construction de cartes gêographiques. Nouv.Mém.Acad.R.Berlin 1779, 161-210 = Oeuvres IV (1869), 637-692.

NOTEN (3)

32) J. le Rond d'Alembert (1717-1783). Essai d'une nouvelle théorie de la résistance des fluides. Paris 1752, 60 e.v.

33) 1796-1874.

34) Brussel 1828.

35) Recherches statistiques sur le royaume des Pays Bas, Nouv.Mém.Acad.R. Sci.Lett. Bruxelles 5 (1829), VI + 58 p. + 12 tables. Zie ook: Du nombre des crimes et des délits dans les provinces du Brabant méridional, des deux Flandres, du Hainaut et d'Anvers, pendant les années 1826, 1827, 1828. Correspondance Math. 5 (1829), 177-187.

36) Verhandelingen K.Vlaamse Akad.Wet. 28 (1966), No. 88, 52 blz.

37) Paris 1835, Bruxelles 1836. Zie ook: Du système social et les lois qui le régissent, Paris 1848.

38) Sur l'homme... p. 8-10.

39) E.A. den Tex, Over de dwalingen en verderfelijke stellingen, tot welke de voorstelling van den Burgerstaat als werktuig, en deszelfs beschouwing uit enkel materiëele oogpunten leiden. Bijdragen tot Rechtsgeleerdheid en Wetgeving 4 (1829), 9-38.

40) Paris 1825-28, 3 delen.

41) N.H. Abel 1802-29.

42) C.G. Jacobi 1804-51.

43) Crelle 3 (1828), 313-323 = Oeuvres 1 (1881) 444-456.

44) Crelle 1 (1826), 66-84 = Oeuvres (1881), 66-87.

45) Crelle 4 (1829), 131-156 = Oeuvres (1881), 478-507.

46) G. Lejeune Dirichlet (1805-59). Crelle 4 (1829), 157-169 = Werke 1 (1889), 117-132.

47) A.F. Moebius, Der barycentrische Calcul, ein neues Hilfsmittel zur analytischen Behandlung der Geometrie = Werke I (1885), 1-388.

48) J. Plücker (1801-68). Crelle 5 (1829), 1-36 = Ges.Abh.I (1895), 124-138.

49) J.K.Fr. Sturm (1803-1855). Mémoire sur la resolution des équations. Bull. Férussac 11 (1829).

50) Talrijke opstellen in Exercices de Math. 2-4 (1827-29) = Oeuvres (2) 7-9 (1889-91).

51) Janos Bolyai 1802-1860.

52) N. Lobačevski 1793-1856.

53) C.F. Gauss, Disquisitiones generales circa superficies curvas, Comm. Göttingen 6, 1827 (1828) = Werke IV (1873), 217-258 = Ostwalds Klassiker 5.

NOTEN (4)

54) 1793-1841. An essay on the application of mathematical analysis to the theories of electricity and magnetism, Nottingham 1828 = Math. Papers (London 1871), 1-116 = Ostwalds Klassiker 61.

55) Tweede editie van 1884 - ik heb zelfs de datum van de eerste niet kunnen achterhalen.

56) Ten onzent bijv. H. de Vries, De vierde dimensie, Groningen 1915, 2e ed. 1925.
R. Weitzenböck, Over de vierde dimensie, Rede. Groningen 1923.
R. Weitzenböck, Der vierdimensionale Raum. Braunschweig 1929. [Nieuwe uitgebreide editie:] Basel 1956. [Bevat veel gegevens.]
M. Maeterlinck, La vie de l'espace. Paris 1928.
Nothing All (J.G.G. Nottrot), Inzicht in de vierde dimensie, met voorwoord van Ch.H. van Os, Groningen, z.j.
F. Ortt, De supercosmos, den Haag 1949.

57) H. Helmholtz (1821-1894). Über die Thatsachen, die der Geometrie zugrunde liegen, Nachr. Ges. Wiss. Göttingen 1868, 193-221 = Wiss.Abh.II (1883), 618-639. Zie ook H. Freudenthal, Die Grundlagen der Geometrie um die Wende des 19. Jahrhunderts. Math.-phys. Semesterberichte 7 (1960), 1-25.

58) J.C.F. Zöllner (1834-1882). Wissenschaftliche Abhandlungen I, II_1, II_2, III, IV, Leipzig 1878-81.

59) 1849-1925. Vorlesungen über die Entwicklung der Mathematik im 19. Jahrhundert, Berlin 1936, 169-170.

60) F. Klein, Ueber den Zusammenhang der Flächen. Math.Ann. 7 (1876), 476-482, speciaal 478 = Ges.Abh. 2 (1922), 67. Zie ook Zöllner I, 276.

61) Wiss.Abh. I, 726. Zie ook II, 214-215.

62) I, 730.

63) A.A. Michelson 1852-1931.

64) G. Cantor (1845-1918). Ein Beitrag zur Mannigfaltigkeitslehre, Crelle 84, 242-258 = Ges.Abh. 1932, 119-137.

65) L.E.J. Brouwer, Beweis der Invarianz der Dimensionszahl, Math.Ann. 70 (1911), 161-165 = Coll.Works II (1976), 430-434.

66) Vorlesungen über Zahlentheorien 3. Aufl. 1879, 4.Aufl. 1894.

67) L. Kronecker (1839-1914), Über die Charakteristik von Functionensystemen. Monatsber.Akad. Wiss. Berlin 1878, 145-152 = Werke II (1897), 71-82.

NOTEN (5)

68) Henri Poincaré (1854-1912).

69) S. Lie (1842-1899).

70) Inleiding tot het eerste artikel in Math.Ann. 16 (1880), 441-528, waar publicaties elders aan vooraf waren gegaan.

71) J.J. Sylvester (1814-1897).

72) A. Cayley (1821-1895).

73) G. Frobenius (1849-1917), Über das Pfaffsche Problem. Crelle 82 (1877), 230-315. Über homogene totale Differentialgleichungen. Crelle 86 (1979), 1-19. Über die schiefe Invariante einer bilinearen oder quadratischen Form. Crelle 86 (1879), 44-71. Mit Stickelberger: Über Gruppen von vertauschbaren Elementen. Crelle 86 (1879), 217-262.

74) H.C.H. Schubert (1848-1911). Leipzig, 1879.

75) B.L. van der Waerden, De algebraiese grondslagen van de meetkunde van het aantal, Zutphen 1926, en een groot aantal publicaties in Math.Ann. 97 (1927) - 115 (1938).

76) Het uitvoerigste in F. Severi, I fondamenti della geometria enumerativa. Annali di Mat. (4) 19 (1940), 153-242. Vertaling: Grundlagen der abzählenden Geometrie. Wolfenbüttel, z.j.

77) J.C.H. Gerretsen. Inleiding tot een topologische behandeling van de meetkunde van het aantal. Groningen 1939.

78) E. Landau (1877-1938), Math. Zeitsch. 30 (1929), 608-634.

79) Über die Irreduzibilität der Kreisteilungsgleichung. Math. Zeitsch. 29 (1928), 461.

80) I. Schur (1875-1941). Über die Irreduzibilität der Kreisteilungsgleichung. Math. Zeitsch. 29 (1928), 462.

81) Elementarer Beweis eines Satzes von Stickelberger. Math. Zeitsch. 29 (1928), 464-465.

82) H. Lewy, Neuer Beweis des analytischen Charakters der Lösungen elliptischer Differentialgleichungen. Math. Ann. 101 (1929), 609-619.

83) E. Hopf, Zum Problem des Strahlungsgleichgewichts in den äusseren Schichten der Sterne. I.II. Z.f. Physik 46 (1928), 374-382, 49 (1928), 155-161.

84) J. von Neumann (1903-57).

85) Zie J. v. Neumann, Math. Ann. 100 (1928), 295-320 = Coll. Works. VI (1963), 1-28.

NOTEN (6)

86) Zie J. v. Neumann, Zur allgemeinen Theorie des Maszes. Fund. Math. 13 (1929), 73-116 = Coll. Works I (1961), 599-642.

87) C.L. Siegel (1896- ?). Über enige Anwendungen diophantischer Approximationen. Abh. Akad. Berlin 1929, Nr. 1, 70 S.

88) D. Hilbert, Die Theorie der algebraischen Zahlkörper. Bericht, erstattet der Deutschen Mathematiker-Vereinigung. Jahresbericht DMV 4 (1897), 175-546 = Ges.Abh. I (1932), 63-363.

89) G. Julia, Mémoire sur l'itération des fonctions rationnelles. Journal de Math. (8) 1 (1918), 47-245.

90) 1893-1968.

91) 1869-1951.

92) 1885-1955.

93) Theorie der Darstellung kontinuierlicher halbeinfacher Gruppen durch lineare Transformationen I, II, III, Nachtrag. Math. Zeitschr. 23 (1925), 271-309, 24 (1926), 328-395, 789-791 = Ges.Abh. II (1968), 543-647. Integralgleichungen und fastperiodische Funktionen. Math. Ann. 97 (1927), 338-356 = Ges. Abh. III (1968), 338-356. Mit F. Peter: Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuierlichen Gruppe. Math. Ann. 97 (1927), 737-755 = Ges. Abh. III (1968), 58-75.

94) B.L. van der Waerden, Moderne Algebra I, Lpz. 1930.

95) 1871-1928.

96) 1882-1935.

97) W. Krull (1899-1971).

98) 1876-1959.

99) 1904-1969.

100) 1882-1969.

101) 1905- ?

102) Karl Menger (1902-), Dimensionstheorie, Leipzig 1928.

103) Dr. h.c. N^2 (Hubert Cremer). Häufungspunkte. Berlin 1927. Mathematische Schnaderküpfel 17: Und wird mir das ganze Getu hier zu trist, dann kauf ich mir 'ne Kanone und werde Putschist.

104) L.E.J. Brouwer (1881-1966). Mathematik, Wissenschaft und Sprache. Monatsh. Mathematik 36 (1929), 153-164 = Coll.Works I (1975), 417-428.

105) D. Hilbert, Die Grundlagen der Mathematik. Abh. Math. Sem. Hamburg 6 (1928), 65-85 = Ges.Abh. III ...

NOTEN (7)

106) H. Weyl, Diskussionsbemerkungen ... Abh. Math. Sem. Hamburg 6 (1928), 86-88 = Ges. Abh. III (1968), 147-149.

107) 80-81.

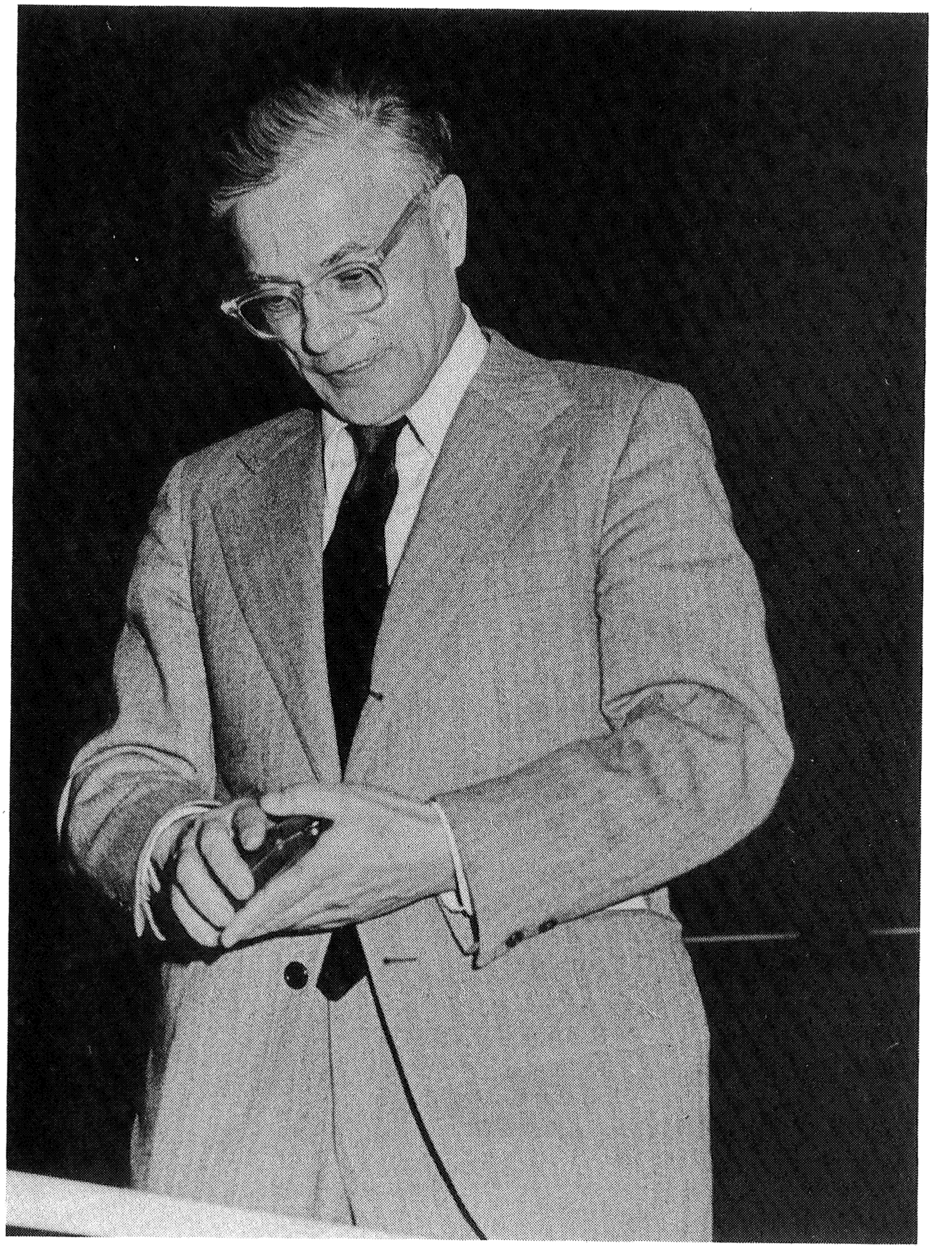
108) Paul Bernays (1888-1977).

109) W. Ackermann (1896-1962).

110) Kurt Gödel (1906-1978). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. Monatshefte Math. Phys. 38 (1931), 173-198.

111) K. Appel & W. Haken, Every planar map is four colorable. Bull. Amer. Math. Soc. 82 (1976), 711-712.

112) R.L. Rivest, A. Shamir, L. Adleman. On digital signatures and public-key cryptosystems. Techn. Memo 82, April 1977, Lab. Computer Science, MIT.



ON THE DEVELOPMENT OF LIE GROUP THEORY

A. Borel

L.E.J. Brouwer Memorial Lecture

In his letter of invitation to give this lecture, Professor Freudenthal pointed out that two choices had been made, first of a field, and then of a mathematician having worked in it. In the present case, the field is Lie groups. Accordingly, I shall devote this lecture to the award winning field, so to say, and take as my main theme some aspects of the development and role in mathematics of Lie groups. In doing so, I shall view "Lie groups" in a rather broad sense, including not only the classical real or complex Lie groups, but linear algebraic groups and p-adic Lie groups as well.

The theory of "finite and continuous groups", later called Lie groups, was built from about 1873 on by the Norwegian mathematician Sophus Lie. It arose out of his work on differential equations and contact transformations, and he had a main goal in mind, namely to develop a Galois theory of differential equations, in which these groups would play the role of the Galois group of an algebraic equation. It seems to me that, from that point of view, Lie groups offer a, by no means unique, example of a theory created for a certain purpose, but not fulfilling it up fully. However, it then went off into many directions, so much so that, one hundred years later, J. Dieudonné was led to write: "Les groupes de Lie sont devenus le centre des mathématiques; on ne peut rien faire de sérieux sans eux" [1). Such a drastic statement was of course quickly challenged and, even to me, seems too sweeping. But I do share the view that Lie groups, as understood here, play a major, even central role in an important and ever spreading part of mathematics.

In only one hour, I cannot present in a comprehensive manner all the evidence to back up such a claim. Without trying to be systematic, I shall

1) J. Dieudonné, "Orientation générale des mathématiques pures en 1973", Gazette des Mathématiciens, Oct. 1974, p. 73-79, Soc. Math. France.

concentrate on some developments which seem to me to have been crucial in increasing the scope of the theory and of its connections with other parts of mathematics. In a topic with such a rich history, even the choice of those is not necessarily unique, and may depend upon one's own perspective. From mine, I would like to single out three:

(1) Global Lie groups and Riemannian symmetric spaces.

(2) Linear algebraic groups.

(3) Buildings and p-adic symmetric spaces.

Before taking them up, I should make some remarks on the first fifty years of the theory of Lie groups.

1. It is a simple matter nowadays to define a real or complex Lie group, as a real or complex analytic manifold G endowed with a group structure such that the map $G \times G \to G$ given by $(x,y) \mapsto x \cdot y^{-1}$ is analytic. But S. Lie could not say that, and this definition differs from his in two respects. First, he considered only transformation groups. The notion of abstract group was not familiar at the time, and even later, when it had become more widespread, F. Klein had some misgivings about putting it in the foreground (Entwicklung der Mathematik I, 335-336). But S. Lie had a notion of isomorphism, called "Gleichzusammensetzung", which focussed attention on the law of composition, and he knew that each isomorphism class could be represented by the group acting on itself by left or right translations. So the difference here is more one of terminology than of substance. The second one is more important: his groups were local; G was in fact a neighborhood of the origin in C^n (mostly, occasionally R^n), the law of composition was defined for elements sufficiently close to the origin, and given by convergent power series. As you know, the thrust of the general theory was to reduce problems on such local groups to algebraic ones on what is now called the Lie algebra of G, that is, the vector space of left invariant vector fields, endowed with the bracket operation of infinitesimal transformations. Lie's approach was analytic, and much influenced by his work on contact transformations. For instance, it seems that an important step for him was the interpretation of the Poisson bracket of two functions as the bracket of two infinitesimal contact transformations. The purely algebraic problems to which his theory led were later solved mostly be other mathematicians, notably by W. Killing and, above all, by E. Cartan.

During this first period, the theory was not purely local, however. Lie and his contemporaries were familiar with the classical groups; namely the special or general linear group $SL_n(C)$ or $GL_n(C)$, the orthogonal and symplectic groups O(n,C) and $Sp_{2n}(C)$, except that they viewed them usually as groups of projective transformations rather than of homogeneous linear ones. Furthermore, a number of investigations were frankly global in character, as for instance Hurwitz's construction of invariants for $SL_n(C)$ or SO(n,C) by integration over the compact groups SU(n) and SO(n) (1897). Also, I should not miss this opportunity to allude to the work of L.E.J. Brouwer on Lie groups and on Brouwer's correspondence with F. Engel (1909-1911), in which Brouwer clearly has a global picture of Lie transformation groups.

2. In spite of this, I would still say that the global theory really got off the ground with H. Weyl's famous Math. Zeitschrift papers (1925-26). There, H. Weyl combined two approaches to representation theory which until then had progressed unaware of one another: the infinitesimal one of E. Cartan and the Frobenius-Schur theory of characters of finite groups, which had been transposed to some classical compact groups by I. Schur, using Hurwitz's integration trick. Particularly striking was the fact that a topological result, the finiteness of the fundamental group of a compact semi-simple group, played a key role in the proof of an algebraic theorem, namely the full reducibility of the finite dimensional representations of the complex semi-simple Lie algebras. This work led a bit later to the Peter-Weyl theorem, and also paved the way for the applications of Lie group representations to physics, which have steadily gained in importance since then.

The work of H. Weyl also had a considerable and almost immediate impact on E. Cartan. At that time, the latter was working on a seemingly unrelated problem, the classification of Riemannian manifolds in which the curvature tensor is invariant under paralleltransport, and had noticed a strange relationship with the classification of simple real Lie algebras he had carried out ten years earlier. His approach had been local. Under the influence of H. Weyl, he recast the question in global terms, explored it further and built up a beautiful theory of semi-simple Lie groups and symmetric spaces in which both were inextricably linked. Apart from flat factors, these spaces are homogeneous spaces of semi-simple groups, and are products of quotients G/K, where either G is compact semi-simple, K the fixed point set of an involution of G and G/K has positive curvature, or G is simple non-compact with finite center, K is a maximal compact subgroup, and G/K has negative curvature. A beautiful illustration of the interplay between groups and differential geometry is the proof of the conjugacy of the maximal compact subgroups of a semisimple group via a fixed point theorem asserting that any compact group of isometries of a complete simply connected Riemannian manifold with negative curvature has a fixed point. For about thirty years, this was the only one.

For G compact, E. Cartan also extended the results of H. Weyl on $L^2(G)$ to $L^2(G/K)$, and began the study of the real homology of G/K via invariant differential forms. This is the origin of the "de Rham theorems", which he conjectured on that occasiom, and of Lie algebra cohomology. Later on (from the late forties on), the topology of compact Lie groups, of the Grassmannians, and representation theory also became prominent in fibre bundle theory, in particular in the study of characteristic classes.

At this point, we can see that Lie group theory, which started as a chapter of analysis and differential equations, then turned to algebra, had become linked with topology and differential geometry. But the significance of this development went beyond this, already because certain symmetric spaces or homogeneous spaces occur in many parts of mathematics. On the compact side, these spaces include in particular such well known ones as the spheres, the projective spaces and the Grassmannians over the real, complex or quaternionic numbers, the Cayley projective plane, and many projective varieties, for instance the flag manifolds (see below), which play a basic role in Schubert's enumerative calculus. Among the non-compact symmetric spaces are to be found the Poincaré upper half-plane and its manifold generalizations: the n-dimensional hyperbolic space, the space $SL_n(R)/SO(n)$ of positive non-degenerate quadratic forms of determinant one on R^n, the space $O(p,q)/(O(p) \times O(q))$ of Hermite minimal majorizing forms of an indefinite quadratic form of index (p,q) on R^{p+q}, the Siegel upper-half plane $Sp_{2n}(R)/U(n)$. This, and the classification of the bounded symmetric domains carried out a bit later by E. Cartan, made it clear that semi-simple groups and symmetric spaces offered a natural framework to study reduction theory with respect to arithmetic groups and automorphic forms in several variables, developed notably by C.L. Siegel in the thirties and the forties. A number of quotients of bounded symmetric domains by arithmetic groups were seen later to parametrize families of abelian varieties, generalizing the relationship between the upper half-plane, the modular group and elliptic curves. This link between semi-simple groups and moduli problems in algebraic geometry was reinforced much later by P. Griffiths' theory of the period mappings for families of smooth projective varieties, whose target spaces are quotients of certain homogeneous spaces of semi-simple groups by arithmetic groups.

The forties also saw the beginning of the study of infinite dimensional unitary representations of semi-simple groups, a topic which grew into a fullfledged harmonic analysis on semi-simple groups, including in particular the spectral decomposition of $L^2(G)$ (the Plancherel formula) by Harish-Chandra. The Weyl-Cartan theory was again an indispensable preliminary to this. In particular a basic step towards the Plancherel formula was the solution of the analogous problem for $L^2(G/K)$, where K is the maximal compact subgroup of G, i.e. the study of the so-called spherical functions on non-compact symmetric spaces.

I could still go on a long way describing further outgrowths of the work of Cartan and Weyl discussed above, but I would now like to pass to a quite different set of ideas, originating in algebraic geometry.

3. Let K be an algebraically closed field. A subgroup G of the general linear group $GL_n(K)$ is said to be algebraic if there exists a set of polynomials in n^2 indeterminates

$$(1) \qquad P_\alpha \in K[(X_{ij})_{1\le i,j\le n}], \qquad (\alpha \in I),$$

with coefficients in K, such that

$$(2) \qquad G = \{g = (g_{ij}) \in GL_n(K) \mid P_\alpha(g_{11}, g_{12}, \ldots, g_{nn}) = 0, \ (\alpha \in I)\}.$$

In analogy with the above definition of a Lie group, one can also proceed more intrinsically and define an affine algebraic group over K as an affine variety over K which is endowed with a group structure such that the map $(x,y) \mapsto x \cdot y^{-1}$ is a morphism of affine varieties. There is no substantial difference between the two notions. If K = C, these groups are complex Lie groups. As such they were already studied in the 19th century, chiefly by E. Picard and L. Maurer. The former had in mind to build up a Galois theory of linear differential equations; this led to the Picard-Vessiot theory. Maurer's motivations were different, in part invariant theory, in part the investigation of Lie groups in which the composition was given by algebraic functions (in suitable coordinates), rather than just by analytic functions. After that, and a Comptes Rendus Note by E. Cartan (1894), the topic seems to have fallen into oblivion for about fifty years. To me this is a mildly curious fact, since the Weyl-Cartan theory pertained first of all to semi-simple groups, which are essentially algebraic. Interest in them was revived in the early forties, first by C. Chevalley, then by E. Kolchin. At that time, the emphasis was on groups over fields as general as possible. Chevalley used an analogue of the exponential mappings of the classical theory to go from the Lie algebra to the group, and was limited to characteristic zero. The first significant results valid in arbitrary characteristic were obtained by E. Kolchin (1948). Interestingly enough, this important step was also motivated by the wish to develop a Galois theory of differential equations, more precisely, a generalization to algebraic differential equations in characteristic p of the Picard-Vessiot theory. Kolchin proved an analogue of Lie's theorem on solvable linear groups now known as the Lie-Kolchin theorem: "every connected solvable algebraic subgroup of $GL_n(K)$ can be put in triangular form", as well as several other results which contained implicitely a rather complete structure theory for solvable groups. However, in the early fifties, the growing importance of certain algebraic homogeneous spaces and the development of abstract algebraic geometry made the need of a more general theory of linear algebraic groups rather widely felt. To go further, some different methods had to be found. They were developed from 1955 on, and led to results and points of view which were new even over C. To try to give an idea of the flavor of those, I would like to sketch the starting point of this theory. For this we may assume that our groundfield K is just C. Let then G be a linear algebraic group which operates on an algebraic variety V, the action being described by a morphism $G \times V \to V$ of algebraic varieties. To make matters simple, assume that $V \subset P_N(C)$ is a projective variety on which G operates by projective transformations, but it could be any variety. Then the "closed orbit lemma" asserts in this case the existence of an orbit $G \cdot x$ $(x \in V)$ which is itself a projective variety. In fact, if $v \in V$, then the orbit map $g \mapsto g \cdot v$ is a rational map of G into V. By a well-known fact, the complement of the image $G \cdot v$ of this map in the smallest algebraic variety $Cl(G \cdot v)$ containing it (its Zariski-closure) is contained in an algebraic variety of strictly smaller dimension. Therefore, if we take $x \in V$ such that $Cl(G \cdot v)$ has the smallest possible dimension, there is no room left and $G \cdot v$ is equal to its closure. (For a general V, this argument shows the existence of a Zariski-closed orbit). As a counterpart to that lemma, think of an irrational line G in a two-dimensional torus T. It acts by translations on T and all the orbits are dense. In fact, this example pertains to real analytic groups, but it can easily be made complex, showing that there is no closed orbit lemma in the complex analytic case. This lemma thus pinpoints a difference between algebraic and analytic groups and, as simple as it is, is the cornerstone of the theory of linear algebraic groups. Assume now G to be connected, commutative. Let x be as above and H the subgroup of G fixing x. Then G/H is on one hand a linear algebraic group, hence an affine variety, and on the other hand a connected projective variety. It is then reduced to a point, hence G has a fixed point on V. Using induction on dimension, one then deduces that any connected solvable group G acting on a projective variety has a fixed point. If $G \subset GL_n(K)$, we may then in particular take for V the variety F_n of full flags in K^n, (i.e. the projective variety whose points are increasing sequences of subspaces $\{V_1 \subset \dots \subset V_{n-1}\}$ in K_n, where $\dim V_i = i$), and deduce that G leaves one flag stable. This is equivalent to saying that G can be put in triangular form, and proves anew the Lie-Kolchin theorem. In this sketch, I have slurred over some technical points, but the proof is in essence valid as is over any algebraically closed field.

I hope this gives some idea of the global arguments which replaced Lie algebra considerations. The theory was rather quickly developed in this framework. A major achievement was the classification of simple algebraic groups by C. Chevalley, which he proved to be independent of K: The simple groups over K are classified by the root systems and lattices, in the same way as over C. It is well-known that the classical groups can be written over Z, i.e. by equations such as (2) above with integral coefficients. C. Chevalley also showed that every type of complex simple group could be so described by conditions over Z, which made sense and defined the corresponding simple group over any K.

4. The next step was to extend the theory to non-algebraically closed groundfields. If k is a subfield of K, let us say that the linear algebraic group $G \subset GL_n(K)$ given by (2) above is defined over k if the ideal of all polynomials $P \in K[(X_{ij})]$ which vanish on G is generated, as an ideal, by elements with coefficients in k. If so, set $G(k) = G \cap GL_n(k)$. If k = R, K = C, then G(k) is a (special kind of) real Lie group. If k is finite, then G(k) is a finite group. In fact, Chevalley's construction over Z yields, for every K, a group defined over the prime field of K, hence over any field. In an earlier paper (1955), Chevalley had performed such an explicit construction starting from complex simple algebras and had shown that the groups G(k) thus obtained were simple, as abstract groups. For k finite, this yielded some new series of finite simple groups, attached to the exceptional simple Lie algebras, the first new finite simple groups since E. Dickson. Some variations of his construction (by R. Steinberg, R. Ree,

M. Suzuki, J. Tits) led to some others. As a result, a solid connection between algebraic groups and finite simple groups was established. As you know, it is rather generally conjectured by the experts on finite simple groups, that apart from those groups and the alternating groups there are only finitely many simple non-commutative finite groups, the so-called sporadic groups. An important part of this classification program is played by the properties and various characterizations of the above simple groups, often called of Lie type or of Chevalley type.

The general study of the groups G(k) was developed from about 1957 on. It led notably to a structure theory of semi-simple, (or, slightly more generally, reductive) groups over arbitrary fields by J. Tits and myself in the early sixties. The notions of Cartan subgroups, roots, Weyl groups, Bruhat decomposition of the classical theory have suitable analogues, but I shall not attempt to describe it more precisely.

5. Tits and I came to it for different reasons. My main motivation was to find in G(k), when k is a number field, subgroups which would allow one to generalize to fundamental domains of arbitrary arithmetic groups the familiar notion of cusp of a fundamental domain of a fuchsian group; J. Tits had been led to this theory by his investigations of various geometries and by the role of the Bruhat decomposition in Chevalley's of 1955 referred to above. Heobtained a far reaching axiomatization which allowed him to give a unified treatment of many known geometries as well as to construct new ones. As you know, according to Klein, a geometry on a space is the study of properties invariant under a given group of transformations. From that point of view, the geometry is governed by the properties of, and the relations between, the isotropy groups of the objects under consideration. Tits turned this around by defining a geometry starting from an aggregate $\mathcal{P}$ of subgroups of a group G with suitable properties. The key notion is that of a (B,N)-pair T, now generally called a Tits system. It consists of two subgroups B,N of G satisfying some rather simple conditions (in fact surprisingly simple, considering the amount of information which can be extracted from them). In particular, $H = B \cap N$ is normal in N and $W = N/H$, called the Weyl group of T, is a Coxeter group. The elements of $\mathcal{P}$, the "parabolic subgroups" of T, are then those which contain a conjugate of B. To T there is associated a simplicial complex X, now called a building. The vertices of X are the maximal elements of $\mathcal{P}$ different from G, and s + 1 vertices $P_0,\dots,P_s$ span an s-simplex if and only if their intersection is in $\mathcal{P}$. The group G operates on the building X by

conjugation and the isotropy groups of the various simplices are just the elements of $\mathcal{P}$. This is the geometry associated to T. If $G = GL_n(k)$, B is the group of upper triangular matrices, N the group of monomial matrices, then $\mathcal{P}$ consists of the stability groups of the flags in k^n (strictly increasing sequences of subspaces of dimension $\neq 0,n$). If we replace any $P \in \mathcal{P}$ by the flag which it stabilizes, then we may view X as the "building of flags in k^n": the vertices are the non-zero proper subspaces of k^n, and vertices $\{V_0,\ldots,V_s\}$ span a simplex if and only if, after renumbering, they form a flag. This complex thus describes the incidence relations among the subspaces of k^n, or of the projective space $P_{n-1}(k)$; by the fundamental theorem of projective geometry, it allows one to recover essentially projective geometry, at any rate for $n \geq 3$.

The Tits system just described has analogues in the groups G(k), where G is semi-simple defined over k. The geometries associated to them include many classical ones, as well as new ones attached to the exceptional groups. This now related algebraic groups with "abstract" geometry, or rather with many geometries such as projective or polar geometry over arbitrary fields. In these first applications of Tits systems and buildings, the Weyl group W had been in most cases a finite euclidean reflection group. It could be in principle also infinite. However, this possibility remained little exploited, until it was noticed that another type of Tits systems was the key to a problem which had been rather baffling until then, the study of maximal compact subgroups of p-adic semi-simple groups.

6. To explain the background to this problem, I have to backtrack a little. I already pointed out that the symmetric spaces offered a natural framework for a general theory of automorphic forms with respect to arithmetic groups, But I was then implicitly referring mainly to its analytic aspects. However, it also has deep arithmetic ones, as exemplified by the work of E. Hecke or C.L. Siegel, for instance. Now in algebraic number theory we are taught to treat as symmetrically as possible all the completions of a number field; a convenient tool for this is the formalism of adeles and ideles, which was extended to algebraic groups in the late fifties. This led to consider, besides real or complex groups, also groups of the form G(k), where k is a non-archimedean local field, say the field Q_p of p-adic numbers, p prime, in particular for semi-simple G.

To study these groups, the first idea was naturally enough to see whether some properties of real groups would carry over, *mutatis mutandis*. After the development of the theory of algebraic groups, it was realized that Cartan's results on the structure of real semi-simple groups were really of two kinds: some, suitably reformulated, could be viewed as special cases of theorems on algebraic groups while the others, such as the conjugacy of maximal compact subgroups and the properties of symmetric spaces, were topological or differential geometric in nature. In other words, some depended on the Zariski topology and the others on the ordinary topology and the C^∞-structures associated to R. The former were now available over non-archimedean fields, and there remained to see whether the others also had some counterpart. With respect to the structures inherited from k, these groups can be viewed as locally compact totally disconnected groups or as Lie groups over k (the definition being the same as over R or C, based on the notion of an analytic manifold over a non-discrete complete valued field). *A priori,* as a topological group, such a group seems quite different from a real or complex group. Still some simple examples led one to hope for the existence of some significant similarities. The first item of business was the determination of the maximal compact subgroups. They were shown to exist in general, and were first determined effectively in some classical cases. They were not necessarily conjugate, but at any rate formed finitely many conjugacy classes. For instance, if $G(k) = SL_n(Q_p)$ these conjugacy classes are represented by the stability groups of the n lattices in Q_p^n spanned over the ring Z_p of p-adic integers by the vectors

$$e_1,\dots,e_i,\ pe_{i+1},\dots,pe_n, \qquad (i = 1,2,\dots,n).$$

It was difficult at first to see any general principle behind this. Furthermore, these subgroups are open, so that the quotient G/K of G by one of them is a countable discrete space, hardly a candidate to play the role of a Riemannian symmetric space. The breakthrough came here with the discovery, by N. Iwahori and H. Matsumoto, that in certain cases the maximal compact subgroups could be described as the maximal elements in the set $\mathcal{P}$ of "parabolic subgroups" associated to a suitable Tits system T in G(k). This was then shown to be a general phenomenon by F. Bruhat and J. Tits. Moreoever, the building X associated to this Tits system supplied a space which turned out to be amazingly analogous to the symmetric spaces of the real theory. In this case, the Weyl group of the Tits system is an infinite reflection group in some affine euclidean space A. The building X is a union of spaces naturally isomorphic to A, endowed with the tesselation defined by the reflection hyperplanes of W. It is contractible, the metrics on the copies of A combine to define a complete metric on T, such that any two points are joined by a unique shortest geodesic segment. There is a metric inequality for geodesic triangles which, in the Riemannian symmetric case, follows from negative curvature, and can be used again to prove that any compact group of isometries has a fixed point. In fact, the proof is valid without change both in the p-adic and the real case. The fixed point theorem is used in an essential way to insure that the maximal elements of $\mathcal{P}$ are exactly the maximal compact subgroups. As in the real case, a geometric fixed point theorem is the key to the description of the maximal compact subgroups.

This theory is one striking evidence of a close analogy one keeps discovering between real and p-adic semi-simple (or reductive) groups. At first, the real case supplied the clues to the p-adic case, but this has since become a two-way street, almost a shuttle, and many results on real groups have been suggested by properties of p-adic groups. This analogy is not formal, but is more in the form of some sort of dictionary, built little by little and still growing. Some remarkable illustrations of this principle are provided by harmonic analysis (for instance the Plancherel theorem or the classification of irreducible admissible representations can be given an essentially common formulation) or by the cohomology of discrete cocompact subgroups.

7. The theory of Lie groups and algebraic groups has now attained a considerable degree of completeness and has found many applications. Its usefulness is nowhere more in evidence than in the present study of automorphic forms and of their connections with arithmetic and algebraic geometry. In the last twelve years or so, Lie groups, algebraic groups, arithmetic groups, have become the framework of a vast program, often referred to as "Langlands' philosophy", in which infinite dimensional representations are brought to bear on the study of Artin L-functions, Hasse-Weil zeta functions of projective varieties, and non-abelian extensions of local or global fields. The full realization of this program does not appear to be in sight, but the old and recent results which illustrate it are so striking and promising that I could not resist alluding to it at this point.

8. For lack of time, I have now to stop this survey, as incomplete as it may be. Still, the entity "Lie groups-algebraic groups" has been related to many parts of mathematics: analysis, differential equations, algebra, topology, differential geometry, fibre bundle theory, arithmetic groups, automorphic forms, algebraic geometry, moduli, finite simple groups, geometry, L-functions, to name the main ones we have met. This may give the impression

that I am trying subrepticiously to prove that, after all, Lie groups are the center of mathematics. But this is not my intention at all. Clearly, some other topics could give rise to a rather similar picture. In fact, I would rather schematize the structure of mathematics by a complicated graph, where the vertices are the various parts of mathematics and the edges describe the connections between them. These connections go sometimes one way, sometimes both ways, and the vertices can act both as sources and sinks. The development of the individual topics is of course the life and blood of mathematics, but, in the same way as a graph is more than the union of its vertices, mathematics is much more than the sum of its parts. It is the presence of those numerous, sometimes unexpected edges, which makes mathematics a coherent body of knowledge, and testifies to its fundamental unity, in spite of its being too vast to be comprehended by one single mind. I hope this lecture has made plausible my belief that Lie group theory is a topic of great vitality and interest both in its own right and by the remarkably important role it has played and is playing in the ongoing process of expansion and unification of mathematics so well described by D. Hilbert in his 1900 Paris address, as a meeting and testing ground for many disciplines and as a starting point, tool, and framework for many incursions into other parts of mathematics.

BIBLIOGRAPHY

A comprehensive bibliography would be out of proportion with the general character of this lecture. I limit myself to a few surveys in which the reader will find many further references.

N. BOURBAKI, *Groupes et algèbres de Lie,* Chap. 2,3, note historique; Act. Sci. Ind. 1349 (1973), Hermann, Paris.

E. CARTAN, *La théorie des groupes finis et continus et l'analysis situs;* Memorial Sci. Math. XLII, 1930, Gauthier-Villars, Paris (Oeuvres complètes, t. 1_2, 1165-1225).

D.V. ALEKSEEVSKI, *Lie groups and homogeneous spaces,* Journal of Soviet Mathematics 4 (1975), 483-539, Plenum Publ. Corp., New York (translated from Itogi Nauk i Tekhniki 11 (1974), 37-123).

N. BOURBAKI, *Groupes et algèbres de Lie,* Chap. 4,5,6, note historique; Act. Sci. Ind 1337 (1968), Hermann, Paris.

J. TITS, *Groupes simples et géométries associées;* Proc. Int. Congress Mathematicians Stockholm 1962, Vol. 1, 197-221.

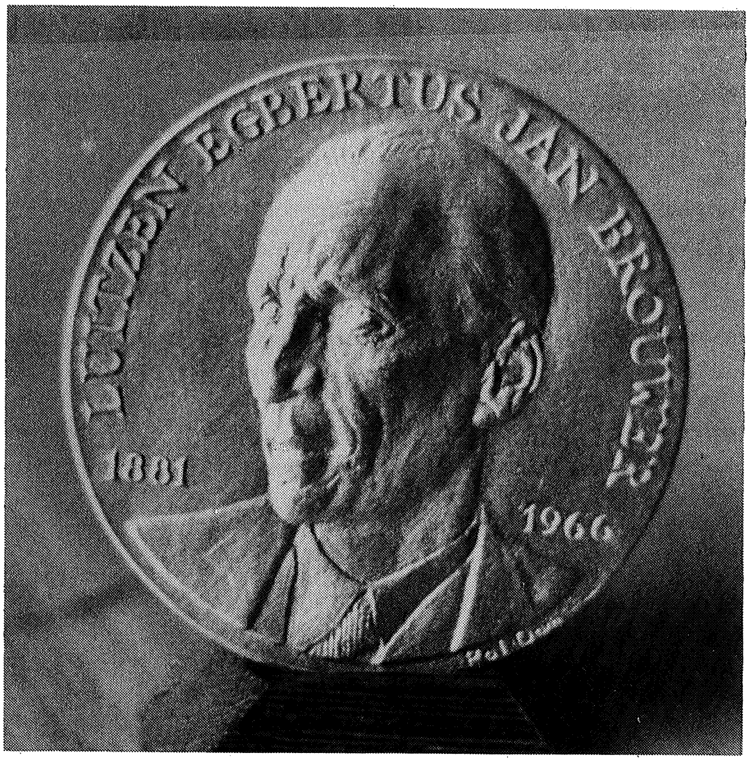
A. BOREL, *Arithmetic properties of linear algebraic groups;* ibid., 10-22.

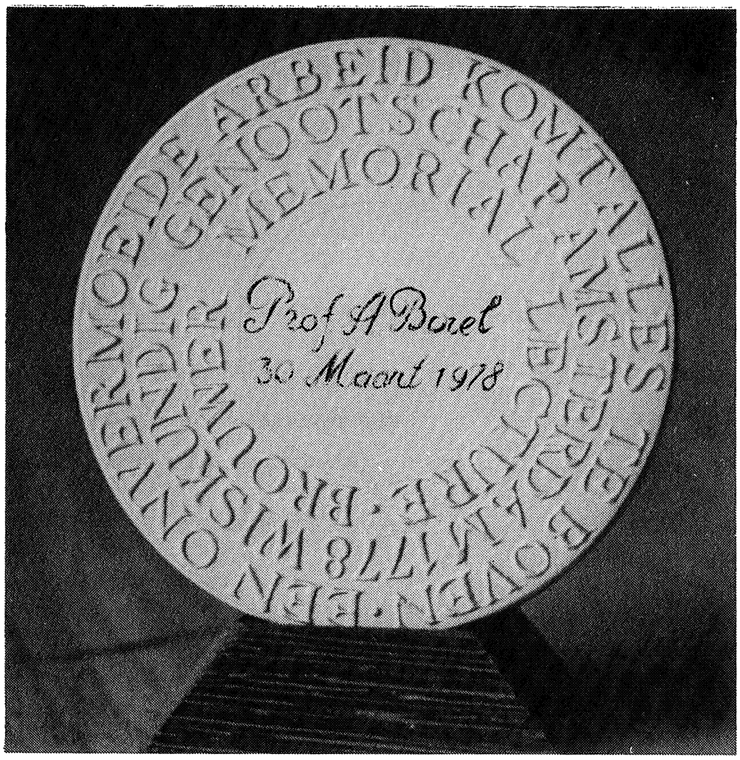
V.P. PLATONOV, *Algebraic groups,* Journal of Soviet Mathematics 4 (1975), 463-482, Plenum Publ. Corp., New York (translated from Itogi Nauk i Tekhniki 11 (1974), 5-36).

J. TITS, *On buildings and their applications;* Proc. Int. Congress Mathematicians Vancouver, 1974, Vol. 1, 209-220.

A. BOREL, *Formes automorphes et séries de Dirichlet, (d'après R.P. Langlands);* Sem. Bourbaki 1974/75, Exp. 466, 40 p., Lecture Notes in Mathematics 514 (1976), 183-222, Springer.

The Institute for Advanced Study
Princeton, N.J. 08540, U.S.A.


LUITZEN EGBERTUS JAN BROUWER
1881
1966


EEN ONVERMOEIDE ARBEID KOMT ALLES TE BOVEN
WISKUNDIG GENOOTSCHAP AMSTERDAM 1778
BROUWER MEMORIAL LECTURE
Prof A Borel
30 Maart 1978

ADDRESS TO PROFESSOR A. BOREL
at the occasion of the Brouwer Memorial Lecture
delivered by Prof.A. Borel March 30, 1978

W.T. van Est

Dear Professor Borel,

After having witnessed your brilliant lecture it is the task of the present speaker to account for the reasons that moved the board of the Wiskundig Genootschap and the Brouwer Lecture committee for selecting the special subject and candidate for the 1978 Brouwer Memorial Lecture.

Although the Brouwer Memorial Lecture has not yet a tradition of long standing, such an account seems natural and has been part of the two previous ceremonies.

Let me recall therefore that there are no specific restrictions imposed on the choice of the topic of the Brouwer Lecture, although it seems to me that it ought to be a field of mathematics of sufficient general interest. Furthermore, the selected speaker should be a mathematician who has made outstanding contributions to the subject involved.

Quite understandably, the first two Brouwer Lectures were in, or at least closely related to, fields in which Brouwer had done paramount work himself, to wit Topology and Foundations of Mathematics. It might seem therefore that in choosing the subject of the 1978 Lecture the board of the Wiskundig Genootschap was again led by a preference for the fields in which Brouwer has worked. It is true that Brouwer did interesting work in the field of Lie groups by solving a special case of Hilbert's fifth problem. However, more relevant reasons for the choice made can be given. I rather think that the choice of Lie groups, algebraic and arithmetic groups, is fully jusitified because of the general interest of the subject in the sense that there is hardly any other field of mathematics in which so many mathematical disciplines meet.

Let us try to bring out again more specifically this cross-road position of the subject of Lie groups, although your lecture has already made this plain enough. *)

The classical work of Sophus Lie, Felix Klein and Elie Cartan has revealed the connections between Lie groups and the classical geometries such as projective geometry, Euclidean geometry, Möbius geometry, etc. on the one hand and differential geometry on the other hand. Klein's Erlanger Programm conceived the classical geometries as a theory of invariant properties relative a suitable group. In Cartan's conception the study of a particular type of differential geometry was the study of principal fibre bundles with structural group from a given class of Lie groups, thus extending the scope of the Erlanger Programm. Hence, via the principal fibre bundles the theory of characteristic classes, a topic which is at the heart of algebraic topology, is directly connected with Lie groups and their topology.

Analysis makes its contacts with Lie groups in various ways. The theory of invariant differential operators on Lie groups constitutes an interesting chapter of the general theory of partial differential operators. The theory of orthogonal polynomials and special functions is connected with Lie group theory via representation theory. Speaking about representation theory, Fourier analysis, automorphic functions and automorphic forms, subjects well rooted in 19th century analysis, come to mind. In the theory of automorphic functions and forms one has to deal with double coset spaces of Lie groups with respect to a maximal compact subgroup and a discrete subgroup. The subject of discrete subgroups is connected in various ways with algebraic number theory.

These few indications to which could have been added more, may suffice to support the general contention about the cross-road position of Lie group theory. In fact Lie group theory and its ramifications is, it seems, an expanding universe, expanding to the past by constantly encompassing classical subjects and expanding to the future in bringing up new topics of mathematical research.

Your work, Professor Borel, reflects faithfully the many-sidedness of the subject. The papers you have written in collaboration with mathematicians working in different fields bring out very clearly the meeting ground character that Lie group theory has for various mathematical disciplines.

---

*) This address having been prepared independently, it was unvoidable that the following account partly overlaps the preceding address.

In every corner of Lie group theory where you worked, you have made fundamental contributions.

In the topology of Lie groups and principal fibre bundles you showed how to extend the Hopf description of the cohomology ring of a group manifold for arbitrary fields as coefficient domains. On the basis of this you were able to calculate the universal characteristic ring in a number of cases. Likewise you made extensive calculations on the cohomology ring of a Lie group manifold. Your joint work with Hirzebruch on characteristic classes of homogeneous spaces is just one example of joining forces from different fields.

The theory of algebraic linear groups advanced by your work considerably beyond the Maurer-Chevalley theory and quoting Chevalley: "C'est au mémoire fondamental de Borel que la théorie des groupes linéaires algébriques sur un corps algébriquement clos de caractéristique quelconque doit l'aspect de doctrine harmonieuse et cohérente qu'elle revêt ajourd'hui.".**)

In joint work with Harish Chandra you showed how the theory of algebraic linear groups could be made to bear on the classical reduction theory of quadratic forms which is a particular instance of the problem of finding a fundamental set of a real algebraic group with respect to an arithmetic subgroup. This approach permitted to unify and to carry further results of a classical branch of research in which among others Hermite, Poincaré and Siegel had left their marks.

One of the surprising by-results of this work is the theorem on the existence of compact Clifford-Klein space forms for every type of symmetric space.

In recent years the arithmetic groups are still in the centre of your interest. The classical question of compactification of fundamental domains has been taken up in joint work with Baily and Serre. In the cohomology of arithmetic groups you obtained deep results, partly in collaboration with Serre, and your work on this topic is still continuing.

Professor Borel I have concentrated on only a few aspects of your work. I have passed over your work on Kählerian and Hermitian geometry, your work on transformation groups and fixpoint theory, your work on Chevalley groups and many special questions in algebraic groups. I have also passed over your

---

**) C.Chevalley: Proc.Int. Congr. of Math. 1958 (Edinburgh), p.65.

occasional excursions into topology (like the Smith fixpoint theorems), into algebraic geometry and complex manifolds. I should have given credit also to the many mathematicians with whom you cooperated. However, this occasion does not call for a meticulous analysis of all the facets of your work.

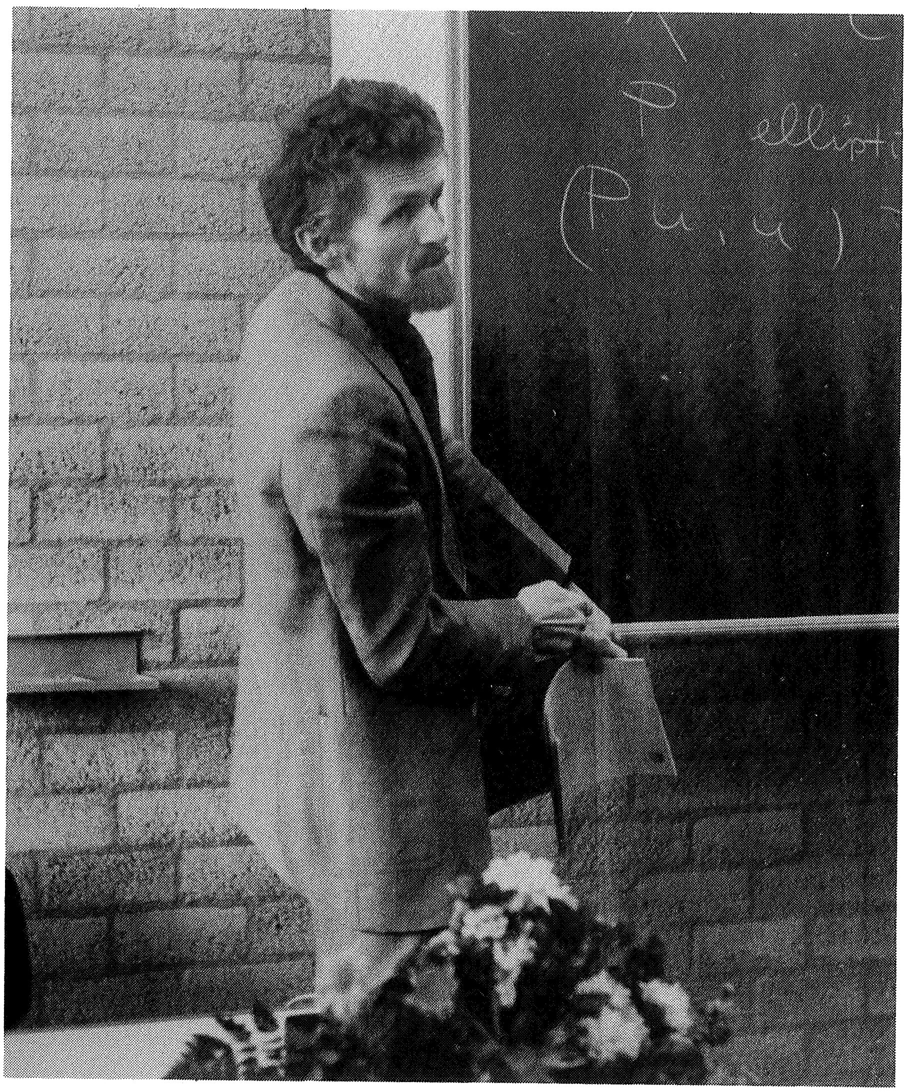
But there is one aspect of your activities that I should like to focus attention on for a moment. These are the many expository talks and papers by which you brought current or classical work of others to the attention of the mathematical public. The Bourbaki Seminar Notes, Springer Lecture Notes, Princeton Mathematical Surveys, Bulletin de la Société Mathématique de France, all bear testimony of these activities by which you rendered a great service to the mathematical community.

