Pacific Journal of Mathematics



IN THIS ISSUE—

David R. Arterburn and Robert James Whitley, *Projections in the space of bounded linear operators* 739
Robert McCallum Blumenthal, Joram Lindenstrauss and Robert Ralph Phelps, *Extreme operators into $C(K)$* 747
L. Carlitz, *A note on multiple exponential sums* 757
Joseph A. Cima, *A nonnormal Blaschke-quotient* 767
Paul Civin and Bertram Yood, *Lie and Jordan structures in Banach algebras* 775
Luther Elic Claborn, *Dedekind domains: Overrings and semi-prime elements* 799
Luther Elic Claborn, *Note generalizing a result of Samuel's* 805
George Bernard Dantzig, E. Eisenberg and Richard Warren Cottle, *Symmetric dual nonlinear programs* 809
Philip J. Davis, *Simple quadratures in the complex plane* 813
Edward Richard Fadell, *On a coincidence theorem of F. B. Fuller* 825
Delbert Ray Fulkerson and Oliver Gross, *Incidence matrices and interval graphs* 835
Larry Charles Grove, *Tensor products over H^*-algebras* 857
Deborah Tepper Haimo, *L^2 expansions in terms of generalized heat polynomials and of their Appell transforms* ... 865
I. Martin (Irving) Isaacs and Donald Steven Passman, *A characterization of groups in terms of the degrees of their characters* 877
Donald Gordon James, *Integral invariants for vectors over local fields* 905
Fred Krakowski, *A remark on the lemma of Gauss* 917
Marvin David Marcus and H. Minc, *A subdeterminant inequality* 921
Kevin Mor McCrimmon, *Norms and noncommutative Jordan algebras* 925
Donald Earl Myers, *Topologies for Laplace transform spaces* 957
Olav Njstad, *On some classes of nearly open sets* 961
Milton Philip Olson, *A characterization of conditional probability* 971
Barbara Osofsky, *A counter-example to a lemma of Skornjakov* 985
Sidney Charles Port, *Ratio limit theorems for Markov chains* 989
George A. Reid, *A generalisation of W^*-algebras* 1019
Robert Wells Ritchie, *Classes of recursive functions based on Ackermann's function* 1027
Thomas Lawrence Sherman, *Properties of solutions of* nth *order linear differential equations* 1045
Ernst Snapper, *Inflation and deflation for all dimensions* 1061
Kondagunta Sundaresan, *On the strict and uniform convexity of certain Banach spaces* 1083
Frank J. Wagner, *Maximal convex filters in a locally convex space* 1087
Joseph Albert Wolf, *Translation-invariant function algebras on compact groups* 1093

Vol. 15, No. 3 **November, 1965**

PACIFIC JOURNAL OF MATHEMATICS

EDITORS

H. SAMELSON
Stanford University
Stanford, California

J. DUGUNDJI
University of Southern California
Los Angeles, California 90007

R. M. BLUMENTHAL
University of Washington
Seattle, Washington 98105

*RICHARD ARENS
University of California
Los Angeles, California 90024

ASSOCIATE EDITORS

E. F. BECKENBACH B. H. NEUMANN F. WOLF K. YOSIDA

SUPPORTING INSTITUTIONS

UNIVERSITY OF BRITISH COLUMBIA
CALIFORNIA INSTITUTE OF TECHNOLOGY
UNIVERSITY OF CALIFORNIA
MONTANA STATE UNIVERSITY
UNIVERSITY OF NEVADA
NEW MEXICO STATE UNIVERSITY
OREGON STATE UNIVERSITY
UNIVERSITY OF OREGON
OSAKA UNIVERSITY
UNIVERSITY OF SOUTHERN CALIFORNIA
STANFORD UNIVERSITY
UNIVERSITY OF TOKYO
UNIVERSITY OF UTAH
WASHINGTON STATE UNIVERSITY
UNIVERSITY OF WASHINGTON
* * *
AMERICAN MATHEMATICAL SOCIETY
CALIFORNIA RESEARCH CORPORATION
SPACE TECHNOLOGY LABORATORIES
NAVAL ORDNANCE TEST STATION

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should by typewritten (double spaced). The first paragraph or two must be capable of being used separately as a synopsis of the entire paper. It should not contain references to the bibliography. No separate author's resumé is required. Manuscripts may be sent to any one of the four editors. All other communications to the editors should be addressed to the managing editor, Richard Arens, at the University of California, Los Angeles, California 90024.

50 reprints per author of each article are furnished free of charge; additional copies may be obtained at cost in multiples of 50.

The *Pacific Journal of Mathematics* is published quarterly, in March, June, September, and December. Effective with Volume 13 the price per volume (4 numbers) is $18.00; single issues, $5.00. Special price for current issues to individual faculty members of supporting institutions and to individual members of the American Mathematical Society: $8.00 per volume; single issues $2.50. Back numbers are available.

Subscriptions, orders for back numbers, and changes of address should be sent to Pacific Journal of Mathematics, 103 Highland Boulevard, Berkeley 8, California.

Printed at Kokusai Bunken Insatsusha (International Academic Printing Co., Ltd.), No. 6, 2-chome, Fujimi-cho, Chiyoda-ku, Tokyo, Japan.

PUBLISHED BY PACIFIC JOURNAL OF MATHEMATICS, A NON-PROFIT CORPORATION

The Supporting Institutions listed above contribute to the cost of publication of this Journal, but they are not owners or publishers and have no responsibility for its content or policies.

* Basil Gordon, Acting Managing Editor until February 1, 1966.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

PROJECTIONS IN THE SPACE OF BOUNDED LINEAR OPERATORS

DAVID ARTERBURN AND ROBERT WHITLEY

Thorp has shown that for X and Y certain Banach spaces of sequences there is no continuous linear projection of the bounded linear operators from X to Y onto the compact linear operators from X to Y. In this paper, this result, as well as related results for the weakly compact linear operators, is demonstrated for cases including $(a)X$ an infinite dimensional abstract L-space and Y an infinite dimensional space whose conjugate contains a countable total set and $(b)X$ a separable B-space and $Y = C(S)$ with S either a metric space containing an infinite number of points or S a compact space which contains a one-to-one convergent sequence.

We recall that a subspace of a Banach space X is said to be complemented (in X) if there is a continuous linear projection of X onto that subspace. In [14] it is shown that for X and Y certain Banach spaces of sequences the subspace $K(X, Y)$ of compact linear operators from X to Y is not complemented in $B(X, Y)$, the space of bounded linear operators from X to Y. Here, we will prove similar results for either X an abstract L-space or Y a space of type $C(S)$ and will also consider projections on the subspace $W(X, Y)$ of weakly compact linear operators mapping X to Y.

All maps will be linear and X and Y will be Banach spaces. Abstract L-spaces are defined in [7, page 394]; $C(S)$ shall be the space of bounded continuous functions on a topological space S and we use the *sup* norm. We recall that a set in X', the conjugate of the Banach space X, is total if the only vector mapped into zero by that set is the zero vector.

Our main results are Theorems 1 and 2 below.

1. THEOREM. *Let $\mathscr{L}$ be an infinite dimensional abstract L-space and let X have a complemented subspace Y. Suppose that Y' contains a countable total set. Then*

(a) *If Y is infinite dimensional, then $K(\mathscr{L}, X)$ is not complemented in $B(\mathscr{L}, X)$. In fact, $K(\mathscr{L}, Y)$ is complemented in $B(\mathscr{L}, Y)$ if and only if these spaces are equal and this happens if and only if Y is finite dimensional.*

Received August 11, 1964. This work was supported by Air Force grant AF-AFOSR 457-63.

(b) *If weak and norm sequential convergence are not equivalent in Y, then $K(\mathscr{L}, X)$ is not complemented in $W(\mathscr{L}, X)$. In fact, $K(\mathscr{L}, Y)$ is complemented in $W(\mathscr{L}, Y)$ if and only if these spaces are equal and this happens if and only if norm and weak sequential convergence are equivalent in Y.*

(c) *If Y is not reflexive, then $W(\mathscr{L}, X)$ is not complemented in $B(\mathscr{L}, X)$. In fact, $W(\mathscr{L}, Y)$ is complemented in $B(\mathscr{L}, Y)$ if and only if these spaces are equal and this happens if and only if Y is reflexive.*

2. THEOREM. *Suppose that S is either a (not necessarily compact) metric space which contains an infinite number of points, or that S is a compact space which contains a one-to-one convergent sequence. Let X have a complemented subspace Y and suppose that Y is separable. Then*

(a) *If Y is infinite dimensional, then $K(X, C(S))$ is not complemented in $B(X, C(S))$. In fact, $K(Y, C(S))$ is complemented in $B(Y, C(S))$ if and only if these spaces are equal and this happens if and only if Y is finite dimensional.*

(b) *If weak and norm sequential convergence are not the same in Y', then $K(X, C(S))$ is not complemented in $W(X, C(S))$. In fact, $K(Y, C(S))$ is complemented in $W(Y, C(S))$ if and only if these spaces are equal and this happens if and only if norm and weak sequential convergence are the same in Y'.*

(c) *If Y is not reflexive, then $W(X, C(S))$ is not complemented in $B(X, C(S))$. In fact, $W(Y, C(S))$ is complemented in $B(Y, C(S))$ if and only if these spaces are equal and this happens if and only if Y is reflexive.*

We remark, in connection with part (b) of both theorems, that weak and norm sequential convergence are the same in l [2, page 137]. In Theorem 2, part (b), the separability of Y is essential, for if $C(S)$ is separable, then $W(m, C(S)) = B(m, C(S))$, since weak* and weak sequential convergence are equivalent in m' [9, Theorem 9, page 168], yet m is not reflexive. It follows from either theorem that $K(l, m)$ is not complemented in $B(l, m)$, a result incorrectly proved in [14].

The above theorems can be extended by use of the following lemma.

3, LEMMA. *Suppose that X_1 and Y_1 are complemented subspaces*

of, respectively, X and Y. Then

(a) *If $K(X, Y)$ is complemented in $B(X, Y)$, then $K(X_1, Y_1)$ is complemented in $B(X_1, Y_1)$.*

(b) *If $K(X, Y)$ is complemented in $W(X, Y)$, then $K(X_1, Y_1)$ is complemented in $W(X_1, Y_1)$.*

(c) *If $W(X, Y)$ is complemented in $B(X, Y)$, then $W(X_1, Y_1)$ is complemented in $B(X_1, Y_1)$.*

Proof. Let P_1 be a projection of X onto X_1 and let P_2 project Y onto Y_1. For case (a), suppose that P is a projection of $B(X, Y)$ onto $K(X, Y)$. Define a map Q on $B(X_1, Y_1)$ by $Q(T) = [P_2 \circ P(T \circ P_1)]|_{X_1}$, where $F|_{X_1}$ is the restriction of a map F to X_1. Then Q is a projection of $B(X_1, Y_1)$ onto $K(X_1, Y_1)$. The other cases are similar.

So, for example, Theorem 1 gives useful information about maps with range in a space which contains a complemented subspace isomorphic to an abstract L-space.

Note that by Lemma 3 it suffices to prove Theorems 1 and 2 under the assumption $X = Y$. We now find canonical subspaces of $\mathscr{L}$ and $C(S)$ and reduce the problem still further.

From Corollary 4, page 221 of [12] we see that any infinite dimensional complemented subspace of an abstract L-space contains a complemented copy of l. So we may assume in the proof of Theorem 1 that $\mathscr{L} = l$.

In [1] Arens has shown that if S_0 is a metrizable compact subspace of a paracompact space S, then there is a projection of $C(S)$ onto a subspace isomorphic to $C(S_0)$; this is a generalization of Borsuk's theorem [4], in which S_0 is a separable closed subspace of a metric space S. From Arens' result we see that if S is a compact space which contains a one to one convergent sequence, then $C(S)$ contains a complemented copy of c. From Borsuk's result, if S is a metric space containing an infinite number of points, then $C(S)$ contains a complemented copy of either m or c. (We remark that a particularly nice proof of Borsuk's theorem is given in [10]). Thus it suffices to prove Theorem 2 for $C(S) = m$ and $C(S) = c$.

We have now reduced the problem to its essentials. We will need the following representation theorems [15]:

Let $T: X \to C(S)$ be a linear operator. Then T defines a function $p: S \to X'$ by $p(s)(x) = Tx(s)$ and p is continuous as a map into (X', X), i.e. into X' with the weak* topology. Then T is continuous if and only if $p(S)$ is bounded and in this case, $\| T \| = \sup \{\| p(s) \|: s \text{ in } S\}$. Also, T is compact if and only if, in addition, $p(S)$ is conditionally

compact (i.e., if and only if the (norm) closure of $p(S)$ is compact) and T is weakly compact if and only if, in addition, $p(S)$ is conditionally compact in the weak topology of X'.

Let $T: l \to X$ be a linear operator. Then, letting e_i denote the characteristic function of the set $\{i\}$, T is continuous if and only if $\{Te_i : i = 1, 2, \cdots\}$ is bounded and in this case $\| T \| = \sup \| Te_i \|$. The map T is compact if and only if the set $\{Te_i : i = 1, 2, \cdots\}$ is conditionally compact and is weakly compact if and only if that set is conditionally compact in the weak topology of X.

The first representation theorem is due to Bartle [3], for compact S, and the second is due to Dunford and Pettis [6].

The following lemma is the backbone of all our proofs. We denote the space of all bounded functions from a set S to a Banach space X by $m(S : X)$ with $\| f \| = \sup \{\| f(s) \| : s \text{ in } S\}$. If the space X is the scalar field we write $m(S)$, which is also called $B(S)$, and if S is a countably infinite set we have the space m of bounded sequences. For any f in $m(S : X)$, the support of f, supt (f), is given by $\{s : f(s) \neq 0\}$.

4. LEMMA. *Let M and N, $N \subseteq M$, be closed subspaces of $m(S : X)$ and let N contain all the functions which have finite support. Suppose that there is a function f in $m(S : X)$ and an uncountable family of functions in $m(S)$, $\{g_a : a \text{ in } A\}$, with the properties:*

(1) $\| g_a \| \leq 1$ *for all a in A,*

(2) fg_a, *the function whose value at s is $f(s)g_a(s)$, is in $M - N$, and*

(3) $\text{supt}\,(g_a) \cap \text{supt}\,(g_b)$ *is finite for $a \neq b$.*

Then $(M/N)'$ does not contain a countable total subset. Hence, if M' contains a countable total subset, then N is not complemented in M.

Proof. Let f_a be the coset in M/N which contains fg_a and note that $f_a \neq 0$. To show that $(M/N)'$ does not contain a countable total subset it will suffice to show that a functional x' in $(M/N)'$ can fail to annihilate only countably many elements in the set $\{f_a : a \text{ in } A\}$, so it will suffice to show that the set $C(n) = \{f_a : | x'(f_a) | \geq 1/n\}$ is finite for each natural number n. To see this let $h_1, h_2, \cdots, h_m$ be in $C(n)$, set $b_i = \overline{x'(h_1)}/| x'(h_i) |$ and let $x = \Sigma b_i h_i$. The critical point is to note that $\| x \| \leq \| f \|$. Then, since $\| x' \| \| f \| \geq | x'(x) | \geq m/n$, we see that $C(n)$ is finite.

If the subspace N is complemented in M we have $M = N \oplus R$ where R is a closed subspace of M. Then, since R' contains a countable

total subset whenever M' does, and M/N is isomorphic to R, we see that $(M/N)'$ contains a countable total subset if M' does.

We use the next lemma in constructing functions g_a which are as described in Lemma 4.

5. LEMMA. *Let I be a countable set. Then there is an uncountable family $\{U_a : a \text{ in } A\}$ of infinite subsets of I with $U_a \cap U_b$ finite for $a \neq b$.*

Proof. See problem 6Q, page 97 of [8].

The above lemmas are a generalization of the method of [16] and the basic idea can be found in [11] and [13].

As we have noted, Theorem 1 is reduced to the following lemma:

6. LEMMA. *Theorem 1 holds in the special case $X = Y$ and $\mathscr{L} = l$.*

Proof. Let $I = \{1, 2, \cdots\}$. By the representation theorem, $B(l, X)$ corresponds to $m(I : X)$ and $K(l, X)[W(l, X)]$ to the subspace of $m(I : X)$ consisting of those functions whose range is [weakly] conditionally compact.

Let $\{U_a : a \text{ in } A\}$ be a family of subsets of I as in Lemma 5 and let g_a be the characteristic function of U_a.

For case (a), suppose that X is infinite dimensional and select a sequence $\{x_i\}$ of points from the unit sphere of X so that $\{x_i\}$ contains no convergent subsequence. We define f in $m(I : X)$ by $f(i) = x_i$. Now we apply Lemma 4 to see that $B(l, X)/K(l, X) = m(I : X)/K$, where K is the subspace of functions with conditionally compact range, is a space whose conjugate contains no countable total set. But $m(I : X)$ does have a conjugate which contains a countable total set, since we are assuming that this is true of $Y = X$; so by Lemma 4 $K(l, X)$ is not complemented in $B(l, X)$.

For case (b) let $\{x_i\}$ be a weakly convergent sequence of points on the unit sphere of X which contains no norm convergent subsequence, assuming that weak and norm sequential convergence are not the same in X, and proceed as above.

For case (c), assume that X is not reflexive and let $\{x_i\}$ be a bounded sequence which contains no weakly convergent subsequences.

That the spaces are equal under the conditions given follows directly. This completes the proof of Theorem 1.

Now Theorem 2 has been reduced to the case $X = Y$ and either $C(S_1) = m$ or $C(S_2) = c$; since S_1 and S_2 are separable Hausdorff spaces

which contain a countably infinite number of isolated points, the following lemma will suffice:

7. LEMMA. *Let S_0 be a separable Hausdorff space which contains a countably infinite number of isolated points. Then Theorem 2 holds if $X = Y$ and $S = S_0$.*

Proof. The proof is quite a bit like the proof of Lemma 6. Let $I = \{s_1, s_2, \cdots\}$ be the countably infinite set of isolated points of $S = S_0$ and let U_a be a family of subsets of I as in Lemma 5. Let g_a be the characteristic function of the set U_a.

For case (a), assume that X is infinite dimensional and choose a sequence $\{x_i'\}$ of elements in X' which converge to zero in the weak* topology of X' and yet contain no norm convergent subsequence. Define f to be zero on $S - I$ and $f(s_i) = x_i'$. Now, via the representation theorem, $B(X, C(S))$ corresponds to the subspace B of $m(S : X')$ which consists of those functions in $m(S : X')$ which are continuous as maps from S to X' with the weak* topology, and $K(X, C(S))$ corresponds to the subspace of B which consists of those functions which have conditionally compact range. So the proof for case (a) will be completed by Lemma 4 if we can show that the function fg_a is continuous as a map from S to (X', X). To see this, suppose that $\{s(\alpha)\}$ is a net in S which converges to s. If s is isolated the net is eventually s and then $\{f(s(\alpha))g_a(s(\alpha))\}$ is eventually $f(s)g_a(s)$ and so fg_a is continuous at s, so we may suppose that s is not isolated. Since s is not isolated, $f(s)g_a(s)$ is zero and so we must show that $\{f(s(\alpha))g_a(s(\alpha))\}$ converges to zero; this net will converge to zero if for each natural number N there is an α_0 such that $s(\alpha)$ is not in $\{s_1, s_2, \cdots, s_N\}$ for $\alpha \geqq \alpha_0$, if there is no such α_0 for some N we find that s is isolated, a contradiction.

For case (b) we assume that weak and norm sequential convergence are not the same in X' and choose a sequence $\{x_i'\}$ which converges weakly to zero but has no norm convergent subsequence. Since x_i' converges weakly to zero it converges to zero in the weak* topology and so fg_a, as above, is continuous.

For case (c), let $\{x_i'\}$ be a sequence which converges to zero in the weak* topology but contains no weakly convergent subsequence. A bit of caution is necessary here. We are assuming that $X = Y$ is separable and so the weak* topology on the unit sphere in X' is metrizable [7, Theorem 1, page 426] and so if weak* and weak sequential convergence are the same, then the sphere is weak sequentially compact and hence weakly sequentially compact, hence weakly compact and so X is reflexive. However, if X is not separable, we may have weak* and weak sequential convergence the same in X' without X being

reflexive; for example, $X = m$ [9, Theorem 9, page 168].

It follows from the representation theorem that the spaces are equal under the given conditions.

This completes the proof of Theorem 2.

There is no known example where $K(X, Y)$ is complemented in $B(X, Y)$ and yet is not equal to $B(X, Y)$, ditto for the subspace $W(X, Y)$ and for $K(X, Y)$ as a subspace of $W(X, Y)$.

A simple case which remains open is whether $K(m, c)$ is complemented in $B(m, c)$. If m had a separable complemented subspace which was infinite dimensional, then Theorem 2 would solve this problem; but m does not [12, Theorem 6, page 221].

Added in proof. The argument following Lemma 3, which relies on references [1] and [4] to show that for certain S the space $C(S)$ contains a complemented subspace isomorphic to either m or c, can be replaced by the elementary Corollary 6 of D. W. Dean's paper Subspaces of $C(H)$ Which Are Direct Factors of $C(H)$ (Proc. Amer. Math. Soc. 16 (1965), 237–242).

We thank E. O. Thorp for his support.

REFERENCES

1. R. Arens, *Extensions of functions on fully normal spaces*, Pacific J. Math. **2** (1952), 11-22.

2. S. Banach, *Opérations Linéaires*, Chelsea, 1955.

3. R. Bartle, *On compactness in fuctional analysis*, Trans. Amer. Math. Soc. **79** (1955), 35-57.

4. K. Borsuk, *Über isomorphie der funktionalräume*, Bull. Int. Acad. Pol. Sci. (1933), 1-10.

5. M. Day, *Normed Linear Spaces*, Springer, 1962.

6. N. Dunford, and B. Pettis, *Linear operations on summable functions*, Trans. Amer. Math. Soc. **47** (1940), 323-392.

7. N. Dunford, and J. Schwartz, *Linear Operators*, Vol. I, Interscience, 1958.

8. L. Gillman, and M. Jerison, *Rings of Continuous Functions*, Van Nostrand, 1960.

9. A. Grothendieck, *Sur les applications linéaires faiblement compactes d'espace du type $C(K)$*, Canadian J. Math. **5** (1953), 129-173.

10. S. Kakutani, *Simultaneous extension of continuous functions considered as a positive linear operation*, Japan J. Math. **17** (1940), 1-4.

11. S. Kakutani, and M. Nakamura, *Banach limits and the Čech compactification of a countable discrete set*, Proc. Imp. Acad. Japan **19** (1943), 224-229.

12. A. Pełczynski, *Projections in certain Banach spaces*, Studia Math. **19** (1960), 209-227.

13. A. Pełczynski, and V. Sudakov, *Remark on non-complemented subspaces of the space $m(S)$*, Colloquim Math. **9** (1962), 85-88.

14. E. Thorp, *Projections onto the subspace of compact operators*, Pacific J. Math. **10** (1960), 693-696.

15. E. Thorp, and R. Whitley, *Operator representation theorems*, Illinois J. Math. (to appear).
16. R. Whitley, *Projecting m onto* c_0, Amer. Math. Monthly (to appear).

MONTANA STATE UNIVERSITY
AND
UNIVERSITY OF MARYLAND

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

EXTREME OPERATORS INTO $C(K)$

R. M. BLUMENTHAL, JORAM LINDENSTRAUSS[1]
AND R. R. PHELPS

If X and Y are real Banach spaces let $S(X, Y)$ denote the convex set of all linear operators from X into Y having norm less than or equal to 1. The main theorem is this: If K_1 and K_2 are compact Hausdorff spaces with K_1 metrizable and if T is an extreme point of $S(C(K_1), C(K_2))$, then there are continuous functions $\phi: K_2 \to K_1$ and λ in $C(K_2)$ with $|\lambda| = 1$ such that $(Tf)(k) = \lambda(k)f(\phi(k))$ for all k in K_2 and f in $C(K_1)$. There are several additional theorems which discuss the possibility of replacing $C(K_1)$ in this theorem by an arbitrary Banach space.

Suppose that K_1 is a compact Hausdorff space and that $C(K_1)$ is the Banach space of real-valued continuous functions on K_1, with supremum norm. Denote by S^* the unit ball of $C(K_1)^*$; then S^* is a weak* compact convex set and therefore the set $\operatorname{ext} S^*$ of its extreme points is nonempty, by the Krein-Milman theorem. Arens and Kelley [1] (cf. [4, p. 441]) showed that these extreme points are precisely those functionals of the form $f \to \lambda f(k) (f \in C(K_1))$, where $k \in K_1$ and $\lambda = 1$ or $\lambda = -1$. [We denote the functional "evaluation at k" by φ_k, so the extreme points of S^* are the functionals $\lambda\varphi_k$, $|\lambda| = 1$.] If we restrict our attention to the "positive face" of S^* (those functionals φ such that $\varphi(1) = 1$), then this is a weak* compact convex subset of S^* and its extreme points are those obtained from $\operatorname{ext} S^*$ by taking $\lambda = 1$. Now, the members of $C(K_1)^*$ can be regarded as continuous linear operators from $C(K_1)$ into $C(K_2)$, where K_2 consists of a single point. Thus, it is natural to consider the possible extension of the above results to the more general situation when K_2 is an arbitrary compact Hausdorff space, and S^* is replaced by the unit ball $S = S(C(K_1), C(K_2))$ of the Banach space $B = B(C(K_1), C(K_2))$ of all bounded linear operators from $C(K_1)$ into $C(K_2)$, with the usual operator norm. Corresponding to the positive face of S^* we have the convex subset S_1 of S, consisting of those T in S such that $T1 = 1$. It was shown by A. and C. Ionescu Tulcea [5] (and generalized in [8]) that $\operatorname{ext} S_1$ consists of the "composition operators", i.e., those of the form $(Tf)(k) = f(\psi(k))$ $(f \in C(K_1), k \in K_2)$ where ψ is a continuous function from K_2 into K_1. It is easily seen that these operators are also extreme in S; more generally, if $\psi: K_2 \to K_1$ is continuous and $\lambda \in C(K_2)$ with $|\lambda| = 1$,

Received February 11, 1965 and in revised form April 23, 1965.
[1] Supported in part by NSF Grant GP. 378.

then every operator T of the form

$$(*) \qquad Tf = \lambda \cdot f \circ \psi\ , \qquad f \in C(K_1)$$

is an extreme point of S. Indeed, if k is a point of K_2, then the functional on $C(K_1)$ defined by $f \to (Tf)(k) = \lambda(k) f(\psi(k))$ is the extreme point $\lambda(k)\varphi_{\psi(k)}$ of $S^* \subset C(K_1)^*$. The above assertion follows easily from this remark. The main result of the present paper is the converse of this assertion (Theorem 1), under the additional hypothesis that K_1 be metrizable. The proof is not elementary, in sharp contrast to the simplicity of the proofs for the corresponding statement for S_1. The metrizability hypothesis may be unnecessary; we use it only to enable us to apply one of Michael's selection theorems [6, 7]. If we consider the case of *complex*-valued continuous functions on K_1 and K_2, then the obvious analogues of the Arens-Kelley and Ionescu Tulcea results are valid ([4, p. 441] and [8]), and it is still true that every operator of the form (*) is extreme. We do not know whether the converse is true, however, even with the metrizability hypothesis.

It is possible to formulate our characterization of the extreme operators in S in another way: An operator T in $S(C(K_1), C(K_2))$ (K_1 metrizable) is extreme in S if and only if the functional $F_T(k): f \to (Tf)(k)$ ($f \in C(K_1)$) is extreme in $S^* \subset C(K_1)^*$ for a dense set of k's in K_2. (Simply use the fact that ext S^* is weak* compact and that $k \to F_T(k)$ is continuous.) In this form, the statement makes sense for operators T from an aribtrary Banach space X into $C(K)$, so that it is conceivable that the following is true: Suppose that X is a Banach space, that K is a compact Hausdorff space and that T is an operator in $S = S(X, C(K))$, the unit ball of bounded operators from X to $C(K)$. Then:

(A) *If T is extreme in S, then the functional $F_T(k)$ is extreme in the unit ball S^* of X^* for a dense set of points k in K.*

The converse of this assertion is clearly true. The assertion itself, however, is false; Theorem 2 shows that (in particular) if K has no isolated points, then there exists a space X and an extreme operator T in $S(X, C(K))$ such that for each k in K, $F_T(k)$ is *not* extreme in S^*. The hypothesis concerning isolated points is unavoidable, since it is easily verified that if T is extreme in $S(X, C(K))$ and k is an isolated point of K, then $F_T(k)$ is extreme in S^*. (Thus, if K has a dense set of isolated points, then assertion (A) is true. This observation let Arterburn and Whitley [2] to conjecture that it is generally true.) Assertion (A) is true if some strong restrictions are made on X; as noted in [2], (A) holds if every point in the boundary of S^* is extreme (i.e. if X^* strictly convex). Assertion (A) is also true if X is three-dimensional (cf. Theorem 3). An example shows that (A) is

no longer true for all four-dimensional spaces.

In order to prove Theorem 1 we make the usual identification between $C(K_1)^*$ and the space of finite signed Borel measures μ on K_1. For f in $C(K_1)$, $\mu(f)$ denotes the integral of f with respect to μ. Recall that $\mu = \mu^+ - \mu^-$ where μ^+, μ^- are nonnegative and have disjoint Borel supports; furthermore, $\|\mu\| = \mu^+(1) + \mu^-(1)$. Now, to each operator T in $B(C(K_1), C(K_2))$ there corresponds a unique function $k \to \mu_k$ from K_2 into $C(K_1)^*$ such that

$$(Tf)(k) = \mu_k(f) \qquad (k \in K_2, f \in C(K_1))$$

and

$$\|T\| = \sup\{\|\mu_k\| : k \in K_2\} .$$

This function is continuous from K_2 into $C(K_1)^*$ in its weak* topology (briefly: $k \to \mu_k$ is weak* continuous). Any weak* continuous function $k \to \nu_k$ for which $\|\nu_k\|$ is bounded defines an operator in B (by means of the above equation) with norm equal to $\sup \|\nu_k\|$. This correspondence between norm bounded weak* continuous functions and operators in B is linear; in particular, we have the following useful fact.

If $k \to \mu_k$ is related to T as above, then T is extreme in $S(C(K_1), C(K_2))$ if and only if the following is true: Whenever $k \to \nu_k$ is weak* continuous and $\|\mu_k \pm \nu_k\| \leqq 1$ for each k in K_2, then $\nu_k = 0$ for all k.

We can now prove our main result.

THEOREM 1. *Suppose that K_1 and K_2 are compact Hausdorff spaces, that K_1 is metrizable, and that T is an operator in $S(C(K_1), C(K_2))$. Then T is an extreme point of this set if and only if there exist continuous functions $\psi: K_2 \to K_1$ and λ in $C(K_2)$, $|\lambda| = 1$, such that*

$$(Tf)(k) = \lambda(k) f(\psi)) \qquad (k \text{ in } K_2, f \text{ in } C(K_1)) .$$

Proof. We have already noted that any operator of the above form is extreme, so suppose T is extreme in $S(C(K_1), C(K_2))$ and let $k \to \mu_k$ be the corresponding weak* continuous map from K_2 into the weak* compact unit ball S^* of $C(K_1)^*$. We distinguish two cases:

Case I. For each k in K_2, either $\mu_k \geqq 0$ or $\mu_k \leqq 0$.

Case II. For some k in K_2, the decomposition $\mu_k = \mu_k^+ - \mu_k^-$ is nontrivial.

We take care of Case I with the help of the Ionescu Tulcea theorem. Note that in this case, $\|\mu_k\| = |\mu_k(1)|$. Furthermore, the

real valued function $k \to |\mu_k(1)|$ is continuous on K_2, hence if μ is any nontrivial measure in S^*, then the function $k \to (1 - \|\mu_k\|)\mu$ is weak* continuous from K_2 into S^*. Since

$$\|\mu_k \pm (1 - \|\mu_k\|)\mu\| \leqq \|\mu_k\| + (1 - \|\mu_k\|)\|\mu\| \leqq 1 ,$$

we see that for each k, $|\mu_k(1)| = \|\mu_k\| = 1$. Thus, if we let $\lambda(k) = \mu_k(1)$, then $\lambda \in C(K_2)$ and $|\lambda| = 1$. Consider the operator $U: C(K_1) \to C(K_2)$ defined by $(Uf)(k) = \lambda(k)^{-1}(Tf)(k)$. It is easily verified that U is extreme in S and that $U1 = 1$. It follows from the Ionescu Tulcea theorem [5; 8, Theorem 1] that there exists a continuous function $\psi: K_2 \to K_1$ such that $Uf = f \circ \psi$, and hence T has the required form.

We next show that if Case II holds, then T can not be extreme. Assume, then, that there exists a point k_0 in K_2 such that $\mu_{k_0} = \mu_{k_0}^+ - \mu_{k_0}^-$ and $\mu_{k_0}^+ \neq 0$, $\mu_{k_0}^- \neq 0$. For each k in K_2, let

$$\Sigma(k) = \{\mu : \mu \in S^*, 0 \leqq \mu \leqq \mu_k^+\} .$$

This is a nonempty convex weak* closed subset of S^*, hence is weak* compact. We shall prove that the map $k \to \Sigma(k)$ from K_2 into the set of all subsets of $C(K_1)^*$ is lower semicontinuous in the sense of Michael [6] (where we take the weak* topology in $C(K_1)^*$). Thus, we must show the following: Given a point k_1 in K_2, a measure μ_1 in $\Sigma(k_1)$ and a weak* neighborhood V of μ_1, then there is a neighborhood U of k_1 such that $V \cap \Sigma(k)$ is nonempty for each k in U. We can assume that $V = \{\mu : \mu \in C(K_1)^*, |\mu(f_i) - \mu_1(f_i)| \leqq 1, f_i \in C(K_1), i = 1, 2, \cdots, n\}$. Suppose that there is a net k_α converging to k_1 such that $V \cap \Sigma(k_\alpha)$ is empty for each α. Since $\mu_{k_\alpha}^+ \in S^*$ for each α we may assume (by taking a subnet if necessary) that $\mu_{k_\alpha}^+ \to \mu_2$ for some $\mu_2 \geqq 0$ in S^*; similarly $\mu_{k_\alpha}^- \to \mu_3 \geqq 0$. Since $\mu_{k_\alpha} \to \mu_{k_1}$ we have $\mu_{k_1} = \mu_2 - \mu_3 = \mu_{k_1}^+ - \mu_{k_1}^-$ and it follows from the definition of the Hahn decomposition that $\mu_2 \geqq \mu_{k_1}^+$. Now, $\mu_{k_1}^+ \geqq \mu_1 \geqq 0$, so by the Radon-Nikodym theorem we can write $d\mu_1 = g_1 d\mu_2$, where g_1 is a Borel function on K_1 with $0 \leqq g_1 \leqq 1$. We can choose a continuous function g on K_1 such that $0 \leqq g \leqq 1$ and

$$\int_{K_1} |g - g_1| \, d\mu_2 \leqq (2 \max\{\|f_i\| : 1 \leqq i \leqq n\})^{-1} .$$

If we define μ_α by $d\mu_\alpha = g d\mu_{k_\alpha}^+$ then $\mu_\alpha \in \Sigma(k_\alpha)$ for each α. Furthermore, for each i we have

$$\mu_\alpha(f_i) = \int_{K_1} f_i d\mu_\alpha = \int_{K_1} f_i g d\mu_{k_\alpha}^+ \to \int_{K_1} f_i g d\mu_2$$

and hence

$$\limsup |\mu_\alpha(f_i) - \mu_1(f_i)| \leqq \int_{K_1} |f_i g - f_i g_1| \, d\mu_2 \leqq 1/2 ,$$

so that μ_α is eventually in $V \cap \Sigma(k_\alpha)$, a contradiction.

Thus, the set-valued map $k \to \Sigma(k)$ is lower semicontinuous and since $C(K_1)$ is separable—this is the only place where we use the metrizability of K_1—we may apply a selection theorem of Michael ([6, Th. 3.2*], [6, Ex. 1.3*] and [7]) to conclude the existence of a continuous selection for this map taking at k_0 the preassigned value $\mu_{k_0}^+$, that is, there exists $\sigma: K_2 \to C(K_1)^*$ which is weak* continuous such that $\sigma_{k_0} = \mu_{k_0}^+$ and $\sigma_k \in \Sigma(k)$ for each k, i.e. $0 \leqq \sigma_k \leqq \mu_k^+$. Similarly, there exists a continuous map $\tau: K_2 \to C(K_1)^*$ such that $\tau_{k_0} = \mu_{k_0}^-$ and $0 \leqq \tau_k \leqq \mu_k^-$ for each k. We define the map $\nu: K_2 \to C(K_1)^*$ by

$$\nu_k = \sigma_k \| \tau_k \| + \tau_k \| \sigma_k \| .$$

Since $\| \sigma_k \| = \sigma_k(1)$, the function $k \to \| \sigma_k \|$ is continuous (similarly for τ) and hence $k \to \nu_k$ is weak* continuous. Furthermore, $\nu_{k_0} \neq 0$. (This last assertion is the reason we needed two cases; in case I, $\nu_k \equiv 0$.) Finally, $\| \mu_k \pm \nu_k \| \leqq 1$ for each k. Indeed,

$$\begin{aligned} \| \mu_k + \nu_k \| &= \| (\mu_k^+ + \sigma_k \| \tau_k \|) \| + \| (\mu_k^- - \tau_k \| \sigma_k \|) \| \\ &= \| \mu_k^+ \| + \| \sigma_k \| \| \tau_k \| + \| \mu_k^- \| - \| \sigma_k \| \| \tau_k \| \\ &= \| \mu_k \| \leqq 1 ; \end{aligned}$$

similarly for $\| \mu_k - \nu_k \|$. It follows that T is not extreme, a contradiction which completes the proof.

It is clear that the above proof would be valid without the hypothesis that K_1 be metrizable if there were a selection theorem for the more general situation. Unfortunately, however, it is known (see e.g. [3]) that there exist compact Hausdorff space K_1 and K_2 (K_1 not metrizable, K_2 metrizable) and a weak* lower semicontinuous map from K_2 into the weak* compact convex subsets of the unit ball of $C(K_1)^*$ which does *not* have a continuous selection.

It is worth noting that we can eliminate the metrizability hypothesis in Theorem 1 if we put an additional hypothesis on the operator T. Namely, *if T is extreme in $S(C(K_1), C(K_2))$ and if the function $k \to \| \mu_k \|$ is continuous, then T is of the form* (*). We will sketch a proof of this fact. Since the proof of Theorem 1 used metrizability only in Case II, we will show that this case cannot occur. First, note that the continuity of $k \to \| \mu_k \|$ implies the weak* continuity of the functions $k \to \mu_k^+$ and $k \to \mu_k^-$. (This may be proved by contradiction, using the weak* compactness of S^*, the weak* continuity of $k \to \mu_k$ and the fact that if $\mu_k = \mu_1 - \mu_2$, $\mu_1, \mu_2 \geqq 0$, then $\mu_1 \geqq \mu_k^+$, $\mu_2 \geqq \mu_k^-$.) It follows that $k \to \nu_k = \mu_k^+(1)\mu_k^- + \mu_k^-(1)\mu_k^+$ is weak* continuous, and $\| \mu_k \pm \nu_k \| \leqq 1$ for each k. Since T is extreme, $\nu_k \equiv 0$; in particular, $0 = \nu_k(1) = 2\mu_k^+(1)\mu_k^-(1)$ for each k, which implies that Case I holds.

The above hypothesis on T is fulfilled in case T is a *compact*

operator; indeed, T is compact if and only if $k \to \mu_k$ is *norm* continuous [4, p. 490]. (Examination of the above proof and that in [8, Theorem 1] shows that in this case it suffices to assume only that T is extreme in the unit ball of compact operators.)

Suppose, now, that X is a Banach space and that T is an extreme point of $S(X, C(K))$ (K compact Hausdorff). As noted in the introduction, the functional $F_T(k)$ is extreme in the unit ball S^* of X^* whenever k is an isolated point of K. The next result shows that for any K there exists a space X such that these are the *only* such functionals which are extreme.

THEOREM 2. *Suppose that K is a compact Hausdorff space. Then there exists a Banach space X and an extreme operator T in $S(X, C(K))$ such that $F_T(k) \in \operatorname{ext} S^*$ (if and) only if k is an isolated point of K.*

Proof. Let M be the compact set of all points in K which are not isolated and let $l_2(M)$ be the Banach space of all real-valued functions y on M such that $\| y \|^2 = \Sigma_{k \in M} y^2(k) < \infty$. Let $X_0 = C(K) \oplus l_2(M)$, with norm $\| (f, y) \| = \max (\| f \|, \| y \|)$ ($f \in C(K)$, $y \in l_2(M)$). As is well known, the conjugate space X_0^* of such a product space is linearly isometric to $C(K)^* \oplus l_2(M)^*$, with $\| (\mu, y^*) \| = \| \mu \| + \| y^* \|$. Since $l_2(M)$ is a Hilbert space, we may identify $l_2(M)^*$ and $l_2(M)$. The space X_0^* in its weak* topology is linearly homeomorphic with the product space $C(K)^* \times l_2(M)^*$ (where we take the weak* topologies in each of these spaces). Let $S_1 = S_{C(K)^*} \times \{0\}$ and $S_2 = \{0\} \times S_{l_2(M)^*}$; these sets are weak* compact and convex, and the set $\operatorname{conv} (S_1 \cup S_2)$ is identical with the unit ball of X_0^*. (Consider $(\mu, y^*) = \| \mu \| (\mu / \| \mu \|, 0) + \| y^* \| (0, y^* / \| y^* \|)$.) For k in M, let $e_k \in l_2(M)$ be the function which is 1 at k, 0 elsewhere. Let $Q = \{\pm(\varphi_k, \pm e_k) : k \in M\} \cup \{\pm(\varphi_k, 0) : k \in M\}$ (as usual, $\varphi_k \in C(K)^*$ is evaluation at k). Note that Q is weak* compact. Indeed, it follows from the definition of $l_2(M)$ that any y in $l_2(M)$ gets arbitrarily small outside finite sets, so if $\{k_\alpha\}$ is a net in M with infinite range, then the set $\{e_{k_\alpha}\}$ has a subnet which converges to 0 in the weak* topology. This fact (together with compactness of M and weak* continuity of $k \to \varphi_k$) makes it possible to show that Q is weak* compact. Let A be the weak* closed convex hull of Q; then A is norm bounded, hence weak* compact, and the set $S' = \operatorname{conv} (S_1 \cup S_2 \cup A)$ is weak* compact and convex. Furthermore, S' is symmetric with respect to 0 and contains the unit ball $\operatorname{conv} (S_1 \cup S_2)$ of X_0^*. From standard results on duality, then, S' is the polar of a convex body in X_0 which defines a norm $\| \cdot \cdot \|'$ on X_0, equivalent to the original norm. Furthermore, if we let X denote X_0 in the norm $\| \cdot \cdot \|'$, then S' is the unit ball of the space X^*. We have defined our space X; to define the operator

T, let $F_T(k) = (\varphi_k, 0)$, $k \in K$. It is clear that F_T is weak* continuous; our next task is to show that T has the required properties.

Suppose, first, that $k \in K$ and that $(\varphi_k, 0) \pm (\mu, y^*) \in S'$. It follows immediately that $\|\varphi_k \pm \mu\| \leqq 1$, so that $\mu = 0$ and hence $(\varphi_k, \pm y^*) \in S'$. We will show that if $k \in K \sim M$, then $y^* = 0$, while if $k \in M$, then $y^* = \lambda e_k$ for some $|\lambda| \leqq 1$. To this end, let $\Phi_k = \{f : f \in C(K), f(k) = 1 = \|f\|\}$. If $f \in \Phi_k$, let $H_f = \{(\nu, z^*) : (\nu, z^*) \in X_0^*, \nu(f) = 1\}$; this is a weak* closed hyperplane and, since $(\nu, z^*) \in S'$ implies $\|\nu\| \leqq 1$ (hence $\nu(f) \leqq 1$), H_f is a supporting hyperplane to S'. In particular, $H_f \cap S'$ is an extremal subset of S'. Since the intersection of extremal sets is an extremal set, the set $J_k = \bigcap_{f \in \Phi_k} H_f$ is an extremal subset of S'; furthermore, it is elear that $(\varphi_k, y^*) \in S' \cap J_k$. Now, since J_k is extremal, $\operatorname{ext}(J_k \cap S') \subset J_k \cap \operatorname{ext} S'$, and from the definition of S' we see that $\operatorname{ext} S' \subset \operatorname{ext} S_1 \cup \operatorname{ext} S_2 \cup \operatorname{ext} A$. Also, since Q is compact, the Milman theorem implies that $\operatorname{ext} A \subset Q$. Since $J_k \cap S_2$ is empty, we have $\operatorname{ext}(J_k \cap S') \subset (J_k \cap \operatorname{ext} S_1) \cup (Q \cap J_k)$. Recall that if $(\nu, z^*) \in J_k$, $\|\nu\| \leqq 1$, then $\nu(f) = 1$ for all $f \in \Phi_k$, so $\nu = \varphi_k$. Thus, $J_k \cap S_1 = \{(\varphi_k, 0)\}$, and $Q \cap J_k = \{(\varphi_k, \pm e_k), (\varphi_k, 0)\}$ if $k \in M$; otherwise $Q \cap J_k$ is empty. Now, $(\varphi_k, y^*) \in J_k \cap S'$ and the latter is the closed convex hull of its extreme points. From what we have just shown, $(\varphi_k, y^*) = (\varphi_k, 0)$ if $k \in K \sim M$, while (φ_k, y^*) is a convex combination of $(\varphi_k, 0)$ and $(\varphi_k, \pm e_k)$ if $k \in M$; these remarks show that y^* has the form indicated earlier. It is clear that $(\varphi_k, 0) \in \operatorname{ext} S'$ if and only if $k \in K \sim M$. It remains to show that if there exists a weak* continuous map G: $K \to X_0^*$ such that $F_T(k) \pm G(k) \in S'$ for each k, then $G = 0$. Such a map would have the form $G(k) = (\mu_k, y_k^*)$, where $k \to \mu_k$, $k \to y_k^*$ are weak* continuous in the appropriate spaces, and we would have $(\varphi_k, 0) \pm (\mu_k, y_y^*) \in S'$. From what we have shown above, $\mu_k = 0$ for all k, $y_k^* = 0$ for $k \in K \sim M$, and $y_k^* = \lambda(k) e_k$ for some $|\lambda(k)| \leqq 1$ if $k \in M$. Since $k \to y_k^*$ is continuous, $y_k^* = 0$ for k in the closure of $K \sim M$. Suppose that k is an interior point of M and choose an infinite net of distinct points k_α in M with $k_\alpha \to k$. We know that $e_{k_\alpha} \to 0$ in the weak* topology, so $y^* k_\alpha = \lambda(k_\alpha) e_{k_\alpha} \to 0$ and hence $y_k^* = 0$. Thus, $y_k^* = 0$ for all k in K and the proof is complete.

Our next result shows that the conjecture of Arterburn and Whitley is true for two or three dimensional spaces X, as well as for certain other finite dimensional spaces.

THEOREM 3. *Suppose that X is a finite dimensional normed linear space such that*

(i) $\dim X \leqq 3$

or

(ii) *the unit ball of X is a polyhedron.*

If K is a compact Hausdorff space and T is an extreme operator in $S(X, C(K))$, then $\{k: k \in K \text{ and } F_T(k) \in \operatorname{ext} S^*\}$ is dense in K.

Proof. We shall give the proof for the case $\dim X = 3$, since the proof for the remaining cases is similar but much simpler. The idea of the proof (in any of the three cases) is this: We show that for each k, $F_T(k)$ must lie on the surface of the unit ball S^* of X^*. We then show that $F_T(k)$ cannot be in the (relative) interior of any face of S^*. In the two dimensional case, this completes the proof, while in the polyhedral case we can apply the same argument to faces of one less dimension, and the proof proceeds by induction. (Recall that in case (ii), the dual ball S^* is *also* a polyhedron.) In the three dimensional case we need a special argument, since the interior of a one dimensional face need not be isolated from other one dimensional faces. Suppose, then, that $\dim X = 3$ and that $T \in \operatorname{ext} S(X, C(K))$. Let $y^* \neq 0$ be a fixed element of S^*. The map $k \to (1 - \| F_T(k) \|)y^*$ is continuous, and $\| F_T(k) \pm (1 - \| F_T(k) \|) y^* \| \leqq \| F_T(k) \| + (1 - \| F_T(k) \|) \| y^* \| \leqq 1$; since T is extreme, this implies that $\| F_T(k) \| = 1$ for each k, i.e., $F_T(k)$ is on the boundary of S^* for each k. We next show that $F_T(k)$ cannot be an interior point of a two dimensional face of S^*, for any k in K. Indeed, if $F_T(k_0)$ were an interior point of such a face, the same would be true for all points of k in some neighborhood of k_0. We could then easily find a continuous real valued function g on K, not identically zero, and $x^* \in X^*$, $x^* \neq 0$, such the $\| F_T(k) \pm g(k)x^* \| \leqq 1$ for each k, contradicting the fact that T is extreme. Suppose, finally, that there exists nonempty open set $U \subset K$ such that $F_T(k) \notin \operatorname{ext} S^*$ for each k in U. From what we have shown it follows that for each x^* in $F_T(U)$ there exists a vector, $u(x^*)$ in X^*, unique up to sign, such that $\| u(x)^* \| = 1$ and $\| x^* \pm \varepsilon u(x^*) \| = 1$ for some $\varepsilon > 0$. For every x^* in $F_T(U)$, let $\lambda(x^*) = \sup \{\varepsilon : \| x^* \pm \varepsilon u(x^*) \| = 1\}$. Suppose that x_n^*, x^* are in $F_T(U)$ and that $x_n^* \to x^*$. The sequence $u(x_n^*)$ has a subsequence converging to some y^* in S^*. If $\lambda(x_n^*) \geqq \delta$ for some $\delta > 0$ and all n we see that $\| x^* \pm \delta y^* \| = 1$ and hence $y^* = \pm u(x^*)$ and $\lambda(x^*) \geqq \delta$. This shows that λ and therefore $\lambda \circ F_T$ are upper semi-continuous; in particular, $\lambda \circ F_T$ has points of continuity and hence there exist a nonempty open subset M of U and an $\varepsilon > 0$ such that $\lambda(x^*) \geqq \varepsilon$ for x^* in $F_T(M)$. We may assume that for some x in X, $| (u(x^*), x) | > 0$ for all x^* in $F_T(M)$ (otherwise replace M by

$$M \cap F_T^{-1}\{x^*; | (u(x^*), x) | > 0\} ,$$

where x is chosen so that this intersection is nonempty). For x^* in $F_T(M)$ let $e(x^*)$ be the sign (i.e. 1 or -1) such that $e(x^*)(u(x^*), x) > 0$. The compactness argument used above shows that the map $x^* \to$

$e(x^*)u(x^*)$ is continuous on $F_T(M)$. Choose a real valued continuous function g on K such that $g \neq 0$, $0 \leqq g \leqq \varepsilon$ and $g(k) = 0$ if $k \notin M$. Then for all k,

$$|| F_T(k) \pm g(k)e(F_T(k))u(F_T(k)) || \leqq 1 .$$

(We define $g(k)e(F_T(k))u(F_T(k))$ to be zero for $k \notin M$.) Since T is extreme, we conclude that the set U must have been empty and the proof is complete.

Our final result is an example which shows that the above theorem fails for a certain four dimensional space X.

EXAMPLE. There exists a four dimensional normed linear space X and an extreme operator T in $S(X, C[0, 1])$ such that $F_T(t) \notin \operatorname{ext} S^*$ for each t in $[0, 1]$.

Proof. Let g be a monotonic increasing function on $[0, 1]$ such that g is discontinuous at every rational point t and continuous elsewhere. Let Σ be the subset of R^4 consisting of the points

$$\pm(\cos t, \sin t, \cos g(t + 0), \sin g(t + 0))$$
$$\pm(\cos t, \sin t, -\cos g(t + 0), -\sin g(t + 0))$$
$$\pm(\cos t, \sin t, \cos g(t - 0), \sin g(t - 0))$$
$$\pm(\cos t, \sin t, -\cos g(t - 0), -\sin g(t - 0)) ,$$

where $0 \leqq t \leqq 1$. The set Σ is symmetric with respect to the origin and it is also compact: If $\{t_n\}$ is a sequence in $[0, 1]$ there is a subsequence $\{t_{n_k}\}$ such that either $t_{n_k} \uparrow t$ or $t_{n_k} \downarrow t$ for some t in $[0, 1]$.

Let S' be the convex hull of Σ; then S' is a compact, convex, symmetric subset of R^4 having nonempty interior, and hence is the unit ball of X^*, where X is the four-dimensional Banach space which has the polar S^0 of S' as its unit ball. Suppose that t is an irrational number in $[0, 1]$. The point $x = (\cos t, \sin t, 0, 0)$ is in S^0; indeed, for each (u, v, w, z) in Σ, $u \cos t + v \sin t \leqq 1$ and equality occurs only if $u = \cos t$, $v = \sin t$ and $(w, z) = \pm(\cos g(t), \sin g(t))$. It follows that if $(\cos t, \sin t, 0, 0) \pm (u, v, w, z) \in S'$, then (since the inner product of these two points with $x = (\cos t, \sin t, 0, 0)$ equals 1) $u = v = 0$ and the point $(\cos t, \sin t, w, z)$ is in the convex hull of the subset of Σ whose inner product with x equals 1, i.e. $w = \lambda \cos g(t)$, $z = \lambda \sin g(t)$, $|\lambda| \leqq 1$.

Let T be the operator from X into $C[0, 1]$ defined by $F_T(t) = (\cos t, \sin t, 0, 0)$, $t \in [0, 1]$. For each t, $F_T(t) \pm (0, 0, \cos g(t), \sin g(t)) \in \Sigma$, and hence $F_T(t) \in S'$ but $F_T(t) \notin \operatorname{ext} S'$. Suppose, however, that there exists $G: [0, 1] \to X^*$ which is continuous and satisfies $F_T(t) \pm G(t) \in S'$ for every t. Then, as we have seen, for every irrational t, $G(t) =$

$(0, 0, \lambda(t) \cos g(t), \lambda(t) \sin g(t))$ with $|\lambda(t)| \leqq 1$. Let t_0 be a rational point in $(0, 1)$. Since G is continuous, $\lim_{t \uparrow t_0} G(t) = \lim_{t \downarrow t_0} G(t)$ and since g is discontinuous at t_0 it follows that if $t_n \to t_0$ with t_n irrational, then $\lambda(t_n) \to 0$. Thus $G(t_0) = 0$, and therefore G vanishes on the rationals in $(0, 1)$ and hence identically on $[0, 1]$. It follows that T is extreme and the proof is complete.

BIBLIOGRAPHY

1. R. F. Arens and J. L. Kelley, *Characterizations of the space of continuous functions over a compact Hausdorff space*, Trans. Amer. Math. Soc. **62** (1947), 499-508.
2. David Arterburn and Robert Whitley, *Extreme linear operators*, dittoed notes.
3. Harry Corson and Joram Lindenstrauss, *Continuous selections with non-metrizable range*, to appear.
4. N. Dunford and J. Schwartz, *Linear Operators*, Vol. 1, New York, 1958.
5. A. Ionescu Tulcea and C. Ionescu Tulcea, *A note on extreme points*, dittoed notes.
6. E. Michael, *Continuous selections* I, Annals of Math. **63** (1956), 361-382.
7. ———, *A linear mapping between function spaces*, Proc. Amer. Math. Soc. **15** (1964), 407-409.
8. R. R. Phelps, *Extreme positive operators and homomorphisms*, Trans. Amer. Math. Soc. **108** (1963), 265-274.

UNIVERSITY OF WASHINGTON

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

A NOTE ON MULTIPLE EXPONENTIAL SUMS

L. CARLITZ

Put

$$S(c) = \sum_{x,y=1}^{p-1} e(x + y + cx'y') ,$$

Where $e(x) = e^{2\pi i/p}$ **and** $xx' \equiv yy' \equiv 1 \pmod{p}$, **Mordell has conjectured that** $S(c) = O(p)$. **The writer shows first, by an elementary argument that** $S(c) = O(p^{3/2})$. **Next he proves, using a theorem of Lang and Weil that** $S(c) = O(p^{11/8})$. **Finally he proves that** $S(c) = O(p^{5/4})$; **the proof makes use of the estimate**

$$\sum_{x=0}^{p-1} \psi(f(x)) = O(p^{1/2}) ,$$

where $\psi(a)$ **is the Legendre symbol and** $f(x)$ **is a polynomial of the fourth degree.**

If we put

$$K(a, b) = \sum_{x=1}^{p-1} e(ax + bx') ,$$

where $ab \not\equiv 0 \pmod{p}$, it is known that

$$|K(a, b)| \leqq 2p^{1/2} . \tag{2}$$

For proof of (2) see [1], [4].

Since

$$\begin{aligned} S &= \sum_{x=1}^{p-1} e(ax) \sum_{y=1}^{p-1} e(by + cx'y') \\ &= \sum_{x=1}^{p-1} e(ax)K(b, cx') , \end{aligned}$$

it follows that

$$|S| \leqq \sum_{x=1}^{p-1} |K(b, cx')| \leqq 2(p - 1)p^{1/2}$$

by (2). Thus, assuming (2), we get

$$S = O(p^{3/2}) . \tag{3}$$

However it is not difficult to prove (3) directly without making use of (2). Put

Received July 28, 1964, and in revised form September 23, 1964, Supported in part by NSF Grant GP-1593.

$$(4) \qquad S(c) = \sum_{x,y=1}^{p-1} e(x + y + cx'y') .$$

There is evidently no loss in generality in taking $a = b = 1$. Then we have

$$\sum_{c=0}^{p-1} |S(c)|^2 = \sum_{c=0}^{p-1} \sum_{x,y=1}^{p-1} \sum_{u,v=1}^{p=1} e\{x + y - uv + c(x'y' - u'v')\}$$
$$= p \sum_{xy \equiv uv \pmod p} e(x + y - u - v) .$$

But

$$\sum_{xy \equiv uv \pmod p} e(x + y - u - v) = \sum_{x,y,u=1}^{p-1} e(x + y - u - xyu')$$
$$= \sum_{y,u=1}^{p-1} e(y - u) \sum_{x=1}^{p-1} e\{x(1 - yu')\}$$
$$= -\sum_{y,u=1}^{p-1} e(y - u) + \sum_{y,u=1}^{p-1} e(y - u) \sum_{x=0}^{p-1} e\{x(1 - yu')\}$$
$$= -1 + p \sum_{y=1}^{p-1} 1 = p^2 - p - 1 ,$$

so that

$$(5) \qquad \sum_{c=0}^{p-1} |S(c)|^2 = p^3 - p^2 - p .$$

It follows at once from (5) that

$$(6) \qquad |S(c)| < p^{3/2} ,$$

so that we have proved (3).

2. Generalizing (4) we define

$$(7) \qquad S_n(c) = \sum_{x_1,\cdots,x_n=1}^{p-1} e(x_1 + \cdots + x_n + cx_1' \cdots x_n') .$$

We shall show that

$$(8) \qquad S_n(c) = O(p^{1/2(n+1)}) .$$

Exactly as above we have

$$(9) \qquad \sum_c |S_n(c)|^2 = p \sum_{x_1,\cdots,x_n} \sum_{y_1,\cdots,y_n} e(x_1 + \cdots + x_n - y_1 - \cdots - y_n) ,$$

where the summation is over all x_j, y_j such that

$$x_1x_2 \cdots x_n \equiv y_1y_2 \cdots y_n , \qquad x_j \not\equiv 0 , \quad y_j \not\equiv 0 \pmod p .$$

Let T_n denote the sum on the right of (9). Then we have

$$\begin{aligned} T_n &= \sum e(x_1 + \cdots + x_n - y_1 - \cdots - y_{n-1} - x_1 \cdots x_n y_1' \cdots y_{n-1}') \\ &= \sum_{\substack{x_1,\cdots,x_{n-1} \\ y_1,\cdots,y_{n-1}}} e(x_1 + \cdots + x_{n-1} - y_1 - \cdots - y_{n-1}) \\ &\qquad \cdot \sum_x e[(1 - x_1 \cdots x_{n-1} y_1' \cdots y_{n-1}')x] . \end{aligned}$$

The inner sum is equal to

$$\begin{cases} p - 1 & (x_1 \cdots x_{n-1} \equiv y_1 \cdots y_{n-1}) \\ -1 & (x_1 \cdots x_{n-1} \not\equiv y_1 \cdots y_{n-1}) , \end{cases}$$

so that

$$T_n = pT_{n-1} - \sum_{\substack{x_1,\cdots,x_{n-1} \\ y_1,\cdots,y_{n-1}}} e(x_1 + \cdots + x_{n-1} - y_1 - \cdots - y_{n-1}) .$$

Hence

$$T_n = pT_{n-1} - 1 . \tag{10}$$

Now

$$T_1 = \sum_{x \equiv y} e(x - y) = p - 1 , \qquad T_2 = p(p-1) - 1 = p^2 - p - 1$$

and generally

$$T_n = p^n - p^{n-1} - \cdots - 1 . \tag{11}$$

Thus (9) becomes

$$\sum_c |S_n(c)|^2 = p^{n+1} - p^n - \cdots - p \tag{12}$$

and (8) follows at once.

It follows from (12) that

$$S_n(c) = o(p^{n/2})$$

cannot hold for all c.

3. Returning to (4) we shall now show that

$$S(c) = O(p^{11/8}) . \tag{13}$$

It is convenient to put

$$S(a, b, c) = \sum_{x,y} e(ax + by + cx'y') .$$

Then

$$\sum_{a=1}^{p-1}\sum_{b=0}^{p-1}\sum_{c=0}^{p-1}|S(a,b,c)|^4 = p^3N, \tag{14}$$

where N denotes the number of solutions of the system

$$\begin{cases} x_1+x_2 \equiv x_3+x_4 \\ y_1+y_2 \equiv y_3+y_4 \\ x_1'y_1'+x_2'y_2' \equiv x_3'y_3'+x_4'y_4' \\ x_1x_2x_3x_4y_1y_2y_3y_4 \not\equiv 0. \end{cases}$$

Eliminating x_4, y_4 it follows that N is the number of solutions of

$$\begin{aligned}(x_1y_1+x_2y_2)x_3y_3(x_1+x_2-x_3)(y_1+y_2-y_3) \\ \equiv x_1y_1x_2y_2[(x_1+x_2-x_3)(y_1+y_2-y_3)+x_3y_3]\end{aligned} \tag{15}$$

such that

$$x_1x_2x_3y_1y_2y_3(x_1+x_2-x_3)(y_1+y_2-y_3) \not\equiv 0. \tag{16}$$

Now by a theorem of Lang and Weil [2] we have

$$N = p^5 + O(p^{5-1/2}),$$

so that (14) becomes

$$\sum_{a=0}^{p-1}\sum_{b=0}^{p-1}\sum_{c=0}^{p-1}|S(a,b,c)|^4 = p^8 + O(p^{15/2}). \tag{17}$$

On the other hand

$$\begin{aligned}\sum_{a=0}^{p-1}\sum_{b=0}^{p-1}\sum_{c=0}^{p-1}|S(a,b,c)|^4 &= |S(0,0,0)|^4 + 3\sum_{a=1}^{p-1}\sum_{b=1}^{p-1}|S(a,b,0)|^4 \\ &\quad + 3\sum_{a=1}^{p-1}|S(a,0,0)|^4 + \sum_{a=1}^{p-1}\sum_{b=1}^{p-1}\sum_{c=1}^{p-1}|S(a,b,c)|^4 \\ &= (p-1)^8 + (p-1)^2 + 3(p-1)^5 + (p-1)^2\sum_{c=1}^{p-1}|S(c)|^4,\end{aligned}$$

so that (17) reduces to

$$\sum_{c=1}^{p-1}|S(c)|^4 = O(p^{11/2}). \tag{18}$$

Clearly (18) implies (13).

4. If an exact formula for

$$\sum_{c=0}^{p-1}|S(c)|^4$$

were available we should presumably be able to prove

$$S(c) = O(p^{5/4}) . \tag{19}$$

In this connection it may be of interest to remark that the sum

$$\sum_{c=0}^{p-1} S^3(c) \tag{20}$$

can be evaluated. Indeed if we put

$$S(a, b, c) = \sum_{x,y} e(ax + by + cx' y') ,$$

then

$$\sum_{a=0}^{p-1} \sum_{b=0}^{p-1} \sum_{c=0}^{p-1} (S(a, b, c))^3 = p^3 N , \tag{21}$$

where N denotes the number of solutions of the system

$$\begin{cases} x_1 + x_2 + x_3 \equiv 0 \\ y_1 + y_2 + y_3 \equiv 0 \\ x_1'y_1' + x_2'y_2' + x_3'y_3' \equiv 0 \\ x_1x_2x_3y_1y_2y_3 \not\equiv 0 . \end{cases} \tag{22}$$

Eliminating x_3, y_3, we find that (22) reduces to

$$x_1(x_1 + x_2)y_1^2 + (x_1^2 + 3x_1x_2 + x_2^2)y_1y_2 + x_2(x_1 + x_2)y_2^2 \equiv 0 \tag{23}$$

together with

$$x_1x_2y_1y_2(x_1 + x_2)(y_1 + y_2) \not\equiv 0 . \tag{24}$$

We may replace (23) by

$$[(x_1 + x_2)y_1 + x_2y_2][x_1y_1 + (x_1 + x_2)y_2] = 0 . \tag{25}$$

If $x_1x_2(x_1 + x_2)y_1 \not\equiv 0$, it is clear from (25) that $y_2 \not\equiv 0$ and $y_1 - y_2 \not\equiv 0$. The two factors in (25) may vanish simultaneously. This will happen when

$$x_1^2 + x_1x_2 + x_2^2 \equiv 0 , \tag{26}$$

that is when -3 is a quadratic residue of p; moreover if x_1, x_2 satisfy (26) with $x_1x_2 \not\equiv 0$ then $x_1 + x_2 \not\equiv 0$. Thus the number of solutions of (26) is equal to

$$\left\{1 + \left(\frac{-3}{p}\right)\right\} \frac{p-1}{2} .$$

If -3 is a nonresidue we find that

$$N = 2(p-1)^2(p-2), \tag{27}$$

while, if -3 is a residue,

$$N = 2(p-1)^2(p-2) - (p-1)^2. \tag{28}$$

For $p = 3$ we have

$$N = 4, \tag{29}$$

for it is evident from (22) that $x_1 \equiv x_2 \equiv x_3$, $y_1 \equiv y_2 \equiv y_3$.

Combining (27) and (28) we have

$$N = 2(p-1)^2(p-2) - \left\{1 + \left(\frac{-3}{p}\right)\right\}\frac{(p-1)^2}{2} \qquad (p > 3). \tag{30}$$

On the other hand, since

$$S(0, 0, 0) = (p-1)^2 S(a, 0, 0) = -(p-1) \qquad (a \not\equiv 0),$$
$$S(a, b, 0) = 1 \qquad (ab \not\equiv 0),$$

we have

$$\sum_{a=0}^{p-1}\sum_{b=0}^{p-1}\sum_{c=0}^{p-1}(S(a, b, c))^3 = (p-1)^6 - 3(p-1)^4 + 3(p-1)^2 + \sum_{a=1}^{p-1}\sum_{b=1}^{p-1}\sum_{c=1}^{p-1}(S(a, b, c))^3$$
$$= (p-1)^6 - 3(p-1)^4 + 3(p-1)^2 + (p-1)^3\sum_{c=1}^{p-1}(S(c))^3.$$

Therefore, using (21) and (30), we get

$$\sum_{c=1}^{p-1}(S(c))^3 = 2p^3(p-2) - (p-1)^4 + 3(p-1)^2 - 3 - \frac{1}{2}\left\{1 + \left(\frac{-3}{p}\right)\right\}. \tag{31}$$

5. We shall now show that

$$S(c) = O(p^{5/4}). \tag{32}$$

With the notation of § 3 we have, as above,

$$\sum_{a=0}^{p-1}\sum_{b=0}^{p-1}\sum_{c=0}^{p-1}|S(a, b, c)|^4 = p^3 N, \tag{33}$$

where N is the number of solutions of the system

$$(34)\qquad \begin{cases} (x_1 + x_2)x_3x_4 \equiv x_1x_2(x_3 + x_4) \\ (y_1 + y_2)y_3y_4 \equiv y_1y_2(y_3 + y_4) \\ x_1y_1 + x_2y_2 \equiv x_3y_3 + x_4y_4 \\ x_1x_2x_3x_4y_1y_2y_3y_4 \not\equiv 0 . \end{cases}$$

Note that we have replaced each x_j, y_j by its reciprocal (mod p).

If we put

$$x_3 = x_1u_1 , \qquad x_4 = x_2u_2 , \qquad y_3 = y_1v_1 , \qquad y_4 = y_2v_2 ,$$

(34) becomes

$$(35)\qquad \begin{cases} (x_1 + x_2)u_1u_2 \equiv x_1u_1 + x_2u_2 \\ (y_1 + y_2)v_1v_2 \equiv y_1v_1 + y_2v_2 \\ x_1y_1 + x_2y_2 \equiv x_1y_1u_1v_1 + x_2y_2u_2v_2 \\ x_1x_2y_1y_2u_1u_2v_1v_2 \not\equiv 0 . \end{cases}$$

Now put $x_2 = x_1x$, $y_2 = y_1y$ and (35) reduces to

$$(36)\qquad \begin{cases} (1 + x)u_1u_2 \equiv u_1 + xu_2 \\ (1 + y)v_1v_2 \equiv v_1 + yv_2 \\ 1 + xy \equiv u_1v_1 + xyu_2v_2 \\ xyx_1y_1u_1v_1u_2v_2 \not\equiv 0 . \end{cases}$$

Finally, eliminating x, y we get the single equation

$$(37)\qquad \frac{(1 - u_1)(1 - v_1)(1 - u_1v_1)}{u_1v_1} + \frac{(1 - u_2)(1 - v_2)(1 - u_2v_2)}{u_2v_2} \equiv 0$$

subject to

$$(38)\qquad x_1y_1u_1v_1u_2v_2 \not\equiv 0 .$$

It should be noted that for fixed u_1, v_1, u_2, v_2 satisfying (37), x, y are uniquely determined by (36) unless $u_1 \equiv u_2 \equiv v_1 \equiv v_2 \equiv 1$; also we find that the forbidden cases $xy \equiv 0$ or xy "infinite" contribute $O(p^2)$.

Let $N'(k)$ denote the number of solutions u, $v \not\equiv 0$ of

$$(39)\qquad (1 - u)(1 - v)(1 - uv) \equiv kuv$$

and let $N(k)$ denote the total number of solutions of (39), so that

$$N(k) = N'(k) + O(1) .$$

Then clearly the number of nonzero solutions of (37) is equal to

$$(40)\qquad \sum_{k=0}^{p-1} N(k)N(-k) + O(p^2) .$$

Let $\psi(a)$ denote the Legendre symbol (a/p). Then for fixed u and k, the number of solutions of (39) is equal to

$$1 + \psi\{(1 + ku - u^2)^2 - 4u(1 - u)^2\} ,$$

so that

$$N(k) = p + \sum_{u=0}^{p-1} \psi(f(k, u)) ,$$

where

$$f(k, u) = (1 + ku - u^2)^2 - 4u(1 - u)^2 . \tag{41}$$

Thus (40) becomes

$$p^3 + 2p \sum_{k=0}^{p-1} \sum_{u=0}^{p-1} \psi(f(k, u)) + \sum_{k=0}^{p-1} \sum_{u=0}^{p-1} \sum_{v=0}^{p-1} \psi(f(k, u))\psi(f(-k, v)) + O(p^2) . \tag{42}$$

Since $f(k, u)$ is a quadratic in k we have

$$\sum_{k=0}^{p-1} \psi(f(k, u)) = -1$$

unless $u(1 - u) \equiv 0$. It follows that

$$\sum_{k=0}^{p-1} \sum_{u=0}^{p-1} \psi(f(k, u)) = O(p^2) . \tag{43}$$

Consider next the sum

$$\sum_{u=0}^{p-1} \psi(f(k, u)) .$$

It is easily seen from (41) that for fixed k, $f(k, u)$ is the square of a polynomial in u only when $k \equiv 0$. We therefore have the estimate

$$\sum_{u=0}^{p-1} \psi(f(k, u)) = O(p^{1/2}) , \tag{44}$$

so that

$$\sum_{k=0}^{p-1} \sum_{u=0}^{p-1} \sum_{v=0}^{p-1} \psi(f(k, u))\psi(f(-k, u)) = O(p^2) . \tag{45}$$

Substituting from (43) and (45) in (42) we see that the number of nonzero solutions (37) is

$$p^3 + O(p^2) .$$

Therefore N, the number of solutions of (34) is

$$p^5 + O(p^4)$$

and (33) becomes

$$\sum_{a=0}^{p-1} \sum_{b=0}^{p-1} \sum_{c=0}^{p-1} |S(a, b, c)|^4 = p^8 + O(p^7) ;$$

since $S(0, 0, 0) = p^2$,

$$S(a, b, c) = S(1, 1, abc) \qquad (abc \not\equiv 0)$$

and there are $(p - 1)^2$ terms $S(a, b, c)$ in the sum that give the same $S(1, 1, c)$, (32) now follows immediately.

Note that, except for (44), the proof is elementary.

REFERENCES

1. L. Carlitz and S. Uchiyama, *Bounds for exponential sums*, Duke Math. J. **24** (1957), 37-41.

2. Serge Lang and Ander Weil, *Number of points of varieties in finite fields*, Amer. J. Math. **76** (1953), 819-827.

3. L. J. Mordell, *On a special polynomial congruence and exponential sum*, Calcutta Mathematical Society Golden Jubilee Commemoration Volume (1958/59), Part I, pp. 29-32.

4. A. Weil, *Some exponential sums*, Proc. Nat. Acad. Sci. **34** (1949), 204-207.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

A NONNORMAL BLASCHKE-QUOTIENT

JOSEPH A. CIMA

We shall call the meromorphic functions of the form $F(z) = B_1(z)/B_2(z)$ Blaschke-quotients, where $B_1(z)$ and $B_2(z)$ are Blaschke products in $|z| < 1$ with zeros at $\{a_n\}$ and $\{b_k\}$ respectively. Although there is a characterization of meromorphic functions which are normal there is no characterization of the Blaschke-quotients which are normal in terms of the non-Euclidean (hyperbolic) distances between the zeros $\{a_n\}$ and $\{b_k\}$. In this paper we show by construction that even if the zeros of a Blaschke-quotient are separated by a positive non-Euclidean distance the Blaschke-quotient need not be normal.

We shall be concerned in this paper with the boundary behavior of meromorphic functions. The concept of a normal function was introduced by K. Noshiro in [2]. Several properties of normal meromorphic functions are developed in a paper of O. Lehto and K. I. Virtanen [2]. Their definition of a normal meromorphic function is:

DEFINITION A meromorphic function $f(z)$ is called normal in a simply-connected domain G if the family $\{f(S(z))\}$ is normal in G (in the sense of Montel), where $z' = S(z)$ denotes an arbitrary one-to-one conformal mapping of G onto itself.

In the same paper they gave the following characterization of a normal meromorphic function:

THEOREM 1. *A nonconstant function $f(z)$, meromorphic in a domain G, is normal if and only if*

$$\rho(f(z))\,|\,dz\,| \leqq C\,d\sigma(z)$$

at every point of G. $\rho(f)$ is the spherical derivative of f, $d\sigma$ is the hyperbolic element of length and C is an absolute constant.

The class of normal functions includes the bounded functions, schlicht functions and functions omitting three values. Moreover, it is known that the class of normal functions interselects the class of functions of bounded characteristic. Functions of bounded characteristic. Functions of bounded characteristic have the following form

$$F(z) = e^{i\lambda}\frac{B_1(z)}{B_2(z)}\exp\left(\frac{1}{2\pi}\int_0^{2\pi}\frac{e^{i\theta}+z}{e^{i\theta}-z}\,d\psi(\theta)\right).$$

Received April 17, 1964 and in revised form August 3, 1964.

$\psi(\theta)$ is a function of bounded variation, λ is real and the $B_i(z)$ are infinite products.

$$B(z, a_n) = B_1(z) = \prod_{n=1}^{\infty} \frac{|a_n|}{a_n}\left(\frac{a_n - z}{1 - \bar{a}_n z}\right),$$

$$\sum_{n=1}^{\infty}(1 - |a_n|) < \infty .$$

For brevity we refer to a quotient of the form $B_1(z)/B_2(z)$ as a Blaschke-quotient. There are functions which are normal which are not of bounded characteristic and functions of bounded characteristic which are not normal. In particular F. Bagemihl and W. Seidel [1] have constructed a holomorphic function of bounded characteristic which is not normal.

If z_1 and z_2 are two points of the unit disk $D = \{z \mid |z| < 1\}$, the non-Euclidean (hyperbolic) distance between z_1 and z_2 is given by the formula

$$\rho(z_1, z_2) = \frac{1}{2}\log\left(\frac{1 + |h|}{1 - |h|}\right), \quad h = \frac{z_1 - z_2}{1 - \bar{z}_1 z_2}$$

The following theorems of F. Bagemihl and W. Seidel [1] show that the non-Euclidean metric is in a sense a natural one for discussing the boundary behavior of a normal function.

THEOREM 2. *Let $\{z_n\}$ be a sequence of points in $D = \{z \mid |z| < 1\}$ which converges to a point $\zeta \in C = \{z \mid |z| = 1\}$ and is such that $\lim \rho(z_n, z_{n+1}) = 0$ as $n \to \infty$, and let $f(z)$ be a normal meromorphic function in D for which $\lim f(z_n) = c$ as $n \to \infty$, where c is finite or infinite. Then $f(z)$ has the angular limit c at ζ.*

THEOREM 3. *Let $\{z_n\}$ be a sequence of points in $D = \{|z| < 1\}$ for which $|z_n| \to 1$ and $f(z)$ is a normal, meromorphic function in D such that $\lim_{n\to\infty} f(z_n) = C$. If $\{z_n'\}$ denotes any sequence of points in D for which $\lim_{n\to\infty} \rho(z_n, z_n') = 0$, then also $\lim_{n\to\infty} f(z_n') = C$.*

One problem remaining in the study of normal functions is to determine those Blaschke-quotients which are normal. That is, if $F(z) = B_1(z)/B_2(z) = B_1(z; a_n)/B_2(z; b_n)$, to determine conditions on the non-Euclidean distances between the zeros of B_1 and B_2 which will make F normal. We note a few immediate results and then construct an example of a Blaschke-quotient such that the zeros of B_1 and those of B_2 are separated by a positive non-Euclidean distance and such that $B_1(z)/B_2(z)$ is not normal.

2. **The construction.** The first comment to be made is that a Blaschke-product has radial limits a.e. That is, $\lim B(re^{i\theta})$, $r \to 1$, exists and equals a complex uni-modular number $e^{i\varphi}$ [3]. The only singularities of $B(z) = B(z; a_n)$ in $|z| \leqq 1$ are those points z_0, $|z_0| = 1$, such that there is a subsequence $\{a_{n_k}\}$ of the zeros of $B(z; a_n)$ with $a_{n_k} \to z_0$ as $k \to \infty$. If we denote by $A(a_n)$ the set of limit points of the set $\{a_n\}$, then we have the following:

LEMMA 1. *Let* $F(z) = B_1(z)/B_2(z) = B_1(z; a_n)/B_2(z; b_n)$. *Then if* $A(a_n) \cap A(b_n)$ *is empty,* F *is normal.*

Proof. The condition of Theorem 1 for $F = B_1/B_2$ with domain the unit disk is

$$\rho(F(z))\,|dz| = \frac{|B_1(z)B_2'(z) - B_1'(z)B_2(z)|}{|B_1(z)|^2 + |B_2(z)|^2}\,|dz| \leqq \frac{C\,|dz|}{(1 - |z|^2)}\,.$$

For $P = e^{i\varphi}$ it suffices to show

$$\overline{\lim_{\substack{z \to P \\ |z|<1}}} \frac{|B_1(z)B_2'(z) - B_1'(z)B_2(z)|}{|B_1(z)|^2 + B_2(z)|^2}(1 - |z|^2) < \infty\,.$$

Now

$$\frac{|B_1(z)B_2'(z) - B_1'(z)B_2(z)|\,(1 - |z|^2)}{|B_1(z)|^2 + |B_2(z)|^2} \leqq \frac{|B_2'(z)|\,(1 - |z|^2) + |B_1'(z)|\,(1 - |z|^2)}{|B_1(z)|^2 + |B_2(z)|^2}\,.$$

The lemma follows from the fact that

$$\underline{\lim_{\substack{z \to P \\ |z|<1}}} (|B_1(z)|^2 + B_2(z)|^2) \geqq 1$$

and a result of W. Seidel and J. L. Walsh [5] which states that $|B_i'(z)|\,(1 - |z|)$ is bounded for $|z| < 1$.

The following lemma is a restatement of Theorem 3.

LEMMA 2. *If* $F(z) = B_1(z; a_n)/B_2(z; b_n)$, *with* $\{a_n\}$ *and* $\{b_n\}$ *disjoint, and if there exist subsequences* $\{a_{n_k}\}$ *and* $\{b_{n_k}\}$, *with* $\rho(a_{n_k}, b_{n_k})$ *tending to zero as* $k \to \infty$, *then* F *is not normal.*

We turn now to the example.

THEOREM 4. *There exists a Blaschke-quotient* $F(z) = B_1(z; a_n)/B_2(z; b_n)$ *with* $\rho(a_n, b_k) \geqq \delta_1 > 0$ *for all positive integers* n *and* k *such that* F *is not normal.*

Proof. The construction proceeds as follows. We exhibit a Blaschke-product $B_2(z; b_n)$ whose zeros lie in an angle at $z = 1$ and which tends to zero in angular approach to 1. We can then construct $B_1(z; a_n)$ such that the zeros $\{a_n\}$ tend to 1 and have the property expressed in the theorem. $B_1(z)$ has the additional property that there is a nontangential path Γ such that $\lim B_1(z)$ for z on Γ is positive. This implies

$$\lim_{\substack{z\to 1\\ z\in\Gamma}} \frac{B_1(z)}{B_2(z)} = \lim_{\substack{z\to 1\\ z\in\Gamma}} \frac{B_1(z; a_n)}{B_2(z; b_n)} = \infty$$

whereas $B_1(a_n)/B_2(a_n) = 0$. Applying Theorem 2, we see that F cannot be normal.

First define B_2 as follows:

$$B_2(z; b_n) = B_2(z; 1 - \frac{1}{n_2} = \prod_{n=1}^{\infty} \frac{\left(1 - \frac{1}{n^2}\right) - z}{1 - \left(1 - \frac{1}{n^2}\right) z}$$

$B_2(z)$ is well defined and holomorphic in $|z| < 1$. The function ρ_n,

$$\rho_n = \rho(b_n, b_{n+1}) = \frac{1}{2} \log \left(\frac{1 + |h_n|}{1 - |h_n|}\right)$$

where

$$|h_n| = \left| \frac{\left(1 - \frac{1}{n^2}\right) - \left(1 - \frac{1}{(n+1)^2}\right)}{1 - \left(1 - \frac{1}{n^2}\right)\left(1 - \frac{1}{(n+1)^2}\right)} \right| = \frac{2n+1}{2n^2 + 2n}$$

tends to zero as $n \to \infty$. Theorem 2 shows that B_2 has angular limit zero at $z = 1$.

For the remainder of the construction we find it useful to do some of the work in a half plane. The linear transformation T,

$$T(z) = i\left(\frac{1+z}{1-z}\right) = \omega = u + iv$$

maps $|z| < 1$ onto $v > 0$, preserving the non-Euclidean metric. The point $z = 1$ corresponds to $\omega = \infty$. As a nontangential path tending to infinity we choose the hypercycle $u = v$, $v > 0$. This curve has fixed non-Euclidean distance $1/2 \log(\sqrt{2} + 1)$ from $u = 0$, $v > 0$. Select points α_n' as follows:

$$\alpha_n' = \alpha_n e^{i\pi/4} = (2e^n - 1)e^{i\pi/4}\,.$$

The corresponding points a_n of $|z| < 1$ are

$$T^{-1}(a_n') = a_n = \frac{a_n' - i}{a_n' + i}.$$

Define $B_1(z)$ by using the points $\{a_n\}$ as the zeros of $B_1(z)$. To show $B_1(z)$ is well defined, it suffices to show $\sum_{n=1}^{\infty}(1 - |a_n|) < \infty$.

$$\sum_{n=1}(1 - |a_n|) = \sum_{n=1}^{\infty}\left(1 - \frac{(\alpha_n^2 + 1) - \sqrt{2}\,\alpha_n}{[(a_n + 1)^2 - 2\alpha_n^2]^{1/2}}\right) \leqq \sum_{n=1}^{\infty}\frac{\sqrt{2}}{\alpha_n}.$$

This last series has terms which are order of magnitude of e^{-n} and so converges.

We note that the curve $u = v$ maps under T^{-1} into a hypercycle h_α having a fixed non-Euclidean distance from the real axis. It also forms the fixed angle $\pi/4$ with the real axis at $z = 1$. Thus the zeros of $B_1(z)$ and those of $B_2(z)$ are separated by a positive non-Euclidean distance.

In the plane $v > 0$, $B_1(z)$ has the form

$$B_1(z) = B_1(T^{-1}(\omega)) = \pi_1(\omega) = \prod_{n=1}^{\infty} e^{ip_n}\left(\frac{\omega - a_n'}{\bar{a}_n' - \omega}\right)$$

where p_n is real. Now select points $i\beta_k = i(2k^2 - 1)$ on $u = 0$, $v > 0$.

$$T^{-1}(i\beta_k) = C_k = 1 - \frac{1}{k^2}.$$

If we show that $\underline{\lim}_{k\to\infty}|\pi_1(i\beta_k)| \geqq \delta > 0$, this implies

$$\underline{\lim_{k\to\infty}}\left|B_1\left(1 - \frac{1}{k^2}\right)\right| \geqq \delta > 0.$$

However, we know $\rho(C_k, C_{k+1})$ tends to zero as $k \to \infty$. Thus by Theorem 3, if $\{t_k\}$ is any real increasing sequence $t_k < 1$, $t_k \to 1$ with $B_1(t_k) \to \alpha$, then we can choose a subsequence $\{C_{n_k}\}$ such that

$$C_{n_k} \leqq t_k < C_{n_k+1}, \quad \rho(C_{n_k}, t_k) < \rho(C_{n_k}, C_{n_k+1}).$$

This implies $B_1(C_{n_k}) \to \alpha$ and so $|\alpha| \geqq \delta$. This proves $\underline{\lim}_{x\to 1}|B_1(x)|$ is positive. It remains to prove $\underline{\lim}_{k\to\infty}|\pi_1(i\beta_k)| \geqq \delta$.

The calculation for $\pi_1(\omega)$ is as follows:

$$|\pi_1(i\beta_k)|^2 = \prod_{n=1}^{\infty}\left|\frac{\beta_k e^{i(\pi/2)} - \alpha_n e^{i(\pi/4)}}{\alpha_n e^{-i(\pi/4)} - \beta_k e^{i(\pi/2)}}\right|$$

$$= \prod_{n=1}^{\infty}\left(1 - \frac{2\sqrt{2}\,\alpha_n\beta_k}{\alpha_n^2 + \beta_k^2 + \sqrt{2}\,\alpha_n\beta_k}\right).$$

Consider the following function of the positive real variables ξ and η,

$$g(\xi, \eta) = \frac{2\sqrt{2}(2e^{\xi} - 1)(2\eta^2 - 1)}{(2e^{\xi} - 1)^2 + (2\eta^2 - 1)^2 + \sqrt{2}(2e^{\xi} - 1)(2\eta^2 - 1)}.$$

Setting $x = [(2e^{\xi} - 1)/2\eta^2 - 1]^2$ we have

$$g(\xi, \eta) = \frac{2\sqrt{2}}{x + \frac{1}{x} + \sqrt{2}} \leqq \frac{2\sqrt{2}}{2 + \sqrt{2}} < 1$$

since $x + (1/x) \geqq 2$ if $x > 0$.

We are considering the product $\prod_{n=1}^{\infty}(1 - g(n, k))$, where $0 \leqq g(n, k) \leqq \sigma < 1$. We can choose a positive number A so large that $1 + x \geqq e^{Ax}$ for $-\sigma \leqq x \leqq 0$. The number $A = -\log(1 - \sigma)/\sigma$ will suffice $(A \geqq 2)$.

For each positive integer k let $N(k) = N$ be the positive integer satisfying $N < 2\log k \leqq N + 1$. $g(y, k)$ is monotone increasing on $1 \leqq y < 2\log k$ and monotone descreasing on $y > 2\log k$, with

$$g(2\log k, k) = \frac{1}{2(2k^2 - 1)}.$$

Thus we have for fixed k

$$\begin{aligned}
\sum_{n=1}^{\infty} g(n, k) &\leqq 2\sqrt{2}(2k^2 - 1)\sum_{n=1}^{\infty}\frac{(2e^n - 1)}{(2e^n - 1)^2 + (2k^2 - 1)^2} \\
&\leqq 2\sqrt{2}(2k^2 - 1)\left[\int_1^N \frac{(2e^y - 1)dy}{(2e^y - 1)^2 + (2k^2 - 1)^2}\right. \\
&\quad + \int_{N+1}^{\infty} \frac{(2e^y - 1)dy}{(2e^y - 1)^2 + (2k^2 - 1)^2} \\
&\quad \left. + \frac{(2e^N - 1)}{(2e^N - 1)^2 + (2k^2 - 1)^2} + \frac{(2e^{N+1} - 1)}{(2e^{N+1} - 1)^2 + (2k^2 - 1)^2}\right] \\
&\leqq 2\sqrt{2}(2k^2 - 1)\left[\int_1^{\infty} \frac{(2e^y)dy}{(2e^y - 1)^2 + (2k^2 - 1)^2} + \frac{2}{2(2k^2 - 1)}\right] \\
&= 2\sqrt{2}\arctan\left(\frac{2e^y - 1}{2k^2 - 1}\right)\Bigg|_1^{\infty} + 2\sqrt{2}.
\end{aligned}$$

Now evaluating the infinite product using the above estimates, we obtain

$$\begin{aligned}
\prod_{n=1}^{\infty}(1 - g(n, k)) &> \exp\left(-A\sum_{n=1}^{\infty} g(n, k)\right) \\
&\geqq \exp\left(-2\sqrt{2}A\left(\frac{\pi}{2} - \arctan\left(\frac{2e - 1}{2k^2 - 1}\right)\right) - 2\sqrt{2}A\right).
\end{aligned}$$

Thus $\underline{\lim}_{k\to\infty} |\pi_1(i\beta_k)|$ is positive. This completes the construction.

The function $F(z) = B_1(z)/B_2(z)$ tends to ∞ on the real axis as $x \longrightarrow 1$ and has a sequence of zeros on the nontangential curve (hypercycle) h_α. Thus F is not normal.

BIBLIOGRAPHY

1. F. Bagemihl, and W. Seidel, *Sequential and continuous limits of meromorphic functions*, Ann. Acad. Sci. Fenn. AI **280** (1960), 1-17.
2. O. Lehto, and K. I. Virtanen, *Boundary behavior and normal meromorphic functions*, Acta Math. **97** (1957), 47-65.
3. R. Nevanlinna, *Eindeutige Analytische Funktionen*, 2nd ed. Berlin, Gottingen, Heidelberg, 1953.
4. K. Noshiro, *Contributions to the theory of meromorphic functions in the unit-circle*, Journal of the Faculty of Science, Hokkaido Imperial University, Series 1, **7** (1939).
5. W Seidel, and J. L. Walsh, *On the derivatives of functions analytic in the unit circle and their radii of univalence and of p-valence*, Trans. Amer. Math. Soc. **52** (1942), 128-216.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

LIE AND JORDAN STRUCTURES IN BANACH ALGEBRAS

PAUL CIVIN AND BERTRAM YOOD

We first consider the theory of Jordan homomorphisms and Jordan ideals in Banach algebras. If B is a B*-algebra or a semi-simple annihilator algebra, any closed Jordan ideal in B is a two-sided ideal. Any Jordan homomorphism of a Banach algebra onto B is automatically continuous. That Jordan homomorphisms are continuous and Jordan ideals are ideals is shown to hold in a number of other situations. We also study the Lie ideals in a semi-simple Banach algebra A. If the center of A is zero and proper closed Lie ideals do not contain their Lie annihilators, then A is direct topological sum of its minimal closed ideals. An H^*-algebra with zero center is an example of such an algebra.

The utility of the study of Jordan isomorphisms in Banach algebra was noted by Kadison [8] in the study of isometrics of B^*-algebras. The Jordan and Lie structures of simple associative rings has been investigated by Herstein in a series of paper (see [3], [4], [5]). Essential use is made of these results in the present work.

2. Pure algebra. Let R be an associative ring. As is well-known [3] we can make R into a Jordan (Lie) ring by introducing the Jordan (Lie) multiplication $x \cdot y = xy + yx ([x, y] = xy - yx)$. For a subset S of R we consider the sets

$$S^J = \{x \in R \mid x \cdot u = 0 \text{ for all } u \in S\},$$
$$S^L = \{x \in R \mid [x, u] = 0 \text{ for all } u \in S\},$$
$$\mathfrak{R}(S) = \{x \in R \mid ux = 0 \text{ for all } u \in S\} \text{ and}$$
$$\mathfrak{L}(S) = \{x \in R \mid xu = 0 \text{ for all } u \in S\}.$$

By an *ideal* in R we mean, unless otherwise specified, a two-sided ideal.

2.1. LEMMA. *Let U be a Lie ideal in R. Then U^J and U^L are Lie ideals.*

Let $x \in U^J$, $u \in U$ and $b \in R$. Since $xu = -ux$, an easy computation shows that $[x, b] \cdot u = [b, u] \cdot x = 0$. Let $y \in U^L$. Since $yu = uy$, we obtain by straightforward calculation that $[[y, b], u] = [y, [b, u]] = 0$.

Received June, 23, 1964. This research was supported by the National Science Foundation grant NSF-G-25219.

2.2. LEMMA. *Let I be an ideal in R. Then*

$$[I^L, R] \subset \mathfrak{L}(I) \cap \mathfrak{R}(I) \subset I^L .$$

Let $y \in I$, $x \in I^L$ and $z \in R$. We get successively the relations $(xz)y = x(zy) = (zy)x = z(yx) = z(xy) = (zx)y$ or $[I^L, R] \subset \mathfrak{L}(I)$. Likewise $[I^L, R] \subset \mathfrak{R}(I)$.

For the remainder of this section, A shall denote a semi-prime alegbra over a field of characteristic $\neq 2$. (A semi-prime algebra is one without nilpotent right or left ideals $\neq(0)$). It follows from the hypothesis that A is semi-prime, that $\mathfrak{L}(I) = \mathfrak{R}(I)$ for any ideal I in A [10] p. 99].

2.3. LEMMA. *Let U be a Lie ideal in A. Then $U^J = \mathfrak{L}(U) = \mathfrak{R}(U)$ is an ideal in A and $U \cap U^J = (0)$.*

Take $x \in U \cap U^J$. Clearly $x^2 = 0$. Also, by Lemma 2.1, $[x, b] \in U \cap U^J$ for each $b \in A$. Then

$$0 = [x, b]^2 = xbxb - xb^2x + bxbx .$$

Multiplying on the left by x and on the right by b, we obtain $(xb)^3 = 0$. Thus xA is a nilpotent right ideal so that $xA = (0)$ and hence $x = 0$. Therefore $U \cap U^J = 0$.

Next let $u \in U$ and $x \in U^J$. The above shows that $[u, x] = 0$ and $u \cdot x = 0$. Thus $ux = xu = 0$. Hence

$$U^J \subset \mathfrak{L}(U) \cap \mathfrak{R}(U) . \tag{1}$$

Next we show that

$$\mathfrak{L}(U) = \mathfrak{R}(U) . \tag{2}$$

For take $x \in \mathfrak{L}(U)$, $u \in U$ and $b \in A$. We have $0 = x[b, u] = xbu$. Hence $(uxb)^2 = 0$. It follows that $ux = 0$ or $\mathfrak{L}(U) \subset \mathfrak{R}(U)$. Likewise $\mathfrak{R}(U) \subset \mathfrak{L}(U)$. A combination of (1) and (2) gives the desired result. The corresponding proposition for Jordan ideals in harder to prove (see Theorem 2.5).

Let U be a Jordan ideal in A. Let K denote the algebraic sum of the ideals in A contained in U. Clearly K is maximal in the set of ideals in A contained in U.

2.4. LEMMA. *The ideal K contains $a \cdot b$ and aAa for each a, b of the Jordan ideal U in A. If $U \neq (0)$ then $K \neq (0)$.*

By a lemma of Herstein [3, Lemma 1], $(a \cdot b)x - x(a \cdot b) \in U$ for each $x \in A$. Since $(a \cdot b) \cdot x \in U$ we see that $(a \cdot b)x$ and $x(a \cdot b)$ lie in U.

If also $y \in A$ then $(a \cdot b)xy + y(a \cdot b)x \in U$. From this we see that the ideal generated by $a \cdot b$ is contained in U. Therefore $a \cdot b \in K$. Also $(a \cdot x) \cdot a \in K$. Using $a^2 \in K$ we obtain $axa \in K$. If $K = (0)$, then $(aA)^2 = (0)$ and $a = 0$ for all $a \in U$.

2.5. THEOREM. *Let U be a Jordan ideal in A. Then $U^J = \mathfrak{L}(U) = \mathfrak{R}(U)$ is an ideal in A and $U \cap U^J = (0)$.*

Let $u \in U, x \in A$ and $z \in \mathfrak{L}(U)$. We have $0 = z(u \cdot x) = (zx)u$. Therefore $\mathfrak{L}(U)$ is an ideal in A. Since $[u\mathfrak{L}(U)]^2 = (0)$ we can conclude that $u\mathfrak{L}(U) = (0)$ or $\mathfrak{L}(U) \subset \mathfrak{R}(U)$. Likewise $\mathfrak{R}(U) \subset \mathfrak{L}(U)$. Thus $\mathfrak{L}(U) \subset U^J$.

Suppose that either a or b lies in U. Then

$$(a \cdot b) \cdot b = a \cdot (b^2) + 2bab \in U .$$

In either case it follows that $bab \in U$. Next let $u \in U$, $x \in A$. Then $(u \cdot x) \cdot (xu) = ux^2u + xu^2x + 2(xu)^2 \in U$. By the preceeding remark we see that $(xu)^2 \in U$. Likewise $(ux)^2 \in U$. Consider now an ideal I in A such that $I \cap U = (0)$. If $z \in I$ then $(uz)^2 \in I \cap U = (0)$. Therefore $uI = (0)$. We then have $(0) = UI = IU$.

Let K be the ideal in A maximal in the set of ideals in A contained in U (see Lemma 2.4). In particular $a^2 \in K$ for $a \in U$. Now $\mathfrak{L}(K)$ is an ideal in A. We show that $\mathfrak{L}(K) \cap U = (0)$. For let $b \in \mathfrak{L}(K) \cap U, x \in U$. Then $(bx)^2 \in U \cap \mathfrak{L}(K)$. This makes $(bx)^4$ an element of $K \cap \mathfrak{L}(K) = (0)$. Therefore $(bA)^4 = (0)$ and $b = 0$.

It follows from the above that $U\mathfrak{L}(K) = \mathfrak{L}(K)U = (0)$. Consequently $\mathfrak{L}(K) \subset \mathfrak{L}(U)$. But, inasmuch as $K \subset U$, we have also that $\mathfrak{L}(K) = \mathfrak{L}(U)$. From Lemma 2.3 we get $K^J = \mathfrak{L}(K) = \mathfrak{L}(U) \subset U^J$. But as $K \subset U$ we see that $K^J = U^J$. Whereas $U \cap \mathfrak{L}(K) = (0)$, we obtain $U \cap U^J = (0)$.

2.6. LEMMA. *If A satisfies the descending chain condition on right ideals, every Jordan ideal in A is an ideal.*

By the proof of Theorem 2.5, bab lies in the Jordan ideal U if a or b lies in U. It follows from [9, Lemma 5] that U is an ideal.

As we shall see below, Jordan ideals are automatically ideals under favorable conditions. If an algebra B (even a Banach algebra) is not semi-simple, this situation does not prevail and even Lemma 2.6 can fail there, as easy examples show.

For S, T subsets of A, by $S \cdot T$ we mean the collection of all finite sums of elements $x \cdot y, x \in S, y \in T$.

2.7. LEMMA. *Let I be an ideal in A. Then $I \cdot A$ is an ideal.*

Also $A \cdot A$ contains A^3.

The relation $a \cdot (b \cdot c) - (a \cdot b) \cdot c = [[a \cdot c], b]$ is readily verified. Suppose at least one of a, b and c is in I. Then $[[a, c], b] \in I \cdot A$. Since I is an ideal, $[a, c] \cdot b \in I \cdot A$. Hence $[a, c]b \in I \cdot A$ and $b[a, c] \in I \cdot A$.

Suppose that $b \in I$. Then $(ac) \cdot b - a[c, b] = (a \cdot b)c \in I \cdot A$. Thus $(a \cdot b)c \in I \cdot A$ and $[b, a]c \in I \cdot A$. Taking the sum and difference of these items, we see that abc and $bac \in I \cdot A$ for all $a, c \in A$. Therefore $(a \cdot b)c$ and likewise $c(a \cdot b) \in I \cdot A$ and $I \cdot A$ is an ideal. The proof shows that $A \cdot A \supset A^3$.

2.8. LEMMA. *A maximal ideal M in A is a maximal Jordan ideal.*

Let π be the natural homomorphism of A onto A/M. If A/M is not a zero algebra, then, by [3, Theorem 1], A/M has only itself and (0) as Jordan ideals. Therefore M is a maximal Jordan ideal. Suppose that A/M is a zero algebra. Since its only ideals are trivial, A/M is one-dimensional. This makes M a maximal linear subspace in A and hence a maximal Jordan ideal.

2.9. THEOREM. *If $A \cdot A = A$, then every maximal Jordan ideal M in A is an ideal.*

Let K be the largest ideal in A contained in M. If K is a maximal ideal in A, then $K = M$ by Lemma 2.8. We show that K is always a maximal ideal. For suppose otherswise. There exists an ideal I in A, $A \neq I$, $K \subset I$, $K \neq I$. If $I \supset M$ we are through and $I \subset M$ is impossible. Therefore $A = I + M$. Then $A \cdot M = I \cdot M + M \cdot M$. Lemma 2.4 gives $M \cdot M \subset K \subset I$ so that $A \cdot M \subset I$. Also $A \cdot A = A \cdot I + A \cdot M \subset I$. By hypothesis, this is impossible. In particular, by Lemma 2.7, the conclusion holds if $A^3 = A$.

2.10. THEOREM. *Suppose that $A \cdot A = A$ and each ideal in A is the intersection of the maximal ideals containing it. Then every Jordan ideal in A is an ideal.*

These conditions are satisfied, for example, if A is biregular in the sense of Arens and Kaplansky [1]. Let U be a Jordan ideal and let K be the largest ideal in A contained in U. If K is a maximal ideal, $K = U$ by Lemma 2.8. Suppose that K is not a maximal ideal and let M be any maximal ideal containing K. We show that $U \subset M$. For suppose otherwise. Clearly $M \not\subset U$ and $A = M + U$. Reasoning as in the proof of Theorem 2.9, we get $A \cdot A \subset M$, which is impossible. Hence

U is contained in every maximal ideal M, $M \supset K$. Therefore $U = K$.

3. **On topological rings.** Now let A be a semi-prime topological algebra over a field of characteristic $\neq 2$.

3.1. THEOREM. *Suppose that* $\mathfrak{L}(I) \neq (0)$ *for every closed ideal* $I \neq A$ *in* A. *Then any closed Jordan ideal* U *in* A *is an ideal in* A.

As in the preceding section we consider the ideal K in A maximal in the set of ideals contained in U. Clearly K is closed. Let $W = K \oplus \mathfrak{L}(K)$. Clearly $x^2 = 0$ for each $x \in \mathfrak{L}(W) = \mathfrak{R}(W)$. Therefore $\mathfrak{L}(W) = (0)$ and W is dense in A. Let $a \in U$, $u \in U$. There exists a net a_α in W such that $a_\alpha \to a$. For each a_α, $a_\alpha = b_\alpha + c_\alpha$, where $b_\alpha \in K$ and $c_\alpha \in \mathfrak{L}(K)$. From the proof of Theorem 2.5, $\mathfrak{L}(K) = \mathfrak{L}(U)$. Then we see that $a_\alpha u = b_\alpha u \in K$ for each α. Consequently $au \in K$. Likewise $ua \in K$, and U is an ideal in A.

The condition $\mathfrak{L}(I) \neq (0)$ for proper closed ideals in A gives the natural two-sided analogue of the annihilator algebras of Bonsall and Goldie [2].

3.2. THEOREM. *Let* B *be a semi-simple topological ring. Suppose that*

(a) *the primitive ideals of* B *are closed*;

(b) $\mathfrak{L}(I) \neq (0)$ *for each closed ideal* $I \neq A$.

Then B *is the direct topological sum of its minimal closed ideals.*

Note that (a) is automatically satisfied if B is a Banach algebra. Suppose first that (0) is a primitive ideal, i.e., B is primitive. Then (b) and $\mathfrak{L}(I)I = (0)$ for a closed ideal I show that all ideals in B other than (0) must be dense. The desired conclusion is then readily seen. We assume then that (0) is not a primitive ideal.

We call an ideal K in B *dual* if $\mathfrak{L}\mathfrak{R}(K) = K$. Our first step is to show that each primitive ideal P is dual. There exists a modular maximal right ideal M in B such that $P = \{a \in B \mid xa \in M \text{ for all } x \in B\}$ and P is the largest ideal in B contained in M. Let j be a left identity for B modulo M, let $b \in \mathfrak{L}\mathfrak{R}(P)$ and suppose that $b \notin P$. Then there exists an element $z \in B$ such that $zb \notin M$. We can write $j = u + kzb + zbx$ where $u \in M$, $x \in B$ and k is an integer. Multiplying on the right by an element $w \in \mathfrak{R}(P) = \mathfrak{L}(P)$, we see that $jw = uw \in M$. Since $jw - w \in M$, we also get $w \in M$. Therefore $P \oplus \mathfrak{R}(P) \subset M$. This is impossible since $\mathfrak{R}(P) \neq (0)$ by hypothesis and $P \oplus \mathfrak{R}(P)$ is a larger ideal than P contained in M. Hence P is dual.

Next take any ideal I in B, where $I \supset P$, $I \neq P$, for a primitive

ideal P. We show that $\mathfrak{L}(I) = (0)$ so that I is dense. Since $I \supset P$, we have $\mathfrak{L}(I) \subset \mathfrak{L}(P)$. Also $(0) = \mathfrak{L}(I)I \subset P$ and $I \not\subset P$. The theory of primitive ideals shows that $\mathfrak{L}(I) \subset P$. Therefore $\mathfrak{L}(I) \subset P \cap \mathfrak{L}(P) = (0)$.

Now let I be any dual ideal in B. We show that I is the intersection of the primitive ideals in B which contain I. Let π be the natural homomorphism of B onto B/I. Since $I \cap \mathfrak{R}(I) = (0)$, π is one-to-one on $\mathfrak{R}(I)$ and $\pi(\mathfrak{R}(I))$ is semi-simple. Let W denote the radical of B/I. Then $W \cap \pi(\mathfrak{R}(I)) = (0) = W\pi(\mathfrak{R}(I))$ by [6, p. 10]. Therefore $\pi^{-1}(W)\mathfrak{R}(I) \subset I$ from which we see that $\pi^{-1}(W)\mathfrak{R}(I) = (0)$. Thus $\pi^{-1}(W) \subset \mathfrak{L}\mathfrak{R}(I) = I$ and $W = (0)$. Inasmuch as B/I is now seen to be semi-simple, I is the intersection of the primitive ideals which contain it.

Next consider a primitive ideals P_0 and let P be any primitive ideal, $P \neq P_0$. By the above and our hypothesis, $P \not\subset P_0$ and $P_0 \not\subset P$ and $\mathfrak{L}(P + P_0) = (0)$. Therefore $\mathfrak{L}(P)\mathfrak{L}(P_0) = (0) = \mathfrak{L}(P_0)\mathfrak{L}(P)$ so that $\mathfrak{L}(P_0) \subset \mathfrak{R}\mathfrak{L}(P) = P$. Also $\mathfrak{R}\mathfrak{L}\mathfrak{R}(P_0) = \mathfrak{R}(P_0) = \mathfrak{L}(P_0)$ so that $\mathfrak{L}(P_0)$ is dual and $\mathfrak{L}(P_0) \not\subset P_0$. Therefore, as $\mathfrak{L}(P_0)$ is the intersection of the primitive ideals containing it, $\mathfrak{L}(P_0) = \cap P$, for $P \neq P_0$, P primitive. It is now clear that the ideals $\mathfrak{L}(P)$, P primitive, are the minimal dual ideals of B.

View $\mathfrak{L}(P_0)$ as a ring. By [6, p. 206], every primitive ideal of $\mathfrak{L}(P_0)$ is the intersection of $\mathfrak{L}(P_0)$ with a primitive ideal of B. Thus $\mathfrak{L}(P_0)$ is a primitive ring. Let $I \neq (0)$ be any closed ideal in B, $I \subset \mathfrak{L}(P_0)$ and let J be the algebraic sum of the ideals $\mathfrak{L}(P)$, $P \neq P_0$, P primitive. We claim that $I + J$ is dense in B. For let $x \in \mathfrak{R}(I + J)$. Then $x \in \mathfrak{R}\mathfrak{L}(P) = P$ for all $P \neq P_0$. Thus $x \in \mathfrak{L}(P_0) \cap \mathfrak{L}(I) = K$, say. Since $IK = (0)$ and $\mathfrak{L}(P_0)$ is a primitive ring, we get $K = (0)$. By (b), $I + J$ is dense in B.

We continue with this notation and show next that $\mathfrak{L}(P_0)\mathfrak{L}(P_0) \subset I$. Let $u \in \mathfrak{L}(P_0)$, $u = \lim (v_\alpha + q_\alpha)$, $v_\alpha \in I$, $q_\alpha \in J$ and let $x \in \mathfrak{L}(P_0)$. Now each $q_\alpha \in P_0$ so that $q_\alpha x = 0$. Therefore $ux = \lim v_\alpha x \in I$. This enables us to see that $\mathfrak{L}(P_0)(P_0 + \mathfrak{L}(P_0)) \subset I$. Inasmuch as $P_0 + \mathfrak{L}(P_0)$ is dense in B, we get $\mathfrak{L}(P_0)B \subset I$. Likewise $B\mathfrak{L}(P_0) \subset I$. From semi-simplicity we see that $B\mathfrak{L}(P_0) \neq (0)$, $\mathfrak{L}(P_0)B \neq (0)$. It follows that $\overline{B\mathfrak{L}(P_0)} = \overline{\mathfrak{L}(P_0)B}$ is a minimal closed ideal in B.

For each primitive ideal P_α, let $Z_\alpha = \overline{B\mathfrak{L}(P_\alpha)}$. It is clear that if $P_\alpha \neq P_\beta$, then $Z_\alpha \cap Z_\beta = (0) = Z_\alpha Z_\beta$. Then the algebraic sum of the Z_α is a direct sum. Let Q be its closure, the direct topological sum of the Z_α. Note that $\mathfrak{R}(Z_\alpha) = \mathfrak{R}(B\mathfrak{L}(P_\alpha)) = \mathfrak{R}\mathfrak{L}(P_\alpha) = P_\alpha$. Therefore $Q = B$. If I is any minimal closed ideal, I must be Z_α, for some α, for otherwise I would annihilate every Z_α.

3.3. COROLLARY. *Under the conditions of Theorem* 3.2. *the primitive ideals of B are the maximal closed ideals of B and every*

ideal in B which is not dense is contained in a primitive ideal. Also B is the direct topological sum of its minimal dual ideals.

Let I be closed ideal in B, $I \neq B$. Then $\mathfrak{L}(I) \neq (0)$ and $\mathfrak{L}(I)I = (0)$. If I were contained in no primitive ideal we would have $\mathfrak{L}(I)$ contained in every primitive ideal, which is imposible.

The proof of Theorem 3.2 shows that a primitive ideal is a maximal closed ideal. Let M be a maximal closed ideal. As just seen, $M \subset P$ for some primitive ideal P, so $M = P$. The proof of Theorem 3.2 also demonstrates that $\mathfrak{L}(P_1)\mathfrak{L}(P_2) = (0)$ if P_1, P_2 are two distinct primitive ideals and that B is the direct topological sum of the $\mathfrak{L}(P)$ which are the minimal dual ideals.

4. Continuity of Jordan homomorphisms. We consider a Jordan homomorphism T defined on a Banach algebra A with range a *dense subset of a semi-simple Banach algebra B.* We seek to show that, under reasonable conditions, such a mapping T is automatically continuous. For an element x in A or B we let $\rho(x)$ denote its spectral radius [10, p. 30].

Two useful identities are noted by Kadison [8, p. 330] for the behavior of Jordan homomorphisms relative to Lie products. These are

$$T([[a, b], c]) = [[Ta, Tb], Tc]$$

and

$$T([a, b]^2) = [Ta, Tb]^2$$

for all $a, b, c \in A$. Hence, as noted in [8], $ab = ba$ makes $[Ta, Tb]$ a central nilpotent element of B. Thus, as B is semi-simple, $[Ta, Tb] = 0$. Then $T(ab) = T(ab + ba)/2 = T(a)T(b) = T(b)T(a)$.

4.1. LEMMA. *For each $x \in A$, $\rho(T(x)) \leqq \rho(x)$.*

From the above discussion we see that if $a \circ b = b \circ a = 0$ that $T(a) \circ T(b) = T(b) \circ T(a) = 0$. From this the result is evident.

By the *separating set* for T we mean the set of $s \in B$ for which there exists a sequence $\{x_n\}$ in A with $\| x_n \| \to 0$ and $\| s - T(x_n) \| \to 0$. By the closed graph theorem, T is continuous if only if $S = (0)$.

Sraightforward arguments show that S is a closed Jordan ideal in B.

4.2. LEMMA. *If B has a left or right identity j, then $j \notin S$.*

Note that j need not lie in $T(A)$. Suppose that $j \in S$ where j is a left identity for B. Then there exists a sequence $\{x_n\}$ in A with

$\| x_n \| \to 0$ and $\| j - T(x_n) \| \to 0$. For each $z \in B$, $(zj)^n = z^n j$. Thus

$$\| (zj)^n \|^{1/n} \leqq \| z^n \|^{1/n} \| j \|^{1/n} .$$

From this we see that $\rho(zj) \leqq \rho(z)$. Clearly $\rho(j) = 1$. Now $j = (j - zj) + zj$ where the summands permute. Therefore

$$1 = \rho(j) \leqq \rho(j - zj) + \rho(zj) \leqq \rho(j - zj) + \rho(z) .$$

Now replace z by $T(x_n)$. We see that

$$\rho(j - T(x_n)j) \leqq \| j - T(x_n) \| \| j \| \to 0$$

and that, by Lemma 4.1, $\rho(Tx_n) \leqq \| x_n \| \to 0$. This yields a contradiction.

4.3. LEMMA. *Each element of S is a two-sided topological divisor of zero in B.*

This is a variation on a result of Rickart [10, p. 72]. Let $s \in S$. Arguments used there show that, if $1 - \lambda s$ is a two-sided topological divisor of zero for arbitrarily large λ, then s is a two-sided topological divisor of zero. We assume, then, that $1 - \lambda s$ is not a two-sided topological divisor of zero for $|\lambda| \geqq k$. Then λs is quasi-regular, $|\lambda| \geqq k$. Suppose that s is not a left topological divisor of zero in B. Then [10, p. 24], B has a left identity j and s is a regular elemement of the algebra jBj; there exists $u \in jBj$ such that $us = su = j$. Then $(u \cdot s)/2 = j \in S$ which is contrary to Lemma 4.2. Consequently s is a left (and similarly right) topological divisor of zero.

4.4. LEMMA. *Let $e \neq 0$ be any idempotent in $T(A)$. Then $e \notin S$.*

Suppose that $e = T(x)$, $e \in S$. There exists a sequence $\{x_n\}$ in A with $\| x_n \| \to 0$ and $\| e - T(x_n) \| \to 0$. Then also $\| e - eT(x_n)e \| \to 0$. From the theory of Jordan homomorphisms (see, for examples [8, p. 329]), $T(xx_nx) = eT(x_n)e$. Now $e - eT(x_n)e$ and $eT(x_n)e$ permute so that, by Lemma 4.1,

$$\begin{aligned} 1 = \rho(e) &\leqq \rho(e - eT(x_n)e) + \rho(eT(x_n)e) \\ &\leqq \| e - eT(x_n)e \| + \rho(T(xx_nx)) \\ &\leqq \| e - eT(x_n)e \| + \| xx_nx \| \to 0 . \end{aligned}$$

Therefore $e \notin S$.

If it is not required that the idempotent e be in $T(A)$ the following weaker conclusion holds.

4.5. LEMMA. *No central idempotent $e \neq 0$ of B lies in S.*

We use the notation of the proof of Lemma 4.4. If $e \in S$, then $1 = \rho(e) \leqq \rho(e - T(x_n)) + \rho(T(x_n)) \leqq \| e - T(x_n) \| + \| x_n \| \to 0$.

As an application we take $B = L_2(G)$ where G is a compact group and the multiplication in $L_2(G)$ is convolution (see [10, p. 330]). Suppose $T(A)$ is dense in $L_2(G)$. Lemma 4.5 shows that S cannot contain a central idempotent $\neq 0$. But in $L_2(G)$, every nonzero ideal contains a central idempotent. Therefore $S = (0)$ and T is continuous.

4.6. THEOREM. *A Jordan homomorphism T of a Banach algebra A onto a dense subset of a strongly semi-simple Banach algebra B is continuous.*

Let M be a modular maximal ideal of B, and let π denote the natural homomorphism of A onto B/M. The mapping πT is a Jordan homomorphism of A onto a dense subset of the simple algebra B/M. The separating set S_1 for the mapping πT is a Jordan ideal of B/M which cannot contain the identity of B/M by Lemma 4.2. But by Herstein's result [3, Theorem 1], S_1 is an ideal of B/M. Therefore $S_1 = (0)$ and πT is continuous. Simple arguments now show that $S \subset M$. Hence $S = (0)$.

4.7. THEOREM. *Let T be a Jordan homomorphism of a Banach algebra A onto a semi-simple Banach algebra B where every nonzero right (left) ideal in B contains a minimal right (left) ideal in B. Then T is con-tinuous.*

Examples of such B are the semi-simple annihilator Banach algebras of Bonsall and Goldie ([10, p. 96] and [2]).

Fix $a, b \in S$. By Lemma 2.4, the ideal generated by $a \cdot b$ lies in S. If the right ideal $\{(a \cdot b)x \mid x \in A\} \neq (0)$, it contains an idempotent $e \neq 0$ contrary to Lemma 4.4. Consequently $(a \cdot b)B = (0)$ and $a \cdot b = 0$ for all $a, b \in S$. For $a \in S$, $x \in B$ we have $(a \cdot x) \cdot a = 0$. Using $a^2 = 0$ we see that $(aB)^2 = (0)$. Hence a lies in the radical of B so that $S = (0)$.

4.8. THEOREM. *Let T be a Jordan homomorphism of a Banach algebra A onto a dense subset of a primitive Banach algebra B with minimal one-sided ideals. Suppose that $T(A) \cap I$ is dense I for some minimal right (left) ideal I. Then T is continuous.*

For the case of an algebra homomorphism see [12, p. 378]. Suppose that I is a right ideal. Let $I_1 = T^{-1}(I)$. If $T(x^2) = 0$ for all $x \in I_1$, then $u^2 = 0$ for each u in a dense subset of I. This would

make $I^2 = (0)$ which is impossible. We can express $I = jB$ where $j^2 = j$. Set $R = \{y \in B \mid jy \in I \cap T(A)\}$. Clearly R is a linear manifold where jR is dense in jB. Thus jRj is dense in jBj. By the Gelfand-Mazur theorem, $jRj = jBj$. Select $x \in I_1$ where $T(x) = jw \neq 0$ and $T(x^2) = jwjw \neq 0$. We have $jwj \neq 0$. Since jBj is a division ring there exists $z \in R$ where $jwjzj = jzjwj = j$. Consequently $(jw)\cdot(jz) \neq 0$. There exists $x_1 \in I_1$ such that $T(x_1) = jz$. Simple computations give $T((x\cdot x_1)/2) = (jw)\cdot(jz)/2 = h$, say, and $T[(x\cdot x_1)^2/4] = h$. Hence h is an idempotent, $\neq 0$, in $T(A) \cap I$. By Lemma 4.4, $h \notin S$ where S is the separating set for T. Consequently $I \not\subset S$.

Denote the socle of B by Σ and let $a, b \in S$. By the arguments of Theorem 4.7, the two-sided ideal Q generated by $a\cdot b$ lies in S. As B is primitive, $\Sigma \subset Q \subset S$ unless $a\cdot b = 0$. Thus, if $a\cdot b \neq (0)$, $h \in S$ which is impossible. Consequently we can reason as in Theorem 4.7 to see that $S = (0)$.

5. **Jordan ideals and homomorphisms in B*-algebras.** Throughout this section U will denote a closed Jordan ideal in a B^*-algebra A. For an element $x \in A$ which is quasi-regular [10, p. 16] we denote its quasi-inverse by x'. Since $x\cdot x' = (x + x')/2$, we see that $x' \in U$ if $x \in U$.

5.1. LEMMA. *If h is self-adjoint and $h^2 \in U$, then $h \in U$.*

For all $\alpha > 0$, $(-\alpha h^2)'$ exists and lies in U. Thus $h\cdot(-\alpha h^2)' \in U$. An argument of Rickart [10, Theorem 4.9.2] shows that for any $x \in A$,

$$x = \lim_{\alpha=\infty} x(-\alpha x^*x)' = \lim_{\alpha=\infty} (-\alpha xx^*)'x \,. \tag{5.1}$$

Applying this to our case, we obtain $2h = \lim h\cdot(-\alpha h^2)' \in U$.

5.2. LEMMA. $U^* = U$.

Let $a \in U$. By Lemma 2.4, the ideal generated by a^2 lies in U. In particular, $[-\alpha(a^2)(a^2)^*]' \in U$ for all $\alpha > 0$. By the proof of Theorem 2.5, $(a^2)^*[-\alpha a^2(a^2)^*]'(a^2)^* \in U$. Formula (5.1) implies that $(a^4)^* \in U$. Since $a^4 \in U$ it follows that $[a^2 + (a^2)^*]^2 = a^4 + a^2\cdot(a^2)^* + (a^4)^* \in U$, and so, by Lemma 5.1, $a^2 + (a^2)^* \in U$. Therefore $(a^2)^* \in U$, so that

$$(a + a^*)^2 = a^2 + a\cdot a^* + (a^2)^* \in U.$$

Another application of Lemma 5.1 shows that $a + a^*$, and thus a^*, lies in U.

5.3. THEOREM. *Any closed Jordan ideal in a B^*-algebras A is*

an ideal of A.

Let U be a closed Jordan ideal of A and let $a \in U$. Suppose $a = h + ik$, where h and k are self-adjoint. Since $a^* \in U$, it follows from Lemma 5.2 that $h \in U$ and $k \in U$. It thus suffices to see that if h and w are self-adjoint and $h \in U$, then hw and wh are in U. Now $(wh)^*(wh) = hw^2h \in U$, since $h \in U$. Also $(wh)(wh)^* = wh^2w \in U$, since $h \in U$. Consequently for $\alpha > 0$, $[-\alpha(wh)^*(wh)]' \in U$, and so $\lim_{\alpha\to\infty}(wh)[-\alpha(wh)^*(wh)]'(wh) \in U$. By formula (5.1), $(wh)^2 \in U$, and so by Lemma 5.2 $(hw)^2 \in U$. Since U is a Jordan ideal, it contains $wh + hw$. But also $[i(wh - hw)]^2 = -\{(wh)^2 - wh^2w - hw^2h + (hw)^2\} \in U$, so again by Lemma 5.1 $wh - hw \in U$, and thus wh and $hw \in U$, so U is an ideal.

Automatic continuity occurs for Jordan $*$-homomorphisms.

5.4. THEOREM. *Let T be a Jordan $*$-homomorphism of a B^*-algebra A onto a dense subset of a B^*-algebra B. Then*

(a) *T is continuous and $\| T \| = 1$,*

(b) *the range of T is B, and*

(c) *the adjoint mapping T' is an isometry.*

Before proceeding with the proof we should remark that these are instances of positive linear maps on operator algebras, which have been extensively studied. See [15].

Let A_0 be the closed $*$-subalgebra generated by a self-adjoint element $h \in A$. Then T is a $*$-homomorphism of the commutative B^*-algebra A_0 into B. Therefore $\| T(h) \| \leqq \rho(T(h)) \leqq \rho(h) \leqq \| h \|$. Consequently if h and k are any two self-adjoint elements of A,

$$\begin{aligned} \| T(h + ik) \| &\leqq \| T(h) \| + \| T(k) \| \leqq \| h \| + \| k \| \\ &\leqq \| h + ik \| + \| h - ik \| = 2 \| h + ik \| . \end{aligned}$$

Thus T is continuous.

