# Pacific Journal of Mathematics

## IN THIS ISSUE—




**Vol. 12, No. 4** **April, 1962**

Mathematical papers intended for publication in the *Pacific Journal of Mathematics* should be typewritten (double spaced), and the author should keep a complete copy. Manuscripts may be sent to any one of the four editors. All other communications to the editors should be addressed to the managing editor, L. J. Paige at the University of California, Los Angeles 24, California.

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.

# ON FUNDAMENTAL PROPERTIES OF A BANACH SPACE WITH A CONE

T. ANDÔ

**1. Introduction.** Normed vector lattices have been investigated from various angles (see [1] Chap. 15 and [7] Chap. 6). On the contrary, it seems that there remain several problems unsolved in the theory of general normed spaces with a cone since the pioneer works of Riesz and Krein, though recently Namioka [8], Schaefer [9] and others made many efforts in analysing and extending the results of Riesz and Krein. In this paper we shall discuss two among them. Let $E$ be a Banach space with a closed cone $K$ (for the terminologies see § 2);

(A) What condition on the dual $E^*$ is necessary and sufficient for that $E = K - K$?

(B) What condition on the dual is necessary and sufficient for the interpolation property of $E$?

Grosberg and Krein [3] dealt with (A) in a reversed form;

(A′) What condition on $E$ is necessary and sufficient for that $E^* = K^* - K^*$ where $K^*$ is the dual cone?

Schaefer ([9], Th. 1.6) obtained a complete answer to (A′) within a scope of locally convex spaces. A result of Riesz gives a half of an answer to (B), while Kreín [6] obtained a complete answer only under the assumption that the cone has an inner point.

The purpose of this paper is to give answers to both (A) and (B) in natural settings. Our starting assumptions consist of the completeness of $E$ and of the closedness of the cone $K$.

After several comments on order properties in § 2, Lemmas in § 3 present algebraic forms to both the property named normality by Krein [5] and that named $(BZ)$-property by Schaefer [9], supported by Banach's open mapping theorem. Then Theorem 1 will produce an answer to (A) via these Lemmas. § 4 is devoted to an answer to (B) under the condition that $E$ is an ordered Banach space. It should be remarked that our main theorems are also valid for $(F)$ spaces, that is, metrisable complete locally convex spaces.

**2. Definitions and consequences.** Let $E$[1] be a real normed space and let $K$ be a cone, that is, a subset of $E$ with the following properties:

(1) $K + K \subset K$,

(2) $\alpha K \subset K$ for all $\alpha \geqq 0$, and

Received February 8, 1962.

[1] Elements of $E$ are denoted by $x, y, a, \cdots, e$, and those of the dual $E^*$ by $f, g, h$. Scalars are denoted by Greek letters. $\theta$ is reserved for the zero element.

(3) $K \cap (-K) = \{\theta\}$. Then the natural partial ordering $\geqq$ is associated with the cone $K$, i.e. $a \geqq b$ in case $a - b \in K$. A subset of the form $\{x; a \leqq x \leqq b\}$ will be called an *interval*. The dual $E^*$ of $E$ is also partially ordered by the *dual cone* $K^* \stackrel{\text{def}}{=} \{f \in E^*; f(x) \geqq 0$ for all $x \in K\}$, though $K^*$ does not always satisfy the condition (3).[2]

The cone $K$ is said to *generate* $E$ or to be a *generating cone* in case every element in $E$ can be written as difference of two in $K$, that is, $E = K - K$. $E$ is said to have the *interpolation property* with respect to $\geqq$ in case $a, b \geqq c, d$ implies the existence of $x$ such that $a, b \geqq x \geqq c, d$. This property is equivalent to the following one named the *decomposition property*: whenever $a, b, x \in K$ and $a + b \geqq x$, there exist $c, d \in K$ such that $x = c + d$, $a \geqq c$ and $b \geqq d$. When for any pair $a, b \in E$ there exists the *supremum* $a \vee b$, $E$ is called a *vector lattice*. A vector lattice has the interpolation property and its cone is generating.

There are several notions connected with the so-called order topology. $E$ is said to be *(o)-complete* in case any upward directed subset with an upper bound (with respect to $\geqq$) has the supremum. When the directed subset in question is restricted to that consisting of countable members, $E$ is said to be *$\sigma$-(o)-complete*. As a less restrictive completeness, $E$ is said to be *quasi-(o)-complete* in case any sequence $\{a_i\}$, such that $\theta \leqq a_1 \leqq a_2 \leqq \cdots \leqq a$ and $a_{i+j} - a_i \leqq \varepsilon_i a$ with $\varepsilon_i \downarrow 0$, has the supremum. In many cases (o)-completeness can be derived from $\sigma$-(o)-completeness. It is clear that if $E$ with the generating cone is (o)-complete and has the interpolation property, it is a vector lattice (cf. [9] Th. 13.2).

Usually a complete normed vector lattice is called a *Banach lattice* in case its norm satisfies the following condition: $|a| \leqq |b|$ implies $\|a\| \leqq \|b\|$ where $|a| \stackrel{\text{def}}{=} a \vee (-a)$. The cone in a Banach lattice is obviously closed. In general, order topology is connected with the norm topology through the closedness of the cone in the following way: if $a_i \leqq a$ $i = 1, 2, \cdots$ and $\lim_{i\to\infty} a_i = x$ then $x \leqq a$, in particular, if $a_1 \leqq a_2 \leqq \cdots$ and $\lim_{i\to\infty} a_i = a$ then $a$ is the supremum of $\{a_i\}$. Thus a Banach lattice is quasi-(o)-complete. In this connection a quasi-(o)-complete Banach space with a closed generating cone will be called an *ordered Banach space*.

**3. Generating cone.** In this section $E$ is a Banach space with a closed cone $K$. First on the ground of Klee's theorem [4] it will be proved that the generating property is equivalent to the stronger one named strict *(BZ)-property* in Schaefer [9] ((3) in Lemma 1 below).

LEMMA 1. *The following conditions are mutually equivalent, where $\alpha$, $\beta$ and $\rho$ are positive constants and $U$ denotes the unit ball of $E$:*

[2] $K^*$ satisfies the condition (3), if and only if $K - K$ is dense in $E$.

(1) *K generates E,*
(2) $(K \cap U - K \cap U)^- \supset \alpha U$ *where* $(\cdot)^-$ *denotes the closure,*
(3) $(K \cap U - K \cap U) \supset \beta U$,
(4) *any* $x \in E$ *admits a decomposition* $x = a - b$ *such that* $a, b \in K$ *and* $\| a \|, \| b \| \leqq \rho \| x \|$.

*Proof.* (1) $\Rightarrow$ (2) follows from the second category of $E$, because $E = K - K = \bigcup_{n=1}^{\infty} n(K \cap U - K \cap U)^-$. In order to see (2) $\Rightarrow$ (3), let $V \overset{\text{def}}{=} K \cap U - K \cap U$ and let $F$ be the subspace generated by $V$. Then on the basis of completeness of $K$, Klee ([4] and [8] Th. 5.5) shows that $F$ is complete under the norm defined by $\| x \|_V = \inf \{| \lambda |; x \in \lambda V\}$. Then (2) shows that under the natural injection of $F$ into $E$ the closure of the image of the unit ball $V$ is a neighborhood of the origin in $E$. A modification of Banach's open mapping theorem (see [2] Chap. I, § 3) yields (3). (3) $\Rightarrow$ (4) and (4) $\Rightarrow$ (1) are trivial.

In the next place quasi-($o$)-completeness will be connected with the property named *normality* in Krein [5] ((3) in Lemma 2 below).

LEMMA 2. *The following conditions are mutually equivalent, where* $\rho$ *is a positive constant*:
(1) *E is quasi-(o)-complete,*
(2) *every interval is bounded in norm,*
(3) $a \leqq x \leqq b$ *implies* $\| x \| \leqq \rho \cdot \max (\| a \|, \| b \|)$,
(4) $(U + K) \cap (U - K) \subset \rho U$.

*Proof.* In order to see (1) $\Rightarrow$ (2), for each $a \in K$ let

$$V_a \overset{\text{def}}{=} \{x: -a \leqq x \leqq a\}$$

and let $F_a$ be the subspace generated by $V_a$. $F_a$ is complete under the norm defined by $\| x \|_a = \inf \{\lambda: -\lambda a \leqq x \leqq \lambda a\}$. In fact, if

$$\| x_{i+1} - x_i \|_a < 1/2^i \qquad (i = 1, 2, \cdots),$$

by the definition of the norm $\theta \leqq y_i \leqq a/2^{i-1}$ where $y_i = x_{i+1} - x_i + a/2^i$. Then quasi-($o$)-completeness implies the existence of the supremum $y$ of the sequence $\{\sum_{i=1}^n y_i\}_n$. Put $x = y + x_1 - a$, then $x - x_i$ is the supremum of the sequence $\{x_n - x_i - a/2^{n-1}\}_{n \geqq i}$ hence $x - x_i \leqq a/2^{i-1}$, and similarly $x - x_i \geqq - a/2^{i-1}$. This means that

$$\| x - x_i \|_a \leqq 1/2^{i-1} \qquad (i = 1, 2, \cdots),$$

hence $\lim_{i \to \infty} x_i = x$. Since $K$ is closed, as remarked in § 2, the natural injection of $F_a$ into $E$ is a closed linear mapping, hence on account of Banach's closed graph theorem (see [2] Chap. I, § 3) it is bounded, i.e. $V_a$ is bounded in $E$. Now every interval is readily proved to be bounded.

(3) follows from (2) via a standard argument (see [8] p. 32). (3) $\Rightarrow$ (4) is trivial. (4) $\Rightarrow$ (1) follows from the closedness of an interval and the completeness of $K$.

Before going into the first theorem, let us recall the definition of polar sets. The *polar set* $A^0$ of $A \subset E$ (resp. $\subset E^*$) is defined by $A^0 = \{f \in E^*; f(x) \leqq 1 \text{ for all } x \in A\}$ (resp. $= \{x \in E; f(x) \leqq 1 \text{ for all } f \in A\}$). For example, $U^0$ is the unit ball of $E^*$ and $K^0 = -K^*$. The bipolar theorem (see [2] Chap. IV, § 1) asserts that (1) $\Gamma(A, B)^0 = A^0 \cap B^0$ where $\Gamma(A, B)$ denotes the convex hull of $A \cup B$, and (2) if $A \ni \theta$ and $B \ni \theta$ are closed convex sets in $E$ (resp. weakly[3], i.e. $\sigma(E^*, E)$, closed convex sets in $E^*$) $(A \cap B)^0 = \Gamma(A^0, B^0)^{w-}$ (resp. $= \Gamma(A^0, B^0)^-$) where $(\cdot)^{w-}$ denotes the weak closure, and (3) if $A$ contains $\theta$ and is a closed convex set in $E$ (resp. weakly closed convex set in $E^*$), $A^{00} = A$. By the way, remark that the weak compactness of $U^0$ and the weak closedness of imply that both $U^0 + K^*$ and $U^0 - K^*$ are weakly closed.

THEOREM 1. (1) *$K$ generates $E$ if and only if $E^*$ is quasi-(o)-complete.*

(2) *$K^*$ generates $E^*$ if and only if $E$ is quasi-(o)-complete.*

*Proof.* (1) First remark the formula: $A + B \supset \Gamma(A, B) \supset \frac{1}{2}A + \frac{1}{2}B$ for any convex sets $A \ni \theta$ and $B \ni \theta$. Now the following chain of equivalences is valid, where $\alpha, \beta, \gamma$ and $\rho$ are positive constants:

$K$ generates $E$

| | | |
|---|---|---|
| $\Longleftrightarrow$ | $(U \cap K - U \cap K)^- \supset \alpha U$ | by Lemma 1 |
| $\Longleftrightarrow$ | $\Gamma(U \cap K, -U \cap K)^- \supset \beta U$ | by the above remark |
| $\Longleftrightarrow$ | $\Gamma(U^0, -K^*)^{w-} \cap \Gamma(U^0, K^*)^{w-} \subset \gamma U^0$ | by the bipolar theorem |
| $\Longleftrightarrow$ | $(U^0 - K^*)^{w-} \cap (U^0 + K^*)^{w-} \subset \rho U^0$ | by the above remark |
| $\Longleftrightarrow$ | $(U^0 - K^*) \cap (U^0 + K^*) \subset \rho U^0$ | by the weak closedness of $U^0 \pm K^*$ |

$\Longleftrightarrow$ $E^*$ is quasi-(o)-complete by Lemma 2. A proof of (2) is similar and is omitted.

The "only if" part of (1) is essentially known (see [8] p. 46), while (2) is a restatement of Grosberg-Krein's theorem [3] in terms of order properties[4].

If $E^*$ is quasi-(o)-complete, in view of Lemma 2 every interval of $E^*$ is bounded in norm and weakly closed, hence weakly compact. Therefore it is readily shown that all the three notions of completeness are the same thing on $E^*$.

**4. Interpolation property.** In this section $E$ is an ordered Banach space. Then Theorem 1 guarantees that $E^*$ is also an ordered Banach

[3] The weak topology always refers to the topology $\sigma(E^*, E)$.

[4] Grosberg-Krein's proof differs essentially from ours.

space. A result of Riesz can be stated as follows (see [8] Th. 6.1): if $E$ is an ordered Banach space with the interpolation property, the dual has the same property, hence by the remark at the end of § 3 it is a vector lattice. In this section the converse will be proved.

LEMMA 3. *Let $E$ be an ordered Banach space. Then the interpolation property can be derived from the following less restrictive one: for any $\varepsilon > 0$ and $a_i \geqq b_j$ in $E$ $(i, j = 1, 2, \cdots, n)$ there exist $x \in E$ and $y \in K$ such that $a_i \geqq x - y$ and $x \geqq b_i (i = 1, 2, \cdots, n)$ and $\| y \| \leqq \varepsilon$.*

*Proof.* Let $a, b \geqq c, d$. We can successively find $x_i \in E$ and $y_i \in K$ ($x_0$ and $y_0$ being disregarded) such that $a, b, x_{i-1} \geqq x_i - y_i$ and $x_i \geqq c, d, x_{i-1} - y_{i-1}$ and $\| y_i \| \leqq 1/2^i$. Then $-y_{i-1} \leqq x_i - x_{i-1} \leqq y_i$, hence by Lemma 2 $\| x_i - x_{i-1} \| \leqq \rho/2^i$ $(i = 1, 2, \cdots)$. The completeness of $E$ implies that $\lim_{i\to\infty} x_i = x$ exists. Since $\lim_{i\to\infty} y_i = \theta$ and $K$ is closed, we can conclude that $a, b \geqq x \geqq c, d$.

Before going into the second theorem, in order to simplify the notations, for each $A \subset E$ (resp. $\subset E^*$) define $A^\# \overset{\text{def}}{=} \{f \in K^*; f(x) \geqq 1$ for all $x \in A\}$ (resp. $\overset{\text{def}}{=} \{x \in K: f(x) \geqq 1$ for all $f \in A\}$). Since $K$ is closed convex, on account of the separation theorem (see [2] Chap. II § 3), for $a \in K$ $\{x; x \geqq a\} = a + K = \{a\}^{\#\#}$.

THEOREM 2. *An ordered Banach space $E$ has the interpolation property, if (and only if) the dual $E^*$ has the same property.*

*Proof.* Suppose that $E^*$ has the interpolation property. It suffices to prove the less restrictive form of the interpolation property for $E$ in Lemma 3. Let $a_i \geqq b_j$ $(i, j = 1, 2, \cdots, n)$. All $b_j$ may be assumed to be in $K$ because $K$ generates $E$. For any $\varepsilon > 0$ and $\gamma > 0$

$$A \overset{\text{def}}{=} (1 + \varepsilon)\Gamma(\{b_i\}^\#; i = 1, 2, \cdots, n)$$

is disjoint from

$$B \overset{\text{def}}{=} \Gamma((a_i - K)^0 \cap \gamma U^0; i = 1, 2, \cdots, n) .$$

Otherwise, since $E^*$ is an ordered Banach space with the interpolation property, in view of Riesz result stated above the second dual $E^{**}$ has the same property, therefore there exists $X \in E^{**}$ such that $a_i \geqq X \geqq b_i$ $(i = 1, 2, \cdots, n)$ where $E$ is imbedded into $E^{**}$ in the natural, linear-order preserving way, then $X(f) \leqq 1$ and $X(f) \geqq 1 + \varepsilon$ for $f \in A \cap B$, because, for example, $f$ can be represented as $f = \sum_{i=1}^n \alpha_i g_i$ such that $g_i \in \{b_i\}^\#$ and $\alpha_i \geqq 0$, $\sum_{i=1}^n \alpha_i = 1 + \varepsilon$, hence

$$X(f) = \sum_{i=1}^n \alpha_i X(g_i) \geqq \sum_{i=1}^n \alpha_i g_i(b_i) \geqq \sum_{i=1}^n \alpha_i = 1 + \varepsilon ,$$

This contradiction proves the expected disjointness. Next we shall prove that $A$ is weakly closed and $B$ is weakly compact. Take, for example, the former and suppose $n = 2$ for the simplicity sake. On account of Banach's theorem (see [2] Chap. IV, § 2) it suffices to prove that

$$\Gamma(\{b_i\}^{\#}; i = 1, 2) \cap \rho U^0$$

is weakly closed for all $\rho > 0$. Suppose that a net $\{\alpha_\lambda f_\lambda + (1 - \alpha_\lambda)g_\lambda\}_\lambda$ converges weakly to some $h$ in $E^*$ where $f_\lambda \in \{b_1\}^{\#}$, $g_\lambda \in \{b_2\}^{\#}$, $0 \leqq \alpha_\lambda \leqq 1$ and $\| \alpha_\lambda f_\lambda + (1 - \alpha_\lambda)g_\lambda \| \leqq \rho$. Since $E^*$ is quasi-(0)-complete, by Lemma 2 $\| \alpha_\lambda f_\lambda \|$ and $\| (1 - \alpha_\lambda)g_\lambda \|$ are uniformly bounded. We may assume that $\{\alpha_\lambda\}_\lambda$ converges to some $\alpha$. If $0 < \alpha < 1$, $\|f_\lambda\|$ and $\|g_\lambda\|$ are uniformly bounded, hence we may even assume that $\{f_\lambda\}_\lambda$ and $\{g_\lambda\}_\lambda$ converge weakly to some $f$ and to some $g$ respectively because of the weak compactness of $U^0$. Since both $\{b_1\}^{\#}$ and $\{b_2\}^{\#}$ are weakly closed, it follows that $h = \alpha f + (1 - \alpha)g$ is in $\Gamma(\{b_1\}^{\#}, \{b_2\}^{\#})$. If $\alpha = 1$, say, we may assume that $\{f_\lambda\}_\lambda$ converges weakly to some $f$ in $\{b_1\}^{\#}$, therefore $h \geqq f$, hence $h \in \{b_1\}^{\#}$ by the definition of $\{b_1\}^{\#}$. Thus the proof of the weak closedness is over.

Now since $A$ is convex, weakly closed and is disjoint from the convex weakly compact set $B$, by the separation theorem (see [2] Chap. II, § 3) there exists $c \in E$ such that $f(c) \geqq 1 > g(c)$ for all $f \in A$ and $g \in B$. From the remark preceding the theorem and by the bipolar theorem it follows that $(1 + \varepsilon)c \geqq b_i (i = 1, 2, \cdots, n)$ and $c \in \bigcap_1^n (a_i - K + 1/\gamma\, U)^-$. Therefore there exist $\{c_i\}_1^n$ such that $c + c_i \leqq a_i$ and $\| c_i \| \leqq 2/\gamma$ $(i = 1, 2, \cdots, n)$. Since the cone of $E$ is generating, by Lemma 1 each $c_i$ admits a decomposition $c_i = d_i - e_i$ with $d_i, e_i \in K$ such that

$$\| d_i \|, \| e_i \| \leqq \rho_1 \| c_i \|$$

where $\rho_1$ is a positive constant. Finally let $x = (1 + \varepsilon)c$ and

$$y = \varepsilon c + \sum_{i=1}^{n} e_i ,$$

then $x - y \leqq a_i$ and $x \geqq b_i\, (i = 1, 2, \cdots n)$ and, for some $\rho_2 > 0$,

$$\| y \| \leqq \varepsilon \| c \| + \sum_{i=1}^{n} \| e_i \| \leqq \varepsilon(\rho_2 \| a_1 \| + 4/\gamma) + 2\rho_1 n/\gamma.$$

Since $\varepsilon > 0$ and $\gamma > 0$ are arbitrary, the expected conclusion has been obtained.

## References

1. G. D. Birkoff, *Lattice theory*, 2nd Ed., New York, 1948.
2. N. Bourbaki, *Espaces vectoriels topologiques* I, II, Paris, 1950, 1953.
3. J. Grosberg and M. G. Krein, *Sur la décomposition des fonctionnelles en composantes positives*, Dokl. Akad. Nauk SSSR **25** (1939), 723-726.

4. V, L. Klee, *Boundedness and continuity of linear functionals*, Duke Math. J., **22** (1955), 263-269.
5. M. G. Krein, *Propriétés fondamentales des ensembles coniques normaux dans l'espace de Banach*, Dokl. Akad. Nauk SSSR **28** (1940), 13-17.
6. ———, *Sur la décomposition minimale d'une fonctionnelle linéaire en composantes positives*, Dokl. Akad. Nauk SSSR **28** (1940), 18-22.
7. H. Nakano, *Modulared semi-ordered linear spaces*, Tokyo, 1950.
8. I. Namioka, *Partially ordered linear topological spaces*, Memoirs of Amer. Math. Soc., **24** (1957).
9. H. H. Schaefer, *Halbgeordnete lokalkonvexe Vektorräume*, I, II, III, Math. Ann., **135** (1958), 115-141; **138** (1959) 254-286; **141** (1960), 113-142.

HOKKAIDO UNIVERSITY
AND
INDIANA UNIVERSITY

# A NOTE ON HYPONORMAL OPERATORS

STERLING K. BERBERIAN

The last exercise in reference [4] is a question to which I did not know the answer: does there exist a hyponormal ($TT^* \leqq T^*T$) completely continuous operator which is not normal? Recently Tsuyoshi Andô has answered this question in the negative, by proving that every hyponormal completely continuous operator is necessarily normal ([1]). The key to Andô's solution is a direct calculation with vectors, showing that a hyponormal operator $T$ satisfies the relation $\| T^n \| = \| T \|^n$ for every positive integer $n$ (for "subnormal" operators, this was observed by P. R. Halmos on page 196 of [6]). It then follows, from Gelfand's formula for spectral radius, that the spectrum of $T$ contains a scalar $\mu$ such that $|\mu| = \| T \|$ (see [9], Theorem 1.6.3.).

The purpose of the present note is to obtain this result from another direction, via the technique of approximate proper vectors ([3]); in this approach, the nonemptiness of the spectrum of a hyponormal operator $T$ is made to depend on the elementary case of a self-adjoint operator, and a simple calculation with proper vectors leads to a scalar $\mu$ in the spectrum of $T$ such that $|\mu| = \| T \|$. This is the Theorem below, and its Corollaries 1 and 2 are due also to Andô. In the remaining corollaries, we note several applications to completely continuous operators.

We consider operators (=continuous linear mappings) defined in a Hilbert space. As in [3], the spectrum of an operator $T$ is denoted $s(T)$, and the approximate point spectrum is $a(T)$. We note for future use that every boundary point of $s(T)$ belongs to $a(T)$; see, for example, ([4], hint to Exercise VIII. 3.4).

LEMMA 1. *Suppose $T$ is a hyponormal operator, with $\| T \| \leqq 1$, and let $\mathscr{M}$ be the set of all vectors which are fixed under the operator $TT^*$. Then,*

(i) *$\mathscr{M}$ is a closed linear subspace,*

(ii) *the vectors in $\mathscr{M}$ are fixed under $T^*T$,*

(iii) *$\mathscr{M}$ is invariant under $T$, and*

(iv) *the restriction of $T$ to $\mathscr{M}$ is an isometric operator in $\mathscr{M}$.*

*Proof.* Since $\mathscr{M} = \{x : TT^*x = x\}$ is the null space of $I - TT^*$, it is a closed linear subspace. The relation $TT^* \leqq T^*T \leqq I$ implies $0 \leqq I - T^*T \leqq I - TT^*$, and from this it is clear that the null space of $I - TT^*$ is contained in the null space of $I - T^*T$. That is, $TT^*x = x$

Received February 27, 1962.

implies $T^*Tx = x$. This proves (ii). (Alternatively, given $TT^*x = x$, one can calculate directly that $\| T^*Tx - x \|^2 \leqq 0$.) If $x \in \mathscr{M}$, that is if $TT^*x = x$, then the calculation $TT^*(Tx) = T(T^*Tx) = Tx$ shows that $Tx \in \mathscr{M}$; moreover, $\| Tx \|^2 = (T^*Tx \mid x) = \| x \|^2$.

LEMMA 2. *Every isometric operator has an approximate proper value of absolute value* 1.

*Proof*. Let $U$ be an isometric operator in a nonzero Hilbert space. Suppose first that the spectrum of $U$ contains 1; since $\| U \| = 1$, it follows that 1 is a boundary point of $s(U)$ (see [4], part (ix) of Exercise VII. 3. 12), hence 1 is an approximate proper value for $U$.

If the spectrum of $U$ does not contain 1, that is if $I - U$ is invertible, we may form the Cayley transform $A$ of $U$; thus,

$$A = i(I + U)(I - U)^{-1} = i(I - U)^{-1}(I + U) .$$

Using the hypothesis $U^*U = I$, let us show that $A$ is self-adjoint. Left-multiplying the relation $(I - U)A = i(I + U)$ by $U^*$, we have $(U^* - I)A = i(U^* + I)$, thus $(I - U)^*A = -i(I + U)^*$. Since $(I - U)^*$ is invertible, with inverse $[(I - U)^{-1}]^*$, we have

$$A = -i[(I - U)^{-1}]^*(I + U)^* = -i[(I + U)(I - U)^{-1}]^* = A^* .$$

It follows that the operators $A + iI$ and $A - iI$ are invertible, and solving the relation $(I - U)A = i(I + U)$ for $U$, we have

$$U = (A - iI)(A + iI)^{-1} = (A + iI)^{-1}(A - iI) .$$

Incidentally, since $U$ is the product of invertible operators, we conclude that $U$ is unitary.

Since $A$ is self-adjoint, we know from an elementary argument that the approximate point spectrum of $A$ is non empty ([7], Theorem 34.2). Let $\alpha \in a(A)$, and let $x_n$ be a sequence of unit vectors such that $\| Ax_n - \alpha x_n \| \to 0$. Define $\mu = (\alpha + i)^{-1}(\alpha - i)$; since $\alpha$ is real, $\mu$ has absolute value 1. It will suffice to show that $\mu$ is an approximate proper value for $U$; indeed, $\| (U - \mu I)x_n \| \to 0$ results from the calculation

$$\begin{aligned} U - \mu I &= (A + iI)^{-1}(A - iI) - (\alpha + i)^{-1}(\alpha - i)I \\ &= (\alpha + i)^{-1}(A + iI)^{-1}[(\alpha + i)(A - iI) - (\alpha - i)(A + iI)] \\ &= 2i(\alpha + i)^{-1}(A + iI)^{-1}(A - \alpha I) , \end{aligned}$$

the fact that $\| (A - \alpha I)x_n \| \to 0$, and the continuity of the operator $2i(\alpha + i)^{-1}(A + iI)^{-1}$.

Incidentally, if $U$ is an isometric operator such that the spectrum of $U$ excludes some complex number $\mu$ of absolute value 1, then $\mu^{-1}U$

is an isometric operator whose spectrum excludes 1. The proof of Lemma 2 then shows that $\mu^{-1}U$ is unitary, hence so is $U$. In other words: the spectrum of a nonnormal isometry must include the unit circle $|\mu| = 1$; indeed, Putnam has shown that the spectrum is the unit disc $|\mu| \leqq 1$ ([8], Corollary 1). The latter result is also an immediate consequence of ([5], Lemma 2.1), and the fact that the spectrum of any unilateral shift operator is the unit disc.

THEOREM. (Andô) *Every hyponormal operator $T$ has an approximate proper value $\mu$ such that $|\mu| = \|T\|$.*

*Proof.* We may assume $\|T\| = 1$ without loss of generality. Since $TT^* \geqq 0$ and $\|TT^*\| = 1$, we know that 1 is an approximate proper value for $TT^*$. Since the property of hyponormality is preserved under $*$-isomorphism, we may assume, after a change of Hilbert space, that 1 is a proper value for $TT^*$ ([3], Theorem 1). Form the nonzero closed linear subspace $\mathscr{M} = \{x : TT^*x = x\}$; according to Lemma 1, $\mathscr{M}$ is invariant under $T$, and the restriction of $T$ to $\mathscr{M}$ is an isometric operator $U$ in the Hilbert space $\mathscr{M}$. By Lemma 2, $U$ has an approximate proper value $\mu$ of absolute value 1. Let $x_n$ be any sequence of unit vectors in $\mathscr{M}$ such that $\|Ux_n - \mu x_n\| \to 0$. Since $Ux_n = Tx_n$, obviously $\mu$ is an approximate proper value for $T$, and $|\mu| = 1 = \|T\|$.

COROLLARY 1. *A generalized nilpotent hyponormal operator is necessarily zero.*

*Proof.* If $T$ is hyponormal, then $s(T)$ contains a scalar $\mu$ such that $|\mu| = \|T\|$. For every positive integer $n$, it follows that $s(T^n)$ contains $\mu^n$ (see [7], Theorem 33.1); then $\|T\|^n = |\mu|^n = |\mu^n| \leqq \|T^n\| \leqq \|T\|^n$, and so $\|T^n\| = \|T\|^n$. If moreover $T$ is a generalized nilpotent, that is if $\lim \|T^n\|^{1/n} = 0$, then $\|T\| = 0$.

COROLLARY 2. *If $T$ is a completely continuous hyponormal operator, then $T$ is normal.*

*Proof.* The proof to be given is essentially the same as Andô's. The proper subspaces of $T$ are mutually orthogonal, and reduce $T$ ([4], Exercise VII. 2.5). Let $\mathscr{M}$ be the smallest closed linear subspace which contains every proper subspace of $T$, and let $\mathscr{N} = \mathscr{M}^\perp$; clearly $\mathscr{N}$ reduces $T$, and the restriction $T/\mathscr{N}$ is a completely continuous hyponormal operator in $\mathscr{N}$ ([4], Exercise VI. 9.18). If the spectrum of $T/\mathscr{N}$ were different from $\{0\}$, it would have a nonzero boundary point $\mu$, hence $\mu$ would be a proper value for $T/\mathscr{N}$ (see [4], Theorem VIII. 3.2); this is impossible since $\mathscr{N}^\perp = \mathscr{M}$ already contains every proper vector for $T$.

We conclude from the Theorem that $T/\mathscr{N} = 0$, and this forces $\mathscr{N} = \{0\}$ (recall that $\mathscr{N}^{\perp}$ contains the null space of $T$). Thus, the proper subspaces of $T$ are a total family, hence $T$ is normal by ([4], Exercise VII. 2.5).

Suppose $T$ is a normal operator whose spectrum (a) has empty interior, and (b) does not separate the complex plane. Wermer has shown that the invariant subspaces of $T$ reduce $T$ ([10], Theorem 7). It is well known that the conditions (a) and (b) are fulfilled by the spectrum of any completely continuous operator. In particular: if $T$ is a completely continuous normal operator, then every invariant subspace of $T$ reduces $T$. A more elementary proof of this may be based on Corollary 2:

COROLLARY 3. *If $T$ is a completely continuous normal operator, and $\mathscr{N}$ is a closed linear subspace invariant under $T$, then $\mathscr{N}$ reduces $T$.*

*Proof.* Indeed, it suffices to assume that $T$ is hyponormal and $\mathscr{N}$ is an invariant subspace such that $T/\mathscr{N}$ is completely continuous. Since $T/\mathscr{N}$ is hyponormal ([4], Exercise VI. 9.10), it follows from Corollary 2 that $T/\mathscr{N}$ is normal, hence $\mathscr{N}$ reduces $T$ by ([4], Exercise VI. 9.9).

Quoting ([4], Theorem VII. 3.1), we have:

COROLLARY 4. *If $T$ is a hyponormal operator, then*

$$\| T \| = LUB\{|(Tx \mid x)| : \| x \| \leqq 1\} .$$

Incidentally, if $T$ is hyponormal, it is clear from Corollary 4 that $\| T^* \| = LUB\{|(T^*x \mid x)| : \| x \| \leqq 1\}$.

COROLLARY 5. *If the completely continuous operator $T$ is semi-normal in the sense of* [8], *then $T$ is normal.*

*Proof.* The definition of semi-normality is that either $TT^* \leqq T^*T$ or $TT^* \geqq T^*T$, in other words, either $T$ or $T^*$ is hyponormal; since both are completely continuous (see [4], Exercise VIII. 1.6), our assertion follows from Corollary 2.

Let us say that an operator $T$ is *nearly normal* in case $T$ commutes with $T^*T$. The structure of nearly normal operators has been determined by Brown, and it is a consequence of his results that a completely continuous nearly normal operator is in fact normal (see the concluding remarks in [5]). This may also be proved as follows. An elementary calculation with square roots shows that a nearly normal operator is hyponormal (see [2], proof of Corollary 1 of Theorem 8); assuming also complete continuity and citing Corollary 2, we have:

COROLLARY 6. *If $T$ is a completely continuous nearly normal operator, then $T$ is normal.*

Finally,

COROLLARY 7. *If $S = T + \lambda I$, where $T$ is a completely continuous operator, and if $S$ is hyponormal, then $S$ is normal.*

*Proof.* Since $S$ is hyponormal, so is $T$ ([4], hint to Exercise VII. 1.6), hence $T$ is normal by Corollary 2; therefore $S$ is normal. So to speak, the $C^*$-algebra of all operators of the from $T + \lambda I$, with $T$ completely continuous, is of "finite class".

We close with an elementary remark about the adjoint of a hyponormal operator: if $T$ is hyponormal, then $s(T^*) = a(T^*)$. For, suppose $\lambda$ does not belong to $a(T^*)$, and let $\mu = \lambda^*$. Then, $(T - \mu I)^* = T^* - \lambda I$ is bounded below ([4], Exercise VII. 3.8), and since $T - \mu I$ is also hyponormal, the relation $(T - \mu I)(T - \mu I)^* \leqq (T - \mu I)^*(T - \mu I)$ shows that $T - \mu I$ is also bounded below. Then $T - \mu I$ is invertible ([4], Exercise VI. 8.11), hence so is $T^* - \lambda I$, thus $\lambda$ does not belong to $s(T^*)$.

## REFERENCES

1. T. Andô, Forthcoming paper in Proc. Amer. Math. Soc.
2. S. K. Berberian, *Note on a theorem of Fuglede and Putnam*, Proc. Amer. Math. Soc., **10** (1959), 175-182.
3. ———, *Approximate proper vectors*, Proc. Amer. Math. Soc., **13** (1962), 111-114.
4. ———, *Introduction to Hilbert space*, Oxford University Press, New York, 1961.
5. A. Brown, *On a class of operators*, Proc. Amer. Math. Soc., **4** (1953), 723-728.
6. P. R. Halmos, *Commutators of operators, II.*, Amer. J. Math., **76** (1954), 191-198.
7. ———, *Introduction to Hilbert space and the theory of spectral multiplicity*, Chelsea, New York, 1951.
8. C. R. Putnam, *On semi-normal operators*, Pacific J. Math., **7** (1957), 1649-1652.
9. C. E. Rickart, *General theory of Banach algbras*, D. van Nostrand, New York, 1960.
10. J. Wermer, *On invariant subspaces of normal operators*, Proc. Amer. Math. Soc., **3** (1952), 270-277.

STATE UNIVERSITY OF IOWA

# ANALYTIC FUNCTIONS WITH VALUES IN A FRECHET SPACE

ERRETT BISHOP

We wish to extend certain results in the theory of analytic functions of several complex variables to the case of analytic functions with values in a Frechet space $F$. To do this, we prove (Theorem 1 below) that such a function $\varphi$ has an expansion of the form

$$(*)\qquad \varphi = \sum_{n=1}^{\infty} P_n \circ \varphi\ ,$$

where $\{P_n\}$ is a sequence of continuous mutually annihilating projections on $F$ whose ranges are all one-dimensional subspaces of $F$. This representation reduces the study of $\varphi$, for many purposes, to the study of the functions $P_n \circ \varphi$, which are essentially scalar-valued analytic functions. We actually prove the stronger (and more useful) result that if $\{\varphi_k\}$ is a sequence of analytic functions with values in $F$ then a single sequence $\{P_n\}$ can be found to give an expansion $(*)$ for every $\varphi_k$. Expansions of vector-valued functions of a different type have been considered by Grothendick [6].

Theorem 1 is applied to generalize Theorem B of H. Cartan [3]. We consider a coherent analytic sheaf $S$ on a Stein manifold $M$ and introduce the notion of the *vectorization* $S_F$ of $S$ (relative to a given Frechet space $F$).

If 0 denotes the sheaf of locally-defined analytic functions and $0_F$ denotes the sheaf of locally-defined analytic functions with values in $F$, then $S_F$ is defined to be the tensor product $S \otimes 0_F$ of the 0-modules $S$ and $0_F$. For the important case of a coherent analytic subsheaf $S$ of the sheaf $0^k$ of locally-defined $k$-tuples of analytic functions, $S_F$ turns out to be canonically isomorphic to the sheaf $S'_F$ determined by assigning to each open set $U$ the module of all $k$-tuples $(f_1, \cdots, f_k)$ of analytic functions from $U$ to $F$ which have the property that for each $u$ in $F^*$ the $k$-tuple $(u \circ f_1, \cdots, u \circ f_k)$ is a cross-section of $S$ over $U$. For instance, if $S$ is the sheaf of all locally-defined analytic functions which vanish on a given analytic set $A$ then it is evident that $S'_F$ is the sheaf of all locally-defined analytic functions with values in $F$ which vanish on $A$.

One of the main results, an extension of Theorem B of [3], will be that the cohomology groups $H^N(M, S_F)$ vanish in all dimensions $N \geqq 1$, where $S_F$ is the vectorization of a coherent analytic sheaf $S$ on a Stein manifold $M$. Using this theorem and the isomorphism of $S_F$ to the sheaf $S'_F$ defined above one could show, for instance, that the usual

Received January 15, 1962. This work was partially supported by the Sloan Foundation.

sheaf—theoretic solutions to Cousin's problems carry over to the case of analytic functions with values in a Frechet space. Special cases were treated by totally different methods in [2], but the techniques of that paper seem to be inadequate to obtain general results.

The proofs are all Banach-space theoretic. That is, only Banach space theory is necessary to obtain the above extension of Theorem B and to prove the necessary facts about vectorizations. We begin with a theorem which is given without proof on p. 278 of Banach [1], who attributes it to H. Auerbach. A proof can be found in Taylor [7]. Since complex Banach spaces are considered here, we give the proof.

THEOREM (Auerbach). *An $n$-dimensional Banach space $B$ has a basis of unit vectors whose dual basis also consists of unit vectors.*

*Proof.* Choose a basis $(b^1, \cdots, b^n)$ of $B$ and for any $x$ in $B$ let $(x_1, \cdots, x_n)$ be the coordinates of $x$ relative to the chosen basis. Let $T$ be the set of all $n$-tuples $(x^1, \cdots, x^n)$ of unit vectors in $B$. For each $(x^1, \cdots, x^n)$ in $T$ let $\alpha(x^1, \cdots, x^n)$ be the absolute value of the determinant $\det(x^i_j)$. Thus $\alpha$ is a continuous function on the compact space $T$. Now $\alpha(x^1, \cdots, x^n) \neq 0$ if and only if $(x^1, \cdots, x^n)$ is a basis. Thus $\alpha$ attains its maximum for $T$ at some point $(y^1, \cdots, y^n)$ in $T$ which is a basis of unit vectors. Let $(u^1, \cdots, u^n)$ be the dual basis in $B^*$. Now $\|u^i\| \geqq 1$ because $\langle y^i, u^i\rangle = 1$. Assume $\|u^i\| > 1$ for some $i$. Thus there exists $t$ in $B$ with $\|t\| = 1$ and $\langle t, u^i\rangle = c > 1$. Thus $\langle t - cy^i, u^i\rangle = 0$, so that $t - cy^i$ is a linear combination of the vectors of the basis $(y^1, \cdots, y^n)$ other than $y^i$. If we let $(z^1, \cdots, z^n)$ be the basis $(y^1, \cdots, y^n)$ with $y^i$ replaced by $t$ it follows that $\alpha(z^1, \cdots, z^n) = c\alpha(y^1, \cdots, y^n)$. Since the basis $(z^1, \cdots, z^n)$ consists of unit vectors this contradicts the choice of $(y^1, \cdots, y^n)$. Thus $\|u^i\| = 1$ for all $i$, and the theorem is proved.

COROLLARY. *If $B_0$ is a finite-dimensional subspace of dimension $n$ of a Banach space $B$ there exist $n$ mutually annihilating projections (idempotent continuous linear operators) on $B$, each of norm 1, whose ranges are one-dimensional subspaces of $B_0$ and whose sum is a projection of $B$ onto $B_0$ of norm at most $n$.*

*Proof.* Let $(y^1, \cdots, y^n)$ be a basis of unit vectors of $B_0$ such that the dual basis $(u^1, \cdots, u^n)$ of $B_0^*$ also consists of unit vectors. Let $v^i$ be an extension of $u^i$ to a linear functional on $B$ of norm 1. The operators $P_1, \cdots, P_n$ on $B$ defined by

$$P_i x = \langle x, v^i\rangle y^i$$

are the desired projections.

We recall that a Frechet space is a locally convex topological linear space $F$ which admits a countable family $\{\|\ \|_k\}$ of continuous semi-norms such that a basis for the neighborhoods of 0 in $F$ is given by the sets

$$\{x \in F : \| x \|_k < 1\} .$$

If $\|\ \|$ is any continuous semi-norm on $F$ it follows that for some $k$ $\| x \| \leqq \| x \|_k$ for all $x$ in $F$. If necessary it may be assumed that $\{\|\ \|_k\}$ is a monotonely nondecreasing sequence of semi-norms, in which case we shall call it a *defining sequence* of semi-norms for $F$.

LEMMA 1. *Let $F$ be a Frechet space with a defining sequence $\{\|\ \|_k\}$ of semi-norms. Let $\{a_n\}$ be a sequence of vectors in $F$, $\{\delta_k\}$ a sequence of nonnegative real numbers, and $\{k_j\}$ a strictly increasing sequence of positive integers. Then there exists a sequence $\{P_n\}$ of mutually annihilating continuous projections on $F$, whose ranges are subspaces of $F$ of dimensions at most* 1, *and a sequence $\{\varepsilon_k\}$, with $0 < \varepsilon_k < \delta_k$ for all $k$, with the following properties. For each positive integer $j$ the operator*

$$Q_j = \sum_{n=1}^{k_j} P_n$$

*is a projection on the subspace $B_j$ of $F$ spanned by the vectors $a_1, \cdots, a_{k_j}$. For each positive integer $n$ the sum*

$$\| a \|_0 = \sum_{k=1}^{\infty} \varepsilon_k \| a \|_k$$

*is finite for $a = a_n$. For each positive integer $j$ and all $n \leqq k_j$ we have $\| P_n \|_0 \leqq (1 + k_1^2) \cdots (1 + k_j^2)$, where*

$$\| P_n \|_0 = \sup \{\| P_n b \|_0 : b \in F, \quad \| b \|_0 = 1\} .$$

*Proof.* We may assume the $\delta_k$ to be so small that $\sum_{k=1}^{\infty} \delta_k \| a_n \|_k < \infty$ for all $n$. By induction we construct a sequence $\{P_n\}$ of mutually annihilating continuous projections, a sequence $\{\varepsilon_k\}$ of positive real numbers, and an increasing sequence $\{N_j\}$ of positive integers such that

(a) $0 < \varepsilon_k < \delta_k$,

(b) For each $j$ the operator $Q_j$ is a projection onto $B_j$,

(c) $\| P_n \|^j < (1 + k_1^2) \cdots (1 + k_i^2)$ for $1 \leqq n \leqq k_i$ and all $i \leqq j$.

We explain what is meant by (c). First of all, $\|\ \|^j$ is the continuous semi-norm on $F$ defined by

$$\| b \|^j = \sum_{k=1}^{N_j} \varepsilon_k \| b \|_k .$$

Secondly, $\| P_n \|^j$ is defined by

$$\| P_n \|^j = \sup \{\| P_n b \|^j : \| b \|^j = 1\} .$$

Assuming that $P_1, \cdots, P_{k_j}$ and $N_1 \cdots, N_j$, and $\varepsilon_1, \cdots, \varepsilon_{N_j}$ have been found with the relevant properties, we show how to continue to the next stage $j+1$. First choose $N_{j+1} > N_j$ so large that $\| \quad \|_{N_{j+1}}$ is a norm (and not merely a semi-norm) on $B_{j+1}$. Choose then $\varepsilon_i$, $N_j < i \leqq N_{j+1}$, so small that $0 < \varepsilon_i < \delta_i$ and $\| P_n \|^{j+1} < (1+k_1^2) \cdots (1+k_i^2)$ for $n \leqq k_j$ and all $i \leqq j$. To see that this can be done, notice that because $\| \quad \|_{N_j}$ is a norm on $B_j$ there exists $r > 0$ so that $r \| a \|^j > \| a \|_m$ for all $a$ in $B_j$ and all $m \leqq N_{j+1}$. Thus

$$\| P_n \|^{j+1} \leqq \sup \{\| P_n b \|^{j+1} : \| b \|^j = 1\} \leqq (1 + \sum_{m=N_j+1}^{N_{j+1}} \varepsilon_m) \| P_n \|^j .$$

Now use (c).

Now let $Q_j'$ be the restriction of $Q_j$ to $B_{j+1}$ and let $I_{j+1}$ be the identity operator on $B_{j+1}$. Thus $I_{j+1} - Q_j'$ is a projection of $B_{j+1}$ onto a subspace $S_{j+1}$. Clearly $B_j$ and $S_{j+1}$ are complementary subspaces of $B_{j+1}$, so that $\dim S_{j+1} \leqq k_{j+1} - k_j$. By the above corollary there exists a projection $E_{j+1}$ with $\| E_{j+1} \|^{j+1} \leqq k_{j+1}$ of $F$ onto $B_{j+1}$. Also by the above corollary there exist mutually annihilating projections $R_n$, $k_j < n \leqq k_{j+1}$, of $S_{j+1}$ onto subspaces of dimensions at most 1 such that $\| R_n \|^{j+1} \leqq 1$ for all $n$ and such that $\Sigma R_n$ is the identity projection of $S_{j+1}$ onto itself. For $k_j < n \leqq k_{j+1}$ we define

$$P_n = R_n(I_{j+1} - Q_j')E_{j+1} .$$

Thus the $P_n$ are mutually annihilating projections for $1 \leqq n \leqq k_{j+1}$. Also $Q_{j+1}$ is a projection onto $B_{j+1}$. Finally for $k_j < n \leqq k_{j+1}$ we have

$$\begin{aligned} \| P_n \|^{j+1} &\leqq \| R_n \|^{j+1} \| I_{j+1} - Q_j' \|^{j+1} \| E_{j+1} \|^{j+1} \\ &\leqq (1 + \sum_{n=1}^{k_j} \| P_n \|^{j+1}) k_{j+1} \\ &< [1 + k_j(1+k_1^2) \cdots (1+k_j^2)] k_{j+1} \\ &\leqq (1+k_1^2) \cdots (1+k_{j+1}^2) . \end{aligned}$$

The same is true for $n \leqq k_j$, by the above construction. Thus the construction has been continued another step. By induction it follows that sequences $\{P_n\}$, $\{N_j\}$, and $\{\varepsilon_k\}$ can be chosen satisfying properties (a), (b), and (c). It is immediate that the sequences $\{P_n\}$ and $\{\varepsilon_k\}$ satisfy the requirements of the lemma.

LEMMA 2. *Let $\{a_n\}$ be a sequence of elements of a Frechet space $F$, $\{\| \quad \|_k\}$ a defining sequence of semi-norms on $F$, and $\{\delta_k\}$ a sequence of positive real numbers. Then there exist a sequence $\{\varepsilon_k\}$ of positive real numbers and a sequence $\{P_n\}$ of mutually annihilating projections on $F$ whose ranges are subspaces of $F$ of dimensions at most 1 having the following properties.*

(i) *$0 < \varepsilon_k < \delta_k$ for all $k$,*

(ii) *For $a = a_n$ the norm $\| a \|_0 = \sum_{k=1}^{\infty} \varepsilon_k \| a \|_k$ is finite for all $n$,*

(iii) *$R_m a_n = a_n$ for all positive integers $m$ and $n$ with $m \geqq 2n$, where $R_m = \sum_{j=1}^{m} P_j$,*

(vi) *For all $t > 1$ and $\varepsilon > 0$ the sum $\sum_{n=1}^{\infty} \| P_n \|_0 t^{-n^{\varepsilon}}$ converges, where $\| P_n \|_0$ is defined as above.*

*Proof.* Define the sequence $\{k_j\}$ by $k_j = 2^j$. Choose the sequences $\{P_n\}$ and $\{\varepsilon_k\}$ as in lemma 1. Clearly (i) and (ii) are satisfied. Now for each positive integer $n$ there is a positive integer $j$ with $2^{j-1} \leqq n < 2^j$. It follows that $a_n \in B_j$. Thus $R_{2^j} a_n = Q_j a_n = a_n$, so that $R_m a_n = a_n$ for all $m \geqq 2^j$ and therefore for all $m \geqq 2n$. This proves (iii).

Now for each $n$ choose $j$ with $2^{j-1} \leqq n < 2^j$. Thus

$$\begin{aligned} \| P_n \|_0 &\leqq (1 + k_j^2)^j = (1 + 2^{2j})^j \\ &\leqq (5n^2)^j \leqq (5n^2)^{\alpha} , \end{aligned}$$

where $\alpha = 1 + \log_2 n$. From this it follows from elementary calculus that (iv) holds, thereby proving the lemma.

LEMMA 3. *Let*

$$\sum_{n_1 \geqq 0, \cdots, n_\alpha \geqq 0} a_i(n_1, \cdots, n_\alpha) z_1^{n_1} \cdots z_\alpha^{n_\alpha}$$

*where $\alpha = \alpha_i$ and $1 \leqq i < \infty$, be a sequence of formal power series with coefficients in a Frechet space $F$. Let $\{\delta_k\}$ be a sequence of positive real numbers. Then there exists a sequence $\{\varepsilon_k\}$ with $0 < \varepsilon_k < \delta_k$ for all $k$ and a sequence $\{P_n\}$ of mutually annihilating continuous projections of $F$ onto subspaces of dimensions at most 1 such that*

(a) *$R_m a_i(n_1, \cdots, n_\alpha) = a_i(n_1, \cdots, n_\alpha)$ whenever $m \geqq 2^{i+2} n^{\alpha}$, where $\alpha = \alpha_i$, $n = n_1 + \cdots + n_\alpha$, and $R_m = \sum_{j=1}^{m} P_j$,*

(b) *$P_m a_i(n_1, \cdots, n_\alpha) = 0$ whenever $m > 2^{i+2} n^{\alpha}$,*

(c) *$\sum_{n=1}^{\infty} \| P_n \|_0 t^{-n^{\varepsilon}} < \infty$ for all $t > 1$ and $\varepsilon > 0$, where $\| \; \|_0$ is defined as above.*

*Proof.* For each $i$ order the coefficients $a_i(n_1, \cdots, n_\alpha)$ into a sequence $\{\alpha_i^k\}_{k=1}^{\infty}$ according to the size of $n$. We now define a sequence $\{a_k\}$ of elements of $F$ which is an ordering of the totality of the $a_i(n_1, \cdots, n_\alpha)$. For $k$ given let $2^i$ be the largest power of 2 dividing $k$ and let $j = 1/2(k2^{-i} + 1)$. Let $a_k = \alpha_i^j$. Now choose the sequences $\{\varepsilon_k\}$ and $\{P_n\}$ as in Lemma 2. Clearly (c) holds. Since (b) is a consequence of (a) we need only check (a). To this end consider a fixed $a_i(n_1, \cdots, n_\alpha)$. Now there exists $j \leqq n^{\alpha}$ with $a_i(n_1, \cdots, n_\alpha) = \alpha_i^j$. In turn $\alpha_i^j = a_k$ for some $k \leqq 2^{i+1} n^{\alpha}$. By (iii) of Lemma 2 it follows that $R_m a_k = a_k$ for $m \geqq 2k$ and therefore for $m \geqq 2^{i+2} n^{\alpha}$, as was to be proved.

We are now prepared to prove a series representation for analytic functions with values in a Frechet space which will be the principal tool in subsequent proofs.

THEOREM 1. *Let $F$ be a Frechet space and let $\{M_i\}$ be a sequence of complex analytic manifolds. For each $i$ let $\varphi_i$ be an analytic function on $M_i$ with values in $F$. Then there exists a sequence of vectors $\{b_n\}$ in $F$ and a sequence $\{P_n\}$ of continuous mutually annihilating projections of $F$ onto one-dimensional subspaces having the following properties. For each $i$ the series $\sum_{n=1}^{\infty} P_n \circ \varphi_i$ converges to $\varphi_i$ on $M_i$. For each $n$ we have $P_n b_n = b_n$, so that $P_n \circ \varphi_i = \varphi_i^n b_n$, for some analytic function $\varphi_i^n$ on $M_i$. For each $i$ the series $\sum_{n=1}^{\infty} \varphi_i^n$ converges absolutely and uniformly on all compact subsets of $M_i$. For each continuous semi-norm $\|\ \ \|$ on $F$ the sequence $\{\| b_n \|\}$ is bounded.*

*Proof.* For each $i$ let $\dim M_i = \alpha = \alpha_i$, so that $M_i$ is coverable by a countable family of analytic homeomorphs $\Gamma$ of the unit polycylinder

$$U^\alpha = \{z = (z_1, \cdots, z_\alpha) : | z_j | < 1, 1 \leqq j \leqq \alpha\} .$$

Thus in the proof of the theorem we may replace the sequence $\{M_i\}$ by the totality of all such $\Gamma$. There is therefore no loss of generality in assuming that each $M_i$ is a polycylinder $U^\alpha$ of dimension $\alpha = \alpha_i$. Let $\{\|\ \ \|_k\}$ be a defining sequence of semi-norms on $F$. Now for each $i$ the analytic function $\varphi_i$ has a power series expansion

$$\varphi_i = \sum_{n_1 \geqq 0, \cdots, n_\alpha \geqq 0} a_i(n_1, \cdots, n_\alpha) z_1^{n_1} \cdots z_\alpha^{n_\alpha}$$

on the polycylinder $M_i = U^\alpha$. This expansion converges absolutely and uniformly on each compact subset of $M_i$ in each semi-norm $\|\ \ \|_k$. By the diagonal process there therefore exist constants $\delta_k > 0$ such that the power series for each $\varphi_i$ converges absolutely and uniformly on each compact subset of $M_i$ in the norm $\sum_{k=1}^{\infty} \delta_k \|\ \ \|_k$, so that in particular this norm is finite for each coefficient $a_i(n_1, \cdots, n_\alpha)$. Now choose the sequences $\{\varepsilon_k\}$ and $\{P_n\}$ as in Lemma 3 relative to the power series expansions of the $\varphi_i$ and to the $\delta_k$ just obtained. Thus the power series for $\varphi_i$ converges absolutely and uniformly on compact subsets of $M_i$ in the norm $\|\ \ \|_0$ defined above. If some of the projections $P_n$ are zero, these may be omitted from the sequence. Thus for each $n$ there is a vector $b_n$ in $F$ with $\| b_n \|_0 = 1$ spanning the range of $P_n$. To show that the sequences $\{P_n\}$ and $\{b_n\}$ have the desired properties, consider a fixed compact subset $T$ of a fixed $M_i$. For each $n$ write

$$\gamma_n = \sum_{n_1 + \cdots + n_\alpha = n} \max \{\| a_i(n_1, \cdots, n_\alpha) z_1^{n_1} \cdots z_\alpha^{n_\alpha} \|_0 : z \in T\} .$$

By the usual convergence criteria we see that there exist $r > 1$ and $c > 0$ such that $r^n\gamma_n < c$ for all $n$.

If $j$ is any positive integer let $k$ be the largest integer such that $2^{i+2}k^\alpha < j$. Thus for each $z$ in $T$ we have

$$\begin{aligned}&\| P_j\varphi_i(z) \|_0 \\ &= \left\| P_j \sum_{n_1+\cdots+n_\alpha \geqq k} a_i(n_1, \cdots, n_\alpha) z_1^{n_1} \cdots z_\alpha^{n_\alpha} \right\|_0 \\ &\leqq \| P_j \|_0 \sum_{n \geqq k} \gamma_n \leqq c \| P_j \|_0 \sum_{n \geqq k} r^{-n} \\ &= c(1 - r^{-1})^{-1} \| P_j \|_0 r^{-k} .\end{aligned}$$

Thus

$$\begin{aligned}\varDelta &= \max \left\{ \sum_{j=1}^{\infty} \| P_j\varphi_i(z) \|_0 : z \in T \right\} \\ &\leqq c(1 - r^{-1})^{-1} \sum_{j=1}^{\infty} r^{-k} \| P_j \|_0 .\end{aligned}$$

Now by the definition of $k$ we see that $k$ is the integral part of $(j2^{-i-2})^{1/\alpha}$, so that $k \geqq j^{1/2\alpha}$ for all $j$ sufficiently large. Thus $\varDelta$ is finite if the sum $\sum_{j=1}^{\infty} r^{-j^\varepsilon} \| P_j \|_0$ converges, where $\varepsilon = (2\alpha)^{-1}$. By the choice of the sequence $\{P_j\}$ this series converges so that $\varDelta$ is finite. Now since $\| b_n \|_0 = 1$,

$$\max \{ | \varphi_i^n(z) | : z \in T \} = \max \{ \| P_n\varphi_i(z) \|_0 : z \in T \} .$$

Therefore the series $\sum_{n=1}^{\infty} \varphi_i^n(z)$ converges absolutely and uniformly on $T$. If $\| \quad \|$ is a continuous semi-norm on $F$ then $\| \quad \| \leqq K \| \quad \|_0$ for some $K > 0$, so that $\{\| b_n \|\}$ is bounded by $K$. Finally, we must show that $\sum_{n=1}^{\infty} P_n \circ \varphi_i$ actually converges to $\varphi_i$ (and not to something else). To see this, note by (a) and (b) of Lemma 3 that $R_m \circ \varphi_i$ and $\varphi_i$ have power series expansions in the coordinates $z_1, \cdots, z_\alpha$ which agree up to terms of total order $n$, whenever $m \geqq 2^{i+2}n^\alpha$. This completes the proof of Theorem 1.

Before giving the definition of the vectorization of an analytic sheaf, we indicate the terminology to be used, following Godement [5]. A presheaf $S$ on a topological space $X$ assigns to each open $U \subset X$ a set $S(U)$ and to each open set $V \subset U \subset X$ a map $r_{VU} : S(U) \to S(V)$ satisfying $r_{WV} \circ r_{VU} = r_{WU}$ for $W \subset V \subset U$. In particular the same terminology will be used if $S$ is a sheaf, that is, a presheaf satisfying axioms (F1) and (F2) on page 109 of [5]. To any presheaf $S$ is canonically associated a sheaf $S'$, and each element $f$ in $S(U)$ gives rise to a unique element in $S'(U)$ which will also be denoted by $f$. If $X$ is a complex analytic manifold a sheaf $S$ on $X$ is called analytic if it is a module over the sheaf 0 of locally defined analytic functions, that is, if for each $U$ the set $S(U)$ is an $0(U)$-module, and if the usual commutation relations between module multiplication and the restriction maps $S(U) \to S(V)$ and $0(U) \to 0(V)$ hold.

DEFINITION 1. Let $S$ be an analytic sheaf on a complex analytic manifold $M$ and let $F$ be a Frechet space. Let $0$ be the sheaf of locally-defined analytic functions on $M$ and let $0_F$ be the sheaf of locally-defined analytic functions on $M$ with values in $F$, where by definition a continuous function $f$ from an open set $U \subset M$ to $F$ is called analytic if $u \circ f$ is analytic for all $u$ in $F^*$. Clearly $0_F$ is an 0-module, i.e., an analytic sheaf. The vectorization $S_F$ of $S$ (relative to $F$) is defined to be the sheaf $S \otimes 0_F$, the tensor product of the 0-modules $S$ and $0_F$. This is defined in [5] as the sheaf determined by the presheaf data

$$U \to S(U) \otimes 0_F(U) ,$$

where $S(U)$ and $0_F(U)$ are considered as $0(U)$-modules, together with the obvious restriction maps.

Note that if $T$ is a continuous linear operator from a Frechet space $F$ into a Frechet space $G$ then the natural homomorphism $T_0$ of $0_F$ into $0_G$ induced by $T$ gives rise to a homomorphism $T' = 1 \otimes T_0$ of $S_F$ into $S_G$. In particular, if $u$ is an element of $F^*$ (and so a continuous linear operator from $F$ into $C$) then $u$ induces a homomorphism of $S_F$ into $S_C$. But $S_C$ is canonically isomorphic to $S$, in virtue of the canonical isomorphism between the $0(U)$-modules $S(U) \otimes 0(U)$ and $S(U)$. (See [5] p. 8.) If we identify $S_C$ with $S$ it follows that each $u$ in $F^*$ induces a homomorphism $u'$ of $S_F$ onto $S$.

DEFINITION 2. If $S$ is an analytic subsheaf of the Cartesian product $0^n$ we define

$$S'_F(U) = \{f \in (0_F(U))^n : u \circ f \in S(U) \text{ for all } u \text{ in } F^*\} .$$

Clearly $S'_F$ so defined is an analytic subsheaf of the Cartesian product $(0_F)^n$.

THEOREM 2. *If $S$ is a coherent analytic subsheaf of $0^n$ then to each $p$ in $U \subset M$ and each $f$ in $S'_F(U)$ there exists a neighborhood $V$ of $p$, functions $H_1, \cdots, H_k$ in $S(V)$ and functions $G_1, \cdots, G_k$ in $0_F(V)$ such that*

$$r_{VU} f = \sum_{m=1}^{k} G_m H_m .$$

*Proof.* Since $S$ is coherent, there exists a neighborhood $V_0 \subset U$ of $p$ and functions $H_1, \cdots, H_k$ in $S(V_0)$ which generate $S$ at each point of $V_0$. We may assume that $\bar{V}_0$ is a compact subset of $U$. Let $V_0 \supset V_1 \supset V_2 \supset \cdots$

be a basis for the neighborhoods of $p$. Let $\Omega$ be the subset of $S(V_0)$ consisting of all elements in $S(V_0)$ which as elements of $(0(V_0))^n$ are bounded on $V_0$. Thus to each $h$ in $\Omega$ there exists $G=(G_1,\cdots,G_k)$ in $(0(V_i))^k$ for some $i$ such that the restriction of $h$ to $V_i$ has the form

$$h=\sum_{i=1}^{k} G_i H_i \ .$$

By choosing $i$ large enough we may assume that

$$\|G\|_i=\sup\{|G_j(q)|:q\in V_i,\ 1\leqq j\leqq k\}$$

is finite. Thus if for each pair $(i, N)$ of positive integers we let $\Omega_{iN}$ be the family of all $h$ in $\Omega$ such that $G$ can be chosen in $(0(V_i))^k$ with $\|G\|_i\leqq N$, we see that $\Omega=\bigcup \Omega_{iN}$ and that each $\Omega_{iN}$ is a closed subset of $\Omega$, where $\Omega$ has the norm defined by

$$\|h\|_0=\sup\{|h_i(q)|:1\leqq i\leqq n,\ q\in V_0\}$$

for each $h=(h_1,\cdots,h_n)\in\Omega\subset(0(V_0))^n$. By the Baire category theorem there exists $(i, N)$ such that $\Omega_{iN}$ has a nonvoid interior. From this it follows as usual that there exists a constant $K>0$ such that for each $h$ in $\Omega$ there exists $G$ in $(0(V_i))^k$ as above with $\|G\|_i\leqq K\|h\|_0$. Now consider $f$ as in the statement of the theorem, so that $f\in S_F'(U)\subset(0_F(U))^n$. By Theorem 1 there exists a sequence of vectors $\{b_j\}$ in $F$ which is bounded in each continuous semi-norm on $F$ and a sequence $\{P_j\}$ of continuous projections on $F$ having one-dimensional ranges such that $\sum_{j=1}^{\infty} P_j\circ f$ converges uniformly to $f$ on all compact subsets of $U$ and such that for each $j$ we have $P_j\circ f=f_jb_j$ with $f_j\in(0(U))^n$, where $\sum_{j=1}^{\infty}|f_j|$ converges uniformly on all compact subsets of $U$. Thus $\sum_{j=1}^{\infty}\|f_j\|_0$ is finite, since $\bar{V}_0\subset U$.

Now for each $j$ there exists $u$ in $F^*$ with $\langle b_j, u\rangle=1$. Thus

$$f_j=u\circ(f_jb_j)=u\circ(P_j\circ f)=(u\circ P_j)\circ f$$

is in $S(U)$ because $f\in S_F'(U)$ and $u\circ P_j\in F^*$. Thus $f_j\in S(U)$ for all $j$. By the above for each $j$ there exists $G^j=(G_1^j,\cdots,G_k^j)$ in $(0(V_i))^k$ such that on $V_i$ we have

$$f_j=\sum_{m=1}^{k} G_m^j H_m \ ,$$

with $\|G^j\|_i\leqq K\|f_j\|_0$. It follows that the series $\sum_{j=1}^{\infty} G^jb_j$ converges uniformly and absolutely on $V_i$ in each continuous semi-norm on $F$. Thus the sum of this series is an element $G=(G_1,\cdots,G_k)$ in $(0_F(V_i))^k$. Thus in the topology of uniform and absolute convergence on compact subsets of $V_i$ in each continuous semi-norm on $F$ we have

$$\begin{aligned} f &= \lim_{t\to\infty} \sum_{j=1}^{t} f_j b_j \\ &= \lim_{t\to\infty} \sum_{j=1}^{t} \sum_{m=1}^{k} G_m^j H_m b_j \\ &= \sum_{m=1}^{k} \Big(\lim_{t\to\infty} \sum_{j=1}^{t} G_m^j b_j\Big) H_m \\ &= \sum_{m=1}^{k} G_m H_m , \end{aligned}$$

as was to be proved.

The following consequence of Theorem 2 will be useful later.

LEMMA 4. *If the element $f$ of $S_F(U)$ has the property that $u'f$ is the zero element of $S(U)$ for all $u$ in $F^*$ then $f = 0$.*

*Proof.* By taking a covering of $U$ by small open sets we reduce to the case in which $f$ has a representation

$$f = \sum_{i=1}^{k} h_i \otimes g_i ,$$

with $h_i$ in $S(U)$ and $g_i$ in $0_F(U)$. Let $R$ be the sheaf on $U$ of relations of $h_1, \cdots, h_k$. Thus for each $u$ in $F^*$ we see that

$$\begin{aligned} 0 = u'f &= \sum_{i=1}^{k} h_i \otimes \langle g_i, u\rangle \\ &= \sum_{i=1}^{k} \langle g_i, u\rangle h_i . \end{aligned}$$

Thus by Definition 2 we see that $g = (g_1, \cdots, g_k) \in R_F'(U)$. By Theorem 2 it follows that each $p$ in $U$ has a neighborhood $V \subset U$ such that there exist $H_1, \cdots, H_t$ in $R(V)$ and $G_1, \cdots, G_t$ in $0_F(V)$ with

$$r_{VU} g = \sum_{j=1}^{t} G_j H_j .$$

Thus for each $i$ with $1 \leqq i \leqq k$ we have

$$r_{VU} g_i = \sum_{j=1}^{t} G_j H_j^i ,$$

where $H_j = (H_j^1, \cdots, H_j^k)$. Therefore on $V$ we have

$$\begin{aligned} f = \sum_{i=1}^{k} h_i \otimes g_i &= \sum_{i=1}^{k} h_i \otimes \Big(\sum_{j=1}^{t} G_j H_j^i\Big) \\ &= \sum_{i=1}^{k} \Big(\sum_{j=1}^{t} h_i \otimes (G_i H_j^i)\Big) \\ &= \sum_{j=1}^{t} \Big(\sum_{i=1}^{k} H_j^i h_i\Big) \otimes G_j = 0 \end{aligned}$$

since $H_j \in R(V)$ for all $j$. This proves Lemma 4.