From this brief exposition it appears quite clearly that you touched on many fields in mathematics. Rather than to a vacillating interest of a restless mind, this is due to the nature of the problems you treated - a nature which is shared by many problems in mathematics, namely the lack of respect for the man-made artificial subdivision of mathematics into different fields, by requiring for their solution techniques and viewpoints borrowed from different mathematical disciplines. These are the built-in safeguards in mathematics against falling apart into disciplines which otherwise would go by a common name for somewhat accidental historical reasons.

The mathematicians who by their work bring out this nature of mathematics are thereby, in my opinion, influencing the development of mathematics for some time to come.

Just as the work of Brouwer has been, and still is, carrying fruit, we think, Professor Borel, that your work will continue to inspire mathematicians of new generations.

This is the account of the reasons for inviting you to deliver the Brouwer Lecture for this occasion. Consider the Brouwer medal also as a token of our high esteem for your work.



HARMONIC ANALYSIS ON SEMI-SIMPLE LIE GROUPS

J.J. Duistermaat

1. INTRODUCTION

My doorway to the subject of the title is the beautiful formula of KOLK and VARADARAJAN [13] for the spectrum of a compact quotient of a symmetric space of non-compact type. This formula fills an important gap in the collection of specific examples where one can determine in what way one might go beyond the generalities about spectra of arbitrary positive elliptic operators on compact manifolds, such as given in DUISTERMAAT and GUILLEMIN [2], for instance.

For the proof Kolk and Varadarajan use a considerable part of the harmonic analysis on semi-simple Lie groups of Harish-Chandra. Trying to understand this from a more differential geometric point of view and using an oscillatory integral trick which I learned from Guillemin at a conference in Durham in 1976, I got proofs of some the basic results which I think are much simpler than the original ones. I do not know yet how far this attempt of "rewriting Harish-Chandra" will succeed, but every result which makes this monumental work more accessible is worth trying.

In this talk I will try to explain the formula of Kolk and Varadarajan, and only at the end give some indications about the theorems of Harish-Chandra for which I believe to have new proofs.

2. INVARIANT OPERATORS ON SYMMETRIC SPACES

Perhaps the simplest description of the symmetric spaces of non-compact type is the one in the spirit of MOSTOW [15]. One starts with the identity component G of an algebraic group of linear transformations in a Euclidean space V (= vector space provided with an inner product h) such that

i) The Lie algebra $\mathfrak{g}$ of G is semi-simple, and

ii) For each $x \in G$ the adjoint x' with respect to h belongs to G.

Write $K = \{x \in G;\ x \text{ is h-orthogonal}\}$ and $P = \{x \in G;\ x \text{ is symmetric and positive with respect to } h\}$.

A lemma of CHEVALLEY [1] says that $x^t \in P$ if $x \in P$, $t \in \mathbb{R}$. It follows that for each $x \in G$ there are uniquely determined $k \in K$, $p \in P$, such that $x = k.p$, namely $p = (x'.x)^{\frac{1}{2}}$, $k = x.p^{-1}$, this is called *Cartan decomposition*. The mapping $(k,p) \mapsto k.p$ is a diffeomorphism: $K \times P \to G$, so that P can be identified with the homogeneous space $K\backslash G$, for the action of G on P one can take $p \mapsto x'.p.x$ $(x \in G)$.

The only reasonable G-invariant system L of lines in P consists of the curves $t \to x'.\exp tX.x$, here $x \in G$ and X belongs to

$$p = \{X \in \mathfrak{g};\ X \text{ is symmetric with respect to } h\}.$$

The length of this curve as t runs from 0 to 1 is defined as $(\mathrm{Tr}\ X^2)^{\frac{1}{2}}$. One easily verifies that each two points of P are joined by a unique line $\in L$, the length of which is called the distance between the two points. This distance function satisfies the triangle inequality, from which it also follows that the lines $\in L$ are the shortest curves between given points. Here the length of an arbitrary curve is defined as the integral of the length of the velocity vector, the length of tangent vectors being determined by assigning to the velocity of the above curve $\in L$ the length $(\mathrm{Tr}\ X^2)^{\frac{1}{2}}$.

The reflection $p \mapsto p^{-1}$ in the identity element is an isometry: $P \to P$. Because G acts transitively on P the reflection at any point of P, reversing the geodesics through that point, is an isometry as well, this is the defining property of *symmetric spaces* according to E. Cartan. Using this one can now give an elementary proof of the fact that each compact group of isometries in P has a fixpoint in P (see FREUDENTHAL and DE VRIES [3], 65.4), which implies that for each compact subgroup U of G there is some $p \in P$ such that $p.U.p^{-1} \subset K$. That is, all maximal compact subgroups of G are conjugate to K, for this reason K is called *the* maximal subgroup of G.

The algebra of G-invariant continuous linear operators: $C^\infty(P) \to C^\infty(P)$ is naturally isomorphic to the convolution algebra $E'(K\backslash G/K)$ of compactly supported distributions u in G which are invariant under left- and right-multiplication by elements of K, the so-called *spherical distributions* on G. The isomorphism consists of assigning to $u \in \mathcal{E}'(K\backslash G/K)$ the convolution operator $u*$ defined by $(u*f)(x) = \int_G u(x.y^{-1})\, f(y)\, dy$, which indeed assigns to each left-K-invariant function on G (element of $C^\infty(P)$) a left-K-invariant function on G. The G-invariant differential operators on P are the $u*$ such that the support of u is contained in K, the G-invariant smoothing operators(=operators: $\mathcal{D}'(P) \to C^\infty(P)$) correspond to the $u \in C_c^\infty(K\backslash G/K) = \mathcal{E}'(K\backslash G/K) \cap C^\infty(G)$.

3. THE HOROSPHERICAL TRANSFORMATION

Let $\mathfrak{a}$ be a maximal abelian sub Lie algebra of $\mathfrak{g}$ such that $\mathfrak{a} \subset \mathfrak{p}$. Then the exponential map is a Lie isomorphism from $(\mathfrak{a},+)$ to a vector subgroup A of G, in geometric terms A is a maximal flat subspace of the symmetric space P which passes through e. The dimension of A, resp. $\mathfrak{a}$ is called the *rank of the symmetric space* P, and will be denoted by r.

The mappings $Y \mapsto [X,Y]$, $X \in \mathfrak{a}$ form a commuting set of symmetric linear mappings: $\mathfrak{g} \to \mathfrak{g}$, so they can be simultaneously diagonalized with real eigenvalues. That is, there is a finite set R of linear forms $\alpha\colon \mathfrak{a} \to \mathbb{R}$, and corresponding linear subspaces $\mathfrak{g}^\alpha$ of $\mathfrak{g}$, such that

i) $[X,Y] = \alpha(X).Y$ for $X \in \mathfrak{a}$, $Y \in \mathfrak{g}^\alpha$, and
ii) $\mathfrak{g} = \sum_{\alpha\in R} \mathfrak{g}^\alpha$, $\dim \mathfrak{g}^\alpha \neq 0$ for all $\alpha \in R$.

The $\alpha \in R$ are called the *rootforms on* $\mathfrak{a}$.

Choose a hyperplane in $\mathfrak{a}^*$ containing no non-zero rootforms, call all rootforms at one side of it the *positive rootforms,* and write

$$\mathfrak{n} = \sum_{\alpha>0} \mathfrak{g}^\alpha.$$

Then $S = \exp(\mathfrak{a}+\mathfrak{n})$ is a connected solvable Lie group, and $\phi\colon (X,Y) \mapsto \exp X.\exp Y$ is a local diffeomorphism: $\mathfrak{a} \times \mathfrak{n} \to S$. S acts locally transitively on $K\backslash G$, hence transitively because $K\backslash G$ is connected. So ϕ induces a covering: $\mathfrak{a} \times \mathfrak{n} \to K\backslash G$, which is a diffeomorphism because $K\backslash G$ is simply connected. It follows that $\psi\colon (k,X,Y) \mapsto k.\exp X.\exp Y$ is a diffeomorphism: $K \times \mathfrak{a} \times \mathfrak{n} \to G$. Writing $N = \exp \mathfrak{n}$ this implies that $\chi\colon (k,a,n) \mapsto k.a.n$ is a diffeomorphism: $K \times A \times N \to G$, its inverse is called the *Iwasawa decomposition of* G.

The nilpotent group N in the Iwasawa decomposition has the following geometric characterization. The set $\mathfrak{a}^+ = \{X \in \mathfrak{a};\ \alpha(X) > 0,\ \forall\alpha > 0\}$ is called the *positive chamber in* $\mathfrak{a}$. For each $X \in \mathfrak{a}^+$ the geodesics γ in P such that the distance between $\gamma(t)$ and $\exp tX$ converges to 0 as $t \to \infty$ are just the geodesics of the form $t \mapsto n'.\exp tX.n$, $n \in N$, that is the geodesics obtained from $t \mapsto \exp tX$ by applying an element of N. For this reason N is called the *horospherical group* of the positive chamber $\mathfrak{a}^+$.

The walls of $\mathfrak{a}^+$ are just the null spaces of the positive rootforms, the group generated by the orthogonal reflections in the walls is finite and is called the *Weyl group* W *of* $\mathfrak{a}$.

For $u \in C_c^\infty(K\backslash G/K)$, $X \in \mathfrak{a}$, write

$$(Au)(X) = e^{\rho(X)} \int_N u(\exp X.n)\, dn, \text{ here } \rho(X) = \tfrac{1}{2} \sum_{\alpha>0} \dim \mathfrak{g}^\alpha.\alpha(X).$$

It is a classical observation by Gelfand that A is an injective homomorphism from the convolution algebra $C_c^\infty(K\backslash G/K)$ to the convolution algebra $C_c^\infty(\mathfrak{a})$, commuting with the reflections $(x \mapsto x^{-1})$ in G, resp. $\mathfrak{a}$. But in fact we have the much stronger theorem of GANGOLLI [4] that A is actually a topological isomorphism between $C_c^\infty(K\backslash G/K)$ and the space $C_c^\infty(\mathfrak{a})^W$ of W-invariant test functions on $\mathfrak{a}$. The inverse is given by $A^{-1}(v) = {}^tA(b*v)$, here ${}^tA$ = transposed of A and b is a tempered distribution on $\mathfrak{a}$. The Fourier transform β of b is called the *Plancherel measure for spherical functions,* it is a rational function of exponential functions explicitly determined by GINDIKIN and KARPELEVIC [7]. These results are based on previous work of HARISH-CHANDRA [9], see also ROSENBERG [16].

A extends to a topological isomorphism between the corresponding distribution spaces $E'(K\backslash G/K)$ and $E'(\mathfrak{a})^W$. Under this isomorphism the spherical distributions with support $\subset K$ correspond to $v \in E'(\mathfrak{a})^W$ with support $\subset \{0\}$, that is $v = D\delta$, here D is a W-invariant partial differential operator in $\mathfrak{a}$ with constant coefficients. Identifying the convolution algebras on both sides with the corresponding algebras of translation invariant operators, one has obtained an isomorphism between the algebra of G-invariant continuous linear operators: $C^\infty(P) \to C^\infty(P)$ and the algebra of translation- and W- invariant operators: $C^\infty(\mathfrak{a}) \to C^\infty(\mathfrak{a})$. Under this isomorphism the G-invariant differential operators on P correspond to the W-invariant differential operators on $\mathfrak{a}$ with constant coefficients, and the G-invariant smoothing operators on P correspond to the convolutions by W-invariant test functions on $\mathfrak{a}$. Because of its relation with the horospherical group N, A is called the *horospherical transformation of the group* G.

Every continuous homomorphism from the convolution algebra $C_c^\infty(\mathfrak{a})^W$ to $\mathbb{C}$ (algebra with the usual multiplication) is equal to testing by a function of the form $\varepsilon_\lambda: X \mapsto e^{\lambda(X)}$ for some $\lambda \in \mathfrak{a}_{\mathbb{C}}^*$. As a typical application of the horospherical transformation one obtains that every continuous homomorphism from the algebra of G-invariant smoothing operators on P to $\mathbb{C}$ is equal to $u* \mapsto \langle u,\phi_\lambda\rangle$, where $\phi_\lambda = {}^tA\varepsilon_\lambda$ is given by $\phi_\lambda(x) = \int_K e^{(\lambda-\rho)(\pi(x,k))}\,dk$, here $\pi: G \to \mathfrak{a}$ is defined as $\psi^{-1}: G \to K \times \mathfrak{a} \times \mathfrak{n}$ followed by the projection to $\mathfrak{a}$. The ϕ_λ are exactly the left- and right- K-invariant common eigenfunctions for all left-K- and right-G-invariant differential operators on G which moreover are equal to 1 at $e \in G$, they are called the *elementary spherical functions on* G. A theorem of HELGASON and JOHNSON [11] states that the bounded elementary spherical functions are exactly the ϕ_λ such that the real part of λ belongs to the convex hull C of the $s^*(\rho)$, $s \in W$.

4. THE SPECTRUM OF $K\backslash G/\Gamma$

Let now Γ be a discrete subgroup of G acting properly and effectively on P with compact orbit space $X = P/\Gamma = K\backslash G/\Gamma$, which then is a compact C^∞ manifold. Regarding functions on X as Γ-invariant functions on P, or as left-K- and right-Γ-invariant functions on G, any G-invariant smoothing operator on P leaves $C^\infty(X)$ invariant, so leads to an operator on X which can be written as $(T_u f)(x) = \int_X K_u(x,y) f(y)\,dy$, $f \in C^\infty(X)$, where

$$(4.1) \qquad K_u(x,y) = \sum_{\gamma\in\Gamma} u(x.\gamma.y^{-1}), \qquad u \in C_c^\infty(K\backslash G/K).$$

The effect of the summation over Γ is that K_u is left-K- and right-Γ-invariant as a function of both x and y, and therefore can be regarded as a C^∞ function on $X \times X$. Note also that for given u with compact support the sum is actually finite because the action of Γ is proper. By a continuous extension to $u \in \mathcal{E}'(K\backslash G/K)$ a similar characterization holds for all continuous linear operators: $C^\infty(X) \to C^\infty(X)$ coming from G-invariant operators on P, this time $K_u \in \mathcal{E}'(X\times X)$.

Because the algebra $\mathfrak{a}$ of operators on X coming from G-invariant operators on P is commutative, invariant under taking adjoints and contains compact operators approximating the identity (namely the smoothing ones), it follows that there is a Hilbert space decomposition $L^2(X) = \sum_{j\in\mathbb{N}} E_j$, such that $n_j = \dim E_j < \infty$ and $T \mid E_j = \mu_j(T).I \mid E_j$ for all $T \in \mathfrak{a}$. But then $u \mapsto \mu_j(T_u)$ is a continuous homomorphism: $(\mathcal{E}'(K\backslash G/K),*) \mapsto (\mathbb{C},.)$ and therefore of the form $\mu_j(T_u) = \langle Au, \varepsilon_{\lambda_j}\rangle$ for some $\lambda_j \in \mathfrak{a}^*_{\mathbb{C}}$, uniquely determined up to the action of the Weyl group. The $\lambda_j \in \mathfrak{a}^*_{\mathbb{C}}/W$, together with their multiplicities n_j, are called the *spectrum of* $K\backslash G/\Gamma$, because they represent the common spectrum of *all* operators on $K\backslash G/\Gamma$ coming from G-invariant operators on $K\backslash G$.

PROPERTIES OF S

i) S is locally finite (if $v \in C^\infty_c(\mathfrak{a})^W$ then $v = Au$ and T_u is a compact operator, hence $\langle v, \varepsilon_{\lambda_j}\rangle \to 0$ as $j \to \infty$).

ii) If $\lambda \in S$ then there exists $s \in W$ such that $-\bar{\lambda} = s^*(\lambda)$.
$(\overline{\mu_j(T_u)} = \mu_j((T_u)^*) = \mu_j(T_{u^*}) = \langle Au^*, \varepsilon_{\lambda_j}\rangle = \overline{\langle Au, \varepsilon_{-\bar{\lambda}_j}\rangle}$, here $u^*(x) = \overline{u(x^{-1})}$.)

iii) If $\lambda \in S$ then $\mathrm{Re}\lambda \in C$=convex hull of the $s^*(\lambda)$, $s \in W$.
(If $v \in E_j$, $v \neq 0$, then $\phi: x \mapsto \langle(\lambda_x)^* v, v\rangle_{L^2(G/\Gamma)} \in C^\infty(K\backslash G/K)$ and is bounded because $(\lambda_x)^*$ is orthogonal in $L^2(G/\Gamma)$, here λ_x = mulplication from th left by x. On the other hand $T_u(\phi) = \mu_j(T_u).\phi$ for all $u \in \mathcal{E}'(K\backslash G/K)$, so $\phi/\langle v,v\rangle$ is an elementary spherical function, which must be ϕ_{λ_j}. So ϕ_{λ_j} is bounded. Apply the theorem of Helgason and Johnson.)

From ii), iii) it follows that S is contained in the union of $i\mathfrak{a}^* = \{\lambda \in \mathfrak{a}^*_{\mathbb{C}};\ -\bar{\lambda} = \lambda\}$ and the"fins" $F_s = \{\lambda \in \mathfrak{a}^*_{\mathbb{C}};\ -\bar{\lambda} = s^*(\lambda),\ \mathrm{Re}\,\lambda \in C\}$, $s \in W$, $s \neq e$. The F_s are strips in real r-dimensional linear subspaces of the complex r-dimensional (= real 2r-dimensional) vector space $\mathfrak{a}^*_{\mathbb{C}}$, with bounded distance to $i\mathfrak{a}^*$. $S \cap i\mathfrak{a}^*$, resp. $S\backslash i\mathfrak{a}^*$ is called the *principal,* resp. *complementary spectrum,* using i) it follows that the complementary spectrum is finite if $r = 1$, but for $r > 1$ it is even unknown whether the complementary spectrum is ever finite. Note that always $\rho \in S$, corresponding to the constant function which obviously is a common eigenfunction. So the complementary spectrum is never void, and in fact contains an extremal element of the set $\mathrm{Re}\,\lambda \in C$.

5. THE FORMULA OF KOLK AND VARADARAJAN

If X is a compact C^∞ manifold and T is an integral operator on X with C^∞ kernel $K(x,y)$, then T, regarded as an operator in $L^2(X)$, is of trace class, and its trace is equal to $\int_X K(x,x)dx$. Because on the other hand the trace is equal to the sum of the eigenvalues, this is an important tool of getting more explicit information about spectra. Applying this to $X = K\backslash G/\Gamma$, $T = T_u$, and using (4.1), this leads to

$$\sum_j n_j . \mu_j(T_u) = \sum_{[\gamma]} \operatorname{vol}(G_\gamma/\Gamma_\gamma) . \int_{G/G_\gamma} u(x.\gamma.x^{-1})dx,$$

for all $u \in C_c^\infty(K\backslash G/K)$. Here G_γ, resp. Γ_γ is the centralizer of γ in G, resp. Γ and $[\gamma]$ denotes the conjugacy class of γ in Γ. The right hand side is obtained from $\int_X \sum_{\gamma\in\Gamma} u(x.\gamma.x^{-1})dx$, which is equal to the sum over the Γ-conjugacy classes δ of $\int_{G/\Gamma} \sum_{\gamma\in\delta} u(x.\gamma.x^{-1})dx$. So the sum of the eigenvalues is expressed as a sum of completely different terms, namely the integrals over the conjugacy classes in G of the elements of Γ, this is the famous *Selberg trace formula* [17].

The very simple idea behind the formula of Kolk and Varadarajan is to write $u = A^{-1}v$, $v \in C_c^\infty(\mathfrak{a})^W$, and consider the summands in both the left- and right-hand side of the Selberg trace formula as functions of v, that is as W-invariant distributions in $\mathfrak{a}$:

$$(5.1) \qquad \sum_j n_j . \varepsilon_{\lambda_j} = \sum_{[\gamma]} \tau_\gamma,$$

here the W-invariant distribution τ_γ in $\mathfrak{a}$ is defined by

$$(5.2) \qquad \langle v, \tau_\gamma \rangle = \operatorname{vol}(G_\gamma/\Gamma_\gamma) . \int_{G/G_\gamma} (A^{-1}v)(x.\gamma.x^{-1})dx.$$

The sum in the left hand side of (5.1) converges only in the distribution sense (the summands being smooth functions), whereas the sum in the right hand side is locally finite. The real work now consists of giving a more explicit description of the distributions τ_γ, and then using the formula (5.1) to obtain as much information as possible about both S and Γ: information about Γ leads to information about S and vice versa, one could even hope that an iteration would lead to more and more refined results.

The only a priori known element of Γ is $\gamma = e$, in this case

$<v,\tau_e> = \mathrm{vol}(G/\Gamma).\ (A^{-1}v)(e) = \mathrm{vol}(G/\Gamma).\ {}^tA(b*v)(e) = \mathrm{vol}(G/\Gamma).<v,b>$, so

(5.3) $\quad \tau_e = \mathrm{vol}(G/\Gamma).b.$

Using that $0 \notin \mathrm{supp}(\tau_\gamma)$ if $\gamma \neq e$, this leads to the following asymptotic result about S. Regard the Fourier transform β of b as a function on $i\mathfrak{a}^*$ (Plancherel measure). Then the number of elements of the principal spectrum (counted with their multiplicities) in big sets far away is asymptotically equal to the measure of such sets with respect to β, whereas the number of elements of the complementary spectrum in these sets is comparatively small. For the detailed estimates, see KOLK [13]. This results proves an old conjecture of GELFAND [6], who disregarded the complementary spectrum. Noting that the eigenvalues of a differential operator on $K\backslash G/\Gamma$ coming from a G-invariant differential operator on $K\backslash G$ are equal to the values of a polynomial p in the points $\lambda \in S$ (p being the Fourier transform of some Au with support $\subset \{0\}$), one obtains asymptotic expansions for the spectrum of such differential operators. These can be compared with the known results for the positive elliptic operators among them. If $r = 1$, the algebra of G-invariant differential operators on $K\backslash G$ is generated by only one element, the Laplace operator, which happens to be positive elliptic, but if $r > 1$ the study of the Laplace operator only gives partial information about S.

As we shall see in the next section, if $r = 1$ then for each $\gamma \neq e$, τ_γ is equal to a real multiple of the Dirac measure (!) at a point $\gamma_{\mathfrak{a}} \in \mathfrak{a}$, called the $\mathfrak{a}$-part of γ. The occurence of these measures has been observed in the case of $G = SL(2,\mathbb{R})$ by LAX and PHILIPS [14]. For $r > 1$ the τ_γ are much more complicated generally.

The only a priori known element of S is ρ, the asymptotic behaviour of ε_ρ at infinity dominates that of the other ε_λ $\lambda \in S$. This can be used to obtain asymptotic estimates for Γ, for instance if $r = 1$ one recovers the results of HUBER [12] and GANGOLLI [5] in a somewhat sharper form. For $r > 1$ results in this direction still await their proper formulation.

6. THE DISTRIBUTIONS τ_γ

For $x \in G$ denote by $\mathrm{Ad}(x)$ the tangent mapping of $y \mapsto x.y.x^{-1}$ at $y = e$, this is a linear mapping: $\mathfrak{g} \to \mathfrak{g}$. The element x is called *regular* if the dimension of the null space of $I - \mathrm{Ad}(x)$, which is equal to the codimension of the conjugacy class of x, is minimal. x is called *semi-simple* if Ad(x) is diagonalizable over $\mathbb{C}$.

If x is regular and semi-simple then $\mathfrak{h} = \ker(I-\mathrm{Ad}(x))$ is an abelian sub Lie algebra of $\mathfrak{g}$, called a *Cartan algebra in* $\mathfrak{g}$. $H = \{y \in G; \mathrm{Ad}(y)|\mathfrak{h} = I|\mathfrak{h}\}$ is a closed abelian (but not necessarily connected) Lie group in G containing x, it is called the *Cartan group in* G corresponding to the Cartan algebra $\mathfrak{h}$, resp. the element x. The union of all Cartan groups in G is exactly equal to the set of all semi-simple elements in G, the regular semi-simple elements form an open dense subset of G.

There are only finitely many conjugacy classes of Cartan algebras in $\mathfrak{g}$, each of which having an element $\mathfrak{h}$ such that

$$\mathfrak{h} = (\mathfrak{h}\cap\mathfrak{k}) + (\mathfrak{h}\cap\mathfrak{p}), \ \mathfrak{h} \cap \mathfrak{p} \subset \mathfrak{a},$$

here $\mathfrak{k}$ is the Lie algebra of the maximal compact subgroup K. Such $\mathfrak{h}$ is said to be *in standard position*. For the corresponding Cartan group H one has $H = (H\cap K).(H\cap P)$, note that the exponential map is a diffeomorphism: $\mathfrak{h} \cap \mathfrak{p} \to H \cap P$. So if $y \in H$ then $y = k.\exp X$, $k \in H \cap K$, $X \in \mathfrak{h} \cap \mathfrak{p}$, X is called the $\mathfrak{a}$-part $y_{\mathfrak{a}}$ of y. If γ is an arbitrary semi-simple element of G then there is a unique conjugacy class of Cartan algebra in standard position $\mathfrak{h}$, such that y, conjugate to γ, belongs to the Cartan group H of $\mathfrak{h}$, and the dimension of $\mathfrak{h} \cap \mathfrak{p}$ is minimal. This $\mathfrak{h}$ is called the *fundamental Cartan algebra of* γ *in standard position,* the element $y_{\mathfrak{a}}$ is determined up to the action of the Weyl group in $\mathfrak{a}$, and is again called the *$\mathfrak{a}$-part* $\gamma_{\mathfrak{a}}$ *of* γ.

If γ is a regular semi-simple element with Cartan algebra in standard position, then τ_γ is equal to integration over the linear variety $\gamma_{\mathfrak{a}} + (\mathfrak{h}\cap\mathfrak{p})^{\perp}$, with density at $\gamma_{\mathfrak{a}} + Y$ equal to $\mu(\gamma,Y)$, if Y runs through the orthogonal complement $(\mathfrak{h}\cap\mathfrak{p})^{\perp}$ of $\mathfrak{h} \cap \mathfrak{p}$ in $\mathfrak{a}$. $Y \mapsto \mu(\gamma,Y)$ is a rational function in exponential functions on $(\mathfrak{h}\cap\mathfrak{p})^{\perp}$ without poles, and its coefficients can in principle be explicitly determined.

There exists one conjugacy class of Cartan algebra $\mathfrak{h}$ in standard position such that $\mathfrak{h} \cap \mathfrak{p} = \mathfrak{a}$. An element having such $\mathfrak{h}$ as fundamental Cartan algebra is called *of Iwasawa type,* it is necessarily regular. If γ is of Iwasawa type then $(\mathfrak{h}\cap\mathfrak{p})^{\perp} = 0$ and it follows that τ_γ is equal to a real multiple of the Dirac measure at $\gamma_{\mathfrak{a}}$. If $r = 1$ then every $\gamma \in \Gamma\backslash\{e\}$ is of Iwasawa type and the description of the τ_γ, $\gamma \in \Gamma$ is finished.

In the general case all elements of Γ are semi-simple, but Γ might contain non-regular elements $\gamma \neq e$. It then belongs to several conjugacy classes of Cartan groups in standard position. *Let* H *be the fundamental one,* let $y \in H$ be conjugate to γ and approximate y with regular elements $y' \in H$. Then there exists a differential operator Q with constant coefficients in $\mathfrak{a}$, only depending on the Cartan algebras involved and not on the specific element γ, such that

$$(6.1) \qquad \tau_\gamma = \tau_y = \lim_{y' \to y,\ y' \text{ regular in } H} Q(\tau_{y'}).$$

It follows that the support of τ_γ again is contained in $\gamma_{\mathfrak{a}} + (\mathfrak{h} \cap \mathfrak{p})^\perp$, but the derivatives in Q transversal to $(\mathfrak{h} \cap \mathfrak{p})^\perp$ make it more singular than just integration against a smooth density in this linear variety. Additional singularities are introduced because the density function of the $\tau_{y'}$ develope singularities due to zero's in the denominators, as $y' \to y$. For more details, see KOLK [13], the complete explicit calculation of all the coefficients however is a formidable work still to be done.

Applying the calculations to $\gamma = e$ one obtains a formula for b, or equivalently for the Plancherel measure β, in a way different from GINDIKIN and KARPELEVIC [7]. On the other hand the general case can be reduced to the case $\gamma = e$ by a passage to the centralizer of γ.

7. BACKGROUND OF HARISH-CHANDRA THEORY

The description of the $\tau_\gamma, \gamma \in \Gamma$ in the previous section is based on the theory of HARISH-CHANDRA [10] of the transformation $f \mapsto F_f^H$ defined by:

$$F_f^H(h) = \Delta(h) \cdot \int_{G/H} f(x.h.x^{-1})dx.$$

Here H is a Cartan group, h a regular element in H, $f \in C_c^\infty(G)$, and Δ a smooth function on the set H' of regular elements in H, which can be chosen in such a way that

a) For every left-and right-invariant differential operator D on G there is an invariant differential operator D_H on H such that $F_{Df}^H = D_H(F_f^H)$ for all $f \in C_c^\infty(G)$.
b) If H is of Iwasawa type and $f \in C_c^\infty(K\backslash G/K)$ then $F_f^H \circ \exp \mid \mathfrak{a}$ is equal to some non-zero factor times $\mathcal{A}f$ ($\mathcal{A}$ = horospherical transformation).

Usually the proof of a) is given by passing to a complex group $G_{\mathbb{C}}$ where the formula is proved by reduction to the maximal compact subgroup of $G_{\mathbb{C}}$ and then using Weyl's character formula for compact groups. I think that an even more transparant proof is obtained by observing that in a complex group every Cartan group is of Iwasawa type. Then apply b) to the horospherical transformation in $G_{\mathbb{C}}$ and use that the horospherical transformation enjoys a property like a) (even better, it is a homomorphism for arbitrary convolutions).

Property a) can be used to show that F_f^H is smooth up to the boundary of H' in H, see VARADARAJAN [18] for a complete proof. For certain boundary points y of H' (namely those for which the centralizer $\mathfrak{z}(y)$ of y in $\mathfrak{g}$ satisfies $\mathcal{D}_{\mathfrak{z}}(y) \cong \mathfrak{sl}(2,\mathbb{R})$ and $\mathfrak{h} \cap \mathcal{D}_{\mathfrak{z}}(y)$ consists of elliptic elements) there is another Cartan group $\underline{H}$ containing y, called the *adjacent Cartan group,* such that

c) $$\lim_{H^+ \ni y^+ \to y} F_f^H(y^+) - \lim_{H^- \ni y \to y} F_f(y^-) = c. \lim_{\underline{H} \ni y \to y} F_f^{\underline{H}}(\underline{y})$$

for some constant $c \neq 0$. Here H^+, H^- denote the two pieces of H' near the boundary point y. The formula c) is called the *jump relation,* it is proved by a reduction to $\mathfrak{sl}(2,\mathbb{R})$, where it expresses the fact that the conjugacy class of $\pm\, \varepsilon\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}$ converges to the conjugacy class of $\pm \begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}$ as $\varepsilon \searrow 0$, whereas the union of these conjugacy classes is equal to the limit of the conjugacy class of $\varepsilon\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$ as $\varepsilon \to 0$.

The final basic result about the F_f^H which is needed is:

d) If H is the fundamental Cartan group of e then there is a differential operator $\tilde{\omega}$ in H such that

$$f(e) = \lim_{H' \ni x \to e} (\tilde{\omega}F_f^H(x) \quad \text{for all } f \in C_c^\infty(G).$$

The proof of HARISH-CHANDRA [10] (see also VARADARAJAN [18]) is long and deep, a much simpler proof using oscillatory integrals has been given for complex groups in GUILLEMIN and STERNBERG [8]. However, using the jump relations, a modification of their proof works also in the general case where more than one conjugacy class of Cartan group may occur. In fact the main problem is then to explain why only the fundamental Cartan group occurs in the formula for f(e).

The computation of the τ_γ for regular and semi-simple γ can be reduced to the problem of determining F_f^H in terms of $\mathcal{A}f$ in the case that $f \in C_c^\infty(K\backslash G/K)$ and $H \subset K$. After the reduction the group G is different from the original one (not every semi-simple Lie group has compact Cartan groups), it has among others the effect that the subspace $(h \cap p)^\perp$ of $\mathfrak{a}$ is replaced by $\mathfrak{a}$. See KOLK [13]. The regular elements in the compact Cartan group H are called *regular elliptic*.

If γ is a regular elliptic element then the restriction π_γ to the conjugacy class of γ of the projection $\pi: G \to \mathfrak{a}$ is a fibration, and a trivial one because $\mathfrak{a}$ is simply connected. This remarkable fact about regular elliptic elements does not seem to have got much attention in the literature. Anyway, because $\mathcal{A}f$ is obtained by integration over the fibers of π and $F_f^H(\chi)$ is obtained by integration over the fibers of π_γ followed by integration over $\mathfrak{a}$ with respect to some smooth density in $\mathfrak{a}$, it follows readily that $F_f^H(\gamma) = \langle \mathcal{A}f, \mu_\gamma \rangle$ for some smooth density μ_γ in . Replacing f by Df and using a), and $\mathcal{A}(Df) = D_\mathfrak{a}(\mathcal{A}f)$ for some differential operator $D_\mathfrak{a}$ on $\mathfrak{a}$ with constant coefficients, one obtains the differential equations $D_H\, \mu_\gamma(X) = {}^tD_\mathfrak{a}\, \mu_\gamma(X)$ for all bi-invariant differential operators D on G. Here D_H, resp. ${}^tD_\mathfrak{a}$ acts on the variable $\gamma \in H$, resp. $X \in \mathfrak{a}$. It follows that the Fourier transform of μ_γ is, as a function of γ, a common eigenfunction of all the operators D_H. Adding the relations at the boundary of H' in H obtained by using the jump relations c) then leads to a determination of the μ_γ.

For non-regular semi-simple elements γ a reduction to the centralizer of γ combined with an application of d) (in its centralizer γ can be translated to the identity element without harm) leads to (6.1), the operator Q being something like the $\tilde{\omega}$ of the centralizer. These comments should at least give a vague idea of the tools which are used in the determination of the τ_γ.

REFERENCES

[1] CHEVALLEY, C., *Theory of Lie Groups,* Princeton University Press 1946.

[2] DUISTERMAAT, J.J. & V.W. GUILLEMIN, *The spectrum of positive elliptic operators and periodic bicharacteristics,* Invent. Math. 29 (1975), 39-79.

[3] FREUDENTHAL, H. & H. DE VRIES, *Linear Lie Groups,* Academic Press 1969.

[4] GANGOLLI, R., *On the Plancherel formula and the Paley-Wiener theorem for spherical functions on semi-simple Lie groups,* Annals of Math. 93 (1971), 150-165.

[5] GANGOLLI, R., *On the length spectra of some compact manifolds of negative curvature,* to appear in Journ. of Diff. Geom..

[6] GELFAND, I.M., *Automorphic functions and the theory of representations,* Proc. Int. Congr. of Math., p. 74-85, Stockholm 1962.

[7] GINDIKIN, S.G. & F.I. KARPELEVIC, *The Plancherel measure for Riemannian symmetric spaces of nonpositive curvature,* Dokl. Akad. Nauk S.S.S.R. 145 (1962), 252-255 = Soviet Math. Dokl. 3 (1962), 962-965.

[8] GUILLEMIN, V.W. & S. STERNBERG, *Geometric Asymptotics,* A.M.S. Math. Surveys No. 14, 1977.

[9] HARISH-CHANDRA, *Spherical functions on semi-simple Lie groups II,* Amer. Journ. of Math. 80 (1958), 553-613.

[10] HARISH-CHANDRA, *Harmonic analysis on real reductive groups, I: the theory of the constant term,* Journ. of Funct. An. 19 (1975), 104-204.

[11] HELGASON, S. & K. JOHNSON, *The bounded spherical functions on symmetric spaces,* Adv. in Math., 3 (1969), 586-593.

[12] HUBER, H., *Zur analytischen Theorie hyperbolischer Raumformen und Bewegungsgruppen I,* Math. Ann. 138 (1959), 1-26.

[13] KOLK, J.A.C., *The Selberg trace formula and asymptotic behaviour of spectra,* Thesis, Utrecht 1977.

[14] LAX, P.D. & R.S. PHILLIPS, *Scattering Theory for Automorphic Functions,* Ann. of Math. Studies 87, Princeton 1976.

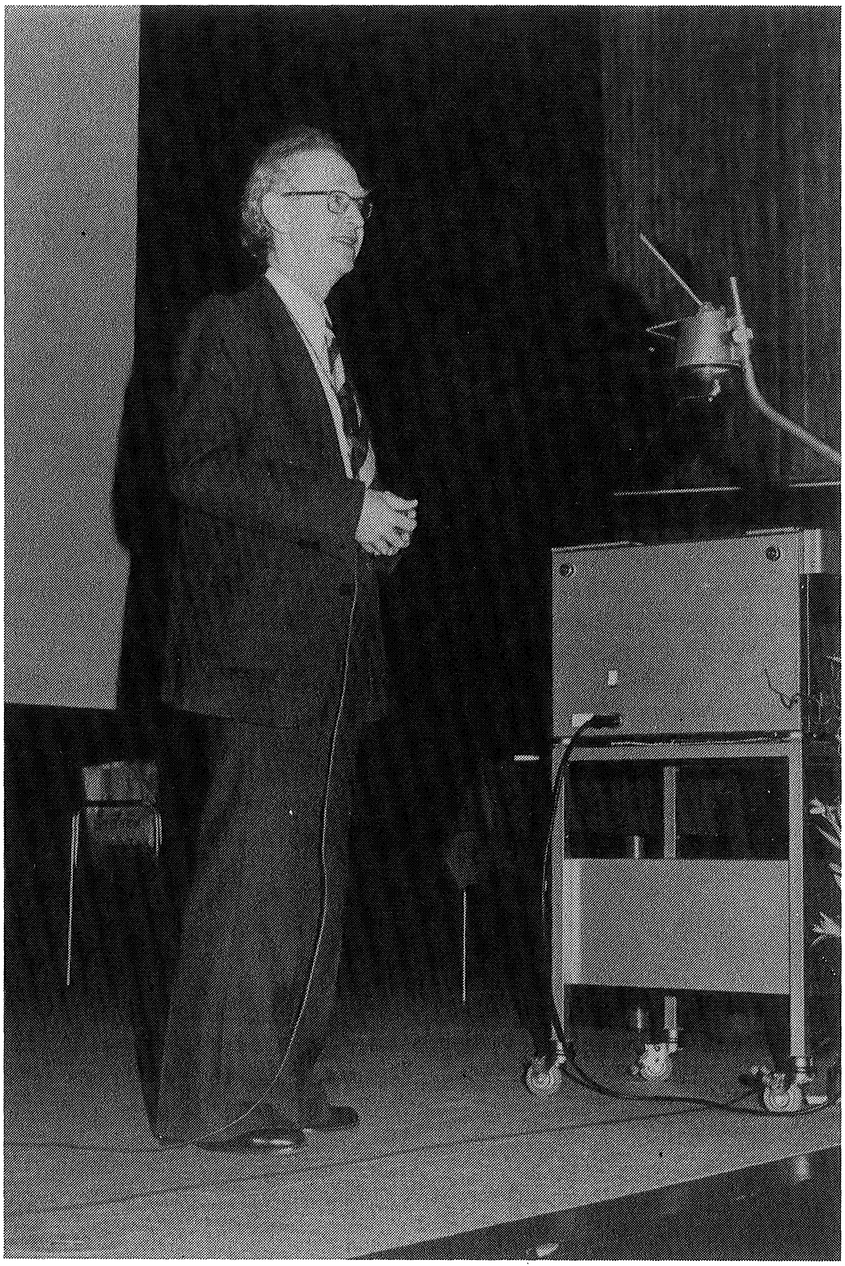
[15] MOSTOW, G.D., *Strong Rigidity of Locally Symmetric Spaces,* Ann. of Math. Studies 78, Princeton 1973.

[16] ROSENBERG, J., *A quick proof of Harish-Chandra's Plancherel theorem for spherical functions on a semi-simple Lie group,* to appear in Proc. A.M.S.

[17] SELBERG, A., *Discontinuous groups and harmonic analysis,* Proc. Int. Congr. of Math., p. 177-189, Stockholm 1962.

[18] VARADARAJAN, V.S., *Harmonic Analysis on Real Reductive Groups*, Springer Lecture Notes in Math. No. 576, 1977.

Utrecht, December 30, 1977.



A SHORT HISTORY OF TRIANGULATION AND RELATED MATTERS

N.H. Kuiper *)

1. TRIANGULATION IN THE WORK OF L.E.J. BROUWER

Real understanding in mathematics means an intuitive simple grasp of a fact. Therefore the urge to understand will seek satisfaction in simplicity of stated theorems, simplicity of methods and proofs, and simplicity of tools. It is this simplicity which can give rise to a sensation of beauty that goes with real understanding. This does not exclude admiration for a proof that is difficult by necessity.

Thus the specific interest of a geometrically-minded mathematician, who deals with figures like curves, surfaces, with structures like metric, group, and with relations like embedding, map, is influenced by this simplicity as well as by the success of methods and tools. Emphasis on existing tools sometimes leads to unnecessary overgrowth. As a consequene the historical development of mathematics is irregular like that of other forms of life and creation. We can see this in the stream of developing mathematics, at the origin of which Brouwer's work on manifolds, related to triangulation, has a prominent place.

POINCARÉ (1895) [1] developed the *analysis situs* (the origin of algebraic topology) of algebraic manifolds V. He showed by examples that the BETTI (1871) numbers do not suffice for a complete topological classification. He defined Betti groups with the help of a division of V into embedded images of convex polyhedra. Aiming at a complete classification of objects like algebraic varieties, by fitting together simple building stones one was led to take as standard parts the embedded images of straight k-simplices of various dimensions $k = 0,1,2,\ldots$, in number space $\mathbb{R}^N$. Any two ought to fit together in a simple way, namely by meeting, if at all

*) The author acknowledges with gratitude the hospitality at the University of Warwick, where he prepared part of this survey in July 1977.

1) Given years refer to dates of publication.

in one common subsimplex. The two parametrizations by barycentric coordinates with respect to the common vertices ought to be the same also. The division of V into such simplices is called a triangulation τ. It is well-defined in so far as it consists of objects, namely simplices, whose only property is thus the dimension k, and with as relations only the incidence at certain vertices between simplices. The division can therefore be described by a "scheme" T consisting of the finite or countable set of vertices, together with the set of those finite subsets that carry a simplex. Nowadays we define a *topological* ($= C^0$) *triangulation* as a homeomorphism $\tau: |T| \to V$ of a simplicial complex $|T|$, the "geometric realisation" of a finite or countable scheme T (realized say in $\mathbb{R}^N$, and consisting of affine simplices), onto a topological space V.

If T and T' are "schemes" and $h: |T'| \to |T|$ is a homeomorphism which sends every simplex of $|T'|$ linearly into a simplex of $|T|$, so that every vertex of T is an image of one vertex of T', and if $\tau: |T| \to V$ is a triangulation, then the triangulation $\tau \circ h: |T'| \to V$ is called a *subdivision* of τ.

The study of the topology of a real algebraic variety or manifold V aims first of all at the definition of invariants of the underlying topological space top(V), and their calculation. At the beginning top(V) was considered too slippery to deal with. Therefore it was replaced by the triangulation $\tau: |T| \to V$, or rather the "scheme" T. The *dimension* of T is n, if n+1 is the maximal number of vertices of simplices of T. Also the *Euler-Poincaré number* is defined in terms of T, and so are the *Betti numbers* from the incidence matrices. But are all properties that are invariant under subdivision of a triangulation topological properties of V? They would be if the following crucial problems had a positive solution.

The triangulation problem. Is there a triangulation for every algebraic variety? for every algebraic manifold? for every topological metrizable manifold?

The Hauptvermutung. This is the affirmation of the following question. Call two triangulations $\tau_1: |T_1| \to V$ and $\tau_2: |T_2| \to V$ TRI-*equivalent* in case there are subdivisions $h_1: |T_1'| \to |T_1|$ and $h_2: |T_2'| \to |T_2|$ for two realisations $|T_1'|$ and $|T_2'|$ of one and the same "scheme" $T = T_1' = T_2'$ *).

*) In DEHN-HEEGARD (1907), the word homeomorph was still used as a synonym of TRI-equivalent between finite simplical or convex-polyhedral complexes. Compare also the definition of pseudo-manifold in SEIFERT-TRELFALL (1934).

Are any two triangulations of a given V TRI-equivalent? (Observe that the composition of homeomorphisms

$$|T_1'| \xrightarrow{h_1} |T_1| \xrightarrow{\tau_1} V \xrightarrow{\tau_2^{-1}} |T_2| \xrightarrow{h_2^{-1}} |T_2'|$$

is only required to be a homeomorphism.)

For many years people wrote inconclusive papers on these two problems.

In 1911 (C,D) two papers of BROUWER on topology appeared, both outstanding in this century. In the first, only five pages long, he proves

The invariance of dimension: If h is a homeomorphism (1-1 continuous map) of an open set $U \subset \mathbb{R}^n$ onto an open set $h(U) \subset \mathbb{R}^m$, then $m = n$.
Brouwer's revolutionary idea and method was to approximate a continuous map f of an n-cube $D \subset \mathbb{R}^n$ (in the case at hand $f = h$), into $\mathbb{R}^m$ by a simplicial (= piecewise linear = PL) map g: i.e. a map linear on each simplex of a triangulation of D by linear simplices.

In his key lemma, the cube has sides of length one, m equals n, $\mathbb{R}^m = \mathbb{R}^n$, and f moves every point of D over a distance at most $d < \frac{1}{2}$. If g is ε-near to f for small $\varepsilon > 0$, then the image of g covers completely a concentric cube D' with sides of length $1-2d-2\varepsilon > 0$, *because*, as he shows, the "*Brouwer degree*", that is the algebraic number of oriented n-simplices covering an image point in D', is almost everywhere *one*. Therefore also the image of f covers such concentric discs D'. A simple argument completes the proof of the topological invariance of dimension.

In the second paper Brouwer defines a closed n-manifold as a topological space V *with* (in our terminology) a finite triangulation $\tau: |T| \to V$, of dimension n, whose simplices at a common vertex meet "like the linear simplices of a star in $\mathbb{R}^n$". This is now called a *Brouwer-triangulated manifold*. He proceeds with the *method of* PL-approximation and defines the *degree* of a continuous map $f: M \to M'$ between closed orientable Brouwer triangulated n-manifolds. Then he proves the *invariance* of the degree *under homotopy* of f, *as well as* the invariance under any *modification of the Brouwer-triangulations* of the topological spaces underlying M and M'. This means that the degree is an invariant of a homotopy class of maps between Brouwer triangulable closed oriented n-manifolds. He applies degree theory to obtain the *Brouwer fixed point theorem*.

The notions and tools in this work were new. The papers are clear now, but they were found hard to understand at the time. Their influence became clear and effective only several years later*). They were fundamental for later algebraic theories of topology. Brouwer assumed triangulations in his definitions of manifolds and he used them in an exemplary way to obtain purely topological results. He must have liked his definition of manifold to be rather constructive**). He also must have been aware of the difficulty of the triangulation problem.

It was only many years later that S.S. CAIRNS (1934) gave in two papers the first proof that a smooth n-manifold (embedded in $\mathbb{R}^N$, respectively abstractly given) has a topological Brouwer-triangulation. BROUWER (1939) presented independently a proof in a lecture for the Wiskundig Genootschap in 1937. This paper had not much impact, also because it had an unusual intuitionistic terminology. It is interesting to observe that neither Cairns nor Brouwer showed interest in C^1-triangulations nor in the Hauptvermutung. FREUDENTHAL (1939), quoting Brouwer, extended the result and gave a proof of the existence of a C^q-triangulation $\tau\colon |T| \to M$ (q-times continuously differentiable on each simplex of $|T|$) for a C^q-manifold M, $q \geq 1$. J.H.C. WHITEHEAD (1940), went further and completed the work by proving uniqueness as well, obtaining the TRI-equivalence of any two C^q-triangulations of M, $q \geq 1$. So he got a kind of smooth Hauptvermutung theorem for smooth manifolds. All TRI-triangulations obtained here were Brouwer triangulations. We denote the class of C^1-equivalence classes of C^1-manifolds by C^1, and the class of TRI-equivalence classes of Brouwer triangulated manifolds (respectively simplical complexes) by PL (respectively TRI). Then the essence of the above theorems is expressed by the existence of a natural map concerning *manifolds*:

(1) $$C^1 \longrightarrow \mathrm{PL} \subset \mathrm{TRI}$$

CAIRNS (1940a) discovered non-Brouwer triangulations of $\mathbb{R}^n$ for $n \geq 3$, that admit Brouwer subdivisions. He (1940b) also proposed the smoothing problem for Brouwer-triangulated n-manifolds and solved it for $n \leq 3$.

This concludes our short commentary on Brouwer's papers of 1911 and 1939 concerning triangulation.

---

*) Early, in the book of H. WEYL (1913) and in the work of J.W. ALEXANDER (1915) who proved the topological invariance of the homology groups.

**) Not quite constructive because it still cannot be decided whether the double cone $\Sigma(\Sigma^3)$ of a Brouwer-triangulated homotopy 3-sphere (manifold) Σ^3 is Brouwer-triangulated, by lack of a solution of the Poincaré conjecture in dimension 3.

2. MANIFOLDS, ALGEBRAIC VARIETIES, AND THEIR TRIANGULATIONS.

We will recall various interesting theorems and facts more or less in the chronological order of their discovery. The main diagram below organizes the problems while giving their relations. Every arrow represents a map between one class of equivalence classes of spaces into another one. The main problems and discoveries concern the injectivity and the surjectivity of these maps. The conclusions often depend on dimension.

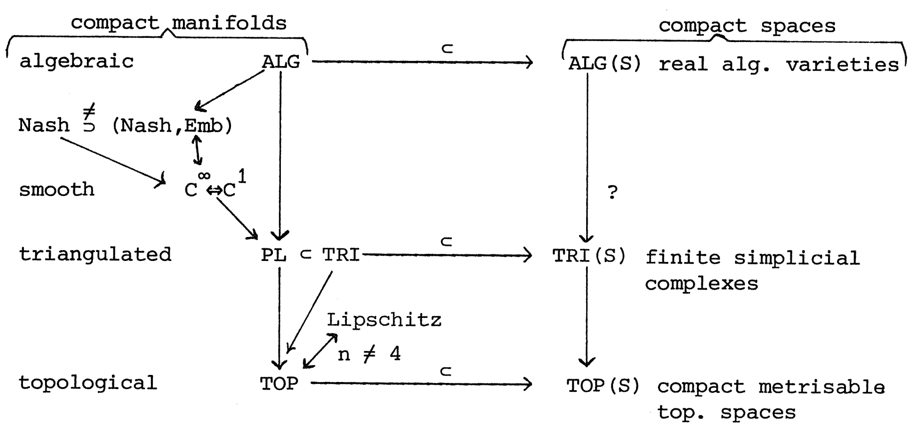
We start from the topological analysis of real algebraic varieties, because this seems, also historically, the most natural problem. It is the study of the *forget* map from equivalence classes of real algebraic varieties to their underlying topological spaces, allowing singularities (as suggested in the notation by the letter S),

$$\mathrm{ALG(S)} \longrightarrow \mathrm{TOP(S)} \tag{2}$$

that arises naturally by "forgetting" part of the structure. For manifolds, for which we delete the above letter S in our notation, this map (2) factorizes with (1) and some natural forget maps to give a *diagram on manifolds*

$$\mathrm{ALG} \longrightarrow C^\infty \longrightarrow C^1 \xrightarrow{(1)} \mathrm{PL} \longrightarrow \mathrm{TOP} \tag{3}$$

This is part of the *main diagram:*



For manifolds of small dimension, the expected existence and uniqueness of triangulations for topological manifolds was obtained for $n = 2$ by T. RADO

(1925) and for n = 3 a quarter of a century later by E. MOISE (1952):

$$n \leq 3: \text{PL} \longleftrightarrow \text{TRI} \longleftrightarrow \text{TOP bijections}$$

PAPAKYRIOKOPOULOS (1943) proved the uniqueness of the TRI-structure of simplicial complexes of dimension 2:

$$n \leq 2: \text{TRI}(S) \longrightarrow \text{TOP}(S) \quad \text{is injective.}$$

Geometric topology of combinatorial (= PL-) manifolds developed slowly. M.H.A. NEWMAN (1926) complained that it could not even be decided whether two subdivisions of a given Brouwer-triangulated manifold were TRI-equivalent. He started the foundations of "geometric topology", a topic much developed by CHR. ZEEMAN (1963). Compare HUDSON (1969) with important later work of M. Cohen in this field.

In the course of time the need for triangulations and a solution of the Hauptvermutung decreased because new homology theories of Vietoris, Cech, Alexander and the singular theory permitted purely topological definitions of invariants, although subdivisions in simplices or cells remained useful for calculating them. A milestone in algebraic topology was the axiomatic theory of EILENBERG and STEENROD (1952), which covered all older (co-) homology theories. Category and functor, notions due to S. Eilenberg and S. Mac Lane appeared as new powerful tools. Naturally algebraic topology, including the fast developing homotopy theory, dominated the field, giving a wealth of new invariants distinguishing spaces, while most people hardly dreamed of the complete classification of manifolds. The results (1) concerning the smooth triangulation of smooth (say C^1-) manifolds were isolated.