Hence the kernel $T^{-1}(0)$ of the mapping T is a closed Jordan ideal in A. By Theorem 5.3, $T^{-1}(0)$ is an ideal and [10, p. 249] $A/T^{-1}(0)$

is a B^*-algebra in the quotient space norm. Let T_0 be the mapping of $A/T^{-1}(0)$ onto B defined by $T_0(x + T^{-1}(0)) = T(x)$ and let π be the natural homomorphism of A onto $A/T^{-1}(0)$. The mapping T_0 restricted to the closed $*$-subalgebra of $A/T^{-1}(0)$ generated by a normal element is an isometry by [10, p. 241]. That T_0 is an isometry on normal elements allows us to use the arguments of [8, Theorem 5] to assert that T_0 is an isometry on $A/T^{-1}(0)$. The range of T_0 and therefore that of T is then all of B, and $\| T(x) \| = \| \pi(x) \|$ for all $x \in A$. Since a natural homomorphism has norm one, $\| T \| = 1$. Now $T = T_0\pi$, so

$T' = \pi' T_0'$. As each of T_0' and π' is an isometry by the theory of normed linear spaces, so is T'.

5.5. THEOREM. *Let A, B be B^*-algebras with identities e_A and e_B, respectively. The Jordan $*$-homomorphisms of A onto B are precisely the continuous linear mappings T of A onto B such that (a) the adjoint T' is isometric, (b) the kernel of T is an ideal of A, and (c) $T(e_A) = e_B$.*

By Theorem 5.4, any Jordan $*$-homomorphism of A onto B has properties (a), (b) and (c).

Suppose T has the properties (a), (b), and (c). Let π be the natural homomorphism of A onto $A/T^{-1}(0)$. Let T_0 be as in the prior theorem. Then $T = T_0\pi$, and $T' = \pi' T_0'$. Since T' is an isometry, so is T_0'. As T_0 is a one-to-one mapping of $A/T^{-1}(0)$ onto B, T_0' is an isometry of the conjugate space of B onto the conjugate space of $A/T^{-1}(0)$. Thus we conclude that T_0 is an isometry of $A/T^{-1}(0)$ onto B. Since $T_0(e_A + T^{-1}(0)) = e_B$, a result of Kadison [8, p. 330] shows that T_0 is a Jordan $*$-isomorphism and thus that $T = T_0\pi$ is a Jordan $*$-homomorphism.

In [8, p. 329] Kadison shows that if T is a Jordan $*$-isomorphism of a B^*-algebra A onto a B^*-algebra B, then T is an isometry. The conventions of that paper require a unit for the algebra. However, a check of the argument involved in the proposition quoted above shows that the identity plays no role whatsoever. The next proposition is in the nature of a converse statement.

5.6. THEOREM. *Let A be a B^*-algebra, and B a Banach algebra with an involution. Let T be a Jordan $*$-isomorphism of A onto B which is an isometry. Then B is a B^*-algebra.*

Let $x \in A$. Then $\| (Tx)^n \| = \| Tx^n \| = \| x^n \|$, so $\rho(Tx) = \rho(x)$ for all $x \in A$, where ρ designates the spectral radius in either algebra.

Suppose $Tx = y \in \text{Rad } B$, the radical of B. Then $y^* \in \text{Rad}\, B$ and consequently $\rho(y \pm y^*) = 0 = \rho(x \pm x^*)$. But $x + x^*$ and $x - x^*$ are self-adjoint and skew elements in a B^*-algebra, so $x \pm x^* = 0$. Consequently $x = 0 = y$, and thus B is semi-simple.

It then follows (see § 4) that the spectrum of $T(x)$ if the same as that of x. Consequently, if $u \in B$ and u is self-adjoint, then its spectrum is real and $\rho(u) = \| u \|$. A result of Yood [13, p. 148] now asserts that there is a bicontinuous $*$-isomorphism σ of B onto a B^*-algebra B_1. The mapping σT of A onto B_1 is a Jordan $*$-isomorphism.

By the quoted result of Kadison, σT is an isometry. Since T is an isometry, so is σ and, therefore, B is a B^*-algebra.

We use the terminology *Jordan involution* for a conjugate linear mapping $x \to x^*$ of period two on a semi-simple complex Banach algebra A where $(x \cdot y)^* = x^* \cdot y^*$, $x, y \in A$. We write $H = \{x \in A \mid x = x^*\}$. Clearly $A = H \oplus iH$ and $\rho(x^*) = \rho(x)$, $x \in A$.

5.7. LEMMA. *Suppose that there is a real normed linear space norm $|\,x\,|$ on H such that $|\,x\,| \leqq \rho(x)$, $x \in H$. Then the Jordan involution $x \to x^*$ is continuous.*

Let H' be the derived set of H. Note that $y^2 \in H$ if $y \in H$. The arguments of [11, Lemma 3.3] show $y^2 = 0$ if $y \in (iH) \cap H'$. Thus $\rho(y) = \rho(iy) = 0$ and $y = 0$ for such y. It follows [11, p. 157] that H is closed. If $x \in A$ and $x = u + iv$, $u, v \in H$ the norm $\|x\|_1 = \|u\| + \|v\|$ is then a complete linear space norm for A topologically equivalent to the given norm so that the Jordan involution is continuous.

We are now able to establish the continuity of a Jordan homomorphism of a complex Banach algebra A onto a B^*-algebra. There need be no involution on A.

5.8. THEOREM. *Let T be a Jordan homomorphism of a complex Banach algebra A onto a B^*-algebra B. Then T is continuous.*

Let P_α be the set of primitive ideals of B, and let π_α be the canonical homomorphism of B onto B/P_α. Then $\pi_\alpha T$ is a Jordan homomorphism of A onto a primitive algebra. By a result of Herstein [4, p. 340], $\pi_\alpha T$ is either a homomorphism or an anti-homomorphism. Since B/P_α is a B^*-algebra, in either case we see that $\pi_\alpha T$ is continuous [10, Theorem 4.1.20]. This shows that $(\pi_\alpha T)^{-1}(0)$ is a closed ideal in A. But

$$T^{-1}(0) = \cap_\alpha (\pi_\alpha T)^{-1}(0) .$$

Therefore $T^{-1}(0)$ is a closed ideal in A. It is not difficult to show that $A/T^{-1}(0)$ is semi-simple.

Let π denote the natural homomorphism of A onto $A/T^{-1}(0)$ and let T_0 be the Jordan isomorphism $T_0(\pi(x)) = T(x)$ of $A/T^{-1}(0)$ onto B. By § 4 we see that $\rho[T(x)] = \rho[T_0(\pi(x))] = \rho[\pi(x)]$ since B is semi-simple.

The natural involution $x \to x^*$ on B induces a Jordan involution on $A/T^{-1}(0)$ by the rule $[T_0^{-1}(w)]^* = T_0^{-1}(w^*)$. Suppose that $T_0^{-1}(h)$ is self-adjoint under this Jordan involution. Then $T_0^{-1}(h) = [T_0^{-1}(h)]^* = T_0^{-1}(h^*)$ so that $h = h^*$ and h is self-adjoint in B. Thus $\rho(T_0^{-1}(h)) = \rho(h) = \|h\|$ for all h self-adjoint in B and $|\,T_0^{-1}(h)\,| = \|h\|$ serves as the auxiliary norm needed for Lemma 5.7. By that result, the Jordan involution

is continuous on $A/T^{-1}(0)$.

To show that T is continuous it is enough to show that T_0 is continuous. For the continuity of T_0 it suffices to show that $\|\pi(x_n)\| \to 0$ and $\| T_0(\pi(x_n)) - T_0(\pi(x_0)) \| \to 0$ imply that $T_0(\pi(x_0)) = T(x_0) = 0$. We can write $\pi(x_n) = \pi(u_n) + i\pi(v_n)$, $n = 0, 1, 2, \cdots$, where each $\pi(u_n)$ and $\pi(v_n)$ is self-adjoint in terms of the Jordan involution of $A/T^{-1}(0)$. Since that involution is continuous, $\| \pi(u_n) \| \to 0$ and $\| \pi(v_n) \| \to 0$. Each $T_0(\pi(u_n)) = T(u_n)$, and each $T(v_n)$, is self-adjoint in the B^*-algebra B. Thus $\| T(u_n) - T(u_0) \| \to 0$ and $\| T(v_n) - T(v_0) \| \to 0$. Also

$$\begin{aligned} \rho(T(u_0)) &= \| T(u_0) \| \leqq \| T(u_n) - T(u_0) \| + \| T(u_n) \| \\ &= \| T(u_n) - T(u_0) \| + \rho(T(u_n)) \\ &= \| T(u_n) - T(u_0) \| + \rho(\pi(u_n)) \to 0 \ . \end{aligned}$$

Therefore $T(u_0) = 0$ so that $\pi(u_0) = 0$. Likewise $\pi(v_0) = 0$ and thus $\pi(x_0) = 0$.

6. **Lie ideals in Banach algebras.** Throughout let B be a complex Banach algebra with center $\mathfrak{Z}$. Suppose that B is semi-simple. If $e^2 = e$ and $eB(Be)$ is a minimal right (left) ideal of B, we call e a *minimal idempotent* of B. By the Gelfand-Mazur Theorem, $eBe = Ke$, where K is the set of complex numbers. If B is primitive, it is known [10, p. 61] that either $\mathfrak{Z} = (0)$ or B has an identity u and $\mathfrak{Z} = Ku$.

Consider the Lie multiplication $[x, y] = xy - yx$ in B. In this topological algebraic setting we modify standard notation [7] as follows. Let $[S, T]$ be the closed linear span of the elements $[s, t]$ where $s \in S$, $t \in T$ and S, T are subsets of B. If $\mathfrak{L}$ is a Lie ideal of B we define $D^k\mathfrak{L}$ inductively by $D^0\mathfrak{L} = \overline{\mathfrak{L}}$ and $D^{k+1}\mathfrak{L} = [D^k\mathfrak{L}, D^k\mathfrak{L}]$. A Lie ideal $\mathfrak{L}$ is called *solvable* if for some positive integer k, $D^k\mathfrak{L} = (0)$. The *Lie radical* of B is defined to be the closed linear span of all the solvable Lie ideals of B.

As in [3] we define $S(\mathfrak{L}) = \{x \in B \mid [x, y] \in \mathfrak{L} \text{ for all } y \in B\}$.

6.1. THEOREM. *Let B be a primitive Banach algebra with socle $S \neq (0)$. Let e be a minimal idempotent and $\mathfrak{L} \neq (0)$ be a Lie ideal of B. Then*

(a) $e\mathfrak{L}e = Ke$.
(b) *If* $[\mathfrak{L}, \mathfrak{L}] = (0)$, *then* $\mathfrak{L} = \mathfrak{Z}$.
(c) *If* $e \in \mathfrak{L}$, *then* $\mathfrak{L} \supset S$.
(d) *If* $\mathfrak{L} \neq \mathfrak{Z}$, *then* $\mathfrak{L}$ *contains* $[x, y]$ *for all* $x \in S$, $y \in B$.

(a) Suppose that $e\mathfrak{L}e \neq Ke$. Then $e\mathfrak{L}e = (0)$. Let $a \in \mathfrak{L}$ and $x \in B$. We have $e(ax - xa)e = 0$ or $eaxe = exae$. Replacing x by xe, we see that $eaxe = 0$. This gives $(eax)^2 = 0$ for all $x \in B$, so that $ea = 0$.

Then $e(ax - xa) = 0 = exa$. From this we get $(BeB)(BaB) = (0)$. As B is primitive it follows that $a = 0$. This makes $\mathfrak{L} = (0)$.

(b) For any $x \in \mathfrak{L}$, $y \in B$ we have $[x, y^2] = [x, y]\cdot y \in \mathfrak{L}$. Since $[[x, y], y] \in \mathfrak{L}$. it follows that $(xy - yx)y$ and $y(xy - yx) \in \mathfrak{L}$. As $[\mathfrak{L}, \mathfrak{L}] = (0)$, x permutes with these elements and with $xy - yx$. Therefore $(xy - yx)yx = (xy - yx)xy$ or $(xy - yx)^2 = 0$. Since $e(xe - ex)^2e = 0$, we obtain $ex^2e = (exe)^2$, $x \in \mathfrak{L}$. Replacing x by $xy - yx$ we see that $[e(xy - yx)e]^2 = 0$. Since $eBe = Ke$, we have $e[\mathfrak{L}, B]e = (0)$. But $[\mathfrak{L}, B]$ is a Lie ideal so that, by (a), $\mathfrak{L} \subset \mathfrak{Z}$. But $\mathfrak{Z}$ is one-dimensional or (0). whence $\mathfrak{L} = \mathfrak{Z}$.

(c) Given $e \in \mathfrak{L}$, we have $[e, [e, x]] \in \mathfrak{L}$ for all $x \in B$. Inasmuch as exe is a scalar multiple of e, we see that $e\cdot x \in \mathfrak{L}$. As $[e, x] \in \mathfrak{L}$, we get $ex \in \mathfrak{L}$ and $[y, ex] \in \mathfrak{L}$ for all $x, y \in B$. But $[y, ex] = yex - exy$ so that $BeB \subset \mathfrak{L}$. By [6, p. 75] we see that $\mathfrak{L} \supset S$.

(d) By [3, p. 282], $S(\mathfrak{L})$ is both a Lie ideal and a subalgebra of B. Consequently, by [3, p. 281], either $\mathfrak{L} \subset S(\mathfrak{L}) \subset \mathfrak{Z}$ (so that $\mathfrak{L} = \mathfrak{Z}$) or $S(\mathfrak{L})$ contains a nonzero ideal of B. In the latter case, $S(\mathfrak{L}) \supset S$ since B is primitive. This yields (d).

6.2. THEOREM. *Let B be a primitive Banach algebra with minimal one-sided ideals. Then the Lie radical of B coincides with $\mathfrak{Z}$.*

It is sufficient to show that any solvable Lie ideal $\mathfrak{L}$ of B is contained in $\mathfrak{Z}$. Suppose $D^0\mathfrak{L} = \bar{\mathfrak{L}}$ and $D^{p+1}\mathfrak{L} = [D^p\mathfrak{L}, D^p\mathfrak{L}]$, $p = 0, 1, \cdots$, where k is the smallest integer such that $D^k\mathfrak{L} = (0)$. We may suppose $k \geqq 1$ and then $[D^{k-1}\mathfrak{L}, D^{k-1}\mathfrak{L}] = (0)$. By Theorem 6.1, we see that $D^{k-1}\mathfrak{L} \subset \mathfrak{Z}$. If $k = 1$ we are through. Otherwise $[D^{k-2}\mathfrak{L}, D^{k-2}\mathfrak{L}] \subset \mathfrak{Z}$. If $\mathfrak{Z} \neq (0)$ then $\mathfrak{Z} = Ku$ where u is the identity of B. But for $x, y \in D^{k-2}\mathfrak{L}$, $xy - yx = \alpha u$, $\alpha \neq 0$ is impossible. For if this relation persists then $sp(xy) = \alpha + sp(yx)$ whereas $sp(xy)$ and $sp(yx)$ agree up to the single value $\{0\}$ and are compact sets. Therefore $[D^{k-2}\mathfrak{L}, D^{k-2}\mathfrak{L}] = D^{k-1}\mathfrak{L} = (0)$ and $D^{k-2}\mathfrak{L} \subset \mathfrak{Z}$. This argument, repeated a finite number of times, leads to $\mathfrak{L} \subset \mathfrak{Z}$.

6.3. COROLLARY. *Let B be a semi-simple Banach algebra with socle $S \neq (0)$ where S is contained in no primitive ideal of B. Then the Lie radical of B is contained in $\mathfrak{Z}$.*

Let $\mathfrak{L}$ be a solvable Lie ideal of B, P be a primitive ideal and π be the natural homomorphism of B onto B/P. By hypothesis there exists a minimal idempotent e of B, $e \notin P$. Now $eBe = Ke$ so that $\pi(e)\pi(B)\pi(e) = K\pi(e)$ from which it follows that $\pi(e)$ lies in the socle

of B/P. Now $\pi(\mathfrak{L})$ is a solvable Lie ideal of B/P so that, by Theorem 6.2, either $\pi(\mathfrak{L}) = (0)$ or $\pi(\mathfrak{L}) = K\pi(u)$ where $\pi(u)$ is an identity for B/P. Hence, for $x \in \mathfrak{L}$, either $x \in P$ or we can write $x = \alpha u + y$, $\alpha \in K$, $y \in P$. In either case we get $[x, w] \in P$ for all $w \in B$. Therefore $[\mathfrak{L}, B]$ is in every primitive ideal or $\mathfrak{L} \subset \mathfrak{Z}$ by semi-simplicity.

There is an additional fact relating to primitive ideals which will be useful in the sequel.

6.4. LEMMA. *Let P be a primitive ideal of a Banach algebra B. Then either $S(P) = P$ or $S(P) = Ku + P$ where u is an identity for B modulo P.*

Let π be the natural homomorphism of B onto B/P. Let $x \in S(P)$ and $y \in B$. Then clearly $\pi(x)\pi(y) = \pi(y)\pi(x)$ or $\pi(x)$ lies in the center $\mathfrak{Z}_1$ of B/P. Now B/P is primitive so either $\mathfrak{Z}_1 = (0)$ or B/P has an identity $\pi(u)$ and $\mathfrak{Z}_1 = K\pi(u)$. If $\mathfrak{Z}_1 = (0)$ clearly $S(P) = P$. If $\mathfrak{Z}_1 = K\pi(u)$ then $P \subset S(P) \subset Ku + P$ so that either $P = S(P)$ or $Ku + P = S(P)$.

7. **Annihilation and Lie annihilation.** In Theorem 3.2 and Corollary 3.3 we obtained structure theorems for a class of algebras where, for each proper closed ideal I, $\mathfrak{L}(I) \neq (0)$. In this section we consider the consequences of a weaker hypothesis on the proper closed ideals for Banach algebras.

Before we investigate the weakened hypothesis, let us note the implication of the statement $\mathfrak{L}(I) = (0)$. In Lemma 2.2, we saw that if I is an ideal in a ring A, then $[I^L, A] \subset \mathfrak{L}(I) \cap \mathfrak{R}(I) \subset I^L$. Thus if $\mathfrak{L}(I) = (0)$, $I^L = \mathfrak{Z}$, the center of A. Suppose that A is semi-prime (so that $\mathfrak{L}(I) = \mathfrak{R}(I)$) and that $\mathfrak{Z}$ contains no nonzero ideal of A. Then if $I^L = \mathfrak{Z}$, $\mathfrak{L}(I) = (0)$. Hence for many of the algebras under consideration in this paper, $I^L = \mathfrak{Z}$ if and only if $\mathfrak{L}(I) = (0)$.

Throughout this section, B denotes a semi-simple Banach algebra, with center $\mathfrak{Z}$, satisfying the condition:

(7.1) *If $\mathfrak{L}(I) = (0)$, for an ideal I, then $I + \mathfrak{Z}$ is dense in B.*

Since $\mathfrak{L}(I + \mathfrak{R}(I)) = (0)$, an immediate consequence of (7.1) is that for every ideal I, $I + \mathfrak{R}(I) + \mathfrak{Z}$ is dense in B. Also, since $I + \mathfrak{Z} \subset S(I)$, it is immediate that $\mathfrak{R}(I) + S(I)$ is dense in B.

7.1. LEMMA. *Let I be a closed ideal of B. Then $\mathfrak{L}\mathfrak{R}(I) \subset S(I)$.*

Let $x \in \mathfrak{L}\mathfrak{R}(I)$, $y \in B$. We can write $y = \lim (u_n + v_n + z_n)$ where $u_n \in I$, $v_n \in \mathfrak{R}(I)$ and $z_n \in \mathfrak{Z}$. Then $xy - yx = \lim (xu_n - u_n x) \in I$.

Therefore $x \in S(I)$.

7.2. LEMMA. *A primitive ideal P for which $\mathfrak{L}(P) \neq (0)$ is a ideal. Also $\mathfrak{L}(P)$ is a minimal dual ideal.*

The first statement is shown by the proof of Theorem 3.2. Let $I \neq (0)$, $I \subset \mathfrak{L}(P)$, where I is dual. Since $I \not\subset P$ and P is a primitive ideal we see that $\mathfrak{L}(I) \subset P$. Thus $I = \mathfrak{R}\mathfrak{L}(I) \supset \mathfrak{L}(P)$ which forces $I = \mathfrak{L}(P)$.

7.3. LEMMA. *Let P be a primitive ideal of B for which $\mathfrak{L}(P) = (0)$. Then P is a modular maximal ideal of B of deficiency one, and there is an identity for B modulo P which is central.*

Proof. By (7.1), $P + \mathfrak{Z}$ is dense. Let π be the natural homomorphism of B onto B/P. Then $\pi(P + \mathfrak{Z}) = \pi(\mathfrak{Z})$ is dense in B/P. Therefore B/P is a commutative primitive Banach algebra, i.e., a field. Thus there exists $u \in \mathfrak{Z}$ such that $B = P + Ku$. Clearly u can be selected to be an identity for B modulo P.

7.4. THEOREM. *Let B be a semi-simple Banach algebra satisfying* (7.1). *Let I be the smallest dual ideal of B which contains all the minimal dual ideals of B, then B/I is commutative.*

Let J be the algebraic sum of all the ideals $\mathfrak{R}(P) \neq (0)$ for P primitive. By Lemma 7.2, $I \supset J$. Also clearly $\mathfrak{R}(J) = \cap P$, for the primitive ideals with $\mathfrak{R}(P) \neq (0)$.

Let $W = \cap M$, where M is a maximal ideal of deficiency one with a relative identity in $\mathfrak{Z}$. For any maximal modular ideal of the stated form $B = M + Ku$, with $u \in \mathfrak{Z}$, so $[B, B] \subset M$. Thus $[B, B] \subset W$. By (7.1), Lemma 7.3, and semi-simplicity, $W \cap \mathfrak{R}(J) = (0)$, so $W \subset \mathfrak{L}\mathfrak{R}(J) \subset I$. Thus $[B, B] \subset I$, and consequently B/I is commutative.

We next consider a hypothesis on the Lie ideals which will enable us to obtain results on the Lie structure of certain Banach algebras.

(7.2) *If $\mathfrak{L}$ is a closed Lie ideal of B and $\mathfrak{L} \supset \mathfrak{L}^L$, then $\mathfrak{L} = B$.*

7.5. THEOREM. *Let B be a semi-simple Banach algebra which satisfies* (7.2). *Then B satisfies the conclusion of Theorem* 7.4.

We show that if (7.2) is satisfied so is (7.1). In view of Lemma 2.2, if $\mathfrak{L}(I) = 0$ for an ideal I then $[I^L, B] = (0)$ or $I^L \subset \mathfrak{Z}$. Thus if $\mathfrak{I} = I + \mathfrak{Z}$, $\mathfrak{I}$ is a Lie ideal and $\mathfrak{I} \supset \mathfrak{Z} \supset I^L = \mathfrak{I}^L$. Consequently (7.2) shows that (7.1) is satisfied.

If the semi-simple Banach algebra B satisfies (7.2) and also

(7.3) *B has zero center*

then (7.1) is satisfied with $\mathfrak{Z} = (0)$, so we are with in the framework of § 3, and in particular Corollary 3.3 holds. Thus a structure holds in terms of minimal associative objects. We next turn our attention to obtaining a structure theorem under these hypotheses in terms of certain Lie ideals.

Any H^*-algebra [10, pp. 272–276] B with zero center satisfies (7.2) and (7.3). Let $\mathfrak{L}$ be a closed Lie ideal of B, and suppose that $\mathfrak{L} \supset \mathfrak{L}^L$. Let I be any minimal closed ideal of B. Direct calculation with the "matrix" representation of I, shows that the only closed Lie ideals in I are (0) and I. Hence $[\mathfrak{L}, I] = (0)$ or $[\mathfrak{L}, I] = I$. In the first instance $I \subset \mathfrak{L}^L \subset \mathfrak{L}$, and in the second $I \subset \mathfrak{L}$ directly. Thus for all minimal closed ideals of B, $\mathfrak{L} \supset I$, so $\mathfrak{L} = B$ and (7.2) is satisfied.

Until otherwise is stated we assume B is a semi-simple Banach algebra satisfying (7.2) and (7.3).

7.6. LEMMA. *B is Lie semi-simple.*

It is sufficient to show that $[\mathfrak{L}, \mathfrak{L}] = (0)$ for a Lie ideal $\mathfrak{L}$ of B implies $\mathfrak{L} = (0)$. Now $[\mathfrak{L}, \mathfrak{L}] = 0$ implies $\mathfrak{L} \subset \mathfrak{L}^L$, and thus that $\mathfrak{L}^{LL} \subset \mathfrak{L}^L$. Condition (7.2) then shows that $\mathfrak{L}^L = B$, whence $\mathfrak{L} \subset \mathfrak{Z} = (0)$ by (7.3).

The Lie semi-simplicity will frequently be used in the form that $\mathfrak{L} \cap \mathfrak{L}^L = (0)$ for any Lie ideal $\mathfrak{L}$ of B.

7.7. COROLLARY. *B contains no modular two-sided ideals.*

If M were a maximal modular ideal of B, then (7.1) implies $B = M \oplus \mathfrak{L}(M)$. Let e be an identity for B modulo M, with $e \in \mathfrak{L}(M)$. Clearly $e \in (\mathfrak{L}(M))^L \cap \mathfrak{L}(M) = (0)$ by the prior lemma. Thus we have a contradiction and there can be no modular ideals in B.

7.8. LEMMA. *If $\mathfrak{L}$ is a Lie ideal of B, then $S(\mathfrak{L}^L) = \mathfrak{L}^L$ and $S(L^-) = \mathfrak{L}^{LL}$.*

Clearly $L^L \subset S(\mathfrak{L}^L)$. Let $x \in S(\mathfrak{L}^L)$ and $y \in \mathfrak{L}$. Then $[x, y] \in \mathfrak{L}^L \cap \mathfrak{L} = (0)$, so $S(\mathfrak{L}^L) \subset \mathfrak{L}^L$.

Since $\mathfrak{L} \subset \mathfrak{L}^{LL}$ and the latter is closed, $S(\mathfrak{L}^-) \subset S(\mathfrak{L}^{LL}) = \mathfrak{L}^{LL}$ by the above. Now (7.2) implies that $\mathfrak{L} + \mathfrak{L}^L$ is dense in B. Let $x \in \mathfrak{L}^{LL}$. Then if $y \in B$, $y = \lim (y_j + z_j)$ with $y_j \in \mathfrak{L}$ and $z_j \in \mathfrak{L}^L$. Consequently, $[x,y] = \lim [x, y_j + z_j] = \lim [x, y_j] \in \mathfrak{L}^-$, and so $\mathfrak{L}^{LL} \subset S(\mathfrak{L}^-)$.

7.9. COROLLARY. *If P is a primitive ideal of B, then $p = p^{LL}$.*

It follows from Lemma 6.4 that $S(P) = P$ unless P is a modular ideal. However, by Corollary 7.7 there are no modular ideals, so Lemma 7.8 yields the desired result.

7.10. LEMMA. *If I is a closed ideal in B, then $(\mathfrak{R}(I))^L = S(I)$.*

Since $\mathfrak{R}(I) \subset I^L$, it follows from Lemma 7.8 that $S(I) = I^{LL} \subset (\mathfrak{R}(I))^L$. Let $x \in (\mathfrak{R}(I))^L$ and $y \in B$. Since the ideal $I + \mathfrak{R}(I)$ is dense in B, $y = \lim (p_n + q_n)$, with $p_n \in I$ and $q_n \in \mathfrak{R}(I)$. Then

$$[x, y] = \lim [x, p_n + q_n] \lim [x, p_n] \in I ,$$

and consequently $(\mathfrak{R}(I))^L \subset S(I)$.

7.11. LEMMA. *Let P be a primitive ideal of B. Then $[\mathfrak{R}(P), \mathfrak{R}(P)]$ is contained in every nonzero closed Lie ideal in $\mathfrak{R}(P)$.*

Let $\mathfrak{J} \neq (0)$ be a closed Lie ideal of B with $\mathfrak{J} \subset \mathfrak{R}(P)$. By the use of two results of Herstein, we see that $S(\mathfrak{J})$ is a ring as well as a Lie ideal [3, p. 282] and therefore [3, p. 281] that either $S(\mathfrak{J})$ is in the center of B or contains a nonzero ideal I of B. The first possibility cannot pertain here since (7.3) holds. Thus we have a nonzero ideal $I \subset S(\mathfrak{J}) = \mathfrak{J}^{LL} \subset (\mathfrak{R}(P))^{LL} = P^L$ by use of Lemmas 7.8, 7.10 and Corollary 7.9. This implies that $\mathfrak{L}(\mathfrak{R}(I)) \cdot \mathfrak{R}(P) \neq 0$, for otherwise $I \subset \mathfrak{L}(\mathfrak{R}(I)) \subset P = \mathfrak{L}(\mathfrak{R}(P))$, and thus $I \subset P \cap P^L = (0)$, which is a contradiction. Consequently $\mathfrak{L}(\mathfrak{R}(I)) \cap \mathfrak{R}(P)$ is a dual ideal $\neq (0)$ contained in the minimal dual ideal $\mathfrak{R}(P)$. Therefore $\mathfrak{R}(P) \subset \mathfrak{LR}(I)$ and $\mathfrak{R}(I) \subset P$. Also $S(I) = I^{LL} \subset \mathfrak{J}^{LLLL} = \mathfrak{J}^{LL} = S(\mathfrak{J}) \subset P^L$. Another application of Lemma 7.10 yields $P^L \subset (\mathfrak{R}(I))^L = S(I)$. However, from $I \supset P^L$, we deduce $S(I) = I^{LL} \subset P^L$, so $P^L = S(I) = S(\mathfrak{J})$. Let D be the largest ideal contained in P^L. The Theorem 1 in [14, p. 156] together with its proof shows that $\mathfrak{J} \supset [D, B]$. Since $D \supset \mathfrak{R}(P)$, $\mathfrak{J} \supset [\mathfrak{R}(P), \mathfrak{R}(P)]$.

7.12. LEMMA. *For any primitive ideal P of B, $[\mathfrak{R}(P), \mathfrak{R}(P)]^L = P$.*

The inclusion $P \subset [\mathfrak{R}(P), \mathfrak{R}(P)]^L$ is immediate. Let $x \in P^L = (\mathfrak{R}(P))^{LL}$ by Lemmas 7.9 and 7.10. For $y \in B$, $y = \lim (p_n + r_n)$ with $p_n \in P$ and $r_n \in \mathfrak{R}(P)$, and $x = \lim (s_m + t_m)$ with $s_m \in P$ and $t_m \in \mathfrak{R}(P)$. Thus

$$\begin{aligned} [x, y] &= \lim_n [x, p_n + r_n] = \lim_n [x, r_n] = \lim_n \lim_m [s_m + t_m, r_n] \\ &= \lim_n \lim_m [t_m, r_n] \in [\mathfrak{R}(P), \mathfrak{R}(P)] . \end{aligned}$$

Consequently $P^L \subset S([P), \Re(P)]) = [\Re(P), \Re(P)]^{LL}$, and so $[\Re(P), \Re(P)]^L \subset P^{LL} = P$. We thus have the asserted equality.

7.13. LEMMA. *Each primitive ideal P of B is a maximal proper closed Lie ideal of B.*

Let $\Im$ be a proper closed Lie ideal of B such that $\Im \supset P$. The hypothesis (7.2) implies that $\Im^L \neq (0)$. Corollary 3.3 and its proof imply that each minimal dual ideal of B is of the form $\Re(Q)$ for Q a primitive ideal of B. If $\Im^L \cap \Re(Q) = (0)$ for all primitive ideals Q, then $[\Im^L, \Re(Q)] \subset \Im^L \cap \Re(Q) = (0)$, so by Corollary 3.3, $[\Im^L, B] = (0)$ and $\Im^L \subset Z = (0)$ which is a contradiction. Thus there exists a primitive ideal Q_0 such that $\Im^L \cap \Re(Q_0) \neq (0)$. Lemma 7.11 then yields $[\Re(Q_0), \Re(Q_0)] \subset \Im^L$. Lemma 7.12 then implies that $\Im \subset \Im^{LL} \subset [\Re(Q_0), \Re(Q_0)]^L = Q_0$. As noted in the proof of Theorem 3.2, the resulting inequality $Q_0 \subset P$ with Q_0 and P primitive ideals implies that $Q_0 = P$ and consequently $\Im = P$.

7.14. LEMMA. *Each minimal closed Lie ideal of B is of the form $[\Re(Q), \Re(Q)]$ for some primitive ideal Q of B.*

Let $\mathfrak{D}$ be a minimal closed Lie ideal of B. As noted in the proof of Lemma 7.13, $\mathfrak{D} \cap \Re(Q) \neq (0)$ for some primitive ideal Q of B; and so $\mathfrak{D} \subset \Re(Q)$. Thus by Lemma 7.11, $[\Re(Q), \Re(Q)] = \mathfrak{D}$ by the minimal nature of $\mathfrak{D}$.

7.15. THEOREM. *If B is a semi-simple Banach algebra which satisfies* (7.2) *and* (7.3), *then B is the direct topological sum of its minimal closed Lie ideals.*

In view of Lemma 7.14, and the fact that any two ideals $\Re(Q)$, with Q primitive, have only zero in common, a sum of minimal closed Lie ideals is a direct sum. Let $\Im$ be the direct topological sum of the minimal closed Lie ideal in B. If $\Im \neq B$, then (7.2) implies $\Im^L \neq (0)$. The argument used in the proof of Lemmas 7.13 and 7.14 shows that there is a primitive ideal Q such that $\Im^L \cap \Re(Q) \neq (0)$. Lemma 7.11 then implies that $[\Re(Q), \Re(Q)] \subset \Im^L$. But then $[\Re(Q), \Re(Q)] \subset \Im^L \cap \Im = (0)$, which is a contradiction.

In the final portion of this section, B will denote a simple Banach algebra with an identity u, which satisfied (7.2) and

(7.4) *The center $\mathfrak{Z}$ of B consists of the complex multiples of the identity u of B.*

In view of the proof of Theorem 7.5, we note that (7.1) is automatically satisfied by B. Thus if $\mathfrak{L}(I) = (0)$, for an ideal I, either I is dense in B or I is a maximal ideal of deficiency one in B.

7.16. LEMMA. *If $\mathfrak{R}(M) \neq (0)$ for each maximal ideal M of B, then each primitive ideal of B is a maximal ideal. Furthermore, B is the direct topological sum of its minimal dual ideals.*

Since B has an identity, each primitive ideal P is contained in a maximal ideal M. If $M \neq P$, then by the argument used in the proof of Theorem 3.2, $\mathfrak{R}(M) = (0)$, which is a contradiction. Thus each primitive ideal is a maximal ideal. If B is a primitive algebra, then B is a simple algebra and thus is the only minimal dual ideal. Otherwise let J be the direct topological sum of the minimal dual ideals $\mathfrak{R}(P)$, where P is a primitive ideal. If J were proper, there would exist a maximal ideal N such that $N \supset J \supset \mathfrak{R}(N) \neq (0)$, which would contradict the semi-simplicity of B. Thus $J = B$.

7.17. LEMMA. *If M is a maximal ideal of B with $\mathfrak{R}(M) = (0)$, then $B = M \oplus Ku$, where K is the complex field. Moreover, M as an algebra satisfies condition* (7.2) *and* (7.3).

By Lemma 7.3 M has deficiency one and thus $B = M \oplus Ku$. Since M is a maximal ideal of B, M is a semi-simple Banach algebra and the center of M is contained in the center of B, so is zero. If $\mathfrak{U}$ is a proper closed Lie ideal of M, and $\mathfrak{U}$ contains $V = \{x \in M : [x, y] = 0, \forall y \in \mathfrak{U}\}$, then $\mathfrak{u} + Ku$ is a proper closed Lie ideal of B and $(\mathfrak{U} + Ku)^L = \mathfrak{U}^L = V + Ku$, so $(\mathfrak{U} + Ku) \supset (\mathfrak{U} + Ku)^L$. The hypothesis (7.2) yields $\mathfrak{U} + Ku = B$, which is a contradiction. Thus M as an algebra satisfies (7.2).

7.18. THEOREM. *Let B be a simple Banach algebra with satisfies* (7.2) *and* (7.4). *The only nonzero closed Lie ideals of B are either B and Ku or are B, Ku and $[B, B]$, in which case $B = [B, B] \oplus Ku$, and B is the direct sum of the minimal closed Lie ideals of B.*

Since B is simple, by a theorem of Herstein [3, p. 282], any closed Lie ideal is either contained in Ku or else contains $[B, B] = \mathfrak{L}$.

Suppose first that $Ku \subset \mathfrak{L}$. From Herstein's theorem applied to $\mathfrak{L}^L$, either (a) $\mathfrak{L}^L \subset Ku \subset \mathfrak{L}$ or (b) $\mathfrak{L} \subset \mathfrak{L}^L$. In case (a), the relation (7.2) yields $[B, B] = B$, so that the only closed Lie ideals are Ku and B. In case (b) $\mathfrak{L}^L \supset \mathfrak{L}^{LL}$ so that, by (7.2), $\mathfrak{L}^L = B$. Thus $\mathfrak{L} \subset \mathfrak{Z}$. In view of the reasoning for Theorem 6.2, $[B, B] = (0)$. Thus B is commutative and

so $B = Ku$. Hence the case $Ku \subset \mathfrak{L}$ leads to the first possibility stated in the theorem.

If $Ku \not\subset [B, B]$, let $\mathfrak{T} = [B, B] \oplus Ku$. Suppose $\mathfrak{T} \neq B$. Another application of Herstein's theorem yields $\mathfrak{T}^L \supset [B, B]$ or $\mathfrak{T}^L \subset Ku \subset \mathfrak{T}$. The latter is in contradiction to (7.2), while the former implies $\mathfrak{T}^L \supset \mathfrak{T}$ so $\mathfrak{T}^{LL} \subset \mathfrak{T}^L$. Thus $\mathfrak{T}^L = B$, which as above shows that B is the complex field. Hence $B = [B, B] + Ku$, in which case B is the direct sum of the minimal closed Lie ideals of B.

7.19. THEOREM. *Let B be a semi-simple Banach algebra which satisfies* (7.2) *and* (7.4). *Suppose that B is not simple, and is not commutative. Then B is the direct topological sum of its minimal closed Lie ideals and is either the direct topological sum of its minimal dual ideals, or the minimal dual ideals and the center.*

Suppose that $\mathfrak{R}(M) \neq (0)$ for each maximal ideal M of B. Then by Lemma 7.16 each primitive ideal of B is a maximal ideal and B is the direct topological sum of its minimal dual ideals. Also, as noted in the proof of Lemma 7.16, each minimal dual ideal has the form $\mathfrak{R}(M)$ for a maximal ideal.

Let M be any maximal ideal of B, and let $\mathfrak{J} \neq (0)$ be a closed Lie ideal, $\mathfrak{J} \subset \mathfrak{R}(M)$. Then, by a result of Herstein [3, p. 281], either $S(\mathfrak{J})$ is in the center of B or there is a nonzero ideal I in $S(\mathfrak{J})$, in which case we may assume I is the largest such ideal. The first possibility cannot occur since it would imply $\mathfrak{J} \subset Ku$, while $\mathfrak{J} \neq (0)$, and $\mathfrak{J} \subset \mathfrak{R}(M)$. Thus the second possibility holds. If $a \in I \cap M$, then for all $x \in B$, $[a, x] \in \mathfrak{J} \subset \mathfrak{R}(M)$, while $[a, x] \in M$, so $[a, x] = 0$ for all $x \in B$. Thus a is central and since $u \notin M$, we conclude $a = 0$. Consequently $I \cap M = (0)$ and $B = M \oplus I$ so that $I = \mathfrak{R}(M)$. By a result of Zuev [14],

$$\mathfrak{J} \supset [I, B] \supset [\mathfrak{R}(M), \mathfrak{R}(M)] .$$

Consequently each Lie ideal $[R(M), R(M)]$ is a minimal closed Lie ideal of B. Let $\mathfrak{U}$ be the direct topological sum of its minimal closed Lie ideals of B. Since B is the direct topological sum of the minimal dual ideals and these have the form $\mathfrak{R}(M)$, we see that $\mathfrak{U} \supset [B, B] + Ku$. Another application of the result of Herstein used above shows that either $S(\mathfrak{U}^L) \subset Ku$ or these exists a nonzero ideal N such that $N \subset S(\mathfrak{U}^L)$. The first possibility implies that $\mathfrak{U}^L = Ku$, whence $\mathfrak{U} \supset \mathfrak{U}^L$ so by (7.2) $\mathfrak{U} = B$. In the second instance, let $a \neq 0$, $a \in N$. Then for all $x \in B$, $[a, x] \in \mathfrak{U}^L \cap \mathfrak{U} = \mathfrak{W}$, from which it is immediate that $\mathfrak{W}^{LL} \subset \mathfrak{W}^L$. Thus by (7.2) $\mathfrak{W}^L = B$, so $\mathfrak{W} \subset Ku$, and by the argument used in Theorem 6.2 we deduce that $a \in Ku$, which implies that $N = S(\mathfrak{U}^L) = B$. We therefore conclude that $[B, B] \subset \mathfrak{U} \cap \mathfrak{U}^L \subset Ku$, so as argued earlier B

is commutative, which is a contradiction.

In the event that $\Re(M) = (0)$ for some maximal ideal M of B, it follows from Lemma 7.17 that $B = M \oplus Ku$ where K is the complex field, and M as an algebra satisfies (7.2) and (7.3). Then by Theorem 7.15 M is the direct topological sum of its minimal closed Lie ideals and by Corollary 3.3 is the direct topological sum of the minimal dual ideals of M_0. Any minimal closed Lie ideal of M is one of B, and any minimal dual ideal is one of B. Since Ku is a minimal closed Lie ideal of B, the theorem follows.

REFERENCES

1. R. Arens and I. Kaplansky, *Topological representation of algebras*, Trans. Amer. Math. Soc. **63** (1948), 457-481.
2. F. Bonsall and A. W. Goldie, *Annihilator algebras*, Proc. London Math. Soc. (3) **4** (1954), 154-167.
3. I. N. Herstein, *On the Lie and Jordan rings of a simple, associative ring*, Amer. J. Math. **77** (1955), 279-285.
4. ———, *Jordan homomorphisms*, Trans. Amer. Math. Soc. **81** (1956), 331-341.
5. ———, *Lie and Jordan structures in simple, associative rings*, Bull. Amer. Math. Soc. **67** (1961), 517-531.
6. N. Jacobson, *Structure of rings*, Amer. Math. Soc. Colloq. Publ. vol. **37**, 1956.
7. ———, *Lie algebras*, New York (1962).
8. R. V. Kadison, *Isometries of operator algebras*, Ann. of Math. **56** (1952), 494-503.
9. I. Kaplansky, *Semi-automorphisms of rings*, Duke Math. J. **14** (1947), 521-527.
10. C. E. Rickart, *General theory of Banach algebras*, Princeton (1960).
11. B. Yood, *Topological properties of homomorphisms between Banach algebras*, Amer, J. Math. **76** (1954), 155-167.
12. ———, *Homomorphisms on normed algebras*, Pacific J. Math. **8** (1958), 373-381.
13. ———, *Faithful *-representations of normed algebras*, Pacific J. Math. **10** (1960), 345-363.
14. I. I. Zuev, *Lie ideals of associative rings*, (Russian) Uspehi Mat. Nauk **18** (1963), no. 1 (109), 115-158.
15. E. Størmer, *Positive linear maps on operator algebras*, Acta. Math. **110** (1963), 233-278.

UNIVERSITY OF OREGON

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

DEDEKIND DOMAINS: OVERRINGS AND SEMI-PRIME ELEMENTS

LUTHER CLABORN

This paper develops two themes: (1) the relation of the class group of a Dedekind domain A to that of an overring B and (2) the question of finding a nonzero, nonunit element x of a Dedekind domain A such that A/xA is regular. We obtain complete results in answer to the first question, giving a corollary concerning the realization of certain groups as class groups. We give various sufficient conditions in answer to the second question; some in terms of the class group, others concerning Dedekind domains which often arise in practice.

In § 1 of the present paper, we study the class group of an overring B of a Dedekind domain A and determine its class group in terms of that of A. We generalize and also strengthen the results of § 1 of an earlier article [1]. Combining several results, we obtain an interesting fact: if G is the class group of a Dedekind domain and G' is a homorphic image of G, then G' is the class group of a suitable Dedekind domain.

Section 2 introduces the question of finding a nonunit x in a Dedekind domain A for which A/xA is a direct sum of fields. Although we obtain no definitive result, various sufficient conditions are given. These require in part the developments of § 1. We also give examples Dedekind domains with "pathological" class groups.

1. We state two well known propositions which we will need by way of background.

PROPOSITION 1.1. Let A be a Dedekind Domain with quotient field F. Let B be a ring such that $A \subseteq B \subset F$. Then $B = \cap A_p$ over those prime ideals P of A for which $B \subseteq A_p$.

PROPOSITION 1.2. Let A be a Dedekind domain with quotient field F. Let B be a ring such that $A \subseteq B \subset F$. Then B is a Dedekind domain.

PROPOSITION 1.3. Let A be a Dedekind domain with quotient field F and let B be a ring such that $A \subseteq B \subset F$. The assignment $I \rightarrow IB$ is a homomorphism of the set of fractionary ideals of A onto the set of fractionary ideals of B.

Received August 27, 1964.

Proof. Let Q be a prime ideal of B and set $P = Q \cap A$. Then $PB = Q$. The inclusion $PB \subseteq Q$ is trivial, while $B_Q = A_p$ implies that $PB_Q = (PB)B_Q = QB_Q$. This yields $PB = Q$ if we know that PB is not contained in another prime ideal Q' of B. But then Q' would also lie over P, which is not the case by Prop. 1.1.

If I is a fractionary ideal of A, then there is a $d \neq 0$ in A such that $dI \subseteq A$. But then clearly $d(IB) \subseteq B$, so IB is a fractionary ideal of B. The mapping is clearly a homomorphism for multiplication. To see that the mapping is onto, let Q be a prime ideal of B. We have seen above that if $P = Q \cap A$, then $PB = Q$. Thus the mapping is onto the prime ideals of B, and these generate the group of fractionary ideals of B.

COROLLARY 1.4. *Let A be a Dedekind domain with quotient field F and let B be a ring such that $A \subseteq B \subset F$. The assignment $I \to IB$ of fractionary ideals of A onto fractionary ideals of B induces a homomorphism $\psi : \bar{I} \to \overline{IB}$ of the class group of A onto that of B.*

Proof. It is sufficient to note that if $I = xA$, then $IB = xB$.

PROPOSITION 1.5. The kernel of ψ is generated by all $\bar{P}_\alpha$, where P_α ranges over all prime ideals such that $P_\alpha B = B$.

Proof. Suppose $P_\alpha B = B$, and let I be a fractionary ideal such that $\bar{I} = \bar{P}_\alpha$, i.e. $I = xP_\alpha$ for $x \in F$. Then $IB = xP_\alpha B = xB$, so $\overline{IB}$ is the identity.

Suppose now that I is a fractionary ideal of A such that $IB = yB$ for $y \in F$. Then $yI^{-1}B = B$, showing that yI^{-1} is a product of primes P_α of A for which $P_\alpha B = B$, and this establishes the assertion.

COROLLARY 1.6. *Let A be a Dedekind domain and $W = \{P_\alpha\}$ be a collection of primes such that $\{\bar{P}_\alpha\}$ does not generate the full class group of A. Then there are an infinite number of prime ideals of A not in the set $\{P_\alpha\}$.*

Proof. Let $B = \bigcap_{P \notin W} A_p$. By Proposition 1.5, B is not a principal ideal domain. Therefore there are an infinite number of prime ideals of B, hence an infinite number of prime ideals of A which are not in W.

COROLLARY 1.7. *Let A be a Dedekind domain with class group*

G. Let H be any subgroup of G. Then there is a Dedekind domain whose class group is G/H.

Proof. In [1], we constructed the Dedekind domain $A' = A[X]_S$, where S denotes the set of all monic polynomials of $A[X]$. We showed that A has the same class group as A [1, Prop. 2.3] and also that A has a prime ideal in every class of the class group [1, Cor. 2–5]. Identify G and H at the class group and a subgroup of the class group of A'. For each class of H, choose a prime P' of A' in the given class. Let W denote the set $\{P'\}$ so chosen. Then $B = \bigcap_{Q \in W} A_Q$ has class group G/H by Proposition 1.5.

2. DEFINITION 2.1. Let A be a Dedekind domain. An element x of A which is not zero and not a unit will be said to be semi-prime if A/xA is a regular ring.

REMARK 2.2. This condition is equivalent to (1) A/xA is a direct sum of fields, or (2) xA is not contained in the square of any prime ideal of A.

In what follows, sufficient conditions will be given for A to contain semi-prime elements. If A has only a finite number of prime ideals, then A is a principal domain and obviously A contains semi-prime elements. This case (A has only a finite number of prime ideals) will be excluded from the developments which follow.

PROPOSITION 2.3. If A has a finite class group, then there are semi-prime elements in A.

Proof. Since we are assuming that A has infinitely many prime ideals, there must be at least one class of the class group containing an infinite set $\{P_i\}$ of the prime ideals. If n is the class number of A, then $P_1 \cdots P_n$ must be principal, say $xA = P_1 \cdots P_n$. x is then a semi-prime element.

PROPOSITION 2.4. Let A be a Dedekind domain, and suppose that every class of the class group (except possibly the principal class) contains a prime ideal. Then A contains a semi-prime element.

Proof. If A is a principal ideal domain, then there is nothing to prove. Otherwise let P be a nonprincipal prime ideal and let Q be a prime in the class of P^{-1}. Then PQ is principal, say $PQ = xA$, and x will be semi-prime unless $P = Q$. We are therefore done unless every class has exponent 2 and there is only one prime in each class.

Choose P to represent one nonprincipal class and Q to represent a different nonprincipal class. Choose a prime ideal R in the class of PQ. Obviously $R \neq P$, $R \neq Q$, while PQR is principal. This gives a semi-prime element in A.

We can actually prove a little more.

PROPOSITION 2.5. Let A be a Dedekind domain, and suppose that for every prime ideal P there is a prime of A in the class of P^{-1}. Then A contains a semi-prime element.

Proof. As in the proof Proposition 2.4 we may assume that every class has exponent 2. The class group of A may therefore be regarded as a vector space over the field with 2 elements. Since the prime ideals of A generate the class group, we may choose a basis $\{\bar{P}_\alpha\}$ for the class group consisting of classes of prime ideals. Let P be any prime ideal of A and let $\bar{P}$ be its class. Let $\bar{P} = \bar{P}_{\alpha_1} \cdots \bar{P}_{\alpha_k}$ be its representation in terms of the given basis. Thus $PP_{\alpha_1} \cdots P_{\alpha_k}$ is principal and we get a semi-prime element unless P is in the set $\{P_\alpha\}$. We may assume then that the set $\{P_\alpha\}$ contains all prime ideals of A. But this contradicts Corollary 1.6, and the proposition is established.

Before giving an example violating the hypothesis of Proposition 2.5, we present a lemma which will be useful in constructing such an example and in a later proof.

LEMMA 2.6. *Let F be a field of characteristic p such that $[F^{1/p} : F] = p$. Let K be a separable extension of F; then $[K^{1/p} : K] = p$.*

Proof. Since K is a separable extension of F, we have $K = F(K^p)$ [3, Thm. 8, p. 69]. Thus $K^{1/p} = F^{1/p}(K)$. But $F^{1/p}$ and K are linearly disjoint [3, Thm. 35, p. 111], so we get $[K^{1/p} : K] = [F^{1/p}(K) : K] = [F^{1/p} : F] = p$.

EXAMPLE 2.7. Let $F' = Z/3Z(a)$ where Z denotes the integers and a is indeterminant. Let F be the separable closure of F' in its algebraic closure. By Lemma 2.6, $[F^{1/3} : F] = 3$. Consider the integral closure A of $F[X]$ in the field $F(X, Y)$, where $Y^3 = aX^3 + X$. It is not difficult to show by a direct computation that $A = F[X, Y]$, but it is easier to notice that since the matrix of partial derivatives of the equation $Y^3 - aX^3 - X$ has always rank 1, $F[X, Y]$ is regular [2, Thm. 1. p. 201]. Over each prime ideal of $F[X]$ there lies only one prime ideal of A and for the relative degree f of the residue field and the ramification index e we have $e = 3$, $f = 1$ or $e = 1$, $f = 3$ [3,

Thm. 22, p. 289]. We show first that for all nonlinear prime elements of $F[X]$, we get $e = 1$, $f = 3$, so these remain principal. Let Q be a prime ideal of $F[X]$ generated by a nonlinear element; $Q = X^q - t$, where q is a power of 3 and $t \in F$. The residue field $F[X]/(X^q - t)$ is $F[t^{1/q}]$, while the residue field relative to A will be $F[t^{1/q}, w]$, where $w^3 = at^{3/q} + t^{1/q}$. Since $[F^{1/3} : F] = 3$, we have $F^{1/3} \subseteq F[t^{1/q}]$, hence $a^{1/3} \in F[t^{1/q}]$. Thus $at^{3/q}$ is a cube in $F[t^{1/q}]$. But $t^{1/q}$ is not a cube in $F[t^{1/q}]$, so $[F[t^{1/q}, w] : F[t^{1/q}]] = 3$. That is, $f = 3$, $e = 1$; thus we see that nonlinear prime elements of $F[X]$ remain prime in A.

For the linear primes $X - t$, $t \varepsilon F$, we get $e = 1$, $f = 3$ if $at^3 + t$ is not a cube in F, while $e = 3$, $f = 1$ if $at^3 + t$ is a cube in F. Certainly we have the latter case at least for $t = 0$. Let P be a prime ideal of A lying over a linear prime ideal of $F[X]$ for which $e = 3$, $f = 1$. Then P is not principal. For if P were principal, say $P = (c_0(X) + c_1(X)Y + c_2(X)Y^2)$ we would get

$$P^3 = (c_0^3(X) + c_1^3(X)(aX^3 + X) + c_2^3(X)(a^2X^6 + 2aX^4 + X^2) .$$

But $P^3 = (x - t)$ for some $t \in F$. Comparing degrees and using the fact that $1, a, a^2$ are independent over F^3, we get a contradiction. Again let P be such a prime and suppose that the class of P^2 (which is the clsss of P^{-1}) contains a prime Q. Q is certainly not principal; therefore Q lies over a linear prime ideal of $F[X]$ and $e = 3$, $f = 1$ for Q. We also get that P^2Q^2 is principal, say

$$P^2Q^2 = (d_0(X) + d_1(X)Y + d_2(X)Y^2) .$$

Cubing, we get

$$(P^3)^2(Q^3)^2 = (d_0^3(X) + d_1^3(X)(aX^3 + X) + d_2^3(X)(a^2X^6 + 2aX^4 + X^2)).$$

On the left side of this equation we have a polynomial of degree 4, while on the right we have a polynomial whose degree is divisible by 3, a contradiction.

PROPOSITION 2.8. Let A be a principal ideal domain and let K be a finite separable extension of the quotient field F of A. Let B be the integral closure of A in K and let C be a ring such that $B \subseteq C \subset K$. Then C contains a semi-prime element.[1]

Proof. There are only a finite number of prime ideals $Q_1, \cdots, Q_k$ of B whose reduced ramification index is greater than 1 [3, Thm. 28, p. 302]. Let $P = \pi A$ be a prime ideal of A not lying under any $Q_1, \cdots, Q_k$. Then πB is a product of distinct primes and is a semi-

[1] The referee has kindly pointed out that this Proposition (and thus the following) hold when B is not necessarily integrally closed.

prime element in B. π will also be a semi-prime element in C unless all prime ideals of B dividing π generate C. The result now follows by Proposition 1.5 and Corollary 1.6.

PROPOSITION 2.9. Let A be the coordinate ring of an algebraic curve over a perfect ground field F. If A is a Dedekind domain, then A contains a semi-prime element.

Proof. $A = F[x_1, \cdots, x_n]$. Choose X in A such that A is integral over $F[X]$; this is possible by [2, Thm. 1, p. 22]. Since A is integrally closed in $K = F(x_1, \cdots, x_n)$, A is the integral closure of $F[X]$ in K. Let K' be the separable closure of $F(X)$ in K, and let A' be the integral closure of $F[X]$ in K. The conclusion holds for A' by Proposition 2.8.

Since $[F(X)^{1/p} : F(X)] = p$, we have $[K'^{1/p} : K'] = p$ by Lemma 2.6. K is a purely inseparable extension of K', so we may break the extension from K' to K into a chain of extensions each of which is pure inseparable of exponent p. This chain can only be

$$K' = K_0 \subset K_0^{1/p} \subset K_0^{1/p^2} \subset \cdots \subset K_0^{1/p^m} = K .$$

But then we have an isomorphism of K onto K' given by $x \to x^{p^m}$ which induces an isomorphim of A onto A'. Since A' contained semi-prime elements, so does A.

REFERENCES

1. L. Claborn, *Dedekind domains and rings of quotients*, Pacific J. Math. **15** (1965), 59-64.

2. S. Lang, *Introduction to Algebraic Geometry*, New York, Interscience Publishers, Inc. (1958).

3. O. Zariski and P. Samuel, *Commutative Algebra*, Vol. I, Princeton, D. Van Nostrand Company (1958).

CORNELL COLLEGE

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

NOTE GENERALIZING A RESULT OF SAMUEL'S

LUTHER CLABORN

Let $C(A)$ denote the class group of a Krull domain A. Samuel has established (a) $C(A) \to C(A[[X_1, \cdots, X_n]])$ is injective, and (b) $C(A) \to C(A[[X_1, \cdots, X_n]])$ is bijective in Case A is a regular U.F.D. This note establishes that $C(A) \to C(A[[X_1, \cdots, X_n]])$ is bijective in Case A is a regular noetherian domain, thus adding a complement to (a) while generalizing (b). A corollary of this is that $A[[X_1, \cdots, X_n]]_S$ is a U.F.D. if A is a regular Noetherian domain and S is the set of nonzero elements of A.

In [1], Samuel gave an example of a nonregular noetherian U.F.D. A such that $A[[X]]$ is not a U.F.D. In this case certainly, the mapping of the class group $C(A)$ into $C(A[[X]])$ is not onto (since a unique factorization domain is characterized by $C(A) = 0$). In the same article, Samuel showed that $A[[X]]$ is a U.F.D. in Case A is a regular U.F.D. Here it is proved that $C(A) \to C(A[[X_1, \cdots, X_n]])$ is one-to-one onto in Case A is a regular noetherian domain. The main tool is the technical Theorem 3 below, which shows that if W is unmixed of height 1 in $A[[X_1, \cdots, X_n]]$, then there is an unmixed height 1 ideal I of A such that IW is principal. From this the result stated above follows directly.

Two lemmas are needed to facilitate the main results.

LEMMA 1. *Let B be a regular noetherian domain and let I and W be two unmixed height 1 ideals of B such that I and W have no associated prime ideals in common. Then $IW = I \cap W$.*

Proof. For each M, B_M is a regular local Noetherian ring, hence a U.F.D., so IB_M and WB_M are both principal ideals. Since I and W have no associated prime ideals in common, neither do IB_M and WB_M. It follows that $IB_M \cap WB_M = IB_M \cdot WB_M$. Thus for each M, $(I \cap W)B_M = IB_M \cap WB_M = IB_M \cdot WB_M = (IW)B_M$. This establishes the lemma.

LEMMA 2. *Let B be a regular noetherian domain and Z be an unmixed height 1 ideal of B. If X is an element of B such that (a) X is in the Jacobson radical of B and (b) $Z : XB = Z$, then $Z + XB$ is unmixed of height 2.*

Proof. Let P be an associated prime ideal of $Z + XB$. Then B_P is a regular local ring, hence is a U.F.D. [3, Thm., p. 406]. ZB_P is

Received July 17, 1964.

principal, so choose T in Z such that $ZB_P = TB_P$. Then PB_P is an associated prime ideal of $(Z + XB)B_P = ZB_P + XB_P = TB_P + XB_P$. But $Z : XB = Z$ implies that $ZB_P : XB_P = ZB_P$, or $TB_P : XB_P = TB_P$, so $\{T, X\}$ is a prime sequence in B_P. This implies that height of $PB_P = 2$ [3, Thm. 2, p. 397]. But height of $P =$ height of PB_P.

THEOREM 3. *Let A be a regular noetherian domain and let $B_n = A[[X_1, \cdots, X_n]]$. If W is any unmixed height 1 ideal of B_n, then there is an unmixed height 1 ideal I of A such that IW is principal and $IW = IB_n \cap W$.*

Proof. Let $B_0 = A$. The theorem will be proved for $n \geqq 0$ by induction on n.

(1) $n = 0$. W is an unmixed height 1 ideal of A. Since A is regular, it is integrally closed, so $W = P_1^{(n_1)} \cap \cdots \cap P_k^{(n_k)}$ where the P_i $i = 1, \cdots, k$ are height 1 prime ideals and $P_j^{(n_i)}$ $i = 1, \cdots, k$ is the n_ith symbolic power of P_i. Choose d an element of A so that $V_{P_i}(d) = n_i$ $i = 1, \cdots, k$ where V_{P_i} denotes the discrete valuation determined by P_i. Then dA can be written

$$dA = P_1^{(n_1)} \cap \cdots \cap P_k^{(n_k)} \cap P_{k+1}^{(n_{k+1})} \cap \cdots \cap P_l^{(n_l)}$$

where the P_{k+j} $j = 1, \cdots, l - k$ are further height 1 prime ideals of A. Let $I = P_{k+1}^{(n_{k+1})} \cap \cdots \cap P_l^{(n_l)}$. Then visibly $I \cap W = dA$ is principal. By Lemma 1, $I \cap W = IW$.

(2) Suppose the theorem has been established for $n - 1 (n \geqq 1)$. Let W be an unmixed height 1 ideal of B_n and write $W = ZX_n^k$ where $Z : X_nB_n = Z$. If $Z = B_n$, the theorem follows trivially. If $Z \neq B_n$, then Z is also unmixed of height 1. Thus $Z + X_nB_n$ is unmixed of height 2 by Lemma 2. Let $Z_0 = Z + X_nB_n/X_nB_n$. Z_0 is unmixed of height 1 in B_{n-1}. By induction, there is an ideal I of A such that $IZ_0 = IB_{n-1} \cap Z_0$ is principal, say $IZ_0 = u_0(X_1, \cdots, X_{n-1}) \cdot B_{n-1}$. Choose an element $u(X_1, \cdots, X_{n-1}, X_n)$ in IZ whose leading coefficient when written as a power series in X_n is $u_0(X_1, \cdots, X_{n-1})$.

Let $f(X_1, \cdots, X_n)$ be any element of $IB_n \cap Z$. Then $f(X_1, \cdots, X_{n-1}, 0)$ is in $IB_{n-1} \cap Z_0$. This implies that $f - g_0u = X_n \cdot f_1$, where f_1 is in B_n and g_0 is in B_{n-1}. Since f and u are both in Z, f_1 is in $Z : X_nB_n = Z$. Clearly f_1 is in IB_n. So repeating, we can find an f_2 in B_n and a g_1 in B_{n-1} such that $f_1 - g_1u = X_n \cdot f_2$. Continuing, we get that

$$f = u \cdot \sum_{i=0}^{\infty} g_i X_n^i$$

showing simultaneously that $IB_n \cap Z$ is principal and that $IB_n \cap Z = IZ$.

To conclude, $IW = IZX_n^k B_n = u \cdot X_n^k B_n$ is principal. If v is in $IB_n \cap W$, from v in W it follows that $v = X_n^k \cdot v'$, where v' is in Z. But then v' is in IB_n so v' is in $IB_n \cap Z = IZ$. This gives that $v = X_n^k v'$ is in $IZX_n^k = IW$, showing that $IB_n \cap W \subseteq IW$. The opposite inclusion is trivial, so the induction is complete.

COROLLARY 4. *The map* $C(A) \to C(A[[X_1, \cdots, X_n]])$ *of the class group of* A *into the class group of* $A[[X_1, \cdots, X_n]]$ *is one-to-one onto if* A *is a regular noetherian domain.*

Proof. Samuel [2, Prop. 1, p. 156 and Prop. 3, p. 138] has shown that the map is one-to-one. Theorem 3 proves that it is onto in the present case.

COROLLARY 5. *Let* A *be a regular noetherian domain. Let* M *be a multiplicative set of* $A[[X_1, \cdots, X_n]]$. *Then* $C(A[[X_1, \cdots, X_n]]_M)$ *is a homomorphic image of* $C(A)$.

Proof. Samuel [2, Prop. 2, p. 157] shows that $C(R) \to C(R_S)$ is always onto. Corollary 4 supplies the rest.

COROLLARY 6. *Let* A *be a regular noetherian domain. Let* S *denote the nonzero elements of* A. *Then* $B' = A[[X_1, \cdots, X_n]]_S$ *is a U.F.D.*

Proof. Let W be an unmixed height 1 ideal of $A[[X_1, \cdots, X_n]]$. Then there is an unmixed height 1 ideal I of A such that IW is principal, say $IW = (U)$. Then $UB' = IB' \cdot WB'$, but $IB' = B'$, so $WB' = UB'$ is principal.

REMARKS. (1) Samuel [2] has established the analogue of Corolary 4 for $A[X_1, \cdots, X_n]$. This implies that Corollaries 5 and 6 also hold for $A[X_1, \cdots, X_n]$, Corollary 6 of course being trivial.

(2) As originally submitted, this note established Theorem 3 and its corollaries only in the case that A is a Dedekind domain. In the original presentation, Corollary 6 was the main tool for the proofs of Corollaries 4 and 5. I wish to express my gratitude to the referee for bringing Samuel's results [2] to my attention and for suggesting the generalization to regular Noetherian domains. Lemma 1 is the only addition necessary to effect the generalization.

REFERENCES

1. P. Samuel, *On unique factorization domains*, Illinois J. **5** (1961), 1-17.
2. ———, *Sur les anneaux factoriels*, Bull. Soc. Math. France, **89** (1961), 155-173.
3. O. Zariski and P. Samuel, *Commutative Algebra*, Vol. II, Princeton, D. Van Nostrand Company (1960).

CORNELL COLLEGE

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

SYMMETRIC DUAL NONLINEAR PROGRAMS

G. B. DANTZIG, E. EISENBERG*, and R. W. COTTLE

Consider a function $K(x, y)$ continuously differentiable in $x \in R^n$ and $y \in R^m$. We form two problems:

PRIMAL: **Find $(x, y) \geqq 0$ and Min F such that**

$$F = K(x, y) - y^T D_y K(x, y),\ D_y K(x, y) \leqq 0$$

DUAL: **Find $(x, y) \geqq 0$ and Max G such that**

$$G = K(x, y) - x^T D_x K(x, y),\ D_x K(x, y) \geqq 0$$

where $D_y K(x, y)$ and $D_x K(x, y)$ denote the vectors of partial derivatives $D_{y_i} K(x, y)$ and $D_{x_j} K(x, y)$ for $i = 1, \cdots, m$ and $j = 1, \cdots, n$. Our main result is the existence of a common extremal solution (x_0, y_0) to both the primal and dual systems when (i) an extremal solution (x_0, y_0) to the primal exists, (ii) K is convex in x for each y, concave in y for each x and (iii) K, twice differentiable, has the property at (x_0, y_0) that its matrix of second partials with respect to y is negative definite.

F and G are related to the conjugate functions considered by Fenchel [6] (see also [9]) and to the Legendre transforms of K.

Special Cases.

A. If we set $K(x, y) = c^T x + b^T y - y^T A x$, we obtain von Neumann's symmetric formulation of primal and dual linear programs: see [3] or [9].

PRIMAL: Find $x \geqq 0$, Min F such that $F = c^T x,\ Ax \geqq b$

DUAL: Find $y \geqq 0$, Max G such that $G = b^T y,\ A^T y \leqq c$.

B. Symmetric dual quadratic programs [2] can be obtained by setting

$$K(x, y) = c^T x + b^T y - y^T A x + \frac{1}{2}(x^T D x - y^T E y).$$

The dual quadratic programs of Dorn [4] result when $E = 0$.

C. If we set $K(x, y) = F_0(x) - \sum_{i=1}^{m} y_i F_i(x)$, we obtain the nonlinear statements of duality by Wolfe [12] and Huard [7]. For y not sign restricted, the problems may be simplified to

Received January 23, 1964. Research supported by Office of Naval Research Contract Nonr-222 (83).

* Died January 28, 1965.

PRIMAL: Find $x \geqq 0$ and Min $F_0(x)$ such that
$F_i(x) = 0,\ i = 1, \cdots, m$.

DUAL: Find $x \geqq 0$, y and Max $K(x, y)$ such that
$D_{x_j}K(x, y) \geqq 0,\ j = 1, \cdots, n$.

A "Weak" Duality Theorem.

With additional conditions on $K(x, y)$ we can state a relation between solutions of the primal and dual systems.

THEOREM 1. *If $K(x, y)$ is convex in x, for each y, and concave in y, for each x, then any (not necessarily extremal) solutions (x, y) of the primal and (u, v) of the dual satisfy the inequality*

$$F(x, y) \geqq G(u, v)$$

where equality holds only if (x, v) is orthogonal to

$$[D_u K(u, v), -D_y K(x, y)] .$$

Proof. By the assumptions of convexity, concavity, and differentiability

$$K(x, v) - K(u, v) \geqq (x - u)^T D_u K(u, v)$$
$$K(x, v) - K(x, y) \leqq (v - y)^T D_y K(x, y) .$$

Subtracting and rearranging, we get

$$[K(x, y) - y^T D_y K(x, y)] - [K(u, v) - u^T D_u K(u, v)]$$
$$\geqq x^T D_u K(u, v) - v^T D_y K(x, y) \geqq 0 .$$

REMARK. The significance of Theorem 1 is apparent when in addition it is also true that $F(x^0, y^0) = G(u^0, v^0)$ for some (x^0, y^0) and (u^0, v^0); in that case, (x^0, y^0) and (u^0, v^0) are extremal solutions of the primal and dual, respectively. This is illustrated in the theorem below.

A Strong Duality Theorem.

Under certain assumptions, we can show that if one of the systems can be extremized, the two systems possess a solution in common. (References on duality, in addition to those mentioned above, are [11], [5], [1].)

THEOREM 2. *If (x^0, y^0) is an extremal solution for the primal where $K(x, y)$ is twice differentiable and $-D_{yy}K(x^0, y^0)$ is positive definite, then (x^0, y^0) satisfies the dual constraints with $F(x^0, y^0) = G(x^0, y^0)$. If, in addition, $K(x, y)$ is convex in x, for each y, and concave in y, for each x, then*

$$\text{Min } F(x, y) = F(x^0, y^0) = G(x^0, y^0) = \text{Max } G(x, y) .$$

Proof. By a result of John [8, Theorem 1], see also Kuhn and Tucker [10], there exist "multipliers" $v_0 \geqq 0$ (scalar) and $v \geqq 0$ (m-vector) such that $(v_0, v) \neq 0$, $v^T D_y(x^0, y^0) = 0$, and

$$\begin{aligned}
&v_0 D_y F(x^0, y^0) + D_{yy} K(x^0, y^0) v \geqq 0 \\
&(y^0)^T [v^0 D_y F(x^0, y^0) + D_{yy} K(x^0, y^{0)} v] = 0 \\
&v_0 D_x F(x^0, y^0) + D_{xy} K(x^0, y^0) v \geqq 0 \\
&(x^0)^T [v_0 D_x F(x^0, y^0) + D_{xy} K(x^0, y^0) v] = 0 .
\end{aligned}$$

Replacing F by its definition, $K(x, y) - y^T D_y K(x, y)$, yields

$$(1) \qquad D_{yy} K(x^0, y^0)(v - v_0 y^0) \geqq 0$$

$$(2) \qquad (y^0)^T D_{yy} K(x^0, y^0)(v - v_0 y^0) = 0$$

$$(3) \qquad v_0 D_x K(x^0, y^0) + D_{xy} K(x^0, y^0)(v - v_0 y^0) \geqq 0$$

$$(4) \qquad (v_0 x^0)^T D_x K(x^0, y^0) + (x^0)^T D_{xy} K(x^0, y^0)(v - v_0 y^0) = 0 .$$

Multiplying (1) by $v^T \geqq 0$, (2) by v_0, and then subtracting, we obtain

$$(v - v_0 y^0)^T D_{yy} K(x^0, y^0)(v - v_0 y^0) \geqq 0 .$$

But since the matrix $- D_{yy} K(x^0, y^0)$ is positive definite, we have $v - v_0 y^0 = 0$. Relations (3) and (4) yield

$$(5) \qquad v_0 D_x K(x^0, y^0) \geqq 0 ,$$

$$(6) \qquad (v_0 x^0)^T D_x K(x^0, y^0) = 0 .$$

Moreover, since $v^T D_y K(x^0, y^0) = 0$, we have

$$(7) \qquad (v_0 y^0)^T D_y K(x^0, y^0) = 0 .$$

But $v_0 > 0$ for otherwise $v = v_0 y^0 = 0$ and $(v_0, v) = 0$, a contradiction. Dividing (5), (6), and (7) by the positive scalar v_0 produces relations which imply that (x^0, y^0) satisfies the dual constraints as well as the equation $F(x^0, y^0) = G(x^0, y^0)$. The final assertion of the theorem is an application of Theorem 1.

REFERENCES

1. A. Charnes, W. W. Cooper and K. Kortanek, *Duality in semi-infinite programs and some works of Haar and Caratheodory*, Management Sci., **10** (1963), 209-228.

2. R. W. Cottle, *Symmetric dual quadratic programs*, Quart. Appl. Math. **21** (1963), 237-243.

3. G. B. Dantzig, *Linear programming and extensions*, Princeton University Press, Princeton University Press, Princeton, New Jersey, 1963.

4. W. S. Dorn, *Duality in quadratic programming*, Quart. Appl. Math. **18** (1960), 155-162.
5. E. Eisenberg, *Duality in homogeneous programming*, Proc. Amer. Math. Soc. **12** (1961), 783-787.
6. W. Fenchel, *Convex cones, sets, and functions*, Lecture Notes, Princeton University, 1953.
7. P. Huard, *Dual programs*, IBM J. Res. Develop. **6** (1962), 137-139.
8. F. John, *Extremum problems with inequalities as subsidiary conditions*, in Studies and Essays, Courant Anniversary Volume, Interscience, New York, 1948.
9. S. Karlin, *Mathematical methods and theory in games, programming and economics*, Vol. I, Addison Wesley, Reading, Massachusetts, 1959.
10. H. W. Kuhn and A. W. Tucker, *Nonlinear programming*, in Proceedings, Second Berkeley Symposium on Mathematical Statistics and Probability, University of California, Berkeley, 1951, 481-492.
11. O. L. Mangasarian, *Duality in nonlinear programming*, Quart. Appl. Math. **20** (1962), 300-302.
12. P. Wolfe, *A duality theorem for nonlinear programming*, Quart. Appl. Math. **19** (1961), 239-244.

UNIVERSITY OF CALIFORNIA, BERKELEY
AND
BELL TELEPHONE LABORATORIES, INC.
HOLMDEL, NEW JERSEY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

SIMPLE QUADRATURES IN THE COMPLEX PLANE

PHILIP J. DAVIS

Given a class S of functions that are Riemann integrable on $[0, 1]$. A quadrature formula $\int_0^1 f(x)dx = \sum_{i=1}^{\infty} a_i f(x_i)$ is called a *simple quadrature for* S if the x_i are distinct and if both the a_i and the x_i are fixed and independent of the particular function of S selected. It is known that if S is too large, for example if $S = C[0, 1]$, a simple quadrature cannot exist. On the other hand, if S is sufficiently restricted, for example the class of all polynomials, then simple quadratures exist.

The present paper investigates further the existence of simple quadratures. It is proved among other things that if S is the class of analytic functions that are regular in the closure of an ellipse with foci at ± 1, a simple quadrature exists for the weighted integral $\int_{-1}^{+1} (1-x^2)^{1/2} f(x)dx$ provided we allow the abscissas x_i to take on complex values.

1. Simple Quadratures. In [3], the author studied the following question. Suppose that there has been given a fairly extensive class S of real functions that are Riemann integrable on $[0, 1]$. Does there exist a quadrature formula of the form

$$\int_0^1 f(x)dx = \sum_{i=1}^{\infty} a_i f(x_i) \tag{1}$$

which is valid for all functions of the class S? The abscissas x_i are assumed distinct, and both the weights a_i and the abscissas x_i are fixed and independent of the particular function of S selected. A quadrature of the form (1) was called a *simple* quadrature to contrast it with quadratures of the form

$$\int_0^1 f(x)dx = \lim_{n\to\infty} \sum_{i=1}^{n} a_{in} f(x_{in}) \tag{2}$$

which allow more freedom than (1) and have accordingly been more frequently investigated. See, e.g., Szegö [9], Chap. 15.

In [3], we found, broadly speaking, that if S is fairly small, simple quadratures exist, while if S has too many functions in it, simple quadratures do not exist. Thus, for instance, there exists a simple quadratures for the class of all polynomials (of unbounded degree), while there cannot exist a simple quadrature for the class of continuous functions. See also Davis [7], Chap. 14, where this question is treated in the framework of weak* convergence.

Received April 25, 1964. This work was supported by the Office of Naval Research Contract Nonr 562 (36).

An interesting intermediate class of functions that possesses a simple quadrature is the class B of entire functions that are finite Fourier Transforms i.e., the class of functions $f(z)$ that have a representation of the form

$$(3) \qquad f(z) = \int_{-\sigma}^{\sigma} e^{izt}\varphi(t)dt, \ \varphi(t)\varepsilon L[-\sigma, \sigma], \ \sigma \text{ fixed.}$$

In the communication literature, B is known as the class of *bandlimited* functions. It is known, see, e.g. Boas [1] p. 220, that an $f(z)$ in B can be represented by means of the cardinal series

$$(4) \qquad f(z) = \sum_{n=-\infty}^{\infty} (-1)^n f\left(\frac{n\pi}{\sigma}\right) \frac{\sin \sigma z}{\sigma z - n\pi}$$

which is uniformly convergent in any bounded set in the complex z-plane. Hence, integrating term by term, we find

$$(5) \qquad \int_0^1 f(x)dx = \sum_{n=-\infty}^{\infty} a_n f(x_n)$$

where

$$(6) \qquad x_n = \frac{n\pi}{\sigma} \text{ and } a_n = (-1)_u \int_0^1 \frac{\sin \sigma x}{\sigma x - n\pi} dx.$$

A similar argument shows that "simple" representations can be found for B for all the usual linear functionals of numerical analysis such as derivatives, etc.

Interesting necessary and sufficient conditions for S to allow a simple quadrature have not yet been determined. In the present paper, therefore, we shall investigate these matters further, considering certain integrals of analytic functions and allowing the abscissas to take on complex values. With this freedom, we shall show that simple quadratures exist for double integrals of analytic functions, for certain single integrals of analytic functions, and for certain inner product integrals associated with Hilbert Spaces of analytic functions.

2. Simple quadratures with complex abscissas. Our work is based upon a device whose use in complex variable theory goes back at least as far as Pompeiu [8], p. 165.

LEMMA. *Let B be a bounded region in the complex plane. Then we can find a sequence of circles C_n: $|z - z_n| < r_n$, $n = 1, 2, \cdots$ such that*

(1) *each C_n is contained in B*

(2) *the circles C_n are nonoverlapping*

(3) *the circles C_n exhaust the area of B:*

$$\sum_{n=1}^{\infty} \pi r_n^2 = \text{area of } B. \tag{7}$$

Another way of putting (7) is that

$$\text{meas}\left(B - \bigcup_{n=1}^{\infty} C_n\right) = 0. \tag{7'}$$

Such a sequence of circles will be called a *complete packing of B.*

Proofs of this lemma are easily given. The reader is referred to the recent paper of O. Wesler [12] where a proof will be found. Wesler also gives a simple geometrical proof, utilizing the Borel-Cantelli lemma, of the interesting theorem, apparently due to S. Mergelyan, that the sum of the circumferences of the circles in a complete packing is infinite: $\sum_{n=1}^{\infty} r_n = \infty$.

The existence of complete packings is not restricted to packings by circles or to two dimensions.

LEMMA. *Let C designate the circle $|z - z_0| < r$. If $f(z)$ is regular in $\bar{C}$, then*

$$\iint_{\sigma} f(z)\,dxdy = \pi r^2 f(z_0). \tag{8}$$

In this lemma, as in what follows, we have not attempted to put the most general conditions on $f(z)$. The above conditions are sufficient.

THEOREM. *Let B be a bounded region of the complex plane and let the circles C_n: $|z - z_n| < r_n$ $(n = 1, 2, \cdots)$ constitute a complete packing of B. Then, if $f(z)$ is regular in B and continuous in $\bar{B}$,*

$$\iint_{B} f(z)dxdy = \pi \sum_{n=1}^{\infty} r_n^2 f(z_n). \tag{9}$$

Proof. In view of (7'),

$$\iint_{B} f(z)dxdy = \sum_{n=1}^{\infty} \iint_{\sigma_n} f(z)dxdy. \tag{10}$$

An application of (8) now yields (9).

Notice that (9) gives us a simple quadrature that is valid for double integrals of analytic functions. There are infinitely many such simple quadrature formulas, each one corresponding to a different complete packing of the region B. Note also that in view of (7) and the boundedness of the sequence $\{f(z_n)\}$, the series in (9) converges absolutely. But in view of the divergence of $\sum_{n=1}^{\infty} r_n$, the convergence

in (9) cannot be too rapid, and serious numerical applications are more than likely excluded.

The special selection of B as the unit circle leads to the following corollary.

COROLLARY. *Let* $C_n: |z - z_n| < r_n$ $(n = 1, 2, \cdots)$ *be a complete packing of the unit circle. Then, if* $f(z)$ *is regular in* C *and continuous in* $\bar{C}$,

$$f(0) = \sum_{n=1}^{\infty} r_n^2 f(z_n). \tag{11}$$

Formula (11) may be regarded as a "simple" representation of the functional $L(f) = f(0)$ in terms of point functionals.

If we select $f(z) = 1/(w - z)$, then (10) leads to

$$\frac{1}{w} = \sum_{n=1}^{\infty} \frac{r_n^2}{w - z_n}. \tag{12}$$

This is Wolff's example of a convergent series of fractions that can be continued analytically into the region of the apparent poles! (Wolff [13]). Choosing $f(z) = ze^{zw}$ leads to $0 = \sum_{k=1}^{\infty} r_k^2 z_k e^{z_k w}$, which is an absolutely convergent, nontrivial representation of 0 as a sum of exponentials. This should be contrasted with the more familiar Dirichlet series where this sort of behavior does not occur.

There is an amusing consequence of this method. As anyone knows (who has played with his change on a bar), it is not possible to find a complete packing of a circle by a finite number of circles other than the trivial identity packing. Proofs are easy to devise, but from our present point of view, this fact emerges as a consequence of interpolation by polynomials in the complex plane! For, suppose that C_k: $|z - z_k| < r_k$, $k = 1, 2, \cdots, n$ were a complete packing of the unit circle. Then, as in (11),

$$f(0) = \sum_{k=1}^{n} r_k^2 f(z_k), \tag{13}$$

and this must certainly be true for all polynomials f. If $z_1, \cdots, z_n$ are distinct from 0, let $f(z)$ be that polynomial which vanishes at $z_1, \cdots, z_n$ and for which $f(0) = 1$. This leads to the contradiction $1 = 0$. If one of the z_n, say z_1, is 0 let $f(z)$ be that polynomial that vanishes at $z_2, \cdots, z_n$, and for which $f(0) = 1$. Then, $r_1^2 = 1$, the trivial case.

Formula (9) can be generalized as follows: Let $C_k: |z - z_k| < r_k$ be a complete packing of B. Then, if $f(z)$ and $g(z)$ are regular in B

and continuous in $\bar{B}$,

$$\frac{1}{\pi}\iint_B f(z)\overline{g(z)}\,dxdy = \sum_{k=1}^{\infty} r_k^2 f(z_k)\overline{g(z_k)} + \frac{1}{2}\sum_{k=1}^{\infty} r_k^4 f'(z_k)\overline{g'(z_k)} + \frac{1}{3(2!)^2}\sum_{k=1}^{\infty} r_k^6 f''(z_k)\overline{g''(z_k)} + \frac{1}{4(3!)^2}\sum_{k=1}^{\infty} r_k^8 f'''(z_k)\overline{g'''(z_k)} + \cdots \tag{14}$$

Proof. We have,

$$f(z) = \sum_{m=0}^{\infty} \frac{f^{(m)}(z_k)\,(z - z_k)^m}{m!}, \qquad \overline{g(z)} = \sum_{m=0}^{\infty} \frac{\overline{g^{(m)}(z_k)}\,\overline{(z - z_k)^m}}{m!},$$

uniformly and absolutely convergent in the closure of C_k. Now, in view of the fact that

$$\iint_{\sigma_k} (z - z_k)^m \overline{(z - z_k)^n}\,dxdy = \frac{\pi}{m+1} r_k^{2m+2}\delta_{mn},$$

we obtain from term by term integration,

$$\frac{1}{\pi}\iint_{\sigma_k} f(z)\overline{g(z)}\,dxdy = r_k^2 f(z_k)\overline{g(z_k)} + \frac{1}{2} r_k^4 f'(z_k)\overline{g'(z_k)} + \frac{r_k^6}{3(2!)^2} f''(z_k)\overline{g''(z_k)} + \frac{r_k^8}{4(3!)^2} f'''(z_k)\overline{g'''(z_k)} + \cdots$$

3. **The use of triangles.** Another type of "simple" interpolation formula can be obtained by employing a triangle.

THEOREM. *Let T be a triangle whose vertices are at w_1, w_2, w_3 and whose area is A. Let C_n: $|z - z_n| < r_n$ be a complete packing of T. Then, if $f''(z)$ is regular in $\bar{T}$,*

$$\frac{f(w_1)}{(w_1 - w_2)(w_1 - w_3)} + \frac{f(w_2)}{(w_2 - w_1)(w_2 - w_3)} + \frac{f(w_3)}{(w_3 - w_1)(w_3 - w_2)} = \frac{\pi}{2A}\sum_{k=1}^{\infty} r_k^2 f''(z_k)\,. \tag{15}$$

Proof. Under the above conditions on f, we have the identity

$$\frac{1}{2A}\iint_T f''(z)\,dxdy = \frac{f(w_1)}{(w_1 - w_2)(w_1 - w_3)} + \frac{f(w_2)}{(w_2 - w_1)(w_2 - w_3)} + \frac{f(w_3)}{(w_3 - w_1)(w_3 - w_2)} \tag{16}$$

See Davis [6].

Now write $\iint_T f''(z)dxdy = \sum_{k=1}^{\infty} \iint_{\sigma_k} f''(z)dxdy$ and proceed as before.

4. **The use of ellipses.** When a weighted (real) integral of an analytic function can be shown to be equal to the double integral of its analytic continuation, then complete packings may be employed to produce a simple quadrature.

Let B be selected to be an ellipse. The ellipse will be normalized as follows. ε_ρ, $(\rho > 1)$, will designate the ellipse whose foci are at $z = \pm 1$ and whose semi-axes are respectively

$$a = \frac{1}{2}(\rho + \rho^{-1}), \qquad b = \frac{1}{2}(\rho - \rho^{-1}). \tag{17}$$

THEOREM. *Let* $C_n : |z - z_n| < r_n$ $(n = 1, 2, \cdots)$ *be a complete packing of the ellipse* ε_ρ. *If* $f(z)$ *is regular in the closure* ε_ρ, *then*

$$\int_{-1}^{+1} (1 - x^2)^{1/2} f(x)dx = 2\pi(\rho^2 - \rho^{-2})^{-1} \sum_{n=1}^{\infty} r_n^2 f(z_n). \tag{18}$$

Proof. We shall show that for all $f(z)$ regular in the closure of ε_ρ, we have the identity

$$\iint_{\varepsilon_\rho} f(z)\,dxdy = \frac{1}{2}(\rho^2 - \rho^{-2}) \int_{-1}^{+1} (1 - x^2)^{1/2} f(x)dx. \tag{19}$$

The quadrature formula (18) will then follow by writing

$$\iint_{\varepsilon_\rho} f(z)dxdy = \sum_{n=1}^{\infty} \iint_{\sigma_n} f(z)dxdy. \tag{20}$$

To show (19), introduce the Tschebyscheff polynomials of the 2nd kind

$$U_n(z) = (1 - z^2)^{-1/2} \sin[(n + 1) \arccos z], \quad (n = 0, 1, \cdots). \tag{21}$$

These polynomials are *doubly* orthogonal. We have, first of all, the real orthogonality

$$\int_{-1}^{+1} (2/\pi)^{1/2} U_m(x)\,(2/\pi)^{1/2} U_n(x)\,(1 - x^2)^{1/2}dx = \delta_{m,n}. \tag{22}$$

But, if $c_{n,\rho}$ are the normalizing constants defined by

$$c_{n,\rho} = 2\left(\frac{n + 1}{\pi}\right)^{1/2} \left(\rho^{2n+2} - \rho^{-2n-2}\right)^{-1/2}, \tag{23}$$

and if

$$p_n^*(z) = c_{n,\rho} U_n(z), \tag{24}$$

then we also have the complex orthogonality

$$\iint_{\varepsilon_\rho} p_m^*(z)\,\overline{p_m^*(z)}\,dxdy = \delta_{mn}. \tag{25}$$

If $f(z)$ is regular in the closure of ε_ρ, then the complex Fourier expansion

$$\begin{aligned} f(z) &= \sum_{n=0}^{\infty} \left(\iint_{\varepsilon_\rho} f(z)\,\overline{p_n^*(z)}\,dxdy\right) p_n^*(z) \\ &= \sum_{n=0}^{\infty} c_{n,\rho}^2 \left(\iint_{\varepsilon_\rho} f(z)\,\overline{U_n(z)}\,dxdy\right) U_n(z) \end{aligned} \tag{26}$$

converges uniformly and absolutely in closed subregions of ε_ρ. But the real Fourier expansion of $f(x)$

$$f(x) = \sum_{n=0}^{\infty} \frac{2}{\pi} \left(\int_{-1}^{+1} f(x) U_n(x)\,(1 - x^2)^{1/2}\,dx\right) U_n(x) \tag{27}$$

also converges uniformly in $[-1, 1]$. Hence, by the uniqueness of the Fourier coefficients, it must follow that

$$\frac{2}{\pi} \int_{-1}^{+1} f(x) U_n(x)\,(1 - x^2)^{1/2}\,dx = c_{n,\rho}^2 \iint_{\varepsilon_\rho} f(z)\,\overline{U_n(z)}\,dxdy \quad n = 0, 1, \cdots. \tag{28}$$

Selecting $n = 0$, and observing that $U_0(x) = 1$, we obtain (19) from (28) and (23).

As an example of an expansion of type (18), write $f(x) = e^{wx}$. Then, (see, e.g. Watson [10], p. 79)

$$\int_{-1}^{+1} (1 - x^2)^{1/2} e^{wx}\,dx = \frac{\pi I_1(w)}{w},$$

where

$$I_1(w) = \frac{1}{2} w + \frac{1}{2!}\left(\frac{1}{2} w\right)^3 + \frac{1}{3!}\left(\frac{1}{2} w\right)^5 + \cdots$$

is the modified Bessel function of the first order. Hence, if $|z - z_n| < r_n$, $n = 1, 2, \cdots$ is a complete packing of ε_ρ, we have, for all w,

$$\frac{I_1(w)}{w} = 2(\rho^2 - \rho^{-2})^{-1} \sum_{n=1}^{\infty} r_n^2 e^{z_n w}. \tag{29}$$

Some further identities that relate to the Tschebyscheff polynomials $U_n(z)$ should be noted. Writing $f(z)U_n(z)$ for $f(z)$ in (19), we obtain

$$\frac{4}{\pi} (\rho^2 - \rho^{-2})^{-1} \iint_{\varepsilon_\rho} f(z) U_n(z)\,dxdy = \frac{2}{\pi} \int_{-1}^{1} (1 - x^2)^{1/2} f(x) U_n(x)\,dx. \tag{30}$$

Hence, from (28), and then from (20),

$$\iint_{\varepsilon_\rho} f(z)U_n(z)dxdy = \frac{\pi c_{n,\rho}^2}{4}(\rho^2 - \rho^{-2})\iint_{\varepsilon_\rho} f(z)\,\overline{U_n(z)}\,dxdy \tag{31}$$

$$= \pi \sum_{k=1}^{\infty} r_k^2 f(z_k)U_n(z_k).$$

The particular selection of $f(z) = U_m(z)$ in (31) yields

$$\iint_{\varepsilon_\rho} U_m(z)U_n(z)dxdy = \frac{\pi c_{n,\rho}^2}{4}(\rho^2 - \rho^{-2})\iint_{\varepsilon_\rho} U_m(z)\,\overline{U_n(z)}\,dxdy \tag{32}$$

$$= \pi \sum_{k=1}^{\infty} r_k^2 U_m(z_k)U_n(z_k) = \frac{\pi}{4}(\rho^2 - \rho^{-2})\delta_{mn}.$$

The first term leads to a complex, but not Hermitian, orthogonality of the polynomials $U_m(z)$, while the 3rd is a point-wise orthogonality.

As a sideline, an inequality for analytic functions can be obtained from (19).

THEOREM. *Let $f(z)$ be analytic in the circle C_r: $|z| \leqq r$, $r > 1$. Let f be real on the real axis and satisfy*

$$Re\, f(z) \geqq 0, \qquad z \in C_r. \tag{33}$$

Then,

$$\frac{r}{\sqrt{r^2 - 1}} f(0) \geqq \frac{2}{\pi}\int_{-1}^{+1}(1 - x^2)^{1/2} f(x) \geqq \frac{\sqrt{r^2 - 1}}{r} f(0). \tag{34}$$

Proof. The circle C_r contains the ellipse ε_ρ where $r = (1/2)(\rho + \rho^{-1})$, and the ellipse, in turn, contains the circle C_s: $|z| \leqq s$, where $s = (1/2)(\rho - \rho^{-1}) = \sqrt{r^2 - 1}$. Now,

$$\pi r^2 f(0) = \iint_{\sigma_r} f(z)dxdy = \iint_{\sigma_r} Re\, f(z)dxdy \geqq \iint_{\varepsilon_\rho} Re\, f(z)dxdy$$

$$= Re\iint_{\varepsilon_\rho} f(z)dxdy = \frac{1}{2}(\rho^2 - \rho^{-2})\int_{-1}^{1}(1 - x^2)^{1/2} f(x)dx$$

$$\geqq \iint_{\sigma_s} Re\, f(z)dxdy = Re\iint_{\sigma_s} f(z)dxdy = \pi s^2 f(0).$$

Other packings of the ellipse ε_ρ by circles lead to other inequalities for the middle term of (34). See also Davis [6], p. 28.

EXAMPLE. Let $0 < w < \pi/2$ be fixed. The function $f(z) = e^{wz}$, $z = x + iy$, is real on the real axis, and $Re\, f(z) = e^{wx}\cos wy \geqq 0$ for $z \in C_r$ with $r = \pi/2w$. We have $r > 1$. Hence, for

$$0 < w < \pi/2, \quad \frac{1}{\sqrt{1-\left(\frac{2w}{\pi}\right)^2}} \geqq \frac{2I_1(w)}{w} \geqq \sqrt{1-\left(\frac{2w}{\pi}\right)^2}.$$

Equality holds for $w = 0$, and the 2nd inequality, of course, is obvious from the Maclaurin expansion of $I_1(w)/w$.

The identity (28) leads to a simple quadrature for inner products of the type $\iint_B f(z)\,\overline{g(z)}\,dxdy$ and which holds under certain regularity conditions.

LEMMA. *Let $\Delta > \rho^3 > \rho > 1$, and suppose that $g(z)$ is regular in ε_Δ. Let*

$$g(z) = \sum_{n=0}^{\infty} b_n p_n^*(z) \tag{35}$$

be the expansion of g in Tschebyscheff polynomials of the second kind. Then, the related function

$$h(z) = \sum_{n=0}^{\infty} (\overline{b_n}/c_{n,\rho}^2) p_n^*(z) \tag{36}$$

is regular in $\varepsilon_{\Delta/\rho}2$, and for all $f(z)$ regular in $\bar{\varepsilon}_\rho$, we have

$$\iint_{\varepsilon_\rho} f(z)\,\overline{g(z)}\,dxdy = \frac{2}{\pi}\int_{-1}^{1} (1-x^2)^{1/2} h(x)\,f(x)dx. \tag{37}$$

Proof. Note that $\varepsilon_\rho \subset \varepsilon_{\Delta/\rho^2} \subset \varepsilon_{\rho^3} \subset \varepsilon_\Delta$. Since $g(z)$ is regular in ε_Δ, we have

$$\limsup_{n\to\infty} |\, b_n \,|^{1/n} \leqq \frac{\rho}{\Delta}. \tag{38}$$

Since

$$\lim_{n\to\infty} \left(\frac{1}{c_{n,\rho}^2}\right)^{1/n} = \rho^2, \tag{39}$$

it follows that

$$\limsup_{n\to\infty} \left| \frac{\overline{b_n}}{c_{n,\rho}^2} \right|^{1/n} \leqq \frac{\rho}{\Delta} \cdot \rho^2 = \frac{\rho}{\Delta/\rho^2}. \tag{40}$$

Therefore $h(z)$ is regular in $\varepsilon_{\Delta/\rho^2}$, the series (36) converging absolutely and uniformly in closed subsets of $\varepsilon_{\Delta/\rho^2}$.

Now we have

$$\begin{aligned} \iint_{\varepsilon_\rho} f(z)\,\overline{g(z)}\,dxdy &= \sum_{n=0}^{\infty} \overline{b_n} \iint_{\varepsilon_\rho} f(z)\,\overline{p_n^*(z)}\,dxdy \\ &= \sum_{n=0}^{\infty} \overline{b_n}\, c_{n,\rho} \iint_{\varepsilon_\rho} f(z)\,\overline{U_n(z)}\,dxdy. \end{aligned} \tag{41}$$

On the other hand,

$$(42)\quad \frac{2}{\pi}\int_{-1}^{+1}\sqrt{1-x^2}\,h(x)\,f(x)dx = \frac{2}{\pi}\int_{-1}^{1}\sqrt{1-x^2}\,f(x)\sum_{n=0}^{\infty}(\overline{b_n}/c_{n,\rho}^2)p_n^*(x)dx$$

$$= \sum_{n=0}^{\infty}\frac{\overline{b_n}}{c_{n,\rho}^2}\,\frac{2}{\pi}\int_{-1}^{1}\sqrt{1-x^2}\,f(x)p_n^*(x)dx$$

$$= \sum_{n=0}^{\infty}\frac{\overline{b_n}}{c_{n,\rho}}\,\frac{2}{\pi}\int_{-1}^{1}\sqrt{1-x^2}\,f(x)U_n(x)dx.$$

Applying (28) to the last integral there is obtained

$$(43)\quad \frac{2}{\pi}\int_{-1}^{+1}\sqrt{1-x^2}\,h(x)\,f(x)dx = \sum_{n=0}^{\infty}\overline{b_n}\,c_{n,\rho}\iint_{\varepsilon_\rho} f(z)\,\overline{U}_n(z)dxdy.$$

Comparing (41) with (43) we obtain (37).

The relevant identities and convergence theorems for the Tschebyscheff polynomials $U_n(x)$ can be found in Szegö [9], p. 59 (4.1.7), p. 238, (9.1.4.).

If we define a linear transformation T_ρ by means of

$$(44)\quad T_\rho(g) = h(z) = \sum_{n=0}^{\infty} c_{n,\rho}^2\,\overline{b_n}\,p_n^*(z),$$

then we can write (37) in the form

$$(45)\quad \iint_{\varepsilon_\rho} f(z)\,\overline{g(z)}\,dxdy = \frac{2}{\pi}\int_{-1}^{+1}\sqrt{1-x^2}\,f(x)T\rho(g)dx.$$

LEMMA. *Let $f(z)$ be regular in $\overline{\varepsilon}_\rho$ and let $g(z)$ be regular in ε_Δ where $\Delta > \rho^3 > \rho > 1$. Then,*

$$(46)\quad \iint_{\varepsilon_\rho} f(z)\,\overline{g(z)}\,dxdy = \frac{4}{\pi(\rho^2-\rho^{-2})}\iint_{\varepsilon_\rho} f(z)T_\rho(g(z))dxdy.$$

Proof. Since $f(z)$ is regular in $\overline{\varepsilon}_\rho$ and since $T_\rho(g)$ is regular in $\varepsilon_{\Delta/\rho^2}$, the product $f(z)T_\rho(g(z))$ is also regular in $\overline{\varepsilon}_\rho$. Hence, by (19) (with the $f(z)$ of this equation set equal to the present $f(z)T_\rho(g(z))$,

$$(47)\quad \frac{2}{\pi}\int_{-1}^{+1}\sqrt{1-x^2}\,f(x)T_\rho(g(x))dx = \frac{4}{\pi(\rho^2-\rho^{-2})}\iint_{\varepsilon_\rho} f(z)T_\rho(g(z))dxdy\,.$$

Combining (45) with (47) we obtain (46).

We can now obtain the following simple quadrature.

THEOREM. *Let $C_n\colon |z-z_n| < r_n$, $n = 1, 2, \cdots$ be a complete packing of the ellipse ε_ρ. Let $g(z)$ be regular in ε_Δ, where $\Delta > \rho^3 > \rho > 1$. Then, for all f that is regular in $\overline{\varepsilon}_\rho$,*

$$\iint_{\varepsilon_\rho} f(z)\overline{g(z)}\,dxdy = \sum_{k=1}^{\infty} a_k f(z_k), \tag{48}$$

where

$$a_k = \frac{4}{\rho^2 - \rho^{-2}}\, r_k^2 (T_\rho(g(z)))_{z=z_k}\,. \tag{49}$$

Proof. Under the above conditions, (46) holds.

Now write $\iint_{\varepsilon_\rho} f(z)T_\rho(g(z))dxdy = \sum_{k=1}^{\infty} \iint_{\sigma_k} f(z)T_\rho(g)dxdy$ and apply the 2nd Lemma of § 2.

This representation should be contrasted with that given in (14). The higher derivatives are no longer present.

5. **The use of lunes.** If B is selected as the area common to two circles we may obtain a simple quadrature formula for another class of analytic functions.

Let B designate the common part of the two circles

$$|z - \sigma i|^2 = 1 + \sigma^2 \text{ and } |z + \sigma i|^2 = 1 + \sigma^2; \sigma > 0, i = \sqrt{-1}.$$

These circles both pass through $z = \pm 1$ and their centers are at $z = \sigma i$ and $z = -\sigma i$ respectively.

THEOREM. *Let* C_n: $|z - z_n| < r_n$, $n = 1, 2, \cdots$ *be a complete packing of the lune* B. *If* $f(z)$ *is regular in* $\bar{B}$ *and has zeros at* $z = +1$ *and* $z = -1$, *then,*

$$\int_{-1}^{+1} f(x)dx = \frac{\pi}{\sigma} \sum_{k=1}^{\infty} \frac{z_k^2 + \sigma^2}{1 - z_k^2}\, r_k^2 f(z_k). \tag{50}$$

Proof. We have already established (see Davis [6]) that if $g(z)$ is regular in $\bar{B}$,

$$\iint_B g(z)\,dxdy = \sigma \int_{-1}^{+1} g(x)\,\frac{1 - x^2}{\sigma^2 + x^2}\,dx. \tag{51}$$

In view of the above conditions on $f(z)$, the function

$$g(z) = \frac{f(z)\,(z^2 + \sigma^2)}{1 - z^2}$$

will be regular in $\bar{B}$. Hence,

$$\sigma^{-1} \iint_B \frac{f(z)\,(z^2 + \sigma^2)}{1 - z^2}\,dxdy = \int_{-1}^{+1} f(x)dx. \tag{52}$$

The simple quadrature (50) now follows by writing $\iint_B g(z)dxdy =$

$\sum_{n=1}^{\infty} \iint_{\sigma_n} g(z)dxdy.$

6. **Final remarks.** In (18) and in (50), we have seen how simplequadratures (with complex abscissas) may be found for $\int_{-1}^{1} (1-x^2)^{1/2} f(x)dx$ and for $\int_{-1}^{1} (1-x^2) f(x)dx$. Regarding these integrals as linear functionals on f, we can pose the following question. For what ilnear functionals L is there a simple representation of the form

$$L(f) = \sum_{n=1}^{\infty} a_n f(z_n) \tag{53}$$

which is valid for every function f of a space F of functions that are regular in a fixed region B? If an expansion of type (53) were demanded for *all* linear functionals in F^*, then we are inquiring whether or not the point functionals $L_n(f) = f(z_n)$ form a basis for the dual space F^*, convergence being taken in the weak* sense of (53). Information on either of these problems, derived from the theory of normed linear spaces or from the theory of functions, would be welcome.

BIBLIOGRAPHY

1. R. P. Boas, *Entire functions*, New York, 1954.
2. T. Carleman, *Sur les series* $\Sigma A_\nu/(z-\alpha_\nu)$, Comptes Rendus, **174** (1922), 588-591.
3. P. J. Davis, *On simple quadratures*, Proc. Amer. Math. Soc. **4** (1953), 127-136.
4. P. J. Davis and J. L. Walsh, *On representations and extensions of bounded linear functionals defined on classes of analytic functions*, Trans. Amer. Math. Soc. **76** (1954), 190-206.
5. P. J. Davis, *Packing inequalities for circles*, Michigan Journal of Math. **10** (1963), 25-31.
6. ———, *Triangle formulas in the complex plane*, Math. of Computation, **18** (1964), 569-577.
7. ———, *Interpolation and approximation*, New York, 1963.
8. D. Pompeiu, *Opera mathematica*, 1959.
9. G. Szegö, *Orthogonal polynomials*, New York, 1939.
10. G. N. Watson, *A Treatise on the theory of Bessel functions*, 2nd Ed., 1952, London and New York.
11. J. Wermer, *Uniform approximation and maximal ideal spaces*, Bull. Amer. Math. Soc. **68** (1962), 298-305.
12. O. Wesler, *An infinite packing theorem for spheres*, Proc. Amer. Math. Soc. **11** (1960), 324-326.
13. J. Wolff, *Sur les series* $\Sigma A_\nu/(z-a_\nu)$, Comptes Rendus, **173** (1921), 1056-1057.

BROWN UNIVERSITY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

ON A COINCIDENCE THEOREM OF F. B. FULLER

E. FADELL

Our objective is the following theorem.

THEOREM (1.1). **Let X denote a space dominated by a finite n-polyhedron K such that $H^n(X; Z)$ is torsion free. Let M be a compact topological n-manifold which is 1-connected and let $f, g: X \to M$ be two given maps. Then, there is a map $g' \sim g: X \to M$ such that f and g' are coincidence free if, and only if, the (rational) Lefschetz coincidence class $L(f, g) = 0$.**

If X is a finite n-polyhedron and M is a triangulated, 1-connected compact manifold, the theorem is implicit in the work of F. B. Fuller ([5], [6]). In this case, one shows that the obstruction to deforming $f \times g: K \to M \times M$ to a map $f' \times g': K \to M \times M - \Delta$, Δ the diagonal of $M \times M$, is zero. Thus, one obtains maps $f' \sim f$, $g' \sim g$, such that f' and g' are coincidence free. Then, as shown in Fuller's thesis [6], one observes that since M is a manifold (local homogeneity is all that is necessary), deforming both f and g to obtain coincidence free maps is equivalent to deforming just one of them to achieve coincidence freeness. We will use a direct approach, employing general obstruction theory for deforming cross sections into fiber subspaces, simple techniques in fiber spaces to relax the conditions on X and the fact that the classical computations relating $H^n(M \times M, M \times M - \Delta)$ and $H^n(M \times M)$ where M is a triangulated manifold remain valid for topological manifolds by employing the techniques of tangent fiber spaces [4] or the recent result of Kister [7] that microbundles are bundles. An immediate application of this theorem is an obstruction theory proof of the following converse of the Lefschetz Fixed Point Theorem for compact 1-connected topological manifolds.

COROLLARY (1.2). *Let M denote a compact 1-connected topological manifold, $f: M \to M$ a given map and L_f the Lefschetz number for f. If $L_f = 0$, there is a map g, $g \sim f$, such that g is fixed point free.*

The corollary for triangulated manifolds, because of our previous remarks, is implicit in Fuller [6]. The triangulated case (which is doubtless classical but hard to find explicitly stated anywhere) may also be shown using the theory of Nielson-Reidemeister fixed point classes and results of Wecken [11]. An alternative proof of the corollary in

Received April 20, 1964. This work was supported by the National Science Foundation under Grant NSF-23849 and the Wisconsin Alumni Research Foundation.

the topological case, without obstruction theory, may be given using Nielson-Reidemeister fixed point classes, results of Weier [12] and the existence of a local index theory for fixed points in the category of ANR's ([2] or [3]). Finally, we remark that (1.1) was discovered by the author before he became aware of Fuller [6] in order to answer a question, raised to him by E. E. Floyd, which is answered by (1.2). (1.2) is easily seen to be false for polyhedra in general.

Singular homology is employed throughout.

2. **Preliminaries.** We recall first the necessary obstruction theory [6].

Let $\xi = (E, p, B, F)$, $\xi_0 = (E_0, p_0, B, F_0)$ denote locally trivial fiber spaces over B such that $E_0 \subset E$, $p_0 = p \mid E_0$ and for each $x_0 \in B$ there is a neighborhood U of x_0 and a homeomorphism of pairs

$$(p^{-1}(U), p_0^{-1}(U)) \xleftarrow{\varphi_U} (U \times F, U \times F_0)$$

such that $p\varphi_U(x, y) = x$, $(x, y) \in U \times F$.

DEFINITION (2.1). (ξ, ξ_0) is called a *locally trivial pair*, with fiber (F, F_0).

Suppose now that (ξ, ξ_0) is a locally trivial pair, as above, such that the fibre (F, F_0) is n-simple. Then, $\pi_n(F, F_0)$ is a local system of groups on B. Suppose further that B is a locally finite simplicial complex, with k-skeleton denoted by B^k, and $f: B \to E$ is a given cross section such that $f(B^{n-1}) \subset E_0$. We paraphrase Theorem 3.2 of [6] as follows.

THEOREM (2.2). *There exists an obstruction class*

$$\bar{d}^n(f) \in H^n(B; \pi_n(F, F_0))$$

such that if $\bar{d}^n(f) = 0$, *then* f *is homotopic (rel* B^{n-2}*) to a cross section* $f': B \to E$ *such that* $f'(B^n) \subset E_0$.

REMARK (2.3). It is understood that when cross sections are homotopic we mean homotopic in the family of cross sections.

Suppose now we add the assumption that $\pi_k(F, F_0) = 0$ for $k < n$ and $\pi_1(B)$ acts trivially on $\pi_n(F, F_0)$. If $g: B \to E$ is a given cross section, then by (2.2) there is a cross section $f: B \to E$, homotopic to g, such that $f(B^{n-1}) \subset E_0$. Lundell [6, p. 167] has computed $\bar{d}^n(f)$ as follows. Designate $\pi_n(F, F_0)$ by π. The Hurewicz homomorphism $h: \pi \to H_n(F, F_0)$ is an isomorphism. Let $\theta^n \in H^n(F, F_0; \pi)$ denote the fundamental class of (F, F_0), i.e., the class corresponding to h^{-1} under the universal coefficient isomorphism $H^n(F, F_0; \pi) \approx \mathrm{Hom}\,(H_n(F, F_0); \pi)$. Consider the diagram

$$\begin{array}{ccc} H^n(E, E_0; \pi) & \xrightarrow{k^*} & H^n(E; \pi) \\ {\scriptstyle i^*}\downarrow & & \downarrow{\scriptstyle f^*} \\ H^n(F, F_0; \pi) & \longrightarrow & H^n(B; \pi) \end{array}$$

where i and k are inclusion maps.

PROPOSITION (2.4) ([6]). $\bar{d}^n(f) = f^*k^*i^{*-1}(\theta^n)$.

Since $f^* = g^*$, $\bar{d}^n(f)$ depends only on g and we may employ the notation $\bar{d}^n(f) = \bar{d}^n(g)$ and state the following.

PROPOSITION (2.5). Given a locally trivial pair (ξ, ξ_0) satisfying all the assumptions imposed above, then a given cross section $g: B \to E$ is homotopic to a cross section $g': B \to E$ such that $g'(B^n) \subset E_0$ if, and only if, $\bar{d}^n(g) = g^*k^*i^{*-1}(\theta^n) = 0$.

Let $\xi = (E, p, B, F)$ and $\xi_0 = (E_0, p_0, B, F_0)$, where (ξ, ξ_0) is a locally trivial pair. Let $\varphi: K \to B$ denote a map. Then, φ induces a locally trivial pair $(\xi', \xi_0') = \varphi^{-1}(\xi, \xi_0)$ over K in the usual manner as follows. $\xi' = (E', p', K, F)$, where

$$E' = \{(x, y) \in K \times E: \varphi(x) = p(y)\},$$
$$p'(x, y) = x.$$

$\xi_0' = (E_0', p_0', K, F_0)$, where

$$E_0' = \{(x, y) \in K \times E_0: \varphi(x) = p_0(y)\}$$
$$p_0'(x, y) = x.$$

Suppose $g: B \to E$ is a given cross section in ξ. Then, $g'(x) = (x, g\varphi(x))$, $x \in K$, defines a corresponding cross section in g'. Letting $\bar{\varphi}(x, y) = y$ we have the following commutative diagram

$$\begin{array}{ccc} E' & \xrightarrow{\bar{\varphi}} & E \\ {\scriptstyle p'}\downarrow\uparrow{\scriptstyle g'} & & {\scriptstyle p}\downarrow\uparrow{\scriptstyle g} \\ K & \xrightarrow{\varphi} & B \end{array}$$

and $\bar{\varphi}(E_0') \subset E_0$.

LEMMA (2.6). *Suppose $\gamma_0: B \to E_0$ is a map (not necessarily a cross section) such that $g \sim \gamma_0$ (as maps). If (ξ, ξ_0) is a fibered pair in the sense of* [1] *(e.g. if B is paracompact), g is homotopic to a cross section $g_0: B \to E_0$.*

Proof. This lemma is a special case of Lemma (5.1) given in § 5.

LEMMA (2.7). *Suppose $\varphi: K \to B$ is a dominating map, i.e., there is a map $\psi: B \to K$ such that $\varphi\psi \sim 1$, and (ξ, ξ_0) is a fibered pair. Then, there is a cross section $g_0: B \to E_0$, $g_0 \sim g$, if, and only if, there is a cross section $g_0': K \to E_0'$, $g_0' \sim g'$.*

Proof. The necessity is simple and hence we confine ourselves to the sufficiency argument. Suppose, therefore, that there is a cross section $g_0': K \to E_0$ such that $g_0' \sim g'$. Let $\gamma = \bar{\varphi} g' \psi$ and $\gamma_0 = \bar{\varphi} g_0' \psi$. Obviously $\gamma \sim \gamma_0$. On the other hand, $\gamma = g\varphi\psi \sim g$. Therefore, $g \sim \gamma_0$, $\gamma_0: B \to E_0$. Lemma (2.6) applies to give the cross section $g_0: B \to E_0$, $g_0 \sim g$.

Continuing with the above notation where (ξ, ξ_0) is locally trivial fibered pair and $g: B \to E$ is a given cross section let us assume further

(A)
- (i) $\pi_i(F, F_0) = 0$ $i < n$, $\pi_n(F, F_0)$ is simple,
- (ii) $\pi_i(B)$ acts trivially on $\pi_n(F, F_0)$,
- (iii) there is a dominating map $\varphi: K \to B$ where K is a finite polyhedron of dimension $\leqq n$.

We have the following commutative diagram where (ξ', ξ_0') is the locally trivial fibered pair induced over K by φ, g' is the cross section in ξ' associated with g as above, and $\pi = \pi_n(F, F_0)$.

(B)
$$\begin{array}{ccccccc} & & H^n(E', E_0'; \pi) & \xrightarrow{k'^*} & H^n(E'; \pi) & \xrightarrow{g'^*} & H^n(K; \pi) \\ & \overset{i^{*\prime}}{\swarrow}\approx & \uparrow{\scriptstyle \bar{\varphi}^*} & & \uparrow{\scriptstyle \bar{\varphi}^*} & & \uparrow{\scriptstyle \varphi^*} \\ H^n(F, F_0; \pi) & & & & & & \\ & \underset{i^*}{\nwarrow}\approx & H^n(E, E_0; \pi) & \xrightarrow{k^*} & H^n(E; \pi) & \xrightarrow{g^*} & H^n(B; \pi) \end{array}$$

i, i', k, k' are inclusion maps.

PROPOSITION (2.8). The cross section $g: B \to E$ is homotopic to a cross section $g_0: B \to E_0$ if, and only if, $g^*k^*i^{*-1}(\theta^n) = 0$, where θ^n is the fundamental class of (F, F_0).

Proof. The assumptions (A) tell us that, via (2.5), the associated cross section $g': K \to E'$ is homotopic to a cross section $g_0': K \to E_0'$ if, and only if, $\bar{d}^n(g') = g'^*k'^*i'^{*-1}(\theta^n) = 0$. Thus (2.7) implies that there is a cross section $g_0: B \to E_0$, $g \sim g_0$ if, and only if, $g'^*k'^*i'^{*-1}(\theta^n) = 0$. But diagram (B) gives

$$\varphi^* g^* k^* i^{*-1}(\theta^n) = g'^* k'^* i'^{*-1}(\theta^n) .$$

Since φ^* is injective the proposition follows.

The preceding material will be applied to the following situation.

Let M denote a topological n-manifold (connected, separable metric). Let $\xi = (M \times M, p, M, M)$, where $p(b, y) = b$ and

$$\xi_0 = (M \times M - \Delta, p_0, M, M - b)$$

where Δ is the diagonal in $M \times M$ and p_0 is the restriction of p. Then, (ξ, ξ_0) is a locally trivial fibered pair with fiber $(M, M - b)$, $b \in M$. Let R denote a principal ideal domain and suppose M is R-orientable, i.e., $\pi_1(M)$ acts trivially on $H_n(M, M - b; R)$ or, alternatively, (ξ, ξ_0) is R-orientable. Then, the inclusion map $i: (M, M - b) \to (M \times M, M \times M - \Delta)$ induces an isomorphism (see [4])

$$i^*: H^n(M \times M, M \times M - \Delta; R) \xrightarrow{\approx} H^n(M, M - b; R) \approx R .$$

Suppose now that $f: X \to M$ is any given map where X is an arbitrary space. Let $E = X \times M$, $E_0 = X \times M - \Delta_f$, where Δ_f is the graph of f. Let $q: E \to X$ be projection on the first factor and $q_0 = q \mid E_0$. For fixed $x \in X$, $q^{-1}(x) = x \times M = M$ and $q_0^{-1}(x) = x \times (M - f(x)) = M - f(x)$. Let $\xi(f) = (E, q, X, M)$, $\xi_0(f) = (E_0, q_0, X, M - f(x))$.

PROPOSITION (2.9). $(\xi(f), \xi_0(f))$ is a locally trivial fibered pair with fiber $(M, M - f(x))$. If (ξ, ξ_0) is R-orientable, so is $(\xi(f), \xi_0(f))$.

Proof. $(\xi(f), \xi_0(f))$ may be identified with $f^{-1}(\xi, \xi_0)$.

REMARK (2.10). There is a natural map $\bar{f}: E = X \times M \to M \times M$ given by $\bar{f}(x, y) = (f(x), y)$ which gives rise to a commutative diagram

$$\begin{array}{ccc} E & \xrightarrow{\bar{f}} & M \times M \\ {\scriptstyle q}\downarrow & & \downarrow{\scriptstyle p} \\ X & \xrightarrow{f} & M \end{array}$$

and $\bar{f}(E_0) = \bar{f}(X \times M - \Delta_f) \subset M \times M - \Delta$.

If M is 1-connected and dim $M \geqq 3$, then $M - b$ is 1-connected and $\pi_k(M, M - b) = 0$ for $k < n$ and $\pi_n(M, M - b)$ is a simple system. If furthermore, X is dominated by a finite polyhedron of dimension $\leqq n$, then $(\xi(f), \xi_0(f))$ satisfies the conditions in (A).

REMARK (2.11). If M is a compact, 1-connected 2-manifold, then $M = S^2$ (the 2-sphere) and the preceding remarks remain valid.

3. Some computations. We discuss first Lefschetz coincidence theory in this setting.

Let M denote an orientable (= Z-orientable) compact topological

n-manifold, R any principal ideal domain. Let $\mu \in H^n(M; R)$, $\bar{\mu} \in H_n(M; R)$ denote fundamental classes corresponding to a given orientation of M. Then for $b \in M$, the inclusion map $j: M \to (M, M - b)$ induces an isomorphism [8]

$$H^n(M; R) \xleftarrow[\approx]{j^*} H^n(M, M - b; R)$$

and we define U_b by $j^*(U_b) = \mu$. Now consider the fibered pair (ξ, ξ_0) where $\xi = (M \times M, p, M, M)$, $\xi_0 = (M \times M - \Delta, p_0, M, M - b)$ (see § 2) which gives rise to the diagram

$$H^n(M, M - b; R) \xleftarrow[\approx]{i^*} H^n(M \times M, M \times M - \Delta; R) \xrightarrow{k^*} H^n(M \times M; R) ,$$

where i and k are inclusions. Define U by $i^*(U) = U_b$ and set $\underline{U} = k^*(U)$. This class $\underline{U}$ is studied by Milnor [7] in case R is a field and M is an R-orientable differentiable manifold. The techniques carry over to the topological situation by using the techiques of tangent fiber spaces [4] or a recent result of J. Kister [7] and we will make use of this fact. For convenience we refer to $\underline{U}$ as the *basic R-class* for M. It is well-defined up to orientation

DEFINITION (3.1). Let M denote an R-orientable compact topological manifold with basic R-class $\underline{U}$. Let $f, g: X \to M$ denote two maps from a space X to M. These maps induce $f \times g: X \times X \to M \times M$. Furthermore, let $\delta: X \to X \times X$ denote the diagonal map. The class

$$\delta^*(f \times g)^*(\underline{U}) = L(f, g; R)$$

is called the *R-Lefschetz coincidence class* for f and g.

REMARK (3.2). When $R = Q$, the field of rationals, $L(f, g; Q)$ is the Lefschetz class usually considered and we distinguish it by the notation $L(f, g)$.

The following theorem in which we employ the notation in (3.1) is immediate.

THEOREM (Lefschetz). *If f and g are coincidence free, then $L(f, g; R) = 0$.*

Let us assume now that R is a field and M is R-orientable as above. Following Milnor [7] arrange a basis

$$(1) \qquad 1 = \alpha_1, \alpha_2, \cdots, \alpha_N = \mu .$$

for $H^*(M; R)$ in increasing order of dimension. We find it also convenient to introduce the notation

(2) K_p = integer subscripts in (1) of p-dimensional elements.

Then, the basic R-class $\underline{U}$ takes on the form

$$(3)\qquad \underline{U} = \Sigma c_{ij}\alpha_i \times \alpha_j = \sum_p \Big(\sum_{i\in K_p} \sum_{j\in K_{n-p}} c_{ij}\alpha_i^p \times \alpha_j^{n-p} \Big) ,$$

where the matrix

$$(4)\qquad C_{p,n-p} = \| c_{ij} \| , \qquad (i,j) \in K_p \times K_{n-p}$$

is square and if

$$(5)\qquad Y_{n-p,p} = \| y_{ij} \| , \quad y_{ij} = \langle \alpha_i \cup \alpha_j, \bar{\mu} \rangle , \quad (i,j) \in K_{n-p} \times K_p ,$$

we have

$$(6)\qquad (-1)^{np} C_{p,n-p} Y_{n-p,p} = I \text{ (an identity matrix)} .$$

Define

$$(7)\qquad \beta_j^{n-p} = (-1)^p \sum_{k\in K_{n-p}} c_{jk}\alpha_k^{n-p} , \qquad j \in K_p .$$

Then it is easily verified that

$$(8)\qquad \langle \alpha_i^p \cup \beta_j^{n-p}, \bar{\mu} \rangle = \delta_{ij} , \qquad i,j \in K_p ,$$

so that the β's form a dual basis. A simple computation now gives the classical formula

$$(9)\qquad L(f,g;R) = \sum_p \Big((-1)^p \sum_{i\in K_p} f^*(\alpha_i^p) \cup g^*(\beta_i^{n-p}) \Big)$$

where f, g are as in (3.1). If $X = M, f = 1$, and R is the field of rationals, another straightforward computation shows

$$(10)\qquad L(1,g) = (-1)^n \Big(\sum_p (-1)^p tr(g_p^*) \Big) \mu = (-1)^n L_g \mu$$

where trg_p^* is the trace of g^* in dimension p and $L_g = \sum_p (-1)^p tr(g_p^*)$ is the usual Lefschetz number for g.

REMARK (3.2). It is clear that the preceding material is applicable to compact topological manifolds which are not necessarily orientable if Z_2 coefficients are employed.