We next give an important characterization of $S_F$ in case $S$ is a coherent analytic subsheaf of $0^n$ for some positive integer $n$.

THEOREM 3. *Let $M$ be a Stein manifold and $S$ a coherent analytic subsheaf of $0^n$. Let $F$ be a Frechet space. For each open $U \subset M$ there is a mapping $\tau(U)$ from $S(U) \otimes 0_F(U)$ into $(0_F(U))^n$ which for each $h = (h_1, \cdots, h_n)$ in $S(U)$ and $g$ in $0_F(U)$ maps $h \otimes g$ onto $gh = (gh_1, \cdots, gh_n)$ in $(0_F(U))^n$. For each such $g$ and $h$ the image $gh$ of $h \otimes g$ actually lies in the subset $S'_F(U)$ of $(0_F(U))^n$. The family of such mappings $\tau(U)$ induces an isomorphism $\tau$ of the sheaf $S_F$ (which was defined above to be the sheaf determined by the presheaf data $U \rightarrow S(U) \otimes 0_F(U))$ onto the sheaf $S'_F$. Thus $S'_F$ and $S_F$ are isomorphic.*

*Proof*. Clearly the map of the Cartesian product $S(U) \times 0_F(U)$ into $(0_F(U))^n$ defined by $(h, g) \rightarrow gh$ induces a group homomorphism of $(S(U), 0_F(U))$—the free abelian group generated by the elements of the Cartesian product $S(U) \times 0_F(U)$—into $(0_F(U))^n$. It is trivial to check that $N(S(U), 0_F(U))$ belongs to the kernel of this map, where $N(S(U), 0_F(U))$ is defined as in [5] p. 8 to be the subgroup of $(S(U), 0_F(U))$ generated by elements of the forms

( i ) $(x_1 + x_2, y) - (x_1, y) - (x_2, y)$

(ii) $(x, y_1 + y_2) - (x, y_1) - (x, y_2)$

(iii) $(ax, y) - (x, ay)$

where $x$, $x_1$, and $x_2$ are in $S(U)$, $y$, $y_1$, and $y_2$ are in $0_F(U)$, and $a \in 0(U)$. Thus this map induces a homomorphism $\tau(U)$ of the quotient $(S(U), 0_F(U))/N(S(U), 0_F(U)) = S(U) \otimes 0_F(U)$ into $(0_F(U))^n$. It is trivial to check that $\tau(U)$ is an $0(U)$-homomorphism. Now with $g$ and $h$ as above and $u$ in $F^*$ we have

$$u \circ \tau(U)(h \otimes g) = u \circ (gh) = (u \circ g)h \in S(U) .$$

Thus $\tau(U)(h \otimes g) \in S'_F(U)$. It follows that the range of $\tau(U)$ is a subset of $S'_F(U)$. It is now clear that the family of mappings $\tau(U)$ induces an 0-homomorphism $\tau$ of $S_F$ into $S'_F$. To show that $\tau$ is one-to-one we must prove

(a) If $\tau(U)(\sum_{i=1}^N h_i \otimes g_i) = 0$ then each $p$ in $U$ has a neighborhood $V$ such that $r_{VU}(\sum_{i=1}^N h_i \otimes g_i) = 0$.

To show that $\tau$ is onto we must prove

(b) If $f \in S'_F(U)$ then each p in $U$ has a neighborhood $V$ such that $r_{VU}f = \tau(V)(\sum_{i=1}^N h_i \otimes g_i)$ for some elements $h_i$ in $S(V)$ and $g_i$ in $0_F(V)$.

We first prove (a). If we let $R$ be the sheaf of relations on $U$ of $h_1, \cdots, h_N$ by the coherence of $R$ there exists a neighborhood $V_0$ of $p$ and elements $r_1 = (r_1^1, \cdots, r_1^N), \cdots, r_n = (r_n^1, \cdots, r_n^N)$ of $R(V_0)$ which generate $R$ at each point of $V_0$. Now

$$\sum_{i=1}^{N} g_i h_i = \tau(U)\Big(\sum_{i=1}^{N} h_i \otimes g_i\Big) = 0 .$$

Thus for each $u$ in $F^*$ we have

$$\sum_{i=1}^{N} (u \circ g_i) h_i = 0$$

so that $(u \circ g_1, \cdots, u \circ g_N) \in R(U)$ for all $u$ in $F^*$. By definition this means that $(g_1, \cdots, g_N) \in R'_F(U)$. Therefore by Theorem 2 we see that there exists a neighborhood $V$ of $p$ and $G = (G_1, \cdots, G_n)$ in $(0_F(V))^n$ such that $(g_1, \cdots, g_N) = G_1 r_1 + \cdots + G_n r_n$. Thus on $V$ we have

$$\begin{aligned}\sum_{i=1}^{N} h_i \otimes g_i &= \sum_{i=1}^{N} h_i \otimes \Big(\sum_{j=1}^{n} G_j r_j^i\Big) \\ &= \sum_{j=1}^{n} \Big(\sum_{i=1}^{N} (r_j^i h_i)\Big) \otimes G_j = 0\end{aligned}$$

since $r_j \in R(V)$ for each $j$. This proves (a).

To prove (b) notice by Theorem 2 that there exists a neighborhood $V$ of $p$, elements $h_1, \cdots, h_N$ in $S(V)$, and elements $g_1, \cdots, g_N$ in $0_F(V)$ such that on $V$ we have

$$f = \sum_{i=1}^{N} g_i h_i = \tau(V)\Big(\sum_{i=1}^{N} h_i \otimes g_i\Big) .$$

This completes the proof of Theorem 3.

We state for future reference a version of a theorem of Banach, first giving a definition.

DEFINITION 3. If $\{g_n\}$ is a sequence of vectors in a Frechet space $F$, the series $\sum_{n=1}^{\infty} g_n$ is called *absolutely convergent* if the series $\sum_{n=1}^{\infty} \|g_n\|$ converges for each continuous semi-norm $\|\ \|$ on $F$.

Notice that a continuous linear transformation from a Frechet space $F$ to a Frechet space $G$ takes absolutely convergent sequences into absolutely convergent sequences.

LEMMA 5. *Let $\sigma$ be a continuous linear map of a Frechet space $F$ onto a Frechet space $G$. Let $\{g_i\}$ be an absolutely convergent sequence from $G$. Then there exists an absolutely convergent sequence $\{f_i\}$ in $F$ such that $\sigma(f_i) = g_i$ for all $i$.*

*Proof.* Let $\{\|\ \|_k\}$ be a defining sequence of semi-norms on $F$. Since the map $\sigma$ is continuous, we see ([1] p. 40) that for each $k$ the set $\sigma\{f : \|f\|_k \leqq 1\}$ contains a neighborhood $\{g : \|g\|'_k \leqq 1\}$ of 0 in $G$, where $\|\ \|'_k$ is some continuous semi-norm on $G$. Thus for each $g$ in

$G$ and each $k$ there exists $f$ in $F$ with $\sigma(f) = g$ and $\|f\|_k \leqq \|g\|'_k$. Now for each $k$ choose $j = j(k)$ such that

$$\sum_{n=j}^{\infty} \|g_n\|'_k < 2^{-k},$$

so that

$$\sum_{k=1}^{\infty} \sum_{n=j(k)}^{\infty} \|g_n\|'_k < \infty .$$

We may assume that $j(1) < j(2) < \cdots$. For each $n$ with $j(k) \leqq n < j(k+1)$ choose $f_n$ in $F$ with $\sigma(f_n) = g_n$ and $\|f_n\|_k \leqq \|g_n\|'_k$. If for each $n$ we let $k(n)$ be the smallest value of $k$ for which $n < j(k+1)$, it follows that

$$\sum_{n=1}^{\infty} \|f_n\|_{k(n)} < \infty .$$

Since for each $t$ we have $\|f_n\|_t \leqq \|f_n\|_k$ for all $k \geqq t$ it follows that

$$\sum_{n=1}^{\infty} \|f_n\|_t$$

is finite for all $t$. This proves the lemma.

THEOREM 4. *If $S$ is a coherent analytic sheaf on a Stein manifold $M$ and if $F$ is a Frechet space then $H^N(M, S_F) = 0$ for all $N \geqq 1$.*

*Proof.* Let $f$ be an element of $H^N(M, S_F)$. Consider a locally finite covering $\{U_i\}$ of $M$ by holomorphically convex open sets $U_i$, so fine that $f$ is represented by an element of $H^N(\{U_i\}, S_F)$. For each finite sequence $K = (i_1, \cdots, i_k)$ of positive integers let $U_K = U_{i_1} \cap \cdots \cap U_{i_k}$. The element $f$ of $H^N(M, S_F)$ can be considered to belong to $H^N(\{U_i\}, S_F)$ and therefore can be represented by a cocycle $f = \{f_I\}$ of $Z^N(\{U_i\}, S_F)$. Here $I$ is any sequence of $N+1$ positive integers, and, for each $I$, $f_I$ is an element of $S_F(U_I)$. Also $\delta f = 0$, where $\delta$ is the coboundary operator from $C^N(\{U_i\}, S_F)$ into $C^{N+1}(\{U_i\}, S_F)$ and $Z^N(\{U_i\}, S_F)$ is the kernel of $\delta$. By choosing the covering $\{U_i\}$ fine enough we may assume that for each $K$ there exist elements $h_{1K}, \cdots, h_{\alpha K}$, with $\alpha$ depending on $K$, in $S(U_K)$ which generate $S$ at each point of $U_K$. This implies ([3], expose XVIII, p. 9) that every $h$ in $S(U_K)$ has a representation of the form $h = \sum_{i=1}^{\alpha} g_i h_{iK}$, with $g_i \in 0(U_K)$. We may also choose the covering $\{U_i\}$ so fine that, for each $I$, $f_I$ can be represented in the form

$$f_I = \sum_{i=1}^{\alpha} h_{iI} \otimes g_{iI}$$

with $h_{iI}$ as above and with $g_{iI}$ in $0_F(U_I)$.

By Theorem 1 there exists a sequence $\{P_n\}$ of continuous mutually annihilating projections on $F$ whose ranges are one dimensional and a sequence $\{b_n\}$ of vectors in $F$ bounded in each continuous semi-norm on $F$ having the following properties. For each $I$ and $i$ the series $\sum_{n=1}^{\infty} P_n \circ g_{iI}$ converges to $g_{iI}$ on $U_I$. For each $I$ and $i$ we have $P_n \circ g_{iI} = g_{iI}^n b_n$, where $g_{iI}^n \in 0(U_I)$. For each $I$ and $i$ the series $\sum_{n=1}^{\infty} g_{iI}^n$ converges absolutely in the Frechet space $0(U_I)$. Now since for each $n$ the projection $P_n$ induces a homomorphism of the sheaf $S_F$ onto itself, the element $\{P_n f_I\}$ of $C^N(\{U_i\}, S_F)$ is in $Z^N(\{U_i\}, S_F)$. Also

$$\begin{aligned} P_n f_I &= \sum_{i=1}^{\alpha} h_{iI} \otimes P_n g_{iI} \\ &= \sum_{i=1}^{\alpha} h_{iI} \otimes g_{iI}^n b_n = \left( \sum_{i=1}^{\alpha} g_{iI}^n h_{iI} \right) \otimes b_n \,. \end{aligned}$$

If for each $n$ and $I$ we let $f_I^n$ be the element $\sum_{i=1}^{\alpha} g_{iI}^n h_{iI}$ of $S(U_I)$ it follows that for each $n$ the element $f^n = \{f_I^n\}^{i=1}$ of $C^N(\{U_i\}, S)$ belongs to $Z^N(\{U_i\}, S)$. It is also clear that $f^n b_n = P_n f$.

Now there exists a natural Frechet space topology on each $S(U)$, described in [4], expose XVII. This topology has the property that for each $h$ in $S(U)$ the map $g \to gh$ of $0(U)$ into $S(U)$ is continuous. We therefore see that for each $I$ the series

$$\sum_{n=1}^{\infty} f_I^n = \sum_{n=1}^{\infty} \left( \sum_{i=1}^{\alpha} g_{iI}^n h_{iI} \right)$$

converges absolutely in $S(U_I)$ because for each $I$ and $i$ the series $\sum_{n=1}^{\infty} g_{iI}^n$ converges absolutely in $0(U_I)$. Now the space $C^N(\{U_i\}, S)$ is the Cartesian product of the Frechet spaces $S(U_I)$, and therefore possesses a Frechet space structure. Moreover $Z^N(\{U_i\}, S)$ is closed in $C^N(\{U_i\}, S)$ and is therefore also a Frechet space. Since for each $I$ the series $\sum_{n=1}^{\infty} f_I^n$ converges absolutely in $S(U_I)$ it follows that $\sum_{n=1}^{\infty} f^n$ converges absolutely in $Z^N(\{U_i\}, S)$. By Theorem B of [3] and Leray's theorem (see [5] p. 213) we see that the coboundary map $\delta$ of the Frechet space $C^{N-1}(\{U_i\}, S)$ into $Z^N(\{U_i\}, S)$ is onto. From [4] we also see that $\delta$ is continuous.

Let $J$ stand for an arbitrary $N$-tuple of positive integers. Thus for each $J$, by the above, there is a continuous homomorphism.

$$\tau_J : (G_1, \cdots, G_\alpha) \to \sum_{i=1}^{\alpha} G_i h_{iJ}$$

of the Frechet space $(0(U_J))^\alpha$ onto the Frechet space $S(U_J)$. These maps induce a continuous homomorphism $\tau$ of the Frechet space $\Phi$ onto the Frechet space $C^{N-1}(\{U_i\}, S)$, where $\Phi$ is defined to be the product $\prod_J (0(U_J))^\alpha$, with $\alpha$ depending as above on $J$, of the Frechet spaces $(0(U_J))^\alpha$. Thus

$$\sigma = \delta \circ \tau$$

is a continuous homomorphism of $\Phi$ onto $Z^N(\{U_i\}, S)$. Since $\sum_{n=1}^{\infty} f^n$ converges absolutely in $Z^N(\{U_i\}, S)$ it follows from Lemma 5 that there exists an absolutely convergent sequence $\{G^n\}$ in $\Phi$ with $\sigma(G^n) = f^n$ for all $n$. For each $n$ write $G^n = \{G^n_J\}$, where

$$G^n_J = (G^n_{1J}, \cdots, G^n_{\alpha J}) \in (0(U_J))^\alpha .$$

Thus for each $J$ we see that the series $\sum_{n=1}^{\infty} G^n_J$ converges absolutely and uniformly on every compact subset of $U_J$, so that the series $\sum_{n=1}^{\infty} G^n_J b_n$ converges absolutely in $(0_F(U_J))^\alpha$ to an element

$$G_J = (G_{1J}, \cdots, G_{\alpha J})$$

in $(0_F(U_J))^\alpha$. Thus for each $i$ and $J$ we have $G_{iJ} = \sum_{n=1}^{\infty} G^n_{iJ} b_n$.

For each $J$ let $e_J$ be the element

$$e_J = \sum_{i=1}^{\alpha} h_{iJ} \otimes G_{iJ}$$

of $S_F(U_J)$. Thus $e = \{e_J\} \in C^{N-1}(\{U_i\}, S_F)$. We shall finish the proof by showing that $\delta e = f$. To this end it is sufficient by Lemma 4 to show $u'(\delta e) = u'(f)$ for all $u$ in $F^*$. We compute:

$$\begin{aligned} u'(e_J) &= \sum_{i=1}^{\alpha} \langle G_{iJ}, u\rangle h_{iJ} \\ &= \sum_{i=1}^{\alpha} \left\langle \sum_{n=1}^{\infty} G^n_{iJ} b_n, u \right\rangle h_{iJ} \\ &= \sum_{n=1}^{\infty} \left( \sum_{i=1}^{\alpha} G^n_{iJ} h_{iJ} \right) \langle b_n, u\rangle \\ &= \sum_{n=1}^{\infty} (\tau_J(G^n_J)) \langle b_n, u\rangle \end{aligned}$$

absolutely in $S(U_J)$. Thus

$$u'(e) = \sum_{n=1}^{\infty} (\tau(G^n)) \langle b_n, u\rangle$$

absolutely in $C^{N-1}(\{U_i\}, S)$. Thus

$$\begin{aligned} u'(\delta e) = \delta(u'(e)) &= \sum_{n=1}^{\infty} (\delta \circ \tau)(G^n) \langle b_n, u\rangle \\ &= \sum_{n=1}^{\infty} \sigma(G^n) \langle b_n, u\rangle = \sum_{n=1}^{\infty} f^n \langle b_n, u\rangle . \end{aligned}$$

Also for each $I$ we have

$$u'(f_I) = \sum_{i=1}^{\alpha} \langle g_{iI}, u\rangle h_{iI}$$

$$= \sum_{i=1}^{\alpha} \left\langle \sum_{n=1}^{\infty} g_{iI}^{n} b_n, u \right\rangle h_{iI}$$
$$= \sum_{n=1}^{\infty} \left( \sum_{i=1}^{\alpha} g_{iI}^{n} h_{iI} \right) \langle b_n, u \rangle = \sum_{n=1}^{\infty} f_I^n \langle b_n, u \rangle .$$

Therefore $u'(f) = \sum_{n=1}^{\infty} f^n \langle b_n, u \rangle$. It follows that $u'(f) = u'(\delta e)$ for all $u$ in $F^*$, so that $f = \delta e$. This completes the proof of Theorem 4.

## REFERENCES

1. S. Banach, *Theorie des operations lineaires*, Warsaw, 1932.
2. E. Bishop, *Some global problems in the theory of functions of several complex variables* Amer. J. of Math., **83** (1961), 479–498.
3. H. Cartan, *Seminaire Ecole Normale Superieure*, 1951–1952.
4. ———, *Seminaire Ecole Normale Superieure*, 1952–1953.
5. R. Godement, *Theorie des faisceaux*, Paris, 1958.
6. A. Grothendieck, *Sur certains espoces de fonctions holomorphes. II.*, Jour. für die reine und angewandte Math., **192** (1953), 77–95.
7. A. E. Taylor, *A geometric theorem and its applications to biorthogonal systems*, Bull. Amer. Math. Soc., **53** (1947), 614–616.

INSTITUTE FOR ADVANCED STUDY
PRINCETON, NEW JERSEY

# EQUICONTINUITY OF SOLUTIONS OF A QUASI-LINEAR EQUATION

S. E. BOHN

On a bounded domain $\Omega$ of the $xy$-plane the equicontinuity of a family of solutions of a linear elliptic partial differential equation is usually demonstrated by showing that the first partial derivatives of solutions are uniformly bounded on compact interior subsets of $\Omega$. Finn [2] uses this same method in showing the equicontinuity for a class of quasi-linear elliptic equations referred to by him as "equations of minimal surface type." However, Finn cites an example which demonstrates that in general bounded collections of solutions of elliptic equations do not have uniformly bounded first partial derivatives on compact interior subsets.

Here we shall consider the question of the equicontinuity of a family of solutions of the quasi-linear equation

$$(1) \qquad L[z] \equiv A(x, y, p, q)r + 2B(x, y, p, q)s + C(x, y, p, q)t = 0$$

where, as usual, $p = z_x$, $q = z_y$, $r = z_{xx}$, $s = z_{xy}$, and $t = z_{yy}$ and where $A$, $B$ and $C$ satisfy a growth condition.

Suppose $D$ to be a domain in the $xy$-plane for which

(i) $A > 0$, $AC - B^2 = 1$, and $A$, $B$, and $C$ are continuous and have continuous first partial derivatives with respect to $p$ and $q$ on $T$ defined by $T \equiv \{(x, y, p, q) : (x, y) \in D \text{ and } -\infty < p, q < +\infty\}$, and

(ii) $(A + C)^2 \leqq (1/125) \log\log (p^2 + q^2 + e) + h(x, y)$ for all $(x, y, p, q) \in T$ where $h(x, y)$ is positive and continuous on $D$.

Henceforth, we shall assume that conditions (i) and (ii) are satisfied whenever reference is made to the equation (1).

THEOREM 1. *Let $\Omega$ be a bounded sub-domain of $D$ with boundary $\omega$ such that $\bar{\Omega} = \Omega + \omega \subset D$. If $\{f_\nu(x, y) : \nu \in \mathscr{A}\}$ is any collection of functions which are continuous and uniformly bounded on $\omega$ and if corresponding to each $f_\nu$ there exists a function $z(x, y; f_\nu)$ which is of class $C^2$ on $\Omega$, is continuous on $\bar{\Omega}$, is a solution of* (1) *on $\Omega$, and is such that $z(x, y; f_\nu) = f_\nu(x, y)$ on $\omega$, then the collection $\{z(x, y; f_\nu) : \nu \in \mathscr{A}\}$ is equicontinuous on $\Omega$.*

In proving Theorem 1 we shall employ a modification of the method used by Serrin [5] and in so doing depend heavily on the following

Received December 20, 1961. This result is contained in the author's doctoral dissertation presented to the University of Nebraska. Sincerest appreciation is expressed to Professor Lloyd K. Jackson for his direction.

principle:

*Maximum Principle* [3]. Let $D$ be any plane domain and consider the function $F(x, y, z, p, q, r, s, t)$ with the following assumptions:

(i) $F$ is continuous in all 8 variables in the region $T$ defined by $T \equiv \{(x, y, z, p, q, r, s, t) : (x, y) \in D$ and $-\infty < z, p, q, r, s, t < +\infty\}$ and

(ii) $F_z, F_p, F_q, F_r, F_s$, and $F_t$ are continuous on $T$, $F_s^2 - 4F_rF_t < 0$, $F_r > 0$, and $F_z \leqq 0$ on $T$.

Let $z_1(x, y)$ and $z_2(x, y)$ be continuous in a bounded and closed subdomain $\mathscr{H} \subset D$ and of class $C^2$ in the interior of $\mathscr{H}$. Furthermore, suppose $z_1(x, y) \leqq z_2(x, y)$ on the boundary of $\mathscr{H}$ and suppose that in the interior of $\mathscr{H}$

$$F(x, y, z_1, z_{1x}, z_{1y}, z_{1xx}, z_{1xy}, z_{1yy}) \geqq 0$$

and

$$F(x, y, z_2, z_{2x}, z_{2y}, z_{2xx}, z_{2xy}, z_{2yy}) \leqq 0 .$$

Then, either $z_1(x, y) < z_2(x, y)$ in the interior of $\mathscr{H}$ or

$$z_1(x, y) \equiv z_2(x, y) \quad \text{on } \mathscr{H} .$$

Suppose $M > 1$ to be a uniform bound of $|f_\nu|$, $\nu \in \mathscr{A}$ on $\omega$. Since constants are solutions of (1) it follows from the Maximum Principle that $|z_\nu(x, y)| \equiv |z(x, y; f_\nu)| < M$ for $(x, y) \in \bar{\Omega}$ and all $\nu \in \mathscr{A}$. Also, suppose $\{z_\nu(x, y) : \nu \in \mathscr{B}\} \equiv \{z_\nu(x, y) : \nu \in \mathscr{A}$ and $z_\nu(x, y) > 0$ on $\bar{\Omega}\}$.

LEMMA 1. *Let $P_0(x_0, y_0)$ be any point of $\Omega$ and suppose $\{K_n\}_{n=0}^{n=\infty}$ is a sequence of closed circular disks each having $P_0(x_0, y_0)$ as its center and $R_n = (1/7)^n R_0$ as its radius where $R_0 \leqq 1$ and $K_0 \subset \Omega$. Then whenever $z_\nu(x, y)$ is a positive solution of (1) there exists a constant $H$, $0 < H < 1$, depending only on $R_0$, $\delta \equiv \max h(x, y)$ where $(x, y) \in \bar{\Omega}$, and $M$ such that for all $\nu \in \mathscr{B}$*

$$z_\nu(x, y) > H[\delta, M, R_0] z_\nu(x_0, y_0) \qquad \text{on } 0 \leqq |P - P_0| \leqq (1/7)R_0$$

*and*

$$z_\nu(x, y) > H[\delta, M, (1/7)^n R_0] z_\nu(x_0, y_0) \geqq H[\delta, M, (1/7)R_0](1/n) z_\nu(x_0, y_0)$$

*on* $0 \leqq |P - P_0| \leqq (1/7)^{n+1} R_0$, $n = 1, 2, 3, \cdots$.[1]

*Proof.* Let $E$ denote the component of the set

$$\{(x, y) \in K_0 : z_\nu(x, y) > (1/2)\, z_\nu(x_0, y_0)\}$$

[1] See Bers and Nirenberg [1] for a proof of a Harnack inequality for solutions of the uniformly elliptic equation (1).

which contains $P_0(x_0, y_0)$. We can apply the Maximum Principle to conclude that $E$ must contain an arc of the circumference of $K_0$. Hence, there is a Jordan arc $\Gamma$ contained in $E$ with one end at $(x_0, y_0)$ and the other end at a point $(x_1, y_1)$ on the circumference of $K_0$ which is such that with the exception of $(x_1, y_1)$ $\Gamma$ is contained in the interior of $E$. Let $K^2$ and $K^3$ be the two closed disks each of which has radius $\sqrt{5}/2\, R_0$ and each of which has the points $(x_0, y_0)$ and $(x_1, y_1)$ on its circumference. Each point $(x, y) \in K^2 \cap K^3$ satisfies at least one of the following conditions:

(a) $(x, y) \in \Gamma \cup bdry(K^2 \cap K^3)$,

(b) $(x, y)$ is in a subdomain of $K^2$ the boundary of which consists of arcs of $\Gamma$ and arcs of the circumference of $K^2$,

(c) $(x, y)$ is in a subdomain of $K^3$ the boundary of which consists of arcs of $\Gamma$ and arcs of the circumference of $K^3$.

Let $K^4$ be the closed disk with center at

$$(x_4, y_4) \equiv \left(\frac{3x_0 + x_1}{4}, \frac{3y_0 + y_1}{4}\right)$$

and radius $(3/4)\, R_0$ and let $(x_2, y_2)$ and $(x_3, y_3)$ be the respective centers of $K^2$ and $K^3$. It is clear that

$$\{(x, y) : (x - x_2)^2 + (y - y_2)^2 \leqq \varepsilon^2(\sqrt{5}/2\, R_0)^2\} \subset \text{comp}\, K_0 \,,$$
$$\{(x, y) : (x - x_3)^2 + (y - y_3)^2 \leqq \varepsilon^2(\sqrt{5}/2\, R_0)^2\} \subset \text{comp}\, K_0 \,,$$

and

$$\{(x, y) : (x - x_4)^2 + (y - y_4)^2 \leqq \varepsilon^2(3/4\, R_0)^2\} \subset \text{interior}\,(K^2 \cap K^3)$$

where $\varepsilon = 1/10$.

Consider the function

$$v(x, y; \xi, \eta; r) \equiv \frac{N(e^{-\alpha\sigma^2} - e^{-\alpha r^2})}{1 - e^{-\alpha r^2}}$$

defined on the region

$$S(\xi, \eta; r) \equiv \{(x, y) : \varepsilon^2 r^2 \leqq \sigma^2 = (x - \xi)^2 + (y - \eta)^2 \leqq r^2\} \cap K_0$$

where $\alpha > 0$ and $N = 1/2\, z_\nu(x_0, y_0)$. In this region

$$v_x^2 + v_y^2 = \frac{4\alpha^2 N^2 \sigma^2 e^{-2\alpha\sigma^2}}{(1 - e^{-\alpha r^2})^2} \leqq \frac{4N^2}{\sigma^2} \leqq \frac{M^2}{\varepsilon^2 r^2} \qquad \text{for all } \nu \in \mathscr{B} \,.$$

Furthermore, $v < N$ on $S(\xi, \eta; r)$, $v = 0$ where $\sigma = r$, and $v > 0$ where $\sigma < r$. If $A$, $B$, and $C$ are evaluated at $(x, y, \gamma v_x, \gamma v_y)$, the following succession of inequalities are valid in $S(\xi, \eta; r)$ where $\gamma$, $0 < \gamma < 1$, is any fixed real number.

$$\begin{aligned} L[\gamma v](1-e^{-\alpha r^2}) &= 2\alpha\gamma N e^{-\alpha\sigma^2}\{2\alpha[A(x-\xi)^2+2B(x-\xi)(y-\eta) \\ &\qquad + C(y-\eta)^2]-(A+C)\} \\ &\geqq 2\alpha\gamma N e^{-\alpha\sigma^2}\left\{\frac{4\alpha(AC-B^2)\sigma^2}{(A+C)+\sqrt{(A+C)^2-4(AC-B^2)}}-(A+C)\right\} \\ &\geqq \frac{2\alpha\gamma N e^{-\alpha\sigma^2}}{(A+C)}\{2\alpha\varepsilon^2 r^2-(A+C)^2\} \\ &\geqq \frac{2\alpha\gamma N e^{-\alpha\sigma^2}}{(A+C)}\left\{2\alpha\varepsilon^2 r^2-\frac{1}{125}\log\log[\gamma^2(v_x^2+v_y^2)+e]-h(x,y)\right\} \\ &\geqq \frac{2\alpha\gamma N e^{-\alpha\sigma^2}}{(A+C)}\left\{2\alpha\varepsilon^2 r^2-\frac{1}{125}\log\log\left[\frac{M^2}{\varepsilon^2 r^2}+e\right]-\delta\right\} \end{aligned}$$

where $\delta \equiv \max h(x, y)$ for $(x, y) \in \bar{\Omega}$. Now $L[\gamma v] \geqq 0$ on $S(\xi, \eta; r)$ if one chooses

$$\alpha \geqq \frac{1}{250\varepsilon^2 r^2}\left\{\log\log\left[\frac{M^2}{\varepsilon^2 r^2}+e\right]+125\delta\right\} .$$

Let

$$v_2(x, y) \equiv v(x, y; x_2, y_2; \sqrt{5}/2\, R_0)$$

and

$$v_3(x, y) \equiv v(x, y; x_3, y_3, \sqrt{5}/2\, R_0) .$$

Let

$$\alpha = \frac{32}{45R_0^2}\log\log[(4M/3\varepsilon\, R_0)^2+e]+125\delta\}$$

and assume that $(x, y)$ is in the interior of $K^2 \cap K^3$ and either $(x, y) \in \Gamma$ or $(x, y)$ satisfies condition (b), then we can apply the Maximum Principle to conclude that $z_\nu(x, y) > v_2(x, y)$. Similarly, if $(x, y)$ is in the interior of $K^2 \cap K^3$ and either $(x, y) \in \Gamma$ or $(x, y)$ satisfies (c), we can conclude that $z_\nu(x, y) > v_3(x, y)$. Thus, for all $(x, y) \in$ interior $(K^2 \cap K^3)$ it follows that

$$z_\nu(x, y) > \min[v_2(x, y), v_3(x, y)] .$$

Now on the circle $(x - x_4)^2 + (y - y_4)^2 = \varepsilon^2(3/4\, R_0)^2$

$$\begin{aligned} \gamma_0 \equiv \min[v_2(x, y), v_3(x, y)] &= N\frac{\exp\left(-\frac{5}{4}\lambda^2\alpha R_0^2\right)-\exp\left(-\frac{5}{4}\alpha R_0^2\right)}{1-\exp\left(-\frac{5}{4}\alpha R_0^2\right)} \\ &> N(1-\lambda^2)\exp\left(-\frac{5}{4}\lambda^2\alpha R_0^2\right) \end{aligned}$$

where $\lambda = [(\sqrt{17} + 3\varepsilon)\sqrt{5}]/10 < 1$. Another application of the Maximum Principle yields $z_\nu(x, y) > (\gamma_0/N)v_4(x, y)$ on $S(x_4, y_4, 3/4\, R_0)$ where

$$v_4(x, y) \equiv v(x, y; x_4, y_4; 3/4\, R_0) .$$

Now the annulus $S(x_4, y_4; 3/4\, R_0)$ contains the disk with center at $(x_0, y_0)$ and radius $1/7\, R_0$. On this disk

$$\begin{aligned} v_4(x, y) \geqq \gamma_1 \equiv N \frac{\exp\left(-\frac{9}{16}\rho^2\alpha R_0^2\right) - \exp\left(-\frac{9}{16}\alpha R_0^2\right)}{1 - \exp\left(-\frac{9}{16}\alpha R_0^2\right)} \\ > N(1-\rho^2)\exp\left(-\frac{9}{16}\rho^2\alpha R_0^2\right) \end{aligned}$$

where $\rho = 11/21$.

Therefore, on the disk with center $(x_0, y_0)$ and radius $1/7\, R_0$.

$$\begin{aligned} z_\nu(x, y) > \frac{\gamma_0\gamma_1}{N} &> \frac{1}{2}(1-\lambda^2)(1-\rho^2)\exp\left[-\left(\frac{5}{4}\lambda^2 + \frac{9}{16}\rho^2\right)\alpha R_0^2\right] z_\nu(x_0, y_0) \\ &> \frac{1}{2}(1-\lambda^2)(1-\rho^2)\exp\left(-\frac{45}{32}\alpha R_0^2\right) z_\nu(x_0, y_0) \\ &> \frac{1}{2}(1-\lambda^2)(1-\rho^2)\exp(-125\delta) \\ &\quad \cdot \exp\{-\log\log[(4M/3\varepsilon R_0)^2 + e]\} z_\nu(x_0, y_0) \\ &> H[\delta, M, R_0] z_\nu(x_0, y_0) \\ &\qquad \text{on } 0 \leqq |P - P_0| \leqq (1/7)\, R_0 \text{ for all } \nu \in \mathscr{B} \end{aligned}$$

where

$$H[\delta, M, R_0] = \frac{1}{2}(1-\lambda^2)(1-\rho^2)\exp(-125\delta)\left\{\log\left[\left(\frac{4M}{3\varepsilon R_0}\right)^2 + e\right]\right\}^{-1} .$$

Now by an inductive argument one concludes that

$$\begin{aligned} H[\delta, M, (1/7)^n R_0] &= \frac{1}{2}(1-\lambda^2)(1-\rho^2)\exp(-125\delta) \\ &\quad \cdot \left\{\log\left[\left(\frac{4M(7)^n}{3\varepsilon R_0}\right)^2 + e\right]\right\}^{-1} \geqq \frac{1}{n} H[\delta, M, 1/7\, R_0] \end{aligned}$$

and

$$z_\nu(x, y) > \frac{1}{n} H[\delta, M, 1/7\, R_0] z_\nu(x_0, y_0) \quad \text{on } 0 \leqq |P - P_0| \leqq (1/7)^{n+1} R_0 ,$$

$n = 1, 2, 3, \cdots$ for all $\nu \in \mathscr{B}$, thus proving the lemma.

LEMMA 2. *Using the assumptions of Lemma* 1

$$z_\nu(x, y) < \frac{1}{H[\delta, M, 7/8\,R_0]}\, z_\nu(x_0, y_0)$$

*for* $0 \leqq |P - P_0| \leqq 1/8\,R_0$ *for all* $\nu \in \mathscr{B}$.

*Proof.* Follows directly from Lemma 1.

It is of interest to note that if $z(x, y)$ is a positive solution of $A(x, y, p, q)r + 2B(x, y, p, q)s + C(x, y, p, q)t = 0$ in a domain $T$, then for any compact $U \subset T$ and compact $S$ properly contained in $U$ there is an $H > 0$ depending only on the bound of $z(x, y)$ on $U$ and the distance from $S$ to the boundary of $U$ such that

$$\frac{1}{H}\, z(x_2, y_2) \leqq z(x_1, y_1) \leqq Hz(x_2, y_2)$$

for any two points $(x_1, y_1)$ and $(x_2, y_2)$ in $S$.

LEMMA 3. *If* $z_\nu(x, y)$ $\nu \in \mathscr{B}$ *is a solution of* (1) *on the interior of a closed circular disk* $K_0$ *of radius* $R_0 \leqq 1$ *with center* $P_0(x_0, y_0)$, *then there exists a continuous decreasing function* $g_{P_0}(r)$, $0 \leqq r < R_0$, $g_{P_0}(0) = 1$, *and a continuous increasing function* $f_{P_0}(r)$, $0 \leqq r < R_0$, $f_{P_0}(0) = 1$ *such that*

$$g_{P_0}(r) z_\nu(x_0, y_0) \leqq z_\nu(x, y) \leqq f_{P_0}(r) z_\nu(x_0, y_0)$$

*for* $0 \leqq |P - P_0| \leqq r$ *where* $g$ *and* $f$ *are independent of* $\nu$.

*Proof.* Define

$$g_{P_0}(r) \equiv \inf_{\nu \in \mathscr{B}} \inf_{|P-P_0| \leqq r} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)}$$

and

$$f_P(r) \equiv \sup_{\nu \in \mathscr{B}} \sup_{|P-P_0| \leqq r} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)}\,.$$

By Lemma 1, Lemma 2, and an argument similar to that used in Kellogg [4] (page 263) $f_{P_0}(r)$ and $g_{P_0}(r)$ exist for each $0 \leqq r < R_0$. Using standard arguments it is clear that

$$(2) \qquad \lim_{r \to r_0^-} \inf_{\nu \in \mathscr{B}} \inf_{|P-P_0| \leqq r} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} = \inf_{\nu \in \mathscr{B}} \inf_{|P-P_0| \leqq r_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)}$$

for $0 < r_0 < R_0$ and

$$(3) \qquad \lim_{r \to r_0^-} \sup_{\nu \in \mathscr{B}} \sup_{|P-P_0| \leqq r} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} = \sup_{\nu \in \mathscr{B}} \sup_{|P-P_0| \leqq r_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)}$$

for $0 < r_0 < R_0$. Also,

$$\liminf_{r\to 0+} \inf_{\nu\in\mathscr{B}} \inf_{|P-P_0|\leqq r} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} = 1 . \tag{4}$$

This follows by observing that whenever $z_\nu(x, y) > 0$ for $0 \leqq |P - P_0| \leqq R_0$

$$z_\nu(x, y) > H[\delta, M, R_0]z_\nu(x_0, y_0) \quad \text{for } 0 \leqq |P - P_0| \leqq 1/7\, R_0$$

and all $\nu \in \mathscr{B}$. This latter inequality implies

$$\begin{aligned} &z_\nu(x, y) - H[\delta, M, R_0]z_\nu(x_0, y_0) \\ &\qquad > H[\delta, M, 1/7\, R_0]\{z_\nu(x_0, y_0) - H[\delta, M, R_0]z_\nu(x_0, y_0)\} \end{aligned}$$

for $0 \leqq |P - P_0| \leqq (1/7)^2R_0$ and all $\nu \in \mathscr{B}$. Thus, for $0 \leqq |P - P_0| < (1/7)^2R_0$ and all $\nu \in \mathscr{B}$

$$z_\nu(x, y) > [1 - (1 - H[\delta, M, R_0])(1 - H[\delta, M, 1/7\, R_0])]z_\nu(x_0, y_0)$$

By induction

$$\begin{aligned} z_\nu(x, y) &> \left[1 - \prod_{i=0}^{n} (1 - H[\delta, M, (1/7)^i R_0])\right] z_\nu(x_0, y_0) \\ &> \left[1 - (1 - H[\delta, M, R_0]) \prod_{i=1}^{n} \left(1 - H[\delta, M, 1/7\, R_0] \frac{1}{i}\right)\right] z_\nu(x_0, y_0) \end{aligned} \tag{5}$$

for $0 \leqq |P - P_0| \leqq (1/7)^{n+1}R_0$ and all $\nu \in \mathscr{B}$. Hence,

$$\begin{aligned} 1 - \inf_{\nu\in\mathscr{B}} &\inf_{|P-P_0|\leqq(1/7)^{n+1}R_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} \\ &< \exp\left(-H[\delta, M, R_0] - H[\delta, M, 1/7\, R_0] \sum_{i=1}^{n} \frac{1}{i}\right) . \end{aligned}$$

(4) then follows by the usual argument.

Suppose $P$ is any point in the circle $0 \leqq |P - P_0| \leqq (1/7)^nR_0/[1 + (1/7)^n]$ and let $K$ be the interior of a closed circular disk of radius $1/[1 + (1/7)^n]R_0$ about $P$. Since $z_\nu(x, y) > 0$ on $0 \leqq |P - P_0| \leqq R_0$ we have $z_\nu(x', y') > 0$ on $0 \leqq |P - P'| \leqq R_0/[1 + (1/7)^n]$ for all $\nu \in \mathscr{B}$. Also

$$\begin{aligned} z_\nu(x', y') - \Big[1 - (1 - H[\delta, M, 7/8\, R_0])^i \\ \cdot \prod_{i=1}^{n} \left(1 - H[\delta, M, 1/8\, R_0] \frac{1}{i}\right)\Big] z_\nu(x, y) > 0 \end{aligned}$$

on

$$0 \leqq |P - P'| \leqq (1/7)^{n+1} \frac{1}{1 + (1/7)^n} R_0 \qquad \text{for all } \nu \in \mathscr{B} .$$

Now $P_0(x_0, y_0)$ is such a point $P'(x', y')$; therefore, for all $\nu \in \mathscr{B}$ and $0 \leqq |P - P_0| \leqq (1/7)^{n+1}[1/(1 + (1/7)^n)]R_0$

$$z_\nu(x_0, y_0) > \Big[1 - (1 - H[\delta, M, 7/8\, R_0]) \cdot \prod_{i=1}^{n} \Big(1 - H[\delta, M, 1/8\, R_0] \frac{1}{i}\Big)\Big] z_\nu(x, y) ,$$

$$\sup_{\nu \in \mathscr{B}} \sup_{|P-P_0| \leq (1/7)^{n+1}[1/(1+(1/7)^n)]R_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} - 1 < \exp\Big(-H[\delta, M, 7/8\, R_0] - H[\delta, M, 1/8\, R_0] \sum_{i=1}^{n} \frac{1}{i}\Big)$$

and we may conclude that

$$\lim_{r \to 0+} \sup_{\nu \in \mathscr{B}} \sup_{|P-P_0| \leq r} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} = 1 . \tag{6}$$

We will now show that

$$\lim_{r \to r_0^+} \sup_{\nu \in \mathscr{B}} \sup_{|P-P_0| \leq r} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} = \sup_{\nu \in \mathscr{B}} \sup_{|P-P_0| \leq r_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} \tag{7}$$

for $0 \leqq r_0 < R_0$.

Suppose the contrary, then since $f_{P_0}(r)$ is increasing $\lim_{r \to r_0^+} f_{P_0}(r) > f_{P_0}(r)$. Hence, there exists an $\varepsilon > 0$ and a decreasing sequence $\{r_n\}$ converging to $r_0$ such that for all positive integers $n$ $f_{P_0}(r_n) - f_{P_0}(r_0) > \varepsilon$. By the definition of supremum there exists for the above $\varepsilon$ and each $n$ a function $z_n(x, y)$ such that

$$\sup_{\nu \in \mathscr{B}} \sup_{|P-P_0| \leq r_n} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} - \sup_{|P-P_0| \leq r_n} \frac{z_n(x, y)}{z_n(x_0, y_0)} \leqq \frac{\varepsilon}{2}$$

and thus,

$$\sup_{|P-P_0| \leq r_n} \frac{z_n(x, y)}{z_n(x_0, y_0)} - \sup_{\nu \in \mathscr{B}} \sup_{|P-P_0| \leq r_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} > \frac{\varepsilon}{2} .$$

By the Maximum Principle

$$\sup_{|P-P_0| \leq r_n} \frac{z_n(x, y)}{z_n(x_0, y_0)}$$

is assumed at some point $P_n(x_n, y_n)$ on $|P - P_0| = r_n$. Hence, there exists a sequence of points $\{P_n(x_n, y_n)\}$ which contains a convergent subsequence which converges to a point $P_0'(x_0', y_0') \in |P - P_0| = r_0$. Suppose our sequence is such without relabeling. Let

$$\varepsilon_1 = \varepsilon \Big/ \sup_{\nu \in \mathscr{B}} \sup_{|P-P_0| \leq r_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} .$$

Therefore,

$$\frac{z_n(x_n, y_n)}{z_n(x_0, y_0)} - \frac{z_n(x_0', y_0')}{z_n(x_0, y_0)} \geqq \frac{z_n(x_n, y_n)}{z_n(x_0, y_0)} - \sup_{\nu\in\mathscr{B}} \sup_{|P-P_0|\leqq r_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)}$$
$$> \frac{\varepsilon_1}{2} \sup_{\nu\in\mathscr{B}} \sup_{|P-P_0|\leqq r_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} .$$

Let us center our attention on the point $P_0'(x_0', y_0')$. Then, using (6), there exists a $\delta_1 > 0$ such that

$$\sup_{\nu\in\mathscr{B}} \sup_{|P-P_0'|\leqq r} \frac{z_\nu(x, y)}{z_\nu(x_0', y_0')} - 1 \leqq \frac{\varepsilon_1}{2} \qquad \text{if } r < \delta_1 .$$

Also, by (4) there exists a $\delta_2 > 0$ such that if $r < \delta_2$

$$1 - \inf_{\nu\in\mathscr{B}} \inf_{|P-P_0'|\leqq r} \frac{z_\nu(x, y)}{z_\nu(x_0', y_0')} \leqq \frac{\varepsilon_1}{2} .$$

Thus, if $|P - P_0'| \leqq \min[\delta_1, \delta_2]$

$$\left|\frac{z_\nu(x, y)}{z_\nu(x_0', y_0')} - 1\right| \leqq \frac{\varepsilon_1}{2} \qquad \text{for all } \nu\in\mathscr{B} .$$

It then follows that if $|P_n - P_0'| \leqq \min[\delta_1, \delta_2]$

$$\frac{\varepsilon_1}{2} \sup_{\nu\in\mathscr{B}} \sup_{|P-P_0|\leqq r_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} < \frac{z_n(x_n, y_n) - z_n(x_0', y_0')}{z_n(x_0, y_0)}$$
$$< \frac{z_n(x_n, y_n) - z_n(x_0', y_0')}{z_n(x_0', x_0')} \cdot \frac{z_n(x_0', y_0')}{z_n(x_0, y_0)}$$
$$< \frac{z_n(x_n, y_n) - z_n(x_0', y_0')}{z_n(x_0', y_0')} \cdot \sup_{\nu\in\mathscr{B}} \sup_{|P-P_0|\leqq r_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)}$$
$$< \frac{\varepsilon_1}{2} \sup_{\nu\in\mathscr{B}} \sup_{|P-P_0|\leqq r_0} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} ,$$

a contradiction. By a similar argument we may conclude

$$(8) \qquad \operatorname*{Lim}_{r\to r_0^+} \inf_{\nu\in\mathscr{B}} \inf_{|P-P_0|\leqq r} \frac{z_\nu(x, y)}{z_\nu(x_0, y_0)} = \inf_{\nu\in\mathscr{B}} \inf_{|P-P_0|\leqq r_0} \frac{z_\nu(x, y)}{z_\nu(y_0, x_0)} ,$$

Hence, by (2), (3), (7), and (8) our lemma is true.

*Proof of Theorem* 1. Recall that for $\nu\in\mathscr{A}$ $|f_\nu| < M$ on $\omega$ and $|z_\nu(x, y)| < M$ on $\bar{\Omega}$. Also, for all $\nu\in\mathscr{A}$, $z_\nu(x, y) + M$ satisfies (1) and $z_\nu(x, y) + M > 0$ on $\bar{\Omega}$.

Let $P_0(x_0, y_0)$ be any point of $\Omega$ and assume $K$ is a closed circular disk whose center is $P_0(x_0, y_0)$ and such that $K \subset \Omega$. Hence, by Lemma 3 there exists positive continuous functions $f_{P_0}(r)$ and $g_{P_0}(r)$ (independent of $\nu$) such that $\lim_{r\to0} f_{P_0}(r) = 1$, $\lim_{r\to0} g_{P_0}(r) = 1$, and on the interior of $K$

$$g_{P_0}(r)[z_\nu(x_0, y_0) + M] \leqq z_\nu(x, y) + M \leqq f_{P_0}(r)[z_\nu(x_0, y_0) + M]$$

and

$$-\,|\,z_\nu(x_0,\,y_0)+M\,|\,|g_{P_0}(r)-1\,|\leqq z_\nu(x,\,y)-z_\nu(x_0,\,y_0)$$
$$\leqq |\,z_\nu(x_0,\,y_0)+M\,|\,|f_{P_0}(r)-1\,|$$

for all $\nu\in\mathscr{A}$. It then follows that since $\{z_\nu(x,y):\nu\in\mathscr{A}\}$ is uniformly bounded on $\bar{\Omega}$ that $\{z_\nu(x,y):\nu\in\mathscr{A}\}$ is equicontinuous on $\Omega$ thus proving Theorem 1.

## References

1. Bers and Nirenberg, *On linear and nonlinear elliptic boundary value problems in the plane*, Convegno Internazionale sulle Equazioni Derivate Parziali, (1954), 141-167.
2. Finn, *On Equations of minimal surface type*, Annals of Math., **60** (1954), 397-416.
3. Hopf, *Elementarie Betrachtungen uber die Losungen partieller Differentialgleichungen zweiter Ordnung vom elliptischen Typus*, Sitzungsberichte Preuss. Akad. Wiss., **19** (1927), 147-152.
4. Kellogg, *Foundation of Potential Theory*, Dover Publications, Inc., 1953.
5. Serrin, *On the Harnack inequality for linear elliptic equations*, Journ. D Analyse Math, **4** (1954-55), 292-308.

Bowling Green State University

# SOME FUNCTION CLASSES RELATED TO THE CLASS OF CONVEX FUNCTIONS

A. M. BRUCKNER AND E. OSTROW

**1. Introduction.** A real-valued function $f$ defined on the positive real line $[0, \infty)$ is said to be convex if for every $x \geqq 0, y \geqq 0$, and $\alpha, 0 \leqq \alpha \leqq 1, f$ satisfies the inequality

$$f[\alpha x + (1 - \alpha)y] \leqq \alpha f(x) + (1 - \alpha)f(y) . \tag{1}$$

Such functions are important in many parts of analysis and geometry and their properties have been studied in detail (see e.g. the expository article Beckenbach [1] which contains an extensive bibliography).

A related class of functions is the class of superadditive functions which satisfy the defining inequality

$$f(x + y) \geqq f(x) + f(y) . \tag{2}$$

These functions, more precisely their negatives which are subadditive, have been studied by Hille and Phillips [5] and R. A. Rosenbaum [7] among others.

In the paper we shall be concerned, in large part, with classes of functions that properly lie between these two classes and which are defined by inequalities which are weaker than (1) but stronger than (2). We obtain a strict hierarchy of classes and various characterizing properties of these classes and study a simple averaging operation that transforms each class into a smaller class.

**2. Definitions and elementary properties of the classes.** We shall restrict our attention generally to functions which are continuous, nonnegative, and for which $f(0) = 0$ unless the contrary is explicitly stated. The requirement of being nonnegative simplifies many proofs which could be given without this assumption by considering the sum of $f$ with a suitably chosen linear function.

DEFINITION 1. Let $f$ be defined on $[0, \infty)$. The average function $F$ of $f$ is the function defined for all $x > 0$ by

$$F(x) = \frac{1}{x}\int_0^x f(t)\,dt , \qquad F(0) = 0 .$$

DEFINITION 2. The function $f$ is said to be starshaped if for each

Received January 10, 1962.

$\alpha, 0 \leqq \alpha \leqq 1$, and all $x$

$$f(\alpha x) \leqq \alpha f(x) .$$

It is easy to see that the set of points lying above the graph of a starshaped function is starshaped with respect to the origin in the usual sense. A function can, of course, be starshaped with respect to any other point on its graph, the definition of this phenomenon being made in an obvious way. The characterization of Lemma 3 below then applies mutatis mutandis. It is not hard to verify that a continuous function is convex if and only if it is starshaped with respect to a set of points dense in its graph.

DEFINITION 3. The function $f$ is said to be convex on the average, starshaped on the average, or superadditive on the average if $F$ is respectively convex, starshaped, or superadditive.

In the sequel we shall use the abbreviation COA for convex on the average. We shall also use the following notation for derivatives:

$$\underline{f}'(x_0) = \underline{\lim}_{h \to 0} \frac{f(x_0 + h) - f(x_0)}{h}, f'_+(x_0) = \underline{\lim}_{h \to 0^+} \frac{f(x_0 + h) - f(x_0)}{h} ,$$

and

$$f'_-(x_0) = \underline{\lim}_{h \to 0^-} \frac{f(x_0 + h) - f(x_0)}{h} .$$

Simple characterizations of the classes are recorded in the following series of lemmas.

LEMMA 1. *A continuous convex function $f$ is left and right differentiable at each point, the one-sided derivatives being increasing functions. Conversely, if any one of the Dini derivatives of a continuous function $f$ is increasing, the function is convex.*

*Proof.* For a proof of the first part see Hardy, Littlewood, and Polya [4]. To prove the converse, let $Df$ denote an increasing Dini derivative of $f$ and let $G$ be an indefinite integral of $Df$. Then $G$ is convex. If $x_0$ is a point of continuity of $Df$, then both $f$ and $G$ are differentiable at $x_0$ and $Df(x_0) = G'(x_0) = f'(x_0)$. Since $Df$ is increasing, it is continuous except on at most a countable set of points. It follows (see Hobson [6]) that $f$ and $G$ differ by at most a constant. Thus $f$ is convex.

The proofs of the next three Lemmas are straightforward and will be omitted.

LEMMA 2. *The function $f$ is* COA *if and only if* $\underline{f}' \geqq 2F'$.

LEMMA 3. *The function $f$ is starshaped if and only if either one the two following conditions is satisfied*:

(i) $f(x)/x$ *is increasing*,

(ii) $\underline{f}'(x) \geqq f(x)/x$ *for all* $x$.

LEMMA 4. *The function $f$ is starshaped on the average if and only if* $f \geqq 2F$.

The inequality $f \geqq 2F$ has the following simple geometric interpretation: Since

$$xF(x) = \int_0^x f(t)\,dt \leqq \frac{x}{2} f(x) ,$$

the area under the graph of $f$ is at each point dominated by the area of the triangle with vertices $(0, 0)$, $(x, 0)$ and $(x, f(x))$.

The inequality $f(x_0) \geqq 2F(x_0)$ can be cast in the form

$$\frac{F(x_0)}{x_0} \leqq \frac{1}{2} \frac{f(x_0)}{x_0} .$$

Since $F(x)/x$ is increasing, we actually obtain the slightly more general result,

$$\frac{F(a)}{a} \leqq \frac{F(x_0)}{x_0} \leqq \frac{1}{2} \frac{f(x_0)}{x_0}$$

for all $a \leqq x_0$. This means geometrically that for $a < x_0$, the area of the triangle cut off from the above mentioned triangle by the line $x = a$ is no smaller than the area under the graph of $f$ from 0 to $a$.

LEMMA 5. *If $f$ is respectively convex, convex on the average, starshaped, or superadditive, then $f$ is a nondecreasing function.*

*Proof*. We have restricted ourselves to nonnegative functions for which $f(0) = 0$. If $f$ is superadditive, then $f(y) = f[x + (y - x)]$

$$\geqq f(x) + f(y - x) \geqq f(x) \quad \text{for } y \geqq x .$$

As we show in Theorem 5, $f$ satisfying any of the other conditions implies that $f$ is superadditive.

If $f$ is merely starshaped on the average, it is clear from the geometric interpretation of $f \geqq 2F$ that $f$ need not be increasing.

Since $f$ is an increasing function provided $f$ belongs to one of the function classes of Lemma 5, $f$ has a finite derivative almost everywhere. For all these classes, $F$ has a continuous derivative for $x > 0$

since $xF'(x) = f(x) - F(x)$. We consider the behavior of $F'$ at the origin in Theorem 7 below.

We now investigate various operations under which our function classes are closed. We have first of all

THEOREM 1. *Let $f$ and $g$ be respectively convex,* COA, *starshaped, starshaped on the average, superadditive, superadditive on the average; then for $a \geqq 0, b \geqq 0, af + bg$ belongs to the same class.*

The proof involves a trivial computation.

The next two theorems consider the behavior of our classes under the operation of pointwise limits.

THEOREM 2. *Let $\{f_n\}$ be a sequence of convex, starshaped, or superadditive functions converging pointwise to a limit function $f$. Then $f$ is respectively convex, starshaped, or superadditive. Moreover, the average functions $F_n$ converge to the average function $F$.*

*Proof.* It is clear that the defining inequalities of these classes are preserved in the limit. The proof of the second statement parallels the proof of the corresponding part of Theorem 3.

THEOREM 3. *Let $\{f_n\}$ be a sequence of* COA *functions converging pointwise to a continuous limit $f$. The limit function is then* COA *and the average functions $F_n$ converge to the average function $F$.*

*Proof.* Let $b > 0$. The sequence $\{f_n\}$ is uniformly bounded on $[0, b]$ by $\sup \{f_n(b)\} = M$. $M$ is finite for $f_n(b) \to f(b)$ and $M$ is a uniform bound because each $f_n$ is an increasing function. By the Lebesgue bounded convergence theorem,

$$\frac{1}{x}\int_0^x f_n(t)\,dt \to \frac{1}{x}\int_0^x f(t)\,dt$$

for each $x \in [0, b]$, that is $F_n(x) \to F(x)$. Since $b$ was arbitrary, this last relation holds for all $x$. The convexity of $F$ follows from the convexity of $F_n$.

In general, however, it is not true that the limit of the average functions is equal to the average of the limit function. If $f_n \to f$ and the averages $F_n \to G$, an easy calculation shows that $F \leqq G$. For functions which are starshaped on the average, we do have the following theorem.

THEOREM 4. *If $\{f_n\}$ is starshaped on the average and $f_n \to f$, then $f$ is starshaped on the average.*

*Proof.* For each $x > 0$, let $T_n^x$ and $T^x$ be the linear functions determined by the origin and the points $(x, f_n(x))$ and $(x, f(x))$. Since $f_n \to f$, $T_n^x \to T^x$. Moreover, the inequality $2F_n \leqq f_n$ is equivalent to

$$\int_0^x f_n(t)\, dt \leqq \int_0^x T_n^x(t)\, dt ;$$

by Fatou's theorem,

$$\int_0^x f(t)\, dt \leqq \varliminf_{n\to\infty} \int_0^x f_n(t)\, dt \leqq \varliminf_{n\to\infty} \int_0^x T_n^x(t)\, dt = \int_0^x T^x(t)\, dt .$$

Thus,

$$\frac{1}{x}\int_0^x f(t)\, dt \leqq \frac{1}{x}\int_0^x T^x(t)\, dt = \frac{1}{2} f(x) ,$$

i.e.

$$F(x) \leqq \frac{1}{2} f(x) ,$$

so $f$ is starshaped on the average.

3. **The hierarchy.** We now consider the inclusion relationships among the six classes.

THEOREM 5. *Let $f$ be a nonnegative continuous function which vanishes at the origin.*

*Consider the following six conditions on $f$:*

(i) *$f$ is convex,*
(ii) *$f$ is* COA,
(iii) *$f$ is starshaped,*
(iv) *$f$ is superadditive,*
(v) *$f$ is starshaped on the average,*
(vi) *$f$ is superadditive on the average.*

*Then the following chain of implications is valid but none of the reverse implications holds*: (i) $\to$ (ii) $\to$ (iii) $\to$ (iv) $\to$ (v) $\to$ (vi).

*Proof.* (i) $\to$ (ii). This will be a consequence of Theorem 10.

(ii) $\to$ (iii).

$$\frac{f(x)}{x} = F'(x) + \frac{F(x)}{x} .$$

Since $F$ is convex, both $F'$ and $F(x)/x$ are increasing. Thus $f(x)/x$ is increasing. It follows from Lemma 3, condition (i), that $f$ is starshaped.

(iii) $\to$ (iv). For $x > 0$ and $y > 0$, we have

$$\frac{f(x)}{x} \leqq \frac{f(x+y)}{x+y}$$

and

$$\frac{f(y)}{y} \leqq \frac{f(x+y)}{x+y} .$$

These inequalities are equivalent to

$$(x+y)f(x) \leqq xf(x+y)$$

and

$$(x+y)f(y) \leqq yf(x+y)$$

which on addition yield $f(x)+f(y) \leqq f(x+y)$.

(iv) $\rightarrow$ (v). We first consider the case in which $f$ is a *polygonal* superadditive function. The general case then follows by a limit argument.

Let $x > 0$ and let $f$ be polygonal of $n$ segments with vertices over the equidistantly spaced points $0, v, 2v, \cdots, nv = x$. Let $T$ be the linear function determined by the origin and the point $(x, f(x))$, i.e. $T(t) = (f(x)/x)t$ for all $t$. Furthermore, let $q(t) = f(t) - T(t)$. The function $q$ is polygonal and superadditive, having its vertices over the same points as $f$, and $q(0) = q(x) = 0$. We will show that $\int_0^x q(t)\,dt \leqq 0$ which suffices for $F$ to be starshaped. Using the linearity of $f$ on the intervals $[kv, (k+1)v]$, we obtain

$$\begin{aligned}\int_0^x q(t)\,dt &= v\sum_{k=1}^{n-1} q(kv)\\ &= \begin{cases} v\displaystyle\sum_{k=1}^{(n-1)/2}[q(kv)+q(n-k)v)] & \text{if } n \text{ is odd ,}\\ vq((n/2)v) + v\displaystyle\sum_{k=1}^{n/2-1}[q(kv)+q((n-k)v)] & \text{if } n \text{ is even .}\end{cases}\end{aligned}$$

Now $q(kv) + q((n-k)v) \leqq q(nv) = q(x) = 0$ for $q$ is superadditive. In either case $\int_0^x q(t)\,dt \leqq 0$.

In the general case let $\{p_n\}$ be a sequence of polygonal functions over equidistantly spaced points such that $p_n \rightarrow f$. Let $T$ be the linear function defined as above related to $f$. Since $\{p_n\}$ is superadditive for each $n$ (see Bruckner [2, THEOREM 8] and $p_n(x) \leqq f(b)$ for all $n$ and all $x \leqq b$, where $b$ is arbitrary, it follows for each $x$ that

$$\int_0^x p_n(t)\,dt \rightarrow \int_0^x f(t)\,dt .$$

Since

$$\int_0^x p_n(t)\,dt \leqq \int_0^x T(t)\,dt\ ,$$

the limit result

$$\int_0^x f(t)\,dt \leqq \int_0^x T(t)\,dt$$

follows.

(v) $\rightarrow$ (vi). This is just the case (iii) $\rightarrow$ (iv) for $F$.

That none of the reverse implications hold is shown by the following examples:

(ii) $\rightarrow$ (i): $f(x) = x^2 - x^3$ is COA on $[0, 4/9]$ but convex only on $[0, 1/3]$.

(iii) $\rightarrow$ (ii):

$$f(x) = \begin{cases} x^2 & 0 \leqq x \leqq 1 \\ x & 1 < x \end{cases}$$

is starshaped on $[0, \infty)$ but COA only on $[0, 1]$.

(iv) $\rightarrow$ (iii): $f(x) = n + (x - n)^2$ for $n \leqq x < n + 1$, $(n = 0, 1, 2, \cdots)$ is superadditive on $[0, \infty)$ but starshaped only on $[0, 1]$.

(v) $\rightarrow$ (iv): Let $f$ be any function that is starshaped on the average without being increasing.

(vi) $\rightarrow$ (v): Let $F$ be any superadditive function which is not starshaped such that $F'$ is continuous. Then $xF(x)$ has a continuous derivative $f(x)$ and $F$ is the average function of $f$.

**4. Behavior for large and small $x$.** Our first theorem in this section shows that superadditive functions are differentiable at the origin. Actually, a weaker hypothesis suffices to give this result.

THEOREM 6. *Let $f$ be a continuous nonnegative function on $[0, c]$, $f(0) = 0$, such that $f((1/n)x) \leqq (1/n)f(x)$ for all $n = 1, 2, 3, \cdots$, and for all $x \in [0, c]$. Then $f$ is differentiable at $x = 0$.*

*Proof.* The hypothesis $f((1/n)x) \leqq (1/n)f(x)$ implies that

$$\overline{\lim_{x\to 0}}\, \frac{f(x)}{x} < \infty\ .$$

Suppose $f$ is not differentiable at the origin. Then there exists an $\varepsilon > 0$ such that $\bar{f}'(0) - \underline{f}'(0) = 3\varepsilon$. Choose $x_0$ so that $f(x_0) < (\underline{f}'(0) + \varepsilon)x_0$ and let $\{y_k\}$ be a sequence such that $y_k \rightarrow 0$ and $f(y_k) > (\bar{f}'(0) - \varepsilon)y_k$ $(k = 1, 2, 3, \cdots)$. Since $f$ is continuous at $x_0$, there is a $\delta > 0$ such that if $|x - x_0| < \delta$, then $f(x) < (\underline{f}'(0) + \varepsilon)x$. Let $y^*$ be a member of the sequence $\{y_k\}$ such that $y^* < \delta$. There is then an integer $N$ such that

$$|\,Ny^* - x_0\,| < \delta;\ \text{hence } f(Ny^*) < (\underline{f}'(0) + \varepsilon)Ny^*\ .$$

However

$$f(y^*) \leqq \frac{1}{N}f(Ny^*) < (\underline{f}'(0) + \varepsilon)y^* < (\bar{f}'(0) - \varepsilon)y^*$$

which contradicts the fact $y^*$ is a member of the sequence $\{y_k\}$. Thus $f$ is differentiable at the origin.

COROLLARY. *If $f$ is superadditive, in particular if $f$ is starshaped,* COA, *or convex, then $f'(0)$ exists.*

THEOREM 7. *Let $f$ be superadditive on the average, and let $F$ be its average function. If $f'(0)$ exists, then $F'$ is continuous at $x = 0$ and $f'(0) = 2F'(0)$.*

*Proof.*

$$F'(x) = \frac{f(x)}{x} - \frac{F(x)}{x}, \qquad x > 0\ .$$

The right member of this equality approaches $f'(0) - F'(0)$ for $F'(0)$ exists by Theorem 6; hence $\lim_{x\to 0} F'(x)$ exists, and because $F'$ is a derivative, this limit must be $F'(0)$. Thus, $F'$ is continuous at $x = 0$ and $2F'(0) = f'(0)$.

Theorem 7 indicates that $2F(x)/x$ is approximately the same as $f(x)/x$ for $x$ near 0, provided $f$ behaves sufficiently well near the origin. The next theorem shows that under suitable hypotheses the same behavior holds for large $x$.

THEOREM 8. *Let $f$ be increasing and starshaped on the average and let $F$ be its average function. Then $\lim_{x\to\infty} f(x)/x$ exists and is equal to $2\lim_{x\to\infty} F(x)/x$.*

*Proof.* Since $F$ is starshaped, the $\lim_{x\to\infty} F(x)/x$ exists.