Of course manifolds existed since GRASSMAN and RIEMANN (1868), and for dimension 2 the notion developed and became "more abstract" in H. WEYL's (1913) *Idee der Riemannschen Fläche*. VEBLEN and WHITEHEAD (1932) formalized the definition of n-manifold M with structure S as follows. M is a connected metrizable topological space covered by images of embedded open $\mathbb{R}^n$-sets given by charts $h_i: U_i \to M$, that are related in their intersections $h_i(U_i) \cap h_j(U_j)$ by homeomorphisms of open sets in $\mathbb{R}^n$, $h_{ij} = h_j^{-1} \circ h_i$, belonging to some pseudo group S. In our present day applications, S can be the pseudogroup of homeomorphisms (TOP), C^1- or C^∞- or analytic diffeomorphisms, piecewise-linear homeomorphisms (PL), locally algebraic homeomorphism (Nash), Lipschitz homeomorphisms, giving rise to most of the entries in our main diagram. As differentiable manifolds, embedded in $\mathbb{R}^N$ as well as abstract, became better understood, in particular under the influence of H. Whitney, it was not difficult to obtain a C^∞-structure, unique but for equivalence on any C^1-manifold:

$$C^\infty \longleftrightarrow C^1 \quad \text{is bijective.}$$

Manifolds being "slippery" bothered mathematicians less and less. It became also clear that PL-manifolds have a Brouwer triangulation, unique up to TRI-equivalence.

J. NASH (1952) proved that every embedded (in $\mathbb{R}^N$) compact C^1- (or C^∞-) manifold can be approximated by a diffeomorphic manifold that is also a component of a real algebraic variety. He also proved that any two mutually diffeomorphic *embedded Nash-manifolds*, are related by a diffeomorphism which is algebraic, and which is locally defined by polynomial equations:

$$(\text{Nash, embedded}) \longleftrightarrow C^\infty \longleftrightarrow C^1 \quad \text{bijections}$$

There passed again a quarter of a century before A. TOGNOLI (1973) proved that every compact C^1-manifold is diffeomorphic to a manifold that is a whole real algebraic variety:

$$\text{ALG} \longrightarrow C^1 \text{ is surjective.}$$

The Veblen-Whitehead definition of manifolds gives a larger class of Nash-manifolds:

$$\text{Nash} \supsetneqq (\text{Nash, embedded}).$$

An example of a non-embeddable Nash structure on the circle is obtained by identifying points in $\mathbb{R}$ by the algebraic relation $x' = x+1$. Any function on the quotient space M yields a periodic function on $\mathbb{R}$ and cannot be algebraic unless it is constant. Hence M cannot be Nash-embedded in $\mathbb{R}^N$. It would be interesting to study all Nash structures on the circle. Perhaps all homogeneous ones admit compatible locally projective structures, as described by KUIPER (1953). The work of J. HUBBARD (1971) suggests that there may be so many non-equivalent Nash-structures that a complete classification is uninteresting. Is there more than one on the two sphere?

J. MILNOR (1956) made the sensational discovery of a manifold M, which is homeomorphic and PL-equivalent to the usual 7-sphere S^7, without being diffeomorphic to it:

$$C^1 \longrightarrow PL \quad \text{is not injective.}$$

This manifold M, a certain S^3-bundle over S^4, is homeomorphic to S^7 because it has a non-degenerate function with exactly two critical points (maximum and minimum). In order to prove M not diffeomorphic to S^7, Milnor used HIRZEBRUCH's (1956) sophisticated theory and calculation of the index of a manifold in terms of Pontrjagin numbers with THOM's (1954) cobordism theory, both powerful and fundamental tools in the further development of manifold theory.

R. THOM (1958) proposed an obstruction theory concerning the introduction of a differential structure (or smoothing) on a PL-manifold. The obstruction was to be in cohomology groups with coefficients in the group Γ_n of smoothings of the n-sphere with its usual PL-structure. As $\Gamma_n = 0$ for $n < 7$ the first obstruction turned out to be in Γ_7, a cyclic group with 28 elements. A very hard case was $\Gamma_4 = 0$, proved by J. CERF (1962). For the groups Γ_n see M. KERVAIRE and J. MILNOR (1963). The ideas of Thom were made into a solid smoothing theory by J. MUNKRES (1960, 1964) and much improved by M. Hirsch. (See M. HIRSCH and B. MAZUR (1974)). M. KERVAIRE (1960) was the first to produce effectively a PL-manifold (of dimension 10) which could not have the structure of a smooth manifold:

$$C^1 \longrightarrow PL \quad \text{is not surjective.}$$

J. EELLS and N. KUIPER (1961) and TAMURA (1961) gave simple examples in the lowest possible dimension 8. These are manifolds that can be obtained by compactifying $\mathbb{R}^8$ by an S^4, as is the case with the smooth quaternion projective plane. Although the PL-structures of the various exotic n-spheres are all the same, this does not mean that each has the same set of smooth triangulations. N. KUIPER (1965) proved that a smooth triangulation with n+1 vertices of a n-sphere exists only for the customary differential structure. A triangulation of an exotic n-sphere requires many more vertices.

We mention as a side remark that the number of vertices e_0 of a triangulation of a closed surface of Euler characteristic χ obeys

$$e_0 \geq \text{minimum}\{k \in \mathbb{Z}:\ 2k \geq 7 + \sqrt{49 - 24\chi}\}$$

and equality can arise for many surfaces, but not for the Klein-bottle ($\chi = 0$, $e_0 > 7$). Compare RINGEL (1974). For the real projective 3-space a triangulation with 11 vertices exists and this seems to be the minimal number possible. E. BRIESKORN (1966) found that a complex algebraic variety with a singularity can have the topology of a manifold in some neighborhood of that singularity. For example the set

$$\{(z_1,\ldots,z_6):\ z_1^2 + z_2^2 + z_3^2 + z_4^3 + z_5^{6k-1} = 0,\quad \sum z_j\bar{z}_j \leq 1\} \subset \mathbb{C}^5\}$$

is homeomorphic to an 8-ball, and its boundary is the seven sphere with exotic differential structure $k\cdot\gamma$, exotic if $k \not\equiv 0$ (28), where γ is the generator of Γ^7. So exotic sphers may have rather simple equations. N. KUIPER (1968) used Brieskorn's examples and "generalized" NASH (1952) to obtain algebraic equations for all non-smoothable PL-8-manifolds. AKBULUT (1976) following TOGNOLI (1973) proved that every PL-8-manifold, (as well as some other PL-manifolds of higher dimensions) can be made into a whole algebraic variety and not only a component. Akbulut and Henry King at present are making progress in obtaining algebraic equations for many more PL-manifolds.

MILNOR (1961) *disproved* the *Hauptvermutung* for simplicial complexes: the one point compactifications of $L(7,1) \times \mathbb{R}^4$ and $L(7,2) \times \mathbb{R}^4$ (concerning lens spaces $L(7,k)$, see H. SEIFERT and W. THRELFALL (1934)) are homeomorphic without being TRI-equivalent

$$\mathrm{TRI}(S) \longrightarrow \mathrm{TOP}(S) \quad \text{is not injective.}$$

The next most important phase in the study of manifolds started with the work of S. SMALE (1961) proving the Poincaré conjecture for dimensions $n \geq 5$. (For n=5 with the help of J. Stallings and Chr. Zeeman). If f is a non-degenerate C^∞-function on a compact manifold M, then for increasing values of t, the submanifold $\{x:\ f(x)\leq t\}$ changes at critical values, and these changes can be realized by attaching handles and thickening them. The Morse relations (See MILNOR (1963)) among the Betti numbers restrict the possible numbers of non-degenerate critical points of various indices on a given manifold M. Smale succeeded, for a function on a manifold M of the homotopy type of S^n in cancelling critical points (and handles) until two remained (maximum and minimum). Therefore M is seen to be homeomorphic as well as PL-equivalent to S^n ($n \geq 5$): the Poincaré conjecture, as well as the *Hauptvermutung* were proved for S^n, $n \geq 5$.

A tremendous activity in manifold theory took place between 1960 and 1970, in which the merging theories for smooth, PL- and topological manifolds developed with new tools like surgery and handlebody theory (h- and s-cobordism theory (see MILNOR (1965)), transversality, microbundles and via homotopy theory to algebraic problems, which were particularly deep and hard for non-simply-connected manifolds (see C.T.C. WALL (1970)). It will be impossible to go into much detail. I might mention S. Novikov and W. Browder as leaders. Compare the contributions on topology in the proceedings of the International Mathematical Congress in Nice, in particular the paper of L. SIEBENMANN (1970). See also the proceedings of *Manifolds Amsterdam* (1970) and R. KIRBY and L. SIEBENMANN (1977). D. SULLIVAN (1967) proved the *Hauptvermutung* for simply connected PL-manifolds of dimension ≥ 6, for which $H_3(M;\mathbb{Z})$ has no 2-torsion. R. KIRBY (1969) made the final breakthrough by proving that every orientable homeomorphism of S^n onto itself is a product of homeomorphisms, each of which is identical on some open set. This was the crucial and longstanding *stable manifold conjecture*. It carried with it the positive answer to the *annulus conjecture*. R. KIRBY and L. SIEBENMANN (1969) (see SIEBENMANN (1970) then solved the triangulation problem and the Hauptvermutung for manifolds of dimension $n \geq 5$. They deduced, using in an essential way results on homotopy tori of C.T.C. Wall and others, that there is exactly *one* well defined (by SIEBENMANN (1970) in a counter example) obstruction in $H^4(M;\pi_3(TOP/0)) = H^4(M;\mathbb{Z}_2)$ to imposing a PL-structure on a topologically closed n-manifold M^n, $n \geq 5$, and, given one PL-structure, the equivalence (= isotopy-) classes of PL-structures biject onto $H^3(M;\mathbb{Z}_2)$. So *for certain topological manifolds no Brouwer-triangulation exists, and for certain* PL*-manifolds the* PL*-Hauptvermutung is false.*

$$PL \longrightarrow TOP \quad \text{is neither injective nor surjective.}$$

It may be true still, and there is hope for the conjecture, that every topological manifold has some triangulation, which of course cannot always be a Brouwer triangulation (=PL). If true then one can hope for algebraic equations as well. R. EDWARDS (1976) constructed triangulations of S^n, $n \geq 5$, with the property that no subdivision is a Brouwer triangulation. So for manifolds:

$$n \geq 5,\ PL \subset TRI \quad \text{is not bijective.}$$

He uses B. MAZUR (1961) and V. POENARU (1960), who constructed long ago a contractible 4-manifold M with boundary ∂M that is not simply-connected, although it necessarily has the homology of S^3. Edwards proved, and this is hard, that the (n-3)-fold suspension

$$W^n = \sum^{n-3}(\partial M) = S^{n-4} * \partial M$$

(obtained by joining every point of S^{n-4} by a line segment to every point of ∂M), for $n \geq 5$, is homeomorphic to S^n. If we triangulate W then naturally $S^{n-4} \subset W$ is triangulated by a sub complex of dimension n-4. Every n-4-simplex in it has a copy of ∂M (and not of S^3) as link, which shows that the triangulation of W is not a Brouwer-triangulation.

In the spirit of our interest in the topology of real algebraic varieties, we mention a real algebraic variety (which is due to C. Gordon), which is a Mazur-Poenaru 3-manifold, that is it bounds a contractible 4-manifold:

$$\partial M = \{(z_1,z_2,z_3) \in \mathbb{C}^3 : z_1^2 + z_2^5 + z_3^7 = 0,\ \sum_{j=1}^3 z_j\bar{z}_j = 1\}$$

Siebenmann observed, and the reader can check that the double suspension of ∂M is

$$V^5 = \sum^2(\partial M) = \{(z_1,z_2,z_3,z_4) \in \mathbb{C}^4 : z_1^2 + z_2^5 + z_3^7 = 0,$$
$$\sum_{j=1}^4 z_j\bar{z}_j = 1\},$$

an algebraic variety, which is homeomorphic to S^5 by Edwards, but whose *natural triangulation is not* PL for the same reasons as above. Observe that the singular curve S^1 with equations $z_1 = z_2 = z_3 = 0$ in the topological 5-sphere V has no normal microbundle. (compare P.S. on p.13).

Let us recall here that S. LOJACIEWICZ (1964) was the first to give an accepted proof that *every real algebraic variety can be triangulated.* In order to define uniqueness one first of all has to distinguish certain triangulations of an algebraic variety to be natural, like smooth triangulations for smooth manifolds, and then to show that any two such natural triangulations are TRI-equivalent. Such a kind of uniqueness proof does not exist in the literature for $n \geq 3$, although there is some hope that it could be deduced from Lojaciewicz's work. (For $n \leq 2$: PAPAKYRIOKOPOULOS (1943).)

It should be noted that the natural singular version of the triangulation conjecture is false: SIEBENMANN (1970), § 3 gave explicit examples of compact locally triangulable spaces that are not traingulable. His example are even locally real algebraic.

Vere recently (as I learned from L. Siebenmann and R. Stern) the triangulation problem for topological manifolds has again much advanced. J. Cannon showed, generalising EDWARDS (1975), that the double suspension $\Sigma^2 W^3$ of *every* homology 3-sphere W^3 is homeomorphic to S^5. With Siebenmann's work this implies that all orientable topological 5-manifolds are triangulable, and there are many of them without any PL-structure.

D. Galewshi and R. Stern, *and independently* T. Matumoto, even define an obstruction element $\tau \in H^5(M;\rho)$ such that the topological manifold M^n, $n \geq 5$, is triangulable if and only if $\tau = 0$; and if M is triangulable there are $|H^5(M;\rho)|$ such triangulations up to "concordance".

Unfortunately, although the group ρ is well defined it is also completely unknown. Even so we can conlude that every simply connected topological 6-manifold can be triangulated. It is also known now that necessary and sufficient for triangulability of all manifolds of dimension ≥ 5 is the existence of a smooth closed homology 3-sphere (manifold) with Rohlin invariant 1 (that is, bounding a parallelizable 4-manifold of index 8) such that the connected sum H # H bounds a homology 4-disc.

Not every compact simplicial complex is homeomorphic to a real algebraic variety. Hardly anything is known about this question. D. SULLIVAN (1971) discovered that in every triangulation of a real algebraic variety the link of a vertex or simplex has even Euler characterstic. For example a double cone on the real projective plane cannot be a real algebraic variety. Compact simplicial complexes of dimension one are algebraic if and only if an even number of edges meet at every vertex. Sullivan's condition is perhaps also sufficient to decide which simplicial complexes of dimension two are algebraic. For higher dimensions the problem is completely open.

Smooth as well as PL-manifolds are Lipschitz manifolds: they can be covered with charts for which the transition functions h_{ij} (see above) obey the condition that locally

$$\frac{\|h_{ij}(x) - h_{ij}(y)\|}{\|x - y\|}$$

is bounded away from 0 and from ∞. D. SULLIVAN (1977) proved that the structure of every topological manifold can be strengthened as much: *every* closed topological *manifold* of dimension $\neq 4$ has a *Lipschitz structure*, and it is unique up to equivalence.

Every PL-structure on S^4 has a unique smoothing and visa versa, but it still remains undecided whether there are more non-equivalent PL-structures on S^4 or closed 4-manifolds in general: The Hauptvermutung and the triangulation conjecture remain open for 4-manifolds. With the Poincaré conjecture for dimensions 3 and 4, the subject of the classification of 3- and 4-manifolds is active, but the main interest in geometry and topology has shifted since 1970 to structures on manifolds like foliations, vector fields, differential equations, Riemannian metrics, functions and maps, their topology and their singularities. The topology of complex algebraic varieties remains very active too.

I mentioned that it was hard for me to do justice to all mathematicians involved in the subject. As it seems appropriate, I will go into some more detail concerning the tremendous development between 1960 and 1970. Several people helped me again to clarify points.

In this decade, 1960 and 1970, the emerging theories of smooth, PL- and topological manifolds were developed using new tools such as surgery and handle body theory (see MILNOR (1965)), transversality, microbundles, and block bundles to transfer geometric questions to homotopy theory and to algebraic questions, which were particularly deep for non-simply-connected manifolds. Following the work of KERVAIRE and MILNOR mentioned above, the powerful general theory of simply connected manifolds was developed by BROWDER and NOVIKOV, and the overall non-simply-connected theory was put into place by WALL.

In a short space one cannot describe all the outstanding contributions made by the many talented mathematicians who worked in this area. Perhaps the most significant achievement was the resolution of the Hauptvermutung and triangulation problem for manifolds. S.P. NOVIKOV contributed the first striking step when he proved the topological invariance of rational Pontrjagin classes. Together with the surgery exact sequence (the "Sullivan sequence"), this already implied the Hauptvermutung for some special cases. By developing a canonical version of Novikov's argument (with the aid of Siebenmann's thesis) LASHOF-ROTHENBERG and SULLIVAN were then able to prove the Hauptvermutung for 4-connected manifolds of dimension ≥ 6. But, CASSON and SULLIVAN (independently) had developed such penetrating (and *complete* in the case of Sullivan) analyses of the classifying space G/PL that appears in the surgery sequence that they were able to extend the proof to cover all simply connected manifolds for which $H_3(M;\mathbb{Z})$ has no 2-torsion.

The final breakthrough began when R. KIRBY showed how to reduce the stable homeomorphisms conjecture to some questions about homotopy tori. This conjecture says that every homeomorphism of $\mathbb{R}^n$ to itself is the product of homeomorphisms, each of which is the identity on some open set, and it also implies the well-known annulus conjecture. But HSIANG-SHANESON and WALL had just classified homotopy tori, and so they could easily resolve the questions of KIRBY in the affirmative.

With the same ideas plus topological immersion theory (LASHOF-ROTHENBERG (1968); LASHOF (1971)), KIRBY and SIEBENMANN (1969) and LASHOF and ROTHENBERG (1969) solved the triangulation and Hauptvermutung problems for $n \geq 5$. KIRBY and SIEBENMANN then deduced, still using the results on homotopy tori in an essential way, that there is exactly one well-defined (by SIEBENMANN (1970) in a counter-example) obstruction in $H^4(M;\pi_3(\mathrm{TOP}/0)) = H^4(M;\mathbb{Z}_2)$ to imposing a PL-structure on a closed topological n-manifold, $n \geq 5$ and, given one PL-structure, the equivalence (= isotopy) classes of PL structures biject onto $H^3(M;\mathbb{Z}_2)$. *So for certain topological manifolds no Brouwer trinagulation exists, and for certain PL-manifolds* (e.g., the torus itself) *the PL-Hauptvermutung is false.*

Institut des Hautes Etudes Scientifiques
Bures-sur-Yvette, France

September 1977

P.S. Here are equations of Siebenmann of a topological 5-manifold that can be triangulated, but has no Brouwer-triangulation or PL-structure:

$$\{(z_1,z_2,z_3,z_4): z_1^2 + z_2^5 + z_3^7 + e(\sum_{j=1}^{3} z_j\bar{z}_j)^5 = 0,\ z_4\bar{z}_4 = 1\} \subset \mathbb{C}^4.$$

REFERENCES

AKBULUT, S., *Algebraic equations for a class of PL-manifolds*. Math. Ann. 231 (1977) p. 19-31.

ALEXANDER, J.W., *A proof of the invariance of certain constants in Analysis situs,* Trans. AMS 16, (1915) p. 148-154.

BETTI, E., Annali di Matematica (2) IV (1871), p. 140-158, *Sopra gli spazi di un numero qualunque di dimensioni*.

BRIESKORN, E., *Beispiele zur Differentialtopologie von Singularitäten,* Inv. Math. 2, (1966) p. 1-14.

BROUWER, L.E.J., *Beweis der Invarianz der Dimensionenzahl,* Math. Ann. 70, (1911) p. 161-165. Complete works II, (1911 C).

BROUWER, L.E.J., *Über Abbildung von Mannigfaltigkeiten,* Math. Ann. 71, (1912), p. 97-115, Complete works II, (1911 D).

BROUWER, L.E.J., *Zum Triangulationsproblem,* Proc. KNAW 41, (1939), p. 701-706, (= Indag Math.).

CAIRNS, S.S., *On the triangulation of regular loci,* Ann. of Math. 35, (1934), p. 579-587.

CAIRNS, S.S., *Triangulation of the manifold of class one,* Bull. AMS 41, (1935), p. 549-552.

CAIRNS, S.S., *Triangulated manifolds which are not Brouwer manifolds,* Ann. of Math. 41 (1940a), p. 792-795.

CAIRNS, S.S., *Homeomorphisms between topological manifolds and analytic manifolds,* Ann. of Math. 41 (1940b), p. 796-808.

CERF, J., *La nullité de* π_0 (Diff. S^3), Séminaire Henri Cartan 1961/63 $\Gamma_4 = 0$, Springer Lecture Notes in Math. 53 (1968).

COHEN, M.M., *A proof of Newman's theorem,* Proc. Cambr. Phil. Soc. 64 (1968) p. 961-963.

DEHN, M. & P. HEEGARD, *Analysis Situs,* Enz der Math. Wiss. III AB 3, (1907), p. 154-222.

EDWARDS, R.D., *The double suspension of a certain homology 3-sphere is* S^5 (1976), Unpublished. See a paper of Giffin in Annals of Mathematics, 1978.

EELLS, J. & N.H. KUIPER, *Manifolds which are like projective planes*, Publ. Math. IHES 14, (1961) p. 5-46.

EILENBURG, S. & N. STEENROD, *Foundations of algebraic topology*, Princeton Univ. Press (1952).

FREUDENTHAL, H., *Die Triangulation der differenzierbaren Mannigfaltigkeiten*, Proc. KNAW 42, (1939), p. 880-901.

HSIANG, W.C. & J. SHANESON, *Fake tori*, in: *Topology of Manifolds*, Proceedings of the 1969 Georgia Conference, Markham Press, 1970.

HIRSCH, M. & B. MAZUR, *Smoothings of piecewise linear manifolds*, Ann. of Math. Studies 80, (1974).

HIRZEBRUCH, F., *Neue topologische Methoden in der algebraischen Geometrie (1956)*, English enlarged edition (1966), Springer Verlag.

HUBBARD, J., with D.R.J. CHILLINGWORTH, *A note on non-rigid Nash structures*, Bull. AMS 77, (1971), p.429-431; and with Douady, A., *Cohomology of Nash sheaves*, (Unpublished).

HUDSON, J.F.P., *Piecewise-linear topology* (1969), Benjamin.

KERVAIRE, M.A., *A manifold which does not admit any differentiable structure*, Comment, Math. Helv. 34, (1960), p. 257-270.

KERVAIRE, M.A. & J. MILNOR, *Groups of homotopy spheres*, Ann. of Math. (2) 77, (1963), p. 504-537.

KIRBY, R., *Stable homeomorphisms and the annulus conjecture*, Ann. of Math. 89, (1969), p. 575-582.

KIRBY, R. & L. SIEBENMANN, *Foundational essays on topological manifolds, smoothings and triangulations*, Ann. of Math. Studies, 88, (1977).

KIRBY, R. & L. SIEBENMANN, *On the triangulation of manifolds and the Hauptvermutung*, Bull. AMS 75, (1969), p. 742-749.

KUIPER, N.H., *Locally projective spaces of dimension one*, Mich. Math. J. 2, (1953), p. 95-97.

KUIPER, N.H., *On the smoothings of triangulated and combinatorial manifolds*, *Differential and combinatorial Topology*, Princeton Univ. Press (1965), p. 3-22.

KUIPER, N.H., *Algebraic equations for nonsmoothable 8-manifolds*, Publ. Math. IHES 33, (1968), p. 139-155.

LASHOF, R.K., *Lees' immersion theorem and the triangulation of manifolds,* Bull. AMS 75 (1969), p. 535-538.

LASHOF, R., *The immersion approach to triangulation and smoothing,* Proceedings of Symposia in Pure Math. XXII, Algebraic Topology A.M.S., Providence, RI, 1971.

LASHOF, R. & M. ROTHENBERG, 1. *Hauptvermutung for manifolds,* in: Proceedings of the Conference on the Topology of Manifolds, Michigan Stat e University (1967), Complementary Series in Math. 13, 81-106, Prindel, Weber and Schmidt, 1968. 2. *Triangulation of Manifolds I, II,* Bull. Amer. Math. Soc. 75 (1969), 750-757.

LOJASIEWICZ, S., *Triangulation of semi-analytic set,* Ann. Scuola Normale Sup. Pisa III, 18, (1964), p. 449-474.

MAZUR, B., *A note on some contractible 4-manifolds,* Ann. of Math. (2) 73, (1961), p. 221-228.

MILNOR, J., *On manifolds homeomorphic to the 7-sphere,* Ann. of Math. 64, (1956), p. 399-405.

MILNOR, J., *Two complexes with are homeomorphic but combinatorially distinct,* Ann. of Math., 74, (1961), p. 575-590.

MILNOR, J., *Morse theory,* Ann. of Math. Study 51, (1963), Princeton Univ. Press.

MILNOR, J., *Lectures on the h-cobordism theorem,* Notes Princeton Univ. Press, (1965).

MOISE, E., *Affine structures on 3-manifolds,* Ann. of Math. 56, (1952), p. 96-114.

MUNKRES, J., *Obstructions to the smoothing of piecewise differentiable homeomorphisms,* Ann. of Math. (2) 72, (1960), p. 521-554.

MUNKRES, J., *Obstructions to imposing differentiable structures,* III, J. of Math. 8, (1964), p. 361-376.

NASH, J., *Real algebraic manifolds,* Ann. of Math. 56, (1952), p. 405-421.

NEWMAN, M.H.A., *On the foundations of combinatorial analysis situs,* Proc. KAVW 21, (1926), p. 611-641.

PAPAKYRIOKOPOULOS, C.D., *A new proof of the invariance of the homology groups of a complex,* Bull. Soc. Math. Grèce 22, (1943), p. 1-154.

POENARU, V., *Les décompositions de l'hypercube en produit topologique,* Bull. Soc. Math. France 88, (1960), p. 113-129.

POINCARÉ, H., *Analysis situs,* Journal de l'Ecole Polytechnique p. 1-121, (1895), complete works VI.

RADO, T., *Ueber den Begriff der Riemannschen Fläche,* Acta. Litt. Scient. Univ. Szegd 2, (1925) p. 101-121.

RIEMANN, B., *Ueber die Hypothesen welche der Geometrie zu Grunde liegen,* Habilitationschrift (1854), Gesammelte Werke (1892), p. 272-287.

RINGEL, G., *Map colour theorem,* Grundlehre der Math. Wiss., Band 209 (1974), Springer Verlag.

ROURKE, C. & B. SANDERSON, *Introduction to piecewise linear topology,* Ergebnisse der Math. 69 (1972), Springer Verlag.

SEIFERT, H. & W. THRELFALL, *Lehrbuch der Topologie* (1934).

SIEBENMANN, L., *Topological manifolds,* Proc. ICM Nice (1970) 2, p. 143-163.

SMALE, S., *Generalized Poincare's conjecture in dimensions greater then four,* Ann. of Math. (2) 74, (1961), p. 391-406.

SULLIVAN, D., *On the Hauptvermutung for manifolds,* Bull. AMS 73, (1967), p. 598-600.

SULLIVAN, D., *Combinatorial invariants of analytic spaces,* Proc. Liverpool singularities symposium I, Springer Lecture Notes 192, (1971), p. 165-168.

SULLIVAN, D., *Every manifold has a Lipschitz structure,* (1977) to appear.

TAMURA, I., *8-manifolds admitting no differentiable structure,* J. Math. Soc. Japan 13, (1961), p. 377-382.

THOM, R., *Quelques propriétés globales des variétés différentiables,* Comm. Math. Helv. 28, (1954), p. 17-86.

THOM, R., *Des variétés triangulées aux variétés différentiables,* Proc. Int. Congr. Math. Edingburgh (1958), p. 248-255.

THOM, R., *Le degré brouwerien en topologie différentielle moderne,* Brouwer Memorial Lecture (1970), Nieuw Archief voor Wiskunde III 21, (1971), p. 10-16.

TOGNOLI, A., *Su una congettura di Nash,* Annali di Sc. Norm. Pisa 27, (1973), p. 176-185.

VEBLEN, D. & J.H.C. WHITEHEAD, *Foundations of differential geometry,* (1932).

WALL, C.T.C., *Surgery on compact manifolds,* Acad. Press (1970).

WEYL, H., *Die Idee der Riemannschen Fläche,* (1913).

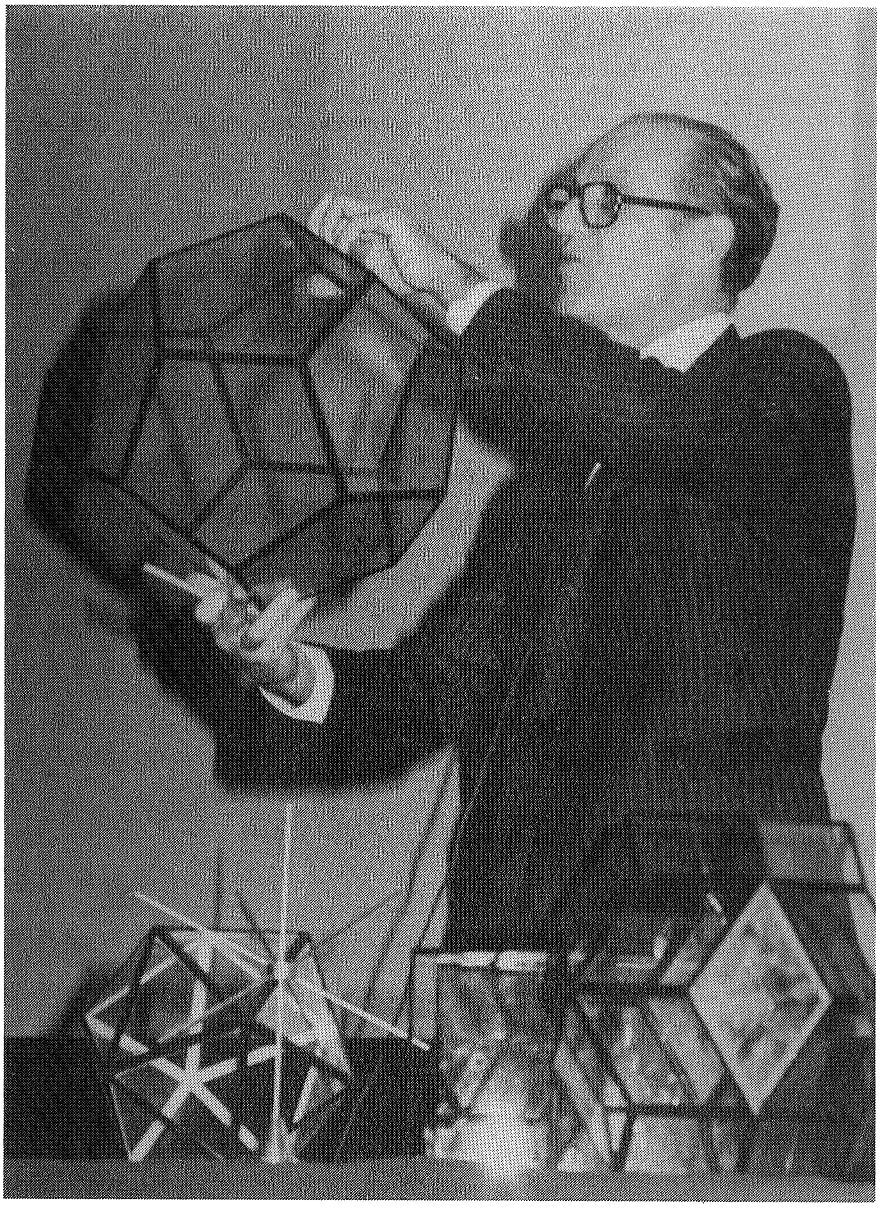
WHITEHEAD, J.H.C., *On* C^1*-complexes,* Ann. of Math. 41, (1940), p. 809-824.

ZEEMAN, C., *Seminar on combinatorial topology,* (1963), Notes IHES, Bures-sur-Yvette and University of Warwick, Coventry.

GENERAL REFERENCES

Manifolds Amsterdam 1970, Springer Lecture Notes 197 (1971).

STEENROD, N., *Mathematical reviews (1940-1967) of papers in algebraic and differential topology,* Part I.

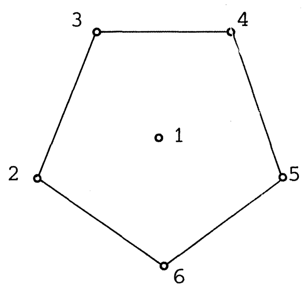


THE PENTAGON

J.J. Seidel

1. INTRODUCTION

My favorite graph is the pentagon graph, or rather the graph on 6 vertices consisting of a pentagon graph and an isolated vertex:



Also the dodecahedron with its 12 pentagonal faces should be of interest. However its dual, the icosahedron, suits me better.

The reason for my preference is the role of the graph as a first case for many combinatorial structures. It is the aim of the present paper to illustrate various aspects of our graph, which serve as a starting point for combinatorial structures such as conference matrices, error-correcting codes, geometric configurations, two-graphs, finite permutation groups, lattices, classical and spherical designs. In doing so, we shall touch several subfields of mathematics and its applications. Our strategy will be to name the sections after these subfields (sometimes with a certain understatement), to start each section with our standard example, and to elaborate a little into the direction of current research.

Note. The original publication of this paper is in the Annals of the New York Academy of Sciences, Second International Conference on Combinational Mathematics, 1978.

2. MATRICES

2.1. First example

The graph of section 1 is described by the matrix

$$C_6 := \begin{bmatrix} 0 & + & + & + & + & + \\ + & 0 & - & + & + & - \\ + & - & 0 & - & + & + \\ + & + & - & 0 & - & + \\ + & + & + & - & 0 & - \\ + & - & + & + & - & 0 \end{bmatrix},$$

whose elements c_{ij} are defined to be -1, 1, 0, according as the vertices i and j are adjacent, nonadjacent, coinciding (for convenience we write + for 1 and - for -1). The first observation is that C_6 is an orthogonal matrix since

$$(C_6)^2 = 5I_6,$$

where I_n denotes the unit matrix of size n.

2.2. Conference matrices

A conference matrix of order n is an $n \times n$ matrix C with elements 0 on the diagonal and ± 1 elsewhere, satisfying $CC^t = (n-1)I$. Necessary conditions for the existence of symmetric conference matrices are $n \equiv 2 \pmod 4$, and $n - 1$ is the sum of 2 squares of integers. They have been constructed for $n - 1 = p^r$, $p \neq 2$ prime, and for other infinite series (including $n = 46$ and $n = 226$), cf. [13], [10], [18]. However, the general problem of existence and uniqueness is unsolved. Conference matrices are used a.o. in network theory and in weighing designs, cf. [1], [20], [24].

2.3. Hadamard matrices

A Hadamard matrix of order n is an $n \times n$ matrix H with elements ± 1 satisfying $HH^t = nI$. A necessary condition for its existence is $n \equiv 0$ (mod 4). It has been conjectured, but never proved, that this condition is also sufficient. A further conjecture is that for all $n \equiv 0 \pmod 4$

there exists a Hadamard matrix of skew type, having the elements $h_{ii} = -1$, and $h_{ji} = -h_{ij}$ for all $i \neq j$ from 1 to n. Here are a few examples of Hadamard matrices:

$$H_4 = \begin{bmatrix} - & + & + & + \\ - & - & + & - \\ - & - & - & + \\ - & + & - & - \end{bmatrix}, \quad H_{2n} = \begin{bmatrix} C_n + I_n & C_n - I_n \\ \\ C_n - I_n & -C_n - I_n \end{bmatrix},$$

where C_n is a symmetric conference matrix of the order n.

3. CODING THEORY

3.1. First example

Consider the elements of the 6×12 matrix

$$[C_6 \quad I_6]$$

as elements of the ternary field $\mathbb{F}_3$. The 6 row vectors of this matrix generate a linear subspace of dimension 6 of the vector space of dimension 12 over $\mathbb{F}_3$. The 3^6 vectors [code words] of this linear subspace [linear code] differ pairwise in 6, 9 or 12 positions [have Hamming distance 6, 9 or 12]. Indeed, the Hamming distances are divisable by 3 since

$$[C_6 \quad I_6] \begin{bmatrix} C_6 \\ \\ I_6 \end{bmatrix} = [O_6]$$

over $\mathbb{F}_3$, and it is easy to check that Hamming distance 3 does not occur. It follows that the linear $(3^6,12,6)$-code is 3-error-detecting and 2-error-correcting. The linear $(3^6,11,5)$ Golay code, obtained by deleting any one coordinate, is 2-error-correcting and perfect. This means that the spheres of radius 2 around the vectors of the code are disjoint, and that the vectors of the code and the spheres exhaust the vector space:

$$3^{11} = 3^6(1+2\times 11+4\times\binom{11}{2}).$$

These codes are related to the Mathieu groups M_{12}, M_{11}, and to the Steiner system 5-(12,6,1) cf. [4].

3.2. Symmetry codes. For $q \equiv -1 \pmod 6$, a symmetry code over $\mathbb{F}_3$ is defined by its generator matrix

$$[C_{q+1} \quad I_{q+1}],$$

where C_{q+1} is a conference matrix. These codes have been introduced and investigated by PLESS [19], cf. also [4]. Examples for q = 5,11,17,23,29 yield 5-designs.

4. DESCRIPTIVE GEOMETRY

4.1. First example

Since the matrix C_6 of section 2 satisfies $C_6^2 = 5I_5$, its eigenvalues are $\sqrt{5}$ and $-\sqrt{5}$, each of multiplicity 3. Hence

$$\frac{1}{2\sqrt{5}} (C_6+I_6\sqrt{5})$$

has the eigenvalues 1 and 0, each of multiplicity 3. It follows that we may write

$$\frac{1}{2\sqrt{5}} (C_6+I_6\sqrt{5}) = HH^t, \quad \text{with } H^tH = I_3,$$

for some 6×3 matrix H and its transposed H^t. Indeed, by the spectral theorem, cf. for instance [14], any $n \times n$ symmetric matrix G may be brought into diagonal form Λ as follows:

$$G = S\Lambda S^t, \qquad S \text{ orthogonal.}$$

Now in the case that the eigenvalues of G are 1^d and 0^{n-d} we write, without loss of generality,

$$\Lambda = \begin{bmatrix} I_d & O \\ & \\ O & O_{n-d} \end{bmatrix}, \quad S = [H \quad K]$$

with $n \times d$ matrix H. It then follows that

$$G = HH^t \quad \text{and} \quad H^tH = I_d.$$

In our example the geometric interpretation is as follows. The 6 row vectors of H represent 6 vectors in 3-space, which are the orthogonal projections of the 6 vectors of an orthonormal frame in 6-space. The 6 row vectors of H have equal length $1/\sqrt{2}$, and are at mutual angles ϕ or $\pi - \phi$, with $\cos \phi = 1/\sqrt{5}$. These 6 vectors span 6 lines which are equiangular, that is, the angle between each pair of lines is the same. We shall soon see how these vectors and lines can be realized in 3-space.

4.2. Eutactic stars

In a real vector space V of finite dimension d a finite spanning set of vectors is called a eutactic star if its vectors are the orthogonal projections into V of the vectors of an orthonormal frame in a vector space U which contains V as a subspace. It is known [23] that a set of vectors forms a eutactic star if and only if the matrix of the inner products of the vectors has exactly 2 distinct eigenvalues (essentially, the argument is the one given in the example). By use of this criterion it is not difficult to construct eutactic stars, cf. [23]. The notion of eutactic star goes back to SCHLÄFLI, [7] p.261. Essentially, the existence of a eutactic star expresses the possibility for "axonometry", a projection method in descriptive geometry. Indeed, by axonometric projection, the geometry in space is described in the plane, and the 3 vectors of an orthonormal frame are projected onto a eutactic star.

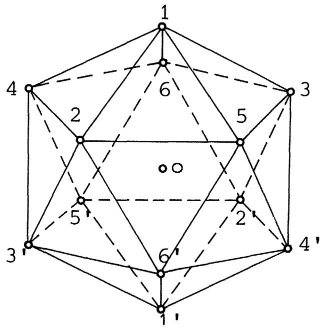
4.3. Root systems

The root systems provide examples of eutactic stars. For instance, the exceptional root system E_8 consists of 240 vectors in Euclidean 8-space, 112 of type $(\pm 1)^2 0^6$, and 128 of type $(\pm\frac{1}{2})^8$ with an even number of minus signs. The inner products of these vectors are ± 2, ± 1, 0. It turns out (cf. [3]) that the 240×240 matrix of the inner products has 2 eigenvalues, one of which is 0. Therefore, the vectors of E_8 are the projections, onto 8-space, of an orthonormal frame in 240-space, up to a factor.

5. SOLID GEOMETRY

5.1. First example

The regular icosahedron provides a natural setting for the graph of section 1 and its matrix C_6. To explain this, we first observe that the 6 lines connecting the antipodal pairs of vertices of the regular icosahedron are equiangular. These diagonals provide the realization in 3-space announced in section 4. Indeed, up to orthogonal transformations they constitute the only set of 6 lines in 3-space whose mutual angles are the same [12].

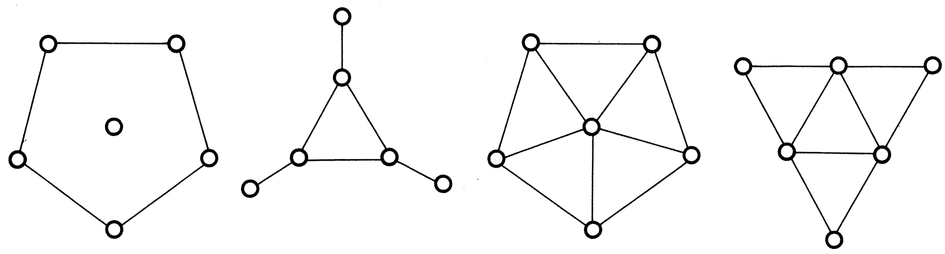


Along each of the 6 diagonals we select a vector of length $1/\sqrt{2}$. As a first choice we consider the vectors 01, 02, 03, 04, 05, 06. The matrix of their inner products is precisely

$$\frac{1}{2\sqrt{5}}\,(C_6+I_6\sqrt{5}),$$

hence these 6 vectors represent the matrix C_6 and our standard graph. The edges [non-edges] of the graph correspond to the pairs of vectors at obtuse [acute] angle.

By a different choice of the 6 vectors different graphs may be obtained. For instance, the vectors 01, 02, 03, 04, 05, 06' yield the second graph of the figure below. Switching from 06 to 06' has the effect of multiplying by -1 the 6-th row and column of the matrix C_6. The new graph is said to be related to the old graph by switching with respect to vertex 6. One can switch with respect to any vertex, and with respect to any subset of the vertices. The graphs thus obtained are said to form a switching class of graphs. In our example any graph in the switching class is isomorphic to one of the following 4 (unlabelled) graphs:



5.2. Equiangular lines

A set of equiangular lines in Euclidean space of dimension d is a set of lines (through one point) any pair of which has the same angle ϕ. To each set of equiangular lines there is attached a switching class of graphs. Conversely, let Γ be any graph on n vertices, let C denote its ($\pm$) adjacency matrix, and let n-d be the multiplicity of the smallest eigenvalue λ of C. Then Γ defines a switching class of graphs, and a set of equiangular lines in $\mathbb{R}^d$ at $\cos\phi = -1/\lambda$. Here are some results, [17], [16], [21] pp.299-307, for the maximum number n(d) of equiangular lines in $\mathbb{R}^d$ at angle ϕ:

d	= 2	3	4	5	6	7	.	.	14	15	21	22	23	.	.	41
n(d)	= 3	6	6	10	16	28	.	.	28	36	126	176	276	.	.	276
$(\cos\phi)^{-1}$	= 2	$\sqrt{5}$	3	3	3	3	.	.	3	5	5	5	5	.	.	5

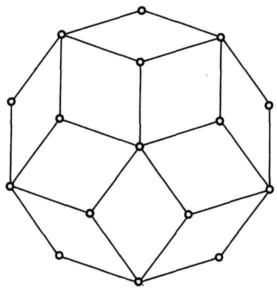
where, for d = 14 and d ≥ 24, the numbers are still subject to some uncertainly. For sufficiently large d we have

$$d\sqrt{d} \le n(d) \le \tfrac{1}{2}d(d+1).$$

6. CRYSTALLOGRAPHY

6.1. First example

The rhombic triacontahedron is a polyhedron in 3-space whose 30 faces are congruent rhombs. Its 60 edges fall into 6 sets of 10 parallel edges, parallel to the 6 diagonals of a regular icosahedron. Here is an incomplete drawing, a view from above (from below) in which the 10 edges perpendicular to the plane of the drawing are missing ([8] p.63).



If the rhombic triacontahedron is contracted in the direction of any one edge, so that the corresponding zone disappears, then a rhombic icosahedron is obtained. This polyhedron is again represented by our drawing, now in a more complete way (views from above and from below). Repeating this process of contraction with respect to a second parallel set of edges, we obtain a rhombic dodecahedron. Its edges are parallel to 4 of the 6 diagonals of the regular icosahedron, hence its rhombs have angles ϕ satisfying $\cos\phi = \pm 1/\sqrt{5}$.

Fedorov, in the course of his crystallographic investigation in 1883, enumerated zonohedra (solids bounded by centrally symmetric polygons). Now his rhombic dodecahedron is different from the polyhedron of the same name explained above. Indeed, the edges of the official rhombic dodecahedron are parallel to the 4 diagonals of the cube, which are equiangular with $\cos\phi = \frac{1}{3}$. It was only in 1960 that Bilinski found the new type of rhombic dodecahdron (cf. [8]).

6.2. Polyhedra and lattices

For a star consisting of n vectors $e_1,\ldots,e_n$ in Euclidean d-space we consider the vectors

$$z_1e_1 + \ldots + z_ne_n.$$

Restriction to $0 \le z_i \le 1$, $i = 1,2,\ldots,n$, leads to polyhedra; restriction to integer z_i leads to lattices (which may or may not be dense). Special properties of the star (such as eutaxy, cf. section 4) yield special polyhedra and lattices. Therefore, the study of stars is useful for the theory of polytopes [7], and of lattices and quadratic forms [6].

6.3. The Korkine - Zolotareff, and the Leech lattice

Following J. McKay we consider the lattice in $\mathbb{R}^{2k}$ consisting of the integral linear combinations of the columns of

$$B_{2k} := \frac{1}{2}\begin{bmatrix} 4I_k & H_k - I_k \\ & \\ O_k & I_k \end{bmatrix},$$

where H_k is a Hadamard matrix of skew type of order k having the constant diagonal $-I_k$, cf. section 2.3. This lattice is unimodular, that is, $\det B_{2k} = 1$. Now

$$B^tB = \begin{bmatrix} 4I & H-I \\ & \\ H^t-I & \frac{1}{4}(k+4)\,I \end{bmatrix}$$

implies the following, for all columns of B, whence for all vectors x of the lattice. For $k \equiv 0 \pmod 4$ all inner products are integral. For $k = 4$ all (x,x) are even. For $k = 12$ the minimum of (x,x) equals 4.

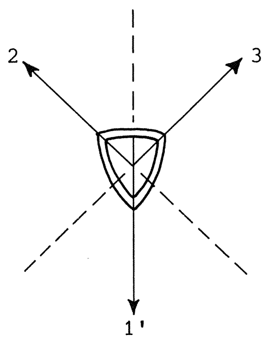
For $k = 4$ we obtain the lattice in $\mathbb{R}^8$ which was first discovered by Korkine and Zolotareff in 1873, cf. [6]. The 240 minimal vectors, with $(x,x) = 2$, form the exceptional root system E_8 mentioned in section 4.3.

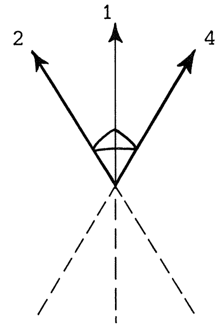
For $k = 12$ we obtain the lattice in $\mathbb{R}^{24}$ discovered by Leech in 1967. The 196560 minimal vectors, with $(x,x) = 4$, are pairwise antipodal and form $\binom{28}{5}$ lines at angles ϕ with $\cos\phi \in \{0,\frac{1}{2},\frac{1}{4}\}$. Their automorphism group is Conway's simple group of the order 4157776806543360000, cf. [5].

7. TOPOLOGY

7.1. First example

In the set of the 6 diagonals of the regular icosahedron two kinds of triples may be distinguised. An odd triple of diagonals is a triple spanned by 3 vectors at obtuse angles, such as 01', 02, 03. An even triple of diagonals is spanned by 3 vectors at acute angles, such as 01, 02, 04. Any triple of diagonals either carries vectors of the obtuse type (odd triple) or of the acute type (even triple).





For any graph in the switching class, the 3 vertices corresponding to any odd [even] triple carry an odd [even] number of edges.

Among the 20 triples of diagonals there are 10 odd triples. Any diagonal is contained in 5 odd triples, and any pair of diagonals is contained in 2 odd triples. Furthermore, each 4 diagonals contain an even number of odd triples. We restate these facts in a more sophisticated language. The 6 diagonals and their odd triples form a 2 - (6,3,2) block design. Furthermore, the characteristic function of the odd triples is a 2-cocycle modulo 2 of the simplex. Indeed, defining $g(x,y,z) = 1$ or 0 according as $\{x,y,z\}$ is an odd or an even triple, we have for each quadruple $\{x,y,z,t\}$ of diagonals the following relation:

$$\delta g(x,y,z,t) := g(x,y,z) + g(x,y,t) + g(x,z,t) + g(y,z,t) \equiv 0 \pmod 2.$$

7.2. Two-graphs

A two-graph (V,Δ) is a pair of a set V and a subset Δ of the set V_3 of all unordered triples from V, such that every 4-subset of V contains an even number of triples from Δ. A two-graph is regular whenever each pair of elements of V is contained in the same number of triples from Δ.

For a survey of two-graphs we refer to [22]. Two-graphs, switching classes of graphs, and sets of equiangular lines are equivalent notions. For n vertices, the number of nonisomorphic two-graphs equals the number of nonisomorphic Euler graphs.

7.3. Cohomology over $\mathbb{F}_2$

The statements of the last paragraph are a consequence of the following isomorphisms between the $\mathbb{F}_2$ vector spaces of 2-cocycles (two-graphs), of cosets of 1-cochains modulo 1-cocycles (switching classes of graphs), and of dual 1-cycles (Euler graphs):

$$Z^2 \cong C^1/Z^1 \cong Z_1^*.$$

We refer to [2] for details.

8. PERMUTATION GROUPS

8.1. First example

It is well-known [15] that the group of the rotations (direct orthogonal transformations) which map the regular icosahedron onto itself is isomorphic to A_5, the alternating group on 5 letters of order 60. The elements of this group also map the set of the 6 diagonals of the icosahedron onto itself, and the group acts 2-transitively on the diagonals. However, for any set of 6 vectors, one along each diagonal, there is only a subgroup which maps the set onto itself. For the set 01,02,03,04,05,06 mentioned in section 5, this subgroup is isomorphic to the dihedral group D_5 of order 10; for the set 01,02,03,04,05,06' to D_3 of order 6.

This translates in terms of two-graphs and graphs as follows. The group of the permutations of the 6 vertices which map odd triples onto odd triples (the automorphism group of the two-graph) is isomorphic to A_5, and acts 2-transitively on the vertices. The group of the permutations of the vertices of any graph in the switching class which map edges onto edges (the automorphism group of the graph) is isomorphic to either D_5 or D_3.

The above illustrates that D_5 and D_3 are subgroups of A_5. Furthermore, the two-graph (the set of the 6 diagonals) provides a structure which admits a doubly transitive representation for the finite simple group A_5. Finally we remark that the group of all orthogonal transformations (so including reflections) which map the icosohedron onto itself has 120 elements. This group is the direct product of $\{\pm I\}$ and A_5.

8.2. Finite simple groups

The notion of regular two-graphs was proposed by Graham Higman as a combinatorial setting for the doubly transitive representation of certain sporadic simple groups. Referring to the results in 5.2 above, and to [22], we mention the following examples of groups acting as an automorphism group of a regular two-graph: Sp(6,2) on regular two-graphs with 28 and with 36 vertices, U(3,q) on $q^3 + 1$ vertices, the Higman-Sims group on 176 vertices, and the Conway ·3 group on 276 vertices. In fact, the last group is characterized by this property, since the non-trivial regular two-graph on 276 vertices is unique (up to taking complements), cf. [11].

8.3. Cohomology of groups

Recently CAMERON [2] defined 2 invariants associated with a group of automorphisms of a two-graph, in terms of one-dimensional and two-dimensional cohomology groups, respectively. In our example the non-triviality of the first invariant expresses the fact that the two-graph has more automorphisms than any graph in its switching class. In our example the triviality of the second invariant expresses the fact that the automorphism group of the two-graph (the 6 diagonals) acts as an automorphism group on the icosahedron, that is, the semi-direct product of $\{\pm I\}$ by A_5 is a direct product.

9. TRIGONOMETRY

9.1. First example

Let P denote the vertex set of a regular pentagon in the plane, inscribed in the unit circle U. Let

$$p_i = (\cos \phi_i, \sin \phi_i), \qquad i = 1,\ldots,5$$

denote the vertices of P, and their coordinates with respect to an orthonormal frame. The elementary addition formula

$$\cos k(\phi_i - \phi_j) = \cos k\phi_i \cos k\phi_j + \sin k\phi_i \sin k\phi_j,$$

when summed over $i,j = 1,\ldots,5$, yields

$$\sum_{i,j=1}^{5} \cos k(\phi_i-\phi_j) = \left(\sum_{i=1}^{5} \cos k\phi_i\right)^2 + \left(\sum_{i=1}^{5} \sin k\phi_i\right)^2$$

for any $k \in \mathbb{N}$. We observe that the left hand side is invariant under rotations of the pentagon; this is not necessarily true for each of the 2 summands at the right hand side. Now it is easy to see that the left hand side equals zero for $k \not\equiv 0 \pmod 5$, hence

$$\sum_{i=1}^{5} \cos k\phi_i = 0 = \sum_{i=1}^{5} \sin k\phi_i, \qquad \text{for } k = 1,2,3,4.$$

Remarkably, these relations remain valid if the pentagon is rotated, as a consequence of our observation above.