4. Proof of Theorem (1.1). The necessity is immediate and we confine ourselves to the "if part."

Let (ξ, ξ_0) denote the fibered pair in § 3, i.e., $\xi = (M \times M, p, M, M)$, $\xi_0 = (M \times M - \Delta, p_0, M, M - b)$. The given map $f: X \to M$ induces the fiber space $(\xi(f), \xi_0(f))$ over X (§ 2) which satisfies conditions (A)

(§ 2). The fiber of $(\xi(f), \xi_0(f))$ over $x \in X$ is $(M, M-b)$, $b = f(x)$. Letting $\xi(f) = (E, q, X, M)$, $\xi_0(f) = (E_0, q_0, X, M-b)$, we have the commutative diagram

$$\begin{array}{ccc} E & \xrightarrow{\bar{f}} & M \times M \\ {\scriptstyle q}\downarrow & & \downarrow{\scriptstyle p} \\ X & \xrightarrow{f} & M \end{array}$$

where $\bar{f}(x, y) = (f(x), y)$. The other given map $g: X \to M$ induces a cross section $\tilde{g}: X \to E$ by $\tilde{g}(x) = (x, g(x))$, $x \in X$. We note at this point that $\bar{f}\tilde{g} = (f \times g)\delta$, where $\delta: X \to X \times X$ is the diagonal map and $f \times g: X \times X \to M \times M$ is the product map. Our objective is to deform the cross section $\tilde{g}$ into $E_0 = X \times M - \Delta_f$. Consider the commutative diagram

$$\begin{array}{ccccc} & & H^n(E, E_0; \pi) & \xrightarrow{\tilde{k}^*} & H^n(E; \pi) \\ & \overset{\tilde{i}^*}{\swarrow} & & & \uparrow{\scriptstyle \bar{f}^*} \\ H^n(M, M-b; \pi) & & \uparrow{\scriptstyle \bar{f}^*} & & \\ & \underset{i^*}{\nwarrow} & & & \\ & & H^n(M \times M, M \times M - \Delta; \pi) & \xrightarrow{k^*} & H^n(M \times M; \pi) \end{array} \qquad \begin{array}{ccc} & \overset{\tilde{g}^*}{\searrow} & \\ & & \\ H^n(X; \pi) & & \end{array}$$

where i, $\tilde{i}$, k, $\tilde{k}$ are inclusions and $\pi = \pi_n(M, M-b) \approx Z$. If θ^n is the fundamental class of $(M, M-b)$, we need to show, by (2.8), that $\tilde{g}^*\tilde{k}^*\tilde{i}^{*-1}(\theta^n) = 0$. We may choose an orientation $\mu \in H^n(M; \pi)$ of M so that $j^*(\theta^n) = \mu$ where $j: M \to (M, M-b)$ is the inclusion map. This orientation determines

$$U \in H^n(M \times M, M \times M - \Delta; \pi) \quad \text{and} \quad \underline{U} \in H^n(M \times M; \pi)$$

such that $\underline{U} = k^*i^{*-1}(\theta^n)$. The corresponding π-Lefschetz class is given by $\delta^*(f \times g)^*(\underline{U}) = L(f, g; \pi)$. On the other hand

$$\begin{aligned} \tilde{g}^*\tilde{k}^*\tilde{i}^{*-1}(\theta^n) &= \tilde{g}^*\tilde{k}^*\tilde{i}^{*-1}i^*(U) = \tilde{g}^*\tilde{k}^*\bar{f}^*(U) \\ &= \tilde{g}^*\bar{f}^*k^*(U) = \tilde{g}\bar{f}^*(\underline{U}) = \delta^*(f \times g)^*(\underline{U}) \,. \end{aligned}$$

Therefore, $L(f, g; \pi) = \tilde{g}^*\tilde{k}^*i^{*-1}(\theta^n)$ and hence the integral Lefschetz class is precisely the obstruction class in question. Now, if Q is the field of rationals, the inclusion homomorphism $l: \pi \to Q$ induces $l^*: H^n(X; \pi) \to H^n(X; Q)$. It is clear that $l^*(L(f, g; \pi)) = L(f, g; Q) = L(f, g)$, where the Lefschetz class $L(f, g)$ is chosen in terms of the orientation μ. But the assumption that $H^n(X; \pi)$ is torsion free together with the fact that X is dominated by a finite polyhedron implies that l^* is injective. Therefore, $L(f, g) = 0$ implies $L(f, g; \pi) = 0$, which in turn implies $\tilde{g}\tilde{k}^*\tilde{i}^{*-1}(\theta^n) = 0$ and the theorem follows.

Since a compact topological n-manifold is always dominated by a finite polyhedron of dimension n, we have the following corollary, which is a converse to the Lefschetz Fixed Point Theorem.

COROLLARY (4.1). *Let M denote a compact, 1-connected topological manifold, $f: M \to M$ a given map and L_f the Lefschetz number for f. If $L_f = 0$, there is a map g, $g \sim f$, such that g is fixed point free.*

Proof. By (10) of § 3, $L(1, f) = (-1)^n L_f \mu$, where $n = \dim M$. If $L_f = 0$, then $L(1, f) = 0$ and $f \sim g$ such that 1 and g have no coincidences, i.e., g has no fixed points.

Corollary (4.1) may be reformulated as follows. Let α denote a homotopy class of maps $M \to M$. Call α *fixed point free* if α contains a fixed point free representative Also define L_α by $L_\alpha = L_f$, $f \in \alpha$.

COROLLARY (4.2). *If M is a compact, 1-connected topological manifold and $\alpha: M \to M$ is a homotopy class of maps, then α is fixed point free if, and only if, $L_\alpha = 0$.*

5. Appendix.

LEMMA (5.1). *Let (ξ, ξ_0) denote a locally trivial fibered pair where $\xi = (E, p, B, F)$, $\xi_0 = (E_0, p_0, B, F_0)$. Suppose we are given a commutative diagram of maps*

$$\begin{array}{ccc} & & E \\ & \overset{F}{\nearrow} & \big\downarrow p \\ X & \xrightarrow{f} & B \end{array}$$

and a map $G: X \to E_0$ such that $F \sim G$. Then, there is a homotopy $H: X \times I \to E$ such that $H_1 = F$, $pH_t = f$, $0 \leqq t \leqq 1$, and $H_0: X \to E_0$.

Proof. Since (ξ, ξ_0) is a fibered pair in the sense of [1], there is a regular lifting function λ for ξ which works simultaneously for ξ_0, i.e., if $e_0 \in E_0$ and $\omega \in B^I$, then $\lambda(e_0, \omega)$ is a path in E_0. By assumption there is a homotopy $K: X \times I \to E$ such that $K_0 = G$ and $K_1 = F$. K induces a map $\tilde{K}: X \to E^I$ in the usual manner with $\tilde{K}(x)(t) = K(x, t)$, $0 \leqq t \leqq 1$. Let $\alpha(x) \in B^I$ denote the path $p(x)$, and observe that $\alpha(x)(t) = p(K(x, t))$, $0 \leqq t \leqq 1$, and, in particular, $\alpha(x)(1) = f(x)$. For $0 \leqq s \leqq t$, let $\alpha(x)_s$ denote the path given by $\alpha(x)_s(t) = \alpha(x)(s + t - st)$, $0 \leqq t \leqq 1$. $\alpha(x)_s$ is a path with initial point $p(H(x, s))$ and terminal point $f(x)$. The required homotopy H is given by

$$H(x, s) = \lambda[K(x, s)\alpha(x)_s](1) , \qquad 0 \leqq s \leqq 1 .$$

Bibliography

1. G. Allaud and E. Fadell, *A fiber homotopy extension theorem*, Trans. Amer. Math. Soc. **104** (1962), 239-251.
2. D. G. Bourgin, *Modern Algebraic Topology*, MacMillan (1963).
3. F. E. Browder, *On the fixed point index for continuous mappings of locally connected spaces*, Summa Brasil. Math. **4** (1960), 253-293.
4. E. Fadell, *Generalized normal bundless for locally flat imbeddings*, to appear in Trans. Amer. Math. Soc. **114** (1965), 488-513.
5. F. B. Fuller, *The homotopy theory of coincidences*, Ann. of Math. **59** (1954), 219-226.
6. ———, *The Homotopy Theory of Coincidences*, Princeton Thesis, 1951.
7. J. Kister, *Microbundles are bundles*, Ann. Math. **80**(1964), 190-199.
8. A. T. Lundell, *Obstruction theory of principal fiber bundles*, Trans. Amer. Math. Soc. **97** (1960), 161-192.
9. J. Milnor, *Lectures on Characteristic Classes* (Mimeographed Notes), Princeton, 1957.
10. D. Puppe, *Topologie II*, Universität Bonn, 1960.
11. F. Wecken, *Fixpunktklassen III*, Math. Annalen **118** (1941-43), 544-577.
12. J. Weier, *Fixpunkttheorie in topologischen Mannigfaltigkeiten*, Math. Zeit. **59** (1953), 171-190.

THE INSTITUTE FOR ADVANCED STUDY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

INCIDENCE MATRICES AND INTERVAL GRAPHS

D. R. FULKERSON AND O. A. GROSS

According to present genetic theory, the fine structure of genes consists of linearly ordered elements. A mutant gene is obtained by alteration of some connected portion of this structure. By examining data obtained from suitable experiments, it can be determined whether or not the blemished portions of two mutant genes intersect or not, and thus intersection data for a large number of mutants can be represented as an undirected graph. If this graph is an "interval graph," then the observed data is consistent with a linear model of the gene.

The problem of determining when a graph is an interval graph is a special case of the following problem concerning (0, 1)-matrices: When can the rows of such a matrix be permuted so as to make the 1's in each column appear consecutively? A complete theory is obtained for this latter problem, culminating in a decomposition theorem which leads to a rapid algorithm for deciding the question, and for constructing the desired permutation when one exists.

Let $A = (a_{ij})$ be an m by n matrix whose entries a_{ij} are all either 0 or 1. The matrix A may be regarded as the incidence matrix of elements $e_1, e_2, \cdots, e_m$ vs. sets $S_1, S_2, \cdots, S_n$; that is, $a_{ij} = 0$ or 1 according as e_i is not or is a member of S_j. For certain applications, one of which will be discussed below, it is of interest to know whether or not one can order the elements in such a way that each set S_j consists of elements that appear consecutively in the ordering. In terms of the incidence matrix A, the question is whether there is an m by m permutation matrix P such that the 1's in each column of PA occur in consecutive positions. We shall describe a computationally efficient method of answering this question, and of determining such a P when one exists.

Given a family of sets $S_1, S_2, \cdots, S_n$, one can form the *intersection graph* of the family by associating a vertex of the graph with each set and joining two distinct vertices with an edge if their corresponding sets have a nonempty intersection. Conversely, any finite graph can of course be viewed as the intersection graph of a family of sets (in many ways). If each set can be taken as an interval on the real line, the graph is called an *interval graph*. Interval graphs have been investigated in [7, 5, 3]. The problem posed above is closely related to that of determining whether a given graph is an interval

Received April 21, 1964.

graph. Necessary and sufficient conditions for this are known [7, 3]. But our problem appears to be more general, and our approach is quite different.

The study of interval graphs was stimulated in part by a biological application concerning the fine structure of genes [1, 7]. A basic problem, posed in [1], is to decide whether or not the subelements of genes are linked together in a linear order. A way of approaching this problem is also described in [1]. Briefly, it is as follows. For certain microorganisms, there are a standard form and mutants, the latter arising from the standard form by alteration of some connected part of the genetic structure. Experiments can be devised for determining whether the blemished parts of two mutant genes intersect or not. Thus the mathematical problem becomes: Given a large number of mutants together with information as to when the blemished portions of pairs of mutants intersect, to decide whether this information is compatible with a linear model of the gene or not. One example, analyzed in [1], shows intersection data for 145 mutants of a certain virus, for which it was found that a linear model was adequate.

In attacking the combinatorial problem posed at the beginning of this section, it does not suffice to consider just the intersection graph of the sets $S_1, S_2, \cdots, S_n$. Instead, we focus attention primarily on a more restricted graph, the *overlap graph* (§ 3), two sets being said here to overlap if they have a nonempty intersection which is properly included in each. The connected components of the overlap graph partition the m by n incidence matrix A into m-rowed submatrices $A_1, A_2, \cdots, A_p$. Thus we can take $A = (A_1, A_2, \cdots, A_p)$, where each A_k corresponds to a component of the overlap graph. Then our main theorem (Theorem 4.1) asserts that if there are permutations P_k, $k = 1, 2, \cdots, p$, such that P_kA_k has consecutive 1's in each column, there is a permutation P such that PA has this property also. This somewhat surprising result, coupled with the fact that one can describe a simple and direct construction (§ 5) for testing whether such P_k exist, provides an answer to the existence question for a general incidence matrix A.

2. A basic theorem. Let A be a (0, 1)-matrix. We say that A has the *consecutive* 1's *property* (for columns) provided there is a permutation matrix P such that the 1's in each column of PA occur consecutively. We note to begin with that the intersection graph of A does not contain enough information to decide whether A has the property or not. For example, the matrices

$$A_1 = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad A_2 = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

have the same intersection graph, a triangle, but A_2 does not have the consecutive 1's property. The first question that naturally comes up, then, is how much information about A is needed to decide whether it has the property or not. Do we need to know A itself, or will something less suffice? Theorem 2.1 below provides a partial answer to this question; it shows that a knowledge of the matrix A^TA is enough. Here A^T denotes the transpose of A.

THEOREM 2.1. *Let A and B be $(0, 1)$-matrices satisfying*

$$A^TA = B^TB . \tag{2.1}$$

Then either both A and B have the consecutive 1's property or neither does. Moreover, if A and B have the same number of rows and A has the consecutive 1's property, then there is a permutation P such that $B = PA$.

Proof. The first part of the theorem follows trivially from the second. For assume (2.1) and let A be m by n, B be m' by n, with $m > m'$. Then we may adjoin $m - m'$ rows of 0's to B, thereby obtaining an m by n matrix C satisfying $A^TA = C^TC$. The second assertion of the theorem then implies $C = PA$, for some permutation P. Consequently both A and C (hence both A and B) have the property or neither does.

Let A and B be m by n. The second part of the theorem is obviously valid for $n = 1$, and we proceed by induction on n. Suppress the first column vector a of A and the first column vector b of B, and call the resulting matrices A_1 and B_1. Thus

$$A = (a, A_1) , \tag{2.2}$$

$$B = (b, B_1) . \tag{2.3}$$

Clearly A_1 has the consecutive 1's property and $A_1^TA_1 = B_1^TB_1$. Hence, by the induction hypothesis, we can permute the rows of B to obtain a matrix $\tilde{B}$ satisfying

$$\tilde{B} = (\tilde{b}, A_1) , \tag{2.4}$$

the column vector $\tilde{b}$ being a permutation of b.

We shall finish the proof by showing that, corresponding to each row vector of $\tilde{B}$, there is an equal row vector of A. We begin by noting that if the column vectors a of (2.2) and $\tilde{b}$ of (2.4) have common 1's (common 0's), the corresponding rows of A and $\tilde{B}$ are equal and can be paired off against each other. Having done this, consider the submatrices of A and $\tilde{B}$ consisting of the remaining rows. Call these matrices A^* and B^*. Thus we may write

$$A^* = (a^*, A_1^*) , \tag{2.5}$$

$$B^* = (b^*, A_1^*) . \tag{2.6}$$

Note that A^* inherits the consecutive 1's property from A, since suppression of rows does not destroy the property, and hence we can permute the rows of A^* to obtain a matrix

$$\bar{A} = (\bar{a}, \bar{A}_1) \tag{2.7}$$

in which the 1's in each column occur consecutively. Applying the same permutation to B^* yields

$$\bar{B} = (\bar{b}, \bar{A}_1) . \tag{2.8}$$

Now $\bar{a}$ and $\bar{b}$ are complementary (0, 1)-vectors, that is, $\bar{b}$ is obtained from $\bar{a}$ by interchanging 0's and 1's. We also have

$$\bar{A}^T\bar{A} = \bar{B}^T\bar{B} , \tag{2.9}$$

since $\bar{A}$ and $\bar{B}$ are obtained from A and B, respectively, by suppressing certain equal rows and permuting rows. In particular, calculating the inner product of $\bar{a}$ with itself and $\bar{b}$ with itself, we have

$$\bar{a}\cdot\bar{a} = \bar{b}\cdot\bar{b} , \tag{2.10}$$

and hence $\bar{a}$ and $\bar{b}$ have the same number of 1's (hence also of 0's), so that the number of 1's in $\bar{a}$ is equal to the number of 0's in $\bar{a}$. Thus $\bar{A}$ and $\bar{B}$ have the following appearance:

$$\bar{A} = \left[\begin{array}{c|c} \begin{matrix}0\\ \vdots \\ 0\end{matrix} & E \\ \hline \begin{matrix}1\\ \vdots \\ 1\end{matrix} & F \\ \hline \begin{matrix}0\\ \vdots \\ 0\end{matrix} & G \end{array}\right] , \qquad \bar{B} = \left[\begin{array}{c|c} \begin{matrix}1\\ \vdots \\ 1\end{matrix} & E \\ \hline \begin{matrix}0\\ \vdots \\ 0\end{matrix} & F \\ \hline \begin{matrix}1\\ \vdots \\ 1\end{matrix} & G \end{array}\right] . \tag{2.11}$$

The matrices E and G in (2.11) have together the same number of rows that F does, and we shall show that the rows of $\bar{A}$ corresponding to F (corresponding to E and G) can be paired with equal rows of $\bar{B}$ corresponding to E and G (corresponding to F).

To this end, let E and G have k and l rows, respectively, and let R_p denote the p by p permutation matrix which reverses order,

$$R_p = \begin{bmatrix} & & 1 \\ & \iddots & \\ 1 & & \end{bmatrix}. \tag{2.12}$$

We assert that

$$F = \begin{bmatrix} R_k E \\ R_l G \end{bmatrix}. \tag{2.13}$$

For consider an arbitrary column vector c of $\bar{A}_1$ and write

$$c = \begin{bmatrix} e \\ f \\ g \end{bmatrix}, \tag{2.14}$$

corresponding to the partition (2.11) of $\bar{A}_1$. From (2.9) we have

$$\bar{a} \cdot c = \bar{b} \cdot c ,$$

and thus the number of 1's in f is equal to the number of 1's in e and g. It follows from this and the fact that the 1's in c occur consecutively that

$$f = \begin{bmatrix} R_k e \\ R_l g \end{bmatrix}. \tag{2.15}$$

This establishes (2.13) and finishes the proof of Theorem 2.1.

We note the following corollary of Theorem 2.1.

COROLLARY 2.2. *Let A and B be (0, 1)-matrices satisfying* (2.1). *If A has the consecutive* 1's *property and has no rows of* 0's, *then there is a permutation P such that*

$$PB = \begin{bmatrix} A \\ 0 \end{bmatrix}. \tag{2.16}$$

Proof. Let A be m by n and let B be m' by n. Suppose $m' < m$. Then we may adjoin $m - m'$ rows of 0's to B, obtaining a matrix C satisfying $A^T A = C^T C$. By Theorem 2.1, C is a row permutation of A, violating the assumption that A has no zero rows. Consequently $m' \geqq m$. The conclusion now follows by adjoining $m' - m$ rows of 0's to A.

It is of course not true in general that equation (2.1) implies $B = PA$ for m by n (0, 1)-matrices A and B. A simple example is

$$A = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

It would be interesting to know conditions on A^TA in order that A have the consecutive 1's property. Although we have not been able to give a complete answer to this question in the sense of finding some nice set of necessary and sufficient conditions, we shall establish a theorem in § 4 which reduces the question to the consideration of (0, 1)-matrices having connected overlap graphs. Section 5 then provides a simple construction for testing whether or not such a matrix has the property.

3. The overlap graph and component graph. We rephrase the definition of "overlap" in terms of inner products of (0, 1)-vectors. Let a and b be (0, 1)-vectors having m components. Their inner product $a \cdot b$ satisfies

$$0 \leqq a \cdot b \leqq \min (a \cdot a, b \cdot b) . \tag{3.1}$$

If strict inequality holds throughout (3.1), that is, if $a \cdot b$ satisfies

$$0 < a \cdot b < \min (a \cdot a, b \cdot b) , \tag{3.2}$$

we say that a and b *overlap*. We also say that a and b are *disjoint* if

$$a \cdot b = 0 , \tag{3.3}$$

and that a *contains* b if

$$a \cdot b = b \cdot b . \tag{3.4}$$

Now let A be an m by n (0, 1)-matrix having column vectors $a_1, a_2, \cdots, a_n$. It is convenient, and imposes no loss of generality in studying the consecutive 1's property, to assume that $a_j \neq 0$, $j = 1, 2, \cdots, n$, and that $a_i \neq a_j$ for $i \neq j$. Henceforth we frequently make these assumptions and refer to such an A as *proper*.

There are various graphs one can associate with a (0, 1)-matrix A that are meaningful insofar as the consecutive 1's property is concerned. For instance, we can take vertices $x_1, x_2, \cdots, x_n$ corresponding to the columns $a_1, a_2, \cdots, a_n$ of A, and put in the following edges, some being directed, others undirected: an edge $[x_i, x_j]$ directed from x_i to x_j if a_i contains a_j; an undirected edge (x_i, x_j) joining x_i and x_j if a_i and a_j overlap. (If we go on to make this a "weighted" graph by associating with each edge (vertex) the appropriate inner product, then Theorem 2.1 shows that we have included sufficient information to decide whether A has the consecutive 1's property). We shall not, however, deal primarily with this full graph $\mathscr{F} = \mathscr{F}(A)$, but shall instead work mostly with two graphs derived from it, one of these being an undirected graph, the other a directed graph.

The first of these is the graph obtained from $\mathscr{F}$ by including only

the undirected edges (x_i, x_j) corresponding to overlapping column vectors a_i and a_j. We call this the *overlap graph* of A and denote it by $\mathscr{G} = \mathscr{G}(A)$.

The second of these graphs is obtained from $\mathscr{F}$ as follows: Let $X_1, X_2, \cdots, X_p$ be the connected components of $\mathscr{G}(A)$, considered as vertices of a graph $\mathscr{D} = \mathscr{D}(A)$. Direct an edge $[X_i, X_j]$ from vertex X_i to vertex X_j if there is an x in X_i and a y in X_j such that $[x, y]$ is an edge from x to y in $\mathscr{F}$. We call $\mathscr{D}$ the *component graph* of A. The component graph $\mathscr{D}$ is obtained by condensing in $\mathscr{F}$ the vertices of a connected component of $\mathscr{G}$ to a single vertex.

We give an example illustrating these concepts. Let

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

Thus

$$A^TA = \begin{bmatrix} 6 & 5 & 5 & 0 & 0 & 2 & 2 & 3 \\ 5 & 8 & 5 & 2 & 2 & 2 & 2 & 3 \\ 5 & 5 & 5 & 0 & 0 & 2 & 2 & 3 \\ 0 & 2 & 0 & 2 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 1 & 2 & 0 & 0 & 0 \\ 2 & 2 & 2 & 0 & 0 & 2 & 1 & 1 \\ 2 & 2 & 2 & 0 & 0 & 1 & 2 & 0 \\ 3 & 3 & 3 & 0 & 0 & 1 & 0 & 3 \end{bmatrix}.$$

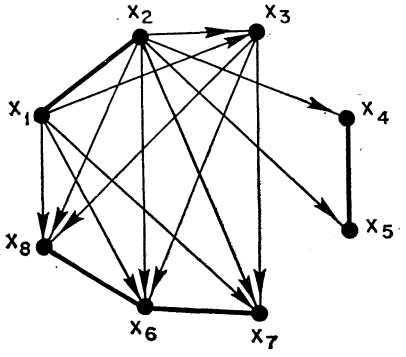


Fig. 3.1. The full graph $\mathscr{F}(A)$

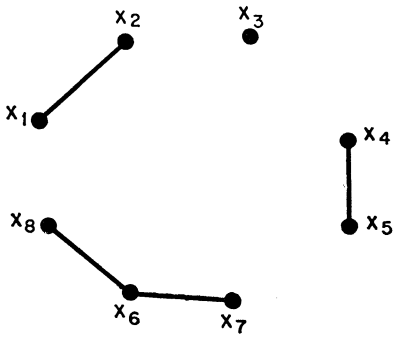


Fig. 3.2. The overlap graph $\mathscr{G}(A)$

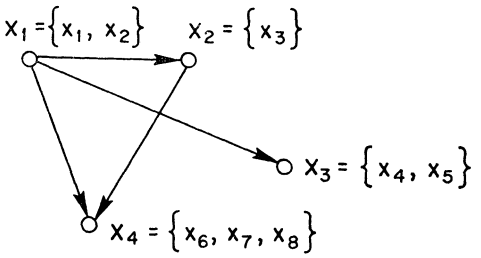


Fig. 3.3. The component graph $\mathscr{D}(A)$

The graphs $\mathscr{F}$, $\mathscr{G}$, and $\mathscr{D}$ for A are shown in Figs. 3.1, 3.2, and 3.3, respectively.

Notice in the example that the component graph $\mathscr{D}$ is acyclic (has no directed cycles) and transitive (if $[X, Y]$ and $[Y, Z]$ are edges, then $[X, Z]$ is an edge). We now prove that this is generally so.

THEOREM 3.1. *The component graph $\mathscr{D}(A)$ of a proper (0, 1)-matrix A is acyclic and transitive.*

Proof. We first show that $\mathscr{D}$ is transitive. Let X, Y, Z be components of $\mathscr{G}$ (vertices of $\mathscr{D}$) such that $[X, Y]$ and $[Y, Z]$ are edges in $\mathscr{D}$. Hence there are vertices x, y, y', z of $\mathscr{F}$ with x in X, y and y' in Y, and z in Z, such that $[x, y]$ and $[y', z]$ are directed edges of $\mathscr{F}$. Moreover, since y and y' are in the same component of $\mathscr{G}$, there is a chain $(y, y_1), (y_1, y_2), \cdots, (y_k, y')$ of undirected edges joining y to y' in $\mathscr{G}$ (hence in $\mathscr{F}$). Thus the matrix A has corresponding columns a, b, b', c and $b_1, b_2, \cdots, b_k$ such that a contains b, b' contains c, b and b_1 overlap, b_1 and b_2 overlap, $\cdots$, b_k and b' overlap. But if a contains b and b overlaps b_1, then either a contains b_1 or a overlaps b_1. The latter of these alternatives is impossible, for otherwise x and y_1 would be in the same component of the overlap graph, contrary to choice. Thus a contains b_1. We may now repeat the argument with a, b_1, b_2 in place of a, b, b_1 to show that a contains b_2, and so on, finally deducing that a contains b'. Since b' contains c, we see that a contains c, and hence $[x, z]$ is a directed edge in $\mathscr{F}$. Consequently $[X, Z]$ is a directed edge in $\mathscr{D}$, and $\mathscr{D}$ is transitive.

To show that $\mathscr{D}$ is acyclic, it suffices, since $\mathscr{D}$ is transitive, to show that both $[X, Y]$ and $[Y, X]$ cannot be edges of $\mathscr{D}$. Suppose they were edges of $\mathscr{D}$. Then, as above, the matrix A would have columns $a_1, a_2, \cdots, a_k$ and $b_1, b_2, \cdots, b_l$ with the following relations holding: a_1 contains b_1; b_l contains a_k; a_1 overlaps $a_2, \cdots, a_{k-1}$ overlaps a_k; b_1 overlaps $b_2, \cdots, b_{l-1}$ overlaps b_l; and no a_i overlaps any b_j. It then follows as above that a_1 contains b_l. Similarly we deduce that b_l contains a_1. Thus $a_1 = b_l$. Since a_1 and b_l are necessarily distinct

columns of A, this violates our assumption that A has no pair of equal columns. Consequently $\mathscr{D}$ is acyclic. This completes the proof of Theorem 3.1.

Theorem 3.1 shows that $\mathscr{D}$ is the graph of a partial ordering. This partial ordering is special in the sense that an element can have at most one immediate predecessor, as Theorem 3.2 below shows. Consequently, if we omit from $\mathscr{D}$ every edge whose existence is implied by transitivity, the resulting graph is simply a collection of rooted trees. (For example, in Fig. 3.3, omitting the edge $[X_1, X_4]$ produces a single tree with root X_1).

THEOREM 3.2. *Each vertex of the component graph $\mathscr{D}(A)$ of a proper (0, 1)-matrix A has at most one immediate predecessor.*

Proof. It suffices to show that if $[X, Z]$ and $[Y, Z]$ are edges of $\mathscr{D}$, then either $[X, Y]$ or $[Y, X]$ is an edge. Thus, let $[X, Z]$ and $[Y, Z]$ be edges of $\mathscr{D}$. It follows that A has columns $a, b, c_1, c_2, \cdots, c_k$ with the following relations holding: a contains c_1, b contains c_k, and successive pairs of c's overlap. Thus a contains c_k. Since b also contains c_k and a and b do not overlap, then either a contains b or b contains a. Consequently, either $[X, Y]$ or $[Y, X]$ is an edge of $\mathscr{D}$.

In § 4 we shall use the structure of the component graph to prove the decomposition theorem mentioned in § 1. This structure will also be exploited later in developing a complete algorithm for arranging the rows of a matrix A to make its 1's appear consecutively in each column, when this is possible. For these purposes, we note here another fact about the full graph $\mathscr{F}(A)$ that has not been stated explicitly, although its proof is contained in that of Theorem 3.1.

THEOREM 3.3. *Let X and Y be components of the overlap graph $\mathscr{G}(A)$ of a proper (0, 1)-matrix A, such that there is an x in X and y in Y with $[x, y]$ an edge of $\mathscr{F}(A)$. Then $[x, y']$ is an edge of $\mathscr{F}(A)$ for arbitrary y' in Y.*

Theorem 3.3 shows that each column of A corresponding to Y plays exactly the same role with respect to columns of A corresponding to X. That is, each "Y-column" will be contained in certain of the "X-columns" and disjoint from the others. This pattern is the same for every Y-column.

4. The decomposition theorem. For an arbitrary (0, 1)-matrix A, we can rearrange columns if necessary and write

$$A = (A_1, A_2, \cdots, A_p) , \tag{4.1}$$

where each submatrix A_k, $k = 1, 2, \cdots, p$, corresponds to a component X_k of the overlap graph $\mathscr{G}(A)$. We term (4.1) an *overlap decomposition* of A. It is of course unique, apart from the ordering of the submatrices. We refer to these submatrices as *overlap components* of A, or briefly, *components* of A. If A has just one component, we say that A is *connected*.

THEOREM 4.1. *A (0, 1)-matrix A has the consecutive 1's property if and only if each of its components has the property.*

Proof. The necessity is obvious. To prove sufficiency, we may assume that A is proper and proceed by induction on the number p of components of A, the case $p = 1$ being trivial.

Before going to the induction step, we first remark that if a matrix has the consecutive 1's property and has equal rows, such rows can be brought together in a permuted form which makes the 1's in each column consecutive. To see this, it suffices to observe two things. First, if a row is deleted from a matrix having the consecutive 1's property, the resulting matrix has the property. On the other hand, if a deleted row is one of a pair of equal rows, then it can be reinserted adjacent to its counterpart in the permuted matrix, and the 1's in each column are still consecutive.

To establish the induction step, let A be a proper matrix having p components, each of which has the consecutive 1's property, and write

$$A = (A_1, A_2, \cdots, A_p) , \tag{4.2}$$

where A_p corresponds to a minimal element in the partial ordering given by $\mathscr{D}(A)$. Now delete A_p from A. By the induction assumption, the matrix

$$A' = (A_1, A_2, \cdots, A_{p-1}) \tag{4.3}$$

has the consecutive 1's property, and hence, by the above remark, there is a permutation P such that PA' has consecutive 1's in each column and like rows of PA' occur together. Consider

$$PA = (B_1, B_2, \cdots, B_p) = B . \tag{4.4}$$

We may select a topmost and bottommost 1 in B_p, and write

$$B = \begin{bmatrix} * & * & \cdots & * & 0 \\ \bar{B}_1 & \bar{B}_2 & \cdots & \bar{B}_{p-1} & \bar{B}_p \\ * & * & \cdots & * & 0 \end{bmatrix} , \tag{4.5}$$

where the first and last rows of $\bar{B}_p$ each contain at least one 1. Now consider the matrix

$$(\bar{B}_1, \bar{B}_2, \cdots, \bar{B}_{p-1}) . \tag{4.6}$$

We assert that all rows of (4.6) are identical. To see this, observe first that since X_p is a minimal element in the partial ordering, each X_k, $k = 1, 2, \cdots, p-1$, either dominates X_p (there is an edge from X_k to X_p in $\mathscr{D}(A)$) or is unrelated to X_p (there is no edge from X_k to X_p or from X_p to X_k). Suppose X_k is unrelated to X_p. Then, since the first and last rows of $\bar{B}_p$ contain a 1, it follows that the first and last rows of $\bar{B}_k$ are all 0's, for otherwise there would be a column of A_p and a column of A_k which are not disjoint. On the other hand, suppose X_k dominates X_p. It then follows from Theorem 3.3 that the first and last rows of $\bar{B}_k$ are equal. Hence the first and last rows of (4.6) are equal, and consequently, by the selection of P, all rows of (4.6) are equal.

We may now permute the rows of $\bar{B}_p$ to make its 1's consecutive in each column. Such a permutation merely shuffles like rows in (4.5), and thus A has the consecutive 1's property.

Any component of A which has no more than two columns obviously has the consecutive 1's property. We may thus state the following corollary of Theorem 4.1.

COROLLARY 4.2. *If each component of a* $(0, 1)$*-matrix* A *has at most two columns, then* A *has the consecutive* 1*'s property.*

5. Testing a connected matrix for the consecutive 1's property. In this section we shall describe a construction for deciding whether a connected matrix A has the consecutive 1's property. The essential idea is not to search explicitly for a permutation P which rearranges A, but rather to attempt to build up a permuted form of A, column by column, imposing necessary inner-product requirements on the columns of the configuration.

Let a, b, c be distinct column vectors of A such that a overlaps b and b overlaps c. If A has the consecutive 1's property, then there is a row permutation of A which contains the configuration C_1 or C_2 described below (and illustrated in Figs. 5.1 and 5.2) according as

$$a \cdot c < \min (a \cdot b, b \cdot c) \tag{5.1}$$

or

$$a \cdot c \geqq \min (a \cdot b, b \cdot c) \tag{5.2}$$

holds. The configuration C_1 is obtained by writing down $a \cdot a$ 1's con-

secutively in a column, then $b \cdot b$ consecutive 1's in a second column so that these two strings of 1's overlap (at the bottom of the first, say) on $a \cdot b$ 1's, then a third string of $c \cdot c$ 1's, overlapping the second string in the same way that the second overlapped the first (at its bottom) on $b \cdot c$ 1's. The configuration C_2 differs from C_1 only by making the third string of 1's overlap the second opposite to the manner in which the second overlapped the first.

a	b	c
1		
1	1	
1	1	
1	1	
1	1	1
	1	1
	1	1
		1

Fig. 5.1. The Configuration C_1

a	b	c
1		1
1	1	1
1	1	1
1	1	1
1	1	
	1	
	1	

Fig. 5.2. The Configuration C_2

Note that while we have two choices in placing the second string of 1's relative to the first (top or bottom of the first), it is immaterial which of these we select. Moreover, having selected one, there is no further latitude in positioning the third string of 1's, since its position relative to the second is determined by (5.1) or (5.2) and the position of the second relative to the first. After constructing the configuration C_1 or C_2, we can then check the inner product of its first and third columns. If this is not equal to $a \cdot c$, we know immediately that A does not have the consecutive 1's property.

Three column vectors of A which can be ordered so that the first and second overlap and the second and third overlap will be termed a *rigid triple* of A. Our construction for a connected A will be based on singling out a spanning subtree of $\mathscr{G}(A)$ and then positioning strings of 1's, using rigid triples obtained from this tree.

Instead of describing the computation formally, we shall consider an example which will make the process clear. Let

$$A = \begin{array}{c} \begin{matrix} ① & ② & ③ & ④ & ⑤ & ⑥ \end{matrix} \\ \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{bmatrix} \end{array}.$$

A spanning subtree of the overlap graph $\mathscr{G}(A)$ is shown in Fig. 5.3 below.

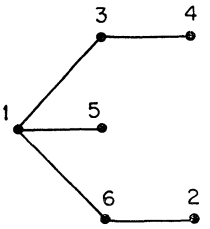


Fig. 5.3. A Spanning Subtree of $\mathscr{G}(A)$

On the first step, we may position the rigid triple 1, 3, 5 to obtain the configuration of Fig. 5.4 below.

①	③	⑤
		1
1		1
1		1
1	1	1
1	1	
	1	

Fig. 5.4

The inner products of this configuration agree with the corresponding ones of A, and we proceed to add column 6, positioning it relative to 1 and 5 (5, 1, 6 form a rigid triple), to obtain the configuration shown in Fig. 5.5.

①	③	⑤	⑥
		1	
1		1	
1		1	
1	1	1	
1	1		1
	1		1

Fig. 5.5

Checking the inner product of column 6 with all others in the configuration shows that we have no contradiction yet, and so we add column 4, positioning it relative to 1 and 3, to obtain the configuration of Fig. 5.6.

①	③	⑤	⑥	④
		1		
1		1		
1		1		1
1	1	1		1
1	1		1	1
	1		1	

Fig. 5.6

Again all inner products of column 4 with preceding columns of the configuration check out properly. We then add column 2, positioning it relative to 1 and 6, obtaining the configuration of Fig. 5.7.

①	③	⑤	⑥	④	②
		1			
1		1			
1		1		1	
1	1	1		1	1
1	1		1	1	1
	1		1		

Fig. 5.7

Since the inner products of column 2 with all preceding columns in the configuration of Fig. 5.7 agree with the corresponding inner products of columns of A, Corollary 2.2 shown that the matrix

$$B = \begin{array}{c} \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \end{array},$$

obtained from Fig. 5.7 by rearranging columns and filling in 0's so that B has no zero rows, is a row permutation of A.

A complete test for a general matrix A is now clear. First determine the connected components of A. This can be easily done in such a way that corresponding spanning subtrees are automatically obtained. Simply select an arbitrary column and find all columns it overlaps; then select one of these and determine any new columns it overlaps, and so on. When no new columns can be determined, a spanning subtree for

one component has been found. The configuration building process described above can then be applied to each component. (In actual computation, one would build the configuration simultaneously with the determination of a connected component.)

If one measures the efficiency of this method by calculating an upper bound on the number of times that it is necessary to compute (or look up) the inner product of two (0, 1)-vectors, it is not difficult to see that if A has n columns, such a bound is of the order $O(n^2)$.

6. Fitting the components together. If no contradiction in inner products is encountered in the construction outlined in the preceding section, one can go on to find the desired row permutation of A by fitting the various connected submatrices together in an appropriate way, using the partial ordering of components given by $\mathscr{D}(A)$. (We assume that A is proper.) The proof of Theorem 4.1 indicates how to proceed. Again we shall not describe the process in detail, but shall illustrate it with the example of § 3.

The configuration-building procedure of § 5 produces the matrices

$$B_1 = \begin{array}{c} \begin{matrix} ① & ② \end{matrix} \\ \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 1 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \end{array}, \quad B_2 = \begin{array}{c} \begin{matrix} ③ \end{matrix} \\ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \end{array}, \quad B_3 = \begin{array}{c} \begin{matrix} ④ & ⑤ \end{matrix} \\ \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{bmatrix} \end{array}, \quad B_4 = \begin{array}{c} \begin{matrix} ⑥ & ⑦ & ⑧ \end{matrix} \\ \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \end{array}.$$

To fit these together, we first determine the maximal elements in $\mathscr{D}(A)$. Here there is just one, corresponding to B_1, and we have:

$$\begin{array}{c} \begin{matrix} ① & ② & ③ & ④ & ⑤ & ⑥ & ⑦ & ⑧ \end{matrix} \\ \left[\begin{array}{cc|cccccc} 1 & 0 & & & & & & \\ 1 & 1 & & & & & & \\ 1 & 1 & & & & & & \\ 1 & 1 & & & & & & \\ 1 & 1 & & & * & & & \\ 1 & 1 & & & & & & \\ 0 & 1 & & & & & & \\ 0 & 1 & & & & & & \\ 0 & 1 & & & & & & \end{array}\right] \end{array}.$$

(If there were more than one maximal element, the determined part of the configuration at this point would appear as a direct sum.)

Next suppress the maximal elements in $\mathscr{D}(A)$ and look for the new maximal elements in the reduced graph. Here there are two, corresponding to B_2 and B_3. By Theorem 3.3, each column of $B_2(B_3)$ is contained in certain columns of B_1 and disjoint from the remaining, and this pattern is the same for all columns of $B_2(B_3)$. Hence it suffices to test one column of $B_2(B_3)$ to find this pattern. We can do this by referring to A^TA or $\mathscr{T}(A)$. The single column of B_2 is contained in both columns of B_1, whereas columns of B_3 are disjoint from the first column of B_1 and contained in the second. We now fit in B_2 by sliding its nonzero part (the part between a topmost and bottommost 1 of B_2) to the top of the consecutive group of equal rows of B_1 consisting of (1, 1)-pairs. Similarly, we fit in B_3 by sliding its nonzero part to the top of the consecutive group of (0, 1)-rows of B_1. This gives the configuration

①	②	③	④	⑤	⑥	⑦	⑧
1	0	0	0	0			
1	1	1	0	0			
1	1	1	0	0			
1	1	1	0	0			
1	1	1	0	0		*	
1	1	1	0	0			
0	1	0	1	0			
0	1	0	1	1			
0	1	0	0	1			

.

The new reduced component graph consists of the single element corresponding to B_4, and we see from $\mathscr{T}(A)$ that each column of B_4 is contained in columns 1, 2, 3 and is disjoint from columns 4 and 5. Consequently we slide the nonzero part of B_4 to appear at the top of the consecutive group of (1, 1, 1, 0, 0)-rows in the configuration. This yields a row permutation of A whose 1's appear consecutively in each column:

①	②	③	④	⑤	⑥	⑦	⑧
1	0	0	0	0	0	0	0
1	1	1	0	0	0	1	0
1	1	1	0	0	1	1	0
1	1	1	0	0	1	0	1
1	1	1	0	0	0	0	1
1	1	1	0	0	0	0	1
0	1	0	1	0	0	0	0
0	1	0	1	1	0	0	0
0	1	0	0	1	0	0	0

.

Note that at each stage in solving the example, we automatically had equal rows occurring together (see the proof of Theorem 4.1) and hence did not need to permute rows of the configuration before fitting in another piece. It is not difficult to show inductively that this will always be the case, provided the prescription of sliding the nonzero part of the new piece to the top of the group of like rows which precede it is followed. A relevant fact in making such a proof is that the nonzero part of the new piece contains no zero rows. This follows from connectedness and consecutivity of 1's, and justifies our terminology.

7. **Application to interval graphs.** The methods developed in preceding sections can be applied to the problem of determining when a graph is an interval graph. As noted in § 1, various sets of necessary and sufficient conditions that a graph be an interval graph are known. Moreover, computational methods based on some of these have been described [7, 3]. The procedure we shall outline appears to be more efficient. In addition, a specific representation in terms of intervals can be easily produced if desired.

We begin by reviewing certain concepts and results that will be relevant. The first of these is that of a rigid circuit graph [2, 7]. An undirected graph[1] $\mathscr{G}$ is a *rigid circuit graph* if every circuit of $\mathscr{G}$ with more than three vertices has a chord. Here a chord is an edge not in the circuit which joins two vertices of the circuit. It is easy to see that an interval graph must be a rigid circuit graph [7]. We also recall the following basic result about rigid circuit graphs [2]: Every rigid circuit graph $\mathscr{G}$ has a simplicial vertex. Here a vertex s of $\mathscr{G}$ is *simplicial* if all the neighbors of s in $\mathscr{G}$ form a simplex or clique in $\mathscr{G}$ (i.e,, each pair of neighbors of s is joined by an edge.) Thus s together with its neighbors is also a clique in $\mathscr{G}$.

Rigid circuit graphs have been characterized in various ways [2, 7]. For our purposes, the following procedure for testing the rigid circuit property will be appropriate:[2] Search for a simplicial vertex in $\mathscr{G}$; if one is found, supress it and repeat the procedure in the reduced graph. It follows that $\mathscr{G}$ is a rigid circuit graph if and only if this process terminates in the deletion of all vertices of $\mathscr{G}$. For if $\mathscr{G}$ is a rigid circuit graph, then $\mathscr{G}$ contains a simplicial vertex, and deletion of vertices maintains the rigid circuit property. Conversely if $\mathscr{G}$ has a circuit with more than three vertices which has no chord, then no

[1] All graphs considered in this section are finite and undirected, and have no multiple edges or loops.

[2] This simple test has apparently not been noted before. See [7], where a more complicated test is used.

vertex of this circuit can be deleted in the process.

Figures 7.1 and 7.2 show examples of rigid circuit graphs, the vertices being numbered in such a way that successively suppressing the next vertex in the numbering fulfills the test.

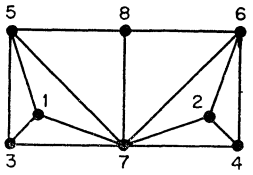


Fig. 7.1

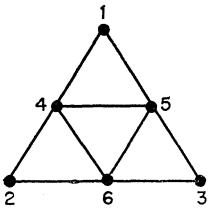


Fig. 7.2

We shall return to rigid circuit graphs in a moment, but at this point we need to introduce a certain incidence matrix which can be associated with an arbitrary graph $\mathscr{G}$. We call this matrix the *dominant clique* vs. *vertex* matrix. As will be obvious, it specifies $\mathscr{G}$ completely. Let $\mathscr{G}$ be any graph. We can consider the family of all sets of vertices which form cliques in $\mathscr{G}$ and partially order these by set inclusion. The maximal elements in this partial ordering will be termed the *dominant cliques* of $\mathscr{G}$. Since two vertices of $\mathscr{G}$ are joined by an edge if and only if they belong to some dominant clique, the dominant clique vs. vertex incidence matrix characterizes $\mathscr{G}$.

Such incidence matrices for the graphs of Figs. 7.1 and 7.2 are shown in Figs. 7.3 and 7.4, respectively.

1	2	3	4	5	6	7	8
1	0	1	0	1	0	1	0
0	1	0	1	0	1	1	0
0	0	0	0	1	0	1	1
0	0	0	0	0	1	1	1

Fig. 7.3

1	2	3	4	5	6
1	0	0	1	1	0
0	1	0	1	0	1
0	0	1	0	1	1
0	0	0	1	1	1

Fig. 7.4

For rigid circuit graphs, the determination of the dominant cliques can be carried out in conjunction with the test for the rigid circuit property described above. Upon deleting a simplicial vertex, simply list it together with its neighbors in the reduced graph. If the original graph $\mathscr{G}$ has n vertices, this yields a set of n cliques. The dominant ones of these n cliques are the dominant cliques of $\mathscr{G}$. For these are certainly cliques of $\mathscr{G}$. Moreover, if $\mathscr{C}$ is an arbitrary clique in $\mathscr{G}$, consider the first vertex of $\mathscr{C}$ which is deleted in the process. At this stage we list a clique $\mathscr{C}'$ that includes $\mathscr{C}$.

The theory and methods we have developed for studying the consecutive 1's property can now be applied to interval graphs via Theorem 7.1 below.

THEOREM 7.1. *A graph $\mathscr{G}$ is an interval graph if and only if the dominant clique vs. vertex incidence matrix of $\mathscr{G}$ has the consecutive* 1*'s property.*

Proof. Let D be the dominant clique vs. vertex incidence matrix of $\mathscr{G}$, so arranged that the 1's in each column occur consecutively. Suppose the first and last 1's in the j-th column of D occur in positions f_j and l_j, respectively. Then $\mathscr{G}$ is the intersection graph of the intervals $[f_j - 1, l_j]$, $j = 1, 2, \cdots, n$.

Conversely, let $\mathscr{G}$ be the intersection graph of a set of n (closed) intervals I_j, $j = 1, 2, \cdots, n$. Let the distinct endpoints $e_1, e_2, \cdots, e_m$ of these intervals be ordered so that $e_1 < e_2 < \cdots < e_m$. Then the m by n incidence matrix $A = (a_{ij})$, defined by setting $a_{ij} = 1$ or 0 according as e_i is or is not in I_j, has consecutive 1's in each column. Moreover, D is obtained from A by deleting rows. Hence D has the consecutive 1's property.

The incidence matrix of Fig. 7.3 has the consecutive 1's property; that of Fig. 7.4 does not. (A rearranged form of Fig. 7.3 is shown in Fig. 7.5). Consequently the graph of Fig. 7.1 is an interval graph (its "intervals" being displayed in Fig. 7.5), but that of Fig. 7.2 is not.

$$\begin{array}{cccccccc} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \left[\begin{array}{cccccccc} 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 & 1 & 0 \end{array}\right] \end{array}$$

Fig. 7.5

8. **Consecutive 1's and total unimodularity.** A (0, 1)-matrix A is said to be totally unimodular [6] if each of its square submatrices has determinant ± 1 or 0. It follows from known sufficient conditions for the total unimodularity property that if A has the consecutive 1's property, then A is totally unimodular [6]. This fact can also be proved directly without difficulty by induction on the number of rows in a square matrix having the consecutive 1's property. The total unimodularity property has significant implications in linear inequality theory. In particular, if A is totally unimodular, and if b is a vector having integral components, then the convex polyhedron defined by the linear inequalities

$$\begin{aligned} Au &\geqq b\,, \\ u &\geqq 0 \end{aligned} \tag{8.1}$$

has all integral extreme points. That is, if u is an extreme solution

of (8.1), then u has integral components. Consequently, the linear program

$$\begin{aligned} Au &\geqq b\ , \\ u &\geqq 0\ , \\ &\text{minimize } c\cdot u \end{aligned} \tag{8.2}$$

always has integral solution vectors u, provided A is totally unimodular and b is integral. In (8.2), c is a given real vector.

We can apply these facts and Theorem 7.1 to derive a certain combinatorial duality theorem for interval graphs which, though similar in some respects to known results about rigid circuit graphs [2, 4], does not hold for the latter. To state this theorem, we first require a definition. We shall say that a set of vertices in a graph $\mathscr{G}$ *represents* all dominant cliques of $\mathscr{G}$ provided each dominant clique contains some vertex of the set. Our interest is in a minimum representation, that is, a set of vertices which represents all dominant cliques and has minimum cardinality over all such sets.

THEOREM 8.1. *Let $\mathscr{G}$ be an interval graph. Then the minimum number of vertices of $\mathscr{G}$ required to represent all dominant cliques is equal to the maximum number of dominant cliques that are mutually disjoint.*

Proof. Let D be the dominant clique vs. vertex incidence matrix of $\mathscr{G}$. In view of Theorem 7.1 and the preceding discussion, the problem of finding a minimum representation of dominant cliques can be posed as that of solving the linear program

$$\begin{aligned} Du &\geqq 1\ , \\ u &\geqq 0\ , \\ &\text{minimize } 1\cdot u\ . \end{aligned} \tag{8.3}$$

Here 1 is a vector all of whose components are unity. The dual of (8.3) is

$$\begin{aligned} D^T w &\leqq 1\ , \\ w &\geqq 0\ , \\ &\text{maximize } 1\cdot w\ . \end{aligned} \tag{8.4}$$

Since D is totally unimodular, so is D^T, and thus the program (8.4) also has (0, 1)-solution vectors w. Consequently (8.4) asks for the maximum number of dominant cliques that are pairwise disjoint. Thus Theorem 8.1 follows from the duality theorem for linear inequalities.

It can be seen similarly that if $\mathscr{G}$ is an interval graph, then the

maximum number of vertices, no two of which are in the same dominant clique (i.e., the maximum number of vertices, no two of which are joined by an edge) is equal to the minimum number of dominant cliques that cover all vertices (i.e., the minimum number of cliques that cover all vertices). This theorem is in fact known to be true more generally for rigid circuit graphs [2, 4]. But Theorem 8.1 is false for rigid circuit graphs. The graph of Fig. 7.2 is a case in point, since the minimum number of vertices required to represent all dominant cliques is two, whereas the maximum number of mutually disjoint dominant cliques is one.

REFERENCES

1. S. Benzer, *On the topology of the genetic fine structure,* Proc. Nat. Acad. Sci. U.S.A. **45** (1959), 1607.

2. G. A. Dirac, *On rigid circuit graphs,* Abhandlungen Mathematischen Seminar Universität Hamburg **25** (1961), 72.

3. P. C. Gilmore and A. J. Hoffman, *A characterization of comparability graphs and of interval graphs,* in Can. J. Math. **16** (1964), 539.

4. A. Hajnal and J. Suranyi, *Über die Auflösung von Graphen in vollständige Teilgraphen,* Ann. Univ. Sci. Budapest, Sect. Math. **1** (1958), 113.

5. G. Hajos, *Über eine Art von Graphen,* Internationale Math. Nachrichten, **11** (1957), 65.

6. A. J. Hoffman and J. B. Kruskal, *Integral boundary points of convex polyhedra,* Annals of Math. Study 38, Princeton University Press, Princeton, New Jersey, (1956), 223.

7. C. G. Lekkerkerker and J. Ch. Boland, *Representation of a finite graph by a set of intervals on the real line,* Fund. Math. Polska Akad. Nauk, **51** (1962), 45.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

TENSOR PRODUCTS OVER H*-ALGEBRAS

Larry C. Grove

Throughout, A, B, and C denote (semi-simple) H^*-algebras whose respective decompositions into minimal closed ideals are $A = \Sigma \oplus A_\alpha$, $B = \Sigma \oplus B_\beta$, and $C = \Sigma \oplus C_\gamma$. It is assumed that A is a right C-module and B is a left C-module. We define a tensor product $A \otimes_C B$ that is again an H^*-algebra, and show that it is isometric and isomorphic with an ideal in $A \otimes B \otimes C$. As a corollary, $A \otimes_C B$ is strongly semi-simple if A, B, and C are each strongly semi-simple. The converse to the corollary is shown to be false. When A, B, and C are closed ideals in some H^*-algebra, with ordinary multiplication as the module action, then $A \otimes_C B$ is shown to be isomorphic with the direct sum of all the one-dimensional ideals in $A \cap B \cap C$. When $A = L^2(G)$, $B = L^2(H)$, and $C = L^2(K)$, for suitable related compact groups G, H, and K, then the module actions are defined, and $A \otimes_C B$ can be constructed. When $G = H = K$, it is shown that $A \otimes_C B \cong L^2(G/N)$, where N is the closure of the commutator subgroup of G. A conjecture is stated that would generalize this result to the case where K is a closed subgroup of $G \cap H$.

Since $A \otimes_C B$ will be represented in terms of ordinary tensor products $A \otimes B$ of H^*-algebras, the requisite facts concerning $A \otimes B$ are stated here (details may be found in [2]).

$A \otimes B$ is the Hilbert space completion of the space $A \otimes' B$ of all conjugate bilinear functionals T on $A \times B$ of the form $T = \sum_{i=1}^{n} a_i \otimes b_i$, where $T(a, b) = \Sigma (a_i, a)(b_i, b)$ (see [3]). We define $(a \otimes b)(c \otimes d) = ac \otimes bd$, and extend by linearity and continuity to multiplication on $A \otimes B$. Then

I. $A \otimes B$ is an H^*-algebra and each $A_\alpha \otimes B_\beta$ may be identified with a closed ideal in $A \otimes B$.

II. $A \otimes B = \Sigma \otimes (A_\alpha \otimes B_\beta)$ is the decomposition of $A \otimes B$ into minimal closed ideals.

III. $A \otimes B$ is strongly semi-simple (see [5], p. 59) if and only if both A and B are strongly semi-simple.

1. Tensor products.

DEFINITION. $F_C(A, B)$ will denote the collection of all finite formal

Received February 19, 1965 and in revised form April 21, 1965. Most of the material in this paper forms part of the author's thesis, written at the University of Minnesota under the direction of Professor B.R. Gelbaum.

sums of the form

$\sum_{i=1}^{n} c_i(a_i, b_i)$, with $a_i \in A$, $b_i \in B$, and $c_i \in C$; i.e. $F_0(A, B)$ is the free C-module generated by $A \times B$.

$F_0(A, B)$ becomes an algebra and a pseudo-inner product space if the operations are defined by the formulas:

$$(c(a, b))\cdot(c'(a', b')) = cc'(aa', bb') ,$$
$$\lambda\Sigma c_i(a_i, b_i) = \Sigma(\lambda c_i)(a_i, b_i), \lambda \text{ complex, and}$$
$$(c(a, b), c'(a', b')) = (c, c')(a, a')(b, b')$$

(the first and third must be extended by linearity). The positive semi-definiteness of the pseudo-inner product follows from the fact that $(c(a, b), c'(a', b')) = (a \otimes b \otimes c, a' \otimes b' \otimes c')$; the other properties required of an inner product obviously hold.

Let I_1' be the ideal in $F_0(A, B)$ spanned by the set of all elements of the following forms:

(1) $$c(a_1 + a_2, b) - c(a_1, b) - c(a_2, b) ,$$

(2) $$c(a, b_1 + b_2) - c(a, b_1) - c(a, b_2) ,$$

(3) $$(c_1 + c_2)(a, b) - c_1(a, b) - c_2(a, b) ,$$

(4) $$\lambda c(a, b) - c(\lambda a, b) , \text{ and}$$

(5) $$\lambda c(a, b) - c(a, \lambda b)$$

for arbitrary $a, a_i \in A$; $b, b_i \in B$; $c, c_i \in C$; and complex numbers λ. Let I_2' be the ideal in $F_0(A, B)$ generated by the set of all elements of the forms:

(6) $$c_1c_2(a, b) - c_1(ac_2, b) , \text{ and}$$

(7) $$c_1c_2(a, b) - c_1(a, c_2b)$$

for arbitrary $a \in A$, $b \in B$, and $c_i \in C$. Then let $I' = I_1' \vee I_2' = I_1' + I_2'$, the ideal generated by the set of all elements of the forms (1)–(7).

PROPOSITION 1. $I_1' = \{X \in F_0(A, B): (X, X) = 0\}$.

Proof. Straightforward computations show that $(X, Y) = 0$ if X is of one of the forms (1)–(5) and $Y = c'(a', b')$. Extending by linearity we have immediately that $(X, Y) = 0$ for all $X \in I_1'$, $Y \in F_0(A, B)$. Suppose then that $X = \sum_{i=1}^{n} c_i(a_i, b_i)$ and that $(X, X) = 0$. It must be shown that $X \in I_1'$.

If $\{c_i\}_{i=1}^{n}$ is not linearly independent, then we may assume that $c_n = \sum_{i=1}^{n-1} \lambda_i c_i$, and so

$$X = \sum_{i=1}^{n-1} c_i(a_i, b_i) + \Big(\sum_{i=1}^{n-1} \lambda_i c_i\Big)(a_n, b_n)$$
$$= \sum_{i=1}^{n-1} c_i(a_i, b_i) + \sum_{i=1}^{n-1} c_i(\lambda_i a_n, b_n)$$
$$+ \Big[\Big(\sum_{i=1}^{n-1} \lambda_i c_i\Big)(a_n, b_n) - \sum_{i=1}^{n-1} c_i(\lambda_i a_n, b_n)\Big] .$$

The expression in brackets is clearly an element of I_1', call it γ_1. Thus we have

$$X = \sum_{j=1}^{2} \sum_{i=1}^{n-1} c_i(a_{ij}, b_{ij}) + \gamma_1 ,$$

where $a_{i1} = a_i$, $a_{i2} = \lambda_i a_n$, $b_{i1} = b_i$, $b_{i2} = b_n$. Repeating the process as many times as is necessary we obtain

$$X = \sum_{j=1}^{2^p} \Big(\sum_{i=1}^{n-p} c_i(a_{ij}, b_{ij})\Big) + \gamma_p ,$$

where $\gamma_p \in I_1'$ and $\{c_i\}_{i=1}^{n-p}$ is linearly independent. Then, for each fixed index i, by using an argument similar to the one above, we can write

$$\sum_{j=1}^{2^p} c_i(a_{ij}, b_{ij}) = \sum_{k=1}^{2^{q(i)}} \Big(\sum_{j=1}^{2^{p-q(i)}} c_i(a_{ij}, b_{ijk})\Big) + \gamma_{iq(i)} ,$$

where $\gamma_{iq(i)} \in I_1'$ and $\{a_{ij}: j = 1, \cdots, 2^p - q(i)\}$ is linearly independent. As a result, we have

$$X = \sum_{i=1}^{n-p} \sum_{j=1}^{2^{p-q(i)}} \sum_{k=1}^{2^{q(i)}} c_i(a_{ij}, b_{ijk}) + \gamma ,$$

where $\{c_i\}$ is linearly independent, $\{a_{ij}\}$ is linearly independent for each fixed i, and $\gamma \in I_1'$.

Fix any pair $< i, j >$ of indices. By the Hahn-Banach Theorem and the Riesz Theorem there exist $a' \in A$ and $c' \in C$ such that

$$|| c' || = || a' || = 1, \ (c_i, c') = d_i > 0, (a_{ij}, a') = d_{ij} > 0 ,$$

$(c_{i'}, c') = 0$ if $i' \neq i$, and $(a_{ij'}, a') = 0$ if $j' \neq j$. Since $F_0(A, B)$ is a pseudo-inner product space, the Schwarz inequality holds. Thus if we let $b' = \Sigma\{b_{ijk}: k = 1, \cdots, 2^{q(i)}\}$, we have

$$| (X, c'(a', b') | \leqq (X, X)(c'(a', b'), c'(a', b')) = 0 .$$

On the other hand,

$$(X, c'(a', b')) = \textstyle\sum_{m,n,k}(c_m, c')(a_{mn}, a')(b_{mnk}, b')$$
$$= d_i d_{ij} || b' ||^2 = 0 ,$$

so that $b' = 0$. If we now write

$$\begin{aligned}\sum_k c_i(a_{ij}, b_{ijk}) &= c_i(a_{ij}, \sum_k b_{ijk}) + [\sum_k c_i(a_{ij}, b_{ijk}) - c_i(a_{ij}, \sum_k b_{ijk})] \\ &= c_i(a_{ij}, 0) + \gamma'_{ij} ,\end{aligned}$$

where γ'_{ij} is the expression in brackets, which is clearly an element of I'_1, then we have

$$X = \sum_{i,j} c_i(a_{ij}, 0) + \gamma' ,$$

where $\gamma' = \sum_{i,j}\gamma'_{ij}$, and so $X \in I'_1$.

$F_0(A, B)$ is a pseudo-normed space, with $\| X \|^2 = (X, X)$. Let us denote by $\mathscr{T}_0(A, B)$ its pseudo-normed completion, i.e. the collection of all Cauchy sequences from $F_0(A, B)$. Define a mapping

$$\varphi: F_0(A, B) \to A \otimes B \otimes C$$

as follows:

$$\varphi(\Sigma c_i(a_i, b_i)) = \Sigma a_i \otimes b_i \otimes c_i .$$

It is immediate that φ is a linear, homogeneous, multiplicative isometry, and that its range is dense. Thus φ can be extended to an isometric homomorphism on $\mathscr{T}_0(A, B)$ onto $A \otimes B \otimes C$. Note that $\| XY \| \leqq \| X \| \| Y \|$ for all $X, Y \in F_0(A, B)$, since $A \otimes B \otimes C$ is a Banach algebra. Thus the operations defined on $F_0(A, B)$ can be extended to $\mathscr{T}_0(A, B)$, as usual.

Let I_1, I_2, and I denote the closures, in $\mathscr{T}_0(A, B)$, of I'_1, I'_2, and I', respectively. It is obvious from Proposition 1 that $I_1 = \{X \in \mathscr{T}_0(A, B): \| X \| = 0\}$, i.e. i.e. I_1 is the closure of (0). Thus I_1 is a subset of every closed subspace of $\mathscr{T}_0(A, B)$, which means, in particular, that $I = I_2$. In other words, I can be described quite simply as the closed ideal of $\mathscr{T}_0(A, B)$ generated by the collection of all elements of the forms (6) and (7).

DEFINITION. $A \otimes_0 B$, the tensor product of A and B, over C, is the quotient algebra $\mathscr{T}_0(A, B)/I$.

$A \otimes_0 B$ is a normed space (as is always the case when a pseudo-normed space is factored by a closed subspace). We proceed to identify it with an ideal in $A \otimes B \otimes C$. Let $D = \varphi(I)$ and define a map $\gamma: A \otimes_0 B \to (A \otimes B \otimes C)/D$ by the formula $\gamma(X + I) = \varphi(X) + D$. It is clear that γ is linear, and since $\gamma(I) = \varphi(0) + D = D$, γ is well defined; it is multiplicative since φ is multiplicative. Finally, γ is an isometry. For if $T = X + I \in A \otimes_0 B$, then

$$\begin{aligned}\| \gamma T \| &= \| \varphi X + D \| = \inf \{\| \varphi X + Z \|: Z \in D\} \\ &= \inf \{\| \varphi X + \varphi Y \|: Y \in I\} \\ &= \inf \{\| X + Y \|: Y \in I\} = \| T \| ,\end{aligned}$$

since φ is an isometric homomorphism.

Since D is a closed ideal in the H^*-algebra $A\otimes B\otimes C$, $(A\otimes B\otimes C)/D$ is isomorphic and isometric with the closed ideal $D^{\perp}$, which we shall denote by E. We summarize the foregoing information in the next theorem.

THEOREM. *There is an isometric isomorphism from $A\otimes_{\sigma}B$ into $A\otimes B\otimes C$; its range is the closed ideal E which is the orthogonal complement of the closed ideal D generated by all elements of the forms*

(i) $a\otimes b\otimes c_1c_2 - ac_2\otimes b\otimes c_1$,

(ii) $a\otimes b\otimes c_1c_2 - a\otimes c_2b\otimes c_1$.

Consequently, $A\otimes_{\sigma}B$ is an H^*-algebra; its minimal closed ideals can be identified with those minimal closed ideals $A_\alpha\otimes B_\beta\otimes C_\gamma$ of $A\otimes B\otimes C$ that are orthogonal to D.

COROLLARY. *If A, B, and C are strongly semi-simple, then $A\otimes_{\sigma}B$ is strongly semi-simple.*

The following proposition provides means by which it is easy to construct examples for which the converse to the above corollary is false.

PROPOSITION 2. If $A_\alpha\otimes B_\beta\otimes C_\gamma$ is a minimal closed ideal in E, then C_γ is of dimension one.

Proof. Choose a canonical basis $\{a_{ij}\otimes b_{kl}\otimes c_{mn}\}$ for $A_\alpha\otimes B_\beta\otimes C_\gamma$ (see [2]). Since $a_{ij}\otimes b_{kl}\otimes c_{mn}\in E$, it must be orthogonal to

$$a_{ij}\otimes b_{kl}\otimes c_{mp}c_{pn} - a_{ij}c_{pn}\otimes b_{kl}\otimes c_{mp}\ .$$

If the dimension of C_γ were greater than one, then it would be possible to choose $n\neq p$, and we would have

$$\begin{aligned} 0 &= (a_{ij}\otimes b_{kl}\otimes c_{mn},\ a_{ij}\otimes b_{kl}\otimes c_{mn} - a_{ij}c_{pn}\otimes b_{kl}\otimes c_{mp} \\ &= \|a_{ij}\|^2\,\|b_{kl}\|^2\,\|c_{mn}\|^2\ , \end{aligned}$$

since $(c_{mn}, c_{mp}) = 0$. This, of course, is a contradiction.

COROLLARY. *If C has no one-dimensional minimal ideals, then $A\otimes_{\sigma}B = (0)$.*

2. **Examples.** Perhaps the easiest method of obtaining examples of H^*-algebras A, B, and C related as above is to let A, B, and C be

closed ideals in some H^*-algebra $\mathscr{A}$. The structure of $A \otimes_\sigma B$, under such circumstances, is described in the next proposition.

PROPOSITION 3. Suppose that A, B and C are closed ideals in an H^*-algebra $\mathscr{A}$. If A and B are viewed as C-modules with ordinary multiplication in $\mathscr{A}$ as the module action, then $A \otimes_\sigma B$ is isomorphic with the direct sum of all the one-dimensional minimal ideals in $A \cap B \cap C$. The isomorphism is an isometry if and only if the identity of each one-dimensional minimal ideal in $A \cap B \cap C$ has norm one.

Proof. Choose a canonical basis $\{u_{pq}^\delta\}$ for $\mathscr{A}$. Then $\{a_{ij}\} = A = \cap \{u_{pq}^\delta\}$, $\{b_{kl}^\beta\} = B \cap \{u_{pq}^\delta\}$, and $\{c_{mn}^\gamma\} = C \cap \{u_{pq}^\delta\}$ are canonical bases for A, B, and C, respectively and $\{a_{ij}^\alpha \otimes b_{kl}^\beta \otimes c_{mn}^\gamma\}$ is a canonical basis for $A \otimes B \otimes C$. If $a_{ij}^\alpha \otimes b_{kl}^\beta \otimes c_{mn}^\gamma \in E$, then, by Proposition 2, $c_{mn}^\gamma = c^\gamma$ is the identity of a one-dimensional minimal ideal. If $\alpha \neq \gamma$, then

$$a_{ij}^\alpha \otimes b_{kl}^\beta \otimes c^\gamma c^\gamma - a_{ij}^\alpha c^\gamma \otimes b_{kl}^\beta \otimes c^\gamma = a_{ij}^\alpha \otimes b_{kl}^\beta \otimes c^\gamma \in D .$$

Similarly, if $\beta \neq \gamma$, then $a_{ij}^\alpha \otimes b_{kl}^\beta \otimes c^\gamma \in D$. Thus if an element of a canonical basis is to be in E it must be of the form $c^\gamma \otimes c^\gamma \otimes c^\gamma$. Relatively straightforward computations show that each such basis element is orthogonal to D, and the proof is completed.

Suppose now that G, H, and K are compact groups, and that $\theta: K \to G$ and $\varphi: K \to H$ are continuous homomorphisms. Then $\theta(K)$ and $\varphi(K)$ are closed subgroups of G and H, respectively, $L^2(G)$ and $L^2(H)$ become modules over $L^2(K)$, with the module action defined by:

$$g * k(x) = \int_K g(x(\theta z)^{-1}) k(z) dz ,$$

$$k * h(y) = \int_K k(z) h((\varphi z)^{-1} y) dz ,$$

for all $g \in L^2(G)$, $h \in L^2(H)$, $k \in L^2(K)$, $x \in G$, and $y \in H$ (all integrations are with respect to normalized Haar measures). If we let $A = L^2(G)$, $B = L^2(H)$, $C = L^2(K)$, then $A \otimes_\sigma B$ is a well-defined H^*-algebra. As was remarked in [2], $A \otimes B \otimes C$ can be identified with $L^2(G \times H \times K)$, and so, by the Theorem of §1, $A \otimes_\sigma B$ can be identified with a closed ideal J in $L^2(G \times H \times K)$. At one extreme, suppose θ and φ map K onto the identities of G and H, respectively. It is not difficult to see that in this case $A \otimes_\sigma B$ can be identified with $L^2(G \times H)$.

At what might be considered another extreme, suppose that G and H are closed subgroups of some compact group, that K is a closed subgroup of $G \cap H$, and that θ and φ are the inclusion maps. Define an equivalence relation on $G \times H \times K$ as follows: $(x, y, z) \sim (u, v, w)$

if and only if $F(x, y, z) = F(u, v, w)$ for all $F \in J$. Then $M = \{(x, y, z): (x, y, z) \sim (e, e, e)\}$ is a closed normal subgroup of $G \times H \times K$, and its cosets are the equivalence classes of $\sim$. All functions $F \in J$ are thus constant on the cosets of M, providing a mapping ψ from J to $L^2((G \times H \times K)/M)$. The map ψ is an isometric isomorphism and its image is an ideal. On the basis of the Tannaka Duality Theorem (see [4], p. 439) it seems reasonable to conjecture that ψ is surjective, so that $A \otimes_0 B$ can be identified with $L^2((G \times H \times K)/M)$. The conjecture has not been settled in general, but let us consider the very special case where $G = H = K$. Then, by Proposition 3, $A \otimes_0 B$ can be identified with the direct sum of all one-dimensional minimal ideals in $L^2(G)$, which in turn is isomorphic and isometric with $L^2(G/N)$, where N is the closure of the commutator subgroup of G. Since G/N and $(G \times G \times G)/M$ are isomorphic via the mapping $xN \to (x, e, e)M$, the conjecture is verified in this special case.

REFERENCES

1. B. Gelbaum, *Tensor products over Banach algebras*, to appear in Trans. Amer. Math. Soc.
2. L. Grove, *A generalized group algebra for compact groups*, to appear in Studia Mathematica, vol. **26**, Fascicule 1.
3. F. Murray, and J. von Neumann. *On rings of operators I*, Ann. of Math. **37** (1936), 116-229.
4. M. Naimark, *Normed Rings*, P. Noordhoff, Groningen, the Netherlands, 1959.
5. C. Rickart, *General Theory of Banach Algebras*, Van Nostrand, New York, 1960.

UNIVERSITY OF MINNESOTA AND DARTMOUTH COLLEGE

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

L^2 EXPANSIONS IN TERMS OF GENERALIZED HEAT POLYNOMIALS AND OF THEIR APPELL TRANSFORMS

DEBORAH TEPPER HAIMO

The object of this paper is to characterize functions which have L^2 expansions in terms of polynomial solutions $P_{n,\nu}(x, t)$ of the generalized heat equation

$$(*) \qquad \left[\frac{\partial^2}{\partial x^2} + \frac{2\nu}{x}\frac{\partial}{\partial x}\right]u(x, t) = \frac{\partial}{\partial t}u(x, t) .$$

and in terms of the Appell transforms $W_{n,\nu}(x, t)$ of the $P_{n,\nu}(x, t)$. H^* denotes the C^2 class of functions $u(x, t)$ which, for $a < t < b$, satisfy (*) and for which

$$u(x, t) = \int_0^\infty G(x, y; t - t')u(y, t')d\mu(y),$$

$$d\mu(x) = 2^{(1/2)-\nu}\left[\Gamma\left(\nu + \frac{1}{2}\right)\right]^{-1}x^{2\nu}dx ,$$

for all $t, t', a < t' < t < b$, the integral converging absolutely, where $G(x, y; t)$ is the source solution of (*). The principal results are the following:

THEOREM. **Let $u(x, t) \in H^*$, $-\sigma \leqq t < 0$, and**

$$u(x, t)[G(x; -t)]^{\frac{1}{2}} \in L^2$$

for each fixed t $-\sigma \leqq t < 0$, $0 \leqq x < \infty$. Then, for $-\sigma \leqq t < 0$,

$$\lim_{N\to\infty}\int_0^\infty G(x; -t)\left|u(x, t) - \sum_{n=0}^{N} a_n P_{n,\nu}(x, -t)\right|^2 d\mu(x) = 0 ,$$

and

$$\int_0^\infty G(x; -t)\,|\,u(x, t)\,|^2\,d\mu(x) = \sum_{n=0}^{\infty}|\,a_n\,|^2\,b_n^{-1}\,t^{2n} ,$$

where

$$b_n = [2^{4n}\,n!\,]^{-1}\frac{\Gamma\left(\nu + \frac{1}{2}\right)}{\Gamma\left(\nu + \frac{1}{2} + n\right)},$$

and

$$a_n = b_n\int_0^\infty u(y, t)W_{n,\nu}(y, -t)d\mu(y) .$$

Received January, 16, 1965. The research was supported by a National Science Foundation Fellowship at Harvard University.

THEOREM. **If** $u(x,t)\in H^*$, $0<t\leqq\sigma$, **and if**

$$u(ix,t)[G(x;t)]^{(1/2)}\in L^2$$

for each fixed t, $0<t\leqq\sigma$, $0\leqq x<\infty$, **then, for** $0<t\leqq\sigma$,

$$\lim_{N\to\infty}\int_0^\infty G(x;t)\left|u(ix,t)-\sum_{n=0}^{N}a_nP_{n,\nu}(x,-t)\right|^2d\mu(x)=0\,,$$

and

$$\int_0^\infty G(x;t)\,|\,u(ix,t)\,|^2\,d\mu(x)=\sum_{n=0}^{\infty}|\,a_n\,|^2\,b_n^{-1}\,t^{2n}\,,$$

where b_n **is given above and**

$$a_n=b_n\int_0^\infty u(ix,t)W_{n,\nu}(x,t)d\mu(x)\,.$$

THEOREM. **If** $u(x,t)\in H^*$, $0<\sigma\leqq t$, **and if**

$$u(x,t)[G(ix;t)]^{(1/2)}\in L^2$$

for each fixed t, $0<\sigma\leqq t$, $0\leqq x<\infty$, **then, for** $0<\sigma\leqq t$,

$$\lim_{N\to\infty}\int_0^\infty G(ix;t)\left|u(x,t)-\sum_{n=0}^{N}a_nW_{n,\nu}(x,t)\right|^2d\mu(x)=0\,,$$

and

$$\int_0^\infty G(ix;t)\,|\,u(x,t)\,|^2\,d\mu(x)=\sum_{n=0}^{\infty}t^{-2n}\,b_n^{-1}(2t)^{-2\nu-1}\,|\,a_n\,|^2\,,$$

where b_n **is given above, and**

$$a_n=b_n\int_0^\infty u(x,t)P_{n,\nu}(x,-t)d\mu(x)\,.$$

The theory is an extension, in part, of recent results of P. C. Rosenbloom and D. V. Widder.

1. **Preliminary results.** The generalized heat polynomial $P_{n,\nu}(x,t)$ is a polynomial defined by

$$P_{n,\nu}(x,t)=\sum_{k=0}^{n}2^{2k}\binom{n}{k}\frac{\Gamma\left(\nu+\frac{1}{2}\right)}{\Gamma\left(\nu+\frac{1}{2}+n-k\right)}x^{2n-2k}t^k\,, \tag{1.1}$$

ν a fixed positive number. Note that when $\nu=0$, $P_{n,0}(x,t)=v_{2n}(x,t)$, the ordinary heat polynomials defined in [8; p. 222]. For $t>0$, $P_{n,\nu}(x,t)$ has the following integral representation.

$$P_{n,\nu}(x,t)=\int_0^\infty y^{2n}G(x,y;t)d\mu(y), \tag{1.2}$$

$$d\mu(y)=2^{(1/2)-\nu}\left[\Gamma\left(\nu+\frac{1}{2}\right)\right]^{-1}x^{2\nu}dx\,.$$

As may readily be verified, for $-\infty < x,\ t < \infty$, $P_{n,\nu}(x, t)$ satisfies the generalized heat equation

$$\Delta_x u(x, t) = \frac{\partial}{\partial t} u(x, t) , \tag{1.3}$$

where $\Delta_x f(x) = f''(x) + (2\nu/x) f'(x)$. We denote by H the class of all C^2 functions which satisfy (1.3). The source solution of (1.3) is given by $G(x; t)$, where

$$G(x, y; t) = \left(\frac{1}{2t}\right)^{\nu+\frac{1}{2}} \exp\left(-\frac{x^2 + y^2}{4t}\right) \mathscr{I}\left(\frac{xy}{2t}\right) , \tag{1.4}$$

with $\mathscr{I}(z) = C_\nu z^{(1/2)-\nu} I_{\nu-(1/2)}(z)$, $C_\nu = 2^{(1/2)-\nu}\Gamma(\nu + (1/2))$, $I_r(z)$ being the Bessel function of imaginary argument of order r, and where $G(x; t) = G(x, 0; t)$. For a detailed study of the properties of $G(x, y; t)$ see [1].

Corresponding to the generalized heat polynomial $P_{n,\nu}(x, t)$ is its Appell transform $W_{n,\nu}(x, t)$ defined by

$$W_{n,\nu}(x,t) = G(x,t)P_{n,\nu}\left(\frac{x}{t}, -\frac{1}{t}\right),\ t > 0,\ n = 0, 1, 2, \cdots , \tag{1.5}$$

which is also a solution of (1.3). It follows readily from the definition of $P_{n,\nu}(x, t)$ that

$$W_{n,\nu}(x, t) = t^{-2n}G(x, t)P_{n,\nu}(x, -t),\ t > 0,\ n = 0, 1, 2, \cdots . \tag{1.6}$$

The importance of $P_{n,\nu}(x, t)$ and $W_{n,\nu}(x, t)$ in our theory is that they form a biorthogonal system on $0 \leqq x < \infty$. We have, for $t > 0$,

$$\int_0^\infty W_{n,\nu}(x, t)P_{m,\nu}(x, -t)d\mu(x) = \frac{1}{b_n}\delta_{mn} , \tag{1.7}$$

where

$$b_n = \Gamma\left(\nu + \frac{1}{2}\right)\Big/\left[2^{4n}n!\Gamma\left(\nu + \frac{1}{2} + n\right)\right] . \tag{1.8}$$

A consequence of (1.7) is a fundamental generating function for the biorthogonal set $P_{n,\nu}(x, -t)$, $W_{n,\nu}(x, t)$. We have, for $0 \leqq x, y < \infty$, $-s < t < s$, $s > 0$,

$$G(x, y; s + t) = \sum_{n=0}^{\infty} b_n W_{n,\nu}(y, s)P_{n,\nu}(x, t) . \tag{1.9}$$

2. Inversion.

For $t > s$, let us set

$$\mathscr{K}(x, y; s, t) = \sum_{n=0}^{\infty} b_n\left(\frac{t}{s}\right)^{(\nu/2)+(1/4)} e^{(x^2/8t)-(y^2/8s)} W_{n,\nu}(x,t)P_{n,\nu}(y, -s) , \tag{2.1}$$

where b_n is defined by (1.8). Then, as a consequence of the definitions and of (1.9), we have

$$\mathscr{K}(x, y; s, t) = \left(\frac{t}{s}\right)^{(\nu/2)+(1/4)} e^{-(x^2(t-s))/(8t(t+s))}\, G(x\sqrt{2s/(t+s)},\, y\sqrt{(t+s)/2s};\, t-s)\,. \tag{2.2}$$

From the well known properties of $G(x, y; t)$ — see [1; § 4] — the following results are immediate.

LEMMA 2.1.

$$\text{(a)}\quad \mathscr{K}(x, y; s, t) \geqq 0,\ 0 \leqq x, y < \infty,\ s < t\,, \tag{2.3}$$

$$\text{(b)}\quad \lim_{y\to\infty} \mathscr{K}(x, y; s, t) = 0,\ 0 \leqq x < \infty,\ s < t\,, \tag{2.4}$$

$$\text{(c)}\quad \lim_{s\to t^-} \mathscr{K}(x, y; s, t) = 0 \text{ } \textit{uniformly } 0 \leqq x, y < \infty, \tag{2.5}$$

$$|y - x| \geqq \delta > 0,\ \delta \textit{ any fixed positive number.}$$

(d) *For x fixed, $0 \leqq x < \infty$,*

$$\lim_{s\to t-} \int_a^b \mathscr{K}(x, y; s, t)\, d\mu(y) = \begin{cases} 1\,, & 0 \leqq a < x < b \leqq \infty\,, \\ 0\,, & 0 \leqq a \leqq b < x < \infty\,, \\ 0\,, & 0 \leqq x < a < b \leqq \infty\,. \end{cases} \tag{2.6}$$

It is now easy to establish the following fundamental inversion theorem.

THEOREM 2.2. *If φ belongs to $L^1(0, \infty)$ and is continuous at x, then*

$$\lim_{s\to t^-} \int_0^\infty \mathscr{K}(x, y; s, t)\varphi(y)\, d\mu(y) = \varphi(x)\,. \tag{2.7}$$

3. The Huygens property. A function $u(x, t)$ is said to have the Huygens property for $a < t < b$ if and only if $u(x, t) \in H$ there and for every t, t', $a < t' < t < b$,

$$u(x, t) = \int_0^\infty G(x, y; t - t')u(y, t')\, d\mu(y)\,, \tag{3.1}$$

the integral converging absolutely. We denote the class of all functions with the Huygens property by H^*. Functions of class H^* have a complex integral representation as given in the following result.

LEMMA 3.1. *If $u(x, t) \in H^*$, $a < t < b$, then for $a < t < t' < b$,*

$$u(x, t) = \int_0^\infty G(ix, y; t' - t)u(iy, t')d\mu(y) \,. \tag{3.2}$$

The fact that $P_{n,\nu}(x, t) \in H^*$ for $-\infty < t < \infty$, and $W_{n,\nu}(x, t) \in H^*$ for $0 < t < \infty$ enables us to conclude that certain integrals involving functions of H^* are constant. A general result was proved in [5], but we state here the specific forms required in this theory.

THEOREM 3.2. *If* $u(x, -t) \in H^*$ *for* $0 < t < \infty$, *then*

$$\int_0^\infty u(x, -t)W_{n,\nu}(x, t)d\mu(x) \tag{3.3}$$

is a constant.

THEOREM 3.3. *If* $u(x, t) \in H^*$ *for* $0 < t < \infty$, *then*

$$\int_0^\infty u(ix, t)W_{n,\nu}(x, t)d\mu(x) \tag{3.4}$$

is a constant.

THEOREM 3.4. *If* $u(x, t) \in H^*$ *for* $0 < t < \infty$, *then*

$$\int_0^\infty u(x, t)P_{n,\nu}(x, -t)d\mu(x) \tag{3.5}$$

is a constant.

4. L^2 **expansions.** We establish criteria for a function $u(x, t)$ so that the series $\sum_{n=0}^\infty a_n P_{n,\nu}(x, -t)$ converges in mean, with weight functions $G(x, -t)$, to $u(x, t)$.

THEOREM 4.1. *Let* $u(x, t) \in H^*$ *for* $-\sigma \leqq t < 0$, *and*

$$u(x, t)[G(x, -t)]^{1/2} \in L^2$$

for $-\sigma \leqq t < 0$, $0 \leqq x < \infty$. *Then, for* $-\sigma \leqq t < 0$,

$$\lim_{N\to\infty} \int_0^\infty G(x, -t) \left| u(x, t) - \sum_{n=0}^N a_n P_{n,\nu}(x, -t) \right|^2 d\mu(x) = 0 \tag{4.1}$$

and

$$\int_0^\infty G(x, -t) \,|\, u(x, t) \,|^2 \, d\mu(x) = \sum_{n=0}^\infty \frac{|\, a_n \,|^2}{b_n} \, t^{2n} \,, \tag{4.2}$$

where b_n *is given by* (1.8) *and*

$$a_n = b_n \int_0^\infty u(y, t)W_{n,\nu}(y, -t)d\mu(y) \,. \tag{4.3}$$

Proof. For t fixed, let $\phi(x, t)$ be a continuous function vanishing outside a finite interval and such that, for $\varepsilon > 0$,

$$\int_0^\infty | u(x, -t)[G(x, t)]^{1/2} - \phi(x, t) |^2 \, d\mu(x) < \varepsilon, \quad 0 < t \leqq \sigma . \tag{4.4}$$

Now set

$$\psi_n(x, t) = P_{n,\nu}(x, -t)[G(x, t)]^{1/2} , \qquad 0 < t \leqq \sigma . \tag{4.5}$$

Then, by (2.1), we have

$$\mathscr{K}(x, y; s, t) = \sum_{n=0}^\infty b_n t^{-2n} \psi_n(x, t)\psi_n(y, s) , \tag{4.6}$$

where b_n is defined by (1.8). Hence

$$\begin{aligned}\int_0^\infty \mathscr{K}(x, y; s, t)\phi(y, t)d\mu(y) &= \int_0^\infty \phi(y, t)d\mu(y) \sum_{n=0}^\infty b_n t^{-2n}\psi_n(x, t)\psi_n(y, s) \\ &= \sum_{n=0}^\infty b_n t^{-2n}\psi_n(x, t)\int_0^\infty \psi_n(y, s)\phi(y, t)d\mu(y) .\end{aligned}$$

If we set

$$A_n(t) = b_n t^{-2n}\int_0^\infty \psi_n(y, t)\phi(y, t)d\mu(y) , \tag{4.7}$$

and apply Theorem 2.2, we find that

$$\sum_{n=0}^\infty A_n(t)\psi_n(x, t) = \lim_{s \to t^-} \int_0^\infty \mathscr{K}(x, y; s, t)\phi(y, t)d\mu(y) = \phi(x, t) . \tag{4.8}$$

If we multiply both sides of (4.8) by $\phi(x, t)d\mu(x)$ and integrate between 0 and ∞, we obtain

$$\sum_{n=0}^\infty A_n(t)\int_0^\infty \psi_n(x, t)\phi(x, t)d\mu(x) = \int_0^\infty \phi^2(x, t)d\mu(x) ,$$

or, by (4.7),

$$\sum_{n=0}^\infty \frac{t^{2n}}{b_n} A_n^2(t) = \int_0^\infty \phi^2(x, t)d\mu(x) . \tag{4.9}$$

Now, let

$$c_n(t) = b_n t^{-2n}\int_0^\infty u(y, -t)[G(y, t)]^{1/2}\psi_n(y, t)d\mu(y) . \tag{4.10}$$

Consider

$$I = \int_0^\infty \left\{u(x, -t)[G(x, t)]^{1/2} - \sum_{k=0}^n c_k(t)\psi_k(x, t)\right\}^2 d\mu(x) . \tag{4.11}$$

Since, by 1.7, we have

$$\int_0^\infty \psi_n(x,t)\psi_m(x,t)d\mu(x) = \frac{t^{2n}}{b_n}\delta_{mn}\,, \tag{4.12}$$

with b_n given in (1.8), it follows that

$$\begin{aligned}
I &= \int_0^\infty [u(x,-t)]^2 G(x,t)d\mu(x) - \sum_{k=0}^n c_k^2(t)\frac{t^{2k}}{b_k} \\
&\leqq \int_0^\infty [u(x,-t)]^2 G(x,t)d\mu(x) + \sum_{k=0}^n \frac{t^{2k}}{b_k}[A_k(t)-c_k(t)]^2 - \sum_{k=0}^n c_k^2(t)\frac{t^{2k}}{b_k} \\
&= \int_0^\infty [u(x,-t)]^2 G(x,t)d\mu(x) + \sum_{k=0}^n \frac{t^{2k}}{b_k}A_k^2(t) - 2\sum_{k=0}^n \frac{t^{2k}}{b_k}A_k(t)c_k(t) \\
&= \int_0^\infty \left\{u(x,-t)[G(x,t)]^{1/2} - \sum_{k=0}^n A_k(t)\psi_k(x,t)\right\}^2 d\mu(x) \\
&\leqq 2\int_0^\infty \{u(x,-t)[G(x,t)]^{1/2} - \phi(x,t)\}^2 d\mu(x) \\
&\qquad + 2\int_0^\infty \left\{\phi(x,t) - \sum_{k=0}^n A_k(t)\psi_k(x,t)\right\}^2 d\mu(x)\,.
\end{aligned}$$

By (4.4), we have

$$\begin{aligned}
I &< 2\varepsilon + 2\int_0^\infty \phi^2(x,t)d\mu(x) + 2\int_0^\infty \sum_{k=0}^n A_k^2(t)\psi_k^2(x,t)d\mu(x) \\
&\qquad - 4\int_0^\infty \phi(x,t)d\mu(x)\sum_{k=0}^n A_k(t)\psi_k(x,t) \\
&< 2\varepsilon + 2\int_0^\infty \phi^2(x,t)d\mu(x) + 2\sum_{k=0}^n A_k^2(t)\frac{t^{2n}}{b_n} \\
&\qquad - 4\sum_{k=0}^n A_k(t)\int_0^\infty \phi(x,t)\psi_k(x,t)d\mu(x) \\
&< 2\varepsilon + 2\left\{\int_0^\infty \phi^2(x,t)d\mu(x) - \sum_{k=0}^n A_k^2(t)\frac{t^{2n}}{b_n}\right\}\,.
\end{aligned}$$

It follows, therefore, by (4.9), that if n is sufficiently large, $I < 4\varepsilon$. Hence

$$\lim_{N\to\infty}\int_0^\infty \left|u(x,-t)[G(x,t)]^{1/2} - \sum_{k=0}^N c_k(t)\psi_k(x,t)\right|^2 d\mu(x) = 0\,, \tag{4.13}$$

or, by (4.5), we have (4.1) with $c_k(t) = a_k$. Theorem 3.4 establishes the fact that a_k is independent of t.

Parseval's equation (4.2) follows since

$$\begin{aligned}
\int_0^\infty G(x,t)\,|\,u(x,-t)\,|^2\,d\mu(x) &= \int_0^\infty \left|\sum_{n=0}^\infty c_n(t)\psi_n(x,t)\right|^2 d\mu(x) \\
&= \sum_{n=0}^\infty |\,a_n\,|^2\frac{t^{2n}}{b_n}\,,
\end{aligned}$$

with the last equality a result of (4.12).

An example illustrating the theorem is given by $u(x, t) = e^{a^2t}\mathscr{J}(ax)$. This function satisfies the hypotheses for $-\infty < t < 0$ and we find that

$$\int_0^\infty G(x, t)\mathscr{J}^2(ax)e^{-2a^2t}d\mu(x) = \mathscr{J}(2a^2t)\,, \qquad 0 < t < \infty\,, \tag{4.14}$$

whereas

$$\begin{aligned}\sum_{n=0}^{\infty} |\, a_n\,|^2 \frac{t^{2n}}{b_n} &= \sum_{n=0}^{\infty} b_n(a^2t)^{2n}2^{4n} \\ &= \mathscr{J}(2a^2t)\,, \qquad 0 < t < \infty\,,\end{aligned} \tag{4.15}$$

since

$$\begin{aligned}a_n &= b_n\int_0^\infty e^{-a^2t}\mathscr{J}(ay)W_{n,\nu}(y, t)d\mu(y)\,, \qquad 0 < t < \infty \\ &= (2a)^{2n}b_n\,.\end{aligned} \tag{4.16}$$

Although, in this example, $u(x, t) \in H^*$ for $-\infty < t < \infty$, the expansion (4.1) does not hold in the extended strip. Note that, in this case, the requirement that $u(x, t)[G(x, -t)]^{1/2}$ be in L^2 fails for $0 < t < \infty$. A modification of Theorem 4.1 when $u(x, t) \in H^*$ for $0 < t \leqq \sigma$ is given by the following result.

THEOREM 4.2. *If* $u(x, t) \in H^*$ *for* $0 < t \leqq \sigma$, *and if*

$$u(ix, t)[G(x, t)]^{1/2} \in L^2$$

for each fixed t, $0 < t \leqq \infty$, $0 \leqq x < \infty$, *then for* $0 < t \leqq \sigma$,

$$\lim_{N\to\infty}\int_0^\infty G(x, t)\left|u(ix, t) - \sum_{n=0}^{N} a_nP_{n,\nu}(x, -t)\right|^2 d\mu(x) = 0\,, \tag{4.17}$$

and

$$\int_0^\infty G(x, t)\,|\,u(ix, t)\,|^2\,d\mu(x) = \sum_{n=0}^{\infty}\frac{t^{2n}}{b_n}\,|\,a_n\,|^2\,, \tag{4.18}$$

where b_n *is given by* (1.8) *and*

$$a_n = b_n\int_0^\infty u(ix, t)W_{n,\nu}(x, t)d\mu(x)\,, \qquad 0 < t \leqq \sigma\,. \tag{4.19}$$

Proof. As in the preceding proof, we have

$$\lim_{N\to\infty}\int_0^\infty \left|u(ix, t)[G(x, t)]^{1/2} - \sum_{n=0}^{N} c_n(t)\psi_n(x, t)\right|^2 d\mu(x) = 0\,,$$

with

$$c_n(t) = b_n t^{-2n} \int_0^\infty u(iy, t)[G(y, t)]^{1/2} \psi_n(y, t) d\mu(y) .$$

Hence (4.17) holds with $c_n(t) = a_n$, which, by Theorem 3.5, is independent of t. Further,

$$\int_0^\infty G(x, t) \mid u(ix, t) \mid^2 d\mu(x) = \int_0^\infty \left| \sum_{n=0}^\infty c_n(t) \psi_n(x, t) \right|^2 d\mu(x)$$
$$= \sum_{n=0}^\infty \frac{t^{2n}}{b_n} \mid a_n \mid^2$$

which is the Parseval equation (4.18).

The example of the preceding theorem satisfies these hypotheses for $0 < t < \infty$, and we have, for $0 < t < \infty$,

$$\int_0^\infty G(x, t) e^{2a^2 t} \mathscr{F}^2(iax) d\mu(x) = \mathscr{F}(2a^2 t) ,$$

whereas

$$a_n = b_n \int_0^\infty e^{a^2 t} \mathscr{F}(iax) W_{n,\nu}(x, t) d\mu(x) ,$$

so that

$$\sum_{n=0}^\infty \frac{t^{2n}}{b_n} \mid a_n \mid^2 = \mathscr{F}(2a^2 t) .$$

Criteria for expansions in terms of $W_{n,\nu}(x, t)$ are given in the following result.

THEOREM 4.3. *If* $u(x, t) \in H^*$ *for* $0 < \sigma \leqq t$, *and if*

$$u(x, t)[G(ix, t)]^{1/2} \in L^2$$

for each fixed t, $0 \leqq \sigma < t$, $0 \leqq x < \infty$, *then for* $0 < \sigma \leqq t$,

$$\lim_{N \to \infty} \int_0^\infty G(ix, t) \left| u(x, t) - \sum_{n=0}^N a_n W_{n,\nu}(x, t) \right|^2 d\mu(x) = 0 , \tag{4.20}$$

and

$$\int_0^\infty G(ix, t) \mid u(x, t) \mid^2 d\mu(x) = \sum_{n=0}^\infty \frac{t^{-2n}}{b_n} (2t)^{-2\nu-1} \mid a_n \mid^2 , \tag{4.21}$$

where b_n *is given by* (1.8) *and*

$$a_n = b_n \int_0^\infty u(x, t) P_{n,\nu}(x, -t) d\mu(x) \qquad \sigma \leqq t < \infty , \tag{4.22}$$

Proof. Again, as in Theorem 4.1, since $u(x, t)[G(ix, t)]^{1/2} \in L^2$, we have

$$\lim_{N\to\infty}\int_0^\infty \left| u(x,t)[G(ix,t)]^{1/2} - \sum_{n=0}^{N} c_n(t)\psi_n(x,t)\right|^2 d\mu(x) = 0 , \tag{4.23}$$

with

$$c_n(t) = b_n t^{-2n}\int_0^\infty u(x,t)[G(ix,t)]^{1/2}\psi_n(x,t)d\mu(x) . \tag{4.24}$$

Now, (4.23) can be written in the form

$$\lim_{N\to\infty}\int_0^\infty G(ix,t)\left| u(x,t) - \sum_{n=0}^{N} c_n(t)(2t)^{\nu+(1/2)}t^{2n}W_{n,\nu}(x,t)\right|^2 d\mu(x) = 0 ,$$

with (4.24) becoming

$$c_n(t) = b_n t^{-2n}(2t)^{-\nu-(1/2)}\int_0^\infty u(x,t)P_{n,\nu}(x,-t)d\mu(x) .$$

Hence, if we set $a_n = c_n(t)t^{2n}(2t)^{\nu+(1/2)}$, a_n is independent of t, by Theorem 3.6, and (4.20) is established. Moreover, Parseval's formula is

$$\begin{aligned}\int_0^a u(ix,t)\,|\,u(x,t)\,|^2\,d\mu(x) &= \sum_{n=0}^{\infty}|\,c_n^2(t)\,|^2\,\frac{t^{2n}}{b_n}\\ &= \sum_{n=0}^{\infty} t^{-2n}(2t)^{-2\nu-1}\,\frac{|\,a_n\,|^2}{b_n} .\end{aligned}$$

Note that the function $u(x,t) = G(x,k;t)$ satisfies the conditions of the theorem for $0 < t < \infty$. In this case, we have

$$a_n = b_n k^{2n} ,$$

and hence

$$\sum_{n=0}^{\infty} t^{-2n}(2t)^{-2\nu-1}\,\frac{|\,a_n\,|^2}{b_n} = \left(\frac{1}{2t}\right)^{2\nu+1}\mathcal{F}\left(\frac{k^2}{2t}\right) ,$$

whereas

$$\int_0^\infty G(ix;t)\,|\,G(x,k;t)\,|^2\,d\mu(x) = \left(\frac{1}{2t}\right)^{2\nu+1}\mathcal{F}\left(\frac{k^2}{2t}\right) .$$

BIBLIOGRAPHY

1. F. M. Cholewinski and D. T. Haimo, *The Weierstrass-Hankel convolution transform*, J. d'Analyse Math. (to appear).
2. A. Erdelyi et al., *Higher transcendental functions*, vol. II, 1953.
3. D. T. Haimo, *Expansions in terms of generalized heat polynomials and of their Appell transforms*, J. Math. Mech. (to appear).
4. ———, *Integral equations associated with Hankel convolutions*, Trans. Amer. Math. Soc. (to appear).
5. ———, *Generalized temperature functions*, Duke Math. J. (to appear).
6. ———, *Functions with the Huygens property*, Bull. Amer. Math. Soc. **71** (1965), 528–532.

7. I. I. Hirschman, Jr., *Variation diminishing Hankel transforms*, J. d'Analyse Math. **8** (1960-61), 307-336.
8. P. C. Rosenbloom and D. V. Widder, *Expansions in terms of heat polynomials and associated functions*, Trans. Amer. Math. Soc. **92** (1959), 220-266.
9. E. C. Titchmarch, *The theory of Fourier integrals*, 1937.
10. G. N. Watson, *A treatise on the theory of Bessel functions*, 2nd ed., Cambridge, 1958.

SOUTHERN ILLINOIS UNIVERSITY AND HARVARD UNIVERSITY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

A CHARACTERIZATION OF GROUPS IN TERMS OF THE DEGREES OF THEIR CHARACTERS

I. M. ISAACS AND D. S. PASSMAN

The purpose of this paper is to study properties of groups which are related to the degrees of their absolutely irreducible characters and in particular to the biggest such degree.

Let p be a fixed prime. We say group G has r.x.e (representation exponent e) if the degrees of all the absolutely irreducible characters of G divide p^e. If G has a subinvariant abelian subgroup whose index divides p^e then by Ito's Theorem, G has r.x.e. We show conversely that if G has r.x.e then G has a subinvariant abelian subgroup whose index divides p^{4e}. While we do not obtain the best possible value for the exponent in the above bound, we do show that it is essentially a linear function of e.

We can obtain information about somewhat larger subgroups. We show that a group G with r.x.e has a subgroup H of index p^e with $[H : \mathfrak{Z}(H)] \leqq p^{3e(e+2)}$, where $\mathfrak{Z}(H)$ is the center of H. The latter bound is by no means best possible. However we show by example that a similar result cannot hold in general for subgroups of index less than p^e.

We study the case $e = 1$ in more detail and completely characterize all such groups. This generalizes a result of Amitsur which discusses the $p = 2$ situation. We prove that G has r.x.1 if and only if (i) G is abelian, (ii) G has a normal abelian subgroup of index p or (iii) $[G : \mathfrak{Z}(G)] = p^3$.

The previous results apply to rather special groups. We consider the more general case now. We say group G has r.b.n (representation bound n) if the degrees of all the absolutely irreducible characters of G are $\leqq n$. If G has an abelian subgroup A with $[G : A] \leqq n$ then as is easily seen G has r.b.n. Conversely we show here that there is a finite valued function h with the property that if G has r.b.n then G has an abelian subgroup A with $[G : A] \leqq h(n)$. This result can be viewed as an analog to Jordan's Theorem for complex linear groups of degree n.

The analogy between groups with r.b.n and linear groups of degree n can be carried further. We show that if G has r.b.n and if p is a prime with $p > n$ then G has a normal abelian Sylow p-subgroup.

Finally we discuss some extensions of the above results to infinite discrete groups.

1. Preliminary remarks.

The main tool of this paper is the character theory of finite groups. We record here our notation and

Received March 17, 1964.

nomenclature.

Let χ be an irreducible character of a group G and let φ be an irreducible character of a subgroup H. φ induces a character φ^* of G and χ restricts down to a character $\chi \mid H$ of H. The Frobenius Reciprocity Theorem ([5] Theorem 38.8) then states

$$[\chi, \varphi^*]_G = [\chi \mid H, \varphi]_H$$

where [,] denotes the inner product over the appropriate group.

Suppose $H \triangleleft G$ (H is normal in G). Then G acts on the irreducible characters of H by conjugation. That is for $g \in G$ and $x \in H$, $\varphi^g(x) = \varphi(gxg^{-1})$. The subgroup T fixing a given irreducible character φ is called the inertia group of φ. Clearly $T \supseteq H$. If $t = [G : T]$ (the index of T in G) then φ has precisely t distinct conjugates

$$\varphi = \varphi_1, \varphi_2, \cdots, \varphi_t .$$

If χ is a constituent of φ^* with multiplicity a then ([5] Theorem 49.7)

$$\chi \mid H = a(\varphi_1 + \varphi_2 + \cdots + \varphi_t) .$$

Let $H \triangleleft G$. Each character of G/H can be viewed in a natural way as a character of G with kernel containing H. Conversely every such character of G comes from a unique one of G/H in this manner. In general we will use the same symbol to denote the character whether viewed in G or G/H. The precise situation will always be clear from context.

Finally we denote by $\boldsymbol{C}[G]$ the group algebra of G over the complex numbers $\boldsymbol{C}$.

We mention here a few results which yield information about the degrees of the irreducible characters of G from a knowledge of the structure of the group.

PROPOSITION 1.1. (Ito). Let A be an abelian subinvariant subgroup of G. Then the degrees of the irreducible characters of G all divide $[G : A]$.

Usually this is stated with A normal. However the more general statement is true as can be seen from Theorem 53.17 of [5].

LEMMA 1.2. *Let H be a normal subgroup of G and φ an irreducible character of H with conjugates $\varphi = \varphi_1, \varphi_2, \cdots, \varphi_t$. Let $\chi_1, \chi_2, \cdots, \chi_r$ be the distinct irreducible constituents of φ^*. Then if $\chi_i \mid H = a_i(\varphi_1 + \varphi_2 + \cdots + \varphi_t)$ we have $[G : H] = a_1^2 t + a_2^2 t + \cdots + a_r^2 t$. In particular for each i, $a_i^2 t \leqq [G : H]$.*

Proof. We have $\deg \chi_i = a_i t \deg \varphi$. By Frobenius Reciprocity, χ_i occurs in φ^* with multiplicity a_i. Hence

$$[G : H] \deg \varphi = \deg \varphi^* = \Sigma a_i \deg \chi_i = (\Sigma a_i^2 t) \deg \varphi$$

and the result follows.

PROPOSITION 1.3. Let A be an abelian subgroup of G. Then the degrees of the irreducible characters of G are $\leqq [G : A]$.

Proof. Let χ be any character of G and let λ be a constituent of $\chi \mid A$. Then χ is a constituent of λ^* and hence $\deg \chi \leqq \deg \lambda^* = [G : A]$.

This result is true more generally without the assumption that G is finite (see Corollary 1 of [1]).

PROPOSITION 1.4. Let $Z = \mathfrak{Z}(G)$ be the center of G. Then the degrees of the irreducible characters of G are $\leqq [G : Z]^{1/2}$.

Proof. Let χ be any character of G and $\chi \mid Z = a \Sigma_1^t \lambda_i$. Since Z is the center of G we must have $t = 1$. Thus

$$\deg \chi = a = (a^2 t)^{1/2} \leqq [G : Z]^{1/2}$$

and the result follows.

This paper is concerned with the converse problem, namely to obtain information about the structure of the group from the degrees of its characters. We use the above propositions as guide lines to indicate the directions in which to look.

2. **Extra-special groups.** We study here the structure and characters of groups E with the following properties.

DEFINITION 2.1. We say a group E is extra-special if

(i) E is a non-abelian group containing a nontrivial normal abelian subgroup.

(ii) All nontrivial normal subgroups of E contain the commutator subgroup E'.

We mention first how these extra-special groups occur. Let G be any non-abelian group. Let K be normal in G and maximal with respect to G/K being non-abelian. Then clearly $E = G/K$ satisfies (ii). If for some reason (for example the solvability of G) we can guarantee that E has a nontrivial normal abelian subgroup, then E is extra-

special.

PROPOSITION 2.2. Let E be an extra-special group. Then E has the properties of one of the following two cases:

Case P: (i) E is a p-group for some prime p.

(ii) $\mathfrak{Z}(E)$ is cyclic and E' has order p and is contained in $Z = \mathfrak{Z}(E)$.

(iii) if β is any nonlinear absolutely irreducible character of E then β vanishes off Z and $\deg \beta = [E:Z]^{1/2}$.

Case Q: (iv) E is not a p-group for any prime p.

(v) $E' = Q$ is a normal, elementary abelian Sylow q-subgroup of E with E/Q cyclic.

(vi) every nonlinear character of E has degree $[E:Q]$.

(vii) if λ is a nonprincipal character of Q then λ has $[E:Q]$ distinct conjugates in E. For any linear character λ of Q and any element $g \in E$-Q there exists a character μ of Q with $\lambda = \mu^g\bar{\mu}$.

Proof. First suppose E is a p-group so we have (i) of Case P. Let Y be a cyclic subgroup of order p contained in the center $Z = \mathfrak{Z}(E)$. Since Y is normal in E we have $Y \supseteq E'$. But $E' > 1$ so we in fact have $Y = E'$ and $|E'| = p$. Thus Z has one and only one such cyclic subgroup of order p and so Z is cyclic. This yields (ii).

Now let β be a nonlinear character of E and let $g \notin Z$. Since g is not central there is an element h with $(g, h) = g^{-1}h^{-1}gh \neq 1$. But $(g, h) \in E' \subseteq Z$ so say $(g, h) = z$. Then $h^{-1}gh = gz$. Since β is a class function we have $\beta(g) = \beta(h^{-1}gh) = \beta(gz)$. Now $z \in Z$ so z is represented by the scalar matrix λI in the representation associated with β. If $\lambda = 1$ then z and hence E' is in the kernel of the representation. Since this is not the case $\lambda \neq 1$. But clearly $\beta(gz) = \lambda\beta(g)$ so we have $\beta(g) = \lambda\beta(g)$ and hence $\beta(g) = 0$. Finally for any $z \in Z$ we have clearly $|\beta(z)| = b$ where $b = \deg \beta$. Hence

$$1 = [\beta, \beta] = |G|^{-1} \Sigma_G \beta(g)\overline{\beta(g)} = |G|^{-1} \Sigma_Z \beta(g)\overline{\beta(g)} = b^2 |Z| / |G| .$$

Thus $b^2 = [G:Z]$ and (iii) follows.

Now we assume that E is not a p-group for any prime p. Let A be the nontrivial normal abelian subgroup of E guaranteed by Definition 2.1. If $|A|$ has two distinct prime factors say p and q then $\mathfrak{S}_p(A)$ (the Sylow p-subgroup of A) and $\mathfrak{S}_q(A)$ are characteristic in A and hence normal in E. Thus $\mathfrak{S}_p(A) \supseteq E'$ and $\mathfrak{S}_q(A) \supseteq E'$ so $E' = 1$, a contradiction. Hence A is a q-group for some prime q. Let Q be a Sylow q-subgroup of E containing A. Since $A \supseteq E'$ we see that Q is normal in E. Now E is solvable so by a theorem of P. Hall ([7] Theorem 9.3.1) Q has a complement B which is abelian since $B \simeq E/Q$.

We show first that Q is elementary abelian. Let $\Phi(Q)$ be the Frattini subgroup of Q. Suppose that $\Phi(Q) > 1$. Then since Φ is characteristic in Q it is normal in E and hence $\Phi(Q) \supseteq E'$. Thus $(B, Q) \subseteq \Phi(Q)$ so B has trivial action on $Q/\Phi(Q)$, the Frattini quotient of Q. But B is a q'-group so ([7] Theorem 12.2.2) B centralizes Q and $E = B \times Q$. Hence since $B > 1$, $E' \subseteq B \cap Q = 1$, a contradiction. Thus $\Phi(Q) = 1$ and Q is elementary abelian.

Now $\mathfrak{C}(Q) \supseteq Q$ where $\mathfrak{C}(Q)$ is the centralizer of Q. Suppose $\mathfrak{C}(Q) > Q$. Then if C is a complement for Q in $\mathfrak{C}(Q)$ we have $\mathfrak{C}(Q) = Q \times C$ and so $C = \mathfrak{S}_{q'}(\mathfrak{C}(Q))$. Thus C is normal in E and again $E' \subseteq Q \cap C = 1$, a contradiction. Thus Q is its own centralizer. Let $g \notin Q$ and set $D = \langle Q, g \rangle$. Now $\mathfrak{Z}(D) \subseteq Q$ and this is clearly the set of fixed points in Q under the action of g. Note since Q is abelian we have a well-defined action of E/Q on Q. Since $D \supset Q \supseteq E'$, D and hence $\mathfrak{Z}(D)$ is normal in E. If $\mathfrak{Z}(D) > 1$ then $\mathfrak{Z}(D) \supseteq E' \supseteq D'$ and D is nilpotent. But this would imply that g centralizes Q, a contradiction. Thus $\mathfrak{Z}(D) = 1$ and so g acts fixed point free on Q. Hence E/Q acts fixed point free on Q and E is a Frobenius group with kernel Q. Since E/Q is abelian, it is cyclic ([4] Theorem V, page 336). Let $x, y \in Q$ and $g \notin Q$. Then $x^g x^{-1} = y^g y^{-1}$ implies $(xy^{-1})^g = xy^{-1}$ so that $x = y$. Hence the $|Q|$ elements $x^g x^{-1}$ for all $x \in Q$ are all distinct. Thus $Q = E'$ and (v) follows.

Let λ be a linear character of Q. Suppose for some $g \notin Q$ we have $\lambda^g = \lambda$. Then for all $x \in Q$, $\lambda(x) = \lambda(gxg^{-1})$ so that $\lambda(gxg^{-1}x^{-1}) = 1$. But as above $gxg^{-1}x^{-1}$ takes on all elements of Q as x does. So λ is in fact the principal character of Q. Hence any non-principal character has $[E : Q]$ distinct conjugates. Since the characters $\mu^g\bar{\mu}$ are all distinct (vii) follows. Finally let β be a nonlinear character of E. Since $Q = E'$, $\beta \mid Q$ has a nonprincipal character as a constituent. But this has $[E : Q] = t$ distinct conjugates. Hence $\beta \mid Q = a \sum_1^t \lambda_i$. Since $a^2 t \leqq [E : Q]$, $a = 1$ and $\deg \beta = at = [E : Q]$. This completes the proof.

We remark that in the proof of (iii) above we actually obtained the more general result:

LEMMA 2.3. *Let F be a class 2 group with $|F'| = p$. If β is an irreducible nonlinear character of F then* $\deg \beta = [F : \mathfrak{Z}(F)]^{1/2}$.

3. Groups with r.x.e.

DEFINITION 3.1. Let p be a fixed prime. We say a group G has r.x.e (representation exponent e) if the degrees of all absolutely irreducible characters of G divide p^e. We say a group G has r.x. (e, s)

if G has r.x.e and $\mathfrak{S}_p(G)$ has r.x.s.

Let f be the function with the following property. If G has r.x.e then G has a subinvariant abelian subgroup whose index in G divides $p^{f(e)}$. Moreover f is the smallest such function.

Let g be the function with the following property. If G has r.x.(e, s) then G has a subinvariant abelian subgroup whose index in G divides $p^{g(e,s)}$. Moreover g is the smallest such function.

There is of course no apriori reason to assume that finite valued functions f and g exist. We show below that they do exist and give bounds for their values.

THEOREM A 3.2. *The functions f and g exist and satisfy*

$$f(0) = 0 \qquad 2e \leqq f(e) \leqq 4e - \log_2 4e$$
$$g(e, 0) = e \qquad e + \min(e, s) \leqq g(e, s) \leqq e + 3s - \log_2 4s\ .$$

This yields in particular

$$f(0) = 0,\quad f(1) = 2,\quad f(2) = 4 \text{ or } 5$$
$$g(e, 0) = e,\quad g(e, 1) = e + 1,\quad e + 2 \leqq g(e, 2) \leqq e + 3\ .$$

We will show in a later paper that $f(2) = 4$.

Let φ be an irreducible character of $\mathfrak{S}_p(G)$ and χ a constituent of φ^*. Then φ is a constituent of $\chi \mid \mathfrak{S}_p(G)$ so we have $\deg \varphi \leqq \deg \chi$. If G has r.x.e then $\deg \chi \leqq p^e$. But $\mathfrak{S}_p(G)$ is a p-group so the degrees of all its characters are powers of p and so $\mathfrak{S}_p(G)$ has r.x.e. Thus G has r.x.(e, e) and it clearly suffices to consider only the function g.

LEMMA 3.3. *Let $N \triangleleft G$. If G has* r.x.(e, s) *then so does N and G/N.*

Proof. Let φ be an irreducible character of N and χ a constituent of φ^*. Then since N is normal $\chi \mid N = a \Sigma_1^t \varphi_i$ where the φ_i are the t distinct conjugates of φ. Hence $\deg \chi = at \deg \varphi$. Since $\deg \chi$ divides p^e so does $\deg \varphi$ and N has r.x.e.

Choose a Sylow p-subgroup of N and extend it to one of G. Thus $\mathfrak{S}_p(N) \subseteq \mathfrak{S}_p(G)$. Let φ be a character of $\mathfrak{S}_p(N)$ and χ a constituent of φ^* (induction to $\mathfrak{S}_p(G)$). Then φ is a constituent of $\chi \mid \mathfrak{S}_p(N)$ and so $\deg \varphi \leqq \deg \chi$. Since $\deg \varphi$ is a power of p and $\deg \chi \leqq p^s$, the first result follows. The second is obvious.

PROPOSITION 3.4. Let G have r.x.(e, s). Then G has a normal abelian Hall p'-subgroup $\mathfrak{H}(G)$.

Proof. We show first by induction on $|G|$ that G has a normal

$\mathfrak{H}(G)$. If G is a p'-group the result is of course trivial. So we assume that p divides $|G|$. From the equation

$$|G| = [G : G'] + \Sigma_i x_i^2$$

where the x_i^2 are the degrees of the nonlinear irreducible characters of G we can conclude that p divides $[G : G']$. Let K be the complete inverse image in G of a subgroup of index p in the abelian group G/G'. Then $K \triangleleft G$ so by the previous lemma K has r.x.(e, s). By induction K has a normal Hall p'-subgroup $\mathfrak{H}(K)$. Clearly $\mathfrak{H}(K)$ is a Hall p'-subgroup of G and since it is characteristic in K, it is normal in G. Hence G has a normal $\mathfrak{H}(G)$.

Since $\mathfrak{H}(G)$ has r.x.(e, s) all its characters have degrees which are powers of p. But $|\mathfrak{H}(G)|$ is prime to p so all the characters are linear and $\mathfrak{H}(G)$ is abelian. This completes the proof.

LEMMA 3.5. *Let G be an arbitrary group. Suppose $K \triangleleft G$ with $E = G/K$ being an extra-special case P group. Let Z be the complete inverse image in G of the center of E. Let β be a nonlinear character of E viewed in G. Given any absolutely irreducible character φ of Z, if χ is a constituent of φ^* and if χ_1 is a constituent of $\chi\beta$ then*

$$(\deg \chi)\ (\deg \chi_1) \geqq p^m t(\deg \varphi)^2$$

where t is the number of distinct conjugates of φ and $\deg \beta = p^m$.

Proof. If $\deg \beta = p^m$ then by (iii) of Proposition 2.2 we have $[G : Z] = p^{2m}$. Also $\beta \mid Z = p^m\lambda$ where λ is a linear character of Z/K. Moreover since Z/K is central in G/K, λ is invariant under the action of G. That is $\lambda^x = \lambda$ for all $x \in G$.

Let χ be a constituent of φ^*. Then since $Z \triangleleft G$

$$\chi \mid Z = a\,\Sigma_1^t \varphi_i \qquad \varphi_1 = \varphi\ .$$

Now $\chi\beta$ vanishes off Z by (iii) again and

$$\chi\beta \mid Z = ap^m\,\Sigma_1^t \varphi_i\lambda\ .$$

Also $(\varphi\lambda)^*$ vanishes off Z and

$$(\varphi\lambda)^* \mid Z = (p^{2m}/t)\,\Sigma_1^t \varphi_i\lambda\ .$$

The first is clear. The second follows since $(\varphi\lambda)^x = \varphi^x\lambda^x = \varphi^x\lambda$ and $\deg(\varphi\lambda)^* = [G : Z] \deg \varphi\lambda = p^{2m} \deg \varphi\lambda$. Thus we have

$$\chi\beta = atp^{-m}(\varphi\lambda)^*.$$

Say $(\varphi\lambda)^* = \Sigma\, a_i\chi_i$. Then $\chi\beta = \Sigma\, a_i atp^{-m}\chi_i$ and hence $a_i at \geqq p^m$. In

particular $a_1at \geqq p^m$. By reciprocity $\chi_1 \mid Z = a_1 \Sigma_1^t \varphi_i\lambda$ so we have

$$\deg \chi = at \deg \varphi, \deg \chi_1 = a_1 t \deg \varphi\lambda = a_1 t \deg \varphi .$$

Hence

$$(\deg \chi)\ (\deg \chi_1) = aa_1t^2(\deg \varphi)^2 \geqq p^m t(\deg \varphi)^2$$

and the result follows.

We are now ready to prove the theorem.

Proof of Theorem A. The main difficulty is the upper bound. We show this by induction on $\mid G \mid$.

Since $G/\mathfrak{H}(G) \simeq \mathfrak{S}_p(G)$ the second parameter in r.x.(e, s) can be viewed as a bound for the characters of $G/\mathfrak{H}(G)$ and this is the approach we take. These of course correspond to the characters of G containing $\mathfrak{H}(G)$ in their kernel. Since the upper bound for the function g is always nonnegative we see that the result holds if G is abelian.

So we assume G is nonabelian. Choose $K \vartriangleleft G$ maximal with respect to G/K being nonabelian. Such a subgroup of course exists. Now $E = G/K$ always has a normal abelian subgroup by Proposition 3.4, hence E is extra-special. By Propositions 3.4 and 2.2 we see that E is either a p-group or a Frobenius group where E/E' is a cyclic p-group. This yields two cases to study. We discuss the latter first.

Case Q. Let Q be the complete inverse image in G of E', the Sylow q-subgroup of E. Then $[G : Q] = [E : E'] = p^m$ for some integer m. Let λ be a fixed linear character of Q which corresponds to a nonprincipal linear character of $E' = Q/K$.

Now let θ be any character of Q and let $T(\theta)$ be its inertia group. The product $\theta\lambda$ is also an absolutely irreducible character of Q. Suppose both $T(\theta) > Q$ and $T(\theta\lambda) > Q$. Since $G/Q \simeq E/E'$ is a cyclic p-group there is a unique minimal subgroup of G containing Q. Hence we can find $x \notin Q$ with $x \in T(\theta) \cap T(\theta\lambda)$. Then

$$\theta\lambda = (\theta\lambda)^x = \theta^x\lambda^x = \theta\lambda^x$$

so $\theta = \theta\lambda^x\overline{\lambda}$. But $x \notin Q$ so $\mu = \lambda^x\overline{\lambda}$ is not the principal character of E'. If N denotes its kernel in Q then $Q > N \supseteq K$, $N \vartriangleleft Q$ and since $\theta = \theta\mu$, θ vanishes off N. Say $\theta \mid N = a \Sigma_1^t \varphi_i$ with $\varphi_1 = \varphi$. Then deg $\theta = at \deg \varphi$. Since θ vanishes of N

$$1 = [\theta, \theta]_Q = (\mid N \mid / \mid Q \mid)\ [\theta \mid N, \theta \mid N]_N = (\mid N \mid / \mid Q \mid)\, a^2t$$

so $a^2t = [Q : N]$. Now a and t divide deg θ, a power of p. But they also divide $[Q : N]$ which is prime to p. Hence $a = t = 1$ and $[Q : N] =$

1, a contradiction.

Thus either $T(\theta) = Q$ or $T(\theta\lambda) = Q$ or both. Since $\deg \theta\lambda = \deg \theta$, it sufficies to assume $T(\theta) = Q$. Let χ be a constituent of θ^*. Then $\chi \mid Q = b \Sigma \theta_i$ where the θ_i are the p^m distinct conjugates of θ. So $p^m \deg \theta$ divides $\deg \chi$. Since G has r.x.e, Q therefore has r.x.(e-m). But Q also has r.x.(e, s) by Lemma 3.3. Hence Q has r.x.(e-m, s). Since $|Q| < |G|$ we have by induction a subinvariant abelian subgroup A of Q of index p^n with $n \leqq e\text{-}m + 3s - \log_2 4s$ (or $n \leqq e\text{-}m$ if $s = 0$). Then A is subinvariant in G with index p^{m+n} and here

$$m + n \leqq e + 3s - \log_2 4s \quad (\text{or } m + n \leqq e) .$$

Thus induction goes through in this case.

Case P. Let Z be the complete inverse image in G of the center of E. Let β be a nonlinear character of E viewed in G so that $\deg \beta = p^m$. If φ is a character of Z then by Lemma 3.5

$$(\deg \varphi)^2 \leqq 1/p^m t \ (\deg \chi) \ (\deg \chi_1) \leqq p^{2e-m}/t .$$

If in fact φ is a character of $Z/\mathfrak{H}(Z)$ then since $\mathfrak{H}(Z) = \mathfrak{H}(G)$ and β is a character of $G/\mathfrak{H}(G)$ we see that so are both χ and χ_1. Hence

$$(\deg \varphi)^2 \leqq 1/p^m t \ (\deg \chi) \ (\deg \chi_1) \leqq p^{2s-m}/t .$$

Thus clearly Z has r.x.(e-$m/2$, s-$m/2$).