Let $\alpha$ be such that $0 < \alpha < 1$ and let $M = \overline{\lim}_{x\to\infty}(f(x)/x)$. Then

$$\begin{aligned}\frac{F(x)}{x} &= \frac{1}{x^2}\int_0^x f(t)\,dt \\ &= \frac{1}{x^2}\int_0^{\alpha x} f(t)\,dt + \frac{1}{x^2}\int_{\alpha x}^x f(t)\,dt\end{aligned}$$

$$\geqq \frac{\alpha F(\alpha x)}{x} + \frac{1-\alpha}{x} f(\alpha x) .$$

It follows that

$$\lim_{x\to\infty} \frac{F(x)}{x} \geqq \alpha^2 \lim_{x\to\infty} \frac{F(x)}{x} + \alpha(1-\alpha)M .$$

This last inequality holds for all $\alpha$, $0 < \alpha < 1$ so

$$\lim_{x\to\infty} \frac{F(x)}{x} \geqq \sup_{0<\alpha<1} \frac{\alpha M}{1+\alpha} = \frac{M}{2} .$$

On the other hand, since $F$ is starshaped, $f(x) \geqq 2F(x)$ for all $x$ so that

$$\lim_{x\to\infty} \frac{F(x)}{x} \leqq \frac{1}{2} \varliminf_{x\to\infty} \frac{f(x)}{x} .$$

It follows that $\lim_{x\to\infty} (f(x)/x)$ exists and equals $2 \lim_{x\to\infty} (F(x)/x)$.

COROLLARY. *Let $f$ be increasing and starshaped on the average with average function $F$. Then the three functions $f(x)/x$, $F'(x)$, and $F(x)/x$ simultaneously are bounded or unbounded.*

*Proof.* This follows directly from the identity

$$\frac{f(x)}{x} = F'(x) + \frac{F(x)}{x}$$

and the preceding theorem.

**5. Minimal extensions.** We suppose in this section that $f$ is defined initially on an interval $[0, c]$. We shall consider in this section the problem of extending $f$ in a minimal way to $[0, \infty)$ while staying within the same class. We start with

DEFINITION 4. Let $f$ be convex (COA, starshaped, superadditive) on $[0, c]$. Suppose $\hat{f}$ is a function defined on $[0, \infty)$ with the following properties:

(i) $\hat{f} = f$ on $[0, c]$,

(ii) $\hat{f}$ is convex (COA, starshaped, superadditive) on $[0, \infty)$,

(iii) if $g$ is any function on $[0, \infty)$ satisfying (i) and (ii), then $g(x) \geqq \hat{f}(x)$ for all $x$;

then $\hat{f}$ is said to be the minimal convex (COA, starshaped, superadditive) extension of $f$.

We restrict our definition to functions which are at least superadditive for minimal extensions of functions in the larger two classes are not, in general, continuous.

It is well known that if $f$ is convex on $[0, c]$, there exists a convex extension of $f$ to $[0, \infty)$ precisely when $f'_-(c) < \infty$. In this case, the minimal convex extension of $f$ is linear on $[c, \infty)$ with slope $f'_-(c)$. When $f$ is starshaped, it is clear that the minimal starshaped extension of $f$ to $[0, \infty)$ is the linear function with slope $f(c)/c$. For superadditive functions the situation is much more complicated and has been studied in detail in Bruckner [2], where it is shown that the minimal extension does exist and, roughly speaking, behaves about as well as $f$.

The following theorem states the corresponding result for functions that are COA on $[0, c]$.

THEOREM 9. *Suppose $f$ is* COA *on $[0, c]$ with average function $F$. Define $\hat{f}$ by the equations*

$$\hat{f}(x) = \begin{cases} f(x) & 0 \leqq x \leqq c \\ 2F'_-(c)x + f(c) - 2F'_-(c)x & x > c\ ; \end{cases}$$

*then $\hat{f}$ is the minimal* COA *extension of $f$ to $[0, \infty)$. If $\hat{F}$ is the average function of $\hat{f}$, then $\hat{F}$ is the minimal convex extension of $F$ to $[0, \infty)$.*

*Proof.* For $x \geqq c$, we have

$$\hat{F}(x) = \frac{1}{x}\int_0^c f(t)\,dt + \frac{1}{x}\int_c^x [2F'_-(c)t + f'(c) - 2F'_-(c)c]\,dt\ .$$

It is easy to check that $\hat{F}(c) = F(c)$ and that for $x > c$, $\hat{F}''(x) = 0$ and $\hat{F}'(x) = \hat{F}'_-(c)$ so that $\hat{F}$ is the minimal convex extension of $F$ to $[0, \infty)$. Thus $\hat{f}$ is a COA extension of $f$ to $[0, \infty)$. Let now $g$, with average function $G$, be any COA extension of $f$ to $[0, \infty)$ and let $x > c$. Since $G$ is convex, $G'$ is increasing so

$$G'(x) \geqq G'_-(c) = F'_-(c) = \hat{F}'(x)\ .$$

Thus

$$\underline{g}'(x) \geqq 2G'(x) \geqq 2\hat{F}'(x) = \hat{f}'(x)\ .$$

Since $\hat{f}$ and $g$ agree at $c$ and $\underline{g}' \geqq \hat{f}'$, $g(x) \geqq \hat{f}(x)$ so $\hat{f}$ is indeed the minimal COA extension of $f$.

If a function is convex on $[0, c]$, then it has extensions of each of the four types mentioned above. It is interesting to compare these various extensions. As an example, consider the function $f(x) = x^2$ on $[0, 1]$. Its minimal convex extension is linear with slope 2, the minimal COA extension is linear with slope 4/3, and the minimal starshaped extension is linear with slope 1. In contrast, the minimal superadditive extension is not linear. It is given by the function $\hat{f}(x) = n + (x - n)^2$ for $n \leqq x < n + 1, n = 1, 2, 3, \cdots$ (see Bruckner [2], p 1155).

**6. Tests for convexity on the average.** In this section we shall consider conditions that are necessary and/or sufficient that a function be COA. Similar tests concerning superadditive functions are found in Bruckner [3]. We begin with the following lemma.

LEMMA 6. *Let $f_c$ be the function such that*

$$f_c(x) = \begin{cases} 0 & 0 \leqq x \leqq c \\ f(x - c) & x > c\,. \end{cases}$$

*If $f$ is* COA, *then $f_c$ is* COA.

*Proof.* Let $F_c$ be the average function of $f_c$. We shall show that $\underline{f}'_c(x) \geqq 2F'_c(x)$, $x \geqq c$. Since $\underline{f}'_c(x) = \underline{f}'(x - c)$ for $x \geqq c$, it suffices to show that $\underline{f}'(x - c) \geqq 2F'_c(x)$. This last inequality will be a consequence of the inequality $F'(x - c) \geqq F'_c(x)$.

Defining

$$A(x) = \int_0^{x-c} f(t)\,dt\,,$$

we have that

$$F(x - c) = \frac{1}{x - c}\int_0^{x-c} f(t)\,dt = \frac{A(x)}{x - c}$$

and

$$\begin{aligned} F_c(x) &= \frac{1}{x}\int_0^x f_c(t)\,dt = \frac{1}{x}\int_c^x f(t - c)\,dt \\ &= \frac{1}{x}\int_0^{x-c} f(t)\,dt = \frac{A(x)}{x}\,. \end{aligned}$$

It thus suffices to show that

$$[A(x)(x - c)^{-1}]' \geqq [A(x)x^{-1}]',$$

the "$'$" denoting differentiation with respect to $x$. This last inequality is equivalent to

$$A'(x) \geqq \frac{2x - c}{x(x - c)}A(x)\,,$$

which is, on replacing $A(x)$ by $\int_0^{x-c} f(t)\,dt$ and simplifying, equivalent to the relation

$$f(x-c) \geqq \frac{2x-c}{x} F(x-c) .$$

Since $f$ is starshaped, $f$ is superadditive; hence $f$ is starshaped on the average. Thus, by Lemma 4,

$$f(x-c) \geqq 2F(x-c) \geqq \frac{2x-c}{x} F(x-c)$$

which proves the lemma.

DEFINITION 5. Let $f$ be defined on [0, a]. The functions $f_1, f_2, \cdots, f_n$ defined on $[0, a_1], [0, a_2], \cdots, [0, a_n]$ respectively form a decomposition of $f$ provided

(i) $f_i(0) = 0 \qquad i = 1, \cdots, n$

(ii) $a_1 + a_2 + \cdots + a_n = a$ and $a_i > 0$ for $i = 1, \cdots, n$

(iii) $$f(x) = \begin{cases} f_1(x) & 0 \leqq x \leqq a_1 \\ f_2(x-a_1) + f_1(a_1) & a_1 < x \leqq a_1 + a_2 \\ \cdots\cdots & \\ f_n(x - a_1 - a_2 - \cdots - a_{n-1}) + f_1(a_1) + \cdots + f_{n-1}(a_{n-1}) & \\ & a - a_n < x \leqq a . \end{cases}$$

In this case we write $f = f_1 \wedge f_2 \wedge \cdots \wedge f_r$.

THEOREM 10. *Let $f_1$ and $f_2$ be* COA *on $[0, a_1]$ and $[0, a_2]$ respectively and let $f = f_1 \wedge f_2$ on $[0, a_1 + a_2]$. Let $\hat{f}_1$ be the minimal* COA *extension of $f_1$. A necessary and sufficient condition that $f$ be* COA *is that $f \geqq \hat{f}_1$ on $[0, a_1 + a_2]$.*

*Proof*. The necessity is obvious. As to the sufficiency let $\hat{F}_1$ be the average function of $\hat{f}_1$. For $x \in [0, a_1 + a_2]$, write

$$F(x) = \hat{F}_1(x) + [F(x) - \hat{F}_1(x)] = \hat{F}_1(x) + \frac{1}{x}\int_0^x [f(t) - \hat{f}_1(t)]\, dt .$$

Consider $g(t) = f(t) - \hat{f}_1(t)$. $g(t) = 0$ on $[0, a_1]$ so there is an $h$ defined on $[0, a_2]$ such that $g(t) = h_{a_1}(t)$ for $t \in [0, a_1 + a_2]$. On $[a_1, a_1 + a_2]$, $-\hat{f}_1$ is linear. Since $f_2$ is COA on $[0, a_2]$, $h$ is COA on $[0, a_2]$ being the sum of COA functions. It follows from Lemma 6 that $g$ is COA on $[0, a_1 + a_2]$. Its average function is therefore convex and so $F$ is the sum of convex functions; hence convex.

THEOREM 11. *Let $f_1, \cdots, f_n$ be* COA *on $[0, a_1], \cdots, [0, a_n]$ respectively and let $f = f_1 \wedge f_2 \wedge \cdots \wedge f_n$. Furthermore let $\hat{f}_k$ be the mimimal* COA *extension of $f_k$, $(k = 1, \cdots, n)$. Then $f$ is* COA *on $[0, a_1 + \cdots + a_n]$*

*if* $f_k \wedge f_{k+1} \wedge \cdots \wedge f_n \geqq \hat{f}_k$ *for each* $k = 1, 2, \cdots, n$.

*Proof.* The proof is an induction argument using the sufficiency part of Theorem 10.

We now return to the proof of the first part of Theorem 5, namely the proof of the statement: if $f$ is convex, then $f$ is COA.

*Proof.* Let us assume first that $f$ is a polygonal function on $[0, c]$. If $f$ has only one segment, then $f$ is linear so the theorem is trivially true.

Supposing, by induction, that the theorem holds for polygonal functions with $n$ segments, let $f$ be polygonal with $(n + 1)$ segments. Let $f_n$ be the polygonal function which agrees with $f$ on the first $n$ segments of $f$ and let $\hat{f}_n$ be the minimal convex extension of $f_n$ to $[0, c]$. Thus $\hat{f}_n$ is convex and polygonal with $n$ segments and so is COA. On the last segment $f$ is linear and $f \geqq \hat{f}_n$. By Theorem 10, $f$ is COA on $[0, c]$.

The general situation follows immediately by Theorem 3. Let $\{p_n\}$ be a sequence of convex polygonal functions approximating $f$. The $\{p_n\}$ are thus COA and so their limit function $f$ is COA on $[0, c]$. Since $c$ is arbitrary, this concludes the proof.

## BIBLIOGRAPHY

1. E. F. Beckenbach, *Convex functions*, Bull. Amer. Math. Soc., **54** (1948), 439-460.

2. A. M. Bruckner, *Minimal superadditive extensions of superadditive functions*, Pacific J. Math., **10** (1960), 1155-1162.

3. ———, *Tests for the superadditivity of functions*, Proc. Amer. Math Soc., **13** (1962), 126-130.

4. G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities*, 2nd Edition, Cambridge, 1952, p. 91.

5. E. Hille and R. S. Phillips, *Functional Analysis and semi-groups*, Amer. Math. Soc. Colloquium Publications, vol. XXXI, Ch. 7.

6. E. H. Hobson, *Theory of functions of a real variable*, vol. 1, 3rd edition, Cambridge 1927, p. 364.

7. R. A. Rosenbaum, *Subadditive functions*, Duke Math. J., **17** (1950), 227-247.

UNIVERSITY OF CALIFORNIA,
SANTA BARBARA

# LIMITS AND BOUNDS FOR DIVIDED DIFFERENCES ON A JORDAN CURVE IN THE COMPLEX DOMAIN

J. H. CURTISS

**1. Introduction.** Let $S_{n+1} = \{z_1, z_2, \cdots, z_{n+1}\}$ be a set of $n + 1$ complex numbers and let $f$ be a function on a set containing $S_{n+1}$ to the complex numbers. The divided difference $d_n = d_n(f \mid z_1, z_2, \cdots, z_{n+1})$ of order $n$ formed for the function $f$ in the points[1] $S_{n+1}$ is defined in a recursive manner as follows:

$$\begin{aligned} d_1 &= d_1(f \mid z_1, z_2) = \frac{f(z_1) - f(z_2)}{z_1 - z_2} \\ d_2 &= d_2(f \mid z_1, z_2, z_3) = \frac{d_1(f \mid z_1, z_2) - d_1(f \mid z_3, z_2)}{z_1 - z_3} \\ &\vdots \\ d_n &= d_n(f \mid z_1, z_2, \cdots, z_{n+1}) \\ &= \frac{d_{n-1}(f \mid z_1, z_2, \cdots, z_n) - d_{n-1}(f \mid z_{n+1}, z_2, \cdots, z_n)}{z_1 - z_{n+1}} . \end{aligned}$$

The definition requires further discussion when the points in $S_{n+1}$ are not all distinct. We shall suppose that they are distinct unless provision is explicitly made for coincidences.

It can be proved by induction [7, p. 15] that if

$$\omega_{n+1}(z) = (z - z_1)(z - z_2) \cdots (z - z_{n+1}) ,$$

then

$$d_n = \sum_{k=1}^{n+1} \frac{f(z_k)}{\omega'_{n+1}(z_k)} , \tag{1.1}$$

where the prime denotes differentiation of $\omega_{n+1}(z)$ with respect to $z$. This formula shows that $d_n$ is a symmetric function of $z_1, z_2, \cdots, z_{n+1}$.

The divided differences of a function given on the real line play a prominent role in the mathematics of computation. Their counterparts in the complex plane have appeared in various classical studies of approximation by complex polynomials. The formal algebra of complex

Received January 10, 1962. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49 (638)-862.

[1] We use the words "points" and "numbers" interchangeably in referring to the arguments in divided differences. This follows the practice in interpolation theory. It is consistent within this terminology to speak of "coincident points" $z_k$.

divided differences is of course much the same as for the real case, but the analytical properties of complex divided differences, such as asymptotic behavior and representability by integrals, are in some cases quite different. It would appear that these analytical properties have not received much attention in the literature, although some of them seem interesting.

A primary motivation for the present paper was the need to establish that under certain smoothness hypotheses on a function $f$ given on a Jordan curve $C$, the divided difference of $f$ of a fixed order formed in points on $C$ is uniformly bounded in modulus for all choices of the points. This property was required in a study of complex interpolation in random points [**1**]. The existence of the bound is proved in § 2 below for the case in which $C$ is the unit circle. The extension to more general Jordan curves appears in § 3. In § 4, the asymptotic behavior of successive divided differences of order $n$ formed in $n+1$ points on a Jordan curve, $n = 1, 2, \cdots$, which in their totality become everywhere dense in a certain way on the curve, is investigated. It is found that the behavior to be expected in cases important in the theory of complex interpolation is that the $n$th divided difference multiplied by the $(n+1)$th power of the transfinite diameter, or capacity, of the curve $C$ in question approaches the limit $\int_C f\,dz/(2\pi i)$. Section 4 is essentially self-contained and can be read separately.

As was mentioned above, the impetus for this study came from a particular application. It is hoped that the results may turn out to be useful in other directions. However, the general spirit in which this paper is written is that of interest in the subject for itself alone, and the possible applications will not be considered further.

**2. An upper bound for the modulus of a divided difference formed on the unit circle.** If the numbers $z_k$ are all real numbers, and if $f$ is continuous on a closed interval $\bar{I}$ of the real line containing $S_{n+1}$ and possesses an $n$th derivative $f^{(n)}$ at each point of the corresponding open interval $I$, then by elementary calculus [7, p. 24] it can be shown that there exists a number $x_0$ in $I$ such that $d_n = f^{(n)}(x_0)/n!$. Thus if $|f^{(n)}|$ is uniformly bounded everywhere on $I$, so also is $|d_n|$ for all choices of $S_{n+1}$ on $\bar{I}$. Again, with real points $S_{n+1}$, if $f^{(n-1)}$ is absolutely continuous [8, pp. 364 ff.] on $\bar{I}$, then the iterated integral on the right side of the following formula (in which we define $z_{n+2}$ as meaning $z_1$),

$$(2.1)\qquad d_n(f \mid z_1, z_2, \cdots, z_{n+1}) = \int_0^1\int_0^{y_1}\cdots\int_0^{y_{n-1}} f^{(n)}\Big[z_2 + \sum_{j=1}^{n} y_j(z_{j+2} - z_{j+1})\Big]dy_1\,dy_2 \cdots dy_n\,,$$

has meaning for all choices of $S_{n+1}$ in which the points $z_j$ are distinct, and indeed can be used to extend the definition of $d_n$ to cases involving confluent points. It is easily shown by induction that the formula is true [7, pp. 17–18]. Thus it follows with $f^{(n-1)}$ absolutely continuous and $|f^{(n)}|$, where it exists, uniformly bounded on $I$, that $|d_n|$ is also uniformly bounded for all choices of $S_{n+1}$ on $\bar{I}$ such that completion of the definition of $d_n$ through (2.1) is possible. If $M$ is the least upper bound of $|f^{(n)}|$ on $I$, then $|d_n| \leqq M/n!$.

The formula (2.1) is no longer generally valid when the numbers $S_{n+1}$ are not all real, and the derivation of a bound for $|d_n|$ in terms of a given bound for $|f^{(n)}|$ is not so readily accomplished. In the remainder of the section we shall consider this problem in the case in which $S_{n+1}$ lies on the unit circle in the complex plane.

In the development, we shall use a complex-variable type of interpretation of the derivatives of a function $g$ given on the circle $C: |z|=1$. The symbol $g^{(1)}(z_1)$ will mean

$$\lim_{z \to z_1} d_1(g \mid z, z_1) = \lim_{z \to z_1} \frac{g(z) - g(z_1)}{z - z_1}, \quad |z| = 1, |z_1| = 1,$$

provided of course that the limit exists. Higher derivatives $g^{(k)}$ are to be defined recursively. The circle $C$ can be parametrized in a one-to-one manner by the equation $z = e^{i\theta}$, with $\alpha \leqq \theta < \alpha + 2\pi$, where $\alpha$ is chosen arbitrarily. If this is done, then

$$g^{(1)}(z) = \frac{dg}{d\theta} \cdot \frac{d\theta}{de^{i\theta}} = \frac{dg}{d\theta} \cdot \frac{1}{ie^{i\theta}}.$$

The chief result is this:

THEOREM 2.1 *Let the function $f$ be given on $C: |z| = 1$ together with its first $n-1$ derivatives $f^{(1)}, f^{(2)}, \cdots, f^{(n-1)}$. Let the points $S_{n+1}$ lie on $C$ and be distinct. Then if $f^{(n-1)}$ satisfies the Lipschitz condition:*

$$|f^{(n-1)}(z) - f^{(n-1)}(t)| \leqq \lambda |z - t|, \lambda > 0,$$

*for all $z$ and $t$ on $C$, it follows that*

$$(2.2) \quad |d_n(f \mid z_1, z_2, \cdots, z_{n+1})| \leqq \frac{\frac{\pi}{2}\left(\frac{\sqrt{\pi}}{2}\right)^{n-1}\lambda_n}{\Gamma\left(\frac{n+1}{2}\right)}$$

*uniformly for all such $S_{n+1}$, where $\lambda_n$ is the least upper bound of*

$$|f^{(n)}(z)|, |z| = 1.$$

The symbol $\Gamma$ in (2.2) refers to the Gamma Function.

The hypothesis on $f^{(n-1)}$ has the implication that $f^{(n-1)}$ be absolutely continuous on $|z|=1$, and that it therefore be the indefinite integral of a derivative $f^{(n)}$ existing everywhere on $|z|=1$ with the possible exception of a set of Lebesgue measure zero. Moreover the implication is that $|f^{(n)}|$, on the set where it exists, is bounded and its least upper bound $\lambda_n$ does not exceed $\lambda$.

Our proof involves integral representations, and it is important to be explicit about the integral calculus to be used.

Consider two points $e^{i\alpha_2}$ and $e^{i\alpha_1}$ on the unit circle. The complex line integral of a function $g$ given on the unit circle extended over either one of the two arcs of the circle joining these points, directed *from* $e^{i\alpha_2}$ *to* $e^{i\alpha_1}$, is to be defined as a Lebesgue integral with respect to the parameter $\theta$ in the parametrization $z=e^{i\theta}$. That is, if $A$ is the chosen directed arc, then

$$\int_A g(z)dz = \int_{\alpha_2}^{\alpha_1} g(e^{i\theta})ie^{i\theta}d\theta \ . \tag{2.3}$$

If $g$ is continuous in a neighborhood of $e^{i\alpha_1}$, then

$$\frac{d}{de^{i\alpha_1}}\int_A g(z)dz = \frac{g(e^{i\alpha_1})ie^{i\alpha_1}}{ie^{i\alpha_1}} = g(e^{i\alpha_1}) \ . \tag{2.4}$$

The notation for the integral on the right side of (2.3) is ambiguous in that it does not indicate which one of the two possible directed arcs $A$ is being integrated over. However in the sequel we shall be dealing only with complex line integrals on $|z|=1$ which are independent of the path of integration. Such an integral extended over either arc directed from $z_2$ to $z_1$, $|z_1|=|z_2|=1$, will be denoted by

$$\int_{z_2}^{z_1} g(z)dz \ .$$

If the two arcs joining $z_2$ to $z_1$ are are of equal length, then $\alpha_2=\alpha_1\pm\pi$, so the variation of $\theta$ in (2.3) is over a closed interval of length $\pi$ of which one endpoint is $\alpha_2$. If the two paths are not of equal length, then the shorter one corresponds under $z=e^{i\theta}$ to an interval of values of $\theta$ of which one endpoint is $\alpha_2$ and the other one, say $\alpha'$, is such that $z_1=e^{i\alpha_1}=e^{i\alpha'}$ and $|\alpha_2-\alpha'|<\pi$. (For example, if $\alpha_1$ and $\alpha_2$ are restricted to the interval $[0, 2\pi]$ and if $\alpha_2>\alpha_1$, $\alpha_2-\alpha_1>\pi$, then we take $\alpha'=2\pi+\alpha_1$.)

We shall now drop the parentheses around superscripts indicating derivatives of functions, but it is to be understood that superscripts can also be exponents when the context requires, as in $(z-z_2)^k$. In the case of divided differences, a derivative superscript will always indicate a partial derivative with respect to the first argument when the notation in the first paragraph of the Introduction is being used. That is

$$d_m^k = d_m^k(f \mid z_1, z_2, \cdots, z_{m+1}) \frac{\partial^{(k)} d_m(f \mid z_1, z_2, \cdots, z_{m+1})}{\partial z_1^{(k)}} .$$

To prove Theorem 2.1 we need two lemmas, of which the first is as follows.

LEMMA 2.1 *Let the function $f$ given on $C\colon |z| = 1$ be such that its* $(n-1)$st *derivative exists everywhere on $C$ and is absolutely continuous. Then*

$$(2.5) \qquad d_1^{h-1}(f \mid z_1, z_2) = \frac{\int_{z_2}^{z_1} (t - z_2)^{h-1} f^h(t) dt}{(z_1 - z_2)^h} ,$$

$$|z_1| = 1, \quad |z_2| = 1, \quad z_1 \neq z_2, \qquad h = 1, 2, \cdots, n .$$

*The integral is independent of the path of integration on $C$.*

The absolute continuity of $f^{n-1}$ implies the absolute continuity of $f, f^2, \cdots, f^{n-2}$, and so implies that each of these functions including $f^{n-1}$ is the indefinite integral of its derivative.

In the case $h = 1$, with $z_1 = e^{i\alpha_1}, z_2 = e^{i\alpha_2}$,

$$\frac{\int_{z_2}^{z_1} f'(t)\, dt}{z_1 - z_2} = \frac{\int_{\alpha_2}^{\alpha_1} f'(e^{i\theta}) i e^{i\theta} d\theta}{z_1 - z_2}$$

$$= \frac{f(z_1) - f(z_2)}{z_1 - z_2}$$

$$= \frac{\int_{\alpha_2}^{\alpha_1 \pm 2\pi} f'(e^{i\theta}) i e^{i\theta} d\theta}{z_1 - z_2} .$$

The second and fourth members of the equation show that whether $\alpha_1 > \alpha_2$ or $\alpha_1 < \alpha_2$, the integral is independent of the path. Thus the Lemma is true for $h = 1$.

Suppose now that (2.5) gives a valid representation of $d_1^{k-1}, 1 \leqq k < h$, with the integral independent of the path and $z_1 \neq z_2$. Then using (2.4), we have after a brief computation

$$(2.6) \qquad d_1^k = \frac{\partial d_1^{k-1}}{\partial z_1} \frac{(z_1 - z_2)^k f^k(z_1) - k \int_{z_2}^{z_1} (t - z_2)^{k-1} f^k(t) dt}{(z_1 - z_2)^{k+1}} .$$

Because of the absolute continuity of $f^k$, integration by parts is valid in the integral in (2.6), with $f^k$ to be differentiated and $(t-z_2)^{k-1}$ to be integrated with respect to $t$. We thereby immediately obtain (2.5) with $h = k+1$, and the integral is again independent of the path. The premises of the induction are true for $k = 1$, so this establishes the Lemma.

LEMMA 2.2 *Let the function $g$ be given at all points on $|z| = 1$ with the possible exception of a set of Lebesgue measure zero; let $|g(z)| \leqq M$ where defined on $|z| = 1$, and let $g$ be such that*

$$I_k(z_1, z_2) = \int_{z_2}^{z_1} (t-z_2)^k g(t)\,dt, \quad |z_1| = |z_2| = 1, \quad k \geqq 0\ ,$$

*is independent of the path of integration. Then*

$$\left|\frac{I_k(z_1, z_2)}{(z_1-z_2)^{k+1}}\right| \leqq \frac{\sqrt{\pi}}{2} \cdot \frac{\Gamma\left(\dfrac{k+1}{2}\right)}{\Gamma\left(\dfrac{k+2}{2}\right)} M, \ k = 0, 1, 2, \cdots\ , \tag{2.7}$$

*for all $z_1$ and $z_2$, $z_1 \neq z_2$, on $|z| = 1$.*

When $k = 0$, the right side of (2.7) reduces to $\pi M/2$.

For the proof, we make the shorter arc joining $z_2 = e^{i\alpha_2}$ and $z_1 = e^{i\alpha_1}$ (or either of the two arcs if they are equal in length) correspond under $z = e^{i\theta}$ to a $\theta$-interval $[\alpha_2, \alpha']$ or $[\alpha', \alpha_2]$, where $\alpha'$ is such that $z_1 = e^{i\alpha'}$ and $|\alpha_2 - \alpha'| \leqq \pi$. Thus

$$I_k(z_1, z_2) = \int_{\alpha_2}^{\alpha'} (e^{i\theta} - e^{i\alpha_2})^k g(e^{i\theta}) ie^{i\theta} d\theta\ .$$

For the case $k = 0$, we use the inequality $|\sin\theta| \geqq |2\theta/\pi|$, $-\pi/2 \leqq \theta \leqq \pi/2$, which is merely an expression of the fact that $\sin\theta$ is convex on $[0, \pi/2]$. We also use the identity $e^{i\alpha} - e^{i\beta} = 2ie^{i(\alpha+\beta)/2} \sin[(\alpha-\beta)/2]$. Then since $|\alpha'-\alpha_2|/2 \leqq \pi/2$, it follows that $|z_1 - z_2| = 2\sin[(\alpha'-\alpha_2)/2]| \geqq (2/\pi)\,|2(\alpha'-\alpha_2)/2|$. Now $|I_0(z_1, z_2)| \leqq M|\alpha' - \alpha_2|$, so the inequality (2.7) follows at once for $k = 0$.

For $k \geqq 1$ we have (recalling the restriction on $|\alpha' - \alpha_2|$),

$$\begin{aligned} \frac{|I_k(z_1, z_2)|}{|z_1 - z_2|^{k+1}} &\leqq \frac{M2^k\left|\displaystyle\int_{\alpha_2}^{\alpha'}\left|\sin^k\left(\frac{\theta-\alpha_2}{2}\right)\right| d\theta\right|}{2^{k+1}\left|\sin^{k+1}\dfrac{\alpha'-\alpha_2}{2}\right|} \\ &= \frac{M}{2}\, \frac{\left|\displaystyle\int_{\alpha_2}^{\alpha_1} \sin^k\left|\frac{\theta-\alpha_2}{2}\right| d\theta\right|}{\sin^{k+1}\left|\dfrac{\alpha'-\alpha_2}{2}\right|}\ . \end{aligned} \tag{2.8}$$

We make the substitution $\beta = |\theta - \alpha_2|/2$ in the integral and let $\gamma = |\alpha' - \alpha_2|/2 \leqq \pi/2$. By examination of the various cases corresponding to $k$ even or odd and $\alpha' < \alpha_2$, $\alpha' > \alpha_2$ we find that the righthand member of (2.8) is always equal to

$$M \int_0^{\gamma} \frac{\sin^k \beta \, d\beta}{\sin^{k+1}\gamma} = MS(\gamma) .$$

Inspection of its derivative shows that $S(\gamma)$ increases steadily with $\gamma$ on the interval $0 \leqq \gamma \leqq \pi/2$. The value of $S(\pi/2)$ is given by the well known formula

$$\int_0^{\pi/2} \sin^k \theta d\theta = \frac{\sqrt{\pi}}{2} \frac{\Gamma\left(\frac{k+1}{2}\right)M}{\Gamma\left(\frac{k+2}{2}\right)} .$$

Thus

$$\frac{|I_k(z_1, z_2)|}{|z_1 - z_2|^{k+1}} \leqq MS\left(\frac{\pi}{2}\right) = \frac{\sqrt{2}}{2} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k+2}{2}\right)} M ,$$

as was to be proved.

Now let $g$ in Lemma 2.2 be $f^n$ and $M$ be $\lambda_n$, where $f$ is the function appearing in the Theorem. The two lemmas establish that

$$d_1^{n-1}(f \mid z_1, z_2) = \frac{\lambda_n \sqrt{\pi}\, \Gamma\left(\frac{n}{2}\right)}{2\Gamma\left(\frac{n+1}{2}\right)} , \quad z_1 \neq z_2 . \tag{2.9}$$

As a function of $z_1$, $d_1^{n-1}$ is continuous for $z_1 \neq z_2$ and uniformly bounded in modulus. Consider next $d_1^{n-2}(f\, z_1, z_2)$ as a function of $z_1$. This function has a continuous derivative for $z_1 \neq z_2$, which is moreover uniformly bounded in modulus. Therefore this function is absolutely continuous in $z_1$ for $z_1$ on any closed arc of the unit circle not containing $z_2$. But the uniform boundedness of $|d_1^{n-1}|$ implies that $d_1^{n-2}(f \mid z_1, z_2)$ is of uniformly bounded variation in $z_1$ on the entire unit circle with the point $z_2$ deleted. By a well-known theorem [8, p. 372, Ex. 6] it follows that as $z_1$ approaches $z_2$ from either side, $d_1^{n-2}$ approaches a limit; and if the limit is the same for approach from either side, then when the definition of $d_1^{n-2}$ is completed by this limit at $z_1 = z_2$ the function $d_1^{n-2}$ will be an absolutely continuous function for all $z_1$ on $|z_1| = 1$. To investigate the limit, we write (2.5) in the form

$$d_1^{n-2} = \left[\frac{\displaystyle\int_{\alpha_2}^{\alpha_1}(e^{i\theta} - e^{i\alpha_2})^{n-2}f^{n-1}(e^{i\theta})ie^{i\theta}d\theta}{\left(\sin\dfrac{\alpha_1-\alpha_2}{2}\right)^{n-1}}\right]\cdot\frac{1}{(2ie^{i(\alpha_1+\alpha_2)/2})^{n-1}}\,.$$

The limit as $z_1 \to z_2$, $\alpha_1 \to \alpha_2$, of the expression in square brackets can be evaluated by l'Hospital's rule used with (2.4) and with the fact that $f^{n-1}$ is continuous. We find that for $z_1$ approaching $z_2$ on either side, there is the unique limit

$$\lim_{z_1\to z_2} d_1^{n-2}(f \mid z_1, z_2) = \frac{f^{n-1}(z_2)}{n-1}\,.$$

Thus with proper completion of the definition of $d_1^{n-2}$ at $z_1 = z_2$, this function is an absolutely continuous function of $z_1$ on $|z_1| = 1$. Similarly we can complete the definition of $d_1^h(f \mid z_1, z_2)$ for $h = n-3, n-4, \cdots, 1, 0$, so that the resulting function is in each case an absolutely continuous function of $z_1$ on $|z_1| = 1$. *We assume henceforth without change in notation that for each relevant value of $h$, the proper extension of the definition of $d_1^h$ at $z_1 = z_2$ has been made.*

What this establishes is that the completed first order divided difference $d_1(f \mid z_1, z_2)$, as a function of $z_1$, together with its first $n-2$ partial derivatives with respect to $z_1$, have the same smoothness and integrability properties as does $f$ and its first $n-1$ derivatives. That is to say, $d_1, d_1^2, \cdots, d_1^{n-2}$ are absolutely continuous functions of $z_1$ and moreover the derivative of $d_1^{n-2}$, where it exists, is uniformly bounded in modulus.

The absolute continuity of the derivatives permits the inductive argument which we used to establish (2.5) to be used again to prove that

$$d_2^{h-1}(f \mid z_1, z_2, z_3) = \frac{\partial^{h-1}}{\partial z_1^{h-1}} d_1(d_1(f \mid z, z_2) \mid z_1, z_3)$$

$$= \frac{\displaystyle\int_{z_3}^{z_1}(t - z_3)^{h-1}d_1^h(f \mid t, z_2)\,dt}{(z_1 - z_3)^h}$$

$$|z_1| = 1,\ |z_2| = 1,\ z_1 \neq z_3,\ h = 1, 2, \cdots, n-1\,.$$

By (2.9) and Lemma 2.2, $d_2^{n-2}$ as a function of $z_1$ is uniformly bounded in modulus. (It is not important at the moment to know how the bound depends on $z_2$ and $z_3$.) The definitions of $d_2^{n-3}, d_2^{n-4}, \cdots, d_2$ can now be completed by continuity at $z_1 = z_3$ so that in each case the resulting function of $z_1$ is absolutely continuous on $|z_1| = 1$. Again we assume without change of notation that the proper extensions have been made.

Proceeding in this way, we establish the chain of equations

$$d_k^{n-k} = \frac{\int_{z_{k+1}}^{z_1} (t - z_{k+1})^{n-k} d_{k-1}^{n-k+1}(f \mid t, z_2, \cdots, z_k)\, dt}{(z_1 - z_{k+1})^{n-k+1}} \tag{2.10}$$

$$k = 1, 2, \cdots, n\,,$$

in which $d_n^0 = d_n$, $d_0^n = f^n$. Theorem 2.1 now can be proved by back substitution into (2.10), beginning with (2.9) and using Lemma 2.2 at each stage. Thus to start with, at least for $z_1 \neq z_2$ and $z_1 \neq z_3$,

$$| d_2^{n-2} | \leqq \frac{\sqrt{\pi}}{2} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \left[ \frac{\lambda_n \sqrt{\pi}\, \Gamma\left(\frac{n}{2}\right)}{2\Gamma\left(\frac{n+1}{2}\right)} \right].$$

Similarly,

$$| d_3^{n-3} | \leqq \left( \frac{\sqrt{\pi}\, \Gamma\left(\frac{n-2}{2}\right)}{2\Gamma\left(\frac{n-1}{2}\right)} \right) \left( \frac{\sqrt{\pi}\, \Gamma\left(\frac{n-1}{2}\right)}{2\Gamma\left(\frac{n}{2}\right)} \right) \left[ \frac{\lambda_n \sqrt{\pi}}{2} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} \right]$$

$$= \left( \frac{\sqrt{\pi}}{2} \right)^3 \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n+1}{2}\right)} \lambda_n\,,$$

and so forth. We finally find that

$$| d_n^0 | \leqq \frac{\pi}{2} \left( \frac{\sqrt{\pi}}{2} \right)^{n-1} \frac{\lambda_n}{\Gamma\left(\frac{n+1}{2}\right)}\,,$$

as stated in the conclusion of the Theorem.

It is clear from the proof that under the hypotheses of Theorem 2.1 on $f$, it is possible to extend the definition of $d_n$ by continuity so as to admit point sets $S_{n+1}$ in which coincidences occur, and then (2.2) will still be valid. However we shall not study this question in detail under the hypotheses of Theorem 2.1 on $f$, which were chosen as being natural to achieve boundedness of $| d_n |$ in the case of distinct points. (The boundedness of $d_1 = | f(z_1) - f(z_2) | / | z_1 - z_2 |$ is equivalent to a Lipschitz condition on $f$.)

The method of proof with only slight modifications can be used to establish the following result:

THEOREM 2.2 *Under the hypothesis of Theorem* 2.1 *concerning* $f$, *and with the added hypothesis that $f^{(n)}$ is continuous on some open arc of $|z|=1$ containing the point $z_1$, the following equation completes the definition of $d_n$ by continuity for the case in which all the points $S_{n+1}$ coincide at $z_1$:*

$$(2.11) \qquad d_n(f \mid z_1, z_1, \cdots, z_1) = \frac{f^{(n)}(z_1)}{n!} .$$

*If $f^{(n)}$ is everywhere continuous on $|z|=1$, then* (2.2) *is valid after proper completion of the definition of $d_n$ for all choices of $S_{n+1}$ without restrictions as to coincidences.*

We conclude this section with two comments. In the first place it is clear that by repeated back-substitution into (2.10), a single formula for $d_n$ in terms of $f^{(n)}$ involving repeated integration can be written out. It would be somewhat similar in appearance to a variant of (2.1) which appears in [5, p. 18, ex. 7].

In the second place, it may be that for $n \geqq 2$ the bound in (2.2) can be improved. For $n=1$ it is the best bound possible, as can be seen from this trivial example: Let $f$ be real and let its graph over a period in the $(\theta, f(e^{i\theta}))$-plane be a line segment joining $(0,0)$ to $(\pi, \pi)$ and another line segment joining $(\pi, \pi)$ to $(2\pi, 0)$. For this function the maximum of $d_1$ is $\pi/2$, the least upper bound of $|f'|$ is one, and the right side of (2.2) is $(\pi/2)\cdot 1$, which is as small as it can be. However the general bound was derived through Lemma 2.2 in which the two points $z_1$ and $z_2$ were placed at opposite ends of a diameter to obtain the numerical appraisal. Such wide-apart spacing is of course not possible for the case of three or more points on the unit circle. The bound given by (2.1) in the real case under the hypothesis of Theorem 2.1 is $\lambda_n/n!$, which is much smaller than that in (2.2).

**3. Boundedness of the modulus of a divided difference formed on a general Jordan curve.** A generalization of Theorem 2.1 to the case in which the unit circle is replaced by a more general Jordan curve is not hard to derive. In doing so, for simplicity we shall not try to keep track of the structure of the upper bound, and shall suppress various details in the proof.

A Jordan curve is homeomorphic to a circle. It can be represented by a parametric equation $z = \phi(\theta)$, where $\phi$ is continuous in the real variable $\theta$ with period $2\pi$, and where for each given point $z$ on the curve, any two solutions of $z = \phi(\theta)$ differ by an integral multiple of $2\pi$. Our considerations here will be restricted to Jordan curves such that the first derivative $d\phi/b\theta = \psi(\theta)$ exists for all $\theta$ and is continuous, and $\psi(\theta) \neq 0$ for all $\theta$. Such a Jordan curve will be said to be "admissible". (Presumably in what follows the definition of admissibility can be slightly relaxed.)

LEMMA 3.1. *If $z = \phi(\theta)$ is a parametric equation of an admissible Jordan curve, then there exist numbers $m$ and $M$, $0 < m < M$, such that*

$$m \leqq \left| \frac{\phi(\theta_1) - \phi(\theta_2)}{e^{i\theta_1} - e^{i\theta_2}} \right| \leqq M$$

*for all $\theta_1$ and $\theta_2$.*

The divided difference appearing in the above inequality is to be interpreted as meaning $\psi(\theta_2)/ie^{i\theta_2}$ when $\theta_1 = \theta_2$.

The existence of an upper bound $M$ follows from Lemma 2.2 with $k = 0$ and $g(e^{i\theta}) = \psi(\theta)/ie^{i\theta}$. The existence of the lower bound can be established by an elementary indirect argument which we omit[2].

As in the unit circle case, it is convenient to interpret the derivatives of a function on a Jordan curve to the complex numbers as limits of complex-variable difference quotients. Specifically for any function $g$ given on a Jordan curve $C$, the symbol $g'(z_1)$ means $\lim_{z \to z_1} d_1(g \mid z, z_1)$, $z$ and $z_1$ on $C$, and higher derivatives are to be defined recursively. If $C$ is admissible, then

$$g'(z) = \frac{d(g(\phi(\theta)))}{d\theta} \cdot \frac{1}{\psi(\theta)} .$$

Integrals are to be defined as in (2.3) with $ie^{i\theta}$ in that formula replaced by $\psi(\theta)$. With this replacement and with $e^{i\alpha_1}$ replaced by $\phi(\alpha_1)$, (2.4) is is valid. An integral over $C$ with limits of integration $z_1$ and $z_2$ which is independent of the path will be written as

$$\int_{z_2}^{z_1} g(z)\, dz .$$

The notation implies of course that the arc over which the integration takes place is directed from $z_2$ to $z_1$. The derivatives of divided differences are always partials with respect to the first apparent argument.

The generalization of Theorem 2.1 is as follows:

THEOREM 3.1. *Let the function $f$ be given on an admissible Jordan curve $C$,* together with $f^1, f^2, \cdots, f^{n-1}$. *Let $f^{n-1}$ satisfy the Lipschitz condition*

$$| f^{n-1}(z) - f^{n-1}(t) | \leqq \lambda \,| z - t |, \lambda > 0 ,$$

*for all $z$ and $t$ on $C$. Let the points $S_{n+1} = \{z_1, z_2, \cdots, z_{n+1}\}$ lie on $C$ and be distinct. Then there exists a constant $M$ depending only on $n$, $\lambda$, and $C$, and independent of $S_{n+1}$, such that $| d_n(f \mid z_1, z_2, \cdots, z_{n+1}) | \leqq M$.*

[2] Various related but deeper results may be found in [**6**, Section 2.5].

The proof starts with a generalization of Lemma 2.1.

LEMMA 3.2. *Let the function $f$, given on an admissible Jordan curve $C$, be such that its $(n-1)$th derivative exists everywhere on $C$ and is absolutely continuous as a function of $\theta$, $z = \phi(\theta)$. Then*

$$d_1^{h-1}(f \mid z_1, z_2) = \frac{\int_{z_2}^{z_1}(t - z_2)^{h-1} f^h(t)\, dt}{(z_1 - z_2)^h},$$

*for $z_1$ and $z_2$ on $C$, $z_1 \neq z_2$, $h = 1, 2, \cdots, n$, and the integral is independent of the path of integration on $C$.*

The argument used to establish Lemma 2.1 carries over to Lemma 3.2 with only minor changes, and will not be restated here.

LEMMA 3.3. *Let the function $G$ be given on the admissible Jordan curve $C$ with the possible exception of a set of Lebesgue measure zero; let $|G|$ be bounded on $C$ and be such that*

$$J_k(z_1, z_2) = \int_{z_2}^{z_1}(t - z_2)^k G(t)\, dt$$

*is independent of the path of integration on $C$ for all $z_1$ and $z_2$ on $C$. Then for each $k$, $k = 0, 1, \cdots$, there exists a constant $M_k$, depending only on $G$ and $C$, and such that*

$$\left| \frac{J_k(z_1, z_2)}{(z_1 - z_2)^{k+1}} \right| \leqq M_k$$

*for all $z_1$ and $z_2$ on $C$, $z_1 \neq z_2$.*

The Lebesgue measure in the theorem means measure on the $\theta$-line after the transformation $\phi^{-1}: z \to \theta$.

To prove this lemma, we let $z_1 = \phi(\alpha_1)$, $z_2 = \phi(\alpha_2)$, $t = \phi(\theta)$, and write

$$(3.1) \qquad \frac{J_k(z_1, z_2)}{(z_1 - z_2)^{k+1}}$$
$$= \left[\frac{e^{i\alpha_1} - e^{i\alpha_2}}{\phi(\alpha_1) - \phi(\alpha_2)}\right]^{k+1} \frac{\int_{\alpha_2}^{\alpha_1}(e^{i\theta} - e^{i\alpha_2})^k g(e^{i\theta}) i e^{i\theta}\, d\theta}{(e^{i\alpha_1} - e^{i\alpha_2})^{k+1}},$$

where

$$(3.2) \qquad g(e^{i\theta}) i e^{i\theta} = G(\phi(\theta)) \left[\frac{\phi(\theta) - \phi(\alpha_2)}{e^{i\theta} - e^{i\alpha_2}}\right]^k \psi(\theta).$$

By Lemma 3.1, the quantities in the square brackets in (3.1) and (3.2) are both uniformly bounded in modulus for all $\theta$, $\alpha_1$, and $\alpha_2$, $\theta \neq \alpha_2$, $\alpha_1 \neq \alpha_2$. For any fixed $\alpha_2$, $g(e^{i\theta})$ as given by (3.2) is integrable, and its modulus is uniformly bounded for all $\theta$ and $\alpha_2$. The integral in (3.1), considered as an integral over an arc of the unit circle, is independent of the path of integration. Thus the hypotheses of Lemma 2.2 are satisfied by $g$ as given by (3.2). The truth of Lemma 3.3 now follows immediately.

Theorem 3.1 can now be proved by the use of Lemmas 3.2 and 3.3 in the same way that Theorem 2.1 was proved. The hypotheses on $f$ and $C$ imply that $f^{n-1}(e^{i\theta})$ is an absolutely continuous function of $e^{i\theta}$ and of $\theta$, and that its derivative with respect to $\phi(\theta)$, where it exists, is uniformly bounded in modulus. The same is true for its derivative with respect to $\theta$. (These facts follow from the existence of numbers $\lambda_1$ and $\lambda_2$ such that with $z = e^{i\theta}$, $t = e^{i\alpha}$,

$$|f^{n-1}(z) - f^{n-1}(t)| \leqq \lambda\, |\phi(\theta) - \phi(\alpha)| \leqq \lambda_1\, |e^{i\theta} - e^{i\alpha}| \leqq \lambda_2\, |\theta - \alpha|\,.$$

(Here we used Lemma 3.1 in passing from the second member to the third member of the chain.) The functions $f, f^1, \cdots, f^{n-2}$ are also absolutely continuous and have uniformly bounded derivatives.

We can now re-establish the recursion formulas (2.10), which look exactly the same as before and so will not be repeated here. The integrals in (2.10) are of course now complex line integrals over $C$. Thereafter by back-substitution, using Lemma 3.3 at each stage, we establish the existence of the bound for $|d_n|$.

The analogous generalization of Theorem 2.2 is also valid. The proper definition of $d_n$ for confluent points is again given by (2.11). It is worth noting that what gives simplicity to our results and minimizes the restrictions on $C$ is the complex-variable type of definition which we are using for derivatives of functions given on $C$.

**4. Some asymptotic properties of divided differences formed on a Jordan curve.** In this section we shall be considering an infinite sequence of divided differences

$$d_1(f\,|\,z_{11}, z_{12}),\ d_2(f\, z_{21}, z_{22}, z_{23}),\ \cdots,\ d_n(f\,|\,z_{n1}, z_{n2}, \cdots, z_n, z_{u+1}),\ \cdots\ ,$$

formed for a function $f$ given on a Jordan curve $C$ in the $z$-plane. Do there exist sequences of point sets $S_{n+1} = \{z_{n1}, z_{n2}, \cdots, z_{n,\ n+1}\}$, $n=1, 2, \cdots$ such that $\lim_{n\to\infty} d_n$ exists for all functions $f$ belonging to an interestingly wide class; and if so, what is this limit?

Let $D$ be the region interior to the curve $C$, let $K$ be the unlimited region exterior to $C$, and let $\bar{D}$ be $D \cup C$. There exists an analytic function

$$z = \chi(w) = cw + c_0 + \frac{c_1}{w} + \frac{c_2}{w^2} + \cdots, c > 0\,, \tag{4.1}$$

univalent for $|w| > 1$, which maps $|w| > 1$ conformally onto $K$ so that the points at infinity in the $z$-plane and $w$-plane correspond. According to the Osgood-Taylor-Carathéodory Theorem, $\chi(w)$ can be extended in a continuous and one-to-one manner onto $|w| = 1$, and $\chi(e^{i\theta}) = \phi(\theta)$ then gives a parametric equation for $C$ of the type considered above in § 3. The number $c$ is called the transfinite diameter (Robin's constant, capacity) of $C$.

If a function $f$ is analytic on $C$, it is also analytic in a region (perhaps multiply connected) which contains $C$ in its interior. Let $w = Re^{i\theta}$ in (4.1). There is a largest value of $R$, say $\rho > 1$, such that $f$ is analytic at every point of the intersection of $C \cup K$ with the region interior to the Jordan curve $C_R$: $z = \chi(Re^{i\theta})$, $0 \leqq \theta \leqq 2\pi$. (See [9, p. 79].) A curve such as $C_R$ is called a level curve of the map given by (4.1).

With $z_{nk} = \chi(e^{i\theta_{nk}})$, $0 \leqq \theta_{nk} < 2\pi$, let $N_n(\theta)$ be the number of elements of the set $\{\theta_{n1}, \theta_{n2}, \cdots, \theta_{n,n+1}\}$ falling into the closed interval $[0, \theta]$. The numbers $\theta_{nk}$, $k = 1, \cdots, n + 1$, $n = 1, 2, \cdots$ are said to be equidistributed on $[0, 2\pi]$ if $\lim_{n\to\infty} N_n(\theta)/(n + 1) = \theta/2\pi$; and when this happens, the corresponding sequence of point sets $S_{n+1}$, $n = 1, 2, \cdots$, is said to be equidistributed on $C$.

Our first result is as follows:

THEOREM 4.1. *Let $f$ be analytic on $\bar{D}$ and let the sequence $\{S_{n+1}\}$ be equidistributed on $C$. Let $\rho$ be the largest value of $|w|$ in the map (4.1) such that $f$ is analytic interior to the level curve $C_{|w|}$. Then for any $R$, $1 < R < \rho$, there exists a constant $M$ depending on $f$, $R$, and $C$, but not on $n$, such that*

$$|d_n(f \mid z_{n1}, \cdots, z_{n,n+1})| \leqq \frac{M}{(cR)^{n+1}} .$$

*Thus* $\lim_{n\to\infty} c^{n+1} d_n = 0$.

To prove this, we use the formula [5, p. 11]

$$d_n = \frac{1}{2\pi i}\int_{C_R} \frac{f(t)}{\omega_{n+1}(t)}\, dt, \qquad 0 < R < \rho , \tag{4.2}$$

where $\omega_{n+1}(z) = (z - z_{n1})(z - z_{n2}) \cdots (z - z_{n,n+1})$. This formula can be used to complete the definition of $d_n$ by continuity for the case of confluent points $z_{nk}$. We then refer to a classical result of L. Fejér [4], [9, pp. 167 ff]: *If $\{S_{n+1}\}$ is equidistributed on a Jordan curve $C$, then*

$$\lim_{n\to\infty} |\omega_{n+1}(z)|^{1/(n+1)} = c\,|w|, z = \chi(w) ,$$

*uniformly for $z$ on any closed subset of $K$.* This implies that if $z$ lies on $C_R$ and $R_1$ is such that $1 < R_1 < R$, then for all $n$ sufficiently large,

$$\frac{c^{n+1}}{|\,\omega_{n+1}(z)\,|} \leqq \frac{1}{R_1^{n+1}}\ .$$

Letting $M_R$ be the maximum of $|f(z)|$ on $C_R$, and $L_R$ be the length of $C_R$, we appraise (4.2) as follows:

$$c^{n+1}\,|\,d_n \leqq \frac{M_R L_R}{2\pi R_1^{n+1}} \tag{4.3}$$

and the theorem follows from this.

THEOREM 4.2. *Let $C$ be rectifiable, let $f$ be analytic on $C$, and let $S_{n+1}$ be the transform under* (4.1) *of $n+1$ distinct points equally spaced on $|\,w\,| = 1$. Then*

$$\lim_{n\to\infty} c^{n+1} d_n = \frac{1}{2\pi i}\int_C f(t)\,dt\ . \tag{4.4}$$

This is consistent with Theorem 4.1, because the sequence $\{S_{n+1}\}$ in Theorem 4.2 is equidistributed and the integral in (4.4) would be zero if $f$ were analytic on $\bar{D}$.

To prove the theorem we use a generalization of (4.2),

$$d_n = \frac{1}{2\pi i}\left(\int_{C_R} + \int_{C'}\right)\frac{f(t)}{\omega_{n+1}(t)}\,dt\ , \tag{4.5}$$

which is easily established by the calculus of residues. Here $C_R$, $R>1$, is a level curve of (4.1) and $C'$ is a suitably chosen rectifiable curve lying in $D$. The curves $C_R$ and $C'$ are chosen so that $f$ is analytic on the closed annular region bounded by $C_R$ and $C'$. Integration on $C'$ is in the opposite sense to that on $C_R$.

The appraisal given by (4.3) is valid for the first integral in (4.5), and it shows that this integral vanishes in the limit. The following result of the author [2] is available for the second integral: *From the hypotheses of Theorem* 4.2 *on $C$ and $S_{n+1}$ it follows that*

$$\lim\nolimits_{n\to\infty} \omega_{n+1}(z)/c^{n+1} = -1$$

*uniformly for $z$ on any closed subset of $D$.* This implies that

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{2\pi i}\int_{C'}\frac{c^{n+1}f(t)}{\omega_{n+1}(t)}\,dt \\
&\quad = \frac{1}{2\pi i}\int_{C'}\lim_{n\to\infty}\frac{c^{n+1}f(t)}{\omega_{n+1}(t)}\,dt \\
&\quad = -\frac{1}{2\pi i}\int_{C'} f(t)\,dt = \frac{1}{2\pi i}\int_C f(t)\,dt\ ,
\end{aligned}$$

which completes the proof.

Generalizations of the above theorems to the case in which $D$ is replaced by a finite number of mutually exterior Jordan regions can be developed by the methods to be found in Walsh's book [9, Chap. VII].

The results from which the above two theorems are derived were originally established in studying the convergence of sequences of polynomials found by interpolation to the function $f$ on $C$. Let $L_{n+1}(z) = L_{n+1}(f; z \mid S_{n+1})$ be the (unique) polynomial in $z$ of degree at most $n$ which is determined by the condition that it shall coincide with $f(z)$ at each of the points $S_{n+1}$, assumed to be distinct. Then from the standard formula

$$L_{n+1}(\alpha) = \omega_{n+1}(\alpha) \sum_{k=1}^{n+1} \frac{f(z_{nk})}{\omega'_{n+1}(z_{nk})(\alpha - z_{nk})}$$

it is seen by comparison with (1.1) that

$$L_{n+1}(g; \alpha \mid S_{n+1}) = \omega_{n+1}(\alpha)\, d_n(f \mid z_{n1}, \cdots, z_{n,n+1}) \tag{4.6}$$

where $f(z) = g(z)/(\alpha - z)$. The following result of the author [2], [3] is relevant: *Let the curve $C$ be such that $\chi'(w)$ is nonvanishing and of bounded variation for $|w| = 1$. Let $g$ be bounded and integrable in the sense of Riemann on $C$. Let the points $S_{n+1}$ be the transforms under* (4.1) *of distinct points equally spaced on the unit circle. Then*

$$\lim_{n\to\infty} L_{n+1}(g; \alpha \mid S_n) = \frac{1}{2\pi i}\int_C \frac{g(t)}{t - \alpha}\, dt$$

*uniformly for $\alpha$ on any closed subset of $D$.*

We now write (4.6) in the form

$$c^{n+1} d_n = \left[ - \frac{c^{n+1}}{\omega_{n+1}(\alpha)} \right] [- L_{n+1}(g; \alpha \mid S_{n+1})] \, . \tag{4.7}$$

If $\alpha$ is a fixed point of $D$ and $f$ is bounded and Riemann integrable on $C$, then so is $g$ and conversely. We recall that $-\omega_{n+1}(\alpha)/c^{n+1}$ tends to unity at each point of $D$ as $n$ becomes infinite. It follows from these facts that the limiting value of (4.7) is

$$\lim_{n\to\infty} c^{n+1} d_n = \frac{-1}{2\pi i}\int_C \frac{g(t)}{t - \alpha}\, dt = \frac{1}{2\pi i}\int_C f(t)\, dt \, .$$

We summarize formally:

THEOREM 4.3. *If the points $S_{n+1}$ are transforms under* (4.1) *of distinct points equally spaced on the unit circle, and if $C$ is such that $\chi'$ is nonvanishing and of bounded variation for $|w| = 1$, and if $f$ is* *bounded and integrable in the sense of Riemann on C, then*

$$\lim_{n\to\infty} c^{n+1} d_n = \frac{1}{2\pi i}\int_C f(t)\, dt \ .$$

## References


1. J. H. Curtiss, "*Interpolation in equidistributed points on the unit circle*", to appear in the Pacific Journal of Mathematics.

2. ———, *Riemann sums and the fundamental polynomials of Lagrange interpolation*" Duke Math. **8** (1941), 634-646.

3. ———, *Interpolation in regularly distributed points*, Trans. Amer. Math. Soc., **38** (1935), 458-473.

4. L. Fejér, "*Interpolation und konforme Abbildung, Gottinger* Nachrichten, (1918), 319-331.

5. Milne-Thomson, L. M., *The Calculus of Finite Differences*, MacMillan and Co., London, 1933.

6. W. E. Sewell, *Degree of Approximation by Polynomials in the Complex Domain*, Princeton University Press, Princeton, 1942.

7. J. F. Steffensen, *Interpolation*, Williams and Wilkins, Baltimore, 1927.

8. E. C. Titchmarsh, *Theory of Functions*, Oxford University Press, Oxford, 1932.

9. J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, Second edition, Amer. Math. Soc., Providence, 1956.


UNIVERSITY OF MIAMI

# DIMENSIONAL INVERTIBILITY

P. H. DOYLE AND J. G. HOCKING

We report here upon another aspect of our continuing investigation of invertibility (see [5, 6]) and its applications in the theory of manifolds.

All spaces considered here are separable and metric.

A separable metric space $X$ will be said to be *$k$-invertible*, $0 \leqq k \leqq \dim X$, if for each nonempty open set $U$ and each compact proper subset $C$ of dimension $\leqq k$, there is a homeomorphism $h$ of $X$ onto itself such that $h(C)$ lies in $U$. Then we say that $X$ is *strongly $k$-invertible* if for each nonempty open set $U$ and each *closed* proper subset $C$ of dimension $\leqq k$, there is a homeomorphism $h$ of $X$ onto itself such that $h(C)$ lies in $U$.

Clearly, "strongly $k$-invertible" implies "$k$-invertible" and the two properties coincide in compact spaces. If $\dim X = n$, then "invertible" and "strongly $n$-invertible" are equivalent but, for instance, $E^n$ is $n$-invertible and not invertible. We remark that $k$-invertibility is a strong form of near-homogeneity and says that compact $k$-dimensional subsets are "small under homeomorphisms." In the case of an $n$-manifold, $k$-invertibility is equivalent to the condition that every compact set of dimension $k$ lie in an open $n$-cell.

We first collect some results on 0-invertible spaces, most of these results being simple generalizations of theorems to be found in [5]. The first of these requires no proof here.

THEOREM 1. *The orbit of any point in a 0-invertible space is dense in the space.*

THEOREM 2. *Each orbit in a 0-invertible space is itself 0-invertible.*

*Proof.* Let 0 be the orbit of any point in a 0-invertible space $X$. Let $U$ be an open subset of 0 and $C$ be a compact 0-dimensional proper subset of 0. Then there is an open set $V$ in $X$ such that $V \cap 0 = U$ and, by 0-invertibility, there is a space homeomorphism $h$ such that $h(C)$ lies in $V$. But by definition of 0 as an orbit, $h(C)$ also lies in 0, hence $h(C)$ lies in $V \cap 0 = U$.

COROLLARY. *Each 0-invertible space is a union of disjoint, dense homogeneous, 0-invertible subspaces.*

Received February 21, 1962.

THEOREM 3. *If $X$ is* 0-*invertible and contains a nondegenerate connected open set, then $X$ is connected.*

*Proof.* If $U$ is a nondegenerate open connected set in $X$, let $p$ be any point in $U$.

For each point $x$ in $X$, there is a space homeomorphism $h_x$ such that $h_x(x \cup p) = h_x(x) \cup h_x(p)$ lies in $U$. Thus $X$ is a union $\bigcup_x h_x^{-1}(U)$ of connected sets, each containing the point $p$.

COROLLARY. *If $X$ is* 0-*invertible and is locally connected at any point, then $X$ is connected or $X$ is the* 0-*sphere.*

THEOREM 4. *If $X$ is* 0-*invertible and is locally Euclidean at any point, then $X$ is a manifold.*

*Proof.* If $X$ contains an open cell $U$ as an open set, then $X$ is connected by Theorem 3 and, as in the proof of Theorem 3, $h_x^{-1}(U)$ is an open cell neighborhood of the point $x$ for each point $x$ in $X$.

THEOREM 5. *If $X$ is strongly* 0-*invertible and contains an open set with compact closure, then $X$ is compact.*

*Proof.* Let $U$ be an open set in $X$ with compact closure $\bar{U}$. Given any infinite set $A$ in $X$ such that $A$ has no limit point, the set $A$ contains an infinite sequence $\{a_n\}$ having no limit point in $X$. But then the sequence $\{a_n\}$ can be carried into $U$ by a space homeomorphism $h$ in view of strong 0-invertibility. In $\bar{U}$, the sequence $\{h(a_n)\}$ has a limit point. This contradiction shows that $X$ is compact.

COROLLARY. *A locally compact, strongly* 0-*invertible space is compact.*

Every 2-manifold is 0-invertible and every compact 2-manifold is strongly 0-invertible because any compact 0-dimensional set in a 2-manifold lies in an arc in the manifold. In higher dimensions, however, 0-invertibility has more force. The following result is an interesting characterization of the 3-sphere.

THEOREM 6. *A strongly* 0-*invertible* 3-*manifold is $S^3$.*

*Proof.* We employ the characterization of R. H. Bing [1] and show that every polygonal simple closed curve in such a 3-manifold lies in an open 3-cell. Let $M^3$ be a strongly 0-invertible 3-manifold and let $J$ be a polygonal simple closed curve in $M^3$. A sufficiently thin tubular neighborhood of $J$ may be chosen to be a polyhedral solid torus $T$ in $M^3$. Since every longitudinal simple closed curve in $T$ is isotopic to $J$, if we can show that there is such a curve which lies in an open 3-cell, the proof will be complete.

Using the solid torus $T$ as the 0th stage, we construct a "necklace of Antoine" $N$ in $M^3$. By the assumption of 0-invertibility, the compact 0-dimensional set $N$ lies in an open 3-cell in $M^3$. Hence there is a standard decomposition $M^3 = P^3 \cup C$, where $P^3$ is an open 3-cell and $C$ is a nonseparating continuum of dimension $\leqq 2$ (see [7]), such that $N \cap C$ is empty. Since $N$ and $C$ are compact, there is a positive distance between $N$ and $C$. Thus there is some stage, say the $k$th, in the construction of $N$ such that the residual set $C$ fails to meet each solid torus in the $k$th stage.

Now we add a 2-disk spanning the hole in each solid torus in the $k$th stage of the construction of $N$. This results in a connected set consisting of alternately "orthogonal" disks with disjoint solid toroidal rims in the interior of each solid torus in the $(k-1)$st stage. Call these sets $L_i^{(k-1)}$, $i = 1, 2, \cdots, n^{k-1}$, where $n \geqq 3$. There are two cases to consider: (1) In each of the sets $L_i^{(k-1)}$ we can find a simple closed curve passing longitudinally around the hole in the corresponding solid torus in the $(k-1)$st stage and not meeting the residual set $C$ or (2) for some set $L_j^{(k-1)}$, $C$ meets every longitudinal simple closed curve on $L_j^{(k-1)}$.

In case (2), the residual set $C$ does not meet the solid toridal rims of the disks in $L_j^{(k-1)}$ but $C$ must meet at least one of the spanning disks in such a way that no arc from one solid torus of a linking pair to the other can be drawn in the spanning disk without meeting $C$. Thus $C$ must separate some spanning disk $D$ into components, one of which meets the solid torus spanned by $D$ and another of which meets one of the solid tori linked with that spanned by $D$. This is impossible. For, in such a case, any longitudinal simple closed curve in the linking solid torus would be linked with $C$ while lying in the complement of $C$ which contradicts the assumption that $M^3 - C = P^3$ is an open 3-cell.

Case (1) reduces to the following situation: Each solid torus in the $(k-1)$st stage of the construction of the necklace $N$ contains a longitudinal simple closed curve lying in the open 3-cell $P^3$ and these curves are linked just as are the solid tori in the $(k-1)$st stage. We can now replace the solid tori in the $(k-1)$st stage by thinner ones where necessary so that the entire $(k-1)$st stage lies in the open 3-cell $P^3$. The spanning disks are now added to these tori to obtain the sets $L_i^{(k-2)}$, $i = 1, 2, \cdots, n^{k-2}$, and the argument above can be repeated. The finite regression is now obvious. The contradiction in case (2) at each step forces us back to the first stage in the construction of the necklace $N$. But then the same argument produces a longitudinal simple closed curve $J'$ in the original solid torus $T$ such that $J' \cap C$ is empty. By our remark above $J$ and $J'$ are isotopic and since $J'$ lies in an open 3-cell, so does $J$.

COROLLARY. *Every polygonal simple closed curve in a 0-invertible 3-manifold lies in an open 3-cell.*

*Proof.* The argument for Theorem 6 goes through in this case, too, because the residual set $C$ is closed and there is still a positive distance between $C$ and a necklace $N$ in the complement of $C$.

Imposing a natural restriction upon the manifold permits us to generalize, not Theorem 6, but its corollary.

THEOREM 7. *In a 0-invertible, combinatorial $n$-manifold, every polygonal simple closed curve lies in an open $n$-cell. (Hence such manifords are simply connected.)*

*Proof.* Let $M^n$ be a 0-invertible, combinatorial $n$-manifold and let $J$ be a polygonal simple closed curve in $M^n$. In the combinatorial $n$-manifold, a sufficiently thin tubular neighborhood of $J$ will be a polyhedral solid $n$-torus $T$ (a homeomorph of the product of an $(n-1)$ disk and the unit circle). In the interior of $T$ we construct a Cantor set $N$ by the method of Blankenship [2]. Then, with the appropriate changes in dimension, the remainder of the proof is identical to that of Theorem 6.

A natural conjecture at this point concerns $k$-invertibility and the vanishing of the homotopy group $\pi_{k+1}(M^n)$. Such a conjecture is fruitless, however, in view of the following result.