9.2. Spherical designs

The considerations above carry over to d-dimensional space. Again the main tool is the addition formula, which now is formulated in terms of Gegenbauer (ultraspherical) polynomials in one, and harmonic polynomials in d variables. Let Hom(k) denote the linear space of the homogeneous polynomials in d variables of total degree k, restricted to the unit sphere U. Let Harm(k) denote the subspace of the harmonic polynomials, satisfying Laplace's equation.

A spherical t-design is a finite subset P of the unit sphere U in $\mathbb{R}^d$ having the following property, for $k = 1,\dots,t$:

$$\sum_{p \in P} h(p) = 0, \qquad \text{for all } h \in \mathrm{Harm}(k).$$

An equivalent formulation is that, for all $f \in \mathrm{Hom}(k)$,

$$\sum_{p \in P^{\sigma}} f(p) \text{ is independent of orthogonal transformations } \sigma \text{ of } \mathbb{R}^d.$$

We refer to [9] for further information. Some examples of spherical t-designs follow: the 5 vertices of the pentagon with $t = 4$, the 12 vertices of the icosahedron with $t = 5$, the 240 vectors of the root system E_8 with $t = 7$, the 196560 minimal vectors of the Leech lattice in $\mathbb{R}^{24}$ with $t = 11$. Spherical 2-designs are spherical eutactic stars which are balanced.

10. MECHANICS

10.1. First example

The phenomenon described in the first example of section 9 admits an interpretation in terms of classical mechanics. Put equal masses on the vertices of the regular pentagon P, then the first, second, third and fourth moments, with respect to the centre of P, of the mass system are independent of rotations of P. Indeed, the k^{th} moments are defined as

$$\sum_{i=1}^{5} (\cos \phi_i)^a (\sin \phi_i)^b, \qquad a + b = k,\ a,b \in \mathbb{N},$$

and each $(\cos \phi)^a (\sin \phi)^b$ is a linear combination of $\cos \phi$, $\sin \phi$, $\cos 2\phi$, $\sin 2\phi, \ldots, \cos k\phi$, $\sin k\phi$. For $k = 1$ this means that the centre of gravity of the mass system is in the centre of the unit circle U. For $k = 2$ this means that the inertia ellips is a circle. It follows, that for $k = 1,2,3,4$ the k^{th} moments of the mass system equal the corresponding k^{th} moments of the circle on which the mass is homogeneously spread out, that is,

$$\frac{1}{5} \sum_{i=1}^{5} (\cos \phi_i)^a (\cos \phi_i)^b = \frac{1}{\pi} \int_U (\cos \phi)^a (\sin \phi)^b d\phi,$$

for $a + b = k \in \{1,2,3,4\}$. This equation may be phrased as the tensor equation below, for $t = 4$ and the present P and U.

10.2. Symmetric tensors

An equivalent definition for the notion of a spherical t-design is the following.

A finite subset P of the unit sphere U in Euclidean space of dimension d is a spherical t-design whenever the following holds for $k = 1,\ldots,t$:

$$\frac{1}{|P|} \sum_{p \in P} \otimes^k p = \frac{1}{|U|} \int_U \otimes^k p \; d\omega(p).$$

The equivalence with the definition in section 9 is evident. Indeed, if the coordinates of p are $(x_1,\ldots,x_d)$, then the $\frac{1}{2}d(d+1)$ components of $p \otimes p$ are $x_i x_j$, $i \le j$ form 1 to d, and the $\binom{d+k-1}{k}$ components of $\otimes^k p$ are the monomials of degree k in $x_1,\ldots,x_d$.

REFERENCES

[1] BELEVITCH, V., *Conference networks and Hadamard matrices,* Ann. Soc. Sci. Bruxelles, 82 (1968), 13-32.

[2] CAMERON, P.J., *Cohomological aspects of two-graphs,* Math. Zeitschr., to be published.

[3] CAMERON, P.J., J.M. GOETHALS, J.J. SEIDEL & E.E. SHULT, *Line graphs, root systems, and elliptic geometry,* J. Algebra 43 (1976), 305-327.

[4] CAMERON, P.J. & J.H. VAN LINT, *Graph theory, coding theory and block designs,* London Math. Soc. Lecture note series 19, Cambridge Univ. Press, 1975.

[5] CONWAY, J.H., *Three lectures on exceptional groups,* Chapter 7 of *Finite simple groups,* ed. M.B. Powell and G. Higman, Acad. Press (1971), 215-247.

[6] COXETER, H.S.M., *Extreme forms,* Canad. J. Math. 3 (1951), 391-441.

[7] COXETER, H.S.M., *Regular polytopes,* third edt., Dover, New York, 1973.

[8] COXETER, H.S.M., *The classification of zonohedra,* in *Twelve geometric essays,* South. Ill. Univ. Press, Carbondale, 1968, 54-74.

[9] DELSARTE, P., J.M. GOETHALS & J.J. SEIDEL, *Spherical codes and designs, Geometriae Dedicata,* to be published.

[10] GOETHALS, J.M. & J.J. SEIDEL, *Orthogonal matrices with zero diagonal,* Canad. J. Math. 19 (1967), 1001-1010.

[11] GOETHALS, J.M. & J.J. SEIDEL, *The regular two-graph on 276 vertices,* Discr. Math. 12 (1975), 143-158.

[12] HAANTJES, J., *Equilateral point-sets in elliptic two- and three- dimensional spaces,* Nieuw Arch. Wisk. 22 (1948), 355-362.

[13] HALL, M., *Combinatorial theory,* Ginn, Blaisdell, 1967.

[14] HALMOS, P.R., *Finite-dimensional vector spaces,* Van Nostrand, Princeton, 1958.

[15] KLEIN, F., *Vorlesungen über das Ikosaeder und ......,* Leipzig, 1884.

[16] LEMMENS, P.W.H. & J.J. SEIDEL, *Equiangular lines,* J. Algebra 24 (1973), 494-512.

17 LINT, J.H. Van & J.J. SEIDEL, *Equilateral point sets in elliptic geometry,* Nederl. Akad. Wetensch. Proc. A 69 (= Indag. Math. 28) (1966), 335-348.

18 MATHON, R., *Symmetric conference matrices of order* $pq^2 + 1$, Canad. J. Math., to be published.

19 PLESSE, V., *Symmetry codes over* GF(3) *and new* 5-*designs,* J. Combin. Theory 12 (1972), 119-142.

20 RAGHAVARAO, D., *Constructions and combinatorial problems in design of experiments,* Wiley, 1971.

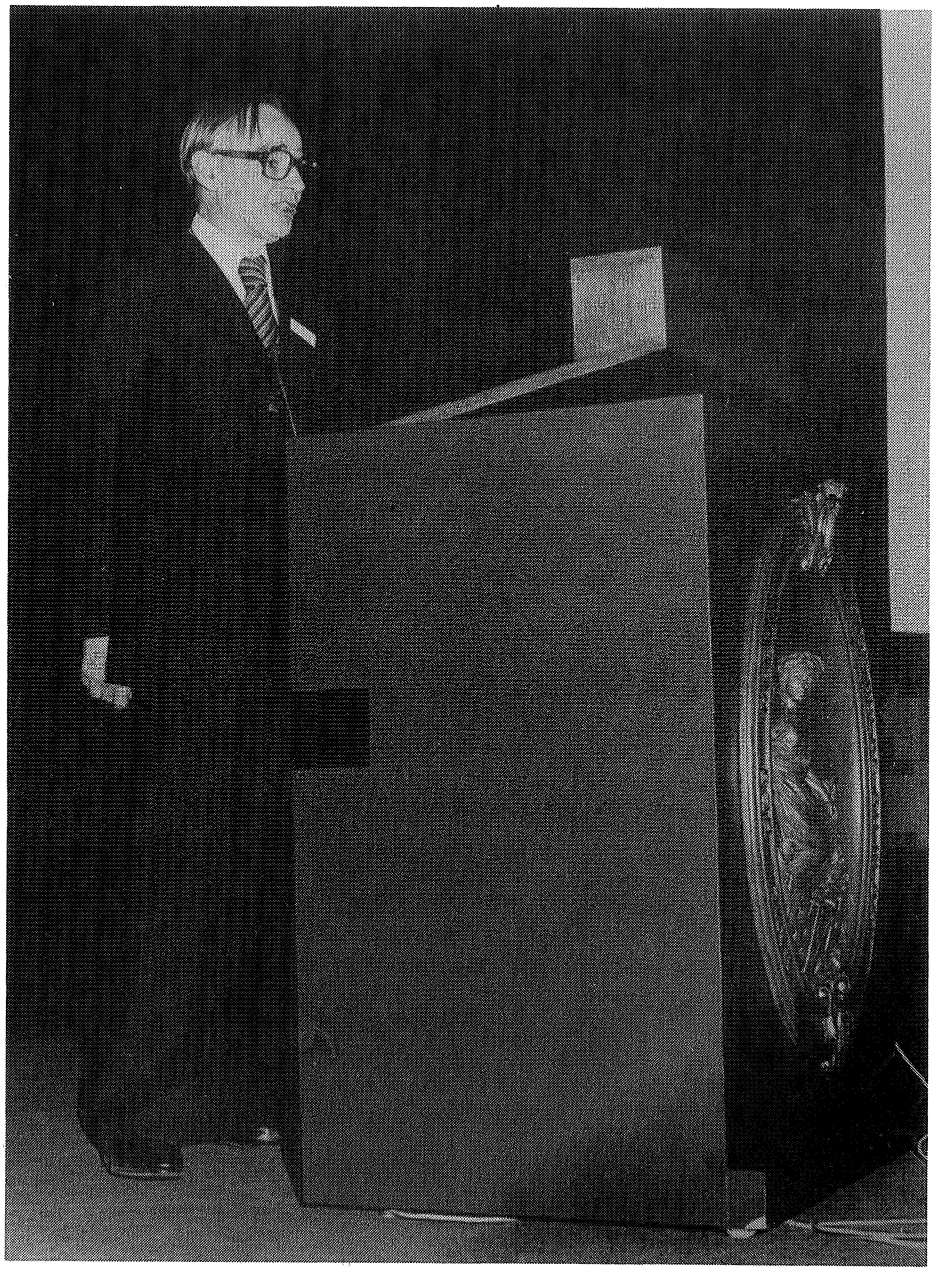
21 ROUSE BALL, W.W. & H.S.M. COXETER, *Mathematical recreations and essays,* twelfth edit., Univ. Toronto Press, 1974.

22 SEIDEL, J.J., *A survey of two-graphs,* Proc. Intern. Colloqu. Teorie Combinatorie (Roma 1973), Tomo I, Accad. Naz. Lincei, Roma 1976, 481-511.

23 SEIDEL, J.J., *Eutactic stars,* Proc. 5th Hungarian Colloqu. on Combinatorics (Keszthely 1976), to be published.

24 SLOANE, N.J.A. & M. HARWIT, *Masks for Hadamard transform optics, and weighing designs,* Appl. Optics 15 (1976), 107-114.

Technological University
Eindhoven, The Netherlands



ALGEBRAIC GEOMETRY IN HOLLAND BETWEEN 1918 AND 1928; REMINISCENCES FROM MY STUDENT DAYS

by B.L. van der Waerden

Under this title Professor B.L. van der WAERDEN presented a plenary lecture. Professor van der Waerden, who began his mathematics studies in Amsterdam in 1919, described in his lecture how his lifelong involvement with algebraic geometry originated in the general interest in this field among Dutch mathematicians during his student days. He gave an impression of his teachers in Amsterdam, their interests and their approach to mathematics, and proceeded with a discussion of some concrete problems from algebraic geometry which drew his attention. Professor van der WAERDEN's reconstruction of his early work and the ideas and motivation behind it were highly illuminating. This early research soon led him to Göttingen and Emmy NOETHER; the great influence of Emmy NOETHER on Van der WAERDEN's own work was described in detail.

The program committee, on inviting professor van der WAERDEN for this plenary lecture, agreed that his talk would not be published. However, a tape recording of the lecture (which was held in the Dutch language) is preserved in the archives of the WISKUNDIG GENOOTSCHAP.

LAPLACE INTEGRALS, FACTORIAL SERIES AND SINGULAR DIFFERENTIAL EQUATIONS

B.L.J. BRAAKSMA

0. INTRODUCTION

In this paper we are concerned with singular differential equations

$$(0.1) \qquad x^{1-p}\frac{dy}{dx} = f(x,y),$$

where $p \in \mathbb{N}$, $p > 0$, $y \in \mathbb{C}^n$ and $f(x,y) \in \mathbb{C}^n$ is a polynomial in the components y_j, $j = 1,\dots,n$ of y, with coefficients that are holomorphic in a region $S = \{x \in \mathbb{C} : |x| > R \text{ and } \alpha < \arg x < \beta\}$ $(\alpha < \beta,\ 0 < R)$. Then ∞ is a singular point of the differential equation of rank at most p. Let

$$(0.2) \qquad f(x,y) = \sum_{|\nu|=0}^{r} a_\nu(x)\, y^\nu,$$

where $\nu = (\nu_1,\dots,\nu_n) \in \mathbb{N}^n$, $|\nu| = \nu_1 + \dots + \nu_n$, $y^\nu = y_1^{\nu_1} \dots y_n^{\nu_n}$. We assume asymptotic expansions for the coefficients

$$(0.3) \qquad a_\nu(x) \sim \sum_{k=0}^{\infty} a_{\nu,k}\, x^{-k} \text{ as } x \to \infty \text{ in } S.$$

The construction of solutions of (0.1) near the singular point ∞ often consists of two parts: in the first part one constructs formal solutions and in the second part one shows that there are analytic solutions associated with the formal solutions. We will consider this analytic part mainly.

In several cases (0.1) has formal solutions

$$(0.4) \qquad y(x) = x^\lambda \sum_{k=0}^{\infty} c_k\, x^{-k}$$

where $\lambda \in \mathbb{C}$, $c_k \in \mathbb{C}^n$ if $k = 0,1,2,\dots$. This means that substitution of the formal series (0.3) in (0.2) and then substitution of the formal series for f and y in (0.1) gives an identity. If $p = 0$ and the coefficients a_p in (0.2) are holomorphic in ∞ then the formal series (0.3) and (0.4) converge and y represents a holomorphic solution according to a theorem of Briot and Bouquet. However, if $p \geq 1$ the formal series (0.4) for y need not converge, even if the series (0.3) converge. On the other hand, in several cases one may show that on a suitable subsector of S there exists a solution y(x) which has the righthand side of (0.4) as asymptotic expansion. For example, if $a_0(x) = 0$ and

$$(0.5) \qquad \sum_{|\nu|=1} a_\nu(x) y^\nu = A(x) y$$

where A(x) is an $n \times n$ matrix which is nonsingular if $x \in S$, $|x|$ sufficiently large, and $\lambda = -1$ then such a solution exists in a subsector of S with central angle less than π/p (cf. WASOW [15, p.58])

Here we are concerned with cases where the coefficients a_ν are representable as Laplace integrals in sectors with central angle at least π/p. We consider two classes of Laplace integrals A_1 and A_2 which will be defined in sect. 1. If the coefficients a_ν belong to A_j and there exists a formal solution (0.4) of (0.1) then under certain conditions there exists a holomorphic solution y of (0.1) such that $x^{-\lambda} y$ is of class A_j and the righthand side of (0.4) is the asymptotic expansion of y as $x \to \infty$ on S (cf. theorems 1 and 2 and their corollaries).

The class A_2 of Laplace integrals consists of functions which admit convergent factorial series expansions. The problem whether there exists a factorial series solution of (0.1) corresponding to a formal solution (0.4) is important since if the factorial series solution exists it may be calculated directly from the formal series (0.4). Hence in these cases we may calculate the solution to any degree of approximation from the formal series (0.4) which in general diverges.

Solutions of (0.1) in the form of Laplace integrals have been studied by Poincaré, Birkhoff, Horn, Trjitzinski, Turrittin and others. Following HORN ([3],[4],[5]) we transform the differential equation (0.1) by means of

$$(0.6) \qquad y(x) = \int_0^\infty e^{-x^p t} \varphi(t) dt + y_0$$

into a singular Volterra integral equation for φ. We show that a solution of this integral equation exists in a suitable Banach space of holomorphic functions of exponential type in a sector. This leads to a solution of (0.1) with the desired properties (cf. sect. 2 for the linear case and sect. 3 for the nonlinear case).

In the nonlinear case formal power series solutions of the form (0.4) do not always exist, though there may be formal solutions of the form

$$(0.7) \qquad y(x) = \sum_{k=0}^{\infty} c_k x^{-\lambda_k}, \; \lambda_k \to \infty \text{ as } k \to \infty.$$

We show in sect. 3 that analytic solutions in the form (0.6) may be associated with these solutions.

In sect. 4 we give applications of sect. 2 and 3. Here we show when formal solutions of (0.1) exist to which correspond analytic solutions by means of the theorems in sect. 2 and 3. Furthermore an application to canonical forms of linear equations is given. The results are related to work by HORN [3],[4], MALMQUIST [10], TURRITTIN [11] and IWANO [7] (cf. also WASOW [15, ch. 11]).

The linear case of (0.1) has been investigated by W.A. Harris Jr. and myself in [1], where also functional differential equations of a certain type are considered.

1. LAPLACE INTEGRALS AND FACTORIAL SERIES

We shall consider the differential system (0.1) with coefficients that belong to a class of Laplace integrals. We use two classes of Laplace integrals. They are defined as follows:

DEFINITION 1. Let p be a positive integer, $\theta_1 \le \theta_2$, $\mu \ge 0$ and define

$$(1.1) \qquad G_1 = \{x \in \mathbb{C} : \exists\, \theta \in [\theta_1, \theta_2] \text{ such that } \mathrm{Re}(x^p e^{i\theta}) > \mu\}.$$

Then $A_1(\theta_1, \theta_2, \mu, p)$ is the set of analytic functions $F\colon G_1 \to \mathbb{C}^n$ such that

$$(1.2) \qquad F(x) = F_0 + \int_0^{\infty e^{i\theta}} e^{-x^p t} f(t)dt \equiv F_0 + L_p f(x), \; x \in G_1,$$

where $F_0 \in \mathbb{C}^n$ and f has the following properties:

(i) Let $S_1 = \{t \in \mathbb{C}: \theta_1 \le \arg t \le \theta_2\}$, then $t^{1-\frac{1}{p}} f \in C(S_1, \mathbb{C}^n)$ and if $\theta_1 < \theta_2$, then f is analytic in the interior $\overset{0}{S}_1$ of S_1 (S_1 contains 0).

ii) $f(t) = \mathcal{O}(e^{\mu_1|t|})$ as $t \to \infty$ on S_1 for all $\mu_1 > \mu$.

iii) $f(t) \sim \sum_{m=1}^{\infty} f_m t^{\frac{m}{p}-1}$ as $t \to 0$ on S_1 where $f_m \in \mathbb{C}^n$, $m = 1,2,\ldots$

The class $A_2(\omega,\mu)$ where $\omega \in \mathbb{C}$, $\omega \neq 0$, $\mu \ge 0$ is a subset of $A_1(\theta,\theta,\mu,1)$ where $\theta = -\arg \omega$. It is defined as follows:

<u>DEFINIITON 2</u>. Let $G_2 = \{x \in \mathbb{C}: \mathrm{Re}\,(xe^{i\theta}) > \mu\}$. Then $A_2(\omega,\mu)$ is the set of all functions $F: G_2 \to \mathbb{C}^n$ with the representation (1.2), $\theta = -\arg \omega$, $p = 1$, such that $F_0 \in \mathbb{C}^n$ and f has the following properties:

i) Let $S_2(\omega)$ be the component of $\{t \in \mathbb{C} : |1 - e^{-\omega t}| \le 1\}$ that contains the ray: $\arg t = \theta$. Then $f : S_2(\omega) \to \mathbb{C}^n$ is continuous and analytic in $\overset{0}{S}_2(\omega)$.

ii) $f(t) = \mathcal{O}(e^{\mu_1|t|})$ as $t \to \infty$ on $S_2(\omega)$ for all $\mu_1 > \mu$.

For short we will often write A_1 or A_2 for the classes defined above. Moreover, we will use a similar definition for matrix functions.

It is well known (cf. DOETSCH [2, p.45, 174] that $F \in A_1(\theta_1,\theta_2,\mu,p)$ implies

$$(1.3) \qquad F(x) \sim F_0 + \sum_{m=1}^{\infty} \Gamma(\tfrac{m}{p}) f_m x^{-m} \qquad \text{as } x \to \infty$$

on any closed subsector of $G_1: -\frac{1}{2}\pi - \theta_2 - \varepsilon \le \arg x^p \le \frac{1}{2}\pi - \theta_1 - \varepsilon$, $\varepsilon > 0$. For short we shall say in this case that (1.3) holds on closed subsectors of G_1. Conversely, if F is analytic on a closed sector G such that $G_1 \subset \overset{0}{G}$ and (1.3) holds on G, then F belongs to $A_1(\theta_1,\theta_2,\mu,p)$ for some $\mu \ge 0$.

If $F \in A_2(\omega,\mu)$, then F is representable by a factorial series

$$(1.4) \qquad F(x) = F_0 + \sum_{m=0}^{\infty} \frac{m!F_{m+1}}{\frac{x}{\omega}(\frac{x}{\omega}+1)\ldots(\frac{x}{\omega}+m)}, \; x \in G_2,$$

where $F_m \in \mathbb{C}^n$ if $m \in \mathbb{N}$ (cf. DOETSCH [2, p.221]).

Conversely. if (1.4) holds, then F has a Laplace integral representation (1.2) with $\theta = -\arg \omega$, $p = 1$ under somewhat weaker conditions on f than in definition 2: $f(t) = \mathcal{O}(e^{\mu_1|t|})$ as $t \to \infty$ on $|\mathrm{Im}\,\omega t| \le \frac{\pi}{2} - \varepsilon$ for all $\mu_1 > \mu$, $0 < \varepsilon < \frac{\pi}{2}$ and f is analytic in $\overset{0}{S}_2(\omega)$.

If $F \in A_2(\omega,\mu)$, then (1.3) with $p = 1$ holds as $x \to \infty$ on any closed subsector of G_2: $|\arg xe^{i\theta}| \le \frac{\pi}{2} - \varepsilon$ $(0 < \varepsilon \le \frac{\pi}{2})$. Conversely, if $F \in A_2(\omega,\mu)$ and (1.3) with $p = 1$ holds as $x \to \infty$ on $|\arg xe^{i\theta}| \le \frac{\pi}{2} - \varepsilon$, then we may construct the factorial series (1.4) from the asymptotic series: we may expand each term in (1.4) in an asymptotic power series, comparison with (1.3) now gives a recursion formula for the F_{m+1}. Alternatively we may write x^{-m} as a factorial series; substitution in (1.3) and comparison with (1.4) gives also a recursion formula for F_m. For the explicit form of this formula cf. WASOW [15, p.330]. In this way we sum asymptotic series for functions in $A_2(\omega,\mu)$ by factorial series. This is a useful property since factorial series converge uniformly in half planes. This property will be used in the sequel where we encounter formal power series solutions which under certain conditions are asymptotic expansions of solutions in $A_2(\omega,\mu)$ and consequently may be summed to any degree of approximation by factorial series.

If $m > 1$, then $S_2(m\omega) \subset S_2(\omega)$ and so $A_2(\omega,\mu) \subset A_2(m\omega,\mu)$. Consequently factorial series (1.4) also are representable by factorial series (1.4) on G_2 with parameter $m\omega$ instead of ω if $m > 1$.

If $F_1, F_2 \in A_j$ then also $F_1F_2 \in A_j$ since $f_1 * f_2$ satisfies the condition of definition j if f_1 and f_2 do so.

2. THE LINEAR EQUATION

We now consider the differential system in the case that it is linear and that it is a coupled system of a system with a singularity of the first kind and a system with a singularity of the second kind.

To formulate this we partition $n \times n$-matrices along the n_1-th row and column $(0 \le n_1 \le n)$:

$$M = \begin{pmatrix} M^{11} & M^{12} \\ & \\ M^{21} & M^{22} \end{pmatrix},$$

where M^{jh} is an $n_j \times n_h$ matrix, $n_2 = n - n_1$. A corresponding splitting of vectors $f = \begin{pmatrix} f^1 \\ f^2 \end{pmatrix}$ after the n_1-th component will be used.

Now consider the system

$$x^{1-p}\frac{dy}{dx} = A(x)y + b(x), \tag{2.1}$$

where p is a positive integer, and concerning A and b we assume either case 1: $A, b \in \mathcal{A}_1(\theta_1,\theta_2,\mu,p)$ or case 2: $p = 1$ and $A, b \in \mathcal{A}_2(\omega,\mu)$.

Then we have representations

$$A(x) = A_0 + \mathcal{L}_p\alpha(x),\ b(x) = b_0 + \mathcal{L}_p\beta(x) \tag{2.2}$$

and asymptotic expansions

$$A(x) \sim \sum_{m=0}^{\infty} A_m x^{-m},\ b(x) \sim \sum_{m=0}^{\infty} b_m x^{-m} \quad \text{as } x \to \infty \tag{2.3}$$

in closed subsectors of G_1 in case 1 and G_2 in case 2.

We assume

$$\begin{cases} A_m^{11} = 0,\ A_m^{12} = 0,\ b_m^1 = 0 \text{ if } m = 0,1,\dots,p-1;\ A_0^{21} = 0, \\ A_0^{22} + ptI \text{ is nonsingular on } S_j \text{ in case } j. \end{cases} \tag{2.4}$$

Then we have

<u>THEOREM 1</u>. *Suppose* $\sum_0^\infty c_m x^{-m}$ *is a formal solution of* (2.1). *Then there exists an analytic solution* y *of* (2.1) *which belongs to* $\mathcal{A}_1(\theta_1,\theta_2,\mu,p)$ *in case* 1 *and to* $\mathcal{A}_2(\omega,\mu)$ *in case* 2 *such that*

$$y(x) \sim \sum_0^\infty c_m x^{-m} \tag{2.5}$$

as $x \to \infty$ *on any closed subsector of* G_1 *in case* 1 *and of* G_2 *in case* 2. *The solution* y *with these properties is unique.*

<u>REMARK</u>. In case 2 we may sum the formal solution $\Sigma_0^\infty c_m x^{-m}$ to a convergent factorial series which satisfies (2.1) on $\mathrm{Re}(xe^{i\theta}) > \mu$ (cf. sect. 1).

<u>PROOF</u>. Let $u = \sum_0^{N-1} c_m x^{-m}$, a partial sum of the formal solution. Then

$$x^{1-p}\frac{du}{dx} = A(x)u + b(x) - c(x),$$

where $c \in A_j$ and $c^1(x) = \mathcal{O}(x^{-p-N})$, $c^2(x) = \mathcal{O}(x^{-N})$ as $x \to \infty$ on closed subsectors of G_j. Hence with $y - u = v$ we get $x^{1-p}\frac{dv}{dx} = A(x)\, v + c(x)$ as equation equivalent to (2.1).

So it is sufficient to prove the theorem in case $b_h^1 = 0$, $h = 0,\dots,$ $p + N - 1$, $b_h^2 = 0$, $h = 0,1,\dots,N-1$ for a sufficiently large integer N. We assume this latter condition from now on or equivalently by (2.2) (cf. (1.3))

$$\beta(t) \sim \sum_{m=N}^{\infty} \beta_m\, t^{\frac{m}{p}-1} \quad \text{as } t \to 0 \text{ in } S_j,\ \beta_h^1 = 0 \text{ if } N \le h \le N + p - 1. \tag{2.6}$$

We seek a solution y of (2.1) which is $\mathcal{O}(x^{-N})$ as $x \to \infty$, and which belongs to A_j. If $y = L_p\varphi$ is of class A_j then

$$x^{1-p}\frac{dy}{dx} = L_p(-pt\varphi),\ A(x)y = L_p(A_0\varphi + \alpha * \varphi). \tag{2.7}$$

Hence (2.1) has a solution $y = L_p\varphi$ of class A_j iff $-pt\varphi = A_0\varphi + \alpha * \varphi + \beta$, and φ satisfies the conditions in definition j for f(j = 1 or 2). This equation for φ is a singular Volterra integral equation.

If $t^{1-\frac{1}{p}} v \in C(S_j,\mathbb{C}^n)$ we define

$$Tv = -\ (A_0 + p\,t\,I)^{-1}(\alpha * v). \tag{2.8}$$

With

$$\psi = -\ (A_0 + p\,t\,I)^{-1}\beta \tag{2.9}$$

the equation for φ is equivalent to

$$\varphi = T\varphi + \psi. \tag{2.10}$$

The assumptions on A_0 imply that

$$(A_0 + p\,t\,I)^{-1} = \operatorname{diag}\{p^{-1}t^{-1}I_{n_1},\ (A_0^{22} + p\,t\,I_{n_2})^{-1}\} \tag{2.11}$$

and that

$$(A_0^{22} + p\,t\,I_{n_2})^{-1} \text{ and } t(A_0 + p\,t\,I)^{-1} \tag{2.12}$$

are uniformly bounded on S_j. So if $n_1 > 0$ then T is singular in $t = 0$.

We solve (2.10) in a Banach space V_N of functions $v \in C(S_j, \mathbb{C}^n)$ such that v is analytic in S_j^0, $t^{1-\frac{N}{p}} v$ is continuous in S_j and

$$(2.13) \qquad \|v\|_N = \sup_{t \in S_j} \left| t^{1-\frac{N}{p}} v(t) \right| e^{-\mu_1 |t|} < \infty.$$

Here μ_1 is a fixed number, $\mu_1 > \mu$ where μ is the parameter μ in $A_1(\theta_1, \theta_2, \mu, p)$ or $A_2(\omega, \mu)$. It is clear that V_N is a Banach space with norm $\|.\|_N$. A similar definition will be used for matrix-valued functions.

Since $b \in A_j$, it follows from (2.2) that $\beta(t) = \mathcal{O}(e^{\mu_1 |t|})$ as $t \to \infty$ in S_j. Using (2.9),(2.6),(2.11) and (2.12) we deduce $\psi \in V_N$.

Next we show that T maps V_N into V_N. From the assumption $A \in A_j$, (2.2) and (2.4) we deduce that $\alpha^{1h} \in V_p$, $\alpha^{2h} \in V_1$, $h = 1,2$. Since

$$(2.14) \qquad t^{k-1} * t^{m-1} = B(k,m)\, t^{k+m-1} \text{ if } \operatorname{Re} k > 0,\ \operatorname{Re} m > 0,$$

we see that $t^{-\frac{N}{p}} (\alpha * v)^1$ and $t^{1-\frac{N+1}{p}} (\alpha * v)^2$ are continuous on S_j and analytic in S_j^0 if $v \in V_N$. Moreover, if $t \in S_j$ then

$$\left| t^{-1} (\alpha * v)^1(t) \right| \le (\|\alpha^{11}\|_p + \|\alpha^{12}\|_p) e^{\mu_1 |t|} \|v\|_N \left| t^{-1} (1 * t^{\frac{N}{p}-1}) \right|.$$

Hence, by (2.11)

$$(2.15) \qquad \|\{(A_0 + p\,t\,I)^{-1} (\alpha * v)\}^1\|_N \le \frac{1}{N} (\|\alpha^{11}\|_p + \|\alpha^{12}\|_p) \|v\|_N.$$

Similarly

$$\left| (\alpha * v)^2(t) \right| \le \|\alpha\|_1 e^{\mu_1 |t|} \|v\|_N \left| t^{\frac{1}{p}-1} * t^{\frac{N}{p}-1} \right|$$

and therefore

$$(2.16) \qquad \|\{(A_0 + p\,t\,I)^{-1} (\alpha * v)\}^2\|_N \le \|\alpha\|_1 \|v\|_N\, B(\tfrac{1}{p}, \tfrac{N}{p}) \sup_{t \in S_j} \left| t^{\frac{1}{p}} (A_0^{22} + p\,t\,I_{n_2})^{-1} \right|.$$

Hence with (2.8) and (2.12) we see that T maps V_N into V_N and that there exists a constant K independent of N such that

$$\|T\| \le K\,\Gamma(\tfrac{N}{p})\,\{\Gamma(\tfrac{N+1}{p})\}^{-1}.$$

Choosing N_0 sufficiently large we see that T is a contraction on V_N if $N \ge N_0$. Consequently there exists a unique solution of (2.10) in V_N if $N \ge N_0$. Now going backwards we easily verify that $y = L_p\varphi$ satisfies (2.1) and $y = \mathcal{O}(x^{-N})$ as $x \to \infty$ in closed subsectors of G_j.

Hence, if $N \ge N_0$ and $\mu_1 > \mu$ there exists a unique solution $y = c_0 + L_p\varphi$ of the original equation (2.1), without assuming (2.6), such that φ is analytic in S_j^0, $t^{1-\frac{1}{p}}\varphi$ is continuous in S_j and

$$\varphi(t) = \sum_{m=1}^{N-1} \frac{c_m}{\Gamma(m/p)}\, t^{\frac{m}{p}-1} + \mathcal{O}(t^{\frac{N}{p}-1}) \text{ as } t \to 0 \text{ in } S_j,$$

$$\varphi(t) = \mathcal{O}(e^{\mu_1|t|}) \text{ as } t \to \infty \text{ in } S_j.$$

Now the uniqueness implies that φ does not depend on N. Hence we have a unique solution $y \in A_j$ with parameter μ_1 instead of μ such that (2.5) holds. By variation of μ_1 we see that this solution y belongs to the class A_j with parameter μ. $\square$

COROLLARY. *We make the same assumptions as in theorem* 1 *except that the cases* 1 *and* 2 *are modified as follows: Assume*

$$\text{(2.17)} \qquad A(x) = \sum_{h=0}^{p-1} x^{-h}\,\tilde{A}_h(x^p),\; b(x) = \sum_{h=0}^{p-1} x^{-h}\,\tilde{b}_h(x^p),$$

where $\tilde{A}_h$, $\tilde{b}_h$, $h = 0,1,\dots,p-1$, *are of class* $A_1(\theta_1,\theta_2,\mu,1)$ *in case* 1 *and of class* $A_2(\omega,\mu)$ *in case* 2. *Then, if* $\sum_0^\infty c_m x^{-m}$ *is a formal solution of* (2.1), *there exists an analytic solution* $y(x) = \sum_{h=0}^{p-1} x^{-h}\,\tilde{y}_h(x^p)$ *where* $\tilde{y}_h$ *is of class* A_j *and* (2.5) *holds as* $x \to \infty$ *in* $-\frac{\pi}{2} - \theta_2 + \varepsilon \le p \arg x \le \frac{\pi}{2} - \theta_1 - \varepsilon\,(\varepsilon > 0)$, *where* $\theta_1 = \theta_2 = -\arg\omega$ *in case* 2.

This may be shown using a rank reduction scheme of TURRITTIN [12]: substitute

$$x = \xi^{1/p},\; u(\xi) = (\tilde{y}_0^T(\xi),\dots,\tilde{y}_{p-1}^T(\xi))^T,$$
$$v(\xi) = (\tilde{b}_0^T(\xi),\dots,\tilde{b}_{p-1}^T(\xi))^T.$$

Then (2.1) is equivalent to

$$\frac{du}{d\xi} = M(\xi)u + v(\xi), \tag{2.18}$$

$$\text{where } M(\xi) \sim \sum_{0}^{\infty} M_m \xi^{-m},\ M_0 = \frac{1}{p}\begin{pmatrix} A_0^0 & & & \\ A_0^1 & \ddots & & \\ \vdots & \ddots & \ddots & \\ A_0^{p-1} & \cdots & A_0^1 & A_0^0 \end{pmatrix}.$$

From (2.4) we may deduce that M_0 has n_1p rows of zeros and that 0 is eigenvalue of M_0 with multiplicity n_1p. Hence M_0 is similar to diag$\{0,M_0^{22}\}$ where M_0^{22} is nonsingular. So we may apply theorem 1 to (2.18) and the result follows. □

3. THE NONLINEAR CASE

We now consider the nonlinear analogue of theorem 1. We may write (0.1) in the form

$$x^{1-p}\frac{dy}{dx} = A(x)\ y + \sum_{|\nu|\neq 1}^{r} b_\nu(x)y^\nu. \tag{3.1}$$

We assume either

case 1: A, $b_\nu \in A_1(\theta_1,\theta_2,\mu,p)$, or

case 2: $p = 1$, A, $b_\nu \in A_2(\omega,\mu)$, where $|\nu| = 0,2,\ldots,r$.

Then we have Laplace representations as in (2.2) with (2.3):

$$b_\nu(x) \sim \sum_{m=0}^{\infty} b_{\nu m}x^{-m} \text{ as } x \to \infty \text{ in closed subsectors of } G_j. \tag{3.2}$$

We assume

$$\begin{cases} A_m^{11} = 0,\ A_m^{12} = 0,\ b_{\nu m}^1 = 0 \text{ if } m = 0,\ldots,p-1 \text{ and} \\ |\nu| = 0,2,\ldots,r;\ A_0^{21} = 0, \\ A_0^{22} + ptI \text{ is nonsingular on } S_j \text{ in case } j. \end{cases} \tag{3.3}$$

If these assumptions are satisfied we have

<u>THEOREM 2</u>. *Suppose* (3.1) *possesses a formal solution* $\sum_1^\infty c_m x^{-m}$. *Then there exists an analytic solution* $y(x)$ *of* (3.1) *such that* (2.5) *holds as* $x \to \infty$ *on closed subsectors of* G_j *in case* j *and there exists a real number* $\mu_1 > \mu$ *such that* $y \in A_1(\theta_1, \theta_2, \mu_1, p)$ *in case* 1 *and* $y \in A_2(\omega, \mu_1)$ *in case* 2.

<u>PROOF</u>. Let $u = \sum_1^{N-1} c_m x^{-m}$. Then (3.1) may be transformed into an equivalent equation for $v = y-u$. This equation is of the same type as (3.1) but with different coefficients $A(x)$, $b_\nu(x)$ which still satisfy the conditions (3.2), (3.3) etc. above. Now

$$(3.4) \qquad b_0^1(x) = \mathcal{O}(x^{-N-p}),\ b_0^2(x) = \mathcal{O}(x^{-N})$$

as $x \to \infty$ in closed subsectors of G_j. Hence it is sufficient to prove theorem 2 in the case that (3.4) holds with a suitably large N. Then $b_{00} = 0$ and (2.6) holds with β replaced by β_0 and β_m by β_{0m}.

In the same way as in the proof of theorem 1 we see that $y = L_p\varphi$ is a solution of (3.1) which belongs to A_j iff there exists a solution φ of (2.10) with

$$(3.5) \qquad \begin{aligned} Tv &= -(A_0 + p\,t\,I)^{-1}\{\alpha * v + \sum_{|\nu|=2}^{r} (b_{\nu 0}v^{*\nu} + \beta_\nu * v^{*\nu})\}, \\ \psi &= -(A_0 + p\,t\,I)^{-1}\beta_0, \end{aligned}$$

and φ satisfies the conditions for f in definition j of sect. 1. In (3.5) $v^{*\nu}$ denotes the convolution of ν_1 factors v_1, ν_2 factors v_2 etc.

We solve (2.10) in a Banach space W_N which arises from the space V_N used in sect. 2 by replacing μ_1 by 0 and S_j by $\tilde{S}_j = S_j \cap \overline{\Delta(0;\varepsilon)}$ where $\Delta(0;\varepsilon)$ is the disc in $\mathbb{C}$ with center 0 and radius $\varepsilon > 0$. As in sect. 2 we may show $\psi \in W_N$ using (2.6), (2.12) and (3.5).

From (2.14) we deduce (2.15) and (2.16) for $v \in W_N$ in the same way as in sect. 2. Similarly we have using (3.3) and (2.2)

$$(3.6) \qquad \begin{cases} \|\{(A_0 + p\,t\,I)^{-1}(\beta_\nu * v_h)\}^1\|_N \le \frac{1}{N}\|\beta_\nu^1\|_p \|v_h\|_N \\ \text{and} \\ \|\{(A_0 + p\,t\,I)^{-1}(\beta_\nu * v_h)\}^2\|_N \le \\ \qquad \le \|\beta_\nu\|_1 \|v_h\|_N\, B(\frac{1}{p},\frac{N}{p}) \sup_{t\in\tilde{S}_j} |t^{\frac{1}{p}}(A_0^{22} + p\,t\,I_{n_2})^{-1}|. \end{cases}$$

If v_0 and w_0 are scalar functions belonging to the corresponding space W_N of scalar functions then we have in view of (2.14)

$$(3.7)\qquad |(v_0 * w_0)(t)| \le \|v_0\|_N \|w_0\|_N \, B(\tfrac{N}{p},\tfrac{N}{p}) \, t^{\frac{2N}{p}-1} \quad \text{if } t \in \tilde{S}_j,$$

and $v_0 * w_0 \in W_N$. If v and w are n-vector functions in W_N and $|\nu| \ge 2$ then $v^{*\nu} - w^{*\nu}$ may be written as a sum of terms $(v_j - w_j) * \psi$, where ψ is the convolution of some components of v and w and so $\psi \in W_N$. Hence we may infer from (3.7) that $v^{*\nu}$ satisfies a Lipschitz condition on sets $\{v \in W_N: \|v\|_N \le K_0\}$ and that the Lipschitz constant tends to zero as ε tends to zero (cf. definition of $\tilde{S}_j$). In view of (3.6), (3.3) and (2.12) the same holds for the term

$$(A_0 + p\,t\,I)^{-1} \sum_{|\nu|=2}^{r} (b_{\nu 0} v^{*\nu} + \beta_\nu * v^{*\nu})$$

in (3.5).

Because of (2.15) we now choose an integer $N_0 > 2\{\|\alpha^{11}\|_p + \|\alpha^{12}\|_p\}$. Next we choose an integer $N \ge N_0$ and define

$$K = \sup\{|t^{1-\frac{N}{p}}\psi(t)| : t \in S_j, |t| \le 1\}, B_\varepsilon = \{v \in W_N : \|v\|_N \le 2K\}.$$

Combining (3.5) with (2.15), (2.16) and the remark above we see that T is a contraction if ε is chosen sufficiently small. Hence there exists a unique solution φ of (2.10) on $\tilde{S}_j$ if $N \ge N_0$.

Now suppose the solution φ of (2.10) with (3.5) is known on $S_j' = S_j \cap \bar{\Delta}(0;\rho_0)$ for some $\rho_0 > 0$. Choose $t_0 \in S_j'$ with $\frac{1}{2}\rho_0 < |t_0| < \rho_0$. Let $S_j^- = S_j' \cap (S_j - t_0)$ and $S_j^+ = S_j^- + t_0$. Here $S_j - t_0$ means the set S_j translated over $-t_0$.

We transform the integral equation (2.10) on S_j^+ using the following decomposition of $v * w$ for scalar functions v,w continuous on $S_j' \cup S_j^+$ and analytic in its interior.

$$(v*w)(t_0+t_1) = \{v(t_0+.)*w\}(t_1) + \{w(t_0+.)*v\}(t_1) +$$

$$+\int_{t_1}^{t_0} w(t_0+t_1-\tau)v(\tau)d\tau, \qquad t_1 \in S_j^-,$$

where the paths of integration $[0,t_1]$ and $[t_0,t_1]$ belong to S_j'. Applying this decomposition successively to the convolutions in (3.5) we may show that

$$(Tv)(t_0+t_1) = -(A_0+p(t_0+t_1)I)^{-1}\{v(t_0+\cdot)*\chi_1\}(t_1) + \chi_2(t_1)\},$$

if $t_1 \in S_j^-$, where χ_1 and χ_2 are functions continuous in S_j^- and analytic in $(S_j^-)^0$ which only depend on the values of v in S_j'.

Hence (2.10) may be transformed into a linear Volterra integral equation for $v = \varphi(t_0+\cdot)$ in S_j^-. It has a unique solution which is continuous in S_j^- and analytic in $(S_j^-)^0$. Hence it coincides with the known solution $\varphi(t_0+\cdot)$ on $S_j^- \cap (S_j'-t_0)$. By variation of t_0 we get a unique solution of (2.10) on $S_j \cap \bar{\Delta}\ (0;3/2\ \rho_0)$, hence on S_j, which is continuous on S_j and analytic in S_j^0.

Now we estimate φ on S_j. Let $\mu_0 > \mu$ and

$$(3.8) \qquad g(\rho) = \sup\{|\varphi(t)|: t \in S_j,\ |t| = \rho\}, \qquad \text{if } \rho \geq 0.$$

Since $A, b_\nu \in \mathcal{A}_j$ we have with (2.2):

$$|\alpha(t)|,\ |\beta_\nu(t)| \leq K\ \rho^{\frac{1}{p}-1} \exp(\mu_0\rho), \qquad \text{if } t \in S_j,\ |t| = \rho.$$

Here K is some positive constant. Now use this and (3.8) in (3.5) and (2.10). Then we see

$$(3.9) \qquad g(\rho) < (T_1 g)(\rho),$$

where

$$(T_1 g)(\rho) = M\{e^{\mu_0\rho} + \sum_{m=2}^{r} g^{*m}(\rho) + \sum_{m=1}^{r} (e^{\mu_0\rho}\rho^{\frac{1}{p}-1}) * g^{*m}(\rho)\}$$

if $\rho \geq 1$ for some constant M. We choose M such that also $\sup\{g(\rho): \rho \in [0,1]\} < M$. Then (3.9) holds on $\mathbb{R}_+$.

Following Walter [14,p.17] we first solve $v = T_1 v$. If $w = \mathcal{L}_1 v$, then

$$w(x) = M(x-\mu_0)^{-1} + M\sum_{m=2}^{r} w^m(x) + M\Gamma(\tfrac{1}{p})(x-\mu_0)^{-\frac{1}{p}} \sum_{m=1}^{r} w^m(x).$$

This equation has a unique solution w in a neighbourhood V of ∞ which is analytic in $x^{1/p}$, positive for $x > 0$, $x \in V$ and

$$w(x) = Mx^{-1} + \mathcal{O}(x^{-1-\frac{1}{p}}) \qquad \text{as } x^{\frac{1}{p}} \to \infty.$$

Let V contain the half plane Re $x \geq \mu_1$ where $\mu_1 > \mu_0$. Then

$$v(\rho) = (\mathcal{L}_1^{-1} w)(\rho) = M + \frac{1}{2\pi i} \int_{\mu_1 - i\infty}^{\mu_1 + i\infty} e^{\rho x}\{w(x) - Mx^{-1}\}dx, \qquad \rho \geq 0.$$

It follows that v is real-valued, $v(0) = M$ (cf. [2,p.174]) and $v(\rho) = \mathcal{O}(\exp \mu_1 \rho)$ as $\rho \to +\infty$. In particular we have $g(0) < v(0)$.

Suppose there exists $\rho_0 > 0$ such that $0 < g(\rho) < v(\rho)$ if $0 < \rho < \rho_0$ and $g(\rho_0) = v(\rho_0)$. Then (3.9) implies

$$g(\rho_0) < (T_1 g)(\rho_0) < (T_1 v)(\rho_0) = v(\rho_0),$$

which gives a contradiction. Hence $g < v$ on $\mathbb{R}_+$ and so

$$(3.10) \qquad |\varphi(t)| \leq K_0 \exp(\mu_1 |t|), \qquad \text{if } t \in S_j,$$

for some constant K_0. Consequently $y = L_p \varphi$ exists on G_j and y is a solution of (3.1). In the same way as in the proof of theorem 1 we may show $y \in \mathcal{A}_j$ with parameter μ_1 instead of μ and (2.5) as $x \to \infty$ on G_j. □

COROLLARY. *Similarly to the corollary of theorem* 1 *we may modify the cases* 1 *and* 2 *in theorem* 2 *as follows. Assume*

$$A(x) = \sum_{h=0}^{p-1} x^{-h} \tilde{A}_h(x^p), \; b_\nu(x) = \sum_{h=0}^{p-1} x^{-h} \tilde{b}_{\nu h}(x^p),$$

where $\tilde{A}_h$, $\tilde{b}_{\nu h}$, $h = 0,\dots,p-1$ *are of class* $\mathcal{A}_1(\theta_1,\theta_2,\mu,1)$ *in case* 1 *and of class* $\mathcal{A}_2(\omega,\mu)$ *in case* 2. *Then, if the other assumptions of theorem* 2 *are satisfied, there exists an analytic solution* $y(x) = \sum_{h=0}^{p-1} x^{-h} \tilde{y}_h(x^p)$ *where* $\tilde{y}_h \in \mathcal{A}_1(\theta_1,\theta_2,\mu_1,1)$ *in case* 1, $\tilde{y}_h \in \mathcal{A}_2(\omega,\mu_1)$ *in case* 2 *for some* $\mu_1 > \mu$ *and* (2.5) *holds as* $x \to \infty$ *in* $-\frac{1}{2}\pi - \theta_2 + \varepsilon \leq p \arg x \leq \frac{1}{2}\pi - \theta_1 - \varepsilon (\varepsilon > 0)$, *where in case* 2: $\theta_1 = \theta_2 = -\arg \omega$.

REMARK. The case that (3.1) possesses a formal solution $y = \sum_0^\infty c_m x^{-m}$ with $c_0 \neq 0$ may be reduced to the case of theorem 2 with $c_0 = 0$ by substituting $v = y - c_0$. Then v satisfies an equation of the form (3.1) with new coefficients A and b_ν for which the conditions of theorem 2 have to be verified.

In several cases (3.1) has formal solutions of the form

$$(3.11) \qquad y(x) = \sum_{|k|=1}^{\infty} d_k x^{-k.\kappa},$$

where $k = (k_0,\ldots,k_g) \in \mathbb{N}^{g+1}$ (g a positive integer), $\kappa = (1,\kappa_1,\ldots,\kappa_g)$, $\operatorname{Re}\kappa_j > 0$ if $j = 1,\ldots,g$ and $k.\kappa = k_0 + k_1\kappa_1 + \ldots + k_g\kappa_g$ (we refer to sect. 4 for cases where such formal solutions exist). In these cases we have the following generalization of theorem 2.

THEOREM 3. *Suppose the coefficients* A *and* b_ν *in* (3.1) *belong to* $A_1(\theta_1,\theta_2,\mu,p)$ *with* (2.3), (3.2) *and* (3.3). *Suppose* (3.1) *possesses a formal solution* (3.11) *as above.*

Then there exists a real number $\mu_1 > \mu$ *and an analytic solution* $y = L_p\varphi$ *of* (3.1) *on* $\tilde{G}_1$ *where* $\tilde{G}_1 = \{x \in \mathbb{C} : \exists\, \theta \in [\theta_1,\theta_2]$ *such that* $\operatorname{Re}(x^p e^{i\theta}) > \mu_1\}$, *such that*

$$y(x) \sim \sum_{|k|=1}^{\infty} d_k\, x^{-k.\kappa} \text{ as } x \to \infty \text{ on closed subsectors of } \tilde{G}_1.$$

PROOF. The proof is similar to that of theorem 2. Let the set of numbers $k.\kappa$ with $|k| \geq 1$ be arranged in order of increasing magnitude of their real parts to the sequence $\lambda_0,\lambda_1,\lambda_2,\ldots$. Hence $0 < \operatorname{Re}\lambda_0 \leq \operatorname{Re}\lambda_1 \leq \ldots$ and $\operatorname{Re}\lambda_m \to \infty$ as $m \to \infty$. Then (3.11) may be rewritten as $y(x) = \sum_0^\infty c_m x^{-\lambda_m}$.

Let $u = \sum_0^{N-1} c_m x^{-\lambda_m}$, where N is chosen in such a way that $\operatorname{Re}\lambda_{N-1} < \operatorname{Re}\lambda_N$. Then (3.1) may be transformed into an equivalent equation for $v = y - u$. This equation is of the same type as (3.1) but with new coefficients $A(x)$ and $b_\nu(x)$ which need not belong to A_1 anymore. However, we still have $A = A_0 + L_p\alpha$, $b_\nu = b_{\nu 0} + L_p\beta_\nu$, where the new functions α and β_ν are analytic in S_1^0,

$$\alpha(t) \sim \sum_0^\infty \tilde{\alpha}_m\, t^{\frac{1}{p}\lambda_m - 1}, \quad \beta_\nu(t) \sim \sum_0^\infty \tilde{\beta}_{\nu m}\, t^{\frac{1}{p}\lambda_m - 1}$$

as $t \to 0$ in S_1 and $t^{1-\frac{1}{p}\lambda_0}\alpha(t)$, $t^{1-\frac{1}{p}\lambda_0}\beta_\nu(t)$ are continuous in S_1. Moreover, (3.3) implies $\tilde{\alpha}_m^{11} = 0$, $\tilde{\alpha}_m^{12} = 0$, $\tilde{\beta}_{\nu m}^{1} = 0$ if $\operatorname{Re}\lambda_m < p$, and

A_0^{21} and A_0^{22} are the same as before. Analogous to (3.4) we now have

$$b_0^1(x) = \mathcal{O}(x^{-\lambda_N - p}), \quad b_0^2(x) = \mathcal{O}(x^{-\lambda_N})$$

as $x \to \infty$ in closed subsectors of G_1.