If m is odd we do slightly better. If $\deg \varphi = p^r$ then $r \leqq e - m/2$ implies $r \leqq e - (m + 1)/2$. Hence we have:

(*) If m is odd G has a normal subgroup Z with $[G : Z] = p^{2m}$ and such that Z has r.x.(e-$(m + 1)/2$, s-$(m + 1)/2$) .

Now let m be even. Let $B > Z$ with $[B : Z] = p$. Since B/K is normal in G/K, B is normal in G. We show that B has r.x.(e-$m/2$, s-$m/2$). Both the e and the s terms follow in the same manner so it suffices to just discuss the e term here. Suppose B has a character θ of degree $> p^{e-m/2}$. Let $\theta \mid Z = a \Sigma_1^r \varphi_i$. Since $a^2 r \leqq [B : Z] = p$ by Lemma 1.2 and both a and r are powers of p it follows that $a = 1$ and $r = 1$ or p. But $\deg \varphi_1 \leqq p^{e-m/2}$ by the above so we must have $r = p$ and $\deg \varphi_1 = p^{e-m/2}$. Now φ_1 has at least p conjugates in B. Hence if φ_1 has t conjugates in G we have $t \geqq r = p$. But finally

$$p^{2e-m} = (\deg \varphi_1)^2 \leqq p^{2e-m}/t \leqq p^{2e-m-1},$$

a contradiction. Thus we have:

(**) If m is even G has a normal subgroup B with $[G : B] = p^{2m-1}$ and such that B has r.x.(e-$m/2$, s-$m/2$).

Suppose m is odd. By induction and (*) Z has a subinvariant abelian subgroup A with $[Z:A] = p^n$ where

$$n \leqq (e\text{-}(m+1)/2) + 3(s\text{-}(m+1)/2) - \log_2 4(s\text{-}(m+1)/2).$$

But A is subinvariant in G with $[G:A] = p^{n+2m}$ and

$$\begin{aligned} n + 2m &\leqq e + 3s - 2 - \log_2 4(s\text{-}(m+1)/2) \\ &= e + 3s - \log_2 4(4s - 2m - 2)\,. \end{aligned}$$

Clearly $s \geqq m$ since $\deg \beta = p^m$ so $4s - 2m - 2 \geqq 2s - 2 \geqq s$ for $s \geqq 2$ and the result follows here. If $s = m = 1$ then $n \leqq e\text{-}1$ so

$$n + 2m \leqq e + 1 = e + 3s - \log_2 4s$$

so the result follows.

Finally let m be even. By induction and (**) B has a subinvariant abelian subgroup A with $[B:A] = p^n$ where

$$n \leqq (e\text{-}m/2) + 3(s\text{-}m/2) - \log_2 4(s\text{-}m/2)\,.$$

But A is subinvariant in G with $[G:A] = p^{2m-1+n}$ and

$$\begin{aligned} 2m - 1 + n &\leqq e + 3s - 1 - \log_2 4(s\text{-}m/2) \\ &= e + 3s - \log_2 4(2s\text{-}m) \\ &\leqq e + 3s - \log_2 4s \end{aligned}$$

since $s \geqq m$. Thus the result follows and the upper bound in Theorem A is proved.

For the lower bound we need the following.

LEMMA 3.6. (i) $f(1) \geqq 2$
(ii) $f(e_1 + e_2) \geqq f(e_1) + f(e_2)$
(iii) $g(e, s) \geqq e - s + f(s)$ *for* $e \geqq s$.

Proof. Let G_i be a group with r.x.e_i which has no subinvariant abelian subgroup with index $p^{f(e_i)-1}$. Then $G_1 \times G_2$ has r.x.$(e_1 + e_2)$ and has no subinvariant abelian subgroup whose index divides $p^{f(e_1)+f(e_2)-1}$. Hence $f(e_1 + e_2) \geqq f(e_1) + f(e_2)$ and (ii) is proved.

Let G be a group with r.x.s which has no subinvariant abelian subgroup with index dividing $p^{f(s)-1}$. Let Q be an elementary abelian q-group with p^{e-s} generators and let H be the semi-direct product of Q with an element of order p^{e-s} where the element merely cyclically permutes a basis of Q. Clearly Q is a maximal abelian subgroup of H. By Proposition 1.1, H has r.x.$(e\text{-}s, 0)$ since $\mathfrak{C}_p(H)$ is cyclic. Then $G \times H$ has r.x.(e, s) but has no subinvariant abelian subgroup whose index divides $p^{e-s+f(s)-1}$. Thus $g(e, s) \geqq e\text{-}s + f(s)$ and (iii) follows.

Part (i) is a consequence of Example 3.7 below.

EXAMPLE 3.7. Let F_1 be the group of order p^6 generated by elements u, v, w, x, y, z all of period p with the additional relations: $x, y,$ and z are central and

$$(u, v) = x, \qquad (v, w) = y, \qquad (w, u) = z .$$

It is not hard to see that F_1 has no abelian subgroup of index p and that $[F_1 : \mathfrak{Z}(F_1)] = p^3$. By Proposition 1.4, F_1 has r.x.1.

We can now finish the proof of Theorem A. Since $f(1) \geqq 2$ we have by (ii) of Lemma 3.6 and induction that $f(e) \geqq 2e$. Then by (iii) of Lemma 3.6 if $e \geqq s$ we have $g(e, s) \geqq e\text{-}s + f(s) \geqq e\text{-}s + 2s = e + s$. If G has r.x.(e, s) and $s > e$ then clearly G has r.x.(e, e) so that $g(e, s) = g(e, e)$ in this case and the result follows.

While we do not have the exact value for the function f, the following may be of interest.

PROPOSITION 3.8. There exists a real number c with $2 \leqq c \leqq 4$ and $f(e) \sim ce$.

Proof. Set $\mu(e) = f(e)/e - f(1)/1$. By Theorem A, $2 \geqq \mu(e) \geqq 0$. Clearly it suffices to show that $\mu(e)$ has a limit as $e \to \infty$. By Lemma 3.6, $f(e_1) + f(e_2) \leqq f(e_1 + e_2)$ so we obtain

(*) $\qquad \mu(e + 1) \geqq e/(e + 1)\mu(e)$

(**) $\qquad \mu(ne) \geqq \mu(e)$ for all positive integers n.

Fix $e \geqq 1$ and let x be any positive integer. Say $(n + 1)e \geqq x > ne$. Then by (*) and (**).

$$\mu(x) \geqq \left(\frac{x - 1}{x}\right)\left(\frac{x - 2}{x - 1}\right) \cdots \left(\frac{ne}{ne + 1}\right)\mu(ne)$$

$$= \left(\frac{ne}{x}\right)\mu(ne) \geqq \frac{n}{n + 1}\,\mu(e) .$$

As $x \to \infty$, $n \to \infty$ and so $\liminf \mu(x) \geqq \mu(e)$. Since this is true for all e, $\liminf \mu(x) \geqq \limsup \mu(x)$ and by the boundedness of μ the limit exists.

We will need an additional result for our later study of groups with r.x.1. Since the abelian group occurs on the first step in this case in the inductive proof of Theorem A we obtain

LEMMA 3.9. *Let G have* r.x.1. *Then G has a normal abelian*

subgroup whose index divides p^2.

4. **Large centers.** Let G be a group with r.x.e. In the previous section we showed that some subgroup of index p^n with $2e \leqq n \leqq 4e$ is abelian. We now study subgroups of index p^e.

THEOREM B 4.1. *Let G be a group with r.x.e. Then G has a subgroup H of index p^e with $[H : \mathfrak{Z}(H)] \leqq p^{3e(e+2)}$.*

We first need certain counting formulas.

DEFINITION 4.2. Let A be an abelian subgroup of an arbitrary group G. Group A acts on G by conjugation and we call the orbits of this action the A-classes of G. Clearly G is partitioned into mutually disjoint A-classes. We denote by $r(A)$ the number of such classes.

If A is normal then G acts by conjugation on the linear characters of A. We denote by $s(A)$ the number of orbits in this action.

LEMMA 4.3. *Let G be a group with r.x.e. and let A be an abelian subgroup. Let $k_1, k_2, \cdots, k_r$ denote the A-classes of G with $r = r(A)$ and let x_j be a representative of k_j. Then*

(i) $|k_j| = [A : \mathfrak{C}_A(x_j)]$

(ii) $|k_1| + |k_2| + \cdots + |k_r| = |G|$

(iii) $r(A) \geqq |G|/p^e$ *with strict inequality if G is nonabelian.*

If A is normal we have in addition

(iv) $r(A) = [G : A]s(A)$.

Proof. Formulas (i) and (ii) are clear. Let $\boldsymbol{C}[G]$ be the group algebra of G over the complex numbers and let A be naturally embedded. We consider $\mathscr{C}(A)$, the centralizer of A in the group algebra. Let $\hat{k}_j$ denote the sum of the elements of k_j in $\boldsymbol{C}[G]$. It is clear that $\hat{k}_1, \hat{k}_2, \cdots, \hat{k}_r$ form a basis for $\mathscr{C}(A)$ so that $r(A) = \dim \mathscr{C}(A)$.

We now compute this dimension in another manner. $\boldsymbol{C}(G)$ is a direct sum of full matrix rings over the complex numbers. Say

$$\boldsymbol{C}[G] = \cdot \sum_{i=0}^{e} b_i[\boldsymbol{C}]_{p^i}$$

since G has r.x.e. In any such direct summand $[\boldsymbol{C}]_{p^i}$ the projection of A can be diagonalized. Thus $\mathscr{C}(A)$ contains all diagonal matrices in each term. Hence

$$\dim \mathscr{C}(A) \geqq \sum b_i p^i.$$

But

$$|G| = \sum b_i p^{2i} \leqq p^e \sum b_i p^i$$

with strict inequality if G is non-abelian since then $e \geqq 1$ and $b_0 \geqq 1$. Thus $r(A) = \dim \mathscr{C}(A) \geqq |G|/p^e$ and (iii) follows.

If A is normal we can compute the dimension more precisely. We consider one of the direct summands of the group algebra and assume A is diagonalized in it. Let χ be the character of this representation with $\chi | A = a \sum_1^t \lambda_i$. We write $a = a(\chi)$ and $t = t(\chi)$ to indicate their dependence on χ. It is easy to see that the dimension of the centralizer of A is equal to $a^2 t$. Hence

$$\dim \mathscr{C}(A) = \sum_\chi a(\chi)^2 t(\chi) .$$

In view of Lemma 1.2, we obtain by grouping these terms

$$r(A) = \dim \mathscr{C}(A) = [G : A]s(A)$$

and the proof is complete.

LEMMA 4.4. *Let G be a group with r.x.e and let A be an abelian subgroup. Suppose $[G : \mathfrak{C}(A)] \geqq p^{e+1}$. Then there exists an element $x \notin \mathfrak{C}(A)$ with*

$$[A : \mathfrak{C}_A(x)] < (p^{e+1} - 1)/(p - 1) = p^e + p^{e-1} + \cdots + 1 < p^{e+1}.$$

If G is a p-group then $[A : \mathfrak{C}_A(x)] \leqq p^e$.

Proof. Let $n = \min \{[A : \mathfrak{C}_A(x)] \mid x \notin \mathfrak{C}(A)\}$. We show that $n < (p^{e+1} - 1)/(p - 1)$.

If $g \in \mathfrak{C}(A)$ then the A-class containing g has only one element and there are $|\mathfrak{C}(A)|$ such classes. If $g \notin \mathfrak{C}(A)$ then the A-class containing g has at least n members by definition of n. Hence there are at most $(|G| - |\mathfrak{C}(A)|)/n$ such classes. By the previous lemma we have

$$|\mathfrak{C}(A)| + (|G| - |\mathfrak{C}(A)|)/n \geqq r(A) > |G|/p^e$$

since clearly G is non-abelian. Let $a = [G : \mathfrak{C}(A)]$. Then the above equation becomes

$$1 + (a - 1)/n > a/p^e$$

or $n/p^e < (a - 1)/(a - p^e) \leqq (p^{e+1} - 1)/(p^{e+1} - p^e)$ since $a \geqq p^{e+1}$. Hence $n < (p^{e+1} - 1)/(p - 1)$. If G is a p-group then n must be a power of p and so we have $n \leqq p^e$ in this case. This completes the proof.

Proof of Theorem B. We show by induction on i with $0 \leqq i \leqq 3e$ that G has a subgroup H_i of index p^{4e-i} with $[H_i : \mathfrak{Z}(H_i)] \leqq p^{i(e+2)}$. By

Theorem A we have the group H_0.

Suppose we have H_i with $i < 3e$. Let $A = \mathfrak{Z}(H_i)$. Suppose first that $[G : \mathfrak{C}(A)] \leqq p^e$. Then we can choose H_{i+1} with $\mathfrak{C}(A) \supseteq H_{i+1} > H_i$ with $[H_{i+1} : H_i] = p$. This follows from Proposition 3.4 and basic properties of p-groups. Then $\mathfrak{Z}(H_{i+1}) \supseteq \mathfrak{Z}(H_i)$ and

$$[H_{i+1} : \mathfrak{Z}(H_{i+1})] \leqq [H_{i+1} : H_i][H_i : \mathfrak{Z}(H_i)] \leqq p\ p^{i(e+2)} \leqq p^{(i+1)(e+2)}.$$

Now assume that $[G : \mathfrak{C}(A)] > p^e$. Since $\mathfrak{C}(A) \supseteq H_i$ and H_i has index p^{4e-i} we see that $[G : \mathfrak{C}(A)] \geqq p^{e+1}$. Then by Lemma 4.4 we can find an element $x \in G$ with $x \notin \mathfrak{C}(A)$ and hence $x \notin H_i$ with $[A : \mathfrak{C}_A(x)] \leqq p^{e+1}$. Let H_{i+1} be a subgroup with $\langle H_i, x\rangle \supseteq H_{i+1} > H_i$ and $[H_{i+1} : H_i] = p$. Then $\mathfrak{Z}(H_{i+1}) \supseteq \mathfrak{C}_A(x)$ so

$$\begin{aligned}[H_{i+1} : \mathfrak{Z}(H_{i+1})] &\leqq [H_{i+1} : H_i][H_i : \mathfrak{Z}(H_i)][\mathfrak{Z}(H_i) : \mathfrak{C}_A(x)] \\ &\leqq p\ p^{i(e+2)}p^{e+1} = p^{(i+1)(e+2)}\end{aligned}$$

and the induction follows.

With $i = 3e$ we have a subgroup $H = H_{3e}$ of index p^e with $[H : \mathfrak{Z}(H)] \leqq p^{3e(e+2)}$.

The bound given in the above theorem is certainly not best possible. However we can show by example that a result of this nature cannot be extended to subgroups of index $< p^e$. We first need another example of a group with r.x.1.

EXAMPLE 4.5. Let A be an elementary abelian p-group with pn generators. Let F_2 be the semi-direct product of A by an element of order p which permutes a basis of A in cycles of length p. By Proposition 1.1, F_2 has r.x.1. Also $[F_2 : \mathfrak{Z}(F_2)] = p\ p^{(p-1)n}$. Hence by choosing n sufficiently large we can make this index as large as we please.

Now set $G = F_2 \times F_2 \times \cdots \times F_2$ the direct product of e copies of F_2. Then G has r.x.e. Any subgroup H of G of index $< p^e$ must project fully onto one of the direct summands. Hence there can be no bound on $[H : \mathfrak{Z}(H)]$.

The following two propositions are more general than we will need to study groups of r.x.1. However they are of interest in their own right.

PROPOSITION 4.6. Let G be an arbitrary group. Let q be the smallest prime for which $\mathfrak{S}_q(G)$ is noncyclic. Given any n nonidentity elements $g_1, \cdots, g_n \in G$ with $n \leqq q$ then G has an absolutely irreducible representation χ with $\chi(g_i) \neq 1$ for $i = 1, 2, \cdots, n$. Moreover $n \leqq q$ is best possible.

Proof. We first show this for abelian G by induction on $|G|$. If G is cyclic the result is of course clear. So assume that G is not cyclic. Then G has a subgroup of type (q, q). This group has $q+1$ disjoint subgroups of order q. Since there are only $n \leqq q$ elements there is some subgroup K of order q not containing any of the g_i. Then $g_i \to \bar{g}_i \neq 1$ in G/K. If $\bar{q}$ is the smallest prime with $\mathfrak{S}_{\bar{q}}(G/K)$ noncyclic then $n \leqq q \leqq \bar{q}$. So by induction there is a linear character λ of G/K with $\lambda(\bar{g}_i) \neq 1$. View λ as a character of G and the result follows.

Now let G be arbitrary but suppose that all the g_i commute. Set $A = < g_i \mid i = 1, 2, \cdots, n >$. Then A is abelian and for any prime p, $\mathfrak{S}_p(A) \subseteqq \mathfrak{S}_p(G)$ for some Sylow p-subgroup of G. Thus if $\tilde{q}$ is the smallest prime with $\mathfrak{S}_{\tilde{q}}(A)$ noncyclic we have $n \leqq q \leqq \tilde{q}$. Then by the previous result A has a linear character λ with $\lambda(g_i) \neq 1$ for all i. Let χ be a constituent of λ^*. The λ is a constituent of $\chi \mid A$ so the g_i are not in the kernel of χ.

Finally let G be arbitrary. We prove the result by induction on n. If all the g_i commute and in particular if $n = 1$ the result follows by the above. Suppose then that say g_n and g_{n-1} do not commute. Set $h = (g_n, g_{n-1})$ so that $h \neq 1$. Then by induction we can find an absolutely irreducible representation $\mathfrak{x}$ of G with $\mathfrak{x}(g_i) \neq 1$ for $i = 1, 2, \cdots, (n-2)$ and $\mathfrak{x}(h) \neq 1$. But $\mathfrak{x}(h) = (\mathfrak{x}(g_n), \mathfrak{x}(g_{n-1})) \neq 1$ so certainly $\mathfrak{x}(g_n) \neq 1$ and $\mathfrak{x}(g_{n-1}) \neq 1$.

To see that $n \leqq q$ is best possible let G be abelian of type (q, q). If $g_1, \cdots, g_{q+1}$ are generators of the $q+1$ proper subgroups of G then one is always in the kernel of every linear character. This completes the proof.

PROPOSITION 4.7. Let G be a group with r.x.e and suppose all primes q dividing $|G|$ are larger than e. Let $A_0 < A_1 < \cdots < A_m$ be a series of normal abelian subgroups of G such that

$$\mathfrak{C}(A_0) > \mathfrak{C}(A_1) > \cdots > \mathfrak{C}(A_m) .$$

Then $m \leqq e$.

Proof. It suffices to show that no such series exists with $m = e+1$. So suppose $m = e+1$ in the above and set $N_i = \mathfrak{C}(A_i)$. Then by assumption N_i centralizes A_i but not A_{i+1} so we have $1 \neq (N_i, A_{i+1})$. Choose $x_i \in (N_i, A_{i+1})$ with $x_i \neq 1$ for $i = 0, 1, \cdots, e$. These are $e+1 \leqq q$ nonidentity elements of G for all q dividing $|G|$. Thus by Proposition 4.6, G has an absolutely irreducible representation $\mathfrak{x}$ with $\mathfrak{x}(x_i) \neq 1$ for $i = 0, 1, \cdots, e$.

Set $\bar{N}_i = \mathfrak{x}(N_i)$, $\bar{A}_i = \mathfrak{x}(A_i)$ and since A_{e+1} is abelian we can assume

it is represented by diagonal matrices. Let $\mathscr{L}(\bar{A}_i)$ denote the linear subspace of matrices spanned by $\bar{A}_i$. If χ is the the character of $\mathfrak{x}$ with say $\chi \mid A_i = a_i \sum_1^{t_i} \lambda_j$ since $A_i \triangle G$ then clearly $\dim \mathscr{L}(\bar{A}_i) = t_i$ which is a power of p since t_i divides the degree of $\mathfrak{x}$. Set $t_i = p^{s_i}$.

Since $A_{i+1} > A_i$ we have $\mathscr{L}(\bar{A}_{i+1}) \supseteq \mathscr{L}(\bar{A}_i)$. But since $(\bar{N}_i, \bar{A}_{i+1}) \neq 1$, $\bar{N}_i$ centralizes $\mathscr{L}(\bar{A}_i)$ but not $\mathscr{L}(\bar{A}_{i+1})$. Hence $\mathscr{L}(\bar{A}_{i+1}) > \mathscr{L}(\bar{A}_i)$ and so $s_{i+1} > s_i$ for $i = 0, 1, \cdots, e$. Since $s_0 \geqq 0$ we have $s_{e+1} \geqq e + 1$. But G has r.x.e so $\deg \mathfrak{x} \leqq p^e$ and since $\mathscr{L}(\bar{A}_{e+1})$ is a space of diagonal matrices $\dim \mathscr{L}(\bar{A}_{e+1}) = p^{s_{e+1}} \leqq p^e$, a contradiction. Thus the result follows.

THEOREM C 4.8. *Group G has r.x.1 if and only if G is one of the following types:*

(i) *An abelian group*

(ii) *A group with a normal abelian subgroup of index p*

(iii) *A group with center of index p^3.*

Proof. Let us assume that G is a non-abelian group which has no normal abelian subgroup of index p. If G has r.x.1 then by Lemma 3.9 G has a normal abelian subgroup A of index p^2. Clearly $A = \mathfrak{C}(A)$.

Let K be a subgroup of G with $G > K > A$. Since G/A is of order p^2, K is normal in G. Clearly $\mathfrak{Z}(K) < A$. Set $A_0 = \mathfrak{Z}(G)$, $A_1 = \mathfrak{Z}(K)$ and $A_2 = A$. Then $A_0 \subseteqq A_1 < A_2$. Suppose that $A_0 \neq A_1$. Then we have $A_0 < A_1 < A_2$ and

$$\begin{array}{ccccc} \mathfrak{C}(A_0) & > & \mathfrak{C}(A_1) & > & \mathfrak{C}(A_2) \\ \| & & \| & & \| \\ G & & K & & A \end{array} .$$

Clearly the hypotheses of Proposition 4.7 are satisfied by G and so this cannot occur. Hence $\mathfrak{Z}(K) = \mathfrak{Z}(G)$.

Let $H = \mathfrak{H}(G)$. We show now that H is central. In view of the above it suffices to show that $\mathfrak{C}(H) > A$. Let $P = \mathfrak{S}_p(A)$. Then $A = P \times H$ and $P \triangle G$. Now the group $\bar{G} = G/P$ has r.x.(1, 0) since its Sylow p-subgroup has order p^2. Thus by Theorem A, $\bar{G}$ has an abelian subgroup of index p. Let B be its complete inverse image in G. Then $B \supseteqq H$ but $B \nsubseteqq A$. Choose $x \in B\text{-}A$ with $x^p \in A$. Since $\bar{B}$ is abelian, $(x, H) \subseteqq P$. But H is normal so $(x, H) \subseteqq H$. Hence $(x, H) = 1$ and $H \subseteqq \mathfrak{Z}(K) = \mathfrak{Z}(G)$ where $K = \langle A, x \rangle$.

Thus $G = H \times \mathfrak{S}_p(G)$ and it suffices to assume that G is a p-group. Now by Lemma 4.4 since $[G : A] = p^2$ there is an element $x \notin \mathfrak{C}(A) = A$ with $[A : \mathfrak{C}_A(x)] = p$. We can assume that $x^p \in A$. Then with $K = \langle A, x \rangle$ we have $[A : \mathfrak{Z}(K)] = [A : \mathfrak{Z}(G)] = p$. Hence $[G : \mathfrak{Z}(G)] = p^3$.

Conversely by Propositions 1.1 and 1.4 any group of type (i), (ii) or (iii) has r.x.l.

That all of the above possibilities occur can be seen by examples. Type (i) is clear. Group F_2 is an example of type (ii) which is neither (i) nor (iii). Finally F_1 is type (iii) but not (i) or (ii).

5. Groups with r.b.n. The results of the previous sections apply to rather special groups. We consider the more general case now.

DEFINITION 5.1. We say group G has r.b.n (representation bound n) if the degrees of all the absolutely irreducible characters of G are $\leqq n$.

Let h_1 be the function with the following property. If G has r.b.n then G has a subinvariant abelian subgroup whose index is at most $h_1(n)$. Moreover h_1 is the smallest such function.

Let h_2 (respectively h_3) be the corresponding function where we restrict G to the class of solvable (respectively nilpotent) groups.

We show here that the h_i exist and obtain bounds for each. This may be viewed as an analog of Jordan's Theorem ([14] Satz 200) which we state below.

PROPOSITION 5.2. (Jordan). There exists a function J with the following property. If G is a complex linear group of degree n then G has a normal abelian subgroup with index $\leqq J(n)$.

We assume that the function J above is best possible. One of the known bounds for J is

$$J(n) \leqq n!\ 12^{n(\pi(n+1)+1)}$$

where $\pi(n+1)$ is the number of primes $\leqq n+1$.

We prove

THEOREM D 5.3.
$$h_1(n) \leqq J(2n)$$
$$h_2(n) \leqq n^{3/2 \log_2 2n}$$
$$h_3(n) \leqq n^4.$$

The third case is an immediate corollary of Theorem A. Let G be nilpotent so that $G = P_1 \times P_2 \cdots \times P_r$ where P_i is a Sylow p_i-subgroup of G. Suppose the degree of the biggest character of P_i is $p_i^{e_i}$. Then the biggest character of G has degree $\prod_i p_i^{e_i} \leqq n$. Now by Theorem A, P_i has a subinvariant abelian subgroup A_i with $[P_i : A_i] \leqq p_i^{4e_i}$. If $A = A_1 \times A_2 \times \cdots \times A_r$ then A is a subinvariant abelian subgroup of G with index $\leqq \prod p_i^{4e_i} \leqq n^4$. Hence we need only

consider the first two cases.

We use a variant on the techniques of § 2. We first need a few simple facts about the function J.

LEMMA 5.4. *For all* $n \geqq 1$, $J(n) \geqq n^3$. *For all* n_1, n_2 *we have* $J(n_1)J(n_2) \leqq J(n_1 + n_2)$.

Proof. The symmetric group S_{n+1} has a faithful irreducible character of degree n. Since for $n \geqq 4$, S_{n+1} has no normal abelian subgroup we have $J(n) \geqq (n+1)!$ for $n \geqq 4$. But then

$$J(n) \geqq (n+1)! \geqq (n+1)n(n-1)2 = 2n(n^2-1) \geqq n^3.$$

Thus we need only consider $n = 1,2,3$. Case $n = 1$ is clear.

Let H be a linear group of degree 2. Then $H/\mathfrak{Z}(H)$ is a collineation group of degree two. These are known to be ([12] Chapter X) (1) cyclic or dihedral, (2) order < 60 and (3) the alternating group A_5 of order 60. From this we conclude immediately that $J(2) = 60$. Hence $60 = J(2) > 2^3$ and $J(3) \geqq J(2) = 60 > 3^3$. The first result follows.

Let G_i be a linear group of degree n_i having no normal abelian subgroup of index less than $J(n_i)$ for $i = 1, 2$. Then $G = G_1 \times G_2$ is a linear group of degree $n_1 + n_2$ having no normal abelian subgroup of index less than $J(n_1)J(n_2)$. Hence $J(n_1 + n_2) \geqq J(n_1)J(n_2)$ and the result follows.

LEMMA 5.5. *Given* $K \triangleleft G$ *with* G/K *non-abelian. If* G *has r.b.n then* K *has r.b.*$(n/2)$.

Proof. Let φ be an absolutely irreducible character of K and let χ be a constituent of φ^*. Set $\chi \mid K = \psi$. Then $\psi = a\sum_1^t \varphi_i$ with $\varphi_1 = \varphi$. Now $\psi^* = [G:K]\chi$ on K and $\psi^* = 0$ off K. Hence if ρ denotes the character of the regular representation of G/K in G we have $\psi^* = \chi\rho$. If $\{\beta_i\}$ is the set of absolutely irreducible characters of G/K in G with $b_i = \deg \beta_i$ then $\rho = \sum b_i\beta_i$.

We compute $[\psi^*, \psi^*]$. By Frobenius reciprocity this is equal to

$$[\psi, \psi^* \mid K]_K = [G:K]\ [\psi, \psi]_K .$$

On the other hand $\psi^* = \chi\rho = \sum b_i\chi\beta_i$ so we have

$$[G:K]\ [\psi, \psi]_K = [\psi^*, \psi^*] = \sum_{ij} b_i b_j\, [\chi\beta_i, \chi\beta_j] .$$

Suppose for some nonlinear character $\beta = \beta_i$ we have that $\chi\beta$ is irreducible. Then

$$n \geqq \deg \chi\beta = (\deg \chi)(\deg \beta) \geqq 2 \deg \chi .$$

So $\deg \chi \leqq n/2$. But $\deg \varphi \leqq \deg \chi$ so the result follows here.

If for some nonlinear $\beta = \beta_i$ we have that $\chi\beta$ is reducible then $[\chi\beta, \chi\beta] > 1$. Hence

$$[G : K]\ [\psi, \psi]_K > \Sigma\, b_i^2 = [G : K]$$

so that $[\psi, \psi]_K > 1$. Now $[\psi, \psi]_K = a^2 t > 1$ implies that $at \geqq 2$. But

$$n \geqq \deg \chi = at \deg \varphi \geqq 2 \deg \varphi$$

so $n/2 \geqq \deg \varphi$ in this case also. Hence K has r.b.$(n/2)$ and the result follows.

We are now ready to prove the theorem.

Proof of Theorem D. The proof is by induction on $|G|$. If G is abelian the result is trivial so we assume that G is non-abelian. Choose $K \triangleleft G$ maximal with respect to G/K being non-abelian. If $E = G/K$ has a nontrivial normal abelian subgroup then E is extra-special and we have two possibilities by Proposition 2.2. Thus there are three cases to consider: Case P, Case Q and Case R, the latter being the possibility that E has no nontrivial normal abelian subgroup. Note Case R cannot occur if G is solvable.

We point out the obvious fact that the r.b.n property (as well as solvability) is inherited by subgroups and quotient groups.

Case P: Let φ be a character of Z as in Lemma 3.5. Then

$$(\deg \varphi)^2 \leqq 1/p^m t\ (\deg \chi)(\deg \chi_1)$$

so $\deg \varphi$ is properly smaller than at least one of $\deg \chi$ and $\deg \chi_1$. Since $\deg \varphi$ divides both $\deg \chi$ and $\deg \chi_1$ we have, if say $\deg \varphi < \deg \chi$, that $\deg \varphi \leqq 1/2 \deg \chi \leqq n/2$. Thus Z has r.b.$(n/2)$. Now $[G : Z] = p^{2m} \leqq n^2$ since p^m is the degree of a character of G. Thus

(*) G has a normal subgroup Z of index $\leqq n^2$ with r.b.$(n/2)$.

Case Q: Let Q be the inverse image in G of E' the Sylow q-subgroup of E. Now $[G : Q]$ is the degree of a character of G by (vi) of Proposition 2.2. Hence $[G : Q] \leqq n$. If Q has r.b.$(n/2)$ then

(**) G has a normal subgroup Q of index $\leqq n$ with r.b.$(n/2)$.

We assume now that this is not the case. Let θ be a character of Q of degree $> n/2$. If χ is a constituent of θ^* then $\chi \mid Q = a \Sigma_1^t \theta_i$ and $\deg \chi = at \deg \theta$. But $\deg \chi \leqq n$ so $a = t = 1$. Hence for all $x \in G - Q$ we have $\theta^x = \theta$. Let μ be any linear character of Q

corresponding to one of E'. Then $\theta\mu$ has degree $> n/2$ so also $(\theta\mu)^x = \theta\mu$. Thus

$$\theta\mu = (\theta\mu)^x = \theta^x\mu^x = \theta\mu^x$$

and $\theta = \theta\mu^x\bar{\mu}$. Hence θ vanishes off the kernel of $\mu^x\bar{\mu}$. Now given any linear character λ of Q/K we can find a character μ with $\lambda = \mu^x\bar{\mu}$ by (vii) of Proposition 2.2. Hence θ vanishes off the kernel of λ. But the intersection of all these kernels is K and so θ vanishes off K.

Let $\theta \mid K = a\,\Sigma_1^t \varphi_i$. Then $[\theta \mid K, \theta \mid K]_K = a^2t$. On the other hand since θ vanishes off K

$$1 = [\theta, \theta]_Q = (|K|/|Q|)\ [\theta \mid K, \theta \mid K]_K = (|K|/|Q|)\, a^2t\ .$$

Hence

$$[Q : K] = a^2t \leqq (at \deg \varphi_1)^2 = (\deg \theta)^2 \leqq n^2.$$

So $[G : K] = [G : Q][Q : K] \leqq n^3$. Since G/K is non-abelian, K has r.b.$(n/2)$ by Lemma 5.5 so we have

(***) G has a normal subgroup K of index $\leqq n^3$ with r.b.$(n/2)$.

Case R: Suppose $E = G/K$ has no normal abelian subgroup. Let χ be a nonlinear character of E with kernel N. If $N > 1$ then $N \supseteqq E'$ and so χ would be linear. Hence $N = 1$ so E is a faithful linear group of degree $m \leqq n$ since G has r.b.n. But then E has a normal abelian subgroup of index $\leqq J(m) \leqq J(n)$ by Proposition 5.2. Hence $[G : K] \leqq J(n)$. Since E is non-abelian, K has r.b.$(n/2)$ by Lemma 5.5. Thus

(****) G has a normal subgroup K of index $\leqq J(n)$ with r.b.$(n/2)$.

We now apply induction. If G is given to be solvable then so is every quotient and subgroup. Thus only Cases P and Q occur. Then (*), (**), and (***) imply that G has a normal subgroup N of index $\leqq n^3$ and with r.b.$(n/2)$. By induction N has subinvariant abelian subgroup A with

$$[N : A] \leqq (n/2)^{3/2 \log_2 n}.$$

Then A is subinvariant in G with

$$[G : A] \leqq n^3(n/2)^{3/2 \log_2 n} = n^{3/2 \log_2 2n}$$

and the bound for h_2 follows.

In the general case all four possibilities can occur. Since by Lemma 5.4, $J(n) \geqq n^3$ we see that G has a normal subgroup N of

index $\leqq J(n)$ with r.b.$(n/2)$. By induction N has a subinvariant abelian subgroup A with $[N:A] \leqq J(n)$. Then A is subinvariant in G with

$$[G:A] = [G:N][N:A] \leqq J(n)J(n) \leqq J(2n)$$

by Lemma 5.4 again. This completes the proof.

6. **Normal Sylow p-subgroups.** The analogy between groups with r.b.n and linear groups of degree n can be carried further. We show

THEOREM E 6.1 *Let G be a group with r.b.n. If p is a prime with $p > n$ then $\mathfrak{S}_p(G)$ is normal and abelian.*

If $n = 1$ the result is trivial. Thus we assume that $n \geqq 2$ and hence $p > 2$. Let G be a group of order $g = p^a g_0$ with g_0 prime to p and $a \geqq 1$. Let $\delta = \varepsilon\delta_0$ be a primitive gth root of unity over the rational numbers $\boldsymbol{Q}$ with ϵ a primitive p^ath root and δ_0 a g_0th root. Then as is well known, the Galois group of $\boldsymbol{Q}(\delta)/\boldsymbol{Q}(\delta_0)$ is cyclic of order $p^{a-1}(p-1)$. Let σ be an element of order $p-1$ in this group. Then σ is an automorphism of $\boldsymbol{Q}(\delta)$ which fixes $\boldsymbol{Q}(\delta_0)$ and cyclically permutes the $p-1$ primitive pth roots of unity. Moreover as is well known σ permutes the characters of G.

LEMMA 6.2. *Let G be a faithful linear group with character χ such that χ is fixed under the action of σ. Then* $\deg \chi \geqq p-1$. *If* $\deg \chi = p-1$ *and G has trivial center then G has no elements of order pq with q prime to p.*

Proof. Let x be an element of order p in G. Its eigenvalues (in the matrix representation) are all pth roots of unity. Since $\chi^\sigma(x) = \chi(x)$ we see that $\chi(x)$ is rational. From the form of the minimal polynomial of a primitive pth root μ over $\boldsymbol{Q}$ we see that x has at least $p-1$ eigenvalues. Hence $\deg \chi \geqq p-1$.

If $\deg \chi = p-1$ then the eigenvalues of x are $\mu, \mu^2, \cdots, \mu^{p-1}$. Clearly we can assume that x is in diagonal form so that

$$x = \operatorname{diag}(\mu, \mu^2, \cdots, \mu^{p-1}) .$$

Let y be an element of G of order q where q is prime to p. Suppose $xy = yx$. Since all the diagonal entries of x are distinct, y is also diagonal. Say $y = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_{p-1})$ and so each λ_i is a qth root of unity. Thus

$$\Sigma_i \lambda_i \mu^i = \chi(xy) = \chi^\sigma(xy) = \Sigma_i \sigma(\lambda_i \mu^i) = \Sigma_i \lambda_i \sigma(\mu^i)$$

since σ fixes $\boldsymbol{Q}(\delta_0)$.

Suppose $\sigma(\mu) = \mu^s$. Then with the subscripts taken modulo p we have

$$\Sigma_i \lambda_{is} \mu^{is} = \Sigma_i \lambda_i \mu^i = \Sigma_i \lambda_i \sigma(\mu^i) = \Sigma_i \lambda_i \mu^{is}$$

or

$$\Sigma_i (\lambda_i - \lambda_{is}) \mu^{is} = 0 .$$

Since the μ^{is} are linearly independent over $\boldsymbol{Q}(\delta_0)$ we have $\lambda_i = \lambda_{is}$ for each i. Hence $\lambda_1 = \lambda_s = \lambda_{s^2} = \cdots$ and all of the λ_i are equal since σ has order $p - 1$. Thus y is central.

Suppose z is an element of G of order pq. Then $x = z^q$ has order p and $y = z^p$ has order q. Moreover x and y commute. By the above $y \in \mathfrak{Z}(G) = 1$, a contradiction. Thus the lemma is proved.

Suppose $\sigma(\delta) = \delta^u$. Since σ is an automorphism u is prime to $g = |G|$. Now σ fixes $\boldsymbol{Q}(\delta_0)$ so we see that $u \equiv 1 \bmod g_0$. Let α be the permutation on the elements of G defined by $\alpha(x) = x^u$. If x is p-regular ($x^{g_0} = 1$) then $\alpha(x) = x$. Clearly α permutes the conjugacy classes of G. If χ is any character of G then $\chi^\sigma(x) = \chi(\alpha(x))$. It follows ([3], § 6 and 7) that the number of irreducible characters fixed by σ is equal to the number of classes fixed by α. Since both σ and α have order $p - 1$, the number of transitivity classes in the respective actions of groups $\langle\sigma\rangle$ and $\langle\alpha\rangle$ on the characters and classes are equal.

LEMMA 6.3. *Let G have r.b.$(p-1)$. Suppose $G' > 1$ is the unique normal subgroup of G and assume that $[G:G']$ divides p. Then p is prime to $|G|$.*

Proof. Suppose to the contrary that $|G| = g = p^a g_0$. Now G has $t = [G:G']$ linear characters and clearly all nonlinear characters are faithful. If all nonlinear characters of G have degree $p - 1$ we have

$$g = t + b(p-1)^2 .$$

But $p - 1$ being the degree of an irreducible character must divide g. Thus $p - 1$ divides $t = 1$ or p, a contradiction for $p > 2$.

Since G has r.b.$(p - 1)$ some nonlinear faithful character of G has degree $k < p - 1$. Thus by Feit's theorem ([6]), $[G : \mathfrak{O}_p(G)]$ is divisible by p to at most the first power. Here $\mathfrak{O}_p(G)$ is the maximal normal p-subgroup of G. Since G is not a p-group (having r.b.$(p-1)$ it would be abelian) we see that $\mathfrak{O}_p(G) = 1$. Hence p divides $|G|$ to the first power only and so $\mathfrak{S}_p(G)$ is cyclic of order p.

Since G has at least two p-regular classes, including the identity class, α fixes at least two classes and hence σ fixes at least two charac-

ters. Now σ fixes the principal character but no other linear character since $[G : G'] = 1$ or p. Hence σ fixes a nonlinear faithful irreducible character χ and Lemma 6.2 applies. Since G has r.b.$(p-1)$, χ has degree $p-1$ and thus G has no elements of order pq with q prime to p. Note $\mathfrak{Z}(G) = 1$ since otherwise G would be a cyclic extension of $G' \subseteqq \mathfrak{Z}(G)$ and hence would be abelian.

This implies that every p-singular (order divisible by p) class consists of elements of order p. Thus if x and y are p-singular elements of G then x is conjugate to some power of y by the second Sylow theorem. Hence all p-singular classes form a single transitivity class under the action of $\langle\alpha\rangle$. We conclude then that the total number of $\langle\alpha\rangle$-transitivity classes is equal to the number of α-fixed classes plus one and the same is therefore true for σ. Thus all irreducible characters moved by σ are $\langle\sigma\rangle$-conjugate. Therefore their number $r > 1$ divides $p-1$ and their degrees are all equal to k. Hence we have

$$g = t + rk^2 + b(p-1)^2.$$

Since r divides $p-1$ and $p-1$ divides g we see that r divides t. But $t = 1$ or p so $r = 1$, a contradiction, and the result follows.

Proof of Theorem E. If G has a normal Sylow p-subgroup then the latter is clearly abelian. We show by induction on $|G|$ that $\mathfrak{S}_p(G) \triangleleft G$.

We can clearly assume that p divides $|G|$ and that G is non-abelian. By Lemma 6.3, G has a normal subgroup N with $G > N > 1$ and $[G : N] \neq p$. Suppose that p divides $|G/N|$. Now G/N has r.b.n so G/N has a normal Sylow p-subgroup by induction. If G/N is not p-group then $M = N\mathfrak{S}_p(G/N)$ is normal in G, $G > M > 1$ and p divides $|M|$. If G/N is a p-group then since $|G/N| \neq p$ we can again find such a normal subgroup M. If p does not divide $|G/N|$ then p divides $|N|$. Hence G has a proper normal subgroup M with p dividing $|M|$.

Now $G > M$ so by induction $\mathfrak{S}_p(M)$ is normal in M. Since $\mathfrak{S}_p(M)$ is then in fact characteristic, it is normal in G. Set $K = G/\mathfrak{S}_p(M)$. Then $|K| < |G|$ so K has a normal Sylow p-subgroup. Hence $\mathfrak{S}_p(G) = \mathfrak{S}_p(M)\mathfrak{S}_p(K)$ is normal in G. This completes the proof.

7. Infinite groups with r.b.n. We show here that Theorem D holds for arbitrary (not necessarily finite) discrete groups.

DEFINITION 7.1. Let G be an arbitrary discrete group and let $\boldsymbol{C}[G]$ be its group algebra over the complex numbers. If $\mathscr{T}$ is an irreducible representation of the algebra then $\mathscr{T}(\boldsymbol{C}[G]) = P$ is primitive and hence is isomorphic to a dense set of linear transformations

over D, the commuting ring of $\mathscr{T}$ ([10] page 28). Let L be the center of D. If $\dim_L P < \infty$ then we say $\mathscr{T}$ is finite and since P is central simple over L ([10] page 122) we have $\dim_L P = m^2$. We set $m = \deg \mathscr{T}$, the degree of $\mathscr{T}$.

If G is finite then $\boldsymbol{C}$ is always the commuting ring of every irreducible representation. Thus the degree defined above agrees with the usual one in this case.

We say G has r.b.n if all irreducible representations $\mathscr{T}$ of $\boldsymbol{C}[G]$ are finite and for each $\deg \mathscr{T} \leqq n$.

The following lemma was obvious for finite groups.

LEMMA 7.2. *If G has r.b.n then so does every subgroup and quotient group.*

Proof. The fact for quotient groups is clear. Now since $\boldsymbol{C}[G]$ is semi-simple ([13] Theorem 5.2), a result of Amitsur ([1] Theorem 1) yields the fact that G has r.b.n if and only if $\boldsymbol{C}[G]$ satisfies the polynomial identity $[x_1, x_2, \cdots, x_{2n}] = 0$. But if H is a subgroup of G then $\boldsymbol{C}[G] \supseteq \boldsymbol{C}[H]$ and so the result follows.

PROPOSITION 7.3. A finitely generated linear group is a subdirect product of finite groups.

Proof. Let $G \subseteq [R]_m$ the ring of $m \times m$ matrices over field R. Let S be the prime field contained in R. Then $S = \boldsymbol{Q}$ if R has characteristic zero and $S = GF(p)$ if R has characteristic p. Let $g_1, g_2, \cdots, g_r$ generate G as a semi-group, that is we include with each generator of G its inverse. Each g_i is represented as a matrix over R. Let X denote the set of all the matrix entries for all the g_i. This is of course a finite set.

Given nonidentity $g \in G$ we must show that G has a normal subgroup N of finite index with $g \notin N$. Since $g \neq 1$ choose x to be a nonzero entry of $g - 1$ (subtraction in $[R]_m$). Now $X \cup \{x\}$ is a finite set so by the Extension Theorem for Places ([2] Contention, page 49) there exists a place ϕ of R into $\bar{S}$, the algebraic closure of S, union ∞, the ideal point, with $\phi(X \cup \{x\}) \subseteq \bar{S}$ and $\phi(x) \neq 0$. Now ϕ clearly induces a homomorphism of G into $[\bar{S}]_m$ with g not in the kernel. Since G is finitely generated it is clear that the image of G is in fact contained in $[T]_m$ where T is a finite algebraic extension of S. If $S = GF(p)$ then T is a finite field and so the image of G is a finite group. Hence if N denotes the kernel of the homomorphism we have $[G : N] < \infty$ and $g \notin N$. Thus the result follows here.

Hence it suffices to assume that $S = \boldsymbol{Q}$. Then T is a finite alge-

braic extension of $\boldsymbol{Q}$ and hence $\mathrm{Int}(T)$, the ring of algebraic integers in T, is a Dedekind domain. Now every element of T is of the form α/n where $\alpha \in \mathrm{Int}(T)$ and n is a rational integer. Let π denote the set of primes in the rational integers which divide the denominators of the $\phi(X)$ which of course generate all of the entries of $\phi(G)$ as a ring. Here we have let ϕ denote the natural extension of the place to the group homomorphism. Now π is clearly a finite set. Let $\phi(x) = \alpha/n$ so $\alpha \neq 0$. Then α is contained in only a finite number of prime ideals of $\mathrm{Int}(T)$. Thus we can find a prime ideal $\mathfrak{p}$ of $\mathrm{Int}(T)$ with $\alpha \notin \mathfrak{p}$ and such that $\mathfrak{p}$ does not divide any of the primes in π. Then clearly $\phi(G) \subseteq [U]_m$ where $U = \mathrm{Int}(T)_{\mathfrak{p}}$ is $\mathrm{Int}(T)$ localized at $\mathfrak{p}$. The map $\theta : U \to U/\mathfrak{p}U$ induces a homomorphism of $\phi(G)$ whose image is finite since $U/\mathfrak{p}U$ is a finite field. But by choice of $\mathfrak{p}$, $\phi(x)$ is not in the kernel of θ. Thus $\theta\phi$ is a homorphism of G into a finite group with g not in the kernel. This completes the proof.

We will have need for the following lemma of M. Hall ([11] page 56): If G is a finitely generated group then G has only a finite number of subgroups of any given finite index. The following proposition was proved in [8]. We offer a shorter proof here due to G. Glauberman.

PROPOSITION 7.4. Let G be an arbitrary group. Then G has an abelian subgroup with index at most m if and only if every finitely generated subgroup of G has such an abelian subgroup.

Proof. If A is abelian with $[G : A] \leqq m$ then for any subgroup H of G we have

$$m \geqq [G : A] \geqq [G \cap H : A \cap H] = [H : A \cap H] .$$

Hence $A \cap H$ is an abelian subgroup of H with index at most m.

Conversely let us assume that every finitely generated subgroup of G has an abelian subgroup of index at most m. For each finite subset α of G let $G_\alpha = \langle \alpha \rangle$ be the group generated by the elements in α. Let $\mathscr{A}_\alpha$ be the set of abelian subgroups of G_α and let $m_\alpha = \min\{[G_\alpha : A] \mid A \in \mathscr{A}_\alpha\}$. By assumption $1 \leqq m_\alpha \leqq m$ for each α. Choose α_0 such that $m_0 = m_{\alpha_0}$ is the largest of the m_α's and set $G_0 = G_{\alpha_0}$.

Let $A_1, A_2, \cdots, A_r$ be the abelian subgroups of G_{α_0} having index m_0 in G_{α_0}. By the above mentioned lemma of M. Hall there are only a finite number of these. We show that for some $i = 1, 2, \cdots,$ or r both $[G : \mathfrak{C}(A_i)] \leqq m_0$ and $\mathfrak{C}(A_i)$ is abelian. This will of course yields the result. Suppose not. Then for each i choose α_i to consist of two noncommuting elements of $\mathfrak{C}(A_i)$ if the latter is non-abelian or $m_0 + 1$ elements in distinct right cosets of $\mathfrak{C}(A_i)$ if $[G : \mathfrak{C}(A_i)] > m_0$. Let

$\alpha = \alpha_0 \cup \alpha_1 \cup \cdots \cup \alpha_r$. This is a finite set so let A_α be an abelian subgroup of G_α with $[G_\alpha : A_\alpha] = m_\alpha$. Now

$$m_0 \geqq m_\alpha = [G_\alpha : A_\alpha] \geqq [G_\alpha \cap G_0 : A_\alpha \cap G_0] = [G_0 : A_\alpha \cap G_0] \geqq m_0 .$$

The last inequality follows since $A_\alpha \cap G_0$ is an abelian subgroup of G_0. Thus $[G_0 : A_\alpha \cap G_0] = m_0 = m_\alpha$ and $A_\alpha \cap G_0 = A_i$ for some i. Say $A_\alpha \cap G_0 = A_1$.

Since A_α is abelian we have $A_\alpha \subseteqq \mathfrak{C}_{G_\alpha}(A_1)$. On the other hand

$$[G_\alpha : \mathfrak{C}_{G_\alpha}(A_1)] \geqq [G_\alpha \cap G_0 : \mathfrak{C}_{G_\alpha}(A_1) \cap G_0] = [G_0 : A_1] = m_0 = m_\alpha$$

since A_1 is its own centralizer in G_0. Thus $A_\alpha = \mathfrak{C}_{G_\alpha}(A_1)$. Now $\alpha_1 \subseteqq G_\alpha$. Hence if $\mathfrak{C}(A_1)$ were non-abelian then α_1 would contain noncommuting elements in $\mathfrak{C}_{G_\alpha}(A_1) = A_\alpha$. But A_α is abelian so this is not the case. On the other hand if $[G : \mathfrak{C}(A_1)] > m_0$ then G_α would contain $m_0 + 1$ elements in different right cosets of $\mathfrak{C}(A_1)$ and hence of $\mathfrak{C}_{G_\alpha}(A_1) = A_\alpha$. But $[G_\alpha : A_\alpha] = m_0$ so we have a contradiction here and the result follows.

We now consider the extension of Theorem D to arbitrary groups.

DEFINITION 7.5. Let $\tilde{h}_1$ be the function with the following property. If G is a not necessarily finite, discrete group with r.b.n then G has a subinvariant abelian subgroup A with $[G : A] \leqq \tilde{h}_1(n)$. Moreover $\tilde{h}_1$ is the smallest such function.

Let $\tilde{h}_2$ (respectively $\tilde{h}_3$) be the corresponding functions when we restrict G to the class of solvable (respectively nilpotent) groups.

THEOREM F 7.6. *The functions $\tilde{h}_i$ are finite valued. In fact $\tilde{h}_i(n) = h_i(n)$ for $i = 1, 2, 3$.*

Proof. First we notice that the property of being solvable or nilpotent is inherited by subgroups and quotient groups. Let $\tilde{h}$ be the function of interest here and let h be the corresponding function for finite groups. Clearly $\tilde{h}(n) \geqq h(n)$.

Let G have r.b.n. If F is a finitely generated subgroup of G then by Lemma 7.2, F has r.b.n. If we know that F has an abelian subgroup to index $\leqq h(n)$ for all such F then the result follows by Proposition 7.4. Hence it suffices to assume that G is finitely generated.

Let G be a finitely generated group with r.b.n. Since $C[G]$ is semi-simple, G is clearly a subdirect product of the groups $\mathscr{T}(G)$ for all irreducible representations $\mathscr{T}$ of the algebra. But $\mathscr{T}$ is finite so $\mathscr{T}(G)$ is a finitely generated linear group. By Proposition 7.3, $\mathscr{T}(G)$ and hence G is a subdirect product of finite groups.

Now let $A_1, A_2, \cdots, A_r$ be the subgroups of G of index $\leqq h(n)$. By M. Hall's lemma there are only a finite number of these. We show that for some i, A_i is abelian. Suppose not. Choose nonidentity $x_i \in A_i'$, the commutator subgroup of A_i, for $i = 1, 2, \cdots, r$. Since G is a subdirect product of finite groups we can find a normal subgroup N of G with $[G : N] < \infty$ and $x_i \notin N$ for all i. Now G/N is a finite group with r.b.n and hence G/N has an abelian subgroup of index $\leqq h(n)$. Let B be its complete inverse image in G. Then $[G : B] \leqq h(n)$ so $B = A_i$ for some i. Say $B = A_1$. Now B/N is abelian so $B' \subseteqq N$. But by choice of N, $A_1' \nsubseteqq N$. This is the required contradiction. Thus G has an abelian subgroup of index $\leqq h(n)$ and hence $\tilde{h}(n) \leqq h(n)$ and the result follows.

References

1. S. A. Amitsur, *Groups with representations of bounded degree II,* Illinois J. Math. **5** (1961), 198-205.
2. E. Artin, *Elements of Algebraic Geometry,* mimeographed notes, New York University, 1955.
3. R. Brauer, *On the connection between the ordinary and modular characters of groups of finite order,* Ann. of Math. **42** (1941), 926-935.
4. W. Burnside, *Theory of Groups Finite Order,* 2nd ed., Dover, New York, 1955.
5. C. W. Curtis and I. Reiner, *Representation Theory of Finite Groups and Associative Algebras,* Interscience, New York, 1962.
6. W. Feit, *Groups which have a representation of degree less than p-1,* Trans. Amer. Math. Soc. **112** (1964), 287-303.
7. M. Hall, Jr., *Theory of Groups,* Macmillan, New York, 1959.
8. I. M. Isaacs and D. S. Passman, *Groups with representations of bounded degree,* Canad. J. Math. **16** (1964), 299-309.
9. ———, *Groups whose irreducible representations have degrees dividing p^e,* Illinois J. Math. **8** (1964), 446-457.
10. N. Jacobson, *Structure of Rings,* Amer. Math. Soc. Colloquium Publications, vol. XXXVII.
11. A. G. Kurosh, *The Theory of Groups,* vol. II, 2nd English ed., Chelsea, New York 1960.
12. G. A. Miller, H. F. Blichfeldt and L. E. Dickson, *Theory and Applications of Finite Groups,* Dover, New York, 1961.
13. C. Rickart, *Uniqueness of norm in Banach algebras,* Ann. of Math. **51** (1950), 615-628.
14. A. Speiser, *Die Theorie der Gruppen von Endlicher Ordnung,* 4th ed., Birkhauser, Basel, 1956.

University of Chicago
University of California, Los Angeles

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

INTEGRAL INVARIANTS FOR VECTORS OVER LOCAL FIELDS

D. G. JAMES

This paper considers isometric invariants of vectors in lattices (quadratic forms) over the ring of integers in a local field for the prime 2. By extending the notion of order to vectors in the lattice we obtain a set of invariants which enable the general vector to be decomposed into a sum of simple vectors. The lengths of these simple vectors are invariant modulo certain powers of 2 and these lengths together with the original invariants form a complete set for the 2-adic integers. In the special case where there are no one dimensional, orthogonal sublattices (improper quadratic forms) the invariants form a complete set for all local fields.

Let F be a local field, that is a field complete with respect to a discrete, non-archimedean valuation with a finite residue class field, and R the ring of integers in F. Denote by $\nu(x)$ the order of the element x in F and by π a fixed prime in R; we may assume that $\nu(\pi) = 1$. Let V be a nonsingular lattice over R, i.e. a finite dimensional, torsion-free R-module with scalar product $\alpha \cdot \beta \in R$. An isometry φ of V is a one-to-one, linear transformation of V satisfying $\varphi(\alpha) \cdot \varphi(\beta) = \alpha \cdot \beta$. We investigate the isometric invariants of vectors $\alpha \in V$ in the unramified dyadic case, $\nu(2) = 1$. The p-adic case (p odd) has been considered by Rosenzweig [7] and an alternative approach by Ankeny (on the following lines) is included in the author's Ph. D. thesis [1].

In § 1 we develop the necessary structure theorems for V, much of which will also be found in [4], [5] and [6]; § 2 gives a complete set of invariants in the improper case; while § 3 treats the general case, but here we restrict ourselves to the 2-adic numbers, the results being more complex.

Frequent use is made of the following generalization of Hensel's lemma (see [3, p. 29]): if $f(x)$ is a polynomial with coefficients in R and $x_0 \in R$ is such that $\nu(f(x_0)) > 2\nu(f'(x_0))$, then $f(x)$ has a root $x \in R$ with $\nu(x - x_0) \geqq 1$.

1. We extend the notion of order to $\alpha \in V$ by defining $\nu(\alpha) = \min \nu(\alpha \cdot \beta)$, the minimum being taken over all $\beta \in V$. For each $r = 0, 1, 2, \cdots$ we denote by $V(r)$ the sublattice $\{\alpha \in V \mid \nu(\alpha) \geqq r\}$ and define $\nu_r(\alpha) = \min \nu(\alpha \cdot \beta)$, then minimum now being taken over all

Received August 20, 1964.

$\beta \in V(r)$. Then

$$\nu_r(\alpha) \leqq \nu_{r+1}(\alpha) \leqq \nu_r(\alpha) + 1 .$$

We write $V = U \oplus W$ if U and W are orthogonal sublattices together spanning V. Denote by $(\xi_1, \cdots, \xi_s)$ the sublattice spanned by $\xi_1, \cdots, \xi_s \in V$ and by U' the orthogonal complement of the sublattice U.

LEMMA 1. *Let $\xi \in V$. Then $V = (\xi) \oplus (\xi)'$ if and only if $\nu(\xi^2) = \nu(\xi)$.*

Proof. Clearly $(\xi) \oplus (\xi)' \subseteq V$. Take $\alpha \in V$. Then

$$\alpha = (\alpha \cdot \xi/\xi^2)\xi + (\alpha - (\alpha \cdot \xi/\xi^2)\xi) \in (\xi) \oplus (\xi)' .$$

This gives the reverse inclusion. Notice that $\alpha \cdot \xi/\xi^2 \in R$ since $\nu(\alpha \cdot \xi) \geqq \nu(\xi) = \nu(\xi^2)$.

LEMMA 2. *If $\lambda, \mu \in V$ satisfy $\nu(\lambda) = \nu(\mu) = \nu(\lambda \cdot \mu) < \nu(\lambda^2)$ then $V = (\lambda, \mu) \oplus (\lambda, \mu)'$.*

Proof. Let $\alpha \in V$ and write $\alpha = (k_1\lambda + k_2\mu) + (\alpha - k_1\lambda - k_2\mu)$ where $k_1 = ((\alpha \cdot \mu)(\lambda \cdot \mu) - (\alpha \cdot \lambda)\mu^2)k_3^{-1}$, $k_2 = ((\alpha \cdot \lambda)(\lambda \cdot \mu) - (\alpha \cdot \mu)\lambda^2)k_3^{-1}$ and $k_3 = (\lambda \cdot \mu)^2 - \lambda^2\mu^2$. Since $\nu(k_3) = 2\nu(\lambda)$ it follows that $k_1, k_2 \in R$. From $\lambda \cdot (\alpha - k_1\lambda - k_2\mu) = \mu \cdot (\alpha - k_1\lambda - k_2\mu) = 0$ we have $\alpha \in (\lambda, \mu) \oplus (\lambda, \mu)'$.

LEMMA 3. *If $V = (\xi) \oplus (\lambda, \mu)$ where $\nu(\xi) = \nu(\lambda) = \nu(\mu) = \nu(\lambda \cdot \mu)$ then there exist $\alpha, \beta, \gamma \in V$ such that $V = (\alpha) \oplus (\beta) \oplus (\gamma)$.*

Proof. We may assume by Lemma 1 that $\nu(\lambda^2) > \nu(\lambda)$ and $\nu(\mu^2) > \nu(\mu)$. Let $\alpha = \xi + \lambda$, $\beta = \xi - (\xi^2/\lambda \cdot \mu)\mu$ so that $\alpha \cdot \beta = 0$. Since $\nu(\alpha^2) = \nu(\alpha)$ and $\nu(\beta^2) = \nu(\beta)$ we are finished by Lemma 1.

We can now establish the main result on the structure of V. Let H_e denote a hyperbolic plane of the form (λ, μ) where $\lambda \cdot \mu = \pi^e$, $\nu(\lambda^2) > \nu(\lambda) = e$ and $\nu(\mu^2) > \nu(\mu) = e$. We call a sublattice of the form $V_e = H_e \oplus \cdots \oplus H_e$ *improper* and a sublattice of the form $V_e = (\xi_1) \oplus \cdots \oplus (\xi_t)$ with $\nu(\xi_i^2) = \nu(\xi_i) = e$, *proper*.

PROPOSITION 1. (O'Meara [5]) Let V be a nonsingular lattice over a local field. Then

$$V = V_{e_1} \oplus V_{e_2} \oplus \cdots \oplus V_{e_m} \qquad e_1 < e_2 < \cdots < e_m$$

with the sublattices V_{e_i}, $1 \leqq i \leqq m$, either proper or improper. Fur-

thermore, e_i, dim V_{e_i} and the forms of the V_{e_i}, $1 \leqq i \leqq m$, are invariants of V.

Proof. Choose $\alpha \in V$ to minimize $\nu(\alpha)$. Since the valuation is discrete there exists $\beta \in V$ such that $\nu(\alpha) = \nu(\beta) = \nu(\alpha \cdot \beta)$. If $\nu(\alpha^2) = \nu(\alpha)$ we split off the vector α using Lemma 1 and if $\nu(\alpha^2) > \nu(\alpha)$ we split off the plane (α, β). Proceeding in this manner we obtain the stated structure for V, after using Lemma 3 to get the correct form for the sublattices V_{e_i}.

We now establish the invariance of e_i and $\dim V_{e_i}$, $1 \leqq i \leqq m$. Let $\xi_1, \cdots, \xi_n$ and $\eta_1, \cdots, \eta_n$ be two bases of V arranged so that $\nu(\xi_i) \leqq \nu(\xi_{i+1})$ and $\nu(\eta_i) \leqq \nu(\eta_{i+1})$, $1 \leqq i \leqq n-1$. It suffices to show that $\nu(\xi_i) = \nu(\eta_i)$, $1 \leqq i \leqq n$. Let d be the first discrepancy and suppose that $\nu(\xi_d) > \nu(\eta_d) = f$. Then expressing η_i in terms of the ξ_i we see that

$$\begin{aligned} \eta_1 &= a_{11}\xi_1 + \cdots + a_{1d-1}\xi_{d-1} + \alpha_1 \\ &\cdots \\ \eta_d &= a_{d1}\xi_1 + \cdots + a_{dd-1}\xi_{d-1} + \alpha_d \,. \end{aligned}$$

In these equations $a_{ij} \in R$ and $\nu(\alpha_i) > f$. Eliminating the ξ_i we see that there exist $c_i \in R$, with at least one c_i a unit, such that $\sum_{i=1}^{d} c_i\eta_i = \sum_{i=1}^{d} c_i\alpha_i$. But $\nu(\Sigma c_i\eta_i) \leqq f$ and $\nu(\Sigma c_i\alpha_i) > f$ gives the required contradiction.

Finally we show that the form of V_{e_i} is invariant. Let

$$V = V_{e_1} \oplus \cdots \oplus V_{e_m} = V^*_{e_1} \oplus \cdots \oplus V^*_{e_m}$$

be two decompositions satisfying the conditions of the proposition and $\xi_1, \cdots, \xi_n$ the basis in the first case and $\eta_1, \cdots, \eta_n$ in the second. Suppose that $V_e = (\xi_h) \oplus \cdots \oplus (\xi_k)$ is proper. To prove that V_e^* is proper it is sufficient to show the existence of a vector λ orthogonal to $\eta_1, \cdots, \eta_{h-1}$ with $\nu(\lambda^2) = \nu(\lambda) = e$. If $\xi_h = \sum_{i=1}^{n} a_i\eta_i$ take $\lambda = \sum_{i=h}^{n} a_i\eta_i$; the required properties of λ follow from those of ξ_h.

Lattices with the same invariants above are said to be of the same *type*. A basis splitting V into the sum of lines and planes, as in Proposition 1, is called a *canonical basis*.

PROPOSITION 2. If $\alpha, \beta \in V$ satisfy $\nu(\alpha) = \nu(\beta) = \nu(\alpha \cdot \beta) = r$, $\nu(\alpha^2) \geqq r + 2$ and $\nu(\beta^2) \geqq r + 1$ then there exist $\lambda, \mu \in V$ such that $V = (\lambda, \mu) \oplus (\lambda, \mu)'$ where $\lambda^2 = \mu^2 = 0$, $\lambda \cdot \mu = \pi^r$ and $\alpha = \frac{1}{2}\pi^{-r}\alpha^2\lambda + \mu$.

Proof. Put $\lambda = x\alpha + \beta$ where x is a root of the equation $\alpha^2x^2 + 2\alpha \cdot \beta x + \beta^2 = 0$ (a root congruent to $x_0 = -\beta^2/2\alpha \cdot \beta$ exists by

Hensel's lemma). Then $\lambda^2 = 0$ and, multiplying by a unit if necessary, $\alpha \cdot \lambda = \pi^r$. Put $\mu = -\frac{1}{2}\pi^{-r}\alpha^2\lambda + \alpha$.

A hyperbolic plane (λ, μ) as in Proposition 2 is called *totally isotropic*.

PROPOSITION 3. (O'Meara [6]) Let $H_i = (\lambda_i, \mu_i)$, $i = 1, 2$, be two totally isotropic hyperbolic planes with $\nu(\lambda_1) = \nu(\lambda_2) = r$. Then an isometry $\varphi: H_1 \rightarrow H_2$ extends to an isometry of V.

Proof. It is sufficient to show that the orthogonal complements of H_1 and H_2 are isometric. Assume first that $\lambda_1 = \lambda_2$. Let $\eta_1, \cdots, \eta_{n-2}$ be a basis of H_1'. Let $\zeta_i = \eta_i - \pi^{-r}(\eta_i \cdot \mu_2)\lambda_1$, $1 \leqq i \leqq n-2$, so that we have $\zeta_i \cdot \zeta_j = \eta_i \cdot \eta_j$, $\zeta_i \cdot \lambda_2 = \zeta_i \cdot \mu_2 = 0$, $1 \leqq i, j \leqq n-2$. Since $\zeta_1, \cdots, \zeta_{n-2}$ is a basis of H_2' this case is finished. Now we assume there exist $\alpha \in H_1$, $\beta \in H_2$ such that $\nu(\alpha \cdot \beta) = r$. From symmetry we may take $\nu(\lambda_1 \cdot \lambda_2) = r$ and then the sublattice (λ_1, λ_2) is totally isotropic by Proposition 2. By the first part we now have that H_1' and H_2' are both isometric to $(\lambda_1, \lambda_2)'$. Finally we assume that $\nu(\alpha \cdot \beta) > r$ for all $\alpha \in H_1$, $\beta \in H_2$. The sublattice $(\lambda_1, \mu_1 + \mu_2)$ is now totally isotropic and its orthogonal complement is isometric to H_1' and H_2' by the second part of the proof.

For $\alpha \in V$ with $\nu(\alpha^2) = \nu(\alpha) = r$ we define $N_r(\alpha)$ to be 0 if there exists $\beta \in V(r)$ such that $\alpha \cdot \beta = 0$ and $\nu(\beta^2) = r$, and to be 1 otherwise. A vector $\alpha \in V$ is called *imprimitive* if $\alpha = \pi\beta$ with $\beta \in V$, otherwise it is *primitive*.

PROPOSION 4. Let $V = (\xi) \oplus (\xi)' = (\eta) \oplus (\eta)'$. Then there exists an isometry φ of V such that $\varphi(\xi) = \eta$ if and only if $\xi^2 = \eta^2$, $\nu(\xi) = \nu(\eta) = r$ and $N_r(\xi) = N_r(\eta)$.

Proof. For the general case see O'Meara [5, Theorems 5.1, 5.2]. See also Jones [2] for a discussion of related results in the 2-adic case.

The rest of this section will be devoted to constructing isometric invariants of vectors.

We call $r(\neq 0)$ a *critical index* of $\alpha \in V$ if

$$\nu_{r-1}(\alpha) = \nu_r(\alpha) < \nu_{r+1}(\alpha) ;$$

0 is a *critical index* of $\alpha \in V$ if

$$\nu(\alpha) < \nu_1(\alpha) .$$

The critical indices can only be the orders of basis vectors (the e_i of

Proposition 1) and hence a vector has only a finite number of critical indices. They are isometric invariants.

We say that the vectors $\lambda_1, \cdots, \lambda_m$ are *fully orthogonal* if $V = V_1 \bigoplus \cdots \bigoplus V_m$ where $\lambda_i \in V_i$, $1 \leqq i \leqq m$.

PROPOSITION 5. Let α have critical indices at $r_1, r_2, \cdots, r_s$ where $r_1 < r_2 < \cdots < r_s$ and let $\nu_{r_j}(\alpha) = r_j + b_j$. Then

$$r_1 + b_1 < r_2 + b_2 < \cdots < r_s + b_s, \quad b_1 > b_2 > \cdots > b_s$$

and $\alpha = \sum_{j=1}^{s} \pi^{b_j}\lambda_j$ where λ_j are fully orthogonal, primitive vectors. Each λ_j has only one critical index and

$$\nu(\lambda_j) = \nu_{r_j}(\lambda_j) = r_j\,, \qquad 1 \leqq j \leqq s\,.$$

Proof. Let $\xi_1, \cdots, \xi_n$ be a canonical basis of V, so that $\alpha = \sum_{i=1}^{n} a_i\xi_i$. We will use induction on the number of critical indices. Let $\lambda_1 = \pi^{-b_1}\sum' a_i\xi_i \in V$, the sum being over those ξ_i for which either $\nu(\xi_i) \leqq r_1$, or $\nu(\xi_i^2) = \nu(\xi_i)$ and $\nu(a_i) \geqq b_1$, or $\nu(\xi_i^2) > \nu(\xi_i) = \nu(\xi_i \cdot \xi_{i+1})$ and $\nu(a_i) \geqq b_1, \nu(a_{i+1}) \geqq b_1$. λ_1 now satisfies the conditions stated and the rest of α is a vector with critical indices at $r_2, \cdots, r_s$.

COROLLARY. *There exist* $\beta_i \in V, 1 \leqq i \leqq s$, *such that* $\beta_i \cdot \lambda_j = 0$ $(i \neq j)$ *and* $\nu(\lambda_i) = \nu(\beta_i) = \nu(\beta_i \cdot \lambda_i)$.

The λ_j are called *critical components* of α. In general they are not uniquely determined. However, the r_j and b_j are invariant and the lengths of the λ_j are invariant modulo certain powers of π.

PROPOSITION 6. Suppose $\alpha = \sum_{j=1}^{s} \pi^{b_j}\lambda_j = \sum_{j=1}^{s} \pi^{b_j}\mu_j$ are two decompositions as in Proposion 5 with $\nu(\lambda_j) = \nu(\mu_j) = r_j$. Then

$$\left(\pi^{-b_t}\sum_{j=1}^{t} \pi^{b_j}\mu_j\right)^2 \equiv \left(\pi^{-b_t}\sum_{j=1}^{t} \pi^{b_j}\lambda_j\right)^2 (\mathrm{mod}\ \pi^{r_{t+1}+b_{t+1}-b_t})\,,$$
$$1 \leqq t \leqq s-1\,.$$

Proof. Fix $t, 1 \leqq t \leqq s-1$. By Corollary to Proposition 5 there exist β_i such that $\nu(\beta_i \cdot \lambda_i) = \nu(\beta_i) = r_i$, $\beta_i \cdot \lambda_j = 0$ $(i \neq j)$. Then

$$\alpha \cdot \beta_i = \pi^{b_i}\lambda_i \cdot \beta_i \equiv \sum_{j=1}^{t} \pi^{b_j}\mu_j \cdot \beta_i (\mathrm{mod}\ \pi^{r_{t+1}+b_{t+1}})\,, \qquad 1 \leqq i \leqq t\,.$$

We can therefore find $x_i \in R$ such that $\nu(x_i - 1) \geqq r_{t+1} + b_{t+1} - r_i - b_i$ and $x_i\lambda_i \cdot \beta_i = \pi^{-b_i}\sum_{j=1}^{t} \pi^{b_j}\mu_j \cdot \beta_i$. This shows that the vector $\eta = \sum_{j=1}^{t} \pi^{b_j}(\mu_j - x_j\lambda_j)$ is orthogonal to all the vectors $\beta_1, \cdots, \beta_t$. Furthermore

$$(1) \qquad \sum_{j=1}^{t} x_j \pi^{b_j} \lambda_j + \eta = \sum_{j=1}^{t} \pi^{b_j} \mu_j \ .$$

We now have

$$\alpha = \sum_{j=1}^{t} \pi^{b_j} \lambda_j + \sum_{j=t+1}^{s} \pi^{b_j} \lambda_j = \eta + \sum_{j=1}^{t} x_j \pi^{b_j} \lambda_j + \sum_{j=t+1}^{s} \pi^{b_j} \mu_j \ .$$

Since $\nu((x_j - 1)\pi^{b_j}\lambda_j) \geqq r_{t+1} + b_{t+1}$ for $1 \leqq j \leqq t$, we have

$$\pi^{-b_t}\eta \in V(r_{t+1} + b_{t+1} - b_t) \ .$$

The result is now obtained by squaring (1).

Notice that it is not necessary to assume that the μ_j are orthogonal. Let $D_t(\alpha) = (\pi^{-b_t} \sum_{j=1}^{t} \pi^{b_j}\lambda_j)^2$ and $d_t(\alpha) = r_{t+1} + b_{t+1} - b_t$ for $1 \leqq t \leqq s-1$. We have established that $D_t(\alpha) (\mathrm{mod}\, \pi^{d_t(\alpha)})$, $1 \leqq t \leqq s-1$, are invariants of the vector α and hence also isometric invariants.

COROLLARY. *If V does not contain any vectors ξ such that $\nu(\xi^2) = \nu(\xi) = d_t(\alpha)$ then $D_t(\alpha)$ modulo $\pi^{d_t(\alpha)+1}$ is actually invariant.*

Under certain conditions an even stronger condition can be established.

Let $\alpha = \sum_{j=1}^{s} \pi^{b_j}\lambda_j$ be as in Proposition 5. Define $N_{r_j}(\alpha)$ to be 0 if there exists a $\beta \in V(r_j)$ such that $\beta \cdot \lambda_j = 0$ and $\nu(\beta^2) = r_j$, otherwise let $N_{r_j}(\alpha)$ be 1. $N_{r_j}(\alpha)$, $1 \leqq j \leqq s$, are independent of the choice of the critical components (left to the reader) and hence are isometric invariants.

We will show that in certain cases the above isometric invariants form a complete set.

We conclude this section with a simple observation. Let $W_i \subseteqq V = (\xi_1, \cdots, \xi_n)$ and $\dim W_i = \dim V$, $i = 1, 2$. Then for some $k \geqq 0$ we have $\pi^k V = (\pi^k\xi_1, \cdots, \pi^k\xi_n) \subseteqq W_1$. Let $\varphi: W_1 \to W_2$ be an isometry. Let $\alpha \in W_1$ be primitive in W_1 but imprimitive in V, i.e. $\alpha = \pi\beta$ where $\beta \notin W_1$, $\beta \in V$. Then if φ is to extend to an isometry of V, $\varphi(\alpha)$ must be imprimitive in V, $\varphi(\alpha) = \pi\beta'$ say, and then we must have $\varphi(\beta) = \beta'$. Thus by considering $\pi^k\xi_i \in W_1$ we can determine whether φ extends to an isometry of V or not.

2. We now give a complete set of invariants for the improper case. We assume that V contains no vectors ξ such that $\nu(\xi^2) = \nu(\xi)$. Call such a lattice V *fully improper*.

PROPOSITION 7. Let α and β be vectors with one critical index, at r. Then there exists an isometry φ of V such that $\varphi(\alpha) = \beta$ if and only if $\alpha^2 = \beta^2$ and $\nu(\alpha) = \nu(\beta)$.

Proof. We may assume α and β are primitive so that $\nu(\alpha) = \nu(\beta) = r$. If $\nu(\alpha) = \nu(\alpha \cdot \beta)$ then by Lemma 2 $V = (\alpha, \beta) \oplus (\alpha, \beta)'$. Define an isometry φ by $\varphi(\alpha) = \beta$, $\varphi(\beta) = \alpha$ and identity on $(\alpha, \beta)'$. We may therefore assume $\nu(\alpha \cdot \beta) > \nu(\alpha)$. Now choose $\rho \in V(r)$ such that $\nu(\alpha \cdot \rho) = \nu(\beta \cdot \rho) = \nu(\alpha) = \nu(\rho)$. If $\nu(\rho^2) > \nu(\rho) + 1$, then (α, ρ) and (β, ρ) are totally isotropic, and we are finished by Proposition 3. If $\nu(\rho^2) = \nu(\rho) + 1$, put $\gamma = x\rho + \alpha$ where $x\rho^2 + 2\rho \cdot \alpha = 0$, so that $\alpha^2 = \beta^2 = \gamma^2$. Now as above, we can find isometries φ_1, φ_2 such that $\varphi_1(\alpha) = \gamma$, $\varphi_2(\gamma) = \beta$. Put $\varphi = \varphi_2\varphi_1$.

For vectors with more than one critical index we use the invariants of Proposition 6 to adjust the lengths of the critical components and then split them off in orthogonal sublattices. The proof is then complete by induction. We need first another result.

PROPOSITION 8. Let $V = (\lambda_1, \eta) \oplus (\lambda_2, \xi)$ where $\nu(\lambda_1) = \nu(\lambda_1 \cdot \eta) = \nu(\eta) = r_1$ and $\nu(\lambda_2) = \nu(\xi) = \nu(\lambda_2 \cdot \xi) = r_2$. If $\alpha = \pi^c\lambda_1 + \lambda_2$ where $c \geqq 1$, and if $\nu(y) \geqq r_2 - r_1 - c > 0$, then there exists an isometry φ of V such that

$$\varphi(\alpha) = \beta = \pi^c(\lambda_1 + y\eta) + z\xi - \lambda_2$$

for some $z \in R$.

Proof. Let $z \in R$ be a root of the equation

$$\frac{1}{2}z^2\xi^2 - z\lambda_2 \cdot \xi + \pi^{2c}y\left(\lambda_1 \cdot \eta + \frac{y}{2}\eta^2\right) = 0 .$$

Then $\nu(z) \geqq c$ and $\alpha^2 = \beta^2$. Put

$$\gamma_1 = \pi^c\left(\lambda_1 + \frac{1}{2}y\eta\right) + \frac{1}{2}z\xi ,$$

$$\gamma_2 = \frac{1}{2}\pi^c y\eta + \frac{1}{2}z\xi - \lambda_2 ,$$

$$\gamma_3 = \eta - \left(\frac{\eta^2}{\eta \cdot \lambda_1}\lambda_1\right)\lambda_1 ,$$

$$\gamma_4 = \xi - \left(\frac{\xi^2}{\xi \cdot \lambda_2}\lambda_2\right)\lambda_2 .$$

It follows that $\gamma_1 - \gamma_2 = \alpha$, $\gamma_1 + \gamma_2 = \beta$ and $\gamma_1 \cdot \gamma_2 = \gamma_1 \cdot \gamma_4 = \gamma_2 \cdot \gamma_3 =$

$\gamma_3 \cdot \gamma_4 = 0$. Define an isometry φ on the sublattice $(\gamma_1, \gamma_2, \gamma_3, \gamma_4)$ of V by

$$\varphi(\gamma_1) = \gamma_1 , \qquad \varphi(\gamma_2) = -\gamma_2 ,$$
$$\varphi(\gamma_3) = \gamma_3 , \qquad \varphi(\gamma_4) = -\gamma_4 .$$

By considering the vectors in this sublattice that are imprimitive in V we see that φ extends uniquely to an isometry of V. Since $\varphi(\alpha) = \beta$ we are finished.

PROPOSITION 9. Let V be a fully improper lattice. Take $\alpha, \beta \in V$. Then there exists an isometry φ of V such that $\varphi(\alpha) = \beta$ if and only if

1. $\alpha^2 = \beta^2$
2. α and β have the same critical indices $r_1, \cdots, r_s$
3. $\nu_{r_i}(\alpha) = \nu_{r_i}(\beta) = r_i + b_i$ $\qquad 1 \leqq i \leqq s$
4. $D_t(\alpha) \equiv D_t(\beta) (\mathrm{mod}\, \pi^{d_t(\alpha)+1})$ $\qquad 1 \leqq t \leqq s-1$.

Proof. The necessity of the conditions follows from § 1. Notice that the invariant $N_{r_i}(\alpha)$ is always 1 in this case. For the sufficiency we use induction on s. The case $s = 1$ is Proposition 7. Take $s > 1$. Let $\alpha = \sum_{i=1}^{s} \pi^{b_i} \lambda_i$ and $\beta = \sum_{i=1}^{s} \pi^{b_i} \mu_i$ be decompositions into fully orthogonal critical components. Let η be such that $\nu(\eta) = \nu(\eta \cdot \lambda_1) = \nu(\lambda_1)$ and $\eta \cdot \lambda_i = 0$ $(i \neq 1)$. Then from 4, with $t = 1$, we have

$$\lambda_1^2 \equiv \mu_1^2 (\mathrm{mod}\, \pi^{r_2 + b_2 - b_1 + 1}) .$$

Take $y \in R$ such that $(\lambda_1 + y\eta)^2 = \mu_1^2$; this equation has a solution y such that $\nu(y) \geqq r_2 + b_2 - r_1 - b_1$. By Proposition 8, taking $c = b_1 - b_2 \geqq 1$, there exists an isometry of V mapping $\pi^{b_1}\lambda_1 + \pi^{b_2}\lambda_2$ into $\pi^{b_1}(\lambda_1 + y\eta) + \pi^{b_2}(z\xi - \lambda_2)$, i.e. mapping α into a vector α_1 whose first critical component has the same length as the first critical component of β. We now split off the first critical component of α_1 into a binary sublattice. We then proceed to adjust the lengths of the remaining critical components of α.

In this way we can map $\alpha = \Sigma \pi^{b_i} \lambda_i$ into a vector $\alpha^* = \Sigma \pi^{b_i} \lambda_i^*$ with $\lambda_i^{*2} = \mu_i^2$, $1 \leqq i \leqq s$. Those λ_i^*, μ_i satisfying $\nu(\mu_i^2) \geqq \nu(\mu_i) + 2$ can be imbedded in totally isotropic, binary lattices and mapped into each other using Propositions 2 and 3. We may therefore assume that $\nu(\mu_i^2) = \nu(\mu_i) + 1$ for all i.

We now restrict ourselves to studying $\alpha = \Sigma \pi^{b_i} \lambda_i$ and $\beta = \Sigma \pi^{b_i} \mu_i$ with $\lambda_i^2 = \mu_i^2$ and $\nu(\lambda_i^2) = \nu(\lambda_i) + 1$. By Proposition 7 we may assume $\mu_1 = \lambda_1$ after applying an isometry to α. Take $\rho \in V$ such that $\nu(\mu_1 \cdot \rho) = \nu(\mu_1) = \nu(\rho)$ and $\rho \cdot \mu_i = 0$, $2 \leqq i \leqq s$. Using Lemma 2 we can split off (μ_1, ρ). However, we will not have $\rho \cdot \lambda_i = 0$ in general.

We apply a further isometry θ to α keeping λ_1 fixed and mapping λ_i, $2 \leqq i \leqq s$, into vectors orthogonal to ρ. Then splitting off (λ_1, ρ) we are finished by induction.

We construct θ as follows. Let $V = (\lambda_1, \eta_1) \oplus (\lambda_2, \xi_1) \oplus U$ where λ_i, $3 \leqq i \leqq s$, are in the sublattice U. We may assume that $\nu(\eta_1^2) = \nu(\eta_1) + 1$ and $\nu(\xi_1^2) = \nu(\xi_1) + 1$, for otherwise we have totally isotropic, binary lattices and these may be concelled. Thus we may assume $\xi_1^2 = \lambda_2^2$. Let

$$\lambda_2' = x\pi^{r_2 - r_1}\left(\lambda_1 - \left(\frac{\lambda_1^2}{\lambda_1 \cdot \eta_1}\eta_1\right)\right) + y\lambda_2 + \xi_1 ,$$

where x and y are chosen so that $\lambda_2'^2 = \lambda_2^2$ and $\lambda_2' \cdot \rho = 0$ (the equations for x and y have solutions in R). Then $\lambda_2' \cdot \lambda_1 = 0$ and $\nu(\lambda_2' \cdot \lambda_2) = \nu(\lambda_2)$. Thus $V = (\lambda_1, \gamma) \oplus (\lambda_2, \lambda_2') \oplus U$ for some $\gamma \in V$. We now take an isometry mapping λ_2 into λ_2' and acting as the identity on $(\lambda_2, \lambda_2')'$. Similarly we can map λ_i, $3 \leqq i \leqq s$, into vectors orthogonal to ρ.

3. We now consider the general case but restrict ourselves to R being the ring of 2-adic integers.

PROPOSITION 10. Let α and β be vectors with one critical index, at r. Then there exists an isometry φ of V such that $\varphi(\alpha) = \beta$ if and only if $\alpha^2 = \beta^2$, $\nu(\alpha) = \nu(\beta)$ and $N_r(\alpha) = N_r(\beta)$.

Proof. We may assume α and β primitive so that $\nu(\alpha) = \nu(\beta) = r$. If $\nu(\alpha^2) = \nu(\alpha)$ we are finished by Proposition 4. Assume $\nu(\alpha^2) > \nu(\alpha)$. As in Proposition 7 we take $\rho \in V(r)$ such that $\nu(\alpha \cdot \rho) = \nu(\beta \cdot \rho) = \nu(\alpha)$. We need only consider the case $\nu(\rho^2) = \nu(\rho)$. Then we may assume that $\alpha = x\rho + \eta$, $\beta = \rho + \xi$ with $\eta \cdot \rho = \xi \cdot \rho = 0$. We now wish to find $\lambda, \mu \in V$ such that $\alpha = \lambda + \mu$, $\lambda \cdot \mu = 0$ and $\lambda^2 = \rho^2$. The necessary equations can be solved, but only because $\nu(x) = 0$ so that for the 2-adics $\nu(x^2 - 1) \geqq 3$. We then construct an isometry φ such that $\varphi(\lambda) = \rho$ and $\varphi(\mu) = \xi$ using Proposition 4.

This result need not be true if we go beyond the 2-adics; further invariants are necessary as is shown by an example latter.

To treat vectors with more than one critical component we must establish some results analogous to Proposition 8. The situation is now more complex and the particular invariants depend on the structure of V. For there to exist an isometry mapping α into β we need the following invariants:

(i) $\alpha^2 = \beta^2$

(ii) α and β have the same critical indices $r_1, \cdots, r_s$

(iii) $N_{r_i}(\alpha) = N_{r_i}(\beta)$, $\quad 1 \leqq i \leqq s$

(iv) $\nu_{r_i}(\alpha) = \nu_{r_i}(\beta) = r_i + b_i$, $\quad 1 \leqq i \leqq s$

(v) $D_t(\alpha) \equiv D_t(\beta) (\mathrm{mod}\, \pi^{d_t(\alpha)+f_t})$, $\quad 1 \leqq t \leqq s-1$,

where $f_t = 0, 1$ or 2 depending on the structure of V.

These conditions, with $f_t = 2$, are always sufficient, but usually not necessary. The conditions with $f_t = 1$ are necessary, provided there do not exist any $\xi_t \in V$ satisfying $\nu(\xi_t^2) = \nu(\xi_t) = d_t(\alpha)$ for $1 \leqq t \leqq s-1$, and also sufficient provided further that $b_{i-1} \neq b_i + 1$ or $r_{i+1} + b_{i+1} \neq r_i + b_i + 1$ for any i. The proofs of these remarks are similar to the proof of Proposition 9 using results analogous to Proposition 8 to adjust the lengths of the critical components.

PROPOSITION 11. Let $V = (\lambda_1) \oplus (\lambda_2)$ where $\nu(\lambda_1) = r_1$, $\nu(\lambda_2) = r_2$ and let $c \geqq 1$ and $r_2 - r_1 - c \geqq 1$. If $\alpha = \pi^c \lambda_1 + \lambda_2$ then there exists an isometry φ of V such that

$$\varphi(\alpha) = \beta = \pi^c x \lambda_1 - y \lambda_2$$

for some $y \in R$, provided either

(a) $\nu(x-1) \geqq r_2 - r_1 - c + 1$,

or

(b) $\nu(x-1) \geqq r_2 - r_1 - c \geqq 2$, $c \geqq 2$.

Proof. Let y be a root of the equation

$$y^2 = 1 + \pi^{2c}(1 - x^2)\frac{\lambda_1^2}{\lambda_2^2}, \qquad y \equiv 1 (\mathrm{mod}\, 4) .$$

This equation has a root $y \in R$ by (a) or (b) using Hensel's lemma. Put

$$\gamma_1 = \frac{1}{2}\pi^c(1 + x)\lambda_1 + \frac{1}{2}(1 - y)\lambda_2$$

and

$$\gamma_2 = \frac{1}{2}\pi^c(1 - x)\lambda_1 + \frac{1}{2}(1 + y)\lambda_2 .$$

Then $\gamma_1 + \gamma_2 = \alpha$, $\gamma_1 - \gamma_2 = \beta$ and $\gamma_1 \cdot \gamma_2 = 0$. For (a) we have $\nu(\gamma_2^2) = \nu(\gamma_2)$, so that $V = (\gamma_2) \oplus (\gamma_2)'$ and we take φ such that $\varphi(\gamma_2) = -\gamma_2$ and φ is the identity on $(\gamma_2)'$. Then $\varphi(\alpha) = \beta$. For (b) we take the isometry on (γ_1, γ_2) defined by $\varphi(\gamma_1) = \gamma_1$, $\varphi(\gamma_2) = -\gamma_2$ and extend this to V through the imprimitive vectors as in Proposition 8.

PROPOSITION 12. Let $V = (\lambda_1) \oplus (\lambda_2, \xi)$ where $\nu(\lambda_1) = r_1$, $\nu(\lambda_2) = \nu(\xi) = \nu(\lambda_2 \cdot \xi) = r_2$, $\nu(\lambda_2^2) > \nu(\lambda_2)$ and $\nu(\xi^2) > \nu(\xi)$. If $\alpha = \pi^c \lambda_1 + \lambda_2$ with

$c \geqq 1$, then there exists an isometry φ of V such that

$$\varphi(\alpha) = \beta = \pi^c x\lambda_1 + y\xi - \lambda_2$$

for some $y \in R$, provided either $\nu(x-1) \geqq r_2 - r_1 - c + 1$ or $\nu(x-1) \geqq r_2 - r_1 - c \geqq 2$.

PROPOSITION 13. Let $V = (\lambda_1, \eta) \oplus (\lambda_2)$ where $\nu(\lambda_1) = \nu(\eta) = \nu(\lambda_1 \cdot \eta) = r_1$, $\nu(\lambda_1^2) > \nu(\lambda_1)$, $\nu(\eta^2) > \nu(\eta)$ and $\nu(\lambda_2) = r_2$. If $\alpha = \pi^c\lambda_1 + \lambda_2$ with $c \geqq 1$ and $r_2 - r_1 - c > 0$, then there exists an isometry φ of V such that

$$\varphi(\alpha) = \beta = \pi^c(\lambda_1 + y\eta) - z\lambda_2$$

for some $z \in R$, provided that either $\nu(y) \geqq r_2 - r_1 - c + 1$ or $\nu(y) \geqq r_2 - r_1 - c$ and $c \geqq 2$.

The proofs of Propositions 12 and 13 are similar to those of Propositions 8 and 11. These results now enable us to establish the results stated for isometric invariants in the 2-adic case above.

Suppose that $\alpha = \sum_{i=1}^{s} \pi^{b_i}\lambda_i$ and $\beta = \sum_{i=1}^{s} \pi^{b_i}\mu_i$ satisfy conditions (i)–(v) above (in some case of f_t). If $\nu(\lambda_1^2) = \nu(\lambda_1)$, then from (v) $\lambda_1^2 \equiv \mu_1^2 (\mathrm{mod}\, \pi^{d_1(\alpha)+f_1})$ so that $\lambda_1^2 = x\mu_1^2$ with $\nu(x-1) \geqq d_1(\alpha) + f_1 - r_1$. Provided that $d_1(\alpha) + f_1 - r_1 \geqq 3$, we may take $\lambda_1^2 = x^2\mu_1^2$ with $\nu(x-1) \geqq d_1(\alpha) + f_1 - r_1 - 1$. Then, applying an isometry as in Proposition 11 or 12, we may assume that $\lambda_1 = \mu_1$, etc. Likewise, if $\nu(\lambda_1^2) > \nu(\lambda_1)$, we use suitable combinations of the methods given above.

The presence of vectors satisfying $\nu(\xi_t^2) = \nu(\xi_t) = d_t(\alpha)$ for certain t enable us to use the weakest form of (v) with $f_t = 0$, but because of the proliferation of cases we proceed no further with this.

To extend the above beyond the 2-adics requires further invariants as is shown by the following example:

$$V = (\xi_1) \oplus (\xi_2) \oplus (\xi_3)$$

$$\xi_1^2 = \xi_2^2 = 1\,, \qquad \xi_3^2 = 2$$

$$\alpha = \xi_1 + \xi_2\,, \qquad \beta = a\xi_1 + a\xi_2 + b\xi_3$$

where the residue class field has four elements (i.e. $ab = 1$, $a^2 = b$, $b^2 = a$ and $a + b = 1$). Although (i)–(v) are satisfied no isometry of V maps α into β. To show this look at the conditions imposed upon the image of ξ_3 modulo 4.

REFERENCES

1. D. G. James, *The extension of isometries of sublattices over local fields*, M.I.T. Ph. D. Thesis, 1963.
2. B. W. Jones, *The arithmetic theory of quadratic forms*, Carus Monographs, No. 10, Mathematical Association of America, 1950.
3. S. Lang, *Algebraic Numbers*, Addison-Wesley, 1964.
4. O. T. O'Meara, *Introduction to quadratic forms*, Springer-Verlag, 1963.
5. ———, *Quadratic forms over local fields*, Amer. J. Math. **77** (1955), 87-116.
6. ———, *Witt's theorem and the isometry of lattices*, Proc. Am. Math. Soc. **7** (1956), 9-22.
7. S. Rosenzweig, *An analogy of Witt's theorem for modules over the ring of p-adic integers*, M.I.T. Ph. D. Thesis, 1958.

UNIVERSITY OF AUCKLAND

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

A REMARK ON THE LEMMA OF GAUSS

FRED KRAKOWSKI

Let R be the ring of integers of some algebraic number field K and $\mathfrak{P} = R[x_0, \cdots, x_r, y_0, \cdots, y_s]$, where the x_i's and y_j's are indeterminates. Call two ideals of $\mathfrak{P}$ equivalent, if after substitution of the indeterminates by arbitrary elements of R they always yield identical ideals in R. For example, consider the ideal I generated by the coefficients of the product of the two polynomials $f(t) = \sum_{i=0}^{r} x_i t^i$ and $g(t) = \sum_{j=0}^{s} y_j t^j$. According to the so-called Lemma of Gauss, I is equivalent to the product J of the ideals $(x_0, \cdots, x_r)$ and $(y_0, \cdots, y_s)$.

The object of this note is to show that the ideal I has the following minimal property: It has the smallest number of generators, namely $r + s + 1$, among all ideals in $\mathfrak{P}$ which are equivalent to J in the above sense.

LEMMA 1. *For every nonconstant polynomial $f \in R[t]$, t an indeterminate, there exist infinitely many prime ideals $P \subset R$, such that the congruence $f(x) \equiv 0 \pmod{P}$ has a solution $x \in R$.*

Proof. Denote by $f_1, \cdots, f_m$ the polynomials conjugate to f over the rationals and let $f = f_1$. Consider their product $F = f_1 \cdots f_m$. The coefficients of F are rational integers and thus there is an infinite sequence of rational primes $p_1, p_2, \cdots$ and corresponding rational integers $x_1, x_2, \cdots$, such that $F(x_i) \equiv 0 \pmod{p_i}$, $i = 1, 2, \cdots$ (see e.g. [1], p. 33).

Let now L be a normal extension of the rationals containing K. For each p_i choose a prime ideal $P_i \subset L$ containing p_i. Then $F(x_i) \equiv 0 \pmod{P_i}$. Since $(p_i, p_j) = (1)$ for $i \neq j$, we also have $P_i \neq P_j$. Thus there exist infinitely many prime ideals of L which divide numbers of the sequence $F(x_i)$, $i = 1, 2, \cdots$.

Assume now there exist only finitely many prime ideals in R, say $Q_1, \cdots, Q_k$, such that the congruence $f(x) \equiv 0 \pmod{Q_j}$ has a solution in R for $j = 1, \cdots, k$. Denote by $Q_1', \cdots, Q_k'$ the ideals in L generated by $Q_1, \cdots, Q_k$. A prime ideal of L containing $F(x_i)$ would then have to be also a divisor of some Q_j' or of an ideal conjugate to Q_j', because $F(x_i)$ is the product of the conjugate elements $f_1(x_i), \cdots, f_m(x_i)$. It would follow that there are only finitely many prime ideals of L containing numbers of the sequence $F(x_1), F(x_2), \cdots$, which is a contradiction. This proves the lemma.

The next lemma gives a necessary condition which is satisfied by

Received June 2, 1964. The author is grateful to Professor Ernst Specker for many helpful conversations.

equivalent ideals of a polynomial ring over R. Denote by R^n the set of n-tuples of elements of R. If $t_1, \cdots, t_n$ are indeterminates and $I = (f_1, \cdots, f_r) \subset R[t_1, \cdots, t_n]$, $a \in R^n$, let $I_a = (f_1(a), \cdots, f_r(a))$. Further let C stand for the field of complex algebraic numbers, C^n for the n-dimensional affine space over C and V_I for the algebraic variety in C^n defined by the ideal I.

LEMMA 2. *Let I and J be ideals of $R[t_1, \cdots, t_n]$ and suppose that for all $a \in R^n$ we have $I_a = J_a$. Then $V_I = V_J$.*

Proof. Let $f_1, \cdots, f_r$ be a basis of I and $g_1, \cdots, g_s$ a basis of J. Suppose $V_I \neq V_J$ and assume there is a point $\alpha = \langle \alpha_1, \cdots, \alpha_n \rangle$ of V_I not contained in V_J. We must show that there exists a n-tuple $a \in R^n$, such that $I_a \neq J_a$.

Now $f_i(\alpha) = 0$, $i = 1, \cdots, r$ but, say, $g_1(\alpha) \neq 0$. $K(\alpha_1, \cdots, \alpha_n)$ is a separable algebraic extension of K, and let θ be a primitive element. We then have $\alpha_i = h_i(\theta)$, $i = 1, \cdots, n$, where h_i is a polynomial whose coefficients may be assumed, without loss of generality, to be integers of R. Also let $p(t)$ be a polynomial in $R[t]$, of which θ is a root and which is irreducible in $K[t]$.

Put $F_i(t) = f_i(h_1(t), \cdots, h_n(t))$, $i = 1, \cdots, r$ and $G_1(t) = g_1(h_1(t), \cdots, h_n(t))$. Since $f_i(\alpha) = 0$, $i = 1, \cdots, r$ and $g_1(\alpha) \neq 0$, we have $F_i(\theta) = 0$, $i = 1, \cdots, r$ and $G_1(\theta) \neq 0$. Hence there are polynomials $q_i(t) \in R[t]$ and elements $s_i \in R$, $i = 1, \cdots, r$, with $s_i F_i(t) = p(t) q_i(t)$, $i = 1, \cdots, r$. On the other hand, since $p(t)$ is irreducible and $G_1(\theta) \neq 0$, $p(t)$ and $G_1(t)$ are relatively prime in $K[t]$, and there are polynomials $A(t)$, $B(t) \in R[t]$, such that

$$A(t)p(t) + B(t)G_1(t) = c \,,$$

where $c \in R$ and $c \neq 0$.

By Lemma 1 there are infinitely many prime ideals P in R, such that the congruence $p(x) \equiv 0 \pmod{P}$ has a solution in R. Each one of the numbers $s_1, \cdots, s_r$ and c is contained in only a finite number of prime ideals. Hence there is a prime ideal $P \subset R$ and an element $x \in R$, such that $p(x) \equiv 0 \pmod{P}$, but $s_i \not\equiv 0 \pmod{P}$, $i = 1, \cdots, r$ and $c \not\equiv 0 \pmod{P}$. Therefore $B(x)\, G_1(x) \not\equiv 0 \pmod{P}$. If we now let $a = \langle h_1(x), \cdots, h_n(x) \rangle$, then $a \in R^n$ and we get $g_1(a) = G_1(x) \not\equiv 0 \pmod{P}$ and thus also $J_a \not\equiv 0 \pmod{P}$. On the other hand, since $s_i \notin P$, it follows that $F_i(x) \equiv 0 \pmod{P}$, hence $f_i(a) \equiv 0 \pmod{P}$, $i = 1, \cdots, r$ and thus $I_a \equiv 0 \pmod{P}$. Therefore $I_a \neq J_a$, which was to be shown.

COROLLARY. *If for all $a \in R^n$ we have $I_a = (1)$, then $V_I = \phi$.*

LEMMA 3. *Consider polynomials* $f_1, \cdots, f_k \in R[t]$. *Assume that for all nonzero elements* $r \in R$ *the* k *numbers* $f_1(r), \cdots, f_k(r)$ *generate the same ideal* $I \subset R$. *Then we also have* $I = (f_1(0), \cdots, f_k(0))$.

Proof. If D is an ideal in R, $D \supset (f_1(0), \cdots, f_k(0))$ and r is an arbitrary nonzero element of D, then $f_i(r) \in D$ for $i = 1, \cdots, k$. Since $I = (f_1(r), \cdots, f_k(r))$, we have $I \subset D$.

Conversely, if $D \supset I$ and $r \in D$, $r \neq 0$, then $f_i(r) \equiv f_i(0) \pmod{D}$. Since $f_i(r) \in D$, also $f_i(0) \in D$ for all i and hence $(f_1(0), \cdots, f_k(0)) \subset D$. This proves the lemma.

LEMMA 4. *Let* $f_1, \cdots, f_k$ *be arbitrary and* $g_1, \cdots, g_m$ *homogeneous linear polynomials in* $R[t_1, \cdots, t_n]$. *Assume that for all* $a \in R^n$ *we have*

$$(f_1(a), \cdots, f_k(a)) = (g_1(a), \cdots, g_m(a)) .$$

Also denote by $h_1, \cdots, h_k$ *the subpolynomials of* $f_1, \cdots, f_k$ *formed by their linear terms. Then* $(h_1(a), \cdots, h_k(a)) = (g_1(a), \cdots, g_m(a))$ *for all* $a \in R^n$.

Proof. Since $(g_1(0), \cdots, g_m(0)) = (f_1(0), \cdots, f_k(0)) = (0)$, we have $f_1(0) = \cdots = f_k(0) = 0$. Thus $f_i = h_i +$ terms of degree $\geqq 2$, $i = 1, \cdots, k$. Take a fixed n-tuple $a \in R^n$ and let $r \in R$ be arbitrary but $\neq 0$. Then

$$\begin{aligned}(f_1(ra), \cdots, f_k(ra)) &= (rh_1(a) + r^2(\cdots), \cdots, rh_k(a) + r^2(\cdots)) \\ &= (r)(h_1(a) + r(\cdots), \cdots, h_k(a) + r(\cdots)) \\ &= (g_1(ra), \cdots, g_m(ra)) = (r)(g_1(a), \cdots, g_m(a)) .\end{aligned}$$

R being an integral domain, we get

$$(h_1(a) + r(\cdots), \cdots, h_k(a) + r(\cdots)) = (g_1(a), \cdots, g_m(a))$$

for all nonzero $r \in R$. By Lemma 2 therefore

$$(h_1(a), \cdots, h_k(a)) = (g_1(a), \cdots, g_m(a)) ,$$

which was to be proved.

THEOREM. *Consider in* $R[x_0, \cdots, x_r, y_0, \cdots, y_s]$ *the ideal* $J = (x_0, \cdots, x_r)(y_0, \cdots, y_s)$ *and suppose* I *is an ideal such that for all* $a \in R^{r+s+2}$ *we have* $I_a = J_a$. *Then the number of elements in a basis of* I *is at least* $r + s + 1$.

Proof. Let $f_1, \cdots, f_n$ be a basis of I and let I' be the ideal generated by the subpolynomials $b_1, \cdots, b_n$ of $f_1, \cdots, f_n$, which are linear with respect to $x_0, \cdots, x_r$ and with respect to $y_0, \cdots, y_s$. Since also the generators of J are bilinear and for all $a \in R^{r+s+2}$ we have $I_a = J_a$, by Lemma 3, we also have $I'_a = J_a$ for all a.

Now the ideal J has only trivial zeroes in C^{r+s+2}, either all $x_i = 0$ or all $y_j = 0$. On the other hand, if $n \leqq r + s$, it follows from a theorem of Macaulay (see [2], p. 54) that I' has a nontrivial zero in C^{r+s+2}. By Lemma 2 this cannot happen. Hence $n \geqq r + s + 1$.

References

1. A. Scholz, *Einführung in die Zahlentheorie*, Sammlung Göschen Band 1131, Berlin 1939.

2. F. S. Macaulay, *Algebraic Theory of Modular Systems*, Cambridge Tracts in Math. 19, Cambridge 1916.

University of California, Davis

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

A SUBDETERMINANT INEQUALITY

MARVIN MARCUS[1] AND HENRYK MINC[2]

Let A be an n-square positive semi-definite hermitian matrix and let $D_m(A)$ denote the maximum of all order m principal subdeterminants of A. In this note we prove the inequality

$$(D_m(A))^{1/m} \geqq (D_{m+1}(A))^{1/(m+1)}, \qquad m = 1, \cdots, n-1,$$

and discuss in detail the case of equality. This result is closely related to Newton's and Szász's inequalities.

Let $A = (a_{ij})$ be an n-square positive semi-definite hermitian matrix with eigenvalues $\lambda_1 \geqq \cdots \geqq \lambda_n \geqq 0$. We introduce some notation. For $1 \leqq m \leqq n$ let $Q_{m,n}$ denote the set of all $\binom{n}{m}$ sequences $\omega = (\omega_1, \cdots, \omega_m)$, $1 \leqq \omega_1 < \omega_2 < \cdots < \omega_m \leqq n$. Let $A[\omega \mid \omega]$ denote the m-square submatrix of A whose (i, j) entry is $a_{\omega_i \omega_j}$, $i, j = 1, \cdots, m$.

THEOREM. *If A is a positive semi-definite hermitian matrix then*

$$\max_{\alpha \in Q_{m\ n}} (\det (A[\alpha \mid \alpha]))^{1/m} \geqq \max_{\omega \in Q_{m+1,n}} (\det (A[\omega \mid \omega]))^{1/m+1}, \qquad m = 1, \cdots, n-1. \tag{1}$$

Equality holds for a given m if and only if either A has rank less than m or $A[\omega^0 \mid \omega^0]$ is a multiple of the identity, where the sequence $\omega^0 \in Q_{m+1,n}$ is one that satisfies

$$\det (A[\omega^0 \mid \omega^0]) = \max_{\omega \in Q_{m+1\ n}} \det A[\omega \mid \omega]. \tag{2}$$

There are two classical results that are closely related to the inequalities (1). These are Szász's inequalities and the Newton inequalities. Szász proved that [1, p. 119]

$$\left(\prod_{\alpha \in Q_{m\ n}} (\det (A[\alpha \mid \alpha]))^{1/\binom{n}{m}} \right)^{1/m} \geqq \left(\prod_{\omega \in Q_{m+1,n}} (\det (A[\omega \mid \omega]))^{1/\binom{n}{m+1}} \right)^{1/(m+1)}. \tag{3}$$

Newton's inequalities [1, p. 106] state that if $E_m(A)$ is the mth elementary symmetric function of the nonnegative numbers $\lambda_1, \cdots, \lambda_n$ then

Received August 27, 1964.

[1] The research of this author was supported by N. S. F. Grant G. P. 1085.

[2] The research of this author was supported by U. S. Air Force Grant No. AF-AFOSR-432-63.

$$(4) \qquad \left(E_m(A)\Big/\binom{n}{m}\right)^{1/m} \geqq \left(E_{m+1}(A)\Big/\binom{n}{m+1}\right)^{1/(m+1)} .$$

However,

$$(5) \qquad E_m(A) = \sum_{\alpha \in Q_{m,n}} \det (A[\alpha \mid \alpha])$$

and hence (4) can be written

$$(6) \qquad \left(\sum_{\alpha \in Q_{m,n}} \det (A[\alpha \mid \alpha])\Big/\binom{n}{m}\right)^{1/m} \geqq \left(\sum_{\omega \in Q_{m+1,n}} \det (A[\omega \mid \omega])\Big/\binom{n}{m+1}\right)^{1/(m+1)} .$$

Notice that (3) compares the geometric mean of the principal subdeterminants of order m with the geometric mean of the principal subdeterminants of order $m+1$. Also (6) makes the same kind of comparison for the arithmetic means of these quantities. The result (1) compares the maxima of the two sets of subdeterminants.

To prove the theorem we state and prove a preliminary lemma.

LEMMA. *If A is a positive semi-definite n-square hermitian matrix then*

$$(7) \qquad \max_{\alpha \in Q_{m,n}} \det (A[\alpha \mid \alpha]) \geqq (\det (A))^{m/n} , \qquad 1 \leqq m \leqq n .$$

Equality holds if and only if either the rank of A is less than m or A is a multiple of the identity matrix.

Proof. We use some properties of the compound matrix of A, denoted by $C_m(A)$. The essential facts concerning $C_m(A)$ are [1, pp. 17, 24, 70]:

(i) $\det (C_m(A)) = (\det (A))^{\binom{n-1}{m-1}}$ (Sylvester-Franke theorem);

(ii) if A is positive semi-definite hermitian, so is $C_m(A)$;

(iii) the characteristic roots of $C_m(A)$ are the $\binom{n}{m}$ products

$$\prod_{i=1}^{m} \lambda_{\omega_i}, \quad \omega \in Q_{m,n} .$$

We want to prove that

$$(8) \qquad \max_{\alpha \in Q_{m,n}} \det (A[\alpha \mid \alpha]) \geqq (\det (A))^{m/n} .$$

If we apply the Hadamard determinant theorem [1, p. 114] to $C_m(A)$ then we conclude from (i)

$$(9) \qquad \prod_{\alpha \in Q_{m,n}} \det (A[\alpha \mid \alpha]) \geqq \det (C_m(A)) = (\det (A))^{\binom{n-1}{m-1}} .$$

If for every $\alpha \in Q_{m,n}$, $\det(A[\alpha \mid \alpha])$ were strictly less than $(\det(A))^{m/n}$ then from (9) we would conclude that

$$(10) \qquad (\det(A))^{\binom{n-1}{m-1}} < ((\det(A))^{m/n})^{\binom{n}{m}} = (\det(A))^{\binom{n-1}{m-1}},$$

a contradiction. Thus (8) holds. If (8) were equality suppose first that not all $\det(A[\alpha \mid \alpha])$, $\alpha \in Q_{m,n}$ are equal. Then from (9) we would obtain the same contradiction (10). Thus for equality to hold in (8)

$$\det(A[\alpha \mid \alpha]) = (\det(A))^{m/n}$$

for all $\alpha \in Q_{m,n}$. This means that all the main diagonal elements of $C_m(A)$ are equal. If this common value is 0 then A has rank at most $m-1$. If the common value is nonzero then (9) is equality throughout and as we know from the case of equality in the Hadamard determinant theorem $C_m(A)$ is a multiple of the identity. Thus from (iii) we know that the characteristic roots

$$\prod_{i=1}^{m} \lambda_{\alpha_i}, \quad \alpha \in Q_{m,n}, \quad m < n,$$

are equal. But then it follows that $\lambda_1 = \cdots = \lambda_n$ and hence A is a multiple of the identity, completing the proof of the lemma.

To prove the inequality (1) we apply the lemma to submatrices. Let ω^0 be a sequence in $Q_{m+1,n}$ for which

$$(11) \qquad \det(A[\omega^0 \mid \omega^0]) = \max_{\omega \in Q_{m+1,n}} \det(A[\omega \mid \omega]).$$

For $\alpha \in Q_{m,n}$ and α a subsequence of ω^0, i.e., $\alpha \subset \omega^0$, we know that $A[\alpha \mid \alpha]$ is an m-square submatrix of $A[\omega^0 \mid \omega^0]$. Hence, by the lemma,

$$(12) \qquad \max_{\alpha \in Q_{m,n}, \alpha \subset \omega^0} \det(A[\alpha \mid \alpha]) \geqq (\det(A[\omega^0 \mid \omega^0]))^{m/(m+1)}.$$

Thus a fortiori

$$(13) \qquad \max_{\alpha \in Q_{m,n}} \det(A[\alpha \mid \alpha]) \geqq (\det(A[\omega^0 \mid \omega^0]))^{m/(m+1)}.$$

Applying (11) we obtain the inequality (1) from (13).

If equality holds in (1) then (12) must be equality as well. Therefore either the rank of $A[\omega^0 \mid \omega^0]$ is less than m or $A[\omega^0 \mid \omega^0]$ is a multiple of the $(m+1)$-square identity matrix. If the former is the case then $\det(A[\omega^0 \mid \omega^0]) = 0$ and hence, since (13) is equality, every mth order principal subdeterminant of A is 0. Thus the rank of A is less than m.

REFERENCE

1. Marvin Marcus and Henryk Minc, *A Survey of Matrix Theory and Matrix Inequalities*, Allyn and Bacon, Boston, 1964.

UNIVERSITY OF CALIFORNIA, SANTA BARBARA

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

NORMS AND NONCOMMUTATIVE JORDAN ALGEBRAS

K. McCRIMMON

Roughly speaking, a norm on a nonassociative algebra is a nondegenerate form Q satisfying $Q(M_x y) = m(x)Q(y)$ for all x, y in the algebra where M_x is a linear transformation having something to do with multiplication by x and where m is a rational function; taking $M_x = L_x$ or $M_x = U_x = 2L_x^2 - L_{x^2}$ we get the forms Q satisfying $Q(xy) = Q(x)Q(y)$ or $Q(U_x y) = Q(x)^2 Q(y)$ investigated by R. D. Schafer. This paper extends the known results by proving that any normed algebra $\mathfrak{A}$ is a separable noncommutative Jordan algebra whose symmetrized algebra $\mathfrak{A}^+$ is a separable Jordan algebra, and that the norm is a product of irreducible factors of the generic norm. As a consequence we get simple proofs of Schafer's results on forms admitting associative composition and can extend his results on forms admitting Jordan composition to forms of arbitrary degree q rather than just $q = 2$ or 3. We also obtain some results of M. Koecher on algebras associated with ω-domains. In the process, simple proofs are obtained of N. Jacobson's theory of inverses and some of his results on generic norms. The basic tool is the differential calculus for rational mappings of one vector space into another. This affords a concise way of linearizing identities, and through the chain rule and its corollaries furnishes methods not easily expressed "algebraically".

Algebras having some sort of "norm" have appeared in various investigations. R. D. Schafer proved in [12] that any algebra $\mathfrak{A}$ with a nondegenerate form Q admitting associative composition $Q(xy) = Q(x)Q(y)$ is a separable alternative algebra. In [11] he proved that if $\mathfrak{A}$ is commutative and has a form Q of degree 2 or 3 admitting Jordan composition $Q(U_x y) = Q(x)^2 Q(y)$, where $U_x = 2L_x^2 - L_{x^2}$, then it is a separable Jordan algebra. In the applications of Jordan algebras to several complex variables [9] M. Koecher considered domains in a real vector space on which a positive homogeneous real-analytic function ω was defined satisfying $\omega(H_x y) = \det H_x \cdot \omega(y)$, where H_x was essentially the Hessian of $\log \omega$ at x. He associated with such an ω-domain a real semisimple Jordan algebra $\mathfrak{A}$ in which $H_x = U_x^{-1}$. In all these cases the algebra was a separable noncommutative Jordan algebra and the norm Q (or ω) was essentially a product of the irreducible factors of the generic norm of $\mathfrak{A}$.

Received April 7, 1964. This reserch was supported by a fellowship from the National Aeronautics and Space Administration and consists of part of the author's doctoral dissertation written at Yale University under Professor Nathan Jacobson.

The present paper originated in an attempt to prove algebraically that the algebra of an ω-domain is a Jordan algebra. Professor Nathan Jacobson suggested that the resulting proof (Lemma 1.3) could be extended to an arbitrary field and might yield at the same time a uniform derivation of Schafer's results, and this led to a general investigation of normed algebras. Speaking very roughly, a norm on a nonassociative algebra $\mathfrak{A}$ is a form Q satisfying $Q(M_x y) = m(x)Q(y)$ where M_x is a linear transformation on $\mathfrak{A}$ having something to do with multiplication by x, and where m is a rational function.

In this paper we will make a systematic study of normed algebras. The basic tool is the standard differential calculus for rational mappings of one vector space into another. This affords a concise way of linearizing identities, and through the chain rule and its corollaries furnishes methods not easily expressed "algebraically".

The paper is divided into three parts, the first of which is devoted to proving that all normed algebras are separable noncommutative Jordan algebras. After recalling the fundamental results of the differential calculus we make precise the definition of a form Q admitting composition on an algebra. With such a Q we associate an associative symmetric bilinear form called the trace of Q. We define Q to be nondegenerate if its trace form is nondegenerate, and show that this agrees with Schafer's definition of nondegeneracy in the cases of interest. A norm on an algebra is then a nondegenerate form admitting composition. The results of Schafer and Koecher follow easily from the main theorem that every normed algebra is a separable noncommutative Jordan algebra. Our definition of nondegeneracy has the advantage that bilinear forms are easier to work than q-linear forms, hence we can extend Schafer's results on forms admitting Jordan composition to forms of arbitrary degree q rather than just $q = 2, 3$. It also allows us to obtain his results when the base field has more than q elements, which is a weaker hypothesis than his condition that the characteristic is 0 or is greater than q.

The second part of the paper is devoted to characterizing the norm of a normed algebra. We first extend N. Jacobson's theory of inverses in Jordan algebras to noncommutative Jordan algebras; the proof of the main properties of these is simpler than his. We next prove a lemma to the effect that under fairly general conditions if Q admits some kind of composition then so do all its irreducible factors. Using this and a technique of N. Jacobson's we can easily derive the basic properties of the generic norm. Applying these results to normed algebras we show that the norm of any normed algebra is a product of irreducible factors of the generic norm.

The last part of the paper is devoted to the work of Koecher [9] and N. Jacobson [6] on isotopes and the group of norm-preserving

transformations of a commutative Jordan algebra.

CHAPTER I

1. Some conventions. In this paper we will always work over a field Φ of characteristic $p \neq 2$ (p may be zero); $|\Phi|$ will denote the cardinality of Φ. *In Chapter I we will always assume that Φ is infinite* (except in § 5), *and that $\mathfrak{X}, \mathfrak{Y}, \mathfrak{Z}$ are finite-dimensional vector spaces over Φ.* "Algebra" will always mean nonassociative algebra with identity.

We briefly recall the well-known facts about the differential calculus for rational mappings over an infinite field [2, pp. 21–37]. Relative to bases $\{x_1, \cdots, x_n\}$ and $\{y_1, \cdots, y_m\}$ for the vector spaces $\mathfrak{X}$ and $\mathfrak{Y}$ a *rational mapping* $F: \mathfrak{X} \to \mathfrak{Y}$ has the form

$$x = \Sigma \xi_i x_i \xrightarrow{F} y = \Sigma \eta_j y_j$$

where $\eta_j = F_j(\xi_1, \cdots, \xi_n)$ are rational expressions in $\Phi(\xi_1, \cdots, \xi_n)$. The value $F(x)$ is defined only for those x in the Zariski-open subset of $\mathfrak{X}$ where the denominators of the components F_j of F don't vanish. A rational mapping $F: \mathfrak{X} \to \Phi$ is called a *rational function.* We will always use x, y, z to denote the "independent variables" of rational mappings on $\mathfrak{X}$, while u, v, w and a, b, c will denote fixed vectors in $\mathfrak{X}$.

If F is a rational mapping of $\mathfrak{X}$ into $\mathfrak{Y}$ then ∂F denotes the differential of F and $\partial F|_x$ the differential of F at a point $x \in \mathfrak{X}$; the latter is a linear map from $\mathfrak{X}$ into $\mathfrak{Y}$, and we denote by $\partial_u F|_x$ its value $\partial F|_x(u)$ at a vector $u \in \mathfrak{X}$. Relative to bases $\{x_i\}$ for $\mathfrak{X}$ and $\{y_j\}$ for $\mathfrak{Y}$ the matrix of $\partial F|_x$ is the Jacobian $(\partial_j F_i|_x)$ where $\partial_j F_i|_x$ is the formal partial derivative of $F_i(\xi_1, \cdots, \xi_n)$ with respect to the indeterminate ξ_j evaluated at x. $x \to \partial_u F|_x$ is a rational mapping $\partial_u F$ of $\mathfrak{X}$ into $\mathfrak{Y}$, and the map $F \to \partial_u F$ is just partial derivation ∂_u in the direction u.

As an example, in this notation the *chain rule*, which is fundamental in the sequel, becomes

$$\partial\{F \circ G\}|_x = \partial F|_{G(x)} \circ \partial G|_x$$

or

$$\partial_u\{F \circ G\}|_x = \partial_v F|_{G(x)} \quad \text{for} \quad v = \partial_u G|_x .$$

The *logarithmic derivative*

$$\partial_u \log F = F^{-1}\partial_u F$$

of a rational function is well defined even though there is no function

$\log F$, and the usual rules hold:

$$\partial \log \{F \cdot G\} = \partial \log F + \partial \log G$$

$$\partial_u \partial_v \log F = \partial_v \partial_u \log F = F^{-2}\{F \cdot \partial_u \partial_v F - \partial_u F \cdot \partial_v F\} .$$

If F is a homogeneous mapping of degree q, ie. $F(\lambda x) = \lambda^q F(x)$, then the *Euler differential equations* imply.

$$\partial_x F|_x = qF(x) .$$

Finally, if F is a rational function such that $\partial_u F = 0$ for all $u \in \mathfrak{X}$ then $F \in \Phi(\xi_1^p, \cdots, \xi_n^p)$ relative to any basis for $\mathfrak{X}$ (p always denotes the characteristic of Φ; if $p = 0$ the condition is that $F \in \Phi$).

2. Forms admitting composition. By a *form* on a vector space $\mathfrak{X}$ we mean a homogeneous polynomial function; *throughout the rest of this chapter Q will denote a form on $\mathfrak{X}$ of degree $q > 0$.* If $\mathfrak{A}$ is an algebra on $\mathfrak{X}$ with identity c we say a form Q *admits composition* on $\mathfrak{A}$ if there are two rational mappings $E: x \to E_x$, $F: x \to F_x$ of $\mathfrak{X}$ into Hom $(\mathfrak{X}, \mathfrak{X})$ satisfying

(1.1)
- (a) $E_c = F_c = I$
- (b) $\partial_u E|_c = \alpha L_u$, $\partial_u F|_c = \beta R_u$ for $0 \neq \alpha, \beta \in \Phi$ where L_u, R_u are left and right multiplications by $u \in \mathfrak{A}$
- (c) $Q(E_x y) = e(x)Q(y)$, $Q(F_x y) = f(x)Q(y)$ for some rational functions e, f on $\mathfrak{X}$ wherever all mappings involved are defined.

Note that Q admits composition on any extension algebra $\mathfrak{A}_\Omega$, Ω an extension field of Φ.

For example, if $\mathfrak{A}$ admits associative composition $Q(xy) = Q(x)Q(y)$ we may take $E_x = L_x$, $F_x = R_x$, $\alpha = \beta = 1$, $e = f = Q$. If $\mathfrak{A}$ is commutative and admits Jordan composition $Q(U_x y) = Q(x)^2 Q(y)$, where $U_x = 2L_x^2 - L_{x^2}$, we may take $E_x = F_x = U_x$, $\alpha = \beta = 2$, $e = f = Q^2$. Indeed,

$$\begin{aligned} \partial_u U_x|_c &= 2\partial_u L_x^2|_c - L(\partial_u x^2|_c) \\ &= 2\{L_u L_c + L_c L_u\} - L_{cu+uc} = 2L_u = 2R_u . \end{aligned}$$

For typographical reasons we will often write $E(x)$, $L(x)$ in place of E_x, L_x etc.