THEOREM 8. *Let $A^{n+1} = S^n \times E^1$, $n \geqq 2$. Then $A^{n+1}$ is an $(n-1)$-invertible manifold (and clearly $\pi_n(A^{n+1})$ is not trivial).*

*Proof.* Assume that $A^{n+1}$ is imbedded in $E^{n+1}$ as the region between two concentric spheres. Then $\bar{A}^{n+1}$ is a closed annulus and there is a map $h$ from $\bar{A}^{n+1}$ onto $S^{n+1}$ such that $h \mid A^{n+1}$ is a homeomorphism and $h$ carries the two components of $\bar{A}^{n+1} - A^{n+1}$ into a pair of points $a$ and $b$.

If $N$ is any compact $(n-1)$-dimensional set in $A^{n+1}$, then $h(N)$ is a compact $(n-1)$-dimensional set in $S^{n+1} - (a \cup b)$. Since $h(N)$ does not separate $S^{n+1}$, there is a polygonal arc $J$ in $S^{n+1} - h(N)$ from $a$ to $b$ and $S^{n+1} - J$ is an $(n+1)$-cell. Whence $h^{-1}(S^{n+1} - J)$

is an $(n+1)$-cell in $A^{n+1}$ containing $N$ and therefore $A^{n+1}$ is $(n-1)$-invertible.

The next result is a slight generalization of our characterization theorem [4].

THEOREM 9. *The only strongly* $(n-1)$*-invertible* $n$*-manifold is* $S^n$.

*Proof*. If $M^n$ is strongly $(n-1)$-invertible, then $M^n$ is compact. Choose any standard decomposition $M^n = P^n \cup C$. Since $C$ is a continuum of dimension $\leqq n-1$ and $P^n$ is an open $n$-cell, there is a space homeomorphism carrying $C$ into $P^n$. Then Corollary 1 of Theorem 2 in [7] applies to show that $M^n$ is an $n$-sphere.

THEOREM 10. *The only* $(n-1)$*-invertible, noncompact* $n$*-manifold is* $E^n$.

*Proof*. Let $M^n$ be an $(n-1)$-invertible, noncompact $n$-manifold. Since $M^n$ is locally compact, it is a union $\bigcup_{j=1}^{\infty} A_j$ where we may choose $A_1$ to be a closed $n$-cell and where $A_j$ is compact and lies in the interior of $A_{j+1}$ for each $j$ (Theorem 2.60 of [8]). Let $U$ be an open $n$-cell in $A_1$ with bi-collored boundary. Each set $BdA_j$ has dimension $\leqq n-1$ and hence there is a homeomorphism $h_j$ of $M^n$ onto itself such that $h_j(BdA_j)$ lies in $U$.

We claim that $h_j(A_j)$ also lies in $U$. For $BdA_j$ separates $M^n$ and if $h_n(A_n)$ does not lie in $U$, then $h_n(M^n - A_n)$ must lie in $U$. But then $\overline{h_j(M^n - A_j)} = h_j(\overline{M^n - A_j})$ is compact whence $M^n = (M^n - A_j) \cup A_j$ is the union of two compact sets and is compact. This contradiction proves that $h_j(A_j)$ lies in $U$.

From here we see that $\{h_j^{-1}(U)\}$ is a sequence of open $n$-cells. We may select a monotone increasing subsequence inductively (or else all $A_j$ lie in some $h_j^{-1}(U)$ which completes the proof). Therefore $M^n$ is the union of a monotone increasing sequence of $n$-cells and, in view of [3], $M^n = E^n$.

To finish this report, we collect some immediate consequences of the Poincare duality and the Hurewicz theorem.

THEOREM 11. *Let* $M^n$ *be a compact, triangulated, orientable,* $k$*-invertible* $n$*-mainfold. Then the homotopy groups* $\pi_p(M^n)$ *are trivial for* $1 \leqq p \leqq k$.

COROLLARY 1. *If* $M^n$ *is as in Theorem* 11, *then* $M^n$ *has trivial integral homology groups in dimensions* $1, 2, \cdots, k$ *and* $n-k, \cdots,$

$n-1$.

COROLLARY 2. *If $M^n$ is as in Theorem 11, and if $k \geqq [n/2]$ (the largest integer in $n/2$), then $M^n$ is a homotopy sphere.*

Recent results of Stallings [9] and Zeeman [10] provide immediate proofs of the following result.

THEOREM 12. *A strongly $[n/2]$-invertible polyhedral n-manifold, $n \geqq 5$, is an n-sphere.*

## REFERENCES

1. R. H. Bing, *Necessary and sufficient conditions that a 3-manifold be $S^3$*, Annals of Math., **68** (1958). 17-37.
2. W. A. Blankenship, *Generalization of a construction of Antoine*, Annals of Math., **53** No. 2 (1951), 276-297.
3. M. Brown, *The monotone union of open n-cells is an open n-cell*, Proc. Amer. Math Soc., **12** No. 5 (1961), 812-814.
4. P. H. Doyle and J. G. Hocking, *A characterization of euclidean n-space*, Mich. Math. J., **7** (1960), 199-200.
5. ———, *Invertible spaces*, Amer. Math. Monthly (to appear)
6. ———, *Continuously invertible spaces*, Pacific J. Math. (to appear).
7. ———, *A decomposition theorem for manifolds*, Proc. Amer. Math. Soc. (to appear).
8. J. G. Hocking and G. S. Young, *Topology*, Addison-Wesley, Reading Mass (1961).
9. J. Stallings, *Polyhedral homotopy spheres*, Bull. Amer. Math. Soc., **66** (1960), 485-488.
10. E. C. Zeeman, *The generalized Poincare conjecture*, Bull. Amer. Math. Soc., **67** (1961), 270.

MICHIGAN STATE UNIVERSITY

# BLOCK DIAGONALLY DOMINANT MATRICES AND GENERALIZATIONS OF THE GERSCHGORIN CIRCLE THEOREM

DAVID G. FEINGOLD AND RICHARD S. VARGA

**1. Introduction.** The main purpose of this paper is to give generalizations of the well known theorem of Gerschgorin on inclusion or exclusion regions for the eigenvalues of an arbitrary square matrix $A$. Basically, such exclusion regions arise naturally from results which establish the nonsingularity of $A$. For example, if $A = D + C$ where $D$ is a nonsingular diagonal matrix, then Householder [7] shows that $\|D^{-1}C\| < 1$ in some matrix norm is sufficient to conclude that $A$ is nonsingular. Hence, the set of all complex numbers $z$ for which

$$\|(zI - D)^{-1}C\| < 1$$

evidently contains no eigenvalues of $A$. In a like manner, Fiedler [4] obtains exclusion regions for the eigenvalues of $A$ by establishing the nonsingularity of $A$ through comparisons with $M$-matrices.[1] Our approach, though not fundamentally different, establishes the nonsingularity of the matrix $A$ by the generalization of the simple concept of a diagonally dominant matrix. But one of our major results (§ 3) is that these new exclusion regions can give significant improvements over the usual Gerschgorin circles in providing bounds for the eigenvalues of $A$.

**2. Block diagonally dominant matrices.** Let $A$ be any $n \times n$ matrix with complex entries, which is partitioned in the following manner:

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,N} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,N} \\ \vdots & & & \vdots \\ A_{N,1} & A_{N,2} & \cdots & A_{N,N} \end{bmatrix}, \tag{2.1}$$

where the diagonal submatrices $A_{i,i}$ are square of order $n_i$, $1 \leqq i \leqq N$. For reasons to appear in § 3, the particular choice $N = 1$ of

$$A = [A_{1,1}] \tag{2.1'}$$

will be useful. Viewing the square matrix $A_{i,i}$ as a linear transformation of the $n_i$-dimensional vector subspace $\Omega_i$ into itself, we associate with this subspace the vector norm $\|\boldsymbol{x}\|_{\Omega_i}$, i.e., if $\boldsymbol{x}$ and $\boldsymbol{y}$ are elements of

Received April 11, 1962.

[1] For the definition of an $M$-matrix, see § 4 or [8].

$\Omega_i$, then

$$\begin{cases} \|\boldsymbol{x}\|_{\Omega_i} > 0 \text{ unless } \boldsymbol{x} = \boldsymbol{O} ; \\ \|\alpha\boldsymbol{x}\|_{\Omega_i} = |\alpha| \, \|\boldsymbol{x}\|_{\Omega_i} \text{ for any scalar } \alpha ; \\ \|\boldsymbol{x} + \boldsymbol{y}\|_{\Omega_i} \leqq \|\boldsymbol{x}\|_{\Omega_i} + \|\boldsymbol{y}\|_{\Omega_i} , \quad 1 \leqq i \leqq N . \end{cases} \tag{2.2}$$

The point here is that we can associate *different* vector norms with different subspaces $\Omega_i$. Now, similarly considering the rectangular matrix $A_{i,j}$ for any $1 \leqq i, j \leqq N$ as a linear transformation from $\Omega_j$ to $\Omega_i$, the norm $\|A_{i,j}\|$ is defined as usual by

$$\|A_{i,j}\| \equiv \sup_{x \in \Omega_j, x \neq O} \frac{\|A_{i,j}\boldsymbol{x}\|_{\Omega_i}}{\|\boldsymbol{x}\|_{\Omega_j}} . \tag{2.3}$$

Note that if the partitioning in (2.1) is such that all the matrices $A_{i,j}$ are $1 \times 1$ matrices and $\|\boldsymbol{x}\|_{\Omega_i} \equiv |x|$, then the norms $\|A_{i,j}\|$ are just the moduli of the single entries of these matrices. As no confusion arises, we shall drop the subscripts on the different vector norms.[2]

DEFINITION 1. Let the $n \times n$ matrix $A$ be partitioned as in (2.1). If the diagonal submatrices $A_{j,j}$ are nonsingular, and if

$$(\|A_{j,j}^{-1}\|)^{-1} \geqq \sum_{\substack{k=1 \\ k \neq j}}^{N} \|A_{j,k}\| \quad \text{for all } 1 \leqq j \leqq N , \tag{2.4}$$

then $A$ is *block diagonally dominant*, relative to the partitioning (2.1). If strict inequality in (2.4) is valid for all $1 \leqq j \leqq N$, then $A$ is *block strictly diagonally dominant*, relative to the partitioning of (2.1).

It is useful to point out that the quantity appearing on the lefthand side of (2.4) can also be characterized form (2.3) by

$$(\|A_{j,j}^{-1}\|)^{-1} = \inf_{x \in \Omega_j, x \neq O} \left( \frac{\|A_{j,j}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \right) , \tag{2.5}$$

whenever $A_{j,j}$ is nonsingular. With (2.5), we can then define $(\|A_{j,j}^{-1}\|)^{-1}$ by continuity to be zero whenever $A_{j,j}$ is singular.

In the special case that all the matrices $A_{i,j}$ are $1 \times 1$ matrices and $\|x\| = |x|$, then (2.4) can be written as

$$|A_{j,j}| \geqq \sum_{\substack{k=1 \\ k \neq j}}^{N} |A_{j,k}| \quad \text{for all } 1 \leqq j \leqq N , \tag{2.4'}$$

which is the usual definition of diagonal dominance.

As an example of a matrix which is block strictly diagonally dominant, consider the case $n = 4$, $N = 2$ of

[2] Later, we shall use the notation $\|\boldsymbol{x}\|_p$ to denote the $l_p$-norm $\|\boldsymbol{x}\|_p \equiv (\sum_i |x_i|^p)^{1/p}$.

$$A = \left[\begin{array}{cc|cc} 0 & 1 & 0 & 0 \\ 2/3 & 0 & 1/3 & 0 \\ \hline 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 1 & 0 \end{array}\right], \tag{2.6}$$

where we choose the vector norms $||x||_\infty \equiv \max_j |x_j|$. In this case,

$$(||A_{1,1}^{-1}||)^{-1} = (||A_{2,2}^{-1}||)^{-1} = \tfrac{2}{3}\,, \quad \text{and } ||A_{1,2}|| = ||A_{2,1}|| = \tfrac{1}{3}\,.$$

Obviously, $A$ is not diagonally dominant in the sense of (2.4′).

DEFINITION 2. The $n \times n$ partitioned matrix $A$ of (2.1) is *block irreducible* if the $N \times N$ matrix $B = (b_{i,j} \equiv ||A_{i,j}||)$, $1 \leqq i, j \leqq N$, is *irreducible*, i.e., the directed graph of $B$ is strongly connected.[3]

THEOREM 1. *If the partitioned matrix $A$ of* (2.1) *is block strictly diagonally dominant, or if $A$ is block irreducible and block diagonally dominant with inequality holding in* (2.4) *for at least one $j$, then $A$ is nonsingular.*

*Proof.* The extension to the case where $A$ is block irreducible and block diagonally dominant with strict inequality for at least one $j$ is easy, so we consider for simplicity only the case when $A$ is block strictly diagonally dominant. Suppose that $A$ is singular, i.e., there exists a nonzero vector $\boldsymbol{W}$ with

$$A\begin{bmatrix} W_1 \\ \vdots \\ W_N \end{bmatrix} = \boldsymbol{O}\,; \tag{2.7}$$

here, we have partitioned $\boldsymbol{W}$ conformally with respect to the partitioning of (2.1). But this is equivalent to

$$\sum_{\substack{j=1 \\ j\neq i}}^{N} A_{i,j} W_j = -A_{i,i} W_i\,, \qquad 1 \leqq i \leqq N\,. \tag{2.8}$$

Since $\boldsymbol{W}$ is a nonzero vector, normalize $\boldsymbol{W}$ so that $||W_j|| \leqq 1$ for all $1 \leqq j \leqq N$, and assume that equality is valid for some $r$, i.e., $||W_r|| = 1$ where $1 \leqq r \leqq N$. Thus, from (2.3)

[3] Equivalently, there exists *no* $N \times N$ permutation matrix $P$ such that $PBP^T = \left[\begin{smallmatrix} C & D \\ O & E \end{smallmatrix}\right]$, where $C$ and $E$ are square nonvoid submatrices. For strongly connected directed graphs, see for example [**6**].

$$(2.8') \qquad \| A_{r,r} W_r \| = \| \sum_{\substack{j=1 \\ j \neq r}}^{N} A_{r,j} W_j \| \leqq \sum_{\substack{j=1 \\ j \neq r}}^{N} \| A_{r,j} \| \cdot \| W_j \| \leqq \sum_{\substack{j=1 \\ j \neq r}}^{N} \| A_{r,j} \| .$$

But as $A_{r,r}$ is nonsingular by hypothesis, then putting $A_{r,r} W_r = Z_r$,

$$\| A_{r,r} W_r \| = \frac{\| A_{r,r} W_r \|}{\| W_r \|} = \frac{\| Z_r \|}{\| A_{r,r}^{-1} Z_r \|} \geqq (\| A_{r,r}^{-1} \|)^{-1} ,$$

using (2.3). This combined with (2.8′) gives a contradiction to the assumption (2.4) that $A$ is block strictly diagonally dominant, which completes the proof for the block strictly diagonally dominant case.

Actually, we can regard Theorem 1 as the block analogue of the well known Hadamard theorem on determinants, since Theorem 1 reduces to this result in the case that all the matrices $A_{i,j}$ of (2.1) are $1 \times 1$ matrices and $\| x \| \equiv | x |$. It should be pointed out that the result of Theorem 1 itself is a special case of a more general result by Ostrowski [10, Theorem 3, p. 185], and Fiedler [4].

As stated in the introduction, the above theorem leads naturally to a block analogue of the *Gerschgorin Circle Theorem.* If $I$ is the $n \times n$ identity matrix which is partitioned as in (2.1), and $I_j$ is the $n_j \times n_j$ identity matrix, suppose that

$$(2.9) \qquad (\| (A_{j,j} - \lambda I_j)^{-1} \|)^{-1} > \sum_{\substack{k=1 \\ k \neq j}}^{N} \| A_{j,k} \| \quad \text{for all } 1 \leqq j \leqq N .$$

Thus, we have from Theorem 1 that $A - \lambda I$ is nonsingular. Hence, if $\lambda$ is an eigenvalue of $A$, then $A - \lambda I$ cannot be block strictly diagonally dominant, which gives us

THEOREM 2. *For the partitioned matrix $A$ of* (2.1), *each eigenvalue $\lambda$ of $A$ satisfies*

$$(2.10) \qquad (\| (A_{j,j} - \lambda I_j)^{-1} \|)^{-1} \leqq \sum_{\substack{k=1 \\ k \neq j}}^{N} \| A_{j,k} \|$$

*for at least one $j$, $1 \leqq j \leqq N$.*

We again remark that if the partitioning of (2.1) is such that all the diagonal submatrices are $1 \times 1$ matrices and $\| x \| \equiv | x |$, then Theorem 2 reduces to the well known Gerschgorin Circle Theorem.

**3. Inclusion regions for eigenvalues.** In Theorem 2, we saw that each eigenvalue $\lambda$ of an arbitrary $n \times n$ complex matrix $A$ necessarily satisfied (2.10) for *at least* one $j$, $1 \leqq j \leqq N$.

DEFINITION 3. For the partitioned $n \times n$ matrix $A$ of (2.1), let the

*Gerschgorin set* $G_j$ be the set of all complex numbers $z$ such that

$$(3.1) \qquad (\|(A_{j,j} - zI)^{-1}\|)^{-1} \leqq \sum_{\substack{k=1 \\ k \neq j}}^{N} \|A_{j,k}\| , \qquad 1 \leqq j \leqq N .$$

Thus, from (2.5), we conclude that the Gerschgorin set $G_j$ always contains the eigenvalues of $A_{j,j}$, independent to the magnitude of the right side of (3.1) and independent of the vector norms used. Next, it is clear that each Gerschgorin set $G_j$ is closed and bounded. Hence, so is their union

$$(3.2) \qquad G = \bigcup_{j=1}^{N} G_j .$$

Thus, we can speak of the *boundary of* $G$, as well as the boundary of each $G_j$. By Theorem 2, all the eigenvalues of $A$ lie in $G$. Can any eigenvalue $\lambda$ of $A$ lie on the boundary of $G$? This can be answered trivially for the particular partitioning of (2.1′). In this case, the right-hand side of (3.1) is vacuously zero, and from (3.1), we see that the set $G$ is a *finite point set* consisting *only* of the eigenvalues of $A$. In this case, Theorem 2 gives *exact* information about the eigenvalues of $A$.

It is interesting that Theorem 2 can be strengthened by the assumption that $A$ is block irreducible, which is the analogue of a well known result of Taussky [11].

THEOREM 3. *Let the partitioned matrix $A$ of* (2.1) *be block irreducible, and let $\lambda$ be an eigenvalue of $A$. If $\lambda$ is a boundary point of $G$, then it is a boundary point of each set $G_j$, $1 \leqq j \leqq N$.*

*Proof.* Since $\lambda$ is an eigenvalue of $A$, then $\sum_{j=1}^{N} A_{i,j} W_j = \lambda W_i$, and if $\| W_j \| \leqq \| W_r \| = 1$, then as before

$$(3.3) \qquad (\|(A_{r,r} - \lambda I_r)^{-1}\|)^{-1} \leqq \sum_{\substack{j=1 \\ j \neq r}}^{N} \|A_{r,j}\| \, \|W_j\| \leqq \sum_{\substack{j=1 \\ j \neq r}}^{N} \|A_{r,j}\| .$$

But as $\lambda$ is a boundary point of $G$, equality must hold throughout (3.3), showing that $\lambda$ is a boundary point of $G_r$. Moreover, if $\|A_{r,j}\| \neq 0$, then $\|W_j\| = 1$, and we can repeat the argument with $r$ replaced by $j$. In this way, we conclude that $\lambda$ is a boundary point of $G_j$. From the irreducibility of $A$, the argument can be extended to every index $j$, $1 \leqq j \leqq N$, which completes the proof. A similar argument can be applied to complete the proof of Theorem 1.

Another familiar result of Gerschgorin can also be generalized. The proof, depending on a continuity argument, follows that given in [13, p. 287].

THEOREM 4. *If the union $H = \bigcup_{j=1}^{m} G_{p_j}$, $1 \leqq p_j \leqq N$, of $m$ Gerschgorin* *sets is disjoint from the remaining $N-m$ Gerschgorin sets for the partitioned matrix $A$ of* (2.1), *then $H$ contains precisely $\sum_{j=1}^{m} n_{p_j}$ eigenvalues of $A$.*

The previous example of the matrix of (2.1′) indicated that sharper inclusion regions for the eigenvalues of a matrix $A$ may be obtained from the generalized form of Gerschgorin's Theorem 2. To give another illustration, consider the partitioned matrix

$$A = \left[\begin{array}{cc|cc} 4 & -2 & -1 & 0 \\ -2 & 4 & 0 & -1 \\ \hline -1 & 0 & 4 & -2 \\ 0 & -1 & -2 & 4 \end{array}\right] = \left[\begin{array}{c|c} A_{1,1} & A_{1,2} \\ \hline A_{2,1} & A_{2,2} \end{array}\right]. \tag{3.4}$$

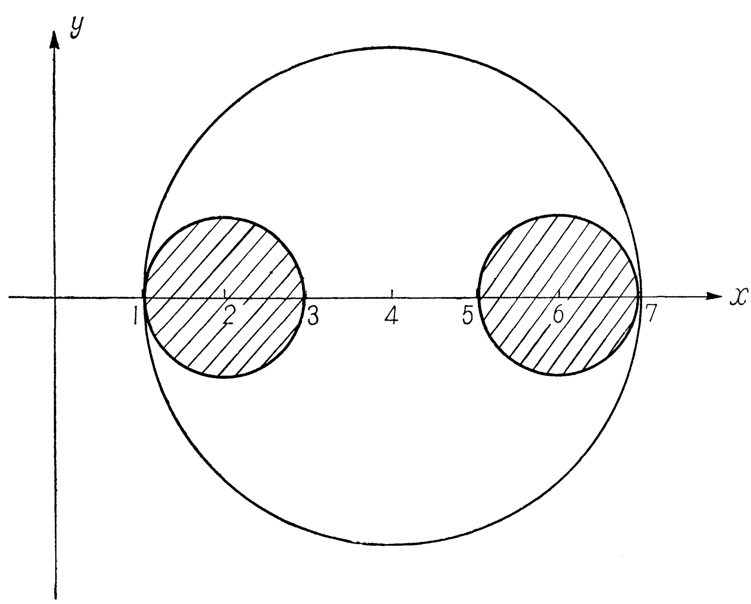
Employing now the vector norm $\|\boldsymbol{x}\|_2 \equiv (\sum_i |x_i|^2)^{1/2}$, it is apparent that $\|A_{1,2}\| = \|A_{2,1}\| = 1$. On the other hand, direct computation shows that

$$(\|(A_{i,i} - zI_i)^{-1}\|^{-1} = \min\{|6-z|, |2-z|\}, \qquad i = 1, 2.$$

By definition, the set $G_1$ then consists of the points $z$ for which

$$|6-z| \leqq 1, \qquad |2-z| \leqq 1,$$

so that $G_1$ is itself the union of two *disjoint* circles. The same is true for $G_2$, since $G_2 = G_1$, as shown in the figure below. The usual Gerschgorin circles



for the matrix $A$ of (3.4) are all given by the single circle $|4-\lambda| \leqq 3$, which is a circle of radius 3, with center at $z = 4$, as shown above. From this figure, we conclude that the block Gerschgorin result *can* give significant improvements over the usual Gerschgorin circles in providing bounds for eigenvalues. For the matrix $A$ of (3.4), its eigenvalues are

$$\lambda_1 = 1, \lambda_2 = 3, \lambda_3 = 5, \lambda_4 = 7 .$$

Note, again from the figure above, that Theorem 3 applies in this case.

At this point, we remark that the previous example was such that each Gerschgorin set $G_j$ consisted of the union of circles. This is a special case of

THEOREM 5. *Let the partitioned matrix $A$ of* (2.1) *be such that its diagonal submatrices $A_{j,j}$ are all normal. If the Euclidean vector norms $\|\boldsymbol{x}\|_2$ are used for each subspace $\Omega_j$, $1 \leqq j \leqq N$, then each Gerschgorin set $G_j$ is the union of $n_j$ circles.*

*Proof.* Let the eigenvalues of $A_{j,j}$ be $\sigma_l$, $1 \leqq l \leqq n_j$. Since $A_{j,j}$ is normal, we can write $(\|(A_{j,j} - zI_j)^{-1}\|)^{-1} = \min_l |\sigma_l - z|$, which, combined with Definition 3, completes the proof.

It is quite simple to obtain the block analogues of well known results on inclusion regions for eigenvalues of $n \times n$ complex matrices. As a first example, the result of *A.* Brauer [2] on ovals of Cassini easily carries over.

THEOREM 6. *Let the $n \times n$ complex matrix $A$ be partitioned as in* (2.1). *Then, all the eigenvalues of $A$ lie in the union of the* $[N(N-1)]/2$ *point sets $C_{i,j}$ defined by*

$$(3.5)\quad (\|(A_{i,i} - zI_i)^{-1}\| \cdot \|(A_{j,j} - zI_j)^{-1}\|)^{-1} \leqq \left(\sum_{\substack{l=1\\ l\neq i}}^{N} \|A_{i,l}\|\right)\left(\sum_{\substack{l=1\\ l\neq j}}^{N} \|A_{j,l}\|\right)$$

*where $1 \leqq i, j \leqq N$ and $i \neq j$. Moreover, if $A$ is block irreducible, and $\lambda$ is an eigenvalue of $A$ not in the interior of $\bigcup_{i\neq j} C_{i,j}$, then $\lambda$ is a boundary point of each of the point sets $C_{i,j}$.*

Other obvious remarks can be made. Clearly, replacing $A$ by $A^T$ leaves the eigenvalues of $A$ invariant. Thus, rows sums can be replaced by column sums in the definition (2.4) of diagonal dominance, and many results using both row and column sums admit easy generalizations. As an illustration, we include the following known [4] generalization of a result by Ostrowski [9].

THEOREM. 7. *Let the $n \times n$ complex matrix $A$ be partitioned as in* (2.1), *and define*

$$R_j \equiv \sum_{\substack{k=1 \\ k \neq j}}^{N} \|A_{j,k}\| ; \quad C_j \equiv \sum_{\substack{k=1 \\ k \neq j}}^{N} \|A_{k,j}\| , \qquad 1 \leqq j \leqq N . \tag{3.6}$$

*Then, for any* $\alpha$ *with* $0 \leqq \alpha \leqq 1$, *each eigenvalue* $\lambda$ *of* $A$ *satisfies*

$$(\|(A_{j,j} - \lambda I_j)^{-1}\|)^{-1} \leqq R_j^{\alpha} C_j^{1-\alpha} \tag{3.7}$$

*for at least one* $j$, $1 \leqq j \leqq N$.

Also, the important result of Fan and Hoffman [3] carries over with ease.

THEOREM 8. *Let the* $n \times n$ *complex matrix* $A$ *be partitioned as in* (2.1). *Let* $p > 1$, *and* $1/p + 1/q = 1$. *If* $\alpha > 0$ *satisfies*

$$\sum_{i=1}^{N} \left\{ \frac{\left(\sum_{j \neq i} \|A_{i,j}\|\right)^q}{\left(\sum_{j \neq i} \|A_{i,j}\|^p\right)^{q/p}} \right\} \leqq \alpha^q (1 + \alpha^q) , \tag{3.8}$$

(whenever $0/0$ occurs on the left-hand side, we agree to put $0/0 = 0$), then every eigenvalue $\lambda$ of $A$ satisfies at least one of the following relations:

$$(\|(A_{j,j} - \lambda I_j)^{-1}\|)^{-1} \leqq \alpha \left( \sum_{\substack{k=1 \\ k \neq j}}^{N} \|A_{j,k}\|^p \right)^{1/p} , \qquad 1 \leqq j \leqq N . \tag{3.9}$$

We wish to emphasize that, unlike the cases previously treated where all the matrices $A_{i,j}$ of (2.1) are $1 \times 1$ matrices, these new inclusion regions now *depend* on the vector norms used. It seems reasonable, at least theoretically, to minimize these inclusion regions by considering all possible vector norms to produce optimum results. Similarly, there is a great deal of flexibility in the manner in which the matrix $A$ is partitioned, and this perhaps can be used to advantage.

**4. Another generalization.** Another result, due again to Taussky [12], states that if an $n \times n$ matrix $A = (a_{i,j})$ is strictly diagonally dominant in the usual sense of (2.4′) with positive real diagonal entries $a_{i,i}$ $1 \leqq i \leqq n$, then the eigenvalues $\lambda_j$ of $A$ satisfy

$$Re \lambda_j > 0 , \quad 1 \leqq j \leqq n . \tag{4.1}$$

Based on our previous results, we now give a generalization of this result which depends upon the use of *absolute norms* [1]. By this, we mean the following. First, if $\boldsymbol{x}$ is a column vector with complex components $x_i$, let $|\boldsymbol{x}|$ denote the vector with components $|x_i|$. If

$$\|\boldsymbol{x}\| = \||\boldsymbol{x}|\| \tag{4.2}$$

for all vectors $\boldsymbol{x}$, then the norm is an absolute norm.[4] This is equivalent

[4] Clearly, the $l_p$-norms of footnote 2 are absolute norms.

[1] to the property that if $|\boldsymbol{y}| \geqq |\boldsymbol{x}|$, i.e., each component of $|\boldsymbol{y}| - |\boldsymbol{x}|$ is a nonnegative real number, then

$$\|\boldsymbol{y}\| \geqq \|\boldsymbol{x}\| . \tag{4.2$'$}$$

Next, if $B = (b_{i,j})$ is a real $m \times m$ matrix with $b_{i,j} \leqq 0$ for all $i \neq j$, and if $B$ is nonsingular with $B^{-1} \equiv (r_{i,j})$ such that $r_{i,j} \geqq 0$ for all $1 \leqq i$, $j \leqq m$, then $B$ is said to be an *M-matrix* [8].

THEOREM 9. *Let the $n \times n$ complex matrix $A$ be partitioned as in* (2.1), *and let $A$ be block strictly diagonally dominant* (*or block irreducible and block diagonally dominant with strict inequality in* (2.4) *for at least one $j$*). *Further, assume that each submatrix $A_{j,j}$ is an M-matrix, $1 \leqq j \leqq N$, and the vector norms for each subspace $\Omega_j$ are absolute norms. If $\lambda$ is any eigenvalue of $A$, then*

$$Re \lambda > 0 . \tag{4.3}$$

*Proof.* For simplicity, we shall consider again only the case where $A$ is block strictly diagonally dominant. Let $z$ be any complex number with $Re z \leqq 0$. If $A_{j,j}^{-1} \equiv (r_{k,l})$, and $(A_{j,j} - zI_j)^{-1} \equiv (s_{k,l}(z))$, it follows [8] from the assumption that $A_{j,j}$ is an $M$-matrix that

$$|s_{k,l}(z)| \leqq r_{k,l} , \qquad 1 \leqq k, l \leqq n_j . \tag{4.4}$$

Next, with (4.4) and the assumption of absolute norms, it follows from (4.2) and (4.2$'$) that

$$\frac{\|(A_{j,j} - zI_j)^{-1}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} \leqq \frac{\|A_{j,j}^{-1}|\boldsymbol{x}|\|}{\||\boldsymbol{x}|\|} ,$$

so that from (2.3),

$$(\|(A_{j,j} - zI_j)^{-1}\|)^{-1} \geqq (\|A_{j,j}^{-1}\|)^{-1} .$$

In other words, for any $z$ with $Re z \leqq 0$, then the matrix $A - zI$ continues to be block strictly diagonally dominant, and hence nonsingular. Thus, if $\lambda$ is an eigenvalue of $A$, then $Re \lambda > 0$, which completes the proof.

## BIBLIOGRAPHY

1. F. L. Bauer, J. Stoer, and C. Witzgall, *Absolute and monotonic norms*, Numerische Math., **3** (1961), 257-264.

2. A. Brauer, *Limits for the characteristic roots of a matrix II*, Duke Math. J., **14** (1947), 21-26.

3. Ky Fan and A. J. Hoffman, *Lower bounds for the rank and location of the eigenvalues of a matrix, Contributions to the Solution of Systems of Linear Equations and the Determination of Eigenvalues*, edited by Olga Taussky, Nat. Bur. of Standards Appl. Math. Series **39** (1954), 117-130.

4. Miroslav Fiedler, *Some estimates of spectra of matrices, Symposium on the Numerical Treatment of Ordinary Differential Equations, Integral and Integro-Differential Equations,* Rome, 1960, Birkhauser Verlag, (1961), 33-36.

5. Miroslav Fiedler and Vlastimil Pták, *Some inequalities for the spectrum of a matrix* Mat. Fyz. Časopis. Slovensk. Akad. Vied, **10** (1960), 148-166.

6. Frank Harary, *On the consistency of precedence matrices,* J. Assoc. Comput. Mach., **7** (1960), 255-259.

7. Alston S. Householder, *On the convergence of matrix iterations,* J. Assoc. Comput. Mach, **3** (1956), 314-324.

8. A. M. Ostrowski, *Über die Determinanten mit überwiegender Hauptdiagonale,* Comment. Math. Helv., **10** (1937), 69-96.

9. A. M. Ostrowski, *Über das Nichtverschwinden einer Klasse von Determinanten und die Lokalisierung der charakteristischen Wurzeln von Matrizen,* Compositio Math., **9** (1951), 209-226.

10. A. M. Ostrowski, *On some metrical properties of operator matrices and matrices partitioned into blocks,* Journal of Math. Anal. and Appl., **2** (1961), 161-209.

11. Olga Taussky, *Bounds for characteristic roots of matrices,* Duke Math. J., **15** (1948). 1043-1044.

12. ———, *A recurring theorem on determinants,* Amer. Math. Monthly, **56** (1949), 672-676.

13. ———, *Some topics concerning bounds for eigenvalues of finite matrices, Survey of Numerical Analysis,* edited by John Todd, McGraw-Hill, (1962), 279-297.

# A GENERALIZED SOLUTION OF THE BOUNDARY VALUE PROBLEM FOR $y'' = f(x, y, y')$.

LEONARD FOUNTAIN AND LLOYD JACKSON

**1. Introduction.** In a paper in 1922 Perron [16] presented a new method of attacking the boundary value problem for Laplace's equation. This method consisted of employing the existence of solutions of the boundary value problem for small circles and the existence in the large of subharmonic and superharmonic functions to demonstrate the existence of a solution of the boundary value problem in the large. Since then Perron's methods have been generalized and applied to more general elliptic partial differential equations, for example, Tautz [17], Beckenbach and Jackson [2], Inoue [11], Jackson [12].

Subharmonic functions bear the same relationship to harmonic functions that convex functions bear to solutions of $y''(x) = 0$. In a paper in 1937 Beckenbach [3] introduced the idea of generalized convex functions. Since then a number of other mathematicians, for example, Bonsall [4], Green [7, 8], and Peixoto [15], have studied subfunctions with respect to solutions of second order ordinary differential equations. These subfunctions are special cases of Beckenbach's generalized convex functions and, if they have sufficient smoothness, are solutions of second order differential inequalities. Solutions of second order differential inequalities appear in many papers concerned with the existence of a solution of the boundary value problem for the equation

$$y'' = f(x, y, y') , \tag{1}$$

for example, Nagumo [14], Babkin [1]. However, the Perron method of systematically exploiting the properties of subfunctions and superfunctions in studying the boundary value problem does not appear to have been applied to equation (1). This paper consists of such a study.

In § 2 we list some properties of solutions of (1) most of which are known. In § 3 we define subfunctions and superfunctions and give some of the properties of these functions that will be needed in the subsequent sections. Most of these properties are analogues of classical properties of convex functions as given for example in [9; Chapt. III]. In § 4 the Perron method is used to obtain a "generalized" solution of the boundary value problem. Finally, in § 5 some conditions are given which are sufficient to guarantee that the "generalized" solution of § 4

Received February 6, 1961, and in revised form March 28, 1962. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF49 (638)-506. Reproduction in whole or in part is permitted for any purpose of the United States Government.

is the solution of the boundary value problem in the usual sense.

**2. Some basic lemmas.** In this section we shall list the basic results concerning equation (1) which will be required in the subsequent sections.

Let $R$ be the region in three dimensional Euclidean space defined by

$$R \equiv [(x, y, z): a \leqq x \leqq b, |y| + |z| < +\infty]$$

where a and $b$ are finite. We shall assume throughout this paper that $f(x, y, z)$ is continuous on $R$. Various other assumptions will be made from time to time concerning $f(x, y, z)$. The first of these are as follows:

$A_1$: $f(x, y, z)$ is a nondecreasing function of $y$ for each fixed $x$ and $z$.

$A_2$: $f(x, y, z)$ satisfies a Lipschitz condition with respect to $y$ and $z$ on each compact subset of $R$.

However, unless we specifically state these or other assumptions we will be assuming only the continuity of $f(x, y, z)$.

By a solution of the boundary value problem,

$$y'' = f(x, y, y')$$
$$y(x_1) = y_1\ , \qquad y(x_2) = y_2$$

where $a \leqq x_1 < x_2 \leqq b$, we shall mean a function $y(x)$ which is of class $C^{(2)}$ and is a solution of (1) on $(x_1, x_2)$, which is continuous on $[x_1, x_2]$, and which assumes the given boundary values at $x_1$ and $x_2$.

We shall also be interested in the special case of equation (1) when $y'$ is not present, that is, equation

$$y'' = f(x, y)\ . \tag{2}$$

We shall always assume $f(x, y)$ is continuous on

$$R^* \equiv [(x, y): a \leqq x \leqq b, |y| < +\infty]\ .$$

LEMMA 1. *Given any $M > 0$ and $N > 0$ there is a $\delta(M, N) > 0$ such that the boundary value problem*

$$y'' = f(x, y, y')$$
$$y(x_1) = y_1\ , \qquad y(x_2) = y_2$$

*has a solution of class $C^{(2)}$ on $[x_1, x_2]$ for any points $(x_1, y_1)$ and $(x_2, y_2)$ with $x_1, x_2 \in [a, b]$, $|x_1 - x_2| \leqq \delta$, $|y_1| \leqq M$, $|y_2| \leqq M$ and $|(y_1 - y_2)/(x_1 - x_2)| \leqq N$.*

LEMMA 2. *Given any $M > 0$ there is a $\delta(M) > 0$ such that the boundary value problem*

$$y'' = f(x, y)$$
$$y(x_1) = y_1 , \qquad y(x_2) = y_2$$

*has a solution of class* $C^{(2)}$ *on* $[x_1, x_2]$ *for any points* $(x_1, y_1)$ *and* $(x_2, y_2)$ *with* $x_1, x_2 \in [a, b]$, $|x_1 - x_2| \leqq \delta$, $|y_1| \leqq M$, *and* $|y_2| \leqq M$.

LEMMA 3. *Let* $M > 0$, $N > 0$ *be fixed and let* $\delta(M, N)$ *be as in Lemma* 1. *Then given any* $\varepsilon > 0$ *there is an* $\eta$, $0 < \eta \leqq \delta(M, N)$, *such that, for any points* $(x_1, y_1)$ *and* $(x_2, y_2)$ *with* $x_1, x_2 \in [a, b]$, $|x_1 - x_2| \leqq \eta$, $|y_1| \leqq M$, $|y_2| \leqq M$, *and* $|(y_1 - y_2)/(x_1 - x_2)| \leqq N$, *there is a solution* $y(x)$ *of* (1) *of class* $C^{(2)}$ *on* $[x_1, x_2]$ *with* $y(x_1) = y_1$, $y(x_2) = y_2$, *and with*

$$|y(x) - \omega(x)| \leqq \varepsilon$$

*and*

$$|y'(x) - \omega'(x)| \leqq \varepsilon$$

*on* $[x_1, x_2]$ *where* $\omega(x)$ *is the linear function with* $\omega(x_1) = y_1$ *and* $\omega(x_2) = y_2$. *An analogous statement with* $N$ *and* $|(y_1 - y_2)/(x_1 - x_2)| \leqq N$ *omitted is valid with respect to solutions of* (2).

*Proof.* Lemmas 1 and 2 can be proved by using the Schauder-Tychonoff fixed point theorem [6; p. 456]. Let

$$G(x, t) \equiv \begin{cases} \dfrac{(x_1 - t)(x_2 - x)}{x_2 - x_1} & \text{on } x_1 \leqq t \leqq x \leqq x_2 \\ \dfrac{(x_1 - x)(x_2 - t)}{x_2 - x_1} & \text{on } x_1 \leqq x \leqq t \leqq x_2 . \end{cases}$$

Let $B$ be the Banach space $C^{(1)}[x_1, x_2]$ with norm $\|y\| \equiv \max|y(x)| + \max|y'(x)|$. For a function which satisfies a Hölder condition with exponent $0 < \alpha < 1$ on $[x_1, x_2]$ let

$$h_\alpha(g) \equiv \sup\left[\frac{|g(r_1) - g(r_2)|}{|r_1 - r_2|^\alpha} : r_1, r_2 \in [x_1, x_2], r_1 \neq r_2\right].$$

Let $K$ be the set of all functions $u(x)$ in $B$ which are such that $u'(x)$ satisfies a Hölder condition with exponent $\alpha$ on $[x_1, x_2]$, $u(x_1) = u(x_2) = 0$, and $\|u\| + h_\alpha(u') \leqq \max[M, N]$. Then $K$ is a compact convex subset of $B$. It can be shown that there is a $\delta(M, N) > 0$ such that the mapping $F(u) = w$ defined by

$$w(x) = \int_{x1}^{x_2} G(x, t)f(t, u(t) + \omega(t), u'(t) + \omega'(t))dt$$

is a continuous mapping of $K$ into itself provided $|x_1 - x_2| \leqq \delta(M, N)$, $|y_1| \leqq M$, $|y_2| \leqq M$, $|(y_1 - y_2)/(x_1 - x_2)| \leqq N$, and $\omega(x)$ is the linear function with $\omega(x_1) = y_1$ and $\omega(x_2) = y_2$. If $u_0(x)$ is the fixed point of the mapping, $y(x) \equiv u_0(x) + \omega(x)$ is a solution of the boundary value problem. Lemma 3 is an immediate consequence of the boundedness of $f(t, u(t) + \omega(t), u'(t) + \omega'(t))$ for $a \leqq t \leqq b$, $u \in K$, $|\omega(t)| \leqq M$, and $|\omega'(t)| \leqq N$.

LEMMA 4. *If $y_0(x)$ is a solution of* (1) *of class $C^{(2)}$ on $[x_1, x_2] \subset [a, b]$ and if $a < x_2 < b$, then there is a $\delta > 0$ such that $x_2 + \delta \leqq b$ and a solution $y(x)$ of* (1) *of class $C^{(2)}$ on $[x_1, x_2 + \delta]$ with $y(x) \equiv y_0(x)$ on $[x_1, x_2]$. A similar statement applies at $x_1$ in case $a < x_1$.*

*Proof.* This is an immediate consequence of well known results [5; p. 15] concerning continuation of solutions.

LEMMA 5. *If $f(x, y, z)$ satisfies condition $A_2$ and if $y_0(x)$ is a solution of* (1) *of class $C^{(2)}$ on $[x_1, x_2] \subset [a, b]$, then for all sufficiently small $|m|$ there are solutions $y(x)$ of* (1) *of class $C^{(2)}$ on $[x_1, x_2]$ satisfying $y(x_1) = y_0(x_1)$, $y'(x_1) = y_0'(x_1) + m$, and $|y(x) - y_0(x)| \leqq |m| e^{2k(x_2 - x_1)}$ on $[x_1, x_2]$ where $k$ is a constant independent of $m$. A similar statement applies if the change in slope is made at $x_2$ instead of $x_1$.*

*Proof.* This lemma is an immediate consequence of well known results [5; p. 22] concerning the continuity of solutions with respect to initial conditions.

**3. Subfunctions and superfunctions.** In this section we define and develop some of the properties of subfunctions and superfunctions with respect to the solutions of an arbitrary but fixed equation (1). These definitions and properties will of course apply also to equation (2), however, Theorem 2 will apply only to equation (2).

We shall use a capital letter $I$ to represent a subinterval of the basic interval $[a, b]$. $I$ may be open, closed, or half-open. $\bar{I}$ is the closure of $I$, $I^0$ the interior of $I$, and $I'$ the complement of $I$.

DEFINITION 1. A real valued function $s$ defined on $I$ is said to be a subfunction on $I$ in case $s(x) \leqq y(x)$ on $[x_1, x_2]$ for any $[x_1, x_2] \subset I$ and any solution $y$ of (1) on $[x_1, x_2]$ with $s(x_1) \leqq y(x_1)$ and $s(x_2) \leqq y(x_2)$.

DEFINITION 2. A real valued function $S$ defined on $I$ is said to be a superfunction on $I$ in case $S(x) \geqq y(x)$ on $[x_1, x_2]$ for any $[x_1, x_2] \subset I$ and any solution $y$ of (1) on $[x_1, x_2]$ with $S(x_1) \geqq y(x_1)$ and $S(x_2) \geqq y(x_2)$.

We shall state our results in terms of subfunctions with obvious analogous results, which we shall not bother to state, holding for superfunctions. When we wish to refer to a result concerning superfunctions we shall simply refer to the corresponding statement concerning subfunctions.

THEOREM 1. *If $s$ is a subfunction on $I$, then the right-hand and left-hand limits, $s(x_0+0)$ and $s(x_0-0)$, exist at every $x_0 \in I^0$ and the appropriate one-sided limits exist at the endpoints of $\bar{I}$. These limits may be infinite.*

*Proof.* It will suffice to consider one case. Assume that $s(x_0-0)$ does not exist at $x_0 \in I^0$. Then there exist finite real numbers $c$ and $d$ such that

$$\liminf_{x \to x_0 -} s(x) \leqq c < d \leqq \limsup_{x \to x_0 -} s(x)$$

We can pick two sequences $\{a_n\}_{n=1}^{\infty} \subset I$ and $\{b_n\}_{n=1}^{\infty} \subset I$ with the following properties:

(i) $\lim a_n = \lim b_n = x_0$,

(ii) $a_n < b_n < a_{n+1}$ for each $n \geqq 1$,

and

(iii) $\lim s(a_n) = \limsup_{x \to x_0 -} s(x)$

and

$$\lim s(b_n) = \liminf_{x \to x_0 -} s(x).$$

Let $\varepsilon = (d-c)/4$ and pick $N_1 > 0$ such that

$$s(a_n) > d - \varepsilon ,$$

and

$$s(b_n) < c + \varepsilon$$

for $n \geqq N_1$.

By Lemmas 1, 2, and 3 there is an $n_0 \geqq N_1$ such that the boundary value problem

$$y'' = f(x, y, y')$$
$$y(b_{n_0}) = y(b_{n_0+1}) = \frac{c+d}{2}$$

has a solution $y(x)$ with $|y(x) - (c+d)/2| < \varepsilon$ on $[b_{n_0}, b_{n_0+1}]$. Then since $s$ is a subfunction,

$$s(b_{n_0}) < c + \varepsilon < \frac{c+d}{2} = y(b_{n_0}) \,,$$

and

$$s(b_{n_0+1}) < c + \varepsilon < \frac{c+d}{2} = y(b_{n_0+1}) \,,$$

it follows that we must have $s(a_{n_0+1}) \leqq y(a_{n_0+1})$. However,

$$s(a_{n_0+1}) > d - \varepsilon = \frac{c+d}{2} + \varepsilon > y(a_{n_0+1}) \,.$$

From this contradiction we conclude that $s(x_0 - 0)$ exists.

COROLLARY 1. *If $s$ is a subfunction on $I$, then $s(x_0) \leqq \max\,[s(x_0+0), s(x_0-0)]$ at every $x_0 \in I^0$.*

*Proof.* If either $s(x_0+0) = +\infty$ or $s(x_0-0) = +\infty$, the given inequality obviously holds. If $s(x_0+0) < +\infty$ and $s(x_0-0) < +\infty$, the same type of argument as was used in proving the Theorem can be employed to show that $s(x_0) > \max\,[s(x_0+0), s(x_0-0)]$ is not possible. Since $s(x)$ has a finite real value at every point of $I$, this shows that we cannot have simultaneously $s(x_0+0) = -\infty$ and $s(x_0-0) = -\infty$ at any $x_0 \in I^0$.

COROLLARY 2. *If $s$ is a bounded subfunction on $I$, then $s$ has at most a countable number of discontinuities on $I$.*

*Proof.* This is a known consequence of the existence of onesided limits everywhere on $I$, for example, see [10; p. 300].

THEOREM 2. *If $s$ is bounded function on $I$ and is a subfunction with respect to the solutions of a differential equation* (2), *then $s$ is continuous on $I^0$.*

*Proof.* By Theorem 1 $s(x_0-0)$ and $s(x_0+0)$ exist at every $x_0 \in I^0$ and $s(x_0) \leqq \max\,[s(x_0+0), s(x_0-0)]$. To be specific assume $s(x_0-0) \geqq s(x_0+0)$. First assume $s(x_0) < s(x_0-0)$ and let $|s(x)| \leqq M$ on $I$. Then by Lemma 2 for $[x_1, x_0] \subset I$ and $|x_1 - x_0| \leqq \delta(M)$ the boundary value problem

$$y'' = f(x, y)$$
$$y(x_1) = s(x_1)\,, \qquad y(x_0) = s(x_0)$$

has a solution $y(x)$. Then, since $s(x) \leqq y(x)$ on $[x_1, x_0]$, $s(x_0-0) \leqq y(x_0-0) = y(x_0) = s(x_0)$. Thus we have a contradiction and we conclude that $s(x_0) =$

$s(x_0 - 0)$.

Now assume $s(x_0 - 0) - s(x_0 + 0) = k > 0$. By Lemma 3 there is an $\eta > 0$ such that $[x_0 - \eta, x_0 + \eta] \subset I$, $\eta \leqq \delta(M)$, and such that, for any $[x_1, x_2] \subset I$ with $|x_1 - x_2| = \eta$, the boundary value problem

$$y'' = f(x, y)$$
$$y(x_1) = s(x_1)\ , \qquad y(x_2) = s(x_2)$$

has a solution $y(x; x_1, x_2)$ with

$$|y(x; x_1, x_2) - \omega(x; x_1, x_2)| < k/4$$

on $[x_1, x_2]$ where $\omega(x; x_1, x_2)$ is the linear function with $\omega(x_1) = s(x_1)$ and $\omega(x_2) = s(x_2)$. Now take $[x_1, x_2] \subset I$ such that $|x_1 - x_2| = \eta$, $x_1 < x_0 < x_2$, $|s(x_2) - s(x_0 + 0)| < k/4$, and $(2M/\eta)|x_2 - x_0| < k/4$. Then, since $|\omega'(x; x_1, x_2)| \leqq 2M/\eta$, it follows that

$$|\omega(x_0; x_1, x_2) - s(x_0 + 0)| < k/2.$$

Consequently,

$$|y(x_0; x_1, x_2) - s(x_0 + 0)| < 3k/4$$

which means that

$$s(x_0) = s(x_0 - 0) > y(x_0; x_1, x_2)\ .$$

This contradicts the fact that $s$ is a subfunction and we conclude that $s$ is continuous on $I^0$.

We shall see a little later that Theorem 2 is not true for equation (1) even if conditions $A_1$ and $A_2$ are assumed in addition to the continuity of $f(x, y, z)$.

For the following theorems the proofs are the same as the corresponding theorems for convex functions.

THEOREM 3. *If $\{s_\alpha: \alpha \in A\}$ is any collection of subfunctions on $I$ bounded above at each point of $I$, then $s_0$ defined by*

$$s_0(x) \equiv \sup_{\alpha \in A} s_\alpha(x)$$

*is a subfunction on $I$.*

THEOREM 4. *Let $s_1$ be a subfunction on $I$ and $s_2$ a subfunction on $[x_1, x_2] \subset \bar{I}$. Assume further that $s_2(x_i) \leqq s_1(x_i)$ for $i = 1, 2$ in case $x_i \in I^0$. Then $s$ defined on $I$ by*

$$s(x) \equiv \begin{cases} s_1(x) & \text{for } x \notin [x_1, x_2] \\ \max\,[s_1(x), s_2(x)] & \text{for } x \in [x_1, x_2] \end{cases}$$

is a subfunction on $I$.

For a function $g$ and a point $x_0$ at which $g(x_0 - 0)$ or $g(x_0 + 0)$ exist we define

$$d^- g(x_0) \equiv \limsup_{y \to x_0 -} \frac{g(x) - g(x_0 - 0)}{x - x_0}$$

$$d_- g(x_0) \equiv \liminf_{x \to x_0 -} \frac{g(x) - g(x_0 - 0)}{x - x_0}$$

$$d^+ g(x_0) \equiv \limsup_{x \to x_0 +} \frac{g(x) - g(x_0 + 0)}{x - x_0}$$

$$d_+ g(x_0) \equiv \liminf_{x \to x_0 +} \frac{g(x) - g(x_0 + 0)}{x - x_0} .$$

THEOREM 5. *If $s$ is a bounded subfunction on $I \subset [a, b]$ with $\bar{I} = [x_1, x_2]$, then $d^- s(x_0) = d_- s(x_0)$ for all $x_1 < x_0 \leqq x_2$ and $d^+ s(x_0) = d_+ s(x_0)$ for all $x_1 \leqq x_0 < x_2$.*

*Proof.* It suffices to consider one case. Assume that $x_1 \leqq x_0 < x_2$ and that $d^+ s(x_0) \neq d_+ s(x_0)$. Then there is a finite number $c$ such that

$$d_+ s(x_0) < c < d^+ s(x_0) .$$

There is a $\delta > 0$ such that $[x_0, x_0 + \delta] \subset [x_1, x_2]$ and such that the initial value problem

$$y'' = f(x, y, y')$$
$$y(x_0) = s(x_0 + 0) , \qquad y'(x_0) = c$$

has a solution $y(x)$ of class $C^{(2)}$ on $[x_0, x_0 + \delta]$. It is clear that this leads to a contradiction of the fact that $s$ is a subfunction on $I$. We conclude that $d^+ s(x_0) = d_+ s(x_0)$.

COROLLARY. *If $s$ is a bounded subfunction on $I$, then $s$ has a finite derivative almost everywhere on $I$.*

For a function $g$ defined on $I$ and $x_0 \in I^0$ we will employ the notation:

$$\bar{D} g(x_0) \equiv \limsup_{\delta \to 0} \frac{g(x_0 + \delta) - g(x_0 - \delta)}{2\delta} ,$$

$$\underline{D} g(x_0) \equiv \liminf_{\delta \to 0} \frac{g(x_0 + \delta) - g(x_0 - \delta)}{2\delta} .$$

THEOREM 6. *If $s$ is a subfunction of class $C^{(1)}$ on $I$, then $\underline{D} s'(x) \geqq f(x, s(x), s'(x))$ on $I^0$.*

*Proof.* Let $x_0 \in I^0$ and choose a $\delta_0 > 0$ such that $[x_0 - \delta_0, x_0 + \delta_0] \subset I$.

Let $|s(x)| \leqq M$ and $|s'(x)| \leqq N$ on $[x_0 - \delta_0, x_0 + \delta_0]$.

Given $\varepsilon > 0$ there is a $\rho > 0$ such that

$$f(x, y, z) \geqq f(x_0, s(x_0), s'(x_0)) - \varepsilon$$

for $|x - x_0| < \rho$, $|y - s(x_0)| < \rho$, $|z - s'(x_0)| < \rho$. Now choose a $\delta_1 > 0$ such that $|\omega(x;\delta) - s(x_0)| < \rho/2$ and $|\omega'(x;\delta) - s'(x_0)| < \rho/2$ on $[x_0 - \delta, x_0 + \delta]$ for all $0 < \delta \leqq \delta_1$ where $\omega(x;\delta)$ is the linear function with $\omega(x_0 - \delta) = s(x_0 - \delta)$ and $\omega(x_0 + \delta) = s(x_0 + \delta)$.

By Lemmas 1 and 3 there is a $\delta_2 > 0$ with $2\delta_2 \leqq \min[2\delta_0, 2\delta_1, \delta(M, N)]$ such that for any $0 < \delta \leqq \delta_2$ the boundary value problem

$$y'' = f(x, y, y')$$
$$y(x_0 - \delta) = s(x_0 - \delta)\,, \qquad y(x_0 + \delta) = s(x_0 + \delta)$$

has a solution $y(x;\delta)$ with

$$|y(x;\delta) - \omega(x;\delta)| < \rho/2$$

and

$$|y'(x;\delta) - \omega'(x;\delta)| < \rho/2$$

on $[x_0 - \delta, x_0 + \delta]$. Hence, for $0 < \delta \leqq \delta_2$, $|y(x;\delta) - s(x_0)| < \rho$, and $|y'(x;\delta) - s'(x_0)| < \rho$ on $[x_0 - \delta, x_0 + \delta]$. Then, since $s$ is a subfunction on $I$ we have for any $0 < \delta \leqq \delta_2$

$$\frac{s'(x_0 + \delta) - s'(x_0 - \delta)}{2\delta} \geqq \frac{y'(x_0 + \delta;\delta) - y'(x_0 - \delta;\delta)}{2\delta} = y''(\xi;\delta)$$

where $x_0 - \delta < \xi < x_0 + \delta$. Hence,

$$\frac{s'(x_0 + \delta) - s'(x_0 - \delta)}{2\delta} \geqq f(\xi, y(\xi;\delta), y'(\xi;\delta)) \geqq f(x_0, s(x_0), s'(x_0)) - \varepsilon$$

for all $0 < \delta \leqq \delta_2$. From which we conclude

$$\underline{D}s'(x_0) \geqq f(x_0, s(x_0), s'(x_0))\,.$$

Under more stringent conditions on the function $f(x, y, z)$, Peixoto [15; p. 564] gives $s''(x) \geqq f(x, s(x), s'(x))$ as a necessary and sufficient condition for a function of class $C^{(2)}$ to be a subfunction. Theorem 6 generalizes the necessity part of this result. The condition $s'' \geqq f(x, s, s')$ is not sufficient to guarantee that a function of class $C^{(2)}$ be a subfunction without having more than just continuity of $f(x, y, z)$. As a matter of fact continuity of $f(x, y, z)$ and condition $A_1$ are still not enough. To see this we observe that, if $s'' \geqq f(x, s, s')$ is a sufficient condition for a function of class $C^{(2)}$ to be a subfunction, then a solution of the boundary value problem when it exists is unique. The boundary value problem $y'' = |y'|^{1/3}$, $y(-1) = y(+1) = 4\sqrt{6}/45$ has both $y(x) \equiv 4\sqrt{6}/45$ and $y(x) = 4\sqrt{6}/45\,|x|^{5/2}$ as solutions.

The following Theorem gives conditions on $f$ which are adequate to insure that a function satisfying $s'' \geqq f(x, s, s')$ is a subfunction. It embodies a maximum principle which must be known; however, since we are not aware of a reference for it in this form, we will include a proof.

THEOREM 7. *Assume that $f(x, y, z)$ satisfies conditions $A_1$ and $A_2$ and that the functions $u(x)$ and $v(x)$ satisfy the following conditions:*

(i) *$u$ and $v$ are both continuous on $\bar{I}$ and of class $C^{(1)}$ on $I^0$,*

(ii) *$\underline{D}u'(x) \geqq f(x, u(x), u'(x))$ and $\bar{D}v'(x) \leqq f(x, v(x), v'(x))$ on $I^0$, and*

(iii) *$u(x) - v(x) \leqq M$, where $M \geqq 0$, at the endpoints of $\bar{I}$. Then either $u(x) - v(x) < M$ on $I^0$ or $u(x) - v(x) \equiv M$ on $\bar{I}$.*

*Proof.* We will assume $M = 0$ since the case where $M > 0$ can be reduced to this one by replacing $v(x)$ by $v(x) + M$.

Now assume that the statement of the Theorem is false. Then there are functions $u$ and $v$ satisfying the hypotheses of the Theorem with $u(x) - v(x) \not\equiv 0$ on $\bar{I}$ but with $u(x) - v(x) \geqq 0$ at some points of $I^0$. Let $N = \max[u(x) - v(x)]$ on $\bar{I}$. Because of the continuity of $u(x) - v(x)$ there is an $x_0 \in I^0$ and an interval $[x_1, x_2] \subset I^0$ such that $x_1 < x_0 < x_2$, $u(x_0) - v(x_0) = N$, and $u(x) - v(x) < N$ either on $x_0 < x \leqq x_2$ or on $x_1 \leqq x < x_0$. Assume that $u(x) - v(x) < N$ on $x_0 < x \leqq x_2$ to be specific.

Let $M_1 > 0$ be such that $|u(x)| + |u'(x)| \leqq M_1$ and $|v(x)| + |v'(x)| \leqq M_1$ on $[x_1, x_2]$ and let $F$ be the set $[(x, y, z)\colon x_1 \leqq x \leqq x_2, |y| + |z| \leqq M_1]$. By hypothesis there is a $k > 0$ such that

$$|f(x, y_1, z_1) - f(x, y_2, z_2)| \leqq k[|y_1 - y_2| + |z_1 - z_2|]$$

for all $(x, y_1, z_1)$ and $(x, y_2, z_2)$ in $F$.

Define the functions $w_1(x)$ and $w_2(x)$ as follows:

$$w_1(x) = \begin{cases} \dfrac{f(x, u(x), u'(x)) - f(x, u(x), v'(x))}{u'(x) - v'(x)} & \text{for } u'(x) \neq v'(x) \\ 0 & \text{for } u'(x) = v'(x)\,, \end{cases}$$

and

$$w_2(x) = \begin{cases} \dfrac{f(x, u(x), v'(x)) - f(x, v(x), v'(x))}{u(x) - v(x)} & \text{for } u(x) \neq v(x) \\ 0 & \text{for } u(x) = v(x)\,. \end{cases}$$

Then it is clear that $|w_1(x)| \leqq k$ and $|w_2(x)| \leqq k$ on $[x_1, x_2]$. Because of the assumed condition $A_1$ on $f$, $w_2(x) \geqq 0$ on $[x_1, x_2]$.

Choose $x_3 \in I^0$ such that $x_2 < x_3$ and define $h(x; \alpha)$ by

$$h(x; \alpha) = e^{-\alpha(x-x_3)^2} - e^{-\alpha(x_0-x_3)^2}$$

where $\alpha > 0$ fixed is chosen large enough that

$$L[h] \equiv h''(x) - w_1(x)h'(x) - w_2(x)h(x) > 0$$

on $[x_1, x_2]$.

Since $u(x_2) - v(x_2) < N$ we can choose $\eta > 0$ such that $u(x_2) - v(x_2) + \eta h(x_2) < N$. Then, if $g(x) \equiv u(x) - v(x) + \eta h(x)$, we have $g(x_1) < N$, $g(x_0) = N$, and $g(x_2) < N$. It follows that $g(x)$ has a maximum $N_1 \geqq N \geqq 0$ at a point $x_4$ with $x_1 < x_4 < x_2$. It follows that $\underline{D}g'(x_4) \leqq 0$. However,

$$\begin{aligned} \underline{D}g'(x_4) &\geqq \underline{D}[u'(x_4) + \eta h'(x_4)] - \bar{D}v'(x_4) \\ &\geqq \underline{D}u'(x_4) - \bar{D}v'(x_4) + \eta h''(x_4) > w_2(x_4)N_1 \geqq 0 \,. \end{aligned}$$

We have arrived at a contradiction and the Theorem is established.

COROLLARY 1. *Let $f(x, y, y')$ satisfy conditions $A_1$ and $A_2$. Then the solution of the boundary value problem*

$$\begin{gathered} y'' = f(x, y, y') \\ y(x_1) = y_1,\ y(x_2) = y_2 \end{gathered}$$

*for $[x_1, x_2] \subset [a, b]$, if it exists, will be unique.*

COROLLARY 2. *If in the statement of Theorem 7 we assume only that $f(x, y, y')$ satisfies condition $A_1$ but strengthen the assumptions concerning $u$ and $v$ by assuming that at least one of the differential inequalities is a strict inequality for $x_1 < x < x_2$, then $u(x) - v(x) < M$ for $x_1 < x < x_2$.*

COROLLARY 3. *If in Theorem 7 $\bar{I} = [c, d]$, $u(c) = v(c)$, and $u(d) > v(d)$, then $u(x) - v(x)$ is nondecreasing on $[c, d]$. If $u(c) > v(c)$ and $u(d) = v(d)$, then $u(x) - v(x)$ is nonincreasing on $[c, d]$.*

THEOREM 8. *Let $s(x)$ be continuous on $I$ and of class $C^{(1)}$ on $I^0$. Then, if $f(x, y, z)$ satisfies $A_1$ and $A_2$ and if $\underline{D}s'(x) \geqq f(x, s(x), s'(x))$ on $I^0$, it follows that $s(x)$ is a subfunction on $I$. If $f(x, y, z)$ satisfies condition $A_1$ and $\underline{D}s'(x) > f(x, s(x), s'(x))$ on $I^0$, $s(x)$ is a subfunction on $I$.*

*Proof.* Let $[x_1, x_2] \subset I$ and let $y(x)$ be a solution of (1) on $[x_1, x_2]$ with $s(x_i) \leqq y(x_i)$ for $i = 1, 2$. Then that $s(x) \leqq y(x)$ on $[x_1, x_2]$ follows from Theorem 7 or Corollary 2 of Theorem 7.

THEOREM 9. *Let $s(x)$ be a continuous subfunction and $S(x)$ a continuous superfunction on $[x_1, x_2]$ with $s(x_i) \leqq S(x_i)$, $i = 1, 2$. Assume that at least one of $s(x)$ and $S(x)$, say $S(x)$, is of class $C^{(1)}$ on $x_1 < x < x_2$. Then, if $f(x, y, z)$ satisfies $A_1$ and $A_2$, $s(x) \leqq S(x)$ on $[x_1, x_2]$. If $f(x, y, z)$ satisfies $A_1$ and $\bar{D}S'(x) < f(x, S(x), S'(x))$ on $x_1 < x < x_2$, then again $s(x) \leqq S(x)$ on $[x_1, x_2]$.*

*Proof.* Assume that the statement of the Theorem is false. Then $s(x) > S(x)$ for some points $x$ with $x_1 < x < x_2$. Let $M = \max[s(x) - S(x)]$ on $[x_1, x_2]$ and let $x_0$ be such that $x_1 < x_0 < x_2$, $s(x_0) - S(x_0) = M$, and $s(x) - S(x) < M$ for $x_0 < x \leqq x_2$. By Lemma 1 there is a $\delta > 0$ such that $x_1 < x_0 - \delta < x_0 + \delta < x_2$ and such that the boundary value problem

$$\begin{gathered} y'' = f(x, y, y') \\ y(x_0 - \delta) = S(x_0 - \delta) + M, y(x_0 + \delta) = S(x_0 + \delta) + M \end{gathered}$$

has a solution $y_1(x)$ of class $C^{(2)}$ on $[x_0 - \delta, x_0 + \delta]$. If $f(x, y, z)$ satisfies $A_1$ and $\bar{D}S'(x) < f(x, S(x), S'(x))$ on $x_1 < x < x_2$, it follows from Corollary 2 of Theorem 7 that $y_1(x_0) < S(x_0) + M$. Furthermore, since $s(x)$ is a subfunction and $s(x_0 \pm \delta) \leqq y_1(x_0 \pm \delta)$, we have $s(x_0) \leqq y_1(x_0) < S(x_0) + M$. From this contradiction we conclude that $s(x) \leqq S(x)$ on $[x_1, x_2]$.