We now proceed as in the proof of theorem 2 with small modifications. The role of the number N is now played by λ_N. □

4. APPLICATIONS

I A formal nontrivial solution $\sum_0^\infty c_m x^{-m}$ of (2.1) with $b \equiv 0$ exists if (2.4) holds, $A_p^{12} = 0$, $A_p^{11} c = 0$ for some $c \neq 0$ and $A_p^{11} + m\,I$ is nonsingular for $m = 1,2,\ldots$. Now $c_0^1 = c$ and theorem 1 applies.

A formal solution $\sum_0^\infty c_m x^{-m}$ of the inhomogeneous equation (2.1) exists if (2.4) holds and $A_p^{11} + m\,I$ is nonsingular for $m = 0,1,2,\ldots$.

II The homogeneous equation (2.1), $b \equiv 0$, has a formal solution $\sum_0^\infty c_m x^{\lambda-m}$ if (2.4) holds, $A_p^{12} = 0$, $A_p^{11} c = \lambda c$ for some $c \neq 0$, $A_p^{11} + (m-\lambda) I$ is nonsingular for $m = 1,2,\ldots$. Now we may transform to the previous case by the substitution $y(x) = x^\lambda z(x)$. Then we may deduce that the formal solution $\sum_0^\infty c_m x^{-m}$ corresponds to an analytic solution of the new differential equation. Hence the original equation (2.1) has a solution $y(x)$ such that $x^{-\lambda} y(x) \in \mathcal{A}_j$ in case j of theorem 1. A corresponding result follows from the corollary of theorem 1 (cf. [1]).

III Assume $A_p^{12} = 0$ and there is no pair of eigenvalues of A_p^{11} which differ by a positive integer. If the hypotheses concerning the coefficient A in theorem 1 are satisfied and $b \equiv 0$, then there exists an $n \times n_1$ matrix solution $U(x)\, x^{A_p^{11}}$ of (2.1) where $U \in \mathcal{A}_j$, $U \to \begin{pmatrix} I_{n_1} \\ 0 \end{pmatrix}$ as $x \to \infty$ in closed subsectors of G_j. A corresponding statement holds if the assumptions concerning A in theorem 1 are replaced by those in the corollary of theorem 1.

The proof may be given by introducing the differential equation for U and applying application I to this equation (cf. [1]).

<u>REMARK</u>. If the condition $A_m^{12} = 0$, $m = 0,1,\ldots,p$, used above, is not satisfied we may transform (2.1) to an equation where this condition is satisfied without violating the other assumptions made above (cf. WASOW [15, ch.5]).

The applications above contain the results of TURRITTIN [11] concerning factorial series as solutions of (2.1) with $b \equiv 0$. However, we get here as halfplane of convergence of these solutions in case 2 the same halfplane where the coefficients A and b are representable by factorial series, whereas Turrittin gets a smaller halfplane (cf. also VII below).

IV The conditions of theorem 2 are satisfied if besides (3.3) we have

i) $b^1_{op} = 0$, $b^2_{00} = 0$.

ii) if $n_1 \geq 1$, then there is at most one positive integer which is eigenvalue of $-A^{11}_p$; if ℓ is such an eigenvalue, then $b^1_{op+h} = 0$, $b^2_{oh} = 0$ for $h = 1,\dots,\ell-1$ and $A^{12}_p(A^{22}_0)^{-1}b^2_{o\ell} = b^1_{op+\ell}$.

Now it is easy to see that a formal solution $\sum_1^\infty c_m x^{-m}$ of (3.1) exists.

V The conditions of theorem 3 are satisfied if (3.3) holds and

i) $b^1_{op} = 0$, $b^2_{00} = 0$.

ii) if $n_1 \geq 1$, then $\kappa_1,\dots,\kappa_g$ are eigenvalues of $-A^{11}_p$ with $\mathrm{Re}\,\kappa_j > 0$, $j = 1,\dots,g$ and $k_0 + k_1\kappa_1 + \dots + k_g\kappa_g$ is not an eigenvalue of $-A^{11}_p$ if $k_0,\dots,k_g \in \mathbb{N}$ and $k_1 + \dots + k_g \geq 2$ or $k_1 + \dots + k_g = 1$, $k_0 > 0$.

iii) condition ii) of IV holds and $\ell \in \{\kappa_1,\dots,\kappa_g\}$.

An algebraic computation shows that in this case formal solutions of the form (3.11) of (3.1) exist. The number of solutions we obtain in this way, depends on the dimensions of the nullspaces of $A^{11}_p + \kappa_j I$, $j = 1,\dots,g$. We refer to HUKUHARA [6] who computed all formal solutions of systems (0.1).

Cases like those above have been studied extensively by IWANO (cf. [7]). He considers cases where $f(x,y)$ in (0.1) need not be a polynomial in y and he assumes that all eigenvalues of A^{11}_p have negative real part.

The treatment in sect. 3 can be extended to cases where $f(x,y)$ is not a polynomial in y. Also cases where the formal solutions contain logarithms like in IWANO [8] can be treated by using solutions $u = L_p\varphi$ where the expansion of φ near the origin contains logarithmic terms.

VI We may apply IV to obtain a block diagonalization for linear systems

$$x^{1-p}\frac{dy}{dx} = C(x)y \tag{4.1}$$

analogous to WASOW [15, theorem 12.2] and MALMQUIST [10, sect. 3].

<u>THEOREM 4</u>. *Suppose* C *is of class* A_j, $C(x) \sim \sum_0^\infty C_m x^{-m}$, $C_0 = \mathrm{diag}\{C^{11}_0, C^{22}_0\}$. *Let* $\lambda_1,\dots,\lambda_r$ *be the eigenvalues of* C^{11}_0 *and* $\lambda_{r+1},\dots,\lambda_n$ *those of* C^{22}_0.

Assume $\pm \frac{1}{p}(\lambda_g - \lambda_h) \notin S_j$ *if* $g \leq r < h$.

Then there exists a transformation $y = T(x)z$ *which takes* (4.1) *into*

$$(4.2) \qquad x^{1-p}\frac{dz}{dx} = \tilde{C}(x)z,\ \tilde{C}(x) = \mathrm{diag}\{\tilde{C}^{11}(x), \tilde{C}^{22}(x)\}$$

such that T *and* $\tilde{C}$ *are of class* A_j *with a value of the parameter* μ *which differs from the parameter* μ *in the class* A_j *to which* C *belongs. Moreover,* $\tilde{C}_0^{11} = C_0^{11}$, $\tilde{C}_0^{22} = C_0^{22}$. *If* $C^{12}(x)$, $C^{21}(x) = O(x^{-N})$, *then* $T(x) = I + O(x^{-N})$.

PROOF. First we substitute $y = Q(x)w$, where $Q(x) = \begin{pmatrix} I & Q^{12}(x) \\ 0 & I \end{pmatrix}$.
Then (4.1) is transformed into

$$(4.3) \qquad x^{1-p}\frac{dw}{dx} = D(x)w,\ D^{12}(x) \equiv 0,$$

iff

$$(4.4) \qquad x^{1-p}\frac{d}{dx}Q^{12} = C^{11}(x)Q^{12} - Q^{12}C^{22}(x) - Q^{12}C^{21}(x)Q^{12} + C^{12}(x).$$

Now $C_0^{11}Q^{12} - Q^{12}C_0^{22}$ defines a linear transformation in the linear space of matrices Q^{12} with eigenvalues $\lambda_g - \lambda_h$, $g = 1,\ldots,r$; $h = r+1,\ldots n$ (cf. WASOW [15, sect. 11]). Hence we may use application IV with $n_1 = 0$, and a solution of (4.4) exists in A_j with a different value of the parameter μ. In a similar way we may transform (4.3) to the form (4.2). □

A special case of theorem 4 with $j = 2$ has been given by TURRITTIN [13]. Theorem 12.2 in WASOW [15] corresponds to theorem 4 with $j = 1$. However, the sectors considered by Wasow have central angle at most π/p.

VII Linear differential equations (4.1) may be transformed to a canonical form by a linear transformation of y with polynomial coefficients and a change of variable $x \mapsto x^q$. The canonical form is

$$(4.5) \qquad x^{1-p}\frac{dy}{dx} = A(x)y + x^{-1-p}B(x)y,$$

where the positive integer p need not be the same as in (4.1) and

$$(4.6) \qquad A(x) = \mathrm{diag}\{P_1(\tfrac{1}{x})\,I_{n_1} + x^{-p}\tilde{A}_1, \ldots, P_k(\tfrac{1}{x})\,I_{n_k} + x^{-p}\tilde{A}_k\}.$$

Here $P_1,\ldots,P_k$ are polynomials of degree (p - 1) at most, which are mutually distinct, and $\tilde{A}_1,\ldots,\tilde{A}_k$ are constant Jordan matrices such that any two eigenvalues of $\tilde{A}_j$ do not differ by a positive integer $(j = 1,\dots,k)$. $B(x)$ is analytic and has an expansion $\sum_0^\infty B_k x^{-k}$. One of the polynomials P_j may be identically zero.

This canonical form is due to MALMQUIST [10] and TURRITTIN [11] and is a refinement of canonical forms of Hukuhara and Birkhoff. It has been studied recently by LEVELT [9] from an algebraic point of view.

Replacing y in (4.5) by $E_h(x)y$ where $E_h(x) = \exp \int_0^x P_h(\xi^{-1})\xi^{p-1}\, d\xi$, we get (4.5) with P_g replaced by $P_g - P_h$, $g = 1,\dots,k$. Let

$$P_g(x) - P_h(x) \sim c_{gh}\, x^{p_{gh}} \text{ as } x \to 0,\ c_{gh} \neq 0 \text{ if } g \neq h. \tag{4.7}$$

Now assume $p_{gh} = \ell$, is independent of g, if $g \neq h$. Then we may apply III: Suppose $B \in A_1(\theta_1,\theta_2,p-\ell,\mu)$ or

$$B(x) = \sum_{m=0}^{p-\ell-1} x^{-m}\, \tilde{B}_m(x^{p-\ell}),\ \tilde{B}_m \in A_2(\omega,\mu) \text{ if } m = 0,\dots,p-\ell-1,$$

and $c_{gh} \notin S_j$ if $g \neq h$. Then there exists an $n \times n_h$ matrix solution $U_h(x)x^{\tilde{A}_h}E_h(x)$ of (4.5), where U_h belongs to the same class as B, and U_h has rank n_h. This includes results of TURRITTIN [11, thm III]. Malmquist has shown in [10, ch. 3] that there exists an analytic transformation $y = T(x)z$ which transforms (4.5) into $x^{1-p}\frac{dz}{dx} = A(x)z$, which may be integrated easily. This reduction may also be performed using theorem 4 and its analogue theorem 12.2 in WASOW [15], and the remark after definition 2 in sect. 1. The result is: Suppose in the notation of (4.7):
$\beta_0 - \alpha_0 > \frac{\pi}{q} - \frac{\pi}{p}$, $\arg c_{gh} \notin (\alpha_0 + \frac{\pi}{2q} - \frac{\pi}{2p},\ \beta_0 + \frac{\pi}{2p} - \frac{\pi}{2q})$ if $q = p - p_{gh}$, $g \neq h$, and $q > q_0 \geq 0$, $\beta_0 - \alpha_0 \leq (\pi/q_0) - \pi/p$ for some q_0.

Suppose $B(x)$ is analytic in the sector S: $-\frac{\pi}{2p} - \beta_0 < \arg x < \frac{\pi}{2p} - \alpha_0$ for sufficiently large $|x|$ and $B(x) \sim \sum_0^\infty B_m x^{-m}$ as $x \to \infty$ on closed subsectors of S. Then the transformation mentioned above exists and $T(x) \sim I + \sum_1^\infty T_m x^{-m}$ as $x \to \infty$ on closed subsectors of S.

REFERENCES

[1] BRAAKSMA, B.L.J. & W.A. HARRIS, Jr., *Laplace integrals and factorial series in singular functional differential systems.* To appear in Applicable Mathematics.

[2] DOETSCH, G., *Handbuch der Laplace Transformation,* Band II. Birkhäuser Verlag, Basel, 1955.

[3] HORN, J., *Integration linearer Differentialgleichungen durch Laplacesche Integrale und Fakultätenreihen.* Jahresber. Deutsch. Math. Ver., 24 (1915) 309-329; 25 (1917) 74-83.

[4] HORN, J., *Verallgemeinerte Laplacesche Integrale als Lösungen linearer und nichtlinearer Differentialgleichungen.* Jahresber. Deutsch. Math. Ver., 25 (1917) 301-325.

[5] HORN, J., *Laplacesche Integrale, Binomialkoeffizientenreihen und Gammaquotientenreihen in der Theorie der linearen Differentialgleichungen.* Math. Zeitschr., 21 (1924) 85-95.

[6] HUKUHARA, M., *Intégration formelle d'un système des équations différentielles non linéaires dans le voisinage d'un point singulier.* Ann. Mat. Pura Appl., (4) 19 (1940) 35-44.

[7] IWANO, M., *Analytic expressions for bounded solutions of non-linear ordinary differential equations with an irregular type singular point.* Ann. Mat. Pura Appl., (4) 82 (1969) 189-256.

[8] IWANO, M., *Analytic integration of a system of nonlinear ordinary differential equations with an irregular type singularity.* Ann. Mat. Pura Appl., (4) 94 (1972) 109-160.

[9] LEVELT, A.H.M., *Jordan decomposition for a class of singular differential operators.* Arkiv för Math., 13 (1975) 1-27.

[10] MALMQUIST, J., *Sur l'étude analytique des solutions d'un système d'équations différentielles dans le voisinage d'un point singulier d'indétermination,* II. Acta Math., 74 (1941) 1-64.

[11] TURRITTIN, H.L., *Convergent solutions of ordinary linear homogeneous differential equations in the neighborhood of an irregular singular point.* Acta Math., 93 (1955) 27-66.

[12] TURRITTIN, H.L., *Reducing the rank of ordinary differential equations.* Duke Math. J., 30 (1963) 271-274.

[13] TURRITTIN, H.L., *Solvable related equations pertaining to turning point problems,* in *Asymptotic Solutions of Differential Equations and their Applications.* Edited by C.H. Wilcox, John Wiley, New York, 1964, 27-52.

[14] WALTER, W., *Differential- and Integral Inequalities.* Springer Verlag, Berlin, 1970.

[15] WASOW, W., *Asymptotic expansions for ordinary differential equations.* Interscience Publishers, New York, 1965.

ON PREDICTION

J.S. Cramer

1. The distinguishing characteristic of the *randomizer* introduced by Hemelrijk elsewhere in this volume is that the outcome of this device is completely unpredictable (HEMELRIJK 1978). Yet it is implied that most people would respond to such a machine by attempting the impossible, thus testifying to the widespread belief that prediction is in itself a good and useful activity. Scientists quote correct prediction as the ultimate aim and test of scientific pursuit, and policymakers call for predictive studies so that they can ward off undesirable developments by anticipatory action. So far, however, very little attention has been paid to the nature and problems of prediction.

In the present paper we start off from Hemelrijk's analysis. Since he has convincingly shown that prediction is pointless in the case of a randomizer, we must change the terms of reference in order to retrieve the subject. We next consider a number of examples which refer to the weather and to social and economic phenomena. Most of the arguments are wellknown, and we do not advance beyond tentative conclusions.

2. We use the word *predictor* indiscriminately for a method or a person. In Hemelrijk's definition a randomizer is completely unpredictable in the sense that all predictors perform equally badly, i.e. no two predictors differ systematically in their number of successes. In the case of a true six-faced die, for instance, all predictors that are restricted to the integers 1, 2, .. 6 in the long run score successes in one sixth of all cases.

Three aspects of this definition deserve closer examination.

3. The first point is that all predictors use the *available information* in a sensible, rational or perhaps optimal manner; it is not enough that the possible outcomes of the randomizer are known, but this knowledge must also be used. In the case of the six-faced die we cannot admit a predictor who mistakenly tries to predict random digits from the full range 0,1,..9. This method is definitely inferior to calling the integers from 1 to 6, and the definition would not hold.

4. The second point is that we must consider the *systematic* performance of predictors, presumably by continued repetition of the extended experiment which consists of matching a prediction with the randomizer's outcome. For any finite number of experiments the scores of various predictors will generally differ, and the definition would never apply unless we are prepared to disregard differences that are not statistically significant*).

5. Thirdly the performance of a predictor is measured by the number of correct predictions. In order to define these unambiguously, the range of continuous variables must be divided into discrete intervals. This is largely a matter of conventional precision: in practice the prediction that retail prices will rise by 7 percent is readily understood to mean 'between 6.5 and 7.5 percent'.

By a natural extension of this criterion the *systematic* performance of a predictor would correspond to the limiting frequency of successes in an ever increasing number of repetitions. We return to this point shortly.

6. At this stage we change the frame of reference. We no longer envisage prediction as a means for defining the probabilistic model, as Hemelrijk does, but consider predictions that are prompted by practical or commercial considerations instead. Such predictions are used as a substitute for

*) It is a moot point whether one may use such fairly advanced statistical concepts in constructing the basis of the probability model, but we do not pursue this question further.

information when decisions have to be taken in uncertain situations. We retain the first two points that have been discussed - viz. the use of all information and the need for a systematic view of prediction - but not the third which identifies prediction performance with the frequency of success.

7. We thus move to *decision theory*. Predictions are no longer classified as correct or erroneous, but it is acknowledged that some errors are worse then others. This is not so in games of chance that offer 'all or nothing', but it certainly applies to weather forecasts or to economic predictions that lead to action by farmers, businessmen or the government. The central variable is the prediction error; if this is zero the correct decision and the appropriate action are taken, but any nonzero error will lead to some *loss*, and this loss is generally related to the size of the error.

The prediction is thus part of a sequence that runs as follows:

. predict;

. act as if the prediction were true;

. suffer some loss due to prediction error.

The best predictor is the predictor that causes least loss. As before, however, we cannot go by a single case or even by any finite number of repetitions. The prediction error is a random variable, and so is the loss, if only because the predictions are matched against the outcome of a statistical experiment. Once more we must decide on a systematic characteristic. The standard practice is to take the *expected value* of the loss associated with a predictor as a measure of its performance.

8. This passage to mathematical expectations is often justified intuitively by referring to the average over a large number of repetitions of the extended experiment, which now consists of the sequence we have just given. Upon suitable definition this yields identically distributed losses, and their average will converge to the expected loss by Khintchine's theorem.

This line of reasoning has the serious drawback that it raises the vexed question of repetition under identical conditions. The prediction of the 1978 unemployment rate, for instance, is so strongly affected by the unique circumstances of 1977 that a repetition of the experiment is inconceivable. The prediction of a strong rise of unemployment would moreover be invalidated by remedial action by the public authorities. We cannot observe the prediction error of such self-defeating prophecies, and even if they are capable of repetition we cannot assess the corresponding loss, nor examine its long-term average.

We therefore prefer to introduce the expected loss as a criterion without this particular justification. It must be regarded as an axiom, or it can again be derived from other axioms that together form a reasonable framework for the utility evaluation of both certain and uncertain events (VON NEUMANN & MORGENSTERN 1947).

9. In order to complete the model we must still specify the loss as a function of prediction error. We do not even discuss the alternatives that readily come to mind and at once opt for the square. This can be regarded as a quadratic approximation in the neighbourhood of the point of zero prediction error and zero loss, further restricted by imposing symmetry. It follows that the expected loss associated with a predictor is proportional to its mean square prediction error.

If this much is accepted, several conclusions can be drawn. First, the best predictor is the expected value of the random variable under consideration. Secondly, if this is known (as in the case of a randomizer with given N), there is still a nonzero loss associated with the extended experiment, and this minimum is proportional to the variance of the predicament. This irreducible term reflects the unpredictable character of all nondegenerate random phenomena, and thus confirms Hemelrijk's view that random variables may be defined by the impossibility of prediction.

In the third place completely specified experiments with known mean outcome are restricted to classroom usage. Whenever predictions are called for in practice, they must be obtained from empirical observations that are themselves random. We thus predict or estimate the nonrandom mean by a random variable, and this of course adds to the expected loss by a term that is again proportional to the mean square error. Hence least mean square error predictors are at a premium.

10. The quest for estimators and predictors which have minimum mean square errors, or which are at least unbiased or consistent, is a major field of statistical theory. But in the real world of weather forecasts or the

somewhat less real world of social and economic predictions estimation is only part of the problem. In its most elaborate form prediction proceeds in three stages, viz.

. the *specification of a model* and the identification of variable parameters;

. the *estimation* of these parameters;

. the *determination of additional information* which is needed for the particular application in view.

Examples may once more be taken from Hemelrijk's paper. The statement that a six-faced die is cast constitutes a model specification. Estimation and hypothesis testing is involved in determining whether the die is true or biased. The statement 'it is not six' would constitute additional information that conditions the prediction.

11. We now consider some practical examples against the background of this scheme. Although we occasionally refer to least squares regression we pay very little attention to estimation; it can safely be left to the textbooks. There are no textbooks of model specification nor of the determination of additional information, yet both present major obstacles.

Several examples are not up to scientific standards. The three stages which we have distinguished inextricably merge, vital assumptions are not clearly stated, and the underlying theory is not explained. The conditions of practical prediction should however be borne in mind. It is a dire necessity, rarely undertaken for pleasure, and an unrewarding activity. Its practitioners are on the whole despondent and apologetic. If anyone can do better, let him try.

12. The statement that real national income will grow by 3% per annum from 1980 to 2000 is a *prophecy*. This is an apparently unconditional prediction without any formal justification by a model or estimation. The predictor's method is introspection, nurtured by experience and perhaps tempered by discussions which will elicit some, but not all, of the implicit assumptions.

13. Prophecies offer a primitive example of the *historical methods*, whereby predictions are generated entirely from past observations of the variable concerned. There is no explicit causal model and the predictions

therefore are not conditional upon some outside variable.

This description covers a wide range of methods of varying sophistication. The simplest example is the prediction that things will remain as they are. If the same reasoning is applied to first differences, trend fitting and other types of extrapolation are in order. Random disturbances must be added to these mechanistic processes in order to permit an interpretation of the observed values, and if these random variables are themselves incorporated into a dynamic process matters rapidly become quite complicated. At this extreme we find time-series analysis as practised by Box and Jenkins. The two stages of model specification and parameter estimation are fully developed; by the nature of the method, the need for the third stage does not arise. There are no outside variables that must be determined before the prediction can be made (BOX & JENKINS 1970).

By their unconditional character time series predictions and prophecies perform the same function. The former draw nowadays heavily on the computer, whereas the latter are based on human reason and intuition. There is a tendency to use the former for short-term purposes - i.e. when the prediction period is short relatively to the observation period, in other words when we have long series of observations - and the latter for long-term prediction. It is not clear what underlying common model would explain this preference.

14. *Analogue* is used, together with other techniques, in preparing weather forecasts. The distribution of atmospheric pressure over a region or the development of the atmospheric circulation over a recent period is coded and then matched against past records by computer or by expert opinion. The subsequent weather conditions observed in similar cases then provide a basis for the forecast (SCHUURMANS 1973).

This is a clear example of the historical method of prediction, and the underlying model to the effect that similar conditions will produce similar results is clear*).

15. All historical predictions illustrate the basic assumption of prediction that certain elements are constant and will occur in the future

---

*) The technique is regarded as commonplace, and there is little written documentation available.

as they have occurred in the past. This may apply to the phenomenon considered itself, or to its rate of change, or again to the implicit causal mechanism in the case of the weather. The same basic assumption is of course made when causal models are used, and the future course of events follows from relations that have been observed in the past.

Historical prediction moreover ensures that the prediction is *feasible* in the sense that the same state of affairs (or something very much like it) has already occurred in the past. This rules out disastrous errors of model specification as in the case of a six-faced die that is mistaken for a random digit generator. In the case of composite predictions like weather forecasts or predictions of social and economic conditions it also ensures that the various elements are not mutually inconsistent. In the field of social phenomena especially we know so little of their interrelations or even of their mutual compatibility that there is a natural tendency to stay close to observed situations. Exactly the same argument explains why time-series analysts largely confine their models to stationary series with a constant variance.

16. It will be clear that the basic assumption of some sort of continuity of the real world is a matter of belief. Some social scientists have argued that in their field prediction is impossible since regularities and relations that have been observed may no longer hold in the future. In the same vein it is observed that social phenomena are unpredictable because they depend on human behaviour (HENNIPMAN 1945, p. 449).

It must be conceded that because of the absence of experiments the limits of social developments have never been explored; in this field the inconceivable may well happen. On the other hand such reservations are fruitless when one is called upon to provide a prediction. Social phenomena and natural phenomena do differ in this respect, but this is not a matter of principle but merely the reflection of the vast difference in knowledge of these two fields. Whenever a relation is well established in the social sciences (and such cases are known), the objection that it may no longer operate in the future is no longer heard, and there is a general feeling of relief that there is at least something we can use.

17. *Causal methods,* however simple, are generally held to be superior to the historical methods; but here also some unexpected problems may arise,

and disturb the neat scheme of three orderly stages that we have introduced above. A recent example is the backward extension of the time series of the Dutch winter climate by de Vries (DE VRIES 1977). Annual series of winter temperature have earlier been constructed from 1735 onwards by Labrijn (LABRIJN 1945), and these overlap the period from 1633 to 1839 that horse-drawn boats were commonly used as public transport between Amsterdam, Haarlem and Leyden. The ledgers of the carriers survive and show for each year how many scheduled trips had to be cancelled because of ice. De Vries establishes a relation of winter temperature (Y) to the number of cancellations (X) by least-squares regression and uses it to obtain estimates of winter temperature in the preceding century.

This example has some interesting aspects. At first sight the model specification is wrong since it is evident that Y is the cause of X, and not the other way around. The regression equation is the wrong way around, and from the viewpoint of estimation inappropriate. Yet it may be shown that the prediction of expected winter temperature, conditional upon the observed number of cancellations, has at least some desirable statistical properties (JOHNSTON 1972, p.290).

The second point is that in this case the value of the conditioning variable is known, and no uncertainty attaches to this information. When it comes to a prediction of the future we are not so fortunate.

18. The last point constitutes the major weakness of causal models as predictors. Consider for instance a survey of the drinking habits of individuals as a function of their income and the nervous stress to which they are subjected. In a least squares regression analysis of such data both variables may considerably contribute to a systematic explanation of alcohol consumption, and if they are orthogonal (as may well be the case) these contributions are additive. The individual variation in drink may therefore be split up in three separate components, viz. variation due to income differences, variation due to stress differences, and residual variation which must be regarded as random in the absence of further explanatory variables.

When it comes to prediction, however, this analysis is of little help unless we know the values of the conditioning variables. If the drinking habits of a given individual are at issue, the prediction will be much reduced in variance if we know his income and his nervous stress;

but if we have to predict these variables in turn, the analysis does not contribute to the prediction at all, and it is easy to show that we arrive at exactly the same result that would follow from a survey that is restricted to drink consumption alone. In the absence of additional information about the conditioning variables causal models yield exactly the same predictions as historical methods.

19. In the example just given the immediate solution is of course to find out the income, or the stress, or both of the individual concerned; but then we may as well establish his drinking habits directly. If different predictors are to be fairly compared such cheating must be avoided, and the only way to do so is to make them predict a future event. I am convinced that this is the main reason that even classroom examples require predictions to be made before the die is cast.

In the actual practice of weather forecasts and economic predictions the need for prediction arises of course precisely because the future course of events is at stake. Yet causal models are used widely in both fields. We shall briefly examine why this is so.

20. Ideally the problem may be resolved by a *dynamic* model specification which explicitly allows for a time lag between cause and effect. Since it is firmly believed that effects follow cause, it should be possible to link the future course of events ultimately entirely to present conditions. For this classical view of the final end of scientific analysis we may again return to Laplace. This idea is nowadays dismissed on theoretical grounds, and there is no question about it that it is of no practical important whatever. Even the most elaborate economic models, for instance, do not constitute a completely dynamic system. In this case a major reason is that a complete model would require a complete coverage of the world economic system, as well as an extension to political and social phenomena. In practice causal models are restricted to a single domain.

The result is that the actual predictors in this field are usually a mixture of causal models and of predictors of the conditioning variables based on even the crudest of historical methods. The fact that the latter are usually introduced as hypotheses and not as predictions does not alter the fact that they are materially the same. Thus the provision in the early post-war economic forecasts that they were conditional on the assumption of no worldwide armed conflict meant in fact that no war was expected. The difference in wording served only to distinguish this prophecy from the conditional causal model prediction that followed.

21. From this point of view, then, causal models are superior to historical predictors only in so far as they rely on present and known conditions; to the extent that these are supplemented by prophecies and historical predictions of future conditions, the result is no better than these basic ingredients. It is true that causal models ensure the mutual consistency of the composite predictions that they yield, but as we have seen most historical methods implicitly contain provisions to the same end.

22. In conclusion it would seem that the various predictors vary in their degree of explicit elaboration rather than in fundamental methodological traits. As the underlying models are uncovered their similarity appears, and it must regretfully be concluded that we are as ignorant as ever of the future.

REFERENCES

BOX, G.E.P. & G.M. JENKINS (1970), *Time series analysis,* San Francisco: Holden-Day.

HEMELRIJK, J. (1978), *Rules for building statistical models,* This volume, p.

HENNIPMAN, P. (1945), *Economisch motief en economisch principe,* Amsterdam: Noord-Hollandsche Uitgevers Maatschappij.

JOHNSTON, J. (1972), *Econometric methods,* New York: McGraw Hill.

LABRIJN, A. (1945), *Het klimaat in Nederland gedurende de laatste twee en een halve eeuw,* K.N.M.I., Utrecht.

VON NEUMANN, J. & O. MORGENSTERN (1947), *Theory of games and economic behaviour,* Princeton, University Press.

SCHUURMANS, C.J.E. (1973), *A 4-year experiment in long range weather forecasting, using circulating analogues,* Meteorologische Rundschau, 26 (1973), p. 1-4.

DE VRIES, J. (1977), *Histoire du climat et economie,* Annales, Economies, Sociétés, Civilisations, 2 (1977), p. 198-226.

INTERPRETING INTUITIONISTIC LOGIC

D. van Dalen

Although formalizations of intuitionistic logic were put forward only in the late twenties, there must have been a certain amount of agreement on the meaning of the connectives in intuitionistic mathematics. In Brouwer's writings the use of logical symbolism is almost absent, his dislike for the formal aspects of mathematics even induced him to use a cumbersome private terminology (e.g. "absurdity of the absurdity of .."). So it would be too much to expect a clear cut notion of meaning of the logical notions; one has to read between the lines to discover Brouwer's views on such notions as "negation", "implication", "disjunction".

The development of intuitionistic logic was, in a sense, faithful to the spirit of the times. First, there were some explorations along formal lines in the papers of Glivenko and Kolmogoroff, culminating in Heyting's formalization (for the history of that part of intuitionistic logic, see Troelstra's contribution to *Two decades of Mathematics in the Netherlands*).

Of no less importance for the study of intuitionistic logic was Heyting's interpretation of the logical constants, now known as the *proof-interpretation*. In 1930 he published his *Sur la logique intuitioniste*, in which he propounded his interpretation, mainly to clarify the properties of the negation. The interpretation is further elaborated in [13] and [14].

Heyting's point of departure is, following Brouwer: "A holds" (or, A is affirmed) if we have a proof for A, where a "proof" is understood to be a construction of a suitable sort. In terms of proofs the logical constants can be assigned the following meaning:

a proof p of $A \wedge B$ is a pair of proofs p_1, p_2 such that p_1 proves A and p_2 proves B,

a proof p of $A \vee B$ consists of a construction which selects one of the formulas A and B and yields a proof for that particular formula,

a proof p of $A \rightarrow B$ is a construction which assigns to each proof q of A a proof $p(q)$ of B, plus a verification that p indeed satisfies these conditions,

a proof p of $\neg A$ is a proof of $A \to \bot$, where $\bot$ is some contradiction,

a proof p of $\forall x\, A(x)$ is a construction which assigns to each object a (of the domain of discourse) a proof $p(a)$ of $A(\bar{a})$, plus a verification that p indeed satisfies these conditions,

a proof p of $\exists x\, A(x)$ is a construction which selects an object a (of the domain of discourse) and yields a proof of $A(\bar{a})$.

Note that negation has been reduced to implication, quite in accordance with Brouwer's practice. The hard cases are here, and in most interpretations, implication (negation) and universal quantification. In the case of the implication we are faced with a construction which operates on construction (so, possibly on itself), which lends an impredicative character to the whole interpretation. The extra clause, requesting a verification that p does what it should do, is not particularly surprising. Anybody who has written programs, or built Turing machines, knows that there is always the supplementary task of proving the correctness. Fortunately, this correctness proof is of a simpler nature (although it may be tedious); in the particular case of intuitionistic logic this is so because of the fundamental assumption (Kreisel) that "p is a proof of A" is decidable.

The proof-interpretation contains in a nutshell the underlying idea of many other interpretations. It is extremely useful as a heuristic guiding principle in intuitionistic logic, and it probably is the most realistic representation of actual practice in intuitionism.

KREISEL has at several occastions, [24], [25], tried to find a proper codification of Heyting's proof-interpretation. In one direction there is his proposal for a theory of constructions, worked out by GOODMAN [12], in another direction there is the project of a suitable arithmetization of the interpretation, which so far has born no (significant) fruits (for a result on the implicational fragment cf. [45]).

Already in 1932 Kolmogoroff presented a related interpretation, in which formulas are interpreted as problems. The clauses for the various connectives are similar to Heyting's clauses, e.g. a solution of the problem $A \to B$ consists of a method for finding a solution of problem B, given a solution of problem A.

Both Heyting and Kolmogoroff's interpretation were fundamental in nature, i.e. they were intended as the "true" meaning of intuitionistic logic. Of the two, clearly Heyting's interpretation is foundationally the more important one. It is closest to Brouwer's conceptions of consequence, negation, etc., and it lends itself to technical elaborations.

Semantics may also play a less lofty role, it can be used as a tool in the study of formal systems. In this sense it was already used by Heyting in his basic paper *Die formalen Regeln der intuitionistischen Logik*, to establish independence of the axioms. This kind of semantics - ad hoc truth tables - certainly had no claim to capture the meaning of intuitionistic logical constants. The truth table method was already well-established by that time, among others through the investigations of Post, Bernays, Łukasiewicz and Tarski. The power of this method was convincingly demonstrated in 1935 by Jaskowski, who designed an increasing family of truth tables such that intuitionistic propositional logic was complete with respect to its join, i.e. derivability in Heyting's system corresponded exactly to truth in all of JASKOWSKI's tables [16].

In the early thirties Heyting's formalization of intuitionistic logic made itself felt. Those who, for whatever reason, did not feel like extracting the true intuitionistic logic from Brouwer's writings had a place to turn to. Here was a codification of intuitionistic logic by a regular student of Brouwer! Evidently this formalization could be relied upon. As a matter of fact all subsequent investigations, e.g. by Gentzen, Gödel, Jaskowski, Tarski, are based on Heyting's system.

The formal laws of intuitionistic logic soon suggested certain analogies in topology and lattice theory. TARSKI [40] and STONE [38] observed that the algebra of open subsets of a topological space under certain operations behaves in a manner, similar to intuitionistic propositional logic.

Define for open subsets A, B of a space X

$$A \wedge B := A \cap B; \quad A \vee B := A \cup B; \quad A \to B = B \cup \mathrm{Int}(A^c); \quad \neg A := \mathrm{Int}(A^c),$$

where Int is the interior-operator. Then a simple calculation shows that all (substitution instances of) theorems of Heyting's system are valid, i.e. take the space X as a value. Tarski shows even more: Heyting's propositional logic is complete for the topological interpretation, i.e. validity in all topological spaces is equivalent to derivability in Heyting's system.

In analogy to Boolean algebra an algebraic version of intuitionistic logic was studied by various authors, Tarski, Birkhoff, McKinsey, Stone and others (cf. [34]). This generalization of Boolean algebra went under several names: Brouwerian algebra, closure algebra, interior algebra, pseudo-Boolean algebra and Heyting algebra. Under the influence of the topos theorists the name *Heyting algebra* seems to stick.

It took a while before Mostowski in 1948 extended the topological interpretation to predicate calculus. Mostowski showed the soundness of the interpretation (i.e. derivable formulas hold in all complete Heyting algebras) and the completeness was proved by Rasiowa and Henkin, independently.

In order to extend the interpretation to predicate logic one needs two more clauses for the quantifiers.

Let us assume that to each closed atom A a value $[\![A]\!]$ in $O(X)$, the algebra of open sets of X, is assigned, and that we have a given universe of individual objects, D. The values of sentences is inductively defined by:

$$
\begin{aligned}
[\![A \wedge B]\!] &:= [\![A]\!] \cap [\![B]\!]; \\
[\![A \vee B]\!] &:= [\![A]\!] \cup [\![B]\!]; \\
[\![A \to B]\!] &:= [\![B]\!] \cup \mathrm{Int}([\![A]\!]^c); \\
[\![\neg A]\!] &:= \mathrm{Int}[\![A]\!]^c; \\
[\![\forall x\, A(x)]\!] &:= \mathrm{Int} \bigcap\{[\![A(\bar{d})]\!] \mid d \in D\}; \\
[\![\exists x\, A(x)]\!] &:= \bigcup\{[\![A(\bar{d})]\!] \mid d \in D\}.
\end{aligned}
$$

Scott observed in [36a] that the topological interpretation can be motivated in a Heyting-like manner.
For an extensive treatment of this interpretation see [34].

Whereas Heyting and Kolmogoroff tried to capture the true meaning of the logical constants, the algebraico-topological school of logicians were of a more modest (indifferent) sort. They were only interested in the formal properties of the logic and the interpretation was simply a tool. The foundational value of these researches was rather doubtful to intuitionists, if only because in the metamathematics classical (transfinite) reasonings were freely used. A closer investigation by KREISEL, [22], and SCOTT, [36], in 1958, 1957 showed that the completeness of propositional logic with respect to the topological interpretation could be shown in an intuitionistic metamathematics.

In the meantime some new interpretations of intuitionistic logic had been proposed. E.W. BETH introduced in 1956 a class of models, now known as Beth models [1], for which he intended "to prove that these calculi are complete with respect to intuitionistic arguments". In this claim Beht was somwhat overly optimistic as KREISEL showed later [23] [26], but Beth was rather close to success as later developments showed. Beth models are, as a matter of fact, a special kind of topological models, but they have a better heuristic motivation, It is a whim of history that very soon Beth's semantics was overshadowed by a similar semantics, that of Kripke. Even to the extent that genuine instances of Beth's semantics were called after Kripke (Kripke-Joyal).

Before we go on with this kind of "semantic" interpretation, let us turn to interpretations of a more proof-theoretical nature. Kleene, who was the creator of recursion theory as a theory in its own right, conjectured that there should be a fairly close connection between recursion theory and Brouwer's intuitionism. Kleene exploited the idea that statements, e.g. $\exists x\, A(x)$, are incomplete, in the sense that an element a has to be indicated such that $A(\bar{a})$. Restricting himself to arithmetic, Kleene undertook to code the extra information needed to "complete" a statement A in a natural number (for an account of the development of this idea, see [19]). He expressed this by saying that "the number e *realizes* A".

Realizability for the simple case of Heyting's arithmetic, HA, is inductively defined as follows: (where e is the coding of the pair $((e)_0,(e)_1)$):

$e \;\underline{r}\; A$	if A is a closed, true atomic statement,
$e \;\underline{r}\; A\wedge B$	if $(e)_0 \;\underline{r}\; A$ and $(e)_1 \;\underline{r}\; B$,
$e \;\underline{r}\; A\vee B$	if $(e)_0 = 0$ and $(e)_1 \;\underline{r}\; A$, or $(e)_0 \neq 0$ and $(e)_1 \;\underline{r}\; B$,
$e \;\underline{r}\; A\rightarrow B$	if for all n with $n \;\underline{r}\; A$ $\{e\}(n) \;\underline{r}\; B$
$e \;\underline{r}\; \exists x\, A(x)$	if $(e)_1 \;\underline{r}\; A(\overline{(e)_0})$,
$e \;\underline{r}\; \forall x\, A(x)$	if for all n $\{e\}(n) \;\underline{r}\; A(\bar{n})$.

This definition is perfectly acceptable from an intuitionistic point of view, and we may even note a certain similarity with the proof-interpretation, e.g. in the case of the implication: there is an algorithm (even a recursive one) that transforms any information needed for A in information needed for B. Let us note, however, that Kleene explicitly states that the proof-interpretation did not help him to find the correct realizability clause for the implication [19].

Realizability differs in nature from semantic interpretations on several accounts. For one, it essentially uses the natural numbers, which makes it suitable for arithmetic (and extensions), but less so for general predicate logic. For another, it heavily relies on Church's Thesis: "each algorithm is a partial recursive function" (at least in motivation), which puts its stamp on the class of realizable ($\underline{r}$-valid) sentences. This appears, e.g., from the fact that the first order version of Church's Thesis is realizable. Hence the realizability interpretation is not faithful. For, assume that $\forall x \exists y\, A(x,y) \to \exists e(\{e\} \text{ is total} \wedge \forall x\, A(x,\{e\}x))$ is derivable in HA, then it is also derivable in Peano's arithmetic, PA, and so every definable set would be recursive, contradiction.

A reason for considering realizability as a part of proof theory is that it gives a kind of translation of arithmetic in arithmetic (using the clauses shown above), hence its usefulness for studying aspects of provability in arithmetics, cf. [19],[41]. Troelstra has shown that Kleene's original realizability is characterized by an extension of Church's Thesis, ECT_0, in the sense that A is realizable in HA iff it is a theorem of HA + ECT_0, [41] p.196. The notion of realizability has been generalized in various directions, e.g. to cover analysis. For an extensive treatment see [41]. One generalization, by LÃUCHLI [28], can be considered as a mixture of semantics and proof theory. Indeed, predicate calculus is complete for it.

GÖDEL's paper *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes* (1958), [11],[41], also gave rise to a number of interpretations of arithmetic (and extensions), mainly in arithmetic of all finite types (and extensions). Like realizability these so-called *functional interpretations* belong rather to proof theory than to semantics (like realizability, these interpretations are not faithful, in particular Gödel's Dialectica Translation changes the meaning of the connectives).

Let us now return to the semantic interpretations. In 1965 KRIPKE presented a new semantics for intuitionistic logic, [27], which turned out, in most respects, to be superior to the existing interpretations. Kripke provided a heuristic motivation for his models (cf. also [37]) which may be illuminating. Since it is most conveniently expressed in terms of knowledge, let us use the picture of the idealized mathematician M. Following Kripke we analyse the meaning of the logical connectives in terms of the knowledge of M at various points in time. We schematically represent the various possible stages of M's knowledge in time as a partially ordered set (usually a tree). Write $\alpha < \beta$ for "stage α (in time) precedes stage β", and $\alpha \Vdash A$ for "M knows A at stage α". We suppose that M does not "loose" knowledge, i.e. if $\alpha \Vdash A$ and $\alpha \leq \beta$, then $\beta \Vdash A$.

Now we consider the logical connectives:

M knows $A \wedge B$ at α if he knows both A and B at α.

M knows $A \vee B$ at α if he knows A at α or if he knows B at α.

M knows $A \rightarrow B$ at α if at any future stage β his knowledge of A implies his knowledge of B.

M knows $\neg A$ if at no future stage β he will know A.

To handle predicate logic we also have to consider the universe of individuals, so we assume that M not only collects knowledge, but also constructs (creates) individuals. The set of individuals constructed at stage α is denoted by $D(\alpha)$. We now continue:

M knows $\exists x\, A(x)$ at stage α if there is an individual $a \in D(\alpha)$ such that he knows $A(\bar{a})$ at stage α.

M knows $\forall x. A(x)$ at stage α if he knows $A(\bar{a})$ for all $a \in D(\beta)$ for all future stages β.

Note that the interpretation takes into account the potential character of M's state of knowledge and of the universe of individuals (in $\forall$, $\rightarrow$, $\neg$). For a proper inductive definition of the relation $\Vdash$ we also need a clause for the atomic statements (basic facts). We simply assume that $\alpha \Vdash A$ for atomic A if M knows (has evidence for) A at stage α, in a sense not to be specified here (one might think of "having a construction for" in the case of arithmetic, or "insight based on introspection", etc.).

Let us now return to Beth's semantics and apply a similar consideration. In Beth's case M is more liberal, his knowledge at certain stages induces truth at certain other stages. M, so to speak, has more grip on the future. This appears, for instance, from the fact that the universe of individuals is fixed, no new individuals are created. So we have a fixed domain D. Let $\alpha \Vdash A$ denote "A is true" at stage α. The clauses for $\rightarrow$, $\neg$ and $\wedge$ are identical to the above clauses. Consider the remaining clauses:

$\alpha \Vdash A$ for atomic A if M, starting at stage α, eventually will know A. To be precise, if for each totally ordered path P through α there is a $\beta \in P$ such that M knows A at stage β.

One may think of M conducting his research. At each stage there may be various ways to continue it in the future. Now $\alpha \Vdash A$, if, no matter how he continues his research, he will always find A. Similar clauses give us the truth of disjunctions and existential statements:

$\alpha \Vdash A \vee B$ if for every path P through α there is a β such that $\beta \Vdash A$ or $\beta \Vdash B$.

$\alpha \Vdash \exists x\, A(x)$ if for every path through α there is a β such that $\beta \Vdash A(\bar{a})$ for some $a \in D$.

Finally the interpretation of $\forall$ is given by a simpler clause, because of the constant domain:

$\alpha \Vdash \forall x\, A(x)$ if $\alpha \Vdash A(\bar{a})$ for all $a \in D$.

Prima facie there is no reason to prefer Kripke models to Beth models, but a little experimenting shows that Kripke models are technically more convenient. For example, Beth models with a finite set of states satisfy the law of the excluded middle; hence for even quite simple counter examples one needs infinite models. On the other hand, Kripke models have the finite model property for a considerable class of logics, i.e. if A is not derivable, then there is a finite Kripke counter model. Kripke demonstrated the superior flexibility by using his semantics for a proof of the undecidability of monadic intuitionistic predicate calculus.

As we already stated, there is a simple connection between Beth models and topological models. Consider a Beth model, with a tree as set of states. Let T be the topological space of paths of the tree and (the set of paths P passing through) nodes as basic opens. Define $[\![A]\!] = \{P \mid \exists \alpha \in P\ (\alpha \Vdash A)\}$, then we get a topological model.

There is a simple connection between Kripke models and Beth models: each Kripke (tree) model can be recursively transformed into a (tree) Beth model, [27]. The converse, however, does not hold, there is no algorithm to transform Beth models into elementarily equivalent Kripke models, [10]. This phenomenon relates to the problem of complexity of Kripke (or Beth) models. It is a classical result that classical predicate logic is complete for models, definable by Δ^0_2 formulae in arithmetic. By means of translation of Kripke semantics in classical semantics one easily obtains a similar result for intuitionistic logic and Kripke models, however with a detour via the completeness theorem. Recently TROELSTRA [44] gave a direct proof of this fact and some related matters, e.g. a decidable (consistent) theory has a recursive Kripke (Beth) modela and a Σ^0_1 theory has a Δ^0_2 Kripke model with a recursive underlying tree.

In the sixties the Kripke-Beth semantics had slightly run out of steam. Beth and Kripke (and Aczel and Thomason) had shown their semantics to be complete for Heyting's predicate calculus, but Kreisel had pointed out that these completeness proofs were not intuitionistically acceptable. In a paper with Verena Dyson he showed that one needed an extra principle: Markov's principle [6], and in [23] he showed, following Gödel, that completeness for internal interpretations (Dummett) implies Markov's principle. The matter is that, whereas classical logic has a unique intended interpretation, it is far from clear what (if any) is the intended interpretation of intuitionistic logic. We have argued that Heyting's proof-interpretation is the best candidate for an intended semantics. But so far no convincing codifications have been presented to give it a status comparable to the two-valued interpretation for classical logic. The internal interpretation is the intuitionistic analogue of Tarski's semantics: Interprete statements in an inhabited domain, with relations (and operations) as in classical models, using intuitionistic reasoning at the meta level. The Gödel-Kreisel result tells us that completeness for this semantics can only be obtained under Markov's principle:

$$\neg\neg\, \exists x\, A(x) \rightarrow \exists x\, A(x)$$

for primitive recursive A.

In a later paper KREISEL even showed that for the internal interpretation completeness conflicts with Church's Thesis, [26],[2]. This seemed to be the end of semantics in an intuitionistic meta theory.

On a more technical level, Kripke semantics had been fruitful in the domain of pure logic, but it barely influenced the subjects that by nature constitute the hard core of intuitionism, to wit arithmetic and analysis. So semantics seemed to be a nice pastime for logicians, but unsuitable for fundamental issues.

On both issues things changed in the seventies. In 1974 VELDMAN presented a modification of Kripke's semantics for which an intuitionistically correct completeness proof indeed could be given [46]. The main innovation consists of allowing falsum ($\perp$) to be true at some stages. The technical details are beyond this survey, for detailed expositions see [5], [42], H. DE SWART adapted the procedure to Beth models [39].

Let us now briefly consider the impact of semantic methods on arithmetic and analysis. In his thesis (1972) SMORYNSKI considers Kripke models of HA, and uses them to obtain proof theoretic results, cf. [37]. In view of the fact that proof theory has always been a rather exclusive field, well-suited to the connoisseur of the finer details, it is surprising how elegantly and smoothly Smorynski managed to obtain significant results.

Instead of going into the model theory for first-order arithmetic, let us move on to second-order arithmetic. Already in 1968 SCOTT presented a model for intuitionistic real analysis in the framework of the topological models, [36a]. Noting that the "function-objects" determined Dedekind Cuts in the rationals, when one fixes a point in the underlying topological space X, he showed that the reals in the model are precisely the continuous maps from X to $\mathbb{R}$. By considering the sheaf of continuous functions with open domains one is led to a generalization of models over a sheaf, [8], [33], [35]. Scott's model demonstrated that at least a model theoretic approach to second-order arithmetic was feasible. Moreover, the model had, intuitionistacally speaking, realistic properties, e.g. it satisfied Brouwer's continuity theorem. In 1973 JOAN MOSCHOVAKIS published a topological model for an intuitionistic theory of choice sequences [32]. She had adapted Scott's approach by considering continuous mappings from Baire space into itself instead of real-valued continuous mappings. The model theoretic approach proved, so to speak, complementary to the proof theoretic approaches. Whereas the proof theoretic interpretations (e.g. realizability) introduce, as a rule, a certain measure of constructiveness, the semantic models tend to move the other way. In particular Moschovakis' model satifies Kripke's schema (which rules out Church's Thesis). From the properties of the model Moschovakis concluded that intuitionistic analysis (on the basis of Kleene's system *FIM* [20]) was consistent with relativized dependent choice (RDC), weak continuity without parameters (WC^c) and WC!, i.e.

$\forall\xi\exists!x\, A(\xi,x) \rightarrow \forall\xi\exists!x\exists y\forall\eta(\bar{\xi}y = \bar{\eta}y \rightarrow A(\eta,x))$, bar-induction (BI) and Kripke's schema (KS)

At about the same time VAN DALEN adapted Scott's model to second-order arithmetic with species variables, *HAS* [3]. This model had, roughly speaking, the same features as Moschovakis' model. Instead of continuity (WC) it satisfied a corresponding version UP! of Troelstra's *uniformity principle* UP, i.e.

$\forall X\exists x\, A(X,x) \rightarrow \exists x\forall X\, A(X,x)$.

A Kripke model for *HAS* was presented by Smorynski. He just took a tree with a constant numerical domain N, and with species given by "growing" families of subsets of N. Again the superior flexibility of the Kripke models was demonstrated in a paper by DE JONG & SMORYNSKI [17]. Smorynski's model is a bit less intuitionistic, in the sense that it satisfies Markov's Principle (MP) and the independence of premiss principle (IP_0). Moreover, this model, and all second-ordermodels mentioned in this survey, have a classical first-order part, i.e. for arithmetical formulae the excluded third holds. In [17] it is shown (among other things) that:

(i) HAS has the existential definability properties for the numerical and species cases: $\mathit{HAS} \vdash \exists X\, A(X) \Longleftrightarrow \exists S\, \mathit{HAS} \vdash A(S)$;

(ii) HAS is closed under the extended uniformity rule (EUR):
$\mathit{HAS} \vdash \forall X(\neg A(X) \to \exists y\, B(X,y)) \Rightarrow \mathit{HAS} \vdash \exists y \forall X(\neg A(X) \to B(X,y))$;

(iii) HAS is closed under the extended rule of Church (ECR):
$\mathit{HAS} \vdash \forall x(\neg A(x) \to \exists y\, B(x,y)) \Rightarrow \mathit{HAS} \vdash \exists e \forall x(\neg A(x) \to B(x,\{e\}(y)))$.

In 1975 VAN DALEN adapted the usual Beth models to construct a model of a theory of choice sequences [4]. In fact the model turned out to be isomorphic to Moschovakis' model. In addition to the results present in [32], it was shown that general weak continuity fails in the model. The Beth model technique, however, turned out to be well-suited for varying the class of choice sequences. Using Cohen forcing a model of the theory of lawless sequences (LS) was obtained, together with some independence results. Although the consistency of LS was since long established, KREISEL, cf. [42], it was highly gratifying to have at least a concrete model of it. In the same paper an interpretation of the theory of the creative subject was obtained, thus settling the consistency problem for it. So far all of the above models were studied in a classical metamathematics.