A more complicated example is the following. Suppose $\mathfrak{A}$ is *quasi-associative*, that is, there is an extension Ω of Φ such that $\mathfrak{A}_\Omega$ is obtained by defining $xy = \lambda x \cdot y + (1 - \lambda) y \cdot x$ for some $\lambda \in \Omega$ where $\tilde{\mathfrak{A}} = (\mathfrak{X}_\Omega, \cdot)$ is an associative algebra on $\mathfrak{X}_\Omega$ (if $\mathfrak{A}$ itself is associative we assume $\tilde{\mathfrak{A}} = \mathfrak{A}$, $\lambda = 1$). Suppose Q is a form on $\mathfrak{X}$ whose extension to $\mathfrak{X}_\Omega$ satisfies $Q(x \cdot y) = Q(x)Q(y)$ identically (such as the generic norm of $\mathfrak{A}$, since

$\mathfrak{A}_\Omega$ and $\tilde{\mathfrak{A}}$ have the same generic norm; see the Corollary to Theorem 2.5). Set $\varphi = \lambda(1-\lambda)$, $E_x = L_x + \varphi U_{x-c}$, $F_x = R_x + \varphi U_{x-c}$, $\alpha = \beta = 1$, $e(x) = f(x) = Q(\varphi(x-c)^2 + x)$ where $U_x = \frac{1}{2}(L_x + R_x)^2 - \frac{1}{2}(L_{x^2} + R_{x^2})$. It is known [1, p. 583] that $\varphi \in \Phi$,[1] so E, F, e, f are all rational mappings on $\mathfrak{X}$ (not just $\mathfrak{X}_Q$). Clearly $E_c = F_c = I$. It is also easily checked that $U_x y = x \cdot y \cdot x$ in terms of the multiplication in $\tilde{\mathfrak{A}}$. Then

$$\partial_u \{U_{x-c}\}|_c (y) = u \cdot y \cdot (c-c) + (c-c) \cdot y \cdot u = 0 ,$$

so $\partial_u E_x|_c = L_u$, $\partial_u F_x|_c = R_u$. Finally,

$$\begin{aligned} Q(E_x y) &= Q(xy + \varphi U_{x-c} y) \\ &= Q(\lambda x \cdot y + (1-\lambda) y \cdot x + \lambda(1-\lambda)(x-c) \cdot y \cdot (x-c)) \\ &= Q(\{\lambda(x-c) + c\} \cdot y \cdot \{(1-\lambda)(x-c) + c\}) \\ &= Q(\lambda(x-c) + c) Q((1-\lambda)(x-c) + c) Q(y) \\ &\qquad\qquad \text{(by assumption on } Q) \\ &= Q(\lambda(1-\lambda)(x-c)^2 + (x-c) + c) Q(y) \\ &= e(x) Q(y) . \end{aligned}$$

Similarly $Q(F_x y) = f(x) Q(y)$, so Q admits composition on $\mathfrak{A}$.

The following weak form of the open mapping theorem is well known.

LEMMA 1.1. *If $F: \mathfrak{X} \to \mathfrak{Y}$ is a rational mapping whose differential ∂F is surjective at a point $x \in \mathfrak{X}$ then there is no nonzero rational function $G: \mathfrak{Y} \to \Phi$ which vanishes on the range $F(\mathfrak{X})$ of F.*

Proof. If there were such a G there would be one which was a polynomial; since Φ is infinite, the hypotheses would remain valid over a perfect extension Ω of Φ. But there the nonexistence of such a polynomial follows from [7, p. 268].

Suppose Q admits composition on an algebra $\mathfrak{A}$ with identity c. The mapping $\tilde{E}: x \to E_x c$ has differential αI at c since

---

[1] In fact, if $L_x, R_x, \tilde{L}_x, \tilde{R}_x$ denote the multiplications in $\mathfrak{A}, \tilde{\mathfrak{A}}$ then

$$L_x = \lambda \tilde{L}_x + (1-\lambda)\tilde{R}_x , \quad R_x = \lambda \tilde{R}_x + (1-\lambda)\tilde{L}_x , \quad L_x + R_x = \tilde{L}_x + \tilde{R}_x .$$

Associativity of $\tilde{\mathfrak{A}}$ means $[\tilde{L}_u, \tilde{R}_v] = 0$ for all u, v. Hence

$$\begin{aligned} [L_x, R_y] &= \lambda(1-\lambda)\{[\tilde{L}_x, \tilde{L}_y] + [\tilde{R}_x, \tilde{R}_y]\} \\ &= \lambda(1-\lambda)[\tilde{L}_x + \tilde{R}_x, \tilde{L}_y + \tilde{R}_y] \\ &= \lambda(1-\lambda)[L_x + R_x, L + R_y] . \end{aligned}$$

If we can find $x, y \in \mathfrak{A} \subset \mathfrak{A}_\Omega$ with $[L_x, R_y] \neq 0$ we can conclude that $\varphi = \lambda(1-\lambda) \in \Phi$; otherwise, $[L_x, R_y] = 0$ identically and $\mathfrak{A}$ is associative, so by assumption $\lambda = 1$, $\varphi = 0 \in \Phi$.

$$\partial\tilde{E}\,|_c\,(u) = \partial_u\tilde{E}\,|_c = (\partial_u E_x\,|_c)c = \alpha L_u c = \alpha u\ .$$

If $Q(c) = 0$ then $Q(\tilde{E}(x)) = Q(E_x c) = e(x)Q(c) = 0$ would imply $Q = 0$ by Lemma 1.1, contradicting $\deg Q > 0$. Thus if Q admits composition we necessarily have $Q(c) \neq 0$; in particular, we can always normalize Q so that $Q(c) = 1$ if we wish.

We define a rational mapping $x \to \tau_x$ of $\mathfrak{X}$ into the space of symmetric bilinear forms on $\mathfrak{X}$ by

$$\tau_x(u, v) = -\partial_u\partial_v \log Q\,|_x\ .$$

The form τ_x is defined whenever $Q(x) \neq 0$. We just saw $Q(c) \neq 0$, so $\tau = \tau_c$ is defined; τ is the *trace form* of Q. If Q is the generic norm, τ is the generic trace, and if $Q(x) = \det L_x$ or $\det U_x$ for $U_x = 2L_x^2 - L_{x^2}$ then $\tau = tr$ or $2tr$ where tr is the usual trace form $tr(u, v) = \operatorname{trace} L_{uv}$.

LEMMA 1.2. *If Q admits composition on $\mathfrak{A}$ then for those x, y where all functions involved are defined we have*

$$\partial_{E(x)v} \log Q\,|_{E(x)y} = \partial_v \log Q\,|_y \tag{1.2}$$

$$\tau_{E(x)y}(E_x u, E_x v) = \tau_y(u, v) \tag{1.3}$$

$$\tau_y(L_u y, v) = \partial_{L(u)v} \log Q\,|_y \tag{1.4}$$

$$\tau(L_u w, v) + \tau(w, L_u v) = \partial_w\partial_u\partial_v \log Q\,|_c \tag{1.5}$$

and dually with E, L replaced by F, R. Also, τ is an associative form:

$$\tau(uv, w) = \tau(u, vw)\ . \tag{1.6}$$

Proof. For (1.2), first observe that $E_x: y \to E_x y$ is linear in y, so $\partial E_x\,|_y = E_x$ where ∂ is applied to functions of y. Then using the chain rule we have

$$\begin{aligned}
\partial_{E(x)v} \log Q\,|_{E(x)y} &= \partial \log Q\,|_{E(x)y}\,(E_x v) \\
&= \partial \log Q\,|_{E(x)y} \circ \partial E_x\,|_y\,(v) \\
&= \partial \log \{Q \circ E_x\}\,|_y\,(v) = \partial_v\{\log e(x)Q\}\,|_y \\
&= \partial_v \log e(x)\,|_y + \partial_v \log Q\,|_y \\
&= \partial_v \log Q|_y\ .
\end{aligned}$$

Regarding (1.2) as a function of y and applying $\partial_u\,|_y$, the right side becomes $\partial_u\partial_v \log Q\,|_y = -\tau_y(u, v)$, while the left side yields

$$\partial_u\{\{\partial_{E(x)v} \log Q\} \circ E_x\}\,|_y = \partial_w\{\partial_{E(x)v} \log Q\}\,|_{E(x)y}$$

by the chain rule where $w = \partial_u E_x\,|_y = E_x u$; thus the left side finally

reduces to $\partial_{E(x)u}\partial_{E(x)v}\log Q\,|_{E(x)y} = -\tau_{E(x)y}(E_xu, E_xv)$. Equating gives (1.3).

Now regard (1.2) as a function of x and apply $\partial_u\,|_c$. The differentiation is routine, but we will go through it in detail this once. The left side of (1.2) is $F(E_xv, E_xy)$ if $F(z, y) = \partial_z \log Q\,|_y$, and

$$\partial_u F(E_xv, E_xy)\,|_c = \partial_u F(E_xv, E_cy)\,|_c + \partial_u F(E_cv, E_xy)\,|_c$$

by the usual rule for differentiating a function of two variables. F is linear in the first variable so we can move the partial inside; hence the first term is $F(\partial_u\{E_xv\}\,|_c, y) = F(\alpha L_uv, y) = \alpha\partial_{L(u)v}\log Q\,|_y$. By the chain rule the second term reduces to $\partial_w F(v, x)\,|_{E(c)y} = \partial_w\partial_v \log Q\,|_y = -\tau_y(w, v)$ for $w = \partial_u\{E_xy\}\,|_c = \alpha L_uy$. Now the right side of (1.2) is independent of x, so applying $\partial_u\,|_c$ gives 0. Thus

$$\alpha\partial_{L(u)v}\log Q\,|_y - \tau_y(w, v) = 0\ ,$$
$$\alpha\partial_{L(u)v}\log Q\,|_y = \alpha\tau_y(L_uy, v)\ .$$

Canceling α gives (1.4).

Similarly, applying $\partial_w\,|_c$ to (1.4) as a function of y gives

$$\tau_c(L_uw, v) + \partial_w\{\tau_y(u, v)\}\,|_c = \tau(L_uw, v) - \partial_w\partial_u\partial_v \log Q\,|_c$$

for the left side and $\partial_w\partial_{L(u)v}\log Q\,|_c = -\tau(w, L_uv)$ for the right side. Equating gives (1.5).

Interchanging u and v in the dual of (1.5) we get

$$\tau(R_wu, v) + \tau(u, R_wv) = \partial_u\partial_v\partial_w \log Q\,|_c\ .$$

The latter is symmetric in u, v, w so comparing with (1.5) we get $\tau(u, vw) = \tau(w, uv)$, proving (1.6).

3. Nondegenerate forms. We call a form Q on an algebra $\mathfrak{A}$ with identity c *nondegenerate* if the trace form $\tau = \tau_c$ is a nondegenerate bilinear form.

Assume for the moment that the characteristic of Φ is 0 or $p > q$, and let Q be an arbitrary form of degree q on a vector space $\mathfrak{X}$. Then $q!$ is not zero in Φ; we claim

$$[x_1, \cdots, x_q] = 1/q!\,\partial_{x_1}\cdots\partial_{x_q}Q\,|_c$$

is a symmetric q-linear form with $Q(x) = [x, \cdots, x]$. Clearly it is a symmetric and multilinear. Note that since Q is of degree q, $\partial_{x_1}\cdots\partial_{x_q}Q$ is of degree 0, ie. constant, so $[x_1, \cdots, x_q] = 1/q!\,\partial_{x_1}\cdots\partial_{x_q}Q\,|_x$ for any x. From the Euler equations we have

$$q![x, \cdots, x] = \overbrace{\partial_x \cdots \partial_x}^{q} Q\,|_x = 1!\overbrace{\partial_x \cdots \partial_x}^{q-1} Q\,|_x = \cdots$$
$$= (q-1)!\overbrace{\partial_x}^{1} Q\,|_x = q!\,Q(x)\,. \tag{1.7}$$

Thus $Q(x) = [x, \cdots, x]$, and hence $[x_1, \cdots, x_q]$ is the unique symmetric q-linear form obtained by polarizing Q. R. D. Schafer has called Q nondegenerate if there is no $u \neq 0$ such that $[x, \cdots, x, u] = 0$ for all x (or equivalently, by linearizing, $[x_1, \cdots, x_{q-1}, u] = 0$ for all x_k). As in (1.7) we see $[x, \cdots, x, u] = 0$ for all x if and only if $1/q\,\partial_u Q\,|_x = 0$, so Schafer's condition of nondegeneracy is that $\partial_u Q = 0 \Rightarrow u = 0$. Since the characteristic is 0 or $p > q$, this means that Q is not independent of any variables in the sense that there is no basis $\{x_1, \cdots, x_n\}$ for $\mathfrak{X}$ relative to which $Q \in \Phi[\xi_1, \cdots, \xi_{n-1}]$. Certainly this is a reasonable restriction.

If Q is nondegenerate in our sense it is always nondegenerate in Schafer's, since $\partial_u Q = 0$ would imply

$$\tau(u, v) = -\partial_v \partial_u \log Q\,|_c = -\partial_v\{Q^{-1}\partial_u Q\}\,|_c = 0$$

for all v. We claim the converse is true if Q is a form admitting composition on an algebra $\mathfrak{A}$ (still assuming the characteristic is 0 or $p > q$). Suppose such a Q is degenerate in our sense, so τ has a nonzero radical $\mathfrak{R}$. τ is associative by (1.6), so $\mathfrak{R}$ is an ideal. As we remarked after Lemma 1.1, we may assume $Q(c) = 1$. Then $\partial_u Q\,|_c = Q(c)^{-1}\partial_u Q\,|_c = \partial_u \log Q\,|_c = \tau(c, u)$ (taking $y = v = c$ in (1.4)) so as in (1.7) $[c, \cdots, c, u] = 1/q\,\partial_u Q\,|_c = 1/q\,\tau(c, u) = 0$ if $u \in \mathfrak{R}$. Thus we have proven the relation $[c, \cdots, c, y, \cdots, y, u] = 0$ for all $u \in \mathfrak{R}$, $y \in \mathfrak{A}$ when the number of y's is 0. Assume it proven when there are m y's. Applying $\partial_{x_1} \cdots \partial_{x_q}$ to $Q(E_x y) = e(x)Q(y)$ as a function of y gives

$$[E_x x_1, \cdots, E_x x_q] = e(x)[x_1, \cdots, x_q]\,.$$

By our induction hypothesis we conclude

$$[E_x c, \cdots, E_x c, E_x y, \cdots, E_x y, E_x u] = 0$$

for $u \in \mathfrak{R}$ if there are m y's. If we apply $\partial_y\,|_c$ to this as a function of x we get three groups of terms. Applying $\partial_y\,|_c$ to $E_x u$ gives the term $[c, \cdots, c, y, \cdots, y, u^*]$ where there are m y's and where

$$u^* = \partial_y\{E_x u\}\,|_c = \alpha L_y u = \alpha y u\,;$$

but $u^* \in \mathfrak{R}$ since $\mathfrak{R}$ is an ideal, so the term vanishes by the induction hypothesis. Applying $\partial_y\,|_c$ to the $E_x y$'s gives the term $m[c, \cdots, c, y, \cdots, y, y^*, u]$ where there are $m - 1$ y's and where $y^* = \alpha y y$. But this is just the

result of applying $\partial_{y^*}|_y$ to the induction hypothesis as a function of y, so it vanishes too. Thus the final term $(q-1-m)$ $[c, \cdots, c, c^*, y, \cdots, y, u]$ obtained from the $E_x c$'s must vanish. But $c^* = \alpha y$, so $[c, \cdots, c, y, \cdots, y, u] = 0$ when there are $m+1$ y's, and the induction is complete. Hence $[y, \cdots, y, u] = 0$ for all $u \in \mathfrak{R}$, $y \in \mathfrak{A}$ and Q is degenerate in Schafer's sense.

4. **Normed algebras.** A *norm* on an algebra $\mathfrak{A}$ is a nondegenerate form Q admitting composition on $\mathfrak{A}$. If $\mathfrak{A}$ has a norm it is called a *normed algebra*. Since a nondegenerate form remains nondegenerate on any extension of the base field, $\mathfrak{A}_\Omega$ is normed for all extensions Ω of Φ.

We saw in § 2 that the algebras studied by Schafer [11, 12] were normed algebras. Another important example of a normed algebra is the following. Suppose $\mathfrak{X}$ is just a vector space, with no algebra in sight. Let Q be a form on $\mathfrak{X}$, $c \in \mathfrak{X}$ a point where the trace form $\tau_c(u, v) = -\partial_u \partial_v \log Q|_c$ is nondegenerate. Then each bilinear form τ_x is obtained from $\tau = \tau_c$ as

$$\tau_x(u, v) = \tau(H_x u, v)$$

where $H: x \to H_x$ is a rational mapping of $\mathfrak{X}$ into Hom $(\mathfrak{X}, \mathfrak{X})$ defined for all x with $Q(x) \neq 0$. H depends on the choice of basepoint c. $H_c = I$, and the H_x's are self-adjoint relative to τ since each τ_x is symmetric. Consider the symmetric trilinear form

$$\sigma(u, v, w) = -\tfrac{1}{2}\partial_u\{\tau_x(v, w)\}|_c = \tfrac{1}{2}\partial_u\partial_v\partial_w \log Q|_c ;$$

since τ is nondegenerate $\sigma(u, v, w) = \tau(L_u v, w)$ for some $L_u \in$ Hom $(\mathfrak{X}, \mathfrak{X})$ where $u \to L_u$ is linear. Then $u \cdot v = L_u v$ defines a bilinear pairing on $\mathfrak{X}$. From the definition of σ we see $L_u = -\tfrac{1}{2}\partial_u H_x|_c$; H is homogeneous of degree -2 in x since τ_x is, so the Euler equations imply $\partial_c H_x|_c = -2H_c$, and $L_c = H_c = I$. Since τ is symmetric the pairing is commutative, so $u \cdot c = c \cdot u = u$. If Q admits composition in the sense that $Q(H_x y) = h(x)Q(y)$ for some rational function h whenever both sides are defined we define the *Koecher algebra* $\mathfrak{A}(Q, c)$ to be $\mathfrak{A} = (\mathfrak{X}, \cdot)$. Then $\mathfrak{A}$ is a commutative algebra with c as identity. Taking $E = F = H$, $\alpha = \beta = -2$, $e = f = h$ we see that $\mathfrak{A}$ is a normed algebra. Such algebras were first studied by M. Koecher [9, pp. 39–43] in connection with ω-domains. We will see later there is a close connection between normed algebras in general and Koecher algebras (Theorem 2.9).

From now on let $\tau = \tau_c$ be the trace form of a norm Q on an algebra $\mathfrak{A}$. We define H_x for $Q(x) \neq 0$ as above by $\tau(H_x u, v) = \tau_x(u, v)$. Since τ is nondegenerate, each linear functional $\partial \log Q|_x$ can be obtained from τ as

$$\partial_v \log Q\,|_x = \tau(x\#, v)$$

for some vector $x\#$. $\#$ is a rational mapping of $\mathfrak{X}$ into itself defined whenever $Q(x) \neq 0$. We have $\partial_u \tau(x\#, v)\,|_x = \partial_u \partial_v \log Q\,|_x = -\tau_x(u, v) = -\tau(H_x u, v)$, so $\partial_u \#\,|_x = -H_x u$.

LEMMA 1.3. *If $\mathfrak{A}$ is a normed algebra then for those x for which all mappings involved are defined we have*

$$L_x^* = R_x\,, \quad R_x^* = L_x \qquad (1.8)$$

(* *the adjoint relative to τ*)

$$H_x R_x = L_{x\#}\,, \quad H_x L_x = R_{x\#} \qquad (1.9)$$

$$H_x, R_x, L_{x\#} \text{ commute}\,; \quad H_x, L_x, R_{x\#} \text{ commute} \qquad (1.10)$$

$$x\#\cdot x = x\cdot x\# = c \qquad (1.11)$$

$$x\#\cdot x^2 = x^2\cdot x\# = x \qquad (1.12)$$

$$U_x H_x = I \text{ where } U_x = R_x(R_x + L_x) - R_{x^2} = L_x(L_x + R_x) - L_{x^2} \qquad (1.13)$$

$$\mathfrak{A} \text{ has no nonzero ideals } \mathfrak{B} \text{ with } \mathfrak{B}^2 = 0\,. \qquad (1.14)$$

Proof. By our assumptions it suffices to pass to the algebraic closure Ω of Φ and prove the result there.

(1.8) is just the associativity (1.6) of τ.

$$\begin{aligned}\tau(H_x L_u x, v) &= \tau_x(L_u x, v)\\ &= \partial_{L(u)v} \log Q\,|_x = \tau(x\#, L_u v) = \tau(L_u^* x\#, v)\end{aligned}$$

from (1.4) so that

$$H_x R_x u = H_x(u\cdot x) = H_x L_u x = L_u^* x\# = R_u x\# = L_{x\#} u\,;$$

dually $H_x L_x = R_{x\#}$, so (1.9) is proven. Hence we have

$$R_x H_x = L_x^* H_x^* = (H_x L_x)^* = R_{x\#}^* = L_{x\#} = H_x R_x\,,$$

so R_x and H_x commute, hence both commute with their product $L_{x\#}$. The dual result holds, so (1.10) is proven.

(1.11) is a bit longer. First,

$$\begin{aligned}\partial_u\{x\#\cdot x\}\,|_x &= \{\partial_u x\#\,|_x\}\cdot x + x\#\cdot\{\partial_u x\,|_x\}\\ &= -H_x u\cdot x + x\#\cdot u = \{-R_x H_x + L_{x\#}\}u = 0\end{aligned}$$

by (1.9), (1.10) for all u, x. Hence

$$0 = \partial_x \tau(x\#\cdot x, v) = \partial_u \tau(x\#, x\cdot v) = \partial_u\{\partial_{x\cdot v} \log Q\,|_x\}$$

for all u, v, x. Thus relative to a basis for $\mathfrak{X}$ $\partial_{x\cdot v} \log Q|_x$ is in $\Omega(\xi_1^p, \cdots, \xi_n^p)$ where p is the characteristic of Ω; since Ω is perfect, it must be a pth power. If $Q = \prod Q_i^{q_i}$ for Q_i distinct irreducible factors then

$$\begin{aligned}\partial_{x\cdot v} \log Q &= \Sigma q_i \partial_{x\cdot v} \log Q_i \\ &= \Sigma q_i Q_i^{-1} \partial_{x\cdot v} Q_i = R^{-1} \Sigma q_i R_i \partial_{x\cdot v} Q_i = R^{-1} S\end{aligned}$$

for $R = \prod Q_i$, $R_i = Q_i^{-1} R$, S a polynomial. But in reduced form $R^{-1}S$ must be a pth power, and the denominator of the reduced form is composed of some of the distinct irreducible factors Q_i, so it can be a pth power only by being constant. Hence $\tau(x\sharp\cdot x, v) = \partial_{x\cdot v} \log Q|_x$ itself is a polynomial. Yet it is homogeneous of degree 0 since by the chain rule

$$\begin{aligned}\partial_{\lambda x\cdot v} \log Q|_{\lambda x} &= \partial_{x\ v} \log \{Q \circ \lambda\}|_x \\ &= \partial_{x\cdot v} \{\log \lambda^q + \log Q\}|_x = \partial_{x\cdot v} \log Q|_x ,\end{aligned}$$

so it must be constant. Thus $\tau(x\sharp\cdot x, v) = \tau(c\sharp\cdot c, v) = \tau(c, v)$ since $c\sharp\cdot c = L_{c^\sharp} c = H_c R_c c = c$ by (1.9). This holds for all v, so by nondegeneracy $x\sharp\cdot x = c$ wherever $x\sharp$ is defined. The dual result is established similarly, so (1.11) is finished.

(1.12) follows from (1.10), (1.11) by

$$x\sharp\cdot x^2 = L_{x^\sharp} R_x x = R_x L_{x^\sharp} x = R_x(x\sharp\cdot x) = R_x c = x$$

and dually. (1.13) holds since for all u

$$\begin{aligned}Iu = u = \partial_u x|_x &= \partial_u \{x\sharp\cdot x^2\}|_x \\ &= (-H_x u)\cdot x^2 + x\sharp\cdot(x\cdot u + u\cdot x) \\ &= \{-R_{x^2} H_x + L_{x\sharp}(L_x + R_x)\}u \\ &= \{-R_{x^2} H_x + H_x R_x(L_x + R_x)\}u \\ &= \{R_x(R_x + L_x) - R_{x^2}\} H_x u = U_x H_x u\end{aligned}$$

by (1.9), (1.10) and dually.

For (1.14), suppose $u, v \in \mathfrak{B}$ where $\mathfrak{B}$ is an ideal with $\mathfrak{B}^2 = 0$. Then $\tau(u, v) = \tau(c, u\cdot v) = 0$, so $\tau(\mathfrak{B}, \mathfrak{B}) = 0$. If $Q(x) \neq 0$ then H_x is defined, and by (1.13) $U_x H_x = I$, so U_x is invertible. Since $\mathfrak{B}$ is an ideal and U_x is composed of multiplications, $U_x \mathfrak{B} \subset \mathfrak{B}$; by nonsingularity $U_x \mathfrak{B} = \mathfrak{B} = H_x \mathfrak{B}$. Thus $\tau_x(u, v) = \tau(H_x u, v) \in \tau(\mathfrak{B}, \mathfrak{B}) = 0$. If $\{x_1, \cdots, x_n\}$ is a basis for $\mathfrak{X}$ with $\{x_{m+1}, \cdots, x_n\}$ a basis for $\mathfrak{B}$, and if $\tilde{\Omega}$ denotes the algebraic closure of $\Omega(\xi_1, \cdots, \xi_m)$ then $\partial_u \partial_v \log Q|_x = -\tau_x(u, v) = 0$ for $u, v \in \mathfrak{B}$ implies $\partial_v \log Q \in \tilde{\Omega}(\xi_{m+1}^p, \cdots, \xi_n^p)$. Repeating the argument of (1.11) with all factorization taking place over $\tilde{\Omega}$ we see $\partial_v \log Q$ must be a polynomial in $\xi_{m+1}, \cdots, \xi_n$ with coefficients in $\tilde{\Omega}$. But for $v \in \mathfrak{B}$, $\deg Q > \deg \partial_v Q = \deg \{Q\cdot \partial_v \log Q\} = \deg Q + \deg \partial_v \log Q \geqq \deg Q$ unless

$\partial_v \log Q$ is the zero polynomial (where deg means the degree as a polynomial in $\xi_{m+1}, \cdots, \xi_n$). Hence we must have $\partial_v \log Q = 0$. Then

$$\tau(w, v) = -\partial_w \partial_v \log Q \mid_c = 0$$

for all $w \in \mathfrak{A}$, and by nondegeneracy $v = 0, \mathfrak{B} = 0$.

Notice that the proof would be greatly simplified if we assumed the characteristic of Φ to be 0 or greater than q. At least in this case, formulas (1.8) to (1.13) arise quite natually when one starts differentiating the relations (1.1). In fact, they appear (disguised) in the work of O. S. Rothaus [10, p. 210] on the differential geometry of ω-domains.

5. **Norms and noncommutative Jordan algebras.** Recall that a *noncommutative Jordan algebra* [11] is a nonassociative algebra in which $L_x, R_x, L_{x^2}, R_{x^2}$ commute for each x. An algebra is *separable* if it is a direct sum of simple ideals with separable centers, or equivalently if it is semisimple and remains so under any extension of the base field.

THEOREM 1.1. *If $\mathfrak{A}$ is a normed algebra then it is a separable noncommutative Jordan algebra, and the symmetrized algebra $\mathfrak{A}^+$ is a separable commutative Jordan algebra.*

Proof. We can apply Lemmas 1.2 and 1.3. By (1.6), τ is a nondegenerate associative symmetric bilinear form, and by (1.14) $\mathfrak{A}$ has no ideals $\mathfrak{B} \neq 0$ with $\mathfrak{B}^2 = 0$, so by Dieudonne's theorem $\mathfrak{A}$ is semisimple. Any extension of $\mathfrak{A}$ remains normed, hence semisimple, so $\mathfrak{A}$ is separable. If $Q(x) \neq 0$, from (1.13), (1.8), and the selfadjointness of H_x we get

$$\begin{aligned} R_x(R_x + L_x) - R_{x^2} = H_x^{-1} &= (H_x^{-1})^* \\ &= \{L_x(L_x + R_x) - L_{x^2}\}^* \\ &= (R_x + L_x)R_x - R_{x^2}, \end{aligned}$$

hence R_x commutes with L_x. From (1.10) we see H_x, L_x, R_x generate a commutative algebra of linear transformations which contains U_x by (1.13), hence also R_{x^2} and L_{x^2}. Thus $R_x, L_x, R_{x^2}, L_{x^2}$ commute for those x where $Q(x) \neq 0$; since Φ is assumed to be infinite this set is dense, so they commute for all x, and $\mathfrak{A}$ is a noncommutative Jordan algebra.

In $\mathfrak{A}^+$ we have $L_x^+ = \frac{1}{2}(L_x + R_x) = R_x^+$, so $\mathfrak{A}^+$ is a commutative Jordan algebra. τ is still a nondegenerate associative form for $\mathfrak{A}^+$, and averaging (1.13) shows

$$U_x = \tfrac{1}{2}(R_x + L_x)^2 - \tfrac{1}{2}(R_{x^2} + L_{x^2}) = 2L_x^{+2} - L_{x^2}^+ = U_x^+$$

is composed of multiplications in $\mathfrak{A}^+$, so the proof of (1.14) carries over verbatim. Thus $\mathfrak{A}^+$ is semisimple by Dieudonne's theorem; the same holds for all extensions of $\mathfrak{A}^+$, and $\mathfrak{A}^+$ is separable.

It is known [1, p. 585 and argument on pp. 590–593] that this implies the simple summands of $\mathfrak{A}$ are either commutative Jordan algebras, quasiassociative algebras, or are of degree 2. We will see in Theorems 2.8 and 2.4 that the generic norm of a separable commutative Jordan algebra is nondegenerate and admits Jordan composition $N(U_x y) = N(x)^2 N(y)$ and therefore such algebras are normed. If $\mathfrak{A}$ is a separable quasiassociative algebra and Ω the extension of Φ such that $\mathfrak{A}_\Omega = \tilde{\mathfrak{A}}(\lambda)$ for $\tilde{\mathfrak{A}}$ associative, $\lambda \in \Omega$, then $\tilde{\mathfrak{A}}$ must be separable too and Theorem 2.8 shows that the generic norm is nondegenerate. By the remarks at the beginning of § 2 we see that $\mathfrak{A}$ is a normed algebra. However, it is not known if all separable flexible algebras of degree 2 are normed, so the converse of Theorem 1.1 is incomplete.

THEOREM 1.2. *If Q is a form on a vector space $\mathfrak{X}$, $c \in \mathfrak{X}$ a point where the trace form τ_c is nondegenerate and relative to which $Q(H_x y) = h(x)Q(y)$ whenever both sides are defined (where $\tau_c(H_x u, v) = \tau_x(u, v)$ and h is some rational function) then the Koecher algebra $\mathfrak{A}(Q, c)$ is a separable commutative Jordan algebra. If normalized, Q admits Jordan composition $Q(U_x y) = Q(x)^2 Q(y)$ for $U_x = 2L_x^2 - L_{x^2}$.*

Proof. We have observed before that $\mathfrak{A}$ is normed; since it is commutative, Theorem 1.1 shows that it is a separable commutative Jordan algebra. (1.13) shows that $U_x = H_x^{-1}$, so $Q(U_x y) = h(x)^{-1}Q(y)$. If Q is normalized, putting $y = c$ gives $Q(x^2) = Q(U_x c) = h(x)^{-1}Q(c) = h(x)^{-1}$, so $Q(U_x y) = Q(x^2)Q(y)$. Putting $y = x^2$ gives $Q((x^2)^2) = Q(x^2)^2$, so $Q(z^2) = Q(z)^2$ if z is of the form x^2. But the differential of $x \to x^2$ at c is $2I$ since $\partial_u x^2 |_c = 2u$; by Lemma 1.1, $Q(z^2) = Q(z)^2$ for all z. Thus $Q(U_x y) = Q(x)^2 Q(y)$.

REMARK. This is related to a result of Koecher's [8, Satz 5]. We note also that the above result holds over the field of real numbers if Q is a positive homogeneous real-analytic function ω on an open subset $\mathfrak{Y}$ of $\mathfrak{X}$ such that $c \in \mathfrak{Y}$ is a point where the Hessian of $\log \omega$ is nondegenerate and $\omega(H_x y) = \det H_x \cdot \omega(y)$ for all $x, y \in \mathfrak{Y}$. The Koecher algebra can be defined as before and the formulas of Lemmas 1.2 and 1.3 remain valid on $\mathfrak{Y}$. We can conclude that L_y commutes with L_{y^2} for $y \in \mathfrak{Y}$, and the commutativity extends to all $x \in \mathfrak{X}$ by analytic continuation. This gives an algebraic proof that the algebra of an ω-domain is a Jordan algebra [9, p. 44].

5. **Some results of R. D. Schafer.** In this section we will ont

assume that Φ is infinite, so we cannot immediately apply the methods of the differential calculus as previously formulated. A polynomial $Q \in \Phi[\xi_1, \cdots, \xi_n]$ together with a choice of basis $\{x_1, \cdots, x_n\}$ for a vector space $\mathfrak{X}$ determine in a canonical way a function on $\mathfrak{X}$; if this induced function $Q(u) = Q(\mu_1, \cdots, \mu_n)$ vanishes for all $u = \mu_1 x_1 + \cdots + \mu_n x_n$ in $\mathfrak{X}$ and if $|\Phi| > \deg Q$ then by the usual specialization theorem we conclude Q must be the zero polynomial. For $u = \Sigma \mu_i x_i$ we set $\partial_u Q = \Sigma \mu_i \partial_i Q$ where ∂_i is formal partial derivation with respect to the indeterminate ξ_i. Then we can define a bilinear form

$$\begin{aligned}\tau_x(u, v) &= -\partial_u \partial_v \log Q(x) \\ &= Q(x)^{-2}\{Q(x)\partial_u \partial_v Q(x) - \partial_u Q(x)\partial_v Q(x)\}\end{aligned}$$

whenever $Q(x) \neq 0$. Note that if Φ is infinite these definition agree with the usual ones of the differential calculus.

THEOREM 1.3. *Let $\mathfrak{A}$ be an algebra with identity c. If $\{x_1, \cdots, x_n\}$ is a basis for $\mathfrak{A}$ and $Q \in \Phi[\xi_1, \cdots, \xi_n]$ a homogeneous polynomial of degree q such that the trace form τ_c is defined and nondegenerate and $Q(ab) = Q(a)Q(b)$ for all $a, b \in \mathfrak{A}$, and if $|\Phi| > q$ then $\mathfrak{A}$ is a separable alternative algebra. Conversely, if $\mathfrak{A}$ is a separable alternative algebra and $\{x_1, \cdots, x_n\}$ a basis then the generic norm $N(\xi_1, \cdots, \xi_n)$ is nondegenerate and $N(ab) = N(a)N(b)$.*

Proof. The assumption that $|\Phi| > q$ implies that $Q(xy) = Q(x)Q(y)$ holds under extension of Φ to an infinite field Ω. As we have noted before, taking $E_x = L_x$, $F_x = R_x$ makes $\mathfrak{A}_\Omega$ a normed algebra. Hence by Theorem 1.1 it is separable. By (1.3), $\tau(u, v) = \tau_{E(x)c}(E_x u, E_x v) = \tau_x(L_x u, L_x v)$ so $I = L_x^* H_x L_x = R_x H_x L_x = R_x L_x H_x$ by (1.8), (1.10). But form (1.13) $I = U_x H_x$; thus $R_x L_x = U_x = R_x(R_x + L_x) - R_{x^2}$, whence $R_x^2 = R_{x^2}$. Dually $L_x^2 = L_{x^2}$, so $\mathfrak{A}_\Omega$ is alternative (again, strictly speaking this has been proved only for those x where $Q(x) \neq 0$, but since Ω is infinite this set is dense and hence the identities hold everywhere). Since $\mathfrak{A}_\Omega$ is a separable alternative algebra, so is $\mathfrak{A}$.

Conversely, if $\mathfrak{A}$ is a separable alternative algebra then by Theorem 2.8 and the Corollary to Theorem 2.5 the generic norm has nondegenerate trace form and $N(xy) = N(x)N(y)$.

THEOREM 1.4. *Let $\mathfrak{A}$ be an algebra with identity c. If $\{x_1, \cdots, x_n\}$ is a basis for $\mathfrak{A}$ and $Q \in \Phi[\xi_1, \cdots, \xi_n]$ a homogeneous polynomial of degree q such that the trace form τ_c is defined and nondegenerate and $Q(\{aba\}_1) = Q(a)^2 Q(b) = Q(\{aba\}_2)$ for all $a, b \in \mathfrak{A}$ where $\{aba\}_1 = 2a(ba) - ba^2$, $\{aba\}_2 = 2(ab)a - a^2 b$, and if $|\Phi| > 2q$ then*

(a) *$\mathfrak{A}$ is a noncommutative Jordan algebra*

(b) $\mathfrak{A}$ *and* $\mathfrak{A}^+$ *are separable*

(c) $(L_a - R_a)^2 = 0$ *for all* $a \in \mathfrak{A}$.

Conversely, if $\mathfrak{A}$ *satisfies* (a), (b), (c) *and if* $\{x_1, \cdots, x_n\}$ *is a basis then the generic norm* $N(\xi_1, \cdots, \xi_n)$ *has nondegenerate trace form and admits the above compositions.*

Proof. For the first part we may pass to an infinite extension Ω of Φ by our assumptions on $|\Phi|$. Let $E_x = 2L_xR_x - R_{x^2}$, $F_x = 2R_xL_x - L_{x^2}$, $\alpha = \beta = 2$, $e = f = Q^2$; then $E_e = F_e = I$, $\partial_u E_x|_e = 2L_u$, $\partial_u F_x|_e = 2R_u$, and $Q(E_xy) = Q(x)^2Q(y) = Q(F_xy)$ since $\{aba\}_1 = E_ab$, $\{aba\}_2 = F_ab$. By (1.1) we see that Q is a norm on $\mathfrak{A}_\Omega$. By Theorem 1.1, $\mathfrak{A}_\Omega$ and $\mathfrak{A}_\Omega^+$ are separable noncommutative and commutative Jordan algebras respectively. From (1.3) we have $\tau(u, v) = \tau_{E(x)c}(E_xu, E_xv)$ so $I = E_x^* H_{E(x)c} E_x = F_xH_{x^2}E_x$ by (1.8), and $E_xF_x = H_{x^2}^{-1}$. As in the proof of Theorem 1.1, $H_x^{-1} = U_x^+$; now $\mathfrak{A}_\Omega^+$ is a commutative Jordan algebra, so $U_{x^2}^+ = (U_x^+)^2$ (see (2.1) below). Thus $E_xF_x = (U_x^+)^2$. Applying $\partial_c|_x$ we get

$$\begin{aligned}\partial_c\{E_xF_x\}|_x &= \{2L_x + 2R_x - 2R_x\}F_x + E_x\{2R_x + 2L_x - 2L_x\} \\ &= 2\{L_xF_x + E_xR_x\}\end{aligned}$$

and $\partial_c(U_x^+)^2|_x = 2L_x^+U_x^+ + 2U_x^+L_x^+$; hence $L_xF_x + E_xR_x = L_x^+U_x^+ + U_x^+L_x^+$. Applying $\partial_c|_x$ again, the left side becomes

$$\begin{aligned}F_x + E_x + 4L_xR_x &= 4L_xR_x + 4R_xL_x - (L_{x^2} + R_{x^2}) \\ &= 8L_xR_x - 2L_{x^2}^+\end{aligned}$$

and the right side becomes $2U_x^+ + 4L_x^{+2} = 8L_x^{+2} - 2L_{x^2}^+$, so equating gives $L_x^{+2} = L_xR_x$, $(L_x + R_x)^2 = 4L_xR_x$, and finally $(L_x - R_x)^2 = 0$. Since these results hold for $\mathfrak{A}_\Omega$ they hold for $\mathfrak{A}$.

Conversely, suppose $\mathfrak{A}$, $\mathfrak{A}^+$ are separable; since the generic norm of $\mathfrak{A} = \bigoplus \mathfrak{A}_i$ is $N = \prod N_i$ it suffices to consider simple algebras $\mathfrak{A}$, and as we remarked after Theorem 1.1 such an $\mathfrak{A}$ is either a commutative Jordan algebra, a quasiassociative algebra, or is of degree 2. In the first case $E_x = F_x = U_x$, and the result follows from Theorems 2.4 and 2.8 below. Suppose $\mathfrak{A}$ is a separable quasiassociative algebra with $(L_a - R_a)^2 = 0$. Since this identity is of degree 2 it is valid on the extension $\mathfrak{A}_\Omega$, Ω the algebraic closure of Φ. Now $\mathfrak{A}_\Omega = \tilde{\mathfrak{A}}(\lambda)$ for $\lambda \in \Omega$, so if $\tilde{R}_a$, $\tilde{L}_a$ denote the multiplications in the associative algebra $\tilde{\mathfrak{A}}$ we have

$$\begin{aligned}L_a = \lambda\tilde{L}_a + (1 - \lambda)\tilde{R}_a\,, \quad R_a = \lambda\tilde{R}_a + (1 - \lambda)\tilde{L}_a\,, \\ (L_a - R_a)^2 = (2\lambda - 1)^2(\tilde{L}_a - \tilde{R}_a)^2\,.\end{aligned}$$

If $\lambda = \frac{1}{2}$, $\mathfrak{A}$ is a Jordan algebra, and we just saw the result holds in that case. Otherwise we must have $(\tilde{L}_a - \tilde{R}_a)^2 = 0$. This is impossible

in matrix algebras of degree > 1 (take a, b such that $ab = b \neq 0$, $ba = 0$), and since $\tilde{\mathfrak{A}}$ is separable it is a direct sum of matrix algebras; hence these must be of degree 1, and since Ω is algebraically closed $\tilde{\mathfrak{A}}$ is a direct sum of fields, so again $\mathfrak{A}$ is a commutative Jordan algebra.

It remains to consider the case where $\mathfrak{A}$ is a separable flexible algebra of degree 2 satisfying (b) and (c); these hypotheses remain valid on an infinite extension Ω of Φ, and it will suffice to show the extension of N is nondegenerate and admits the given composition on $\mathfrak{A}_\Omega$. Now (b) and Theorem 2.8 imply N is nondegenerate (since $\mathfrak{A}_\Omega$ and $\mathfrak{A}_\Omega^+$ have the same generic norm), so we only have to show N admits E_x, F_x as composition. Since $\mathfrak{A}_\Omega$ is flexible,

$$R_x(R_x + L_x) - R_{x^2} = L_x(L_x + R_x) - L_{x^2} = 2L_x^{+2} - L_{x^2}^+ = U_x^+$$

by averaging (see (2.6) below). Hence $E_x = U_x^+ + R_x(L_x - R_x)$, $F_x = U_x^+ + L_x(L_x - R_x)$. Write $\mathfrak{A}_\Omega = \Omega c \oplus \mathfrak{M}$ where $\mathfrak{M} = c^\perp$ is the orthogonal complement of Ωc under τ; since τ is the generic trace and $\mathfrak{A}_\Omega^+$ is of degree 2, $xy + yx \in \Omega c$ if $x, y \in \mathfrak{M}$. Thus if $xy = \alpha c + z$ $(z \in \mathfrak{M})$ then $yx = \beta c - z$. By flexibility,

$$\alpha x + zx = (xy)x = x(yx) = \beta x - xz\ ,$$
$$(\beta - \alpha)x = zx + xz \in \Omega c\ .$$

But $x \in \mathfrak{M}$, so $\beta = \alpha$. Then $\tau(c, [x, y]) = \tau(c, 2z) = 0$. Since

$$[\lambda c + x, \mu c + y] = [x, y]$$

we have $\tau(c, [u, v]) = 0$ for any $u, v \in \mathfrak{A}_\Omega$. Since

$$E_x w = U_x^+ w + (L_x - R_x)R_x w = U_x^+ w + [x, wx]$$

we have

$$\tau(c, E_x w) = \tau(c, U_x^+ w)\ .$$

Next,

$$\begin{aligned}\tau(\{L_x - R_x\}y, y) &= \tau(\{R_y - L_y\}x, R_y^+ c)\\ &= \tau(R_y^+\{R_y - L_y\}x, c) = \tau(\{R_y - L_y\}R_y^+ x, c)\\ &= \tau([R_y^+ x, y], c) = 0\end{aligned}$$

by the above since Theorem 2.4 and Lemma 1.2 show τ is associative for $\mathfrak{A}_\Omega^+$. Polarizing gives

$$(L_x - R_x) + (L_x - R_x)^* = 0\ ,$$

hence $(L_x - R_x)^*(L_x - R_x) = 0$ by (c). Now (c) also implies

$$L_x(L_x - R_x) = R_x(L_x - R_x) = L_x^+(L_x - R_x)\ ;$$

all operators commute here, so

$$\begin{aligned} E_x^* E_x &= \{U_x^+ + L_x^+(L_x - R_x)^*\}\{U_x^+ + L_x^+(L_x - R_x)\} \\ &= U_x^{+2} + U_x^+ L_x^+\{(L_x - R_x) + (L_x - R_x)^*\} \\ &\quad + L_x^{+2}(L_x - R_x)^*(L_x - R_x) \\ &= U_x^{+2} = U_x^{+*} U_x^+ \end{aligned}$$

and finally

$$\tau(E_x u, E_x v) = \tau(U_x^+ u, U_x^+ v) .$$

But from $N(c) = 1$ and $\deg N = 2$ we get

$$\begin{aligned} \tau(u, v) &= N(c)^{-2}\{\partial_u N|_c \, \partial_v N|_c - N(c)\partial_u\partial_v N|_c\} \\ &= \tau(c\#, u)\tau(c\#, v) - \partial_u N|_v . \end{aligned}$$

By (1.11), $c\# = c\# \cdot c = c$; therefore

$$\begin{aligned} \partial_u\{N \circ E_x\}|_v &= \partial_{E(x)u} N|_{E(x)v} \\ &= \tau(c, E_x u)\tau(c, E_x v) - \tau(E_x u, E_x v) \\ &= \tau(c, U_x^+ u)\tau(c, U_x^+ v) - \tau(U_x^+ u, U_x^+ v) \\ &= \partial_{U+(x)u} N|_{U+(x)v} = \partial_u\{N \circ U_x^+\}|_v \end{aligned}$$

for all u, v; from $\partial_u\{N \circ E_x - N \circ U_x^+\}|_v = 0$ we see $N \circ E_x - N \circ U_x^+$ is in $\Omega[\xi_1^p, \cdots, \xi_n^p]$. But it is homogeneous of degree 2, so $N \circ E_x = N \circ U_x^+$, and by Theorem 2.4 $N(E_x y) = N(U_x^+ y) = N(x)^2 N(y)$. Similarly for F_x, so N admits composition.

These theorems are due to Schafer [11, 12]. He assumed that Q was given as a function $Q: x \to [x, \cdots, x]$ on $\mathfrak{A}$ where $[, \cdots,]$ was a q-linear form. It is easy to see that a basis $\{x_1, \cdots, x_n\}$ for $\mathfrak{A}$ determines a unique homogeneous polynomial $Q \in \Phi[\xi_1, \cdots, \xi_n]$ of degree q such that the functional relation $Q(x) = [x, \cdots, x]$ remains valid under all extensions of Φ, so our assumptions are essentially equivalent to his. His formulation in terms of functions has the advantage of being intrinsic, but it requires that Φ have characteristic 0 or $p > q$, while the above proofs hold if only $|\Phi| > q$ or $|\Phi| > 2q$.

Schafer proved Theorem 1.4 only for the special cases $q = 2, 3$.[2] Different choices for the noncommutative ternary compositions lead to slightly different classes of algebras; for example,

$$E_x = 2L_x^2 - L_{x^2} , \quad F_x = 2R_x^2 - R_{x^2}$$

leads to the algebras satisfying $(L_x - R_x)^2(L_x^2 - L_{x^2}) = 0$, which includes all alternative algebras.

[2] Professor L. Paige informs me that Mrs. E. Papousek has independently extended Schafer's results to arbitrary q.

When $p = 0$ or $p > q$ analogous results hold even if the dimension of $\mathfrak{A}$ is infinite. The differential calculus, Lemma 1.2, and the discussion of nondegeneracy carry over straightforwardly; in the analogue of Lemma 1.3 we cannot define H_x or $x\sharp$, so the notation is less convenient, but the proof is not much longer. Although we cannot draw any conclusions about separability, we can conclude that an infinite dimensional normed algebra is a noncommutative Jordan algebra (Theorem 1.1) and from this we can get Schafer's result that infinite-dimensional algebras admitting associative or Jordan composition are alternative or Jordan algebras respectively (Theorems 1.3 and 1.4). However, it is conjectured that all normed algebras are necessarily finite-dimensional, so for convenience we have restricted ourselves to the finite-dimensional case.

CHAPTER II

1. Inverses. In this chapter we do not assume that Φ is infinite, and in this first section we do not assume that the algebras are finite-dimensional. Let $\mathfrak{A}$ be a commutative Jordan algebra; then we have the following identities [5, pp. 1155–1156]:

$$U_{U(x)y} = U_x U_y U_x \qquad (U_x = 2L_x^2 - L_{x^2}) \tag{2.1}$$

$$[L_x, L_{yz}] + [L_y, L_{zx}] + [L_z, L_{xy}] = 0 \tag{2.2}$$

$$L_z L_y L_x + L_x L_y L_z + L_{(xz)y} = L_{yz} L_x + L_{xz} L_y + L_{yx} L_z \tag{2.3}$$

$$L_{x^n} = 2L_{x^{n-1}} L_x - L_{x^{n-2}} U_x \qquad (n \geqq 2)\,. \tag{2.4}$$

We say that an element $a \in \mathfrak{A}$ is *regular* with *inverse* b if $ab = c$ and $a^2 b = a$ (where c is the identity of $\mathfrak{A}$).

THEOREM 2.1. *Let $\mathfrak{A}$ be a commutative Jordan algebra with identity c. Then for $a \in \mathfrak{A}$ the following are equivalent:*

(a) *a is regular*

(b) *c is in the range of U_a*

(c) *U_a is an invertible transformation.*

The inverse is unique; if a has inverse b then b has inverse a and $U_a U_b = U_b U_a = I$. In this case $L_b = U_a^{-1} L_a$, all L_{ak}, L_{bj} commute, and $\Phi(a, b)$ is a commutative associative algebra.

Proof. If U_a is an invertible transformation then clearly c is in its range; conversely, if c is in the range of U_a, say $U_a d = c$, then by (2.1) $I = U_c = U_{U(a)d} = U_a U_d U_a$, so U_a is invertible.

If U_a is invertible and $b = U_a^{-1} a$ then $U_a c = a^2 = L_a a = L_a U_a b =$

$U_aL_ab = U_a(ab)$ (U_a commutes with L_a since L_{a^2} commutes with L_a in a Jordan algebra). Since U_a is one-to-one, $ab = c$. We get $U_aa = U_a(a^2b)$ in the same way, so $a^2b = a$ (which also follows from

$$a = U_ab = 2a(ab) - a^2b = 2a - a^2b) .$$

Hence a is regular with inverse b.

If a is regular with inverse b, then $L_{ab} = I$, so (2.2) shows $[L_a, L_{b^2}] = [L_b, L_{a^2}] = 0$. Thus

$$\begin{aligned} U_ab^2 &= 2L_a^2(L_{b^2}c) - L_{a^2}(b^2) = 2L_{b^2}(L_a^2c) - a^2b^2 \\ &= a^2b^2 = L_{a^2}L_bb = L_bL_{a^2}b = L_ba = c \end{aligned}$$

is in the range of U_a.

$U_ab = 2a(ab) - a^2b = a$ shows the inverse b is uniquely determined as $U_a^{-1}a$; it also shows $U_a = U_{U(a)b} = U_aU_bU_a$ by (2.1), so $U_aU_b = U_bU_a = I$. Then b has inverse $U_b^{-1}b = U_ab = a$.

In this case $[L_b, L_{a^2}] = [L_b, U_a] = 0$ (since $U_a = U_b^{-1}$), so L_b commutes with $2L_a^2 = U_a + L_{a^2}$. Thus putting $x = y = a, z = b$ in (2.3) yields $L_a = (2L_a^2 - L_{a^2})L_b = U_aL_b$, $L_b = U_a^{-1}L_a$. Hence U_b, L_b are in the commutative algebra of linear transformations generated by L_a, L_{a^2}, U_a^{-1}; by (2.4) this includes all L_{a^k}, L_{b^j}, so they all commute. Then it is easy to see that $\Phi(a, b)$ is a commutative associative algebra.

This theorem is due to N. Jacobson [4, 5]. The above proof is simpler than his; the simplification comes from repeated use of the "fundamental formula" (2.1), which was unknown when he first proved the theorem.

Now let $\mathfrak{A}$ be a (possibly infinite-dimensional) noncommutative Jordan algebra. The following identities hold [1, pp. 573–575]:

$$[L_x, R_z] = [R_x, L_z] \tag{2.5}$$

$$R_{yx} + L_xL_y = L_{xy} + R_xR_y \tag{2.6}$$

$$R_xL_y(R_x + L_x) + L_{yx^2} = R_{x^2}L_y + L_{yx}(R_x + L_x) . \tag{2.7}$$

We define $a \in \mathfrak{A}$ to be *regular* with *inverse* b if $ab = ba = c, a^2b = ba^2 = a$ (c the indentity of $\mathfrak{A}$).

THEOREM 2.2. *If $\mathfrak{A}$ is a noncommutative Jordan algebra then $a \in \mathfrak{A}$ has an inverse b if and only if b is the inverse of a in the commutative Jordan algebra $\mathfrak{A}^+$. Then $L_b = U_a^{-1}R_a$, $R_b = U_a^{-1}L_a$ where $U_a = R_a(R_a + L_a) - R_{a^2} = L_a(L_a + R_a) - L_{a^2} = U_a^+$. All $L_{a^k}, R_{a^k}, L_{b^j}, R_{b^j}$ commute, and $\Phi(a, b)$ is a commutative associative algebra.*

Proof. Clearly if $ab = ba = c, a^2b = ba^2 = a$ then a has inverse b in $\mathfrak{A}^+$. Conversely, suppose a has b as inverse in $\mathfrak{A}^+$; then $ab + ba =$

$2c$, $a^2b + ba^2 = 2a$. From (2.6)

$$\begin{aligned}(L_{a^2} - R_{a^2})b &= (L_a^2 - R_a^2)b \\ &= (L_a - R_a)(L_a + R_a)b = 2(L_a - R_a)c = 0\end{aligned}$$

so $a^2b = ba^2 = a$. Then $ba = b(a^2b) = (ba^2)b = ab$, so $ab = ba = c$, and b is the inverse of a in $\mathfrak{A}$.

By (2.5), $[L_a, R_b] = [R_a, L_b]$; by (2.6), $ab = ba = c$ implies $R_aR_b = L_aL_b$, $R_bR_a = L_bL_a$, so $[R_a, R_b] = [L_a, L_b]$. Then

$$[R_a + L_a, L_b] = \tfrac{1}{2}[R_a + L_a, R_b + L_b] = 2[L_a^+, L_b^+] = 0$$

from Theorem 2.1. Hence putting $x = a$, $y = b$ in (2.7) yields

$$R_a = \{R_a(R_a + L_a) - R_{a^2}\}L_b = U_aL_b\ .$$

Now (2.6) implies $R_a^2 - R_{a^2} = L_a^2 - L_{a^2}$, so averaging gives

$$\begin{aligned}U_a &= R_a(R_a + L_a) - R_{a^2} \\ &= L_a(L_a + R_a) - L_{a^2} = 2L_a^{+2} - L_{a^2}^+ = U_a^+\ .\end{aligned}$$

By Theorem 2.1, U_a^+ is invertible, so $L_b = U_a^{-1}R_a$. $R_b = U_a^{-1}L_a$ follows by duality. It is standard that R_x, L_x, R_{x^2} L_{x^2} generate all R_{x^k}, L_{x^k} (by induction using (2.7), its dual, and power-associativity) so the rest of the proof proceeds as in Theorem 2.1.

COROLLARY. *An algebraic element a in a noncommutative Jordan algebra has an inverse if and only if it has an inverse in the associative algebra $\Phi[a]$.*

Notice that formulas (1.9) to (1.13) of Lemma 1.3 show again that in a normed algebra $x\# = x^{-1}$, $H_x = U_x^{-1}$, $L_{x^{-1}} = U_x^{-1}R_x$, $R_{x^{-1}} = U_x^{-1}L_x$, and U_x, L_x, R_x, $L_{x^{-1}}$, $R_{x^{-1}}$ all commute.

2. Standard properties of the generic norm. In this section again all algebras are finite-dimensional, but Φ is not assumed to be infinite. We will give alternate proofs of the results of N. Jacobson [3] concerning the generic norm. After first proving a general lemma about polynomials admitting some kind of composition, an application of a technique involving the theory of inverses and the Hilbert Nullstellensatz (due to N. Jacobson [3, p. 37]) yields the desired results in a direct fashion.

LEMMA 2.1. *If $Q \in \Phi[\eta_1, \cdots, \eta_m]$ is a polynomial and $M_1, \cdots, M_m$ are rational expressions in $\Phi(\xi_1, \cdots, \xi_n, \eta_1, \cdots, \eta_m)$ which are linear in the indeterminates η_j and such that*

(a) *$Q(M_1(\xi, \eta), \cdots, M_m(\xi, \eta)) = m(\xi)Q(\eta)$ for some rational expres-*

sion $m(\xi_1, \cdots, \xi_n) \in \Phi(\xi_1, \cdots, \xi_n)$

(b) $M_i(c, \eta) = \eta_i$ *for some* $c = (c_1, \cdots, c_n)$ *where* $c_i \in \Phi$ *then each irreducible factor* Q_i *of* Q *admits the composition*

$$Q_i(M_1(\xi, \eta), \cdots, M_m(\xi, \eta)) = m_i(\xi)Q_i(\eta)$$

for some rational expression m_i.

Proof. For convenience let $M(\xi, \eta)$ denote $(M_1(\xi, \eta), \cdots, M_m(\xi, \eta))$. Let $Q = \prod Q_i^{q_i}$ be the factorization of Q into distinct irreducible factors. From (a) we have

$$\prod Q_i(M(\xi, \eta))^{q_i} = Q(M(\xi, \eta)) = m(\xi)Q(\eta) = m(\xi) \prod Q_i(\eta)^{q_i} .$$

Since each Q_i divides the right side and is irreducible as a polynomial over $\Phi(\xi_1, \cdots, \xi_n)$ it must divide some $Q_j(M(\xi, \eta))$, say $Q_j(M(\xi, \eta)) = m_i(\xi, \eta)Q_i(\xi)$ where m_i is rational in ξ and a polynomial in η. From (b), specializing $\xi \to c$ gives $Q_j(\eta) = m_i(c, \eta)Q_i(\eta)$. Since the Q_k are distinct and irreducible, $i = j$; then $Q_i(M(\xi, \eta)) = m_i(\xi, \eta)Q_i(\eta)$ and the assumed linearity of the $M_k(\xi, \eta)$ in η imply m_i is of degree 0 in η, $m_i(\xi, \eta) = m_i(\xi)$. Hence $Q_i(M(\xi, \eta)) = m_i(\xi)Q_i(\eta)$ as desired.

As a corollary, if Q is homogeneous we may take $n = 1$, $M(\lambda, \eta) = \lambda\eta$; then from

$$Q(M(\lambda, \eta)) = Q(\lambda\eta) = \lambda^q Q(\eta)$$

we conclude by the lemma that each irreducible factor Q_i satisfies $Q_i(\lambda\eta) = m_i(\lambda)Q_i(\eta)$. But $\prod m_i(\lambda) = \lambda^q$, so $m_j(\lambda) = \lambda^{q_i}$ and we see that each irreducible factor of a form is again a form. Of course, this is obvious anyway.

If $\mathfrak{A}$ is a strictly power-associative algebra over Φ (ie. remains power-associative under any extension of the base field) the *generic minimum polynomial* $m_\xi(\lambda)$ of $\mathfrak{A}$ is the mimimum polynomial of the generic element $\xi = \xi_1 x_1 + \cdots + \xi_n x_n$ of $\mathfrak{A}_\Sigma$, $\Sigma = \Phi(\xi_1, \cdots, \xi_n)$, relative to some basis $\{x_1, \cdots, x_n\}$ for $\mathfrak{A}$ over Φ. Specializing $\xi \to x$ gives a polynomial $m_x(\lambda)$ which has the same irreducible factors as the minimum polynomial $\mu_x(\lambda)$ of x in $\mathfrak{A}$ (this and the following assertions are found in [3, pp. 27–28]). The *generic norm* is the constant term $N(\xi_1, \cdots, \xi_n)$ of $m_\xi(\lambda)$; it is a homogeneous polynomial in the ξ_i. As a polynomial function the generic norm is independent of the basis for $\mathfrak{A}$, and the generic norm of an extension algebra $\mathfrak{A}_\Omega$ is just the natural extension of N to $\mathfrak{A}_\Omega$. Note, however, that N is given initially by a polynomial and a choice of basis rather than by a polynomial function (compare with Theorems 1.3, 1.4).

THEOREM 2.3. *If $\mathfrak{A}$ is a commutative Jordan algebra, the generic norm $N(x)$ has the same irreducible factors as $K(x) = \det U_x$.*

Proof. It suffices to prove everything over the algebraic closure Ω of Φ: the K, N of $\mathfrak{A}_\Omega$ are just the natural extensions of those of $\mathfrak{A}$, so if the former have the same irreducible factors over Ω the latter must have the same irreducible factors over Φ. By Theorem 2.1 and the Corollary to Theorem 2.2 $K(x) = 0 \Leftrightarrow x$ has no inverse in $\Omega[x]$. Since $\Omega[x]$ is associative, x has no inverse $\Leftrightarrow$ the constant term of $\mu_x(\lambda)$ is zero. But $\mu_x(\lambda)$ and $m_x(\lambda)$ have the same irreducible factors, so their constant terms vanish on the same set. The constant term of $m_x(\lambda)$ is $N(x)$, so $K(x) = 0 \Leftrightarrow N(x) = 0$. Because Ω is algebraically closed we can apply the Hilbert Nullstellensatz to conclude that K and N have the same irreducible factors.

THEOREM 2.4. *If $\mathfrak{A}$ is a commutative Jordan algebra, all normalized irreducible factors M of the generic norm admit Jordan composition $M(U_x y) = M(x)^2 M(y)$.*

Proof. By the fundamental formula (2.1), $U_{U(x)y} = U_x U_y U_x$, so taking determinants gives $K(U_x y) = K(x)^2 K(y)$. Since this holds for an infinite extension of Φ, it becomes an identity between polynomials relative to a choice of basis for $\mathfrak{A}$, and since $U_e = I$ we can apply Lemma 2.1 and Theorem 2.3 to conclude that each irreducible factor M of K or N admits composition $M(U_x y) = m(x)M(y)$. From $M(e) = 1$ we can conclude $M(U_x y) = M(x)^2 M(y)$ as in the proof of Theorem 1.2.

THEOREM 2.5. *Let $\mathfrak{A}$ be a strictly power-associative algebra, $\mathfrak{B}$ a subalgebra. Then the restriction $N|_{\mathfrak{B}}$ of the generic norm N of $\mathfrak{A}$ to $\mathfrak{B}$ is a polynomial on $\mathfrak{B}$ having the same irreducible factors as the generic norm $N_{\mathfrak{B}}$ of $\mathfrak{B}$. If M is any normalized factor of N then $M(xy) = M(x)M(y)$ if x, y are contained in an associative subalgebra.*

Proof. For the first assertion, if $x \in \mathfrak{B}$ then

$$N|_{\mathfrak{B}}(x) = 0 \Longleftrightarrow N(x) = 0 \Longleftrightarrow x \text{ has no inverse in } \Phi[x];$$

since $\Phi[x] \subset \mathfrak{B} \subset \mathfrak{A}$ the same reasoning shows $N_{\mathfrak{B}}(x) = 0 \Leftrightarrow x$ has no inverse in $\Phi[x]$. Thus $N|_{\mathfrak{B}}(x) = 0 \Leftrightarrow N_{\mathfrak{B}}(x) = 0$; this remains valid on extension of Φ to its algebraic closure, so as in Theorem 2.3 we conclude $N|_{\mathfrak{B}}, N_{\mathfrak{B}}$ have the same irreducible factors.

Next, let M be a normalized irreducible factor of N and assume $\mathfrak{B}$ is an associative subalgebra of $\mathfrak{A}$. If $D(x)$ is the determinant of

left multiplication by x on $\mathfrak{B}$ then the associativity $L_{xy} = L_x L_y$ of $\mathfrak{B}$ gives $D(xy) = D(x)D(y)$. Since this remains valid on an infinite extension of Φ, relative to a basis for $\mathfrak{B}$ it becomes an identity between polynomials. Applying Lemma 2.1 we see that every irreducible factor of D admits composition. But by the same argument as before, $N|_{\mathfrak{B}}$ has the same irreducible factors as D. Hence $M|_{\mathfrak{B}}$, as a factor of $N|_{\mathfrak{B}}$, admits composition $M|_{\mathfrak{B}}(xy) = m(x)M|_{\mathfrak{B}}(y)$; taking $y = c$, the assumption that M is normalized gives $m(x) = M|_{\mathfrak{B}}(x)$. Hence if x, y are in an associative subalgebra $\mathfrak{B}$ we have

$$M(xy) = M|_{\mathfrak{B}}(xy) = M|_{\mathfrak{B}}(x)M|_{\mathfrak{B}}(y) = M(x)M(y) .$$

COROLLARY. *If $\mathfrak{A}$ is an alternative algebra, all normalized irreducible factors M of the generic norm admit associative composition $M(xy) = M(x)M(y)$.*

THEOREM 2.6. *Let $\mathfrak{A}$ be an alternative or Jordan algebra. If $\{x_1, \cdots, x_n\}$ is a basis and $Q \in \Phi[\xi_1, \cdots, \xi_n]$ a homogeneous polynomial of degree q such that $Q(ab) = Q(a)Q(b)$ or $Q(U_a b) = Q(a)^2 Q(b)$ respectively for all $a, b \in \mathfrak{A}$, and if $|\Phi| > q$ or $|\Phi| > 2q$ then in either case Q is a product of irreducible factors of the generic norm N.*

Proof. The assumptions on $|\Phi|$ guarantee that we can pass to the algebraic closure of Φ and prove the result there. We saw $Q(c) \neq 0$ in § 2 of Chapter I (c the identity of $\mathfrak{A}$). If $N(x) \neq 0$ then x has an inverse y in $\Phi[x]$, $xy = U_x y = c$. Thus $Q(c)$ equals $Q(x)Q(y)$ or $Q(x)^2 Q(y)$, and in either case $Q(x)$ cannot be 0. Thus $Q(x) = 0 \Rightarrow N(x) = 0$, and the Hilbert Nullstellensatz yields the result.

THEOREM 2.7. *The generic norm of a simple alternative or Jordan algebra is irreducible.*

Proof. We refer to [3, pp. 33 and 39].

THEOREM 2.8. *The generic norm of an alternative or Jordan algebra $\mathfrak{A}$ with identity c has nondegenerate trace form τ_c if and only if $\mathfrak{A}$ is separable.*

Proof. By the Corollary to Theorem 2.5 or by Theorem 2.4 N admits R_x, L_x or U_x on $\mathfrak{A}_\Omega$, Ω an infinite extension of Φ. Thus if τ_c is nondegenerate $\mathfrak{A}_\Omega$ is normed, so is separable by Theorem 1.1, and hence $\mathfrak{A}$ is too. Conversely, assume $\mathfrak{A}$ is separable; it will suffice to prove N_Ω is nondegenerate for Ω the algebraic closure of Φ. Then $\mathfrak{A}_\Omega = \bigoplus \mathfrak{A}_i$ for $\mathfrak{A}_i$ simple alternative or Jordan algebras respectively, and $N_\Omega = \prod N_i$, $\tau = \bigoplus \tau_i$, so it will suffice to prove N_i is nondegen-

erate. We are thus reduced to considering the case of a simple algebra $\mathfrak{A}$ over an algebraically closed field Ω. By Lemma 1.2 τ is a well-defined associative biliner form, hence by simplicity it is either nondegenerate or the zero form. Assume $\tau = 0$; from (1.4) with $y = c$, $E_x = L_x$ or U_x we get $\tau_x(L_x u, L_x v) = \tau(u, v) = 0$ or $\tau_{x^2}(U_x u, U_x v) = \tau(u, v) = 0$ respectively. Then $\tau_x = 0$ or $\tau_{x^2} = 0$ on the sets where L_x or U_x are nonsingular. Since Ω is infinite, $\tau_z = 0$ for z in a dense set in either case, so $\tau_x = 0$ for all x; in other words $\partial_u \partial_v \log N = 0$ for all u, v. Thus $\partial_v \log N \in \Omega(\xi_1^p, \cdots, \xi_n^p)$; but it is homogeneous of degree -1, so it is zero, and $\partial_v N = N \partial_v \log N = 0$ for all v. Then $N \in \Omega[\xi_1^p, \cdots, \xi_n^p]$; since Ω is perfect, N is a pth power, which contradicts Theorem 2.7. Thus $\tau \neq 0$, so it must be nondegenerate.

4. Applications to normed algebras. In this section we use the foregoing results to characterize all possible norms on a normed algebra. Throughout the section we will assume that Φ is infinite.

LEMMA 2.2. *If $Q \in \Phi[\eta_1, \cdots, \eta_m]$ is a polynomial and $M_1, \cdots, M_m$ are rational expressions in $\Phi(\xi_1, \cdots, \xi_n, \eta_1, \cdots, \eta_m)$ which are linear in the indeterminates η_k and such that*

(a) *$\partial_k \log Q(M_1(\xi, \eta), \cdots, M_m(\xi, \eta)) = \partial_k \log Q(\eta)$ for all k where ∂_k is formal partial derivation with respect to η_k*

(b) *$M_i(c, \eta) = \eta_i$ for some $c = (c_1, \cdots, c_n)$ where $c_i \in \Phi$*

then $Q = Q'Q''$ where each irreducible factor Q_i of Q' admits the composition $Q_i(M_1(\xi, \eta), \cdots, M_m(\xi, \eta)) = m_i((\xi Q_i(\eta)$ for some rational expression m_i and where $Q'' \in \Phi[\eta_1^p, \cdots, \eta_m^p]$.

Proof. For convenience we set $M(\xi, \eta) = (M_1(\xi, \eta), \cdots, M_m(\xi, \eta))$. Let $Q = \prod Q_i^{q_i}$ be the factorization of Q as in Lemma 2.1 Suppose $Q_1, \cdots, Q_j$ are the Q_i which admit composition. Write $Q = Q'Q''$ for $Q' = \prod_{i \leq j} Q_i^{q_i}$. Then $Q'(M(\xi, \eta)) = m'(\xi)Q'(\eta)$, so

$$\begin{aligned} \partial_k \log Q'(M(\xi, \eta)) &= \partial_k \log m'(\xi) + \partial_k \log Q'(\eta) \\ &= \partial_k \log Q'(\eta) . \end{aligned}$$

Since $\partial_k \log Q = \partial_k \log Q' + \partial_k \log Q''$, subtraction from (a) gives

$$\partial_k \log Q''(M(\xi, \eta)) = \partial_k \log Q''(\eta) .$$

Now

$$\begin{aligned} \partial_k \log Q'' &= \sum_{i>j} q_i \partial_k \log Q_i \\ &= \sum q_i Q_i^{-1} \partial_k Q_i = R^{-1} \sum q_i R_i \partial_k Q_i = R^{-1} S \end{aligned}$$

where $R = \prod_{i>j} Q_i$, $R_i = Q_i^{-1} R$. In reduced form the denominator of

$\partial_k \log Q''$ thus consists of some of the distinct irreducible factors Q_i of R with $i > j$, while the denominator of

$$\begin{aligned}\partial_k \log Q''(M(\xi, \eta)) &= R(M(\xi, \eta))^{-1} \sum q_i R_i(M(\xi, \eta)) \partial_k Q_i(M(\xi, \eta)) \\ &= R(M(\xi, \eta))^{-1} T(\xi, \eta)\end{aligned}$$

considered as a rational expression in η with coefficients in $\Phi(\xi_1, \cdots, \xi_n)$ is just $R(M(\xi, \eta)) = \prod_{i>j} Q_i(M(\xi, \eta))$. If Q_i is a factor of the former denominator it must be a factor of the latter; then as in Lemma 2.1 we see that Q_i admits composition. This contradicts our assumption that Q_i does not admit composition for $i > j$, so no such Q_i exists, and the denominator of $R^{-1}S$ in reduced form must be a constant. Then $\partial_k \log Q''$ is a polynomial; we noticed after Lemma 2.1 that each Q_i is homogeneous, so $\partial_k \log Q''$ is homogeneous of degree -1, and thus must be identically zero. Then $\partial_k Q'' = Q'' \partial_k \log Q'' = 0$ for all k, and $Q'' \in \Phi[\eta_1^p, \cdots, \eta_m^p]$.

THEOREM 2.9. *If Q is a norm on a normed algebra $\mathfrak{A}$ with identity c then Q is a product of irreducible factors of the generic norm N of $\mathfrak{A}$. The symmetrized algebra $\mathfrak{A}^+$ is the Koecher algebra $\mathfrak{A}(Q, c)$. In particular, any commutative normed algebra is the Koecher algebra of its norm.*

Proof. Q remains a norm on $\mathfrak{A}_\Omega$, Ω the algebraic closure of Φ, and N remains the generic norm, so it suffices to prove the result over Ω.

We have

$$\begin{aligned}\partial_u \partial_v \log \{Q \circ H_x\} \mid_y &= \partial_{H(x)u} \partial_{H(x)v} \log Q \mid_{H(x)y} \\ &= -\tau_{H(x)y}(H_x u, H_x v) \\ &= -\tau(H_x^* H_{H(x)y} H_x u, v) .\end{aligned}$$

Now $H_x^* = H_x$ and by averaging (1.13) $H_x^{-1} = 2L_x^{+2} - L_{x^2}^+ = U_x^+$, so from the fundamental formula (2.1) for the commutative Jordan algebra $\mathfrak{A}^+$

$$\begin{aligned}H_x H_{H(x)y} H_x &= (U_x U_{H(x)y} U_x)^{-1} \\ &= (U_{U(x)H(x)y})^{-1} = U_y^{-1} = H_y .\end{aligned}$$

Therefore

$$\begin{aligned}\partial_u \partial_v \log \{Q \circ H_y\} \mid_y &= -\tau(H_y u, v) \\ &= -\tau_y(u, v) = \partial_u \partial_v \log Q \mid_y ,\end{aligned}$$

and

$$\partial_u\{\partial_v \log \{Q \circ H_x\} - \partial_v \log Q\}\,|_y = 0\ .$$

The term in braces is homogeneous of degree -1, so taking $u = y$ gives $\partial_v \log \{Q \circ H_x\}\,|_y = \partial_v \log Q\,|_y$ by Euler's equations. Since Ω is infinite, the relation $\partial_v \log Q(H_x y) = \partial_v \log Q(y)$ becomes a relation between polynomials relative to a basis for $\mathfrak{A}$, and we can apply Lemma 2.2 to write $Q = Q'Q''$ where all the irreducible factors Q_i of Q' admit H_x as composition and where Q'' is a pth power (since Ω is perfect). If Q_i admits H_x it admits $U_x = H_x^{-1}$; if we normalize it as in Theorem 1.2 we can apply Theorem 2.6 to conclude Q_i is an irreducible factor of N. Let $Q = \prod Q_i^{q_i}$ be the factorization of Q. If p does not divide q_i then $Q_i^{q_i}$ is not a pth power, hence not all of it can appear in Q'', so Q_i must appear in Q' and hence be an irreducible factor of N. To complete the proof we need only show that for each j we can alter the exponents so that $\tilde{Q} = \prod Q_i^{\tilde{q}_i}$ is still a norm but p does not divide $\tilde{q}_j$. By Lemma 2.1, all Q_i admit E_x, F_x of (1.1), so any $\tilde{Q}$ does too; hence $\tilde{Q}$ will be a norm as soon as its trace form $\tilde{\tau}$ is nondegenerate.

If $\mathfrak{A}_\Omega = \bigoplus \mathfrak{A}_i$ is the decomposition of $\mathfrak{A}_\Omega$ into simple ideals guaranteed by Theorem 1.1 then $N = \prod N_i$, N_i the generic norm of $\mathfrak{A}_i$. By Theorem 2.7 the N_i are the irreducible factors of N over Ω, so if p does not divide q_i then Q_i is an N_j. If p divides q_i then $\partial Q_i^{q_i} = 0$ and hence Q_i does not contribute to the trace form τ of Q. Thus τ is a linear combination of the traces of the N_j. The latter are concentrated on the $\mathfrak{A}_j$'s, so by the nondegeneracy of τ all the N_j must appear among the Q_i with q_i not divisible by p. We may renumber so that $Q_i = N_i$ for $1 \leqq i \leqq m$. Thus the Q_j for $j > m$ are the ones p divides q_j. By Theorem 2.8 the trace form τ_i of $Q_i = N_i$ is nondegenerate on $\mathfrak{A}_i$. By Lemma 1.2, τ_j for $j > m$ is a (perhaps degenerate) associative bilinear form on $\mathfrak{A}_\Omega$, so $\tau_j = \bigoplus_{i \leqq m} \lambda_i \tau_i$ for some $\lambda_i \in \Omega$ (because any associative form on the simple algebra $\mathfrak{A}_i$ over the algebraically closed field Ω is a scalar multiple of the given nondegenerate form τ_i). If we pick integers $\bar{q}_i$ such that $\bar{q}_i 1 + \lambda_i \neq 0$ in Ω and set $\bar{Q} = \prod_{i \leqq m} Q_i^{\bar{q}_i}$ then $\tilde{Q} = Q_j \bar{Q}$ has trace form $\tilde{\tau} = \bigoplus_{i \leqq m} (\bar{q}_i + \lambda_i)\tau_i$ which is nondegenerate. The exponent of Q_j in $\tilde{Q}$ is 1, so $\tilde{Q}$ is the desired norm. (As a matter of fact, since this implies Q_j is an irreducible factor of N it is one of the N_i and hence there actually is no Q_j for $j > m$).

For the last statement of the theorem, observe that the Koecher algebra is defined because Q admits H_x. Multiplication in $\mathfrak{A}^+$ is $L_u^+ = \frac{1}{2}(L_u + R_u)$, so

$$\begin{aligned}\tau(L_u^+ v, w) &= \tfrac{1}{2}\tau(\{L_u + R_u\}v, w)\\ &= \tfrac{1}{2}\{\tau(L_u v, w) + \tau(v, L_u w)\}\\ &= \tfrac{1}{2}\partial_u\partial_v\partial_w \log Q\,|_e\end{aligned}$$

by (1.8) and (1.5). But this defines the multiplication in the Koecher algebra, so $\mathfrak{A}^+ = \mathfrak{A}(Q, c)$.

Note that the norms on $\mathfrak{A}$ are precisely all $Q = \prod N_i^{n_i}$ where no n_i is divisible by p (in particular no n_i is 0).

CHAPTER III

1. Isotopes. Throughout this chapter $\mathfrak{A}$ will denote a finite-dimensional commutative Jordan algebra (not necessarily semisimple and Φ not necessarily infinite). We first recall the standard results about isotopes [5], which are all consequences of MacDonald's Theorem. Let u be any element of $\mathfrak{A} = (\mathfrak{x}, \cdot)$. We define a new multiplication on $\mathfrak{x}$ by $x \cdot_u y = (x \cdot u) \cdot y + (u \cdot y) \cdot x - (x \cdot y) \cdot u$. The algebra $\mathfrak{A}^{(u)} = (\mathfrak{x}, \cdot_u)$ has an identity if and only if u is regular, in which case the identity is u^{-1}. In this case we say $\mathfrak{A}^{(u)}$ is the *u-isotope* of $\mathfrak{A}$; $\mathfrak{A}^{(u)}$ is again a Jordan algebra, and the operator $U_x^{(u)}$ is given by

$$U_x^{(u)} = U_x U_u \,. \tag{3.1}$$

The relation of isotopy is symmetric and transitive.

THEOREM 3.1. *Let Q be a form on a finite-dimensional vector space $\mathfrak{x}$ over an infinite field Φ and $c \in \mathfrak{x}$ a point where the Koecher algebra $\mathfrak{A} = \mathfrak{A}(Q, c)$ is defined: thus the trace form τ_c is nondegenerate and $Q(H_x y) = h(x)Q(y)$. If $\tilde{c} \in \mathfrak{x}$ is any point where $\tau_{\tilde{c}}$ is nondegenerate then the Koecher algebra $\tilde{\mathfrak{A}} = \mathfrak{A}(Q, \tilde{c})$ is defined and is the u-isotope of $\mathfrak{A}$ for $u = \tilde{c}^{-1}$. Thus the Koecher algebras corresponding to different basepoints are isotopic.*

Proof. Let $\tau = \tau_c$, $\tilde{\tau} = \tau_{\tilde{c}}$ where $\tau_x(u, v) = -\partial_u \partial_v \log Q|_x$ as usual; then $\tau(H_y u, v) = \tau_y(u, v) = \tilde{\tau}(\tilde{H}_y u, v) = \tau_{\tilde{c}}(\tilde{H}_y u, v) = \tau(H_{\tilde{c}} \tilde{H}_y u, v)$ implies $\tilde{H}_y = H_{\tilde{c}}^{-1} H_y$ by the assumed nondegeneracy of $\tau, \tilde{\tau}$. Then $Q(\tilde{H}_x y) = Q(H_{\tilde{c}}^{-1} H_x y) = h(\tilde{c})^{-1} h(x) Q(y) = \tilde{h}(x) Q(y)$ and $\tilde{\mathfrak{A}}$ is defined. From (1.13) we see $\tilde{U}_y = \tilde{H}_y^{-1} = H_y^{-1} H_{\tilde{c}} = U_y U_{\tilde{c}}^{-1}$. Since $\tilde{\tau}$ is defined we must have $Q(\tilde{c}) \neq 0$, so we know $u = \tilde{c}^{-1}$ exists and equals $\tilde{c}^{\#}$ by (1.11), (1.12). Thus $\mathfrak{A}^{(u)}$ is a Jordan algebra with identity $u^{-1} = \tilde{c}$ and $U_y^{(u)} = U_y U_u = U_y U_{\tilde{c}^{-1}} = U_y U_{\tilde{c}}^{-1}$ using (3.1) and Theorem 2.1. But $\tilde{\mathfrak{A}}$ has the same identity and $\tilde{U}_y = U_y U_{\tilde{c}}^{-1} = U_y^{(u)}$, so $\tilde{\mathfrak{A}} = \mathfrak{A}^{(u)}$.

COROLLARY. *If $\mathfrak{A}, \tilde{\mathfrak{A}}$ are commutative normed algebras on $\mathfrak{x}, \tilde{\mathfrak{x}}$ over an infinite field Φ with norms $Q, \tilde{Q}$ respectively, and if $W: \mathfrak{x} \to \tilde{\mathfrak{x}}$ is a one-to-one linear transformation such that $\tilde{Q}(Wx) = Q(x)$, then W is an isomorphism of $\mathfrak{A}$ with an isotope of $\tilde{\mathfrak{A}}$.*

Proof. W carries $\mathfrak{A}$ in the natural way isomorphically onto an algebra $\bar{\mathfrak{A}}$ on $\bar{\mathfrak{X}}$ with identity $\bar{c} = Wc$ and norm $\bar{Q} = Q \circ W^{-1}$. By Theorem 2.9, $\bar{\mathfrak{A}}$ is the Koecher algebra $\mathfrak{A}(\bar{Q}, \bar{c})$ of $\bar{Q}$ with basepoint $\bar{c}$, and $\tilde{\mathfrak{A}}$ is the Koecher algebra $\mathfrak{A}(\tilde{Q}, \tilde{c})$ where $\tilde{c}$ is the identity of $\tilde{\mathfrak{A}}$. But $\bar{Q}(Wx) = Q(x) = \tilde{Q}(Wx)$ and Φ is infinite, so $\bar{Q} = \tilde{Q}$; by Theorem 3.1 $\bar{\mathfrak{A}}$ is the $\bar{c}^{-1}$ isotope of $\tilde{\mathfrak{A}}$. Hence W is an isomorphism of $\mathfrak{A}$ with an isotope of $\tilde{\mathfrak{A}}$.

COROLLARY. *If $\mathfrak{A}$ is a separable Jordan algebra with generic norm N, $|\Phi| > \deg N$, then a one-to-one linear transformation W of $\mathfrak{A}$ into itself is an automorphism of $\mathfrak{A}$ if and only if $Wc = c$ and $N(Wa) = N(a)$ for all $a \in \mathfrak{A}$.*

Proof. Clearly any automorphism satisfies the conditions. Conversely, any such W satisfies $N(Wx) = N(x)$ on an infinite extension by our assumption on $|\Phi|$; if we take $\mathfrak{A} = \tilde{\mathfrak{A}}$, $Q = \tilde{Q} = N$ the proof of the preceding corollary and the fact that $(Wc)^{-1} = c^{-1} = c$ show that W is an isomorphism of $\mathfrak{A}$ with $\mathfrak{A}^{(c)} = \mathfrak{A}$.

Note that $Q(Wa) = Q(a)$ need not hold for an automorphism W if Q is an arbitrary norm on the separable algebra $\mathfrak{A}$.

2. **The group $\mathscr{G}(\mathfrak{A})$.** We here generalize some results of Koecher [9, pp. 70–73]; remember that $\mathfrak{A}$ need not be semisimple and Φ need not be infinite. Let $\mathscr{G}(\mathfrak{A})$ be the set of all nonsingular linear transformations W on $\mathfrak{A}$ such that

$$U_{Wx} = WU_xW^*$$

for some linear transformation W^* and all x, and let $\mathscr{U}(\mathfrak{A})$ be the group generated by the U_x for x regular (see Theorem 2.1). One easily verifies that $\mathscr{G}(\mathfrak{A})$ is a group, $W^* \in \mathscr{G}(\mathfrak{A})$ for all $W \in \mathscr{G}(\mathfrak{A})$, and (by (2.1)) $\mathscr{U}(\mathfrak{A})$ is a subgroup; since

$$\begin{aligned} WU_xW^{-1} &= U_{Wx}W^{*-1}W^{-1} \\ &= U_{Wx}(WU_cW^*)^{-1} = U_{Wx}U_{Wc}^{-1} \in \mathscr{U}(\mathfrak{A}) \end{aligned}$$

it is a normal subgroup. If $W \in \mathscr{G}(\mathfrak{A})$ then

$$U_{U(Wx)y} = U_{WU(x)W^*y} = WU_xW^*U_yW^{**}U_xW^* ;$$

but from (2.1) $U_{U(Wx)y} = U_{Wx}U_yU_{Wx} = WU_xW^*U_yWU_xW^*$, so equating gives $W^{**} = W$. $(W_1W_2)^* = W_2^*W_1^*$ is clear, so * is an involution on $\mathscr{G}(\mathfrak{A})$ which leaves the elements U_x fixed. Geometrically, $W^* = \sharp \circ W^{-1} \circ \sharp$ where $\sharp$ is the mapping $x \longrightarrow x^{-1} = U_x^{-1}x$:

$$\begin{aligned} W^*x &= W^*U_{x^{-1}}^{-1}x^{-1} \\ &= W^*(WU_{W^{-1}x^{-1}}W^*)^{-1}x^{-1} = U_{W^{-1}x^{-1}}^{-1}W^{-1}x^{-1} \\ &= (W^{-1}x^{-1})^{-1} = \{\# \circ W^{-1} \circ \#\}x \ . \end{aligned}$$

Thus $W^*v = u \Leftrightarrow Wu^{-1} = v^{-1}$.

THEOREM 3.2. *Two isotope $\mathfrak{A}^{(u)}, \mathfrak{A}^{(v)}$ are isomorphic under a linear transformation W if and only if $W \in \mathscr{G}(\mathfrak{A})$ and $W^*v = u$. The group of automorphisms of $\mathfrak{A}$ is the subgroup of $\mathscr{G}(\mathfrak{A})$ fixing the identity c.*

Proof. Suppose $W: \mathfrak{A}^{(u)} \to \mathfrak{A}^{(v)}$ is an isomorphism. Then W takes the identity u^{-1} of $\mathfrak{A}^{(u)}$ onto the identity v^{-1} of $\mathfrak{A}^{(v)}$, so $Wu^{-1} = v^{-1} \Rightarrow W^*v = u$, and by the multiplicative property

$$WL_x^{(u)} = L_{Wx}^{(v)}W \Rightarrow WU_x^{(u)} = U_{Wx}^{(v)}W \Rightarrow WU_xU_u = U_{Wx}U_vW$$

(by (3.1)) $\Rightarrow U_{Wx} = WU_xW^*$ for $W^* = U_uW^{-1}U_v^{-1}$ and all x, so $W \in \mathscr{G}(\mathfrak{A})$.

Conversely, if $W \in \mathscr{G}(A)$ and $W^*v = u$ then for all x $u = W^*v = (U_x^{-1}W^{-1}U_{Wx})v$, so $Wx \cdot_v Wx = U_{Wx}v = WU_xu = W(x \cdot_u x)$; by commutativity this linearizes to $W(x \cdot_u y) = Wx \cdot_v Wy$, and W is an isomorphism.

Applying this to the case $u = v = c$ and noting that $W^*c = c \Leftrightarrow Wc = c$ since $c^{-1} = c$, we see W is an automorphism of $\mathfrak{A} = \mathfrak{A}^{(c)}$ if and only if $W \in \mathscr{G}(\mathfrak{A})$ and $Wc = c$.

Isotopes are not always isomorphic; for example, the algebras $\mathfrak{H}(\mathfrak{D}_n, \gamma)$ for different γ are isotopic but not necessarily isomorphic. However, if Φ is algebraically closed then there is essentially only one $\mathfrak{H}(\mathfrak{D}_n, \gamma)$. This is a special case of the following general result.

THEOREM 3.3. *If Φ is algebraically closed, all isotopes $\mathfrak{A}^{(u)}$ are isomorphic.*

Proof. Since Φ is algebraically closed, every regular element has a regular square root [3, p. 43]. If $\mathfrak{A}^{(u)}$ is an isotope, u is regular by definition, so there is a regular $v \in \mathfrak{A}$ with $v^2 = u$. Then $U_vc = v^2 = u$, and $U_v = U_v^* \in \mathscr{U}(\mathfrak{A}) \subset \mathscr{G}(\mathfrak{A})$ since v is regular, so by Theorem 3.2 U_v is an isomorphism of $\mathfrak{A}^{(u)}$ onto $\mathfrak{A}^{(c)} = \mathfrak{A}$.

This can also be seen without recourse to Theorem 3.2. Let $v^2 = u$ as above; by bilinearity and commutativity it will suffice to demonstrate $U_v(x \cdot_u x) = (U_vx) \cdot (U_vx)$ in order to prove U_v is an isomorphism of $\mathfrak{A}^{(u)}$ onto $\mathfrak{A}$. But $x \cdot_u x = 2x \cdot (x \cdot u) - x^2 \cdot u = U_xu$, so $U_v(x \cdot_u x) = U_vU_xu = U_vU_xv^2 = U_vU_xU_vc = U_{U(v)x}c = (U_vx)^2$ by (2.1).

As a corollary of this result we get a quick proof of the following proposition [3, p. 43].

COROLLARY. *The generic norm of an isotope* $\mathfrak{A}^{(u)}$ *is* $N^{(u)}(x) = N(u)N(x)$.

Proof. It suffices to pass to the algebraic closure. There we can find a square root v of u, and by the above U_v is an isomorphism of $\mathfrak{A}^{(u)}$ onto $\mathfrak{A}$. Under an isomorphism corresponding elements clearly have the same generic norm, so

$$N^{(u)}(x) = N(U_v x) = N(v)^2 N(x) = N(v^2)N(x) = N(u)N(x) .$$

COROLLARY. *If* Φ *is algebraically closed, every* $W \in \mathscr{G}(\mathfrak{A})$ *can be written* $W = UV$ *where* $U \in \mathscr{U}(\mathfrak{A})$ *and* V *is an automorphism.*

Proof. Since $U_{Wc} = WU_c W^*$ is nonsingular, Wc is regular, and hence as above has a regular square root v. Then $U = U_v \in \mathscr{U}(\mathfrak{A})$, and $V = U^{-1}W$ has $V \in \mathscr{G}(\mathfrak{A})$ and $Vc = U^{-1}Wc = U_v^{-1}v^2 = c$, so V is an automorphism by Theorem 3.2.

Suppose now that $\mathfrak{A}$ is separable over an infinite field Φ with generic norm N. We claim that $\mathscr{G}(\mathfrak{A})$ is just the group of normpreserving transformations W such that $N(Wx) = \omega N(x)$ identically with $\omega \in \Phi$. If $N(Wx) = \omega N(x)$ we can apply the differential calculus as in (1.3) to get $\tau_{Wx}(Wu, Wv) = \tau_x(u, v)$, hence $W^* H_{Wx} W = H_x$, $H_{Wx} = W^{*-1} H_x W^{-1}$, $U_{Wx} = WU_x W^*$ and $W \in \mathscr{G}(\mathfrak{A})$; note that * here is the adjoint relative to the nondegenerate form τ. Conversely, if $W \in \mathscr{G}(\mathfrak{A})$ then by Theorem 3.2 W is an isomorphism of $\mathfrak{A}^{(u)}$ onto $\mathfrak{A}$ for $u = W^*c$. Then $N(Wx) = N^{(u)}(x) = N(u)N(x)$ by the first Corollary, and W is normpreserving.

The group $\mathscr{G}(\mathfrak{A})$ has been computed by N. Jacobson [6, I, II, III] for central simple $\mathfrak{A}$; in each case it is easy to compute W^*.

3. The Lie algebra of $\mathscr{G}(\mathfrak{A})$. As a final example of the usefulness of the differential calculus we give short proofs of some results of N. Jacobson [3, pp. 42, 47, 48] on the Lie algebra of the algebraic group $\mathscr{G}(\mathfrak{A})$. We will assume throughout the section that Φ is infinite. We say a polynomial Q has a linear transformation W as *Lie invariant* if

$$\partial_{Wx} Q \,|_x = 0$$

identically. Such W form a Lie algebra $\mathscr{L}(\mathfrak{A}, Q)$ of linear transformations. Indeed, applying $\partial_y \,|_x$ to the defining relation gives

$$\partial_{W_y}Q\,|_x + \partial_y\partial_{W_x}Q\,|_x = 0\ ;$$

if $W, V \in \mathscr{L}(\mathfrak{A}, Q)$ then

$$\partial_{WV_x}Q\,|_x = -\partial_{V_x}\partial_{W_x}Q\,|_x = -\partial_{W_x}\partial_{V_x}Q\,| = \partial_{VW_x}Q\,|_x\ ,$$

and $\partial_{[W,V]_x}Q\,|_x = 0$ implies $[W, V] \in \mathscr{L}(\mathfrak{A}, Q)$.

THEOREM 3.4. *If N is the generic norm of a Jordan algebra $\mathfrak{A}$ then $L_a \in \mathscr{L}(\mathfrak{A}, N)$ if and only if a has trace zero, ie. $\tau(c, a) = 0$. If $\mathfrak{A}$ is separable then*

$$\mathscr{L}(\mathfrak{A}, N) = \mathscr{L}(\mathfrak{A}') \oplus \mathscr{D}(\mathfrak{A})$$

where $\mathscr{L}(\mathfrak{A}')$ is the space of multiplications L_a by elements of trace zero and $\mathscr{D}(\mathfrak{A})$ is the Lie algebra of derivations.

Proof. For the first assertion it will suffice to prove $\partial_{L(a)y}N\,|_y = N(y)\tau(c, a)$. We apply $\partial_a\,|_c$ to $N(U_xy) = N(x)^2N(y)$ as a function of x. Using the chain rule the left side becomes

$$\partial_a N(U_xy)\,|_c = \partial_b N\,|_{U(c)y} = \partial_b N\,|_y$$

for $b = \partial_a\{U_xy\}\,|_c = 2L_ay$. Using $N(c) = 1$ and the Euler equations the right side becomes

$$\begin{aligned} 2N(c)\partial_a N\,|_c\, N(y) &= 2N(y)N(c)^{-1}\partial_a N\,|_c = 2N(y)\partial_a \log N\,|_c \\ &= -2N(y)\partial_c\partial_a \log N\,|_c = 2N(y)\tau(c, a)\ . \end{aligned}$$

Equating gives $2\partial_{L(a)y}N\,|_y = 2N(y)\tau(c, a)$.

For the second assertion, let $W \in \mathscr{L}(\mathfrak{A}, N)$, $Wc = a$. Since $\mathfrak{A}$ is now assumed separable, τ is nondegenerate by Theorem 2.8. Hence we can define $\#$ as in § 4 of Chapter I, and

$$0 = N(x)^{-1}\partial_{W_x}N\,|_x = \partial_{W_x}\log N\,|_x = \tau(x\#, Wx)\ .$$

Putting $x = c$ we get $\tau(c, a) = 0$, so $L_a \in \mathscr{L}(\mathfrak{A}') \subset \mathscr{L}(\mathfrak{A}, N)$. Therefore $D = W - L_a \in \mathscr{L}(\mathfrak{A}, N)$, $Dc = 0$. We will prove D is a derivation. As above $0 = \tau(x\#, Dx)$, so $0 = \partial_y\{\tau(x\#, Dx)\}\,|_x = \tau(x\#, Dy) - \tau(H_xy, Dx)$. By nondegeneracy $D^*x\# = H_xDx$, and $U_xD^*x\# = Dx$ by (1.13). From $U_{x^2} = U_x^2$ (by (2.1)) and $(x^2)\# = x^{-2} = U_x^{-1}c = H_xc$ (by (1.11)-(1.13)) we get

$$\begin{aligned} D(x\cdot x) - 2x\cdot Dx &= U_{x^2}D^*(x^2)\# - 2L_xU_xD^*x\# \\ &= U_x\{U_xD^*H_xc - 2L_xD^*x\#\} \\ &= -U_x\{\partial_c\{U_xD^*x\#\}\,|_x\} \\ &= -U_x\{\partial_c Dx\,|_x\} = -U_xDc = 0\ ; \end{aligned}$$

linearizing shows D is a derivation. Hence

$$W = L_a + D \in \mathscr{L}(\mathfrak{A}') + \mathscr{D}(\mathfrak{A}) ,$$

and we have shown $\mathscr{L}(\mathfrak{A}, N) \subset \mathscr{L}(\mathfrak{A}') + \mathscr{D}(\mathfrak{A})$; but $\mathscr{L}(\mathfrak{A}') \subset \mathscr{L}(\mathfrak{A}, N)$, and $\mathscr{D}(\mathfrak{A}) \subset \mathscr{L}(\mathfrak{A}, N)$ is known [14], so $\mathscr{L}(\mathfrak{A}, N) = L(\mathfrak{A}') + D(\mathfrak{A})$. Finally, $\mathscr{L}(\mathfrak{A}') \cap \mathscr{D}(\mathfrak{A}) = 0$ since $L_a c = a$, $Dc = 0$, so the sum is direct.

References

1. A. A. Albert, *Power-associative rings*, Trans. Amer. Math. Soc. **64** (1948), 552-593.
2. C. Chevalley, *Théorie des Groupes de Lie*, Tome II, Hermann, Paris, 1951.
3. N. Jacobson, *Generic norm of an algebra*, Osaka Math. J. **15** (1963), 25-50.
4. ———, *A theorem on the structure of Jordan algebras*, Proc. Nat. Acad. Sci. U.S.A. **42** (1956), 140-147.
5. ———, *A coordinatization theorem for Jordan algebras*, Proc. Nat. Acad. Sci. U.S.A. **48** (1962), 1154-1160.
6. ———, *Some groups of linear transformations defined by Jordan algebras I*, J. Reine Angew. Math. **201** (1959), 178-195; *II*, ibid. **204** (1960), 74-98; *III*, ibid. **207** (1961), 61-95.
7. ———, *Lie Algebras*, Interscience, New York, 1962.
8. M. Koecher, *Eine Charakterisierung der Jordan Algebren*, Math. Ann. **148** (1962), 244-256.
9. ———, *Jordan Algebras and their Applications*, Multilithed notes, University of Minnesota, 1962.
10. O. S. Rothaus, *Domains of positivity*, Abh. Math. Seminar Hamburg **24** (1960), 189-235.
11. R. D. Schafer, *On cubic forms permitting a new type of composition*, J. Math. Mech. **10** (1961), 159-174.
12. ———, *On forms of degree n permitting composition*, J. Math. Mech. **12** (1963), 777-792.
13. ———, *Noncommutative Jordan algebras*, Proc. Amer. Math. Soc. **6** (1955), 472-475.
14. J. Tits, *A theorem on generic norms of strictly power associative algebras*, Proc. Amer. Math. Soc. **15** (1964), 35-36.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

TOPOLOGIES FOR LAPLACE TRANSFORM SPACES

DONALD E. MYERS

In this paper four topologies are compared:
(i) an L_2-type topology on the space of functions having bilateral transforms,
(ii) an L_1 and (iii) an L_2-type topology on the space of transforms, and
(iv) finally that of one form of convergence of compact subsets for the space of analytic functions. It is shown that sequential convergence in (i) implies (iii) and (iv) and (ii) implies (i) and (iv) and hence (iii).

In an earlier paper [2]; the author used equivalence classes of analytic functions to construct an imbedding space for Schwartz Distributions. The mechanism for constructing the mapping was the bilateral Laplace Transform, in this way the traditional approach to operational calculus was preserved. In that paper a topology was imposed on the imbedding space from the space of analytic functions. We now obtain some additional results about the possible topologies defined on the space of analytic functions.

THEOREM 1. *Let $F_j(t), j = 0, 1, 2, \cdots, -\infty < t < \infty$ be real valued functions such that for each j*

$$d_1(F_j) = \left[\int_0^{\infty} | e^{-\sigma_1 t} F_j(t) |^2 \, dt\right]^{1/2} < \infty$$

and

$$d_2(F_j) = \left[-\int_0^{-\infty} | e^{-\sigma_2 t} F_j(t) |^2 \, dt\right]^{1/2} < \infty$$

where $-\infty < \sigma_1 < \sigma_2 < \infty$. If

$$d(F_j - F_0) = d_1(F_j - F_0) + d_2(F_j - F_0) \to 0$$

as $j \to \infty$ then

(i) *$f_j(z) \to f_0(z)$ uniformly on compact subsets of $\sigma_1 < R(z) < \sigma_2$ where $f_j(z) = \int_{-\infty}^{\infty} e^{-zt} F_j(t) dt$*

(ii) *$\| f_j - f_0 \|_x \to 0$*

where

Received April 24, 1964. This research was partially supported by NSF Contract Grant NSF GP-1951.

$$\| f_j \|_x = \left[\int_{-\infty}^{\infty} | f_j(x + iy) |^2 \, dy \right]^{1/2}$$

for each

$$\sigma_1 < x < \sigma_2 .$$

Proof. From [3], pp 245, $d_1 < \infty$, $d_2 < \infty$ implies the existence of $f_j(z)$ with the integral defining $f_j(z)$ converging absolutely for $\sigma_1 < R(z) < \sigma_2$. Rewrite this integral as

$$f_j(z) = \int_{-\infty}^{0} e^{-t(z-\sigma_2)} e^{-\sigma_2 t} F_j(t) dt + \int_{0}^{\infty} e^{-t(z-\sigma_1)} e^{-t\sigma_1} F_j(t) dt .$$

By the Cauchy-Schwarz Inequality

$$| f_j(z) | \leqq \left[\frac{1}{2[\sigma_2 - R(z)]} \int_{-\infty}^{0} | e^{-\sigma_2 t} F_j(t) |^2 \, dt \right]^{1/2} + \left[\frac{1}{2[R(z) - \sigma_1]} \int_{0}^{\infty} | e^{-\sigma_1 t} F_j(t) |^2 \, dt \right]^{1/2} .$$

Let

$$g_K(z) = \max_{z \in K} \left[\sqrt{\frac{1}{2[R(z) - \sigma_1]}} , \sqrt{\frac{1}{2[\sigma_2 - R(z)]}} \right]$$

for K an arbitrary subset of $-\sigma_1 < R(z) < \sigma_2$. Now

$$\max_{z \in K} | f_j(z) - f_0(z) | \leqq g_K(z) d(F_j - F_0)$$

and (i) is established.

To prove (ii) we use an additional consequence of the theorem from [3], pp 254, namely for each j

$$\int_{-\infty}^{\infty} | e^{-xt} F_j(t) |^2 \, dt = \int_{-\infty}^{\infty} | f_j(x + iy) |^2 \, dt$$

for $\sigma_1 < x < \sigma_2$. Since

$$\int_{0}^{\infty} | e^{-xt} F_j(t) |^2 \, dt \leqq [d_1(F_j)]^2$$

$$\int_{-\infty}^{0} | e^{-xt} F_j(t) |^2 \, dt \leqq [d_2(F_j)]^2$$

it follows that

$$\| f_j - f_0 \|_x^2 \leqq [d_1(F_j - F_0)]^2 + [d_2(F_j - F_0)]^2 .$$

By definition $d_1(F_j) \geqq 0$, $d_2(F_j) \geqq 0$, therefore $d(F_j - F_0) \to 0$ implies $d_1(F_j - F_0) \to 0$ and $d_2(F_j - F_0) \to 0$ and (ii) is proved.

Although it is clear that $d(F)$ is a norm on the set of functions with $d < \infty$ and $\|f\|_x$ is a norm on a subset of the functions analytic for $\sigma_1 < R(z) < \sigma_2$, the topology of uniform convergence on compact sets gives only a countably normed space or a Fréchet matric. (See [1], pg. 139.)

THEOREM 2. *Let $f_j(z)$, $j = 0, 1, 2, \cdots$ be functions analytic for $\sigma_1 < R(z) < \sigma_2$ such that for each j*

$$\int_{-\infty}^{\infty} |f_j(x + iy)|\, dy < \infty, \quad \sigma_1 < x < \sigma_2$$

and $\lim_{|y|\to\infty} f_j(x + iy) = 0$ uniformly on closed subintervals of $\sigma_1 < x < \sigma_2$. If $\|f_j - f\|_x \to 0$ as $j \to \infty$ for each $\sigma_1 < x < \sigma_2$. then

(i) *$d'(F_j - F_0) \to 0$ as $j \to \infty$ where*

$$d'(F_j - F_0) = d_1'(F_j - F_0) + d_2'(F_j - F_0)$$

and

$$F_j(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{xt} e^{ity} f_j(x + iy)dy\,, \tag{1}$$

$\sigma_1 < x < \sigma_2$. (*The prime denotes that the metric is the same as in Theorem* 1, *except that σ_1 is replaced by x_1 and σ_2 by x_2 where $\sigma_1 < x_1 < x_2 < \sigma_2$.*)

(ii) *$f_j(z) \to f_0(z)$ uniformly on compact subsets of $x_1 < R(z) < x_2$.*

Proof. From [3], pg. 265 the hypotheses of the theorem are sufficient to insure that each $f_i(z)$ is a bilateral transform and also the validity of the inversion formula (1).

Since $|f_j(x + iy)| \to 0$ as $|y| \to \infty$ each $|f_j(x + iy)|$ is bounded for $\sigma_1 < x < \sigma_2$ if $|f_j(x + iy)| \leqq M_j(x)$ then

$$\int_{-\infty}^{\infty} |f_j(x + iy)|^2\, dy \leqq M_j(x) \int_{-\infty}^{\infty} |f_j(x + iy)|\, dy < \infty\,.$$

Consider, for $\sigma_1 < x < x_1$

$$\int_0^{\infty} |e^{x_1 t} F_j(t)|^2\, dt = \frac{1}{(2\pi)^2} \int_0^{\infty} \left| e^{-t(x_1 - x)} \int_{-\infty}^{\infty} e^{ity} f_j(x + iy)dy \right|^2 dt$$

$$\leqq \frac{1}{(2\pi)^2} \|f_j\|_x^2 \int_0^{\infty} e^{-2t(x_1 - x)} dt$$

but $0 < x_1 - x$ so that

$$C_1 = \frac{1}{2\pi} \int_0^{\infty} e^{-2t(x_1 - x)} dt < \infty$$

and $d_1'(F_j) \leqq C_1 \| f_j \|_x$. Likewise

$$\int_{-\infty}^{0} | e^{-x_2 t} F_j(t) |^2 \, dt \leqq \frac{1}{(2\pi)^2} \| f_j \|_{x'}^2 \int_{-\infty}^{0} e^{t(x'-x_2)} dt$$

for $x_2 < x' < \sigma_2$, or

$$d_2'(F_j) \leqq C_2 \| f_j \|_{x'}, \quad \text{where} \quad C_2 = \frac{1}{2\pi} \int_{-\infty}^{0} e^{t(x'-x_2)} dt \, .$$

If $\| f_j - f_0 \|_x \to 0$ as $j \to \infty$ for all $\sigma_1 < x < \sigma_2$ then

$$\begin{aligned} d'(F_j - F_0) &= d_1'(F_j - F_0) + d_2'(F_j - F_0) \\ &\leqq C_1 \| f_j - f_0 \|_x + C_2 \| f_j - f_0 \|_{x'} \end{aligned}$$

or $d'(F_j - F_0) \to 0$.

Part (ii) follows by applying Theorem 1 to part (i).

In a subsequent paper the author expects to extend these to multidimensional Laplace Transforms and to Laplace Transforms on Locally Compact Abelian Groups.

References

1. W. Thron, *Introduction to the Theory of Functions of a Complex Variable*, John Wiley, New York, 1953.
2. D. E. Myers, *An imbedding space for Schwartz distributions*, Pacific J. Math. **11** (1961), 1467-1477.
3. D. V. Widder, *The Laplace Transform*, Princeton, 1946.

University of Arizona

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

ON SOME CLASSES OF NEARLY OPEN SETS

OLAV NJÅSTAD

The open sets in a topological space are those sets A for which $A^0 \supset A$. Sets for which $A^{0-0} \supset A$ -"α-sets"- or $A^{0-} \supset A$ -"β-sets"- may naturally be considered as more or less "nearly open". In this paper the structure of these sets and classes of sets are investigated, and some applications are given.

Topologies determining the same class of α-sets also determine the same class of β-sets, and vice versa. The class of β-sets forms a topology if and only if the original topology is extremally disconnected. The class of α-sets always forms a topology, and topologies generated in this way-"α-topolgies"- are exactly those where all nowhere dense sets are closed.

The class of all topologies which determine the same α-sets is convex in the ordering by inclusion, the α-topology being its finest member. Most topologies ordinary met with are the coarsest members of their corresponding classes; in particular this is the case for all regular topologies.

All topologies determining the same α-sets also determine the same continuous mappings into arbitrary regular spaces.

Since the paper was submitted, it has come to our attention that Freud, in [5], has investigated problems which have a certain connection with parts of those here treated.

1. Let $\mathscr{T}$ be a topology (identified with its class of open sets) on a set E, and let ^{0} and $^-$ denote interior and closure with respect to $\mathscr{T}$. We shall call a set A with the property $A^{0-0} \supset A$ an *α-set* (with respect to $\mathscr{T}$), and we shall denote the class of all such sets $\mathscr{T}^\alpha$. A set B with the property $B^{0-} \supset B$ (or equivalently $B^{0-} = B^-$) shall be called a *β-set* (with respect to $\mathscr{T}$), and the class of all such sets denoted $\mathscr{T}^\beta$. A class consisting of exactly all the α-sets (resp. β-sets) of some topology shall be called an *α-structure* (resp. *β-structure*). Evidently $\mathscr{T} \subset \mathscr{T}^\alpha \subset \mathscr{T}^\beta$. *We notice that every nonempty β-set has a nonempty interior.* If all sets of the family $\{B_i\}_{i \in I}$ are β-sets, then

$$\bigcup_{i \in I} B_i \subset \bigcup_{i \in I} B_i^{0-} \subset (\bigcup_{i \in I} B_i^0)^- \subset (\bigcup_{i \in I} B_i)^{0-} ,$$

that is: A β-structure is closed with respect to arbitrary unions.

We shall now characterize $\mathscr{T}^\alpha$ in terms of $\mathscr{T}^\beta$:

Received September 18, 1963 and in revised form April 28, 1964.

PROPOSITION 1. Let $\mathscr{T}$ be a topology on E. $\mathscr{T}^{\alpha}$ consists of exactly those sets A for which $A \cap B \in \mathscr{T}^{\beta}$ for all $B \in \mathscr{T}^{\beta}$.

Proof. Let $A \in \mathscr{T}^{\alpha}$, $B \in \mathscr{T}^{\beta}$, $x \in A \cap B$ and let U be an open neighbourhood of x. Clearly $U \cap A^{0-0}$ too is an open neighbourhood of x, so $V = (U \cap A^{0-0}) \cap B^{0}$ is nonempty. Since $V \subset A^{0-}$ this implies

$$U \cap (A^{0} \cap B^{0}) = V \cap A^{0} \neq \phi \ .$$

It follows that

$$A \cap B \subset (A^{0} \cap B^{0})^{-} = (A \cap B)^{0-} \ ,$$

i.e. $A \cap B \in \mathscr{T}^{\beta}$.-Conversely, let $A \cap B \in \mathscr{T}^{\beta}$ for all $B \in \mathscr{T}^{\beta}$. Then in particular $A \in \mathscr{T}^{\beta}$. Assume $x \in A \cap \mathsf{C}(A^{0-0})$($\mathsf{C}$ denoting complement). Then $x \in B^{-}$, where $B = \mathsf{C}(A^{0-})$. Clearly $\{x\} \cup B \in \mathscr{T}^{\beta}$, and consequently

$$A \cap (\{x\} \cup B) \in \mathscr{T}^{\beta} \ .$$

But

$$A \cap (\{x\} \cup B) = \{x\} \ ,$$

hence $\{x\}$ is open. As $x \in A^{0-}$, this implies $x \in A^{0-0}$, contrary to assumption. Thus $x \in A$ implies $x \in A^{0-0}$, and $A \in \mathscr{T}^{\alpha}$. This completes the proof.

-Thus we have found that $\mathscr{T}^{\alpha}$ is completely determined by $\mathscr{T}^{\beta}$, i.e.: all topologies with the same β-structue also determine the same α-structure, explicitly given by Prop. 1. We shall see that conversely all topologies with the same α-structure determine the same β-structure, so that $\mathscr{T}^{\beta}$ is completely determined by $\mathscr{T}^{\alpha}$.

PROPOSITION 2. Every α-structure is a topology.

Proof. $\mathscr{T}^{\beta}$ contains the empty set and is closed with respect to arbitrary unions. A standard result gives that the class of those sets A for which $A \cap B \in \mathscr{T}^{\beta}$ for all $B \in \mathscr{T}^{\beta}$ constitutes a topology, hence the proposition.

-Henceforth we shall also use the term *α-topology* for α-structure. Two topologies determining the same α-structure shall be called α-equivalent, and the equivalence classes shall be called *α-classes*.

We may now characterize $\mathscr{T}^{\beta}$ in terms of $\mathscr{T}^{\alpha}$ in the following way:

PROPOSITION 3. Let $\mathscr{T}$ be a topology on E. Then $\mathscr{T}^{\beta} = \mathscr{T}^{\alpha\beta}$, and hence α-equivalent topologies determine the same β-structure.

Proof. Let cl. and int. denote closure and interior with respect to $\mathcal{T}^\alpha$. If $x \in B \in \mathcal{T}^\beta$ and $x \in A \in \mathcal{T}^\alpha$, then $A^{0-0} \cap B^0 \neq \phi$ since A^{0-0} is a neighbourhood of x. So certainly B^0 meets A^{0-} and therefore (being open) meets A^0, proving $A \cap B^0 \neq \phi$, and *a fortiori* $A \cap \text{int. } B \neq \phi$. This means

$$\text{cl. int. } B \supset B ,$$

i.e.: $B \in \mathcal{T}^{\alpha\beta}$. -On the other hand let $A \in \mathcal{T}^{\alpha\beta}$, $x \in A$ and $x \in V \in \mathcal{T}$. As $V \in \mathcal{T}^\alpha$ and $x \in \text{cl. int. } A$, we have

$$V \cap \text{int. } A \neq \phi ,$$

and there exists a nonempty set $W \in \mathcal{T}$ such that $W \subset V \cap \text{int. } A \subset A$. In other words:

$$V \cap A^0 \neq \phi ,$$

and $x \in A^{0-}$. Thus we have verified $\mathcal{T}^{\alpha\beta} \subset \mathcal{T}^\alpha$, and the proof is complete. -Combining Prop. 1 and Prop. 3 we get $\mathcal{T}^{\alpha\alpha} = \mathcal{T}^\alpha$, or

COROLLARY 1. *A topology $\mathcal{T}$ is an α-topology if and only if $\mathcal{T} = \mathcal{T}^\alpha$. Thus an α-topology belongs to the α-class of all its determining topologies, and is the finest topology of this class.*

Evidently $\mathcal{T}^\beta$ is a topology if and only if $\mathcal{T}^\alpha = \mathcal{T}^\beta$. In this case $\mathcal{T}^{\beta\beta} = \mathcal{T}^{\alpha\beta} = \mathcal{T}^\beta$, or

COROLLARY 2. *If a β-structure $\mathcal{B}$ is a topology, then $\mathcal{B} = \mathcal{B}^\alpha = \mathcal{B}^\beta$.*

Before proceeding with our discussion we consider the following

EXAMPLE. Let $\mathcal{R}$ be the ordinary topology on the set of real numbers R. Clearly the complement of a nowhere dense, not closed set (such sets certainly exist) is a nonopen α-set. So $\mathcal{R} \neq \mathcal{R}^\alpha$. Among the β-sets are all intervals, so $\mathcal{R}^\alpha \neq \mathcal{R}^\beta$, and $\mathcal{R}^\beta$ is not a topology.

2. We proceed to give some results on the structure of α-topologies.

PROPOSITION 4. The α-sets with respect to a given topology are exactly those sets which may be written as a difference between an open set and a nowhere dense set.

Proof. If $A \in \mathcal{T}^\alpha$ we have

$$A = A^{0-0} - (A^{0-0} - A) ,$$

where $A^{0-0} - A$ clearly is nowhere dense. Conversely, if $A = B - N$, $B \in \mathscr{T}$, N nowhere dense, we easily see that $A^{0-} \supset B$ and consequently

$$A^{0-0} \supset B \supset A .$$

So the proof is complete.

COROLLARY. *A topology is an α-topology if and only if all nowhere dense sets are closed.*

For an α-topology may clearly be characterized as a topology where the difference between an open set and a nowhere dense set is again an open set, and this evidently is equivalent to the condition stated.

PROPOSITION 5. Topologies which are α-equivalent determine the same class of nowhere dense sets.

Proof. Let N be a nowhere dense set with respect to $\mathscr{T}^\alpha$. For every nonempty $A \in \mathscr{T}$ there is a nonempty $B \in \mathscr{T}^\alpha$ such that $B \subset A$ and $B \cap N = \phi$. As $B^0 \neq \phi$, it follows that N is not dense in A with respect to $\mathscr{T}$, and consequently that N is nowhere dense with respect to $\mathscr{T}$.

Conversely, let N be nowhere dense with respect to $\mathscr{T}$. N^- contains no nonempty set from $\mathscr{T}$. As cl. $N \subset N^-$ (cl. denotes closure with respect to $\mathscr{T}^\alpha$), cl. N contains no nonempty set from $\mathscr{T}^\alpha$. So N is nowhere dense with respect to $\mathscr{T}^\alpha$, which completes the proof.

The converse of Prop. 5 is naturally not true, since both the coarsest and the finest topology on a set are α-topologies, and neither of them possesses any nonempty nowhere dense sets.

We observe that every β-set- and *a fortiori* every α-set- is the union of an open set and a nowhere dense set. This together with the corollary to Prop. 4 implies

COROLLARY 1. *If the topology $\mathscr{T}$ has the property that all open sets are $\mathscr{F}_\sigma$-sets, then the topology $\mathscr{T}^\alpha$ has the same property.*

A topology where no nonempty open set is of the first category is called a Baire-topology (cf. [4, p. 109]). From what is said above immediately follows

COROLLARY 2. *A topology $\mathscr{T}$ is a Baire-topology if and only only if $\mathscr{T}^\alpha$ is a Baire-topology.*

We shall comment on these corollaries later on.

A set A is called *minimally bounded* with respect to the topology $\mathscr{T}$ if $A^{0-} \supset A, A^{-0} \subset A$ (cf. [1, p. 101]). (In the case of *open* sets, minimal boundedness coincides with *regularity* in the sense of [2, p. 176]). Clearly this means $A \in \mathscr{T}^{\beta}, \mathsf{C}A \in \mathscr{T}^{\beta}$.

PROPOSITION 6. Topologies which are α-equivalent determine the same class of minimally bounded sets, and the same class of minimally bounded open sets.

Proof. The first assertion is obvious. Let $A \in \mathscr{T}^{\alpha}, \mathsf{C}A \in \mathscr{T}^{\beta}$. Then we have

$$A^{0-0} \supset A , \qquad (\mathsf{C}A)^{0-} \supset \mathsf{C}A .$$

It follows that

$$A^{0-0} = A^{-0} = A .$$

So $A \in \mathscr{T}$, and the proposition follows.

We recall that a topology is called *extremally disconnected* if the closure of every open set is open.

PROPOSITION 7. If the β-structure $\mathscr{B}$ is a topology, all topologies $\mathscr{T}$ for which $\mathscr{T}^{\beta} = \mathscr{B}$ are extremally disconnected. If $\mathscr{B}$ is not a topology, no $\mathscr{T}$ for which $\mathscr{T}^{\beta} = \mathscr{B}$ is extremally disconnected. In particular: Either all or none of the topologies of an α-class are extremally disconnected.

Proof. Let $\mathscr{T}^{\beta} = \mathscr{B}$, and suppose there is an $A \in \mathscr{T}$ such that $A^{-} \notin \mathscr{T}$. Let $x \in A^{-} - A^{-0}$. With

$$B = \{x\} \cup A^{-0}, C = \mathsf{C}(A^{-0})$$

we have

$$B^{0-} \supset A^{-0-} = A^{-} \supset \{x\}, C^{0-} = \mathsf{C}(A^{-0}) = C \supset \{x\} .$$

Hence both B and C are in $\mathscr{T}^{\beta}$. The intersection $B \cap C = \{x\}$ is not open since $x \in A^{-} - A^{-0}$, hence not a β-set. So $\mathscr{B} = \mathscr{T}^{\beta}$ is not a topology.

Now suppose $\mathscr{B}$ is not a topology, and $\mathscr{T}^{\beta} = \mathscr{B}$. There is a $B \in \mathscr{T}^{\beta}$ such that $B \notin \mathscr{T}^{\alpha}$. Asssume $B^{0-} \in \mathscr{T}$. Then

$$B^{0-0} = B^{0-} \supset B ,$$

i.e. $B \in \mathscr{T}^\alpha$, contrary to assumption. Thus we have produced an open set whose closure is not open, which completes the proof.

COROLLARY. *A topology $\mathscr{T}$ is extremally disconnected if and only if $\mathscr{T}^\beta$ is a topology.*

Proof evident.

We next give a result on the continuous mappings of the topological space $(E, \mathscr{T})$ into a topological space $(F, \mathscr{U})$.

PROPOSITION 8. All topologies of a given α-class on E determine the same class of continuous mappings into an arbitrary regular topological space $(F, \mathscr{U})$.

Proof. Let $\mathscr{T}$ be a topology on E and assume f continuous from $(E, \mathscr{T}^\alpha)$ into the regular space $(F, \mathscr{U})$. Let $x \in E$, and let U be a closed neighbourhood of $f(x)$. There is a $V \in \mathscr{T}^\alpha$ such that $x \in V$ and $f(V) \subset U$. We may write $V = A - N$, where $A \in \mathscr{T}$, N is nowhere dense (Prop. 4), $N \subset A$. Let $y \in N$, and suppose $f(y) \notin U$. Then there exists a $W \in \mathscr{T}^\alpha$ such that $y \in W \subset A$ and $f(W) \subset \mathsf{C}U$. But as N is nowhere dense we have $W \not\subset N$. That is: $W \cap V \neq \phi$, contrary to assumption. Thus $f(A) \subset U$, hence f is continuous with respect to $\mathscr{T}$ and $\mathscr{U}$ and the proof is complete.

In this connection we mention that a mapping of the space $(E, \mathscr{T})$ into the space $(F, \mathscr{U})$ is called *quasicontinuous* if for every $x \in E$, every neighbourhood V of x and U of $f(x)$ there exists a nonempty $W \in \mathscr{T}$ such that $W \subset V$ and $f(W) \subset U$ (cf. [6, p. 184]). It is easily seen that this is equivalent to the condition: For every $U \in \mathscr{U}$, $f^{-1}(U) \in \mathscr{T}^\beta$. Hence we immediately get

PROPOSITION 9. All topologies of a given α-class on E determine the same class of quasicontinuous mappings into an arbitrary topological space $(F, \mathscr{U})$.