Now assume that $f(x, y, z)$ satisfies $A_1$ and $A_2$ and that we know only that $S(x)$ is of class $C^{(1)}$ on $x_1 < x < x_2$ which implies $\bar{D}S'(x) \leqq f(x, S(x), S'(x))$ on $x_1 < x < x_2$. Then, if $y_1(x)$ is again the solution of the above boundary value problem, we have $y_1(x_0) \leqq S(x_0) + M$. By Lemma 5 and Theorem 7 there is an $m > 0$ such that the initial value problem

$$\begin{aligned} y'' &= f(x, y, y') \\ y(x_0 - \delta) &= y_1(x_0 - \delta) \\ y'(x_0 - \delta) &= y_1'(x_0 - \delta) - m \end{aligned}$$

has a solution $y_2(x)$ of class $C^{(2)}$ on $[x_0 - \delta, x_0 + \delta]$ with $y_2(x) < y_1(x)$ on $(x_0 - \delta, x_0 + \delta]$ and

$$y_1(x_0 + \delta) > y_2(x_0 + \delta) > s(x_0 + \delta) .$$

Then we have $s(x_0) \leqq y_2(x_0) < y_1(x_0) \leqq S(x_0) + M$ which is again a contradiction. Thus we have $s(x) \leqq S(x)$ on $[x_1, x_2]$.

COROLLARY. *Let $M > 0$ be a constant and assume that $f(x, y, z)$*

*satisfies* $A_1$ *and* $A_2$. *Then, if* $S(x)$ *is a continuous superfunction on* $I$, $S(x)+M$ *is also, and, if* $s(x)$ *is a continuous subfunction on* $I$, $s(x)-M$ *is also.*

THEOREM 10. *Assume that* $f(x, y, z)$ *satisfies conditions* $A_1$ *and* $A_2$, *that* $y(x)$ *is a solution of* (1) *of class* $C^{(2)}$ *on* $[x_1, x_2] \subset [a, b]$, *and that* $s(x)$ *is a subfunction on* $[x_1, x_2]$. *Assume further that there is an* $x_0$, $x_1 < x_0 < x_2$, *at which either* $s(x_0) = y(x_0)$, $s(x_0+0) = y(x_0)$, *or* $s(x_0-0) = y(x_0)$. *Then, if* $s(x_1) \leqq y(x_1)$, $s(x) \geqq y(x)$ *on* $x_0 < x \leqq x_2$. *If* $s(x_2) \leqq y(x_2)$, $s(x) \geqq y(x)$ *on* $x_1 \leqq x < x_0$.

*Proof.* Follows immediately from Lemma 5, Theorem 7, and the definition of subfunctions.

COROLLARY. *If* $f(x, y, z)$ *satisfies conditions* $A_1$ *and* $A_2$, *if* $y(x)$ *is a solution of* (1) *of class* $C^{(2)}$ *on* $[x_1, x_2] \subset [a, b]$, *and if* $s(x)$ *is a subfunction on* $[x_1, x_2]$ *with* $s(x_i) \leqq y(x_i)$, $i = 1, 2$, *then either* $\max[s(x), s(x+0), s(x-0)] < y(x)$ *on* $x_1 < x < x_2$ *or* $s(x) \equiv y(x)$ *on* $[x_1, x_2]$.

In the papers mentioned in the introduction which deal with generalized convex functions it is assumed that for any two points $(x_1, y_1)$, $(x_2, y_2)$, $x_1 \neq x_2$ in the strip $a \leqq x \leqq b$, $|y| < +\infty$ the boundary value problem has a unique solution which is defined throughout $a \leqq x \leqq b$. This leads to the conclusion that subfunctions and superfunctions are continuous in the interiors of their intervals of definition. With the assumptions we make this conclusion cannot be drawn. Consider the equation

$$y'' = -18x(y')^4 \tag{3}$$

which is such that $f(x, y, z)$ is continuous everywhere and satisfies $A_1$ and $A_2$. The function $g$ defined by $g(x) = x^{1/3} + 1$ for $0 < x \leqq 1$, $g(x) = x^{1/3}$ for $-1 \leqq x < 0$, and $g(0) = g_0$ where $0 \leqq g_0 \leqq 1$ is simultaneously a subfunction and superfunction on $[-1, +1]$ with respect to solutions of (3).

**4. A generalized solution of the boundary value problem.** In the previous Section the existence "in the small" of solutions for the initial value problem and the boundary value problem for (1) was used in discussing some of the properties of subfunctions and superfunctions with respect to solutions of (1). In this section we use the Perron method of using subfunctions and superfunctions to deal with the boundary value problem "in the large" for equation (1). Throughout this section we shall be dealing with the boundary value problem:

$$y'' = f(x, y, y')$$
$$y(a) = \alpha , \qquad y(b) = \beta . \tag{4}$$

DEFINITION 3. The function $\varphi(x)$ is said to be an under-function with respect to the boundary value problem (4) in case $\varphi(x)$ is a subfunction on $[a, b]$ with $\varphi(a) \leqq \alpha$ and $\varphi(b) \leqq \beta$.

DEFINITION 4. The function $\psi(x)$ is said to be an over-function with respect to the boundary value problem (4) in case $\psi(x)$ is a superfunction on $[a, b]$ with $\psi(a) \geqq \alpha$ and $\psi(b) \geqq \beta$.

In most of the results obtained in this section we shall require the following additional hypothesis:

$A_3$: $f(x, y, z)$ is such that with respect to the boundary value problem (4) there is an under-function which is continuous on $[a, b]$ and there is an over-function which is continuous on $[a, b]$ and is of class $C^{(1)}$ on $(a, b)$.

DEFINITION 5. Let $\{\varphi\}$ represent the collection of all under-functions with respect to boundary value problem (4) which are continuous on $[a, b]$. Then we define $H(x)$ by

$$H(x) \equiv \sup [\varphi(x): \varphi \in \{\varphi\}]$$

for each $x \in [a, b]$.

THEOREM 11. *If $f(x, y, y')$ satisfies conditions $A_1$, $A_2$, and $A_3$, $H(x)$ is a bounded subfunction on $[a, b]$.*

*Proof.* By $A_3$, $\{\varphi\}$ is nonnull and there is an over-function $\psi_0$ continuous on $[a, b]$ and of class $C^{(1)}$ on $(a, b)$. By Theorem 9 $\varphi(x) \leqq \psi_0(x)$ on $[a, b]$ for each $\varphi \in \{\varphi\}$, consequently, $H(x) \leqq \psi_0(x)$ on $[a, b]$. By Theorem 3 $H(x)$ is a subfunction and, if $\varphi_0 \in \{\varphi\}$, $\varphi_0(x) \leqq H(x) \leqq \psi_0(x)$ so that $H$ is bounded on $[a, b]$.

THEOREM 12. *If $f(x, y, y')$ satisfies $A_1$, $A_2$, and $A_3$, then $H(x)$ is a superfunction on $[a, b]$.*

*Proof.* Assume that $H$ is not a superfunction. Then there exists $[x_1, x_2] \subset [a, b]$ and a solution $y(x)$ of (1) on $[x_1, x_2]$ such that $H(x_i) \geqq y(x_i)$, $i = 1, 2$, but $H(x) < y(x)$ for some $x$ with $x_1 < x < x_2$. Let $x_0$, $x_1 < x_0 < x_2$, be such that $y(x_0) - H(x_0) = \varepsilon > 0$. By the definition of $H$ there are continuous under-functions $\varphi_1$ and $\varphi_2$ such that $H(x_1) -$

$\varphi_1(x_1) \leqq \varepsilon/4$ and $H(x_2) - \varphi_2(x_2) \leqq \varepsilon/4$. Now define $\varphi_3$ on $[a, b]$ as follows:

$$\varphi_3(x) \equiv \begin{cases} \max[\varphi_1(x), \varphi_2(x)] & \text{for } x \notin [x_1, x_2] \\ \max[\varphi_1(x), \varphi_2(x), y(x) - \varepsilon/2] & \text{for } x \in [x_1, x_2] \, . \end{cases}$$

Then by Theorems 3 and 4 and the Corollary of Theorem 9 $\varphi_3$ is a continuous under-function. However, $\varphi_3(x_0) \geqq y(x_0) - \varepsilon/2 = H(x_0) + \varepsilon/2$ which is impossible. It follows that $H$ is a superfunction on $[a, b]$.

COROLLARY. *For each* $x \in (a, b)$ $H(x) = \min[H(x+0), H(x-0)]$.

*Proof*. Since $H$ is a superfunction, it follows from Corollary 1 of Theorem 1 that $H(x) \geqq \min[H(x+0), H(x-0)]$. Since $H$ is lower semicontinuous on $[a, b]$, $H(x) \leqq \min[H(x+0), H(x-0)]$.

Theorem 13. *If* $f(x, y, y')$ *satisfies* $A_1$, $A_2$, *and* $A_3$, *then* $H$ *is a solution of* (1) *on an open subset of* $[a, b]$ *the complement of which is of measure* 0.

*Proof*. Let $x_0 \in (a, b)$ be a point at which $H'(x_0)$ exists. Then there is a $\delta_0 > 0$ such that $[x_0 - \delta_0, x_0 + \delta_0] \subset [a, b]$ and such that for all $\delta$ with $0 < \delta \leqq \delta_0$ we have

$$\left| \frac{H(x_0 + \delta) - H(x_0 - \delta)}{2\delta} \right| \leqq |H'(x_0)| + 1 \, .$$

Let $\delta(M, N) > 0$ be as in Lemma 1 with $M = \sup|H(x)|$ on $[a, b]$ and $N = |H'(x_0)| + 1$. Then for $2\delta = \min[2\delta_0, \delta(M, N)]$ the boundary value problem

$$y'' = f(x, y, y')$$

$$y(x_0 - \delta) = H(x_0 - \delta), \; y(x_0 + \delta) = H(x_0 + \delta)$$

has a solution $y(x)$ of class $C^{(2)}$ on $[x_0 - \delta, x_0 + \delta]$. Since $H$ is simultaneously a subfunction and superfunction, $H(x) \equiv y(x)$ on $[x_0 - \delta, x_0 + \delta]$. The result then follows as a consequence of the Corollary of Theorem 5.

In Theorem 5 we proved that, if $s$ is a bounded subfunction on $[a, b]$, then $d^+s(x) = d_+s(x)$ on $a \leqq x < b$ and $d^-s(x) = d_-s(x)$ on $a < x \leqq b$. In view of these equalities we introduce the additional notation:

$$Ds(x_0 + 0) \equiv \lim_{x \to x_0+} \frac{s(x) - s(x_0 + 0)}{x - x_0}$$

and

$$Ds(x_0 - 0) \equiv \lim_{x \to x_0-} \frac{s(x) - s(x_0 - 0)}{x - x_0} \, .$$

THEOREM 14. *If $f(x, y, y')$ satisfies $A_1$, $A_2$, and $A_3$, then $DH(x+0) = DH(x-0)$ for all $x \in (a, b)$. Let $E$ be the set of points in $(a, b)$ at which $H$ does not have a finite derivative. If $x \in E$ is a point of continuity of $H$, either $DH(x+0) = DH(x-0) = +\infty$ or $DH(x+0) = DH(x-0) = -\infty$. If $H(x+0) > H(x-0)$, $DH(x+0) = DH(x-0) = +\infty$. If $H(x+0) < H(x-0)$, $DH(x+0) = DH(x-0) = -\infty$.*

*Proof.* Since for $x \in (a, b)$ and $x \notin E$ $DH(x+0) = DH(x-0) = H'(x)$, we need consider only the points of $E$.

First we observe that it follows from the argument used in the proof of Theorem 13 that, if $x_0 \in E$ is a point of continuity of $H$, $DH(x_0+0)$ and $DH(x_0-0)$ cannot both be finite. Assume that $x_0 \in E$ is a point of continuity of $H$, that $DH(x_0+0) = +\infty$, but that $DH(x_0-0) \neq +\infty$. It follows that there is an $N > 0$ and a $\delta_0 > 0$ such that $\omega(x) \equiv H(x_0) + N(x - x_0) < H(x)$ for $0 < x_0 - x \leqq \delta_0$. By Lemmas 1 and 4 there is a $\delta_1$, $0 < \delta_1 \leqq \delta_0$, a $\delta_2 > 0$, and a solution $y(x)$ of (1) of class $C^{(2)}$ on $[x_0-\delta_1, x_0+\delta_2]$ with $y(x_0) = H(x_0)$ and $y(x_0-\delta_1) = \omega(x_0-\delta_1) < H(x_0-\delta_1)$. Applying Theorem 10 we conclude that $H(x) \leqq y(x)$ on $[x_0, x_0+\delta_2]$. This implies that $DH(x_0+0) \leqq y'(x_0)$ which contradicts the assumption that $DH(x_0+0) = +\infty$. We conclude that $DH(x_0-0) = +\infty$. The other possibilities at a point of continuity can be dealt with in a similar way.

Now assume that $H(x_0+0) > H(x_0-0)$ and that $DH(x_0+0) \neq +\infty$. Then by the same type of argument as was used above we can conclude that there exist $\delta_1 > 0$, $\delta_2 > 0$, and a solution $y(x)$ of (1) of class $C^{(2)}$ on $[x_0-\delta_1, x_0+\delta_2]$ satisfying $y(x_0) = H(x_0+0)$ and $y(x_0+\delta_2) > H(x_0+\delta_2)$. We can then again apply Theorem 10 to conclude that $H(x) \geqq y(x)$ on $x_0 - \delta_1 \leqq x < x_0$ from which it follows that $H(x_0-0) \geqq y(x_0) = H(x_0+0)$. This contradicts the assumption that $H(x_0-0) < H(x_0+0)$ and we conclude that $DH(x_0+0) = +\infty$.

The remainder of the proof concerning the points of discontinuity is similar to this and will be omitted.

Next we consider the behavior of $H$ at the endpoints of the interval $[a, b]$.

THEOREM 15. *Assume that $f(x, y, y')$ satisfies $A_1$, $A_2$, and $A_3$. Then, if $DH(a+0) \neq +\infty$, $H(a+0) = H(a)$. If $H(a+0) < \alpha$, $DH(a+0) = -\infty$. If $DH(a+0)$ is finite, $H(a+0) = H(a) = \alpha$. Similar statements apply at $x = b$.*

*Proof.* The proof will be omitted since the methods used in it are very similar to those used in the proofs of Theorems 12 and 14.

If $f(x, y, y')$ satisfies conditions $A_1$, $A_2$, and $A_3$, and if the boundary value problem (4) has a solution, $H(x)$ is that solution. On the basis of the properties of the function $H(x)$ it seems reasonable to refer to

$H(x)$ as a "generalized solution" of the boundary value problem. Usually by a generalized solution of a second order differential equation on an interval one means a function which has an absolutely continuous first derivative and which satisfies the differential equation almost everywhere on the interval. The function $H(x)$ may not even be continuous on $[a, b]$. Consider the boundary value problem $y'' = -18x(y')^4$, $y(-1) = -1$, $y(+1) = +2$. Here conditions $A_1$ and $A_2$ are obviously fulfilled and, since $\psi(x) \equiv +2$ is an over-function and $\varphi(x) \equiv -1$ is an under-function, condition $A_3$ is satisfied. In this case $H(x) = x^{1/3}$ for $-1 \leqq x \leqq 0$ and $H(x) = x^{1/3} + 1$ for $0 < x \leqq +1$.

We terminate this Section by considering the function $H(x)$ with respect to the boundary value problem:

$$\begin{aligned} y'' &= f(x, y) \\ y(a) = \alpha\ , &\quad y(b) = \beta\ . \end{aligned} \tag{5}$$

THEOREM 16. *Assume that $f(x, y)$ satisfies $A_1$ and $A_3$ with the additional assumption that there is an over-function $\psi$ with respect to the boundary value problem* (5) *such that $\psi$ is continuous on $[a, b]$, is of class $C^{(1)}$ on $(a, b)$, and satisfies $\bar{D}\psi'(x) < f(x, \psi(x))$ on $(a, b)$. Then the function $H(x)$, defined in the same manner as above, is again bounded on $[a, b]$ and is simultaneously a subfunction and a superfunction with respect to solutions of* (2). *In this case $H(x)$ is of class $C^{(2)}$ and is a solution of* (2) *on $[a, b]$.*

*Proof.* The proof that $H(x)$ is bounded and is simultaneously a subfunction and a superfunction on $[a, b]$ is exactly as given in Theorems 11 and 12 with one exception. Since we do not now have the Corollary of Theorem 9 available, we must give a separate proof that, if $y(x)$ is a solution of (2) on $[x_1, x_2] \subset [a, b]$ and $M \geqq 0$, then $y(x) - M$ is a subfunction with respect to (2) on $[x_1, x_2]$. To see that this is the case assume that $y_1(x)$ is a solution of (2) on $[x_3, x_4] \subset [x_1, x_2]$ with

$$y(x_3) - M = y_1(x_3)\ ,$$
$$y(x_4) - M = y_1(x_4)\ ,$$

and
$$y(x) - M > y_1(x) \quad \text{on} \quad x_3 < x < x_4\ .$$

Because of condition $A_1$ we then have

$$y''(x) - y_1''(x) = f(x, y(x)) - f(x, y_1(x)) \geqq 0$$

on $(x_3, x_4)$. This implies that $y(x) - y_1(x)$ is convex on $[x_3, x_4]$ which in turn implies that $y(x) - y_1(x) \leqq M$ on $[x_3, x_4]$. Thus it is not possible that $y(x) - y_1(x) > M$ on $(x_3, x_4)$. It follows that $y(x) - M$ is a subfunction on $[x_1, x_2]$.

Since $H(x)$ is bounded and is simultaneously a subfunction and a superfunction, we can apply Lemma 2 to conclude that $H(x)$ is of class $C^{(2)}$ and a solution of (2) on $[a, b]$.

**5. Existence theorems for a solution of the boundary value problem.** In this concluding Section we consider the question of determining additional conditions on $f(x, y, y')$ which will suffice to guarantee that $H(x)$ be a solution of the boundary value problem (4). Some of the results are known and we are merely giving new proofs of them, others appear to be new.

THEOREM 17. *Assume that $f(x, y, y')$ satisfies $A_1$, $A_2$, and $A_3$, that $\psi(x)$ is an over-function continuous on $[a, b]$ and of class $C^{(1)}$ on $(a, b)$, and that $\varphi(x)$ is a continuous underfunction. Assume that there is a function $h(t)$ positive and continuous for $t \geqq 0$ such that $|f(x, y, y')| \leqq h(|y'|)$ for $a \leqq x \leqq b$, $\varphi(x) \leqq y \leqq \psi(x)$, $|y'| < +\infty$ and such that*

$$\int_0^\infty \frac{tdt}{h(t)} = +\infty .$$

*Then $H(x)$ is the solution of boundary value problem* (4). *Nagumo* [14].

*Proof.* Let $x_0 \in (a, b)$ be a point at which $H'(x_0)$ exists. By Theorem 13 there is an open interval containing $x_0$ in which $H$ is a solution of (1). Let $(c, d) \subset [a, b]$ be the maximal such interval. Then, if $N$ is chosen so that

$$\int_{|H'(x_0)|}^{N} \frac{tdt}{h(t)} = \max \psi(x) - \min \varphi(x) ,$$

we will have $|H'(x)| \leqq N$ on $(c, d)$. It follows from Theorems 14 and 15 that $c = a$, $d = b$ and that $H$ is the solution of the boundary value problem.

THEOREM 18. *If $f(x, y)$ is continuous for $a \leqq x \leqq b$, $|y| < +\infty$, and satisfies $A_1$, then the boundary value problem* (5) *has a unique solution for each $\alpha$ and $\beta$. Babkin* [1].

*Proof.* Let $\omega(x)$ be the linear function with $\omega(a) = \alpha$ and $\omega(b) = \beta$. Define the functions $u(x)$ and $v(x)$ on $[a, b]$ by

$$u''(x) = |f(x, \omega(x))| + 1$$
$$u(a) = u(b) = 0 ,$$

and

$$v''(x) = -|f(x, \omega(x))| - 1$$
$$v(a) = v(b) = 0 .$$

Tnen it is not difficult to verify that $\psi_0(x) = v(x) + \omega(x)$ is an over-function of class $C^{(2)}$ satisfying $\psi_0''(x) < f(x, \psi_0(x))$ on $[a, b]$, and $\varphi_0(x) = u(x) + \omega(x)$ is an under-function of class $C^{(2)}$ satisfying $\varphi_0''(x) > f(x, \varphi_0(x))$ on $[a, b]$. The hypotheses of Theorem 16 are satisfied so that we can conclude that $H(x)$ is of class $C^{(2)}$ and is a solution of (2) on $[a, b]$. Since $\varphi_0(x) \leqq H(x) \leqq \psi_0(x)$ on $[a, b]$, $H(a) = \alpha$ and $H(b) = \beta$ so that $H(x)$ is a solution of boundary value problem (5). It follows from the proof of Theorem 16 that it is unique.

THEOREM 19. *Let $f(x, y, y')$ satisfy $A_1$, $A_2$, and $A_3$, and assume that there is a continuous function $g(x, y)$ such that $g(x, y) \leqq f(x, y, y')$ for all $(x, y, y') \in R$. Then $H(x)$ is of class $C^{(2)}$ for all $a < x < b$. If in addition $g(x, y)$ is nondecreasing as a function of $y$ for each fixed $x$, $H(x)$ is continuous on $[a, b]$.*

*Proof.* Let $[x_1, x_2] \subset [a, b]$ and let $S(x)$ be a solution of

$$y'' = g(x, y) \tag{6}$$

on $[x_1, x_2]$ with $H(x_1) \leqq S(x_1)$ and $H(x_2) \leqq S(x_2)$. Then $S''(x) = g(x, S(x)) \leqq f(x, S(x), S'(x))$ on $(x_1, x_2)$, hence by Theorem 8 $S(x)$ is a superfunction on $[x_1, x_2]$. Then from Theorem 9 and the fact that $\varphi(x_1) \leqq S(x_1)$ and $\varphi(x_2) \leqq S(x_2)$ we conclude that $\varphi(x) \leqq S(x)$ on $[x_1, x_2]$ for each continuous under-function $\varphi$. From this we conclude that $H(x) \leqq S(x)$ on $[x_1, x_2]$ and that $H$ is a subfunction with respect to solutions of (6). By Theorem 2 $H(x)$ is continuous on $(a, b)$.

Assume that $H$ does not have a finite derivative at $x_0$, $a < x_0 < b$. Assume that $DH(x_0 + 0) = +\infty$. By Lemma 2 there is a $\delta > 0$ such that the boundary value problem

$$y'' = g(x, y)$$
$$y(x_0) = H(x_0) , \qquad y(x_0 + \delta) = H(x_0 + \delta)$$

has a solution $y(x)$ of class $C^{(2)}$ on $[x_0, x_0 + \delta]$. Since $H(x) \leqq y(x)$ on $[x_0, x_0 + \delta]$, $DH(x_0 + 0) \leqq y'(x_0)$ which contradicts the assumption that $DH(x_0 + 0) = +\infty$. Similarly $DH(x_0 - 0) = -\infty$ is not possible. It follows from Theorem 14 that $H(x)$ has a finite derivative at each point of $(a, b)$, therefore, by Theorem 13 $H(x)$ is of class $C^{(2)}$ and is a solution of (1) on $a < x < b$.

If $g(x, y)$ is nondecreasing in $y$, it follows from Theorem 18 that the boundary value problem

$$y'' = g(x, y)$$

$$y(a) = \alpha\ , \qquad y(b) = \beta$$

has a solution $\psi(x)$ of class $C^{(2)}$ on $[a, b]$. $\psi(x)$ is an over-function with respect to the boundary value problem (4). It suffices to consider the endpoint $x = a$. Since $H(x) \leqq \psi(x)$, $H(a+0) \leqq \psi(a) = \alpha$. If $H(a+0) = \alpha$, $DH(a+0) \leqq \psi'(a)$ so that $DH(a+0) \neq +\infty$. It follows from Theorem 15 that $H(a+0) = H(a)$. If $H(a+0) < \alpha$, we again apply Theorem 15 to obtain $H(a+0) = H(a)$. We conclude that $H(x)$ is continuous on $[a, b]$.

COROLLARY *If $f(x, y, y')$ satisfies conditions $A_1$ and $A_2$, if there is a continuous function $g(x, y)$ nondecreasing in $y$ for each fixed $x$ and satisfying $g(x, y) \leqq f(x, y, y')$ on $R$, and if there exists a continuous under-function $\varphi(x)$ with $\varphi(a) = \alpha$ and $\varphi(b) = \beta$, then the boundary value problem* (4) *has a unique solution.*

If $f(x, y, y')$ satisfies the hypotheses of Theorem 19 including the assumption that $g(x, y)$ is nondecreasing in $y$ for each fixed $x$, then $H(x)$ is continuous on $[a, b]$ and of class $C^{(2)}$ on $(a, b)$. Furthermore, $DH(a+0) \neq +\infty$ and $DH(b-0) \neq -\infty$. As a consequence of Theorem 15 we could conclude that $H(x)$ is the solution of the boundary value problem if it could be shown that $DH(a+0) \neq -\infty$ and $DH(b-0) \neq +\infty$. This would be the case, for example, if for some $N > 0$ and some $\delta > 0$ $H''(x) < N$ on $a < x \leqq a + \delta$ and on $b - \delta \leqq x < b$.

As an illustration of these remarks consider the boundary value problem:

$$y'' = (1 + x^2)y^3 + e^{y'^2 \sin x}$$
$$y(-\pi/2) = \alpha\ , \qquad y(3\pi/2) = \beta\ .$$

The hypotheses of Theorem 19 are satisfied. $\varphi(x) \equiv \min[-1, \alpha, b]$ is an under-function and $\psi(x) \equiv \max[1, \alpha, \beta]$ is an over-function. In the intervals $-\pi/2 < x \leqq 0$ and $\pi \leqq x < 3\pi/2$ $H''(x)$ is bounded above, consequently, this boundary value problem always has a unique solution.

We conclude the paper with a final result in this direction.

THEOREM 20. *Assume that $f(x, y, y')$ satisfies $A_1$ and $A_2$, and that there is a continuous function $g(x, y)$ which is nondecreasing in $y$ and is such that $g(x, y) \leqq f(x, y, y')$ for all $(x, y, y') \in R$. Assume further that there exist functions $\psi$, $\varphi$, and $h$ such that*

(i) *$\psi(x)$ is continuous on $[a, b]$, is of class $C^{(1)}$ on $(a, b)$, and is an over-function with respect to boundary value problem* (4),

(ii) *$\varphi(x)$ is a continuous under-function with respect to the boundary value problem,*

*and* (iii) *$h(t)$ is positive and continuous for $t \geqq 0$, $|f(x, y, y')| \leqq h(|y'|)$*

*for* $a \leqq x \leqq b$, $\varphi(x) \leqq y \leqq \psi(x)$, $|y'| < +\infty$, *and*

$$\int_c^\infty \frac{tdt}{h(t)} > \max \psi(x) - \min \varphi(x) ,$$

*where*

$$c = \max\left|\frac{|\psi(b) - \varphi(a)|}{b-a}, \frac{|\psi(a) - \varphi(b)|}{b-a}\right| .$$

*Then* $H(x)$ *is a solution of boundary value problem* (4).

*Proof.* By Theorem 19, $H(x)$ continuous on $[a, b]$ and is of class $C^{(2)}$ on $(a, b)$. Since $\varphi(x) \leqq H(x) \leqq \psi(x)$ on $[a, b]$, $|(H(b) - H(a))/(b-a)| = |H'(x_0)| \leqq c$ for some $a < x_0 < b$. If $N$ is chosen so that $\int_c^N tdt/h(t) = \max \psi(x) - \min \varphi(x)$ then $|H'(x)| \leqq N$ on $(a, b)$. It follows from Theorem 15 that $H(x)$ is a solution of the boundary value problem.

As an illustration of this Theorem consider the boundary value problem

$$y'' = x^2y^3 + (y')^4$$
$$y(0) = \alpha , \qquad y(1) = \beta .$$

If $|\alpha| < M$ and $|\beta| < M$, the hypotheses of Theorem 20 are satisfied with $g(x, y) = x^2y^3$, $\psi(x) \equiv \max[|\alpha|, |\beta|]$, $\varphi(x) \equiv \min[-|\alpha|, -|\beta|]$ and $h(t) = t^4 + M^3$. Hence, by Theorem 20 the boundary value problem has a solution for $|\alpha| < M$ and $|\beta| < M$ if

$$\int_{2M}^\infty \frac{tdt}{t^4 + M^3} \geqq 2M .$$

The largest $M > 0$ for which this inequality is satisfied is the positive root of $\pi/2 - \text{Arctan } 4M^{1/2} = 4M^{5/2}$.

There are few existence theorems for the boundary value problem that do not impose more stringent conditions than Theorem 20 does on the rate of growth of $f(x, y, y')$ with respect to $y'$. In the cases in which it is applicable Theorem 20 seems to give stronger results than other known theorems.

A different method of obtaining existence theorems for the boundary value problem (4) via existence theorems "in the small" was recently given by Kamenskii [13].

## REFERENCES

1. B. N. Babkin, *Solution of a boundary value problem for an ordinary differential equation of second order by Caplygin's method,* Akad. Nauk S.S.S.R. Prikl. Math. Meh., **18** (1954), 239-242.

2. E. F. Beckenbach and L. K. Jackson, *Subfunctions of several variables*, Pacific J. Math., **3** (1953), 291-313.

3. E. F. Beckenbach, *Generalized convex functions*, Bull. Amer. Math. Soc., **43** (1937), 363-371.

4. F. F. Bonsall, *The characterization of generalized convex functions*, Quart. J. Math., Oxford Ser. (2) **1** (1950), 100-111.

5. Coddington and Levinson, *The Theory of Ordinary Differential Equations*, McGraw-Hill, 1955.

6. Dunford and Schwartz, *Linear Operators I*, Interscience Publishers, 1958.

7. J. W. Green, *Approximately convex functions*, Duke Math. J., **19** (1952), 499-504.

8. ———, *Generalized convex functions*, Proc. Amer. Math. Soc., **4** (1953), 391-396.

9. Hardy, Littlewood, and Polya, *Inequalities*, Cambridge University Press, 1952.

10. Hobson, *The Theory of Functions of a Real Variable and the Theory of Fourier Series*, Cambridge University Press, 1921.

11. M. Inoue, *Dirichlet problem relative to a family of functions*, J. Inst. Polytech., Osaka City Uni., **7** (1956), 1-16.

12. L. K. Jackson, *Subfunctions and the Dirichlet problem*, Pacific J. Math., **8** (1958), 243-255.

13. G. A. Kamenskii, *A two point boundary value problem for a non-linear second order differential equation and some theorems on intermediate values*, Doklady Akad. Nauk S.S.S.R. **139** (541-543).

14. Mitio Nagumo, *Über die Differentialgleichung* $y'' = f(x, y, y')$, Proc. Physics-Math. Soc., Japan Ser. 3 **19** (1937), 861-866.

15. M. M. Peixoto, *Generalized convex functions and second order differential inequalities*, Bull. Amer. Math. Soc., **55** (1949), 563-572.

16. O. Perron, *Eine neue Behandlung der ersten Randwertaufgabe fur* $\Delta u = 0$, Math. Zeit., **18** (1923), 42-54.

17. G. Tautz, *Zur Theorie der elliptischen Differentialgleichungen II*, Math. Ann., **118** (1943), 733-770.

# RINGS IN WHICH SEMI-PRIMARY IDEALS ARE PRIMARY

ROBERT W. GILMER

Every ring considered in this paper will be assumed to be commutative and to have a unit element. An ideal $A$ of a ring $R$ will be called semi-primary if its radical $\sqrt{A}$ is prime. That a semi-primary ideal need not be primary is shown by an example in [3; p. 154]. This paper is a study of rings $R$ satisfying the following condition: (*) Every semi-primary ideal of $R$ is primary. The ring $Z$ of integers clearly satisfies (*). More generally, if $A$ is a semi-primary ideal of a ring $R$ such that $\sqrt{A}$ is a maximal ideal of $R$, then $A$ is primary. [3; p. 153]. Hence, every ring having only maximal nonzero prime ideals satisfies (*).

An ideal $A$ of a ring $R$ is called $P$-primary if $A$ is primary and $P = \sqrt{A}$. If ring $R$ satisfies (*), then $A$ is $\sqrt{A}$-primary if and only if $\sqrt{A}$ is prime. Some well-known properties of a ring $R$ satisfying (*) are listed below.

*Property* 1. If $R$ satisfies (*) and $A$ is an ideal of $R$, then $R/A$ satisfies (*). [3; p. 148].

*Property* 2. If $R$ satisfies (*), if $A$ and $B$ are ideals of $R$ such that $A \subseteqq B \subseteqq \sqrt{A}$, and if $A$ is $\sqrt{A}$-primary then $B$ is $\sqrt{A}$-primary. [3; p. 147].

THEOREM 1. *If ring $R$ satisfies (*) and $P$, $A$, and $Q$ are ideals of $R$ such that $P$ is prime, $P \subset A$, and $Q$ is $P$-primary, then $QA = Q$.*

*Proof.* Since $\sqrt{QA} = P$, $QA$ is $P$-primary. Thus $Q \cdot A \subseteqq QA$ and $A \nsubseteqq P$ imply that $Q \subseteqq QA \subseteqq Q$. Hence $QA = Q$ as asserted.

THEOREM 2. *If $P$ is a nonmaximal prime ideal in a ring $R$ satisfying (*) and if $Q$ is $P$-primary, then $Q = P$.*

*Proof.* We let $P_1$ be a proper maximal ideal properly containing $P$. If $p_1 \in P_1$ such that $p_1 \notin P$ and if $p \in P$, then $Q \subseteqq Q + (pp_1) \subseteqq P$. By property 2, $Q + (pp_1)$ is $P$-primary. Since $pp_1 \in Q + (pp_1)$ and $p_1 \notin P$, $p \in Q + (pp_1)$. Then for some $q \in Q$, $r \in R$, $p(1 - rp_1) = q$. Now $1 - rp_1 \notin P_1$ since $P_1 \subset R$ so that $1 - rp_1 \notin P$. Thus $p \in Q$ and $P \subseteqq Q \subseteqq P$. Hence $P = Q$ and our proof is complete.

Received December 28, 1961.

COROLLARY 2.1. *If ring $R$ satisfies (*), if $P_1$ and $P_2$ are prime ideals of $R$ with $P_1 \subset P_2$, and if $Q$ is $P_2$-primary, then $P_1 \subset Q$.*

*Proof.* Since $\sqrt{QP_1} = P_1$, $QP_1$ is $P_1$-primary. By Theorem 2, $P_1 = QP_1 \subseteqq Q$. Now $Q$ is $P_2$-primary so that $P_1 \neq Q$. Hence $P_1 \subset Q$.

COROLLARY 2.2. *If ring $R$ satisfies (*) and $P$ is a nonmaximal prime ideal of $R$, then $P$ is idempotent.*

*Proof.* The ideal $P^2$ has radical $P$ and is therefore $P$-primary. By Theorem 2, $P^2 = P$.

THEOREM 3. *If $R$ is a ring satisfying (*), if $d$ is not a zero divisor or unit of $R$, and if $P$ is a minimal prime ideal of $(d)$, then $P$ is maximal in $R$.*

*Proof.* Suppose that $P$ is not maximal in $R$. Let $M$ denote the complement of $P$ in $R$. We define $A$ to be the set of all those elements $x$ of $R$ such that there exists $m \in M$ such that $xm \in (d)$. Since $P$ is prime, $A$ is an ideal and $A \subseteqq P$. We wish to show that $P = A$. Thus if $p \in P$ and if $N$ is the set of all elements of $R$ of the form $p^k m$ where $k$ is a nonnegative integer and $m \in M$, then $N$ is a multiplicatively closed set containing $M$ and $p$ and hence properly containing $M$. Because $P$ is a minimal prime ideal of $(d)$, $M$ is a maximal multiplicatively closed subset of $R$ not meeting $(d)$. [2; p. 106]. Therefore $N \cap (d) \neq \phi$ so that there exists an integer $k > 0$ and an element $m$ of $M$ such that $p^k m \in (d)$. That is, $p^k \in A$ so that $p \in \sqrt{A}$. Hence $P \subseteqq \sqrt{A} \subseteqq \sqrt{P} = P$ which implies $P = \sqrt{A}$. This means that $A$ is $P$-primary. Under the assumption that $P$ is nonmaximal, we conclude that $P = A$ by Theorem 2. Now $P$ is also a minimal prime ideal of $(d^2)$ so that if $B$ is the set of elements $y$ of $R$ such that $ym \in (d^2)$ for some $m \in M$, we likewise have $P = B$. Since $d \in P$, there exist $m \in M$ and $r \in R$ such that $dm = rd^2$. The element $d$ is not a zero divisor so that $m = rd \in (d) \subseteqq P$ which is a contradiction to our choice of $m$. Therefore $P$ is maximal as the theorem asserts.

COROLLARY 3.1. *If ring $R$ satisfies (*) and if $P$ is a proper prime ideal of $R$ containing a nonzero divisor $d$, then $P$ is maximal in $R$.*

*Proof.* There is a minimal prime ideal $P_1$ of $(d)$ contained in $P$. [1; p. 9]. By Theorem 3, $P_1$ is maximal. Hence $P$ is also maximal.

COROLLARY 3.2. *If $J$ is an integral domain satisfying (*), then nonzero proper prime ideals of $J$ are maximal.*

COROLLARY 3.3. *If ring $R$ satisfies (*) and if $P$ is a proper prime ideal of $R$, then $P$ is either maximal or minimal.*

*Proof.* Suppose that $P$ is not minimal and let $P_1$ be a prime ideal properly contained in $P$. Now $R/P_1$ is an integral domain satisfying (*) by property 1. By Corollary 3.2, $P/P_1$ is maximal in $R/P_1$. Thus $P$ is maximal in $R$. [3; p. 151].

THEOREM 4. *If ring $R$ satisfies (*) and $P$ is a finitely generated nonmaximal prime ideal of $R$ then $P$ is a direct summand of $R$. If $P_1$ is a prime ideal not containing $P$, then $P$ and $P_1$ are relatively prime.*

PROOF. By Corollary 2.2, $P = P^2$. Since $P$ is finitely generated, there exists an element $e \in P$ such that $(1 - e)P = (0)$. [3; p. 215]. Evidently $e^2 = e, P = (e)$ and $R = P \oplus (1 - e)$. Now $e(1 - e) \in P_1$ and $e \notin P_1$ so that $1 - e \in P_1$. Therefore $1 = e + (1 - e) \in P + P_1$ so that $P$ and $P_1$ are relatively prime.

THEOREM 5. *If the Noetherian ring $S$ satisfies (*), $S$ is a finite direct sum of Noetherian primary rings and Noetherian integral domains in which nonzero proper prime ideals are maximal. Conversely if $T$ is a finite direct sum of Noetherian primary rings and Noetherian integral domains in which nonzero proper prime ideals are maximal, then $T$ is a Noetherian ring satisfying (*).*

*Proof.* Since $S$ is Noetherian, every ideal of $S$ is finitely generated. Let $(0) = Q_1 \cap \cdots \cap Q_s$ be an irredundant representation of (0) as an intersection of greatest primary components where $P_i = \sqrt{Q_i}$. If $P_1, P_2, \cdots, P_k$ are the nonmaximal prime ideals of $S$ in this collection, $P_i = Q_i$ for $1 \leqq i \leqq k$ by Theorem 2. If $1 \leqq i < j \leqq s, P_i + P_j = S$. This follows from Theorem 4 if $P_i$ and $P_j$ are nonmaximal. If $P_j$, say, is maximal, then $P_j \not\supseteq P_i$ by Corollary 2.1, for $Q_j \not\supseteq P_i$ from the irredundance of the representation. Therefore, $P_i + P_j = S$. Thus the $P_i$'s, and hence the $Q_i$'s, are pairwise relatively prime. [3; p. 177]. This means that $S \cong S/P_1 \oplus \cdots \oplus S/P_k \oplus S/Q_{k+1} \oplus \cdots \oplus S/Q_s$. [3; p. 178]. Each $S/P_i$ in this representation is a Noetherian integral domain in which nonzero prime ideals are maximal. Since $Q_j$ for $k + 1 \leqq j \leqq s$ is $P_j$-primary with $P_j$ maximal, $S/Q_j$ is a Noetherian primary ring. [3; p. 204].

The converse follows from elementary facts concerning the ideal theory in a finite direct sum since it is apparent that each summand satisfies (*).

We give the following example of ring which is not a finite direct sum of indecomposable summands and which satisfies (*).

Let $S = \sum_{i=1}^{\infty w} Z_i$, where each $Z_i$ is the ring of integers and $\sum_{i=1}^{\infty w}$ designates the weak direct sum. Let $R = S + Z$ be the usual extension of $S$ to a ring with unit element. [2; p. 87]. Clearly $S$ is a prime ideal of $R$, as is $I_p = S + pZ$ for every prime $p$ of $Z$. In fact, each $I_p$ is a maximal ideal of $R$. It is easy to show that there is no prime ideal $P$ between $S$ and $I_p$.

Next, assume that $P$ is a prime ideal of $R$ that does not contain all of $S$. Then some $e_k \notin P$, where $e_k$ is the unity of $Z_k$. However, since $e_j e_k = 0$ for every $j \neq k$, evidently $Z_k \subset P$ for every $j \neq k$. By the same reasoning, $(1 - e_k)R \subseteqq P$. As before, it is easily shown that either $P = (1 - e_k)R$ or $P = (1 - e_k)R + pe_k R$ for some prime $p$.

Knowing precisely what the prime ideals of $R$ are, it is just a routine matter to check that $R$ satisfies (*).

The author is not able to give necessary and sufficient conditions which he feels are adequate that an arbitrary ring satisfy (*). The condition of Corollary 3.3, while necessary, is not sufficient to imply that a ring satisfy (*) as is shown by the following example.

If $S$ is the ring of polynomials in two indeterminates $X$ and $Y$ over a field $K$, then every nonzero proper prime ideal of $S$ has height 1 or 2. [4; p. 193]. Therefore if $A = (XY)$ and if $R = S/A$, $R$ is a Noetherian ring in which every prime ideal is maximal as minimal. The nonmaximal prime ideal $(X)/A$ of $R$, however, is not idempotent so that $R$ does not satisfy (*).

## BIBLIOGAPHY


1. W. Krull, *Idealtheorie,* (New York, 1948).
2. Neal H. McCoy, *Rings and Ideals,* (Menasha, Wisconsin, 1948).
3. O. Zariski, and Pierre Samuel, *Commutative Algebra.* V.I. (Princeton, 1958).
4. O. Zariski, and Pierre Samuel, *Commutative Algebra.* V. II. (Princeton, 1960).

# $K$-POLAR POLYNOMIALS

RUTH GOODMAN

**1. Introduction.** The complex polynomials

$$f(z) = \sum_{j=0}^{n} \binom{n}{j} a_j z^j , \qquad g(z) = \sum_{j=0}^{n} \binom{n}{j} b_j z^j \tag{1}$$

are called apolar if their coefficients satisfy the condition

$$\sum_{j=0}^{n} (-1)^j \binom{n}{j} a_{n-j} b_j = 0 .$$

A well known property of apolar polynomials is given [1] by

GRACE'S THEOREM. *If the polynomials $f(z)$ and $g(z)$ are apolar, then every circular domain containing all the zeros of one polynomial also contains at least one zero of the other.*

The term "circular domain" is used here to denote any region into which the circle $|z| \leqq 1$ can be transformed by a nonsingular linear fractional transformation

$$w = (ax + b)/(cx + d) ;$$

that is, a circular domain is a closed interior of a circle, a closed exterior of a circle, or a closed half plane.

It is natural to ask whether similar but more stringent conditions on the coefficients of (1) will insure that every circular domain containing all the zeros of one polynomial also contains at least $k$ zeros of the other when $k$ is integer greater than unity. We show here that this is the case. Our results can be stated more easily if we first make the

DEFINITION. The polynomials (1) are called $k$-polar ($1 \leqq k \leqq n$, $k$ an integer) if their coefficients satisfy the $k^2$ conditions

$$\sum_{j=0}^{n-k+1} (-1)^j \binom{n-k+1}{j} a_{s-j} b_{j+h} = 0 \tag{2}$$

$$(h = 0, \cdots, k-1; s = n, \cdots, n-k+1) .$$

We shall show that $k$-polarity of the polynomials (1) is sufficient to insure that the desired relation between their zeros does hold.

It is apparent that when $k$ is relatively large in comparison with

Received February 7, 1962.

$n$ there is only a restricted class of polynomials $f(z)$ for which $k$-polar polynomials $g(z)$ can exist. We shall show that when $2k+1 \geqq n$ the $k$-polarity of the polynomials (1) is both necessary and sufficient for them to have a common, repeated zero such that the multiplicities, $p$ and $q$, with which this zero occurs in the two polynomials satisfy the inequalities $p \geqq k$, $q \geqq k$, $p+q \geqq n+k$.

**2. The polar derivative.** To prove our principal results, we shall need a lemma concerning the $(n-1)$st degree polynomial

$$f_\zeta(z) = nf(z) + (\zeta - z)f'(z) = n\sum_{j=0}^{n-1}\binom{n-1}{j}(a_{j+1}\zeta + a_j)z^j .$$

This polynomial is called the "polar derivative of $f(z)$" or the "derivative of $f(z)$ with respect to $\zeta$". It can be obtain [2] from $f(z)$ as follows:

By the linear transformation

$$z = L(w) = (aw+b)/(cw+d) \qquad (bc - ad = 1) \tag{3}$$

transform $f(z)$ into the polynomial

$$F(w) = (cw+d)^n f(L(w)) ; \tag{4}$$

then to the derivative $F'(w)$ apply the inverse transformation $w = L^{-1}(z)$, obtaining $f_\zeta(z)$. If $c \neq 0$, then $\zeta = a/c = L(\infty)$; if $c = 0$, then $\zeta = \infty = L(\infty)$.

We shall refer to the polynomial $F(w)$ defined by (4) as the transform by (3) of the polynomial $f(z)$. It is important to observe [2] that the zeros of the transform $F(w)$ are the transforms by $w = L^{-1}(z)$ of those of $f(z)$.

LEMMA 1. *Let the nth degree polynomial $f(z)$ have $n-k$ zeros in $|z|<1$ and $k$ zeros in $|z|>r$, where $r>1$. Then there is a point $\zeta$ (not unique) such that $f_\zeta(z)$ has exactly $k-1$ zeros in $|z|>r$.*

*Proof.* Form $F(w)$ by applying to $f(z)$ the transformation

$$z = L(w) = (\zeta w - 1)/(w - \zeta) \qquad (1 < \zeta < r) ,$$

which takes $|z|<1$ into $|w|<1$ and takes $|z|>r$ into the circle

$$K_2\colon\ |w - C_2| < R_2 , \qquad C_2 = \frac{\zeta(r^2-1)}{r^2-\zeta^2} , \qquad R_2 = \frac{r(\zeta^2-1)}{r^2-\zeta^2} .$$

Now $F(w)$ has $k$ zeros in $K_2$ and $n-k$ zeros in $|w|<1$. Since the maximum modulus of these latter $n-k$ zeros is less than unity, we can choose $\mu<1$ such that these zeros also lie in $|w|<\mu$. Let $\rho =$

$(1+\mu)/2$. The circle

$$K_1\colon\ |w-(\rho-1)|<\rho$$

contains the circle $|w|<\mu$; for the line segment connecting $w=-\mu$ and $w=\mu$ is a diameter of $|w|<\mu$ and is contained in the line segment connecting $w=-1$ and $w=\mu$, which is a diameter of $K_1$. Thus $K_1$ contains $n-k$ zeros of $F(w)$. Applying the Walsh two circle theorem [5] to $K_1$ and $K_2$, we find that the zeros of $F'(w)$ lie in $K_1$, $K_2$, and the third circle

$$K\colon\ |w-C|<R\,,\quad C=\frac{(n-k)C_2+k(\rho-1)}{n}\,,\quad R=\frac{(n-k)R_2+k\rho}{n}\,.$$

Furthermore, it is an immediate consequence of the two circle theorem that if the boundaries of $K$ and $K_2$ do not intersect then there are exactly $k-1$ zeros of $F'(w)$ in $K_2$. The condition for the non-intersection of these two circles is

$$C_2-C>R_2+R\,.$$

This condition is equivalent to

$$n(C_2-C-R_2-R)=kC_2-R_2(2n-k)-k(2\rho-1)>0\,,$$

and this last inequality is equivalent to

$$\phi(\zeta)=k(r^2-1)\zeta-(2n-k)r(\zeta^2-1)-k(2\rho-1)(r^2-\zeta^2)>0\,.$$

Now

$$\phi(1)=2k(r^2-1)(1-\rho)>0\,,$$

since $r>1$ and $\rho<1$. Since $\phi(\zeta)$ is a real, continuous function of $\zeta$, it follows that $\phi(\zeta)>0$ in an interval $1\leqq\zeta\leqq 1+\varepsilon$, where $\varepsilon>0$. For any value of $\zeta$ in this interval, $K$ and $K_2$ do not intersect and $F'(w)$ has exactly $k-1$ zeros in $K_2$. Now the zeros of $f_\zeta(z)$ are the transforms by $z=L(w)$ of those of $F'(w)$. Hence exactly $k-1$ of them lie in the transform of $K_2$, that is, in $|z|>r$.

## 3. Properties of the $k$-polarity conditions.

To prove our principal results, we shall need to establish first some properties of the $k$-polarity conditions.

LEMMA 2. *For $k=1,\cdots,n+1$, the polynomials* (1) *can be written in the form*

$$f(z)=\sum_{j=0}^{k-1}\binom{k-1}{j}z^j f_{k,j},$$

where

$$f_{k,j} = f_{k,j}(z) = \sum_{i=0}^{n-k+1} \binom{n-k+1}{i} a_{i+j} z^i \qquad (j = 0, \cdots, k-1) .$$

The functions $f_{k,j}$ satisfy the relation

$$z f_{k+1,j+1} + f_{k+1,j} = f_{k,j} .$$

*Proof.* We show first the property of the functions $f_{k,j}$ which is stated last in the lemma. Using the definition of $f_{k,j}$ and a well known property of the binomial coefficients, we write

$$\begin{aligned}
z f_{k+1,j+1} + f_{k+1,j} &= \sum_{i=0}^{n-k} \binom{n-k}{i} a_{i+j+1} z^{i+1} + \sum_{i=0}^{n-k} \binom{n-k}{i} a_{i+j} z^i \\
&= \sum_{i=0}^{n-k+1} \binom{n-k}{i-1} a_{i+j} z^i + \sum_{i=0}^{n-k} \binom{n-k}{i} a_{i+j} z^i \\
&= a_{n-k+1+j} z^{n-k+1} + \sum_{i=1}^{n-k} \left[ \binom{n-k}{i-1} + \binom{n-k}{i} \right] a_{i+j} z^i + a_j \\
&= \sum_{i=0}^{n-k+1} \binom{n-k+1}{i} a_{i+j} z^i \\
&= f_{k,j} .
\end{aligned}$$

The proof of the first part of the lemma is by induction. It is true when $k = 1$, since $f_{1,0}$ reduces at once to $f(z)$. For any $k > 1$ we have

$$\begin{aligned}
\sum_{j=0}^{k} \binom{k}{j} z^j f_{k+1,j} &= z^k f_{k+1,k} + \sum_{j=1}^{k-1} \left[ \binom{k-1}{j-1} + \binom{k-1}{j} \right] z^j f_{k+1,j} + f_{k+1,0} \\
&= \sum_{j=0}^{k-1} \binom{k-1}{j} z^{j+1} f_{k+1,j+1} + \sum_{j=0}^{k-1} \binom{k-1}{i} z^j f_{k+1,j} \\
&= \sum_{j=0}^{k-1} \binom{k-1}{j} z^j (z f_{k+1,j+1} + f_{k+1,j}) \\
&= \sum_{j=0}^{k-1} \binom{k-1}{j} z^j f_{k,j} .
\end{aligned}$$

If the first part of the lemma is true when $k$ is replaced by $k-1$, then the last expression above is equal to $f(z)$. It follows that the lemma is true for all values of $k$.

LEMMA 3. *The polynomials* (1) *are $k$-polar if and only if the polynomials $f_{k,j}$ and $g_{k,i}$ are apolar for all $i = 0, \cdots, k-1$ and $j = 0, \cdots, k-1$.*

*Proof.* The proof is immediate, since applying the apolarity condition to all $f_{k,j}$ and $g_{k,i}$ yields conditions (2) at once.

LEMMA 4. *The k-polarity conditions* (2) *are invariant under non-singular linear transformations of the polynomials* (1).

*Proof.* Since any non-singular linear transformation is equivalent to a succession of transformations of the forms $z = \gamma w (\gamma \neq 0)$, $z = 1/w$, $z = w + \gamma$, the lemma can be established by showing the invariance of (2) for each of these special forms.

Each sum in (2) is invariant under magnifications and rotations. For applying $z = \gamma w$ to both $f(z)$ and $g(z)$ replaces $a_{s-j}$ by $\gamma^{s-j}a_{s-j}$ and $b_{j+h}$ by $\gamma^{j+h}b_{j+h}$, whence each term of the sum is multiplied by $\gamma^{s-j}\gamma^{j+h} = \gamma^{s+h}$. The sum, therefore, remains equal to zero.

Under the transformation $z = 1/w$, the polynomials (1) are carried into

$$F(w) = \sum_{j=0}^{n}\binom{n}{j}A_j w^j \text{ and } F(w) = \sum_{j=0}^{n}\binom{n}{j}B_j w^j ,$$

where $A_j = a_{n-j}$ and $B_j = b_{n-j} (j = 0, \cdots, n)$. The entire set of conditions (2) is invariant under this transformation. For we have

$$\begin{aligned}
&\sum_{j=0}^{n-k+1}(-1)^j\binom{n-k+1}{j}A_{s-j}B_{j+h} \\
&= \sum_{j=0}^{n-k+1}(-1)_j\binom{n-k+1}{j}a_{n-s+j}b_{n-h-j} \\
&= \sum_{j=n-k+1}^{0}(-1)^{n-k+1-j}\binom{n-k+1}{n-k+1-j}a_{2n-s-k+1-j}b_{k-h-1+j} \\
&= (-1)^{n-k+1}\sum_{j=0}^{n-k+1}(-1)^j\binom{n-k+1}{j}a_{s'-j}b_{h'+j} ,
\end{aligned}$$

where $s' = 2n - s - k + 1$ and $h' = k - h - 1$, so that $s'$ takes on the values $n - k + 1, \cdots, n$ and $h'$ takes on the values $k - 1, \cdots, 0$. Hence satisfaction of (2) by $f(z)$ and $g(z)$ insures satisfaction of (2) by $F(w)$ and $G(w)$.

To prove the invariance of (2) under translations, we first make use of Lemma 2 and show that if $f(z)$ is transformed into $F(w)$ by $z = w + \gamma$, then each polynomial $F_{k,j}(w)$ is a linear combination of the polynomials $f_{k,j}(w + c) (j = 0, \cdots, k - 1)$. Precisely, we show that the equations

$$(5) \qquad F_{k,j}(w) = \sum_{h=0}^{k-j-1}\binom{k-j-1}{h}\gamma^h f_{k,j+h}(w+\gamma) \qquad (j = 0, \cdots, k-1)$$

hold for every $k = 1, \cdots, n + 1$. The proof is by induction on $k$. We show first that the desired relations hold for the highest value of $k$, that is, $k = n + 1$. When $k = n + 1$, the equations defining $f_{k,j}$ and $F_{k,j}$ reduce to $f_{n+1,j} = a_j$ and $F_{n+1,j} = A_j$, so that (5) becomes

$$A_j = \sum_{h=0}^{n-j} \binom{n-j}{h} \gamma^h a_{j+h} \qquad (j = 0, \cdots, k-1) .$$

To see that this holds, we find $A_j$ by collecting the coefficients of the powers of $w$ in the polynomial $f(w+\gamma)$. We have

$$\begin{aligned} F(w) = f(w+\gamma) &= \sum_{i=0}^{n} \binom{n}{i} a_i (w+\gamma)^i \\ &= \sum_{i=0}^{n} \binom{n}{i} a_i \sum_{j=0}^{i} \binom{i}{j} \gamma^{i-j} w^j \\ &= \sum_{j=0}^{n} \binom{n}{j} w^j \sum_{i=j}^{n} \frac{\binom{n}{i}\binom{i}{j}}{\binom{n}{j}} \gamma^{i-j} \\ &= \sum_{j=0}^{n} \binom{n}{j} w^j \sum_{i=j}^{n} \binom{n-j}{i-j} a_i \gamma^{i-j} , \end{aligned}$$

so that

$$A_j = \sum_{i=j}^{n} \binom{n-j}{i-j} a_i \gamma^{i-j} = \sum_{h=0}^{n-j} \binom{n-j}{h} \gamma^h a_{j+h} .$$

Thus equations (5) hold when $k = n+1$. Next, we assume that they hold for general index $k+1$ and show that they also hold for index $k$. For convenience, we shall temporarily let $\phi_{k,j}$ denote $f_{k,j}(w+\gamma)$. ($F_{k,j}$ will denote $F_{k,j}(w)$ as usual.) Using the property of $F_{k,j}$ and $f_{k,j}$ established in Lemma 2 and assuming that equations (5) hold for $k+1$, we can write

$$\begin{aligned} F_{k,j} &= wF_{k+1,j+1} + F_{k+1,j} \\ &= w \sum_{h=0}^{k-j-1} \binom{k-j-1}{h} \gamma^h \phi_{k+1,j+h+1} + \sum_{h=0}^{k-j} \binom{k-j}{h} \gamma^h \phi_{k+1,j+h} \\ &= w \sum_{h=0}^{k-j-1} \binom{k-j-1}{h} \gamma^h \phi_{k+1,j+h+1} \\ &\quad + \sum_{h=0}^{k-j-1} \left\{ \binom{k-j-1}{h-1} + \binom{k-j-1}{h} \right\} \gamma^h \phi_{k+1,j+h} \\ &\quad + \phi_{k+1,j} + \gamma^{k-j} \phi_{k+1,k} \\ &= w \sum_{h=1}^{k-j-1} \binom{k-j-1}{h} \gamma^h \phi_{k+1,j+h+1} \\ &\quad + \sum_{h=0}^{k-j-1} \binom{k-j-1}{h} \gamma^{h+1} \phi_{k+1,j+h+1} \\ &\quad + \sum_{h=0}^{k-j-1} \binom{k-j-1}{h} \gamma^h \phi_{k+1,j+h} \\ &= \sum_{h=0}^{k-j-1} \binom{k-j-1}{h} \gamma^h \{ (w+\gamma) \phi_{k+1,j+h+1} + \phi_{k+1,j+h} \} \\ &= \sum_{h=0}^{k-j-1} \binom{k-j-1}{h} \gamma^h \phi_{k,j+h} . \end{aligned}$$

Thus equations (5) hold for $k = n+1, \cdots, 1$.

We have now established that each polynomial $F_{k,j}(w)$ is a linear combination of the polynomials $f_{k,j}(w+\gamma)$. To finish the proof of the invariance of (2) under translations, we recall the known facts (i) that apolarity is invariant under translations of the polynomials [1] and (ii) that if $E_1$ and $E_2$ are two sets of polynomials such that every polynomial of $E_1$ is apolar to every polynomial of $E_2$, then any linear combination of polynomials from $E_1$ is apolar to any linear combination of polynomials [1] from $E_2$. By Lemma 3, the $k$-polarity of $f(z)$ and $g(z)$ implies the apolarity of each polynomial in the set $E_1$: $\{f_{k,k-1}(w), \cdots, f_{k,0}(w)\}$ to each polynomial in the set $E_2$: $\{g_{k,k-1}(w), \cdots, g_{k,0}(w)\}$. Property (i) therefore implies that all polynomials of $E_1'$: $\{f_{k,k-1}(w+\gamma), \cdots, f_{k,0}(w+\gamma)$ are apolar to all polynomials of $E_2'$: $\{g_{k,k-1}(w+\gamma), \cdots, g_{k,0}(w+\gamma)\}$. We have just shown that each polynomial $F_{k,j}(w)$ is a linear combination of polynomials from $E_1'$ and each $G_{k,j}(w)$ is a linear combination of polynomials from $E_2'$. Thus property (ii) implies the apolarity of all the $F_{k,j}(w)$ to all the $G_{k,j}(w)$. Lemma 3 now gives the $k$-polarity of $F(w)$ and $G(w)$.

For convenience, we shall denote the repeated polar derivative $f_{\zeta_1\cdots\zeta_s}(z)$ as $f(z; \zeta, s)$.

LEMMA 5. *Let $k \geqq 2$ and $1 \leqq s \leqq k-1$. The $k$-polarity of $f(z)$ and $g(z)$ is necessary and sufficient for the $(k-s)$-polarity of the repeated polar derivatives $f(z; \zeta, s)$ and $g(z; \eta, s)$ for arbitrary points $\zeta_1, \cdots, \zeta_s$ and $\eta_1, \cdots, \eta_s$.*

*Proof.* It suffices to make the proof for $s = 1$, since re-application of this proof will then establish the lemma for all values of $s$ concerned. Letting $\phi(z) = f(z; \zeta, 1)$ and $\psi(z) = g(z; \eta, 1)$, we have

$$\phi(z) = \sum_{j=0}^{n-1} \binom{n-1}{j} (a_{j+1}\zeta_1 + a_j) z^j ,$$

whence

$$\begin{aligned} \phi_{k-1,j}(z) &= \sum_{i=0}^{n-k+1} \binom{n-k+1}{i} (a_{i+j}\zeta_1 + a_{i+j}) z^i \\ &= \zeta_1 f_{k,j+1}(z) + f_{k,j}(z) \qquad (j = 0, \cdots, k-2) . \end{aligned}$$

Similarly,

$$\psi_{k-1,j}(z) = \eta_1 g_{k,j+1}(z) + g_{k,j}(z) \qquad (j = 0, \cdots, k-2) .$$

The $k$-polarity of $f(z)$ and $g(z)$ implies the apolarity of both $f_{k,j+1}(z)$ and $f_{k,j}(z)$ to both $g_{k+1,j}(z)$ and $g_{k,j}(z)$. Thus $\phi_{k-1,j}(z)$ and $\psi_{k-1,j}(z)$, which are linear combinations of these polynomials, are apolar. The $(k-1)$-polarity of $\phi(z)$ and $\psi(z)$ now follows at once from Lemma 3.

If, on the other hand, $f(z;\zeta,1)=f_{\zeta_1}(z)$ and $g(z;\eta,1)=g_{\eta_1}(z)$ are $(k-1)$-polar for arbitrary values of $\zeta_1$ and $\eta_1$, then, in particular, both $f_0(z)$ and $f_\infty(z)$ are $(k-1)$-polar to both $g_0(z)$ and $g_\infty(z)$. For convenience, denote $f(z;\zeta,1)$ by $\phi(z;\zeta_1)$ and $g(z;\eta,1)$ by $\psi(z;\eta_1)$. We have

$$\phi(z;0)=f_0(z)=n\sum_{j=0}^{n-1}\binom{n-1}{j}a_jz^j\,,$$
$$\phi(z;\infty)=f_\infty(z)=n\sum_{j=0}^{n-1}\binom{n-1}{j}a_{j+1}z^j\,,$$

whence

$$\phi_{k-1,j}(z;0)=n\sum_{i=0}^{(n-1)-(k+1)-1}\binom{n-k+1}{i}a_{i+j}z^i=f_{k,j}(z)\,,$$
$$\phi_{k-1,j}(z,\infty)=n\sum_{i=0}^{n-k+1}\binom{n-1}{i}a_{i+j+1}z^i=f_{k,j+1}(z)$$
$$(j=0,\cdots,k-2)\,.$$

Similarly,

$$\psi_{k-1,j}(z;0)=g_{k,j}(z)\,,$$
$$\psi_{k-1,j}(z;\infty)=g_{k,j+1}(z)\qquad(j=0,\cdots,k-2)\,.$$

The $(k-1)$-polarity of $\phi(z;0)$ and $\phi(z;\infty)$ to $\psi(z;0)$ and $\psi(z;\infty)$ implies the apolarity of all the $\phi_{k-1,j}(z;0)$ and $\phi_{k-1,j}(z;\infty)$ to all the $\psi_{k-1,j}(z;0)$ and $\psi_{k-1,j}(z;\infty)$ for $j=0,\cdots,k-2$. The apolarity of all the $f_{k,j}(z)$ to all the $g_{k,j}(z)$ for $j=0,\cdots,k-1$ now follows at once. This, in turn, implies the $k$-polarity of $f(z)$ and $g(z)$.

LEMMA 6. *Let the nth degree polynomials $f(z)$ and $g(z)$ be k-polar. Let $\zeta_1,\cdots,\zeta_{n-k+1}$ be the zeros of any one of the polynomials $g_{k,k-1}(z),\cdots,$ $g_{k,0}(z)$, and let all these zeros be finite. Then $f(z;\zeta,n-k+1)$ vanishes identically.*

*Proof.* If $\zeta_1,\cdots,\zeta_{n-k+1}$ are the zeros of

$$g_{k,h}(z)=\sum_{i=0}^{n-k+1}\binom{n-k+1}{i}b_{i+h}z^i\,,$$

then their elementary symmetric functions can be expressed in terms of the coefficients. Let $S_0^{(m)}=1$ and for $i=1,\cdots,m$ let $S_i^{(m)}$ denote the sum of all possible products of $\zeta_1,\cdots,\zeta_m$ taken $i$ at a time. (Note that $b_{n-k+1+h}\neq 0$ since it is the leading coefficient of $g_{k,h}(z)$ and all the zeros of this polynomial are finite.) We have

$$S_1^{(n-k+1)}=(-1)^i\binom{n-k+1}{i}\frac{b_{n-k+1+h-i}}{b_{n-k+1+h}}\qquad(i=0,\cdots,n-k+1)\,.$$

Thus we can write

$$\begin{aligned} &b_{n-k+1+h} \sum_{i=0}^{n-k+1} a_{j+i} S_i^{(n-k+1)} \\ &\sum_{i=0}^{n-k+1} (-1)^i \binom{n-k+1}{i} a_{j+i} b_{n-k+1+h-i} \qquad (j = 0, \cdots, k-1) . \end{aligned}$$

Now for each value of $j$, the last expression above is the left side of one of the conditions (2). Consequently the $k$-polarity of $f(z)$ and $g(z)$ gives

$$\sum_{i=0}^{n-k+1} a_{j+i} S_i^{(n-k+1)} = 0 \qquad (j = 0, \cdots, k-1) .$$

Now it is known [3] that $f(z; \zeta, t)$ can be written in the form

$$f(z; \zeta, t) = \frac{n!}{(n-t)!} \sum_{j=0}^{n-t} \binom{n-t}{j} \sum_{i=0}^{t} a_{j+i} S_i^{(t)} z^j .$$

For $t = n - k + 1$, we have just shown that the sum which appears in the coefficient of each $z^j$ vanishes. Consequently, we have $f(z; \zeta, n-k+1) \equiv 0$, as we wanted to show.

**4. *K*-polar polynomials.** We are now ready to prove our principal results.

THEOREM 1. *If the polynomials $f(z)$ and $g(z)$ are $k$-polar, then every circular domain containing all the zeros of one polynomial also contains at least $k$ zeros of the other.*

*Proof.* The proof will be by induction on $k$. For $k = 1$, this theorem is simply Grace's theorem.

Assume that the theorem holds for $k = m$, and let $f(z)$ and $g(z)$ be $(m+1)$-polar. Let $C$ be a closed circular domain containing all the zeros of $g(z)$ and exactly $s$ zeros of $f(z)$. Then $C$ is contained in an open circular domain $C'$ whose closure also contains exactly $s$ zeros of $f(z)$. Since $k$-polarity is invariant under linear transformations, we can take $|z| > 1$ as $C'$. Then for a suitable $r > 1$, all the zeros of $g(z)$ and exactly $s$ zeros of $f(z)$ lie in $|z| > r$, while $n - s$ zeros of $f(z)$ lie in $|z| < 1$. By Lemma 1, therefore, there is a point $\zeta$ such that exactly $s - 1$ zeros of $f_\zeta(z)$ lie in $|z| > r$. Also, by Laguerre's theorem [2], all the zeros of $g_\eta(z)$ lie in $|z| > r$ whenever $\eta$ lies in $|z| \leqq r$. By Lemma 5, the $(m+1)$-polarity of $f(z)$ and $g(z)$ implies the $m$-polarity of $f_\zeta(z)$ and $g_\eta(z)$ for all values of $\zeta$ and $\eta$. Consequently, the assumption that the theorem holds for $k = m$ implies that the circular domain $|z| > r$, which contains all the zeros of $g_\eta(z)$, must contain at least $m$ zeros of $f_\zeta(z)$. Since we know that this domain contains exactly $s-1$ zeros of $f_\zeta(z)$, we have $s-1 \geqq m$. That is, $s \geqq m+1$, so that the theorem holds for $k=m+1$.