TROELSTRA has taken up this matter in his *Choice Sequences* [42], where he rebuilt the Moschovakis' model using an observation of Kreisel, that validity in the particular topological model should be equivalent to intuitionistic validity in all universes U_α (of continuous projections of α (lawless)). The results here are, as to be expected, essentially weaker than in the classical approach. In the case of LS, however, Troelstra found an intuitionistically acceptable model construction that yielded all the results obtained by the (classical) forcing method. His method is based on his earlier observation, that one lawless sequence codes a (countable) universe of sequences satisfying LS, cf. [43],[42]. So, given a lawless sequence we can construct intuitionistically acceptable models of subclasses of choice sequences. Indeed, lawless sequences seem to be prominent in the foundations of intuitionistic mathematics, as appears, for instance from their role in the various completeness theorems (cf. [5],[42]).

Recently S. Weinstein used a Kripke model approach in order to study analysis with sequences. His [47] establishes a number of metamathematical results e.g. DP and ED for some systems and also the independence of Kripke's strong schema from the weak version KS^-.

Intuitionistic logic has lately found applications in a, at first sight rather unlikely field: category theory. The fact is that certain categories, namely topoi, reflect higher-order intuitionistic logic. As an intermediate step let us have a look at Fourman and Scott's sheave models. In Scott's original model for intuitionistic analysis, reals were interpreted as continuous mappings from Baire space into $\mathbb{R}$. A next step is: consider the sheaf of continuous maps from opens of a topological space X into $\mathbb{R}$. The sections behave like reals on there domains. So, it seemed reasonable to consider relations and operations "locally". In particular each element could be assigned a "domain of existence", or "extent": Ea := the domain of the section a. Generalising this idea to arbitrary sheaves Fourman and Scott designed a logic with an existence predicate and partial elements, [7]. This innovation in logic is plausible from an intuitionistic point of view. It is well-known that certain operations, e.g. the inverse on $\mathbb{R}$, are by nature partial. Classically this is not particularly interesting since every operation can be made total. For constructivists partial operations are, however, essential. So this logic draws the correct constructivistic conclusion by making all elements partial. Fourman gave in his thesis a detailed analysis of the sheave semantics, including higher-order logic and completeness theorems. Scott observed that this logic is the ideal framework for a description operator: $Ix \cdot A(x)$ (the unique x with the property A). Independently category theorists had discovered the need for intuitionistic logic. In topoi (categories with finite products, and a subobject classifier) one can mimick a good part of ordinary set theory; put otherwise the category of sets is paradigmatic for topoi. In particular a topos contains an object Ω, which can be viewed as the collection of truth values of the topos. In classical set theory Ω consists of the two truth values 0 and 1. In general, however, Ω turns out to be a complete Heyting algebra. Hence "the logic" of a topos is in general intuitionistic, cf. [8]. There is an immediate connection with the logic of sheaves: the category of sheaves (over a space X) is a topos, and its Ω is the algebra of open sets of X. In topos theory many familiar notions ty up. The reinterpretation of logical connectives given by Beth and Kripke, is in a natural way present in topoi. There is even a bonus for using this "internal logic" of a topos; whereas products, monics, equalizers, etc. are defined by some universal property in the language of category theory, one can in the internal logic characterize these concepts in a straightforward, set-like, way, e.g. $A \overset{p}{\leftarrow} C \overset{q}{\rightarrow} B$ is a product if $\forall x \in A\ \forall y \in B\ \ \exists! z \in C\ \ (p(z) = x \wedge q(z) = y)$ holds in the topos (with topos-semantics). The initiator and main creator of the category-logic school is W. Lawvere, who not only studied categories with an open mind for its logical aspects, but also pointed out that topoi and their logic transcend the normal set theory by codifying the concept of "variable set", known to us already from the models of Smorynski and Van Dalen, and inherently present in intuitionism. Categorical logic (or logical category theory) is a fast exploding chapter of mathematics. The reader may consult [8],[30] or the forthcoming proceedings of the Durham conference on *Applications of Sheaf Theory to Logic Algebra and Analysis*.

The usefulness of these abstract methods has been demonstrated, e.g. by MULVEY [33] who obtained traditional algebraic results by using intuitionistic proofs in the framework of the sheaf interpretation. ROUSSEAU [35] aplied similar techniques in the case of complex analysis. In quite another direction Fourman and Hyland used the sheaf interpretation over complete Heyting algebras to construct a model of intuitionistic analysis without the fan theorem [9].

The foregoing shows that the interpretations of intuitionistic logic have come a long way since Heyting's first proposals. Nonetheless, the basic problem remains: to give a satisfactory description of the intended interpretation of intuitionistic logic.

REFERENCES

[1] BETH, E.W., *Semantic construction of intuitionistic logic,* Meded. Kon. Ned. Ak. van Wet., vol. 19, no 11, 1956.

[2] DALEN, D. VAN, *Lectures on intuitionism* [31], pp.1-94.

[3] DALEN, D. VAN, *A model for* HAS, Fund. Math. (82), 1974, pp.167-174.

[4] DALEN, D. VAN, *An interpretation of intuitionistic analysis,* Ann. Math. Logic, 1977.

[5] DUMMETT, M., *Elements of intuitionism,* Oxford, 1977.

[6] DYSON, V.M. & G. KRIESEL, *Analtsis of Beth's semantic construciton of intuitionistic logic,* Stanford Techn. Report no. 3, 1961.

[7] FOURMAN, M., *Connections between category theory and logic,* Thesis 1974, Oxford.

[8] FOURMAN, M., *The logic of topoi,* Handbook of Math. Logic (ed. J. Barwise), Amsterdam, 1977.

[9] FOURMAN, M.P. & J.M.E. HYLAND, *Sheaf models for analysis,* (to appear).

[10] GABBAY, D., *Completeness properties of Heyting's predicate calculus with respect to R.E - models,* J.S.L. (41) 1976, pp.81-94.

[11] GÖDEL, K., *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes,* Dialectica (12) 1958, pp.280-287.

[12] GOODMAN, N.D., *A theory of constructions equivalent to arithmetic* [18], pp.101-120.

[13] HEYTING, A., *Sur la logique intuitioniste,* Ac. Royale de Belgique, Bull. classe Sc. 16 (1930), pp.957-963.

[14] HEYTING, A., *Die intuitionistische Grundlegung der Mathematik,* Erkenntnis 2 (1931), pp.106-115.

[15] HEYTING, A., *Intuitionism, An introduction,* Amsterdam, 1956.

[16] JAŚKOWSKI, S., *Recherches sur le système de la logique intuitioniste,* Actes Congr. Int. Phil. Sci. VI, Paris 1936, pp.58-61.

[17] JONG, D.H. DE & C. SMORYNSKI, *Kripke models and the theory of species,* Ann. Math. Logic (9) 1976, pp.157-186.

[18] KINO, A. MYHILL, J. & R.E. VESLEY (ed.), *Intuitionism and proof theory,* Amsterdam, 1970.

[19] KLEENE, C.S., *Realizability: a retrospective survey* [31], pp.95-112 (1973).

[20] KLEENE, C.S. & R.E. VESLEY, *The foundations of intuitionistic mathematics,* Amsterdam, 1965.

[21] KOLMOGOROFF, A.N., *Zur Deutung der intuitionistischen Logik,* Math. Ztschr. (35) 1932, pp.58-65.

[22] KREISEL, G., *A remark on free choice sequences and the topological completeness proofs,* J.S.L. (23) 1958, pp.369-388.

[23] KREISEL, G., *On weak completeness of intuitionistic predicaet logic,* J.S.L. (27) 1962, pp.139-158.

[24] KREISEL, G., *Foundations of intuitionistic logic,* Logic Methodology and Phil. of Science, Stanford, 1962, pp.198-210.

[25] KREISEL, G., *Mathematical logic, Lectures on modern mathematics,* 3rd Ed. T.L. Saaty, New York, 1965, pp.95-195.

[26] KREISEL, G., *Church's Thesis: a kind of reducibility axiom of constructive mathematics* [18], pp.121-150.

[27] KRIPKE, S., *Semantical analysis of intuitionistic logic. Formal systems and recursive function theory,* Ed. J.N. Crossley, M.A.E. Dummett.

[28] LÄUCHLI, H., *An abstract notion of realizability for which intuitionistic predicate calculus is complete* [18], pp.227-234.

[29] LAWVERE, F.W., *Variable quantities and variable structures in topoi.*

[30] MAKKAI, M. & G.E. REYES, *First-order categorical logic,* Berlin, 1977.

[31] MATHIAS, A.R.D. & R.H. ROGERS, *Summer school in mathematical logic,* Berlin, 1973.

[32] MOSCHOVAKIS, J.R., *A topological interpretation of second-order intuitionistic arithmetic,* Comp. Math. (26), 1973, pp.261-276.

[33] MULVEY, C., *Intuitionistic algebra and representations of rings,* Mem. Am. Math. Soc. (148), 1974, pp.3-57.

[34] RASIOWA, H. & R. SIKORSKI, *The mathematics of Metamathematics,* Warszawa, 1963.

[35] ROUSSEAU, C., *Théorie des topos et analyse complexe,* Thèse, Montréal 1977.

[36] SCOTT, D.S., *Completeness proofs for the intuitionistic sentential calculus. Summaries of talks,* Summer Inst. Symb. Logic, Cornell Univ. 1957.

[36a] SCOTT, D.S., *Extending the top. interpr. I,II,* Comp. Math. (20) 1968, pp.194-210; [18], pp.235-255.

[37] SMORYNSKI, C., *Applications of Kripke models in* [41], pp.324-391.

[38] STONE, M.M., *Topological representations of distributive lattices and Brouwerian logics,* Časopis pro Ďestováni Mathematiky a Fysiki (1937)-'38), pp.1-25.

[39] SWART, H. De, *Another intuitionistic completeness theorem,* J.S.L. (41) 1976, pp.664-662.

[40] TARSKI, A., *Der Aussagenkalkül und die Topologie,* Fund. Math. (31) 1938, pp.103-134.

[41] TROELSTRA, A.S. (ed.), *Metamathematical investigation of intuitionistic arithmetic and analysis,* Berlin, 1973.

[42] TROELSTRA, A.S., *Choice sequences. A chapter of intuitionistic mathematics,* Oxford, 1977.

[43] TROELSTRA, A.S., *Appendix to D. van Dalen's "An interpretation of intuitionistic analysis",* Ann. Math. Logic, 1977.

[44] TROELSTRA, A.S., *Some remarks on the complexity of Henkin-Kripke models* (to appear).

[45] TROELSTRA, A.S., *The interplay between logic and math.: Intuitionism* (to appear).

[46] VELDMAN, W., *An intuitionistic completeness theorem for intuitionistic predicate logic,* J.S.L. (41) 1976, pp.159-167.

[47] WEINSTEIN, S., *Some Applications of Kripke Models to Formal Systems of Intuitionistic Analysis,* Ph.D. Thesis, 1977.

FUNCTIONAL INTERPRETATIONS OF HEYTING'S ARITHMETIC IN ALL FINITE TYPES

Justus Diller, Münster, Germany

Functional interpretations have proved to be a useful tool in proof theory. We are here concerned with "first-order" interpretations where the interpreting functionals are of finite type, e.g. modified realisability mr of KREISEL 1959 and Dialectica interpretation D of GÖDEL 1958 which was extended to classical analysis using bar recursive functionals by SPECTOR 1962. TROELSTRA 1973, III, §§ 4 to 8, gives many applications of mr and D.

We want to analyze common principles underlying these interpretations. This analysis starts with a natural deduction motivation for the $\wedge$-interpretation of D-Nahm 1974: Given a deduction Π of a prenex formula $\exists y \forall z\ B$ from an assumption $\exists v \forall w\ A$, we should be able to construct from Π, for any given v, an object y such that, for any z, every path in Π leading upwards to an occurence of the assumption $\exists v \forall w\ A$ gives us an object w which in this path is the reason for A to imply B; as different paths in Π may produce different objects w, the deduction Π as a whole only gives us a set W of objects (reasons) w such that Π contains a proof of B from assumptions A for all $w \in W$. In short, from Π we obtain a proof of

$$\forall v \exists y \forall z \exists \text{set } W(\wedge w \in W\ A \rightarrow B).$$

The size of the set W depends on the ramification of the deduction Π. D-Nahm 1974 considers only finite ramifications. As we shall see in § 3, decisions about the size of these "sets of reasons" W give special functional interpretations. As long as we leave this decision open, we are led to an extension HA^{ω}_{ϵ} of Heyting's arithmetic of finite types HA^{ω} by a fragment of set theoretical language and a restricted universal quantifier $\wedge w \in W$.

Such extensions and general functional interpretations for arithmetic and analysis have been investigated at Münster in the last years. STEIN 1976 and RATH 1978 give natural deduction systems for HA^{ω}_{ϵ} leading finally to a general and strong form of MINC' 1974 theorem: Given a closed deduction of a formula $\exists x\, B[x]$, normalization and all special cases of the general interpretation lead to the same term c, but different proofs of $B[c]$.

We give a Hilbert-style version of HA^{ω}_{ϵ} in § 1 which is fairly brief to state, easy to handle, but probably weaker than RATH's 1978 natural deduction system. In § 2, we introduce the general interpretation J using the above motivation for the translation $(A \to B)^J$, and we prove the characterization and interpretation theorem for J. Our J corresponds to I' of RATH 1978, Ch. I, section 4.2, and to K = 0 of STEIN 197?, § 2. In § 3, we show that known interpretations like mr, D,$\wedge$, and the n-interpretations of STEIN 1976, I, § 2 are special cases of J, by expanding arbitrary models of HA^{ω} to models of HA^{ω}_{ϵ} in which J is equivalent to the given special interpretation.

An extension of these results to systems of analysis is in preparation.

§ 1. HEYTING's ARITHMETIC OF FINITE TYPES WITH SET TERMS

We start with a formulation of the theory HA^{ω}_{ϵ}.

Inductive definition 1.1 of types.

1. o is a type;
2. If σ and τ are types, then $(\sigma\to\tau)$ is a type;
3. If τ is a type, then $\bar{\tau}$ is a type.

A type is called *linear*, if it is defined by 1 and 2 only; it is called a *set type*, if it is of the form $\bar{\tau}$.
$\rho\to\sigma\to\tau$ stands for $(\rho\to(\sigma\to\tau))$. If σ and τ are type tuples $\sigma_1,\dots,\sigma_m$ and $\tau_1,\dots,\tau_n$ resp., then $\sigma\to\tau$ stands for the type tuple $\sigma_1\to\dots\to\sigma_m\to\tau_1,\dots,\sigma_1\to\dots\to\sigma_m\to\tau_n$, and $\bar{\sigma}$ stands for $\overline{\sigma_1},\dots,\overline{\sigma_m}$.

DEFINITION 1.2. A *type structure* is a family $(M_\sigma \mid \sigma \text{ is a type})$ of non-empty sets as follows:

(i) $M_{\sigma\to\tau}$ is a set of (non-extensional) functions from M_σ into M_τ where different elements of $M_{\sigma\to\tau}$ may have the same graph $\subset M_\sigma \times M_\tau$;
(ii) $M_{\bar{\tau}}$ is a set of subsets of M_τ.

Similarly, a *linear type structure* is a family $(M_\sigma \mid \sigma \text{ is a linear type})$ of non-empty sets such that (i) holds.

DEFINITION 1.3. *Basic terms* are

1. the numeral O of type o;
2. the function symbol suc of type $o\to o$;
3. for any type τ, a recursor R_τ of type $\tau\to(o\to\tau\to\tau)\to o\to\tau$;
4. for any type τ, denumerably many variables $x^\tau, y^\tau, \ldots$ of type τ.

Inductive definition 1.4 of terms and their types.

1. Any basic term is a term of the same type;
2. If a and b are terms of type $\sigma\to\tau$ and σ resp., then a(b) is a term of type τ;
3. If a is a term of type τ and x^σ is a variable of type σ, then $(\lambda x^\sigma.a)$ is a term of type $(\sigma\to\tau)$;
4. If a is a term of type τ, then $\{a\}$ is a term of type $\bar{\tau}$;
5. If a and b are terms of type $\bar{\tau}$, then $a \cup b$ is a term of type $\bar{\tau}$;
6. If a and b are terms of type $\bar{\sigma}$ and $\bar{\tau}$ resp., and if the variable x^σ is not free in a, then $\bigcup_{x^\sigma \in a} b$ is a term of type $\bar{\tau}$.

Free and bound occurrences of variables and "b is substitutable for x in a" are defined as usual. $a_x[b]$ is the result of substituting b for every free occurrence of x in a. For this we write a[b], if confusion cannot arise.

A *functional* is a term without free occurrences of variables. A *numeral* is a functional O, 1 := suc O, 2 := suc 1, etc.

Inductive definition 1.5 of formulas.

1. If a and b are terms of the same type, then a = b is a formula;
2. If A and B are formulas, then $(A \wedge B)$ and $(A \to B)$ are formulas;
3. Let a be a term of type $\bar{\sigma}$, x^σ a variable of type σ not free in a, B a formula not containing a quantifier $\forall$ or $\exists$. Then $\wedge x^\sigma \in a\, B$ is a formula;
4. If A is a formula and x^σ is a variable, then $\forall x^\sigma A$ and $\exists x^\sigma A$ are formulas.

A formula is *quantifier-free* (qf), if it does not contain a quantifier $\forall$ or $\exists$. The restricted quantifier $\wedge\, x \in a$ (read: for all elements x of the set a) is not considered a quantifier in the proper sense, though it binds the variable x.

Abbreviations. Let $a \equiv a_1,\ldots,a_n$ and $b \equiv b_1,\ldots,b_m$ be term tuples of type tuples ρ and σ resp. If ρ is of the form $\sigma\to\tau$, then a(b) or ab stands for the term tuple $a_1(b_1)\ldots(b_m),\ldots,a_n(b_1)\ldots(b_m)$ of type tuple τ. If $\rho \equiv \sigma$, (hence n = m), then a = b stands for the formula tuple $a_1=b_1,\ldots,a_n=b_n$. If $\rho \equiv \bar{\sigma}$ and x is a tuple of variables $x_1,\ldots,x_n$ of type tuple σ, then $\wedge\, x \in a$

stands for the string $\wedge x_1 \in a_1 \ldots \wedge x_n \in a_n$. Similarly for quantifier strings $\forall x$ and $\exists x$.

Axioms and rules for propositional logic.

A1. $A \to A$

A2. $A,\quad A \to B \vdash B$

A3. $A \to B,\quad B \to C \vdash A \to C$

A4. $A \wedge B \to A$

A5. $A \wedge B \to C \vdash A \to B \to C$

A6. $A \to B \to C \vdash A \wedge B \to C$

A7. $A \to B,\quad A \to C \vdash A \to B \wedge C$

Axioms and rules for the restricted quantifier.

B1. $\wedge x \in \{a\}\ B \to B_x[a]$

B2. $\wedge x \in a \cup bC \to \wedge x \in aC$

$\wedge x \in a \cup bC \to \wedge x \in bC$

B3. $\wedge y \in \bigcup_{x \in a} bC \to \wedge x \in a \wedge y \in bC$ (x not free in C)

B4. $\wedge x \in a\ (B[x] \to C[x]) \to \wedge y \in aB[y] \to \wedge z \in aB[z]$

(x,y,z new variables)

B5. $A \to B \vdash A \to \wedge x \in aB$ (x not free in A).

Equality axioms.

C1. $\lambda x.a[x] = \lambda y.a[y]$ (x, y new variables)

C2. $(\lambda x.a[x])(b) = a[b]$ (b substitutable for x in a)

C3. $R_\tau ab0 = a$ (a,b of types τ, $\sigma \to \tau \to \tau$ resp.)

$R_\tau ab(\mathrm{suc}\ t) = bt(R_\tau abt)$

C4. $a = b \to F[a] \to F[b]$ (a,b substitutable in F)

Rule of induction.

(IND) $F[0],\ F[x] \to F[\mathrm{suc}\ x] \vdash F[t]$ (t substitutable for x in F)

Axioms and rules for quantification.

Q1. $\forall x\ A[x] \to A[b]$

Q2. $A[b] \to \exists x\ A[x]$ (b substitutable for x in A)

Q3. $A \to B \vdash A \to \forall xB$

Q4. $B \to A \vdash \exists xB \to A$ (x not free in A).

This concludes the definition of the theory HA^ω_ϵ, a Hilbert-style version of Heyting's arithmetic in all finite types, extended by a restricted quantifier $\wedge x \in a$. Its quantifier free fragment is denoted by T_ϵ, an extension of Gödel's 1958 theory T of primitive recursive functionals of finite types by $\wedge x \in a$. To be precise, T_ϵ has the same types and terms as HA_ϵ^ω, its formulas are the qf formulas of HA_ϵ^ω, and its axioms and rules are those of HA_ϵ^ω without Q1 to Q4.

We define $\neg A$ by $A \to 0=1$, $A \leftrightarrow B$ by $(A \to B) \wedge (B \to A)$, and in the presence of the quantifier $\exists$, we define $A \vee B$ by

$$\exists x((x=0 \to A) \wedge (x \neq 0 \to B)).$$

Heyting's arithmetic of finite types HA^ω is the subsystem of HA_ϵ^ω with linear types only, terms defined according to definition 1.4.1 to 1.4.3 only, formulae without restricted quantifier and hence without axioms and rules B. An equivalent formulation is denoted $N\text{-}HA^\omega$ in TROELSTRA 1973, I, § 6. It is outlined in D-Nahm 1974 that the usual rules of intuitionistic logic, including substitution rule and deduction theorem, hold in T_ϵ and in HA_ϵ^ω. These rules are not destroyed by the new axioms and rules B1 to B5. In particular, we have:

A8. $B \vdash A \to B$ by A4, A5, A2;

B5'. $B \vdash \wedge x \in aB$ by A8, B5, $0=0$, A2;

B6. $\wedge y \in a\, B[y] \to \wedge z \in a\, B[z]$ (y,z new) by A1, B5', B4, A2;

B7. $A \to B \to C \vdash A \to \wedge x \in aB \to \wedge x \in aC$ (x not free in A) by B5, B4, A3 (the formula A may also be missing);

B8. $A \to B \to C \vdash \wedge x \in aA \to \wedge x \in aB \to \wedge x \in aC$ by B7, B4, A3;

B9. $\wedge x \in aA \wedge \wedge x \in aB \leftrightarrow \wedge x \in a(A \wedge B)$ from $A \to B \to A \wedge B$ by B8, A6, and by A4, B7, A7.

Standard properties of primitive recursive functions are provable in T_ϵ just as in T (cf. TROELSTRA 1973, 1.6.9). We use functions $+$, $\dot{-}$ etc. and write $s \leq t$ for $s \dot{-} t = 0$, and $s < t$ for $\mathrm{suc}\ s \dot{-} t = 0$.

By D-SCHÜTTE 1971, simultaneous recursion can be derived in T_ϵ: for any type tuple τ, recursor tuples R_τ are definable in T_ϵ such that the tuples of recursion equations C3 are derivable. Of course, the same holds for simultaneous λ-abstraction.

<u>LEMMA 1</u>. $T_\epsilon \vdash \wedge x \in aA \wedge \wedge y \in bB \to \wedge x \in a \wedge y \in b(A \wedge B)$

(x not free in B,b, and y not free in A).

PROOF. From $A \to B \to (A \wedge B)$, we obtain by B7

$$A \to \wedge y \in bB \to \wedge y \in b(A \wedge B).$$

We interchange the premisses, apply B7 to $\wedge x \in a$, and interchange premisses again.

Inductive definition 1.6 of zero-functionals 0^τ for every type τ.

1. $0^0 := 0;$
2. $0^{\sigma\to\tau} := (\lambda x^\sigma.0^\tau);$
3. $0^{\bar{\tau}} := \{0^\tau\}$

We have $\wedge x \in 0^{\bar{\tau}}\ A[x] \to A[0^\tau]$ by B1 and $0^{\sigma\to\tau}(b) = 0^\tau$ by C1. A converse of lemma 1 holds only in the following form:

LEMMA 2. $T_\epsilon \vdash \wedge x \in a \cup 0 \wedge y \in b \cup 0(A \wedge B) \to \wedge x \in aA \wedge \wedge y \in bB$ (x not free in B,b, and y not free in A).

PROOF. By B2 and B1, $\wedge y \in b \cup 0(A \wedge B) \to A$. By B2 and B7, this gives the first half. Similarly, B2 and B9 give $\wedge y \in b \cup 0(A \wedge B) \to \wedge y \in bB$, and by B7, B2, and B1, we get the second half.

LEMMA 3. *The following extended induction rule is admissible in* T_ϵ:

$$B[0,z],\ \wedge w \in a\ B[x^0,w] \to B[\mathrm{suc}\ x^0,z] \vdash B[t,z],$$

where z,w *are disjoint tuples of variables of type tuple* τ, a *is a term tuple of type tuple* $\bar{\tau}$, *and* t *is substitutable in* B.

PROOF. Let $W := \lambda x^0 z.a$, and define by simultaneous recursion (cf. D-SCHÜTTE 1971, Satz 3)

$$g\ 0\ t\ z = \{z\}$$

$$g(\mathrm{suc}\ x)tz = \bigcup_{u \in gxtz} W(t \dot{-} \mathrm{suc}\ x)u.$$

From the first assumption, we obtain by B5'

(1) $\wedge u \in gttz\ B[0,u].$

By B3 and B6, the definition of g gives

$$\wedge u \in g(\mathrm{suc}\ x)tz\ B[t \dot{-} \mathrm{suc}\ x,\ u] \rightarrow \wedge u \in gxtz\ \wedge w \in W(t \dot{-} \mathrm{suc}\ x)\ u\ B[t \dot{-} \mathrm{suc}\ x, w].$$

From this and a substitution instance of the second assumption we get by B7 and A3

$$\wedge u \in g(\mathrm{suc}\ x)tz\ B[t \dot{-} \mathrm{suc}\ x,\ u] \rightarrow \wedge u \in gxtz\ \ B[\mathrm{suc}(t \dot{-} \mathrm{suc}\ x), u].$$

Since $x < t \rightarrow \mathrm{suc}(t \dot{-} \mathrm{suc}\ x) = t \dot{-} x$, this gives

(2) $$x < t \rightarrow \wedge u \in g(\mathrm{suc}\ x)tz\ B[t \dot{-} \mathrm{suc}\ x, u] \rightarrow \wedge u \in gxtz\ B[t \dot{-} x, u].$$

By an induction "from x=t down to x=0" (formally, we first substitute $t \dot{-} \mathrm{suc}\ x$ for x in (2) and then apply induction) it follows from (1) and (2)

$$\wedge u \in g0tz\ B[t \dot{-} 0, u],$$

that is $\wedge u \in \{z\}\ B[t,u]$, which implies $B[t,z]$ by B1.

§ 2. A GENERAL FUNCTIONAL INTERPRETATION OF HA^{ω}_{ϵ}

We assign a prenex formula A^J to any formula A of HA^{ω}_{ϵ} using the motivation given in the introduction (cf. D-VOGEL 1975, pp.56-57).

Inductive definition 2.1 of a translation J.

J maps any formula A of HA^{ω}_{ϵ} into a formula A^J of HA^{ω}_{ϵ} of the form $\exists v \forall w\ A_J$, where A_J is a qf formula, and v,w are disjoint tuples of variables distinct from the variables free in A. A^J is uniquely determined only up to the names of bound variables.

1. $(a=b)^J$ is $a=b$

Let $A^J := \exists v \forall w\ A_J$ and $B^J := \exists y \forall z\ B_J$ with disjoint tuples v,w,y,z be already defined. Then:

2. $(A \wedge B)^J$ is $\exists v,y \forall w,z\ (A_J \wedge B_J)$
3. $(A \rightarrow B)^J$ is $\exists Y,W\ \forall v,z\ (\wedge w \in Wvz\ A_J \rightarrow B_J[Yv,z])$
4. $(\wedge x \in aB)^J$ is $\wedge x \in aB$ (B is qf)
5. $(\forall uA)^J$ is $\exists V\ \forall u,w\ A_J[u,\ Vu,w]$
6. $(\exists uA)^J$ is $\exists u,\ v\ \forall w\ A_J$

In 3, the tuple W has the same length as the tuple w. If w is empty, i.e. if $A^J \equiv \exists v\, A_J$, then

$$(A\to B)^J \text{ is } \exists Y\, \forall v,z\ (A_J \to B_J[Yv,z]).$$

In particular, if A is qf, then A^J is A, and

$$(A \to B)^J \text{ is } \exists y\, \forall z\ (A \to B_J).$$

For disjunction, this gives

$$(A \vee B)^J \text{ is } \exists x,v,y\, \forall w,z\ ((x=0 \to A_J) \wedge (x\neq 0 \to B_J)).$$

A formula A coincides with its translation A^J (up to names of bound variables), iff A is of the form $\exists v\, \forall w\, B$ with B qf. Hence A^{JJ} always is A^J. These properties get lost if $\vee$ is taken as primitive.

We consider the following axiom schemata:

Axiom of choice for types σ,τ.

$$AC_{\sigma\tau}\quad \forall x^\sigma\, \exists y^\tau\, A[x,y] \to \exists Y^{\sigma\to\tau}\, \forall x\, A[x,Yx]$$

AC denotes the union of all schemata $AC_{\sigma\tau}$.

Independence of premisses

$$IP_\epsilon\quad (\forall wA \to \exists yB) \to \exists y(\forall wA \to B)$$

where the variables of the tuple y are not free in A, and A is a qf formula.

Markov's principle

$$M_\epsilon\quad (\forall wA \to B) \to \exists W(\wedge w \in WA \to B)$$

where the variables of the tuple W are not free in B, and A,B are qf formulae. $HA_\epsilon^{\omega+}$ denotes the theory $HA_\epsilon^\omega + AC + IP_\epsilon + M_\epsilon$.

LEMMA 4. *In* HA_ϵ^ω *are derivable:*

1. $(a=b)^J \leftrightarrow a=b$

2. $(A \wedge B)^J \leftrightarrow A^J \wedge B^J$
3. $(\wedge x \in aB)^J \leftrightarrow \wedge x \in a\ B^J$
4. $(\exists u\ B)^J \leftrightarrow \exists u\ B^J$
5. $(\forall u\ B)^J \rightarrow \forall u\ B^J$
6. $(A \rightarrow B)^J \rightarrow A^J \rightarrow B^J$; furthermore
7. $HA^{\omega}_{\epsilon} + AC \vdash \forall u\ B^J \rightarrow (\forall u\ B)^J$
8. $HA^{\omega+}_{\epsilon} \vdash (A^J \rightarrow B^J) \rightarrow (A \rightarrow B)^J$

PROOF. 1 to 4 are obvious, because in 3 B is qf. We consider 6 and 8. Using the notation of definition 2.2, the following are derivable in HA^{ω}_{ϵ}:

$$(A \rightarrow B)^J \rightarrow \forall v\ \exists y\ \forall z\ \exists W\ (\wedge w \in WA_J \rightarrow B_J)$$

$$\exists W(\wedge w \in WA_J \rightarrow B_J) \rightarrow (\forall wA_J \rightarrow B_J)$$

$$\forall z(\forall wA_J \rightarrow B_J) \leftrightarrow (\forall wA_J \rightarrow \forall zB_J)$$

$$\exists y(\forall wA_J \rightarrow \forall zB_J) \rightarrow (\forall wA_J \rightarrow B^J)$$

$$\forall v(\forall wA_J \rightarrow B^J) \leftrightarrow (A^J \rightarrow B^J).$$

This gives 6. The implications missing to prove 8 are just AC, M_{ϵ}, and IP_{ϵ}.

DEFINITION 2.2. Let A^J be $\exists v\ \forall w\ A_J[v,w]$ with qf A_J. The formula A is called *J-interpretable* in T_{ϵ}, if there is a term tuple b in T_{ϵ} containing free only the free variables of A such that $T_{\epsilon} \vdash A_J[b,w]$. The term tuple b is then called (a tuple of) *interpreting terms* for A.

Obviously, we then have $HA^{\omega}_{\epsilon} \vdash A^J$.

LEMMA 5. *If* A^J *and* B^J *coincide up to names of bound variables, then* $A \rightarrow B$ *is J-interpretable in* T_{ϵ}.

PROOF. Let A^J be $\exists v\ \forall w\ C[v,w]$ and let B^J be $\exists y\ \forall z\ C[y,z]$ with C qf. By B1, we have

$$\wedge w \in \{z\}\ C[v,w] \rightarrow C[v,z].$$

If we put $Y := \lambda v.\ v$ and $W := \lambda vz.\{z\}$, it follows by equality axioms C2 and C4 that Y,W are interpreting terms of $A \rightarrow B$.

THEOREM 1, *characterization of the translation J.*

1. *The theories* $HA^{\omega+}_{\epsilon}$ *and* $HA^{\omega}_{\epsilon} + \{A \leftrightarrow A^J \mid A \text{ formula of } HA^{\omega}_{\epsilon}\}$ *are equivalent.*
2. *If A is J-interpretable in* T_{ϵ}, *then* $HA^{\omega+}_{\epsilon} \vdash A$.

PROOF. $HA_\epsilon^{\omega+} \vdash A \leftrightarrow A^J$ follows from lemma 4 immediately by induction on the length of A. Conversely, any instance D of the schemata AC, IP_ϵ, and M_ϵ is an implication $A \to B$ to which lemma 5 applies. Therefore, D is J-interpretable in T_ϵ and $HA_\epsilon^\omega \vdash D^J$, hence $HA_\epsilon^\omega + (D \leftrightarrow D^J) \vdash D$. This proves 1 which immediately implies 2.

THEOREM 2, *interpretation theorem. Any theorem of* $HA_\epsilon^{\omega+}$ *is J-interpretable in* T_ϵ.

Proof by induction on deductions. Axioms A1 and AC, IP_ϵ, M_ϵ and similarly C4 are J-interpretable by lemma 5. The qf axioms B1 to B4 and C1 to C3 are their own J-translations. A2 is almost a special case of A3.

A3. $A \to B,\ B \to C \vdash A \to C$. We assume that for all variables that occur free only in B, the corresponding zero functionals have been substituted. By induction hypothesis, we are given term tuples Y_1, W_1, Y_2, W_2 such that

$$\wedge w \in W_1vz\ A_J[v,w] \to B_J[Y_1v,z]$$

$$\wedge z \in W_2yx\ B_J[y,z] \to C_J[Y_2y,x].$$

We substitute Y_1v for y in the second formula and apply B7 to the first formula, writing a for $W_2(Y_1v)x$, and obtain

$$\wedge z \in a \wedge w \in W_1vz\ A_J[v,w] \to \wedge z \in a\ B_J[Y_1v,z]$$

$$\wedge z \in a\ B_J[Y_1v,z] \to C_J[Y_2(Y_1v),x].$$

By an instance of B3 and two applications of A3, we have

$$\wedge w \in \bigcup_{z\in a} W_1vz\ A_J[v,w] \to C_J[Y_2(Y_1v),x].$$

If we choose $W := \lambda vx. \bigcup_{z\in a} W_1vz$ and $Y := \lambda v.Y_2(Y_1v)$, we see by equality inferences, that

$$\wedge w \in Wvx\ A_J[v,w] \to C_J[Yv,x],$$

and hence Y,W are interpreting terms of $A \to C$.

A4. $A \wedge B \to A$. We look for tuples of functionals Y,W, and W' such that

$$\wedge w \in Wvv'z\ \wedge w' \in W'vv'z\ (A_J[v,w] \wedge B_J[v',w'] \to A_J[Yvv',z]).$$

Such functionals are $Y := \lambda vv'.v$, $W := \lambda vv'z.\{z\}$, and an arbitrary W', e.g. $W' = 0$, or if an empty set $\emptyset$ is available, $W' := \lambda vv'z.\emptyset$.

A5. $A \wedge B \to C \vdash A \to B \to C$. By induction hypothesis, we have term tuples Y,W, and W' such that

$$\bigwedge w \in Wvv'z \bigwedge w' \in W'vv'z\ (A_J \wedge B_J) \to C_J[Yvv',z].$$

By lemma 1, we see that Y,W,W' are at the same time interpreting terms of $A \to B \to C$.

A6. $A \to B \to C \vdash A \wedge B \to C$. By induction hypothesis, we have term tuples Y,W, and W' such that

$$\bigwedge w \in Wvv'z\ A_J \to \bigwedge w' \in W'vv'z\ B_J \to C_J[Yvv',z].$$

By lemma 2, this implies

$$\bigwedge w \in Wvv'z \cup 0 \bigwedge w' \in W'vv'z \cup 0 (A_J \wedge B_J) \to C_J[Yvv',z].$$

Hence Y, $\lambda vv'z.(Wvv'z \cup 0)$, $\lambda vv'z.(W'vv'z \cup 0)$ are interpreting terms of $A \wedge B \to C$.

A7. $A \to B$, $A \to C \vdash A \to B \wedge C$. By induction hypothesis, we have term tuples Y,W and Y', W' such that

$$\bigwedge w \in Wvz\ A_J \to B_J[Yv,z] \quad \text{and}$$

$$\bigwedge w \in W'vz'\ A_J \to C_J[Y'v,z'].$$

Then by B2, A3 and A7

$$\bigwedge w \in (Wvz \cup W'vz')A_J \to B_J[Yv,z] \wedge C_J[Y'v,z'].$$

Hence, $Y,Y',\lambda vzz'.(Wvz \cup W'vz')$ are interpreting terms of $A \to B \wedge C$.

B5. In the presence of quantifiers, B5 may be replaced by $\forall xB \to \bigwedge x \in aB$, because this, Q2 and A3 give B5. This, however, has a as interpreting term, since B is qf.

(Ind) $F[0]$, $F[x] \to F[\text{suc } x] \vdash F[t]$. By induction hypothesis and C2, there are tuples v_o, Y, and W such that

$$F_J[0,v_o,z] \text{ and}$$

$$\wedge w \in Wxvz\ F_J[x,v,w] \to F_J[\text{suc } x,\ Yxv,z].$$

By simultaneous recursion, we define a term tuple f by

$$f0 = v_0 \quad \text{and} \quad f(\text{suc } x) = Yx(fx).$$

Substituting fx for v, we get by C4

$$F_J[0,f0,z] \quad \text{and}$$

$$\wedge w \in Wx(fx)z\ F_J[x,fx,w] \to F_J[\text{suc } x,\ f(\text{suc } x),z].$$

With $F_J[x,fx,z]$ as $B[x,z]$ and $Wx(fx)z$ as a, these are the premisses of lemma 3. Hence, by lemma 3, $F_J[t,ft,z]$ is derivable, and ft are interpreting terms of $F[t]$.

The interpretation of Q1 to Q4 does not present any difficulties. This finishes the proof of theorem 2. We state some simple corollaries to the theorem.

<u>COROLLARY 2.1</u>. *Let* Φ *be a set of new basic functionals and* Γ *a set of formulae in the language of* $T_\epsilon + \Phi$. *Then any theorem of* $HA_\epsilon^{\omega+} + \Gamma$ *is* J*-interpretable in* $T_\epsilon + \Gamma$.

The additional axioms in Γ are interpretable by themselves, because they are quantifier free.

This corollary is the starting point for functional interpretations of intuitionistic analysis. In the principal application, Φ is the set of bar recursive functionals, and Γ is the set of their recursion equations.

<u>COROLLARY 2.2</u>. $HA_\epsilon^{\omega+}$ *is a conservative extension of* T_ϵ. *If* T_ϵ *is consistent, then* $HA_\epsilon^{\omega+}$ *is consistent.*

<u>PROOF</u>. If A is a qf theorem of $HA_\epsilon^{\omega+}$, then $A \equiv A^J \equiv A_J$. Hence, by the theorem, $T_\epsilon \vdash A$. For $A \equiv 1 = 0$, this is the relative consistency of $HA_\epsilon^{\omega+}$.

As we shall see, HA_ϵ^{ω} is conservative over HA^{ω}, but it is an open question whether T_ϵ is conservative over Gödel's qf theory T.

<u>COROLLARY 2.3</u>. HA_ϵ^{ω} *is closed under Markov's rule rule-*M_ϵ*: If* $\vdash \forall w\ A \to B$ *and* A,B *are* qf, *then there is a term tuple* W *such that* $\vdash \wedge w \in W\ A \to B$.

This is simply a special case of the theorem.

The theorem, as it stands, on the other hand, implies closures of HA^{ω}_{ϵ} under rule-AC and rule-IP_{ϵ} only for qf formulae A,B. To obtain these closure results in general, one introduces the *hybrid* Jq of the translation J which is inductively defined like J except for the following two cases:

$$(A \to B)^{Jq} \text{ is } \exists Y,W\ \forall v,z\ (\wedge w \in Wvz\ A_{Jq} \wedge A \to B_{Jq}[Yv,z]),$$

$$(\exists uA)^{Jq} \text{ is } \exists u,v\ \forall w\ (A_{Jq} \wedge A).$$

Following the proof of theorem 2, it is an easy exercise to prove a Jq-interpretation theorem for HA^{ω}_{ϵ} only, in fact with the same interpreting terms as before. In a slightly different context, STEIN 197? gives a unified treatment for a general functional interpretation and its hybrid. Because the conclusions of AC and IP_{ϵ} are existential formulae, and because $\vdash (\exists uA)^{Jq} \to \exists uA$, the Jq-interpretation theorem immediately implies closure of HA^{ω}_{ϵ} under rule-AC and rule-IP_{ϵ}.

§ 3. SPECIAL MODELS

We introduced the notion of type structure in definition 1.2. A type structure carries little information in the sense that only application has a fixed meaning in it: If f,c are (names for) elements of $M_{\sigma\to\tau}$, M_{σ} resp., then the term f(c) denotes in M_{τ} the value of the function f at argument c. The other basic functionals and operations are not yet interpreted. We close that gap by introducing models.

<u>DEFINITION 3.1.</u> A *classical model* of HA^{ω}_{ϵ} is a type structure $A = (M_{\tau} | \tau$ a type) together with a (semantic) interpretation of basic functionals, λ-abstraction and set forming operations in A such that A satisfies the axioms and rules of HA^{ω}_{ϵ}: there is $0 \in M_{o}$, suc $\in M_{o\to o}$, and for any τ, $R_{\tau} \in M_{\sigma}$ with $\sigma = \tau\to(o\to\tau\to\tau)\to o\to\tau$; if $a[x^{\sigma}]$ is a term built from x^{σ} and (names of) elements of A such that $a[c] \in M_{\tau}$ for every $c \in M_{\sigma}$, then $\lambda x^{\sigma}.a[x^{\sigma}]$ denotes in A one of the functions $f \in M_{\sigma\to\tau}$ for which $f(c) = a[c]$ for all $c \in M_{\sigma}$.

Furthermore, for any $c \in M_{\tau}$, there is $\{c\} \in M_{\bar\tau}$, and for $a,b \in M_{\bar\tau}$, there is $a \cup b \in M_{\bar\tau}$, and if $f \in M_{\sigma\to\bar\tau}$ and $a \in M_{\bar\sigma}$, then $\bigcup_{x\in a} f(x) \in M_{\bar\tau}$.

An (intuitionistic) *model of* HA^{ω}_{ϵ} is a Kripke tree $A = (K,\le,(A_{\alpha}|\alpha \in K),(\varphi^{\alpha}_{\beta} \mid \alpha \le \beta))$ of classical models A_{α} with homomorphismus $\varphi^{\alpha}_{\beta} : A_{\alpha} \to A_{\beta}$ (for details, see OSSWALD 1969, ch.I, § 2, also KRIPKE 1965).

A *classical model of* HA^{ω} is now a linear type structure with the above properties, but without set forming operations, and a *model* of HA^{ω} is a Kripke tree of classical models of HA^{ω}.

By correctness and completeness, the theorems of HA^{ω} are exactly the formulae valid in all models of HA^{ω} (cf. OSSWALD 1969, ch.I, § 3). As the use of sets and the restricted quantifier in the theory HA^{ω}_{ϵ} is weak, though the properties of sets in the models defined above are quite substantial, the theory HA^{ω}_{ϵ} is certainly correct with respect to these models, but it is presumably not complete.

In this section, we expand models A of HA^{ω} in certain standard ways to special models A^E of HA^{ω}_{ϵ} by essentially defining in terms of A what the classes M^E_{τ} of the type structure A^E are to be. We first describe the general pattern of these expansions.

REMARKS.

(i) Any type τ has either the form $\rho\to o$ or the form $\rho\to\bar{\sigma}$ where ρ is a type tuple uniquely determined by τ.

(ii) If a class M^E_{σ} or M^E_{τ} is a singleton, then any $f \in M^E_{\sigma\to\tau}$ has only one value $f(c)$ in M^E_{τ}, independent of the argument $c \in M^E_{\sigma}$, amd we may well identify f with its unique value $f(c)$ and put $M^E_{\sigma\to\tau} = M^E_{\tau}$ without danger of confusion.

DEFINITION 3.2. Let T be a new symbol. An *expanding function* is a map

$$E : \{\text{types}\} \to \{\text{linear types}\} \cup \{T\}$$

such that

1. $E\, o = o$;
2. $E(\sigma\to\tau) = E\sigma\to E\tau$, if $E\sigma$ and $E\tau$ are linear types;
3. $E(\sigma\to\tau) = E\tau$, if $E\sigma = T$ or $E\tau = T$.

If $A = (M_{\tau}|\tau$ linear type) is a model of HA^{ω}, we define an *E-expansion* $A^E = (M^E_{\tau}|\tau$ a type) as follows:

$$M^E_{\tau} = M_{E\tau}, \text{ if } E\tau \text{ is a linear type;}$$

$$M^E_{\rho\to\bar{\sigma}} = \{M_{\sigma}\}, \text{ if } E\bar{\sigma} = T.$$

0, suc have the same meaning in A and in A^E; if $E\tau$ is linear, R_{τ} denotes in A^E the same object as $R_{E\tau}$ in A. If $E\sigma$, $E\tau$ are linear, and $a[c] \in M_{E\tau}$ for all $c \in M_{E\sigma}$, then $\lambda x^{\sigma}.a[x^{\sigma}]$ denotes in A^E the same element of $M^E_{\sigma\to\tau} = M_{E\sigma\to E\tau}$

as $\lambda x^{E\sigma}.a[x^{E\sigma}]$ does in A.

An expanding function E is the identity on linear types and hence A^E is in fact an expansion of A. To obtain a particular E-expansion of a given model A of HA^{ω}, we first have to define E for set types $\bar{\tau}$ and then fix a meaning for the set forming operations and the restricted quantifier. We want to show that in suitable classes of E-expansions, the translation J becomes equivalent to known functional interpretations.

3.1. <u>The trivial expansion: modified realisability</u>

As universe of type $\bar{\tau}$, we take the singleton whose only element is the universe of type τ. In such type structures, we always have

$$\wedge x^{\tau} \in a\ B \leftrightarrow \forall x^{\tau} B.$$

Following the lines described above, this canonically determines the meaning of the set forming operations.

<u>DEFINITION 3.3</u>. Let the map U be the expanding function with

$$U(\bar{\tau}) = T \quad \text{for all types } \tau.$$

If A is a (classical) model of HA^{ω}, we call the U-expansion A^U of A the *trivial expansion* of A.

Given A, the trivial expansion is in fact uniquely determined by this definition: the set forming operations leading to terms of type $\bar{\tau}$, necessarily yield the only object of type $\bar{\tau}$, namely M^U_{τ}.

<u>LEMMA 6</u>. *Let* A *be a classical model of* HA^{ω}. *For any functional* a *of* HA^{ω}_{ϵ} *of type* τ *there is an* a^U *such that*

(i) *if* $U\tau = T$, a^U *is the empty string of symbols,*

(ii) *if* $U\tau$ *is linear,* a^U *is a functional of* HA^{ω} *of type* $U\tau$ *denoting the same object in* $M_{U\tau}$ *as* a.

<u>PROOF</u>.

(i) is a matter of convenience, and

(ii) follows by a straight forward induction on the length of functionals a that are formed also from constants of A^U, using (i).

THEOREM 3. *If $\mathcal{A}$ is a (classical) model of* HA^{ω}, *then its trivial expansion* $\mathcal{A}^{U}$ *is a (classical) model of* HA^{ω}_{ϵ} *in which hold*

(3) $\exists x^{\tau}B \leftrightarrow B$ and $B \leftrightarrow \forall x^{\tau}B$, if $U\tau = T$,

(4) $\wedge x^{\sigma}\epsilon a\ B \leftrightarrow \forall x^{\sigma}B$.

PROOF. Let $\mathcal{A}$ be a classical model. If $U\tau = T$, M^{U}_{τ} is a singleton, and hence (3) holds. If τ is $\bar{\sigma}$, then $M^{U}_{\tau} = \{M^{U}_{\sigma}\}$, so that any term a of type $\bar{\sigma}$ is interpreted as the universe M^{U}_{σ} of type σ, and hence (4) holds.

Axioms and rules of groups A, Q, Ind hold in $\mathcal{A}^{U}$ just as they do in $\mathcal{A}$. Because of (4), B1 to B5 hold in $\mathcal{A}^{U}$, as they reduce to valid laws on $\forall$. Equality axioms C hold in $\mathcal{A}^{U}$ either trivially, if the equations are of a type τ with $U\tau = T$, or otherwise by lemma 6 and since $\mathcal{A}^{U}$ is a U-expansion of $\mathcal{A}$.

If $\mathcal{A}$ is an intuitionistic model of HA^{ω}, any $\mathcal{A}^{U}_{\alpha}$ is a classical model of HA^{ω}_{ϵ} by the preceding, and any homomorphism φ^{α}_{β} extends in the obvious way to a homomorphism from $\mathcal{A}^{U}_{\alpha}$ to $\mathcal{A}^{U}_{\beta}$. This proves the theorem.

COROLLARY 3.1. HA^{ω}_{ϵ} *is a conservative extension of* HA^{ω}.

PROOF. If $HA^{\omega} \nvdash A$, there is a model $\mathcal{A}$ of HA^{ω} in which A is not valid. This model has a trivial expansion $\mathcal{A}^{U}$ which by theorem 3 is a model of HA^{ω}_{ϵ}. Hence $HA^{\omega}_{\epsilon} \nvdash A$.

We now turn to a comparison of the general interpretation J and modified realisability. For formulae A of HA^{ω}, v mr A ("the tuple v modified realizes the formula A") is defined in KREISEL 1959 and in TROELSTRA 1973, § 3.4. In our context we may say:

DEFINITION 3.4. For formulae A of HA^{ω}_{ϵ}, A^{mr} (A" is modified realisable") is inductively defined like A^{J}, except that the clause for implication now reads:

$$(A\to B)^{mr} \text{ is } \exists Y\ \forall v\ (\forall w\ A_{mr}[v,w] \to \forall z\ B_{mr}[Yv,z]).$$

If we write v mr A for $\forall w\ A_{mr}[v,w]$, this is in fact the definition of TROELSTRA 1973.

THEOREM 4. *Let* A *be a formula of* HA^{ω}_{ϵ} *and let* A^{J} be $\exists v\ \forall w\ A_{J}$. *If $\mathcal{A}$ is a model of* HA^{ω}, *the following are valid in the trivial expansion* $\mathcal{A}^{U}$ *of* $\mathcal{A}$:

$$\forall v(\forall w\ A_J[v,w] \leftrightarrow v \text{ mr } A), \text{ hence}$$
$$A^J \leftrightarrow A^{mr}.$$

PROOF. By induction on the length of A. By lemma 4 and because mr and J have the same definitional clauses for equality, conjunction, restricted and existential quantification, the induction goes through for these cases. We consider only the case of implication. Let B^J be $\exists y\ \forall z\ B[y,z]$. We then have for all Y,W:

$$\forall v,z\ (A\to B)_J \leftrightarrow \forall v,z\ (\wedge w \in Wvz\ A_J \to B_J[Yv,z])$$

$$\leftrightarrow \forall v,z\ (\forall w\ A_J \to B_J[Yv,z]) \quad \text{by (4)}$$

$$\leftrightarrow \forall v\ (\forall w\ A_J \to \forall z\ B_J[Yv,z])$$

$$\leftrightarrow \forall v\ (v \text{ mr } A \to Yv \text{ mr } B) \quad \text{by ind.hyp.}$$

$$\leftrightarrow Y \text{ mr } (A\to B) \quad \text{by defn. of mr}$$

$$\leftrightarrow Y,W \text{ mr } (A\to B)$$

since by lemma 6, Y and Y,W denote the same tuple of objects in $\mathcal{A}^U$.

In the class of trivial expansions, J-interpretation and modified realisability are equivalent. Furthermore, the same objects in these models interpret and modified realize a given formula. Because of (4), Markov's principle M_ϵ becomes a tautology, and IP_ϵ becomes $IP_{\exists\text{-free}}$, independence of premisses in the case where $\exists$ (and $\vee$) does not occur in the formula A.

COROLLARY 4.1, *characterization of mr. In* HA^ω, *the schemata* $A \leftrightarrow A^{mr}$ *and* $AC + IP_{\exists\text{-free}}$ *are equivalent.*

PROOF. By the above and theorem 1, the equivalence is valid in all trivial expansions $\mathcal{A}^U$ and hence in all models $\mathcal{A}$ of HA^ω. By completeness, it holds in HA^ω.

COROLLARY 4.2, *mr-interpretation theorem. Any theorem of* $HA^\omega + AC + IP_{\exists\text{-free}}$ *is modified realizable in* HA^ω.

PROOF. If A is a theorem of $HA^\omega + AC + IP_{\exists\text{-free}}$, then by the above, A is a theorem of $HA_\epsilon^{\omega+} + (4)$. By theorem 2 and lemma 4, A has interpreting terms b in $T_\epsilon + (4)^J$. But $(4)^J \leftrightarrow (4)$ in HA^ω_ϵ, hence

$$HA^\omega_\epsilon + (4) \vdash \forall w\, A_J[b,w].$$

Let A be any model of HA^ω. Then A^U is a model of $HA^\omega_\epsilon + (4)$ by theorem 3, and by theorem 4, b mr A holds in A^U and hence in A. By completeness, $HA^\omega \vdash$ b mr A.