3. We shall now deduce some properties of the order structure of the α-classes. We first recall that every class contains a finest element, namely its associated α-topology. Next we prove the following

PROPOSITION 10. Every α-class is convex in the set of topologies on E ordered by inclusion. That is: $\mathscr{T} \subset \mathscr{U} \subset \mathscr{T}^\alpha$ implies $\mathscr{T}^\alpha = \mathscr{U}^\alpha$.

Proof. We assume $\mathscr{T} \subset \mathscr{U} \subset \mathscr{T}^\alpha$, and denote interior and closure in $\mathscr{T}$ (resp. $\mathscr{U}$) by ^{0}, $^-$ (resp. int., cl.). Using $\tilde{}$ to denote closure

in $\mathscr{T}^\alpha$, we have (for all A)

$$\text{cl. int. } A \supset (\text{int. } A)^\sim \supset A^{0\sim} .$$

Now $A^{0\sim}$ is the closure of an open set in $\mathscr{T}^\alpha$, hence minimally bounded in $\mathscr{T}^\alpha$ and so also in $\mathscr{T}$. Hence

$$A^{0\sim} \supset A^{0\sim-0} \supset A^{0-0} ,$$

so that

$$\text{cl. int. } A \supset A^{0-0} .$$

It follows that

$$\text{int. cl. int. } A \supset A^{0-0} .$$

Hence, if $A \in \mathscr{T}^\alpha$, we have *a fortiori* that

$$\text{int. cl. int. } A \supset A .$$

Thus

$$\mathscr{U} \subset \mathscr{T}^\alpha \subset \mathscr{U}^\alpha .$$

Applying this result to the inclusion $\mathscr{U} \subset \mathscr{T}^\alpha \subset \mathscr{U}$ we get

$$\mathscr{T}^\alpha \subset \mathscr{U}^\alpha \subset \mathscr{T}^{\alpha\alpha} = \mathscr{T}^\alpha .$$

So the equality $\mathscr{T}^\alpha = \mathscr{U}^\alpha$ is proved.

We shall now investigate in some detail the problem of a coarsest topology in the α-class, and in this connection give conditions for an α-topology to be the only topology of its class. To this end we make the following definition: We shall say that the topology $\mathscr{T}$ is *quasi-regular* at a point x if the point has a fundamental system of minimally bounded open neighbourhoods, and call the topology quasi-regular if it is quasi-regular at all points. It is easily seen that an equivalent condition for quasi-regularity at the point x is that it has a fundamental system of interiors of closed neighbourhoods. Clearly every topology satisfying the separation axiom 0_{III} of [3, p. 54] is quasi-regular, in particular every regular topology. With this terminology we may characterize those topologies which are the coarsest members of their classes in the following way:

PROPOSITION 11. A topology $\mathscr{T}$ is the coarsest topology of its α-class if and only if it is quasi-regular at all points which are closed with respect to $\mathscr{T}^\alpha$.

Proof. All topologies $\mathscr{U}$ for which $\mathscr{U}^\alpha = \mathscr{T}^\alpha$ determine the same class of minimally bounded open sets, and therefore the neighbourhood filter with respect to $\mathscr{T}$ is coarser than that with respect to $\mathscr{U}$ at points where $\mathscr{T}$ is quasi-regular. Now assume that $\mathscr{T}$ is quasi-regular at all points which are closed with respect to $\mathscr{T}^\alpha$, and let x be a point where $\mathscr{T}$ is not quasi-regular. Then $\{x\}$ is not nowhere dense, i.e. int. cl. $\{x\} \neq \phi$ (int. and cl. denote interior and closure with respect to $\mathscr{T}^\alpha$). Every nonempty α-set contained in int. cl. $\{x\}$ clearly contains x. Now every such set contains a nonempty set from $\mathscr{U}$, which of course also contains x. It follows that the neighbourhood filter with respect to $\mathscr{U}$ constitutes a fundamental system of neighbourhoods with respect to $\mathscr{T}^\alpha$. This means that all topologies of the α-class possess the same neighbourhood filter at x. Thus the sufficiency of the condition is demonstrated.

Now let $\{x\}$ be closed with respect to $\mathscr{T}^\alpha$, and suppose that $\mathscr{T}$ (and hence $\mathscr{T}^\alpha$) is not quasi-regular at x. Let $\mathscr{V}(x)$ denote the filter generated by all minimally bounded open neighbourhoods of x, and $\mathscr{V}(y)$ the neighbourhood filter of y with respect to $\mathscr{T}^\alpha$ for $y \neq x$. It is easily verified that $\mathscr{V}(z)$ satisfies the requirements $V_I - V_{IV}$ of [3, p. 11–12] for all $z \in E$. Thus the filters $\mathscr{V}(z)$ determine a topology $\mathscr{U}$ on E, which is not finer than $\mathscr{T}$, since the neighbourhood filter of x with respect to $\mathscr{U}$ is strictly coarser than that with respect to $\mathscr{T}$. Obviously $\mathscr{U} \subset \mathscr{T}^\alpha$.

Let $^-$ and ^{0} denote closure and interior with respect to $\mathscr{U}$. Closures with respect to $\mathscr{T}^\alpha$ and with respect to $\mathscr{U}$ of sets containing x clearly are equal. Now assume $A \in \mathscr{T}^\alpha$. Our object is to prove $A \in \mathscr{U}^\alpha$. Suppose this is false; then $A \notin \mathscr{U}$, so A is an open neighbourhood of x with respect to $\mathscr{T}^\alpha$. Clearly $A^0 = A - \{x\}$. Now from $x \in A^{0-}$ would follow

$$A^{0-} = \text{cl. } A \supset \text{int. cl. } A \supset A \supset \{x\} .$$

Since int. cl. A is minimally bounded it belongs to $\mathscr{U}$, and we would have

$$A^{0-0} \supset \text{int. cl. } A \supset A$$

which means $A \in \mathscr{U}^\alpha$. Thus we conclude that $x \notin A^{0-}$. Then there is a $B \in \mathscr{U}$ containing x such that $B \cap (A - \{x\}) = \phi$. Since both A and B are contained in $\mathscr{T}^\alpha$, $A \cap B \in \mathscr{T}^\alpha$. But $A \cap B = \{x\}$, and $\{x\}$ is closed with respect to $\mathscr{T}^\alpha$. This would mean that $\mathscr{T}^\alpha$ is quasi-regular at x, contrary to assumption. So we may conclude that $A \in \mathscr{U}^\alpha$. Thus we have shown that $\mathscr{T}^\alpha \subset \mathscr{U}^\alpha$. From $\mathscr{U} \subset \mathscr{T}^\alpha \subset \mathscr{U}^\alpha$ we conclude $\mathscr{U}^\alpha = \mathscr{T}^{\alpha\alpha} = \mathscr{T}^\alpha$ (Prop. 10). Thus if $\mathscr{T}$ is not quasi-regular at all

points which are closed with respect to $\mathscr{T}^\alpha$, it is not the coarsest topology of its α-class. This completes the proof.

COROLLARY 1. *A quasi-regular topology $\mathscr{T}$ is the coarsest topology of its α-class.*

Obvious.

COROLLARY 2. *If the topology $\mathscr{T}$ is not an α-topology, then its α-topology is not quasi-regular.*

Obvious.

We see that from every topology which is itself not an α-topology is deduced an α-topology which in a certain sense is rather pathological. This may be of some interest, since $\mathscr{T}^\alpha$ also inherits some "nice" properties from $\mathscr{T}$. Thus consider the most usual Baire-spaces: those which are locally compact or deduced from a complete metric structure. Many of these are not α-topologies; in these cases the construction of the associated α-topology gives Baire-spaces which are not even uniformizable (cor. 2 to Prop. 5). And to every perfectly normal topology is associated an (α-) topology which may be far from being perfectly normal, but which nevertheless has the property that every open set is an $\mathscr{F}_\sigma$-set (Cor. 1 to Prop. 5). According to Prop. 8 these two topologies determine the same realvalued continuous functions, only the open sets of the first occurring as inverse images of open sets in R.

An α-class does not always possess a coarsest topology, as is shown by the following

EXAMPLE. Let E be an infinite set, and let $\mathscr{T}$ consist of the empty set and all complements of finite sets. This topology clearly is an α-topology (and even a β-structure), not quasi-regular at any point. Here it is easily seen that the only topology which is coarser than all topologies of the α-class determined by $\mathscr{T}$ is the coarsest topology on E. This topology itself does not belong to the class.

PROPOSITION 12. An α-topology $\mathscr{T}$ is the only topology of its α-class if and only if it is quasi-regular at all closed points.

Proof evident.

As examples of α-topologies which are the only topologies of their α-classes we may mention the discrete topology and the coarsest

topology on E (both of which are quasi-regular). We have seen that the α-topology deduced from the ordinary topology $\mathscr{R}$ on R is distinct from $\mathscr{R}$, so this topology is not quasi-regular.

REFERENCES

1. G. Aumann, *Reelle Funktionen*, Berlin-Göttingen-Heidelberg, 1954.
2. G. Birkhoff, *Lattice Theory*, 2. ed., New York 1948.
3. N. Bourbaki, *Topologie Générale*, Ch. I-II, Act. Sci. et Ind. 858-1142, Paris 1951.
4. N. Bourbaki, *Topologie Générale*, Ch. IX, Acta. Sci. et Ind. 1045, Paris 1958.
5. G. Freud, *Ein Beitrag zu dem Satze von Cantor und Bendixson*, Acta. Math. Acad. Sci. Hung. **9** (1958), 333-336.
6. S. Kempisty, *Sur les fonctions quasicontinues*, Fund. Math. **19** (1932), 184-197.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

A CHARACTERIZATION OF CONDITIONAL PROBABILITY

MILTON PHILIP OLSON

The conditional probability functions relative to a sub-σ-field $\mathscr{B}$ are shown to constitute a vector-valued measure on the σ-field of the probability space, and it is proved that the conditional expectation relative to $\mathscr{B}$ of an $\mathscr{L}_1$ random variable X is the integral of X with respect to this vector-valued measure. A complete characterization is given of those vector-valued measures which are conditional probabilities. This machinery is illustratively applied to give alternative derivations of results of Moy and Rota on the characterization of conditional expectation operators.

Probabilists often regret that, in general, conditional probabilities do not define probability measures almost everywhere. This defect arises naturally from the fact that conditional probabilities are Radon-Nikodym derivatives of certain set functions; and, hence, they are defined only up to equivalence. It seems advantageous to relinquish the concept of conditional probabilities as point functions, almost everywhere determined, and consider them as they are—elements of a function space. Part of the awkwardness of conditional probabilities then disappears: The conditional probabilities form a vector-valued measure such that conditional expectation of an integrable function is its integral with respect to this conditional probability measure.

Throughout this paper $(\Omega, \mathscr{A}, \mu)$, will be a fixed probability space and $\mathscr{B}$ a fixed sub-σ-field of $\mathscr{A}$. $\mathscr{L}_1(\Omega, \mathscr{A}, \mu)$, or simply $\mathscr{L}_1$, denotes the Banach space of all complex-valued, μ-integrable, $\mathscr{A}$-measurable functions on Ω. $\mathscr{E}^{\mathscr{B}}$ will be the conditional expectation operator in $\mathscr{L}_1$ relative to $\mathscr{B}$. I_A is the indicator of the set A.

In § 1, we show that if $\varphi^{\mathscr{B}}(A) = \mathscr{E}^{\mathscr{B}} I_A$ for $A \in \mathscr{A}$, then $\phi^{\mathscr{B}}$ is a vector-valued measure on $\mathscr{A}$ with values in $\mathscr{L}_1$. Furthermore, we show that if $X \in \mathscr{L}_1$, then

$$\mathscr{E}^{\mathscr{B}} X = \int X d\varphi^{\mathscr{B}} , \tag{0.1}$$

the notion of integral here being defined as in [1, p. 323].

Over the last ten years, several characterizations of $\mathscr{E}^{\mathscr{B}}$ have appeared. The papers of Moy [3] and Rota [4] characterized $\mathscr{E}^{\mathscr{B}}$ in

Received March 5, 1964. This paper was prepared with the partial support of the Office of Naval Research, Contract Nonr-222 (43). This paper in whole or in part may be reproduced for any purpose of the United States Government.

terms of its properties as a linear operator; Sidak [5] characterized it by means of its action on $\mathscr{L}_2$. In § 2, our characterization of conditional probability will be completed by identifying those countably additive measures on $\mathscr{A}$ with values in $\mathscr{L}_1$ which are conditional probabilities. For such a measure, a formula like (0.1) will hold, and the corresponding sub-σ-field will be determined.

In § 3, this characterization of conditional probability will be used to obtain the characterization of conditional expectation of Moy [3] and of Rota [4] using a theorem on the representation of operators on a Lebesgue space. In a subsequent publication, further results in [3] and [4] as well as some in [5] will be obtained as applications of our characterization of conditional probabilities.

The author wishes to express his appreciation to Professor Edward W. Barankin for continued inspiration and encouragement in this line of thought, as well as for numerous simplifications and technical improvements in the results.

1. Representation of conditional expectation as an integral with respect to a vector-valued measure. Let the notational conventions introduced above continue to hold. By definition [1, p. 318], a function ψ on the σ-field $\mathscr{A}$ to $\mathscr{L}_1$ is called a vector-valued measure if $\psi(\phi) = 0$ and ψ is countably additive on disjoint sets belonging to $\mathscr{A}$. If $\{A_i\}_{i=1}^{\infty} \subset \mathscr{A}$ such that $A_i \cap A_j = \phi$ when $i \neq j$, then we have from properties of conditional expectations (see [2, p. 347]),

$$\varphi^{\mathscr{B}}\left(\sum_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \varphi^{\mathscr{B}}(A_i) \qquad \text{a.s.} \tag{1.1}$$

Since the partial sums $\sum_{i=1}^{n} \varphi^{\mathscr{B}}(A_i)$ are uniformly bounded a.s. by 1, it follows that their integrals are uniformly continuous with respect to μ. Therefore, be the $\mathscr{L}_r$-convergence theorem [2, p. 163],

$$\lim_{n\to\infty} \sum_{i=1}^{n} \varphi^{\mathscr{B}}(A_i) = \sum_{i=1}^{\infty} \varphi^{\mathscr{B}}(A_i)$$

in $\mathscr{L}_1$-norm. Noting (1.1), countable additivity follows. $\varphi^{\mathscr{B}}(\phi) = 0$ a.s. is immediate. Thus, $\varphi^{\mathscr{B}}$ is a vector-valued measure.

The total variation of a vector-valued measure need not be finite. Its role in the theory of integration with respect to a vector-valued measure ψ is assumed by a finite positive set function $|||\psi|||$ called the semi-variation of ψ. In general, $|||\psi|||$ is not additive. $|||\psi|||$ is additive if and only if $|||\psi|| = \nu[\psi]$, the total variation of ψ (see [1, p. 320]).

By definition,

$$|||\varphi^{\mathscr{B}}(A)||| = \sup_{\mathscr{B}, \{\alpha_i\}} \left\| \sum_{i=1}^{n} \alpha_i \varphi^{\mathscr{B}}(A_i) \right\|, \qquad A \in \mathscr{A} \tag{1.2}$$

where $\mathscr{P}$ is the collection of all finite partitions of the set A by elements of $\mathscr{A}$ and the α_i are complex numbers such that $|\alpha_i| \leqq 1$. Recall that

$$\begin{aligned} \|\varphi^{\mathscr{B}}(A)\| &= \int |\varphi^{\mathscr{B}}(A)|\, d\mu \\ &= \int \mathscr{E}^{\mathscr{B}} I_A d\mu = \int_A d\mu = \mu(A)\,, \qquad A \in \mathscr{A}. \end{aligned} \tag{1.3}$$

Therefore, for any partition $\{A_i\}_{i=1}^n$ of A

$$\begin{aligned} \left\| \sum_{i=1}^n \alpha_i \varphi^{\mathscr{B}}(A_i) \right\| &\leqq \sum_{i=1}^n |\alpha_i| \, \|\varphi^{\mathscr{B}}(A_i)\| \\ &\leqq \sum_{i=1}^n \mu(A_i) = \mu(A) \end{aligned} \tag{1.4}$$

since the numbers $\{\alpha_i\}_{i=1}^n$ all have moduli less than one. But equality of the first and last members of (1.4) is achieved when $\alpha_1 = \cdots = \alpha_n = 1$, so that $|||\varphi^{\mathscr{B}}||| = \nu(\mu) = \mu$.

Now let $X \in \mathscr{L}_1$; we shall show that X is integrable with respect to $\varphi^{\mathscr{B}}$ (see [1, p. 323]) and that its integral over Ω is $\mathscr{E}^{\mathscr{B}} X$. Let $\{X_n\}_1^\infty$ be a sequence of simple functions converging μ a.s. to X and such that $|X_n| \uparrow X$. It follows that $X_n \to X$ $\varphi^{\mathscr{B}}$ a.e. since $|||\varphi^{\mathscr{B}}||| = \mu$. Then [2, p. 348], $\mathscr{E}^{\mathscr{B}} X_n \to \mathscr{E}^{\mathscr{B}} X$ μ a.s. Since

$$|\mathscr{E}^{\mathscr{B}} X_n| \leqq \mathscr{E}^{\mathscr{B}} |X_n| \leqq \mathscr{E}^{\mathscr{B}} |X| \in \mathscr{L}_1\,, \tag{1.5}$$

the integrals of the $\mathscr{E}^{\mathscr{B}} X_n$ are uniformly continuous and therefore

$$\mathscr{E}^{\mathscr{B}} X_n \longrightarrow \mathscr{E}^{\mathscr{B}} X \qquad \text{in } \mathscr{L}_1\text{-norm.} \tag{1.6}$$

But if $X_n = \sum_{i=1}^r \alpha_i I_{A_i}$, we have

$$\begin{aligned} \int X_n d\varphi^{\mathscr{B}} &= \sum_{i=1}^r \alpha_i \varphi^{\mathscr{B}}(A_i) \\ &= \sum_{i=1}^r \alpha_i \mathscr{E}^{\mathscr{B}} I_{A_i} = \mathscr{E}^{\mathscr{B}}\left(\sum_{i=1}^r \alpha_i I_{A_i}\right) = \mathscr{E}^{\mathscr{B}} X_n\,. \end{aligned} \tag{1.7}$$

By (1.6) and (1.7), it follows that

$$\lim_{n\to\infty} \int X_n d\varphi^{\mathscr{B}} = \mathscr{E}^{\mathscr{B}} X \qquad \text{in } \mathscr{L}_1\text{-norm.} \tag{1.8}$$

Since (1.8) holds for any function in $\mathscr{L}_1$, it follows that

$$\lim_{n\to\infty} \int I_A X_n d\varphi^{\mathscr{B}} = \mathscr{E}^{\mathscr{B}} I_A X \quad \text{in } \mathscr{L}_1\text{-norm} \tag{1.9}$$

for all $A \in \mathscr{A}$.

These last two limit assertions establish, according to definition [1, p. 323], the integrability of X and the formula

$$\int_A X d\varphi^{\mathscr{B}} = \mathscr{E}^{\mathscr{B}} I_A X , \qquad A \in \mathscr{A}. \tag{1.10}$$

In particular, (0.1) holds as asserted.

We gather together the results of this section in the form of a theorem.

THEOREM 1. *Let $(\Omega, \mathscr{A}, \mu)$ be a probability space and $\mathscr{B}$ a sub-σ-field of $\mathscr{A}$. Then the conditional probability functions relative to $\mathscr{B}$, denoted $\varphi^{\mathscr{B}}(A)$ for $A \in \mathscr{A}$, form a vector-valued measure on $\mathscr{A}$ with values in $\mathscr{L}_1(\Omega, \mathscr{A}, \mu)$ having semi-variation μ. Furthermore, if $X \in \mathscr{L}_1$*

$$\mathscr{E}^{\mathscr{B}} X = \int X d\varphi^{\mathscr{B}} . \tag{1.11}$$

2. $\mathscr{L}_1$-valued measures which are conditional probabilities. In this section, we determine the $\mathscr{L}_1(\Omega, \mathscr{A}, \mu)$-valued measures on $\mathscr{A}$ which are conditional probabilities.

THEOREM 2. *Let ψ be an $\mathscr{L}_1(\Omega, \mathscr{A}, \mu)$-valued measure on $\mathscr{A}$. Then ψ is the vector-valued conditional probability measure relative to some sub-σ-field $\mathscr{B}$ if and only if the following three conditions are satisfied:*

$$\begin{aligned} &\text{(i)} \quad \psi(A) \geqq 0 \ \text{ a.s.} \qquad A \in \mathscr{A} , \\ &\text{(ii)} \quad \psi(\Omega) = 1 \ \text{ a.s.}, \\ &\text{(iii)} \quad \| I_A \psi(B)\psi(C) \| = \| \psi(A) I_B \psi(C) \| = \| \psi(A)\psi(B) I_C \| \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad A, B, C \in \mathscr{A} . \end{aligned} \tag{2.1}$$

If ψ has these three properties, then it is conditional probability relative to the σ-field

$$\mathscr{B} = \{A \in \mathscr{A} \mid \psi(A) = I_A \text{ a.s.}\} . \tag{2.2}$$

Proof. Since $0 \leqq \psi(A) \leqq 1$ a.s. for all $A \in \mathscr{A}$, the values of ψ are equivalence classes of μ-essentially bounded functions. Consequently the products shown in (iii) above are all μ-integrable; and, hence, their norms are finite.

The conditions of the theorem are necessary. Let $\varphi^{\mathscr{B}}$ be the conditional probability measure relative to a sub-σ-field $\mathscr{B}$, then $\varphi^{\mathscr{B}}$ is known to satisfy conditions (i) and (ii) above. For any $A, B, C \in \mathscr{A}$,

$$\| I_A \cdot \varphi^{\mathscr{B}}(B) \cdot \varphi^{\mathscr{B}}(C) \| = \int_A \varphi^{\mathscr{B}}(B) \cdot \varphi^{\mathscr{B}}(C) d\mu$$

$$
\begin{aligned}
&= \int_\Omega \mathscr{E}^{\mathscr{B}} \mid I_A \cdot \varphi^{\mathscr{B}}(B) \cdot \varphi^{\mathscr{B}}(C) \mid d\mu \\
&= \int_\Omega (\mathscr{E}^{\mathscr{B}} I_A) \cdot \varphi^{\mathscr{B}}(B) \cdot \varphi^{\mathscr{B}}(C) d\mu \\
&= \int_\Omega \varphi^{\mathscr{B}}(A) \cdot \varphi^{\mathscr{B}}(B) \cdot \varphi^{\mathscr{B}}(C) d\mu \ .
\end{aligned} \tag{2.3}
$$

The right-hand side is symmetric in A, B, and C. Interchanging first A with B and then A with C, (iii) is proved for $\varphi^{\mathscr{B}}$. If $B \in \mathscr{B}$, then I_B is $\mathscr{B}$-measurable and $\varphi^{\mathscr{B}}(B) = I_B$. If $A \in \mathscr{A}$, but $A \notin \mathscr{B}$, then I_A is not $\mathscr{B}$-measurable and $\varphi^{\mathscr{B}}(A) = \mathscr{E}^{\mathscr{B}} I_A \neq I_A$. Thus, $\mathscr{B}$ is completely characterized by (2.2). This completes the proof of necessity.

It remains to prove sufficiency. Let ψ be an $\mathscr{L}_1$-valued measure on $\mathscr{A}$ satisfying conditions (i), (ii), and (iii); and let

$$\mathscr{B} = \{A \in \mathscr{A} \mid \psi(A) = I_A\} \ .$$

By letting $C = \Omega$, so that $\psi(C) = 1$ in (iii), we have

$$\int_A \psi(B) d\mu = \int_B \psi(A) d\mu \qquad A, B \in \mathscr{A} \ . \tag{2.4}$$

This equation can be used to prove that $\mathscr{B}$ is closed under intersections. Let $A, B \in \mathscr{B}$ and $C \in \mathscr{A}$. Then (2.4) implies

$$\int_\sigma \psi(A \cap B) d\mu = \int_\Omega \psi(C) I_{A \cap B} d\mu \ . \tag{2.5}$$

By repeated application of (iii) and various definitions, we have

$$
\begin{aligned}
\int_\sigma \psi(A \cap B) d\mu &= \int_\Omega \psi(C) I_A I_B d\mu \\
&= \int_\Omega \psi(C) \psi(A) I_B d\mu \\
&= \int_\Omega I_\sigma \psi(A) \psi(B) d\mu \\
&= \int_\sigma I_A \cdot I_B d\mu \\
&= \int_\sigma I_{A \cap B} d\mu \ .
\end{aligned} \tag{2.6}
$$

Since this holds for all $C \in \mathscr{A}$, we conclude that $\psi(A \cap B) = I_{A \cap B}$ and that $A \cap B \in \mathscr{B}$. $\mathscr{B}$ is closed under complements: Let $A \in \mathscr{B}$. Then $\psi(A) = I_A$; also $\psi(\Omega) = 1 = I_\Omega$. By the additivity of ψ,

$$1 = \psi(\Omega) = \psi(A) + \psi(A^c) = I_A + \psi(A^c) \ .$$

Therefore $\psi(A^c) = 1 - I_A = I_{A^c}$, and $A^c \in \mathscr{B}$. Thus, $\mathscr{B}$ is a field.

To prove that $\mathscr{B}$ is a σ-field, it suffices to prove that $\mathscr{B}$ is closed under countable disjoint unions. Let $\{A_n\}_{n=1}^{\infty} \subset \mathscr{B}$ be a disjoint family. Then by the countable additivity of ψ,

$$\begin{aligned} \psi\left(\bigcup_{n=1}^{\infty} A_n\right) &= \sum_{n=1}^{\infty} \psi(A_n) \\ &= \sum_{n=1}^{\infty} I_{A_n} \\ &= I_{\bigcup_{n=1}^{\infty} A_n} . \end{aligned} \tag{2.7}$$

So, $\bigcup_{n=1}^{\infty} A_n \in \mathscr{B}$. Therefore $\mathscr{B}$ is a σ-field.

If $B \in \mathscr{B}$ and $A \in \mathscr{A}$, then (2.4) implies

$$\begin{aligned} \int_B \psi(A) d\mu &= \int_A \psi(B) d\mu \\ &= \int_A I_B d\mu \\ &= \mu(A \cap B) . \end{aligned} \tag{2.8}$$

This is the defining relation for conditional probability relative to $\mathscr{B}$ *provided* that $\psi(A)$ is a $\mathscr{B}$-measurable function.

In reality, the statement above is imprecise: $\psi(A)$ is an equivalence class of functions in $\mathscr{L}_1$. It suffices to prove that $\psi(A)$ contains one $\mathscr{B}$-measurable function. (Implicitly, we are using the fact that if $\mathscr{B}$ does not contain the entire collection of null sets in $\mathscr{A}$, these null sets can be adjoined to $\mathscr{B}$ in a harmless way.) From each equivalence class $\psi(A)$, we select a representative. Continuing the imprecision mentioned above, we will also denote this representative by $\psi(A)$, and show that $C_{A,\alpha} = \{\omega \mid \psi(A)(\omega) \leqq \alpha\} \in \mathscr{B}$ for any $A \in \mathscr{A}$ and any real number α. $(C^c_{A,\alpha} = \{\omega \mid \psi(A)(\omega) > \alpha\})$. This will prove that $\psi(A)$ is $\mathscr{B}$-measurable.

Let $B = C^c_{A,\alpha}$ and $C = C_{A,\alpha}$ in (iii) above. Then,

$$\begin{aligned} \int_A \psi(C^c_{A,\alpha})\psi(C_{A,\alpha}) d\mu &= \int_{\sigma_{A,\alpha}} \psi(A)\psi(C^c_{A,\alpha}) d\mu \\ &\leqq \alpha \int_{\sigma_{A,\alpha}} \psi(C^c_{A,\alpha}) d\mu , \end{aligned} \tag{2.9}$$

and

$$\begin{aligned} \int_A \psi(C^c_{A,\alpha})\psi(C_{A,\alpha}) d\mu &= \int_{\sigma^c_{A,\alpha}} \psi(A)\psi(C_{A,\alpha}) d\mu \\ &\geqq \alpha \int_{\sigma^c_{A,\alpha}} \psi(C_{A,\alpha}) d\mu . \end{aligned} \tag{2.10}$$

Therefore,

$$\alpha \int_{\sigma^c_{A,\alpha}} \psi(C_{A,\alpha})d\mu \leqq \int_A \psi(C^c_{A,\alpha})\psi(C_{A,\alpha})d\mu \leqq \alpha \int_{\sigma_{A,\alpha}} \psi(C^c_{A,\alpha})d\mu \,. \tag{2.11}$$

By (2.4), the extremal members of inequality (2.11) are equal, and we have

$$\alpha \int_{\sigma^c_{A,\alpha}} \psi(C_{A,\alpha})d\mu = \int_A \psi(C^c_{A,\alpha})\psi(C_{A,\alpha})d\mu = \int_{\sigma^c_{A,\alpha}} \psi(A)\psi(C_{A,\alpha})d\mu \,. \tag{2.12}$$

Or,

$$\int_{\sigma^c_{A,\alpha}} (\psi(A) - \alpha)\psi(C_{A,\alpha})d\mu = 0 \,. \tag{2.13}$$

But on $C^c_{A,\alpha}$, $\psi(A) - \alpha > 0$ μ a.s. Thus, since $\psi(C_{A,\alpha}) \geqq 0$ μ a.s., $\psi(C_{A,\alpha}) = 0$ μ a.s. on $C^c_{A,\alpha}$. Using (iii) again with $A = B = \Omega$ and $C = C_{A,\alpha}$, we have

$$\int_\Omega \psi(C_{A,\alpha})d\mu = \int_{\sigma_{A,\alpha}} 1 d\mu = \int_{\sigma_{A,\alpha}} I_{\sigma_{A,\alpha}} d\mu = \mu(C_{A,\alpha}) \,. \tag{2.14}$$

But

$$\int_\Omega \psi(C_{A,\alpha})d\mu = \int_{\sigma_{A,\alpha}} \psi(C_{A,\alpha})d\mu + \int_{\sigma^c_{A,\alpha}} \psi(C_{A,\alpha})d\mu = \int_{\sigma_{A,\alpha}} \psi(C_{A,\alpha})d\mu \,. \tag{2.15}$$

Comparing (2.14) and (2.15),

$$\int_{\sigma_{A,\alpha}} \psi(C_{A,\alpha})d\mu = \int_{\sigma_{A,\alpha}} I_{\sigma_{A,\alpha}} d\mu \,. \tag{2.16}$$

On $C^c_{A,\alpha}$, $\psi(C_{A,\alpha}) = I_{\sigma_{A,\alpha}} = 0$ μ a.s. On $C_{A,\alpha}$, $0 \leqq \psi(C_{A,\alpha}) \leqq 1$ by (i) and $I_{\sigma_{A,\alpha}} = 1$, so $I_{\sigma_{A,\alpha}} - \psi(C_{A,\alpha}) \geqq 0$ μ a.s. But (2.16) implies

$$\int_\Omega (I_{\sigma_{A,\alpha}} - \psi(C_{A,\alpha}))d\mu = 0 \,. \tag{2.17}$$

Thus, $\psi(C_{A,\alpha}) = I_{\sigma_{A,\alpha}}$ μ a.s. and $C_{A,\alpha} \in \mathscr{B}$ for all $A \in \mathscr{A}$ and α real. This concludes the proof of Theorem 2.

3. Characterization of conditional expectation in $\mathscr{L}_p, 1 \leqq p < \infty$. Throughout this section, T will be a continuous linear map of $\mathscr{L}_p$ into $\mathscr{L}_p$. As usual, it will be convenient to use the same symbol for a function and its equivalence class. Thus, the statement "X is bounded

by $M > 0$" means that the equivalence class of X consists of functions whose μ-essential supremum is not greater than M.

In [3], among more general results, Moy established the following which we now rephrase and state.

Moy's result: Let T be a continuous linear transformation of $\mathscr{L}_1$ into $\mathscr{L}_1$ having the following four properties:

(i) if X is bounded, then TX is bounded;

(ii) if X and Y are bounded, then

$$T(X \cdot TY) = TX \cdot TY ;$$

(iii) $$T1 = 1 ;$$

(iv) T is a contraction, that is, for every $X \in \mathscr{L}_1$

$$\| TX \| \leqq \| X \| .$$

Then, we have for every $X \in \mathscr{L}_1$,

$$TX = \mathscr{E}^{\mathscr{B}} X ,$$

where $\mathscr{B}$ is the σ-field of all sets $A \in \mathscr{A}$ such that $T(I_A \cdot X) = I_A \cdot TX$ for all bounded X.

In [4], Rota made the following definition which we will use below.

DEFINITION 1. An averaging operator T in $\mathscr{L}_p(\Omega, \mathscr{A}, \mu)$ (where p is a fixed real number $1 \leqq p \leqq \infty$) is a linear operator in $\mathscr{L}_p(\Omega, \mathscr{A}, \mu)$ with the following three properties:

(i) T is a contraction operator:

$$\| TX \|_{\mathscr{L}_p} \leqq \| X \|_{\mathscr{L}_p} \qquad \text{for } X \in \mathscr{L}_p ; \tag{3.1}$$

(ii) if Y is of class $\mathscr{L}_p$ and X is an essentially bounded function on $(\Omega, \mathscr{A}, \mu)$, then the function $(TX) \cdot (TY)$ is of class $\mathscr{L}_p$ and

$$T(X \cdot TY) = (TX) \cdot (TY) ; \tag{3.2}$$

(3.3) (iii) $$T1 = 1 .$$

We note that the operator T in Moy's result is an averaging operator, following a standard argument of approximation of a function in $\mathscr{L}_1$ by a sequence of bounded measurable functions. In the case $1 < p < \infty$, Rota [4] proved the following result which we rephrase and state.

Rota's Result: Let T be an averaging operator in $\mathscr{L}_p(\Omega, \mathscr{A}, \mu)$ (for fixed p). Then there exists a unique sub-σ-field $\mathscr{B} \subset \mathscr{A}$ such

that $TX = \mathscr{E}^{\mathscr{B}}X$ for $X \in \mathscr{L}_p$.

In the proof, Rota defines $\mathscr{B}$ as the smallest σ-field containing all the σ-fields of inverse images of the Borel sets by bounded functions Y fixed under T. It is clear that this procedure will yield a σ-field $\mathscr{B}$ which is completed with respect to μ-null sets, just as Moy's construction of $\mathscr{B}$ does. The above statement must be interpreted in this light. Comparing the recipe for construction of $\mathscr{B}$ in the two results, we see that Rota's criterion for membership in $\mathscr{B}$ is, on the surface, weaker than Moy's. On the other hand, if $TI_A = I_A$, then $TI_A \cdot X = I_A \cdot X$ for X bounded and $T(TI_A)\cdot X) = T(I_A \cdot X)$. By the smoothing property (3.2) of averaging operators, $T((TI_A)\cdot X) = TI_A \cdot TX = T(I_A \cdot X)$. Hence $I_A \cdot TX = T(I_A \cdot X)$ for all X bounded, and the two recipes define the same family of sets.

A theorem can now be stated which contains both Moy's result and Rota's. This general theorem will be proved using a theorem on the representation of operators on a Lebesgue space and the characterization of conditional probabilities given in § 2.

THEOREM 3. *Let T be an averaging operator in $\mathscr{L}_p(\Omega, \mathscr{A}, \mu)$ (where p is a fixed real number $1 \leqq p < \infty$). Then there exists a unique sub-σ-field $\mathscr{B} \subset \mathscr{A}$ completed with respect to the μ-null sets of $\mathscr{A}$ such that $TX = \mathscr{E}^{\mathscr{B}}X$ for $X \in \mathscr{L}_p$. Furthermore,*

$$\mathscr{B} = \{A \in \mathscr{A} \mid TI_A = I_A\} .$$

Proof. The first part of the proof consists in showing that $T^*X = TX$ for all X bounded. A brief explanation of the meaning of this "self-adjointness" relation is given for the sake of completeness: As is well known for $1 \leqq p < \infty$, $\mathscr{L}_p{}^*$ is isometrically isomorphic to $\mathscr{L}_q$, where q is determined by the relation $1 = 1/p + 1/q$. Every linear operator $T: \mathscr{L}_p \to \mathscr{L}_p$ determines a mapping $T^*: \mathscr{L}_q \to \mathscr{L}_q$. Since $\mathscr{L}_\infty \subset \mathscr{L}_p \cap \mathscr{L}_q$ as sets for $1 \leqq p < \infty$, it makes sense to compare the action of T and T^* on bounded functions. To avoid confusion, the norm will be shown explicitly: thus, $\| \cdot \|_{\mathscr{L}_p}$.

The fact that $\| TX \|_{\mathscr{L}_p} \leqq \| X \|_{\mathscr{L}_p}$ and $\| T1 \|_{\mathscr{L}_p} = \| 1 \|_{\mathscr{L}_p} = 1$ implies

$$\| T \|_{\mathscr{B}(\mathscr{L}_p)} = 1 , \tag{3.4}$$

where $\| \cdot \|_{\mathscr{B}(\mathscr{L}_p)}$ is operator norm in $\mathscr{B}(\mathscr{L}_p)$, the bounded linear operators on $\mathscr{L}_p$.

Rota [3, p. 58] has shown that for $1 < p < \infty$ the averaging operator hypotheses imply that $T^*1 = 1$. In the case $p = 1$, let $Y_0 = T^*1$. Then,

$$\int Y_0 d\mu = \int T^*1 \cdot 1 d\mu = \int 1 \cdot T1 d\mu = 1 . \tag{3.5}$$

Further, since $T^*: \mathscr{L}_\infty \to \mathscr{L}_\infty$, $T^*1 = Y_0 \in \mathscr{L}_\infty$; and

$$\begin{aligned} \| Y_0 \|_{\mathscr{L}_\infty} &= \sup_{\substack{X \in \mathscr{L}_1 \\ \|X\|_{\mathscr{L}_1} \leqq 1}} \left| \int Y_0 \cdot X d\mu \right| \\ &= \sup_{\substack{X \in \mathscr{L}_1 \\ \|X\|_{\mathscr{L}_1} \leqq 1}} \left| \int T^*1 \cdot X d\mu \right| \\ &= \sup_{\substack{X \in \mathscr{L}_1 \\ \|X\|_{\mathscr{L}_1} \leqq 1}} \left| \int TX d\mu \right| \leqq \| T \|_{\mathscr{B}(\mathscr{L}_1)} = 1 . \end{aligned} \tag{3.6}$$

This means $-1 \leqq Y_0 \leqq +1$ μ a.s.; and, taken together with (3.5), this means $Y_0 = 1$ μ a.s. Hence, we have for $1 \leqq p < \infty$

$$T^*1 = 1 . \tag{3.7}$$

Since $\int 1 \cdot TX d\mu = \int T^*1 \cdot X d\mu$, this gives immediately

$$\int TX d\mu = \int X d\mu , \qquad X \in \mathscr{L}_p . \tag{3.8}$$

And so, finally, for any bounded X and $A \in \mathscr{A}$, we may calculate

$$\begin{aligned} \int_A TX d\mu &= \int I_A \cdot TX d\mu = \int T(I_A \cdot TX) d\mu && \text{(by 3.8)} \\ &= \int (TI_A) \cdot (TX) d\mu && \text{(by 3.2).} \end{aligned} \tag{3.9}$$

But

$$\begin{aligned} \int_A T^*X d\mu &= \int (T^*X) \cdot I_A d\mu = \int X \cdot (TI_A) d\mu \\ &= \int T(X \cdot TI_A) d\mu && \text{(by 3.8)} \\ &= \int (TX) \cdot (TI_A) d\mu && \text{(by 3.3).} \end{aligned} \tag{3.10}$$

Therefore, since $\int_A T^*X d\mu = \int_A TX d\mu$ for all $A \in \mathscr{A}$, we can conclude

$$T^*X = TX \qquad \text{for all } X \text{ bounded.} \tag{3.11}$$

In the next part of the proof, the principal instrument will be a theorem on the representation of operators on a Lebesgue space whose general form Dunford and Schwartz [1, p. 540] ascribe to Kantorovich and Vulich:

Result of Kantorovich and Vulich: Let $(\Omega, \mathscr{A}, \mu)$ be a measure space and let T be a continuous linear map of the Banach space $\mathfrak{X}$ into $\mathscr{L}_1(\Omega, \mathscr{A}, \mu)$. Then there is a uniquely determined function $\psi(\cdot)$ on $\mathscr{A}$ to $\mathfrak{X}^*$ such that

(i) for each $X \in \mathfrak{X}$, the set function $\psi(\cdot)[X]$ is μ-continuous and countably additive on $\mathscr{A}$;

(ii) for every $X \in \mathfrak{X}$, we have

$$TX = \frac{d[\psi(\cdot)X]}{d\mu} \tag{3.12}$$

(iii) the norm of T satisfies the relations

$$\sup_{A \in \mathscr{A}} \| \psi(A) \|_{\mathfrak{X}^*} \leqq \| T \|_{0p} \leqq 4 \sup_{A \in \mathscr{A}} \| \psi(A) \|_{\mathfrak{X}^*} . \tag{3.13}$$

Conversely, if the function $\psi(\cdot)$ on $\mathscr{A}$ to $\mathfrak{X}^*$ satisfies (i), then (ii) defines an operator T on $\mathfrak{X}$ to $\mathscr{L}_1$ whose norm satisfies (iii).

Furthermore, T is weakly compact if and only if $\psi(\cdot)$ is countably additive on $\mathscr{A}$ in the strong topology of $\mathfrak{X}^*$.

The proof of this important theorem can be found in [1, p. 498]. To apply the result above to the operator T, we choose to consider T as a continuous linear map of $\mathscr{L}_p (1 \leqq p < \infty, p$ fixed) into $\mathscr{L}_1$. Thus, $\mathfrak{X}$ is $\mathscr{L}_p$ and $\mathfrak{X}^*$ can be identified with $\mathscr{L}_q$. If $\psi(\cdot)$ is the function given in the result above, then for any $X \in \mathscr{L}_1$ and $A \in \mathscr{A}$, we have

$$\begin{aligned} \int_A TX d\mu &= \int_A \frac{d\psi(\cdot)[X]}{d\mu} d\mu = \psi(A)[X] \\ &= \int \psi(A) \cdot X d\mu . \end{aligned} \tag{3.14}$$

On the other hand, $T = T^*$ on bounded functions, so

$$\begin{aligned} \int_A TX d\mu &= \int I_A \cdot (TX) d\mu \\ &= \int (T^* I_A) \cdot X d\mu = \int (T I_A) \cdot X d\mu . \end{aligned} \tag{3.15}$$

Hence, for each $A \in \mathscr{A}$, putting $X = I_B$, we get

$$\int_B T I_A d\mu = \int_B \psi(A) d\mu \qquad B \in \mathscr{A}. \tag{3.16}$$

And therefore,

$$T I_A = \psi(A) , \qquad A \in \mathscr{A}. \tag{3.17}$$

Rota shows [4, p. 58] that an averaging operator in $\mathscr{L}_p, 1 < p < \infty$

maps bounded measurable functions into bounded measurable functions. For $p = 1$, $\psi(A) \in \mathscr{L}_\infty$; and, hence $TI_A = \psi(A)$ is a bounded measurable function for $1 \leqq p < \infty$. $\psi(\phi) = TI_\phi = 0$. A simple domination argument shows that if $\{A_j\}_{j=1}^\infty$ is a disjoint family, $\sum_{j=1}^n I_{A_j} \to I_{\cup_{j=1}^\infty A_j}$ in $\mathscr{L}_p$-norm. Therefore

$$T\left(\sum_{j=1}^n I_{A_j}\right) = \sum_{j=1}^n \psi(A_j) \longrightarrow \psi\left(\bigcup_{j=1}^\infty A_j\right) \quad \text{in } \mathscr{L}_p\text{-norm.} \tag{3.18}$$

Thus, ψ is a vector-valued measure on $\mathscr{A}$ to the bounded measurable functions on $\mathscr{A}$, which is countably additive in the $\mathscr{L}_p$-topology $(1 \leqq p < \infty)$. A fortiori, ψ is $\mathscr{L}_1$-countably additive. Since $TI_\Omega = \psi(\Omega) = 1$, condition (ii) in (2.1) is satisfied. Furthermore, for $A \in \mathscr{A}$

$$\mu(A) = \int (T^*I) \cdot I_A d\mu = \int \psi(A) d\mu \,. \tag{3.19}$$

But,

$$\begin{aligned} \int \psi(A) d\mu &\leqq \int |\, \psi(A) \,|\, d\mu \leqq \int |\, TI_A \,|\, d\mu \\ &\leqq \| TI_A \|_{\mathscr{L}_1} \leqq \| T \|_{\mathscr{B}(\mathscr{L}_1)} \cdot \| I_A \|_{\mathscr{L}_1} \leqq \mu(A) \,, \end{aligned} \tag{3.20}$$

since $\| T \|_{\mathscr{B}(\mathscr{L}_p)} = 1, 1 \leqq p < \infty$. Therefore, equality holds throughout (3.20), so

$$\int [|\, \psi(A) \,| - \psi(A)] d\mu = 0 \,. \tag{3.21}$$

Because $|\, \psi(A) \,| - \psi(A) \geqq 0$, we have

$$\psi(A) = |\, \psi(A) \,| \geqq 0 \,. \tag{3.22}$$

Thus, condition (ii) in (2.1) is satisfied.

The smoothing property (3.2) implies T is idempotent: $T^2X = T(1 \cdot TX) = T1 \cdot TX = TX$. Now let $A, B, C \in \mathscr{A}$, using (3.17), (3.2), (3.8), the fact that T maps bounded measurable functions into bounded measurable functions, and idempotence, we have

$$\begin{aligned} \int_A \psi(B)\psi(C) d\mu &= \int_A (TI_B) \cdot (TI_C) d\mu \\ &= \int I_A \cdot T(I_B \cdot (TI_C)) d\mu \\ &= \int T(I_A \cdot T(I_B \cdot (TI_C))) d\mu \\ &= \int (TI_A) \cdot T(I_B \cdot (TI_C)) d\mu \end{aligned} \tag{3.23}$$

$$= \int (TI_A) \cdot (TI_B) \cdot (T^2 I_C) d\mu$$

$$= \int (TI_A) \cdot (TI_B) \cdot (TI_C) d\mu \, .$$

Clearly, the result is the same if A is interchanged with B or A is interchanged with C. Thus, condition (iii) in (2.1) is satisfied; and $\psi(\cdot)$ is a conditional probability in $\mathscr{L}_1$, by Theorem 2. The integral with respect to $\psi(\cdot)$ of $X \in \mathscr{L}_1$ is $\mathscr{E}^{\mathscr{B}} X$, where $\mathscr{B}$ is the sub-σ-field defined by

$$\mathscr{B} = \left\{ A \in \mathscr{A} \mid \psi(A) = \int I_A d\psi = I_A \right\} .$$

Thus, $\mathscr{B}$ consists of those sets in $\mathscr{A}$ whose indicator functions are invariant under $\mathscr{E}^{\mathscr{B}}$. It is well known [2, p. 348] that $\mathscr{E}^{\mathscr{B}}$ defines an $\mathscr{L}_p$-continuous linear transformation ($1 \leqq p < \infty$). Then,

$$\mathscr{E}^{\mathscr{B}} I_A = \psi(A) = TI_A \, , \tag{3.24}$$

and so the restriction of $\mathscr{E}^{\mathscr{B}}$ to $\mathscr{L}_p$ agrees with the $\mathscr{L}_p$-continuous operator T on a generating set; hence, it agrees on all of $\mathscr{L}_p$. We conclude that T is $\mathscr{E}^{\mathscr{B}}$ on $\mathscr{L}_p$. This, with the fact then that $\mathscr{B} = \{A \in \mathscr{A} \mid TI_A = I_A\}$, finishes the proof of Theorem 3.

This theorem shows that the set of averaging operators in $\mathscr{L}_r$ and $\mathscr{L}_s$ are in one-to-one correspondence $1 \leqq r, s < \infty$. Moreover, every averaging operator T on $\mathscr{L}_r$ has a unique $\mathscr{L}_s$-continuous extension $1 \leqq s \leqq r < \infty$. In a subsequent publication, the author intends to relate these facts to the work of Sidak [5].

REFERENCES

1. N. Dunford and J. Schwartz, *Linear Operators Part I: General Theory*, New York, 1958.
2. M. Loève, *Probability Theory*, 3rd ed., Princeton, 1963.
3. S. C. Moy, *Characterizations of conditional expectation as a transformation on function spaces*, Pacific J. Math. 4 (1954), 47-63.
4. G. C. Rota, *On the representation of averaging operators*, Rendiconti del Seminario Matematico della Universita di Padova, Vol. XXX, 1960, pp. 52-64.
5. Z. Sidak, *On relations between strict-sense and wide-sense conditional expectations*, Theory of Probability and Its Applications, Vol. II, 1957, pp. 267-272.

UNIVERSITY OF CALIFORNIA, BERKELEY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

A COUNTER-EXAMPLE TO A LEMMA OF SKORNJAKOV

B. L. OSOFSKY

In his paper, Rings with injective cyclic modules, translated in Soviet Mathematics 4 (1963), p. 36–39, L. A. Skornjakov states the following lemma: If a cyclic R-module M and all its cyclic submodules are injective, then the partially ordered set of cyclic submodules of M is a complete, complemented lattice.

An example is constructed to show that this lemma is false, thus invalidating Skornjakov's proof of the theorem: Let R be a ring all of whose cyclic modules are injective. Then R is semi-simple Artin. The theorem, however, is true. (See Osofsky [4].)

The theorem, however, is true. (See Osofsky [4].)

In this paper, all rings have identity and all modules are unital left modules. ${}_R\mathfrak{M}$ will denote the category of R-modules, and ${}_RM$ will signify $M \in {}_R\mathfrak{M}$.

Let Q be a commutative, left self injective, regular, non-Artin ring, and let I be a maximal ideal of Q which is not a direct summand of ${}_QQ$. (For example, let Q be a direct product of fields, and I a maximal ideal containing their direct sum.) Let $N = Q \oplus Q/I$. We observe the following:

1. ${}_QN$ **is injective.** Q is injective by hypothesis, and Q/I is a simple module over the commutative regular ring Q; hence injective by a theorem of Kaplansky. (See [5].)

2. ${}_QM \subseteqq {}_QN$ **is a direct summand of** ${}_QN$ **if and only if** ${}_QM$ **is finitely generated.** If ${}_QM$ is a direct summand of ${}_QN$, ${}_QM$ is generated by the projections of $(1, 0 + I)$ and $(0, 1 + I)$. If ${}_QM$ is finitely generated, and π is the projection of N onto $(Q, 0 + I)$, then $\pi({}_QM)$ is finitely generated. Hence $\pi({}_QM)$ is a direct summand of ${}_QQ$. (See von Neumann [6].) Say $Q = \pi({}_QM) \oplus K$. Since $\pi({}_QM)$ is projective (it is a direct summand of Q), ${}_QM = (\pi({}_QM))' \oplus (\text{Ker } \pi \cap {}_QM)$. Since Q/I is simple, $Q/I = (\text{Ker } \pi \cap {}_QM) \oplus K_2$ where $K_2 = 0$ or Q/I. Then $N = M \oplus K \oplus K_2$.

3. **The direct summands of** N **do not form a lattice.** In particular, $Q(1, 0 + I) \cap Q(1, 1 + I) = (I, 0 + I)$ is not a direct summand

Received August 5, 1964 and in revised form March 18, 1965. The author gratefully acknowledges support from the National Science Foundation under grant GP-1741. The author wishes to thank the referee for simplifying and clarifying her original example, and for several very constructive suggestions on the presentation.

of $(Q, 0 + I)$, hence not of N.

N is not a counter-example to Skornjakov's lemma, since N is not cyclic. However, properties 1, 2 and 3 are preserved under category isomorphisms. For we have:

PROPOSITION. ${}_RM$ is finitely generated $\Leftrightarrow$ the union of a linearly ordered chain of proper submodules is proper.

Proof. $\Rightarrow$ Let $M = \sum_{i=1}^{n} Rx_i$, and let $\{N_\mu\}$ be a linearly ordered chain of submodules whose union is M. If $x_i \in N_{\mu_i}$, then $\{x_i \mid i = 1, \cdots, n\} \subseteq N_\nu$, where $\nu = \max\{\mu_i \mid 1 \leqq i \leqq n\}$. Then $M = N_\nu$.

$\Leftarrow$ Given ${}_RM$, let $\aleph$ be the smallest cardinal such that M has a generating set of cardinality $\aleph$. Index such a generating set $\{x_\mu\}$ by $\{\mu \mid \mu < \Omega\}$, where Ω is the first ordinal of cardinality $\aleph$. Then $\{\sum_{\nu \leqq \mu} Rx_\nu\}$ is a linearly ordered chain of submodules whose union is M. If Ω is a limit ordinal (that is, if $\aleph$ is infinite), then each $\sum_{\nu \leqq \mu} Rx_\nu$ is generated by less than $\aleph$ elements; hence proper.

Thus M finitely generated corresponds to the categorical property that the collection of nonepimorphic monomorphisms into M is inductive under the ordering: $f \leqq g$ if and only if there is an h with $f = gh$.[1]

Let $R = \mathrm{Hom}_Q(Q \oplus Q, Q \oplus Q)$. By Morita [3], Theorem 3.4, the functor $\mathrm{Hom}_Q(Q \oplus Q, \ ) : {}_Q\mathfrak{M} \to {}_R\mathfrak{M}$ is a category isomorphism. Hence ${}_RM = \mathrm{Hom}_Q(Q \oplus Q, N)$ has properties 1, 2, 3. Moreover, if $K = \{\lambda \in R \mid (Q \oplus Q)\lambda \subseteq (0, I)\}$, then M is isomorphic to R/K since ${}_Q(Q \oplus Q)$ projective implies the natural map from $R = \mathrm{Hom}_Q(Q \oplus Q, Q \oplus Q) \to \mathrm{Hom}_Q(Q \oplus Q, Q \oplus Q/I) = M$ is an epimorphism. Hence M is cyclic, and as in 2, every direct summand of M is cyclic. Thus M is the required counter-example.

We conclude with the observation that the technique used in 2 gives us a categorical equivalence to regular rings which is closer to the usual definition than Auslander's theorem that R is regular if and only if the global flat dimension of R is 0. (See Auslander [1].)

$P \in {}_R\mathfrak{M}$ is a progenerator if it is finitely generated, projective, and every $M \in {}_R\mathfrak{M}$ is an epimorphic image of a direct sum of copies of P.

PROPOSITION. The following are equivalent:

[1] Although the categorical definition of finitely generated appears in H. Bass, *The Morita theorems*, University of Oregon (mimeographed notes), the author found no proof in the literature that this is equivalent to the module definition, and so is including this proof for completeness.

(a) R is regular.

(b) Every finitely generated submodule of a projective module is a direct summand.

(c) There is a progenerator $P \in {}_R\mathfrak{M}$ such that every finitely generated submodule of P is a direct summand.

Proof. (b) $\Rightarrow$ (a) (See von Neumann [6].)

(a) $\Rightarrow$ (c) R is a progenerator with the required properties.

(c) $\Rightarrow$ (b) Let N be a projective module, M a finitely generated submodule.

Let P be the progenerator of condition (c). Then there is an epimorphism $f: \Sigma \bigoplus P_i \to N$. Since N is projective, this splits and $\Sigma \bigoplus P_i = N' \bigoplus \ker f$, where $N' \approx N$. Thus M is a finitely generated submodule of $\Sigma \bigoplus P_i$, and if it is a direct summand of $\Sigma \bigoplus P_i$, it is a direct summand of N.

Since M is finitely generated, M is contained in a finite direct sum $\sum_{j=1}^{n} P_j$. If $n = 1$, M is a direct summand of P by hypothesis, and hence a direct summand of $\Sigma \bigoplus P_i$. Now assume any finitely generated submodule of $\sum_{j=1}^{n-1} P_j$ is a direct summand. Let π_n be the projection of $\sum_{j=1}^{n} P_j$ onto P_n. Then $\pi_n(M)$ is a direct summand of P_n, say $P_n = \pi_n(M) \bigoplus K_1$. Ker $\pi_n \cap M$ is a direct summand of M, hence finitely generated. Then by the induction hypothesis, $\sum_{j=1}^{n-1} P_j = (\text{Ker } \pi_n \cap M) \bigoplus K_2$. Then $\sum_{j=1}^{n} P_j = K_1 \bigoplus K_2 \bigoplus M$, so M is a direct summand of $\Sigma \bigoplus P_i$, and hence of N.

REFERENCES

1. M. Auslander, *On regular group rings*, Proc. Amer. Math. Soc. **8** (1957), 658-664.
2. H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton University Press, 1956.
3. K. Morita, *Duality for modules and its applications to the theory of rings with minimum condition*, Tokyo Kyoiku Daigaku **6** (1958), 83-142.
4. B. L. Osofsky, *Rings all of whose finitely generated modules are injective*, Pacific J. Math. **14** (1964), 645-650.
5. A. Rosenberg and D. Zelinsky, *Finiteness of the injective hull*, Math. Zeit. **70** (1959), 372-380.
6. J. von Neumann, *On regular rings*, Proc. Nat. Acad. Sc. (USA) **22** (1936), 707-713.

RUTGERS, THE STATE UNIVERSITY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

RATIO LIMIT THEOREMS FOR MARKOV CHAINS

SIDNEY C. PORT

In an irreducible, recurrent, Markov chain, with integer states, let $N_n(A)$ be the occupation time of A by time n, where A is a finite set of states. Our principal concern in this paper is to investigate various "ratio limit theorem" for $P_x(N_n(A) = k)$. Criteria are given for various ratio limits to exist. The limits (when they exist) are shown to be expressible in terms of an integral over the set of states E completed with its dual recurrent boundary $\hat{B}$. Applications are given to several specific Markov chains.

Throughout this paper $\{X_n\}$ will be an irreducible, recurrent Markov chain with states in a denumerable set E and with nth step transition probabilities P_{xy}^n. For convenience we may take E to be the integers. For discrete time Markov processes, ratio limits for the quantities $P_x(N_n(A) = k)$ were first investigated by Kac [4] for certain special cases of partial sums of independent random variables with a common distribution. Recently these quantities have been intensively studied by Kesten and Spitzer [9] for the irreducible chains formed by the successive partial sums of independent, identically distributed, integer-lattice-valued random vectors in r dimensions. They show the remarkable fact that in *all* such chains, for any two states x, y, any integer $k \geqq 0$, and any finite nonempty set A, the limits

$$\lim_{n\to\infty} \frac{P_x(N_n(A) = k)}{P_y(N_n(\{y\}) = 0)} \tag{1.1}$$

exist and they explicitly find their values.

Now, in general, limits (1.1) exist in very few recurrent chains and we shall have to be content with much weaker types of ratio limits if we want results of any generality. The weakest form of these ratio limits asserts that for any two states x, y, and any finite nonempty set A,

$$\lim_{t\to 1^-} \frac{\sum_{n=0}^{\infty} t^n P_x(N_n(A) = k)}{\sum_{n=0}^{\infty} t^n P_y(N_n(\{y\}) = 0)} \tag{1.2}$$

exists for all $k \geqq 0$. Although limits (1.2) exist in every positive-recurrent chain, there are null-recurrent chains in which these weak

Received September 15, 1964.

limits fail to hold even for sets A having just one point. In § 3 of this paper, we will show that a necessary and sufficient condition for the limits in (1.2) to exist for all finite sets A, arbitrary states x, y, and arbitrary nonnegative integers k, is that for some state, say 0, the limits

$$\lim_{t\to 1^-} \sum_{n=0}^{\infty} t^n (P_{00}^n - P_{x0}^n) \tag{1.3}$$

exist for all states x. We will also show (in § 3) that if the series

$$\sum_{n=0}^{\infty} (P_{00}^n - P_{x0}^n) \tag{1.4}$$

converges for all states x, then we may conclude that the Doeblin-type ratio limits,

$$\lim_{R\to\infty} \frac{\sum_{n=0}^{R} P_x(N_n(A) = k)}{\sum_{n=0}^{R} P_y(N_n(\{y\}) = 0)} \tag{1.5}$$

exist and have the same value as those in (1.2).

In § 4 we establish an interesting representation for these limits by using the boundary theory for recurrent chains. In § 5 we investigate several conditions under which the strong ratio limits (1.1) exist. We conclude the paper, in § 6, with the application of the results of the previous sections to several specific examples.

2. **Notation.** In this section we shall introduce the notation to be used throughout the remainder of the paper.

Letters A, B etc., will denote nonempty subsets of E.

$$V_A = \inf\{k > 0 : X_k \in A\}$$

$$\delta_A(x) = \begin{cases} 1 \text{ if } x \in A, \\ 0 \text{ if } x \notin A, \end{cases}$$

$$N_n(A) = \begin{cases} \sum_{k=1}^{n} \delta_A(X_k) \text{ (the occupation time of } A \text{ by time } n), & n > 0 \\ 0, \quad n = 0 \end{cases}$$

$$F_A^n(x, y) = P_x(V_A = n, X_n = y).$$

As defined above, $F_A^n(x, y) = 0$ if $y \notin A$ and $F_A^0(x, y) = 0$.

$$Q_n(x; A) = P_x(V_A > n) = P_x(N_n(A) = 0),$$

$$\Pi_A(x, y) = \sum_{n=1}^{\infty} F_A^n(x, y),$$

$$P_{xy}^A = \begin{cases} \Pi_A(x, y), & x \in A, \\ 0, & x \notin A; \end{cases}$$

P_{xy}^{A} is the transition matrix of the Markov chain, "restricted to A."

$$H_A(x, y) = \begin{cases} \Pi_A(x, y) , & x \notin A , \\ \delta_{xy} , & x \in A ; \end{cases}$$

$H_A(x, y)$ is the "harmonic measure" of A.

$$g_A(x, y) = \sum_{n=0}^{\infty} P_x(V_A > n, X_n = y) ;$$

$g_A(x, y)$ is the expected number of visits to y among $(X_0, X_1, \cdots)$ before the first entrance into A among $(X_1, X_2, \cdots)$. As defined above, $g_A(x, y) = \delta_{xy}$ if $y \in A$.

$$G_A(x, y) = \begin{cases} g_A(x, y) , & x \notin A , \\ 0 , & x \in A ; \end{cases}$$

$G_A(x, y)$ is the "Green's function" of A.

$$ {}_zP_{xy}^{n} = \begin{cases} P_x(V_{\{z\}} \geqq n, X_n = y) , & n > 0 , \\ \delta_{xy} \text{ if } x \neq z ; \ 0 \text{ if } x = z, n = 0 , \end{cases}$$

$$ {}_zP_{xy}^{*} = \sum_{n=0}^{\infty} {}_zP_{xy}^{n} ,$$

$$ \pi_y = {}_0P_{0y}^{*} .$$

(Recall that π_y is the unique stationary measure of the chain with $\pi_0 = 1$. See [1] Theorem 7 p. 50)

$$K_A(x, y) = g_A(x, y)\pi_y^{-1} ,$$

$$\Pi_A^r(x, y) = \begin{cases} \delta_{xy}, r = 0 , \\ \sum_t \Pi_A(x, t)\Pi_A^{r-1}(t, y), r > 0 , \end{cases}$$

$$C_{xy}^{n} = P_{yy}^{n} - P_{xy}^{n} ,$$

$$\tilde{C}_{xy} = \lim_{t \to 1^-} \sum_{n=0}^{\infty} C_{xy}^{n} t^n ,$$

provided the limit on the right exists.

$$C_{xy}^{*} = \sum_{n=0}^{\infty} C_{xy}^{n} ,$$

provided the series converges.

The dual chain to P is the chain with transition matrix $\hat{P}_{xy} = (\pi_y/\pi_x)P_{yx}$. Quantities which refer to the dual chain will be denoted by $\hat{}$. For example $\hat{G}_A(x, y)$ is the Green's function for A for the dual chain. Finally, for any random variable Z,

$$E_xZ = E(Z \mid X_0 = x) .$$

3. **Weak ratio limits.** We shall commence our investigation with the weak ratio limits for $P_x(V_A > n)$.

THEOREM 3.1. *In order that the limits*

$$\lim_{t\to 1^-} \frac{\sum_{n=0}^{\infty} t^n Q_n(x; A)}{\sum_{n=0}^{\infty} t^n Q_n(y; \{y\})} = M_y(x; A) \tag{3.1}$$

should exist for all states x, y and for all finite nonempty sets A, it is both necessary and sufficient that for some state, say 0, and all states x, that the limit

$$\lim_{t\to 1^-} \sum_{n=0}^{\infty} C_{x0}^n t^n = \tilde{C}_{x0} \tag{3.2}$$

exist. If we further know that the series

$$\sum_{n=0}^{\infty} C_{x0}^n = C_{x0}^* \tag{3.3}$$

converges for all states x, then we may conclude the stronger fact that the Doeblin-type limits

$$\lim_{R\to\infty} \frac{\sum_{n=0}^{R} Q_n(x; A)}{\sum_{n=0}^{R} Q_n(y; \{y\})} = M_y^*(x; A) \tag{3.4}$$

exist for all states x, y and for all finite nonempty sets A. In this latter case $M_y(x; A) = M_y^*(x; A)$. The $M_y(x; A)$ satisfy the following relations:

$$M_y(x; \{z\}) = (\pi_y/\pi_z)[\tilde{C}_{xz} + \delta_{xz}] \tag{3.5}$$

and if $B = A \cup \{z\}$, $z \notin A$,

$$M_y(x; B) = M_y(x; A) - \Pi_B(x, z) M_y(z; A)\ . \tag{3.6}$$

For ease we shall divide the proof into several lemmas.

LEMMA 3.1. *If the limits in* (3.2) *exist for all states x, then for any state y and all states x*

$$\lim_{t\to 1^-} \sum_{n=0}^{\infty} t^n C_{xy}^n = \tilde{C}_{xy} \tag{3.7}$$

exists. Similarly, if series (3.3) *converges for all states x, then for any state y and all states x*

$$\sum_{n=0}^{\infty} C_{xy}^n = C_{xy}^* . \tag{3.8}$$

In this latter case $C_{xy}^* = \tilde{C}_{xy}$.

Proof. It is readily seen that for $n > 0$ and any two states x, y,

$$C_{xy}^n = \sum_{k=1}^{n} C_{x0}^k \, {}_0P_{0y}^{n-k} - \sum_{k=1}^{n} C_{y0}^k \, {}_0P_{0y}^{n-k} + {}_0P_{yy}^n - {}_0P_{xy}^n . \tag{3.9}$$

Since the chain is irreducible, the series

$$\sum_{n=0}^{\infty} {}_0P_{0y}^n, \ \sum_{n=0}^{\infty} {}_0P_{xy}^n, \ \sum_{n=0}^{\infty} {}_0P_{yy}^n ,$$

all converge (See [1], p. 45). Taking generating functions in (3.9) we see that (3.7) follows from (3.2) by Abel's theorem. Similarly, if series (3.3) converges, then the convergence of series (3.8) follows from (3.9) by Merten's theorem ([3], p. 228). Finally, the equality of C_{xy}^* and $\tilde{C}_{xy}$, when the former exists, is a direct consequence of Abel's theorem.

Next we will show that the return times to fixed states are always "weakly" asymptotic.

LEMMA 3.2.[1] *In any irreducible, recurrent Markov chain we have, for any two states z and y,*

$$\lim_{R\to\infty} \frac{\pi_z \sum_{n=0}^{R} Q_n(z; \{z\})}{\pi_y \sum_{n=0}^{R} Q_n(y; \{y\})} = 1 . \tag{3.10}$$

Proof. Assume $y \neq z$. Familiar generating-function relation for the tail of a power series show that for $|t| < 1$,

$$\frac{\sum_{n=0}^{\infty} t^n Q_n(z; \{z\})}{\sum_{n=0}^{\infty} t^n Q_n(y; \{y\})} = \frac{\sum_{n=0}^{\infty} t^n P_{yy}^n}{\sum_{n=0}^{\infty} t^n P_{zz}^n} = R_{yz}(t) . \tag{3.11}$$

Now from a "last entrance decomposition," we see that

$$P_{yy}^n = {}_zP_{yy}^n + \sum_{k=0}^{n} P_{yz}^k \, {}_zP_{zy}^{n-k} ,$$

and thus

$$\sum_{n=0}^{\infty} t^n P_{yy}^n = \sum_{n=0}^{\infty} t^n \, {}_zP_{yy}^n + \sum_{n=0}^{\infty} t^n P_{yz}^n \sum_{n=0}^{\infty} t^n \, {}_zP_{zy}^n .$$

[1] Added in proof. Under the assumptions of the lemma, we also have that

$$\sum_{n=0}^{R} \sum_{x\in B} \pi(x) P_x(N_n(B) = k) \sim \sum_{n=0}^{R} \pi(y) P_y(N_n(\{y\}) = 0)$$

for any integer $k \geqq 0$ and any finite nonempty set B.

Using the above relation and well-known generating-function relations (see [1], p. 53 eqns. (1) and (2)), we see that $R_{yz}(t)$ may be written as

$$R_{yz}(t) = \sum_{n=0}^{\infty} t^n {}_zP_{yy}^n \left[1 - \sum_{n=1}^{\infty} F_{\{z\}}^n(z, z)t^n\right] + \sum_{n=1}^{\infty} F_{\{z\}}^n(y, z)t^n \sum_{n=0}^{\infty} t^n {}_zP_{zy}^n . \tag{3.12}$$

Now $R_{yz}(t)$ is a power series in t which converges at $t = 1$ to the value π_y/π_z. To see this, observe that the second term on the right in (3.12) is the product of two power series with positive coefficients which converge at $t = 1$ to the values 1 and ${}_zP_{zy}^*$, respectively. Thus the product series converges at $t = 1$ to ${}_zP_{zy}^*$. On the other hand, the first term on the right in (3.12) is the product of a power series with positive coefficients which converges at $t = 1$ to ${}_zP_{yy}^*$, and a power series which converges at $t = 1$ to 0. Thus Merten's theorem [*op. cit.*] implies that the product series converges at 1 to 0. We have thus shown that $R_{yz}(1) = {}_zP_{zy}^*$. But by [1] (Corollary 1, p. 49, and Theorem 7, p. 50) we have ${}_zP_{zy}^* = \pi_y/\pi_z$.

Denote the coefficient of t^n in $R_{yz}(t)$ by r_{yz}^n. Then from (3.11) we easily obtain

$$\sum_{n=0}^{R} Q_n(z; \{z\}) = \sum_{n=0}^{R} Q_n(y; \{y\}) \sum_{j=0}^{R-n} r_{yz}^j . \tag{3.13}$$

Since $\sum_{j=0}^{\infty} r_{yz}^j$ converges, and since

$$\lim_{R\to\infty} \frac{Q_R(y, \{y\})}{\sum_{n=0}^{R} Q_n(y; \{y\})} = 0 ,$$

we have, by a well-known Abelian theorem on Norlünd summability (see [3; p. 64[),

$$\lim_{R\to\infty} \frac{\sum_{n=0}^{R} Q_n(z; \{z\})}{\sum_{n=0}^{R} Q_n(y; \{y\})} = \sum_{n=0}^{\infty} r_{yz}^n = \pi_y/\pi_z . \tag{3.14}$$

This completes the proof.

As a corollary of the proof we have the following:

COROLLARY 3.1. *In any irreducible, recurrent Markov chain, the series* $R_{yz}(t)$ *defined in* (3.11) *converges at* $t = 1$ *to* π_y/π_z.

We of course have that the weaker, Abelian version, is also universally valid.

COROLLARY 3.2. *In any irreducible, recurrent Markov chain,*

$$\lim_{t\to 1^-} \frac{\sum_{n=0}^{\infty} Q_n(z;\{z\})t^n}{\sum_{n=0}^{\infty} Q_n(y;\{y\})t^n} = \pi_y/\pi_z \,. \tag{3.15}$$

Proof. This follows directly from Corollary 3.1 (by Abel's theorem). We are now in a position to complete the proof of Theorem 3.1.

Proof of Theorem 3.1. We shall proceed by induction on the number of points in A. Let us first assume that the series (3.3) converges for all states x and show that this leads to the conclusion that the limits in (3.4) exist for all states x, y and for all finite non-empty sets A. Suppose $A = \{y\}$. If $x = y$, then obviously the limit in (3.4) exists and has the value 1; so assume that $x \neq y$. Since for $x \neq y$ we have

$$\sum_{n=0}^{\infty} Q_n(x, y)t^n = (1-t)^{-1}\Big[1 - \sum_{n=1}^{\infty} t^n F^n_{\{y\}}(x, y)\Big] = (1-t)^{-1}\left[1 - \frac{\sum_{n=0}^{\infty} t^n P^n_{xy}}{\sum_{n=0}^{\infty} t^n P^n_{yy}}\right]$$

we see that for $x \neq y$,

$$\sum_{n=0}^{\infty} t^n Q_n(x;\{y\}) = \sum_{k=0}^{\infty} C^k_{xy} t^k \sum_{n=0}^{\infty} t^n Q_n(y;\{y\}) \,, \tag{3.16}$$

and thus

$$Q_n(x;\{y\}) = \sum_{k=0}^{n} C^k_{xy}\, Q_{n-k}(y;\{y\}) \,. \tag{3.17}$$

Consequently,

$$\frac{\sum_{n=0}^{R} Q_n(x;\{y\})}{\sum_{n=0}^{R} Q_n(y;\{y\})} = \frac{\sum_{n=0}^{R} Q_n(y;\{y\}) \sum_{j=0}^{R-n} C^j_{xy}}{\sum_{n=0}^{R} Q_n(y;\{y\})} \,. \tag{3.18}$$

By Lemma 3.1,

$$\lim_{R\to\infty} \sum_{j=0}^{R} C^j_{xy} = C^*_{xy} \,,$$

and thus, by the same Abelian theorem as used in the proof of the previous lemma, we see that the limit as $R \to \infty$ in (3.18) exists and has the value C^*_{xy}.

From Lemma 3.2, we then have

$$\lim_{R\to\infty} \frac{\sum_{n=0}^{R} Q_n(x, \{z\})}{\sum_{n=0}^{R} Q_n(y; \{y\})} = (\pi_y/\pi_z)[C^*_{xz} + \delta_{xz}] ,$$

which establishes the existence of the limits in (3.4) for all sets A having exactly one point. Suppose we have established the existence of these limits for all sets A having exactly $r > 0$ points. Let $B = A \cup \{z\}$, where $z \notin A$ but is otherwise arbitrary. It is readily seen that

$$Q_n(x; B) = Q_n(x; A) - \sum_{k=1}^{n} F^k_B(x, z) Q_{n-k}(z; A) , \tag{3.19}$$

and thus

$$\frac{\sum_{n=0}^{R} Q_n(x; B)}{\sum_{n=0}^{R} Q_n(y; \{y\})} = \frac{\sum_{n=0}^{R} Q_n(x; A)}{\sum_{n=0}^{R} Q_n(y; \{y\})} - \frac{\sum_{n=0}^{R} Q_n(z; A) \sum_{j=1}^{R-n} F^j_B(x, z)}{\sum_{n=0}^{R} Q_n(y; \{y\})} . \tag{3.20}$$

By the induction assumption, the first term on the right converges to $M^*_y(x; A)$. If we multiply and divide the second term on the right by $\sum_{n=0}^{R} Q_n(z; A)$, and apply the Abelian theorem mentioned above, then we may conclude that the second term on the right in (3.20) converges to $\Pi_B(x, z)M^*_y(z; A)$. Thus the limit in (3.20) exists and has value $M^*_y(x; A) - \Pi_B(x, z)M^*_y(z; A)$. By induction, we then have that the limit in (3.4) exists for all finite sets A, and moreover, that relation (3.6) holds.

Now let us assume that the limits in (3.2) exist for all states x, and show this leads to the conclusion that the limits (3.1) exist as required.

From (3.16), Lemma 3.1, and Corollary 3.2 we have at once that the limits in (3.1) exist for all sets A having a single point, and moreover, that $M_y(x, \{z\}) = (\pi_y/\pi_z)[\tilde{C}_{xz} + \delta_{xz}]$. Suppose we have established the existence of the limits in (3.1) for all sets A having exactly $r > 0$ points. Again let $B = A \cup \{z\}, z \notin A$. From (3.19) we have

$$\frac{\sum_{n=0}^{\infty} Q_n(x; B)t^n}{\sum_{n=0}^{\infty} Q_n(y; \{y\})t^n} = \frac{\sum_{n=0}^{\infty} Q_n(x; A)t^n}{\sum_{n=0}^{\infty} Q_n(y; \{y\})t^n} - \frac{\sum_{k=1}^{\infty} F^k_B(x, z)t^n \sum_{n=0}^{\infty} Q_n(z; A)t^n}{\sum_{n=0}^{\infty} Q_n(y; \{y\})t^n}$$

and thus, by the induction assumption and Abel's theorem, we have that the limit on the right, as $t \to 1^-$, exists, and has the value $M_y(x; A) - \Pi_B(x, z)M_y(z; A)$. Hence, by induction, we have that the limits in (3.1) exist for all finite sets A. This establishes the sufficiency portion of Theorem 3.1.

Now suppose we know that the limits in (3.1) exist for all finite sets A and for all states x, y; then, in particular, they exist for $A = \{0\}$. From (3.16) we see at once that this implies the existence of the limits in (3.2) for all states x, and thus condition (3.2) is necessary.

Finally, a simple Abelian-type argument shows that $M_y(x; A) = M_y^*(x; A)$ whenever the latter exists. This completes the proof of Theorem 3.1.

Before proceeding further let us make some comments on the preceding results.

Whenever series (3.3) converges for all states x, then as noted in the proof of the above theorem, Abel's theorem gives us that the limits in (3.2) exist for all states x. On the other hand, if the limits in (3.2) exist for all states x, then series (3.3) may fail to converge but we do have that it is at least $(C; 1)$ summable.[2] To see this, observe that if

$$S_{xy}^n = \sum_{k=0}^{n-1} C_{xy}^k , \tag{3.21}$$

then, for $x \neq y$,

$$S_{xy}^n = \sum_{r=1}^{n} F_{\{y\}}^r(x, y)[E_y N_n(\{y\}) - E_y N_{n-r}(\{y\})] + Q_n(x; \{y\}) E_y N_n(\{y\}) ,$$

and thus $S_{xy}^n \geqq 0$. Consequently, whenever the limits

$$\lim_{t \to 1^-} \sum_{n=0}^{\infty} C_{xy}^n t^n = \tilde{C}_{xy}$$

exist, we have by a well-known Tauberian theorem (see [3], p. 154),

$$\lim_{n \to \infty} n^{-1} \sum_{r=0}^{n-1} S_{xy}^r = \tilde{C}_{xy} .$$

If the chain is positive recurrent, then $E_x V_A < \infty$ for all nonempty sets A. But $\sum_{n=0}^{\infty} Q_n(x; A) = E_x V_A$, and thus the limits (3.4) exist in *all* positive recurrent chains. Moreover, in this case, $M_y(x; A) = (E_y V_{\{y\}})^{-1} E_x V_A$. Since series (3.3) need not converge in a periodic positive recurrent chain, we see that this condition is not necessary for the existence of the limits in (3.4) (at least in the positive recurrent case). In this regard let us point out that the series (3.3) does converge for all states x in every *aperiodic* positive recurrent chain (see [6])

Under the same conditions as Theorem 3.1, the following extension also holds.

THEOREM 3.2. *If the limits* (3.2) *exist for all states x, then for any nonempty finite set A, any two states x, y, and any nonnegative*

[2] This fact will be needed in § 4.

integer k,

$$\lim_{t\to 1^-} \frac{\sum_{n=0}^{\infty} P_x(N_n(A) = k)t^n}{\sum_{n=0}^{\infty} P_y(N_n(\{y\}) = 0)t^n} = \sum_z \Pi_A^k(x, z)M_y(z; A) . \tag{3.22}$$

Moreover, if series (3.3) *converges for all states* x, *then*

$$\lim_{R\to\infty} \frac{\sum_{n=0}^{R} P_x(N_n(A) = k)}{\sum_{n=0}^{\infty} P_y(N_n(\{y\}) = 0)} = \sum_z \Pi_A^k(x, z)M_y(z; A) . \tag{3.23}$$

Proof. For brevity we shall only prove the assertion in (3.23). The proof of (3.22) is very similar. We proceed by induction on k. For $k = 0$ the assertion in (3.23) is just that of Theorem 3.1, and thus (3.23) holds for $k = 0$.

Suppose we have established the result for all $k \leqq k_0$. Now,

$$P_x(N_n(A) = k_0 + 1) = \sum_z \sum_{j=1}^{n} F_A^j(x, z)P_z(N_{n-j}(A) = k_0) , \tag{3.24}$$

and thus

$$\frac{\sum_{n=0}^{R} P_x(N_n(A) = k_0 + 1)}{\sum_{n=0}^{R} P_y(N_n(\{y\}) = 0)} = \sum_{z\in A} \frac{\sum_{n=0}^{R} P_z(N_n(A) = k_0) \sum_{j=1}^{R-n} F_A^j(x, z)}{\sum_{n=0}^{R} P_y(N_n(\{y\}) = 0)} . \tag{3.25}$$

Since we may write each term on the right in the above expression as

$$\frac{\sum_{n=0}^{R} P_z(N_n(A) = k_0) \sum_{j=1}^{R-n} F_A^j(x, z)}{\sum_{n=0}^{R} P_z(N_n(A) = k_0)} \cdot \frac{\sum_{n=0}^{R} P_z(N_n(A) = k_0)}{\sum_{n=0}^{R} P_y(N_n(\{y\}) = 0)} ,$$

and since $\sum_{j=1}^{\infty} F_A^j(x, z) = \Pi_A(x, z)$,

$$\lim_{R\to\infty} \frac{P_z(N_R(A) = k_0)}{\sum_{n=0}^{R} P_z(N_n(A) = k_0)} = 0 ,$$

we see (by the induction assumption and the same Abelian theorem as used in the previous proofs) that the limit, as $R \to \infty$, in (3.25) exists and has the value

$$\sum_{z\in A} \Pi_A(x, z) \sum_{u\in A} \Pi_A^{k_0}(z, u)M_y(u; A) = \sum_{u\in A} \Pi_A^{k_0+1}(x, u)M_y(u; A) .$$

Thus, the assertion of the theorem is true for $k = k_0 + 1$, and thus,

by induction, the theorem holds for all values of k. This completes the proof.

Observe that in the special case in which $A = \{y\}$ and $k > 0$, the limits in (3.23) have the value 1. For this special case we can show, by the same methods as used to establish the result in general, that this result holds in *every* recurrent chain.

The same arguments we used to establish the above results enable us to show that limits of the more general expressions

$$\frac{\sum_{n=0}^{R} P_x(N_n(\{y_i\}) = k_i,\ 1 \leqq i \leqq r)}{\sum_{n=0}^{R} P_y(N_n(\{y\}) = 0)}$$

exist, and to compute their values. Since these results are quite complicated to write down, however, we shall not pursue these generalizations.

4. **Representation.** By using the boundary theory for recurrent chains, as developed by Kemeny and Snell in [8], we may establish interesting representations for the limits found in the last section. For convenience, we shall summarize below that portion of this theory which we shall need. For details we refer the reader to [8].

We shall be interested in the dual, i.e., exit boundary, of the chain. If we choose a state y as a "taboo" state, then a boundary point ξ corresponds to a sequence $\{t_n\}$ of states such that $|t_n| \to \infty$ and the limits

$$\lim_{n\to\infty} K_{\{y\}}(x, t_n)$$

exist for all states x. The set of boundary points, $\hat{B}$, so obtained does not depend on which state y is chosen as the taboo state, and $E^* = E \cup \hat{B}$ is a compact metric space which gives a discrete topology to E. For each x and A, the functions $K_A(x, \cdot)$, $\hat{G}_A(\cdot, x)$, and $\hat{H}_A(\cdot, x)$ can be extended to E^* as continuous functions. For a fixed y, let ${}_yQ_{xt} = P_{xt}$ if $t \neq y$, and let ${}_yQ_{xy} = 0$. Then ${}_yQ_{xt}$ is the substochastic transition matrix of a transient chain (the yth associated transient chain). A function f is called superregular for ${}_yQ$ if ${}_yQf(x) \leqq f(x)$. If ${}_yQf = f$, then f is called regular for ${}_yQ$. A nonnegative regular function is minimal if every nonnegative regular function $g \leqq f$ is a multiple of f. A point $\xi \in \hat{B}$ is called minimal if $K_{\{y\}}(\cdot, \xi)$ is a minimal regular function for ${}_yQ$. (Again, the minimal points of $\hat{B}$ can be shown to be independent of y.) If h is a nonnegative superregular function for ${}_yQ$, then the transient chain with transition matrix, ${}_yQ_{xt}h(t)h(x)^{-1}$ is called the h-process for ${}_yQ$. The following representation theorem follows by applying the results in § 5 of [8] to the measure $\nu(x) = f(x)\pi_x\pi_y^{-1}$ on the dual chain.

THEOREM 4.1. *Suppose* (i) $f(x) \geqq 0$, (ii)$f(y) = 0$, *and* (iii) $Pf(x) \leqq f(x) + \delta_{xy}$. *Then, there is a unique probability measure* Γ_f, *vanishing on the nonminimal points of* $\hat{B}$, *such that*

$$f(x) + \delta_{xy} = \pi_y \int_{E^*} K_{\{y\}}(x, \xi)\Gamma_f(d\xi) .$$

Finally, if $h(x) = f(x) + \delta_{xy}$, then h is a superregular function for ${}_yQ$. Let $\theta^A(x)$ be the probability that, starting from y, this h process last visits the set A at the point x, ($\theta^A(x)$ depends on y and f). Then for $x \in A$, by results in [8], we have

$$\sum_{t \in A} (I - P^A)_{xt} f(t) = [\theta^A(x) - \delta_{xy}]\pi_y/\pi_x . \tag{4.1}$$

Moreover, if $\{A_n\}$ is a fundamental sequence of sets, i.e., the A_n are monotone increasing and $\cup_n A_n = E$, then for any continuous function g on E^*,

$$\lim_{n \to \infty} \sum_x \theta^{A_n}(x) g(x) = \int_{E^*} g(\xi)\Gamma_f(d\xi) . \tag{4.2}$$

That is, the θ^{A_n} converge weakly to Γ_f.

We are now in a position to establish the following result.

THEOREM 4.2. *Suppose the limits in* (3.2) *exist for all states* x. *Then, for any finite nonempty set* A, *and any two states* x, y *we have, for* $x \notin A$,

$$M_y(x; A) = (\pi_y/\pi_x) \int_{E^*} \hat{G}_A(\xi, x)\beta(d\xi) , \tag{4.3}$$

while for $x \in A$,

$$M_y(x; A) = (\pi_y/\pi_x) \int_{E^*} \hat{H}_A(\xi, x)\beta(d\xi) , \tag{4.4}$$

where β *is a unique probability measure on* E^* *which vanishes on the nonminimal boundary points.*

We shall first establish the desired representation for all sets A having exactly one point. This will be an immediate consequence of (3.5) and the following lemma.

LEMMA 4.1. *For each* y *and all states* x,

$$\tilde{C}_{xy} = (\pi_y/\pi_x) \int_{E^*} \hat{G}_{\{y\}}(\xi, x)\beta(d\xi) , \tag{4.5}$$

where β is a unique probability measure on E^* which vanishes on the nonminimal boundary points.

Note. In [8] a corresponding representation was shown to hold for C^*_{xy}. In the proof given below to establish (4.5), we shall use techniques similar to those used in [8] to establish the result for C^*_{xy}.

Proof. In the remarks following the proof of Theorem 3.1, we showed that

$$\lim_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} S^k_{xy} = \tilde{C}_{xy} \,. \tag{4.6}$$

From Fatou's lemma, we then have

$$\sum_t P_{xt}\tilde{C}_{ty} \leqq \liminf_{n\to\infty} n^{-1} \sum_{k=0}^{n-1} \sum_t P_{xt}S^k_{ty} \leqq \tilde{C}_{xy} + \delta_{xy} \,.$$

Consequently, for each fixed y, we see that $\tilde{C}_{xy}$ satisfies conditions (i)–(iii) of Theorem 4.1, and thus there is a unique probability measure $\beta_{\{y\}}$, vanishing on the nonminimal boundary points, such that

$$\tilde{C}_{xy} + \delta_{xy} = \pi_y \int_{E^*} K_{\{y\}}(x, \xi)\beta_{\{y\}}(d\xi) \,.$$

But since for each $\xi \in E^*$,

$$K_{\{y\}}(x, \xi) = [\hat{G}_{\{y\}}(\xi, x) + \delta_{xy}]\pi_x^{-1} \,, \tag{4.7}$$

we have

$$\tilde{C}_{xy} = (\pi_y/\pi_x) \int_{E^*} \hat{G}_{\{y\}}(\xi, x)\beta_{\{y\}}(d\xi) \,.$$

To complete the proof we must now show that the measure $\beta_{\{y\}}$ is independent of y. In order to do this we may proceed as follows. Let A be any finite set containing x; then,

$$\sum_t \hat{P}_{yt}\hat{H}_A(t, x) = \hat{H}_A(y, x) + \hat{P}^A_{yx} - \delta_{yx}, \; x \in A \,.$$

If we iterate the above relation n times, we obtain the identity

$$\sum_t \hat{P}^{n+1}_{yt}\hat{H}_A(t, x) = \hat{H}_A(y, x) + \sum_{t\in A} \hat{N}^n_{yt}[\hat{P}^A - I]_{tx} \,,$$

where $\hat{N}^n_{yt} = (I + \hat{P} + \cdots + \hat{P}^n)_{yt}$. Now, $\sum_t \pi_t\hat{P}^A_{tx} = \pi_x$, and thus we may rewrite the above identity as

$$\sum_t \hat{P}^{n+1}_{yt}\hat{H}_A(t, x) = \hat{H}_A(y, x) + \sum_{t\in A} \left[\left(\frac{\pi_t}{\pi_y}\right)\hat{N}^n_{yy} - \hat{N}^n_{yt}\right][I - \hat{P}^A]_{tx} \,. \tag{4.8}$$

But

$$(\pi_t/\pi_y)\hat{N}^n_{yy} - \hat{N}^n_{yt} = (\pi_t/\pi_y)S^n_{ty} .$$

Using this identity we obtain from (4.8) the identity

$$\sum_t \frac{(\hat{N}^{n+1} - I)_{yt}H_A(t, x)}{n+1} = \hat{H}_A(y, x) + (\pi_x/\pi_y)\sum_{t\in A}(I - P^A)_{xt}\left[\frac{1}{n+1}\sum_{k=0}^{n} S^k_{ty}\right] .$$

Consequently, from (4.6) we have that the limit, as $n \to \infty$, on the right-hand side of the above expression exists and has the value

$$\hat{H}_A(y, x) + (\pi_x/\pi_y)\sum_{t\in A}(I - P^A)_{xt}\tilde{C}_{ty} .$$

For a fixed x and A, let

$$\varphi_n(y) = \sum_t \frac{\hat{N}^{n+1}_{yt}\hat{H}_A(t, x)}{n+1} .$$

From the above, we then have that $\lim_{n\to\infty}\varphi_n(y) = \varphi(y)$ exists. But

$$\hat{N}^{n+1} - I = \hat{P}\hat{N}^n ,$$

and thus

$$\varphi_n(y) = \sum_t \hat{P}_{yt}\varphi_{n-1}(t)n/n + 1 + \hat{H}_A(y, x)(n+1)^{-1} .$$

By dominated convergence, we then have that $\varphi(y)$ satisfies the relation $\hat{P}\varphi(x) = \varphi(x)$, and as the $\hat{P}$ chain is recurrent, we must have that $\varphi(x)$ is a constant (independent of y).[3] Denote this constant by $\tilde{\lambda}_A(x)$. We have thus established the following identity:

$$[\tilde{\lambda}_A(x) - \hat{H}_A(y, x)]\pi_y/\pi_x = \sum_{t\in A}(I - P^A)_{xt}\tilde{C}_{ty},\ x \in A ,$$

and, in particular, for $y \in A$ we have

$$[\tilde{\lambda}_A(x) - \delta_{yx}]\pi_y/\pi_x = \sum_{t\in A}(I - P^A)_{xt}\tilde{C}_{ty},\ x, y \in A . \tag{4.9}$$

However, from (4.1), for any finite set A containing x,

$$\sum_{t\in A}(I - P^A)_{xt}\tilde{C}_{ty} = (\pi_y/\pi_x)[\theta^A(x) - \delta_{xy}] , \tag{4.10}$$

where $\theta_A(x)$ is the probability that, starting from y, x is the last state to be visited in the set A by the h process for ${}_yQ$ determined by the function $\tilde{C}_{\cdot y} + \delta_{\cdot y} = h(\cdot)$. From (4.9) and (4.10) we see that if $y \in A$,

[3] See Theorem 3 on p. 226 of [6].

then $\theta^A(x) = \tilde{\lambda}_A(x)$. Consequently, (by (4.2)) for any fundamental sequence of sets $\{A_n\}$, we have that $\{\tilde{\lambda}_{A_n}\}$ converges weakly to $\beta_{\{y\}}$, and as the $\tilde{\lambda}_{A_n}$ are independent of y, we must have that $\beta_{\{y\}}$ is too. This completes the proof.

We may now complete the proof of Theorem 4.2. First, observe that from (4.7), we see that (4.3) and (4.4) for the set $A = \{z\}$ are equivalent to the single relation,

$$M_y(x; \{z\}) = \pi_y \int_{E^*} K_{\{z\}}(x, \xi)\beta(d\xi) . \tag{4.11}$$

Now suppose we have established the relation

$$M_y(x; A) = \pi_y \int_{E^*} K_A(x, \xi)\beta(d\xi) , \tag{4.12}$$

for all sets A having exactly $r > 0$ points. Let $B = A \cup \{z\}$, where $z \notin A$ but is otherwise arbitrary. From the relation

$$P_x(V_B > n, X_n = t) = P_x(V_A > n, X_n = t) - \sum_{k=1}^{n} F_B^k(x, z)P_z(V_A > n - k, X_{n-k} = t) ,$$

we obtain the identity

$$g_B(x, y) = g_A(x, y) - \Pi_B(x, z)g_A(z, y) ,$$

and thus for each $\xi \in E^*$,

$$K_B(x, \xi) = K_A(x, \xi) - \Pi_B(x, z)K_A(z, \xi) .$$

Consequently, by the induction assumption and equation (3.6),

$$M_y(x; B) = \pi_y \int_{E^*} K_B(x, \xi)\beta(d\xi) ,$$

which establishes (4.12) for the set B. By induction, we then have that (4.12) holds for all finite nonempty sets.

If $x \notin A$ we have that $g_A(x, t) = G_A(x, t)$, and thus for $x \notin A$, (4.12) becomes (4.3). On the other hand, for any x, t we have

$$g_A(x, t) = \sum_z P_{xz}G_A(z, t) + \delta_{xt} ,$$

and thus

$$K_A(x, t) = \left[\sum_z \hat{G}_A(t, z)\hat{P}_{zx} + \delta_{xt}\right]\pi_x^{-1} .$$

But if $x \in A$, the expression in braces is $\hat{H}_A(t, x)$. Consequently, from (4.12) and the above we obtain (4.4). This completes the proof of

Theorem 4.2.

From the proof of the above theorem we have the following:

COROLLARY 4.1. *For any two states x, y and any finite nonempty set A,*

$$M_y(x; A) = \pi_y \int_{E^*} K_A(x, \xi)\beta(d\xi) , \tag{4.13}$$

provided the limits in (3.2) *exist for all states x.*

COROLLARY 4.2. *Under the same conditions as Theorem* 4.2 *we have $M_y(x; A)\pi_y^{-1}$ is independent of y. Moreover*

$$\sum_{x \in A} \pi_x M_y(x; A) = \pi_y . \tag{4.14}$$

Let us see what the representation in Theorem 4.2 becomes in a positive recurrent chain. If the chain is positive recurrent, then it is readily seen that the representation in Corollary 4.1 is equivalent to the identity

$$(\pi_y E_y V_{\{y\}})^{-1} E_x V_A = \sum_t K_A(x, t)\beta(\{t\}) + \int_{\hat{B}} K_A(x, \xi)\beta(d\xi) .$$

However, it is a well-known and easily verifiable fact that

$$E_x V_A = \sum_t g_A(x, t) = m^{-1} \sum_t K_A(x, t)(m\pi_t) , \tag{4.15}$$

where $m = (\sum_t \pi_t)^{-1}$. By the uniqueness of β we must then have

$$\beta(\hat{B}) = 0, \text{ and } \beta(\{t\}) = m\pi_t .$$

From the above results on the positive recurrent case we see that we may view the resuls of Theorem 4.2 as an extension of the identity (4.15) to those null chains for which the limits in (3.2) exist for all states x.

5. **Strong ratio limits.** Let us now turn our attention to the problem of the strong ratio limits. It is quite easy to give examples of aperiodic recurrent chains (both positive and null) in which series (3.3) converges for all states x (and thus the weak ratio limits (3.4) exist), but in which the strong ratio limits fail to exist. The following is a sufficient condition for null recurrent chains which is frequently satisfied.

THEOREM 5.1. *If the limits* (3.2) *exist for all states x and if for some state, say z_0, we have*

$$\sum_{n=0}^{\infty} t^n Q_n(z_0; \{z_0\}) \sim (1-t)^{-\alpha} L\Big(\frac{1}{1-t}\Big), \; t \to 1^- , \tag{5.1}$$

where $\alpha > 0$ *and* $L(u)$ *is a slowly varying function, then for any state* y,

$$Q_n(y; \{y\}) \sim (\pi_{z_0}/\pi_y)\Gamma(\alpha)^{-1} n^{\alpha-1} L(n) , \tag{5.2}$$

and for any finite nonempty set A, *arbitrary states* x, y, *and arbitrary nonnegative integer* k,

$$\lim_{n\to\infty} \frac{P_x(N_n(A) = k)}{P_y(N_n(\{y\}) = 0)} = \sum_z \Pi_A^k(x, z) M_y(z; A) . \tag{5.3}$$

Proof. We know (see (3.15)) that

$$\lim_{t\to 1^-} \frac{\sum_{n=0}^{\infty} t^n Q_n(y; \{y\})}{\sum_{n=0}^{\infty} t^n Q_n(z_0; \{z_0\})} = \pi_{z_0}/\pi_y ,$$

and thus, if (5.1) holds, we have

$$\sum_{n=0}^{\infty} t^n Q_n(y; \{y\}) \sim (1-t)^{-\alpha} L\Big(\frac{1}{1-t}\Big)\pi_{z_0}/\pi_y, \; t \to 1^- .$$

Since $Q_n(y; \{y\})$ is nonnegative, by Karamata's Tauberian theorem ([3], p. 154) we have

$$\sum_{k=0}^{n} Q_k(y; \{y\}) \sim (\pi_{z_0}/\pi_y)\frac{n^\alpha L(n)}{\Gamma(\alpha+1)}, \; n \to \infty , \tag{5.4}$$

and since $Q_n(y; \{y\})$ is monotone, another Tauberian theorem ([2], p. 517) asserts that from (5.4) we may conclude that (5.2) holds. Now if the limits in (3.2) exist for all x, then by Theorem 3.2 and (5.1) we have

$$\sum_{n=0}^{\infty} t^n P_x(N_n(A) = k) \sim \pi_y^{-1}(1-t)^{-\alpha} L\Big(\frac{1}{1-t}\Big)\pi_{z_0} \sum_z \Pi_A^k(x, z) M_y(z; A) ,$$

and thus,

$$\sum_{n=0}^{\infty} t^n P_x(N_n(A) \leqq k) \sim \pi_y^{-1}(1-t)^{-\alpha} L\Big(\frac{1}{1-t}\Big)\pi_{z_0} \sum_z \sum_{j=0}^{k} \Pi_A^j(x, z) M_y(z; A) .$$

Since $N_n(A)$ is nondecreasing, $P_x(N_n(A) \leqq k)$ is nonincreasing in n for each fixed k. Consequently, by applications of the Tauberian theorems mentioned above, we may conclude

$$P_x(N_n(A) \leqq k) \sim \pi_y^{-1}\Gamma(\alpha)^{-1} n^{\alpha-1} L(n)\pi_{z_0} \sum_z \sum_{j=0}^{k} \Pi_A^j(x, z) M_y(z; A) .$$

Thus we must have

$$P_x(N_n(A) = k) \sim \pi_y^{-1}\Gamma(\alpha)^{-1}n^{\alpha-1}L(n)\pi_{z_0}\sum_z \Pi_A^k(x, z)M_y(z; A) .$$

Now (5.3) follows from (5.2) and the above expression.

REMARK. Again observe that in the special case when $A = \{y\}$, the limit value is 1 for all $k > 0$. If we trace through the above proof for this special case, we will see that we never use the assumption that the limits in (3.2) exist. Thus we have

COROLLARY 5.1. *If* (5.1) *holds for some state* z, *then* (5.2) *holds for any state* y, *and moreover, for any integer* $k > 0$,

$$\lim_{n\to\infty} \frac{P_x(N_n(\{y\}) = k)}{P_y(N_n(\{y\}) = 0)} = 1 . \tag{5.5}$$

Observe that whenever (5.5) holds, we have the curious result that for $0 \leqq j \leqq k$.

$$\lim_{n\to\infty} P_x(N_n(\{y\}) = j \mid N_n(\{y\}) \leqq k) = (k + 1)^{-1} . \tag{5.6}$$

In the case when the dual boundary has only one point we may establish the existence of the strong ratio limits for one point sets with less assumptions. In fact we have the following.

THEOREM 5.2. *Suppose that the dual boundary of the* P *chain has only one point and* $Q_n(y; \{y\}) > 0$ *for all* n. *Then for each fixed* m,

$$\lim_{n\to\infty} \frac{Q_{n+m}(x, \{y\})}{Q_n(y; \{y\})} = \begin{cases} (\pi_y/\pi_x)\mu_{xy}, & x \neq y , \\ 1, & x = y , \end{cases} \tag{5.7}$$

if and only if

$$\lim_{n\to\infty} \frac{Q_{n+1}(y; \{y\})}{Q_n(y; \{y\})} = 1 .^4 \tag{5.8}$$

Proof. Clearly if (5.7) holds, then setting $y = x$ and $m = 1$, we obtain (5.8). To establish the converse, we need only consider the case $m = 0$. If $x = y$, there is nothing to prove; so from now on assume $x \neq y$. Now we have

$$Q_n(x, \{y\}) = \sum_{r=0}^{\infty} F_{\{y\}}^{r+1+n}(x, \{y\}) = \sum_t G_{\{y\}}(x, t)\, {}_yP_{ty}^{n+1} ,$$

[4] If the chain is null recurrent, then $Q_n(y, \{y\}) > 0$ for all n. (See the proof of Theorem 5.3). The author has constructed an example of a recurrent chain in which the dual boundary has exactly 2 points and such that (5.8) holds, but in which the limits in (5.7) *fail* to hold. However, in this example it is not known if condition (3.2) holds, so that the weak limits may also not exist.

and thus,

$$\frac{Q_n(x, \{y\})}{Q_n(y; \{y\})} = \frac{Q_{n+1}(y; \{y\})}{Q_n(y, \{y\})} \sum_t \hat{G}_{\{y\}}(t, x) R_n(t)\, \pi_y/\pi_x \,, \tag{5.9}$$

where

$$R_n(t) = \frac{\hat{P}_{yt}^{n+1}}{\hat{Q}_{n+1}(y; \{y\})} \,. \tag{5.10}$$

Now for $t \neq y$,

$$R_n(t) = P_y(\hat{X}_{n+1} = t \mid \hat{V}_{\{y\}} > n + 1) \,,$$

and thus

$$R_n(t) \geqq 0 \text{ and } \sum_{t \neq y} R_n(t) \doteq 1 \,. \tag{5.11}$$

Since the $\hat{P}$ chain is irreducible, we have, for some m_0, that $\hat{P}_{ty}^{m_0} = \alpha > 0$, and as

$$P_y(n < \hat{V}_{\{y\}} \leqq n + m_0 + 1) \geqq P_y(\hat{V}_{\{y\}} > n + 1, \hat{X}_{n+1} = t)\hat{P}_{ty}^{m_0} \,,$$

we have,

$$\frac{P_y(\hat{V}_{\{y\}} > n + 1, \hat{X}_{n+1} = t)}{\hat{Q}_{n+1}(y; \{y\})} \leqq (1/\alpha)\left[\frac{Q_n(y; \{y\})}{Q_{n+1}(y; \{y\})} - \frac{Q_{n+m_0+1}(y; \{y\})}{Q_{n+1}(y; \{y\})}\right] .$$

From (5.8) and the above inequality, we then have, for each fixed t,

$$\lim_{n \to \infty} R_n(t) = 0 \,. \tag{5.12}$$

The assumption that the dual boundary has only one point is equivalent to the fact that the limit

$$\lim_{|t| \to \infty} \hat{G}_{\{y\}}(t, x) = \mu_{xy} \tag{5.13}$$

exists. From (5.8) and (5.9), we see that to establish (5.7) it is only necessary to show

$$\lim_{n \to \infty} \sum_t \hat{G}_{\{y\}}(t, x) R_n(t) = \mu_{xy} \,.$$

However, this is an easy consequence of (5.11), (5.12), and (5.13).

Although the condititon that the dual boundary have only one point is very restrictive, we shall see in § 6 that it applies to some of the important chains of practical interest. In particular, it was shown by Spitzer [13] to hold for all partial sums of independent, identically distributed, integer-valued random variables except those with mean 0

and finite variance.[5] The above theorem for this class of chains was established by Kesten and Spitzer in [9], and the proof given above was patterned after their proof.

Theorem 5.1 includes certain cases not covered by the present theorem. For example, it was shown in [9] that, for the partial sum of independent, integer-valued random variables with a common distribution having zero mean and finite variance, the conditions of Theorem (5.1) hold; but in this case it was shown that the dual boundary has two points. On the other hand, the present theorem clearly includes cases not covered by Theorem (5.1)—notably, some positive recurrent chains are included in the present result.

If the chain is null recurrent and has only a single point in its dual boundary, then, as was noted in [8], we must have that C^*_{xy} exists for all states x, y. Moreover, in this case we must have that $C^*_{xy} = (\pi_y/\pi_x)\mu_{xy}$. On the other hand, if the chain is positive recurrent, the weak and strong limits need no longer be the same.

Usually it is not easy to verify when condition (5.8) holds. In the null-recurrent case, a simple sufficient condition is the following.

THEOREM 5.3. *If, in a null-recurrent chain, P^n_{yy} is a monotone function of n, then* (5.8) *holds.*

Proof. We must have $Q_n(y; \{y\}) > 0$ for all n. For if $Q_{n_0}(y; \{y\}) = 0$, then $Q_{n+n_0}(y; \{y\}) = 0$ for all $n \geqq 0$, and thus $\sum_{n=0}^{\infty} Q_n(y; \{y\}) < \infty$. But this cannot be true in a null-recurrent chain. Thus we have

$$\frac{Q_n(y; \{y\})}{Q_{n-1}(y; \{y\})} \leqq 1 .$$

Now, it is well-known that

$$1 = \sum_{k=0}^{n} P^{n-k}_{yy} Q_k(y; \{y\}) ,$$

and thus

$$Q_n(y; \{y\}) + \sum_{k=0}^{n-1} (P^{n-k}_{yy} - P^{n-1-k}_{yy}) Q_k(y; \{y\}) = 0 .$$

Since P^n_{yy} is nonincreasing, we have

$$Q_n(y; \{y\}) \geqq Q_{n-1}(y; \{y\})(1 - P^n_{yy}) .$$

The fundamental limit theorem for Markov chains (see [1]) then gives us

[5] Of course, the partial sums must also be such that they constitute an irreducible, recurrent chain.

$$\liminf_{n\to\infty} \frac{Q_n(y;\{y\})}{Q_{n-1}(y;\{y\})} \geqq \lim_{n\to\infty}(1 - P_{yy}^n) = 1 .$$

Let us now extend the result of Corollary 5.1.

THEOREM 5.4. *If for a state* y, $Q_n(y;\{y\}) > 0$ *for all* n *and there is a positive* $\alpha \leqq 1/2$ *such that*

$$\sup_n \frac{Q_{[\alpha n]}(y;\{y\})}{Q_n(y;\{y\})} < \infty \tag{5.14}$$

and (5.8) *holds, then* (5.5) *holds for any* $k > 0$ *and any initial point* x.

Proof. Consider first the case $x = y$. Let V_k denote the time of the kth visit to y.[6] Then clearly,

$$P_y(N_n(\{y\}) = k) = \sum_{r=1}^{n} P_y(V_k = r)Q_{n-r}(y;\{y\}) .$$

By (5.8) we have

$$\lim_{m\to\infty}\lim_{n\to\infty} \frac{\sum_{r=1}^{m} P_y(V_k = r)Q_{n-r}(y;\{y\})}{Q_n(y;\{y\})} = \sum_{r=1}^{\infty} P_y(V_k = r) = 1 .$$

To complete the proof, we must therefore show that for all $k > 0$,

$$\lim_{m\to\infty}\overline{\lim_{n\to\infty}} \sum_{r=m+1}^{n} P_y(V_k = r)Q_{n-r}(y;\{y\})Q_n(y;\{y\})^{-1} = 0 . \tag{5.15}$$

If $x = y$, then clearly (5.5) holds for $k = 0$. Suppose we have already established that (5.5) holds for all $k < k_0$ when $x = y$. Then,

$$\lim_{n\to\infty} \frac{P_y(N_n(\{y\}) < k_0)}{P_y(N_n(\{y\}) = 0)} = \lim_{n\to\infty} \frac{P_y(V_{k_0} > n)}{Q_n(y;\{y\})} = k_0 . \tag{5.16}$$

Now, decompose the sum in (5.15) as follows:

$$\sum_{r=m+1}^{n} = \sum_{r=m+1}^{[\alpha n]} + \sum_{[\alpha n]+1}^{n-m} + \sum_{n-m+1}^{n} .$$

Since

$$\sum_{m+1}^{[\alpha n]} \leqq \frac{Q_{n-[\alpha n]}(y;\{y\})}{Q_n(y,\{y\})} \sum_{r=m+1}^{[\alpha n]} P_y(V_{k_0} = r) ,$$

we obtain from (5.14)

$$\lim_{m\to\infty}\overline{\lim_{n\to\infty}} \sum_{r=m+1}^{[\alpha n]} = 0 .$$

[6] That is, $V_1 = V_{\{y\}}$, and for $k > 1$, $V_k = \inf\{r > V_{k-1}: X_r = y\}$.

Next,

$$\sum_{r=[\alpha n]+1}^{n-m} \leqq Q_m(y; \{y\})\left[\frac{P_y(V_{k_0} > [\alpha n])}{Q_{[\alpha n]}(y, \{y\})} \frac{Q_{[\alpha n]}(y, \{y\})}{Q_n(y, \{y\})}\right]$$

and thus, by (5.14) and (5.16),

$$\lim_{m\to\infty} \overline{\lim_{n\to\infty}} \sum_{r=[\alpha n]+1}^{n-m} = 0 .$$

Finally, since

$$\sum_{n-m+1}^{n} \leqq \frac{P_y(V_{k_0} > n-m)}{Q_{n-m}(y, \{y\})} \frac{Q_{n-m}(y, \{y\})}{Q_n(y, \{y\})} - \frac{P_y(V_{k_0} > n)}{Q_n(y, \{y\})} ,$$

from (5.8) and (5.16) we obtain

$$\lim_{m\to\infty} \lim_{n\to\infty} \sum_{n-m+1}^{n} = 0 .$$

Thus, when $x = y$, (5.15) holds for $k = k_0$, and consequently (5.5) holds for all $k \geqq 0$. Now for an arbitrary x we have for $k > 0$,

$$P_x(N_n(\{y\}) = k) = \sum_{r=1}^{n} F^r_{\{y\}}(x, y) P_y(N_{n-r}(\{y\}) = k-1) ,$$

and thus

$$\lim_{m\to\infty} \lim_{n\to\infty} \frac{\sum_{r=1}^{m} F^r_{\{y\}}(x, y) P_y(N_{n-r}(\{y\}) = k-1)}{Q_n(y, \{y\})} = 1 .$$

To complete the proof, we must therefore show that

$$\lim_{m\to\infty} \overline{\lim_{n\to\infty}} \sum_{r=m+1}^{n} = 0 . \tag{5.17}$$

If $x = y$, then, by what was just shown above, we must have that (5.17) holds. If $x \neq y$, we have

$$F^{n+m_0}_{\{y\}}(y, \{y\}) \geqq {}_yP^{m_0}_{yx} F^n_{\{y\}}(x, y) ,$$

and since $\sum_{n=0}^{\infty} {}_yP^n_{yx} = \pi_x/\pi_y$, we must have ${}_yP^{m_0}_{yx} > 0$ for some m_0. Thus, if we set $\beta = {}_yP^{m_0}_{yx}$, we have

$$\sum_{n=m+1}^{n} \leqq \sum_{r=n+1+m_0}^{n+m_0} \frac{F^r_y(y, \{y\}) P_y(N_{n+m_0-r}(\{y\}) = k-1)}{\beta Q_{n+m_0}(y, \{y\})} \frac{Q_{n+m_0}(y, \{y\})}{Q_n(y, \{y\})} ,$$

and accordingly (5.17) holds in general. This completes the proof.

6. Examples. In this section we shall apply the results of the last section to various specific Markov chains. We note again that

for the important case of partial sums of independent, identically distributed, integer-lattice-valued random vectors in r dimensions, Kesten and Spitzer [9] have shown that the strong ratio limits (1.1) exist for all nonempty finite sets A.

EXAMPLE 1. *Random Walk on the Nonnegative Integers.* The state space of this chain is the nonnegative integers, and its transition matrix is given for $x \geqq 0$ by

$$P_{xy} = \begin{cases} p_x, & y = x + 1 , \\ r_x, & y = x , \\ q_x, & y = x - 1 , \\ 0, & \text{elsewhere,} \end{cases} \tag{6.1}$$

where $p_x + q_x + r_x = 1$. For $x \geqq 0$ we have $q_{x+1}, p_x > 0$, while for $x = 0$ we have $q_0 = 0$. Let

$$\pi_x = \frac{p_0 p_1 \cdots p_{x-1}}{q_1 \cdots q_x} , \tag{6.2}$$

and observe that

$$p_x \pi_x = q_{x+1} \pi_{x+1} . \tag{6.3}$$

We gather together below some essential facts about random walks which we shall need. For details we refer the reader to [5].

Associated with each random walk is a sequence of polynomials $Q_x(t)$ on $[-1, 1]$, and a probability distribution $\Psi(t)$ with support on $[-1, 1]$. Each polynomial $Q_x(t)$ is of exact degree x, and the polynomials satisfy the reccurrence relation

$$\begin{aligned} &tQ_x(t) = p_x Q_{x+1}(t) + r_x Q_x(t) + q_x Q_{x-1}(t), \; x \geqq 0 , \\ &Q_0(t) = 1 , \qquad Q_{-1}(t) = 0 . \end{aligned} \tag{6.4}$$

Moreover, we have

$$P_{xy}^n = \pi_y \int_{-1}^{1} t^n Q_x(t) Q_y(t) d\Psi(t) \tag{6.5}$$

and, in particular,

$$\delta_{xy} = P_{xy}^0 = \pi_y \int_{-1}^{1} Q_x(t) Q_y(t) d\Psi(t) . \tag{6.6}$$

The integrals in the above expressions include the mass m_1 at 1 and m_{-1} at -1. If the walk is *symmetric* (i.e., $r_x = 0$ for all x), then $m_{-1} = m_1$. If the walk is asymmetric, then $m_{-1} = 0$. If the walk is positive recurrent, then $m_1 = (\Sigma_x \pi_x)^{-1}$, while for a walk that is not

positive recurrent we have $m_1 = 0$. From (6.3) and (6.4) we obtain

$$(1 - t)\pi_x Q_x(t) = -\pi_x p_x(Q_{x+1}(t) - Q_x(t)) + \pi_{x-1}p_{x-1}(Q_x(t) - Q_{x-1}(t)) ,$$

and thus

$$-(1 - t)\sum_{z=0}^{y} \pi_z Q_z(t) = \pi_y p_y(Q_{y+1}(t) - Q_y(t)) .$$

Another summation yields

$$Q_{x+1}(t) = 1 - (1 - t)\sum_{y=0}^{x} (p_y\pi_y)^{-1} \sum_{z=0}^{y} \pi_z Q_z(t) . \tag{6.7}$$

Setting $t = 1$ in the above, we see that

$$Q_x(1) = 1 \text{ all } x \geqq 0 . \tag{6.8}$$

We are now in a position to establish the following important property of random walks.

THEOREM 6.1. *In every null-recurrent random walk and in every positive-recurrent asymmetric random walk, the series*

$$C^*_{xy} = \sum_{n=0}^{\infty} (P^n_{yy} - P^n_{xy})$$

converges. Moreover,

$$C^*_{xy} = \begin{cases} m_1\pi_y \sum_{n=x}^{y-1} (p_n\pi_n)^{-1} \sum_{k=0}^{n} \pi_k , & x < y , \\ \pi_y \sum_{k=y}^{x-1} (p_k\pi_k)^{-1}\left\{1 - m_1 \sum_{j=0}^{k} \pi_j\right\} , & x > y , \\ 0 , & x = y . \end{cases} \tag{6.9}$$

Proof. From (6.5) we see that

$$\sum_{n=0}^{N} (P^n_{yy} - P^n_{xy}) = \pi_y \int_{-1}^{1} (1 - t)^{-1}(1 - t^{N+1})Q_y(t)[Q_y(t) - Q_x(t)]d\Psi(t) . \tag{6.10}$$

Suppose $y > x$. Then from (6.7) and (6.10) we have

$$\sum_{n=0}^{N} (P^n_{yy} - P^n_{xy}) = -\pi_y \int_{-1}^{1} (1 - t^{N+1})Q_y(t) \sum_{k=x}^{y-1} (p_k\pi_k)^{-1} \sum_{r=0}^{k} \pi_r Q_r(t)d\Psi(t) .$$

By (6.6), we see that

$$\pi_y \int_{-1}^{1} Q_y(t) \sum_{k=x}^{y-1} (p_k\pi_k)^{-1} \sum_{r=0}^{k} \pi_r Q_r(t) d\Psi(t) = 0 ,$$

and thus

$$\sum_{n=0}^{N} (P_{yy}^n - P_{xy}^n) = \pi_y \int_{-1}^{1} t^{N+1} Q_y(t) \sum_{k=x}^{y-1} (p_k\pi_k)^{-1} \sum_{r=0}^{k} \pi_r Q_r(t) d\Psi(t) . \tag{6.11}$$

From the above, we obtain

$$C_{xy}^* = m_1\pi_y \sum_{n=x}^{y-1} (p_n\pi_n)^{-1} \sum_{k=0}^{n} \pi_k + \lim_{N\to\infty} \pi_y m_{-1}(-1)^{N+1} Q_y(-1) \sum_{n=x}^{y-1} (p_n\pi_n)^{-1} \sum_{k=0}^{n} \pi_k Q_k(-1) ,$$

and thus C_{xy}^* exists if and only if $m_{-1} = 0$. In that case,

$$C_{xy}^* = m_1\pi_y \sum_{n=x}^{y-1} (p_n\pi_n)^{-1} \sum_{k=0}^{n} \pi_k .$$

Now suppose $y < x$. A similar computation to that used above then shows that

$$\sum_{n=0}^{N} (P_{yy}^n - P_{xy}^n) = \pi_y \int_{-1}^{1} (1 - t^{N+1}) Q_y(t) \sum_{k=y}^{x-1} (p_k\pi_k)^{-1} \sum_{j=0}^{k} \pi_j Q_j(t) d\Psi(t) .$$

From (6.6), we obtain

$$\pi_y \int_{-1}^{1} Q_y(t) \sum_{k=y}^{x-1} (p_k\pi_k)^{-1} \sum_{j=0}^{k} \pi_j Q_j(t) d\Psi(t) = \pi_y \sum_{k=y}^{x-1} (p_k\pi_k)^{-1} ,$$

and thus

$$\sum_{n=0}^{N} (P_{yy}^n - P_{xy}^n) = \pi_y \sum_{k=y}^{x-1} (p_k\pi_k)^{-1} - \pi_y \int_{-1}^{1} t^{N+1} Q_y(t) \sum_{k=y}^{x-1} (p_k\pi_k)^{-1} \sum_{j=0}^{k} \pi_j Q_j d\Psi$$

Consequently, taking the limit as $N \to \infty$, for $m_{-1} = 0$ we obtain

$$C_{xy}^* = \pi_y \sum_{k=y}^{x-1} (p_k\pi_k)^{-1} - \pi_y m_1 \sum_{n=y}^{x-1} (p_n\pi_n)^{-1} \sum_{k=0}^{n} \pi_k = \pi_y \sum_{n=y}^{x-1} (p_n\pi_n)^{-1} \left\{ 1 - m_1 \sum_{k=0}^{n} \pi_k \right\} .$$

This establishes the theorem.

As a corollary we have the following result.

COROLLARY 6.1. *Under the same condition as for Theorem* 6.1 *we have, for* $z \neq x$,

$$\lim_{y\to\infty} \pi_y^{-1} G_{\{z\}}(x, y) = \begin{cases} \sum_{n=z}^{x-1} (p_n \pi_n)^{-1}, & z < x, \\ 0, & z > x. \end{cases} \tag{6.12}$$

Proof. By Theorem 22 of [6], for $z \neq x, y$ we have

$$G_{\{z\}}(x, y) = C^*_{zy} - C^*_{xy} + (\pi_y/\pi_z) C^*_{xz}. \tag{6.13}$$

From (6.9), we see that when $y > \max(x, z)$,

$$\begin{aligned} \frac{C^*_{zy} - C^*_{xy}}{\pi_y} &= m_1 \left\{ \sum_{n=z}^{y-1} - \sum_{n=x}^{y-1} \right\} (p_n \pi_n)^{-1} \sum_{k=0}^{n} \pi_k \\ &= \begin{cases} m_1 \sum_{n=z}^{x-1} (p_n \pi_n)^{-1} \sum_{k=0}^{n} \pi_k, & z < x, \\ -m_1 \sum_{n=x}^{z-1} (p_n \pi_n)^{-1} \sum_{k=0}^{n} \pi_k, & z > x. \end{cases} \end{aligned}$$

Thus if $y > \max(x, z)$, we have

$$\pi_y^{-1} G_{\{z\}}(x, y) = \begin{cases} \sum_{n=z}^{x-1} (p_n \pi_n)^{-1}, & x > z, \\ 0, & x < z, \end{cases}$$

which establishes (6.12).

REMARK. Observe that in the null-recurrent case the matrix C^*_{xy} is triangular with 0's on and above the main diagonal. Also observe that in the null-recurrent case the limit in (6.12) is C^*_{xz}/π_z, while in the positive-recurrent case this limit is different from the above expression.

Theorem 6.1 shows that the stronger Doeblin form of weak ratio limits always holds for null recurrent random walks. From Corollary 6.1 we see that Theorem 5.2 is applicable to all random walks except symmetric positive-recurrent walks.

EXAMPLE 2. *The Aperiodic Renewal Chain.* This chain also has the nonnegative integers as its state space. Its transition matrix is given by

$$P_{xy} = \begin{cases} p_{x+1}, & y = x + 1, \\ q_{x+1} = 1 - p_{x+1}, & y = 0, \\ 0, & \text{elsewhere}, \end{cases} \tag{6.14}$$

where $0 < p_x < 1$, $x > 0$.

Let

$$\pi_x = p_1 \cdots p_x , \qquad (\pi_0 = 1) ,$$
$$\sigma_x = \sum_{k=0}^{x-1} \pi_k , \qquad \lim_{x\to\infty} \sigma_x = \sigma_\infty .$$

It was shown in [6] that

$$C^*_{xt} = \begin{cases} \dfrac{\sigma_t}{\sigma_\infty} - \dfrac{\pi_t}{\pi_x}\dfrac{\sigma_x}{\sigma_\infty} , & t \geqq x , \\ \dfrac{\pi_t}{\pi_x}\left(1 - \dfrac{\sigma_x}{\sigma_\infty}\right) + \dfrac{\sigma_t}{\sigma_\infty} , & t < x , \end{cases} \tag{6.15}$$

where $1/\sigma_\infty = 0$ if $\sigma_\infty = \infty$.