THEOREM 2. *For $(n+1)/2 \leqq k \leqq n$, the $k$-polarity of the $n$th degree polynomials $f(z)$ and $g(z)$ is necessary and sufficient for them to have a common, repeated zero whose multiplicities, $p$ and $q$, satisfy the inequalities $p \geqq k$, $q \geqq k$, $p+q \geqq n+k$.*

*Proof.* Suppose that two polynomials have a common repeated root whose multiplicities satisfy the given inequalities. A linear transformation will take the polynomials into

$$z^p\phi(z) = \sum_{i=0}^{n}\binom{n}{i}a_i z^i$$

and

$$z^q\psi(z) = \sum_{i=0}^{n}\binom{n}{i}b_i z^i$$

where $a_0 = \cdots = a_{p-1} = 0$ and $b_0 = \cdots = b_{q-1} = 0$. Now every product $a_ib_j$ which occurs in the $k$-polarity conditions (2) vanishes. For if $a_ib_j$ is to be nonzero, we must have $i \geqq p$ and $j \geqq q$, so that $i+j \geqq p+q$ whence $i+j \geqq n+k$. The maximum value which $i+j$ can assume for any $a_ib_j$ in (2), however, is $n+k-1$. Thus conditions (2) are satisfied and the polynomials are $k$-polar.

Suppose now that $f(z)$ and $g(z)$ are $k$-polar, with $k \geqq (n+1)/2$. We can, if necessary, perform a linear transformation on the polynomials to make $b_n \neq 0$ and $b_0 = 0$; that is, we can make all the zeros $\zeta_1, \cdots, \zeta_{n-k+1}$ of $g_{k,k-1}(z)$ finite and put one of these zeros at the origin. By Lemma 6, $f(z; \zeta, n-k+1) \equiv 0$. Thus [4] either $f(z; \zeta, n-k) \equiv 0$ or $f(z; \zeta, n-k) = c(z-\eta_{n-k+1})^k$. In either event, there is an $h$ in the range $k \leqq h \leqq n$ such that $f(z; \zeta, n-h+1) \equiv 0$ and $f(z; \zeta, n-h) = c(z-\zeta_{n-h+1})^h$. (Note that $f(z; \zeta, 0) = f(z)$.) We can assume that $\zeta_{n-h+1}$ is at the origin, so that $f(z; \zeta, n-h) = cz^h$. By Lemma 5, the $k$-polarity of $f(z)$ and $g(z)$ guarantees the $(k+h-n)$-polarity of $f(z; \zeta, n-h)$ and $g(z; \eta, n-h)$ for arbitrary $\eta_1, \cdots, \eta_{n-h}$. Let

$$f(z; \zeta, n-h) = \sum_{j=0}^{h}\binom{h}{j}A_j z^j ,$$

$$g(z; \eta, n-h) = \sum_{j=0}^{h}\binom{h}{j}B_j z^j .$$

Then we have $A_0 = \cdots = A_{h-1} = 0$, $A_h \neq 0$; and the $(k+h-n)$-polariy conditions which involve $A_h$ reduce to

$$A_hB_0 = \cdots = A_hB_{k+h-n-1} = 0 \ ,$$

whence

$$B_0 = \cdots = B_{k+h-n-1} = 0 \ . \tag{6}$$

We know [3] that

$$B_j = \mu \sum_{i=0}^{n-h} b_{j+i}S_i^{(n-h)} \qquad (j = 0, \cdots, h) \ ,$$

where $\mu = (n!)/(h!)$. Now equations (6) hold for arbitrary values of $\eta_1, \cdots, \eta_{n-h}$. Hence they hold in particular for $\eta_1 = \cdots = \eta_{n-h} = 0$. For these values, we have $S_1^{(n-h)} = \cdots = S_{n-h}^{(n-h)} = 0$, so that $B_0 = \mu b_0$, whence $B_0 = 0$ implies $b_0 = 0$. We can now use $\eta_1 = 1$, $\eta_2 = \cdots = \eta_{n-h} = 0$, so that $S_1^{(n-h)} = 1$, $S_2^{(n-h)} = \cdots = S_{n-h}^{(n-h)} = 0$, $B_0 = \mu b_1$, whence $b_1 = 0$. Using $\eta_1 = \eta_2 = 1$, $\eta_3 = \cdots = \eta_{n-h} = 0$ gives $S_1^{(n-h)} = 2$, $S_2^{(n-h)} = 1$, $S_3^{(n-h)} = \cdots = S_{n-h}^{(n-h)} = 0$, $B_0 = \mu b_2$, whence $b_2 = 0$. It is clear that we can proceed in this way to establish $b_3 = \cdots = b_{n-h} = 0$. We now have $B_1 = \mu b_{n-h+1}S_{n-h}^{(n-h)}$, whence we can conclude that $b_{n-h+1} = 0$. It then follows that $B_2 = \mu b_{n-h+2}S_{n-h}^{(n-h)}$, whence $b_{n-h+2} = 0$. We can proceed in this way to show that successive values of $b_j$ vanish until we arrive at $B_{k+h-n-1} = \mu b_{k-1}S_{n-h}^{(n-h)} = 0$, whence $b_{k-1} = 0$. Thus $g(z)$ has at least a $k$-fold zero at the origin. Let $q$ be the multiplicity of this zero, so that $b_0 = \cdots = b_{q-1} = 0$, $b_q \neq 0$. Since $q \geqq k = 2k - k \geqq n + 1 - k$, it follows that $b_q$ appears as the $b_j$ of highest index in $k$ of the $k$-polarity conditions. Since it is the only nonvanishing $b_j$ in any of these $k$ conditions, they reduce to

$$b_qa_0 = \cdots = b_qa_{k-1} = 0 \ ,$$

whence

$$a_0 = \cdots = a_{k-1} = 0 \ .$$

Thus $f(z)$ has a $p$-fold zero at the origin with $p \geqq k$. To finish the proof, we have left only to show that $p + q \geqq n + k$. Now the product $a_pb_q$ is nonvanishing. If it were to appear in any of the $k$-polarity equations (2), then the indices of every product $a_ib_j$ appearing in the same equation would have to satisfy $i + j = p + q$. But this means that if $i > p$ so that $a_i \neq 0$, then $j < q$ so that $b_j = 0$. Thus, if $a_pb_q$ did appear in any equation of (2), it would be the only non-vanishing product in this equation, whence the equation would not hold. Hence the product $a_pb_q$ cannot appear in any of the equations (2). But every product $a_ib_j$ does appear for which

$$n - k + 1 \leqq i + j \leqq n + k - 1 \ .$$

Therefore, either $p+q<n-k+1$ or $p+q>n+k-1$. But $p+q \geqq k+k \geqq n+1>n+1-k$. Consequently, we must have $p+q>n+k-1$, that is, $p+q \geqq n+k$.

## References

1. J. H. Grace, *The zeros of a polynomial*, Proc. Cambridge Philos. Soc., **11** (1902), 352-357.
2. E. Laguerre, Oeuvres, Paris: Gauthier-Villars, 1898, Vol. 1, pp. 48-66.
3. M. Marden, *The geometry of the zeros of a polynomial in a complex variable*, A. M. S. Math. Surveys No. III, New York: (1949), 38-43.
4. G. Pólya and G. Szegö, *Aufgaben und Lehrsätz aus der Analysis*, Berlin: Julius Springer, Vol. II, (1925), 61-64.
5. J. L. Walsh, *On the location of the roots of the derivative of a polynomial*, C. R. du Congress international des Mathematiciens, Strasbourg (1920), 339-342.

WESTINGHOUSE ELECTRIC CORP.

# ON THE ADDITIVITY OF LATTICE COMPLETENESS

*to the memory of Maurice Audin*

ISRAEL HALPERIN AND MARIA WONENBURGER

**1. Introduction.** It was shown in [**1**, Theorem 4.3] that upper $\aleph$-continuity[1] is additive in the following sense:

(1.1) *Suppose that* $[0, a]$, $[0, b]$ *are upper* $\aleph$*-continuous in a relatively complemented modular lattice. Then* $[0, a \cup b]$ *is upper* $\aleph$*-continuous provided that* $[0, a \cup b]$ *is upper* $\aleph$*-complete.*

But it may happen that $[0, a]$, $[0, b]$ are both upper $\aleph$-complete (both may even be von Neumann geometries with a perspective to $b$) and yet $[0, a \cup b]$ is *not* upper $\aleph$-complete. In fact there are von Neumann rings $\mathscr{R}$ for which the lattice $\bar{R}_{\mathscr{S}}$, with $\mathscr{S} = \mathscr{R}_2$, is not even upper $\aleph_0$-complete (see the Remark preceding Definition 3.1)

With a modest supplementary condition however, additivity of upper $\aleph$-completeness does hold, as we show in this paper.

**2. Terminology and notation.** We shall use the notation of [**1**], [**2**], and [**4**].

I will denote a set of indices $\alpha$ and $\bar{I}$ will denote the cardinal power of $I$.

$\aleph$ will denote an infinite cardinal, $\Omega$ will denote the least ordinal number whose corresponding cardinal power is $\aleph$.

A lattice is called *upper* $\aleph$*-complete* if the union $a = \bigcup (a_\alpha | \alpha \in I)$ exists whenever $\bar{I} \leqq \aleph$, and is called *upper* $\aleph$*-continuous* if for every $b$: $b \cap a = \bigcup ((b \cap \bigcup (a_\alpha | \alpha \in F)) | \text{all finite } F \subset I)$, with dual definitions for *lower* $\aleph$*-completeness* and *lower* $\aleph$-continuity. The lattice is called $\aleph$*-complete*, respectively $\aleph$*-continuous* if it is both upper and lower $\aleph$-continuous.

A complemented modular lattice $L$ is called an $\aleph$*-von Neumann-geometry* if it is $\aleph$-complete and $\aleph$-continuous (irreducibility is *not* assumed).

If we omit the $\aleph$ in any of these designations, this implies that the lattice $L$ has the corresponding $\aleph$-property for all $\aleph$.

If $\mathscr{R}$ is an associative regular ring (not necessarily with unit element) then $\bar{R}_{\mathscr{R}}$ denotes the relatively complemented modular lattice of its principal right ideals, ordered by inclusion. $\mathscr{R}$ is called an $\aleph$*-von Neumann-ring*, respectively a *von Neumann ring*, according as $\bar{R}_{\mathscr{R}}$ is an $\aleph$-von

Received December 28, 1961. Dr. Wonenburger is a postdoctorate Fellow (of the National Research Council of Canada) at Queen's University.

[1] Terminology and notation are explained in section 2 below.

Neumann-geometry, respectively a von Neumann geometry.

In any relatively complemented modular lattice, if $a \geqq b$ then $[a - b]$ will denote an arbitrary (but fixed) element such that $[a - b] \dot{\bigcup} b = a$ (the dot indicates that the summands in the union are independent). We write $a \sim b$ to denote: $a$ is perspective to $b$, and $a \lesssim b$ to denote: $a \sim b_1$ for some $b_1 \leqq b$. Elements $a, b$ are called *completely disjoint*, (notation: $(a, b)P$) if: $a_1 \sim b_1$, $a_1 \leqq a$, $b_1 \leqq b$ together imply $a_1 = 0$.

**3. The additivity of completeness theorem.**
In this section $a, b, c, \cdots x_\alpha, \cdots$ will denote elements in a given relatively complemented modular lattice $L$.

If $[0, a \cup c]$ is upper $\aleph$-complete we shall write $u(a, c, \aleph)$ to mean:

(3.1) *Whenever* $x_\alpha \leqq a \cup c$ *for all* $\alpha \in I$ (*with* $\bar{I} \leqq \aleph$) *and*

$$a \cap (\bigcup(x_\beta | \beta \in F)) = 0$$

*for all finite* $F \subset I$, *then* $a \cap (\bigcup(x_\alpha | \alpha \in I)) = 0$.

It is important to note: if $u(a, c, \aleph)$ holds then $u(a', c', \aleph)$ holds for all $a' \leqq a, c' \leqq c$.

Clearly, if $[0, a \cup c]$ is upper $\aleph$-complete *and* upper $\aleph$-continuous then $u(a, c, \aleph)$ does hold.

Similarly, if $[0, a \cup c]$ is lower $\aleph$-complete we shall write $l(a, c, \aleph)$ to denote:

$\overline{(3.1)}$ *Whenever* $x_\alpha \leqq a \cup c$ *for all* $\alpha \in I$ (*with* $\bar{I} \leqq \aleph$) *and*

$$a \cup (\bigcap(x_\beta | \beta \in F)) = a \cup c$$

*for all finite* $F \subset I$, *then* $a \cup (\bigcap(x_\alpha | \alpha \in I)) = a \cup c$.

It is important to note: if $l(a, c, \aleph)$ holds then $l(a', c', \aleph)$ holds for all $a' \leqq a, c' \leqq c$.

Clearly, if $[0, a \cup c]$ is lower $\aleph$-complete *and* lower $\aleph$-continuous then $l(a, c, \aleph)$ does hold.

LEMMA 3.1. *Suppose that each of* $[0, a \cup b]$, $[0, b \cup c]$, $[0, a \cup c]$ *is upper* $\aleph$-*complete and suppose that* $u(a, c, \aleph)$ *holds. Then* $[0, a \cup b \cup c]$ *is upper* $\aleph$-*complete.*

*Proof.* We may suppose that $\{a, b, c\}$ is an independent set, for if $c, b$ are replaced by $[c - (a \cap c)]$ and $[b - (b \cap (a \cup c))]$ respectively the hypotheses of Lemma 3.1 continue to hold and the conclusion is not changed.

Using transfinite induction, we may suppose that Lemma 3.1 holds for all $\aleph' < \aleph$. We may therefore assume that $x_\alpha$ is given, $\leqq a \cup b \cup c$ for all $0 < \alpha < \Omega$, that $\bigcup(x_\alpha | \alpha \leqq \beta)$ exists for all $\beta < \Omega$ and we need only show that $\bigcup(x_\alpha | \alpha < \Omega)$ exists.

We may suppose $x_\alpha \leqq x_\beta$ for $\alpha \leqq \beta < \Omega$ (by replacing the original $x_\alpha$ by $\bigcup(x_\beta | \beta \leqq \alpha)$ for all ($\alpha < \Omega$).

Set $\bar{x}_0 = \bigcup((x_\alpha \cap (a \cap b)) | \alpha < \Omega)$ (this union exists since, by hypothesis, $[0, a \cup b]$ is upper $\aleph$-complete). Set $\bar{x}_\alpha = \bar{x}_0 \cup x_\alpha$ for $0 < \alpha < \Omega$ and observe that $\bar{x}_\beta \leqq \bar{x}_\alpha$ for all $0 \leqq \beta \leqq \alpha < \Omega$.

Set $y_0 = \bar{x}_0$ and $y_\alpha = [\bar{x}_\alpha - \bigcup(\bar{x}_\beta | 0 \leqq \beta < \alpha)]$ for $0 < \alpha < \Omega$. Then $\bigcup(y_\beta | 0 \leqq \beta < \alpha) = \bigcup(\bar{x}_\beta | 0 \leqq \beta < \alpha)$ for all $0 < \alpha < \Omega$, as may be verified easily by transfinite induction.

Clearly, we need only show that $\bigcup(y_\alpha | 0 \leqq \alpha < \Omega)$ exists. Hence it is sufficient to show that $\bigcup_\alpha y_\alpha$ exists, where (for the rest of this proof) we write $\bigcup_\alpha$ to mean $\bigcup_{0<\alpha<\Omega}$ (note: $0 \leqq \alpha < \Omega$ has been replaced by $0 < \alpha < \Omega$).

Set $u = \big(a \cup (\bigcup_\alpha((a \cup y_\alpha) \cap (b \cup c)))\big) \cap \big(b \cup (\bigcup_\alpha((b \cup y_\alpha) \cap (a \cup c)))\big)$ (this union exists since, by hypothesis, $[0, b \cup c]$ and $[0, a \cup c]$ are upper $\aleph$-complete). We observe that $u \geqq y_\beta$ for all $0 < \beta < \Omega$ since each factor of $u$ has this property: for fixed $\beta$, $a \cup (\bigcup_\alpha((a \cup y_\alpha) \cap (b \cup c))) \geqq a \cup ((a \cup y_\beta) \cap (b \cup c)) = (a \cup y_\beta) \cap (a \cup b \cup c) = a \cup y_\beta \geqq y_\beta$.

We shall show that $u$ is the desired union $\bigcup_\alpha y_\alpha$. It is clearly sufficient to show for every $w$: if $u \geqq w \geqq y_\alpha$ for all $0 < \alpha < \Omega$ then $u \leqq w$.

Since $a \cup y_\alpha \leqq a \cup w$ and $b \cup y_\alpha \leqq b \cup w$ for all $0 < \alpha < \Omega$,

$$\begin{aligned} u &\leqq \big(a \cup ((a \cup w) \cap (b \cup c))\big) \cap \big(b \cup ((b \cup w) \cap (a \cup c))\big) \\ &= (a \cup w) \cap (b \cup w) = w \cup (a \cap (b \cup w)) . \end{aligned}$$

It is therefore sufficient to show that $a \cap (b \cup w) \leqq w$. We shall show that $a \cap (b \cup u) = 0$; this will imply:

$$a \cap (b \cup w) \leqq a \cap (b \cup u) = 0 \leqq w .$$

Now $b \cup u = \big(a \cup b \cup (\bigcup_\alpha(a \cup y_\alpha) \cap (b \cup c)))\big) \cap \big(b \cup (\bigcup_\alpha((b \cup y_\alpha) \cap (a \cup c)))\big)$,

$$\begin{aligned} a \cap (b \cup u) &= a \cap \big(b \cup (\bigcup_\alpha((b \cup y_\alpha) \cap (a \cup c)))\big) \\ &= a \cap ((b \cap (a \cup c)) \cup \big(\bigcup_\alpha((b \cup y_\alpha) \cap (a \cup c)))\big) \\ &= a \cap \big(\bigcup_\alpha((b \cup y_\alpha) \cap (a \cup c))\big) . \end{aligned}$$

Since $u(a, c, \aleph)$ is assumed to hold we need only show:

$$a \cap (\bigcup(((b \cup y_\alpha) \cap (a \cup c)) | \alpha = \alpha_1, \cdots, \alpha_m)) = 0$$

for every finite set of indices $0 < \alpha_1 < \alpha_2 < \cdots < \alpha_m < \Omega$.

Hence it is sufficient to show that

$$a \cap (b \cup (\bigcup(y_\alpha | \alpha = \alpha_1, \cdots, \alpha_m))) = 0 ,$$

and so it is sufficient to show that

$$(a \cup b) \cap (\bigcup(y_\alpha | \alpha = \alpha_1, \cdots, \alpha_m)) = 0 . \tag{3.2}$$

For this purpose, we note: $y_\alpha \cap (\bigcup(y_\beta | 0 \leqq \beta < \alpha) = 0$ for all $0 < \alpha < \Omega$. This implies that $\{y_\alpha | \alpha = 0, \alpha_1, \cdots, \alpha_m\}$ is an independent set and hence $y_0 \cap (\bigcup(y_\alpha | \alpha = \alpha_1, \cdots, \alpha_m)) = 0$. This implies (3.2) since the left side of (3.2) is $\leqq y_0$. Thus Lemma 3.1 is proved.

COROLLARY 1. *Suppose that $[0, a_i \cup a_j]$ is upper $\aleph$-complete for $i, j = 1, \cdots, m$ for some finite integer $m$ and suppose that $u(a_i, a_j, \aleph)$ holds whenever $i < j$. Then $[0, a_1 \cup \cdots \cup a_m]$ is upper $\aleph$-complete.*

*Proof.* If $m \leqq 2$ the conclusion is part of the hypotheses. Suppose that $m > 2$ and that the Corollary is known to hold with $m - 1$ in place of $m$; then Lemma 3.1 can be applied (with $a = a_1$, $b = a_3 \cup \cdots \cup a_m$ and $c = a_2$) to show that the Corollary holds for $m$ itself. By induction on $m$, the Corollary is established.

COROLLARY 2. *Suppose that $[0, a_i \cup a_j]$ is upper $\aleph$-complete and upper $\aleph$-continuous for $i, j = 1, \cdots, m$ for some finite integer $m$. Then $[0, a_1 \cup \cdots \cup a_m]$ is upper $\aleph$-complete and upper $\aleph$-continuous.*

*Proof.* Since upper $\aleph$-continuity of $[0, a_i \cup a_j]$ implies that $u(a_i, a_j, \aleph)$ holds, Corollary 1 shows that $[0, a_1 \cup \cdots \cup a_m]$ is upper $\aleph$-complete. The upper $\aleph$-continuity then follows from [1, Theorem 4.3].

LEMMA 3.2. *Suppose that $a = a_1 \cup a_2 \cup \cdots \cup a_m$ and $a_i \lesssim a_1 \cup \cdots \cup a_{i-1}$ for $1 < i \leqq m$. Then $a$ can be expressed in the form:*

(3.3) $a_1 \dot\cup \bar{a}_2 \dot\cup \cdots \dot\cup \bar{a}_n$ *for some $n \geqq m$ and elements $\bar{a}_2, \cdots, \bar{a}_n$ such that $\bar{a}_i \lesssim a_1$ for all $1 < i \leqq n$.*

*Moreover $\bar{a}_2$ may be taken to coincide with $a_2$ if $a_1 \cap a_2 = 0$.*

*Proof.* Lemma 3.2 holds trivially if $m = 1$ and also if $m = 2$ and $a_1 \cap a_2 = 0$. We may therefore suppose (by induction) that $m > 1$ and that $b = a_1 \cup \cdots \cup a_{m-1}$ has the form (3.3).

We can replace $a_m$ by $[a_m - (a_m \cap b)]$ since the hypotheses of Lemma 3.2 continue to hold and the conclusion is not changed. After this change,

$$a_m \cap b = a_m \cap (a_1 \dot\cup \bar{a}_2 \dot\cup \cdots \dot\cup \bar{a}_n) = 0 .$$

Since $a_m \lesssim a_1 \dot\cup \bar{a}_2 \dot\cup \cdots \dot\cup \bar{a}_n$ there is a perspectivity mapping $\varphi$ of $[0, a_m]$ with $\varphi(a_m) \leqq b$. Then

$$a_m = a_{m,1} \dot\cup a_{m,2} \dot\cup \cdots \dot\cup a_{m,n}$$

where

$$\varphi(a_{m,1}) = \varphi(a_m) \cap a_1 ,$$

and for $1 < i \leqq n$,

$$\varphi(a_{m,i}) = [(\varphi(a_m) \cap (a_1 \dot{\cup} \bar{a}_2 \dot{\cup} \cdots \dot{\cup} \bar{a}_i)) - (\varphi(a_m) \cap (a_1 \dot{\cup} \bar{a}_2 \dot{\cup} \cdots \dot{\cup} \bar{a}_{i-1})] .$$

Obviously, $a_{m,1} \lesssim a_1$. If $i > 1$ then $a_{m,i} \sim \varphi(a_{m,i})$; $\varphi(a_{m,i}) \lesssim \bar{a}_i$; $\bar{a}_i \lesssim a_1$; and $a_{m,i} \cap (\varphi(a_{m,i}) \cup \bar{a}_i \cup a_1) = 0$; these facts imply that $a_{m,i} \lesssim a_1$ (use (2.2) of [1]). The conclusion of Lemma 3.2 now follows at once.

LEMMA 3.3. *Suppose that*

(i) $a = a_1 \cup a_2 \cup \cdots \cup a_m$ *for some finite* $m \geqq 2$,
(ii) $a_2 \sim a_1$,
(iii) $a_i \lesssim a_1 \cup \cdots \cup a_{i-1}$ *for* $2 < i \leqq m$,
(iv) $[0, a_1 \dot{\cup} a_2]$ *is upper* $\aleph$*-complete*,
(v) $u(a_1, a_2, \aleph)$ *holds.*

*Then* $[0, a]$ *is upper* $\aleph$*-complete.*

*Proof.* Applying Lemma 3.2, and using a new $m$ and new elements $a_3, \cdots, a_m$ we may suppose that (i), (iii) hold in the strengthened form: $a = a_1 \dot{\cup} a_2 \dot{\cup} \cdots \dot{\cup} a_m$ and $a_i \lesssim a_1$ for $2 < i \leqq m$.

Suppose that $1 \leqq i < j \leqq m$. If $i \neq 2$ then $a_j \lesssim a_2$ (because of (ii)) and there is a perspectivity mapping $\varphi$ of $[0, a_i \cup a_j]$ with $\varphi(a_i) \leqq a_1$ and $\varphi(a_j) \leqq a_2$. Hence $[0, a_i \cup a_j]$ is upper $\aleph$-complete and $u(a_i, a_j, \aleph)$ holds in this case.

If $i = 2$ there is a perspectivity mapping $\varphi$ of $[0, a_2 \cup a_j]$ with $\varphi(a_2) = a_1$, $\varphi(a_j) = a_j$; the result for $[0, a_1 \cup a_j]$ obtained previously now implies: $[0, a_2 \cup a_j]$ is upper $\aleph$-complete and $u(a_2, a_j, \aleph)$ holds.

Corollary 1 to Lemma 3.1 now applies to these elements $a_1, \cdots, a_m$ and this completes the proof of Lemma 3.3.

COROLLARY. *Suppose that the hypotheses* (i), (ii), (iii), *of Lemma* 3.3 *hold and suppose also that*

(vi) $[0, a_1 \dot{\cup} a_2]$ *is upper* $\aleph$*-complete and upper* $\aleph$*-continuous.*

*Then* $[0, a]$ *is upper* $\aleph$*-complete and upper* $\aleph$*-continuous.*

*Proof.* (vi) implies (iv), (v). Hence $[0, a]$ is upper $\aleph$-complete by Lemma 3.3. Upper $\aleph$-continuity then follows from [1, Theorem 4.3].

LEMMA 3.4. (*Additivity of lower* $\aleph$*-continuity*). *Suppose that* $[0, a_1 \cup \cdots \cup a_m]$ *is lower* $\aleph$*-complete and that* $[0, a_i]$ *is lower* $\aleph$*-*

*continuous for* $i=1,\cdots,m$. *Then* $[0, a_1 \cup \cdots \cup a_m]$ *is lower* $\aleph$*-continuous.*

*Proof.* We may assume that $\{a_1, \cdots, a_m\}$ is an independent set (replace $a_i$ by $[a_i - (a_i \cap (a_1 \cup \cdots \cup a_{i-1}))]$ for $2 \leqq i \leqq m$).

Then $[a_1, a_1 \cup a_2]$ is lower $\aleph$-continuous since it is lattice isomorphic to $[0, a_2]$ under the mapping: $x \rightarrow x \cap a_2$. Similarly $[a_2, a_1 \cup a_2]$ is lower $\aleph$-continuous. By the dual of [1, Theorem 4.3], $[0, a_1 \cup a_2] = ([a_1 \cap a_2, a_1 \cup a_2])$ is lower $\aleph$-continuous. Lemma 3.4 follows by induction on $m$.

LEMMA 3.5. *Suppose that each of* $[0, a \cup b]$, $[0, b \cup c)$, $[0, a \cup c]$ *is lower* $\aleph$*-complete and suppose that* $l(a, c, \aleph)$ *holds. Then* $[0, a \cup b \cup c]$ *is lower* $\aleph$*-complete.*

*Proof.* We may suppose that $\{a, b, c\}$ is an independent set, for if $c, b$ are replaced by $[c - (a \cap c)]$ and $[b - (b \cap (a \cup c))]$ respectively the hypotheses of Lemma 3.5 continue to hold ($l(a, c_1, \aleph)$ is equivalent to $l(a, c, \aleph)$ if $a \cup c_1 = a \cup c$) and the conclusion is not changed.

Now set $B = a \cup c$, $C = b \cup a$, $A = b \cup c$, and $1 = a \cup b \cup c$. We have: $[A \cap B, 1] (= [c, a \cup b \cup c])$ is lower $\aleph$-complete since it is lattice isomorphic to $[0, a \cup b]$ under the mapping $x \rightarrow x \cap (a \cup b)$. Similarly each of $[B \cap C, 1]$, $[C \cap A, 1]$ is lower $\aleph$-complete.

We can now show that $[0, a \cup b \cup c] (= [A \cap B \cap C, 1])$ is lower $\aleph$-complete (by applying the dual of Lemma 3.1) if we can show:

(3.4) *Whenever* $X_\alpha \geqq C \cap A$ *for* $\alpha \in I$ *(with* $\bar{I} \leqq \aleph$*) and* $C \cup (\bigcap(X_\beta | \beta \in F)) = 1$ *for all finite* $F \subset I$, *then* $C \cup (\bigcap(X_\alpha | \alpha \in I)) = 1$.

Since $C \cap A = b$ and $C = a \cup b$, (3.4) can be rewritten:

(3.4)′ *Whenever* $X_\alpha \geqq b$ *for* $\alpha \in I$ *(with* $\bar{I} \leqq \aleph$*) and* $a \cup (\bigcap(X_\beta | \beta \in F)) = a \cup b \cup c$ *for all finite* $F \subset I$ *then* $a \cup (\bigcap(X_\alpha | \alpha \in I)) = a \cup b \cup c$.

Suppose that the hypotheses of (3.4)′ hold and set $x_\alpha = X_\alpha \cap (a \cup c)$. Then $x_\alpha \leqq a \cup c$ for all $\alpha$ and

$$\begin{aligned} &a \cup (\bigcap (x_\beta | \beta \in F)) \\ &= a \cup ((\bigcap(X_\beta | \beta \in F)) \cap (a \cup c)) = (a \cup (\bigcap (X_\beta | \beta \in F))) \cap (a \cup c) \\ &= (a \cup b \cup c) \cap (a \cup c) = a \cup c\,. \end{aligned}$$

Since $l(a, c, \aleph)$ holds, it follows that

$$\begin{aligned} &a \cup (\bigcap (x_\alpha | \alpha \in I)) = a \cup c\,; \ a \cup (\bigcap (X_\alpha | \alpha \in I) \cap (a \cup c)) = a \cup c\,; \\ &a \cup (\bigcap (X_\alpha | \alpha \in I)) \geqq a \cup c \text{ (hence} = a \cup b \cup c)\,. \end{aligned}$$

This means: (3.4)′ does hold. This completes the proof of Lemma 3.5.

COROLLARY 1. *Suppose that* $[0, a_i \cup a_j]$ *is lower* $\aleph$*-complete for* $i, j = 1, \cdots, m$.

*Suppose also that* $l(a_i, a_j, \aleph)$ *holds for all* $i<j$. *Then* $[0, a_1 \cup \cdots \cup a_m]$ *is lower* $\aleph$*-complete.*

*Proof.* This follows from Lemma 3.5 by induction on $m$, just as Corollary 1 to Lemma 3.1 followed from Lemma 3.1.

COROLLARY 2. *Suppose that* $[0, a_i \cup a_j]$ *is lower* $\aleph$*-complete and lower* $\aleph$*-continuous for* $i, j = 1, \cdots, m$. *Then* $[0, a_1 \cup \cdots \cup a_m]$ *is lower* $\aleph$*-continuous.*

*Proof.* Since lower $\aleph$-continuity of $[0, a_i \cup a_j]$ implies that $l(a_i, a_j, \aleph)$ holds, Corollary 1 shows that $[0, a_1 \cup \cdots \cup a_m]$ is lower $\aleph$-complete. The lower $\aleph$-continuity of $[0, a_1 \cup \cdots \cup a_m]$ then follows from Lemma 3.4.

LEMMA 3.6. *Suppose that*

(i) $a = a_1 \cup a_2 \cup \cdots \cup a_m$ *for some finite* $m \geqq 2$,
(ii) $a_2 \sim a_1$,
(iii) $a_i \lesssim a_1 \cup \cdots \cup a_{i-1}$ *for* $2 < i \leqq m$,
(iv) $[0, a_1 \dot\cup a_2]$ *is lower* $\aleph$*-complete,*
(v) $l(a_1, a_2, \aleph)$ *holds.*

*Then* $[0, a]$ *is lower* $\aleph$*-complete.*

COROLLARY. *Suppose that* (i), (ii), (iii) *hold and also*

(vi) $[0, a_1 \dot\cup a_2]$ *is lower* $\aleph$*-complete and lower* $\aleph$*-continuous.*

*Then* $[0, a]$ *is lower* $\aleph$*-complete and lower* $\aleph$*-continuous.*

*Proof.* Lemma 3.6 and its Corollary follow from Lemma 3.5 and Lemma 3.4 just as Lemma 3.3 and its Corollary followed from Corollary 1 to Lemma 3.1 and [1, Theorem 4.3].

THEOREM 3.1. *Suppose that each of* $[0, a_i \cup a_j]$ *is an* $\aleph$*-von Neumann-geometry* (*respectively a von Neumann-geometry*) *for* $i, j = 1, \cdots, m$. *Then* $[0, a_1 \cup \cdots \cup a_m]$ *is an* $\aleph$*-von Neumann-geometry* (*respectively a von Neumann geometry*).

*Proof.* This follows from Corollary 2 to Lemma 3.1 and Corollary 2 to Lemma 3.5.

COROLLARY 1. *Suppose that*

(i) $a = a_1 \cup a_2 \cup \cdots \cup a_m$ *for some finite* $m \geqq 2$,

(ii) $a_2 \sim a_1$,
(iii) $a_i \lesssim a_1 \cup \cdots \cup a_{i-1}$ *for* $2 < i \leqq m$,
(iv) $[0, a_1 \dot{\cup} a_2]$ *is an* $\aleph$*-von Neumann-geometry (respectively a von Neumann-geometry).*

*Then* $[0, a]$ *is an* $\aleph$*-von Neumann-geometry, respectively a von Neumann-geometry.*

*Proof.* This follows from the Corollary to Lemma 3.3 and the Corollary to Lemma 3.6.

COROLLARY 2. *Suppose that* $\mathscr{R}$ *is an* $\aleph$*-von Neumann-ring (respectively a von Neumann-ring). If* $\bar{R}_{\mathscr{R}}$ *has a basis* $x_1, x_2, \cdots, x_m$ *such that* $x_2 \sim x_1$ *and* $x_i \lesssim x_1$ *for* $2 < i \leqq m$, *then* $\mathscr{R}_2$ *is an* $\aleph$*-von Neumann-ring (respectively, a von Neumann-ring).*

*Proof.* By hypothesis, the unit element of the lattice $\bar{R}_{\mathscr{R}}$ is the union $x_1 \dot{\cup} \cdots \dot{\cup} x_m$. The unit element of $\bar{R}_{\mathscr{S}}$, with $\mathscr{S} = \mathscr{R}_2$, can be represented as a union $x_1 \dot{\cup} \cdots \dot{\cup} x_m \dot{\cup} y_1 \dot{\cup} \cdots \dot{\cup} y_m$ with $y_i \sim x_i$ and hence $y_i \lesssim x_1$ for $1 \leqq i \leqq m$. Since $[0, x_1 \dot{\cup} x_2]$ is an $\aleph$-von Neumann geometry (respectively a von Neumann geometry) along with $\bar{R}_{\mathscr{R}}$, Corollary 1 applies and this completes the proof of Corollary 2.

COROLLARY 3. *Suppose that* $\mathscr{R}$ *and* $\mathscr{R}_2$ *are both* $\aleph$*-von Neumann-rings (respectively von Neumann-rings). Then* $\mathscr{R}_n$ *is an* $\aleph$*-von Neumann-ring (respectively a von Neumann-ring) for all finite* $n$.

*Proof.* If $n > 2$ the unit element of $\bar{R}_{\mathscr{S}}$, with $\mathscr{S} = \mathscr{R}_n$, can be expressed as $x_1 \dot{\cup} x_2 \dot{\cup} \cdots \dot{\cup} x_n$ where $x_1$ is the unit element of $\bar{R}_{\mathscr{R}}$, $x_i \sim x_1$ for all $i$, and $[0, x_1 \dot{\cup} x_2] = \bar{R}_{\mathscr{R}_2}$. Theorem 3.1 applies and this completes the proof of Corollary 3.

REMARK. Let $\mathscr{R}$ be the ring of sequences $x = (x^n)$ with all $x^n$ complex numbers and all but a finite number of $x^n$ real, with componentwise addition and multiplication; this example was given by Kaplansky [3, page 526]. This $\mathscr{R}$ is a von Neumann-ring but $\mathscr{R}_2$ is not even upper $\aleph_0$-complete.

DEFINITION 3.1. If $L$ is a relatively complemented modular lattice, then an element $a$ is called Boolean (with respect to $L$) if $b_1 \sim b_2$, $b_1 \leqq a$ together imply $b_1 = b_2$; $a$ is called the *Boolean part* of $L$ (necessarily unique if it exists)[2] if $a$ is Boolean and $a_1 \leqq a$ for every Boolean $a_1$.

[2] This is an abuse of language: properly, $[0, a]$ should be called the Boolean part of $L$.

LEMMA 3.7. *Suppose that $L$ is a relatively complemented modular lattice. If $(a, b)P$ holds then for every $c$ in $L$, $c \cap (a \cup b) = (c \cap a) \cup (c \cap b)$ and $[0, a \cup b]$ is the direct sum of $[0, a]$ and $[0, b]$. On the other hand if $a$ is Boolean then*

(i) *$b \leqq a$ implies that $b$ is Boolean,*
(ii) *$b \cap a = 0$ implies that $(b, a)P$ holds,*
(iii) *$b \geqq a$ implies that the relative complement $[b - a]$ is unique,*
(iv) *$a \cap (b \cup c) = (a \cap b) \cup (a \cap c)$ for all $b, c$ in $L$,*
(v) *$[0, a]$ is a Boolean algebra.*

*Proof.* Suppose that $(a, b)P$ holds and set $d = [(c \cap (a \cup b)) - ((c \cap a) \cup (c \cap b))]$, $d_a = (d \cup b) \cap a$, $d_b = (d \cup a) \cap b$. Then $d \leqq a \cup b$, $d \cap a = d \cap b = 0$, $d_a \dot{\cup} d = (d \cup b) \cap (d \cup a) = d_b \dot{\cup} d$, so $d_a \sim d_b$. Since $d_a \leqq a$, $d_b \leqq b$ and $(a, b)P$ holds, we must have: $d_a = 0$; $b = d_a \cup b = d \cup b$; $d \leqq b$; hence $d = 0$, $c \cap (a \cup b) = (c \cap a) \cup (c \cap b)$. If $c \leqq a \cup b$ then $c = (c \cap a) \cup (c \cap b)$; and if $c = c_1 \cup c_2$ with $c_1 \leqq a$, $c_2 \leqq b$ then $c \cap a = c_1 \cup (c_2 \cap b \cap a) = c_1 \cup 0 = c_1$, $c \cap b = c_2$. This proves that $[0, a \cup b]$ is the direct sum of $[0, a]$ and $[0, b]$.

(i) and (ii) are obvious from the definition of Boolean element.

(ii) asserts that $a$ is in the centre of $L$ as defined in [**1**, (2.5)]. But if $a$ is in the centre of $L$ and $b$ is any element in $L$ with $b \geqq a$ then $a$ is in the centre of $[0, b]$, hence $[b - a]$ is uniquely determined (use [**1**, (2.6)]). This proves (iii).

If $b, c$ are arbitrary elements in $L$, set $b_1 = [b - (a \cap b)]$, $c_1 = [c - (a \cap c)]$. Since $a \cap b_1 = a \cap c_1 = 0$ and $a$ is in the centre of $L$, it follows that $(a, b_1)P$, $(a, c_1)P$, hence $(a, b_1 \cup c_1)P$ (use [**1**, (2,6)]); therefore $a \cap (b_1 \cup c_1) = 0$. By the modular law

$$\begin{aligned} a \cap (b \cup c) &= a \cap (b_1 \cup c_1 \cup (a \cap b) \cup (a \cap c)) \\ &= (a \cap b) \cup (a \cap c) \cup (a \cap (b_1 \cup c_1)) \\ &= (a \cap b) \cup (a \cap c) \end{aligned}$$

and hence (iv) holds.

Thus $[0, a]$ is a distributive complemented lattice, equivalently: a Boolean algebra. This proves (v).

LEMMA 3.8. *Suppose that $L$ has a unit element $1 = a_1 \cup a_2 \cup \cdots \cup a_m$ with $m \geqq 2$, $a_2 \sim a_1$, $a_i \lesssim a_1$ for $2 < i \leqq m$ and $a_1 \cap a_2 = 0$. Then the Boolean part of $L$ exists and is 0.*

*Proof.* By Lemma 3.2 we may assume that $1 = a_1 \dot{\cup} \cdots \dot{\cup} a_m$ with $m \geqq 2$, $a_2 \sim a_1$ and $a_i \lesssim a_1$ for $2 < i \leqq m$.

To prove Lemma 3.8 we may suppose that $a \neq 0$ and we need only exhibit elements $b_1, b_2$ such that $b_1 \leqq a$, $b_1 \sim b_2$, and $b_1 \neq b_2$.

If $a_i \cap a \neq 0$ for any $i$ it suffices to choose this element as $b_1$ since the relations $a_1 \sim a_2$ and $a_i \lesssim a_1$ if $i \neq 1$ imply $b_1 \sim b_2$ for some $b_2 \neq b_1$ (even $b_1 \cap b_2 = 0$).

On the other hand, if $a_i \cap a = 0$ for all $i$, set $b_1 = (a_1 \cup \cdots \cup a_i) \cap a$ where $i$ is the smallest integer for which this element is different from 0 (necessarily $1 < i \leqq m$) and set $b_2 = ((a_1 \cup \cdots \cup a_{i-1}) \dot\cup b_1) \cap a_i$. Then $b_1 \sim b_2$ since $(a_1 \cup \cdots \cup a_{i-1}) \dot\cup b_1 = (a_1 \cup \cdots \cup a_{i-1}) \dot\cup b_2$; and $b_1 \neq b_2$ since $b_2 \leqq a_i$ and $b_1 \cap a_i \leqq a \cap a_i = 0$. This completes the proof of Lemma 3.8.

LEMMA 3.9. *Suppose that $L$ is an upper complete complemented modular lattice and let $a$ be the union of all Boolean elements in $L$. Then $a$ is the Boolean part of $L$.*

*Proof.* We need only show that $a$ is Boolean, that is, we may suppose that $b \leqq a$, that $\varphi$ is a perspective mapping of $[0, b]$, that $b \neq \varphi(b)$ and we need only derive a contradiction. By replacing $b$ by $[b - (b \cap \varphi(b))]$ we may suppose $b \neq 0$ and $b \cap \varphi(b) = 0$.

Now for every $c$: $(\varphi(b \cap c)) \sim (b \cap c)$ and $(\varphi(b \cap c)) \cap (b \cap c) = 0$. If $c$ is Boolean this implies: $b \cap c = 0$, and hence (since $c$ is Boolean) $(b, c)P$ holds. It follows from [1, formula (2.6)] that $(b, a)P$ holds, contradicting the fact that $b \neq 0$ and $b \leqq a$. This contradiction proves Lemma 3.9.

THEOREM 3.2. *Suppose that $L$ is a relatively complemented modular lattice and*

(i) $a = a_0 \cup a_1 \cup a_2 \cup \cdots \cup a_m$ *for some finite* $m \geqq 2$,
(ii) $(a_0, a_1 \cup \cdots \cup a_m)P$ *holds,*
(iii) $a_2 \sim a_1$, $a_2 \cap a_1 = 0$,
(iv) $a_i \lesssim a_1 \cup \cdots \cup a_{i-1}$ *for* $2 < i \leqq m$,
(v) $\varphi$ *is a perspective mapping of* $[0, b]$ *with* $\varphi(b) \leqq a$.

*Let $\pi$ denote one of the properties: to be upper $\aleph$-complete and upper $\aleph$-continuous, or to be lower $\aleph$-complete and lower $\aleph$-continuous. Then $[0, a \cup b]$ has property $\pi$ if both of $[0, a_1 \cup a_2]$ and $[0, a_0 \cup \varphi^{-1}(a_0 \cap \varphi(b))]$ have property $\pi$; if $a_0$ is the Boolean part of $[0, a]$ and $[0, b]$ has a Boolean part $b_0$, it is sufficient that $[0, a_1 \cup a_2]$ and $[0, a_0 \cup b_0]$ should both have property $\pi$.*

*Proof.* Since $(a_0, a_1 \cup \cdots \cup a_m)P$ holds, Lemma 3.7 shows that $\varphi(b) = \varphi(b_1) \dot\cup \varphi(b_2)$ where $b_1 = \varphi^{-1}(a_0 \cap \varphi(b))$ and $b_2 = \varphi^{-1}((a_1 \cup \cdots \cup a_m) \cap \varphi(b))$. Then $(a_0 \cup b_1, a_1 \cup \cdots \cup a_m \cup b_2)P$ holds (use [1, (2.6)]).

By Lemma 3.7, $[0, a \cup b]$ is the direct sum of $[0, a_0 \cup b_1]$ and $[0, a_1 \cup \cdots \cup a_m \cup b_2]$ and has property $\pi$ if each of the summands has it.

Since $b_2 \lesssim a_1 \cup \cdots \cup a_m$, $[0, a_1 \cup \cdots \cup a_m \cup b_2]$ has property $\pi$ if $[0, a_1 \cup a_2]$ has it, by Lemma 3.3 and its Corollary and Lemma 3.6 and its Corollary.

If $a_0$ is the Boolean part of $[0, a]$ then $\varphi(b) \cap a_0$ is Boolean with respect to $[0, a]$, a fortiori Boolean with respect to $[0, \varphi(b)]$. Thus, $b_1$ is Boolean with respect to $[0, b]$. If $[0, b]$ has a Boolean part $b_0$ then $b_1 \leqq b_0$ and $a_0 \cup b_1 \leqq a_0 \cup b_0$, hence $[0, a_0 \cup b_1]$ has property $\pi$ if $[0, a_0 \cup b_0]$ has it.

This proves all parts of Theorem 3.2.

REMARK. If $\mathscr{R}$ is a von Neumann ring then $\mathscr{R}$ has a unique decomposition as a direct sum $\mathscr{R} = \mathscr{B} \oplus \mathscr{R}$ such that $\bar{R}_{\mathscr{B}}$ is the Boolean part of $\bar{R}_{\mathscr{R}}$ and $\bar{R}_{\mathscr{R}}$ has a basis $x_1, x_2, x_3$ with $x_2 \sim x_1$ and $x_3 \lesssim x_1$. Then Theorem 3.2 and Corollary 2 to Theorem 3.1 apply and show that $\mathscr{R}_2$ is a von Neumann ring if and only if $\mathscr{B}_2$ is a von Neumann ring (for details see [2]).

## REFERENCES

1. Ichiro Amemiya and Israel Halperin, *Complemented modular lattices*, Canadian J. of Math., **11** (1959), 481-520.

2. Israel Halperin, *Elementary divisors in von Neumann rings*, Acta Scientiarum Mathematicarum Szeged, to appear.

3. Irving Kaplansky, *Any orthocomplemented complete modular lattice is a continuous geometry*, Annals of Math., **61** (1955), 524-541.

4. John von Neumann, *Continuous Geometry*, Princeton University Press, 1960.

QUEEN'S UNIVERSITY,
KINGSTON, CANADA.

# ARC-WISE CONNECTEDNESS IN SEMI-METRIC SPACES

ROBERT W. HEATH

**1. Introduction.** The Arc Theorem usually encountered is the following: a connected and locally connected Cauchy complete metric[1] space is arc-wise connected [10]. The most general Arc Theorem is Theorem 1 in Chapter II of [14], in which "Cauchy complete metric space" is replaced by "a space satisfying Moore's Axiom 1"—i.e. a "complete Moore space" (equivalent to a complete regular developable space [1]; see also [16] and [15]). Wyman Richardson, in one of F. B. Jones' classes proved the Arc Theorem for strongly complete regular semi-metric spaces (unpublished though the argument differed considerably from Moore's argument). This was not, however, a real generalization because such spaces are Moore spaces (cf. Corollary 4.3 of this paper).

Since most theorems which are true in Moore spaces are true in regular semi-metric[1] spaces, and since the exceptions are "in general those theorems whose validity depends upon that property of Moore spaces which forces the equivalence of perfect and hereditary separability" [7], one might hope that the Arc Theorem could be further generalized by simply replacing "metric" by "regular semi-metric." This paper establishes that the Arc Theorem cannot be generalized directly to Cauchy complete regular semi-metric spaces but can be extended to a somewhat more general class of regular semi-metric spaces then those satisfying Moore's Axiom 1. The examples given show that, even in the presence of such properties as possessing a uniformity and being compactly connected, a regular semi-metric space can be Cauchy complete, connected and locally connected but not be arc-wise connected. Other possible means of extending the Arc Theorem are eliminated by establishing that in the presence of certain topological properties a regular semi-metric space is a Moore space (*e.g.* a strongly complete semi-metric space is a Moore space)—or is even metrizable.

This paper is essentially a dissertation [4] written at the University of North Carolina under the direction of Professor F. B. Jones. The

Received July 22, 1960, and in revised form May 17, 1961, and December 20, 1961. Presented to the American Mathematical Society, June 18, 1960.

[1] A topological space $S$ is said to be *semi-metric* if there is a distance function $d$ for $S$ with respect to which the topology of $S$ is invariant. A *distance function $d$ for $S$* is a function from $S \times S$ to the nonnegative numbers such that, if each of $x$ and $y$ is a point of $S$, then (1) $d(x, y) = 0$ only in case $x = y$ and (2) $d(x, y) = d(y, x)$ [**11**; **18**]. The space is *metric* if the distance function also satisfies (3) $d(x, y) + d(y, x) \geqq d(x, z)$ for each triple $x, y, z$ of points of $S$. Note that every Moore space is regular and semi-metric. The set $U_c(x) = \{y: d(x, y) < c\}$ is referred to herein as a $c$-neighborhood (with respect to $d$) of $x$. *Cauchy complete* is defined as in [11, p. 316]. *Topological space* and *regular* are defined as in [9, pp. 37 and 113]. Terms not defined herein are used as in [14], [11], or [1].

author wishes to thank Professor Jones for his encouragement and direction.

**2. Cauchy complete semi-metric spaces in which the Arc Theorem does not hold true.** The following examples and theorems show that a Cauchy complete regular semi-metric space may be connected and locally connected (and even compactly connected) without being arc-wise connected. Example 2.2 is such a space, some additional properties of which are given in Theorem 2.3. Example 2.5 is such a space which is compactly connected. In the remainder of this section some additional properties of those spaces are pointed out to show that those properties could not be used to extend the Arc Theorem, and there are described some other spaces which are practically indistinguishable from the first two spaces but which are arc-wise connected. The following definition will be useful since weak completeness is equivalent to Cauchy completeness [11, Theorem 2.3].

DEFINITION 2.1. A space $S$ is said to be $\begin{Bmatrix}\text{weakly complete}\\ \text{strongly complete}\end{Bmatrix}$ provided there exists a distance function $d$ such that (1) the topology of $S$ is invarient with respect to $d$ and (2) if $M$ is a nonincreasing sequence of closed sets in $S$ such that, for each $n$, there is a $1/n$-neighborhood of a point $p_n\begin{Bmatrix}\text{in } M_n\\ \text{in } S\end{Bmatrix}$ which contains $M_n$, then $\prod_{n=1}^{\infty} M_n$ contains a point.

EXAMPLE 2.2. Let $S$ consist of the points of $[0, 1] \times [0, 1]$ with a distance function $d$ and a topology defined as follows.

(1) If $x \in S$, $d(x, x) = 0$, and

(2) if $x$ and $y$ are two points of $S$ and $a(x, y)$ is the smallest nonnegative angle (in radians) formed by the line which contains $x$ and is parallel to $X$ (the $x$-axis), or is $X$, and the line which contains $x$ and $y$, then $d(x, y) = |x - y| + a(x, y)$. For each point $p$ of $S$ and each positive number $c$, let the $c$-neighborhood of $p$, $U_c(p) = \{x: d(x, p) < c\}$, be an element of a basis for the topology of $S$.

Clearly the semi-metric space $S$ is weakly complete (hence Cauchy complete), completely regular (hence uniform and, of course, regular; cf. [9]) and separable. That $S$ is connected and locally connected follows from the fact that horizontal line segments in $[0, 1] \times [0, 1]$ have the same relative topology in $S$ as in Euclidean two-space.

Note that, if $(b, c)$ is a point of $S$, if $d > 0$, if $0 < a < \pi/2$, and if $R(b, c; a, d) = [\{(x, y): |x - b| < d,$ and either $|y - c| < |x - b|$ tan $a$ or $y = c\}]\cdot S$—i.e. if $R(b, c; a, d)$ is the point set consisting of $(b, c)$ and all points of $S$ interior to a (horizontally oriented) "bow-tie region" centered on $(b, c)$ and having (horizontal) length $d$ and central angles of magnitude $2a$ radians—then $R(b, c; a, d)$ is an open set in $S$. Furthermore, the collection $\{R(b, c; a, d): (b, c) \in S, d > 0 \text{ and } 0 < a < \pi/2\}$ of all such bow-tie regions in $S$ forms a basis for the topology of $S$. That basis is useful in the proof of Theorem 2.3.

The proof of the following lemma, as well as a more detailed proof of Theorem 2.3, is contained in the proof of Theorem 1 [4, p. 10].

LEMMA. *If $M$ is an arc in $S$ with nondegenerate $x$ and $y$ projections, then there is a subarc $M_1$ of $M$ with nondegenerate $x$ and $y$ projections and whose $x$-projection is a subset of the $x$-projection of $M - M_1$.*

THEOREM 2.3. *There exists a separable, connected, locally connected, weakly complete, completely regular, semi-metric space which is not arc-wise connected.*

*Proof.* Let $S$ be the space of Example 2.2. The space $S$ is not arc-wise connected. For suppose that there is an arc in $S$ with endpoints $(0, 0)$ and $(1, 1)$.

By the above lemma, if $M$ is an arc in $S$ and with endpoints $(0, 0)$ and $(1, 1)$, there is a sequence $\{M_i\}_{i=1}^{\infty}$ of subarcs of $M$ such that, for each $n$, $M_n \supset M_{n+1}$ and $\{x: (x, y) \in M_{n+1}\} \subset \{x: (x, y) \in [M_n - M_{n+1}]\}$. Then, since $M$ is compact, there is a point $(p, q)$ in $\prod_{i=1}^{\infty} M_i$, and, for each $n$, there is a point $(p, q_n)$ in $M_n$ such that $q_n = q_m$ only if $n = m$. Thus $M$ contains an infinite subset without a limit point which violates the compactness of $M$.

DEFINITION 2.4. A space $S$ is said to be *compactly connected* provided that, if $a$ and $b$ are points of $S$, $S$ contains a compact continuum which contains both $a$ and $b$.

There are now two questions to be answered. Is a connected, locally connected, weakly complete, regular (or completely regular) semi-metric space compactly connected? Also, is a connected, locally connected, weakly complete, regular (or completely regular) semi-metric space which is compactly connected necessarily arc-wise connected? The answer to both questions is no. It can be shown that the space of Example 2.2 is not compactly connected by an argument in general following the same outline as the proof of Theorem 2.3—replacing subarc by irreducible subcontinuum and making use of Theorems 32, 39, and 47 from Chapter I of Moore's *Foundations* (for a detailed proof see Theorem 2 [4, p. 13]). Example 2.5 and Theorem 2.6 answer the second question.

EXAMPLE 2.5. Let $K$ be the "polyhedral $\sin 1/x$ curve" in $[0, 1] \times$

[0, 1] which is the union of all horizontal line segment of the form

$$\{x: 0 \leqq \mathrm{Re}\,[x] \leqq 1 \text{ and } \mathrm{Im}\,[x] = 0\}$$

or

$$\{x: 0 \leqq \mathrm{Re}\,[x] \leqq 1 \text{ and } \mathrm{Im}\,[x] = 1/2n\} \qquad \text{for } n = 0, 1, 2, \cdots,$$

and of all vertical line segments of the form

$$\left\{x: \mathrm{Re}\,[x] = 0, \frac{1}{2^{2m+1}} \leqq \mathrm{Im}\,[x] \leqq \frac{1}{2^{2m}}\right\} \qquad \text{for } m = 0, 1, 2, \cdots$$

or

$$\left\{x: \mathrm{Re}\,[x] = 1, \frac{1}{2^{2m}} \leqq \mathrm{Im}\,[x] \leqq \frac{1}{2^{2m-1}}\right\} \qquad \text{for } m = 1, 2, 3, \cdots$$

Let $d$ be the distance function for $[-2, 2] \times [0, 1]$ defined as follows:

(1) if $x$ is a point of $K$ and $y$ is a point of $[-2, 2] \times [0, 1]$, then $d(x, y) = |x - y|$;

(2) if $x$ and $y$ are points of $([-2, 2] \times [0, 1] - K)$ and $a(x, y)$ is the smallest nonnegative angle (in radians) formed by the line $xy$ and the horizontal, then $d(x, y) = |x - y| + a(x, y)$.

Let $S$ be the topological space consisting of the set $[-2, 2] \times [0, 1]$ with the following topology: for each point $p$ of $K$ and each $c > 0$, the circular neighborhood $\{x: x \in S, |x - y| < c\}$ is a region in $S$, and, if $p$ is a point of $S$ but not of $K$, every "bow-tie region" (as defined in Example 2.2) with center at $p$ is a region in $S$. Clearly $S$ is a completely regular semi-metric space which is weakly complete, connected, locally connected, and separable.

THEOREM 2.6. *There is a connected, locally connected, weakly complete, completely regular semi-metric space which is compactly connected but not arc-wise connected.*

*Proof.* Let $S$ be the space defined in Example 2.5. Since $K$ has the same relative topology as it would in the usual plane topology, $K$ is a compact continuum; likewise each horizontal interval contained in $S$ is a compact continuum. Hence, if $a$ and $b$ are points of $S$, the point set $[K + \{x: x \in S \text{ and } \mathrm{Im}\,[x] = \mathrm{Im}\,[a]\} + \{x: x \in S \text{ and } \mathrm{Im}\,[x] = \mathrm{Im}\,[b]\}]$ is a compact continuum which contains $a$ and $b$ and is contained in $S$. Thus $S$ is compactly connected.

From the proof of Theorem 2.3, it is clear that any nondegenerate compact continuum in $S$, other than a horizontal interval, must intersect $K$. Therefore, a compact continuum $L$ in $S$ which contains the points $(2, 1)$ and $(-2, 0)$ must contain $K$; but each point of $\{z: z \in K, 0 < \mathrm{Re}\,[z] < 1$ and $\mathrm{Im}\,[z] = 0\}$ is a nonseparating point of $K$ and is not a boundary point of any component of $L - K$ (and therefore a nonseparating point of $L$). Hence there is no arc in $S$ which contains $(2, 1)$ and $(-2, 0)$.

Returning to the space $S$ of Example 2.2, it is perhaps of interest whether points of that space are connectable by any type of sets bearing some resemblance to arcs. It will be shown in § VI that, if $a$ and $b$ are points of $S$, there is a continuum $M$ in $S$ whose only nonseparating points are $a$ and $b$. It can also be shown that, if $a$ and $b$ are points of $S$, then either $a$ and $b$ are endpoints of an arc in $S$ (in case $a$ and $b$ are on the same horizontal line segment), or $a$ and $b$ are the only nonseparating points of a connected subset $M$ of $S$ which is the graph of a function, namely $M^{-1} = \{(x, y): (y, x) \in M\}$ is a function. The existence of the latter can be established by an argument somewhat similar to the proof of Theorem 5 in [6]. For a detailed proof see [4, pp. 20–24].

Consider now the following two examples each of which is a connected, locally connected, weakly complete, completely regular, semi-metric space which closely resembles Example 2.2, and each of which also is neither a Moore space nor strongly complete (nor complete in any of the "intermediate" senses to be subsequently defined), but each of which *is* arc-wise connected.

EXAMPLE 2.7. Define a distance function $d$ for the points of $[0, 1] \times [0, 1]$ as follows: if $x$ and $y$ are two points of $[0, 1] \times [0, 1]$, then (1) if $\mathrm{Re}\,[x] = \mathrm{Re}\,[y]$, $d(x, y) = 1$, and (2) if $\mathrm{Re}\,[x] \neq \mathrm{Re}\,[y]$, then $d(x, y) = 2(\text{g.l.b.}\,[c: c > 0$ and $y \in \{z: |z - (x + c)| < c$ or $|z - (x - c)| < c\}]) = |x - y|^2/|\mathrm{Re}\,x - \mathrm{Re}\,y| = |x - y| \sec \alpha$, where $\alpha$ is the smallest nonnegative angle formed by the line passing through $x$ and $y$ and the horizontal line through $x$. Note that neighborhoods of radius less than 1 are "bow-ties" formed by tangent circles (the center of such a neighborhood is the point of tangency, and, if the neighborhood has center $x$ and radius $2c$, the centers of the circles are $x - c$ and $x + c$ and the radius of each circle is $c$). The topological space $S$ consisting of the points of $[0, 1] \times [0, 1]$ and regions which are (all) such neighborhoods is a connected, locally connected, weakly complete, completely regular semi-metric space by the same arguments as used in Example 2.2. That $S$ is neither strongly complete nor a Moore space will be more easily seen following subsequent theorems. The space $S$ is arc-wise connected since every nonvertical line segment in $[0, 1] \times [0, 1]$ has the same relative topology in $S$ as it does in the plane topology, and hence is an arc in $S$.

The second example is due to L. F. McAuley [11].

EXAMPLE 2.8. Let $X$ denote the $x$-axis of the Cartesian plane $E^2$. Define a distance function $d$ for the points of $E^2$ as follows: if each of $p$ and $q$ is a point of $E^2$, then

(1) if neither $p$ nor $q$ belongs to $X$ or if both belong to $X$, $d(p, q) = |p - q|$, and

(2) if exactly one of $p$ any $q$ belongs to $X$, $d(p, q) = |p - q| + a$ where $a$ is the nonobtuse angle (measured in radians) between $X$ and the line $L$ determined by $p$ and $q$. Thus a neighborhood of a point of $X$ is a "bow-tie" neighborhood while a neighborhood of a point not on $X$ is a disc (with some distortion in case the neighborhood intersects $X$). The topological space $S$ consisting of the points of $E^2$ and regions which are $d$-neighborhoods is clearly a connected, locally connected, weakly complete, completely regular semi-metric space that is arc-wise connected.

Each of the distance functions defined in Examples 2.2, 2.7, and 2.8 has the following continuity property: (in each case $d$ denotes the distance function for the space $S$) if $x$ and $y$ are point sequences in $S$ which have respective sequential limit points $p$ and $q$ such that $p \neq q$, then $\lim_{n\to\infty} d(x_n, y_n) = d(p, q)$. It is easily shown that, if $S$ is a regular semi-metric space with a distance function which has the above continuity property, then neighborhoods with respect to that function are open sets and the closure of a compact set in $S$ is compact.

Each of the distance functions defined in Examples 2.7 and 2.8 has in addition the following "convexity" property. If $a$ and $b$ are two points of $S$ such that $d(a, b) < 1$ and $n$ is a natural number, then there is a point sequence $a_0, a_1, a_2, \cdots, a_n$ in $S$ such that $a_0 = a, a_n = b$ and, if $0 \leqq i < j < k \leqq n$, $d(a_i, a_j) + d(a_j, a_k) = d(a_i, a_k)$, and $a_{i+1}$ is the only point of $S$ such that $d(a_i, a_{i+1}) = d(a_{i+1}, a_{i+2}) = 1/2d(a_i, a_{i+2})$. That property, plus the properties that neighborhoods are connected open sets and that the closure of a compact set is compact, is a sufficient condition for the (weakly complete, regular semimetric) spaces of Examples 2.7 and 2.8 to be arc-wise connected.

**3. Conditions for semi-metric, developable, and metric spaces.** Among the open questions about semi-metric spaces are the following. Is there a "purely topological" characterization of semi-metric spaces [12], and what "topological" property can be added to a semi-metric space to get a developable [1, p. 180] (or Moore) space [2, p. 64]? The answers to those questions, or at least some uniform characterization, of semi-metric, developable and metric spaces, should be useful in trying to extend Moore's Arc Theorem. The author found the characterization given below by the Conditions A, B and C useful not only for that purpose, but also for easy construction of nondevelopable semimetric spaces as well as nonmetric Moore spaces (in trying to generalize the Arc Theorem by weakening the completeness part of Moore's Axiom 1). For another application see [5].

A second set of topological conditions, A′ and B′, is obtained by weakening Conditions A and B. Condition A′ is the topological axiom used in the Arc Theorem given in § VI; and Theorem 3.5 establishes that a regular $T_1$-space satisfying Conditions A and B′ is a Moore space—from which it follows (in § IV) that a strongly complete regular semi-metric space is a Moore space (Theorem 2.2 in [11] is a corollary to this theorem). Theorem 3.6 establishes that every semi-metric space has a property analogous to that characterizing property of metric spaces pointed out in [8] and to the similar characterizing property of Moore spaces implicit in Moore's Axiom 1.

Throughout this section $Z$ denotes the set of all natural numbers, and a $T_1$-space (also $T_2$, $T_3$, etc.) is as defined in [9, p. 56]. The following definition is also used.

DEFINITION 3.1. A sequence $x$ of points in the space $S$ *converges* to a point $y$ of $S$ only in case every region which contains $y$ contains $x_i$ for all but finitely many values of $i$; $y$ is then called a *sequential limit point* of $x$.

Suppose that $S$ is a $T_1$-space. Consider the following three conditions on a function $g$ from $Z \times S$ to the collection of all open sets in $S$.

CONDITION A. (1) For each point $x$ of $S$, $\{g_n(x)\}_{n=1}^{\infty}$ is a nonincreasing sequence which forms a local base for the topology at $x$. (2) If $y$ is a point of $S$ and $x$ is a point sequence in $S$ such that, for each natural number $m$, $y \in g_m(x_m)$, then $x$ converges to $y$.

CONDITION B. If $y$ is a point of $S$ and $x$ and $z$ are point sequences in $S$ such that, for each $m$, $[y + x_m] \subset g_m(z_m)$, then $x$ converges to $y$.

CONDITION C. If each of $x$ and $y$ is a point of $S$ and $n$ is a natural number such that $x \in g_n(y)$, then $y \in g_n(x)$. [cf. 3, p. 257 and p. 261].

THEOREM 3.2. *A necessary and sufficient condition that a $T_1$-space $S$ be semi-metric is that there is a function $g$, from $Z \times S$ to the open sets of $S$, such that $g$ satisfies Condition A.*

*Proof*. The condition is sufficient. For suppose that there is a function $g$ which satisfies Condition A. Define the function $m$, from $S \times S$ to the natural numbers, as follows: if $x$ and $y$ are two points of $S$, $m(x, y)$ is the smallest natural number $p$ such that $y \notin g_p(x)$. Define a distance function $d$ for $S$ as follows: if $x \in S$, $d(x, x) = 0$; if $x$ and $y$ are two points of $S$, $d(x, y) = \min [1/m(x, y), 1/m(y, x)] =$ the reciprocal of the smallest natural number $p$ such that $y \notin g_p(x)$ and $x \notin g_p(y)$. Clearly, if each of $x$ and $y$ is a point of $S$, $d(x, y) = d(y, x)$; and $d(x, y) = 0$ only if $x = y$. Also limit points are invariant with respect to $d$ (for a proof see [4, p. 30]).