In fact, $\forall w\, A_J[b,w]$ is derived in the $\exists$-free fragment of $HA^\omega_\epsilon + (4)$. To obtain b mr A already in the $\exists$-free fragment of HA^ω, our model theoretic theorem 3 seems to be too crude.

3.2. Expansion by finite sequences: $\wedge$- and Dialectica-interpretation

We now turn to models which prove the Dialectica-interpretation (GÖDEL 1958) and the $\wedge$-interpretation (D-NAHM 1974) to be special cases of J. Again, we start with an arbitrary classical model $A = (M_\tau | \tau$ linear) of HA^ω, and now we want $M_{\bar{\tau}}$ the be the *set of finite sequences* of elements of M_τ. In a context where extensionality is assumed, finite sequences of objects of type τ may be coded into one element of type τ. In an arbitrary classical model A, however, finite sequences c of elements of M_τ are given by a pair consisting of the length l of c and an element g of $M_{o\to\tau}$ such that $c = \langle g0,\dots,g(l-1)\rangle$. For linear $\tau \equiv \sigma\to o$ (σ a type tuple), this pair (l,g) may be coded into one element f of $M_{o\to\tau}$ such that $f00^\sigma = l$ and $f(\text{suc } x) = gx$ for all $x \in M_o$. Since equality in M_τ is not necessarily extensional, we do not have to bother about a refined definition of equality between these codes in $M_{o\to\tau}$. This construction leads naturally to the following

DEFINITION 3.5. Let the map $\wedge$: {types} $\to$ {linear types} be the expanding function with

$$\wedge\bar{\tau} = o \to \wedge\tau.$$

For any classical model $A = (M_\tau|\tau$ linear) of HA^ω, the $\wedge$-expansion then is $A^\wedge = (M_{\wedge\tau}|\tau$ a type). If $f \in M_{\wedge\bar{\tau}}$, the *length* of f is the element $lh(f) \in M_o$ defined by $f(0) = \lambda x^\sigma.lh(f)$ (where $\tau = \sigma\to o$), and we call f a *finite sequence* if $f = f|(lh(f)+1)$ (for the restriction functional $|$, cf. D-SCHÜTTE 1971, def. 8).

1. For $c \in M_{\wedge\tau}$, $\{c\}$ is the finite sequence of length 1 with $\{c\}(1) = c$;
2. For $a,b \in M_{\wedge\bar{\tau}}$, $a \cup b$ is the finite sequence of length $lh(a) + lh(b)$ with

$$(a \cup b)(x) = \begin{cases} a(x) & \text{if } 0 < x \leq lh(a) \\ b(x \dot{-} lh(a)) & \text{if } x > lh(a). \end{cases}$$

3. For $a \in M_{\wedge\bar{\sigma}}$, $f \in M_{\wedge\sigma\to\wedge\bar{\tau}}$, $\bigcup_{u\in a} f(u)$ is the finite sequence of length $\sum_{0<y\leq lh(a)} lh(f(a(y)))$ with

$$(\bigcup_{u\in a} f(u))(x) = f(a(y))(x \dot{-} \sum_{0<z<y} lh(f(a(z)))$$

if $\sum_{0<z<y} lh(f(a(z))) < x \leq \sum_{0<z\leq y} lh(f(a(z)))$.

4. For $a \in M_{\wedge\bar{\tau}}$ and any qf formula $B[u^\tau]$,

(5) $\quad \wedge u^\tau \in a\ B[u^\tau]$ holds iff $\forall x\ (x < lh(a) \to B[a(x+1)])$.

It should be noted that A and $A^\wedge$ differ only as families, but their underlying sets $\{M_\tau | \tau \text{ linear}\}$ are the same.

By D-VOGEL 1975, (1.4) and (1.5), finite sequences f,g are equal, iff $f(x) = g(x)$ for all $x \leq lh(f)$, and $f|(lh(f)+1)$ always is a finite sequence (of length $lh(f)$). Hence the finite sequences formed by set forming operations are uniquely determined by 1,2, and 3 of definition 3.5, every functional of HA^ω_ϵ denotes an element of $A^\wedge$, and by (5), $f \in M_{\wedge\bar{\tau}}$ intuitively denotes the finite sequence $\langle f(1),\ldots,f(lh(f))\rangle$.

<u>LEMMA 7</u>. *For every functional* a *of* HA^ω_ϵ *of type* τ *there is a functional* $a^\wedge$ *of* HA^ω *of type* $\wedge\tau$ *that denotes the same element in* A *as* a *denotes in* $A^\wedge$.

The proof uses the same induction as lemma 6.

<u>THEOREM 5</u>. *If* A *is a (classical) model of* HA^ω, *then its* $\wedge$*-expansion* $A^\wedge$ *is a (classical) model of* HA^ω_ϵ.

<u>PROOF</u>. If A is a classical model of HA^ω, $A^\wedge$ satisfies the axioms and rules of HA^ω. We only have to check B1 to B5. Axioms B1 to B3 are just being taken care of by 1,2, and 3 of definition 3.5, because, using (5), c is an element of the finite sequence $\{c\}$, a and b are subsequences of the finite sequences $a \cup b$, and for any $c \in a$, $b[c] = \lambda u.b[u](c)$ is a subsequence of $\bigcup_{u\in a} b[u]$.

Finally, because of (5), B4 is a special case of $\forall$-distribution and B5 is a special case of Q3.

If $\mathcal{A}$ is an intuitionistic model, then the homomorphisms φ^{α}_{β} between the classical models $\mathcal{A}_{\alpha}$ and $\mathcal{A}_{\beta}$ extend in the obvious way to homomorphisms from $\mathcal{A}^{\wedge}_{\alpha}$ to $\mathcal{A}^{\wedge}_{\beta}$, since $\mathcal{A}_{\alpha}$ and $\mathcal{A}^{\wedge}_{\alpha}$ have the same underlying set.

We now compare the general interpretation J and the $\wedge$-interpretation (D-NAHM 1974). We adapt the $\wedge$-translation slightly to our present context.

DEFINITION 3.6. For formulae A of HA^{ω}_{ϵ}, $A^{\wedge}$ is inductively defined like A^{J}, except that the clause for implication now reads:

$$(A \rightarrow B)^{\wedge} \text{ is } \exists Y,W\, \forall v,z\ (\forall x(x < lh(Wvz) \rightarrow A_{\wedge}[v,Wvz(x+1)]) \rightarrow B_{\wedge}[Yv,z]).$$

THEOREM 6. *For any formula* A *of* HA^{ω}_{ϵ}, *let* A^{J} *be* $\exists v \forall w\, A_{J}$ *and let* $A^{\wedge}$ *be* $\exists v \forall w\, A_{\wedge}$. *If* $\mathcal{A}$ *is a model of* HA^{ω}, *the following are valid in the* $\wedge$*-expansion* $\mathcal{A}^{\wedge}$ *of* $\mathcal{A}$:

$$\forall v \forall w\ (A_{J} \leftrightarrow A_{\wedge}), \quad \text{hence} \quad A^{J} \leftrightarrow A^{\wedge}.$$

PROOF by induction on the length of A. Again, we consider the case of implication only. Let B^{J} be $\exists y \forall z B_{J}[y,z]$. We then have for all Y,W,v,z:

$$(A \rightarrow B)_{J} \leftrightarrow (\wedge w \in Wvz\ A_{J} \rightarrow B_{J}[Yv,z])$$

$$\leftrightarrow (\forall x(x < lh(Wvz) \rightarrow A_{J}[v,Wvz(x+1)]) \rightarrow B_{J}[Yv,z]) \text{ by } (5)$$

$$\leftrightarrow (A \rightarrow B)_{\wedge} \text{ by ind.hyp.}$$

The proof is more direct and gives a stronger result than in the case of trivial expansions. If (5) is taken as a formal definition, $A_{\wedge}$ and A_{J} are the same formulae. This is not surprising, since the $\wedge$-interpretation was the prototype of the J-interpretation and in a way already embodies its general idea.

Let $IP_{\wedge}$ be IP_{ϵ} where A is qf only up to quantifiers $\forall y^{o}(y^{o} < t \rightarrow \ldots)$, and let $M_{\wedge}$ be the corresponding schema:

$$M_{\wedge} \quad (\forall w\ A[w] \rightarrow B) \rightarrow \exists W(\forall x(x < lh(W) \rightarrow A[W(x+1)]) \rightarrow B)$$

where the variables of the tuple W are not free in B, and A,B are qf up to quantifiers $\forall y^{o}(y^{o} < t \rightarrow \ldots)$.

COROLLARY 6.1, *characterization of* $\wedge$. *In* HA^{ω}, *the schemata* $A \leftrightarrow A^{\wedge}$ *and* $AC + IP_{\wedge} + M_{\wedge}$ *are equivalent.*

PROOF. Let $\mathcal{A}$ be a model of HA^{ω}. In its $\wedge$-expansion $\mathcal{A}^{\wedge}$, IP_{ϵ} and M_{ϵ} are equivalent to $IP_{\wedge}$ and $M_{\wedge}$ by (5), and the schemata $A \leftrightarrow A^{J}$ and $A \leftrightarrow A^{\wedge}$ are equivalent by theorem 6. Hence, the above equivalence holds in $\mathcal{A}^{\wedge}$ and therefore in $\mathcal{A}$. By completeness, it holds in HA^{ω}.

COROLLARY 6.2, *$\wedge$-interpretation theorem. Any theorem of* $HA^{\omega} + AC + IP_{\wedge} + M_{\wedge}$ *is $\wedge$-interpretable in* HA^{ω}.

The proof is like that of corollary 4.2, except that it uses (5) and theorems 5,6, instead of (4) and theorems 3,4.

We turn to the Dialectica-interpretation of GÖDEL 1958. It is the following special case of the general interpretation:

DEFINITION 3.7. For formulae A of HA^{ω}_{ϵ}, A^{D} is inductively defined like A^{J}, except that the clause for implication now reads:

$$(A \to B)^{D} \text{ is } \exists Y,W\, \forall v,z\, (A_{D}[v,Wvz] \to B_{D}[Yv,z]).$$

Hence for A in HA^{ω}, A_{D} is qf in the strict sense, and $A^{D\wedge}$ and A^{D} are identical. The finite sequence occurring in $(A \to B)_{\wedge}$ is reduced to one element. As is well known, the Dialectica-interpretation does not work for the theory HA^{ω} as it stands, but only for related theories in which all prime formulae are equivalent to equations between terms of type o. However, in these theories, D and $\wedge$ are equivalent. This follows from theorem 3.5.10 in TROELSTRA 1973 (cf. also YASUGI 1963) which we quote as

THEOREM 7. *If the formula* A *of* HA^{ω} *is built up from equations of type* o, *then*

$$HA^{\omega} + AC + IP_{\wedge} + M_{\wedge} \vdash A \leftrightarrow A^{D}.$$

We are interested in two consequences of this theorem.

COROLARRY 7.1. *If* A *is as above, then*

$$HA^{\omega} \vdash A^{\wedge} \leftrightarrow A^{D}.$$

PROOF. $\wedge$-interpreting $A \leftrightarrow A^{D}$, corollary 6.2 gives $HA^{\omega} \vdash (A \leftrightarrow A^{D})^{\wedge}$ which implies $HA^{\omega} \vdash A^{\wedge} \leftrightarrow A^{D\wedge}$ by lemma 4. But $A^{D\wedge} \equiv A^{D}$.

CORLLARY 7.2. *Any theorem of* $HA^{\omega} + AC + IP_{\wedge} + M_{\wedge}$ *that is built up from equations of type* o *is Dialectica interpretable in Gödel's theory* T.

PROOF. If A is such a theorem, then by theorem 7, A^D is such a theorem too. By corollary 6.2, A^D has $\wedge$-interpreting terms, and because of $A^{D\wedge} \equiv A^D$, these are at the same time D-interpreting terms of A.

It is in general not the case that the theorem A itself has the same $\wedge$-interpreting and D-interpreting terms.

The Dialectica interpretation appears here only as (the equivalent of) a restriction of the $\wedge$-interpretation which in turn is a special case of the general interpretation J. This is of course not the place it deserves as D is the only interpretation J where A_J is qf in the strict sense, and since D can be extended to classical analysis (cf. SPECTOR 1962). But D is also the only interpretation that requires something like the equivalence of qf formulas with equations of type o (cf. the discussion in TROELSTRA 1973, remark 3.5.6). It is exactly the $\wedge$-interpretation that is close enough to D for corollary 7.1, and does not depend on the above requirement.

In 1.6.15, TROELSTRA 1973 formulates a system which he calls HA^ω, the formulae of which are all built up from equations of type o. Theorem 7 with corollaries also holds with his HA^ω instead of ours. RATH 1978, ch. IV, shows that our HA^ω is conservative over Troelstra's.

3.3. Inbetween: n-interpretations

The search for unifying aspects of the known interpretations mr, D, and $\wedge$ led to a hierarchy of interpretations, one for each $n > 0$, that fill a gap between $\wedge$ (corresponding to $n = 0$) and mr (corresponding to $n = \infty$). They were introduced in STEIN 1976 and are treated syntactically in the forthcoming STEIN 197?. We mainly describe the expansions in which J becomes equivalent to an n-interpretation.

Inductive definition 3.8 of the degree $d(\tau)$ of a type τ.

1. $d(0) = 0$
2. $d(\sigma\to\tau) = \max(d(\sigma)+1,\ d(\tau))$
3. $d(\bar{\tau}) = d(\tau)$.

If we define *pure types* $0 := o$ and $n + 1 := n \to n$, then the pure type n is of degree n.

For fixed $n > 0$, we assume *pairing functionals* $<,>_\sigma$ of type $\sigma\to(n-1)\to(n-1)$ to be available for any linear type σ with $d(\sigma)<n$, together with inverses j_1 of type $(n-1)\to\sigma$ and j_2 of type $(n-1)\to(n-1)$ such that $j_i<a_1,a_2> = a_i$. If extensionality is assumed, such pairing functionals are definable.

DEFINITION 3.9. For fixed $n > 0$, let the map U_n be the expanding function with

$$U_n(\bar{\tau}) = T \quad \text{for} \quad d(\tau) < n$$

$$U_n(\bar{\tau}) = (n-1) \to U_n(\tau) \quad \text{for } d(\tau) \geq n.$$

(For $n = \infty$, $U_n = U$.) Let A^n be the U_n-expansion of a classical model A of a HA^ω with pairing functionals. For $d(\tau) \geq n$, we define

1. If $c \in M^n_\tau$, then $\{c\} := \lambda x^{n-1}.c$
2. If $a,b \in M^n_{\bar{\tau}}$, then $a \cup b \in M^n_{\bar{\tau}}$ such that

$$(a\cup b)(z) = \begin{cases} a(j_2z) & \text{if } j_1z = 0, \\ b(j_2z) & \text{if } j_1z \neq 0; \end{cases}$$

3. If $a \in M^n_{\bar{\sigma}}$, $f \in M^n_{\sigma\to\bar{\tau}}$, then $\bigcup_{u\in a} f(u)$ is an object $g \in M^n_{\bar{\tau}}$ such that

$$g(z) = f(j_1z)(j_2z) \qquad \text{if } d(\sigma) < n,$$

$$g(z) = f(a(j_1z))(j_2z) \qquad \text{if } d(\sigma) \geq n.$$

4. For $a \in M^n_{\bar{\tau}}$ and any qf formula B,

$$\text{(6)} \qquad \wedge u^\tau \in a\ B[u^\tau] \text{ holds iff } \forall x^{n-1}\ B[a(x^{n-1})].$$

Given A, the U_n-expansion A^n is again uniquely determined, if $a \cup b$ and $\bigcup_{u\in a} f(u)$ are constructed from their defining equations in a standard way. For if $d(\tau) < n$, then M^n_τ is the only object of type $\bar{\tau}$, and we necessarily have

$$\{c\} = a \cup b = \bigcup_{u\in a} f(u) = M^n_\tau.$$

It is now routine to check

THEOREM 8. *If A is a (classical) model of HA^ω with pairing functionals, then its U_n-expansion A^n is a (classical) model of HA^ω_ϵ in which* (3) *and* (4) *for* $d(\sigma) < n$ *hold.*

U_n-expansions are like trivial expansions at small types $\tau(d(\tau) < n)$, and they are related to $\wedge$-expansions at large types $\tau(d(\tau) \geq n)$. In the same way, n-interpretations stand between mr and $\wedge$.

DEFINITION 3.10. For formulae A of HA^{ω}_{ϵ}, A^n is inductively defined like A^J, except that the clause for implication now reads

$$(A \to B)^n \text{ is } \exists Y,W\, \forall v,z\ (\forall w^{<n}\ \forall x^{n-1}\ A_n[v,w^{<n},\ Wvzx^{n-1}] \to B_n[Yv,z]),$$

where $w^{<n}$ is the subtuple of w consisting of the variables in w of degree $< n$.

This n-interpretation is, except for a change of notation, the same as the n_*-interpretation of RATH 1978, Ch. IV, § 1, which is there proved to be equivalent to the n-interpretation of STEIN 1976.

THEOREM 9. *For any formula* A *of* HA^{ω}_{ϵ}, *let* A^J *be* $\exists v\, \forall w\, A_J$ *and let* A^n *be* $\exists v\, \forall w\, A_n$. *If* $\mathcal{A}$ *is a model of* HA^{ω} *with pairing functionals,*

$$\forall v(\forall w\, A_J \leftrightarrow \forall w\, A_n),\ \text{hence } A^J \leftrightarrow A^n$$

hold in its U_n*-expansion* $\mathcal{A}^n$.

The proof is like that of theorem 4.

Let IP_n be IP_{ϵ}, where A is qf up to quantifiers $\forall u^{\tau}$ with $d(\tau) \le n$, and consider the following *weak Markov-principles*

$$M_{n,\sigma}\quad (\forall w^{\sigma}\ A[w] \to B) \to \exists W^{n\to\sigma}\ (\forall x^n\ A[Wx] \to B),$$

where W is not free in B, and A,B are qf up to quantifiers $\forall u^{\tau}$ with $d(\tau) \le n$.

M_n denotes the union of all schemata $M_{n,\sigma}$ with $d(\sigma) > n$.

The schemata IP_n increase in strength with growing n, whereas the schemata M_n decrease. If we go to limits, IP_{ω} is just $IP_{\exists\text{-free}}$, whereas M_{ω} is a tautology.

COROLLARY 9.1. *For* $n \in \mathbb{N}$, *the schemata* $A \leftrightarrow A^{n+1}$ *and* $AC + IP_n + M_n$ *are equivalent in* HA^{ω} *with pairing functionals.*

COROLLARY 9.2. *Any theorem of* $HA^{\omega} + AC + IP_n + M_n$ *with pairing functionals is* (n+1)*-interpretable in* HA^{ω} *with pairing functionals.*

3.4. A remark on weak Markov principles

For $n < d(\sigma)$, the principle $M_{n,\sigma}$ states that to infer B from $\forall w^{\sigma}A$, we do not need A[w] for all w of type σ, but only for all w in the image of a suitable map W of type $n \to \sigma$. Given σ, the schema $M_{n,\sigma}$ is the stronger, the smaller n is. Even $M_{o,\sigma}$, however, is much weaker than $M_\wedge$ or other usual Markov principles.

Let I be a functional interpretation mapping the formulae of HA^ω into themselves. We assume that we have an expanding function E and for any model $\mathcal{A}$ of HA^ω an E-expansion $\mathcal{A}^E$ which is a model of HA^ω_ϵ in which $A^I \leftrightarrow A^J$ holds. This is the situation of all the special interpretations that we studied in this section.

Given k, we ask for the smallest n such that the schema $A \leftrightarrow A^I$ implies $M_{n,k}$. This amounts to asking for the number theoretic function g defined by

$$g(k) := \text{the smallest } n \text{ such that } HA^\omega + \{A \leftrightarrow A^I\} \vdash M_{n,k}.$$

We call g the *type reduction* of I.

For I = mr, g is the identity, and for I = n+1, g is given by

$$(7) \qquad g(k) = \begin{cases} k \text{ for } k \leq n, \\ n \text{ for } k > n. \end{cases}$$

This is already the general situation.

<u>LEMMA 8</u>. *Let* g *be a type reduction. We then have for all natural numbers* k:

1. $g(k) \leq k$;
2. $g(g(k)) = g(k)$;
3. $g(k) \leq g(k+1)$;
4. If $g(k+1) = k+1$, then $g(k) = g(k)$.

<u>PROOF</u>. 1. $M_{k,k}$ holds with $W = \lambda w.w$. 2. Because $A^{JJ} \equiv A^J$, we have $A^{II} \leftrightarrow A^I$ in HA^ω and hence

$$HA^\omega + M_{n,k} + M_{1,n} \vdash M_{1,k}.$$

Putting $n = g(k)$ and $1 = g(n) = g(g(k))$, we get $1 \geq n$ by definition of g, hence $1 = n$ by 1. 3. If we put $proj := \lambda w.w0$, we have $proj(\lambda x^k.v) = v$ and therefore

$$\forall w^{k+1}\, A[proj(w^{k+1})] \to \forall v^k A[v^k], \text{ so that } M_{n,k+1} \vdash M_{n,k}.$$

4. $g(k) = k$ means that there is an E-expansion $\mathcal{A}^E$ with the class $M^E_k \in M^E_{\overline{k}}$. Now if $M^E_{k+1} \in M^E_{\overline{k+1}}$, then

$$\bigcup_{w \in M^E_{k+1}} \{\mathrm{proj}(w)\} = M^E_k \in M^E_{\bar{k}}.$$

The function g is by this lemma bounded by identity, idempotent, monotonous, and the set on which g = id is convex. Such a function is necessarily one of the functions named above, as one easily sees.

THEOREM 10. *Let* I *be as above, and let* g *be the type reduction of* I. *Then there is an* $n \in \mathbb{N} \cup \{\infty\}$ *satisfying* (7).

This suggests that our method of charaterizing functional interpretations via E-expansions essentially yields the interpretations treated above. The known functional interpretations of HA^{ω}, however, may be viewed as one and the same functional interpretation looked at in models of HA^{ω}_{ϵ} that differ only in there semantic interpretation of the restricted universal quantifier.

REFERENCES

J. DILLER and W. NAHM 1974: *Eine Variante zur Dialectica-Interpretation der Heyting-Arithmetik endlicher Typen*. Arch. math. Logik 16, 49-66.

J. DILLER and K. SCHÜTTE 1971: *Simultane Rekursionen in der Theorie der Funktionale endlicher Typen*. Arch. math. Logik 14, 69-74.

J. DILLER and H. VOGEL 1975: *Intensionale Funktionalinterpretation der Analysis*, in: Proof theory symposion Kiel 1974, ded. to K. Schütte, Lecture Notes in Mathematics 500, Springer, Berlin-Heidelberg-New York, 56-72.

K. GÖDEL 1958: *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes*, Dialectica 12, 280-287.

W.A. HOWARD 1968: *Functional interpretation of bar induction by bar recursion*. Comp. Math. 20, 107-124.

G. KREISEL 1959: *Interpretation of analysis by means of constructive functionals of finite type*, in: A. Heyting (ed.), *Constructivity in mathematics*, North-Holland, Amsterdam, 101-128.

S. KRIPKE 1965: *Semantical analysis of intuitionistic logic* I, In: *Formal systems and recursive functions*, North-Holland, Amsterdam, 92-130.

G. MINC 1974: *On* E-*theorems (Russian)* in: *Investigation in constructive mathematics* VI. Zapisky Nauk. Sem. Leningrad, Steklov Inst. 40, 110-118.

H. OSSWALD 1969: *Modelltheoretische Untersuchungen in der Kripke-Semantik*, Arch. math. Logik 13, 3-21.

P. RATH 1978: *Eine verallgemeinerte Funktionalinterpretation der Heyting-Arithmetik endlicher Typen*. Dissertation Münster.

J.R. SHOENFIELD 1967: *Mathematical Logic*. Addison-Wesley, Reading etc.

C. SPECTOR 1962: *Provably recursive functionals: a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics*, in: J.C.E. Dekker (ed.) *Recursive function theory*, Proc. Symp. Pure Math. V. AMS, Providence, R.I., 1-27.

M. STEIN 1976: *Interpretationen der Heyting-Arithmetik endlicher Typen,* Dissertation Münster.

M. STEIN 197?: *Interpretations of Heyting's Arithmetic - an analysis by means of a language with set symbols,* to appear.

A.S. TROELSTRA 1973: *Metamathematical investigation of intuitionistic arithmetic and analysis*. Lecture Notes in Mathematics 344, Springer, Berlin-Heidelberg-New York.

M. YASUGI 1963: *Intuitionistic analysis and Gödel's interpretation*. J. Math. Soc. Japan 15, 101-112.

ON THE INTERPLAY BETWEEN MATHEMATICS AND PROGRAMMING

E.W. Dijkstra

This talk is delivered under confusing circumstances. And the only way I can think of, of assisting you in not getting confused by these complicating circumstances is describing them explicitly in my introduction. The complication is that, while I would prefer to give a completely technical talk, its moral is heavily loaded from a political point of view: it is a technical talk to be delivered against almost overwhelming political odds. In order to make you understand all this we have to go back to about ten years ago, when Programming Methodology became a topic of explicit concern.

In the history of programming October 1968 has been a turning point. In that month a conference on the topic denoted by the newly coined term "software engineering" was held in Garmisch-Partenkirchen. The conference was sponsored by the NATO Science Committee. The conference was remarkable because a large fraction of the participants had positions so high in their local hierarchy that they could afford to be honest. As a result this was the first sizeable conference at which the existence of the so-called "software crisis" was openly admitted. The gloomy atmosphere of doom at that conference has been a severe shock for some of the participants; some left the place very depressed indeed. The majority, however, left the conference with a feeling of relief, some even in a state of great excitement: it had been admitted *at last* that we did not know to program well enough. I myself and quite a few others had been waiting eagerly for this moment, because now at last something could be done about it. For years we had already been worried by the consequences of the proliferation of error-loaded software, but there is very little point in trying to urge the world to mend its ways as long as that world is still convinced that its ways are perfectly adequate. It was at that conference in Garmisch-Partenkirchen, that the whole climate changed. Now, nearly a decade later, we can only conclude that the excitement was fully justified: it was indeed a turning point in the history of programming. Since that conference, programming has never been the same again.

In reaction to the recognition that we did now know how to program well enough, people began to ask themselves how a really competent programmer would look like. What would we have to teach if we wanted to educate a next generation of really competent programmers? This became the central question of the study that later would become known as "programming methodology". A careful analysis of the programmer's task was made, and programming emerged as a task with a strong mathematical flavour. As I have once put it "Programming is one of the hardest branches of applied mathematics because it is also one of the hardest branches of engineering, and vice versa". Why the programming task has such a strong mathematical flavour is something I shall indicate later.

A lower bound for what the adequate education of a really competent programmer should comprise was very convincingly established, but it was *not* an easy message to sell, because it demonstrated by necessity the total inadequacy of the education of what is known as "the average programmer". The world today has about a million "average programmers", and it is frightening to be forced to conclude that most of them are the victims of an earlier underestimation of the intrinsic difficulty of the programmer's task and now find themselves lured into a profession beyond their intellectual capabilities. It is a horrible conclusion to draw, but I am afraid that it is unavoidable.

The conclusion that competent programming required a fair amount of mathematical skills has been drawn on purely technical grounds and, as far as I know, has never been refuted. On emotional grounds which are only too understandable, many people have refused to draw the conclusion, and the conclusion is opposed to, not because its validity is challenged, but because its social consequences are so unpleasant.

The situation is immensely aggravated by changes in attitude towards science and technology in general, that took place during the sixties. In that decade we have seen a growing distrust of technology, a disillusion with science, which by the end of that decade caused political outbursts from which most universities haven't fully recovered yet.

For those who had hoped that the explosive growth of universities and other research establishments would automatically bear fruits in proportion to that growth, the results have indeed been disappointing, because, while the quantity grew, the average quality declined. Browsing through a scientific journal or attending a conference is nowadays rather depressing; there is no denying it: there is just an awfull lot of narrow mediocrity, of downright junk even. Many people seem to have failed to see, that it was not science itself, but only the explosive growth of the institutions that was to blame. Throwing away the child with the bathwater, they have declared war on science in its best tradition. They are openly antiscientific, antiacademic, very much against rigour and formal techniques, and they propose to be agressively creative, gloriously intuitive and nobly interdisciplinary instead. The cruel love of perfection and excellence, that used to characterize the hard sciences, are but elitist relics to be abolished as quickly as possible, and progressive from now onwards shall mean soft. The political slogans of the late sixties cast these views in a jargon that is still alive and still causes confusion.

The result of all this is that the message that "software", in spite of its name, requires a very hard discipline, is in many environments now politically unacceptable, and therefore fought by political means. In characteristically anonymous blurbs in periodicals of the Computer Weekly variety I find myself under political attack. "Dijkstra articulates the voice of reaction" is a mild one. "I am inclined to view Dijkstra [...] as intellectual product of the Germanic system" is much worse. And I arouse the "suspicion that [my] concepts are the product of an authoritarian upbringing" coming as I do from a country having "social philosophies touched by authoritarianism and the welfare state" etc. Nice is also the choice of adjectives when my efforts are described as "directed into turning a noble art into a rigid discipline". The first time I found myself confronted with the opinion that adhering to a formal discipline hampers creativity I was completely baffled, because it is absolutely contrary to my experience and the experience of the people I have worked with. I found the suggestion so ludicrous that I could not place it at all: it is so terribly wrong. Since then I have learned that as symptom of a political attitude it is quite well interpretable.

Having thus -I hope- cleared the sky from political encumbrances, I shall now turn to the technical part of my talk.

Why is programming intrinsically an activity with a strong mathematical flavour? Well, mathematical assertions have three important characteristics.

1) Mathematical assertions are always general in the sense that they are applicable to many -often even infinetely many- cases: we prove something for *all* natural numbers or *all* nondegenerate Euclidean triangles.
2) Besides general, mathematical assertions are very precise. This is already an unusual combination, as in most other verbal activities generality is usually achieved by vagueness.
3) A tradition of more than twenty centuries has taught us to present these general and precise assertions with a convincing power that has no equal in any other intellectual discipline. This tradition is called Mathematics.

The typical program computes a function that is defined for an incredibly large number of different values of its arguement; the assertion that such and such a program corresponds to such and such a function has therefore the generality referred to above.

Secondly: the specification of what a program can achieve for us must be pretty precise, if it is going to be a safe tool to use. Regarded as a tool its usage can only be justified by an appeal to its stated properties, and if those are not stated properly its usage cannot be justified properly. And here we have the second characteristic.

Thirdly: the assertion that such and such a program corresponds to such and such a function, although general and precise, is not much good if it is wrong. If the program is to be regarded as a reliable tool, our least obligation is a convincing case, that that assertion is correct. That program testing does *not* provide such a convincing case is well-known. The theoretically inclined can deduce this from the indeed incredibly large number of different argument values for which the function is typically defined; the more experimentally inclined can conclude this from more than twenty years of experience in which program testing as main technique for quality control has not been able to prevent the proliferation of error-loaded software. The only alternative that I see is the only alternative mankind has been able to come up with for dealing with such problems, and that is a nice convincing argument. And that is what we have always called Mathematics.

Here we touch upon the major shift in the programmer's task that took place during the last ten years. It is no longer sufficient to make a program of which you hope that it is correct -i.e. satisfies its specifications- you must make the program in such a way that you can give a convincing argument for its correctness. Superficially it may seem that this shift has made the task of the poor programmer only more difficult: besides making a program he has to supply a correctness argument as well. It may indeed be hard to supply a nice correctness argument for a *given* program; if, however, one does not add one's correctness concerns as an afterthought, but thinks about the correctness argument right at the start, the correctness concerns have proved to be of great heuristic value. And the wise programmer now develops program and correctness argument hand in hand; as a matter of fact, the development of the correctness argument usually runs slightly ahead of the development of the program: he first decides how he is going to prove the correctness and then designs the program so as to fit the next step of the proof. That's fine.

You may think that I have introduced a more serious difficulty by stating that the programmer should make his program in such a way that he can give "a convincing argument" for its correctness. Convincing to whom? Well, of course, only to those who care. But couldn't those have very, very different notions of what to regard as "convincing"? Has the programmer to provide as many different arguments as there may be people caring about the correctness of his program? That would make his task clearly impossible.

The task is, indeed, impossible as long as we don't distinguish between "conventional" and "convenient". What different people from different parts of the world have been used to varies so wildly, that it is impossible to extract a guiding principle from trying to present your argument in the most "conventional" way: their usual patterns of thinking are most likely inadequate anyhow. About convenience of a notation, about effectiveness of an argument, about elegance of a mathematical proof, however, I observed among mathematicians a very strong consensus -the consensus was, as a matter of fact, much greater than most of the mathematicians I spoke suspected themselves- and it is this consensus among mathematicians that has proved to be a very valuable guiding principle in deciding towards what type of "convincing argument" the programmer should be heading.

Let me now try to sketch to you the type of mathematics involved in arguing about programs. One way of viewing a program is as the rules of behaviour which can be followed by an automatic computer, which is then said "to execute" the program. The process taking place when a computer executes a program is called a "computation", and a computation can be viewed as a time-sequence or a long succession of different machine states. The part of the machine in which its current state is recorded is called the store -or:

the memory-; the store is very large because it must be able to distinguish between a huge number of different states.

In arguing about programs we have to characterize the set of machine states that are possible at various stages of the computational process. Individual states are characterized by the values of variables in very much the same way as the position of a point in a plane can be characterized by the value of its coordinates in a well-chosen coordinate system. There are in this analogy only two differences: while the coordinates in the Euclidean plane are usually viewed as continuous, the variables characterizing the state of the machine are discrete variables that can only take on a finite number of different values. And secondly: while in Euclidean plane geometry two coordinates suffice to fix the position of a point, in computations we typically need thousands or millions of different variables te record the current machine state.

In spite of the fact that that last difference is a drastic one, the analogy is yet a useful one. Everybody familiar with analytic geometry knows how specific figures, lines, circles, ellipses etc. can be characterized by *equations*: the figures are regarded as the subset of the points whose coordinates satisfy the equation. The analogy to the figure in analytic geometry is the subset of possible states at a certain point of progress of the computation, and in analogy to analytic geometry, such a subset is characterized by an equation: the subset comprises all states of the machine in which the values of the variables satisfy that equation.

The analogy can even be carried a little bit further: we all know how the ease with which a proof in analytical geometry can be carried out often depends on the choice of our coordinate system. The program designer has a similar freedom when he chooses the conventions according to which the variables he introduced shall represent the information to be manipulated. He can use this freedom to speed up the computation; he can also use it to simplify the equations characterizing the sets of states he is interested in. If he is lucky, or gifted, or both, his choice of representation serves both goals.

So much for the analogy; now for the difference. The number of variables he is dealing with is much larger than the two coordinates of plane geometry, and the subsets of machine states he needs to characterize have very seldomly an obvious regularity as the straight line, the circle, and the ellipse that analytic geometry is so good at dealing with. This has two immediate consequences.

First of all we need a much richer framework and vocabulary in which we can express the equations than the simple algebraic relations that carry analytic geometry. The framework is provided by the first-order predicate calculus, and the vocabulary by the predicates the programmer thinks it wise to introduce. That the first-order predicate calculus was the most suitable candidate for the characterization of sets of machine states was assumed right at the start; early experiences, however, were not too encouraging, because it only seemed practicable in the simplest cases, and we discovered the second consequence: the large number of variables combined with the likely irregularity of the subsets to be characterized very quickly made most of the formal expressions to be manipulated unmanageably long.

Let me put it in other words. The programmer is invited to apply the first-order predicate calculus; I am even willing to make a stronger statement: not knowing of any other tool that would enable to do the job, the programmer *must* apply the first-order predicate calculus. But he has to do so in an environment in which he is certain to create an unmanageable mess unless he *carefully* tries to avoid doing so (and even then success is not guaranteed!). He has to be constantly consciously careful to keep his notation as adequate and his argument as elegant as possible. And it is only in the last years that we are beginning to discover what that care implies. Let me give you a simple example to give you some feeling for it.

To begin with we consider a finite undirected graph at each vertex of which a philosopher is located: philosophers located at vertices that are directly connected by one edge of the graph are called each other's neighbours and no philosopher is his own neighbour. For the time being the life of a philosopher exists of an endless alternation of two mutually exclusive states, called "thinking" and "tabled".

In our next stage we allow edges to be marked or not, a marked edge meaning that the two philosophers at its two ends are both tabled, more precisely

P1: For any pair (A, B) of neighbours
"both A and B are tabled" = "the edge between A and B is marked".

We assume that the system is started in an initial state in which

1) all edges are unmarked

2) all philosophers are thinking.

As a result, P1 initially holds. Therefore P1 will continue to hold indefinitely, provided no philosopher transition from thinking to tabled introduces a violation of it. This is obviously achieved by associating with these transitions the following "point actions" -where no two different point actions are assumed to take place simultaneously-

T1: < mark the edges connecting you to tabled neighbours and switch from thinking to tabled >

T2: < unmark your marked edges and switch from tabled to thinking >.

The first transition now introduces a mark for every pair of tabled neighbours introduced by it, the second one removes a mark for every pair of tabled neighbours disappearing as a result of it. With these conventions the permanent truth of P1 is guaranteed.

From the above we see that a mark on the edge between the neighbours A and B has either been placed by A or by B. In our next stage we shall indicate which of the two has placed the mark by representing a marked edge between A and B by a directed edge, i.e. by placing an arrow along the edge. In this representation relation P1 is rephrased as

P1: For any pair (A, B) of neighbours
"both A and B are tabled" = "the edge between A and B is directed".

The direction of the arrow is fixed, by rephrasing the transitions as

T1: < direct arrows pointing towards your tabled neighbours and switch from thinking to tabled >

T2: < make all your edges undirected and switch from tabled to thinking >.

We observe that transitions T1 create arrows and only transitions T2 destroy them. More precisely: each arrow is created as an outgoing arrow of its creator, hence,

a philosopher without outgoing arrows remains without outgoing arrows until it performs itself its *own* transition T1.

We now subdivide the state "tabled" into the succession of two substates "hungry" followed by "eating", where the transition is marked by the observation of absence of outgoing arrows, more precisely

"philosopher A is tabled" = "philosopher A is hungry or eating"

and the life of a philosopher now consists of a *cyclic* pattern of transitions

T1: <direct arrows pointing towards your tabled neighbours and switch from thinking to hungry>

T1.5: <observe that you have no outgoing arrows and switch from hungry to eating>

T2: <remove all your incoming arrows and switch from eating to thinking>

and we establish the permanent truth of

P2: For any philosopher A we have
"philosopher A has no outgoing arrows" <u>or</u> "philosopher A is hungry".

In transition T1 the first term P2 may become false, but the second one becomes certainly true; in transition T1.5 the second term becomes false at a moment when the first term is true, a truth that cannot be destroyed by the other philosophers. In T2 the fact that initially the philosopher is eating tells us in combination with P2 that its arrows, if any, must be incoming arrows; hence, removal of your incoming arrows is the same as removal of all your arrows.

Relations P1 and P2 guarantee that no two neighbours can be eating simultaneously: if they were, they would both be tabled, hence there would be an arrow between them (on account of P1), for one of them it would be an outgoing arrow, but P2 excludes that an eating philosopher, which by definition is not hungry, has outgoing arrows.

(In addition we can prove that if the graph is finite and each eating period for each philosopher is finite, then each hungry period for each philosopher will be finite. This follows from the fact that the arrows never form a directed cyclic path.)

The way in which the above argument has been described illustrates one of the aspects of the "care" which is becoming typical for the competent programmer: "step-wise refinement" is one of the catchwords. Note that we have started the argument in terms of the still very simple concepts "tabled" and "marked". Only after the exhaustion of these two concepts, the state "marked" was split up into two mutually exclusive substates as represented by the two possible directions of an arrow along the edge. And only when the consequences of that refinement had been explored, the state "tabled" was subdivided into two mutually exclusive states, viz. "hungry" and "eating".

In the simple example shown such a cautious approach may seem exaggerated, but for the trained programmer it becomes a habit. In a typical program so many different variables are manipulated that the programmer would lose his way in his argument if he tried to deal with them all at once. He has to deal with so many concerns that he would lose his way if he did not separate them fairly effectively. He tries to keep his arguments simple compared to the final program by abstracting from all sorts of details that can be filled in later.

In yet another respect the above argument is typical. I did not tell you the original problem statement, but that was phrased as a synchronization problem, in which no two neighbours were allowed to eat simultaneously. The notion "hungry" has to be invented by the programmer; and then the argument is introduced by abstracting from the difference between "hungry" and "eating", in terms of the notion "tabled" that did not occur in the original problem statement at all. Such abstractions *must* be performed: instead of "tabled" one can say "hungry" or "eating", but the translation of "a pair of tabled neighbours" gives you some hint of the clumsiness thus engendered.

One last detail worth noticing is provided by our arrows. We had to introduce two different forms of marking: we could have done that with colours, say red edges and blue edges, but then we would have lost that my incoming arrows are my neighbours outgoing arrows, and the whole argument would have lost its clarity.

So much for the care needed to keep the arguments manageable: we can summarize it by stating that in programming mathematical elegance is not a dispensable luxury, but a matter of life and death.

In the example sketched the argument could be rendered nicely and compactly essentially thanks to the introduction of the proper nomenclature, but quite often more drastic steps have to be taken. In order to formulate the equations characterizing sets of possible machine states it is quite often necessary to change the program by the insertion of additional operations on so-called "auxiliary variables". They are not necessary for the computation itself, they are hypothetical variables whose values we can view as being changed in the course of the computational process studied. They record some aspect of the progress of the computation that is not needed for the answer, but for the argument justifying the program. Their values can appear in the characterizing equations in terms of which the correctness argument is couched. The introduction of the appropriate auxiliary variables is a next step in the progress of "choosing an adequate nomenclature"; the role of the auxiliary variables in proofs of program correctness is very similar to the role of auxiliary lines or points in geometrical proofs, and their invention requires each time a similar form of creativity. This is one of the reasons why I as a computing scientist can only regret that the attention paid to Euclidean geometry in our secondary school curricula has been so drastically reduced during the last decades.

In a recent correctness proof I had to go still one step further. I had to introduce auxiliary variables, but their values did not occur directly in our characterizing equations: in those equations occurred terms which had to be defined as the minimal solution of two sets of equations in which the auxiliary variables appeared as constants. As far as I am aware, that proof was the first one of its kind, but its discovery was a pure joy. It showed a counterexample to the commonly held but erroneous belief that formal correctness proofs for programs are only belabouring the obvious; it showed how the first-order predicate calculus was an indispensable and adequate tool, but, most important of all, it showed how a careful analysis of the syntactic structure of the predicates quite naturally led to all the additional logical gear to be invented.

In the interplay between mathematics and programming during the last ten years programming as an intellectual discipline has clearly been at the receiving end. A new area of intellectual activity has been discovered to be amenable to mathematical treatment, and thanks to the introduction of mathematical techniques we can now design programs that are an order of magnitude better than the ones we could design ten years ago. In the past the discovery of a new area of applicability of mathematics has always influenced and stimulated mathematics itself, and it is reasonable to wonder about the question what influence on mathematics may be expected this time.

I expect that the influence will be very wholesome. The programmer applies mathematical techniques in an environment with an unprecendented potential for complication; this circumstance makes him methodologically very, very conscious of the steps he takes, the notations he introduces etc. Much more than the average mathematician he is explicitely concerned with the effectiveness of this argument, much more than the average mathematician he is consciously concerned with the mathematical elegance of his argument. He simply has to, if he refuses to be drowned in unmastered complexity. From the programmer's exposure and experience I can expect only one influence on mathematics as a whole: a great improvement of the taste with which formal methods are applied. This improvement may very well turn out to be drastic. In texts about the philosophy of science from the first half of this century it is quite common to encounter a postulated antagonism between formal rigour on the one hand and "understandability" on the other. Already now, whenever I see such a text it strikes me as hopelessly out of date, arguing as it does against formal rigour instead of against ugliness: in those days the two were evidently often regarded as synonymous. And I have some indication that this improvement in taste is not only the dream of an optimist. I have conducted a little experiment with students from all over the world, in which I asked them to prove a nice little theorem from number theory that, although everyone can understand what the theorem states, happens to be unknown: the mathematicians with programming experience did markedly better than the mathematicians without that experience.

RULES FOR BUILDING STATISTICAL MODELS

J. Hemelrijk

"Statistics uses the empirical hypothesis that apparatus ('lotteries') exist, admitting random choices of one among any given number of elements. Such apparatus do not exist in absolute perfection and their degree of perfection can only be defined after development of their theory. Their role is analogous to that of rigid bodies in euclidean geometry and of perfect clocks in dynamics. Empirical interpretation of probability statements is only possible with reference to such random apparatus or to natural phenomena empirically found to behave statistically sufficiently like these".

D. van Dantzig (1957)

1. INTRODUCTION

The use of *mathematical models* is widespread and of an old date, but the general recognition of this fact is comparatively new. The question of how to choose a statistical mathematical model has led to considerable confusion and controversy, and still does. Mathematical statisticians wisely save their skins by using the axiomatic approach, leaving the controversy to others and the confusion to the users of their theory. For axioms, however useful, say nothing about their application. It seems to the author that the time has come to formulate *rules* for the choice of statistical models. In this paper a number of such rules are proposed. They will

---

*) Report nr SW 53 of the Dept. of Math. Statistics of the Mathematical Centre, Amsterdam. This paper appears also in Statistica Neerlandica 32, nr. 3 (1978).

certainly not please everybody, if only because they are formulated from the classical, objectivistic, point of view. They may, however, strengthen and clarify this point of view and help the user of statistical theory in its correct use and interpretation.

The subject is an extensive one, which can only be touched upon in a short paper. Therefore many details have to be taken for granted, the history of the subject is left aside and the controversy between objectivists and subjectivists is ignored.

In general a mathematical model is a *simplification* and an *exactification* of a part of reality. The simplification is necessary because of the extreme complexity of reality and the exactification because of its vagueness. Reality is always a bit out of focus: equality, for instance, is usually approximate equality and therefore not strictly transitive. In a mathematical model transitivity of equality and other desirable properties hold exactly and this makes it possible to develop extensive theories. But one should keep reality and model strictly apart. Confusing the two leads to baffling paradoxes - some of them well known - which can only be solved by disentangling reality and model.

Statistical models are concerned with parts of reality which are subject to uncertainty and which we will call (statistical) experiments. The possible outcomes of a statistical experiment are usually known, but the actual results are in a higher or lesser degree unpredictable. Causality does not seem adequate for analysing such experiments; instead the probabilistic approach is used.

In the following sections rules for using this approach are formulated step by step. These rules are not part of mathematics. They are not theorems nor are they laws of nature. They may be seen as *directions for use* of statistical models. They are certainly not perfect (nothing is) and their use cannot be enforced. But they are useful as a guide for sensible application of statistical methods.

2. RANDOMIZERS

Pure unpredictability in a statistical sense is found in a lottery, or randomizer. Everybody knows what a lottery is, but nevertheless it is suprisingly difficult to give a satisfactory description of its properties. A separate paper would be needed to this end. Let us just point out some properties which we need for later justification of our rules.

A *randomizer* is a machine which can be used repeatedly and which

a) *every time when it is activated gives one of a fixed finite set of results, but*

b) *every time the result to be obtained is completely unpredictable in the sense that any method of prediction is just as bad as any other and knowledge of past results does not help at all.*

The vagueness of this description is a nice and proper example of the vagueness of reality mentioned in section 1. The essence of a randomizer is negative: it is impossible to find two systems of prediction which differ systematically with respect to their number of sucesses. Nobody has succeeded in bringing this negative property of a randomizer in a positive form, except in the framework of a mathematical model.

Consider a randomizer with N possible results: the numbers $0,\dots,N-1$, say. Call this an N-*randomizer*. Then the following properties hold:

c) *If an* N-*randomizer is activated* n *times in succession and the* N^n n-*tuples of possible results are considered as possible outcomes of this composite experiment, then an* N^n-*randomizer is obtained.*

d) *If, with* $M < N$, *an* N-*randomizer is activated until one of the numbers* $0,\dots,M-1$ *is obtained, (skipping all outcomes but the last one) this composite experiment constitutes an* M-*randomizer.*

These properties, follow logically from property b), but they can also be verified empirically if one wishes to take the trouble. It may be remarked that the experimental law of large numbers, the approximate long-run equality of the relative frequencies of the N outcomes of a randomizer, also follows logically from property b). The reverse is not true: periodic processes (e.g. the hours indicated by a watch) obey the experimental law of large numbers but do not have property b). Thus the unpredictability of separate results is more fundamental than the experimental law of large numbers.

3. THE BASIC STATISTICAL MODEL

As far as the statistical model for a randomizer is concerned there is general concensus. The N possible outcomes are assigned equal probabilities 1/N in agreement with the *classical probability definition of Laplace:*

the probability of an event is equal to the ratio of the number of possible outcomes favorable for the event to the total number of possible outcomes.

More precisely, if we call the activation of an N-randomizer: *"drawing at random from* $0,\ldots,N-1$", then the model for one random drawing consist of three elements:

1) *The space of (elementary) events:* $\Omega = \{0,\ldots,N-1\}$.*)
2) *Composite events: all subsets of* Ω.
3) *The Laplace-definition assigning a probability to every event:*

(1) $P(\Omega') \stackrel{\text{def}}{=} N(\Omega')/N \qquad \Omega'\subset\Omega$

with $N(\Omega')$ = the number of element of Ω'.

The basic threefold structure of this model holds for all statistical models, though usually in a more complicated form. It is also completely in harmony with the axiomatic set-up. We call this model a *finite symmetric probability space* and our first rule is:

Rule 1. *For one random drawing we use a finite symmetric probability space as mathematical model.*

Now consider a sequence of n random drawings, resulting in an n-vector of numbers from Ω. According to property c) of section 2 this composite experiment is the same as one random drawing from the N^n possible n-vectors. Thus rule 1 also gives us the model for this sequence of drawings. If one works this out the result is the product probability space of n finite symmetric probability spaces, one for each of the n random drawings. We omit the details; they are well-known to every statistician and we want to hurry on to more important points. But we do remark that the reasoning also holds for a sequence of random drawings from different randomizers and that we arrive thus at our second rule:

Rule 2. *For a sequence of* n *random drawings we use as a model the product probability space of the* n *symmetric probability spaces of the separate drawings.*

Remark that the term "independent" need not yet be introduced at this stage; it is implicit in property b) and emerges explicitly in a natural way when later on conditional probabilities are introduced. At the present stage one

*) "(possible) result", "(possible) outcome" and "elementary event" are used as synomyms.

might say that a randomizer is independent of everything: it walks, like a cat, by itself.

4. THE PRINCIPLE OF EQUIVALENCE

The transition from rule 1 to rule 2 has been accomplished by stating that - according to property b) - n random drawings from $0,\ldots,N-1$ "are the same as" one random drawing from $0,\ldots,N^n-1$ (after numbering the n-vectors in an arbitrary order). This expression "the same as" is not very accurate; the two experiments compared are not the same, but they both have the properties of a randomizer. In a certain sense they are equivalent with respect to their statistical properties. It is worth while to elaborate on this point because it leads us to one of the key-points of our set-up.

Consider two repeatable experiments E' and E" with the same possible outcomes ($\Omega' = \Omega''$) but otherwise possibly very different. Let the following information be supplied:

1) *an accurate description of* E' *and* E",
2) *two sequences of results* A *and* B *from these experiments, however without identification; this means that it is not known whether* A *and* E' *(and* B *and* E"*) belong together or the other way around.*

Additional information is supplied on request:

3) *further details about* E' *and* E",
4) *extensions of the sequences* A *and* B *(again without identification),*
5) *sequences* C' *from* E' *and* C" *from* E".

If in this situation there is no conceivable method of identifying the sequences A and B, then E' and E" are called (statistically) *equivalent*. Their statistical behaviour with respect to the possible outcomes considered, is the same. The generalization to more than two experiments is straightforward and we can now formulate:

The principle of equivalence. *If experiments are equivalent in the sense described above, then the use of the same model for all of them is justified.*

External reasons like practical importance and cost of time and money may lead to the use of different models when, statistically speaking, the use of the same model would be desirable. In this paper, however, we will strictly adhere to the principle of using the same model for equivalent experiments.

A lot more can be said about the concept of equivalence, but a practical example may at this point be more clarifying. Five experiments have, to this end, each been excecuted 221 times. They are

E_1: *recording the last digit of the hodometer*) of the authors car when he left the car for more than half an hour.*

E_2: *recording at the same moments, the last digit of the sub-hodometer, which records the same distance in units of* 100 *m (mod* 10 000*).*

E_3: *recording the last digit of the hodometers of cars in public parking lots.*

E_4: *throwing a blue tensided die carrying the numbers* 0,...,9.

E_5: *throwing a red tensided die with the same numbers.*

For every experiment the results of 221 excecutions were recorded in the order of their observation. The dice were well made and they were thrown in such a way that E_4 and E_5 may be considered to be 10-randomizers. For these two experiments equivalence is clear: from property b) of section 2 it follows that all N-randomizers are equivalent (for any fixed N). It is not very plausible that E_1 and E_2 are equivalent to E_4 and E_5, but E_3 might well be. For although E_3 is much more complicated then E_4 and E_5 it is difficult to imagine why it would be possible to find two systems of predicting the next outcome of E_3 one of which is better than the other. This might well be possible for E_1 and E_2.