If $t > \max(x, y)$, we obtain

$$\frac{C^*_{yt} - C^*_{xt}}{\pi_t} = \frac{1}{\sigma_\infty}\left(\frac{\sigma_x}{\pi_x} - \frac{\sigma_y}{\pi_y}\right)$$

and thus, by (6.13),

$$\lim_{t\to\infty} \pi_t^{-1} G_{\{y\}}(x, t) = \frac{1}{\sigma_\infty}\left(\frac{\sigma_x}{\pi_x} - \frac{\sigma_y}{\pi_y}\right) + \frac{C^*_{xy}}{\pi_y} \tag{6.16}$$
$$= \begin{cases} 0 , & y \geqq x , \\ \dfrac{1}{\pi_x} , & y < x . \end{cases}$$

Again observe that this limit is C^*_{xy}/π_y only in the null case. This establishes the following:

THEOREM 6.2. *In any aperiodic, recurrent renewal chain, C^*_{xy} exists for all states x, y. Moreover if $y \neq x$,*

$$\lim_{t\to\infty} \pi_t^{-1} G_{\{y\}}(x, t) = \begin{cases} 0 , & y > x , \\ \pi_x^{-1} , & y < x . \end{cases} \tag{6.17}$$

The above theorem shows then that all the stronger forms of the weak ratio limits exist in every renewal chain, and that Theorem 5.2 is applicable. It is quite simple, however, to exhibit renewal chains for which some or none of the strong ratio limit results hold. To see this, all we need do is recall the familiar fact ([1], Sec. 8) that given any sequence $\{W_k\}$ of independent, identically distributed positive integer-valued random variables, we may construct a renewal chain such that the successive returns to 0 have waiting times W_k. Since all the conditions for the various strong ratio limits to hold involve conditions on the tails of the distribution of W_1, and since we may

readily construct such distributions which do or do not satisfy these conditions, we see that there are renewal chains for which these strong ratio limits hold, and renewal chains for which they do not hold.

REMARK. The dual of the renewal chain presents another example of a Markov chain with a single boundary point for which series (3.3) always converges.

As our next example, we consider the reflecting-barrier process.

EXAMPLE 3. *Reflecting Barrier Process.* Let U_i be independent, identically distributed, integer-valued random variables assuming positive and negative values and such that $E(e^{i\theta U_1}) = 1$ if and only if $\theta = 2n\pi$, and let

$$T_n = (T_{n-1} + U_n)^+ = \max(0, T_{n-1} + U_n) .$$

It is a familiar fact (see, e.g., [11]) that

$$P(T_n = 0 \mid T_0 = x) = P(M_{n-1} = 0, S_n \leqq -x) , \tag{6.18}$$

where $S_n = U_1 + \cdots + U_n$, and $M_n = \max(0, S_1, \cdots, S_n)$. From (6.18) we see at once that

$$(P_{00}^n - P_{x0}^n) \geqq 0 ,$$

and thus

$$\sum_{n=0}^{\infty} (P_{00}^n - P_{x0}^n)$$

converges. As the T_n process is irreducible, by Lemma 3.1 C_{xy}^* exists for all states x, y. Thus the Doeblin-type ratio limits are always valid for recurrent T_n chains. By results in [12] (see eqn. 4.21) condition (5.1) of Theorem 5.1 is satisfied if and only if

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} P(S_k \leqq 0) = \alpha , \qquad 0 \leqq \alpha < 1 , \tag{6.19}$$

and thus, when the above condition is satisfied, the strong results of Theorem 5.1 are applicable. In particular, if $P(S_k \leqq 0)$ converges to a value <1 (e.g., if U_1 has a symmetric distribution), we see that (6.19) holds.

Since

$$G_{\{0\}}(x, y) = \sum_{n=0}^{\infty} P(S_n = y - x, S_i > -x, 0 \leqq i \leqq n)$$

we see that $G_{\{0\}}(x, y)$ is the same as the Green's functisn for the half

line $(-\infty, 0]$ for the S_n process. From results in [13] (see E3 P. 332) it easily follows that the dual boundary of the T_n process has exactly one point whenever the S_n are recurrent. But if the S_n are recurrent, then the T_n must be null recurrent (see [11]), and since $P_{00}^n = P(M_n = 0)$ is clearly a monotone function of n, we see from Theorems 5.2 and 5.3 that

$$\lim_{n\to\infty} \frac{P_x(V_{\{0\}} > n)}{P_0(V_{\{0\}} > n)} = C_{x0}^* + \delta_{x0} \tag{6.20}$$

whenever the S_n are recurrent. In particular, (6.20) holds whenever $EU_1 = 0$.

REFERENCES

1. K. L. Chung, *Markov Chains with Stationary Transition Probabilities*, Springer-Verlag, Berlin, 1960.
2. G. Doetch, *Handbook Der Laplace Transformation*, Band 1, Birkhauser-Verlag, Basel, Switzerland, 1950.
3. G. H. Hardy, *Divergent Series*, Oxford University Press, New York, 1956.
4. M. A. Kac, *A class of limit theorems*, Trans. Amer. Math. Soc. **84** (1957), 459-471.
5. S. Karlin, and J. McGregor, *Random walks* Illinois J. Math. **3** (1959), 66-81.
6. J. G. Kemeny, and J. L. Snell, *Potentials for denumerable Markov chains*, J. Math. Anal. Appl. **3** (1961), 196-260.
7. ———, *Notes on discrete potential theory*, J. Math. Anal. Appl. **3** (1961), 117-121.
8. ———, *Boundary theory for recurrent Markov chains*, Trans. Amer. Math. Soc. **106** (1963), 495-420
9. H. Kesten and F. Spitzer, *Ratio limit theorems*, J. Analyse Math. **11** (1963), 285-322.
10. S. Orey, *Sums Arising in the theory of Markov chains*, Proc. Amer. Math. Soc. **12** (1961), 847-856.
11. S. C. Port, *On random walks with a reflecting barrier*, Trans. Amer. Math. Soc. 117 (1965), 362-370.
12. ———, *Some theorems on functionals of a Markov chain*, Ann. Math Stat. **35** (1964), 1275-1290.
13. F. Spitzer, *Hitting probabilities, Principles of Random Walk*, Van Nostrand, Princeton N. J., 1964.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

A GENERALISATION OF W^*-ALGEBRAS

G. A. REID

Using the theory of double centralisers due to B. E. Johnson, we define a QW^*-algebra as being a B^*-algebra, A, such that the algebra of double centralisers of each closed *-subalgebra B is contained in a suitable related closed *-subalgebra B_{00}.

After obtaining explicit descriptions of the algebras of double centralisers of commutative and noncommutative B^*-algebras, we prove that in the general noncommutative case a W^*-algebra is a QW^*-algebra, and a QW^*-algebra is an AW^*-algebra, while in the commutative case the QW^* and AW^* conditions are equivalent.

We prove that if A is QW^* then so are its centre, any maximal commutative *-subalgebra, and any subalgebra of the form eAe for e a projection in A.

We shall be concerned with centraliser theory, for the basic details of which reference may be made to Johnson [2], [3].

I should like to take this opportunity of expressing my sincere gratitude to Dr. J. H. Williamson, my research supervisor, for his advice and encouragement.

DEFINITION 1. A *left centraliser* $\mathscr{T}$ of the algebra A is a linear map $\mathscr{T}$ of A into itself such that $\mathscr{T}(xy) = (\mathscr{T}x)y$ for all $x, y \in A$.

A *right centraliser* $\mathscr{S}$ is a linear operator on A such that $\mathscr{S}(xy) = x(\mathscr{S}y)$ for all $x, y \in A$.

A *double centraliser* (the concept is due to Johnson [2]) is a pair of linear operators $(\mathscr{T}, \mathscr{S})$ such that $x\cdot(\mathscr{T}y) = (\mathscr{S}x)\cdot y$ for all $x, y \in A$.

The set of all double centralisers on A is denoted by $Q(A)$.

We will assume throughout that $xA = 0$ or $Ax = 0$ only holds for $x = 0$. We note that this holds for B^*-algebras since $xA = 0 \Rightarrow xx^* = 0 \Rightarrow x = 0$, and $Ax = 0 \Rightarrow x^*x = 0 \Rightarrow x = 0$.

It is not difficult to see that defining $(\mathscr{T}_x, \mathscr{S}_x) \in Q(A)$ for $x \in A$ by $\mathscr{T}_x(y) = xy$, $\mathscr{S}_x(y) = yx$, and algebraic operations in $Q(A)$ by

$$\lambda_1(\mathscr{T}_1, \mathscr{S}_1) + \lambda_2(\mathscr{T}_2, \mathscr{S}_2) = (\lambda_1\mathscr{T}_1 + \lambda_2\mathscr{T}_2, \lambda_1\mathscr{S}_1 + \lambda_2\mathscr{S}_2)$$
$$(\mathscr{T}_1, \mathscr{S}_1)\cdot(\mathscr{T}_2, \mathscr{S}_2) = (\mathscr{T}_1\mathscr{T}_2, \mathscr{S}_2\mathscr{S}_1)$$

Received August 27, 1964.

we have A embedded as a subalgebra of $Q(A)$, which is an algebra with identity. $A = Q(A)$ if and only if A has an identity. Also, for $(\mathscr{T}, \mathscr{S}) \in Q(A)$, $\mathscr{T}$ is a left centraliser and $\mathscr{S}$ is a right centraliser, and either of $\mathscr{T}, \mathscr{S}$ determines the other uniquely.

If A is commutative, the notions of right, left and double centraliser coincide, and for $(\mathscr{T}, \mathscr{S}) \in Q(A)$ we have $\mathscr{T} = \mathscr{S}$.

PROPOSITION 1. If A is a Banach algebra then all double centralisers are continuous.

Proof. Suppose $(\mathscr{T}, \mathscr{S}) \in Q(A)$ and say $x_n \to x$, $\mathscr{T}x_n \to y$. Then

$$\begin{aligned} z\cdot(\mathscr{T}x_n) &= (\mathscr{S}z)\cdot x_n \\ \to z\cdot y \qquad &\to (\mathscr{S}z)\cdot x = z\cdot(\mathscr{T}x)\ . \end{aligned}$$

So $z(y - \mathscr{T}x) = 0$ for all $z \in A$ i.e. $A(y - \mathscr{T}x) = 0$ and so $y = \mathscr{T}x$. Therefore $\mathscr{T}$ is a closed operator on the Banach space A, hence by the Closed Graph Theorem, $\mathscr{T}$ is continuous. Likewise so is $\mathscr{S}$.

We are particularly interested in C^*-algebras and in both the commutative and noncommutative cases explicit descriptions of their centraliser algebras may be given.

By the Gelfand Representation Theorem a commutative B^*-algebra is isometrically isomorphic to the space $C_0(Z)$ of all continuous functions vanishing at infinity on its carrier space, Z, a locally compact Hausdorff space.

PROPOSITION 2. For a locally compact Hausdorff space Z we have $QC_0(Z) = C(Z)$, the space of all bounded continuous functions on Z.

Proof. Certainly any $h \in C(Z)$ defines an element $\mathscr{T}_h$ of $QC_0(Z)$ by $\mathscr{T}_h f = h\cdot f$ for $f \in C_0(Z)$, for

$$f \in C_0(Z),\ h \in C(Z) \Rightarrow hf \in C_0(Z)$$

and

$$h(fg) = (hf)g\ .$$

We clearly have $\|\mathscr{T}_h\| \leqq \|h\|_\infty$. Suppose conversely we are given a centraliser $\mathscr{T}$ on $C_0(Z)$. Then for $f, g \in C_0(Z)$ we have

$$(\mathscr{T}f)g = \mathscr{T}(fg) = \mathscr{T}(gf) = (\mathscr{T}g)f$$

so for $z \in Z$ taking any $f \in C_0(Z)$ such that $f(z) \neq 0$ and defining $h(z) = \mathscr{T}f(z)/f(z)$ we have $h(z)$ well defined independently of f.

Being a quotient of continuous functions, h is continuous at z, for each $z \in Z$. And for any $g \in C_0(Z)$,

$$\mathscr{T}g(z) = \frac{\mathscr{T}f(z)}{f(z)} g(z) = h(z)g(z)$$

so

$$\mathscr{T}g = hg = \mathscr{T}_h g \ .$$

Now by Proposition 1, $\mathscr{T}$ is a bounded operator, so taking $f \in C_0(Z)$ such that $0 \leqq f \leqq 1$ and $f(z) = 1$ we have $h(z) = \mathscr{T}f(z)$ and $|\mathscr{T}f(z)| \leqq ||\mathscr{T}f||_\infty \leqq ||\mathscr{T}|| \, ||f||_\infty = ||\mathscr{T}||$ so $||h||_\infty \leqq ||\mathscr{T}||$ and we see $h \in C(Z)$.

Hence all $\mathscr{T}$ are of the form $\mathscr{T}_h$ and $||\mathscr{T}|| = ||h||_\infty$. So $QC_0(Z) = C(Z)$.

PROPOSITION 3. If A is a C^*-algebra over H, principal identity E, then $Q(A)$ is isometrically isomorphic to

$$\{T \in \mathscr{B}(H) \colon T = ETE, \ TA \cup AT \subset A\} \ .$$

Proof. Recall that the principal identity of a C^*-algebra A is defined to be the orthogonal projection of H onto $M = H \ominus N$ where $N = \{\xi \in H \colon A\xi = 0\}$. Equivalently M is the closure of

$$M_1 = \{T\xi : T \in A, \xi \in H\} \ .$$

Suppose given $(\mathscr{T}, \mathscr{S}) \in Q(A)$, then $\mathscr{T}$ is a bounded left centraliser.

Since A is a C^*-algebra it has an approximate identity (Segal [6]), $(Z_\lambda)_{\lambda \in \Lambda}$ say, so $||Z_\lambda|| = 1$, and $SZ_\lambda \to S$, $Z_\lambda S \to S$ for each $S \in A$. So $\mathscr{T}(Z_\lambda S) \to \mathscr{T}(S)$. But $\mathscr{T}(Z_\lambda S) = \mathscr{T}(Z_\lambda)S = T_\lambda S$ where $T_\lambda = \mathscr{T}(Z_\lambda)$, so $\mathscr{T}(S) = \lim_\lambda T_\lambda S$ and $||T_\lambda|| \leqq ||\mathscr{T}|| \, ||Z_\lambda|| = ||\mathscr{T}||$. For $\xi \in M_1$, $\xi = S\eta$ some $S \in A, \eta \in H$ so $\mathscr{T}(S)\eta = \lim_\lambda T_\lambda S\eta = \lim_\lambda T_\lambda \xi$. Define $T\xi = \lim_\lambda T_\lambda \xi = \mathscr{T}(S)\eta$, then T maps M_1 into M and $||T\xi|| \leqq ||\mathscr{T}|| \, ||\xi||$ so $||T|| \leqq ||\mathscr{T}||$.

So extend T to a map of M into M and define $T = 0$ on $H \ominus M$, so we have $T = ETE$ and $\mathscr{T}(S)\eta = \lim_\lambda T_\lambda S\eta = TS\eta$. Therefore $\mathscr{T}(S) = TS$ and $||\mathscr{T}|| \leqq ||T||$. So $||\mathscr{T}|| = ||T||$.

We have

$$\begin{aligned} (\mathscr{S}S)Z_\lambda &= S(\mathscr{T}Z_\lambda) = STZ_\lambda \\ &\to \mathscr{S}S \qquad\qquad\quad \to ST \ . \end{aligned}$$

So $\mathscr{S}(S) = ST$ for all $S \in A$, and as for $\mathscr{T}$, $||\mathscr{S}|| = ||T||$. Since $TS, ST \in A$ for all $S \in A$ we have $TA \cup AT \subset A$. Conversely given any

T such that $T = ETE$ and $TA \cup AT \subset A$, the maps $S \to TS$, $S \to ST$ both map A into itself and define a double centraliser of A. Hence result.

Denote the set $\{T \in \mathscr{B}(H)\colon T = ETE,\ TA \cup AT \subset A\}$ by $I(A)$, the idealiser of A in $E \cdot \mathscr{B}(H) \cdot E$.

Now let us suppose that B is a closed *-subalgebra of the B^*-algebra A. We define $B_0 = \{x \in A\colon Bx = xB = 0\}$ and $B_{00} = (B_0)_0$. Then B_{00} is a closed *-subalgebra of A containing B. Should it be necessary to make explicit mention of the algebra A we will write $B_0(A)$, etc.

Suppose two elements x_1, x_2 of B_{00} give the same double centraliser on B, so $x_1 y = x_2 y$ and $y x_1 = y x_2$ for all $y \in B$. Then $(x_1 - x_2)B = B(x_1 - x_2) = 0$ so $x_1 - x_2 \in B_0$. But $(x_1 - x_2)^* \in B_{00}$ so we have

$$(x_1 - x_2)^*(x_1 - x_2) = 0$$

and hence $x_1 - x_2 = 0$. So $x_1 = x_2$.

DEFINITION 2. A B^*-algebra A is said to be a *QW^*-algebra* if for each closed *-subalgebra B of A all double centralisers of B are given by elements of B_{00}. We see that for each double centraliser the corresponding element of B_{00} is unique, and so we may briefly say that A is QW^* if and only if $Q(B) \subset B_{00}$ for all closed *-subalgebras B.

We recall the definition of an AW^*-algebra (Kaplansky [4]).

DEFINITION 3. A B^*-algebra A is said to be an *AW^*-algebra* if

(i) every set of orthogonal projections in A has a least upper bound in A.

(ii) every maximal commutative *-subalgebra B of A is generated by its projections.

We also recall that a W^*-algebra is a C^*-algebra, over H say, which is closed in the weak operator topology defined by seminorms $\| T \|_{\xi,\eta} = |\langle T\xi, \eta\rangle|$ for $\xi, \eta \in H$. Denote weak closure by $^{-w}$.

PROPOSITION 4. For A a C^*-algebra, $I(A) \subset A^{-w}$.

Proof. By von Neumann's Double Commutant Theorem, $A^{-w} = \{T \in \mathscr{B}(H)\colon T = ETE,\ T \in A''\}$ where as usual A'' denotes the double commutant of A.

Suppose $T \in I(A)$, $S \in A'$, $R \in A$, then certainly $T = ETE$ and $(ST - TS)R = S(TR) - T(SR) = TRS - TRS = 0$. So $(ST - TS)E = 0$

and therefore $ST = TSE$. Since $T^* \in I(A)$, $S^* \in A'$ we have $S^*T^* = T^*S^*E$ so $TS = EST$. Thus $TS = EST = ETSE = TSE = ST$ and so $T \in A''$. Hence $I(A) \subset A''$.

THEOREM 1. *For a B^*-algebra A, $W^* \Rightarrow QW^* \Rightarrow AW^*$.*

If A is commutative, carrier space Z, then A is $QW^* \Leftrightarrow A$ is $AW^* \Leftrightarrow Z$ is extremally disconnected.

Proof. If A is a W^*-algebra and B is a closed *-subalgebra of A with principal identity E, then since A is W^* we note $E \in A$, and by Proposition 4, $I(B) \subset B^{-w} \subset A^{-w} = A$. Also we easily see that $B_0 = (I - E)A(I - E)$ so $B_{00} = EAE$. Thus $Q(B) \subset B_{00}$ by Proposition 3 and hence A is QW^*.

Suppose now that A is a commutative B^*-algebra, carrier space Z, so by the Gelfand Representation Theorem A is isometrically isomorphic to $C_0(Z)$.

It is well known that A is AW^* if and only if Z is an extremally disconnected compact Hausdorff space.

Suppose A is QW^*, then taking $B = A$ we see that A has an identity, so Z is compact Hausdorff.

Let U be any open dense subset of Z.

Then taking $B = \{f \in C(Z): f = 0 \text{ on } Z \backslash U\} = C_0(U)$, B is a closed *-ideal in A so $Q(B) = C(U) \subset A$.

So any continuous function f on U is extendible to Z. Therefore Z is extremally disconnected (see Gillman and Jerison [1], p. 96).

Now suppose that Z is an extremally disconnected compact Hausdorff space, and suppose B is a closed *-subalgebra of $A = C(Z)$.

Let $(Z_\lambda)_{\lambda \in \Lambda}$ be the sets of constancy of B (see Rickart [5], Ch. 3, § 2), then these form an upper semicontinuous decomposition of Z, so the space of these sets, Z' say, is a compact Hausdorff space and B may be considered as a space of continuous functions on Z'.

B is self-adjoint and separates points of Z', so by the Stone-Weierstrass Theorem, *either* B consists of all continuous functions on Z', in which case B has an identity so $Q(B) = B$, *or* B consists of all continuous functions on Z' vanishing at some point Z_0 of Z'. So $Q(B) =$ all continuous functions on $Z' \backslash \{Z_0\}$.

Given any function on $Z' \backslash \{Z_0\}$ it corresponds to a function f on $Z \backslash Z_0 = Y$ say.

Y is open, so $\bar{Y}$ is a compact open subset of Z, and therefore $\bar{Y}$ is extremally disconnected (Gillman and Jerison [1], p. 23). So there exists an extension of f to $\bar{Y}$, and defining $f = 0$ on $Z \backslash \bar{Y}$ we extend f to a continuons function on Z.

Now since

$$B_0 = \{g \in C(Z): g = 0 \text{ on } Y\}$$
$$= \{g \in C(Z): g = 0 \text{ on } \bar{Y}\}$$

and

$$B_{00} = \{g \in C(Z): g = 0 \text{ on } Z \setminus \bar{Y}\}$$

we therefore have $Q(B) \subset B_{00}$.

So A is QW^* and we have proved our theorem for A commutative.

Now let us return to the general case and suppose A to be QW^*.

(i) Suppose (e_α) is a set of orthogonal projections in A (so $\alpha \neq \beta \Rightarrow e_\alpha e_\beta = 0$).

Let B = closed *-subalgebra of A generated by the e_α's.
= closed linear hull of the e_α's.

Now there exists a unique $e \in B_{00}$ such that $ex = xe = x$ for all $x \in B$ and $e^*, e^2 \in B_{00}$ with

$$e^*x = xe^* = x$$
$$e^2x = xe^2 = x \quad \text{for all } x \in B .$$

So $e^2 = e^* = e$ and thus e is a projection.

Also $ee_\alpha = e_\alpha e = e_\alpha$ all α, so $e \geqq e_\alpha$ all α.

Now suppose f is a projection in A such that $f \geqq$ all e_α. Then $fe_\alpha = e_\alpha f = e_\alpha$ all α, so since all $x \in B$ are limits of linear combinations of the e_α's, we have $fx = xf = x$ for all $x \in B$.

Now

$$y \in B_0 \Rightarrow yfx = yx = 0$$
$$xyf = 0 \qquad \text{all } x \in B \Rightarrow yf \in B_0$$

so for all $y \in B_0$,

$$fey = f0 = 0$$
$$yfe = 0 \qquad \text{thus } fe \in B_{00} .$$

But

$$fex = fx = x$$
$$xfe = xe = x$$

all $x \in B$, so since e is unique, $e = fe$.

So $ef = fe = e$ and $e \leqq f$.

Hence e is a least upper bound in A for the e_α's.

(ii) Suppose B is a maximal commutative *-subalgebra of A. Then by Proposition 5 below, B is QW^*, thus since B is commutative it follows from the above result that B is AW^*, and is a maximal commutative *-subalgebra of itself and therefore generated by its projections.

Thus we have both conditions for A to be AW^*.

The obvious question of interest arising from this theorem is whether or not the QW^* and the AW^* conditions are equivalent in the noncommutative case, but so far we have not been able to settle this problem.

We now prove some results for QW^*-algebras similar to those holding for W^*- and AW^*-algebras. We are indebted to the referee for pointing out case (iv) of Proposition 5 as generalising cases (i) and (ii).

PROPOSITION 5. If A is a QW^*-algebra then so also are the following closed *-subalgebras of A:

(i) the centre Z of A,

(ii) any maximal commutative *-subalgebra of A,

(iii) the subalgebra eAe for any projection e in A,

(iv) S'' for any subset S of A such that $S^* = S$, where S'' is the double commutant of S in A.

Proof. We first prove (iv) from which (i) and (ii) follow immediately.

(iv) Suppose B is a closed *-subalgebra of S''.

Since A is QW^* any double centraliser on B is given by some $x \in B_{00}(A)$.

To prove $x \in B_{00}(S'')$, since $B_0(S'') \subset B_0(A)$, we need only show $x \in S''$.

Let $y \in S'$, $z \in B \subset S''$, then

$$(xy - yx)z = x(yz) - y(xz) = xzy - xzy = 0$$
$$z(xy - yx) = (zx)y - (zy)x = yzx - yzx = 0$$

so $xy - yx \in B_0(A)$.

Now

$$u \in B_0(A) \Rightarrow yuz = 0$$
$$zyu = yzu = 0 \quad \text{all } z \in B \Rightarrow yu \in B_0(A)\ ,$$

and likewise $u \in B_0(A) \Rightarrow uy \in B_0(A)$.

Therefore since $x \in B_{00}(A)$, $xyu = 0$ and $uxy = 0$ for all $u \in B_0(A)$, so $xy \in B_{00}(A)$, and likewise $yx \in B_{00}(A)$. So $(xy - yx)^* \in B_{00}(A)$ and hence $xy - yx = 0$ for all $y \in S'$. Thus $x \in S''$ and the result follows.

(i) We have $Z = A'$, $Z' = A$ so $Z = Z''$, and clearly $Z = Z^*$, so the result follows from (iv).

(ii) Suppose C is a maximal commutative *-subalgebra of A, then by maximality C is closed and $C' = C$, so $C = C''$ and the result follows from (iv).

(iii) Let B be a closed *-subalgebra of eAe, then since A is QW^*

any double centraliser on B is given by some $x \in B_{00}(A)$. Since $B \subset eAe$ we have $y \in B_0(A) \Rightarrow ey, ye \in B_0(A)$ and $x \in B_{00}(A) \Rightarrow exe \in B_{00}(A)$.

But for $z \in A$ we have

$$zexe = (zx)e = zx$$
$$exez = e(xz) = xz$$

so by the uniqueness of x in $B_{00}(A)$ we have $x = exe$.

Thus $x \in eAe$ and so $x \in B_{00}(eAe)$. Hence eAe is QW^*.

References

1. L. Gillman, and M. Jerison, *Rings of Continuous Functions*, van Nostrand, Inc., Princeton, N.J. (1960).

2. B. E. Johnson, *Centralisers in topological algebras*, Ph.D. dissertation, Cambridge. (1961).

3. ———, *An introduction to the theory of centralisers*, Proc. London Math. Soc. (3) **14** (1964), 299-320.

4. I. Kaplansky, *Projections in Banach algebras*, Ann. of Math. **53** (1951), 235-249.

5. C. E. Rickart, *General Theory of Banach Algebras*, van Nostrand, Inc., Princeton, N.J. (1960).

6. I. E. Segal, *Irreducible representations of operator algebras*, Bull. Amer. Math. Soc. **53** (1947), 73-88.

St. John's College, Cambridge

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

CLASSES OF RECURSIVE FUNCTIONS BASED ON ACKERMANN'S FUNCTION

ROBERT W. RITCHIE

Grzegorczyk has defined an increasing sequence of classes $\mathscr{E}^n$ of functions with the properties that $\mathscr{E}^3$ is the class of elementary functions of Csillag-Kalmar and $\cup \mathscr{E}^n$ is the class of primitive recursive functions. Further, $\mathscr{E}^{n+1}$ properly contains $\mathscr{E}^n$, if $n>2$ then $\mathscr{E}^{n+1}$ contains a "universal function" over all one-argument functions in $\mathscr{E}^n$, and a sequence of functions $g_n(x, y)$ in terms of which the $\mathscr{E}^n$ are defined has the property that each $g_{n+1}(x, x)$ (eventually) majorizes all the one-argument functions in $\mathscr{E}^n$.

The functions $g_n(x, y)$ are defined by somewhat artificial nested recursions, and Grzegorczyk poses the following question: "Can the same theorems be proved for classes $\mathscr{F}^n$ as for the classes $\mathscr{E}^n$?" Here $\mathscr{F}^n$ differs from $\mathscr{E}^n$ only in substituting a more natural function $f_n'(x, y)$ for each $g_n(x, y)$ in the definition of the class. In this paper, we answer his question affirmatively. Indeed, we prove that $\mathscr{F}^n = \mathscr{E}^n$ for all $n \geqq 0$, and further, $f'_{n+1}(x, x)$ eventually majorizes all the one-argument functions in $\mathscr{F}^n$.

In the first section below, we define the functions $f_n(x, y)$ (trivial variants of Grzegorczyk's $f_n'(x, y)$) which we shall use in place of $g_n(x, y)$, and develop various properties of these functions. In the second section, we define and study a sequence $\mathscr{C}_n$ of classes of one-argument functions. Each $\mathscr{C}_n$ is defined from $f_n(x, x)$ and other initial functions by composition and pure iteration as studied by Robinson [7]. In the third section, we define $\mathscr{F}_n$ (slight modifications of Grzegorczyk's $\mathscr{F}^n$) and establish various properties of these classes including that $\cup \mathscr{F}_n$ is the class of all primitive recursive functions. In the final section we establish the equality of $\mathscr{F}_n$ and $\mathscr{E}^n$, and then discuss Grzegorczyk's $f_n'(x, y)$ and prove that $\mathscr{F}^n$ also equals $\mathscr{E}^n$.

1. **The functions $f_n(x, y)$.** In [1], Ackermann defines a function of three variables which he shows is not primitive recursive. He obtains this function by considering the functions $x + y$, xy and x^y, observing that each is obtained from the preceding by a recursive definition and generalizing this process. Let us depart slightly from [1] and generalize the sequence of three functions as follows:

Received January 25, 1964. This work was supported by National Science Foundation grant GP-1851.

DEFINITION 1.1. For each $n \geqq 0$, define the function $f_n(x, y)$ by

$$\left.\begin{aligned} f_0(x, y) &= x + 1 \\ f_1(x, y) &= x + y \\ f_2(x, y) &= x \cdot y \\ f_{n+1}(x, 0) &= 1 \\ f_{n+1}(x, y + 1) &= f_n(x, f_{n+1}(x, y)) \end{aligned}\right\} \quad \text{for all } n \geqq 2 .$$

This differs from Ackermann's generalization in that we begin with $x + 1$, not $x + y$, and also we set $f_n(x, 0)$ equal to 1 instead of x for all $n \geqq 4$. The resulting difference between $f_n(x, y)$ and $\varphi(x, y, n)$ as defined in [1, p. 120] (or $\xi(n, y, x)$ as the function is written in [5, p. 272]) is minor as the following, which results from Lemma 1.1, shows

$$\begin{aligned} f_{n+1}(x, y) = \varphi(x, y, n) = \xi(n, y, x) \text{ for } n = 0, 1, 2, \text{ all } x, y ; \\ f_{n+1}(x, y + 1) \leqq \varphi(x, y, n) = \xi(n, y, x) < f_{n+1}(x, y + 2) \\ \text{for } n \geqq 3,\ x \geqq 2, \text{ all } y . \end{aligned}$$

These changes from Ackermann's definition were made here because starting with $f_0(x, y) = x + 1$ aligns the subscripts or our functions with those of Grzegorczyk, and letting $f_n(x, 0)$ be 1 instead of x simply seems more natural to the author.

Let us begin our study of the functions $f_n(x, y)$ by establishing three lemmas which yield information on the rate of growth of these functions. These three lemmas are closely related to Ackermann's results (properties III and IIIB, IV, and V of [1, pp. 121–2]).

LEMMA 1.1. *For all* $n \geqq 1$, $x \geqq 2$, *and* $y \geqq 0$, $f_n(x, y + 1) > f_n(x, y) \geqq y + 1$. *Hence* $f_n(x, z) > f_n(x, y) \geqq y + 1$ *for all* $z > y$.

Proof. For $n = 1$ this simply states that $x + y + 1 > x + y \geqq y + 1$. Let us assume that the result is true for n and establish its validity for $n + 1$. Since $f_{n+1}(x, 1) = x > f_{n+1}(x, 0) = 1$, the result holds for $y = 0$. Let us prove it for all y by induction. Assume $f_{n+1}(x, y + 1) > f_{n+1}(x, y) \geqq y + 1$; then $f_{n+1}(x, y + 2) = f_n(x, f_{n+1}(x, y + 1)) \geqq f_{n+1}(x, y + 1) + 1 > f_{n+1}(x, y + 1)$ by the assumed validity of the theorem at n. Further $f_{n+1}(x, y + 1) > y + 1$, or $f_{n+1}(x, y + 1) \geqq y + 2$ by the assumption immediately above.

LEMMA 1.2. *For all* $n \geqq 0$, $x \geqq 1$, *and* $y \geqq 1$, $f_n(x, y) < f_n(x + 1, y)$; *hence* $f_n(x, y) < f_n(z, y)$ *for all* $z > x$.

Proof. Immediate for $n = 0, 1$ and 2. Let us now assume that we have established the theorem for n, and show then that it holds for $n + 1$. Since for all y $f_{n+1}(1, y) = 1$ and $f_{n+1}(2, y) \leqq 2$ (using

$f_{n+1}(2, 1) = 2$ and Lemma 1.1), we have the result if $x = 0$ or $x = 1$. Thus let us assume $x \geqq 2$. Now $f_{n+1}(x, 1) = x < x + 1 = f_{n+1}(x + 1, 1)$ so the conclusion holds for $y = 1$. Assume it for y. Then it also holds for $y + 1$ since

$$\begin{aligned} f_{n+1}(x + 1, y + 1) &= f_n(x + 1, f_{n+1}(x + 1, y)) \\ &> f_n(x, f_{n+1}(x + 1, y)) \\ &> f_n(x, f_{n+1}(x, y)) \\ &= f_{n+1}(x, y + 1) . \end{aligned}$$

Here we used the assumption of the lemma at n and then the assumption for y at $n + 1$ together with Lemma 1.1. This completes the proof.

LEMMA 1.3. *For all* $n \geqq 2, x \geqq 2, y \geqq 0$; $f_n(x, y) \leqq f_{n+1}(x, y)$.

Proof. $f_{n+1}(x, 0) = 1 \geqq f_n(x, 0)$ for all $n \geqq 2$

$$\begin{aligned} f_{n+1}(x, y + 1) &= f_n(x, f_{n+1}(x, y)) \\ &\geqq f_n(x, y + 1) . \end{aligned}$$

The last step follows since $f_{n+1}(x, y) \geqq y + 1$ and $f_n(x, z) \geqq f_n(x, y + 1)$ for any $z \geqq y + 1$, both by Lemma 1.1.

These three lemmas will be used freely in the work below. Rather than refer to them explicitly, we shall often simply refer to the "monotonicity properties" of the $f_n(x, y)$.

We need one further important property of the functions $f_n(x, y)$. In Theorems 2.2 and 2.3 in the next section, we shall need to know that $f_n(f_n(x, k), l) \leqq f_n(x, m)$ for some m dependent only upon k and l whenever x is sufficiently large. In Theorem 1.1 below we find that we can take $m = f_{n-1}(k, l)$. The proof of this comprises the next five lemmas and we get the companion result

$$f_{n-1}(f_m(x, y), f_m(x, z)) \leqq f_m(x, y + z)$$

for all $m \geqq n$ automatically in the proof. The remainder of this section is devoted exclusively to the sequence of lemmas leading to the proof of Theorem 1.1.

LEMMA 1.4. *For all* $m \geqq 3$ *and all* x, y, *and* z,

$$f_2(f_m(x, y), f_m(x, z)) \leqq f_m(x, y + z) .$$

Proof by induction on m. In the case $m = 3$, the lemma is immediate since $f_3(x, y) = x^y$, and $f_2(x, y)$ is multiplication of x by y. Let us assume that the lemma is true for a fixed m (and for all x, y, and

z). Under this assumption we shall prove that $f_{m+1}(x, y)\cdot f_{m+1}(x, z) \leqq f_{m+1}(x, y + z)$ for all (y, z) with $y \leqq z$ which suffices by symmetry.

First we note that since $f_{m+1}(x, 0) = 1$, we have $f_{m+1}(x, 0)\cdot f_{m+1}(x, z) \leqq f_{m+1}(x, 0 + z)$ for all $z \geqq 0$. We now proceed by induction on y, assuming that we know the result for some y and all $z' \geqq y$ and proving it for $y + 1$ and all $z \geqq y + 1$. We begin this verification by noting that the cases $x = 0$ and $x = 1$ satisfy the desired inequality since $f_{m+1}(0, y)$ is 0 or 1 as y is odd or even, and $f_{m+1}(1, y) = 1$ for all y, so that we may assume $x \geqq 2$. Consider a value of $z \geqq y + 1 > 0$,

$$\begin{aligned} f_{m+1}(x, y + 1)\cdot f_{m+1}(x, z) &= f_m(x, f_{m+1}(x, y))\cdot f_m(x, f_{m+1}(x, z - 1)) \\ &\leqq f_m[x, f_{m+1}(x, y) + f_{m+1}(x, z - 1)] \\ &\leqq f_m[x, f_{m+1}(x, y)\cdot f_{m+1}(x, z)] \\ &\leqq f_m(x, f_{m+1}(x, y + z)) \\ &= f_{m+1}(x, (y + 1) + z) . \end{aligned}$$

Here we have used, in order, the assumption of the lemma for the fixed value m, the fact that $a + b \leqq a(b + 1)$ for all $a, b \geqq 1$ along with the monotonicity of $f_{m+1}(x, z)$ to see that $f_{m+1}(x, y) + f_{m+1}(x, z - 1) \leqq f_{m+1}(x, y)\cdot f_{m+1}(x, z)$, the monotonicity of $f_m(x, w)$, and finally the assumption that $f_{m+1}(x, y)\cdot f_{m+1}(x, z) \leqq f_{m+1}(x, y + z)$ for our fixed value of y and any larger z.

LEMMA 1.5. *For all* $x \geqq 2, k \geqq 0, l \geqq 0$; $f_3(f_3(x, k), l) = f_3(x, k\cdot l)$.

We omit the straightforward inductive proof of this lemma since it merely states that $(x^k)^l = x^{k\cdot l}$.

The preceding two lemmas form the basis of an inductive proof of the two properties we desire and state explicitly in Theorem 1.1 below. The next two lemmas form the inductive parts of the proofs of each of the properties. Unfortunately, we seem to need both properties at each step in the induction, and the details become somewhat involved.

LEMMA 1.6. *Let* $n \geqq 4$ *and* $x \geqq 2$. *Assume that, for all* $k \geqq 2$ *and* $l \geqq 0$, $f_{n-1}(f_{n-1}(x, k), l) \leqq f_{n-1}(x, f_{n-2}(k, l))$ *and further that for all* $m \geqq n$ *and for all* y *and* z

$$f_{n-2}[f_m(x, y), f_m(x, z)] \leqq f_m(x, y + z) .$$

Then, for all $m \geqq n$ *and for all* y *and* z *we have*

$$f_{n-1}[f_m(x, y), f_m(x, z)] \leqq f_m(x, y + z) .$$

Proof by induction on m. First, let us note directly from the equations $f_n(x, 0) = 1$ and $f_n(1, y) = 1$ that the cases $y = 0$ and $z = 0$ are immediate. Hence we shall assume now that both y and z are positive.

Consider the case $n = m$.

$$\begin{aligned} & f_{n-1}\{f_n(x, y), f_n(x, z)\} \\ & \quad = f_{n-1}\{f_{n-1}[x, f_n(x, y - 1)], f_n(x, z)\} \\ & \quad \leqq f_{n-1}\{x, f_{n-2}[f_n(x, y - 1), f_n(x, z)]\} \\ & \quad \leqq f_{n-1}\{x, f_n(x, (y + z) - 1)\} \\ & \quad = f_n(x, y + z) . \end{aligned}$$

Here we used the first hypothesis with $k = f_n(x, y - 1)$ and $l = f_n(x, z)$ and then the second hypothesis, and, of course, monotonicity.

Now let us assume the result for $m - 1$ and prove it for m. As in Lemma 1.4 we shall prove the result for pairs (y, z) with $z \geqq y$ and leave the remaining case to the reader. We have already noted that the result holds for pairs $(0, z)$. Let us now induce on y (still assuming the result for all pairs (y, z) at $m - 1$). Assuming we have the result for all $z \geqq y - 1$ at m, we now establish the result for all pairs (y, z) with $z \geqq y$.

$$\begin{aligned} & f_{n-1}\{f_m(x, y), f_m(x, z)\} \\ & \quad = f_{n-1}\{f_{m-1}[x, f_m(x, y - 1)], f_{m-1}[x, f_m(x, z - 1)]\} \\ & \quad \leqq f_{m-1}\{x, f_m(x, y - 1) + f_m(x, z - 1)\} \\ & \quad \leqq f_{m-1}\{x, f_2[f_m(x, y - 1) + 1, f_m(x, z - 1)]\} \\ & \quad \leqq f_{m-1}\{x, f_{n-2}[f_m(x, y - 1) + 1, f_m(x, z - 1)]\} \\ & \quad \leqq f_{m-1}\{x, f_{n-2}[f_m(x, y), f_m(x, z - 1)]\} \\ & \quad \leqq f_m(x, y + z) . \end{aligned}$$

Here we have used, in order, the assumption of the lemma at $m - 1$, the fact that $a + b \leqq (a + 1)b$, monotonicity ($f_2(a, b) \leqq f_{n-2}(a, b)$ and then $f_m(x, y - 1) < (f_m(x, y))$, the second hypothesis of the lemma, and monotonicity.

LEMMA 1.7. *Let* $n \geqq 4, x \geqq 2, k \geqq 2$ *and* $l \geqq 0$. *Assume that* $f_{n-1}[f_n(x, y), f_n(x, z)] \leqq f_n(x, y + z)$ *for all* y *and* z. *Then*

$$f_n(f_n(x, k), l) < f_n(x, f_{n-1}(k, l)) .$$

Proof by induction on l. If $l = 0$, the lemma is immediate. Assuming it for l, we now establish the case $l + 1$.

$$\begin{aligned} & f_n(f_n(x, k), l + 1) \\ & \quad = f_{n-1}\{f_n(x, k), f_n[f_n(x, k), l]\} \\ & \quad < f_{n-1}\{f_n(x, k), f_n[x, f_{n-1}(k, l)]\} \\ & \quad \leqq f_n\{x, k + f_{n-1}(k, l)\} \\ & \quad \leqq f_n\{x, f_{n-2}[k, f_{n-1}(k, l)]\} = f_n(x, f_{n-1}(k, l + 1)) . \end{aligned}$$

Here we used the assumption of the lemma at l, the hypothesis with $y = k$ and $z = f_{n-1}(k, l)$, and monotonicity to see that $f_2(k, f_{n-1}(k, l)) \leqq f_{n-2}(k, f_{n-1}(k, l))$.

We may now put the preceding four lemmas together into an inductive proof of the following theorem, with which we conclude our preliminary discussion of the functions $f_n(x, y)$.

THEOREM 1.1. *For all $n \geqq 3$, $x \geqq 2$, $k \geqq 2$ and $l \geqq 0$; $f_n(f_n(x, k), l) \leqq f_n(x, f_{n-1}(k, l))$. (In fact the inequality is strict if $n > 3$.) Further, for all y and z and for all $m \geqq n$, $f_{n-1}(f_m(x, y), f_m(x, z)) \leqq f_m(x, y + z)$.*

Proof by induction on n. Lemmas 1.4 and 1.5 establish the two statements for $n = 3$. In the inductive step, we assume both statements for $n - 1$, and establish the second statement for n by Lemma 1.6. Then the first statement follows for n by Lemma 1.7.

2. **The classes $\mathscr{E}_n$ of one-argument functions.** In this section, we shall make use of iteration and a particular set of pairing functions, both due to Robinson, in defining a sequence of classes of one-argument primitive recursive functions. Let us being by stating some definitions.

DEFINITION 2.1. The function $f(x)$ is said to be defined from $g(x)$ by *iteration* (more precisely, "pure iteration with no parameter beginning at 0") if $f(0) = 0$ and $f(x + 1) = g(f(x))$. We shall sometimes write $f(x) = g^x(0)$. The function $f(x^{(n)})$ is said to be defined by the *substitution* of $h_1(x^{(n)}), \cdots, h_m(x^{(n)})$ into $g(y^{(m)})$ if

$$f(x^{(n)}) = g(h_1(x^{(n)}), \cdots, h_m(x^{(n)})) .$$

Here either m or n is allowed to be 0. We have used the notation $x^{(n)}$ for $x_1, \cdots, x_n$, which we shall employ freely below.

We shall make heavy use of the pairing functions J, K and L used in [7, § 4]. These are defined as follows:

$$\begin{aligned} J(x, y) &= ((x + y)^2 + x)^2 + y \\ K(z) &= E[z^{1/2}] \\ L(z) &= E(z) . \end{aligned}$$

Here $E(z)$ is the excess of z over a square ($z - [z^{1/2}]^2$), and $[z^{1/2}]$ is the largest integer not exceeding the square root of z. The important properties of these pairing functions are:

(1) $K(J(x, y)) = x$ and $L(J(x, y)) = y$ (but not necessarily $J(K(z), L(z)) = z$)

(2) $J(0, 0) = 0$, hence $K(0) = L(0) = 0$.

(3) If $L(z + 1) \neq 0$ then $K(z + 1) = K(z)$ and $L(z + 1) = L(z) + 1$ (even in those cases in which z is not a value of $J(x, y)$).

The significance of (3) is the ease with which one can obtain all the pairs $(x, y), \cdots, (x, 0)$ from $J(x, y + 1) = z$. Namely, merely list $(K(z - 1), L(z - 1)), (K(z - 2), L(z - 2), \cdots$ until you reach a pair of the form $(K(z - k), 0)$. This is used in reducing the number of arguments in recursive definitions in [7, §§ 3, 4] (see Lemma 3.1 below).

Let us immediately generalize these functions J, K and L which establish a one-to-one correspondence between pairs of (nonnegative) integers and some integers. These new functions $J^{(n)}, M_1^{(n)}, \cdots, M_n^{(n)}$ will establish a correspondence between n-tuples and some integers, and will be used heavily in § 3, especially in connection with the notion of "associate" defined there. The definitions are inductive and self-explanatory (given our convention that $x^{(n)}$ means $x_1, \cdots, x_n$).

$$\begin{aligned} J^{(1)}(z) &= M_1^{(1)}(z) = z \\ J^{(n+1)}(x^{(n+1)}) &= J[J^{(n)}(x^{(n)}), x_{n+1}] \\ M_i^{(n+1)}(z) &= M_i^{(n)}(K(z)), \text{ for all } 1 \leqq i \leqq n \\ M_{n+1}^{(n+1)}(z) &= L(z) . \end{aligned}$$

It is immediate from these definitions that the following generalizations of (1), (2), and (3) (which will be used in § 3) hold:

(1′) $M_i^{(n)}(J^{(n)}(x^{(n)}) = x_i$ for all $1 \leqq i \leqq n$.

(2′) $J^{(n)}(0, \cdots, 0) = 0$, hence $M_i^{(n)}(0) = 0$ for all $1 \leqq i \leqq n$.

(3′) If $M_n^{(n)}(z + 1) \neq 0$ then $M_i^{(n)}(z + 1) = M_i^{(n)}(z)$ for all $1 \leqq i \leqq n - 1$ and $M_n^{(n)}(z + 1) = M_n^{(n)}(z) + 1$ (even in those cases in which z is not in the range of $J^{(n)}(x^{(n)})$).

We shall use one further notation, this taken essentially from [6]. We shall write "$\langle f(x) = 0 \to g(x); h(x)\rangle$" for "either $g(x)$ if $f(x) = 0$ or $h(x)$ otherwise". Given the functions $0^x = \langle x = 0 \to 1; 0\rangle$ and $\operatorname{sgn}(x) = \langle x = 0 \to 0; 1\rangle$, we can define $\langle f(x) = 0 \to g(x); h(x)\rangle$ as $g(x) \cdot 0^{f(x)} + h(x) \cdot \operatorname{sgn}(f(x))$. Hence we may note that $\langle f(x) = 0 \to g(x); h(x)\rangle$ is definable from $f(x)$, $g(x)$, $h(x)$, 0^x and $\operatorname{sgn}(x)$ by addition, multiplication and substitution of one-argument functions into one-argument functions.

Now we have the necessary background to define the classes $\mathscr{E}_n$ of one-argument functions. These will be defined to include the one-argument functions which we have just discussed and to be closed

under the operations above as well. We make these statements more precise in Theorem 2.1 below.

DEFINITION 2.2. We shall say the function $f(x^{(n)}, y)$ is defined from $g(x^{(n)})$, $h(x^{(n)}, y, z)$ and $j(x^{(n)}, y)$ by *limited recursion* (see [4]) if

$$f(x^{(n)}, 0) = g(x^{(n)})$$
$$f(x^{(n)}, y + 1) = h(x^{(n)}, y, f(x^{(n)}, y))$$
$$f(x^{(n)}, y) \leqq j(x^{(n)}, y) .$$

We shall say that the function $f(x)$ is defined from $h(x)$ and $j(x)$ by *limited iteration* if $f(x) = h^x(0)$ and $f(x) \leqq j(x)$.

DEFINITION 2.3. For each $n \geqq 2$, $\mathscr{E}_n$ is the smallest class of one-argument functions containing $S(x)$, $E(x)$, $[x^{1/2}]$ and $f_n(x, x)$ which is closed under addition ($f(x) = g(x) + h(x)$), multiplication ($f(x) = g(x) \cdot h(x)$), composition of one-argument functions ($f(x) = h(g(x))$) and limited iteration.

THEOREM 2.1. *Each class $\mathscr{E}_n$ contains*

(i) *the identity function* $I(x) = x$,
(ii) *the constant function* $k(x)$ *for each* $k \geqq 0$,
(iii) $K(x)$ *and* $L(x)$ (*in fact* $M_i^{(n)}(x)$ *for each* $n \geqq 1$ *and* $1 \leqq i \leqq n$),
(iv) $\operatorname{sgn}(x) = \langle x = 0 \to 0; 1 \rangle$,
(v) $0^x = \langle x = 0 \to 1; 0 \rangle$, *and*
(vi) $f_i(x, x)$ *for* $0 \leqq i \leqq 2$.

Further, each $\mathscr{E}_n$ is closed under

(vii) *pairing* ($f(x) = J^{(n)}(f_1(x), \cdots, f_n(x))$), *and*
(viii) *conditional expression* ($f(x) = \langle g(x) = 0 \to h(x); j(x) > \ $).

Proof.

(i) $I(x) = S^x(0)$, $I(x) \leqq S(x)$.

(ii) $0(x) = I^x(0)$, $0(x) \leqq I(x)$, and $k(x) = S^k(0(x))$ for each fixed integer k.

(iii) $K(x) = E[x^{1/2}]$, $L(x) = E(x)$ and each $M_i^{(n)}(x)$ is obtained by composition from these.

(iv) $\operatorname{sgn}(x) = 1^x(0)$, $\operatorname{sgn}(x) \leqq 1(x)$

(v) $0^x = E\{S^2[2(x) \cdot \operatorname{sgn}(x)]\}$

(vi) $f_0(x, x) = S(x)$
$f_1(x, x) = I(x) + I(x)$
$f_2(x, x) = I(x) \cdot I(x)$

(vii) $J(g(x), h(x)) = [(g(x) + h(x))^2 + g(x)]^2 + h(x)$, obtained by addition and multiplication from $g(x)$ and $h(x)$

(viii) $\langle g(x) = 0 \rightarrow h(x); j(x)\rangle = h(x)\cdot 0^{g(x)} + j(x)\cdot \operatorname{sgn}(g(x))$ obtained by composition, addition and multiplication.

THEOREM 2.2. *For any $n \geqq 2$ and any $f(x)$ in $\mathscr{G}_n$, there is a positive integer k such that, for every $x \geqq 2$, $f(x) < f_{n+1}(x, k)$.*

Proof. The theorem holds for the basic functions of $\mathscr{G}_n$, namely $S(x) < f_2(x, 2) \leqq f_{n+1}(x, 2)$, $E(x) < x = f_{n+1}(x, 1)$, $[x^{1/2}] < x = f_{n+1}(x, 1)$ and $f_n(x, x) = f_{n+1}(x, 2) < f_{n+1}(x, 3)$. Now assume inductively that $g(x) < f_{n+1}(x, k)$ and $h(x) < f_{n+1}(x, l)$ for k and l both greater than 1 (without loss of generality). We shall show that the four operations in terms of which $\mathscr{G}_n$ is defined preserve the boundedness property. If $f(x)$ is defined by limited iteration the result is obvious. If $f(x) = g(x) + h(x)$ or $g(x)\cdot h(x)$, then, since x, k and l are all at least 2, we have $f(x) < f_{n+1}(x, k)\cdot f_{n+1}(x, l) \leqq f_{n+1}(x, k + l)$ by Lemma 1.4. Finally, if $f(x) = h(g(x))$, we use Theorem 1.1 to see that

$$\begin{aligned} f(x) &< f_{n+1}(f_{n+1}(x, k), l) \\ &\leqq f_{n+1}(x, f_n(k, l)) \end{aligned}$$

as desired.

DEFINITION 2.4. The function $f(x)$ *majorizes* $g(x)$ if there is an integer k such that $f(x) > g(x)$ for all $x \geqq k$.

COROLLARY. *For every $n \geqq 2$, the function $f_{n+1}(x, x)$ majorizes every function in $\mathscr{G}_n$. Hence $\mathscr{G}_n \neq \mathscr{G}_m$ for $m \neq n$.*

The next theorem tells us that iteration (unlimited) does not lead out of the union of the $\mathscr{G}_n$. This, in connection with a result of Robinson [7 Th. 3], shows that $\cup \mathscr{G}_n$ is the class of all one-argument primitive recursive functions. However, we shall obtain that corollary result a little differently at the end of the next section, so we do not state it explicitly here.

We precede the theorem with a lemma which we shall use in its proof.

LEMMA 2.1. *For every $n \geqq 2$ and $k \geqq 1$ the function $h_{n,k}(x) = f_n(k, x)$ is in $\mathscr{G}_m$ for every $m \geqq n$; hence, in particular it is in $\mathscr{G}_n$.*

Proof by induction on n. For $n = 2$, we merely note that $h_{2,k}(x) = k(x)\cdot I(x)$. Let us assume that $h_{n,k}(x)$ is in $\mathscr{G}_m$ for all $m \geqq n$ and all k. Then define $h'_{n+1,k}(x)$ by:

$$h'_{n+1,k}(0) = 0$$
$$h'_{n+1,k}(x+1) = g_{n,k}(h'_{n+1,k}(x))$$
$$h'_{n+1,k}(x) \leqq f_{n+1}(k, x) \leqq f_{n+1}(k+x, k+x) \leqq f_{m+1}(k+x, k+x)$$

where $g_{n,k}(x) = h_{n,k}(\langle x = 0 \to 1; x\rangle)$, a function in $\mathscr{E}_m$ for every $m \geqq n$ by Theorem 2.1 (viii) and $f_{m+1}(k+x, k+x)$ is in $\mathscr{E}_{m+1}$ since it results from substituting $k(x) + I(x)$ into the basic function $f_{m+1}(x, x)$. Thus $h'_{n+1,k}(x)$ is in $\mathscr{E}_m$ for every $m \geqq n+1$, and $h_{n+1,k}(x) = \langle h'_{n+1,k}(x) = 0 \to 1; h'_{n+1,k}(x)\rangle$, is also in $\mathscr{E}_m$ as desired.

THEOREM 2.3. *If $f(x)$ is defined by iteration from $g(x)$, that is, $f(x) = g^x(0)$, and if $g(x)$ is in $\mathscr{E}_n$, then $f(x)$ is in $\mathscr{E}_{n+1}$.*

Proof. Let $g(x)$ be less than $f_{n+1}(x, k)$, for all $x \geqq 2$ and some $k \geqq 2$ (by Theorem 2.2). We shall show that, for any $x \geqq 2$ and $y \geqq 0$, $g^x(y) \leqq f_{n+1}(\bar{y}, f_{n+1}(k, x))$ where $\bar{y}$ is the maximum of y and 2. From this it will follow that $f(x) = g^x(0) \leqq f_{n+1}(2, f_{n+1}(k, x))$, which is, by monotonicity, at most $f_{n+1}(k, f_{n+1}(k, x))$. But $f_{n+1}(k, x)$ is a one-argument function $h(x)$ in $\mathscr{E}_{n+1}$ by the lemma. Our bounding function is then $h^2(x)$, again in $\mathscr{E}_{n+1}$. Thus it suffices to show that $g^x(y) \leqq f_{n+1}(\bar{y}, f_{n+1}(k, x))$. Here $g^x(y)$ is taken to mean the function defined by

$$g^0(y) = y,\ g^{x+1}(y) = g(g^x(y))\ .$$

Now $g^0(y) = y \leqq f_{n+1}(\bar{y}, f_{n+1}(k, 0))$, so let us establish our bound on $g^x(y)$ by induction on x. Assuming that $g^x(y) \leqq f_{n+1}(\bar{y}, f_{n+1}(k, x))$ for a fixed value of x and for all y, we have

$$\begin{aligned} g^{x+1}(y) &= g(g^x(y)) \\ &\leqq f_{n+1}(\overline{g^x(y)}, k) \\ &\leqq f_{n+1}[f_{n+1}(\bar{y}, f_{n+1}(k, x)), k] \\ &\leqq f_{n+1}[\bar{y}, f_n[f_{n+1}(k, x), k]] \\ &= f_{n+1}(\bar{y}, f_{n+1}(k, x+1))\ . \end{aligned}$$

Here we have used the first part of Theorem 1.1 and then the second part of Theorem 1.1 having noted that $k = f_{n+1}(k, 1)$.

COROLLARY. *The union of the classes $\mathscr{E}_n$, $n \geqq 2$, is closed under addition, multiplication, composition of one-argument functions and iteration.*

3. The classes $\mathscr{F}_n$ of primitive recursive functions. In this section we define classes $\mathscr{F}_n$ which are essentially the $\mathscr{F}^n$ introduced by Grzegorczyk (see the final paragraph of § 4). We shall use results

of Robinson to show that, for each $n \geqq 2$, $\mathscr{G}_n$ is the class of all one-argument functions in $\mathscr{F}_n$. In fact, we shall show that $\mathscr{F}_n$ is the class of all functions which have their one-argument "associates" under the pairing functions included in $\mathscr{G}_n$. We shall make this precise in Definition 3.2 and Theorem 3.4 below.

We begin by defining the classes $\mathscr{F}_\pi$.

DEFINITION 3.1. For each $n \geqq 0$, let $\mathscr{F}_n$ be the smallest class of functions containing the successor function $S(x)$, the zero function $0(x)$, the identity functions $U_i^n(x_1, \cdots, x_n) = x_i$, and $f_n(x, y)$ which is closed under substitutions and limited recursion (see Definitions 2.1 and 2.2).

It is an immediate consequence of this definition and the monotonicity of the $f_n(x, y)$ that $f_n(x, y)$ can be defined by limited recursion in every class $\mathscr{F}_m$ where m is at least as large as n. Hence, $\mathscr{F}_n$ is contained in $\mathscr{F}_m$ for every $m \geqq n \geqq 0$. That this containment is proper will follow from the relation between $\mathscr{F}_n$ and $\mathscr{G}_n$ to be established below. We now prove a theorem which, though it consists only in listing several functions in $\mathscr{F}_n$, establishes part of this relationship.

THEOREM 3.1. *The functions* $x + y, x \cdot y, x \dot{-} y, [x^{1/2}], Ex$ *and* $f_n(x, x)$ *are in* $\mathscr{F}_n$ *for every* $n \geqq 2$. *Hence* $\mathscr{G}_n \subset \mathscr{F}_n$ *for every* $n \geqq 2$.

Proof. We define a sequence of functions in $\mathscr{F}_n$ which includes the desired functions. Each definition is either by substitution or limited recursion from functions already known to be in $\mathscr{F}_n$.

$x + 0 = x$ $\qquad$ $x \cdot 0 = 0$

$x + (y + 1) = S(x + y)$ $\qquad$ $x \cdot (y + 1) = x + (x \cdot y)$

$x + y \leqq f_n(S(S(x)), y)$, (see Lemma 1.3); $\qquad$ $x \cdot y \leqq f_n(x, y)$, (by Lemma 1.3)

$P(0) = 0$ $\qquad$ $x \dot{-} 0 = I(x)$ $\qquad$ $0^0 = 1$

$P(x + 1) = I(x) = U_1^1(x)$ $\qquad$ $x \dot{-} (y + 1) = P(x \dot{-} y)$ $\qquad$ $0^{x+1} = 0(0^x)$

$P(x) \leqq I(x)$ $\qquad$ $x \dot{-} y \leqq U_1^2(x, y)$ $\qquad$ $0^x \leqq S(x)$

$$x^2 = I(x) \cdot I(x) f_n(x, x) = f_n(I(x), I(x)) E(x) = x \dot{-} [x^{1/2}]^2$$

$$[0^{1/2}] = 0$$

$$[(x + 1)^{1/2}] = [x^{1/2}] + 0^{([x^{1/2}]+1)^2 \dot{-} (x+1)}$$

$$[x^{1/2}] \leqq I(x) .$$

All the basic functions of $\mathscr{G}_n$ are among these. Hence we need only show that $\mathscr{F}_n$ is closed under the operations used in defining $\mathscr{G}_n$ to complete the proof. But this is immediate since the addition and

multiplication functions are in $\mathscr{F}_n$ and it is closed under substitution and limited recursion.

We wish to show that $\mathscr{G}_n$ is precisely the class of all one-argument functions in $\mathscr{F}_n$. In fact, we shall show that $\mathscr{F}_n$ is just the class of all functions which can be obtained from $\mathscr{G}_n$ by use of the pairing functions alone. A critical step in this process is the replacement of definition by limited recursion with definition by its special case, limited iteration. The replacement of primitive recursion by iteration was carried out in [7, §§ 3 and 4] following earlier work by Péter. We modify that replacement slightly so that it will yield immediate bounds on the functions in question and thus apply to the limited recursions and iterations we are using. Lemma 3.1 follows [7] directly, but we include it here for completeness and for notational convenience. Lemma 3.2 differs slightly from Robinson's work; the difference is introduced so that the $f'(z)$ we obtain can be expressed in terms of the original function for all values of z, thus keeping the bound on $f'(z)$ for all z.

LEMMA 3.1. *Let* $f(x^{(n)}, y)$ *be defined by primitive recursion from* $g(x^{(n)})$ *and* $h(x^{(n)}, y, z)$:

$$f(x^{(n)}, 0) = g(x^{(n)})$$
$$f(x^{(n)}, y + 1) = h(x^{(n)}, y, f(x^{(n)}, y)) .$$

Then $F(z) = f(M_1^{(n+1)}(z), \cdots, M_{n+1}^{(n+1)}(z))$, *and hence* $f(x^{(n)}, y) = F(J^{(n+1)}(x^{(n)}, y))$, *where* $F(z)$ *is defined by*

$$F(0) = g(0^{(n)})$$
$$F(z + 1) = H(z, F(z))$$

where

$$H(z, w) = \langle M_{n+1}^{(n+1)}(z + 1) = 0 \rightarrow g[M_1^{(n+1)}(z + 1), \cdots, M_n^{(n+1)}(z + 1)];$$
$$h(M_1^{(n+1)}(z), \cdots, M_{n+1}^{(n+1)}(z), w)\rangle ,$$

Proof by induction on z. Since $M_i^{(n+1)}0 = 0$, we have the desired result for $z = 0$. Assuming the result for z, we consider $F(z + 1)$. There are two cases according as $M_{n+1}^{(n+1)}(z + 1) = 0$ or not. In the former case the result is immediate by the definition of $f(x^{(n)}, 0)$ as $g(x^{(n)})$. In the latter case the result follows by induction and an observation that, by property (3′) of the pairing functions, $M_i^{(n+1)}(z+1) = M_i^{(n+1)}(z)$ for all $1 \leqq i \leqq n$ and $M_{n+1}^{(n+1)}(z + 1) - 1 = M_{n+1}^{(n+1)}(z)$.

LEMMA 3.2. *Let* $F(z)$ *be defined by primitive recursion*:

$$F(0) = k$$
$$F(z + 1) = H(z, F(z)) .$$

Then $f'(z) = \langle z = 0 \to 0; J(z, F(z))\rangle$ *and hence* $F(z) = \langle z = 0 \to k; L(f'(z))\rangle$ *where* $f'(z)$ *is defined by pure iteration, namely:*

$$f'(0) = 0$$
$$f'(z+1) = J\{K(f'(z)) + 1, H'[K(f'(z)), L(f'(z))]\}$$

with

$$H'(z, w) = \langle z = 0 \to H(0, k); H(z, w)\rangle .$$

Proof. First we define $F'(z)$, a function which agrees with $F(z)$ except when $z = 0$:

$$F'(0) = 0$$
$$F'(z+1) = H'(z, F'(z)) .$$

Then we note that $f'(z) = J(z, F'(z))$ as desired.

These two lemmas combine into the following general theorem.

THEOREM 3.2. *Let* $f(x^{(n)}, y)$ *be defined by primitive recursion from* $g(x^{(n)})$ *and* $h(x^{(n)}, y, z)$, *where* n *is any nonnegative integer (possibly zero). Then there is a function* $f'(z)$ *such that*

$$f'(z) = \langle z = 0 \to 0; J(z, f(M_1^{(n+1)}(z), \cdots, M_{n+1}^{(n+1)}(z))\rangle ,$$

and hence

$$f(x^{(n)}, y) = \langle J^{(n+1)}(x^{(n)}, y) = 0 \to g(0^{(n)}); Lf'(J^{(n+1)}(x^{(n)}, y))\rangle$$

where $f'(x)$ *is defined by pure iteration, namely*

$$f'(0) = 0$$
$$f'(z+1) = H''(f'(z))$$

with

$$H''(w) = J\{K(w) + 1, H'(K(w), L(w))\} ,$$
$$H'(z, w) = \langle z = 0 \to H(0, g(0^{(n)}); H(z, w)\rangle ,$$
$$H(z, w) = \langle M_{n+1}^{(n+1)}(z+1) = 0 \to g[M_1^{(n+1)}(z+1), \cdots, M_n^{(n+1)}(z+1)];$$
$$h(M_1^{(n+1)}(z), \cdots, M_{n+1}^{(n+1)}(z), w\rangle .$$

Now we are prepared to prove the two theorems which relate the classes $\mathscr{T}_n$ and $\mathscr{C}_n$. We begin with the definition of an associate of a function.

DEFINITION 3.2. The function $g(x^{(m)})$ is called *the m-argument associate of* the function $f(x^{(n)})$ if $g(M_1^{(m)}(x), \cdots, M_m^{(m)}(x)) = f(M_1^{(n)}(x), \cdots, M_n^{(n)}(x))$ for all x.

Let us note two important facts about associates. First, if $g(x^{(m)})$ is an associate of $f(x^{(n)})$ it must be the function

$$f(M_1^{(n)}(J^{(m)}(x^{(m)})), \cdots, M_n^{(n)}(J^{(m)}(x^{(m)}))) .$$

Thus there is *at most* one m-argument associate of any given n-argument function. However, there need not be any. For example, the one-argument function $f(x)=x$ does not have a 2-argument associate. For, assuming that $g(x, y)$ is an associate, we have by the above that $g(x, y) = f(J(x, y)) = ((x + y)^2 + x)^2 + y$. But this function $g(x, y)$ does not satisfy the defining condition that, for every x, $g(K(x), L(x))= f(x)$. In particular $f(1) = 1$ but $K(1) = L(1) = 0$, hence $g(K(1), L(1)) = 0$; in fact $g(K(y^4), L(y^4)) = 0$ for all y since $K(y^4) = L(y^4) = 0$. (The reader may show easily that in general an n-argument function $f(x^{(n)})$ will have an m-argument associate if and only if $f(M_1^{(n)}(x), \cdots, M_n^{(n)}(x))= f(M_1^{(n)}(y), \cdots, M_n^{(n)}(y))$ whenever $M_i^{(m)}(x) = M_i^{(m)}(y)$ for all $1 \leqq i \leqq m$. However, we make no use of this fact here.)

We shall show, in the next two theorems, that $\mathscr{T}_n$ is precisely the class of all associates of functions in $\mathscr{G}_n$.

THEOREM 3.3. *The class of all associates of functions in $\mathscr{G}_n$ is closed under substitutions and limited recursion. Further, any function defined by primitive recursion from functions with associates in $\mathscr{G}_n$ has an associate in $\mathscr{G}_{n+1}$.*

Proof. First let us consider substitutions. Assume $g_i(x^{(m)}), 1 \leqq i \leqq n$, and $h(x^{(n)})$ are associates of functions $g_i^*(x)$ and $h^*(x)$ in $\mathscr{G}_k$. Then $f(x^{(m)}) = h(g_1(x^{(m)}), \cdots, g_n(x^{(m)}))$ is an associate of $f^*(x) = h^*(J^{(n)}(g_1^*(x), \cdots, g_n^*(x)))$. But $f^*(x)$ is in $\mathscr{G}_k$ by Theorem 2.1 (vii) and closure under substitutions.

Now let us turn to recursion. Assume $g(x^{(n)})$ and $h(x^{(n)}, y, z)$ are associates of functions $g^*(x)$ and $h^*(x)$ in $\mathscr{G}_k$, and let $f(x^{(n)}, y)$ be defined from $g(x^{(n)})$ and $h(x^{(n)}, y, z)$ by primitive recursion. By Theorem 3.2, $f^*(x) = f(M_1^{(n+1)}(x), \cdots, M_{n+1}^{(n+1)}(x)) = \langle x = 0 \rightarrow g(0^{(n)}); L(f'(x))\rangle$, where the function $f'(x)$ is defined by pure iteration from $H''(w)$ as described in the theorem. But $H''(w)$ is in $\mathscr{G}_k$, as we now show. It will be in $\mathscr{G}_k$ if $H'(K(w), L(w))$ is, by the closure of $\mathscr{G}_k$ under conditional expression. For the same reason, $H'(K(w), L(w))$ will be in $\mathscr{G}_k$ if $H(K(w), L(w))$ is. But this last function is precisely $\langle M_{n+1}^{(n+1)}(K(w) + 1) = 0 \rightarrow g^*(K(K(w) + 1)); h^*(w)\rangle$ which is in $\mathscr{G}_k$. Hence $H''(w)$ is in $\mathscr{G}_k$ so that an application of Theorem 2.3 shows that $f'(x)$ is in $\mathscr{G}_{k+1}$. But then $f^*(x)$ is also in $\mathscr{G}_{k+1}$ as desired.

We have now only to show closure of the associates of $\mathscr{G}_k$ under *limited* recursion. To do so, we reason as follows. If, in fact, $f'(x)$

is bounded by some function in $\mathscr{G}_k$, then $f'(x)$ and hence $f^*(x)$ will also be in $\mathscr{G}_k$. But exactly this is the case if $f(x^{(n)}, y)$ is defined by limited recursion with a bounding function with an associate, say $j^*(x)$, in $\mathscr{G}_k$. For then $f'(x) \leqq J(x, j^*(x))$. This completes the proof of the theorem.

THEOREM 3.4. *For every $n \geqq 2$, the class $\mathscr{F}_n$ is the class of all associates of all functions in $\mathscr{G}_n$.*

Proof. First, we have $\mathscr{G}_n \subset \mathscr{F}_n$ by Theorem 3.1. But every associate of any $f(x)$ in $\mathscr{G}_n$ is obtained by substituting $J^{(k)}(x^{(k)})$ into $f(x)$ by the remarks following the definition of asssociate. Since $\mathscr{F}_n$ is closed under substitutions by definition, all these associates are in $\mathscr{F}_n$. To prove the converse, we note that, by Theorem 3.3, it suffices to show that the basic functions of $\mathscr{F}_n$ are associates of functions in $\mathscr{G}_n$. This is obvious for all the basic functions except $f_n(x, y)$, and we shall prove that $f_n(x, y)$ has an associate in $\mathscr{G}_n$ by induction on n. We first observe that the function $K(x) \cdot L(x)$ is an associate of $f_2(x, y)$ which is in $\mathscr{G}_2$, so that the result holds for $n = 2$. The inductive step follows from Theorem 3.3 since $f_{n+1}(x, y)$ is defined from $1(x)$ and $h(x, y, z) = f_n(U_1^3(x, y, z), U_3^3(x, y, z))$ by primitive recursion.

We conclude this section with three corollaries which spell out the relationship of the classes $\mathscr{G}_n$, $\mathscr{F}_n$, and the class of all primitive recursive functions.

COROLLARY. *For every $n \geqq 2$, $\mathscr{G}_n$ is precisely the class of all one variable functions in $\mathscr{F}_n$.*

Proof. $\mathscr{G}_n$ is the class of all one-variable associates of functions in $\mathscr{G}_n$.

COROLLARY. *For every $n \geqq 2$, $\mathscr{G}_n$ is properly contained in $\mathscr{G}_{n+1}$. For every $n \geqq 0$, $\mathscr{F}_n$ is properly contained in $\mathscr{G}_{n+1}$.*

Proof. As commented immediately after the definition of the classes $\mathscr{F}_n$, each $\mathscr{F}_n$ is contained in $\mathscr{F}_{n+1}$. But, by the corollary to Theorem 2.2, $f_{n+1}(x, x)$ is in $\mathscr{F}_{n+1}$ but not $\mathscr{F}_n$.

COROLLARY. *The union of the increasing sequence of classes $\mathscr{F}_n$ is the class of all primitive recursive functions.*

Proof. The primitive recursive functions can be characterized (see, for example, [3, p. 49]) as the smallest class of functions containing

$S(x)$, $0(x)$ and $U_i^n(x_1, \cdots, x_n)$ which is closed under substitutions and primitive recursion. The very definition of the $\mathscr{F}_n$ provides containment of these basic functions and also closure under substitutions. Closure of the union under primitive recursion is provided by Theorems 3.3 and 3.4.

4. Equality of $\mathscr{F}_n$ and $\mathscr{E}^n$. In this section we shall show that $\mathscr{F}_n = \mathscr{E}^n$ for every $n \geqq 0$. We begin with Grzegorczyk's definition of of $g_n(x, y)$ and $\mathscr{E}^n$.

DEFINITION 4.1. We define $g_n(x, y)$ by setting

$$\left.\begin{aligned} g_0(x, y) &= x + 1 \\ g_1(x, y) &= x + y \\ g_2(x, y) &= (x + 1)(y + 1) \\ g_{n+1}(x, 0) &= g_n(x + 1, x + 1) \\ g_{n+1}(x, y + 1) &= g_{n+1}(g_{n+1}(x, y), y) \end{aligned}\right\} n \geqq 2 .$$

Then we define $\mathscr{E}^n$ as the smallest class of functions containing $S(x)$, $0(x)$, $U_i^n(x_1, \cdots, x_n)$ and $g_n(x, y)$ which is closed under substitutions and limited recursion.

We have written $g_n(x, y)$ for the function which Grzegorczyk denoted by $f_n(y, x)$, and have trivially modified the list of basic functions of $\mathscr{E}^n$ (Grzegorczyk observed that it was sufficient to take $S(x)$, $U_1^2(x, y)$ and $U_2^2(x, y)$ as basic since $0(x)$ and $U_i^n(x_1, \cdots, x_n)$ could be defined from these by substitutions).

Let us include for reference a list of monotonicity properties of the $g_n(x, y)$ which were established in [4, p. 28]. We shall use these frequently below in relating $f_n(x, y)$ to $g_n(x, y)$, and refer to them simply as "[4, p. 28]".

(1) $g_n(x, y) > x$ for $n \geqq 2$

(2) $g_n(x, y + 1) > g_n(x, y)$ for $n \geqq 1$

(3) $g_n(x + 1, y) > g_n(x, y)$ for $n \geqq 0$.

Each inequality is easily proved from the definitions.

Now let us prove a lemma to be used in relating $\mathscr{F}_n$ and $\mathscr{E}^n$. The first assertion of the lemma was noted and used by Grzegorczyk in the case $n = 3$ [4, p. 29] and by Axt for all $n \geqq 3$ [2, p. 58].

LEMMA 4.1. *For every* $n \geqq 3$, $g_n(x, y) = h_n(x, 2^y)$ *where* $h_n(x, y)$ *is defined by*

$$h_n(x, 0) = x$$
$$h_n(x, y + 1) = g_{n-1}(h_n(x, y) + 1, h_n(x, y) + 1) .$$

Further, $f_n(x, y) < g_n(x, y)$ *for every* $n \geqq 2, x, y \geqq 0$.

Proof. A completely straightforward induction on z establishes the identity $h_n(h_n(x, y), z) = h_n(x, y + z)$ for all x, y and z. Then another straightforward induction on y establishes that $g_n(x, y) = h_n(x, 2^y)$ completing the first part of the lemma. Turning to the second assertion, we observe that $f_2(x, y) < g_2(x, y)$ is immediate. Let us assume now that $n \geqq 3$. In case $x = 0$, we have $f_n(0, 0) = 1 < g_{n-1}(1, 1) = g_n(0, 0)$, by [4, p. 28]. Further $f_n(0, y + 1) \leqq 1 < g_n(0, y + 1)$ by another application of [4, p. 28]. In case $x = 1$, we have $f_n(1, y) = 1 < g_{n-1}(1, 1) = g_n(0, 0) < g_n(1, y)$ by [4, p. 28]. Thus we need to establish $f_n(x, y) \leqq g_n(x, y)$ only for $n \geqq 3$, $x \geqq 2$. In these cases we shall in fact show that $f_n(x, y) < h_n(x, y) < g_n(x, y)$. That $h_n(x, y) < g_n(x, y)$ follows immediately from the definition of $h_n(x, y)$, the property ([4, p. 28]) that $g_{n-1}(x, y) > x$ and the first part of this lemma. To see that $f_n(x, y) < h_n(x, y)$ we proceed inductively on y assuming that we already have $f_{n-1}(x, y) < g_{n-1}(x, y)$. We begin by noting that $h_n(x, y) \geqq x$ by induction on y. Now $f_n(x, 0) = 1 < h_n(x, 0)$, and then

$$\begin{aligned} f_n(x, y + 1) &= f_{n-1}(x, f_n(x, y)) \\ &< f_{n-1}(x, h_n(x, y)) \\ &< g_{n-1}(x, h_n(x, y)) \\ &< g_{n-1}(h_n(x, y) + 1, h_n(x, y) + 1) \\ &= h_n(x, y + 1) . \end{aligned}$$

This completes our proof.

Now we establish the equality of $\mathscr{E}^n$ and $\mathscr{F}_n$.

THEOREM 4.1. *For every* $n \geqq 0$, $\mathscr{F}_n = \mathscr{E}^n$.

Proof. For $n = 0, 1$ the two classes have identical definitions. For $n \geqq 2$, to establish equality it will suffice to show that $f_n(x, y)$ is in $\mathscr{E}^n$ and that $g_n(x, y)$ is in $\mathscr{F}_n$. But, by the preceding lemma, $f_n(x, y) < g_n(x, y)$ for all $n \geqq 2$, and this allows us to define $f_n(x, y)$ by limited recursion in $\mathscr{E}^n$. Namely, $f_2(x, y)$ is defined by primitive recursion from $f_1(x, y)$, which we know to be in $\mathscr{E}^1 \subseteqq \mathscr{E}^2$. (That $\mathscr{E}^n \subseteqq \mathscr{E}^{n+1}$ is proved in [4, pp. 33–4]). Since $f_2(x, y)$ is bounded by a function in $\mathscr{E}^2$, $f_2(x, y)$ is then in $\mathscr{E}^2$. Assuming $f_n(x, y)$ is in $\mathscr{E}^n$, we find that $f_{n+1}(x, y)$ is in $\mathscr{E}^{n+1}$ by the same argument. Now we know that $\mathscr{F}_n \subseteqq \mathscr{E}^n$ and need only to define $g_n(x, y)$ within $\mathscr{F}_n$ to complete the proof. We shall proceed by induction on n, beginning at

$n = 2$ where $g_2(x, y) = f_2(x+1, y+1)$, a function in $\mathscr{F}_n$. Assume the result holds for n. By Theorems 3.3 and 3.4 we see that $h_{n+1}(x, y)$ is in $\mathscr{F}_{n+1}$ since it is defined by primitive recursion from functions in $\mathscr{F}_n$; namely $I(x)$ and $g_n(U_3^3(x, y, z) + 1, U_3^3(x, y, z) + 1)$. But then $g_{n+1}(x, y) = h_{n+1}(x, 2^y)$ is in $\mathscr{F}_{n+1}$ since $2^y = f_3(2, y)$ is in $\mathscr{F}_{n+1}$ for all $n \geqq 2$.

* * *

Technically, Grzegorczyk did not pose his question about the classes $\mathscr{F}_n$ based upon the functions $f_n(x, y)$. Instead he asked about classes $\mathscr{F}^n$ whose definition is obtained from that of $\mathscr{F}_n$ by replacing $f_n(x, y)$ by $f'_n(x, y)$. The definition of these functions $f'_n(x, y)$ is in turn obtained from that of $f_n(x, y)$ by deleting the special definition of $f_2(x, y)$ and letting the general case cover that as well. Namely, (except for the fact that we are consistently reversing the order of his variables) Grzegorczyk defined $f'_n(x, y)$ as

$$\left.\begin{aligned} f'_0(x, y) &= x + 1 \\ f'_1(x, y) &= x + y \\ f'_{n+1}(x, 0) &= 1 \\ f'_{n+1}(x, y+1) &= f'_n(x, f'_{n+1}(x, y)) \end{aligned}\right\} \text{for all } n \geqq 1 .$$

This makes $f'_2(x, y) = x \cdot y + 1$, $f'_3(x, y) = x^y + \cdots + x^1 + 1 = (x^{y+1} - 1)/(x - 1)$, and $f'_n(x, y)$ for larger n relatively inexpressible. It is easily seen that $f_n(x, y) \leqq f'_n(x, y)$ for all $n \geqq 2$ and all x and y. We leave the details of the proof, which by now should suggest themselves immediately, to the reader. From this it follows that $f'_{n+1}(x, x)$ majorizes every one-argument function in $\mathscr{F}_n$. The containment of $\mathscr{F}_n$ in $\mathscr{F}^n$ follows since $f_n(x, y)$ is definable within $\mathscr{F}^n$ by limited recursion. But then $\mathscr{F}_n = \mathscr{F}^n$ since $f'_n(x, y)$ is in $\mathscr{F}_n$ by induction on n using Theorems 3.3 and 3.4 and the fact that $f'_{n+1}(x, y)$ is defined from $f'_n(x, y)$ by recursion.

Bibliography

1. Wilhelm Ackermann, *Zum Hilbertschen Aufbau der reellen Zahlen*, Math. Annalen **99** (1928), 118-133.

2. Paul Axt, *Enumeration and the Grzegorczyk hierarchy*, Zeitscharift fur math. Logik und Grundlagen der Math. **9** (1963), 53-65.

3. Martin Davis, *Computability and Unsolvability*, McGraw-Hill, New York, N.Y., (1958), 25+210 pp.

4. Andrej Grzegorczyk, *Some classes of recursive functions*, Rozprawy Matematyczne, Warsaw, (1953), 46 pp.

5. S. C. Kleene, *Introduction to metamathematics*, Van Nostrand, Princeton, N. J., (1952), 10+550 pp.

6. John McCarthy, *Recursive functions of symbolic expressions and their computation by machine*, Part I, Communications of the Assoc. for Computing Machinery **3** (1960), 184-194.

7. R. M. Robinson, *Primitive recursive functions*, Bull. Amer. Math. Soc. **53** (1947), 925-942.

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

PROPERTIES OF SOLUTIONS OF Nth ORDER LINEAR DIFFERENTIAL EQUATIONS

THOMAS L. SHERMAN

The purpose of this paper is to establish zero properties for solutions of the general n^{th} order linear differential equation

$$(1) \qquad L_n y = \sum_{i=0}^{n} r_i(x) y^{(i)} = 0 ,$$

where $r_n(x) \neq 0$ and all the coefficients are continuous.

This work was stimulated by the recent investigations of fourth-order equations by Leighton and Nehari [11], Barrett [3, 4, 5] and Howard [8]. For example, if $\eta_1(a)$ is defined to be the first point $b > a$ for which there exist a solution of

$$(2) \qquad (r(x)y'')'' - p(x)y = 0 \quad (r, p > 0, r \in C_2, p \in C)$$

with four zeros in $[a, b]$ (counting multiplicities), then it is known [11] that there is a solution which vanishes with double zeros at a and $\eta_1(a)$ ($\eta_1(a)$ is termed the first (right) conjugate point of a). If $\mu_1(a)$ is defined to be the first point $b > a$ for which there exists a non-trivial solution of (2) satisfying the conditions $y(a) = y'(a) = (ry'')(b) = (ry'')'(b) = 0$, then it is known [3] that $a < \mu_1(a) < \eta_1(a)$ (a is called the first (left) focal point of $\mu_1(a)$). These results were subsequently extended to general even-order self-adjoint equations by Reid [13] and Hunt [9].

In § 1 we derive some basic results which are used in later sections.

In § 2 we extend the definition of conjugate point found in [11]. Using this definition we obtain generalizations (in a direction different from that of [9] and [13]) of results of [11].

In § 3 we define a notion of focal point for (1) and extend the discussions of [3, 4, 5] to obtain results similar to those of § 2.

In § 4 we develop an eigenvalue relationship from which we can easily obtain a generalization of the following theorem of Leighton and Nehari:

THEOREM. *Equation* (2) *is disconjugate* (*i.e.*, $\eta_1(a)$ *does not exist*)

Received June 12, 1963. Supported, in part, by the National Science Foundation under grant NSF G23719. This paper was presented to the Mathematics Department of the University of Utah in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

on $[a, \infty)$ *if and only if the minimum eigenvalue of the associated problem*

$$(ry'')'' - py = \lambda y\ , \quad y(a) = y'(a) = y(b) = y'(b) = 0\ ,$$

is positive for all $b > a$.

1. **Preliminaries.** We begin our discussion with a number of theorems concerning properties of real functions.

For reference we state the following basic lemma, which was proven by Leighton and Nehari [11].

LEMMA 1. *Let* $u(x)$ *and* $v(x)$ *be of class* C_1 *in* (a, b) *and let* $v(x)$ *be of constant sign* $(\neq 0)$ *in the interval. If* $x = \alpha$ *and* $x = \beta$ $(a < \alpha < \beta < b)$ *are consecutive zeros of* $u(x)$, *then there exists a constant* μ *such that the function* $u(x) - \mu v(x)$ *has at least a double zero in* (α, β).

This lemma has been extended by Azbelev and Chaluk [1] to the case where β is allowed to be a double zero of $u(x)$ and a simple zero of $v(x)$. We shall use the method of the latter together with Lemma 1 to prove

THEOREM 1. *Let* $u(x)$ *be a function such that* $u(x)$ *has a zero of order* $n \geq 1$ *at* $x = a$ *and a zero of order* $m \geq 1$ *at* $x = b$, *and* $u(x)$ *is of constant sign* $(\neq 0)$ *in* (a, b). *Let* $v(x)$ *be a function such that* $v(x)$ *has a zero of order* $n_1 < n$ *at* $x = a$ *and a zero of order* $m_1 < m$ *at* $x = b$, *and* $v(x)$ *is of constant sign* $(\neq 0)$ *in* (a, b). *Further suppose* $u(x)$ *and* $v(x)$ *are both of class* $C_M[a, b]$, *where* $M = \max(n_1, m_1)$. *Then there exists a linear combination* $z(x)$ *of* $u(x)$ *and* $v(x)$ *such that* $z(x)$ *has a double zero in* (a, b).

Proof. We may as well suppose that $u(x)$ and $v(x)$ are both positive on (a, b). Let $c \in (a, b)$ and let

$$w(x) = v(x) - \frac{v(c) + 1}{u(c)} u(x)\ .$$

There exists an $h_1 \in (a, c)$ such that $w(h_1) > 0$, since, by the zero properties of $u(x)$ and $v(x)$, $w^{(n_1)}(a) = v^{(n_1)}(a) > 0$ and $w^{(i)}(a) = 0$ $(i = 0, 1, \cdots, n_1 - 1)$. Similarly there is an $h_2 \in (c, b)$ such that $w(h_2) > 0$, since $(-1)^{m_1} w^{(m_1)}(b) > 0$ and $w^{(i)}(b) = 0$ $(i = 0, 1, \cdots, m_1 - 1)$. Also $w(c) = -1 < 0$; thus $w(x)$ has a zero in (a, c) and a zero in (c, b). Since $w(x)$ is continuous, there exists α and β $(a < \alpha < c < \beta < b)$ such that $w(\alpha) = w(\beta) = 0$ and $w(x) < 0$ on (α, β). Applying Lemma 1 to the functions $w(x)$ and $v(x)$ on (α, β), have our result.

THEOREM 2. *Suppose* $r(x)f(x) \in C_n$ *and* $f(x) \in C_n$ *in a neighborhood of the point* $x = a$. *If* $f^{(i)}(a) = 0$ $(i = 0, 1, \cdots, n)$ *and if* $r(x)$ *is a bounded function in a neighborhood of* a, *then* $(rf)^{(i)}(a) = 0$ $(i = 0, 1, \cdots, n)$.

Proof. The theorem is trivial for $n = 0$. For $n = 1$,

$$\begin{aligned} |(rf)'(a)| &= \left| \lim_{h \to 0} \frac{(rf)(a+h) - (rf)(a)}{h} \right| \\ &= \left| \lim_{h \to 0} \frac{r(a+h)f(a+h) - r(a)f(a)}{h} \right| \\ &\leqq \overline{\lim_{h \to 0}} |r(a+h)| \left| \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \right| \\ &= M |f'(a)| = 0 , \end{aligned}$$

where $M = \overline{\lim}_{x \to a} |r(x)|$. Now using the formula

$$(rf)^{(k)}(a) = \lim_{h \to 0} \frac{\sum_{i=0}^{k} (-1)^i C(k, i)(rf)(a + (k-i)h)}{h^k}$$

we shall show that $(rf)^{(k)}(a) = 0$ $(1 \leqq k \leqq n)$. Applying L'Hospital's rule, and using the zero properties of $f(x)$ at $x = a$ and the boundedness of r at $x = a$, we have

$$\begin{aligned} |(rf)^{(k)}(a)| &= \left| \lim_{h \to 0} \sum_{i=0}^{k} \frac{(-1)^i C(k, i) r(a + (k-i)h) f(a + (k-i)h)}{h^k} \right| \\ &\leqq M \sum_{i=0}^{k-1} \left| \frac{(k-i)^k}{i!\,(k-i)!} \lim_{h \to 0} \frac{f^{(k-1)}(a + (k-i)h) - f^{(k-1)}(a)}{(k-i)h} \right| \\ &= M |f^{(k)}(a)| \sum_{i=0}^{k-1} \frac{(k-i)^k}{i!\,(k-i)!} . \end{aligned}$$

Since $f^{(k)}(a) = 0$, the result follows.

COROLLARY. *Suppose* $f(x) \in C_n$, $r(x)f(x) \in C_n$, *and* $r(x) \in C$ *in a neighborhood of* $x = a$. *If* $f^{(i)}(a) = 0$ $(i = 0, 1, \cdots, n-1)$, $(rf)^{(n)}(a) = 0$, *and* $r(a) \neq 0$, *then* $f^{(n)}(a) = 0$.

Proof. Since $r(x)$ is continuous at $x = a$, we can replace M by $r(a)$, "$\leqq$" by "$=$", and drop the absolute value signs in the inequality in the proof of Theorem 2, and obtain a valid equality; the result is then immediate.

In the next theorem and in the remainder of the paper we shall deal with the equation

$$(1) \qquad L_n y = \sum_{i=0}^{n} r_i(x) y^{(i)}(x) = 0 ,$$

where $r_n(x) \neq 0$ and the coefficients $r_i (i = 0, 1, \cdots, n)$ are real valued continuous functions.

The following basic existence theorem is apparently sufficiently well-known so that references to it in the literature are scarce; we include it here for the sake of completeness.

THEOREM 3. *If* $a_i \in [a, b]$ $(i = 1, \cdots, k \leqq n-1)$ *and* $a_1 < a_2 < \cdots < a_k$, *then there exists a nontrivial solution of* (1) *which satisfies the boundary conditions*

$$\begin{aligned} y^{(i_1)}(a_1) &= 0 \qquad i_1 = n_{11}, n_{12}, \cdots, n_{1p_1} \\ y^{(i_2)}(a_2) &= 0 \qquad i_2 = n_{21}, n_{22}, \cdots, n_{2p_2} \\ &\vdots \\ y^{(i_k)}(a_k) &= 0 \qquad i_k = n_{k1}, n_{k2}, \cdots, n_{kp_k} \end{aligned}$$

where $0 \leqq n_{j1} < n_{j2} < \cdots < n_{jp_j} < n$ $(j = 1, \cdots, k)$ *and* $\sum_{j=1}^{k} p_j \leqq n - 1$.

Proof. Let $y_1(x), y_2(x), \cdots, y_n(x)$ be a fundamental set of solutions of (1). We wish to find a nontrivial set of constants $C_1, \cdots, C_n$ such that $y = C_1 y_1 + \cdots + C_n y_n$ satisfies the above boundary conditions.

Applying the boundary conditions to $y(x)$, we have

$$\begin{aligned} y^{(n_{11})}(a_1) &= C_1 y_1^{(n_{11})}(a_1) + \cdots + C_n y^{(n_{11})}(a_1) = 0 \\ &\vdots \\ y^{(n_{1p_1})}(a_1) &= C_1 y_1^{(n_{1p_1})}(a_1) + \cdots + C_n y_n^{(n_{1p_1})}(a_1) = 0 \\ &\vdots \\ y^{(n_{21})}(a_2) &= C_1 y_1^{(n_{21})}(a_2) + \cdots + C_n y_n^{(n_{21})}(a_2) = 0 \\ &\vdots \\ y^{(n_{2p_2})}(a_2) &= C_1 y_1^{(n_{2p_2})}(a_2) + \cdots + C_n y_n^{(n_{2p_2})}(a_2) = 0 \\ &\vdots \\ y^{(n_{k1})}(a_k) &= C_1 y_1^{(n_{kp_k})}(a_k) + \cdots + C_n y_n^{(n_{kp_k})}(a_k) = 0 \\ &\vdots \\ y^{(n_{kp_k})}(a_k) &= C_1 y_1^{(n_{k1})}(a_k) + \cdots + C_n y_n^{(n_{k1})}(a_k) = 0 . \end{aligned}$$

This is a system of $\sum_{j=1}^{k} p_j$ (i.e., less than n) homogeneous equations in n unknowns $C_1, \cdots, C_n$, and so there always exists a nontrivial solution.

We shall make repeated use of this theorem throughout the remainder of this paper.

2. Zero properties of solutions. In this section we shall be concerned with zero properties of solutions of (1). The principal stimulation for this investigation is the work of Leighton and Nehari [11]. Their concern was with the equation

$$(r(x)y'')'' - p(x)y = 0 , \tag{2}$$

where $r(x) \in C_2$, $p(x) \in C$, and $r(x) > 0$. They defined the first conjugate point of the point a, denoted by $\eta_1(a)$, to be the minimum point $b > a$ such that there is a nontrivial solution $y(x)$ of (2) vanishing at $x = a$, with four zeros, counted in their multiplicities, in $[a, b]$. They showed that if there is a nontrivial solution with four zeros on $[a, \infty)$, then $\eta_1(a)$ exists. It was then shown that, if $p(x) > 0$ in (2), a solution which vanishes four times in $[a, \eta_1(a)]$ has double zeros at $x = a$ and $x = \eta_1(a)$ and does not vanish in $(a, \eta_1(a))$. Barrett [3, 4, 5] then extended this work to equations of the form

$$[(r(x)y'')' + q(x)y']' + p(x)y = 0 ,$$

where $r(x)$, $q(x)$ and $p(x)$ are all continuous. Hunt [9] and Reid [13] in turn extended this work to self-adjoint differential equations of order $2n$, defining conjugate point as the minimum point $b > a$ such that there exists a nontrivial solution with a zero of order n at a and a zero of order n at b.

Hanan [7], using a definition of conjugate point similar to that of Leighton and Nehari, investigated the third order equation

$$y''' + p(x)y'' + q(x)y' + r(x)y = 0 .$$

Investigations similar to Hanan's were undertaken by Azbelev and Chaluk [1].

We shall here be interested in the extension of this work to the equation (1), using the conjugate point concept of Leighton and Nehari. Some of the results obtained in this section have been reported independently by A. Ju. Levin [12] who uses the methods of Green's functions. It should be noted that self-adjointness and assumptions on the signs of the coefficients have been of fundamental importance in the work cited above (except for that of Levin); in general we make no such assumptions here. Following Barrett [2, 3], we make the following definition:

DEFINITION 1. Equation (1) is said to be disconjugate in $[a, \infty)$ if there exists no nontrivial solution of (1) which vanishes at a, and has n zeros, counting multiplicities, in $[a, \infty)$.

We now prove a basic theorem, special cases of which have been

used by Leighton and Nehari and Barrett, and a special case of which was proved by Hanan.

THEOREM 4. *Suppose that for each positive integer k there is an nth order linear boundary operator U_k, operating at points $b_{k1}, \cdots, b_{kj_k}$, such that*

(i) (1) *has a nontrivial solution $y_k(x)$ satisfying $U_k(y_k(x)) = 0$;*

(ii) *there exist points $a_{k1}, \cdots, a_{kn}$ such that* $y_k^{(i-1)}(a_{ki}) = 0$ $(i = 1, \cdots, n)$;

(iii) $\min_i b_{ki} \leqq \min_i a_{ki} < \max_i a_{ki} \leqq \max_i b_{ki}$;

(iv) $\min_i b_{ki} \leqq \min_i b_{ji}$ *and* $\max_i b_{ji} \leqq \max_i b_{ki}$ *if* $k < j$.

Then $\lim_{k\to 0}(\max_i b_{ki} - \min_i b_{ki}) > 0$.

REMARK. If $y_k(x)$ is a nontrivial solution of (1) having n zeros on an interval, counted in accordance with their multiplicities, then repeated application of Rolle's Theorem shows the existence of the desired set of points $a_{k1}, a_{k2}, \cdots, a_{kn}$.

Proof of Theorem 4. Let $A_k = \min_i a_{ki}$, $B_k = \max_i a_{ki}$, $C_k = \min_i b_{ki}$ and $D_k = \max_i b_{ki}$. Conditions (iv) and (iii) assure us that $\{D_k\}$ forms a decreasing sequence bounded below; hence there is a D such that $D_k \to D$. Similarly $\{C_k\}$ forms an increasing sequence bounded above; hence there is a C such that $C_k \to C$ and $D \geqq C$. We need only show that equality does not hold.

Let $\phi_1(x), \phi_2(x), \cdots, \phi_n(x)$ be the fundamental set of solutions of (1) satisfying $\phi_{i+1}^{(j)}(D) = \delta_{ij}$ $(i, j = 0, 1, \cdots, n-1)$. Then $y_k(x)$ can be written

$$y_k(x) = \sum_{i=1}^{n} c_{ik}\phi_i(x) .$$

We normalize $y_k(x)$ by letting

$$\bar{y}_k(x) = \frac{y_k(x)}{\left(\sum_{i=1}^{n} c_{ik}^2\right)^{1/2}} = \sum_{i=1}^{n} d_{ik}\phi_i(x) ,$$

so that $\sum_{i=1}^{n} d_{ik}^2 = 1$. Then $\bar{y}_k(x)$ is a solution of (1) possessing the same zero properties as $y_k(x)$; in particular, $\bar{y}_k(x)$ satisfies condition (ii). Further, on any closed interval about D, $\{\bar{y}_k(x)\}$ is a uniformly bounded and equicontinuous family. Hence there is a uniformly convergent subsequence $\{\bar{y}_{k_i}(x)\}$ whose limit function $\bar{y}(x)$ is a solution of (1).

Now suppose $D = C$; then by condition (iii)

$$A = \lim_{k\to 0} A_k = \lim_{k\to 0} B_k = B .$$

But now by condition (ii), for each $j(0 \leqq j < n)$, A is a limit point of zeros of $\{\bar{y}_{k_i}^{(j)}(x)\}$; hence $\bar{y}^{(j)}(A) = 0$ $(j = 0, 1, \cdots, n-1)$, and so $\bar{y}(x) \equiv 0$. This is impossible, since $\sum_{i=1}^{n} d_{ik}^2 = 1$, which completes the proof.

We are now in a position to make

DEFINITION 2. The first conjugate point $\eta_1(a)$ of the point a is the smallest number $b > a$ such that there exists a nontrivial solution of (1) which vanishes at a and has n zeros, counting multiplicities in $[a, b]$.

We now establish one of the main results of this section.

THEOREM 5. *If* (1) *is not disconjugate, then* $\eta_1(a)$ *exists, and there is a nontrivial solution of* (1) *which has a total of at least* n *zeros at* a *and* $\eta_1(a)$ *and does not vanish in* $(a, \eta_1(a))$.

Proof. The existence of $\eta_1(a)$ is guaranted by Theorem 4. Let $Y = \{y(x) \mid y(x)$ is a nontrivial solution of (1) which has n zeros in $[a, \eta_1(a)]\}$, $R = \{r \mid$ there is a $y(x) \in Y$ with a zero of order r at $x = a\}$, $n_0 = \max r \in R$, $M = \{m \mid$ there is a $y \in Y$ with a zero of order n_0 at $x = a$, and m zeros in $(a, \eta_1(a))\}$, $\bar{m} = \max m \in M$, and let $\phi(x) \in Y$ be a solution with $\bar{m}$ zeros in $(a, \eta_1(a))$ and a zero of order n_0 at $x = a$. Let $a = a_0 < a_1 < \cdots < a_p = \eta_1(a)$ be the zeros of $\phi(x)$ on $[a, \eta_1(a)]$, with respective multiplicities $n_0, n_1, \cdots, n_p$. Then $n_1 + n_2 + \cdots + n_{p-1} = \bar{m}$. Let $\bar{n} = n_0 + \cdots + n_{p-1}$. We claim that $\phi(x)$ exhibits the property asserted. For suppose the theorem false for $\phi(x)$, i.e., suppose $p > 1$. We shall show the existence of a $\psi(x) \in Y$ with a zero of order n_0 at $x = a$ and a zero of order $\bar{m} + 1$ in $(a, \eta_1(a))$, which will contradict the maximality of $\bar{m}$. To this end let us consider a nontrivial solution $z(x)$ of (1) with zeros at $a_0, a_1, \cdots, a_p$ of multiplicitis $n_0 + 1, n_1, \cdots, n_{p-1} - 1$, $n - \bar{n} - 1$ respectively; such a solution exists (note that $p > 1$ implies $n_0 < n - 1$) since only $n - 1$ zeros (counting multiplicities) are specified. Since the zero of $z(x)$ at a_0 is of order $n_0 + 1$, $z(x) \notin Y$, and so $z(x)$ does not vanish on (a_{p-1}, a_p), nor are any of the zeros of multiplicity greater than that specified. Hence we can apply Theorem 1 to $z(x)$ and $\phi(x)$ on (a_{p-1}, a_p) to obtain a linear combination $\psi(x)$ with a double zero at c, say, on (a_{p-1}, a_p). Now $\psi(x)$ has zeros at $a_0, a_1, \cdots, a_{p-2}$, a_{p-1}, c, a_p of multiplicities $n_0, n_1, \cdots, n_{p-2}, n_{p-1} - 1, 2, n - \bar{n} - 1$; i.e., $\psi(x) \in Y$ with a zero at a of order n_0, and with a total multiplicity of $\bar{m} + 1$ of zeros on $(a, \eta_1(a))$, contradicting the maximality of $\bar{m}$; this established the theorem.

COROLLARY. *Let* $\phi(x)$ *be a nontrivial solution of* (1) *with zeros*

of orders n_0 and at least $n - n_0$ at a and $\eta_1(a)$ respectively, where n_0 is maximal. Then $\phi(x)$ is essentially unique i.e., unique except for multiplication by a constant factor). (We note that a similar result holds for a $\psi(x) \in Y$, not vanishing on $(a, \eta_1(a))$, and with a zero of maximal order at $\eta_1(a)$).

Proof. Let $\psi(x)$ be any such solution. By the maximality of n_0 we have $\phi^{(n_0)}(a) \neq 0 \neq \psi^{(n_0)}(a)$. Consider the solution

$$\chi(x) = \psi^{(n_0)}(a)\phi(x) - \phi^{(n_0)}(a)\psi(x) .$$

$\chi(x)$ clearly has a zero of order $n + 1$ at $x = a$ and a zero of order $n - n_0$ at $x = \eta_1(a)$; hence, by the maximality of n_0, $\chi(x) \equiv 0$.

In order to examine the behavior of $\eta_1(a)$ as a function of a we prove the following theorems. The principal theorem, Theorem 7, in this sequence was obtained by Leighton and Nehari [11] for equation (2) and by Hunt [9] for the special definition of η_1 mentioned at the beginning of this section, for an equation more general than that in [11].

THEOREM 6. *Let a be any point for which $\eta_1(a)$ exists; then, for any $b < a$, $\eta_1(b)$ exists.*

Proof. Let a_0 be the largest number $c \geqq a$ such that $\eta_1(c) = \eta_1(a)$ (the existence of a_0 is guaranteed by Theorem 4). Let $\phi(x)$ be a solution such that $\phi(a_0) = 0$ and $\phi(x)$ has n zeros in $[a_0, \eta_1(a)]$, where the order of the zero of $\phi(x)$ at $\eta_1(a)$ is minimal over the set of all such solutions. Let $m_0, m_1, \cdots, m_r$ be the orders of the zeros of $\phi(x)$ at $a_0 < a_1 < \cdots < a_r = \eta_1(a)$ respectively. We are supposing m_r is minimal. Let $M = m_0 + \cdots + m_{r-1} \leqq n - 1$. Clearly $m_r \geqq n - M$.

Now suppose the theorem is not true. Let $\psi(x)$ be a nontrivial solution of (1) such that $\psi(b) = 0$ and $\psi(x)$ has zeros of multiplicities $m_0, m_1, \cdots, m_{r-2}, m_{r-1} - 1, n - M - 1$ at $a_0, a_1, \cdots, a_{r-2}, a_{r-1}, a_r$ respectively. We have specified (counting multiplicities) $n - 1$ zeros of $\psi(x)$ in $[b, \eta_1(a)]$; hence, since we are supposing $\eta_1(b)$ does not exist, $\psi(x)$ has zeros only at these points and only of the multiplicities specified. By Theorem 1, applied to $\phi(x)$ and $\psi(x)$, there must exist a linear combination $\chi(x)$ of $\phi(x)$ and $\psi(x)$ with a double zero in (a_{r-1}, a_r). Further, $\chi(x)$ has zeros of orders $m_0, m_1, \cdots, m_{r-2}, m_{r-1} - 1, n - M - 1$ at $a_0, a_1, \cdots, a_{r-1}, a_r$ respectively. Hence $\chi(x)$ has n zeros on $[a_0, \eta_1(a)]$. Now if $r = 1 = m_0 = m_r$ then $\chi(x)$ has a double zero and is also a solution of a second order equation; hence $\chi(x) \equiv 0$. If $r = 1 = m_0$ and $m_r > 1$ then $\chi(\eta_1(a)) = 0$ and $\chi(x)$ has n zeros in $(a_0, \eta_1(a)]$ which contradicts the maximality of a_0. If $r = 1 = m_r$ and $m_0 > 1$ then $\chi(x)$ has n zeros

in $[a_0, \eta_1(a))$ and vanishes at a_0, which contradicts the fact that $\eta_1(a^0) = \eta_1(a)$. In all other cases $\chi(x)$ has n zeros in $[a_0, \eta_1(a)]$ with a zero of order $n - M - 1 < m_r$ at $\eta_1(a)$, which contradicts the minimality of m_r.

COROLLARY 1. *Let a be any point for which $\eta_1(a)$ exists; then, for any $b < a$, there exists a solution which vanishes at b and which vanishes $n - 1$ times, counting multiplicity, in $[a, \eta_1(a)]$.*

Proof. The function $\psi(x)$ constructed in the proof of Theorem 5 is such a function.

COROLLARY 2. *Let a be any point for point for which $\eta_1(a)$ exists; then no solution has more than $n - 1$ zeros in $[a, \eta_1(a))$.*

Proof. Suppose this is not true. Then there exists a solution with n zeros in $[a, \eta_1(a))$. Hence there is a $b \in [a, \eta_1(a))$ such that $\eta_1(b) < \eta_1(a)$; clearly such a b cannot equal a. Then, by Corollary 1, there exists a solution $\phi(x)$ which vanishes at a and which vanishes $n - 1$ times, counting multiplicity, in $[b, \eta_1(b)]$. Hence $\phi(x)$ vanishes at a and has n zeros in $[a, \eta_1(b)]$, which contradicts the definition of $\eta_1(a)$.

It should be noted that, for $n = 2$, Corollaries 1 and 2 combine to give the classicial Sturm separation theorem.

COROLLARY 3. *$\eta_1(a)$ is a nondecreasing function of a.*

THEOREM 7. *$\eta_1(a)$ is an increasing function of a.*

Proof. Suppose the theorem is not true. We know by Corollary 3 that $\eta_1(a)$ is a nondecreasing function of a. Thus there exist $a, \eta_1(a), b$ and $\eta_1(b)$ such that $a < b$ and $\eta_1(a) = \eta_1(b)$. Hence, by Theorem 5 and Corollary 3, for every $x \in [a, b]$, $\eta_1(x)$ exists and equals $\eta_1(a)$. For each $\alpha \in [a, b]$ let $\phi_\alpha(x)$ be the essentially unique solution which vanishes at least n times at α and $\eta_1(a)$, where the order of the zero at $\eta_1(a)$ is maximal over all such solutions, and $\phi_\alpha(x) \neq 0$ for $x \in (\alpha, \eta_1(a))$, as in the corollary to Theorem 5.

Let $S_i = \{\alpha \mid \phi_\alpha(x)$ has a zero of exactly order i at $x = \eta_1(a)\}$, $(i = 1, \cdots, n - 1)$. Clearly $\bigcup_{i=1}^{n-1} S_i = [a, b]$. Hence there is a subinterval $[c, d] \subseteq [a, b]$ and an $i (1 \leqq i < n)$ such that S_i is dense in $[c, d]$ (see Simmons [14], page 74). Pick the maximum such i and denote it by m; denote the subinterval $[c, d]$ corresponding to this value of i by $[a_m, b_m]$.

Let $\{y_1, y_2, \cdots, y_n\}$ be the fundamental set of solution satisfying $y_{i+1}^{(j)}(\eta_1(a)) = \delta_{ij}$. Now the condition that S_m is dense in $[a_m, b_m]$ is equivalent to the condition that the determinant

$$W_{m+1} = \begin{vmatrix} y_{m+1} & y_{m+2} & \cdots & y_n \\ y'_{m+1} & y'_{m+2} & \cdots & y'_n \\ \vdots & & & \\ y_{m+1}^{(n-m-1)} & y_{m+2}^{(n-m-1)} & \cdots & y_n^{(n-m-1)} \end{vmatrix}$$

vanish at a dense set of points in $[a_m, b_m]$. This follows since $W_{m+1}(\alpha) = 0$ is the condition for a solution with a zero of order m at $\eta_1(a)$ to have a zero of order $n - m$ at a point $\alpha \neq \eta_1(a)$. Further, since W_{m+1} is continuous, we have $W_{m+1} \equiv 0$ in $[a_m, b_m]$. Moreover from the maximality of m we know that

$$W_{m+2} = \begin{vmatrix} y_{m+2} & y_{m+3} & \cdots & y_n \\ y'_{m+2} & y'_{m+3} & \cdots & y'_n \\ \vdots & & & \\ y_{m+2}^{(n-m-2)} & y_{m+3}^{(n-m-2)} & \cdots & y_n^{(n-m-2)} \end{vmatrix} \neq 0$$

on some subinterval $[a'_m, b'_m] \subseteq [a_m, b_m]$.

Now consider the differential equation

$$\begin{vmatrix} y & y_{m+2} & \cdots & y_n \\ y' & y'_{m+2} & \cdots & y'_n \\ \vdots & & & \\ y^{(n-m-1)} & y_{m+2}^{(n-m-1)} & \cdots & y_n^{(n-m-1)} \end{vmatrix} = 0$$

formed by substituting y for y_{m+1} in the equation $W_{m+1} \equiv 0$. This is an equation of order $n - m - 1$ with non-vanishing leading coefficient W_{m+2} on the interval $[a'_m, b'_m]$. The solutions are $y_{m+2}, y_{m+3}, \cdots, y_m$, a total of $n - m - 1$ solutions. These are linearly independent, since $W_{m+2} \neq 0$ on $[a'_m, b'_m]$; hence they form a solution basis. But y_{m+1} is also a solution, since $W_{m+1} \equiv 0$ on $[a'_m, b'_m]$; therefore $y_{m+1}, y_{m+2}, \cdots, y_n$ are linearly dependent on $[a'_m, b'_m]$, which contradicts the fact that $y_1, \cdots, y_n$ are solutions of (1) which are independent on $[a, b]$. This completes the proof.

3. Zero properties of derivatives of solutions. We now turn our attention to the behavior of the derivatives of solutions. The principal motivation for this type of investigation is the work of Barrett [2, 3, 4, 5], in which are references to earlier work in this field. Barrett [2] discussed the behavior of solutions of the equation

$$(p(x)y')' + f(x)y = 0$$

in relation to the boundary conditions $y(a) = y'(b) = 0$; the minimum such $b > a$ for which these conditions are satisfied was denoted by $\mu_1(a)$, and a was called a (left) focal point of the point $\mu_1(a)$. This discussion was extended to the equation (2), with a focal point being defined in terms of the first point $b > a$, for which conditions of the type $y(a) = y'(a) = y''(b) = y'''(b) = 0$ are satisfied. This was in turn extended by Hunt [9] and Reid [13] to more general even order self-adjoint differential equations of the form

$$[r_n(x)y^{(n)}]^{(n)} + \sum_{i=0}^{n-1} [(-1)^{n+i+1} r_i(x)y^{(i)}]^{(i)} = 0 .$$

Their definition of focal point involved the conditions

$$\begin{aligned} y(a) = y'(a) = \cdots = y^{(n-1)}(a) = 0 \\ = (r_n y^{(n)})(b) = (r_n y^{(n)})'(b) = \cdots = (r_n y^{(n)})^{(n-1)}(b) . \end{aligned}$$

It should be noted that the use of the terms "focal point" and "conjugate point" stems from considerations in the calculus of variations, in which field the majority of the investigations have taken place (see for example Reid [13] and the references in that paper).

In the present paper, we propose to extend the definition of focal point in a different direction, to which the techniques of differential equations seem more applicable. To this end we make the following definition.

DEFINITION 3. Suppose that there exists a point $c > a$, an integer $k(0 < k < n)$, and a nontrivial solution $\phi(x)$ of (1) which has a zero of order k at $x = a$ and such that $\phi^{(k)}(x)$ vanishes $n - k$ times, counting multiplicities, in $(a, c]$. Then the minimum such c, over all such solutions for all possible k, which exists by Theorem 4, will be denoted by $\tilde{\mu}_1(a)$.

DEFINITION 4. Suppose that there exists a point $c > a$, an integer $k(0 < k < n)$, and a nontrivial solution $\phi(x)$ of (1) such that $\phi(x)$ has a zero at $x = a$ of order k and such that $\phi^{(k)}(x)$ has a zero at $x = c$ of order $n - k$. By Theorem 4 there is a minimum such c, over all such solutions for all possible k, to be denoted by $\mu_1(a)$; a will then be called the first (left) focal point of $\mu_1(a)$.

THEOREM 8. *If $\tilde{\mu}_1(a)$ exists, then $\mu_1(a)$ exists and $\mu_1(a) = \tilde{\mu}_1(a)$.*

Proof. Let $U = \{u(x) \mid u(x)$ is a nontrivial solution of (1) and there exists a $k(0 < k < n)$ such that $u(x)$ has a zero of order k at $x = a$ and such that $u^{(k)}(x)$ has $n - k$ zeros (counting multiplicities) in $(a, \tilde{\mu}_1(a)]\}$,

$K = \{k \mid$ there is a $u(x) \in U$ with a zero of order k at $x = a\}$, $m_0 = \max k \in K$, $P = \{p \mid$ there is a $u(x) \in U$ with a zero of order m_0 at $x = a$, and such that $u^{(m_0)}(x)$ has p zeros in $(a, \tilde{\mu}_1(a))\}$, $p_0 = \max p \in P$, and let $\psi(x) \in U$ be a solution with a zero of order m_0 at $x = a$ and such that $\psi^{(m_0)}(x)$ has p_0 zeros in $(a, \tilde{\mu}_1(a))$. Let the zeros of $\psi^{(m_0)}(x)$ be of orders $m_1, m_2, \cdots, m_j$ at $a_1, a_2, \cdots, a_j$ respectively, where $a < a_1 < \cdots < a_j = \tilde{\mu}_1(a)$; by definition of p_0, $m_1 + m_2 + \cdots + m_{j-1} = p_0$.

Now suppose the conclusion does not hold; i.e., suppose $j > 1$. Let $z(x)$ be a nontrivial solution of (1) with a zero of order $m_0 + 1$ at $x = 0$ (note that $z(x) \notin U$) and with $z^{(m_0)}(x)$ having zeros of multiplicities $m_1, m_2, \cdots, m_{j-1}, m_{j-1} - 1, n - m_0 - 1 - p_0$ at $a_1, a_2, \cdots, a_{j-2}, a_{j-1}, a_j$ respectively, except that if $m_{j-1} = 1$ we make no specification at a_{j-1}. We have specified $n - 1$ zeros of $z(x)$ and $z^{(m_0)}$ on $[a, \tilde{\mu}_1(a)]$, $n - m_0 - 1$ zeros of $z^{(m_0)}(x)$ on $[a, \tilde{\mu}_1(a)]$, p_0 zeros of $z^{(m_0)}(x)$ on $[a, \tilde{\mu}_1(a))$, and $p_0 - 1$ zeros of $z^{(m_0)}(x)$ on $(a, \tilde{\mu}_1(a))$. Hence by Rolle's Theorem $z^{(m_0+1)}(x)$ has at least $n - m_0 - 2$ zeros on $[a, \tilde{\mu}_1(a)]$, and, in fact, since $z(x) \notin U$, $z^{(m_0)}(x)$ has exactly $n - m_0 - 2$ zeros on $[a, \tilde{\mu}_1(a)]$. Therefore, $z^{(m_0)}(x)$ does not vanish at any point on (a_{j-1}, a_j), nor are any of the specified zeros of multiplicities greater than that specified. We can now apply Theorem 1 to the functions $z^{(m_0)}(x)$ and $\psi^{(m_0)}(x)$ on the interval $[a_{j-1}, a_j]$, concluding that there is a linear combination $\chi(x)$ of $\psi(x)$ and $z(x)$ such that $\chi^{(m_0)}(x)$ has a double zero in (a_{j-1}, a_j). We now have a nontrivial solution $\chi(x)$ of (1) with a zero of order m_0 at $x = a$, such that $\chi^{(m_0)}(x)$ has at least $n - m_0$ zeros on $[a, \tilde{\mu}_1(a)]$ (hence $\chi(x) \in U$), and $\chi^{(m_0)}(x)$ has at least $p_0 + 1$ zeros on $(a, \tilde{\mu}_1(x))$ which contradicts the maximal property of p_0. Hence $j = 1$ and the conclusion holds.

THEOREM 9. *If $\eta_1(a)$ exists, then $\mu_1(a)$ exists and $\mu_1(a) < \eta_1(a)$.*

Proof. Let $\phi(x)$ be the essentially unique solution of Theorem 5, with its zero at a of order n_0 (where n_0 is maximal over the orders of the zero of all nontrivial solutions vanishing n times at a and $\eta_1(a)$); then its zero at $\eta_1(a)$ is of order at least $n - n_0$. By Rolle's Theorem, $\phi'(x)$ vanishes at least once on $(a, \eta_1(a))$ and at least $n - n_0$ times on $(a, \eta_1(a)]$. After repeated application of Rolle's Theorem we find that $\phi^{(n_0)}(x)$ vanishes at least n_0 times on $(a, \eta_1(a))$ and at least $n - n_0$ times on $(a, \eta_1(a)]$. Hence $\tilde{\mu}_1(a)$ exists, by Theorem 4; clearly $\tilde{\mu}_1(a) \leqq \eta_1(a)$, and by Theorem 8 $\mu_1(a) = \tilde{\mu}_1(a)$.

To show that $\mu_1(a) < \eta_1(a)$, suppose $\mu_1(a) \geqq \eta_1(a)$. Then, by the above, $\mu_1(a) = \tilde{\mu}_1(a) = \eta_1(a)$. Let $R = \{r \mid$ there exists a $u(x) \in U$ with a zero of order r at $x = a$ and such that $u^{(r)}(x)$ vanishes in $(a, \tilde{\mu}_1(a))\}$. By assumption, $\tilde{\mu}_1(a) = \eta_1(a)$; hence, by the first paragraph of this proof, R is nonvoid. Let $r_0 = \max r \in R$. Let $S = \{s \mid$ there exists a

$u(x) \in U$ with a zero of order r_0 at $x = a$ and such that $u^{(r_0)}(x)$ has s zeros (counting multiplicities) in $(a, \tilde{\mu}_1(a))\}$. Let $s_0 = \max s \in S$. We now apply the argument used in the proof of Theorem 8 to obtain a contradiction.

COROLLARY 1. *If $\phi(x)$ is a nontrivial solution of* (1) *with a zero of order k at $x = a$ and such that $\phi^{(k)}(x)$ has a zero of order $n - k$ at $x = \mu_1(a)$, where k is maximal over all such solutions, then $\phi^{(i)}(x) \neq 0$ in $(a, \mu_1(a)]$ $(i = 0, 1, \cdots, k - 1)$ and $\phi^{(k)}(x) \neq 0$ in $(a, \mu_1(a))$.*

Proof. Theorem 8 tells us that $\phi^{(k)}(x) \neq 0$ in $(a, \mu_1(a))$; and if $\phi^{(i)}(x) = 0$ at some point $x \in (a, \mu_1(a)]$, $(0 \leqq i \leqq k - 1)$, Rolle's Theorem allows us to conclude that there is a point $x_0 \in (a, \mu_1(a))$ where $\phi^{(k)}(x_0)$ is zero.

COROLLARY 2. *The conclusions of Theorems* 8 *and* 9 *are valid under the weaker conditions that there exists a nontrivial solution $\phi(x)$ of* (1) *satisfying the following: $\phi(x)$ has a zero of order q at $x = a$, and either*

(i) *$\phi(x)$ has $n - 1$ zeros (counting multiplicities) on $[a, \infty)$ and there exists a $p(1 \leqq p \leqq q)$ such that $\phi^{(p)}(x)$ vanishes at least once past the $(n - 1)^{st}$ zero; or*

(ii) *there exists a $p(0 \leqq p \leqq q)$ such that $\phi^{(p)}(x)$ has $n - q$ zeros (counting multiplicities) on (a, ∞).*

Proof. As in the proof of Theorem 9, Rolle's Theorem guarantees the existence of $\tilde{\mu}_1(a)$. The remainder of the proof is identical with that of Theorem 8.

4. An eigenfunction result. The basic relationships between conjugate points of equation (2) and eigenvalues of the associated boundary value problem were investigated in detail by Leighton and Nehari [11]. Subsequently Howard [8], using variational methods, discussed the relationships between focal points of (2) and eigenvalues of an eigenvalue problem, with focal point boundary conditions, associated with (2), making various assumptions on the coefficients. Barrett [3, 5] extended these results to fourth order equations of more general form. These results were extended by Reid [13] to more general even order self-adjoint differential equations.

After obtaining a general relationship between the solutions of (1) and those of

$$L_n^+ y = \sum_{i=0}^{n} (-1)^i (r_i y)^{(i)} = 0 \tag{3}$$

(the Lagrangian adjoint equation), we shall derive an eigenvalue relationship, from which we will develop a disconjugacy criterion. It should be noted that Theorem 10 is but a part of the more general theory of boundary operators ([10, Chapter 9], [6, Chapter 11]) included here to demonstrate the intimate relationship between boundary value problems associated with equation (1) and those associated with equation (3).

THEOREM 10. *If $u(x)$ is a solution of* (1) *with a zero of order exactly k at $x = a$ and a zero of order $n - k$ at $x = b$, then there exists a nontrivial solution $v(x)$ of* (3) *having a zero of order k at $x = b$ and a zero of order $n - k$ at $x = a$. Further, if k is maximal in the sense of Theorem* 5, *then $v(x)$ is essentially unique (i.e., unique except for multiplication by a constant factor).*

Proof. The operators L and L^+ in the equations (1) and (3) are connected by Green's identity:

$$(vL_n, u) - (u, L_n^+ v) = \int_a^b (vL_n u - uL_n^+ v)dx = \sum_{i=1}^{n} \left(\sum_{j=0}^{i-1} (-1)^j (vr_i)^{(j)}(x) u^{(i-j-1)}(x) \right) \Bigg|_a^b . \tag{4}$$

Let u be the solution of $L_n y = 0$ whose existence we have assumed; let v be a nontrivial solution of $L_n^+ y = 0$ with a zero of order $n - k - 1$ at $x = a$ and a zero of order k at $x = b$. Then, upon applying Theorem 2, (4) becomes

$$0 = (-1)^{n-k} (vr_n)^{(n-k-1)}(a) u^{(k)}(a) .$$

Since $u^{(k)}(a) \neq 0$ by hypothesis, this implies that $(vr_n)^{n-k-1}(a) = 0$; hence, by the corollary to Theorem 2, $v^{(n-k-1)}(a) = 0$, and $v(x)$ has a zero of order $n - k$ at $x = a$.

To prove the second part of the theorem suppose k is maximal and $v(x)$ is not unique; i.e., suppose that there are two nontrivial solutions $v_1(x)$ and $v_2(x)$ of (3) with zeros of order $n - k$ at $x = a$, and zeros of order k at $x = b$. Either $v_1(x)$, $v_2(x)$ or a linear combination of $v_1(x)$ and $v_2(x)$ has a zero of order $k + 1$ at $x = b$ and a zero of order $n - k$ at a. Let $v(x)$ be the nontrivial solution of (3) with a zero of order at least $n - k$ at $x = a$ and a maximal order zero, at $x = b$. Interchanging the roles of u and v in the previous argument, we find that there is a solution $u(x)$ with a zero of order $k + 1$ at $x = a$ and a zero of order $n - k - p$ at $x = b$, where $p > 0$; this contradicts the maximality of k.