The condition is necessary. For suppose that $S$ is a semi-metric space. Define the function $g$ as follows: if $x$ is a point of $S$ and $n$ is a natural number, $g_n(x) =$ interior $[U_{1/n}(x)] = \{y$: for some region $R$, $y \in R \subset U_{1/n}(x)\}$. Clearly $g$ satisfies Condition A.

THEOREM 3.3. *A necessary and sufficient condition that a $T_1$-space $S$ be developable is that there be a function $g$, from $Z \times S$ to the open sets of $S$, such that $g$ satisfied Conditions A and B.*

*Proof.* The condition is sufficient. For suppose that there is a function $g$ satisfying Conditions A and B. For each natural number $i$, let $G_i = \{g_j(x) : x \in S, j \geqq i\}$. The coverings $G_1, G_2, G_3, \cdots$ constitute a development (for proof see [4, p. 32]).

Suppose, conversely, that $G_1, G_2, \cdots$ is a development for $S$. Define the function $g$ as follows: for each point $x$ of $S$ let $g_1(x)$ be some member of $G_1$ which contains $x$, and, if $n$ is a natural number greater than 1, let $g_n(x)$ be a member of $G_n$ such that $x \in g_n(x) \subset g_{n-1}(x)$. Clearly $g$ satisfies Conditions A and B.

THEOREM 3.4. *A necessary and sufficient condition that a $T_1$-space $S$ be metric is that there is a function $g$, from $Z \times S$ to the open sets of $S$ such that $g$ satisfies Conditions A, B, and C.*

*Proof.* The condition is necessary, for suppose that $S$ is a metric space. Define the function $g$, from $Z \times S$ to the open sets of $S$, as follows: for each point $x$ and natural number $n$, $g_n(x) =$ interior $[U_{1/n}(x)]$. It is clear that $g$ satisfies Conditions A, B, and C.

Conversely, let $S$ be a $T_1$-space, and let $g$ be a function which satisfies Conditions A, B, and C. For each $n$, let $G_n = \{g_m(x) : x \in S, m \geqq n\}$. By Moore's metrization theorem [11, p. 325], if $S$ is not metrizable, there are two points $p$ and $q$ and a region $R$ such that, for each $n$, $G_n$ contains members $h$ and $k$ such that $p \in h$, $h \cdot k \neq 0$, and $k \cdot [S - (R - q)] \neq 0$ (i.e., $K \cdot (S - R) \neq 0$ since $S$ is $T_1$). Thus there are a point $p$, a region $R$, and point sequences $x, y$, and $z$ such that, for each $n$, $p \in g_n(x_n)$, $y_n \in g_n(x_n) \cdot g_n(z_n)$, and $g_n(z_n) \cdot [S - R] \neq 0$. By Condition B, $[y_n + p] \subset g_n(x_n)$ $(n = 1, 2, 3, \cdots)$ implies that the sequence $y$ converges to $p$. Therefore, there is an increasing natural number sequence $m$ such that, for each natural number $n$, $y_{m(n)} \in g_n(p)$, so, that, by Condition C, $p \in g_n(y_{m(n)})$. Also $y_{m(n)} \in g_{m(n)}(Z_{m(n)})$ implies that $z_{m(n)} \in g_{m(n)}(y_{m(n)})$—hence that $z_{m(n)} \in g_n(y_{m(n)})$.

Thus, for each $n$, $[p + z_{m(n)}] \subset g_n(y_{m(n)})$, so that, by Condition B, $\{z_{m(n)}\}_{n=1}^{\infty}$ converges to $p$. Therefore there is a subsequence $r$ of $m$ such that, for each natural number $n$, $z_{r(n)} \in g_n(p)$, hence $p \in g_n(z_{r(n)})$. But by supposition there is a point sequence $u$ such that, for each $n$, $u_n \in (S - R)$ and $u_n \in g_n(z_{r(n)})$, so that $u$ converges to $p$; which leads to the contradiction that $p \in \overline{(S - R)}$ while $p$ is contained in the region $R$. Thus $S$ must be metrizable.

Consider Conditions A′ and B′ which are at least formally weaker than A and B respectively.

CONDITION A′. (1) For each point $x$ of $S$, $\{g_n(x)\}_{n=1}^{\infty}$ is a nonincreasing sequence which forms a local base for the topology at $x$. (2) If $y$ is a point of $S$ and $x$ is a point sequence such that, for each natural number $n$, $y \in g_n(x_n)$ and there is a natural number $k$ such that $\overline{g_{n+k}(x_{n+k})} \subset g_n(x_n)$, then the sequence $x$ converges to $y$.

CONDITION B′. If $y$ is a point of $S$, $R$ is an open set containing $y$, and $x$ is a point sequence such that, for each $n$, $y \in g_n(x_n)$ and there is a $k$ such that $\overline{g_{n+k}(x_{n+k})} \subset g_n(x_n)$, then there is a natural number $m$ such that $g_m(x_m) \subset R$.

It will be convenient now to have Condition A, A′, B, B′, and C translated into corresponding conditions on a basis for the space.

DEFINITION 3.5. Suppose that $S$ is a $T_1$-space and that $G$ is a basis for $S$. The basis $G$ satisfies Condition A (A′, B, B′, or C) means that there is a function $g$, from $Z \times S$ to the open sets of $S$, such that $G = \{g_n(x) \colon x \in S, n = 1, 2, 3, \cdots\}$ and $g$ satisfies Condition A (A′, B, B′, or C).

Theorem 3.6 establishes that the Arc Theorem cannot be generalized by simply replacing "Moore space" by "a regular semi-metric space satisfying Condition B′" since such a space is itself a Moore space; moreover it will readily follow from this theorem (cf. Corollary 4.3) that every strongly complete regular semi-metric space is a Moore space, thus eliminating another means of extending the Arc Theorem as well as improving upon Theorem 2.2 of [11].

THEOREM 3.6. *Suppose that $G$ is a basis for a regular $T_1$-space $S$. If $G$ satisfies Condition A and B′, then $S$ has a basis $H$ which satisfies Conditions A and B, hence $S$ is a Moore space.*

*Proof.* Suppose that the regular $T_1$-space $S$ has a basis $G =$

$\{g_m(x): x \in S, m = 1, 2, \cdots\}$ which satisfies Conditions A and B′. Let $\alpha = \{p_1, p_2, \cdots\}$ be a well-ordering of the points of $S$. Define the functions $h$ (from $Z \times S$ to $G$), $r$ (from $\alpha \times Z$ to $Z$), and $n$ (from $\alpha x$ [a subset of $Z$] to $\alpha$) as follows. For each $p_z \in S$, let $h_1(p_z) = g_{r_z[1]}(p_z) = g_1(p_z)$. For each $p_z \in S$ and each natural number $i$ greater than 1:

(Case 1) If there is *not* a point $q$ of $[S - p_z]$ and a natural number $j < i$ such that $p_z \in h_j(q)$ and $h_j(q) \cdot [S - g_{r_z[i-1]+1}(p_z) \neq 0$, then let $h_i(p_z) = g_{r_z[i]}(p_z) = g_{r_z[i-1]+1}(p_z)$; or

(Case 2) otherwise, for each (such) $j < i$, let $p_{n_z[j]}$ be the first member $q$ of $\alpha(q \neq p_z)$ such that $p_z \in h_j(q)$ and $h_j(q) \cdot [S - g_{r_z[i-1]+1}(p_z)] \neq 0$; let $r_z[i]$ be the smallest natural number $m > r_z[i-1]$ such that $\overline{g_m(p_z)} \subset \prod[h_j(p_{n_z[j]}): j < i$, and $j$ is not covered by Case 1]; and let $h_i(p_z) = g_{r_z[i]}(p_z)$.

The basis $H = \{h_i(x): x \in S, i = 1, 2, \cdots\}$ then satisfies Conditions A and B′, since $H$ is a subcollection of $G$ and since, if $x \in g \in G$, then there is an $h \in H$ such that $x \in h \subset g$.

The basis $H$ satisfies Condition B. For if not, then there is a point $x$ of a region $R$ such that, for each $m$, there is a point $q$ such that $x \in h_m(q)$ and $h_m(q) \cdot [S - R] \neq 0$. Let $y$ be the point sequence such that, for each $m$, $y_m$ is the first point $q$ in $\alpha$ such that $x \in h_m(q)$ and $h_m(q) \cdot [S - R] \neq 0$. (It will now be shown that, for each natural number $i$, there is an $m > i$ such that $\overline{h_m(y_m)} \subset h_i(y_i)$.)

If $i$ is a natural number there is a natural number $N_1 > i$ such that, if $m > N_1$, then $y_m \in [h_i(y_i) - y_i]$ (since $y$ converges to $x$ and $x \in h_i(y_i)$) and there is an $N_2$ such that, if $m > N_2$, then $g_{r(y_m)[m-1]+1}(y_m)$ does not contain $h_i(y_i)$ (otherwise, by Condition A, each point of $h_i(y_i)$ would be a sequential limit point of $y$, and $h_i(y_i)$ contains at least two points, namely, $x$ and a point not in $R$); thus there is a natural number $m$ such that

$$y_m \in [h_i(y_i) - y_i] \quad \text{and} \quad h_i(y_i) \cdot [S - g_{r(y_m)[m-1]+1}(y_m)] \neq 0 \, .$$

Moreover, there is no point $q$ such that $q$ precedes $y_i$ in $\alpha$ and $h_m(y_m) \subset h_i(q)$ (since $y_i$ is the first point $a$ in $\alpha$ such that $x \in h_i(a)$ and $h_i(a) \cdot [S - R] \neq 0$, and $h_m(y_m) \subset h_i(q)$ would imply that $h_i(q) \cdot [S - R] \neq 0$); hence there is no point $q$ such that $q$ precedes $y_i$ in $\alpha$ and $y_m \in h_i(q)$ and $h_i(q) \cdot [S - g_{r(y_m)[m-1]+1}(y_m)] \neq 0$ (otherwise $h_i(q)$ would contain $h_m(y_m)$ by definition of $h_m(y_m)$). Therefore $\overline{h_m(y_m)} \subset h_i(y_i)$ since $y_i \neq y_m$.

Hence by Condition B′, there is a natural number $N$ such that if $m > N$, then $h_m(y_m) \subset R$ contrary to the supposition that for each $i$, $h_i(y_i) \cdot [S - R] \neq 0$.

Using the same argument down to the last sentence, which is the first place that Condition B′ is used, Theorem 3.7 below also follows.

THEOREM 3.7. *Suppose that $S$ is a regular semi-metric space. Then there is a basis $H = \{h_n(x): x \in S, n = 1, 2, \cdots\}$ with the property that for each $p \in S$ and for each closed and compact subset $M$ of $S - p$ there is a natural number $N$ such that if $m > N$ and $p \in h_m(x)$ then $h_m(x) \cdot M = 0$.*

**4. Completeness axioms.** Another way to generalize Moore's Arc Theorem is to weaken the completeness used. Three successively weaker completeness axioms (1, 1′, and 1″) are given below. In a Moore space: Completeness Axiom 1 is equivalent to Moore's completeness, which is known to be weaker than strong completeness [11, Example 3.3]; and Completeness Axioms 1′ and 1″ are both equivalent to the completeness in Mrs. Rudin's Axiom 1″ [16, p. 320], and hence weaker than Moore's Completeness [16, p. 324]. In semi-metric spaces Completeness Axiom 1′ is stronger than 1″ (all examples in § 11 satisfy Axiom 1″ but not 1′; also see Corollary 4.3).

A Cauchy (or weakly) complete semi-metric space satisfies Completeness Axiom 1″. In a metric space all of the completenesses mentioned are equivalent [15].

The theorems listed below (for proofs see [4, pp. 35–43]) give the relationships between the three completeness axioms and the topological properties defined in § III. Aside from finding that completeness Axioms 1′ and 1″, which are used in separate arc theorems, are more general than Moore's completeness, the main results obtained in this section are (1) that a strongly complete regular semi-metric space is a Moore space (2) that Cauchy (or weak) completeness is weaker than Moore's completeness and (3) a generalization of Theorem 120 in [14] (also Theorem 6 of [16]).

Suppose that $S$ is a $T_1$-space and $G = \{g_n(x): x \in S, n = 1, 2, 3 \cdots\}$ is a basis for $S$ which satisfies Condition A′. Consider the following completeness axioms for $G$.

*Completeness Axiom* 1. If $M$ is a nonincreasing sequence of closed sets such that, for each $n$, there is a point $x_n$ of $S$ such that $M_n \subset g_n(x_n)$, then $\prod_{n=1}^{\infty} M_n \neq 0$.

*Completeness Axiom* 1′. If $M$ is a nonincreasing sequence of closed sets and $x$ a point sequence in $S$ such that, for each $n$, $M_n \subset g_n(x_n)$ and there is a natural number $k$ such that $\overline{g_{n+k}(x_{n+k})} \subset g_n(x_n)$, than $\prod_{n=1}^{\infty} M_n \neq 0$.

*Completeness Axiom* 1″. If $x$ is a point sequence such that, for each $n$, there is a $k$ such that $\overline{g_{n+k}(x_{n+k})} \subset g_n(x_n)$, then $\prod_{n=1}^{\infty} g_n(x_n) \neq 0$.

Theorem 4.1 shows in particular that Axioms 1′ and 1″ are equivalent in a developable $T_1$-space.

THEOREM 4.1. *Suppose that $S$ is a $T_1$-space with a basis $G$ that satisfies Conditions $A'$ and $B'$. A necessary and sufficient condition that $G$ satisfy Completeness Axiom* 1′ *is that $G$ satisfy Completeness Axiom* 1″.

An immediate corollary to Theorems 4.2 and 3.6 is that a strongly complete regular semi-metric space is a Moore space.

THEOREM 4.2. *Suppose that the $T_2$-space $S$ has a basis $G$ which satisfies Condition $A'$. A necessary and sufficient condition that $G$ satisfy Condition $B'$ and Completeness Axiom* 1″ *is that $G$ satisfy Completeness Axiom* 1′.

COROLLARY 4.3. *If the regular $T_1$-space $S$ has a basis $G$ which satisfies Condition $A$ and Completeness Axiom* 1′, *then $S$ has a basis which satisfies Conditions $A$ and $B$. Hence every strongly complete regular semi-metric space is a complete Moore space.*

The following theorem shows that the space having a basis with some of the completeness properties and another basis with some of the topological properties has a basis with the combined properties.

THEOREM 4.4. *If the $T_1$-space $S$ has a basis $G$ which satisfies Condition $A'$ and one of the Completeness Axioms* 1, 1′, *and* 1″, *and if $S$ has a basis $H$ which satisfies some combination of Conditions $A'$, $A$, $B'$, and $B$, then $S$ has a basis $K$ which satisfies the Completeness Axiom that $G$ satisfies and the combination of Conditions $A'$, $A$, $B'$, and $B$ that $H$ satisfies.*

The next theorem is a generalization of a portion of a theorem due to Moore [14, p. 83, Theorem 120]. (The other part of the theorem also holds in any of the same spaces). Essentially the same proof may be used (see [4, p. 39]).

THEOREM 4.5. *Suppose that the $T_1$-space $S$ has a basis $G$ which satisfies Condition $A'$* (*or $A'$ and $A$, $B'$, or $B$*) *and one of the Completeness Axioms* 1, 1′, *and* 1″. *If $M$ is an inner limiting subset* (i.e. *a $G_\delta$ set*) *of $S$, then there is a basis $H$ for $M$ such that $H$ satisfies the same combination of Conditions $A'$, $A$, $B'$, and $B$ and the Completeness Axiom that $G$ satisfies.*

The completeness defined in Definition 4.6 is clearly weaker in general than Axiom 1′; however, Theorem 4.7 shows that replacing the latter by the former in an arc theorem would not be a real generalization.

DEFINITION 4.6. A basis $G$ for a $T_1$-space $S$ which satisfies Condition A′ is said to satisfy Completeness Axiom 1′ peripherally provided that, if $R \in G$, $M$ is a nonincreasing sequence of closed subsets of $B(R)$, the boundary of $R$, and $x$ is a point sequence in $S$ such that, for each $n$, $M_n \subset g_n(x_n)$ and there is a $k$ such that $\overline{g_{n+k}(x_{n+k})} \subset g_n(x_n)$, then $\prod_{n=1}^{\infty} M_n \neq 0$.

THEOREM 4.7. *If $G$ is a basis for a $T_2$-space $S$, if each member of $G$ is connected, and if $G$ satisfies Condition $A'$ and satisfies Completeness Axiom* 1′ *peripherally, then $G$ satisfies Condition $B'$.*

For a proof see [4, pp. 42–43].

**5. Metrization theorems.** The metrization theorems below serve not only to eliminate certain hypotheses from consideration for generalizing the Arc Theorem, but also to show that the spaces described in § II, which are not even developable (and in some of which the Arc Theorem does not hold) are nonetheless very close to being metrizable.

Theorem 5.1 generalizes a well-known theorem which is included in Theorem 10 [21]. It can also be shown that every semi-metric space contains a dense metric subspace (but not necessarily one which is an inner limiting subset).

The proof of the following lemma is exactly analogous to the proof of Theorem 15 in [14, p. 11].

LEMMA. *Suppose that $S$ is a regular $T_1$-space with a basis $G$ that satisfies Condition $A'$ and the Completeness Axiom* 1″. *No closed subset $M$ of $S$ is the sum of countably many closed sets each of which is contained in the boundary of its complement (in $M$).*

THEOREM 5.1. *If the regular $T_1$-space $S$ has a basis $G$ that satisfies Condition $A'$ and the Completeness Axiom* 1″, *then $S$ contains a dense inner limiting subset $K$ which (with the relative topology) is metrizable and complete.*

*Proof.* Let $H_1$ be a maximal collection of mutually exclusive regions each of which belongs to $\{g_i(x)\colon x \in S,\, i \geqq i\}$. For each $n > 1$, let $H_n$ be a maximal collection of mutually exclusive regions each of which belongs to $\{g_i(x)\colon x \in S,\, i \geqq n$, and there is a region $h$ in $H_{n-1}$ such that $g_i(x) \subset h\}$. Let $K = \prod_{n=1}^{\infty} [H_n^*]$. Since by the above lemma, if $R$ is a region in $G$, $\bar{R}$ is not the sum of countably many closed sets each contained in the boundary of its complement (in $\bar{R}$), so that $\bar{R} \neq \sum_{n=1}^{\infty} [(S - H_n^*) \cdot \bar{R}]$ and $\bar{R} \cdot K \neq 0$, it follows that $K$ is dense in $S$. Clearly $K$ is an inner limiting set and $K$ (with the relative topology) is metrizable since $(\sum_{n=1}^{\infty} H_n) \cdot K$ forms a basis for the relative topology of $K$, and since, for each $n$, the elements of $H_n$ are pairwise disjoint and each member of $H_{n+1}$ is a subset of a member of $H_n$. That $K$ is complete follows from Theorem 4.5.

Theorem 5.2 shows how close the spaces of Examples 2.2, 2.5, 2.7, and 2.8 are to being metrizable. Note that each of those spaces satisfies the hypotheses of the theorem except for being locally peripherally locally compact instead of locally peripherally compact.

THEOREM 5.2. *Suppose that the semi-metric space $S$ has a distance function $d$ such that $d$-neighborhoods are connected sets and such that, if $p$ is a sequential limit point of the point sequence $x$ and $q$ is a sequential limit point of the point sequence $y$ and $\lim_{i \to \infty} d(x_i, y_i) = 0$, then $p = q$. If $S$ is locally peripherally compact, then $S$ is metrizable.*

*Proof.* The theorem will be proved by showing that $S$ satisfies the hypothesis of W. A. Wilson's theorem [20, pp. 361 and 366; also 2, p. 63] that a semi-metric space is metrizable provided that, for every pair $x, y$ of sequences, if $p$ is a sequential limit point of $x$ and $\lim_{i \to \infty} d(x_i, y_i) = 0$, then $p$ is a sequential limit point of $y$. For suppose that $p$ is a point of $S$ and $x$ and $y$ are point sequences such that $p$ is a sequential limit point of $x$ and $\lim_{i \to \infty} d(x_i, y_i) = 0$, but $p$ is not a sequential limit point of $y$. Then there is a region $R$ with compact boundary, $B(R)$, and a sequence $y'_1, y'_3, y'_3, \cdots$ of $y$ such that $R$ contains $p$ but contains none of the points $y'_1, y'_2, y'_3, \cdots$; thus (noting that it may be assumed, without loss of generality, that, for each $i$, $x_i \in R$ and $d(x_i, y'_1) < (1/i)$) each of the connected neighborhoods $U_1(x_1)$, $U_{1/2}(x_2)$, $U_{1/3}(x_3), \cdots$ contains a point of $(S - R)$ and hence contains a point of $B(R)$. Let $z$ be a point sequence such that, for each $n$, $z_n \in U_{1/n}(x_n) \cdot B(R)$. Since $B(R)$ is compact, there is a point $q$ of $B(R)$ such that $q$ is a sequential limit point of asubsequence of $z$, which leads to a contradiction of the hypothesis of the theorem.

By Theorem 5.3 paracompactness (defined along with pointwise paracompactness in [1. p. 177]) is too restrictive in the presence of a completeness axiom slightly stronger than that possessed by the spaces in § II.

THEOREM 5.3. *Suppose that the regular $T_1$-space $S$ has a basis $G$ which satisfies Condition A and Completeness Axiom 1′ peripherally and whose elements are connected sets. If $S$ is pointwise paracompact,* *then $S$ is developable; hence, if $S$ is paracompact, then $S$ is metrizable.*

For a proof of this theorem see [4, pp. 47–48].

**6. An arc theorem.** Theorem 6.2 is more general than R. L. Moore's Arc Theorem [14, p. 86], to the extent that the completeness axiom used is known to be less restrictive than that in the hypothesis of Moore's theorem, and the other properties used are at least formally more general than those used by Moore. To adapt the proof of Moore's theorem to Theorem 6.2, however, requires only fairly minor modifications. Theorem 6.4 establishes that certain semi-metric spaces which are not arc-wise connected (including Example 2.2) are nontheless connectable by closed connected sets which closely resemble arcs.

In the following definition, which will be used in the proof of Theorem 6.2, "simple chain" (or "chain") is as defined in [14, p. 56].

DEFINITION 6.1. Suppose that $G = \{g_i(x): x \in S, i = 1, 2, 3, \cdots\}$ is a basis for the $T_1$-space $S$, that $A$ and $B$ are two points of $S$, and that, for each $n$, $C_n = \{Q[n, 1], Q[n, 2], \cdots, Q[n, m_n]\}$ is a simple chain from $A$ to $B$. The sequence $\{C_1, C_2, C_3, \cdots\}$ *has property $P$ with respect to $G$* if

(1) for each $i$ $\{\overline{Q[1+i, j]}\}_{j=1}^{m_{i+1}}$ is a refinement of $\{Q[i, j]\}_{j=1}^{m_i}$;

(2) for each $i$ and each $j$ such that $1 < j \leqq m_i$ there is a $K$ such that $(\sum_{P<K} Q[i+1, P]) \cdot Q[i, j] = 0$, and such that $(\sum_{P>K} \overline{Q[i+1, P)} \subset (\sum_{q=j}^{m_i} Q[i, q])$ and $\overline{Q[i+1, k+1]} \subset Q[i, j-1] \cdot Q[i, j]$; and

(3) there is a collection of natural numbers $\{k[i, j]: j \leqq m, i = 1, 2, 3, \cdots\}$ and a collection of points $\left\{x_{\substack{ij \\ 1 \leqq j \leqq m_i}}\right\}_{i=1}^{\infty}$ such that, for each $i$ and each $j \leqq m_i$,

(a) $k[i, j] \geqq i$

(b) $Q[i, j] = g_{k(i,j)}(x_{ij})$ and

(c) if $x_{i+1,j} \in Q[i, t]$, then $\overline{Q[i+1, j]} \subset Q[i, t]$.

THEOREM 6.2. *If the connected regular $T_1$-space $S$ has a basis $G$ each of whose elements is connected and which satisfies Condition $A'$ and Completeness Axiom $1'$, then $S$ is arc-wise connected.*

*Proof.* Let $a$ and $b$ be two points of $S$. Moore's proof (in particular Theorem 77 [14, p. 56]) may be applied with slight alterations to the sequence $G$ of open coverings of $S$, such that, for each $n$, $G_n = \{g_k(x): x \in S$ and $k \geqq n\}$ to obtain a sequence, $\{C_i\}_{i=1}^{\infty} = \{Q[i, 1], Q[i, 2], \cdots, Q[i, m_i]\}_{i=1}^{\infty}$, of chains from $a$ to $b$ which has property $P$ with respect to $G$.

Let $M = \prod_{i=1}^{\infty} C_i^* = \prod_{i=1}^{\infty} [\sum_{j=1}^{m_i} \overline{Q[i, j]}]$.

Clearly $M$ is closed.

Also $M$ is compact. For suppose that an infinite subset $K$ of $M$ has no limit point. Then there exists a sequence $\{H_n\}_{n=1}^{\infty}$ of collections of pointsets such that, for each $n$:

(i) each point set in $H_n$ is a link of the chain $C_n$,

(ii) each set in $H_n$ contains an infinite subset of $K$,

(iii) $H_{n+1}$ is a refinement of $H_n$, and

(iv) (by the above property 3c of the sequence $C$) if $x \in H_{n+1}$, $y \in H_n$, and $x \subset y$, then $\bar{x} \subset y$. By theorem 78 [14, p. 56] there is a sequence $h$ such that, for each $n$, $h_n \in H_n$ and $h_{n+1} \subset h_n$, so that, by the above property (iv) of the sequence $H$, $\overline{h_{n+1}} \subset h_n$. For each $i$, $h_i$ is a link of the chain $C_i$, so that, by property 3b (above) of the sequence $C$, there is an increasing natural number sequence $d$ and a point sequence $u$ such that, for each $i$, $h_i = g_{d(i)}(u_i)$, hence, for each $i$, the closed set $(h_i \cdot K)$ is contained in $g_{d(i)}(u_i)$, $(h_i \cdot K) \supset (h_{i+1} \cdot K)$, and $\overline{g_{d(i+1)}(u_{i+1})} \subset g_{d(i)}(u_i)$. Therefore, by Completeness Axiom 1′, there is a point $p$ such that $p \in \prod_{i=1}^{\infty} [K \cdot h_i]$; hence, $p \in K$, and for each $i$, $p \in g_{d(i)}(u_i)$, so that (by Condition $A'$) $p$ is a sequential limit point of a subsequence $\{u_{t(i)}\}_{i=1}^{\infty}$ of the sequence $u$. Consider then the nonincreasing sequence of closed sets, $\{(K - p) \cdot h_{t(i)}\}_{i=1}^{\infty}$, which likewise has the property that, for each $i$, $[(K - p) \cdot h_{t(i)}] \subset g_{d[t(i)]}(u_{t(i)})$. Again, there is a point $q$ such that $q \in (K - p)$ and such that $q$ is a sequential limit point of $\{u_{t(i)}\}_{i=1}^{\infty}$ and $p \neq q$. Thus the assumption that $M$ is not compact leads to a contradiction.

That $M$ is connected follows as in Moore's proof with slight alterations (or see [4, pp. 53–54]); and that each point of $M - (a + b)$ is a separating point of $M$ and that $a$ and $b$ are nonseparating points also follows as in Moore's proof.

COROLLARY. *The connected regular $T_1$-space $S$ is arc-wise connected if $S$ satisfies any one of the following conditions:*

(a) $S$ has a basis $G$ which satisfies Conditions A′ and B′ and Completeness Axiom 1″ and each of whose elements is connected;

(b) $S$ has a basis $G$ which satisfies Condition A′ and satisfies Completeness Axiom 1′ peripherally and each of whose elements is connected;

(c) $S$ is locally connected and satisfies Mary Ellen Estill Rudin's Axiom 1″ [16].

(d) $S$ is a locally connected strongly complete semi-metric space.

"Strong chainability," defined below, is a rather restricted special case of chainability but is useful for showing that certain spaces which are not arc-wise connected are connectable by sets which closely resemble arcs. Roughly speaking a space is strongly chainable with respect to a basis $G$ provided that, for every pair $a, b$ of points of $S$, there is a sequence $C$ of chains from $a$ to $b$ such that

(1) $C$ has Property $P$ with respect to $b$

(2) the "centers" of $C(n)$ are also "centers" of $C(n+1)$ and

(3) the intersection of two adjacent links of $C(n)$ with centers $p$ and $q$ contains a point $y$ such that the subchain from $p$ to $q$ of $C(n+1)$ is contained in $g_n(y)$ and $y$ is a "center" of $C(n+1)$. Theorem 6.2 then establishes that the space of Example 2.2 is connectable by sets having all the properties of an arc but compactness.

DEFINITION 6.3. Suppose that the regular $T_1$-space $S$ has a basis, $G = \{g_n(x): x \in S, n = 1, 2, \cdots\}$, which satisfies Condition A. $S$ is *strongly chainable with respect to* $G$ provided that, if $a$ and $b$ are two points of $S$, there is a sequence, $C = \{Q[i, 1], Q[i, 2], \cdots, Q[i, m_i]\}_{i=1}^{\infty}$ of chains from $a$ to $b$ such that:

(1) $C$ has property $P$ with respect to $G$.

(2) There is a collection, $\{t(i, j): j \leqq m_i, i = 1, 2, \cdots\}$ of natural numbers such that, for each $i$ and each $j \leqq m_i$, $x_{ij} = x_{i+1, t(i,j)}$: and

(3) there is a collection $\{y_{ij}: j < m_i, i = 1, 2, \cdots\}$ of points such that, for each $i$ and each $j < m_i$, (a) $y_{ij} \in Q[i, j] \cdot Q[i, j+1]$ and $\sum_{r=t(i,j)}^{t(i,j+1)} Q[i+1, r] \subset g_i(y_{ij})$ and (b) there is a natural number $r$ such that $y_{ij} = x_{i+1,r}$. For each $i$ and each $j \leqq m_i$, the point $x_{ij}$ (from Definition 6.1 part (3)) will be referred to as the *center* of the link $Q[i, j]$ of $C_i$.

Note that the space of Example 2.2 (§ II), which is not arc-wise connected, is strongly chainable with respect to the basis consisting of all $1/n$ neighborhoods (for $n = 1, 2, \cdots$).

THEOREM 6.4. *Suppose that $S$ is a connected regular $T_1$-space which has a basis, $G = \{g_n(x): x \in S, n = 1, 2, \cdots\}$, that satisfies Condition A and Completeness Axiom 1″ and each of whose elements is connected. If $S$ is strongly chainable with respect to G, then, for each pair of points a and b, there is a continuum M containing a and b such that a and b are the only non-separating points of M.*

*Proof.* Let $a$ and $b$ be two points of $S$; let $C = \{Q[i, 1], Q[i, 2], \cdots, Q[i, m_i]\}_{i=1}^{\infty}$ be a sequence of chains from $a$ to $b$ satisfying Definition 6.3 and, for each $i$ and each $j \leqq m_i$, let $x_{ij}$ be the center of $Q[i, j]$. Denote by $L$ the set of all centers $\{x_{ij}: j \leqq m_i, i = 1, 2, \cdots\}$; and let $M = \prod_{i=1}^{\infty} [C_i[^*$. Clearly $(a + b) \subset M$, $M$ is closed, and $a$ and $b$ are the only nonseparating points of $M$.

Furthermore, $M$ is connected. For suppose that $M = H + K$, $\bar{H} \cdot K = H \cdot \bar{K} = 0$ (where $a \in H$). Because $M$ is closed, each of $H$ and $K$ is a closed set. Also, because $\bar{L} = M$, $L \cdot H \neq 0$ and $L \cdot K \neq 0$. It will now be shown that there is a natural number $n$, a natural number sequence $u$, and point sequences $p$ and $q$ such that, for each $i \geqq n$, $u(i) \leqq m_i$ and

(1) $\overline{Q[i+1, u(i+1)]} \subset Q[i, u(i)]$ and

(2) $p_i \in H$, $q_i \in K$, and $\overline{Q[i+1, u(i+1)]} \subset g_i(p_i) \cdot g_i(q_i)$. Since $a \in H$, since $L \cdot H \neq 0$, and since $L \cdot K \neq 0$, there is a natural number $n$ and a natural number $j < m_n$ such that $x_{nj} \in H$ and $x_{n,j+1} \in K$. If $y_{nj} \in K$, let $u(n) = j$; if $y_{nj} \in H$, let $u(n) = j + 1$. In either case there is a natural number $r$ such that $x_{n+1,r} \in H$, $x_{n+1,r+1} \in K$, and $\overline{[Q[n+1, r]} + \overline{Q[n+1, r+1]]} \subset Q[n, u(n)] \cdot g_n(y_{nj})$; again, if $y_{n+1,r} \in K$, let $u(n+1) = r$, and if $y_{n+1,r} \in H$, let $u(n+1) = r + 1$; and the process may be continued to define sequences $u$, $p$, and $q$ which have the stated properties. Then, by Completeness Axiom 1$''$ there is a point $z$ such that $z \in \prod_{n=1}^{\infty} \overline{Q[n+1, u(n+1)]}$ and $z$ is a sequential limit point of the sequence $p$ in $H$ and of the sequence $q$ in $K$; hence $z \in H \cdot K$ contrary to the assumption that $H \cdot K = 0$.

**7. Summary and questions.** Theorem 2.3 establishes that Moore's Arc Theorem cannot be generalized directly to Cauchy complete regular semi-metric spaces, while Theorem 6.2 shows that it can be generalized to a class of semi-metric spaces somewhat more general than complete Moore spaces—in particular, the completeness axiom used is known to be weaker than that of Moore's Axiom 1. The examples in § II and the theorems establishing certain sufficient conditions for a semi-metric space to be developable or even metrizable given in §§ III, IV and V show rather clearly the limited nature of the progress that can be made towards extending the arc Theorem to semi-metric spaces. For example, Theorems 3.6 and 4.2 establish that every strongly complete regular semi-metric space is a complete Moore space.

The following questions then are suggested:

(1) Can Moore's Arc Theorem be generalized in another direction, such as to complete uniform spaces?

(2) Since the class of strongly complete regular semi-metric spaces properly includes the class of all complete Moore spaces and is properly included in the class of all complete metric spaces, what is a sufficient condition for a complete Moore space—or a weakly complete semi-metric space—to be strongly complete, and what is a sufficient condition for a strongly complete regular semi-metric space to be metrizable?

(3) Is there any reasonable necessary and sufficient condition for a connected and locally connected complete regular semi-metric space to be arc-wise connected?

1. R. H. Bing, *Metrization of topological spaces*, Canadian J. of Math., **3** (1951), 175-186.
2. Morton Brown, *Semi-metric spaces* Summer Institute on Set Theoretic Topology, Madison, Wisconsin, Amer. Math. Soc., (1955), 62-64.
3. L. W. Cohen, *Uniformity in topological spaces*, Lectures in Topology ed. by Wilder and Ayres, University of Michigan Press, (1941), 255-265.
4. R. W. Heath, *Arc-wise connectedness in semi-metric spaces*, doctoral dissertation, University of North Carolina, 1959.
5. ———, *A regular semi-metric space for which there is no semi-metric under which all spheres are open*, Proc. Amer. Math. Soc., **12** (1961), 810-811.
6. F. B. Jones, *Connected and disconnected plane sets and the functional equations* $f(x) + f(y) = f(x + y)$, Bull. Amer. Math. Soc., **48** (1942), 115-120.
7. ———, *Introductory remarks on semi-metric spaces*, Summer Institute on Set Theoretic Topology, Madison, Wisconsin, Amer. Math. Soc. (1955), 58.
8. ———, *R. L. Moore's Axiom 1 and metrization*, Proc. Amer. Math. Soc., **9** (1958), 487.
9. J. L. Kelley, *General Topology*, Princeton: D. Van Nostrand Company, 1955.
10. C. Kuratowski, *Topologie II*, Warsaw, (1950), 184.
11. L. F. McAuley, *A Relation between perfect separability, Completeness, and normality in semi-metric spaces*, Pacific J. Math., **6** (1956), 315-326.
12. ———, *On semi-metric spaces*, Summer Institute on Set Theoretic Topology, Madison, Wisconsin, Amer. Math. Soc., (1955), 58-62.
13. R. L. Moore, *Abstract sets and foundations of analysis situs*, Bull, Amer. Math. Soc., **33** (1927), 141.
14. ———, *Foundations of point set theory*, Amer. Math. Soc. Coll. Publ., **13**, New York: Amer. Math. Soc., 1932.
15. J. H. Roberts, *A property related to completeness*, Bull. Amer. Math. Soc., **38** (1932), 835-838.
16. Mary Ellen Estill Rudin, *Concerning abstract spaces*, Duke Math. J., **17** (1950), 317-327.
17. A. H. Stone, *Paracompactness and product spaces*, Bull. Amer. Math., Soc., **54** (1948), 977-982.
18. C. W. Vickery, *Moore spaces and metric spaces*, Bull. Amer. Math. Soc., **46** (1940), 560-564.
19. G. T. Whyburn, *Analytic topology*, Amer. Math. Soc. Coll. Publ., **28**, New York, 1942.
20. W. A. Wilson, *On semi-metric spaces*, Amer. J. Math., **53** (1931), 361-373.
21. J. N. Younglove, *Concerning dense metric subspaces of certain nonmetric spaces*, Fundamenta Mathematicae, **48** (1959), 15-25.

THE UNIVERSITY OF NORTH CAROLINA AND
THE UNIVERSITY OF GEORGIA

# ON UNIMODULAR MATRICES

I. HELLER AND A. J. HOFFMAN

**1. Introduction and summary.** For the purpose of this note a matrix is called unimodular if every minor determinant equals 0, 1 or $-1$.

I. Heller and C. B. Tompkins [1] have considered a set

$$S = \{u_i, v_j, u_i + v_j, u_i - u_{i*}, v_j - v_{j*}\}$$

where the $u_1, u_2, \cdots, u_m, v_1, v_2, \cdots, v_n$ are linearly independent vectors in $m + n = k$-dimensional space $E$, and have shown that in the coordinate representation of $S$ with respect to an arbitrary basis in $E$ every nonvanishing determinant of $k$ vectors of $S$ has the same absolute value, and that, with respect to a basis in $S$, the vectors of $S$ or of any subset of $S$ are the columns of a unimodular matrix. For the purpose of this note the class of unimodular matrices obtained in this fashion shall be denoted as the class $T$.

A. J. Hoffman and J. B. Kruskal [4] have considered incidence matrices $A$ of vertices versus directed paths of an oriented graph $G$, and proved that:

(i) if $G$ is alternating, then $A$ is unimodular;

(ii) if the matrix $A$ of *all* directed paths of $G$ is unimodular, then $G$ is alternating. The terms are defined as follows. A graph $G$ is oriented if it has no circular edges, at most one edge between any given two vertices, and each edge is oriented. A path is a sequence of distinct vertices $v_1, v_2, \cdots, v_k$ of $G$ such that, for each $i$ from 1 to $k-1$, $G$ contains an edge connecting $v_i$ with $v_{i+1}$; if the orientation of these edges is from $v_i$ to $v_{i+1}$, the path is directed; if the orientation alternates throughout the sequence, the path is alternating. A loop is a sequence of vertices $v_1, v_2, \cdots, v_k$, which is a path except that $v_k = v_1$. A loop is alternating if successive edges are oppositely oriented and the first and last edges are oppositely oriented. The graph is alternating if every loop is alternating. The incidence matrix $A = (a_{ij})$ of the vertices $v_i$ of $G$ versus a set of directed paths $p_1, p_2, \cdots, p_k$ of $G$ is defined by

$$a_{ij} = \begin{cases} 1 & \text{if } v_i \text{ is in } p_j \\ 0 & \text{otherwise .} \end{cases}$$

The class of unimodular matrices thus associated with alternating graphs shall be denoted by $K$.

I. Heller [2] and [3] has considered unimodular matrices obtained

Received April 12, 1962. Research sponsored (for both authors) by the RAND Corporation, in conjunction with the Symposium in Combinatorial Analysis, and (for the first author) by the National Science Foundation.

by representing the edges (interpreted as vectors) of an $n$-simplex in terms of a basis chosen among the edges (in graph theoretical terms: the edges and vertices of the simplex form a complete graph $G$; a basis is a maximal tree in $G$, that is, a tree containing all vertices of $G$), and has shown that:

(i) the matrix representing all edges of the simplex is unimodular and maximal (i.e., will not remain unimodular when a new column is adjoined);

(ii) the columns of every unimodular matrix of $n$ rows and $n(n+1)$ columns represent the edges of an $n$-simplex.

The class of (unimodular) matrices whose columns are among the edges of a simplex shall be denoted by $H$. $H$ can also be defined as a class of incidence matrices: A matrix $A$ belongs to $H$ if there is some oriented graph $F$ without loops such that $A$ is the incidence matrix of the edges of $F$ versus a set of path in $F$. That is,

$$a_{ij} = \begin{cases} 1 & \text{if edge } e_i \text{ is in path } p_j \\ -1 & \text{if } -e_i \text{ is in } p_j \\ 0 & \text{otherwise} . \end{cases}$$

In [2] it has further been shown that:

(iii) there exist unimodular matrices which do not belong to $H$;

(iv) the classes $H$ and $T$ are identical.

The purpose of the present note is to show that the class $K$ is identical with the set of nonnegative matrices of $H$.

2. THEOREM. *If a matrix $A$ of $n$ rows and $m$ columns belongs to $K$ (i.e., $A$ is the incidence matrix of the $n$ vertices of some alternating graph $G$ versus a set of $m$ directed paths in $G$), then $A$ belongs to $H$ (i.e, there is some $n$-simplex $S$ and a basis $B$ among its edges such that the columns of $A$ represent edges of $S$ in terms of $B$). Conversely, every non-negative matrix of $H$ belongs to $K$.*

3. NOTATION. An oriented graph is viewed as a set

$$R = V \cup E\,, \tag{3.1}$$

where $V$ is the set of vertices $A_1, A_2, \cdots, A_n$, and $E$ is the set of oriented edges $e_\nu$, that is certain ordered pairs $(A_i, A_j)$ with $j \neq i$ of elements of $V$, such that at most one of the two pairs $(A_i, A_j), (A_j, A_i)$ is in $E$.

For brevity of notation we define

$$[A_i, A_j] = \{(A_i, A_j), (A_j, A_i)\}\ . \tag{3.2}$$

The origin and endpoint of an edge $e$ are denoted by $\rho e$ and $\sigma e$:

$$\rho(A, B) = A\ , \quad \sigma(A, B) = B\ , \tag{3.3}$$

If $A$ and $B$ are vertices of $R$, the relation $A \prec B$ ($A$ is immediate predecessor of $B$), also written as $B \succ A$, is defined by

$$A \prec B \Longleftrightarrow (A, B) \in R . \tag{3.4}$$

Similarly, if $a,b$ are edges of $R$,

$$a \prec b \Longleftrightarrow \sigma a = \rho b . \tag{3.5}$$

A subset $V'$ of vertices of $R$ defines a subgraph of $R$

$$R(V') = V' \cup E' \tag{3.6}$$

where $(A, B) \in E' \Longleftrightarrow A \in V', B \in V', (A, B) \in E$.

4. *Proof.* Using the graph-theoretical definition of the class $H$, the first half of the theorem shall be proved by showing that to each alternating graph $G$ there is an oriented loopless graph $F$ such that the $K$-matrices associated with $G$ are among the $H$-matrices associated with $F$.

A column of a $K$-matrix is the incidence column $K_p$ of the vertices of $G$ versus a directed path $p$ in $G$; a column of an $H$-matrix is the incidence column $H_q$ of the edges of $F$ versus a path $q$ in $F$. For given $G$ it will therefore be sufficient to show the existence of an $F$ such that

(4.1) to each directed path $p$ in $G$ there is a path $q = \varphi(p)$ in $F$ such that $K_p = H_q$ .

This will be shown by constructing an $F$ and a mapping $\varphi$ of the set of vertices of $G$ onto the set of edges of $F$ in such a way that $\varphi$ satisfies (4.1), or equivalently, that $\varphi$ preserves the relation defined in (3.4) and (3.5), that is, for any two distinct vertices $A, B$ of $G$,

$$A \prec B \text{ (in } G) \Longrightarrow \varphi(A) \prec \varphi(B) \text{ (in } F) . \tag{4.2}$$

The construction of $F$ and $\varphi$ shall now be carried out under the assumption that $G$ is connected. If $G$ is not connected, the same construction can be applied to each component of $G$, yielding an $F$ with an equal number of components.

If $G$ has $n$ vertices, take as the vertices of $F$ a set of $n + 1$ distinct elements $P_0, P_1, \cdots, P_n$.

The $n$ edges $e_1, e_2, \cdots, e_n$ of $F$ are defined successively as follows.

First, choose an arbitrary vertex $A_1$ in $G$, define

$$\varphi(A_1) = e_1 = (P_0, P_1) , \tag{4.3}$$

and note that:

(i) the subgraph $G_1 = G(A_1)$, consisting of the one vertex $A_1$ of $G$, is, trivially, connected;

(ii) the graph $F_1 = \{P_0, P_1, (P_0, P_1)\}$ is connected;

(iii) with respect to $G_1$ and $F_1$, $\varphi$ trivially satisfies (4.2).

Then, assuming $A_\nu \in G$ already chosen and $e_\nu = \varphi(A_\nu)$ defined for $\nu = 1, 2, \cdots, k$ in such a manner that $G_k = G\{A_1, A_2, \cdots, A_k\}$ and $F_k = \{P_0, P_1, \cdots, P_k, e_1, \cdots, e_k\}$ are each connected and $\varphi$ satisfies (4.2) with respect to $G_k$ and $F_k$, choose $A_{k+1} \in G$ such that

$$[A_i, A_{k+1}] \cap G \neq 0 \tag{4.4}$$

for some $i \leqq k$ and define

$$\varphi(A_{k+1}) = e_{k+1} = \begin{cases} (\sigma e_i, P_{k+1}) \text{ when } (A_i, A_{k+1}) \in G \\ (P_{k+1}, \rho e_i) \text{ when } (A_{k+1}, A_i) \in G , \end{cases} \tag{4.5}$$

noting that this definition depends on the choice of $i$ since more than one $i$ may satisfy (4.4).

Obviously, $G_{k+1}$ and $F_{k+1}$ are each connected.

To show that $\varphi$ satisfies (4.2) with respect to $G_{k+1}$ and $F_{k+1}$, let $A_r \prec A_s$ in $G_{k+1}$.

If $r \leqq k$ and $s \leqq h$, (4.2) is satisfied according to the induction's hypothesis.

For $\{r, s\} = \{i, k+1\}$, (4.2) is satisfied by definition (4.5). Namely: for $r = i$, $s = k+1$, (4.5) defines $e_{k+1} = (\sigma e_i, P_{k+1})$, hence $\sigma e_i = \rho e_{k+1}$, which by (3.5) means $e_i \prec e_{k+1}$; similarly for $s = i$, $r = k+1$, (4.5) defines $e_{k+1} = (P_{k+1}, \rho e_i)$, hence $\sigma e_{k+1} = \rho e_i$, which means $e_{k+1} \prec e_i$.

There remains the case $\{r, s\} = \{j, k+1\}, j \neq i, j \leqq k$, with

$$[A_j, A_{k+1}] \cap G_{k+1} \neq 0 , \tag{4.6}$$

that is either $A_j \prec A_{k+1}$ or $A_{k+1} \prec A_j$ in $G_{k+1}$.

In this case $A_{k+1}$, which by (4.4) has an edge in common with $A_i$, now also has an edge in common with $A_j \neq A_i$, thus connecting these two distinct vertices of $G_k$ by the path

$$A_i, A_{k+1}, A_j \tag{4.7}$$

in $G_{k+1}$ but outside $G_k$.

On the other hand, by the induction's hypothesis, $G_k$ is connected. Hence $A_i$ and $A_j$ are connected by a path in $G_k$

$$A_i, A_{t_1}, A_{t_2}, \cdots, A_{t_\lambda}, A_j \tag{4.8}$$

($\lambda = 0$ not a priori excluded).

The paths (4.7) and (4.8) combine to the loop

$$A_{k+1}, A_i, A_{t_1}, A_{t_2}, \cdots, A_{t_\lambda}, A_j, A_{k+1} \tag{4.9}$$

in $G_{k+1}$, which is obviously also a loop in $G$.

Since $G$ is alternating, the loop (4.9) must be alternating. This implies that the number of vertices is even, hence $\lambda = 2\nu + 1$, and that the orientation is either

$$A_{k+1} \prec A_i \succ A_{t_1} \prec A_{t_2} \succ \cdots \prec A_{t_{2\nu}} \succ A_{t_{2\nu+1}} \prec A_j \succ A_{k+1} \tag{4.10}$$

or the opposite.

Now assume first

$$A_{k+1} \prec A_j , \tag{4.11}$$

which implies the orientation (4.10), and consider that part of the loop which is in $G_k$, namely the path (4.8)

(4.10) and the induction's hypothesis that, relative to $G_k$ and $F_k$, $\varphi$ satisfies (4.2), imply

$$e_i \succ e_{t_1} \prec e_{t_2} \succ \cdots \prec e_{t_{2\nu}} \succ e_{t_{2\nu+1}} \prec e_j , \tag{4.12}$$

hence

$$\rho e_i = \sigma e_{t_1} = \rho e_{t_2} = \sigma e_{t_3} = \cdots = \rho e_{t_{2\nu}} = \sigma e_{t_{2\nu+1}} = \rho e_j . \tag{4.13}$$

The definition (4.5) of $e_{k+1}$, in conjunction with $A_{k+1} \prec A_i$ from (4.10), implies

$$\sigma e_{k+1} = \rho e_i . \tag{4.14}$$

This together with (4.13) yields

$$\sigma e_{k+1} = \rho e_j , \quad \text{that is } e_{k+1} \prec e_j , \tag{4.15}$$

which proves that assumption (4.11) implies (4.15).

Similarly, the assumption $A_{k+1} \succ A_j$ yields $e_{k+1} \succ e_j$, by reversing the relation $\prec$ and interchanging $\rho$ and $\sigma$ in the above argument.

This completes the proof that to any connected alternating graph $G$ there exists a connected oriented graph $F$ and a mapping $\varphi$ satisfying (4.2)

That $F$ has no loops (and hence is a tree) is obvious from the fact that its $n + 1$ vertices are connected by $n$ edges. Hence, the incidence matrices of $F$ certainly belong to class $H$.

If $G$ consists of $k$ components, the construction will yield an $F$ consisting of $k$ trees.

This completes the proof of the theorem's first half, namely that every $K$-matrix is an $H$-matrix.

The second half of the theorem, namely that each nonnegative $H$-matrix is a $K$-matrix, is due to J. Edmonds. It will be proved by showing that to each loopless oriented $F$ there is an alternating $G$ and a mapping $\psi$ of the edges of $F$ onto the vertices of $G$ that preserves the relation $\prec$, that is, for any two edges $a, b$ of $F$

$$(4.16) \qquad a \prec b \Longrightarrow \psi(a) \prec \psi(b) .$$

This is achieved by the following simple construction.

If $F$ has $n$ edges $e_1, e_2 \cdots, e_n$, choose a set of $n$ elements $A_1, A_2, \cdots, A_n$ as the vertices of $G$, define $\psi$ by

$$(4.17) \qquad \psi e_i = A_i ,$$

and define the edges of $G$ by

$$(4.18) \qquad (A_i, A_j) \in G \Longleftrightarrow e_i \prec e_j ,$$

that is, $G$ shall have an edge oriented from $A_i$ to $A_j$ if and only if $\sigma e_i = \rho e_j$.

Obviously $\psi$ preserves the relation $\prec$, since (4.18) is equivalent to

$$(4.19) \qquad A_i \prec A_j \Longleftrightarrow e_i \prec e_j .$$

Note that $\prec$ is also preserved by the inverse of $\psi$, that is, in the transition from $G$ to $F$.

Note further that $G$ is oriented (in the sense of the definition given in [4] and cited in § 1 of present note), that is:

(a) each edge of $G$ is oriented, since the edges of $G$ have been defined by (4.18) as oriented edges;

(b) $G$ has no circular edge, since $(A_i, A_i) \in G$ for some $i$ would imply $e_i \prec e_i$, or equivalently $\sigma e_i = \rho e_i$, that is, $e_i$ a circular edge in $F$, contradicting the assumption on $F$;

(c) $G$ has at most one edge between any given two vertices: $(A_i, A_j) \in G$ and $(A_j, A_i) \in G$ for some pair $i, j$, would imply $e_i \prec e_j$ and $e_j \prec e_i$, that is $\sigma e_i = \rho e_j$ and $\sigma e_j = \rho e_i$, hence $e_i$ and $e_j$ would form a 2-loop (with the vertices $\rho e_i$, $\sigma e_i$), again contradicting the assumption on $F$.

Finally, to show that $G$ is alternating, note that, by (4.17) and (4.19), $G$, $F$ and $\varphi = \psi^{-1}$ satisfy the condition (4.1). Thus the incidence matrices (of vertices versus directed paths) associated with $G$ are among the incidence matrices (edges versus paths) associated with $F$, and hence unimodular. Especially then, the incidence matrix of the vertices versus *all* the directed paths of $G$ is unimodular, which, by the Hoffman-Kruskal Theorem (Theorem 4 in [4], cited in § 1 of this note), implies that $G$ is necessarily alternating.

This completes proof of the theorem.

It is worth noting that the last part of the proof (namely that $G$ is alternating) can easily be established without using the result of [4] (which contains more than is needed here).

## References

1. I. Heller and C. Tompkins, *An extension of a theorem of Dantzig's*, Ann. Math. Study No. 38.

2. I. Heller, *On linear systems with integralvalued solutions*, Pacific. J. Math., **7** (1957), 111-111.

3. ———, *Constraint Matrices of Transportation-Type Problems*, Naval Res. Logistics Quarterly, Vol. 4, No. 1.

4. A. J. Hoffman and J. B. Kruskal, *Integral Boundary Points of Convex Polyhedra*, Ann. Math. Study No. 38.

STANFORD UNIVERSITY
IBM RESEARCH CENTER

# DUALITY IN GENERAL ERGODIC THEORY

ROBERT G. HEYNEMAN

**Introduction.** Let $G$ be a semi-group of operators acting on a Banach space $E$. Alaoglu-Birkhoff [1], Eberlein [8], Jacobs [9], deLeeuw and Glicksberg [12], and others have given conditions under which certain orbits (see § 1) in $E$ will contain a single fixed vector under the action of $G$. In general of course, a given orbit may contain many fixed points or none at all; moreover it need not be the case that the 'ergodic' vectors (those whose orbits contain a single fixed point) form a linear subspace as one would wish.

The object of this paper is to show how the introduction of considerations involving the conjugate space $E^*$ under the action of the adjoint semi-group $G^*$ illuminate these matters. We shall see that there is an intimate connection between the existence of fixed points in the orbits of one space and the uniqueness of fixed points in the orbits of the associated space. Our first result in this direction, Theorem 1.3, asserts that if every orbit in one space contains *at least* one fixed point, then every orbit in the other contains *at most* one fixed point.

In § 2 we define what we mean by saying that the semigroup $G$ acts *ergodically* on the space $E$. When this is the case the pathology that arises from the existence of more than one fixed point in a given orbit of $E$ cannot occur. Thus the ergodicity of $G$ on $E$ may be considered as a strong uniqueness requirement on the fixed points of orbits of $E$. When $E$ is reflexive we can then show that this requirement (that $(G, E)$ is ergodic) may be *characterized* by the fact that every orbit of the conjugate space $E^*$ contains at least one fixed point under the adjoint semi-group $G^*$. Indeed whether $E$ is reflexive or not, Theorem 3.1 asserts that the 'ergodic behaviour' of the orbits of one space insures the existence of (at least) one fixed point in any weakly compact orbit of the other.

These results 'explain' and unify many earlier results which were obtained using different specialized techniques. The following two examples are instructive:

(a) $G$ is abelian. As it is quite trivial to verify that abelian semi-groups act ergodically, both $(G, E)$ and $(G^*, E^*)$ are ergodic. Then since $(G, E)$ is ergodic (respectively $(G^*, E^*)$ is ergodic), we see that every weakly compact orbit of $E$ contains at most (respectively at least) one fixed point. Thus weakly compact orbits contain precisely one fixed point (cf. for example [8]).

(b) $G$ is a group acting on a Hilbert space $E$. Here one can show

Received June 2, 1962. This work was supported by the Office of Ordnance Research (Contract DA-04-200-ORD-171).

(Jacobs [9]) that whenever a (bounded) group acts on a Hilbert space, any orbit contains at least one fixed point. Since this applies to $(G^*, E^*)$ as well as to $(G, E)$, we see from the previous discussion that every orbit of $E$ contains exactly one fixed point. Jacobs makes use of the special nature of the hypotheses on $G$ and $E$ in deriving the uniqueness of fixed points in orbits in case (b). Nonetheless the spirit of his argument is akin to ours and suggested the interest of such an investigation as the present one.

In the last section of the paper we describe the relationship between ergodicity and invariant means. We show in particular that if $(G, E)$ is ergodic where $G$ is a semi-group of 'transition operators' on an appropriate space $E$, then $E$ admits a mean which is invariant under $G$.

This paper represents part of the authors doctoral dissertation presented in 1957 at the University of California at Berkeley. In the authors thesis ergodicity was also characterized generally in terms of the notion of convergence due to Birkhoff and Alaoglu [1]. Here we have preferred to proceed independently of all convergence considerations in an entirely self-contained way. The author would like to thank Professor F. Wolf, under whose direction the thesis was written for his generous help and advice.

**1. Fixed points in orbits.** Throughout this paper $G$ will denote a *bounded* semi-group of linear transformations acting on a Banach space $E$. This means simply that $G$ is closed under multiplication and that there is a positive number $M$ such that $||gx|| < M||x||$ for all $x \in E$ and $g \in G$. We will assume that $G$ contains the identity transformation. If $x \in E$, the closed convex hull of the set $\{gx; g \in G\}$ will be referred to as the *orbit* of $x$ and denoted $K(x)$; subsets of this type will frequently be called orbits without specific reference to the generating vector. $\boldsymbol{G}$ will denote the collection of operators on $E$ which are convex combinations of elements of $G$. Then $\boldsymbol{G}$ is a bounded semi-group in its own right with the same bound $M$, the same orbits, and the same fixed points as $G$. Clearly $K(x) = \text{closure}\,\{\boldsymbol{g}x; \boldsymbol{g} \in \boldsymbol{G}\}$. Finally we define $N = \{x \in E; 0 \in K(x)\}$, $F = \{x \in E; gx = x \text{ for all } g \in G\}$, $D = \{x - gx; x \in E \text{ and } g \in G\}$, and $[D] =$ the closed subspace spanned by $D$.

In passing to the action of the adjoint semi-group $G^*$ on the conjugate space $E^*$, the corresponding dual objects are naturally defined. Thus if $\xi \in E^*$, the orbit of $\xi$ will mean the closed convex hull of $\{g^* \xi; g^* \in G^*\}$ and will be denoted again $K(\xi)$. In the same spirit we define $\boldsymbol{G}^*$, $N_*$, $D_*$ and $F_*$.

We use the notation $(x, \xi)$ to express the linkage between a vector $x \in E$ and a vector $\xi \in E^*$. If $S$ is a subset of $E$ and $T$ is a subset of $E^*$, we set $S^\perp = \{\xi \in E^*; (x, \xi) = 0 \text{ for all } x \in S\}$ and $T^\perp = \{x \in E; (x, \xi) = 0$ for all $\xi \in T\}$. Recall that $S^{\perp\perp} = [S]$, the closed subspace spanned by $S$.

The following technical proposition expressing the relationships between the sets we have defined will be used repeatedly throughout this paper.

1.1. PROPOSITION.

1.1.1. *$N \cap F = 0$*

1.1.2. *$D^{\perp} = F_*$ and $D_*^{\perp} = F$*

1.1.3. *$N$ is closed in $E$*

1.1.4. *$D \subset N \subset [D]$*

1.1.5. *If $x_0 \in F$, then $x_0 \in K(x)$ if and only if $x - x_0 \in N$*

1.1.6. *If $[D] \cap F = 0$, then any orbit of $E$ can contain at most one fixed point.*

*Proof.* If $x \in N \cap F$, then $K(x) = \{x\}$ and also $0 \in K(x)$ so that (1) follows. (2) is immediate by virtue of the identity $(x - gx, \xi) = (x, \xi - g^* \xi)$. To prove (3), let $x \in \bar{N}$ and choose $n \in N$ with $\|x - n\| < \varepsilon$. We can then find $\boldsymbol{g} \in \boldsymbol{G}$ with $\|\boldsymbol{g}n\| < \varepsilon$. We now have $\|\boldsymbol{g}x\| \leqq \|\boldsymbol{g}(x - n)\| + \|\boldsymbol{g}n\| < (M + 1)\varepsilon$ so that $x \in N$.

To prove (4), we define $\boldsymbol{g}_n = 1/n\,(1 + g + g^2 + \cdots + g^{n-1})$ where $g \in G$. Then $\boldsymbol{g}_n \in \boldsymbol{G}$. If now $x - gx \in D$, we have $\boldsymbol{g}_n(x - gx) = 1/n\,(x - g^n x) \to 0$ so that $0 \in K(x - gx)$. Thus $D \subset N$. To show that $N \subset [D]$, it will suffice to prove that $(N, F_*) = 0$ for then $N \subset F_*^{\perp} = D^{\perp\perp} = [D]$. But if $n \in N$ and $\xi \in F_*$ we may choose $\boldsymbol{g} \in \boldsymbol{G}$ with $\|\boldsymbol{g}n\| < \varepsilon/\|\xi\|$. We have:

$$|(n, \xi)| = |(n, \boldsymbol{g}^*\xi)| = |(\boldsymbol{g}n, \xi)| \leqq \|\boldsymbol{g}n\|\,\|\xi\| < \varepsilon$$

so that

$$(n, \xi) = 0\ .$$

If $x_0 \in F$, then $\boldsymbol{g}x - x_0 = \boldsymbol{g}(x - x_0)$ so that $\|\boldsymbol{g}x - x_0\| < \varepsilon$ if and only if $\|\boldsymbol{g}(x - x_0)\| < \varepsilon$. This proves (5).

To prove (6), let $x_1, x_2$ be fixed points in the orbit $K(x)$. Then by (5) $n_1 = x - x_1$ and $n_2 = x - x_2$ are both in $N$. Since $N \subset [D]$, this means that $n_1$ and $n_2$ are in $[D]$ so that $n_1 - n_2 = x_2 - x_1 \in [D] \cap F$. In particular if $[D] \cap F = 0$, then $x_1 = x_2$.

1.2 EXAMPLE. In the classical context where $G$ is the bounded semigroup consisting of the powers of a single operator $T$ (where $\|T^k\| \leqq M$, $k = 1, 2, \cdots$), we may identify $N$ with the closed subspace: $\eta = \{x \in E; T_n x = 1/n(x + Tx + \cdots + T^{n-1}x) \to 0\}$. For, take $x - T^k x \in D$. Then $T_n(x - T^k x) = k/n\,T_k(1 - T^n)x \to 0$ so that $D \subset \eta$. Since $\eta$ is closed, this means that $[D] \subset \eta$ and so $N \subset \eta$. But also if $T_n x \to 0$, then $0 \in K(x)$ so that $\eta \subset N$. Thus $N = \eta$.

If now $x \in E$ and $x_0$ is a fixed point in $K(x)$ then by (1.5) $x - x_0 \in N$ and so $T_n x - x_0 = T_n(x - x_0) \to 0$. Consequently $T_n x \to x_0$. We have thus shown: *if the orbit of a vector* $x \in E$ *contains a fixed point* $x_0$, *then* $T_n x$ *converges to* $x_0$. Conversely the identity $(1 - T)T_n x = 1/n(1 - T^n)x$ shows that if $T_n x$ converges to $x_0$, then $x_0$ is a fixed point. (cf. Eberlein [8]).

1.3 THEOREM. *If every orbit of* $E$ *(respectively* $E^*$*) contains at least one fixed point, then any orbit of* $E^*$ *(respectively* $E$*) contains at most one fixed point.*

*Proof.* If the orbit of every vector $x \in E$ contains a fixed point $x_0$, then $x - x_0 \in N$, so that every vector $x$ in $E$ can be expressed as the sum of a vector $x_0$ in $F$ and a vector $x - x_0$ in $N$. Thus $F^\perp \cap N^\perp = 0$. Now since $D \subset N \subset [D]$ one has $N^\perp = D^\perp = F_*$. Also $(F, D_*) = 0$ so that $F^\perp$ contains $[D_*]$. Consequently $F^\perp \cap N^\perp$ contains $[D_*] \cap F_*$ and so $[D_*] \cap F_* = 0$. Applying Proposition 1.1.6 to the adjoint space, the conclusion then follows.

If every orbit of $E^*$ contains at least one fixed point, then by what we have just shown, any orbit of $E^{**}$ contains at most one fixed point. But because of the isometric imbedding of $E$ in $E^{**}$, the orbit of a vector $x \in E$ is the same whether $x$ is considered to lie in $E$ or $E^{**}$. Thus orbits of $E$ contain at most one fixed point.

1.3.1 COROLLARY. *If every orbit in* $E$ *and in* $E^*$ *contains at least one fixed point, then any orbit in* $E$ *or in* $E^*$ *contains precisely one fixed point.*

1.4 EXAMPLES

1.4.1. If $G$ consists of *contractions*[1] on Hilbert space then any orbit $K(x)$ certainly contains at least one fixed point. For if $x_0$ is the (*unique*) element of $K(x)$ having smallest norm, then since $\|gx_0\| \leqq \|x_0\|$ and $gx_0 \in K(x)$, it follows from the defining property of $x_0$ that $gx_0 = x_0$; that is, $x_0$ is a fixed point.

As the same argument applies to the adjoint semi-group $G^*$ (which also consists of contractions on Hilbert space) we conclude by Corollary 1.3.1 that every orbit contains *precisely* one fixed point (Alaoglu-Birkhoff [1]).

1.4.2. More generally Day [5], pointed out that whenever $G$ consists of contractions on a strictly-convex[2] reflexive space, the above argument is still effective and shows that every orbit contains at least one fixed

[1] An operator $T$ is called a contraction if $\|T\| \leqq 1$. It is called an isometry if $\|Tx\| = \|x\|$ for all $x \in E$.

[2] A Banach space is strictly convex is the unit sphere (vectors of norm 1) contains no line segment.

point. Thus if we assume that *both* $E$ and $E^*$ are strictly convex, Corollary 1.3.1 again allows us to conclude that every orbit contains precisely one fixed point. Such is the case, for example, if $E$ is an $L_p$ space for $p > 1$.

1.4.3. If $G$ consists of a (bounded) *group* acting on a Hilbert space $E_1$ then we may define a new norm on $E$ in the following way: $|x|^2 \equiv \sup_{g \in G} \langle gx, gx \rangle$. This clearly defines an equivalent norm relative to which $G$ acts isometrically. Moreover Jacobs [9] has shown that this new norm is strictly convex. Thus by Day's result, every orbit of $E$ contains at least one fixed point. Since $G^*$ is also a group, the same conclusion is valid for orbits of $E^*$; by Corollary 1.3.1 it then follows that every orbit of $E$ contains precisely one fixed point (Jacobs [9]).

1.5. REMARK. We are indebted to the referee for informing us of some unpublished results of C. Ryll-Nardzewski [14]. His results imply the following: if $G$ is a semi-group of *isometries*[1] on a Banach space $E$ then any *weakly-compact* orbit of $E$ contains fixed points. In particular if $E$ is reflexive (and $G$ consists of isometries) then every orbit of $E$ contains at least one fixed point. If in addition $G^*$ also acts as isometries on the (reflexive) space $E^*$, we conclude by Corollary 1.3.1 that every orbit of $E$ contains precisely one fixed point.

In the same way if $G$ is any (bounded) group acting on a reflexive Banach space we may renorm the space as in Example 3 above so that $G$ consists of isometries. The result of Ryll-Nardzewski thus again applies to show that in this case too every orbit contains precisely one fixed point

## 2. Ergodicity and duality.

We now proceed to an examination of the 'good' case where $G$ acts 'ergodically' on $E$.