It is clear that speculations of this kind are not a sufficient basis for deciding about equivalence. The observations themselves, however, may help. And one of the tasks of statistical theory is to provide methods to test equivalence of experiments and the goodness of fit of models to experiments. These methods are indeed available and one of them can be used in our case. In order to confuse the reader the five sequences have been assigned labels A, B, C, D, E at random. Table 1 contains the observations in their original form. It is difficult to draw any conclusions directly from these date. They have been completely recorded in Table 1 in order to enable the reader to play around with them himself. A first step in getting a better survey of the data is to arrange them in a frequency table.

*) The hodometer cumulatively counts the distance covered by the car in km (mod 100 000).

This has been done in table 2, where two columns have been added, one for the well-known test statistic χ^2 and one for the right-hand tail-probability P. Extreme values of P indicate deviations from randomness; the number of degrees of freedom is 9, the hypothesis tested: randomness.

Table 1. Five sequences of observations

A	9 0 4 9 3 3 6 1 8 7 7 5 9 5 3 8 1 8 5 3 9 9 7 9 2 3 5 2 7 0 0 6 4 6 5 5 1 9 4 5 4 7 0 4 2 5 1 7 8 5 6 3 1 4 6 1 3 3 4 7 3 1 9 3 8 9 8 0 6 0 6 5 6 6 4 3 9 0 0 3 8 9 9 0 8 5 3 4 2 5 5 4 7 2 5 2 0 8 7 3 7 0 6 9 6 4 9 1 3 2 3 9 7 0 3 2 6 6 2 8 9 6 8 8 2 8 1 0 8 5 3 3 8 8 0 5 7 7 2 4 7 1 0 6 6 0 5 5 6 6 4 4 9 9 0 7 8 3 7 4 5 2 3 1 3 7 9 4 9 8 4 0 4 6 4 1 3 7 9 3 3 0 8 4 3 8 7 4 4 4 7 9 3 8 8 7 9 1 2 7 4 6 1 8 2 7 7 8 8 7 8 3 3 4 7 8 9 3 2 2 7
B	7 1 2 8 8 1 8 9 5 4 0 6 0 0 0 1 1 6 7 4 7 5 2 8 3 1 3 5 4 6 7 7 2 9 4 9 9 9 1 2 4 7 3 1 6 4 9 7 6 3 8 8 3 6 8 9 6 9 5 9 3 8 5 1 7 9 1 7 3 6 8 1 8 8 2 1 3 0 2 3 0 4 5 0 7 8 0 5 2 9 3 4 3 4 5 1 8 6 1 3 8 5 3 2 0 1 3 9 6 7 9 2 4 6 4 0 4 6 3 9 5 7 8 4 1 6 1 2 2 7 3 9 4 2 6 5 1 9 6 2 8 7 2 8 6 4 5 5 0 7 0 0 4 9 5 1 5 8 7 9 4 8 3 9 3 2 3 2 6 6 8 5 1 5 2 9 8 2 7 7 5 2 0 2 0 6 1 9 5 1 7 4 0 1 7 4 4 2 1 4 0 6 3 6 0 9 8 2 2 5 1 1 8 5 1 6 0 8 7 9 8
C	4 5 7 9 0 2 7 3 7 3 0 1 1 0 9 5 5 4 0 9 2 3 4 1 9 8 8 8 6 0 6 3 5 0 6 3 6 4 3 0 9 6 4 0 1 6 1 5 0 2 3 8 9 7 6 8 7 8 9 6 2 6 0 2 5 5 8 8 3 9 3 2 2 0 5 2 1 2 7 1 0 1 7 8 9 2 0 1 3 8 2 1 1 3 6 1 5 0 4 4 8 3 5 6 7 5 5 9 8 5 7 7 2 3 8 8 1 7 8 0 0 0 0 7 8 4 9 9 9 5 0 1 6 7 1 5 8 2 9 2 3 6 4 9 6 5 5 5 9 1 9 2 6 5 6 1 7 9 7 3 0 4 0 8 2 4 3 2 3 5 7 0 7 0 7 8 3 0 6 7 2 7 1 9 4 9 9 9 9 4 1 9 7 6 9 2 1 5 8 3 7 1 0 2 0 6 6 4 1 2 1 0 0 9 0 6 2 7 1 5 1
D	5 6 6 0 8 2 6 5 3 9 2 6 5 5 4 5 4 8 2 9 5 6 7 7 2 9 9 8 7 3 0 3 5 9 2 9 3 2 2 4 1 5 9 6 2 8 5 1 1 4 7 9 3 3 4 6 4 8 4 7 4 0 4 8 1 5 0 3 7 7 3 1 4 3 6 5 9 3 2 3 9 8 4 8 1 1 4 3 9 2 3 4 3 9 5 9 2 9 6 4 5 9 5 7 3 8 3 6 0 6 2 3 4 5 4 1 6 0 4 7 1 7 3 7 1 2 4 4 4 7 1 5 0 1 3 2 5 9 1 7 0 4 8 7 9 8 3 3 7 0 2 8 7 6 5 9 2 5 1 0 1 9 3 6 5 6 8 8 4 9 2 8 1 9 3 6 5 1 4 2 6 0 3 3 2 7 7 0 4 8 1 5 2 2 5 9 8 2 4 6 1 0 4 7 5 7 5 3 3 7 1 0 5 9 3 6 9 9 8 7 6
E	7 3 9 5 9 5 1 3 4 7 0 0 0 9 5 1 9 6 5 9 0 1 9 6 9 4 4 1 6 3 7 0 0 4 9 0 7 8 6 7 3 3 1 0 8 7 3 2 5 1 9 7 9 1 0 4 7 9 6 2 1 4 2 2 2 9 9 3 5 9 6 2 1 8 1 0 1 9 4 1 0 3 7 6 3 4 3 1 8 8 9 8 8 8 0 5 1 6 5 1 3 8 0 5 8 6 8 2 5 7 6 5 8 2 9 3 0 9 7 5 5 9 3 1 9 4 2 4 4 3 5 5 3 8 9 7 0 1 4 6 2 5 3 4 7 0 1 0 5 9 4 4 4 6 3 0 0 9 4 9 1 3 6 4 8 4 5 8 2 2 3 5 8 8 2 6 8 3 5 9 7 3 8 4 7 0 3 4 7 9 2 3 7 3 3 8 8 7 9 9 2 5 4 1 9 2 1 4 9 2 1 6 0 3 4 6 6 5 3 4 0

Table 2. Frequencies of 0,...,9 in the five sequences

	0	1	2	3	4	5	6	7	8	9	χ^2	P
A	19	14	16	30	25	19	20	27	27	24	11.26	0.26
B	19	26	23	20	21	21	22	21	25	23	1.94	0.992
C	29	25	22	19	14	22	21	23	20	26	6.92	0.65
D	14	20	22	28	26	26	20	22	18	25	7.46	0.59
E	22	22	17	27	25	22	17	18	21	30	7.46	0.59

None of the frequencies in table 2 deviates extremely from its mean 22.1. None of the values P is very small. One, however, pertaining to sequence B, is very close to 1, indicating some source of regularity which cannot be expected in a randomizer. Thus B may well stem from E_1 or E_2. But the result is still very undecisive. Therefore we go one more step in our analysis, aiming straightly at a point where E_1 and E_2 may well be very different from E_3, E_4 and E_5. For every pair of consecutive results, x_1 and x_2 say, we form the difference x_2-x_1 (mod 10). This gives us five new sequences of 220 results each. We need not give a table of these in the form of table 1, because the reader can easily write this down himself. The new sequence A would start with: 1 4 5 4 0 ... This operation applied to successive results of a 10-randomizer gives again a 10-randomizer. This can easily be proved by means of the model implied by rules 1 and 2. It can also be viewed as a property like c) and d) of section 2; the reader can easily verify this by some thinking. On the other hand it is very plausible that this does not hold at all for E_1 and E_2 because the author often travels the distance from home to work by car.

Table 3. Frequencies of 0,...,9 in differencies mod 10

	0	1	2	3	4	5	6	7	8	9	χ^2	P
A	28	26	22	21	18	24	20	29	24	18	4.82	0.89
B	17	20	23	12	19	28	37	28	18	18	21.27	0.012
C	24	29	18	16	25	28	22	18	16	24	9.36	0.40
D	18	15	14	35	49	10	26	11	11	31	68.64	2.6×10^{-9}
E	23	19	22	31	22	20	19	16	23	25	6.82	0.66

The frequencies of the five new series are given in table 3. Now the situation is completely changed. In D the differences 3 and 4 are very predominant and in B the same holds, but less strongly, for 5, 6 and 7. The value of P is very small for D and small for B; there is little doubt that D and B stem from E_1 and E_2, possibly even in this order. Additional information of the types 4) and 5) mentioned above would most probably lead to a decision in this question. Thus our conclusion is that E_4, E_5 and E_6 may well be considered equivalent, but E_1 and E_2 certainly are not equivalent, neither to each other nor to the other three. If the reader would wish to try to identify E_3 among A, C and E, he can provide additional observations of E_3 himself.

Anticipating an objection to the principle of equivalence we may concede that it will never be possible to prove conclusively that two experiments are equivalent. But then, absolute certainly about such things is not part of this life. If experiments are *deemed* equivalent for sufficiently sensible reasons and if observations in sufficient numbers do not contradict this, then the principle can be used. For on the other hand non-equivalence can be proved experimentally to a reasonable degree of certainty, as the example illustrates.

5. PROBABILITY SPACES WITH UNEQUAL PROBABILITIES

To arrive at probability spaces with *unequal probabilities* for the elementary events, the space of events Ω of a symmetric probability space is partitioned into a set of non-overlapping subsets. These, together with their probabilities form a new probability space. The *addition law for exclusive events*, which in the symmetric probability space follows from the Laplace-definition, is carried over to the new probability space and this leads us to finite discrete probability spaces. The principle of equivalence then justifies the use of such a space as a model for experiments where a lack of symmetry does not suggest the use of equal probabilities at all. A simple example: let experiment E' be throwing a loaded six-sided die, E" using an N-randomizer with sufficiently big N with $\Omega = \{0,\dots,N-1\}$ partitioned into six subsets with unequal numbers of elements $n_1,\dots,n_6$, carrying the numbers $1,\dots,6$. The contention is that for suitably chosen N and $n_1,\dots,n_6$ the two experiments are equivalent, thus justifying the use of a discrete probability field for E'. Of course a suitable choice of $n_1,\ldots,n_6$ (and N) would have to depend on observations of E' but that only emphasizes the need of testing the goodness of fit of a model which has been chosen on the basis of rules and practical considerations. We thus arrive at:

Rule 3. If an experiment is equivalent to a suitably chosen partitioned randomizer then we use a discrete probability space as mathematical model.

6. STATISTICAL INDEPENDENCE

At the end of section 3 it was remarked that the independence of successive uses of randomizers is implicit in the properties of a randomizer. It is expressed in property b) by means of the fact that the past does not help to predict the future. This concept must be generalized and more formally expressed:

DEFINITION. *Consider* n *experiments* $E_1,\ldots,E_n$, *each of which separately is adequately described by a completely specified probability space; if knowledge of the results of any part of these experiments (after they have been performed) does not influence the predictability of the results of any of the others, then the experiments are called statistically independent.*

This, again, is a practical concept of considerable vagueness, which needs exactification by means of a mathematical model. It is clear from the definition and the previously formulated rules that the whole sequence $(E_1,\ldots,E_n)$ is equivalent to n random drawings from suitably chosen partitioned randomizers and thus rule 2 indicates the use of the product-space:

Rule 4. If n *statistically independent experiments are each described by a probability space the combined experiment* $(E_1,\ldots,E_n)$ *is described by the product of these probability spaces.*

Omitting, in this rule, the term "completely specified", which figures in the above definition, only means a slight generalization. The term cannot be omitted from the definition: if there are unknown parameters involved previous experiments - independent or not - may supply information about these parameters and thus influence the predictability of the other experiments. This would for instance occur in a sequence of throws of the loaded die used as an example in section 5, where nevertheless successive throws would be independent.

7. CONDITIONAL PROBABILITIES AND COMPOSITE MODELS

From independence to dependence is only one step but a very important one. An example of statistical dependence is found in repetitions of E_1 of section 4; line D of table 3 (which does in fact pertain to E_1) clearly indicates that adding 4 to the previous result (mod 10) is certainly superior as a method of prediction to adding 5. To build models for dependent experiments we need conditional probabilities.

Usually conditional probabilities are introduced in the model by means of a definition. Let Ω_1 and Ω_2 be subsets of the space of events Ω then the *conditional probability* of finding an element of Ω_2 under the condition that an element of Ω_1 occurs is

$$(2) \qquad P(\Omega_2|\Omega_1) \overset{\text{def}}{=} P(\Omega_1 \cap \Omega_2)/P(\Omega_1),$$

where $P(\Omega_1)$ must be positive. This definition in itself says nothing about the way it should be used in applications. We therefore present a justification of (2) based on our rules, which also leads to a new rule giving insight in the way it should be used for model-building.

Consider the following two experiments.

E': *drawing one element at random from* Ω_1 *(using an* $N(\Omega_1)$*-randomizer for the purpose),*

E": *drawing elements at random from* Ω *(by means of an* $N(\Omega)$*-randomizer) until for the first time an element from* Ω_1 *is obtained and considering this element as the outcome of the composite experiment.*

According to property d) of section 2 E' and E" are equivalent and thus we ought to use the same model for both of them. But according to rule 1 the model for E' is a symmetric probability space with Ω_1 as space of events and with the Laplace-definition. This means that we should also use this model for E" and this is exactly what happens. The notation "$|\Omega_1$" is used to indicate the conditioning on Ω_1 in either of the two ways indicated by E' or E". The Laplace-definition applied to E' now leads straight to (2), for according to this definition we have

$$P(\Omega_2|\Omega_1) = N(\Omega_1 \cap \Omega_2)/N(\Omega_1) =$$

$$= \{N(\Omega_1 \cap \Omega_2)/N\}/\{N(\Omega_1)/N\} =$$

$$= P(\Omega_1 \cap \Omega_2)/P(\Omega_1),$$

where the unconditional probabilities pertain to one random drawing from Ω. Note that neither E' nor E" can be performed if $N(\Omega_1) = 0$; thus the reasoning only holds if $P(\Omega_1) > 0$.

The equivalence of E' and E" seems rather evident but the following anecdote shows that this does not hold for everybody*). An advertising agency organized a quiz in order to promote some product. The quiz consisted of some simple questions and the response was overwhelming. Thousands of answers were received and, of course, the prizes had to be awarded at random among the correct solutions. To this end the agency hired a number of working students in order to sift out the wrong answers (which were comparatively few). This took several weeks time and when this work was completed the winners were drawn at random from the correct solutions. This procedure corresponds to E' and it is perfectly correct. How much more simple and less time-consuming it would have been, however, to use procedure E"!

The generalization of (2) to partitioned probability spaces is straightforward. We will skip it. It is also clear that from (2) the *general multiplication law* and the theorem on composite probabilities follow and that statistical independence means that conditional probabilities are equal to the corresponding unconditional ones.

After these preparations we want to formally introduce the use of conditional probabilities in building up models for stepwise experiments. Let $E^{(1)}$ be an experiment with $\Omega^{(1)}$ as its space of events and $P^{(1)}$ as its probability function on $\Omega^{(1)}$, all according to previous rules. Let $E^{(2)}$ be a second experiment with space of events $\Omega^{(2)}$, but depending on the result of $E^{(1)}$ in the following sense: for every $\omega^{(1)} \in \Omega^{(1)}$ an experiment $E^{(2)}_{\omega^{(1)}}$ is given which has a probability function $P^{(2)}_{\omega^{(1)}}$ (on $\Omega^{(2)}$), depending on $\omega^{(1)}$, again in accordance with previous rules. The composite experiment $E = (E^{(1)}, E^{(2)})$ is composed of $E^{(1)}$ and $E^{(2)}_{\omega^{(1)}}$, where $\omega^{(1)}$ is the event realized in $E^{(1)}$. In these circumtances $E^{(2)}$ is called *statistically dependent* on $E^{(1)}$ and E is called a *stepwise composed experiment*. By induction

*) H. Piller, personal communication.

we get any finite number of steps.

The model for E must be in accordance with previous model-rules and to find such a model we again consider two equivalent experiments:

E': *one realisation of* $E^{(2)}_{\omega^{(1)}}$ *for given* $\omega^{(1)}$; *the probability for obtaining* $\omega^{(2)} \in \Omega^{(2)}$ *is then*

$$P^{(2)}_{\omega^{(1)}}(\omega^{(2)}).$$

E": *repeating* E *until* $E^{(1)}$ *gives* $\omega^{(1)}$ *and looking at the result of* $E^{(2)}$ *in that trial. This means:imposing the condition* $\omega^{(1)}$ *on* E, *and thus the model for* E *must be such that the probability for obtaining* $\omega^{(2)} \in \Omega^{(2)}$ *is*

$$P(\omega^{(2)} | \omega^{(1)}).$$

The equivalence of E' and E" now leads to

$$(3) \qquad P(\omega^{(2)} | \omega^{(1)}) = P^{(2)}_{\omega^{(1)}}(\omega^{(2)})$$

and together with the multiplication law, which must also hold in the model for E, we find that we have to build up this model on $\Omega^{(1)} \times \Omega^{(2)}$ by means of

$$(4) \qquad P((\omega^{(1)}, \omega^{(2)})) = P^{(1)}(\omega^{(1)}) P^{(2)}_{\omega^{(1)}}(\omega^{(2)}).$$

This is the only possibility if we want to obey our previous rules and the principle of equivalence. This result can be summarized as follows.

Rule 5. *For stepwise experiments where for every step previous rules lead to a probability space depending on the results of previous steps, a model is built up by means of conditional probability spaces for the steps and by means of the multiplication law for simultaneous probabilities.*

8. CONDITIONAL PROBABILITIES AND INFORMATION

Although some details were glossed over in section 7 the treatment of a seemingly obvious method may seem rather extensive to some readers. But one must be careful as the following example is meant to show. A player throws a good six-sided die and you are to guess the result. This is the situation of section 2: your guess does not really matter as long as it is one of the numbers 1,...,6; the die is a randomizer. But now the player throws the die, has a look at the result without enabling you to do the same and he gives you the following information: "it is not 6". Now let us suppose that he does not lie (that is another game), how does this information influence your guess? I asked a number of statisticians this question and all of them agreed that the guess would now be made as if 1,...,5 were equally probable. This would seem in accordance with property d) of section 2 and with the use of conditional probabilities as in the previous section. But this method may be completely wrong. Suppose the player uses the following *information-policy*: if the result is 1 he says "it is not 6", if it is not 1 he says "it is not 1". Then he will in any case give true information, but it will be very misleading if the result is 1. Equal conditional probabilities are then *not* the right model. Without the information the model was clear: a random drawing from 1,...,6. After the information one is completely muddled. Giving out information should be accompanied by the knowledge of the information-policy.

DEFINITION. *Information about the result of an experiment is only reliable if the receiver knows the information-policy used, i.e. if he knows which information would be given for every possible result of the experiment.*

This boils down to a partitioning of Ω into subsets of elements with the same information. The information then indicates in which of these known subsets the result occurs and then conditional probabilities can be used without fail.

Rule 6. *Information about the result of an experiment can only be expressed in the model by means of specified conditional probabilities if the information is reliable in the sense of the above definition.*

If information is not reliable it is unreliable. This is how some news media and politicians misinform people without actually telling lies: what they say is formally true but if one does not know their information-policy or forgets to keep it in mind, their truth is twisted and may be very misleading.

Lack of knowledge about the information-policy can be incorporated adequately in the model by introducing an unknown partitioning of Ω, i.e. unknown conditions for the conditional probabilities. In our example with the die this would, in case "it is not 6", lead to five possible outcomes with unknown probabilities, which can have only a finite number of different values because there are only a finite number of possible partitionings. Further knowledge about the actual values but also about the actual information-policy could then be gathered by observing repeated independent trials of the same experiment. On the other hand it may be remarked that one may remedy the situation by randomizing ones guess among the numbers 1,...,5. Then at least the probability of a right guess is 1/5. Thus perhaps, one should never read a newspaper without a die or a coin at hand.

9. FINAL REMARKS

Although up till this point we only have finitely many rational probabilities in a probability space the generalization to infinitely many real ones and to continuous probability spaces is of a less fundamental nature. It is all passing to the limit and approximating discrete situations by means of continuous ones for the sake of mathematical convenience and greater generality. So we need not be sorry that the scope of this paper does not allow us to go over all that. It is a pity that the space allotted is too small to talk about some other things like: the interpretation of probabilities in order to go back from the model to reality after the analysis in the model is completed and to the phenomenon that statisticians do not only seek to predict the future, but also the past: the example in section 7 is of that character just as e.g. the method of confidence intervals for unknown parameters. These things are interesting but they will have to wait.

REFERENCES

D. VAN DANTZIG (1957), *Statistical priesthood (Savage on personal probabilities)*, Stat. Neerl. 11, 1-16.

J. HEMELRIJK (1968), *Back to the Laplace definition*, Stat. Neerl. 22, 13-21.

B. VAN ROOTSELAAR and J. HEMELRIJK (1969), *Back to "Back to the Laplace definition"*, Stat. Neerl. 23, 87-89.

SYSTEMS OF CONFIGURATIONS

D.G. Higman

Generic adjacency algebras of systems of homogeneous coherent configurations are defined analogously to the generic Hecke algebras studied by CURTIS, IWAHORI and KILMOYER [1] and can be viewed as generalizations of them. To treat common properties of ordinary and generic adjacency algebras we introduce virtual adjacency algebras. In particular, versions of theorems of Frame and Wielandt for ordinary adjacency algebras as in [4] and for generic adjacency algebras can be obtained as consequences of versions for virtual adjacency algebras. Details of the material outlined here will appear (amongst other things) in [5]. General references for coherent configurations are [3] and [4].

1. SYMMETRIC ALGEBRAS

Let F be a field of characteristic 0 with algebraic closure $\bar{F}$, and let $\mathbb{A}$ be an associative F-algebra with identity element 1 and a fixed F-basis $w_1,\ldots,w_r$ which will be refered to as the *standard basis*. We have $w_i w_j = \sum_{k=1}^{r} a_{ijk} w_k$ with the *structure constants* $a_{ijk} \in F$. A good example to keep in mind is the group algebra of a finite group G with the elements of G as standard basis.

Assume given a linear functional $\zeta\colon \mathbb{A} \to F$ such that the bilinear form $(x,y) = \zeta(xy)$, $x, y \in A$, is symmetric and nondegenerate. We refer ζ as a *virtual trace* on $\mathbb{A}$. Then $\mathbb{A}$ is a symmetric algebra with dual basis $w_1,\ldots,w_r$ defined by $\zeta(\hat{w}_i w_j) = \delta_{ij}$. In the case of a group algebra, if we take ζ to be the trace of the regular representation, then $\hat{g} = \frac{1}{|G|} g^{-1}$.

Now assume that $\mathbb{A}$ is semisimple. Let $\Delta_1,\ldots,\Delta_m$ be the inequivalent absolutely irreducible representations of $\mathbb{A}$ in $\bar{F}$, and let $\zeta_1,\ldots,\zeta_m$ be the corresponding absolutely irreducible characters, with $\zeta_s(1) = e_s$ so $\sum_{s=1}^{m} e_s^2 = r$. Write $\Delta_s(w) = (a_{ij}^s(w))$ for $w \in \mathbb{A}$ and list the r coefficient functions a_{ij}^s, $1 \le i,j \le e_s$, $1 \le s \le m$: $a_1,a_2,\ldots,a_r$. Put $a_{\bar{\lambda}} = a_{ji}^s$ if

$a_{\bar{\lambda}} = a_{ij}^{s}$. For linear functionals ϕ,ψ: $\underline{A}_{\bar{F}} \to \bar{F}$ write $(\phi,\psi) = \sum_{i=1}^{r}\phi(\hat{w}_i)\psi(w_i)$. Then we have the *Schur relations*

$$(1) \qquad (a_\lambda, a_\mu) = \delta_{\lambda\bar{\mu}} \frac{1}{h_\lambda}$$

and the *orthogonality relations*

$$(2) \qquad (\zeta_s, \zeta_t) = \delta_{st} \frac{e_s}{z_s}$$

with $h_\lambda = z_s$ if $a_\lambda = a_{ij}^{s}$.

The constants z_s depend on ζ but not on the particular basis. These are the usual Schur and orthogonality relations in the case of a group algebra.

2. COHERENT CONFIGURATIONS

A *combinatorial configuration* $(X,\mathcal{O})$ consists of a nonempty finite set X and a set $\mathcal{O}$ of nonempty binary relations on X, and $(X,\mathcal{O})$ is *coherent* if

(I) $\mathcal{O}$ is a partition of X^2,

(II) $\alpha \in \mathcal{O}$, $\alpha \cap 1 \neq \emptyset \Rightarrow \alpha \subseteq 1$, where $1 = 1_X$ is the diagonal subset of X^2,

(III) $\alpha \in \mathcal{O} \Rightarrow \alpha^* = \{(y,x) \mid (x,y) \in \alpha\} \in \mathcal{O}$, and

(IV) $\alpha_i, \alpha_j, \alpha_k \in \mathcal{O}$, $(x,y) \in \alpha_k \Rightarrow a_{ijk} := |\{z \in X \mid (x,z) \in \alpha_i$ and $(z,y) \in \alpha_j\}|$ is independent of the choice of $(x,y) \in \alpha_k$.

We also say that a partition $\mathcal{O}$ of X^2 is coherent if conditions (II), (III) and (IV) are satisfied.

If $(X,\mathcal{O})$ is a coherent configuration (abbreviated CC), then $r := |\mathcal{O}|$ is the *rank*, the map $\mathcal{O} \to \mathcal{O}$ $\alpha \to \alpha^*$ is the *pairing*, and the nonnegative integers a_{ijk} are the *intersection numbers*. $(X,\mathcal{O})$ is *symmetric* if the pairing is trivial, i.e., if $\alpha = \alpha^*$ for all $\alpha \in \mathcal{O}$, and *homogeneous* if $1 \in \mathcal{O}$. Symmetric CC's are necessarily homogeneous and are equivalent to *association schemes* as introduced by Bose and Shimamoto.

We may think of a CC as a colored graph in which the vertices are colored as well as the (directed) edges. Conditions (I) through (IV) are readily visualized in these terms.

Our interest in CC's stems from there frequent occurrence in combinatorial and group theoretic situations and the availability of methods to study them. That CC's abound is clear from the following remarks.

(a) Consider the lattice of all partitions of X^2, where $\mathcal{O}_1 \le \mathcal{O}_2$ means that $\mathcal{O}_1$ is a fusion of $\mathcal{O}_2$, i.e., that $\mathcal{O}_2$ is a refinement of $\mathcal{O}_1$. If two partitions $\mathcal{O}_1$ and $\mathcal{O}_2$ are coherent, then so is $\mathcal{O}_1 \wedge \mathcal{O}_2$. Thus any partition of X^2 generates a coherent one (cf. [6]. Weisfeiler's *stable graphs* are equivalent to our CC's).

(b) If G is a finite group acting on X, then $(X, X^2/G)$ is coherent. We say that this configuration is *group induced,* and that it is *induced* or *afforded* by the action of G on X. This is the *group case.*

(c) Let $X = A^n$ be the set of all words of length n in a finite alaphabet A. Define relations $\alpha_0, \alpha_1, \dots, \alpha_n$ on X by $(x,y) \in \alpha_i$ if and only if x and y differ in exactly i coordinates. Thus i, the *Hamming distance* between x and y, is the number of errors if x is transmitted and y received. Then $(X,\mathcal{O})$, $\mathcal{O} = \{\alpha_0, \dots, \alpha_n\}$, is a symmetric CC of rank n + 1 (i.e., an association scheme with n treatments) sometimes called the *Hamming scheme.* Of course $(X,\mathcal{O})$ depends only on $|A|$ and n. Fixing n we obtain a *system* of symmetric CC's of rank n + 1 paramatrized by $|A|$, the term *system* being used in the sense of the definition to be given in §4. For the importance of this and other association schemes in coding theory see Delsarte's work, e.g., [2]. $(X,\mathcal{O})$ is induced by the action of $\sum_A \text{ wr } \sum_n$ on $X = A^n$.

A basic method for studying CC's $(X,\mathcal{O})$ is via their adjacency algebras (which are the Bose-Mesner algebras in the case of association schemes). Let $\mathcal{O} = \{\alpha_1, \dots, \alpha_r\}$ and let w_i be the matrix of α_i. Then $w_i w_j = \sum_{k=1}^r a_{ijk} w_k$, so $\mathbb{A} = \langle w_1, \dots, w_r \rangle_F$ is a subalgebra of F_r called the *adjacency algebra of* $(X,\mathcal{O})$ *over* F, and $w_1, \dots, w_r$ is called the *standard basis* of $\underline{A}$. Here we have an actual trace ζ defined by $\zeta(w) = \text{trace } w$ for $w \in A$, and the dual basis is given by

$$(3) \qquad w_i = \frac{1}{|\alpha_i|} w_{i^*} = \frac{1}{|\alpha_i|} w_i^t$$

where $\alpha_{i^*} = (\alpha_i)^*$. Moreover, $\mathbb{A}$ is semisimple so all the notations of §1 can be applied. The number $v_i = \frac{|\alpha_i|}{|X|}$ is the valency of the graph (X,α_i), and is an intersection number (e.g., in the homogeneous case $v_i = a_{ii^*1}$ if $\alpha_1 = 1$). We find that

$$(4) \qquad \zeta = \sum_{s=1}^{m} z_s \zeta_s .$$

In particular, v_i and z_s are positive integers, called the *valencies* and *multiplicities* respectively for $(X,\mathcal{O})$. It is often of interest to compute the character-multiplicity table:

$$(5)\qquad \begin{array}{c|ccc|c} & w_1 \ldots & w_s & \ldots w_r & \\ \hline \zeta_1 & & \vdots & & z_1 \\ \vdots & & & & \vdots \\ \zeta_s & \ldots & \zeta_s(w_i) & \ldots & z_s \\ \vdots & & \vdots & & \vdots \\ \zeta_m & & & & z_m \end{array}$$

Knowledge of this table is equivalent to knowledge of the intersection numbers a_{ijk}. The crunch in a nonexistence proof often comes from the fact that the z_s (calculated, e.g., by the orthogonality relations (2)) must be rational integers, or from an application of the Krein conditions (cf.[3, 4]).

In the group case $\mathbb{A}$ is the centralizer algebra of the permutation representation and the z_s are the degrees of the irreducible constituents of the permutation character.

3. VIRTUAL ADJACENCY ALGEBRAS

We now suppose that F, $\mathbb{A}$, $w_1,\ldots,w_r$ and a_{ijk} have the same significance as in §1, and that there is given a *pairing* $i \to i^*$ of $\{1,2,\ldots,r\}$, i.e., a permutation such that $i^{**} = i$ and $1^* = 1$. We call $\mathbb{A}$ a *virtual adjacency algebra over* F *with standard basis* $w_1,\ldots,w_r$, *structure constants* a_{ijk}, and *pairing* $i \to i^*$, if

(6) $\qquad a_{i1k} = a_{1ik} = \delta_{ik}$, $1 \le i,k \le r$ (i.e. if w_1 is the identity element of $\underline{A}$),

(7) $\qquad a_{ij1} = a_{ji1} = \delta_{ij^*}\, v_i \quad$ with $v_i \ne 0$ and $n := \sum_{i=1}^{r} v_i \ne 0$,

and

(8) $\qquad \mathbb{A}$ is semisimple.

We refer to the v_i as the *virtual valencies* for $\mathbb{A}$. There is a virtual trace ζ on $\mathbb{A}$ defined by

(9) $\zeta(w_i) = \delta_{i1}n,\ 1 \le i \le r.$

The dual basis is given by

(10) $$\hat{w}_i = \frac{1}{nv_i} w_{i^*}, \qquad 1 \le i \le r,$$

and we find that

(11) $$\zeta = \sum_{s=1}^{m} z_s \zeta_s.$$

In particular, ζ is an actual trace if and only if the z_s, called the *virtual multiplicities*, are positive rational integers.

If $\mathbb{A}$ is the adjacency algebra of a homogeneous CC and the $\alpha_i \in \mathcal{O}$ are so numbered that $\alpha_1 = 1$, then $\mathbb{A}$ satisfies the definition of a virtual adjacency algebra, ζ given by (9) is the actual trace, and the virtual character-multiplicity table is the actual one. A formulation accomodating arbitrary CC's requires an elaboration of the above notation which we will not go into here.

4. SYSTEMS AND THEIR GENERIC ALGEBRAS

Let F be a field of characteristic 0, $i \to i^*$ a pairing of $\{1,2,\ldots,r\}$, and Δ an infinite subset of F. Suppose given for each $q \in \Delta$ a virtual adjacency algebra $\mathbb{A}(q)$ over F with standard basis $w_{q1},\ldots,w_{qr}$ and pairing $i \to i^*$. For $\mathbb{A}(q)$, $q \in \Delta$, we modify the notations of §§1 and 3 as follows:

v_{qi}, $1 \le i \le r$, are the virtual valencies.

$n_q = \sum_{i=1}^{r} v_{qi}$,

ζ_{qs}, e_{qs}, z_{qs}, $1 \le s \le m_q$, are the absolutely irreducible characters, degrees, and the corresponding virtual multiplicites, respectively, and

$\zeta^{(q)} = \sum_{s=1}^{m_q} z_{qs}\, \zeta_{qs}$ is the virtual trace.

The family $\{\mathbb{A}(q)\}_{q\in\Delta}$ will be called a *system of virtual adjacency algebras over* F if

(12) v_{qi} and z_{qs} are positive rational integers for all i and s,

and

(13) the structure constants for the $\underline{\mathbb{A}}(q)$, $q \in \Delta$, are given *generically* in the sense that there exist polynomials $a_{ijk} \in F[u]$, u an indeterminate, such that for all $q \in \Delta$, $a_{ijk}(q)$ are the structure constants for $\mathbb{A}(q)$.

A family $\{(X(q),\mathcal{O}(q))\}_{q\in\Delta}$ of homogeneous CC's of rank r will be called a *system of homogeneous* CC's if the intersection numbers are given generically. This means precisely that the corresponding adjacency algebras form a system as defined above since condition (12) is automatic for actual adjacency algebras.

The *generic adjacency algebra* for a system $\{\mathbb{A}(q)\}_{q\in\Delta}$ is the $F(u)$-algebra $\mathbb{A}$ with basis $w_1,\ldots,w_r$ and the structure constants a_{ijk} of (13). By specialization it follows that

(14) $\mathbb{A}$ is a virtual adjacency algebra over $F(u)$ with standard basis $w_1,\ldots,w_r$ and the given pairing.

<u>NOTE</u>. If we specialize $\mathbb{A}$ at $q \in F$ (or some extension of F), $q \notin \Delta$, the result will be a virtual adjacency algebra, but condition (12) need not be satisfied.

For the generic algebra $\mathbb{A}$ we use the notations of §§1 and 3, and prove

(15) For all $q \in \Delta$, $m_q = m$ and the ζ_{qs} can be so numbered that $\zeta_{qs} = \zeta_{sf^*}$, $e_{qs} = e_s$ and $z_{qs} = f^*(z_s)$, $1 \le s \le m$. Here f^* is a suitable extension of the specialization $F[u] \to F$, $u \to q$, and $\zeta_{sf^*} = f^*(\zeta_s(w_i))$.

For our final results we add the assumption

(16) At most finitely many rational primes are not in Δ.

There are interesting systems which do not satisfy this. Assuming (16), as we do from no on, we have

(17) v_i and z_s are in $\mathbb{Q}[u]$ for all i and s.

The proof that $z_s \in \mathbb{Q}[u]$ follows that in [1] of the corresponding fact for generic Hecke algebras, based on an application of Hilbert's irreducibility criterion.

Now we can state the generic versions of the Theorems of Frame and Wielandt.

(18) Let $Q = n^r \prod_{i=1}^{r} v_i / \prod_{s=1}^{m} (z_s)^{e_s^2}$ (the *generic Frame quotient*). Then $Q \in \mathbb{Q}[u]$, and, if $\zeta_s(w_i) \in \mathbb{Q}[u]$ for all s, i, then $Q = d^2$, $d \in \mathbb{Q}[u]$.

(19) If a primary polynomial q(u) in Q[u] divides z_s for ℓ distinct values $s_1,\ldots,s_\ell$ of s, then q(u) divides nv_i for $e_{s_1}^2 +\ldots+ e_{s_\ell}^2$ distinct values of i.

REMARK. Under suitable conditions, which hold, for example for systems of homogeneous CC's, Q in (18) can be replaced by $Q_0 = n^{-2}Q$.

We prove (18) and (19) by observing that the Schur relations (1) and Wielandt's arguments (cf. [4]) give versions of these results for virtual adjacency algebras. From these we obtain on the one hand the versions for (homogeneous) CC's as in [4], and on the other, the generic versions (18) and (19). The restriction to the homgeneous case is inessential and has been made here only to simplify the exposition.

REFERENCES

[1] CURTIS, C.W., N. IWAHORI and R. KILMOYER, *Hecke algebras and characters of parabolic type of finite groups with* (B,N)*-pairs,* Math. Pub. I.H.E.S. No. 40 (1972), 81-116.

[2] DELAARTE, P., *An algebraic approach to the association schemes of coding theory,* Philips Res. Repts. Suppl. 10 (1973).

[3] HIGMAN, D.G., *Coherent configurations* I, II, Geometiae Dedicata 4 (1975), 1-32 5 (1976), 413-424.

[4] HIGMAN, D.G., *Lectures on permutation groups,* (notes by Wolfgang Hauptmann), Vorlesungen aus dem Mathematischen Institut Giessen. Heft 4 (1977).

[5] HIGMAN, D.G., *Virtual and generic adjacency algebras,* (in preparation).

[6] WEISFEILER, B., *On construction and identification of graphs,* Springer Lecture Notes in Mathematics. Vol. XIX (1976), 187-203.

OTHER TITLES IN THE SERIES MATHEMATICAL CENTRE TRACTS

A leaflet containing an order-form and abstracts of all publications mentioned below is available at the Mathematisch Centrum, Tweede Boerhaavestraat 49, Amsterdam-1005, The Netherlands. Orders should be sent to the same address.

MCT 1 T. VAN DER WALT, *Fixed and almost fixed points*, 1963. ISBN 90 6196 002 9.

MCT 2 A.R. BLOEMENA, *Sampling from a graph*, 1964. ISBN 90 6196 003 7.

MCT 3 G. DE LEVE, *Generalized Markovian decision processes, part I: Model and method*, 1964. ISBN 90 6196 004 5.

MCT 4 G. DE LEVE, *Generalized Markovian decision processes, part II: Probabilistic background*, 1964. ISBN 90 6196 005 3.

MCT 5 G. DE LEVE, H.C. TIJMS & P.J. WEEDA, *Generalized Markovian decision processes, Applications*, 1970. ISBN 90 6196 051 7.

MCT 6 M.A. MAURICE, *Compact ordered spaces*, 1964. ISBN 90 6196 006 1.

MCT 7 W.R. VAN ZWET, *Convex transformations of random variables*, 1964. ISBN 90 6196 007 X.

MCT 8 J.A. ZONNEVELD, *Automatic numerical integration*, 1964. ISBN 90 6196 008 8.

MCT 9 P.C. BAAYEN, *Universal morphisms*, 1964. ISBN 90 6196 009 6.

MCT 10 E.M. DE JAGER, *Applications of distributions in mathematical physics*, 1964. ISBN 90 6196 010 X.

MCT 11 A.B. PAALMAN-DE MIRANDA, *Topological semigroups*, 1964. ISBN 90 6196 011 8.

MCT 12 J.A.TH.M. VAN BERCKEL, H. BRANDT CORSTIUS, R.J. MOKKEN & A. VAN WIJNGAARDEN, *Formal properties of newspaper Dutch*, 1965. ISBN 90 6196 013 4.

MCT 13 H.A. LAUWERIER, *Asymptotic expansions*, 1966, out of print; replaced by MCT 54 and 67.

MCT 14 H.A. LAUWERIER, *Calculus of variations in mathematical physics*, 1966. ISBN 90 6196 020 7.

MCT 15 R. DOORNBOS, *Slippage tests*, 1966. ISBN 90 6196 021 5.

MCT 16 J.W. DE BAKKER, *Formal definition of programming languages with an application to the definition of ALGOL 60*, 1967. ISBN 90 6196 022 3.

MCT 17 R.P. VAN DE RIET, *Formula manipulation in ALGOL 60, part 1*, 1968. ISBN 90 6196 025 8.

MCT 18 R.P. VAN DE RIET, *Formula manipulation in ALGOL 60, part 2*, 1968. ISBN 90 6196 038 X.

MCT 19 J. VAN DER SLOT, *Some properties related to compactness*, 1968. ISBN 90 6196 026 6.

MCT 20 P.J. VAN DER HOUWEN, *Finite difference methods for solving partial differential equations*, 1968. ISBN 90 6196 027 4.

MCT 21 E. WATTEL, *The compactness operator in set theory and topology*, 1968. ISBN 90 6196 028 2.

MCT 22 T.J. DEKKER, *ALGOL 60 procedures in numerical algebra, part 1*, 1968. ISBN 90 6196 029 0.

MCT 23 T.J. DEKKER & W. HOFFMANN, *ALGOL 60 procedures in numerical algebra, part 2*, 1968. ISBN 90 6196 030 4.

MCT 24 J.W. DE BAKKER, *Recursive procedures*, 1971. ISBN 90 6196 060 6.

MCT 25 E.R. PAERL, *Representations of the Lorentz group and projective geometry*, 1969. ISBN 90 6196 039 8.

MCT 26 EUROPEAN MEETING 1968, *Selected statistical papers, part I*, 1968. ISBN 90 6196 031 2.

MCT 27 EUROPEAN MEETING 1968, *Selected statistical papers, part II*, 1969. ISBN 90 6196 040 1.

MCT 28 J. OOSTERHOFF, *Combination of one-sided statistical tests*, 1969. ISBN 90 6196 041 X.

MCT 29 J. VERHOEFF, *Error detecting decimal codes*, 1969. ISBN 90 6196 042 8.

MCT 30 H. BRANDT CORSTIUS, *Excercises in computational linguistics*, 1970. ISBN 90 6196 052 5.

MCT 31 W. MOLENAAR, *Approximations to the Poisson, binomial and hypergeometric distribution functions*, 1970. ISBN 90 6196 053 3.

MCT 32 L. DE HAAN, *On regular variation and its application to the weak convergence of sample extremes*, 1970. ISBN 90 6196 054 1.

MCT 33 F.W. STEUTEL, *Preservation of infinite divisibility under mixing and related topics*, 1970. ISBN 90 6196 061 4.

MCT 34 I. JUHÁSZ, A. VERBEEK & N.S. KROONENBERG, *Cardinal functions in topology*, 1971. ISBN 90 6196 062 2.

MCT 35 M.H. VAN EMDEN, *An analysis of complexity*, 1971. ISBN 90 6196 063 0.

MCT 36 J. GRASMAN, *On the birth of boundary layers*, 1971. ISBN 90 6196 064 9.

MCT 37 J.W. DE BAKKER, G.A. BLAAUW, A.J.W. DUIJVESTIJN, E.W. DIJKSTRA, P.J. VAN DER HOUWEN, G.A.M. KAMSTEEG-KEMPER, F.E.J. KRUSEMAN ARETZ, W.L. VAN DER POEL, J.P. SCHAAP-KRUSEMAN, M.V. WILKES & G. ZOUTENDIJK, *MC-25 Informatica Symposium*, 1971. ISBN 90 6196 065 7.

MCT 38 W.A. VERLOREN VAN THEMAAT, *Automatic analysis of Dutch compound words*, 1971. ISBN 90 6196 073 8.

MCT 39 H. BAVINCK, *Jacobi series and approximation*, 1972. ISBN 90 6196 074 6.

MCT 40 H.C. TIJMS, *Analysis of (s,S) inventory models*, 1972. ISBN 90 6196 075 4.

MCT 41 A. VERBEEK, *Superextensions of topological spaces*, 1972. ISBN 90 6196 076 2.

MCT 42 W. VERVAAT, *Success epochs in Bernoulli trials (with applications in number theory)*, 1972. ISBN 90 6196 077 0.

MCT 43 F.H. RUYMGAART, *Asymptotic theory of rank tests for independence*, 1973. ISBN 90 6196 081 9.

MCT 44 H. BART, *Meromorphic operator valued functions*, 1973. ISBN 90 6196 082 7.

MCT 45 A.A. BALKEMA, *Monotone transformations and limit laws*, 1973. ISBN 90 6196 083 5.

MCT 46 R.P. VAN DE RIET, *ABC ALGOL, A portable language for formula manipulation systems, part 1: The language*, 1973. ISBN 90 6196 084 3.

MCT 47 R.P. VAN DE RIET, *ABC ALGOL, A portable language for formula manipulation systems, part 2: The compiler*, 1973. ISBN 90 6196 085 1.

MCT 48 F.E.J. KRUSEMAN ARETZ, P.J.W. TEN HAGEN & H.L. OUDSHOORN, *An ALGOL 60 compiler in ALGOL 60, Text of the MC-compiler for the EL-X8*, 1973. ISBN 90 6196 086 X.

MCT 49 H. KOK, *Connected orderable spaces*, 1974. ISBN 90 6196 088 6.

MCT 50 A. VAN WIJNGAARDEN, B.J. MAILLOUX, J.E.L. PECK, C.H.A. KOSTER, M. SINTZOFF, C.H. LINDSEY, L.G.L.T. MEERTENS & R.G. FISKER (Eds), *Revised report on the algorithmic language ALGOL 68*, 1976. ISBN 90 6196 089 4.

MCT 51 A. HORDIJK, *Dynamic programming and Markov potential theory*, 1974. ISBN 90 6196 095 9.

MCT 52 P.C. BAAYEN (ed.), *Topological structures*, 1974. ISBN 90 6196 096 7.

MCT 53 M.J. FABER, *Metrizability in generalized ordered spaces*, 1974. ISBN 90 6196 097 5.

MCT 54 H.A. LAUWERIER, *Asymptotic analysis, part 1*, 1974. ISBN 90 6196 098 3.

MCT 55 M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 1: Theory of designs, finite geometry and coding theory*, 1974. ISBN 90 6196 099 1.

MCT 56 M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 2: graph theory, foundations, partitions and combinatorial geometry*, 1974. ISBN 90 6196 100 9.

MCT 57 M. HALL JR. & J.H. VAN LINT (Eds), *Combinatorics, part 3: Combinatorial group theory*, 1974. ISBN 90 6196 101 7.

MCT 58 W. ALBERS, *Asymptotic expansions and the deficiency concept in statistics*, 1975. ISBN 90 6196 102 5.

MCT 59 J.L. MIJNHEER, *Sample path properties of stable processes*, 1975. ISBN 90 6196 107 6.

MCT 60 F. GÖBEL, *Queueing models involving buffers*, 1975. ISBN 90 6196 108 4.

* MCT 61 P. VAN EMDE BOAS, *Abstract resource-bound classes, part 1*. ISBN 90 6196 109 2.

* MCT 62 P. VAN EMDE BOAS, *Abstract resource-bound classes, part 2*. ISBN 90 6196 110 6.

MCT 63 J.W. DE BAKKER (ed.), *Foundations of computer science*, 1975. ISBN 90 6196 111 4.

MCT 64 W.J. DE SCHIPPER, *Symmetric closed categories*, 1975. ISBN 90 6196 112 2.

MCT 65 J. DE VRIES, *Topological transformation groups 1 A categorical approach*, 1975. ISBN 90 6196 113 0.

MCT 66 H.G.J. PIJLS, *Locally convex algebras in spectral theory and eigenfunction expansions*, 1976. ISBN 90 6196 114 9.

* MCT 67 H.A. LAUWERIER, *Asymptotic analysis, part 2.* ISBN 90 6196 119 X.

MCT 68 P.P.N. DE GROEN, *Singularly perturbed differential operators of second order*, 1976. ISBN 90 6196 120 3.

MCT 69 J.K. LENSTRA, *Sequencing by enumerative methods*, 1977. ISBN 90 6196 125 4.

MCT 70 W.P. DE ROEVER JR., *Recursive program schemes: semantics and proof theory*, 1976. ISBN 90 6196 127 0.

MCT 71 J.A.E.E. VAN NUNEN, *Contracting Markov decision processes*, 1976. ISBN 90 6196 129 7.

MCT 72 J.K.M. JANSEN, *Simple periodic and nonperiodic Lamé functions and their applications in the theory of conical waveguides*, 1977. ISBN 90 6196 130 0.

MCT 73 D.M.R. LEIVANT, *Absoluteness of intuitionistic logic*, 1979. ISBN 90 6196 122 x.

MCT 74 H.J.J. TE RIELE, *A theoretical and computational study of generalized aliquot sequences*, 1976. ISBN 90 6196 131 9.

MCT 75 A.E. BROUWER, *Treelike spaces and related connected topological spaces*, 1977. ISBN 90 6196 132 7.

MCT 76 M. REM, *Associons and the closure statement*, 1976. ISBN 90 6196 135 1.

MCT 77 W.C.M. KALLENBERG, *Asymptotic optimality of likelihood ratio tests in exponential families*, 1977 ISBN 90 6196 134 3.

MCT 78 E. DE JONGE, A.C.M. VAN ROOIJ, *Introduction to Riesz spaces*, 1977. ISBN 90 6196 133 5.

MCT 79 M.C.A. VAN ZUIJLEN, *Empirical distributions and rankstatistics*, 1977. ISBN 90 6196 145 9.

MCT 80 P.W. HEMKER, *A numerical study of stiff two-point boundary problems*, 1977. ISBN 90 6196 146 7.

MCT 81 K.R. APT & J.W. DE BAKKER (eds), *Foundations of computer science II*, part I, 1976. ISBN 90 6196 140 8.

MCT 82 K.R. APT & J.W. DE BAKKER (eds), *Foundations of computer science II*, part II, 1976. ISBN 90 6196 141 6.

* MCT 83 L.S. VAN BENTEM JUTTING, *Checking Landau's "Grundlagen" in the AUTOMATH system*, ISBN 90 6196 147 5.

MCT 84 H.L.L. BUSARD, *The translation of the elements of Euclid from the Arabic into Latin by Hermann of Carinthia (?) books vii-xii*, 1977. ISBN 90 6196 148 3.

MCT 85 J. VAN MILL, *Supercompactness and Wallman spaces*, 1977. ISBN 90 6196 151 3.

MCT 86 S.G. VAN DER MEULEN & M. VELDHORST, *Torrix I*, 1978. ISBN 90 6196 152 1.

* MCT 87 S.G. VAN DER MEULEN & M. VELDHORST, *Torrix II*, ISBN 90 6196 153 x.

MCT 88 A. SCHRIJVER, *Matroids and linking systems*, 1977. ISBN 90 6196 154 8.

MCT 89 J.W. DE ROEVER, *Complex Fourier transformation and analytic functionals with unbounded carriers*, 1978.
ISBN 90 6196 155 6.

* MCT 90 L.P.J. GROENEWEGEN, *Characterization of optimal strategies in dynamic games*, . ISBN 90 6196 156 4.

* MCT 91 J.M. GEYSEL, *Transcendence in fields of positive characteristic*, . ISBN 90 6196 157 2.

* MCT 92 P.J. WEEDA, *Finite generalized Markov programming*, .
ISBN 90 6196 158 0.

MCT 93 H.C. TIJMS (ed.) & J. WESSELS (ed.), *Markov decision theory*, 1977.
ISBN 90 6196 160 2.

MCT 94 A. BIJLSMA, *Simultaneous approximations in transcendental number theory*, 1978. ISBN 90 6196 162 9.

MCT 95 K.M. VAN HEE, *Bayesian control of Markov chains*, 1978.
ISBN 90 6196 163 7.

* MCT 96 P.M.B. VITÁNYI, *Lindenmayer systems: structure, languages, and growth functions*, 1978. ISBN 90 6196 164 5.

* MCT 97 A. FEDERGRUEN, *Markovian control problems; functional equations and algorithms*, 1978. ISBN 90 6196 165 3.

MCT 98 R. GEEL, *Singular perturbations of hyperbolic type*, 1978.
ISBN 90 6196 166 1

MCT 99 J.K. LENSTRA, A.H.G. RINNOOY KAN & P. VAN EMDE BOAS, *Interfaces between computer science and operations research*, 1978.
ISBN 90 6196 170 X.

MCT 100 P.C. BAAYEN, D. VAN DULST & J. OOSTERHOFF (Eds), *Proceedings bicentennial congress of the Wiskundig Genootschap, part 1*, 1979.
ISBN 90 6196 168 8.

MCT 101 P.C. BAAYEN, D. VAN DULST & J. OOSTERHOFF (Eds), *Proceedings bicentennial congress of the Wiskundig Genootschap, part 2*, 1979.
ISBN 90 9196 169 6.

MCT 102 D. VAN DULST, *Reflexive and superreflexive Banach spaces*, 1978.
ISBN 90 6196 171 8.

MCT 103 K. VAN HARN, *Classifying infinitely divisible distributions by functional equations*, 1978. ISBN 90 6196 172 6.

* MCT 104 J.M. VAN WOUWE, *Go-spaces and generalizations of metrizability*, . ISBN 90 6196 173 4.

* MCT 105 R. HELMERS, *Edgeworth expansions for linear combinations of order statistics*, . ISBN 90 6196 174 2.

AN ASTERISK BEFORE THE NUMBER MEANS "TO APPEAR"