We now derive an eigenfunction relationship between equations (1) and (3).

THEOREM 11. *Consider the adjoint n^{th} order linear differential operators L_n and L_n^+ and the adjoint problems*

$$Ly = \lambda y\,, \qquad Uy = 0\,, \tag{5}$$

$$L^+y = \lambda y\,, \qquad U^+y = 0\,, \tag{6}$$

where U and U^+ are self-adjoint or regular adjoint linear homogeneous boundary operators and λ is a parameter. Suppose further that Green's function for (5) *has only simple poles. Let the eigenfunctions of* (5) *and* (6) *be $\phi_i(x)$ and $\phi_i^+(x)$ respectively, normalized so that $(\phi_i, \phi_i^+) = 1$. Let $f(x)$ and $g(x)$ be any two functions such that L_ng exists, $Ug = 0$, L_n^+f exists, and $U^+f = 0$. Further let $a_i = (f, \phi_i)$ and $b_i = (g, \phi_i^+)$ and let λ_i be the ith eigenvalue of* (5). *Then*

$$(f, L_ng) = \sum_{i=0}^{\infty} a_ib_i\lambda_i\,.$$

Proof. We freely use the well-known properties of problems (5), (6) (see for instance Ince [5] Chapter 11, and Coddington and Levenson [6] Chapter 12).

$$\begin{aligned}(f, L_ng) &= \Big(\sum_{i=0}^{\infty} a_i\phi_i^+, L_ng\Big) = \sum_{i=0}^{\infty} a_i(\phi_i^+, L_ng)\\ &= \sum_{i=0}^{\infty} a_i(L_n^+\phi_i^+, g) = \sum_{i=0}^{\infty} a_i\Big(\lambda_i\phi_i^+, \sum_{j=0}^{\infty} b_j\phi_j\Big)\\ &= \sum_{i=0}^{\infty} a_i\lambda_i \sum_{j=0}^{\infty} b_j(\phi_i^+, \phi_j)\\ &= \sum_{i=0}^{\infty} a_i\lambda_i \sum_{j=0}^{\infty} b_j\delta_{ij} = \sum_{i=0}^{\infty} a_ib_i\lambda_i\,.\end{aligned}$$

If we specialize this result to the self-adjoint case, letting $f = g$ in (7), we obtain

$$(f, L_nf) = \sum_{i=0}^{\infty} a_i^2\lambda_i\,. \tag{8}$$

Further suppose $L_ny = 0$ assumes the special form

$$L_ny = [r_n(x)y^{(n)}]^{(n)} + \sum_{i=0}^{n-1} [(-1)^{n+i+1}r_i(x)y^{(i)}]^{(i)} = 0 \tag{9}$$

where $r_i(x) > 0$ $(i = 0, 1, \cdots, n)$.

We now state a disconjugacy criterion due to Reid [13].

THEOREM (Reid [13]). *Equation* (9) *is disconjugate in the sense of Reid (i.e., no solution possesses two nth order zeros) on the interval $[a, \infty)$ if and only if for all values b on (a, ∞) and all functions*

$y(x)$ such that $y(x) \in C_{n-1}[a, b]$, $y(x)$ is absolutely continuous on $[a, b]$, $(y^{(n)})^2$ is integrable on $[a, b]$ and $y(x)$ has n^{th} order zeros at $x = a$ and $x = b$, $(y, L_n y) > 0$ (where $L_n y$ is as in equation (9)).

Using this result we prove the following theorem. (Our impression is that Theorem 12 is in the literature, but the author has been unable to find it).

THEOREM 12. *Equation (9) is disconjugate in the sense of Reid if and only if the minimum eigenvalue λ_0 is positive (all functions are assumed to be of the class described in the above theorem of Reid).*

Proof. Suppose (9) is disconjugate. Letting $f(x) = \phi_0(x)$ in (8) and applying the theorem of Reid, we have $0 < (\phi_0, L_n\phi_0) = \lambda_0$, since $(\phi_i, \phi_j) = \delta_{ij}$. The sufficiency is immediate from equation (8).

BIBLIOGRAPHY

1. N. V. Azbelev, and Z. B. Chalyuk, *On the question of the distribution of the zeros of solution of a third order linear differential equation*, Mat. Sbornik, **51** (1960), 475-486 (Russian).
2. J. H. Barrett, *Disconjugacy of second-order linear differential equations with non-negative coefficients*, Proc. Amer. Math. Soc. **10** (1959), 552-561.
3. ———, *Disconjugacy of a self-adjoint differential equation of the fourth order*, Pacific J. Math. **11** (1961), 25-37.
4. ———, *Fourth order boundary value problems and comparison theorems*, Canadian J. Math. **13** (1961), 625-638.
5. ———, *Two-point boundary problems for linear self-adjoint differential equations of the fourth order with middle term*, Duke Math. J. **29** (1962), 543-554.
6. E. A. Coddington, and N. Levinson, *Theory of Ordinary Differential Equations*, McGraw-Hill, New York, 1955.
7. M. Hanan, *Oscillation criteria for third-order linear differential equations*, Pacific J. Math. **11** (1961), 919-944.
8. H. Howard, *Oscillation criteria for fourth-order linear differential equations*, Trans. Amer. Math. Soc. **96** (1960), 296-311.
9. R. W. Hunt, *The behavior of solutions of ordinary, self-adjoint differential equations of arbitrary even order*, Pacific J. Math. **12** (1962), 945-961.
10. E. L. Ince, *Ordinary differential equations*, Dover, New York, 1956.
11. W. Leighton, and Z. Nehari, *On the oscillation of solutions of self-adjoint differential equations of the fourth order*, Trans. Amer. Math. Soc. **89** (1958), 325-377.
12. A. Ju. Levin, *Some questions on the oscillations of solutions of linear differential equations*, Doklady Akad. Nauk. **148** (1963), 512-515 (Russian).
13. W. T. Reid, *Oscillation criteria for self-adjoint differential systems*, Trans. Amer. Math. Soc. **101** (1961), 91-106.
14. G. F. Simmons, *Introduction to topology and modern analysis*, McGraw-Hill, New York, 1963.

UNIVERSITY OF UTAH

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

INFLATION AND DEFLATION FOR ALL DIMENSIONS

ERNST SNAPPER

We assume that a finite group G acts on the left on finite sets X and Y, and that there is given a function $f: X \to Y$. We assume that $f(\sigma x) = \sigma f(x)$ for all $\sigma \in G$ and $x \in X$; and that $f^{-1}(y)$ has the same number h of elements for all $y \in Y$. We show that the cohomology groups $H^r(X; G, A)$ and $H^r(Y; G, A)$ of the permutation representations (G, X) and (G, Y) with values in a G-module A are interrelated by homomorphisms inflation$_r$: $H^r(Y; G, A) \to H^r(X; G, A)$ and deflation$_r$: $H^r(Y; G, A) \to H^r(Y; G, A)$, for all $r \in Z$. The main properties of inf$_r$ (inflation$_r$) and def$_r$ (deflation$_r$) are:

I. For all $r \in Z$, def$_r$inf$_r$: $H^r(Y; G, A) \to H^r(Y; G, A)$ consists of multiplying the elements of $H^r(Y; G, A)$ by h^q, where $q \geqq 1$ and q depends on r.

II. If for some $r \in Z$, $H^r(Y; G, A)$ is uniquely divisible by h, inf$_r$ is a monomorphism and def$_r$ is an epimorphism and $H^r(X; G, A) = \mathrm{im}(\mathrm{inf}_r) \oplus \ker(\mathrm{def}_r)$, where $\oplus$ denotes the direct sum of abelian groups.

III. $H^r(Y; G, A)$ is uniquely divisible by h for all $r \in Z$ in each of the following two cases.

IIIa. A is uniquely divisible by h.

IIIb. $(h, m) = 1$ where m is the index of (G, Y).

We then study the special case where the permutation representations (G, X) and (G, Y) are transitive and where (G, X) is furthermore free of fixed points. Since the classical inflation and deflation mappings fall under this heading, we have now extended these mappings to all of Z. We describe the six mappings inf$_r$ and def$_r$ for $r = 0, \pm 1$ explicity in terms of trace mappings, augmentation ideals and crossed homomorphisms.

G stands for a finite group. For every normal subgroup H of G and G-module A, the inflation (or lift) mapping $H^r(G/H, A^H) \to H^r(G, A)$ is well known for $r \geqq 1$; A^H always denotes the submodule of A whose elements are left fixed by H. Dually, there is available the deflation mapping $H^r(G, A) \to H^r(G/H, A^H)$ for $r \leqq -2$ (see [7]). In the present paper we extend the inflation and deflation mappings to all $r \in Z$. (Z denotes the ring of the rational integers.) We develop the theory for arbitrary permutation representations (see [6] for the cohomology of permutation representations) which includes the case

Received March 5, 1964. This research was supported in part by the National Science Foundation grant NSF-GP722.

that H is not normal.

The fact that the inflation mapping followed by the deflation mapping consists of multiplying by a power of $[H: 1]$ (see Theorem 5.1), indicates that these mappings behave particularly nicely if A is uniquely divisible by $[H: 1]$, or if H is a Hall subgroup of G. These cases are worked out in § 6, 7, 8, 11, 12 and 13 and are needed for the author's forthcoming paper on duality in the cohomology of permutation representations. The study of deflation in dimension 1 brings to the fore natural endomorphisms of the group of crossed homomorphisms from G to A. There is one such endomorphism for each subgroug of G. (see § 15 and 16.)

1. **Inflation for chains.** X stands for a finite set and (G, X) for a permutation representation (see the introduction of [6]); i.e., $\sigma x \in X$ for all $x \in X$ and $\sigma \in G$, and $(\sigma\tau)x = \sigma(\tau x)$ and $1x = x$ for all $\sigma, \tau \in G$; 1 always denotes the unit element of the group under discussion. Let (L, Y) be a second permutation representation of some finite group L acting on some finite set Y, and let $\theta = (\varphi, f)$: $(G, X) \to (L, Y)$ be a morphism of permutation representations (see the introduction of [6]); i.e., $\varphi: G \to L$ is a group homomorphism and $f: X \to Y$ is a function where $f(\sigma x) = \varphi(\sigma) f(x)$ for all $\sigma \in G$ and $x \in X$. The rth chain group $C_r(X; G)$ of the standard complex $C.(X; G)$ of (G, X) is the G-module $Z[X^q]$, where X^q is the cartesian product of X with itself q times; $q = r + 1$ if $r \geqq 0$ and $q = -r$ if $r < 0$ (see § 1 of [6]; the same definitions hold of course for (L, Y).) The function $(x_1, \cdots x_q) \to (f(x_1), \cdots, f(x_q))$ from X^q to Y^q can be extended by linearity to a homomorphism α_r: $Z[X^q] \to Z[Y^q]$ which is a G-homomorphism if we regard the L-module $Z[Y^q]$ as a G-module under φ: $G \to L$. All this gives rise to the diagram:

$$\begin{array}{ccccccccccccc}
\cdots & \xrightarrow{\partial_2} & C_1(X; G) & \xrightarrow{\partial_1} & C_0(X; G) & \xrightarrow{\partial_0} & C_{-1}(X; G) & \xrightarrow{\partial_{-1}} & C_{-2}(X; G) & \xrightarrow{\partial_{-2}} & \cdots \\
 & & \downarrow{\scriptstyle \alpha_1} & & \downarrow{\scriptstyle \alpha_0} & \searrow^{\varepsilon} \quad \nearrow_{\mu} & \downarrow{\scriptstyle \alpha_{-1}} & & \downarrow{\scriptstyle \alpha_{-2}} & & \\
 & & & & & Z & & & & & \\
 & & & & & \nearrow_{\varepsilon'} \quad \searrow^{\mu'} & & & & & \\
\cdots & \xrightarrow[\partial'_2]{} & C_1(Y; L) & \xrightarrow[\partial'_1]{} & C_0(Y; L) & \xrightarrow[\partial'_0]{} & C_{-1}(Y; L) & \xrightarrow[\partial'_{-1}]{} & C_{-2}(Y; L) & \xrightarrow[\partial'_{-2}]{} & \cdots
\end{array}$$

We have primed the differentiation mappings ∂'_r and augmentation mappings ε', μ' of the complex $C.(Y; L)$. We know from § 1 of [6] that $\mu\varepsilon = \partial_0$ and that $\mu'\varepsilon' = \partial'_0$; and § 13 of [6] tells us that $\partial'_r\alpha_r = \alpha_{r-1}\partial_r$ for $r \geqq 1$ and that $\varepsilon'\alpha_0 = \varepsilon$. The reason why one shies away from studying α_r for $r < 0$ is that these commutativity relations fail

for $r < 0$. We show however that they fail by so little that these maps α_r are still very useful for $r < 0$.

We assume for the remainder of this paper that $f^{-1}(y)$ contains the same number of elements for all $y \in Y$, and denote this number by h. This implies of course that $f: X \to Y$ is an epimorphism and hence that α_r, for all $r \in Z$, is an epi. Conversely, if f is an epi *and the permutation representation (G, X) is transitive,* the number of elements in $f^{-1}(y)$ does not depend on y. This follows easily from the fact that for every morphism of permutation representations the partitioning $X = \bigcup f^{-1}(y)$ of X consists of domains of imprimitivity of (G, X). (See § 146 of [2] for domains of imprimitivity.)

We replace the differentiation operator ∂'_r of $C.(Y; L)$ by $h\partial'_r$ if $r < 0$, but leave ∂'_r unchanged for $r < 0$. We also change μ' to $h\mu'$ but leave ε' unchanged. We now show that the following diagram displays a chain mapping of complexes.

(I)

$$\begin{array}{ccccccccccccc}
\cdots & \xrightarrow{\partial_2} & C_1(X;G) & \xrightarrow{\partial_1} & C_0(X;G) & \xrightarrow{\partial_0} & C_{-1}(X;G) & \xrightarrow{\partial_{-1}} & C_{-2}(X;G) & \xrightarrow{\partial_{-2}} & \cdots \\
 & & \downarrow{\scriptstyle \alpha_1} & & \downarrow{\scriptstyle \alpha_0} & \searrow^{\varepsilon} \quad \swarrow_{\mu} & \downarrow{\scriptstyle \alpha_{-1}} & & \downarrow{\scriptstyle \alpha_{-2}} & & \\
 & & & & & Z & & & & & \\
 & & & & & \nearrow_{\varepsilon'} \quad \searrow_{h\mu'} & & & & & \\
\cdots & \xrightarrow[\partial'_2]{} & C_1(Y;L) & \xrightarrow[\partial'_1]{} & C_0(Y;L) & \xrightarrow[h\partial'_0]{} & C_{-1}(Y;L) & \xrightarrow[h\partial'_{-1}]{} & C_{-2}(Y;L) & \xrightarrow[h\partial'_{-2}]{} & \cdots
\end{array}$$

PROPOSITION 1.1. The upper row of diagram (I) is a G-complex and the lower row is an L-complex. The diagram is completely commutative, that is;

(1) $\partial'_r\alpha_r = \alpha_{r-1}\partial_r$ for $r \geqq 1$;
(2) $\varepsilon'\alpha_0 = \varepsilon$;
(3) $\mu\varepsilon = \partial_0$;
(4) $h\mu'\varepsilon' = h\partial'_0$;
(5) $\alpha_{-1}\mu = h\mu'$;
(6) $h\partial'_r\alpha_r = \alpha_{r-1}\partial_r$ for $r \leqq -1$.

The chain mapping $\{\alpha_r, r \in Z\}$ is an epimorphism and a G-mapping if we consider the lower row as a G-complex under $\varphi: G \to L$.

Proof. The upper row is the G-complex $C.(X; G)$. The fact that $C.(Y; L)$ is an L-complex implies immediately that the lower row is also an L-complex. The first three commutativity relations have been discussed above and (4) follows from $\mu'\varepsilon' = \partial'_0$. For (5) we observe that $\alpha_{-1}\mu(1) = \alpha_{-1}\Sigma_{x\in X}x = \Sigma_{x\in X}f(x) = h\Sigma_{y\in Y}y = h\mu'(1)$. For (6) we select $(x_1, \cdots, x_r) \in X^r$ and use the definition of ∂_{-r} of § 1 of [6] to

compute that $\alpha_{-r-1}\partial_{-r}(x_1, \cdots, x_r) = \alpha_{-r-1}(\Sigma_{x\in X}(x, x_1, \cdots, x_r) + \Sigma_{i=1}^{r}(-1)^i \Sigma_{x\in X}(x_1, \cdots, x_i, x, x_{i+1}, \cdots x_r)) = \sum_{x\in X}(f(x), f(x_1), \cdots, f(x_r)) + \Sigma_{i-1}^{r}(-1)^i \Sigma_{x\in X}(f(x_1), \cdots, f(x_i), f(x)f(x_{i+1}), \cdots, f(x_r)) = h\Sigma_{y\in Y}(y, f(x_1), \cdots, f(x_r)) + h\Sigma_{i=1}^{r}(-1)^i \Sigma_{y\in Y}(f(x_1), \cdots, f(x_i), y, f(x_{i+1}), \cdots, f(x_r)) = h\partial'_{-r}(f(x_1), \cdots, f(x_r)) = h\partial'_{-r}\alpha_{-r}(x_1, \cdots, x_r)$. Finally, the fact that α_r: $C_r(X; G) \to C_r(Y; L)$ is an epimorphism and may be regarded as a G-homomorphism has been mentioned previously. Done.

One should be careful to observe that the lower row of diagram (I) may not be acyclic any longer. True, its rth cycle group is the same as the rth cycle group of the acyclic complex $C.(Y; L)$, because $C_r(Y; L) = Z[Y^q]$ is without torsion for all $r \in Z$. However, if $r \leqq -1$, the rth boundary group of the lower row of diagram (I) is hB_r where B_r denotes the rth boundary group of $C.(Y; L)$.

It is convenient to think of the mappings α_r as the "inflation mappings for chains" because, if $r \geqq 1$, α_r gives rise to the customary inflation mapping (see Definition 4.1). If however $r \leqq 0$, either α_r or $h\alpha_r$ is used to define the inflation mapping (same definition).

2. **Deflation for chains.** We define, for every $r \in Z$, a homomorphism β_r: $C_r(Y; L) \to C_r(X; G)$. Again, $C_r(Y; L) = Z[Y^q]$, where $q = r + 1$ if $r \geqq 0$ and $q = -r$ if $r < 0$. The mapping $(y_1, \cdots, y_q) \to \Sigma(x_{i_1}, \cdots, x_{i_q})$, where the summation is over all q-tuples of the cartesian product $f^{-1}(y_1) \times \cdots \times f^{-1}(y_q)$, maps the Z-base of $Z[Y^q]$ into $Z[X^q] = C_r(X; G)$. We define β_r as the extension by linearity of this mapping to $Z[Y^q]$. We observe that β_r is the dual of the mapping α_{-r-1} in the following sense. $C_r(Y; L)$ may be regarded as $\mathrm{Hom}_z(C_{-r-1}(Y; L), Z)$ and similarly, for $C_r(X; G)$. (See § 1 of [6].) If we apply the functor $\mathrm{Hom}_Z(*, Z)$ to the homomorphism α_{-r-1}: $C_{-r-1}(X; G) \to C_{-r-1}(Y; L)$ we obtain the homomorphism β_r: $C_r(Y; L) \to C_r(X; G)$. This observation makes the following proposition into an easy corollary of Proposition 1.1.

PROPOSITION 2.1. The upper row of diagram (II) (see below) is a G-complex and the lower row is an L-complex. The diagram is completely commutative, that is:

(1) $\partial_r\beta_r = \beta_{r-1}h\partial'_r$ for $r \geqq 1$;

(2) $\varepsilon\beta_0 = h\varepsilon'$;

(3) $\mu\varepsilon = \partial_0$;

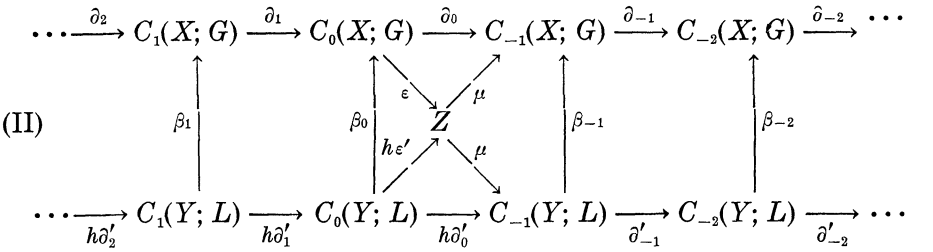
(4) $\mu' h\varepsilon' = h\partial'_0$;

(5) $\beta_{-1}\mu' = \mu$;

(6) $\partial_r\beta_r = \beta_{r-1}\partial'_r$ for $r \leqq -1$.

The chain mapping $\{\beta_r; r \in Z\}$ is a monomorphism and is a G-mapping if we consider the lower row as a G-complex under φ: $G \to L$.

(II)



Observe that the lower rows of diagrams (I) and (II) are not the same but correspond to one another under the functor $\mathrm{Hom}_Z(*, Z)$. It is convenient to think of the mappings β_r as the "deflation mappings for chains" because, if $r \leqq -2$, β_r gives rise to the deflation mapping defined in [7]. If however $r \geqq -1$, either β_r or $h\beta_r$ is used to define the deflation mapping (see Definition 5.1).

PROPOSITION 2.2. $\alpha_r\beta_r - h^{r+1}$ if $r \geqq 0$ and $\alpha_r\beta_r = h^{-r}$ if $r \leqq -1$. Here, h^q denotes the endomorphism of $C_r(Y; L)$ which consists of multiplying its elements by h^q.

Proof. The Z-base of $C_r(Y; L)$ consists of the q-tuples $(y_1, \cdots, y_q) \in Y^q$. Furthermore, $\alpha_r\beta_r(y_1, \cdots, y_q) = \alpha_r\Sigma(x_{i_1}, \cdots, x_{i_q}) = \Sigma(f(x_{i_1}), \cdots, f(x_{i_q}))$ where the summation is over the h^q q-tuples $(x_{i_1}, \cdots, x_{i_q})$ of the cartesian product $f^{-1}(y_1) \times \cdots \times f^{-1}(y_q)$. Hence the last sum is equal to $h^q(y_1, \cdots, y_q)$. Done.

3. Inflation and deflation for cochains. We now have to "hom" diagrams (I) and (II) with modules. Although it is possible to work simultaneously with a G-module and an L-module, we restrict ourselves to the case which is of principal interest for group theory. *We assume for the remainder of this paper that $G = L$ and that φ is the identity mapping of G. Furthermore, A stands for a G-module.*

If we apply the functor $\mathrm{Hom}_G(*, A)$ to the chain complex $C.(X; G)$, we obtain the cochain complex $C^\cdot(X; G, A)$ (see § 2 of [6]). We denote the rth cochain group of $C^\cdot(X; G, A)$ by $C^r(X; G, A)$ and treat the permutation representation (G, Y) in the same way. Hence, under the functor $\mathrm{Hom}_G(*, A)$, the mappings $\alpha_r\colon C_r(X; G) \to C_r(Y; G)$ and $\beta_r\colon C_r(Y; G) \to C_r(X; G)$ become, respectively, mappings $a_r\colon C^r(Y; G, A) \to C^r(X; G, A)$ and $b_r\colon C^r(X; G, A) \to C^r(Y; G, A)$; here, $a_r = \mathrm{Hom}_G(\alpha_r, 1_A)$ and $b_r = \mathrm{Hom}_G(\beta_r, 1_A)$ where 1_A denotes the identity of A. When we apply the same functor to diagrams (I) and (II) we obtain, respectively, diagrams (III) and (IV); and Propositions 1.1, 2.1 and 2.2 give Proposition 3.1.

$$\cdots \xrightarrow{\delta_{-3}} C^{-2}(X;G,A) \xrightarrow{\delta_{-2}} C^{-1}(X;G,A) \xrightarrow{\delta^{-1}} C^0(X;G,A) \xrightarrow{\delta_0} C^1(X;G,A) \xrightarrow{\delta_1} \cdots$$

(III)
$$\begin{array}{ccccccccc} & & a_{-2}\uparrow & & a_{-1}\uparrow & & a_0\uparrow & & a_1\uparrow \\ \cdots & \xrightarrow[h\delta'_{-3}]{} & C^{-2}(Y;G,A) & \xrightarrow[h\delta'_{-2}]{} & C^{-1}(Y;G,A) & \xrightarrow[h\delta'_{-1}]{} & C^0(Y;G,A) & \xrightarrow[\delta'_0]{} & C^1(Y;G,A) \xrightarrow[\delta'_1]{} \cdots \end{array}$$

$$\cdots \xrightarrow{\delta_{-3}} C^{-2}(X;G,A) \xrightarrow{\delta_{-2}} C^{-1}(X,G,A) \xrightarrow{\delta_{-1}} C^0(X;G,A) \xrightarrow{\delta_0} C^1(X;G,A) \xrightarrow{\delta_1} \cdots$$

(IV)
$$\begin{array}{ccccccccc} & & b_{-2}\downarrow & & b_{-1}\downarrow & & b_0\downarrow & & b_1\downarrow \\ \cdots & \xrightarrow[\delta'_{-3}]{} & C^{-2}(Y;G,A) & \xrightarrow[\delta'_{-2}]{} & C^{-1}(Y;G,A) & \xrightarrow[h\delta'_{-1}]{} & C^0(Y;G,A) & \xrightarrow[h\delta'_0]{} & C^1(Y,G,A) \xrightarrow[h\delta'_1]{} \cdots \end{array}$$

PROPOSITION 3.1. All four rows in diagrams (III) and (IV) are complexes of abelian groups, and both diagrams are commutative diagrams. The chain mapping $\{a_r; r \in Z\}$ is a monomorphism, but the chain mapping $\{b_r; r \in Z\}$ is not necessarily an epimorphism. Furthermore, $b_r a_r = h^{r+1}$ if $r \geqq 0$ and $b_r a_r = h^{-r}$ if $r \leqq -1$; here, h^q denotes the endomorphism of $C^r(Y; G, A)$ which consists of multiplying its elements by h^q.

It is clear from the previous sections that it is convenient to think of the mappings a_r and b_r as, respectively, the "inflation mapping" and "deflation mapping" for cochains.

4. **Inflation for cohomology groups.** We denote, as in [6], the rth cocycle group (coboundary group, cohomology group) of the complex $C^\cdot(X; G, A)$ by $Z^r(X; G, A)$, $(B^r(X; G, A)$, $H^r(X, G, A))$; we do of course the same for $C^\cdot(Y; G, A)$. We read immediately from diagram (III) that $a_r(Z^r(Y; G, A)) \subset Z^r(X; G, A)$ for all $r \in Z$; and that $a_r(B^r(Y; G, A)) \subset B^r(X; G, A)$ if $r \geqq 1$. If $r \leqq 0$, a_r may not transform coboundaries into coboundaries (see Example 9.1); this depends on the nature of our morphism $(G, X) \to (G, Y)$ and the G-module A. However, diagram (III) does tell us immediately that $ha_r(B^r(Y; G, A)) \subset B^r(X; G, A)$ and that $ha_r(Z^r(Y; G, A)) \subset Z^r(X; G, A)$ for all $r \in Z$.

The above implies the following for the cohomology groups. The homomorphism ha_r always induces a homomorphism $(ha_r)^*$: $H^r(Y; G, A) \to H^r(X; G, A)$ for all $r \in Z$. The homomorphism a_r induces a homomorphism a_r^*: $H^r(Y; G, A) \to H^r(X; G, A)$ for $r \geqq 1$ but, depending on the morphism $(G, X) \to (G, Y)$ and on A, not for $r \leqq 0$. Whenever a_r^* exists, that's the mapping we want. If however a_r^* does not exist we should not despair but be satisfied with $(ha_r)^*$. The following definition reflects this attitude.

DEFINITION 4.1. Let $r \in Z$. If it happens that $a_r(B^r(Y; G, A)) \subset$

$B^r(X; G, A)$, we call the homomorphism $a_r^*: H^r(Y; G, A) \to H^r(X; G, A)$ the inflation mapping or lift mapping for dimension r. If $a_r(B^r(Y; G, A)) \not\subset B^r(X; G, A)$, we call the homomorphism $(ha_r)^*: H^r(Y; G, A) \to H^r(X; G, A)$ the inflation mapping or lift mapping. We denote the inflation mapping by inf or $\inf_r$.

The above definition gives the customary inflation mapping when $r \geqq 1$. We repeat that, when $r \leqq 0$, it depends on the morphism $(G, X) \to (G, Y)$ and the module A whether $\inf_r = a_r^*$ or $\inf_r = (ha_r)^*$.

REMARK 4.1. One could obviously have proceeded differently. Namely, diagram (III) shows that a_r always induces a homomorphism from the rth cohomology group H^r of the lower row of that diagram into $H^r(X; G, A)$. The groups H^r for $r \leqq 0$ seem to be of no particular interest for group theory which is why we proceeded as in Definition 4.1.

EXAMPLE 4.1. Consider the morphism of permutation representations $(1_G, f)$: $(G, G) \to (G, G/H)$. Here, $X = G$ and the permutation representation (G, G) consists of G acting by left multiplication on itself. Furthermore H is a subgroup of G, *not necessarily normal*, and Y is the set G/H of the left cosets of H. The permutation representation $(G, G/H)$ consists of G acting on these cosets by left multiplication. Finally, $f(\sigma) = \sigma H$ for $\sigma \in G$. The number of elements in $f^{-1}(\sigma H)$ is the order h of H and hence is independent of σH. Consequently, Definition 4.1 applies and $\inf_r: H^r(G/H; G, A) \to H^r(G; G, A)$ is defined for all $r \in Z$. As is well known, $H^r(G; G, A)$ is the classical cohomology group $H^r(G, A)$, and $H^r(G/H; G, A)$ is the relative group $H^r(G: H, A)$ defined in [1]. If $r \geqq 1$, $\inf_r$ coincides with the inflation mapping defined in § 7 of [1]. If H is a normal subgroup of G, $H^r(G: H, A)$ is isomorphic with the classical cohomology group $H^r(G/H, A^H)$ (see the Corollary on page 68 of [1]) and we obtain, if $r \geqq 1$, the customary inflation mapping from $H^r(G/H, A^H)$ into $H^r(G, A)$. We shall frequently come back to this example.

5. **Deflation for cohomology groups.** We read from diagram (IV) that $b_r(B^r(X; G, A)) \subset B^r(Y; G, A)$ for all $r \in Z$; and that $b_r(Z^r(X; G, A)) \subset Z^r(Y; G, A)$ if $r \leqq -2$. If $r \geqq -1$, b_r may not transform cocycles into cocycles. Diagram (IV) also tells us that $hb_r(Z^r(X; G, A)) \subset Z^r(Y; G, A)$ and that $hb_r(B^r(X; G, A)) \subset B^r(Y; G, A)$ for all $r \in Z$.

Consequently, hb_r induces a homomorphism $(hb_r)^*: H^r(X; G, A) \to H^r(Y: G, A)$ for all $r \in Z$. The homomorphism b_r induces a homomorphism $b_r^*; H^r(X; G, A) \to H^r(Y; G, A)$ for $r \leqq -2$ but, depending on the morphism $(G, X) \to (L, Y)$ and the module A, not for $r \leqq -1$.

We proceed as in the case of inflation.

DEFINITION 5.1. Let $r \in Z$. If it happens that $b_r(Z^r(X; G, A)) \subset Z^r(Y; G, A)$, we call the homomorphism $b_r^*: H^r(X; G, A) \to H^r(Y; G, A)$ the deflation mapping for dimension r. If $b_r(Z^r(X; G, A)) \not\subset Z^r(Y; G, A)$, we call the homomorphism $(hb_r)^*: H^r(X; G, A) \to H^r(Y; G, A)$ the deflation mapping. We denote the deflation mapping by def or def_r.

We repeat that, when $r \leqq -2$, $\text{def}_r = (b_r)^*$. If $r \geqq -1$, it depends on the morphism $(G, X) \to (G, Y)$ and the G-module A whether $\text{def}_r = b_r^*$ or $\text{def}_r = (hb_r)^*$. Remark 4.1 applies of course equally well to deflation.

EXAMPLE 5.1. Consider the morphism $(1_G, f)$: $(G, G) \to (G, G/H)$ of Example 4.1. Definition 5.1 defines the deflation mapping def_r: $H^r(G, A) \to H^r(G: H, A)$ for all $r \in Z$. If H is a normal subgroup of G, def_r maps $H^r(G, A)$ into $H^r(G/H, A^H)$; if furthermore $r \leqq -2$, def_r coincides with the deflation mapping studied in [7].

THEOREM 5.1. *Let h^q denote the endomorphism of $H^r(Y; G, A)$ which consists of multiplying its elements by h^q. For each $r \in Z$ there exists an integer $q \geqq 1$, depending on r, such that $\text{def}_r \text{ inf}_r = h^q$.*

Proof. $\text{def}_r \text{ inf}_r$ is equal to $b_r^* a_r^*$ or to $(hb_r)^* a_r^*$ or $b_r^*(ha_r)^*$ or $(hb_r)^*(ha_r)^*$. Proposition 3.1 tells us that $b_r a_r$, $(hb_r)a_r$, $b_r(ha_r)$ and $(hb_r)(ha_r)$ all consist of multiplying the elements of $C^r(Y; G, A)$ by a positive power of h. Done.

We now study various special instances of inflation and deflation. Hereto, we need some material on uniquely divisible modules.

6. Uniquely divisible modules.

In this whole section, $k \in Z$ stands for a fixed, nonzero integer. If F is a module (i.e., an abelian group written additively) we denote the identity mapping of F onto itself by 1_F. Hence, $k1_F$ denotes the endomorphism of F which consists of multiplying its elements by k. As always, F is called divisible by k if $k1_F$ is an epimorphism; and F is called uniquely divisible by k if $k1_F$ is an automorphism.

PROPOSITION 6.1. Let $0 \longrightarrow D \xrightarrow{i} E \xrightarrow{j} F \longrightarrow 0$ be an exact sequence of modules. If two of them are uniquely divisible by k, so is the third.

Proof. Consider the commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & D & \xrightarrow{i} & E & \xrightarrow{j} & F & \longrightarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & D & \xrightarrow{i} & E & \xrightarrow{j} & F & \longrightarrow & 0 \end{array}$$

where the vertical arrows denote, respectively, $k1_D$, $k1_E$ and $k1_F$. We conclude from the "5 lemma" (see [3], page 5) that, if two of the vertical arrows are automorphisms, so is the third. Done.

PROPOSITION 6.2. Let $v: E \to F$ be a homomorphism from the module E to the module F. If E and F are both uniquely divisible by k, so are $\ker(v)$, $\operatorname{coker}(v)$, $\operatorname{im}(v)$ and $\operatorname{coim}(v)$. (Coim stands for coimage.)

Proof. Since E is divisible by k, $\operatorname{im}(v)$ is evidently divisible by k. The fact that, actually, $\operatorname{im}(v)$ is uniquely divisible by k then follows from the fact that $k1_F$ is a monomorphism. This also takes care of $\operatorname{coim}(v) \simeq \operatorname{im}(v)$. We now apply Proposition 6.1 to the exact sequences $0 \to \operatorname{im}(v) \to F \to \operatorname{coker}(v) \to 0$ and $0 \to \ker(v) \to E \to \operatorname{coim}(v) \to 0$ and we are done.

REMARKS 6.1. Propositions 6.1 and 6.2 together say that the category of modules which are uniquely divisible by k is a complete subcategory of the category of abelian groups (see page 138 of [5]). This subcategory is not "épaisse" (same reference) since the additive group of Z is a subgroup of the additive group of the rational numbers; the latter group is uniquely divisible by k but, if $k \neq \pm 1$, the first one is not.

PROPOSITION 6.3. Let E and F be two Λ-modules where Λ is some ring with unit element. If one of the modules is uniquely divisible by k, so is $\operatorname{Hom}_\Lambda(E, F)$.

Proof. Suppose that $k1_E$ is an automorphism. Then, $\operatorname{Hom}_\Lambda(k1_E, 1_F)$: $\operatorname{Hom}_\Lambda(E, F) \to \operatorname{Hom}_\Lambda(E, F)$ is an automorphism, and it consists of course of multiplying the elements of $\operatorname{Hom}_\Lambda(E, F)$ by k. We proceed similarly if $k1_F$ is an automorphism. Done.

We now return to our permutation representation (G, X). Since (G, X) is entirely arbitrary, Lemma 6.1 is valid for all permutation representations.

LEMMA 6.1. *Let A be a G-module which is uniquely divisible by k. Then, $H^r(X; G, A)$ is uniquely divisible by k for all $r \in Z$.*

Proof. Let $r \in Z$. $C^r(X; G, A) = \mathrm{Hom}_G(C_r(X; G), A)$ is uniquely divisible by k by Proposition 6.3. We apply Proposition 6.2 to the homomorphisms $C^{r-1}(X, G, A) \to C^r(X; G, A) \to C^{r+1}(X; G, A)$ and find that the cocycle group $Z^r(X; G, A)$ and the coboundary group $B^r(X, G, A)$ are uniquely divisible by k. Since $H^r(X; G, A)$ is the cokernel of the inclusion mapping $B^r(X; G, A) \to Z^r(X; G, A)$, the same proposition gives the desired result.

REMARK 6.2. Lemma 6.1 gives a cute proof of the well known fact that $H^r(G, A) = 0$ if A is uniquely divisible by the order n of G. Namely, $nH^r = 0$ and, by Lemma 6.1, $nH^r(G, A) = H^r(G, A)$. More generally, if A is uniquely divisible by the index of the permutation representation (G, X), then $H^r(X; G, A) = 0$ for all $r \in Z$. (See Corollary 10.2 of [6].)

7. **The case that $H^r(Y; G, A)$ is uniquely divisible by h.** We recall that the set Y is partitioned into the domains of transitivity $T_1, \cdots, T_u$ of the permutation representation (G, Y). If T_i has m_i elements, the greatest common divisor m of $m_1, \cdots, m_u$ is called *the index of* (G, Y) (see §4 of [6]).

Lemma 7.1. *Let $d = (h, m)$. If A is uniquely divisible by d, then $H^r(Y; G, A)$ is uniquely divisible by h for all $r \in Z$.*

Proof. $H^r(Y; G, A)$ is uniquely divisible by d by Lemma 6.1, and $mH^r(Y; G, A) = 0$ by Corollary 10.2 of [6]. Done.

The following proposition is an immediate corollary of Lemma 7.1.

PROPOSITION 7.1. In each of the following two cases $H^r(Y, G, A)$ is uniquely divisible by h for all $r \in Z$.

(a) A is uniquely divisible by h.

(b) $(h, m) = 1$.

EXAMPLE 7.1. Case (*b*) of Proposition 7.1 is important for Hall subgroups. (*A* subgroup H of G is called a Hall subgroup if the order of H is relatively prime to the index $[G: H]$ of H.) In the morphism $(G, G) \to (G, G/H)$ of Example 4.1, the index of $(G, G/H)$ is the index $[G: H]$; hence, $(h, m) = 1$ if and only if H is a Hall subgroup of G.

THEOREM 7.1. *Let $r \in Z$ and let $H^r(Y; G, A)$ be uniquely divisible by h. Then, inf_r is a monomorphism and def_r is an epimorphism; and $H^r(X; G, A) = im(inf_r) \oplus ker(def_r)$ where $\oplus$ denotes the direct sum of abelian groups.*

Proof. $\mathrm{def}_r\mathrm{inf}_r = h^q$ for some $q \geqq 1$ by Theorem 5.1. Since h^q is an automorphism of $H^r(Y; G, A)$, Theorem 7.1 follows from routine group arguments.

EXAMPLE 7.2. Consider the morphism $(G, G) \to (G, G/H)$ of Example 4.1 and suppose that, for some $r \in Z$, $H^r(G/H; G, A)$ is uniquely divisible by h. Since $mH^r(G/H; G, A) = 0$ where $m = [G: H]$, it is obvious that $m(\mathrm{im}(\mathrm{inf}_r)) = 0$. It may however very well be that $H^r(X; G, A)$, which is equal to $H^r(G, A)$, contains further elements which are annihilated by m. For instance, *if A is uniquely divisible by h, all elements of $H^r(X; G, A)$ are annihilated by m.* This follows from (1) $H^r(X; G, A)$ is divisible by h (it is even uniquely divisible by h by Lemma 6.1); (2) $mhH^r(X; G, A) = 0$ since $H^r(X; G, A) = H^r(G, A)$ and mh is the order of G.

In this connection, it is interesting to recall that Faddeev proved in [4] that, if H is a Hall subgroup of G, and $r \geqq 1$, $\mathrm{im}(\mathrm{inf}_r)$ consists of all the elements of $H^r(G, \mathrm{A})$ which are annihilated by m. We conclude: *Let $r \geqq 1$, let A be uniquely divisible by h and let H be a Hall subgroup of G. Then, inf_r and def_r are both isomorphisms. In particular, $H^r(G, A) \simeq H^r(G: H, A)$.* (This last isomorphism and the fact that inf is an isomorphism also follow from Faddeev's results on the restriction mapping. All one has to observe is that $H^r(H, A) = 0$, since A is uniquely divisible by h.) The author conjectures that this result remains true for $r \leqq 0$.

8. **The case that A is uniquely divisible by h.** We know from Lemma 6.1 that, if A is uniquely divisible by h, Theorem 7.1 may be applied for all $r \in Z$. We now add to this that in this case $\mathrm{inf}_r = a_r^*$ and $\mathrm{def}_r = b_r^*$ for all $r \in Z$. In other words, the factor h in Definitions 4.1 and 5.1 can be omitted. For deflation this is even correct if $h1_A$ is only a monomorphism.

THEOREM 8.1. *If A is uniquely divisible by h, $\mathrm{inf}_r = a_r^*$ for all $r \in Z$. If $h1_A$ is a monomorphism, $\mathrm{def}_r = b_r^*$ for all $r \in Z$.*

Proof. Let A be uniquely divisible by h and select $r \in Z$. We see from diagram (III) that the rth coboundary group of the lower row of that diagram is $B^r(Y; G, A)$ if $r \geqq 1$ and is $hB^r(Y; G, A)$ if $r \leqq 0$. We see from the proof of Lemma 6.1 that $B^r(Y; G, A)$ is uniquely divisible by h and hence $hB^r(Y; G, A) = B^r(Y; G, A)$. Since $\{a_i, i \in Z\}$ is a chain mapping it is now clear that $a_r(B^r(Y; G, A)) \subset B^r(X; G, A)$; hence, by Definition 4.1, $\mathrm{inf}_r = a_r^*$.

Let $h1_A$ be a monomorphism and select $r \in Z$. We see from

diagram (IV) that the rth cocycle group of the lower row of that diagram is $Z^r(Y; G, A)$ if $r \leqq -2$ and is $\ker(h\delta'_r)$ if $r \geqq -1$. Since $h1_A$ is a monomorphism the endomorphism which consists of multiplying the elements of $C^{r+1}(Y; G, A)$ by h is evidently a mono; hence, $\ker(h\delta'_r) = \ker(\delta'_r) = Z^r(Y; G, A)$. Since $\{b_i, i \in Z\}$ is a chain mapping it is now clear that $b_r(Z^r(X; G, A)) \subset Z^r(Y; G, A)$; hence, by Definition 5.1, $\mathrm{def}_r = b_r^*$. Done.

We are now going to study inflation and deflation for dimensions 0, -1, and 1.

9. Inflation in dimension zero. *We restrict ourselves in the remainder of this paper to the morphism* $(1_G, f)$: $(G, G) \to (G, G/H)$ *of example* 4.1. Hence, from now on, $X = G$, $Y = G/H$, $h = [H: 1]$ and $m = [G: H]$ where m is the index of $(G, G/H)$. We denote the order of G by n. The trace mapping $S_{G/H}$: $A^H \to A^G$ is the customary one; we usually write S_G, S_H instead of $S_{G/1}$ or $S_{H/1}$.

We know that there exists an isomorphism j: $A \to C^0(X; G, A)$ given by $(j(a))(1) = a$, where $a \in A$ and 1 is the unit element of G. (See Proposition 4.2 of [6].) The same reference tells us that there exists an isomorphism k: $A^H \to C^0(Y; G, A)$ given by $(k(a))(H) = a$, where $a \in A^H$.

PROPOSITION 9.1. The following diagram commutes

$$\begin{array}{ccc} A & \xrightarrow{j} & C^0(X; G, A) \\ {\scriptstyle i}\uparrow & & \uparrow{\scriptstyle a_0} \\ A^H & \xrightarrow{k} & C^0(Y; G, A) \end{array}$$

where i: $A^H \to A$ is the inclusion mapping.

Proof. Let $a \in A^H$. Then $(ji(a))(1) = i(a) = a$, while $(a_0k(a))(1) = (k(a))(H) = a$. Done.

We conclude that *inflation for* 0*-cochains is nothing but the inclusion mapping* i: $A^H \to A$. Since $Z^0(Y; G, A) = Z^0(X; G, A) = A^G$ (see Proposition 4.1 of [6]) and $i \mid A^G$ is the identity, *inflation for* 0*-cocycle is the identity mapping of* A^G. We have observed in § 4 that we cannot expect that $a_0(B^0(Y; G, A)) \subset B^0(X; G, A)$. Let's see what the situation is.

$B^0(Y; G, A) = S_{G/H}A^H$ and $B^0(X: G, A) = S_GA$ by Proposition 4.3 of [6]. However the inclusion goes the wrong way, that is, $S_GA \subset S_{G/H}A^H$ as follows from $S_GA = S_{G/H}S_HA \subset S_{G/H}A^H$. We conclude from Definition 4.1:

PROPOSITION 9.2. $\inf_0 = a_0^*$ iff $S_{G/H}A^H = S_GA$. In that case, $\inf_0$ is the identity mapping of A^G/S_GA. Otherwise, $\inf_0(a + S_{G/H}A^H) = ha + S_GA$ for all $a \in A^G$.

EXAMPLE 9.1. Let $A = Z$ with trivial G-action. Then, $A^H = A^G = Z$, $S_{g\backslash H}A^H = mZ$, $S_gA - nZ$ and hence, if $H \neq \{1\}$, $S_{g\,H}A^H \neq S_gA$. Furthermore, $A^G/S_{G/H}A^H = Z_m$ (the cyclic group with m elements) and $A^G/S_GA = Z_n$. We see from Proposition 9.2 that $\inf_0: Z_m \to Z_n$ is the natural monomorphism $z + mZ \to hz + nZ$ where $z \in Z$; this is also true if $H = \{1\}$. It is immediate from Proposition 9.2 that in general, *if G acts trivially on A and $h1_A$ is a monomorphism, $\inf_0$ is a monomorphism.*

REMARK 9.1. We always have $hS_{G/H}A^H \subset S_GA \subset S_{G/H}A^H$. The right hand inclusion was observed before Proposition 9.2. The left hand inclusion follows either from $ha_0(B^0(Y; G, A)) \subset B^0(X; G, A)$ (see §4) or from $S_GA \supset S_GA^H = S_{G/H}S_HA^H = hS_{G/H}A^H$.

10. **Deflation in dimension zero.** Let $j: A \to C^0(X; G, A)$ and $k: A^H \to C^0(Y; G, A)$ denote the same isomorphism as in Proposition 9.1.

PROPOSITION 10.1. The following diagram commutes.

$$\begin{array}{ccc} A & \xrightarrow{j} & C^0(X; G, A) \\ {\scriptstyle S_H}\downarrow & & \downarrow{\scriptstyle b_0} \\ A^H & \xrightarrow{k} & C^0(Y; G, A) \end{array}$$

Proof. Let $a \in A$. Then, $(kS_Ha)(H) = S_Ha$, while $(b_0j(a))(H) = j(a)(\Sigma_{\rho \in H}\rho) = \Sigma_{\rho \in H}\rho a = S_Ha$. Done.

We conclude that *deflation for 0-cochains is the trace mapping* $S_H: A \to A^H$. Furthermore, *deflation for 0-cocycles consists of multiplying the elements of A^G by h*, since this is the effect of S_H on A^G. This comes as a mild surprise since it shows that $b_0(Z^0(X; G, A)) \subset Z^0(Y; G, A)$ which, as we observed in §5, can not be expected to be true for all morphisms of permutation representations. We know from the same section that $b_0(B^0(X; G, A)) \subset B^0(Y; G, A)$ which is equivalent to saying that $hS_GA \subset S_{G/H}A^H$; this last inclusion follows from $S_GA \subset S_{G/H}A^H$, observed before Proposition 9.2.

Since $S_GA \subset S_{G/H}A^H \subset A^G$, the natural epimorphism $\gamma: A^G/S_GA \to A^G/S_{G/H}A^H$ is given by $\gamma(a + S_GA) = a + S_{G/H}A^H$, where $a \in A$. It would have been nice if γ had been def_0, but we regretfully conclude from Definition 5.1:

PROPOSITION 10.2. $\mathrm{def}_0 = b_0^*$. Explicitly, $\mathrm{def}_0(a + S_GA) = ha + S_{G/H}A^H$ for all $a \in A^G$; *i.e.*, $\mathrm{def}_0 = h\gamma$.

EXAMPLE 10.1. Let $A = Z$ with trivial G-action. Then, def_0: $Z_n \to Z_m$ is $h\gamma$, where γ: $Z_n \to Z_m$ is the natural epimorphism given by $\gamma(z + nZ) = z + mZ$ for $z \in Z$. It is clear from this example that def_0 may be neither a monomorphism nor an epimorphism.

11. Coboundaries in dimension -1. In order to study inflation in dimension -1 we need some material on the (-1)-coboundaries of the permutation representation $(G, Y) = (G, G/H)$.

Let $\sigma_1, \cdots, \sigma_m$ be a set of representatives for the left cosets of H, i.e., $Y = G/H = \{\sigma_1 H, \cdots, \sigma m H\}$. We assume that the enumeration is such that $\sigma_1 H, \cdots, \sigma_u H$ $(1 \leqq u \leqq m)$ is a set of representatives of the permutation representation $(H, G/H)$. (According to §4 of [6] this means that $(H, G/H)$ has u domains of transitivity and that $\sigma_i H$ belongs to the ith domain.) We shall use the following notation.

NOTATION 11.1. $H_i = H \cap \sigma_i H \sigma_i^{-1}$ and $M_i = H \cap \sigma_i^{-1} H \sigma_i$ for $i = 1, \cdots, u$. Observe that $M_i = \sigma_i^{-1} H_i \sigma_i$.

NOTATION 11.2. $S_i \in Z[H]$ is the sum of a fixed set of representatives for the left cosets of H_i as a subgroup of H; $S_i' \in Z[H]$ is the sum of a fixed set of representatives for the left cosets of M_i as a subgroup of H, where $i = 1, \cdots, u$. Hence the trace mapping $A^{H_i} \to A^H (A^{M_i} \to A^H)$ consists of multiplying the elements of A^{H_i} by S_i (of A^{M_i} by S_i').

We must first get a hold on $C^{-2}(Y; G, A) = \mathrm{Hom}_G(Z[Y^2], A)$.

PROPOSITION 11.1. The permutation representation (G, Y^2) has the pairs $(H, \sigma_i H)$ for $i = 1, \cdots, u$ as a set of representatives.

Proof. Let $1 \leqq i \neq j \leqq u$. Then, $\sigma(H, \sigma_i H) \neq (H, \sigma_j H)$ for all $\sigma \in G$. Namely, $\sigma H = H$ means that $\sigma \in H$ and this implies that $\sigma\sigma_i H \neq \sigma_j H$. Now consider the arbitrary pair $(\sigma H, \tau H)$ of Y^2 where $\sigma, \tau \in G$. Then, $\sigma^{-1}(\sigma H, \tau H) = (H, \sigma^{-1}\tau H)$ and there exists a $\rho \in H$ such that $\rho\sigma^{-1}\tau H = \sigma_i H$ for some $1 \leqq i \leqq u$. Since $\rho\sigma^{-1}(\sigma H, \tau H) = (H, \sigma_i H)$ we are done.

The subgroup of G which leaves the pair $(H, \sigma_i H)$ fixed is the group H_i of Notation 11.1; $i, \cdots, u$. Hence we conclude from §4 of [6] that there exists an isomorphism t: $A^{H_1} \oplus \cdots \oplus A^{H_u} \to C^{-2}(Y; G, A)$ given by: If $a_i \in A^{H_i}$ for $i = 1, \cdots, u$, then $(t(a_1, \cdots, a_u))(H, \sigma_i H) = a_i$.

We can also consider the homomorphism d_{-2}: $A^{H_1} \oplus \cdots \oplus A^{H_u} \to A^H$ given by $d_{-2}(a_1, \cdots, a_u) = \sum_{i=1}^{u} (S_i'(\sigma_i^{-1} a_i) - S_i a_i)$ where again $a_i \in$

A^{H_i} for $i = 1, \cdots, u$. (It is immediate that, if $a_i \in A^{H_i}$, then $\sigma_i^{-1}a_i \in A^{M_i}$.

Finally, since $C^{-1}(Y; G, A) = C^0(Y; G, A)$, there is available the isomorphism k: $A^H \to C^{-1}(Y; G, A)$ of Proposition 9.1.

PROPOSITION 11.2. The following diagram commutes.

$$\begin{array}{ccc} C^{-2}(Y; G, A) & \xrightarrow{\delta_{-2}} & C^{-1}(Y; G, A) \\ {\scriptstyle t}\uparrow & & \uparrow{\scriptstyle k} \\ A^{H_1} \oplus \cdots \oplus A^{H_u} & \xrightarrow[d_{-2}]{} & A^H \end{array}$$

Proof. Let $a_i \in A^{H_i}$ for $i = 1, \cdots, u$. Then $(kd_{-2}(a_1, \cdots, a_u))(H) = d_{-2}(a_1, \cdots, a_u)$. Furthermore, using the formula for δ_{-2} of §1 of [6], $(\delta_{-2}t(a_1, \cdots, a_u))(H) = t(a_1, \cdots, a_u)(\sum_{j=1}^m (\sigma_j H, H) - \sum_{j=1}^m (H, \sigma_j H))$. In order to compute the sum $\sum_{j=1}^m (H, \sigma_j H)$ we consider the permutation representation $(H, \{(H, \sigma_1 H), \cdots, (H, \sigma_m H)\})$. It is immediate that the pairs $(H, \sigma_1 H), \cdots, (H, \sigma_u H)$ also form a set of representatives for this permutation representation. Since H_i is the subgroup of H which leaves $(H, \sigma_i H)$ fixed, $\sum_{j=1}^m (H, \sigma_j H) = \sum_{i=i}^u S_i(H, \sigma_i H)$ and hence $t(a_1, \cdots, a_u)(\sum_{j=1}^m (H, \sigma_j H)) = \sum_{i=1}^u S_i a_i$. In order to compute the sum $\sum_{j=1}^m (\sigma_j H, H)$ we consider the permutation representation $(H, \{(\sigma_1 H, H), \cdots, (\sigma_m H, H)\})$. Since $\sigma_i^{-1}(H, \sigma_i H) = (\sigma_i^{-1}H, H)$ we see easily that the pairs $(\sigma_i^{-1}H, H), \cdots, (\sigma_u^{-1}H, H)$ from a set of representatives for this last permutation representation. Since M_i is the subgroup of H which leaves $(\sigma_i^{-1}H, H)$ fixed, $\sum_{j=1}^m (\sigma_j H, H) = \sum_{i=1}^u S_i'(\sigma_i^{-1}H, H)$ and hence $t(a_1, \cdots, a_u)(\sum_{j=1}^m (\sigma_j H, H)) = \sum_{i=1}^u S_i'(\sigma_i^{-1}a_i)$. We conclude that $(\delta_{-2} t(a_1, \cdots, a_u))(H) = \sum_{i=1}^u (S_i'(\sigma_i^{-1}a_i) - S_i a_i) = d_{-2}(a_1, \cdots, a_u)$. Done.

REMARK 11.1. The above elements $\sigma_1, \cdots, \sigma_u$ are nothing but a set of representatives for the double cosets of H as a subgroup of G. This remark makes it easy to check that our expression $\sum_{i=1}^u (S_i'(\sigma_i^{-1}a_i) - S_i a_i)$ for the (-1)-coboundaries of (G, Y) is equivalent to, although not identical with, the expression $*$ on page 69 of [1].

We denote the kernel of the trace mapping $S_{G/H}$: $A^H \to A^G$ by $\ker(S_{G/H})$. The ideal of $Z[G]$ which has as ideal base the elements $\sigma - 1$, where $\sigma \in G$, is as usual denoted by I.

LEMMA 11.1. $\operatorname{im}(d_{-2}) \subset (IA \cap \ker(S_{G/H}))$.

Proof. The following diagram commutes.

$$\begin{array}{ccccc} C^{-2}(Y;\,G,\,A) & \xrightarrow{\delta_{-2}} & C^{-1}(Y;\,G,\,A) & \xrightarrow{\delta_{-1}} & C^{0}(Y;\,G,\,A) \\ \uparrow t & & \uparrow k & & \uparrow k \\ A^{H_1} \oplus \cdots \oplus A^{H_u} & \xrightarrow[d_{-2}]{} & A^{H} & \xrightarrow[S_{G/H}]{} & A^{H} \end{array}$$

The left hand square commutes by Proposition 11.2; the right hand square commutes by § 4 of [6]. Since k is an isomorphism and $\delta_{-1}\delta_{-2} = 0$ we read from this diagram that $S_{G/H}d_{-2} = 0$, i.e., that $\mathrm{im}(d_{-2}) \subset \ker(S_{G/H})$. We now turn to $\mathrm{im}(d_{-2}) \subset IA$. We observe that the groups H_i and M_i of Notation 11.1 are conjugate (in G) and hence contain the same number, say c_i, of elements. Hence the two decompositions of H into the left cosets of H_i, respectively M_i, both consist of subsets of H with c_i elements. We conclude from Theorem 4 on page 12 of [8] that there exists a common set of representatives for the left cosets of H_i and of M_i as subgroups of H. We now use such a common set of representatives to compute S_i and S_i' of Notation 11.2, and obtain that $S_i = S_i'$. Hence, if $a_i \in A^{H_i}$ for $i = 1, \cdots, u$, $d_{-2}(a_1, \cdots, a_u) = \sum_{i=1}^{u} S_i(\sigma_i^{-1} - 1)a_i \in IA$. Done.

COROLLARY 11.1. *If G acts trivially on A, $\mathrm{im}(d_{-2}) = 0$.*

Proof. G acts trivially on A if and only if $IA = 0$. Done.

12. **Inflation in dimension -1.** The homomorphism a_{-1}: $C^{-1}(Y;\,G,\,A) \to C^{-1}(X;\,G,\,A)$ is identical with the homomorphism a_0: $C^0(Y;\,G,\,A) \to C^0(X;\,G,\,A)$. Consequently, Proposition 9.1 is valid with a_0 replaced by a_{-1}; i.e., $ji = a_{-1}k$. We conclude that *inflation for (-1)-cochains is the inclusion mapping i: $A^H \to A$.* Since $Z^{-1}(Y;\,G,\,A) = \ker(S_{G/H})$ and $Z^{-1}(X;\,G,\,A) = \ker(S_G)$, *inflation for (-1)-cocycles is the inclusion mapping* $\ker(S_{G/H}) \to \ker(S_G)$. (The fact that $\ker(S_{G/H}) \subset \ker(S_G)$ follows from § 4 or from $S_G = S_{G/H}S_H$.) Since $B^{-1}(Y;\,G,\,A) = \mathrm{im}(d_{-2})$ (see Proposition 11.2) and $B^{-1}(X;\,G,\,A) = IA$ we see from Lemma 11.1 that $a_{-1}(B^{-1}(Y;\,G,\,A)) \subset B^{-1}(X;\,G,\,A)$; this could *not* have been predicted from § 4. We conclude from Definition 4.1:

PROPOSITION 12.1. $\mathrm{inf}_{-1} = a_{-1}^*$. Explicity, $\mathrm{inf}_{-1}(a + \mathrm{im}(d_{-2})) = a + IA$ for all $a \in \ker(S_{G/H})$.

The following theorem is crucial for the duality theory of transitive permutation representations.

THEOREM 12.1. *Let $d = (h, m)$. If A is uniquely divisible by d, then $\mathrm{im}(d_{-2}) = IA \cap \ker(S_{G/H})$. This happens for instance in each of the following two cases:*

(a) A is uniquely divisible by h;

(b) H is a Hall subgroup of G.

Proof. We see from Proposition 12.1 that $\ker(\inf_{-1}) = (IA \cap \ker(S_{G/H}))/\mathrm{im}(d_{-2})$. Lemma 7.1 and Theorem 7.1 tell us that $\inf_{-1}$ is a monomorphism if A is uniquely divisible by d. The remainder of Theorem 12.1 follows from Proposition 7.1 and Example 7.1. Done.

13. **Deflation in dimension -1.** The homomorphism b_{-1}: $C^{-1}(X; G, A) \to C^{-1}(Y; G, A)$ is identical with the homomorphism b_0: $C^0(X; G, A) \to C^0(Y; G, A)$. Hence we conclude from Proposition 10.1 that *deflation for (-1)-cochains is the trace mapping* S_H: $A \to A^H$. It follows immediately from $S_G = S_{G/H}S_H$ that $S_H(\ker(S_G)) \subset \ker(S_{G/H})$, which signifies that $b_{-1}(Z^{-1}(X; G, A)) \subset Z^{-1}(Y; G, A)$; this could *not* have been predicted from §5. We conclude from Definition 5.1:

PROPOSITION 13.1. $\mathrm{def}_{-1} = b^*_{-1}$. Explicitly, $\mathrm{def}_{-1}(a + IA) = S_H a + \mathrm{im}(d_{-2})$ for all $a \in \ker(S_G)$.

The following theorem is the dual of Theorem 12.1.

THEOREM 13.1. *In each of the following two cases,* $\mathrm{im}(d_{-2}) + S_H(\ker(S_G)) = \ker(S_{G/H})$.

(a) *A is uniquely divisible by h.*

(b) *H is a Hall subgroup of G.*

Proof. We see from Proposition 13.1 that $\mathrm{im}(\mathrm{def}_{-1}) = [\mathrm{im}(d_{-2}) + S_H(\ker(S_G))]/\mathrm{im}(d_{-2})$. Hence, def_{-1} is an epimorphism if and only if $\mathrm{im}(d_{-2}) + S_H(\ker(S_G)) = \ker(S_{G/H})$. Proposition 7.1, Example 7.1 and Theorem 7.1 tell us that def_{-1} is an epimorphism in each of the cases (a) and (b). Done.

Lemma 13.1. $S_H(IA) \subset \mathrm{im}(d_{-2})$.

Proof. Since $B^{-1}(X; G, A) = IA$ and $B^{-1}(Y; G, A) = \mathrm{im}(d_{-2})$, Lemma 13.1 is equivalent to saying that $b_{-1}(B^{-1}(X; G, A)) \subset B^{-1}(Y; G, A)$. This last inclusion was observed in §5. Done.

14. **Inflation in dimension 1.** We denote by M the additive group of the crossed homomorphisms from G to A; and by M_H the subgroup of M whose elements are zero on H. We know from §6 of [6] that there exists an isomorphism v: $Z^1(Y; G, A) \to M_H$ which is defined by $(vc)(\sigma) = c(H, \sigma H)$ for $c \in Z^1(Y; G, A)$ and $\sigma \in G$. Similarly, the isomorphism w: $Z^1(X; G, A) \to M$ is defined by $(wc)(\sigma) = c(1, \sigma)$ where $c \in Z^1(X; G, A)$, $\sigma \in G$ and 1 is the unit element of G. We denote the inclusion mapping $M_H \to M$ by u and recall from §4 that $a_1(Z^1(Y; G, A)) \subset Z^1(X; G, A)$.

PROPOSITION 14.1. The following diagram commutes.

$$\begin{array}{ccc} Z^1(X; G, A) & \xrightarrow{w} & M \\ {\scriptstyle a_1}\uparrow & & \uparrow{\scriptstyle u} \\ Z^1(Y; G, A) & \xrightarrow[v]{} & M_H \end{array}$$

Proof. Let $c \in Z^1(Y; G, A)$ and $\sigma \in G$. Then, $((wa_1)(c))(\sigma) = (a_1c)(1, \sigma) = c(H, \sigma H)$; and $(uv(c))(\sigma) = (vc)(\sigma) = c(H, \sigma H)$. Done.

We conclude that *inflation for* 1*-cocycles is the inclusion mapping* $u: M_H \to M$. In order to study inflation for 1-coboundaries, we recall from §6 of [6] that $v(B^{-1}(Y; G, A))$ is the subgroup M'_H of M_H which is described as follows: If $g \in M'_H$ and $\sigma \in G$, then $g(\sigma) = (\sigma - 1)a$ for some fixed $a \in A^H$. The subgroup $M' = w(B^1(X; G, A))$ of M is described similarly with A^H replaced by A. Since $M'_H \subset M'$ we see that $a_1(B^1(Y; G, A)) \subset B^1(X; G, A)$ which checks with §4. We conclude from Definition 4.1:

PROPOSITION 14.2. $\inf_1 = a_1^*$. Explicitly, $\inf_1(g + M'_H) = g + M'$ for all $g \in M_H$.

It is well known that $\inf_1: H^1(Y; G, A) \to H^1(X; G, A)$ is always a monomorphism. (see Theorem 7.3 of [1] or Theorem 15.1 of [6].) This also follows from Proposition 14.2 and the observation that $M'_H = M' \cap M_H$.

15. **Endomorphisms of the group of crossed homomorphisms.** Let M and M_H be as in the previous section. In order to study deflation in dimension 1, we define what should be regarded as the natural homomorphism $D: M \to M_H$. If $g \in M$ and $\sigma \in G$ we denote the sum $\Sigma g(\gamma)$, where γ runs through σH, by $s_g(\sigma H)$. In particular $s_g(H) = \Sigma g(\rho)$, where ρ runs through H. We now define the homomorphism $D: M \to M_H$.

DEFINITION 15.1. *If* $g \in M$ *and* $\sigma \in G$, $(D(g))(\sigma) = s_g(\sigma H) - s_g(H)$.

One proves routinely that D is a homomorphism from M into M_H. We observe that $s_g(\sigma H) = \Sigma g(\sigma\rho)$, where ρ runs through H. Using that $g(\sigma\rho) = g(\sigma) + \sigma g(\rho)$, we find:

PROPOSITION 15.1. If $g \in M$ and $\sigma \in G$, $(D(g))(\sigma) = hg(\sigma) + (\sigma - 1)s_g(H)$.

EXAMPLE 15.1. Let G act trivially on A. Then, $M = \mathrm{Hom}(G, A)$ and M_H consists of those homomorphisms from G to A which vanish

on H. If $g \in \mathrm{Hom}(G, A)$, we see from Proposition 15.1 that $D(g) = hg$ and indeed, multiplication by h is the most naive way to change a homomorphism belonging to $\mathrm{Hom}(G, A)$ into one which is zero on H.

We now prepare for the study of $\ker(D)$.

PROPOSITION 15.2. If $g \in M$, $S_H(s_g(H)) = 0$.

Proof. Let $\rho \in H$. Then, $\rho s_g(H) = \Sigma \rho g(\gamma)$ where γ runs through H. Since $g(\rho\gamma) = g(\rho) + \rho g(\gamma)$, this last sum equals $-hg(\rho) + sg(H)$. Consequently, $S_H(s_g(H)) = -hs_g(H) + hs_g(H) = 0$. Done.

We know from §6 of [6] that the homomorphism δ_0': $C^0(Y; G, A) \to Z^1(Y; G, A)$ may be interpreted as the homomorphism δ_0': $A^H \to M_H$, where $(\delta_0'(a))(\sigma) = (\sigma - 1)a$ for $a \in A^H$ and $\sigma \in G$. Similarly, the homomorphism δ_0: $C^0(X; G, A) \to Z^1(X; G, A)$ may be interpreted as the homomorphism δ_0: $A \to M$, where $(\delta_0(a))(\sigma) = (\sigma - 1)a$ for $a \in A$ and $\sigma \in G$. We also recall from §10 that the homomorphism b_0: $C^0(X; G, A) \to C^0(Y; G, A)$ may be interpreted as the homomorphism S_H: $A \to A^H$.

PROPOSITION 15.3. The following diagram commutes.

$$\begin{array}{ccc} A & \xrightarrow{\delta_0} & M \\ {\scriptstyle S_H}\downarrow & & \downarrow{\scriptstyle D} \\ A^H & \xrightarrow[\delta_0']{} & M_H \end{array}$$

Proof. Let $a \in A$ and $\sigma \in G$. Then $(\delta_0' S_H(a))(\sigma) = (\sigma - 1)S_H(a)$. Furthermore, denoting $\delta_0(a) = g$, $(D\delta_0(a))(\sigma) = (Dg)(\sigma) = hg(\sigma) + (\sigma - 1)s_g(H) = h(\sigma - 1)a + (\sigma - 1)\Sigma(\rho - 1)a$ where ρ runs through H. Since $\Sigma(\rho - 1)a = S_H(a) - ha$, $(D\delta_0(a))(\sigma) = (\sigma - 1)S_H(a)$. Done.

If K is a subgroup of M we denote the larger subgroup $\{g \mid g \in M, hg \in K\}$ by $K{:}\, h$. We continue the investigation of the diagram of Proposition 15.3.

PROPOSITION 15.4. $h\ker(D) \subset \delta_0(\ker(S_H)) \subset \ker(D)$. If $h1_A$ is a monomorphism, $\ker(D) = \delta_0(\ker(S_H)){:}\, h$.

Proof. The inclusion $\delta_0(\ker(S_H)) \subset \ker(D)$ is read immediately from the commutative diagram of Proposition 15.3. In order to show that $h\ker(D) \subset \delta_0(\ker)S_H))$, we select $g \in \ker(D)$ and show that $hg \in \delta_0(\ker(S_H))$. That is, we prove that for all $\sigma \in G$, $hg(\sigma) = (\sigma - 1)a$ for some fixed $a \in \ker(S_H)$. We see from Proposition 15.1 that $hg(\sigma) = (\sigma - 1)(-s_g(H))$ and from Proposition 15.2 that $-s_g(H) \in \ker(S_H)$.

The first line of Proposition 15.4 has now been proved. We conclude from it that $\ker(D) \subset \delta_0(\ker(S_H))$: $h \subset \ker(D)$: h. If $h1_A$ is a monomorphism, $h1_M$ is a monomorphism and hence $\ker(D) = \ker(D)$: h. Done.

REMARK 15.1. We shall see in the next section that the homomorphism hD: $M \to M_H$ is precisely the deflation for 1-cocycles. Clearly, $\ker(hD) = \ker(D)$: h and hence we have good information about the kernel of the deflation mapping.

16. **Deflation in dimension 1.** One proves easily that the isomorphism v: $Z^1(Y; G, A) \to M_H$ of Proposition 14.1 has as inverse the isomorphism v': $M_H \to Z^1(Y; G, A)$ defined by: If $g \in M_H$ and σ, $\tau \in G$, then $(v'(g))(\sigma H, \tau H) = g(\tau) - g(\sigma)$. (The proof uses that $g \in M_H$ if and only if $g \in M$ and g is constant on the left cosets of H.) We shall regard v' as a monomorphism v': $M_H \to C^1(Y; G, A)$. Similarly, we have the monomorphism w': $M \to C^1(X; G, A)$ defined by: If $g \in M$ and $\sigma, \tau \in G$, then $(w'(g))(\sigma, \tau) = g(\tau) - g(\sigma)$.

PROPOSITION 16.1. The following diagram commutes.

$$\begin{array}{ccc} M & \xrightarrow{w'} & C^1(X; G, A) \\ {\scriptstyle hD}\downarrow & & \downarrow{\scriptstyle b_1} \\ M_H & \xrightarrow[v']{} & C^1(Y; G, A) \end{array}$$

Proof. Let $g \in M$ and $\sigma, \tau \in G$. Then, using Definition 15.1, $(v' hD(g))(\sigma H, \tau H) = (hD(g))(\tau) - (hD(g))((\sigma) = h(s_g(\tau H) - s_g(\sigma H))$. Furthermore $(b_1 w'(g))(\sigma H, \tau H) = w'(g)(\Sigma(\sigma\rho, \tau\gamma))$, where the summation is over all pairs $(\rho, \gamma) \in H \times H$. Consequently, $(b_1 w'(g))(\sigma H, \tau H) = \Sigma(g(\tau\gamma) - g(\sigma\rho)) = hs_g(\tau H) - hs_g(\sigma H)$. Done.

We conclude that *deflation for 1-cocycles is the mapping* hD: $M \to M_H$. We see that $b_1(Z^1(X: G, A)) \subset Z^1(Y; G, A)$ which could *not* have been predicted from §5. In order to study deflation for 1-coboundaries we return to the groups M' and M'_H of §14.

PROPOSITION 16.2. $D(M') \subset M'_H$.

Proof. We read from the diagram of Proposition 15.3 that $D\delta_0(A) = \delta'_0 S_H(A)$. Since $\delta_0(A) = M'$ and $\delta'_0 S_H(A) \subset \delta'_0(A^H) = M'_H$, we are done.

It follows trivially from Proposition 16.2 that $hD(M') \subset M'_H$, i.e., that $b_1(B^1(X; G, A)) \subset B^1(Y; G, A)$ which checks with §5. We conclude from Definition 5.1:

PROPOSITION 16.3. $\mathrm{def}_1 = b_1^*$. Explicitly, $\mathrm{def}_1(g + M') = hD(g) + M'_H$ for all $g \in M$.

REMARK 16.1. Proposition 16.2 shows that D induces a homomorphism D^*: $H^1(X; G, A) \to H^1(Y; G, A)$, given by $D^*(g + M') = D(g) + M'_H$ for all $g \in M$. Evidently, D^* is the natural mapping from $H^1(X; G, A)$ into $H^1(Y; G, A)$ and $\mathrm{def}_1 = hD^*$. The factor h is pure waste; and that, in times of deflation!

REFERENCES

1. I. T. Adamson, *Cohomology Theory for Non-Normal Subgroups and Non-Normal Fields*, Proc. Glascow Math., Assoc. **2**(1954), 66-76.
2. W. Burnside, *Theory of Groups of Finite Order*, Dover Publications, second edition (1955).
3. H. Cartan and S. Eilenberg, *Homological Algebra*, Princeton University Press (1956)
4. D. K. Faddeev, *On Homology Theory For Finite Groups of Operators*, Izv. Akad. Nauk SSSR. Ser. Mat., **19**(1955), 193-200 (Russian).
5. A. Grothendieck, *Sur Quelques Points D'Algébre Homologique*, Tôhoku Math. J., **9** (1957), 119-221.
6. E. Snapper, *Cohomology of Permutation Representations. I. Spectral Sequences*, J. Of Math. and Mech., **13**(1964), 133-161.
7. E. Weiss, A *Deflation Map*, J. of Math. and Mech., **8**(1959), 309-329.
8. H. Zassenhaus, *The Theory of Groups*, Chelsea Publishing Company, (1949).

DARTMOUTH COLLEGE

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

ON THE STRICT AND UNIFORM CONVEXITY OF CERTAIN BANACH SPACES

K. SUNDARESAN

Let (X, S, μ) be a σ-finite non-atomic measure space let N be a real valued continuous convex even function defined on the real line such that

(1) $N(u)$ is nondecreasing for $u \geqq 0$,

(2) $\lim_{u\to\infty} N(u)/u = \infty$,

(3) $\lim_{u\to 0} N(u)/u = 0$.

Let L_N be the set of all real valued μ-measurable functions f such that $\int_X N(f)d\mu < \infty$. It is known that if there exists a constant k such that $N(2u) \leqq kN(u)$ for all $u \geqq 0$ then L_N is a linear space; in fact, L_N is a B-Space if a norm $||\cdot||$ is defined by setting

$$(*) \qquad ||f|| = \inf\left\{1/\zeta \,\middle|\, \zeta > 0, \int_X N(\eta, f)d\mu \leqq 1\right\}.$$

Denoting the B-space $(L_N, ||\cdot||)$ by L_N^* it is proposed to obtain the necessary and sufficient conditions in order that L_N^* may be (1) Strictly Convex (2) Uniformly Convex.

The linear space L_N admits another norm $|||\cdot|||_{(N)}$ known as the Orlicz norm defined by setting

$$|||f|||_{(N)} = \sup \int_X |fg|\, d\mu$$

for such that $\int_X M(|g|)d\mu \leqq 1$, M being the function complementary to N in the sense of Young. For a discussion of this class of Banach spaces we refer to Mazur and Orlicz [2]. Convexity properties of the Orlicz norm have been studied in Milnes [3].

The space L_N^* may be considered as a modulared linear space defined in Nakano [4]. A nonnegative extended real valued function m defined on a linear space is called a *modular* if

(i) $m(0) = 0$;

(ii) for any $x \in L$ there exists $\xi > 0$ such that $m(\xi x) < \infty$;

(iii) $m(\xi x) = 0$ for all $\xi > 0$ implies $x = 0$;

(iv) $m(x) = \sup_{0\leqq\xi<1} m(\xi x)$;

(v) m is convex (i.e., $\alpha \geqq 0$, $\beta \geqq 0$, $\alpha + \beta = 1$, $x, y \in L$ imply $m(\alpha x + \beta y) \leqq \alpha m(x) + \beta m(y)$).

Received January 20, 1965. Research work supported by N.S.F. Grant GP-378.

The *modulared linear space* may be considered as a normed linear space if a norm $||\cdot||$ is defined by setting

$$(**) \qquad ||x|| = \inf\{1/\xi \mid \xi > 0 \text{ and } m(\xi x) \leqq 1\}.$$

We note that the linear space L_N is a modulared space if

$$m(f) = \int_X N(f)d\mu,$$

and the norm $||\cdot||$ defined by (**) is the same as the norm defined in *. In fact, the modulared space L_N is a *finite modulared* space, meaning that $m(f) < \infty$, for all $f \in L_N$.

A Banach space B is said to be strictly convex if $x, y \in B$, $||x|| = ||y|| = ||(x+y)/2|| = 1$ imply $x = y$. It is uniformly convex if to each ε, $0 < \varepsilon \leqq 2$, there corresponds a $\delta(\varepsilon) > 0$ such that conditions $||x|| = ||y|| = 1$, $||x - y|| \geqq \varepsilon$ imply that $||x + y|| < 2 - \delta(\varepsilon)$.

We shall start by characterizing the strict convexity of L_N^*.

LEMMA 1. *The modulared norm defined in* (**) *associated with a finite modulared space is strictly convex if and only if* $m(x) = m(y) = m\{(x+y)/2)\} = 1$ *imply* $x = y$.

The proof is an easy consequence of the fact that in a finite modulared space, $m(x) = 1$ if and only if $||x|| = 1$ where $||\cdot||$ is the related modulared norm.

THEOREM. *The Banach space* L_N^* *is strictly convex if and only if the N-function N is strictly convex; i.e.,*

$$N\left(\frac{u+v}{2}\right) < \frac{1}{2}[N(u) + N(v)]$$

for all real u, v *such that* $u \neq v$.

Proof. Let N be a strictly convex N-function. Let $f, g \in L_N^*$ such that

$$m(f) = m(g) = m\left(\frac{f+g}{2}\right) = 1.$$

By definition of m it follows that

$$\int_X \left[\frac{N(f) + N(g)}{2}\right] - N\left(\frac{f+g}{2}\right)\Big] d\mu = 0.$$

whence the convexity of N together with the restrictions on f, and g imply that $f = g$ a.e. Thus by Lemma 1, L_N^* is strictly convex.

To prove the "only if" part, let L_N^* be strictly convex. If possible let N be not strictly convex so that there exist $a, b \geqq 0$ $a \neq b$ such that $N\{(a+b)/2\} = 1/2\,[N(a) + N(b)]$. The continuity of N together with the condition $\lim_{u \to 0} N(u)/u = 0$ imply that N is linear on the interval $[a, b]$ and $a \neq 0$, $b \neq 0$. For $u \in [a, b]$ let $N(u) = pu + q$, where p and q are reals.

Since μ is a nonatomic positive measure there exist pairwise disjoint measurable sets A, B, C of arbitrarily small measure such that

$$\mu(A) = \mu(B) = \mu(C) .$$

Let us define functions f, g as follows. Let $f(x) = a$ for $x \in A$, $f(x) = b$ for $x \in B$, and $f(x) = 0$ for all $x \notin A \cup B$. Let $g(x) = b$ for $x \in A$, $g(x) = a$ for $x \in B$, and $g(x) = 0$ for $x \notin A \cup B$, and $g(x) = 0$ for $x \notin A \cup B$. Then

$$m(f) = \int_X N(f)d\mu = [p(a+b) + 2q]\mu(A) ,$$

$$m(g) = \int_X N(g)d\mu = [p(a+b) + 2q]\mu(B) ,$$

$$m\left(\frac{f+g}{2}\right) = \frac{1}{2}\,[m(f) + m(g)] ,$$

and $m(f) = m(g) = m\{(f+g)/2\}$. By a suitable choice of A, B, C we can assume that

$$m(f) = m(g) = m\left(\frac{f+g}{2}\right) = K < \frac{1}{2} .$$

Now let h be a function on X defined by setting

$$h(x) = 0 \text{ if } X \in C , \qquad h(x) = t \quad \text{if} \quad x \in C$$

where t is such that $N(t)/\mu(C) = 1 - K$. Let $f_1 = h + f$, and $g_1 = h + g$; since $h \wedge f = 0 = h \wedge g$, we obtain

$$m(f_1) = m(h) + m(f) = (1 - K) + K = 1 .$$

Similarly $m(g_1) = 1$, and further

$$m\left(\frac{f_1 + g_1}{2}\right) = m\left(\frac{f + g}{2} + h\right) = m\left(\frac{f+g}{2}\right) + m(h) = 1 .$$

Thus we have $f_1 \in L_N^*$, $g_1 \in L_N^*$ and $m(f_1) = m(g_1) = m\{(f_1 + g_1/2)\} = 1$; however $f_1 \neq g_1$. Thus L_N^* is not strictly convex, a contradiction.

We next proceed to characterize the uniform convexity of L_N^*.

It is known [5] that in a modulared semiordered linear space, the modular norm is uniformly convex if and only if the associated norm

is uniformly convex. The modulared linear spaces L_N are modulared semiordered linear spaces under the natural pointwise ordering, and the above two norms are respectively the norms $\|\cdot\|_{(N)}$ and $|||\cdot|||_{(N)}$.

With this remark we conclude that the Theorem 8 in Milnes [3] which characterizes the uniform convexity of the norm $|||\cdot|||_{(N)}$ also characterizes the uniform convexity of the norm $\|\cdot\|_{(N)}$.

I wish to thank Professor Victor L. Klee for his valuable suggestions during the preparation of the manuscript, and the referee for his suggestions and corrections which lead to a revision of the manuscript.

REFERENCES

1. Krasnoselikii and Rutickii, *Convex functions and Orlicz spaces*, Translation by L. F. Boron, Nordoff Ltd., 1961.
2. Mazur and Orlicz, *On some classes of linear spaces*, Stud. Mathem. **17** (1957), 97-119.
3. H. W. Milnes, *Convexity in Orlicz spaces*, Pacific J. Math. **7** (1957), 1451-1458.
4. H. Nakano, *Topology and Linear Topological Spaces*, Maruzen and Co., 1951, pp. 204.
5. Ando Tsyuoshi, *Convexity and evenness in modulared semi-ordered linear spaces*, J. Fac. Sci. Hokkaido Univ. Ser 1.14, (1955), 59-95.

UNIVERSITY OF WASHINGTON
AND
CARNEGIE INSTITUTE OF TECHNOLOGY

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

MAXIMAL CONVEX FILTERS IN A LOCALLY CONVEX SPACE

F. J. WAGNER

Let E $[\mathscr{T}]$ be a locally convex space, $\mathfrak{B}$ a saturated covering of E by bounded sets, and E' the topological dual of $E[\mathscr{T}]$. Let $\mathscr{T}_{\mathfrak{B}}$ be the topology on E' of uniform convergence on sets of $\mathfrak{B}$ and E'' the topological dual of $E'[\mathscr{T}_{\mathfrak{B}}]$. We assume E'' has the natural topology $\mathscr{T}_n$—that of uniform convergence on the equicontinuous sets of E'.

This article includes the following: (1) an intrinsic characterization for a bounded convex set B of E of the closure $\bar{B}$ of B in E''; (2) an intrinsic characterization of the closure $\bar{E}$ of E in E''; and (3) necessary and sufficient conditions that $\bar{E}$ be E''.

The spaces β. Let $\mathfrak{M}$ be the class of all closed convex neighborhoods[1] of 0 in $E[\mathscr{T}]$, and $B \in \mathfrak{B}$. A filter $\mathfrak{F}$ on B is called a *convex filter* if, for every $F \in \mathfrak{F}$, there exist $M, N \in \mathfrak{M}$ and $\chi \in E$ such that $\mathring{M} \supset N$, $F \supset (M + \chi) \cap B$, and $(N + \chi) \cap B \in \mathfrak{F}$. Clearly if $\mathfrak{F}$ and $\mathfrak{G}$ are two convex filters on B, such that every set of $\mathfrak{F}$ meets every set of $\mathfrak{G}$, then the least upper bound filter of $\mathfrak{F}$ and $\mathfrak{G}$ on B is also convex. Furthermore:

LEMMA 1. *For $M, N \in \mathfrak{M}$, if $\mathring{M} \supset N$, then there exists $K \in \mathfrak{M}$ such that $\mathring{M} \supset K \supset \mathring{K} \supset N$.*

Proof. If p and q are the distance functions of M and N, then $1/2(p + q)$ is the distance function of such a K.

THEOREM 1. *A convex filter $\mathfrak{F}$ on B is a maximal convex filter on B if and only if, for every two closed convex bodies K and L of E such that $\mathring{K} \supset L$, either $K \cap B \in \mathfrak{F}$ or $B \backslash L \in \mathfrak{F}$.*

Proof. Assume $\mathfrak{F}$ is maximal and let K and L be as above, and let $B \backslash L \notin \mathfrak{F}$. Let $x \in \mathring{L}$ and define a sequence $\{M_n\}$ in $\mathfrak{M}$ so that

$$\mathring{K} - x \supset M_1 \supset \mathring{M}_1 \supset L - x \quad \text{and} \quad \mathring{M}_n \supset M_{n+1} \supset \mathring{M}_{n+1} \supset L - x \quad (n \geqq 1).$$

Then the filter $\mathfrak{G}$ on B with base $\{(M_n + x) \cap B \mid n = 1, 2, 3, \cdots\}$ is

Received July 8, 1964 and in revised form January 11, 1965. Supported by National Science Foundation grant NSF G-24865.

[1] The notation and definitions are principally those of Gottfried Köthe, Topologische Lineare Räume I, Springer-Verlag, Berlin, 1960.

convex and $K \cap B \in \mathfrak{G} \subset \mathfrak{F}$.

Conversely let $\mathfrak{F}$ and $\mathfrak{G}$ be two convex filters on B such that $\mathfrak{F}$ is strictly weaker than $\mathfrak{G}$. Let $G \in \mathfrak{G}$, $M, N \in \mathfrak{M}$, and $x \in E$ such that $G \notin \mathfrak{F}$, $\mathring{M} \supset N$, $G \supset (M + x) \cap B$, and $(N + x) \cap B \in \mathfrak{G}$. Then neither $(M + x) \cap B$ nor $B \backslash (L + x) \in \mathfrak{F}$.

REMARKS 1. For every $x \in B$, $\mathfrak{V}_B(x) = \{V \cap B \subset B \mid V$ a neighborhood of x in $E\}$ is a maximal convex filter on B.

2. For a maximal convex filter $\mathfrak{F}$ on B, there is $x \in B$ such that $\mathfrak{F} = \mathfrak{V}_B(x)$ if and only if $\mathfrak{F}$ has nonempty intersection.

LEMMA 2. *Every maximal convex filter on B is a weak Cauchy filter.*

Proof. Let $\mathfrak{F}$ be a maximal convex filter on B,

$$u \in E',\ M = \{x \in E \mid |ux| \leqq 1/2\} \quad \text{and} \quad N = \{x \in E \mid |ux| \leqq 1/4\}.$$

Then $M, N \in \mathfrak{M}$ and $\mathring{M} \supset N$. Since B is weakly precompact, there exist $x_1, x_2, \cdots, x_n \in E$ such that $\bigcup_{i=1}^{n} (N + x_i) \supset B$, and so $(M + x_i) \cap B \in \mathfrak{F}$ for some $1 \leqq i \leqq n$. For $x, y \in (M + x_i) \cap B$, we have $|ux - uy| \leqq 1$.

For a maximal convex filter $\mathfrak{F}$ on B and $u \in E'$, let $\mathfrak{F}(u)$ denote the limit of the restriction of u to B according to the filter $\mathfrak{F}$.

LEMMA 3. *For every maximal convex filter $\mathfrak{F}$ on B, the mapping $u \to \mathfrak{F}(u)$ on E' is linear and $\mathscr{T}_{\mathfrak{B}}$ continuous.*

Proof. Linearity is easily proved. Also let V be the polar set of the absolutely convex hull of $2B$, $u \in V$, and $F \in \mathfrak{F}$ such that $|ux - \mathfrak{F}(u)| \leqq 1/2$ for every $x \in F$. Then, for such an x, we have $|\mathfrak{F}(u)| \leqq |\mathfrak{F}(u) - ux| + |ux| \leqq 1$.

We shall denote by $\beta = \beta_B$ the set of all maximal convex filters on B. By Lemma 3 there is a mapping π_B from β_B into E'' such that $\pi_B(\mathfrak{F})(u) = \mathfrak{F}(u)$ for every $\mathfrak{F} \in \beta_B$ and $u \in E'$.

THEOREM 2. *If either $\mathscr{T}$ is the weak topology or B is convex, then π_B is a one-to-one mapping of β_B onto the $\mathscr{T}_n$-closure $\bar{B}$ of B in E''.*

Proof. For $\mathfrak{F} \in \beta_B$, $\pi_B(\mathfrak{F})$ is in the weak closure of B in E''. For

given $u_1, \cdots, u_n \in E'$ and $\varepsilon > 0$, let $F_1, \cdots, F_n \in \mathfrak{F}$ such that $|u_i x - \mathfrak{F}(u_i)| \leqq \varepsilon$ $(1 \leqq i \leqq n)$ and $x \in \bigcap_{i=1}^n F_i$. Then $|\mathfrak{F}(u_i) - u_i x| \leqq \varepsilon$, $(1 \leqq i \leqq n)$.

Also, if B is convex, $\pi_B(\mathfrak{F})$ is in the $\mathscr{T}_n$-closure $\bar{B}$ of B in E''. Suppose the contrary. Then there is a continuous real linear functional w on E'' and a real number r such that $w(\pi_B(\mathfrak{F})) < r$ and $wz > r$ for every $z \in \bar{B}$.

Assume first that E is a real vector space. Let u be the restriction of w to E', so $u \in E$. Let $F \in \mathfrak{F}$ such that $|ux - \mathfrak{F}(u)| < r - w(\pi_B(\mathfrak{F}))$ for every $x \in F$. Then, for such an x, we have $wx = ux - \mathfrak{F}(u) + \mathfrak{F}(u) < r$. But $x \in B$.

Now let E be a complex vector space. Then there is a complex linear functional v on E'' such that $w = \Re v$. Let u be the restriction of v to E and $F \in \mathfrak{F}$ such that $|ux - \mathfrak{F}(u)| \leqq r - w(\pi_B(\mathfrak{F}))$ for every $x \in F$. Then for such an x we have $wx = \Re(vx) = \Re(ux - \mathfrak{F}(u)) + \Re(\mathfrak{F}(u)) < r$. Again, we have a contradiction.

Thus $\pi_B(\beta_B) \subset \bar{B}$ if $\mathscr{T}$ is the weak topology or B is convex.

On the other hand, if $z \in \bar{B}$, then:

$$\mathfrak{V}_B(z) = \{V \cap B \subset B \mid V \text{ a neighborhood of } z \text{ in } E''[\mathscr{T}_n]\} \in \beta_B$$

and $\pi_B(\mathfrak{V}_B(z)) = z$. Let V be a neighborhood of z in $E''[\mathscr{T}_n]$, and let U and W be closed convex neighborhoods of 0 in $E''[\mathscr{T}_n]$ such that $\mathring{U} \supset W$ and $U + U \subset V - z$. Let $\chi \in (U + z) \cap (-\mathring{W} + z) \cap B$, $M = U \cap E$, and $N = V \cap E$. Then $M, N \in \mathfrak{M}$ and $\mathring{M} \supset N$, $V \supset (M + \chi) \cap B$, and $(N + \chi) \cap B = (W + \chi) \cap B \in \mathfrak{V}_B(z)$. Thus $\mathfrak{V}_B(z)$ is convex.

Let K and L be closed convex bodies of E such that $\mathring{K} \supset L$. Let $x \in \mathring{L}$, $M = K - x$, and $N = L - x$. Either $z \in$ interior $M^{\circ\circ} + x$——in which case $K \cap B = (M + x) \cap B = (M^{\circ\circ} + x) \cap B \in \mathfrak{V}_B(z)$——or $z \notin N^{\circ\circ} + x$——in which case $E''\backslash(N^{\circ\circ} + x)$ is a neighborhood of z in E'' and so $B\backslash L = [E''\backslash(N^{\circ\circ} + x)] \cap B \in \mathfrak{V}_B(z)$. Thus $\mathfrak{V}_B(z) \in \beta_B$.

Finally, let $u \in E'$, $\varepsilon > 0$, and $F \in \mathfrak{V}_B(z)$ such that $|ux - \mathfrak{V}_B(z)(u)| \leqq \varepsilon/2$ for every $x \in F$. Let $V = \{w \in E'' \mid |wu - zu| \leqq \varepsilon/2\}$. Then, for $x \in F \cap V$, we have $|\mathfrak{V}_B(z)(u) - zu| \leqq |\mathfrak{V}_B(z)(u) - ux| + |ux - zu| \leqq \varepsilon$. Therefore, $\pi_B(\mathfrak{V}_B(z))(u) = zu$ for $u \in E''$, and so $\pi_B(\mathfrak{V}_B(z)) = z$.

REMARK. Thus $\pi_B(\mathfrak{V}_B(z)) = z$ for $z \in \bar{B}$ and $\mathfrak{F} = \mathfrak{V}_B(\pi_B(\mathfrak{F}))$ for $\mathfrak{F} \in \beta_B$.

COROLLARY 1. *If either $\mathscr{T}$ is the weak topology or B is convex, then every maximal convex filter on B is a $\mathscr{T}$-Cauchy filter.*

COROLLARY 2. *If either $\mathscr{T}$ is the weak topology or B is convex,*

then for every $\mathfrak{F} \in \beta_B$ *and* $M \in \mathfrak{M}$, *there exist* $x \in B$ *such that* $(M + x) \cap B \in \mathfrak{F}$.

Proof. Let $F \in \mathfrak{F}$ such that $F - F \subset M$ and $x \in F$.
For $M \in \mathfrak{M}$ and $x \in B$ we define:

$$\nu_B(M, x) = \{\mathfrak{F} \in \beta_B \mid (\mathring{M} + x) \cap B \in \mathfrak{F}\}$$
$$\mu_B(M, x) = \{\mathfrak{F} \in \beta_B \mid \pi_B(\mathfrak{F}) \in \text{interior } M^{\circ\circ} + x\} .$$

For $M, N \in \mathfrak{M}$ and $x, y \in B$, if $z \in (\mathring{M} + x) \cap (\mathring{N} + y) \cap B$ and $K = (M + x - z) \cap (N + y - z)$, then $\nu_B(M, x) \cap \nu_B(N, y) = \nu_B(K, z)$ and $\mu_B(M, x) \cap \mu_B(N, y) = \mu_B(K, z)$. Hence the class of all sets of the form $\nu_B(M, x)$ and the class of all sets of the form $\mu_B(M, x)$ (for $M \in \mathfrak{M}$ and $x \in B$) form bases of topologies, called the ν- and μ-topologies respectively, on β_B.

THEOREM 3. *If* $\pi_B(\beta_B) \subset \bar{B}$ *(in particular if either* $\mathscr{T}$ *is the weak topology or* B *is convex), then* ν- *and* μ-*topologies coincide and* π_B *is a homeomorphism of* β_B *onto* $\bar{B}$.

Proof. If $\pi_B(\beta_B) \subset \bar{B}$, then, for $M \in \mathfrak{M}$ and $x \in B$, we have $\mu_B(M, x) \subset \nu_B(M, x)$, and so the identity mapping of β_B with the μ-topology onto β_B with the ν-topology is continuous.

Also π_B from β_B with the ν-topology onto $\bar{B}$ is continuous. Let $\mathfrak{F} \in \beta_B$ and V a neighborhood of $\pi_B(\mathfrak{F})$ in $E''[\mathscr{T}_n]$. Let U be a closed convex neighborhood of 0 in E'' such that $U + U \subset V - \pi_B(\mathfrak{F})$, $M = U \cap E$, and $x \in (\mathring{U} + \pi_B(\mathfrak{F})) \cap B$. Then $(\mathring{M} + x) \cap B \in \mathfrak{B}_B(\pi_B(\mathfrak{F})) = \mathfrak{F}$, and so $\mathfrak{F} \in \nu_B(M, x)$. Also if $\mathfrak{G} \in \nu_B(M, x)$, there is a neighborhood W of $\pi_B(\mathfrak{G})$ such that $W \cap B = (M + x) \cap B = (U + x) \cap B$ so

$$\pi_B(\mathfrak{G}) \in \overline{W \cap B} \subset U + x \subset U + U + \pi_B(\mathfrak{F}) \subset V .$$

Finally π_B^{-1} from $\bar{B}$ onto β_B with the μ-topology is continuous by the definition of the sets μ.

COROLLARY 1. *If either* $\mathscr{T}$ *is the weak topology or* B *is convex, then* B *is closed in* $E''[\mathscr{T}_n]$ *if and only if every maximal convex filter on* B *has nonempty intersection.*

COROLLARY 2. B *is weakly compact if and only if every maximal weakly convex filter on* B *has nonempty intersection.*

2. The space η. Let $\mathfrak{A}$ denote the class of all convex sets of $\mathfrak{B}$ and $\alpha = \bigcup_{B \in \mathfrak{A}} \beta_B$ the topological union of the spaces β_B. Let π be

the continuous function from α into $E''[\mathscr{T}_n]$ defined by $\pi(\mathfrak{F}) = \pi_B(\mathfrak{F})$ if $\mathfrak{F} \in \beta_B$. For $A, B \in \mathfrak{A}$ such that $A \subset B$, define a mapping g_{BA} from β_A into β_B by $g_{BA}(\mathfrak{F}) = \mathfrak{V}_B(\pi_A(\mathfrak{F}))$ (for $\mathfrak{F} \in \beta_A$). Then $g_{BA} = \pi_B^{-1}\pi_A$ and consequently is a homeomorphism of β_A into β_B. Also, if $A \subset B \subset C$, then $g_{CA} = g_{CB}\, g_{BA}$.

THEOREM 4. *Let $A, B \in \mathfrak{A}$ such that $A \subset B$, and let $\mathfrak{F} \in \beta_A$ and $\mathfrak{G} \in \beta_B$. The following three conditions are equivalent;*

(a) $\mathfrak{G} = g_{BA}(\mathfrak{F})$;
(b) $\pi(\mathfrak{F}) = \pi(\mathfrak{G})$;
(c) *Every set of $\mathfrak{G}$ contains a set of $\mathfrak{F}$.*

Proof. $\mathfrak{F} = \mathfrak{V}_A(\pi_B(\mathfrak{F}))$, $\mathfrak{G} = \mathfrak{V}_B(\pi_B(\mathfrak{G}))$, and $g_{BA}(\mathfrak{F}) = \mathfrak{V}_B(\pi_B(\mathfrak{F}))$. Hence (a) and (b) are equivalent. Also (b) implies (c): Given $G \in \mathfrak{G}$ there is a neighborhood V of $\pi(\mathfrak{G}) = \pi(\mathfrak{F})$ such that $G = V \cap B \supset V \cap A \in \mathfrak{F}$. Also (c) implies (b): If $\pi(\mathfrak{F}) \neq \pi(\mathfrak{G})$, then $\pi(\mathfrak{F})$ and $\pi(\mathfrak{G})$ have disjoint neighborhoods V and W in E'', and so $W \cap A$ is a set of $\mathfrak{G}$ containing no set of $\mathfrak{F}$.

COROLLARY. *Let A and $B \in \mathfrak{A}$, $\mathfrak{F} \in \beta_A$, and $\mathfrak{G} \in \beta_B$. The following three conditions are equivalent:*

(a) $\pi(\mathfrak{F}) = \pi(\mathfrak{G})$.
(b) *There exists $C \in \mathfrak{A}$ such that $C \supset A \cup B$ and $g_{CA}(\mathfrak{F}) = g_{CB}(\mathfrak{G})$.*
(c) *There exists $C \in \mathfrak{A}$ and $\mathfrak{H} \in \beta_C$ such that $C \supset A \cup B$ and every set of $\mathfrak{H}$ contains a set of $\mathfrak{F}$ and a set of $\mathfrak{G}$.*

Now let R be the equivalence relation $\pi(\mathfrak{F}) = \pi(\mathfrak{G})$ on α, η the quotient space α/R, ρ the canonical mapping of α onto η, and σ the mapping from η into E'' such that $\pi = \sigma\rho$.

THEOREM 5. *σ is a homeomorphism of η onto the $\mathscr{T}_n$-closure $\bar{E}$ of E in E''.*

Proof. We need only prove $\sigma(\eta) = \pi(\alpha) \supset \bar{E}$. Consider the dual system $\langle E', \bar{E}\rangle$. Since every $u \in E'$ is uniformly continuous on E, the topology induced on $\bar{E}$ by $\mathscr{T}_n$ is admissible for this dual system. For $z \in \bar{E}$, there is a closed absolutely convex set $B \in \mathfrak{B}$ such that $|zu| \leqq 1$ for every $u \in B^\circ$. Hence, $z \in B^{\circ\circ}$ = the closure of B in any admissible topology = the $\mathscr{T}_n$-closure $\bar{B}$ of B.

For $B \in \mathfrak{A}$, the weakest topology on β_B for which every function of the form $\mathfrak{F} \to \mathfrak{F}(u)$ (for $u \in E'$) is continuous will be called the *weak topology* of β_B. Clearly β_B in the weak topology is homeomorphic

with $\bar{B}$ in the topology induced on $\bar{B}$ by the weak-star topology of E''.

THEOREM 6. *The following three conditions are equivalent*:

(a) $\bar{E} = E''$;

(b) $\bar{B}$ *is weak-star compact for every* $B \in \mathfrak{A}$;

(c) β_B *is weakly compact for every* $B \in \mathfrak{A}$.

Proof. Clearly (b) and (c) are equivalent. Also (a) implies (b); by the Alaoglu—Bourbaki theorem, for $B \in \mathfrak{A}$, the weak-star closure of B in $E'' = \bar{E}$ is weak-star compact; but since $\mathscr{T}_n$ is an admissible topology for the dual systm $\langle E', \bar{E}\rangle$, this weak-star closure is $\bar{B}$. Finally (b) implies (a): regarding $\mathfrak{B}$ as a total class of bounded subsets of $\bar{E}$, by the Mackey-Arens theorem $\mathscr{T}_{\mathfrak{B}}$ is an admissible topology for the dual system $\langle E', \bar{E}\rangle$, and so $E'' = \bar{E}$.

THEOREM 7. *For* $B \in \mathfrak{A}$, β_B *is weakly compact if and only if for every maximal weakly-convex filter* $\mathfrak{F}$ *on* B, *there is a maximal* $\mathscr{T}$*-convex filter on* B *which is stronger than* $\mathfrak{F}$.

Proof. Let β_B^w be the space of all maximal weakly convex filters on B and π_B^w the homeomorphism of β_B^w into E'' with the weak-star topology. In general $B \subset \pi_B(\beta_B) = \bar{B} \subset$ weak-star closure of $B = \pi_B^w(\beta_B^w)$.

If β_B is weakly compact, then $\pi_B^w(\beta_B^w) = \pi_B(\beta_B) = \bar{B}$. So, for $\mathfrak{F} \in \beta_B^w$, $\pi_B^w(\mathfrak{F}) \in \bar{B}$ and hence $\mathfrak{B}_B(\pi_B^w(\mathfrak{F})) \in \beta_B$ is stronger than $\mathfrak{F}$.

Conversely, let $\mathfrak{F} \in \beta_B^w$ and $\mathfrak{G} \in \beta_B$ stronger than $\mathfrak{F}$. Then $\pi_B^w(\mathfrak{F}) = \pi_B(\mathfrak{G})$, and so $\pi_B^w(\beta_B^w) \subset \pi_B(\beta_B)$.

COROLLARY. $\bar{E} = E''$ *if and only if, for every* $B \in \mathfrak{A}$ *and every maximal weakly-convex filter* $\mathfrak{F}$ *on* B, *there is a* $\mathscr{T}$*-convex filter on* B *stronger than* $\mathfrak{F}$.

3. Acknowledgement. The author wished to express his gratitude to Professor Ky Fan for the help and encouragement given him in the work presented here.

UNIVERSITY OF CINCINNATI

PACIFIC JOURNAL OF MATHEMATICS
Vol. 15, No. 3, 1965

TRANSLATION-INVARIANT FUNCTION ALGEBRAS ON COMPACT GROUPS

JOSEPH A. WOLF

Let X be a compact group. $\mathfrak{C}(X)$ denotes the Banach algebra (point multiplication, sup norm) of continuous complex-valued functions on X. A is any closed subalgebra of $\mathfrak{C}(X)$ which is stable under right and left translations and contains the constants. It is shown, by means of the Peter-Weyl Theorem and some multilinear algebra, that the condition (*) *every representation of degree* 1 *of* X *has finite image* is necessary and sufficient that every possible A be self-adjoint. If X is connected, then (*) means that X is a projective limit of semisimple Lie groups; if X is a Lie group, then (*) means that X is semisimple. The Stone-Weierstrass Theorem then gives a quick classification of all possible algebras A on an arbitrary connected semisimple Lie group X.

In an earlier paper we characterized the compact groups on which every closed translation-invariant[1] function space is self adjoint [1, Theorem 4.1]. An application of the Stone-Weierstrass Theorem [1, § 7] resulted in a classification of the closed translation-invariant function algebras on the connected Lie groups which satisfied the conditions of the characterization. Those Lie groups are the compact connected semisimple Lie groups with no simple component locally isomorphic to $SU(n)$ $(n > 2)$, to $SO(4n + 2)$, nor to E_6.

In this paper we give a direct characterization of the compact groups on which every closed translation-invariant function algebra is self adjoint. For compact connected Lie groups, the characterization is that the group be semisimple. Many of these groups have closed translation-invariant function spaces which are not self adjoint. Finally we classify the closed translation-invariant function algebras on compact simple simply connected Lie groups, as an example of a general enumeration procedure.

2. **Notation.** Let X be a compact group. D denotes the set of all (equivalence classes of) irreducible finite dimensional representations of X. Let $\beta \in D$; then W^β denotes the representation space, and β^* is the representation of X on the dual space of W^β induced by β. β^* is called the *contragredient* of β. $E^\beta = W^\beta \otimes W^{\beta^*}$ is identified with the space of matrix functions (coefficients) of β. Let $C(X)$ denote the

Received April 20, 1964. Partially supported by National Science Foundation Grant GP-812.

[1] Stable under both right and left translations.

set of continuous complex-valued functions on X with sup norm and point multiplication. Then $C(X)$ is a commutative Banach algebra. The Peter-Weyl Theorem says that every element of $C(X)$ has a Fourier development along $\{E^\beta\}_{\beta\in D}$, i.e.,

$$(*) \qquad C(X) \sim \sum_{\beta\in D} E^\beta .$$

The *translation group* $T(X)$ of X is the group consisting of all transformations

$$(u, v)\colon x \longrightarrow uxv^{-1} , \qquad u, v, x \in X ,$$

of X. A subset $S \subset C(X)$ is called *translation-invariant* if $T(X)\cdot S = S$. The closed translation-invariant subspaces of $C(X)$ are just the subspaces spanned by sets of the form $\{E^\beta\}_{\beta\in F}$ with $F \subset D$.

If X is a compact connected Lie group endowed with a Riemannian metric invariant under both right and left translations, then X is a Riemannian symmetric space and $T(X)$ is the largest connected group of isometries.

A subspace of $C(X)$ is called *self-adjoint* if it is closed under complex conjugation of functional values.

Let α and β be representations of a group Y, with finite dimensional representation spaces V and W. Then $V \otimes W$ denotes the tensor product. If $\{v_i\}$ is a basis of V and $\{w_j\}$ is a basis of W, then $\{v_i \otimes w_j\}$ is a basis of $V \otimes W$. We have matrix functions defined by

$$\alpha(y)\cdot v_i = \sum_u \alpha_{iu}(y)\cdot v_u$$
$$\beta(y)\cdot w_i = \sum_t \beta_{jt}(y)\cdot w_t .$$

Now let $\alpha \otimes \beta$ be the representation of Y on $V \otimes W$ given by

$$(\alpha \otimes \beta)(y)\cdot v \otimes w = \{\alpha(y)\cdot v\} \otimes \{\beta(y)\cdot w\} .$$

$\alpha \otimes \beta$ is given on the basis $\{v_i \otimes w_j\}$ by

$$(\alpha \otimes \beta)(y)\cdot v_i \otimes w_j = \sum_{u,t} \alpha_{iu}(y)\beta_{jt}(y) v_u \otimes w_t .$$

Let α be a representation of a group Y on a finite dimensional space V. The symmetric group S_r on r letters acts on $V \otimes \cdots \otimes V$ (r times) by permutation of the factors. The *alternation* $\Lambda^r(V)$ consists of all elements $p \in V \otimes \cdots \otimes V$ (r times) such that $\sigma(p) = (\operatorname{sign}\sigma)p$ for all $\sigma \in S_r$. If $\{v_1, \cdots, v_n\}$ is a basis of V, then $\Lambda^r(V)$ has a basis consisting of all $v_{i_1} \wedge v_{i_2} \wedge \cdots \wedge v_{i_r}$ where $1 \leqq i_1 < i_2 < \cdots < i_r \leqq n$ and

$$v_{i_1} \wedge \cdots \wedge v_{i_r} = \sum_{\sigma\in S_r} (\operatorname{sign}\sigma) v_{i_{\sigma(1)}} \otimes \cdots \otimes v_{i_{\sigma(r)}} .$$

For example, if $r = 2$ then $v_1 \wedge v_2 = v_1 \otimes v_2 - v_2 \otimes v_1$; $v_1 \wedge \cdots \wedge v_n$ spans $\Lambda^n(V)$. The action of $\alpha \otimes \cdots \otimes \alpha$ on $V \otimes \cdots \otimes V$ commutes with the action of S_r. Thus $\alpha \otimes \cdots \otimes \alpha$ (r times) preserves the subspace $\Lambda^r(V)$ of $V \otimes \cdots \otimes V$. The representation on that subspace, the so-called rth *alternation of* α, is denoted $\Lambda^r(\alpha)$. If $\alpha(y)\cdot v_i = \sum_j \alpha_{ij}(y)\cdot v_j$, then one checks that $\Lambda^r(\alpha y)\cdot v_{(i)} = \sum_{(j)} A_{(i)(j)}(y)\cdot v_{(j)}$ where $v_{(i)} = v_{i_1} \wedge \cdots \wedge v_{i_r}$ and

$$A_{(i)(j)} = \begin{vmatrix} \alpha_{i_1 j_1} & \alpha_{i_1 j_2} & \cdots & \alpha_{i_1 j_r} \\ \cdots & \cdots & \cdots & \cdots \\ \alpha_{i_r j_1} & \alpha_{i_r j_2} & \cdots & \alpha_{i_r j_r} \end{vmatrix}.$$

3. The characterization.

Our main result is:

3.1. THEOREM. *Let X be a compact topological group. Then every closed translation-invariant subalgebra of $C(X)$ is self adjoint, if and only if, every representation of degree 1 of X has finite image.*

3.2. COROLLARY. *Let X be a compact connected group. Then every closed translation-invariant subalgebra of $C(X)$ is self adjoint, if and only if, X is an inverse limit of semisimple Lie groups.*

3.3. COROLLARY. *Let X be a compact connected Lie group. Then every closed translation-invariant subalgebra of $C(X)$ is self adjoint, if and only if, X is semisimple.*

The proof of Theorem 3.1 is based on two simple lemmas.

3.4. LEMMA. *Let λ be a finite dimensional representation of a group Y. Then λ^* is equivalent to the representation*

$$y \to (\det\cdot\lambda(y))^{-1}\cdot\Lambda^n(\lambda(y))$$

where $n + 1$ is the degree of λ.

Proof. Let V be the representation space and choose a nonzero element $e \in \Lambda^{n+1}(V)$. If g is a linear transformation of V, then the determinant of g is given by $\Lambda^{n+1}(g)\cdot e = (\det\cdot g)e$. If $w \in \Lambda^n(V)$, we define a linear functional f_w on V by $w \wedge v = f_w(v)\cdot e$; $w \to f_w$ is an isomorphism of $\Lambda^n(V)$ onto the dual space V^* of V. Let ν be the representation $y \to (\det\cdot\lambda(y))^{-1}\cdot\Lambda^n(\lambda y)$. Then

$$\begin{aligned} \{f_{\nu(y)\cdot w}(\lambda(y)\cdot v)\}e &= \nu(y)w \wedge \lambda(y)v \\ &= (\det\cdot\lambda(y))^{-1}\cdot\Lambda^{n+1}(\lambda y)(w \wedge v) \\ &= w \wedge v = f_w(v)\cdot e\,. \end{aligned}$$

Thus the isomorphism $w \to f_w$ induces an equivalence of ν with λ^*.

3.5. LEMMA. *Let λ and μ be finite dimensional representations of a group Y. Then every matrix function of $\Lambda^r(\lambda)$ is a linear combination of products of r matrix functions of λ, and every matrix function of $\lambda \otimes \mu$ is a linear combination of products of matrix functions of λ with matrix functions of μ.*

Proof. Obvious from the definitions of $\Lambda^r(\lambda)$ and $\lambda \otimes \mu$.

3.6. *Proof of Theorem* 3.1. Suppose first that X has a representation π of degree 1 with infinite image. Then $\pi(X)$ is a circle group consisting of the unimodular complex numbers, and $\pi^n: x \to \pi(x)^n$ is another representation of degree 1. Let A be the subalgebra of $C(X)$ spanned by the E^{π^n}, $n \geqq 0$. Then A is a closed translation-invariant subalgebra of $C(X)$; $\pi \in A$ and $\bar{\pi} \notin A$, so A is not self adjoint.

Now suppose that every representation of degree 1 of X has finite image and let A be a closed translation-invariant subalgebra of $C(X)$. Then A is spanned by subspaces $E^\beta \subset C(X)$, say for β running over a subset $D_A \subset D$, and we must prove that A contains the adjoint of E^β for every $\beta \in D_A$. In orthonormal dual bases of the representation spaces, $\overline{\beta(x)} = {}^t\beta(x)^{-1}$ for every $x \in X$; thus E^{β^*} is the adjoint of E^β. Now we need only prove that $\beta \in D_A$ implies $\beta^* \in D_A$.

Let $\alpha \in D_A$ and define $\gamma(x) = \det \cdot \alpha(x)$. Then $\gamma = \Lambda_r(\alpha)$ where r is the degree of α, so $\gamma \in D_A$ by Lemma 3.5. $\gamma(X)$ is finite by hypothesis; thus $\gamma(X)$ is cyclic of some finite order n, consisting of the numbers $e^{2\pi ik/n}$. Now $\bar{\gamma} = \gamma^{-1} = \gamma^* = \gamma^{n-1} \in A$, so $\gamma^{-1} \in D_A$. $\Lambda^{r-1}(\alpha) \in D_A$ and D_A is closed under $\otimes$, by Lemma 3.5; thus $\gamma^{-1} \otimes \Lambda^{r-1}(\alpha) \in D_A$. Lemma 3.4 says $\alpha^* = \gamma^{-1} \otimes \Lambda^{r-1}(\alpha)$; thus $\alpha^* \in D_A$.

3.7. *Proof of corollaries.* If X is compact connected Lie group, then $X = X' \cdot T$ where X' is the derived group, T is the identity component of the center of X, and $T \cap X'$ is finite. T is a product of circle groups, X' is semisimple, and the following conditions are equivalent: (i) X is semisimple, (ii) $X = X'$, and (iii) $T = \{1\}$.

Let X be semisimple and let α be a representation of degree 1. $\alpha(X)$ is commutative, so the kernel of α contains $X' = X$. Thus $\alpha(X)$ is finite.

Let X be not semisimple. Then $S = T/T \cap X'$ is a torus of positive dimension, so there is a nontrivial representation β of S of degree 1. If $\pi: X \to X/X' = S$ is the projection, then $\alpha = \beta \cdot \pi$ is a representation of degree 1 of X and $\alpha(X) = \beta(S)$ is infinite.

Corollary 3.3 is proved.

Let X be a compact connected group. If π is a finite dimensional representation, we have the Lie group $\pi(X)$; $\{\pi(X)\}$ is an inverse system of compact Lie groups with X as inverse limit, and $\{\pi(X)\}$ is maximal for this property. A representation α is of degree 1 with infinite image, if and only if $\alpha(X)$ is a circle group. If each $\pi(X)$ is semisimple, then none is a circle group. If some $\pi(X)$, say $\alpha(X)$, is a circle group, then $\beta(X)$ is not semisimple for $\beta > \alpha$, so X cannot be an inverse limit of a subsystem of $\{\pi(X)\}$ which consists of semisimple groups. Corollary 3.2 follows.

4. **Enumeration of self adjoint function algebras.** The enumeration of closed translation-invariant self adjoint function algebras on a compact group is based on the following combination of the Peter-Weyl Theorem and the Stone-Weierstrass Theorem.

4.1. THEOREM. *Let X be a compact topological group, let D be the set of equivalence classes of the irreducible representations of X, and, given $\pi \in D$, let E^π denote the space of matrix functions of π. Then there are one-to-one correspondences between*

(i) *closed self adjoint translation-invariant subalgebras A of $C(X)$ which contain the constants,*

(ii) *subsets $F \subset D$ with the properties* (a) $1_X \in F$, (b) *if $\beta \in F$ then $\beta^* \in F$ and* (c) *if $\alpha, \beta \in F$ then F contains every irreducible summand of $\alpha \otimes \beta$, and*

(iii) *closed normal subgroups $\Gamma \subset X$.*

The correspondences are given by

$$A \sim \sum_{\beta \in F} E^\beta \quad \text{and} \quad A = C(X/\Gamma)$$

where $C(X/\Gamma) \subset C(X)$ by defining $f(x) = f(x\Gamma)$ for $f \in C(X/\Gamma)$.

4.2. COROLLARY. *If X is a compact group for which every representation of degree 1 has finite image, then the closed translation-invariant subalgebras of $C(X)$ which contain the constants, are just the algebras $C(Y)$ where Y is a quotient group of X.*

Corollary 4.2 is an immediate consequence of Theorems 3.1 and 4.1.

4.3. *Proof of Theorem* 4.1. Let $F \subset D$ and define A_F to be the subspace of $C(X)$ spanned by $\{E^\beta\}_{\beta \in F}$. Then A_F is a closed translation-invariant subspace of $C(X)$. Condition (iia) says that A_F contains the constants, condition (iib) says that A_F is self adjoint, and condition (iic) says that A_F is an algebra. Thus $A \sim \sum_{\beta \in F} E^\beta$ gives a one-to-one correspondence between the classes (i) and (ii).

Let $F \subset D$ satisfy conditions (iia, b, c), and define Γ_F to be the intersection of the kernels of the elements of F. Then Γ_F is a closed normal subgroup of X. Let $p: X \to X/\Gamma_F$ be the projection and define D_F to be the set of irreducible representations of X/Γ_F. Then $D_F \subset D$ under $\sigma \to \sigma \cdot p$. Now $A_F \subset C(X/\Gamma_F)$ under $f(x_F) = f(x)$, and $A_F = C(X/\Gamma_F)$ by the Stone-Weierstrass Theorem. It follows that $D_F = F$. Thus $\Gamma \leftrightarrow \{\beta \in D : \Gamma \subset \text{Kernel. } \beta\}$ is a one-to-one correspondence between the classes (iii) and (ii), and the composition of this with our other correspondence is given by $\Gamma \leftrightarrow C(X/\Gamma)$.

4.4. *Application to compact simply connected Lie groups.* Let X be a compact simply connected Lie group. Then $X = X_1 \times \cdots \times X_r$ where the X_i are compact simply connected simple Lie groups. By Corollary 3.3 and Theorem 4.1, the closed translation-invariant subalgebras of $C(X)$ containing the constants are enumerated by the enumeration of the closed normal subgroups $\Gamma \subset X$. X has center $Z = Z_1 \times \cdots \times Z_r$ where Z_i is the center of X_i, and the closed normal subgroups of X are just the groups $\Gamma = X' \cdot Z'$ where

(i) X' is a product of zero or more of the groups X_i and

(ii) Z' is a subgroup of Z.

Thus our enumeration problem is reduced to a knowledge of the centers of the groups X_i. This information is known; it is summarized in the following table for the convenience of the reader. Here $Z_{(m)}$ denotes the cyclic group of order m and the groups listed are all the compact simple simply connected Lie groups.

For example, the number of closed translation-invariant subalgebras

TABLE 1

group	description	center
$SU(n)$ $(n > 1)$	special unitary group in n complex variables	$Z_{(n)}$
Spin (n) $(2 < n \neq 4)$	two sheeted covering group of the rotation group in n real variables	$Z_{(2)}$ (n odd), $Z_{(2)} \times Z_{(2)}$ ($n = 4k$), $Z_{(4)}$ ($n = 4k + 2$)
$Sp(n)$	unitary group in n quaternion variables	$Z_{(2)}$
G_2	automorphism group of the Cayley algebra	{1}
F_4	elliptic group of the Cayley projective plane	{1}
E_6	. . . . .	$Z_{(3)}$
E_7	. . . . .	$Z_{(2)}$
E_8	. . . . .	{1}

of $C(X)$ properly containing the constants is one (just $C(X)$) for $X = G_2$, F_4 or E_8; it is two for $X = \mathrm{Spin}(2k+1)$, $Sp(n)$, E_6 or E_7; it is three for $X = \mathrm{Spin}(4k+2)$ and five for $X = \mathrm{Spin}(4k)$; for $X = SU(n)$ it is the number of divisors of n, counting both 1 and n.

Reference

1. J. A. Wolf, *Self adjoint function spaces on Riemannian symmetric manifolds*, Trans. Amer. Math. Soc. **113** (1964), 299-315.

University of California, Berkeley

Pacific Journal of Mathematics
Vol. 15, No. 3 November, 1965

David R. Arterburn and Robert James Whitley, *Projections in the space of bounded linear operators* 739
Robert McCallum Blumenthal, Joram Lindenstrauss and Robert Ralph Phelps, *Extreme operators into $C(K)$* 747
L. Carlitz, *A note on multiple exponential sums* 757
Joseph A. Cima, *A nonnormal Blaschke-quotient* 767
Paul Civin and Bertram Yood, *Lie and Jordan structures in Banach algebras* 775
Luther Elic Claborn, *Dedekind domains: Overrings and semi-prime elements* 799
Luther Elic Claborn, *Note generalizing a result of Samuel's* 805
George Bernard Dantzig, E. Eisenberg and Richard Warren Cottle, *Symmetric dual nonlinear programs* 809
Philip J. Davis, *Simple quadratures in the complex plane* 813
Edward Richard Fadell, *On a coincidence theorem of F. B. Fuller* 825
Delbert Ray Fulkerson and Oliver Gross, *Incidence matrices and interval graphs* 835
Larry Charles Grove, *Tensor products over H^*-algebras* 857
Deborah Tepper Haimo, *L^2 expansions in terms of generalized heat polynomials and of their Appell transforms* 865
I. Martin (Irving) Isaacs and Donald Steven Passman, *A characterization of groups in terms of the degrees of their characters* 877
Donald Gordon James, *Integral invariants for vectors over local fields* 905
Fred Krakowski, *A remark on the lemma of Gauss* 917
Marvin David Marcus and H. Minc, *A subdeterminant inequality* 921
Kevin Mor McCrimmon, *Norms and noncommutative Jordan algebras* 925
Donald Earl Myers, *Topologies for Laplace transform spaces* 957
Olav Njstad, *On some classes of nearly open sets* 961
Milton Philip Olson, *A characterization of conditional probability* 971
Barbara Osofsky, *A counter-example to a lemma of Skornjakov* 985
Sidney Charles Port, *Ratio limit theorems for Markov chains* 989
George A. Reid, *A generalisation of W^*-algebras* 1019
Robert Wells Ritchie, *Classes of recursive functions based on Ackermann's function* 1027
Thomas Lawrence Sherman, *Properties of solutions of n*th *order linear differential equations* 1045
Ernst Snapper, *Inflation and deflation for all dimensions* 1061
Kondagunta Sundaresan, *On the strict and uniform convexity of certain Banach spaces* 1083
Frank J. Wagner, *Maximal convex filters in a locally convex space* 1087
Joseph Albert Wolf, *Translation-invariant function algebras on compact groups* 1093