2.1 PROPOSITION. *The following conditions are equivalent*:

1. $gN \subset N$ *for any* $g \in G$ (*if* $0 \in K(n)$, *then* $0 \in K(gn)$).
2. *If* $n \in N$ *then* $K(n) \subset N$.
3. $N$ *is a linear subspace of* $E$ (i.e. *if* $0 \in K(x)$ *and* $0 \in K(y)$, *then* $0 \in K(x + y)$).
4. $N = [D]$.

*Proof.* (1) $\Longrightarrow$ (2) since $N$ is closed.

(2) $\Longrightarrow$ (3). Let $x \in N$ and $y \in N$. Choose $g_1 \in G$ such that $\|g_1 x\| < \varepsilon$. Then as $g_1 y \in N$ by (2), we can choose $g_2 \in G$ such that $\|g_2 g_1 y\| < \varepsilon$. We now have:
$\|g_2 g_1(x + y)\| \leqq M\|g_1 x\| + \|g_2 g_1 y\| < (M + 1)\varepsilon$. Thus $x + y \in N$.

(3) $\Longrightarrow$ (4) for $N$ is closed and $D \subset N \subset [D]$.

(4) $\Longrightarrow$ (1). Since $F_*$ is invariant under the action of $\boldsymbol{G}^*$, it follows that $[D] = F_*^\perp$ is invariant under the action of $\boldsymbol{G}$. In particular, if $[D] = N$, then condition (1) is satisfied.

The pair $(G, E)$ will be called *ergodic* if any of the above conditions is satisfied. We will then call a *vector* of $E$ ergodic if its orbit contains a fixed point. By proposition 1.1.5 and proposition 2.1 a vector $x \in E$ is ergodic if and only if it belongs to the *subspace* $R = N \oplus F = [D] \oplus F$. $R$ will be referred to as the *ergodic* subspace. (This nomenclature is in accord with a somewhat unfortunate tradition.)

2.2 EXAMPLES 2.2.1. *If $G$ is abelian, then $(G, E)$ is ergodic.* For if $n \in N$ and $\boldsymbol{g} \in \boldsymbol{G}$ we may choose $\boldsymbol{g}_1 \in \boldsymbol{G}$ with $\|\boldsymbol{g}_1 n\| < \varepsilon$. We then have $\|\boldsymbol{g}_1\boldsymbol{g}n\| = \|\boldsymbol{g}\boldsymbol{g}_1 n\| < M\varepsilon$ so that $\boldsymbol{g}n \in N$. Thus $(G, E)$ is ergodic by proposition 2.1.1.

2.2.2. *If every orbit of $E$ contains precisely one fixed point, then $(G, E)$ is ergodic.* For let $n \in N$ and $\boldsymbol{g} \in \boldsymbol{G}$. Since $K(\boldsymbol{g}n) \subset K(n)$, the fixed point of $K(\boldsymbol{g}n)$ must coincide with that of $K(n)$; that is, $0 \in K(\boldsymbol{g}n)$. Thus $\boldsymbol{g}n \in N$ and so again by proposition 2.1.1 $(G, E)$ is ergodic.

2.2.3. *If $G$ admits a right invariant mean, then $(G, E)$ is ergodic* (Theorem 4.2).

2.2.4. *If every orbit of $E^*$ contains at least one fixed point, then $(G, E)$ is ergodic.* (We will see later—cf. Corollary 3.1.1—that when $E$ is reflexive, $(G, E)$ is ergodic if and only if every orbit of $E^*$ contains at least one fixed point.)

*Proof.* If $(G, E)$ is *not* ergodic, by proposition 2.1.1 there is an $n \in N$ and a $\boldsymbol{g} \in \boldsymbol{G}$ with $\boldsymbol{g}n \notin N$, so that $0 \in K(n)$ but $0 \notin K(\boldsymbol{g}n)$. The Hahn-Banach Theorem then asserts the existence of a functional $\xi \in E^*$ which separates 0 from the closed convex set $K(\boldsymbol{g}n)$; that is, $0 < \alpha \leqq (K(\boldsymbol{g}n), \xi)$ where $\alpha$ is a real number. In particular, $0 < \alpha \leqq (\boldsymbol{G}\boldsymbol{g}n, \xi) = (\boldsymbol{g}n, \boldsymbol{G}^*\xi)$ and so $0 < \alpha \leqq \overline{(\boldsymbol{g}n, \boldsymbol{G}^*\xi)}$. But as $\boldsymbol{g}n$ induces a continuous functional on $E^*$ (in the norm topology on $E^*$) we have:

$$\overline{(\boldsymbol{g}n, \boldsymbol{G}^*\xi)} \supset (\boldsymbol{g}n, \{\overline{\boldsymbol{G}^*\xi}\}) = (\boldsymbol{g}n, K(\xi))$$

so that $0 < \alpha \leqq (\boldsymbol{g}n, K(\xi))$. If then $K(\xi)$ were to contain a fixed point $\xi_0$, we would have $0 < \alpha \leqq (\boldsymbol{g}n, \xi_0) = (n, \xi_0)$ which would contradict the fact that $N$ is perpendicular to $F_*$.

2.3 THEOREM. *If $(G, E)$ is ergodic, then:*

2.3.1. *the ergodic subspace $R$ is closed and strongly invariant* (in the sense that $gx \in R$ if and only if $x \in R$),

2.3.2. *the orbit of a vector $x \in E$ contains precisely one fixed point if $x \in R$ and contains none if $x \notin R$,*

2.3.3. *if $x \in R$ and $p(x)$ is the associated fixed point of $K(x)$ then $x \to p(x)$ defines a (bounded) linear operator $p$ on $R$ such that $pg = gp = p^2 = p$.*[3]

2.3.4. *$F_* \neq 0$ whenever $F \neq 0$; indeed the dimension of $F_*$ is at least as great as the dimension of $F$.*

*Proof.* If $(G, E)$ is ergodic, then $N = [D]$ and so $[D] \cap F = 0$. Then by proposition 1.1.6 orbits of $E$ contain at most one fixed point. Thus if $x \in R$, $K(x)$ contains precisely one fixed point. This gives 2.

By proposition 2.1, $[D] = N$ is invariant under the action of $G$ so that $R = N \oplus F$ is invariant under $G$. Moreover if $gx \in R$, then $K(gx)$ contains a fixed point and since $K(x) \supset K(gx)$, $K(x)$ contains the same fixed point. Thus $x \in R$.

Next we show that $R$ is closed. For $x \in R$, let $p(x)$ denote the unique fixed point in $K(x)$. Then $\|p(x)\| \leqq M\|x\|$ for all $x \in R$. Moreover, if $n \in N$ and $f \in F$, then $p(n + f) = f$ so that $\|f\| = \|p(n + f)\| \leqq M\|n + f\|$.

Suppose then that $x_n \in R$ and that $x_n \to x_0$. Put $x_n = d_n + f_n$ where $d_n \in [D] = N$ and $f_n \in F$. Since $x_n$ is Cauchy and $\|f_n - f_m\| \leqq M\|(d_n - d_m) + (f_n - f_m)\| = M\|x_n - x_m\|$, we conclude that $f_n$ is Cauchy. Thus $f_n \to f_0 \in F$ and so $d_n \to x_0 - f_0 \in N$. Consequently $x_0 \in N \oplus F = R$.

Let $x_1, x_2 \in R$ and $x_i = n_i + f_i$ where $n_i \in N$ and $f_i \in F$. Then $p(x_i) = f_i$ and $p(x_1 + x_2) = p((n_1 + n_2) + (f_1 + f_2)) = f_1 + f_2$ since by Proposition 2.1 $n_1 + n_2 \in N$. Thus $p$ is linear.

Finally in order to prove that $\dim F \leqq \dim F_*$, we may assume that $\dim F_* < \infty$. We then have $F_* = [D]^\perp \cong (E/[D])^*$ so that $\dim F_* = \dim (E/[D]) = \operatorname{codim} [D]$. But if $(G, E)$ is ergodic, $[D] \cap F = 0$, and so in this case $\operatorname{codim} [D] \geqq \dim F$.

This completes the proof.

Part 4 also follows from some results of Yood [15].

Remark. If $E$ is reflexive and both $(G, E)$ and $(G^*, E^*)$ are ergodic, (cf. Corollary 3.1.1) then applying Theorem 2.3.4 to both spaces, we conclude that $\dim F = \dim F_*$. In particular, this equality holds when $G$ is

[3] It is easy to see that $gx \to x_0$ in the sense of Alaoglu-Birkhoff [1] if and only if $x \in R$ and $x_0 = p(x)$.

a group or an abelian semi-group on Hilbert space. When $G$ consists of contractions on Hilbert space, it follows for the same reason that $\dim F = \dim F_*$. But in this case one actually has more; namely $F - F_*$. For if $x \in F$, then $\|x\|^2 = (gx, x) = (x, g^*x) \leqq \|x\| \, \|g^*x\| \leqq \|x\|^2$ and so $\|x\|^2 = \|g^*x\|^2 = (x, g^*x)$. Thus $\|g^*x - x\|^2 = (g^*x - x, g^*x - x) = 0$ so that $g^*x = x$ and $x \in F_*$.

**3. Existence of fixed points.** As the next proof will require us to conduct our arguments in the (non-normed) weak-$*$topology of $E^*$, we mention here some of the relevant background. The definitions and proofs of the results we use concerning topological vector spaces can be found in Bourbaki [3].

We define the *weak-topology* on a Banach space $E$ to be the least fine topology relative to which all the elements of the adjoint space $E^*$ are continuous. Thus the weak topology on the Banach space $E^*$ is defined using the elements of $E^{**}$. By the *weak-$*$ topology* on $E^*$ we mean the least fine topology relative to which the elements of $E$ induce continuous functionals. By definition then, when $E^*$ is endowed with the weak-$*$ topology, any element of $E$ induces a continuous functional on it. One can show that *every* functional on $E^*$ which is continuous in the weak-$*$ topology arises in this way from an element of $E$.

Although the norm topology is in general definitely richer in closed sets than the weak topology, Mazur's theorem asserts that a *convex* closed set is weakly closed. We will also make use of the fact that the unit ball of $E^*$ is compact in the weak-$*$ topology. Finally, as the new topologies on $E$ and $E^*$ are locally convex (that is, every vector possesses a fundamental system of convex neighborhoods), the theorems of the Hahn-Banach type apply. These guarantee in particular the existence of continuous functionals strictly separating a given closed convex subset from a disjoint compact convex subset.

3.1 THEOREM. (a) *If $(G, E)$ is ergodic, then any convex weak-$*$ compact subset of $E^*$ which is invariant under the action of $G^*$, contains a fixed point. In particular, any orbit of $E^*$ which is compact in the weak-$*$ topology contains a fixed point.*

(b) *If $(G^*, E^*)$ is ergodic, then any convex weakly-compact subset of $E$ which is invariant under the action of $G$, contains a fixed point. In particular, any orbit of $E$ which is compact in the weak topology contains a fixed point.*

REMARK. Since by Mazur's theorem orbits are weakly closed, the two assertions of part (b) are actually equivalent. Orbits of $E^*$, on the other hand, need not be weak-$*$ closed so that the first assertion of part (a) is really stronger than the second one.

*Proof of the theorem.* Part (a). Suppose that $W$ is a convex, weak-$*$ compact subset of $E^*$ which is invariant under $G^*$ and yet which does not contain fixed points. Then $W \cap F_* = \phi$. As $F_* = D^\perp$, $F_*$ is weak-$*$ closed. In the space $E^*$ *endowed with the weak-$*$ topology* we may then apply one form of the Hahn-Banach Theorem to the disjoint closed convex set $F_*$ and the compact convex set $W$. This theorem asserts the existence of a weak-$*$-continuous functional on $E^*$ which *strictly* separates $W$ and $F_*$. As we have seen that all such functionals arise from elements in $E$, there is then an $x \neq 0$ in $E$ and a real number $\alpha$ with $(x, F_*) < \alpha < (x, W)$. Since $F_*$ is a subspace, this requires $(x, F_*)$ to be zero and so $x \in F_*^\perp = [D]$.

Thus $0 < \alpha < (x, W)$ where $x \in [D]$. Choosing an arbitrary $\omega_0 \in W$, we have $G^*\omega_0 \subset W$ so that $0 < \alpha < (x, G^*\omega_0) = (Gx, \omega_0)$ and hence $0 < \alpha \leqq \overline{(Gx, \omega_0)}$. But as $\omega_0$ is continuous in the norm topology of $E$, $(K(x), \omega_0) = (\{\overline{Gx}\}, \omega_0) \subset \overline{(Gx, \omega_0)}$ and so $0 < \alpha \leqq (K(x), \omega_0)$. In particular then, we conclude that $0 \notin K(x)$; that is, $x \notin N$. As $x$ was shown to be in $[D]$, this means that $(G, E)$ cannot be ergodic.

Part (b). Since the weak-$*$ topology of $E^{**}$ induces on $E$ (which is naturally imbedded in $E^{**}$) a topology which coincides with the ordinary weak topology of $E$, a subset of $E$ which is weakly compact may be considered a subset of $E^{**}$ which is compact in the weak-$*$ topology on that space. An application of part (a) then gives part (b).

3.1.1 COROLLARY. *If $E$ is reflexive, then $(G, E)$ is ergodic if and only if every orbit of $E^*$ contains at least one fixed point.*

*Proof.* Orbits are bounded and (by Mazur's theorem) weakly closed so that in a reflexive space any orbit is weakly compact. Theorem 3.1 thus gives the forward implication. Example 2.2.4 gives the converse independently of reflexivity.

3.2 REMARKS.

3.2.1. If both $(G, E)$ and $(G^*, E^*)$ are ergodic, then any orbit in either $E$ or $E^*$ can contain at most one fixed point (Theorem 2.3.2). But then by Theorem 3.1, any weakly compact orbit of $E$ or any weak-$*$ compact orbit of $E^*$ must contain precisely one fixed point. Since an abelian semi-group always acts ergodically, these results are valid in particular in the case where $G$ is abelian ($G^*$ is also abelian). We shall see that the same is true when $G$ possesses a two-sided invariant mean (Remark 4.3.1.).

3.2.2. By the corollary of the last theorem we see that when $E$ is reflexive every orbit contains precisely one fixed point if and only if

either (a) both $(G, E)$ and $(G^*E^*)$ are ergodic

or (b) every orbit of $E$ and of $E^*$ contains at least one fixed point.

In the case where $G$ is abelian, (a) is immediate. If $E$ is a Hilbert space and $G$ is either a semi-group of contractions or a (bounded) group, we have seen that (b) holds. In fact using the results of C. Ryll-Nardzewski [14] mentioned in remark 1.5., (b) is valid on *any* reflexive Banach space so long as $G$ is a bounded group or if both $G$ and $G^*$ are each semi-groups of isometries. His proof presumably depends on delicate measure-theoretic machinery. It would be interesting to see if one could prove (a) directly in these cases.

4. **Means and ergodicity.** Let $E$ be the Banach space $B(\Omega)$ of all bounded continuous functions on the completely regular topological space $\Omega$ under sup norm[4]. An element $\xi \in E^*$ is called positive if $\xi(f) \geqq 0$ whenever $f \geqq 0$. In that case it is clear that $\|\xi\| = \xi(1)$. A positive functional $\lambda \in E^*$ is called a *mean* on $E$ if $\lambda(1) = 1$. We then have $\|\lambda\| = 1$ and moreover $|\lambda(f)| \leqq \lambda(|f|) \leqq \|f\|$ for any $f \in E$. Let $P$ be the set of means on $E$. Evidently $P$ is weak-$*$ closed in $E^*$. As $P$ is convex and is contained in the (weak-$*$ compact) unit ball of $E^*$, it follows that $P$ is *weak-$*$ compact.*

An operator $T$ on $E$ is called an *endomorphism* (or a transition-operator) if $Tf \geqq 0$ whenever $f \geqq 0$ and also $T1 = 1$. This is equivalent to requiring $T^*P \subset P$; i.e., that the set of means on $E$ is carried into itself by $T^*$. Finally as $\|f\| \leqq 1$ if and only if $-1 \leqq f \leqq 1$, we see that the norm of an endomorphism is 1.

Suppose now that $G$ is a semi-group of endomorphisms on the Banach space $E$ as above. Since every element of $G$ has norm 1, $G$ is bounded in the sense of our earlier discussion. A mean $\Lambda$ on $E$ is said to be *invariant* (under $G$) if $\Lambda(gf) = \Lambda(f)$ for each $g \in G$, $f \in E$. Thus an invariant mean on $E$ is simply an element of $P \cap F_*$. In general, of course, this set may be empty. However, as a consequence of Theorem 3.1, we have:

4.1 THEOREM. *If $G$ is a semi-group of endomorphisms which acts ergodically on $E$, then $E$ possesses an invariant mean.*

*Proof.* The set $P$ of means on $E$ is a convex weak-$*$ compact subset of $E^*$. The fact that $G$ consists of endomorphisms means that $P$ is

[4] The development in this paragraph could be carried through taking for $E$ what Kakutani [11] has called an abstract $M$-space with unit, but his results show that the generality gained is only formal.

carried into itself by every element of $G^*$. But then by Theorem 3.1 (a), $P$ contains a fixed point for $G$.

4.1.1. COROLLARY. *If $G$ is abelian, then $E$ possesses an invariant mean* (Kakutani [10] and Day [4]).

If the space $\Omega$ is a (topological) semi-group $G$, then there are several ways for $G$ to induce endomorphisms on the space $E = B(G)$. If $g \in G$ and $f \in B(G)$ let us define the operators $L_g$ and $R_g$ on $E$ by: $(L_g f)(g') = f(gg')$ and $(R_g f)(g') = f(g'g)$. We then have $L_{g_1}L_{g_2} = L_{g_2 g_1}$, $R_{g_1}R_{g_2} = R_{g_1 g_2}$ and $R_{g_1}L_{g_2} = L_{g_2}R_{g_1}$ so that $G_L = \{L_g;\, g \in G\}$ and $G_R = \{R_g;\, g \in G\}$ form two semi-groups of endomorphisms on $E$ which commute elementwise. Thus $G_T = \{L_{g_1}R_{g_2};\, g_1, g_2 \in G\}$ also forms a semigroup of endomorphisms on $E$. Corresponding to these three semigroups we obtain the notion of *left*, *right* and *two-sided* invariant means on $E$ (or as we shall say, on $G$). By Corollary 4.1.1 any *abelian* semi-group possesses an invariant mean. The existence of Haar measure (cf. § 4.4) shows that any *compact group* possesses a (unique!) invariant mean.

Heretofore, in the discussion of bounded semi-groups of operators, the *topology* on the semi-group played no role. We might equally well have been dealing with an *abstract* semi-group $G$, together with a bounded representation of $G$ into the multiplicative semi-group of operators on $E$, where we call a representation $\pi$ of $G$ bounded when the image semi-group is bounded. If $G$ is a topological semi-group, we will say that the representation $\pi$ is weakly-continuous if $g \to (\pi(g)x, \xi)$ is a continuous function on $G$ for any $x \in E$, $\xi \in E^*$. For convenience we omit the letter $\pi$ and speak of the continuity of $(gx, \xi)$, the ergodicity of $(G, E)$, etc.

4.2 THEOREM. *Let $\pi$ be a bounded weakly-continuous representation of the topological semi-group $G$ on the Banach space $E$. Then if $G$ admits a right invariant mean, $(G, E)$ is ergodic.*

*Proof.* For $x \in E$, $\xi \in E^*$, let $[x, \xi]$ denote the function in $B(G)$ whose value at $g$ is $(gx, \xi)$. If $\bigwedge$ denotes a right invariant mean on $G$, then we may define a transformation $T\colon E \to E^{**}$ by means of the equation $(Tx, \xi) = \bigwedge([x, \xi])$. Then

$$\begin{aligned} \|Tx\| &= \sup_{\|\xi\| \leq 1} |(Tx, \xi)| = \sup_{\|\xi\| \leq 1} |\bigwedge([x, \xi])| \leqq \sup_{\|\xi\| \leq 1} \|[x, \xi]\| \\ &= \sup_{\|\xi\| \leq 1} \sup_{g} |(gx, \xi)| \leqq M \|x\| \end{aligned}$$

Thus $T$ is continuous.

Observing that $R_g([x, \xi]) = [gx, \xi]$ we have:

$$(Tgx, \xi) = \bigwedge([gx, \xi]) = \bigwedge(R_g[x, \xi]) = \bigwedge([x, \xi]) = (Tx, \xi)$$

for any $\xi \in E^*$ so that $Tgx = Tx$ and consequntly $T$ vanishes on $D$. By the continuity of $T$ then, $T$ vanishes on $[D]$ and so a fortiori on $N$.

But now conversely if $Tx = 0$, we claim that $0 \in K(x)$ so that $x \in N$. For otherwise, we could find $\xi \in E^*$ and a real number $\alpha$ with $0 < \alpha \leqq (K(x), \xi)$. In particular $\alpha \leqq (gx, \xi)$ for all $g \in G$ and so $[x, \xi] \geqq \alpha$. But then $(Tx, \xi) = \bigwedge([x, \xi]) \geqq \alpha > 0$, which contradicts the fact that $Tx = 0$. Thus $Tx = 0$ if and only if $x \in N$, so that $N$ is a linear subspace and $(G, E)$ is ergodic.

4.3.1. Let $G$ be a bounded semi-group of operators on $E$. If $G$ admits a *right* invariant mean when given either the discrete or the uniform operator topology then Theorem 4.2. applies so that $(G, E)$ is ergodic. If instead $G$ admits a *left* invariant mean in either of these topologies then $G^*$ admits a right invariant mean in the same topology so that $(G^*, E^*)$ is ergodic. Thus in this case by Theorem 3.1 any invariant compact convex set of $E$ contains a fixed point (cf. Day [6]). In particular if $G$ admits a two-sided invariant mean in either topology then any compact convex orbit of $E$ must contain precisely one fixed point.

4.3.2. Combining Theorem 4.1 and 4.2, we see that if $G$ is a (topological) semi-group of endomorphisms of the space $E = B(\Omega)$ (and $g \to (gx, \xi)$ is continuous) then whenever $G$ possesses a right invariant mean, $E$ also possesses a mean which is invariant under $G$.

### 4.4. Application to Haar Measure on a Locally Compact Group.

As an amusing application of the fact that *abelian* semi-groups admit invariant means, we give here a construction of Haar measure (or rather of a nontrivial invariant content[5]) on an *arbitrary* locally compact group.

Suppose then that $\mathscr{G}$ is a locally compact group and let $G$ denote the collection of neighbourhoods of the identity $e$ in $\mathscr{G}$. Then $G$ is an abelian semi-group under the operation of *intersection*! Let $\Lambda$ be an invariant mean on $G$. We wish to associate with each compact subset $K$ of the group $\mathscr{G}$ a bounded function $\widetilde{K}$ on $G$ in such a way that $\lambda: K \to \Lambda(\widetilde{K})$ will define a nontrivial invariant content. Let $K_0$ be a fixed *compact* neighborhood of $e$. Then if $S \subset \mathscr{G}$ and the interior of $S$ is nonvoid, define $(K:S)$ as the smallest integer $n$ such that $K$ can be covered by $n$ (left) translates of $S$. We now define the function $\widetilde{K}$ on $G$ by setting $\widetilde{K}(V) = (K: V)/(K_0: V)$ where $V \in G$. As $\widetilde{K}(V) \leqq (K: K_0)$, $\widetilde{K}$ is bounded and we may define $\lambda(K) = \Lambda(\widetilde{K})$. Observing that $(gK: V) = (K: V)$ for $g \in \mathscr{G}$, we have $\widetilde{K} = \widetilde{(gK)}$ and so $\lambda(gK) = \lambda(K)$. Also if $K$ has a nonvoid interior, then $K(V) \geqq 1/(K_0: K)$ so that in this case

[5] cf. Halmos-Measure Theory, Theorem B, p. 254.

$\lambda(K) \geqq 1/(K_0 : K) > 0$. It is clear that in general $\lambda(K_1 \cup K_2) \leqq \lambda(K_1) + \lambda(K_2)$. If, moreover, $K_1$ and $K_2$ are *disjoint*, $(K_1 \cup K_2 : V) = (K_1 : V) + (K_2 : V)$ for all small enough $V \in G$. By virtue of the invariance of $\Lambda$, we then have $\lambda(K_1 \cup K_2) = \lambda(K_1) + \lambda(K_2)$.

## Bibliography


1. L. Alaoglu and G. Birkhoff, *General ergodic theorems*, Ann. of Math., **41** (1940), 293-309.
2. G. Birkhoff, *An ergodic theorem for general semi-groups*, Proc. Nat. Acad. Sci. USA, **25** (1939), 625-627.
3. N. Bourbaki, *Elements de Mathematique*, Livre V. Espaces Vectorielles Topologiques, Paric (1953).
4. M.M. Day, *Ergodic Theorems for Abelian semi-groups*, Trans. Amer. Math Soc., **51** (1942), 399-412.
5. ———, *Reflexive Banach spaces not isomorphic to uniformly convex spaces*, Bull. Amer. Math. Soc., **47** (1941), 313-317
6. ———, *Fixed-point theorems for compact convex sets*, Ill. J. Math., **5** (1961), 585-590.
7. J. Dixmier, *Les Moyennes invariantes dans les semi-groupes et leurs applications*, Acta Scientif. Math., **12** (1950).
8. W.F. Eberlein, *Abstract ergodic theorems and weakly almost periodic functions*, Trans. Amer. Math. Soc., **67** (1949), 217-240.
9. K. Jacobs, *Ein Ergodensatz fur beschrankte Gruppen im Hilbertschen Raum*, Math. Ann., **128** (1954), 340-349.
10. S. Kakutani, *Two fixed point theorems concerning bicompact convex sets*. Proc. Imp. Acad. Tokyo, **14** (1938), 242-245.
11. ———, *Concrete representation of abstract M-spaces*, Ann. Math., **42** (1941), 994-1024.
12. K. deLeeuw and I. Glicksberg, *Applications of almost periodic compactications*, Acta Mathematica **105** (1961), 63-97.
13. J.E.L. Peck, *An ergodic theorem for a non-commutative semi-group of linear operators*, Proc. Amer. Math. Soc., **2** (1951), 414-421.
14. C. Ryll-Nardzewski, *Generalized random ergodic theorems and weakly almost periodic functions* (to appear in Studia Math. and elsewhere).
15. B. Yood, *On fixed points for semi-groups of linear operators*, Proc. Amer. Math. Soc., **2** (1951), 225-233.

# ABELIAN SUBGROUPS OF $p$-GROUPS

CHARLES HOBBY

Let $G$ be a finite $p$-group where $p$ is an odd prime. We say that $G$ has *property* $A_n$ if every abelian normal subgroup of $G$ can be generated by $n$ elements. Further, if $G_n$ denotes the $n$th element in the descending central series of $G$, we say that $G$ has property $A_n(G_n)$ if every abelian subgroup of $G_n$ which is normal in $G$ can be generated by $n$ elements. If $G$ has property $A_1$, then $G$ is cyclic. N. Blackburn [1] found all of the groups which have property $A_2$. It follows from the work of Blackburn that if $G$ has property $A_2$ then the derived group of $G$ is abelian and every subgroup of $G$ has property $A_2$. We shall show that if $G$ has property $A_3$ then every subgroup of $G$ has property $A_3$. There exist groups which have property $A_3$ in which the derived series is arbitrarily long [2] so no analogue of Blackburn's result on the derived group is possible. We next consider groups $G$ which have property $A_n(G_n)$ and show that $G_n$ can be generated by $n$ elements. This leads to the existence of a bound on the derived length of $G$ which depends only on $n$ and the exponent of $G_n$.

We shall use the following notation: $p$ is an odd prime; $G=G_1\supset G_2\supset\cdots$ is the descending central series of $G$; $Z(G)=Z_1(G)\subset Z_2(G)\subset\cdots$ is the ascending central series of $G$; $G^{(k)}$ is the $k$th derived group of $G$; $(H, K)$ is the subgroup of $G$ generated by all elements $(h, k)=h^{-1}k^{-1}hk$ for $h\in H$, $k\in K$; $N\lhd G$ means $N$ is normal in $G$; $N\subset G$ means $N$ is properly contained in $G$; $C_G(N)$ is the centralizer of $N$ in $G$; $H^G$ is the normal subgroup of $G$ generated by $H$; $\mho(G)$ is the subgroup generated by $p$th powers of elements of $G$. $\Omega(G)$ is the subgroup generated by all elements of order $p$ in $G$; $\phi(G)$ is the Frattini subgroup of $G$; $|G|$ is the order of $G$.

If $A\lhd G$ and $A\subset C_G(A)$, then there is a subgroup $B$ of $C_G(A)$ such that $B\lhd G$ and $[B\colon A]=p$. It follows that if a normal subgroup $A$ of $G$ is properly contained in an abelian subgroup $C$ of $G$, then $A$ is properly contained in some abelian normal subgroup $B$ of $G$.

LEMMA 1. *Suppose $A\lhd G$ and $A\subset C$ where $C$ is an elementary abelian subgroup of $G$. Then $G$ contains an elementary abelian normal subgroup $B$ such that $A$ is a subgroup of index $p$ in $B$.*

*Proof.* Suppose $G$ is a group of minimal order for which the lemma is false. Then $C\subset G$, so there is a subgroup $M$ of index $p$ in $G$ which contains $C$. It follows that $M$ contains an elementary abelian normal subgroup $B_1$ such that $[B_1\colon A]=p$. Set $D=M\cap C_G(A)$. Then $B_1\lhd D\lhd G$.

Received May 29, 1962.

Since $(D, B_1) \subseteqq A$ and $(D, A) = 1$, we have $B_1 \subseteqq Z_2(D) \lhd G$. Therefore $B_1^G \subseteqq Z_2(D)$. But $Z_2(D)$ is a regular $p$-group for $p > 2$, so $B_1^G$ has exponent $p$. Let $B$ be a subgroup of $B_1^G$ which is normal in $G$ and which contains $A$ as a subgroup of index $p$. Clearly $B$ is elementary abelian, so the lemma is true for $G$.

THEOREM 1. *If $G$ has property $A_3$ then every subgroup of $G$ has property $A_3$.*

*Proof.* Suppose $G$ is a group of minimal order for which the theorem is false. Then $G$ contains an elementary abelian normal subgroup $A$ of order $p^3$, and there is a subgroup $M$ of index $p$ in $G$ which does not have property $A_3$. It follows that $M$ contains an elementary abelian normal subgroup $D$ of order $p^4$. Let $N$ be a subgroup of order $p^2$ in $A$ which is contained in $M$ and which is normal in $G$. If we let $C = C_G(N)$, then $[G: C] \leqq p$, hence $[D: D \cap C] \leqq p$. Thus we may suppose that $N \subset D$, since otherwise we could choose a new subgroup $D_1$ in $(C \cap D)N$ such that $N \subset D_1 \lhd M$ and $D_1$ is elementary abelian of order $p^4$.

Since $G$ has property $A_3$ it follows from Lemma 1 that $A$ contains the only elements of order $p$ in $C_G(A)$. Therefore $N = D \cap C_G(A)$. It is easy to see that $[C: C_G(A)] \leqq p^2$, thus $C = DC_G(A)$. Therefore, if $d \in D$, $g \in G$, then $g^{-1}dg = d_1c$ for some $d_1 \in D$, $c \in C_G(A)$. We recall that $D$ is an abelian normal subgroup of $M$, and that $M \lhd G$. Thus $D$ and $g^{-1}Dg$ generate a group of class at most two; hence for $p > 2$ the group generated by $D$ and $g^{-1}Dg$ has exponent $p$. Thus it follows from $g^{-1}dg = d_1c$ that $c^p = 1$, whence $c \in A$. Therefore $AD \lhd G$. But $A \cap D = N$, so $[AD: D] = p$. Since $D$ is not normal in $G$, we must have $AD = D(g^{-1}Dg)$ for some element $g \in G$. Therefore $D \cap g^{-1}Dg$ has order at least $p^3$ and is contained in $Z_1(AD)$ which is normal in $G$. Thus $AD$ must contain an element of order $p$ which centralizes $A$ and which does not belong to $A$. This is a contradiction.

THEOREM 2. *If $G$ has property $A_n(G_n)$ then $G_n$ can be generated by $n$ elements.*

*Proof.* Suppose $G$ is a group of minimal order for which the theorem is false. Then $G_n$ is not abelian, so $\phi(G_n) \neq 1$. Let $Z$ be a group of order $p$ in $Z_1(G) \cap \phi(G_n)$. Then $G_n$ and $(G/Z)_n$ have the same number of generators, so $(G/Z)_n$ must contain an elementary abelian subgroup $B/Z$ of order $p^{n+1}$ which is normal in $G/Z$. Let $B$ be the preimage of $B/Z$ in $G$. Then $B \lhd G$, $B$ has order $p^{n+2}$, and $B^{(1)} \subseteqq Z$. Thus $B$ has class at most two, hence is regular for $p > 2$. But $\mho(B) \subseteqq Z$, so $\Omega(B)$ is a group of order at least $p^{n+1}$ which is normal in $G$. Thus there is a subgroup $A$ of $\Omega(B)$ such that $A \lhd G$, $\mho(A) = 1$, and $A$ has order $p^{n+1}$. Let $N$ be a subgroup of index $p$ in $A$ which is normal in $G$. Then $|N| = p^n$ and $N \lhd G$ imply $N \subseteqq Z_n(G)$, whence $N \subseteqq Z_1(G_n)$. Therefore $A$ is abelian, a contradiction.

COROLLARY. *Suppose $G$ has property $A_n(G_n)$, where $G_n$ has exponent $p^m$. Let $k$ be an integer such that $2^k \geqq n$. Then $G^{(k+m)} = 1$.*

*Proof.* By Theorem 2, $G_n$ can be generated by $n$ elements. Therefore [3, Theorem 2] $\phi(G_n) = \Omega(G_n)$. It follows that $G_n^{(m)} = \langle 1 \rangle$ [4, Theorem 2]. In any $p$-group, $G^{(t)} \subseteqq G_{2^t}$. Therefore $G^{(k)} \subseteqq G_n$, whence $G^{(k+m)} = \langle 1 \rangle$.

## REFERENCES

1. N. Blackburn, *Generalizations of certain elementary theorems on p-groups*, Proc. London Math. Soc., (3) **11** (1961), 1-22.
2. C. Hobby, *The derived series of a finite p-group*, Illinois J. Math., **5** (1961), 228-233.
3. ———, *Generalizations of a theorem of N. Blackburn on p-groups*, Illinois J. Math., **5** (1961), 225-227.
4. ———, *A characteristic subgroup of a p-group*, Pacific J. Math., **10** (1960), 853-858.

UNIVERSITY OF WASHINGTON

# THE MINIMUM BOUNDARY FOR AN ANALYTIC POLYHEDRON

KENNETH HOFFMAN AND HUGO ROSSI

**1. Introduction.** If $K$ is a compact subset of a complex analytic manifold $M$, then for each $f$, holomorphic (analytic) in a neighborhood of $K$, the maximum modulus of $f$ over $K$ is attained on the topological boundary of $K$. If the complex dimension of $M$ is greater than 1, it may happen that there are proper closed subsets of the topological boundary on which each holomorphic $f$ attains its maximum modulus. In case there are sufficiently many holomorhic functions on $M$ to separate the points of the manifold, a general result of Šilov [6] states that there is a uniquely determined smallest closed subset of $K$ which has this maximum modulus property. This set is known as the *Šilov boundary* for the ring of functions holomorphic in a neighborhood of $K$.

The Šilov theorem is valid for separating algebras of continuous complex-valued functions on a compact space, and has nothing to do with analyticity as such. Many years earlier, the pioneering work on maximum modulus sets for rings of analytic functions had been done by Bergman [1; 2; 3]. He considered principally domains in $C^n$ which were bounded by a finite number of analytic hypersurfaces, and for these he introduced a *distinguished boundary surface*. For a wide class of such domains, he showed that his distinguished boundary was a smallest maximum modulus set. References to more recent work on these problems may be found in the second author's paper [9].

In this paper we consider the case in which $M$ is a Stein manifold, e.g., a domain of holomorphy in $C^n$, and the compact set $K$ is an *analytic polyhedron*. This means that $K$ has the form

$$K = \{m \in D; |f_j| \leqq 1, j = 1, \cdots, k\}$$

where $f_1, \cdots, f_k$ are holomorphic functions on some open subset $D$ of the manifold $M$. We consider those subsets $S$ of $K$ which have the property that, for every $f$ holomorphic in a neighborhood of $K$, the maximum modulus of $f$ over $K$ is *attained* on the subset $S$. We prove that among all such subsets $S$ there is a smallest one, which we call the *minimum boundary* for the polyhedron. The closure of this minimum boundary is (of course) the Šilov boundary for the ring of functions holomorphic on $K$. While it is difficult to give an explicit description of this Šilov boundary, such a description can be given for the minimum boundary. It is obtained by deleting from $K$ all connected local analytic

Received May 16, 1962.

varieties of positive dimension which are contained in $K$. In terms of the functions $f_j$ which define the polyhedron $K$, the description is as follows. Let $m_0 \in K$, and let $j_1, \cdots, j_r$ be those indices $j$ for which $|f_j(m_0)| = 1$. Then $m_0$ belongs to the minimum boundary if and only if $m_0$ is an isolated point of the set (variety)

$$V = \{m \in D; f_{j_i}(m) = f_{j_i}(m_0), i = 1, \cdots, r\} .$$

In [7], the first author identified the minimum boundary for the ring of functions on $K$ which are uniform limits of functions holomorphic in a neighborhood of $K$. The arguments there made essential use of a fundamental theorem of Bishop: if $A$ is a uniformly closed separating algebra of continuous complex-valued functions on a compact metric space $K(1 \in A)$, there is a smallest subset of $K$ on which each function in $A$ attains its maximum modulus. This minimum boundary for $A$ consists of the points of $K$ at which some function in $A$ "peaks", i.e., those points with the property that there is a function in $A$ which attains its maximum modulus at the point, and at no other point. The description of the minimum boundary for the ring of functions holomorphic on $K$ is exactly the one which was shown in [7] to define the minimum boundary for the uniform closure of the ring. In particular, it results that the minimum boundary for the polyhedron consists of those points of $K$ at which some holomorphic function peaks (over $K$), or, it consists of the peak points for functions which are uniformly approximable by holomorphic functions. We shall use methods from [7], and we shall make essential use of Bishop's general existence theorem. This theorem is not directly applicable to the ring of functions holomorphic in a neighborhood of $K$, since the ring is not generally closed under uniform convergence. However, by using a technique from the second author's paper [**8**], based upon the solution of the second Cousin problem for Stein manifolds, we are able to show that the ring of holomorphic functions has the same peak points as does its uniform closure.

**2. Notation and basic definitions.** A *Stein manifold* is a $d$-dimensional complex analytic manifold $M$ such that

( i ) the global holomorphic (analytic) functions on $M$ separate the points of $M$;

(ii) for each point $m \in M$ there are global holomorphic functions $h_1, \cdots, h_d$ which serve as coordinates in some neighborhood of $m$;

(iii) $M$ is a countable union of compact sets;

(iv) if $K$ is any compact set in $M$, the set of points $m \in M$ such that $|h(m)| \leqq \sup_K |h|$ for every holomorphic function $h$ on $M$ is also compact.

Let $M$ be a Stein manifold. An *analytic polyhedron* in $M$ is a subset $P$ of $M$ such that

(i) $P$ is compact

(ii) $P = \{m \in D; |f_j(m)| \leqq 1, j = 1, \cdots, k\}$, where $D$ is an open subset of $M$ and $f_1, \cdots, f_k$ are holomorphic functions on $D$.

If $P$ is an analytic polyhedron in $M$ we denote by $H(P)$ the set of all functions $f$ on $P$ such that $f$ is the restriction to $P$ of a function holomorphic in some neighborhood of $P$. We denote by $A(P)$ the class of functions on $P$ which can be approximated, uniformly on $P$, by functions in $H(P)$. Both $H(P)$ and $A(P)$ are algebras of continuous functions on $P$. Our task is to prove the existence of a smallest subset $S$ of $P$ such that, for every $h$ in $H(P)$, the maximum modulus of $h$ over $P$ is attained on the set $S$, and then to describe the set $S$ explicitly. For this we need to discuss briefly boundaries for algebras of continuous functions.

Let $X$ be a compact Hausdorff space, and let $A$ be a collection of continuous complex-valued functions on $X$. *A boundary* (*of* $X$) *for* $A$ is a subset $S$ of $X$ such that

$$\max_S |f| = \max_X |f| , \qquad f \in A$$

that is, a subset $S$ of $X$ such that for each $f$ in $A$ the maximum modulus of $f$ over $X$ is *attained* at some point of $S$. If

(a) $A$ is a complex-linear algebra, using pointwise operations

(b) the constant functions are in $A$

(c) the functions in $A$ separate the points of $X$,

then among all *closed* boundaries for $A$ there is a smallest one, i.e., the intersection of all closed boundaries for $A$ is a boundary for $A$. This smallest closed boundary for $A$ we call the *Šilov boundary* for $A$, in honor of G. E. Šilov who first proved its existence [6]. If, in addition to (a), (b), and (c) we have

(d) $A$ is closed under uniform convergence

(e) $X$ is metrizable

then the intersection of *all* boundaries for $A$ is a boundary for $A$. This smallest of all boundaries we shall call the *minimum boundary* for $A$. Its existence was proved by Bishop [4], who also showed that it consists of those points $x \in X$ which are *peak points* for $A$. We call $x$ a peak point for $A$ if there exists an $f$ in $A$ such that $|f(x)| > |f(y)|$ for all points $y$ in $X$ which are different from $x$. Evidently, the Šilov boundary for $A$ is the closure of the minimum boundary for $A$.

Both $H(P)$ and $A(P)$ are algebras of continuous functions on the compact space $P$ which satisfy conditions (a), (b), (c) above. Since $A(P)$ is the uniform closure of $H(P)$, these algebras have the same Šilov boundary. Now $P$ is metrizable, as is easy to see from the countability condition imposed on a Stein manifold. Therefore, there exists a minimum boundary for the uniformly closed algebra $A(P)$. The general function algebra results cited above do not guarantee the existence of a minimum boundary for $H(P)$; thus, as we proceed now to prove the existence of such a boundary, we shall make heavy use of explicit properties of analytic polyhedra.

**3. The minimum boundary.** Now suppose we are given the analytic polyhedron

$$P = \{m \in D; |f_j(m)| \leqq 1, j = 1, \cdots, k\}$$

in the Stein manifold $M$. With each point $m_0$ in the polyhedron $P$ we associate an analytic variety $V_{m_0}$ in the ambient neighborhood $D$ by

$$V_{m_0} = \{m \in D; f_{j_i}(m) = f_{j_i}(m_0), i = 1, \cdots, r\} \tag{3.1}$$

where $j_1, \cdots, j_r$ are those indices $j$ such that $|f_j(m_0)| = 1$.

THEOREM 1. *Let $m_0 \in P$ and suppose $m_0$ is a local peak point for the algebra $A(P)$. Then $m_0$ is an isolated point of the variety $V_{m_0}$ (3.1).*

*Proof.* This is proved in [7; Theorem 4.1]; here, we merely outline the proof. By the statement that $m_0$ is a local peak point for the algebra $A(P)$ we mean that there is a function $h \in A(P)$ and a neighborhood $N$ of $m_0$ such that $f(m_0) = 1$ and $|f(m)| < 1$ for all other points in $N \cap P$. Given such an $f$ and $N$, we may assume that those functions $f_j$ (occurring in the definition of $P$) which are of modulus less than 1 at $m_0$ are of modulus less than 1 on the open set $N$. Then $N \cap V_{m_0} = V$ is an analytic variety in $N$ and this variety is contained in the polyhedron $P$. Since $f$ is a uniform limit on $P$ of functions holomorphic in a neighborhood of $P$, $f$ is 'analytic' on the variety $V$. Also, $f$ has a local maximum over $V$ at the point $m_0$. The maximum modulus principle for analytic varieties then states that $m_0$ is an isolated point of $V$.

THEOREM 2. *Let $m_0 \in P$ and suppose that $m_0$ is an isolated point of the variety $V_{m_0}$ (3.1). Then $m_0$ is a peak point for the algebra $H(P)$.*

*Proof.* Let $g_i = \frac{1}{2}(1 + \overline{f_{j_i}(m_0)}f_{j_i}), i = 1, \cdots, r$. Then the function $g_i$ is bounded by 1 on $P$, and has the value 1 at any point of $P$ where it is of modulus 1. Now let $h = g_1 \cdots g_r$. Then $h$ is bounded by 1 on $P$, has the value 1 at any point of $P$ where it is of modulus 1, and, furthermore, the set of points in $P$ where $h$ has the value 1 is precisely the intersection of $P$ with the analytic variety $V_{m_0}$. Thus, $m_0$ is an isolated point of the set on which $h = 1$. We should also remark that $h$ is holomorphic in the open set $D$ which occurs in the definition of $P$.

Since $M$ is a Stein manifold, we can find functions $h_1, \cdots h_n$, holomorphic on all of $M$, such that the map $m \to (h_1(m), \cdots, h_n(m))$ is biholomorphic on $P$, and the image of $P$ under this map is a polynomial convex subset of $C^n$. See [5]. Now we consider the map

$$\phi(m) = (h_1(m), \cdots, h_n(m), h(m))$$

from $D$ into $C^{n+1}$. This map is biholomorphic, and the set $K = \phi(P)$ is polynomial convex. (polynomial convexity of $K$ means that if $z$ is a point of $C^{n+1}$ which is not in $K$, there exists a polynomial $p$ in $(n+1)$ variables such that $|p(z)| > \sup_K |p|$.) Let $z^0 = \phi(m_0)$. The coordinate function $z_{n+1}$ is bounded by 1 on $K$, is equal to one at any point of $K$ where it is of modulus 1, and $z^0$ is an isolated point of $K \cap \{z_{n+1} = 1\}$.

Choose a neighborhood $U$ of the point $z^0$ such that for any point $z = (z_1, \cdots, z_{n+1})$ in $U \cap K$ which is different from $z^0$ we have $|z_{n+1}| < 1$. By [8; Theorem 2.4] there exists a function $g$, holomorphic in a neighborhood $W$ of $K$, such that $g$ never vanishes on $W - U$ and $g/(1 - z_{n+1})$ is holomorphic and without zeros on $W \cap U$.

On the neighborhood $W \cap U$, the function $g$ has the form $g = (z_{n+1} - 1)k$, where $k$ is holomorhpic and never vanishes on $W \cap U$. In particular $k(z^0) \neq 0$. Thus, by shrinking $U$, we may assume that $k$ has a single-valued logarithm; say $k = e^l$ where $l$ is holomorphic on $W \cap U$. Since $K$ is polynomial convex, it is an intersection of polynomial convex open sets. Therefore, it may be assumed that $W$ is polynomial convex, and hence is itself a Stein manifold. Similarly, a small contraction of $W$ will assure that the part of the intersection of $W$ with the hyperplane $\{z_{n+1} = 1\}$ which lies in $U$ is a closed analytic variety in $W$. Since $W$ is a Stein manifold, there is a holomorphic function $p$ on $W$ such that $p = l$ on that variety [5]. Now let $\tilde{g} = ge^{-p}$. Then $\tilde{g}$ is holomorphic on $W$ and has no zeros on $W - U$. Also $\tilde{g} = (z_{n+1} - 1)ke^{-p}$ on $W \cap U$. Now $ke^{-p} = ke^{-l} = 1$ on that part of the hyperplane $\{z_{n+1} = 1\}$ which lies in $W \cap U$. Thus, on $W \cap U$, $ke^{-p} = 1 + (z_{n+1} - 1)\tilde{k}$. Finally we have a function $\tilde{g}$, holomorphic on $W$, which has no zeros on $W - U$ and has the form

$$\tilde{g} = (z_{n+1} - 1) + (z_{n+1} - 1)^2\tilde{k}$$

on $W \cap U$ ($\tilde{k}$ holomorphic on $W \cap U$).

From $\tilde{g}$ we shall now construct a function holomorphic in a neighborhood of $K$, which has the property that its maximum modulus over $K$ is attained at $z^0$ and at no other point of $K$. This function will be an analytic function of $\tilde{g}$. Thus we shall examine the range of $\tilde{g}$ on $K$. The crucial fact about the set $\tilde{g}(K)$ is that it lies outside a simply connected domain in the plane which has an analytic boundary containing the origin. To see this, we argue as follows. Choose a neighborhood $N$ of $z^0$ such that the function $\tilde{k}$ is bounded on $W \cap N$. We shall then have

$$|\tilde{g} - (z_{n+1} - 1)| \leqq c\,|z_{n+1} - 1|^2 \quad \text{on} \quad W \cap N$$

where $c$ is some positive constant. The range of $(z_{n+1} - 1)$ on $K$ lies in the left half-plane. For each point $w$ in the left halfplane, we consider the disc

$$|\zeta - w| \leqq c\,|w|^2 .$$

It is easy to see that there is an analytic curve $\gamma$ through the origin such that for all points $w$ near the origin this disc lies to the left of $\gamma$. Because, a short computation shows that the envelope of the family of discs is an analytic curve through the origin. We may assume that $N$ is sufficiently small that for any point $z$ in $K \cap N$ the point $w = z_{n+1} - 1$ has this property. Therefore, the range of $\tilde{g}$ on $N \cap K$ lies to the left of $\gamma$. On $K - N$ the function $\tilde{g}$ has no zeros. Choose $\varepsilon > 0$ such that $|\tilde{g}| > \varepsilon$ on $K - N$. If $\varepsilon$ is sufficiently small, the circle $|w| = \varepsilon$ will intersect $\gamma$ in precisely two points. Let $D$ be the domain bounded by $\gamma$ and $|w| = \varepsilon$. Let $\tau$ be the Rieman map of the complement of $D$ onto the unit disc, which carries the origin onto 1. Since $\gamma$ is an analytic curve, $\tau$ will extend analytically across that part of $\gamma$ which is on the boundary of $D$. The composition $\tau \circ \tilde{g} = F$ is then holomorphic on a neighborhood of $K$, is of modulus less than 1 on $K - \{z^0\}$, and $F(z^0) = 1$.

Now we return to the polyhedron $P$ which we mapped holomorphically onto $K$ via the map $\phi$. If we let $f = F \circ \phi$, then $f$ is holomorphic in a neighborhood of $P$, and the maximum modulus of $f$ over $P$ is attained at $m_0$ and at no other point of $P$. Thus $m_0$ is a peak point for the algebra $H(P)$.

COROLLARY. *Let $M$ be a Stein manifold, and Let $P$ be an analytic polyhedron in $M$. Then there is a (unique) smallest subset $S$ of $P$ such that for every function $f$, holomorphic in a neighborhood of $P$, the maximum modulus of $f$ over $P$ is attained on the set $S$. A necessary and sufficient condition that a point $m_0$ in $P$ should belong to this minimum boundary $S$ is any one of the following.*

(i) *$m_0$ is a peak point for the algebra $H(P)$.*

(ii) *$m_0$ is a peak point for the algebra $A(P)$.*

(iii) *$m_0$ is a local peak point for the algebra $H(P)$.*

(iv) *$m_0$ is a local peak point for the algebra $A(P)$.*

(v) *There is no connected local analytic variety of positive dimension which passes through $m_0$ and is contained in $P$.*

(vi) *$m_0$ is an isolated point of the variety $V_{m_0}$, defined by* (3.1).

*Proof.* By Bishop's theorem [4], there is a minimum boundary for the algebra $A(P)$, and it consists of those points of $P$ which are peak point for $A(P)$. In Theorem 1 we showed that any local peak point for the algebra $A(P)$ satisfies (vi). Indeed, the proof showed that the point satisfies (v), which clearly implies (vi). Theorem 2 states that (vi) implies (i). From this it is clear that the six statements about $m_0$ are equivalent. Furthermore, it is evident that the minimum boundary for $A(P)$ is a boundary for $H(P)$; and, since each point of this boundary is a peak point for $H(P)$, this boundary is the smallest boundary for $H(P)$.

## BIBLIOGRAPHY

1. S. Bergman, *Über ausgezeichnete Randfläche in der Theorie der Funktionen von zwei komplexen Veränderlichen*, Math. Ann., **104** (1931), 611.

2. ———, *Über die Veranschanlichung der Kreiskörper und Bereiche mit ausgezeichneter Randfläche*, Jahresber. d. d. Math. Ver., **42** (1933), 238.

3. ———, *Über eine in gewissen Bereichen mit Maximumfläche gültige Integraldarstellung der Functionen zweier komplexer Variabler*, Math. Z., **39** (1934), 76-94, 605-608.

4. E. Bishop, *A minimal boundary for function algebras*, Pacific J. Math., **9** (1959), 629.

5. *Séminaires de H. Cartan*, École Normale Supérieure, Paris, 1951-52.

6. Gelfand, I., Raikov, D., and Šilov, G., *Commutative Normed Rings*, A.M.S. Translations Series 2, vol. 5, 1957.

7. K. Hoffman, *Minimal boundaries for analytic polyhedra*, Rend. del Circ. Mat. di Palermo, ser. 2, **9** (1960), 147.

8. H. Rossi, *The local maximum modulus principle*, Ann. of Math., **71** (1960), 1.

9. ———, *Holomorphically convex sets in several complex variables*, Ann. of Math., **74** (1961), 470.

# THE BERGMAN KERNEL FUNCTION FOR TUBES OVER CONVEX CONES

ADAM KORÁNYI

In this article we determine the Bergman kernel function of the tube domain over an arbitrary convex cone not containing any entire straight line. For homogeneous self-dual cones this problem was solved by O. S. Rothaus ([3], Theorem 2.6). It turns out that his method can also be used in our considerably more general case. In fact, the proofs of our Theorems 1 and 2 follow closely the corresponding proofs of Rothaus; it is only in Lemma 2 that the proof of Rothaus has to be replaced by an essentially different convexity argument.

Let $V$ be an $n$-dimensional real vector space. A set $D \subset V$ is called a cone if $x \in D$ and $\lambda > 0$ imply $\lambda x \in D$. Let $V^*$ be the dual space of $V$. The dual cone $D^*$ of $D$ is defined as the set of all $\alpha \in V^*$ such that $\langle \alpha, x \rangle > 0$ for all $x \in \bar{D}$, $x \neq 0$. We call the cone $D$ regular if it is

(i) open,

(ii) convex,

(iii) nonempty, and

(iv) contains no entire straight line, i.e. $x \in D$ implies $-x \notin D$. It is easy to see that if $D$ is regular then $D^*$ is regular too, and $D^{**} = D$.

We assume that a Euclidean norm $x \to |x|$ is defined on $V$. The dual norm on $V^*$ will likewise be denoted by $\alpha \to |\alpha|$.

LEMMA 1. *If $D$ is a regular cone and $K \subset D$ is a compact set then there exists a number $\rho > 0$ such that $\langle \alpha, x \rangle \geqq \rho |\alpha|$ for all $x \in K$, $\alpha \in \bar{D}^*$.*

*Proof.* The proof is the same as that of [2] Lemma 1. By homogeneity it suffices to prove the assertion for $|\alpha| = 1$. Let $S = \{\alpha \in V^* \mid |\alpha| = 1\}$ be the unit sphere in $V^*$. Now $\langle \alpha, x \rangle$ is a positive continuous function on the compact set $(S \cap \bar{D}^*) \times K$ and thus has a positive minimum $\rho$, finishing the proof.

We define the positive real-valued function $M$ on $D^*$ by

$$M(\alpha) = \int_D e^{-\langle \alpha, x \rangle} dx$$

for all $\alpha \in D^*$. By Lemma 1 the integral converges uniformly on compact sets. As it can immediately be seen, $M$ is a homogeneous function of degree $-n$.

LEMMA 2. *Let $D$ be a regular cone and let $\beta \in \partial D^*$ (the boundary of $D^*$ in $V^*$). Then*

Received May 4, 1962.

$$\lim_{\alpha \to \beta} M(\alpha) = \infty .$$

*Proof.* If $\beta = 0$ the assertion is trivial. Let $\beta \neq 0$. For $\alpha \in D^*$ and $t > 0$ define $H_\alpha(t) = \{x \in D \,|\, \langle \alpha, x \rangle = t\}$ and let

$$V_\alpha(t) = \int_{H_\alpha(t)} dv_\alpha$$

be the volume of $H_\alpha(t)$ ($dv_\alpha$ denotes the volume element of the hyperplane $\{x \,|\, \langle \alpha, x \rangle = t\}$). Clearly we have $V_\alpha(t) = t^{n-1} V_\alpha(1)$ for all $t > 0$. Also

$$\begin{aligned} M(\alpha) &= \int_D e^{-\langle \alpha, x \rangle} dx = \int_0^\infty dt \int_{H_\alpha(t)} e^{-\langle \alpha, x \rangle} dv_\alpha \\ &= \int_0^\infty V_\alpha(t) e^{-t} dt = V_\alpha(1) \Gamma(n) . \end{aligned}$$

Therefore the Lemma will be proved if we show that $\lim_{\alpha \to \beta} V_\alpha(1) = \infty$.

Let $U \subset \bar{D}^*$ be a compact neighborhood of $\beta$ relative to $\bar{D}^*$. Then the set $L$ of all $x \in D$ such that $\langle \alpha, x \rangle < 1$ for all $\alpha \in U$ has an interior. (In fact, if $A$ is a bound for $|\alpha|$ on $U$, it is easy to see that $L$ contains all $x \in D$ such that $|x| < A^{-1}$). Let $K$ be an open sphere contained in $L$; let $c \in D$ be its center and $r > 0$ its radius.

For $\alpha \in U$ let $K_\alpha$ be the $(n-1)$-dimensional sphere of radius $r$ and center $c_\alpha = \langle \alpha, c \rangle^{-1} c$ contained in the hyperplane $\{x \,|\, \langle \alpha, x \rangle = 1\}$. By convexity and by $\langle \alpha, c \rangle^{-1} > 1$ we have $K_\alpha \subset H_\alpha(1)$. Since $|c_\alpha| = \langle \alpha, c \rangle^{-1} |c|$ and since the continuous function $\langle \alpha, c \rangle^{-1}$ is bounded on the compact set $U$, there exists a number $R$ such that

$$(1) \qquad |c_\alpha| \leqq R$$

for all $\alpha \in U$.

Now let $\Omega > R$ be an arbitrarily large number. There exists an element $a \in \bar{D}$, $|a| = 1$ such that $\langle \beta, a \rangle = 0$, for otherwise we would have $\beta \in D^*$. Hence there exists an element $x \in D$, $|x| = 1$ such that $\langle \beta, x \rangle < (R + \Omega)^{-1}$. It follows then that there exists a neighborhood $U(\Omega) \subset U$ of $\beta$ relative to $D^*$ such that $\langle \alpha, x \rangle < (R + \Omega)^{-1}$ for all $\alpha \in U(\Omega)$. Let $x_\alpha = \langle \alpha, x \rangle^{-1} x$. Clearly we have $x_\alpha \in H_\alpha(1)$ and

$$(2) \qquad |x_\alpha| > R + \Omega$$

for all $\alpha \in U(\Omega)$. Now $H_\alpha(1)$ is convex, and thus contains the convex hull $B_\alpha$ of $K_\alpha$ and $x_\alpha$; hence, be (1) and (2),

$$V_\alpha(1) \geqq \int_{B_\alpha} dv_\alpha > \frac{C}{n-2} \Omega$$

for all $\alpha \in U(\Omega)$, $C$ denoting the volume of the $(n-2)$-dimensional sphere of radius $r$. This completes the proof.

Let $V_C = V \oplus iV$ be the complexification of $V$. The tube over $D$ in $V_C$ is the domain $T_D = \{x + iy \mid x \in D, y \in V\}$. For $z = x + iy \in V_C$ and $\alpha \in V^*$ we write $\langle \alpha, z\rangle = \langle \alpha, x\rangle + i\langle \alpha, y\rangle$. We denote by $\mathscr{L}^2(T_D)$ the Hilbert space of holomorphic functions on $T_D$, square integrable with respect to $dxdy$, and by $L^2_M(D^*)$ the Hilbert space of functions on $D^*$ square integrable with respect to $M(\alpha)d\alpha$.

THEOREM 1. *The mapping $\varphi \to f$ defined by*

$$f(z) = \pi^{-n/2}\int_{D^*}\varphi(\alpha)e^{-\langle \alpha, z\rangle}d\alpha \tag{3}$$

*is an isomorphism of $L^2_M(D^*)$ onto $\mathscr{L}^2(T_D)$.*

*Proof.* Let $\varphi \in L^2_M(D^*)$. Then

$$\int_{D^*}|\varphi(\alpha)e^{-\langle \alpha, z\rangle}|\,d\alpha = \int_{D^*}|\varphi(\alpha)|\,e^{-\langle \alpha, x\rangle}d\alpha$$
$$\leqq \left(\int_{D^*}|\varphi(\alpha)|^2\,M(\alpha)d\alpha\right)^{1/2}\left(\int_{D^*}e^{-2\langle \alpha, x\rangle}M(\alpha)^{-1}d\alpha\right)^{1/2}$$

by the Schwarz inequality. The first integral is just $\|\varphi\|^2$, the second is also convergent by Lemma 2 and by the homogeneity of $M$; by Lemma 1 it is even bounded on compact subsets of $D$. Thus (3) converges absolutely and uniformly on compact subsets of $T_D$, and hence represents a holomorphic function. Furthermore, reversing the order of integration (which is possible since the integrand is positive and measurable), and then applying the Plancherel theorem we have

$$\begin{aligned}\|\varphi\|^2 &= \int_{D^*}|\varphi(\alpha)|^2M(\alpha)d\alpha = \int_{D^*}|\varphi(\alpha)|^2d\alpha\int_D e^{-\langle \alpha, x\rangle}dx \\ &= 2^n\int_{D^*}|\varphi(\alpha)|^2d\alpha\int_D e^{-2\langle \alpha, x\rangle}dx = 2^n\int dx\int_{D^*}|\varphi(\alpha)e^{-\langle \alpha, x\rangle}|^2\,d\alpha \\ &= \int_D dx\int_V|f(x+iy)|^2dy = \|f\|^2\,,\end{aligned} \tag{4}$$

which shows that $f \in \mathscr{L}^2(T_D)$ and also that the mapping is an isomorphism.

Remains to show (and this is the more important part) that the isomorphism is onto.

First we prove that there exists a measurable function $\varphi$ on $V^*$ such that

$$f(z) = f(x+iy) = \lim \pi^{-n/2}\int_{V^*}\varphi(\alpha)e^{-\langle \alpha, z\rangle}d\alpha$$

for almost all $x \in D$. In fact, by Fubini's theorem $f(x+iy)$ as a function of $y$ is in $L^2(V)$ for almost all $x$; so the Fourier transform

$$\psi(x, \alpha) = \lim \pi^{-n/2}\int_V f(x + iy)e^{-i\langle\alpha, y\rangle}dy$$

exists. The assertion is that $\psi(x, \alpha) = \varphi(\alpha)e^{-\langle\alpha, x\rangle}$ with some measurable $\varphi$. Let $N \subset D$ be a subset whose distance from $\partial D$ is $d > 0$. Then, by a well-known property of $\mathscr{L}^2$-spaces, $|f(z)| = |f(x + iy)| \leqq C_d\|f\|$ for all $x \in N$, $f \in \mathscr{L}^2(T_D)$. Using this remark the proof of our assertion is the same as that of a similar assertion in [1], p. 128, and will not be reproduced here.

Next we show that $\varphi(\alpha) = 0$ for almost all $\alpha \notin D^*$. In fact, using the Plancherel theorem and reversing the order of integration we obtain

$$\|f\|^2 = 2^n\int_{V^*} d\alpha\int_D |\varphi(\alpha)|^2 e^{-2\langle\alpha, x\rangle}dx .$$

In particular, $\int_D |\varphi(\alpha)|^2 e^{-2\langle\alpha, x\rangle}dx$ exists for almost all $\alpha$ and is integrable. Now if $\alpha \notin D^*$, then $\langle\alpha, x\rangle < 0$ for some $x \in D$ and hence $\int_D e^{-2\langle\alpha, x\rangle}dx$ diverges. Therefore $\varphi(\alpha) = 0$ for almost all such $\alpha$.

Finally we must show that $\varphi \in L^2_M(D^*)$. This however follows at once from the Plancherel theorem through the equalities (4).

THEOREM 2. *The Bergman kernel function of $T_D$ is*

$$K(z, w) = \frac{1}{\pi^n}\int_{D^*} e^{-\langle\alpha, z+\bar{w}\rangle}M(\alpha)^{-1}d\alpha .$$

*Proof.* From Theorem 1 it is clear that, for fixed $w \in T_D$, $K(z, w)$ as a function of $z$ is in $\mathscr{L}^2(T_D)$. Also for fixed $w \in T_D$ and $x \in D$, $K(z, w)$ is in $L^2(V)$ as a function of $y$.

Let $f \in \mathscr{L}^2(T_D)$, then $f$ can be represented in the form (3). Using the Plancherel theorem and then reversing the order of integration (which can be done since the integrand is measurable and the repeated integral in reverse order exists absolutely), we obtain

$$\begin{aligned}\int_{T_D} f(z)\bar{K}(z, w)\,dxdy &= \int_D dx\int_V f(z)\bar{K}(z, w)dy \\ &= 2^n\int_D dx\int_{V^*}\varphi(\alpha)e^{-\langle\alpha, x\rangle}e^{-\langle\alpha, x+w\rangle}M(\alpha)^{-1}d\alpha \\ &= 2^n\int_{V^*} d\alpha\varphi(\alpha)e^{-\langle\alpha, w\rangle}M(\alpha)^{-1}\int_D e^{-2\langle\alpha, x\rangle}dx \\ &= \int_{V^*}\varphi(\alpha)e^{-\langle\alpha, w\rangle}d\alpha = f(w)\end{aligned}$$

for all $w \in T_D$. Owing to the fact that the Bergman kernel is uniquely determined by its reproducing property, the proof is finished.

## REFERENCES

1. S. Bochner and W. T. Martin, *Several Complex Variables,* Princeton University Press, 1948.
2. M. Koecher, *Positivitätsbereiche im $R^n$,* Amer. J. Math., **79** (1957), 575-596.
3. O. S. Rothaus, *Domains of Positivity,* Abh. Math. Semin. Hamburg **24** (1960), 189-235.

UNIVERSITY OF CALIFORNIA, BERKELEY

# THE TIME-DOMAIN ANALYSIS OF A CONTINUOUS PARAMETER WEAKLY STATIONARY STOCHASTIC PROCESS

P. MASANI AND J. ROBERTSON

**1. Introduction.** In this paper we shall give a new, spectral-free, method to obtain the differential innovations and the Wold decomposition of a univariate, continuous parameter, weakly stationary[1], mean-continuous, non-deterministic stochastic process $(f_t, \ -\infty < t < \infty)$. We shall affect a transition from the continuous to the discrete parameter case by systematic use of the infinitesimal generator $iH$ of the shift group $(U_t, \ -\infty < t < \infty)$ of the process, and of the Cayley transform $V$ of the self-adjoint operator $H$(§ 2). Our analysis will be purely in the time-domain.

With the $f_t$-process we shall associate the discrete parameter process $(f'_n)_{n=-\infty}^{\infty}$, where $f'_n = V^n(f_0)$. Since $V$ is unitary, the $f'_n$-process is weakly stationary. Letting $\mathscr{M}_t$, $\mathscr{M}'_n$ be the past and present subspaces of the $f_t$- and $f'_n$-processes, respectively, and $\mathscr{M}_{-\infty}$, $\mathscr{M}'_{-\infty}$ be their remote pasts, we shall show that $\mathscr{M}_0 = \mathscr{M}'_0$ and $\mathscr{M}_{-\infty} = \mathscr{M}'_{-\infty}$ (§ 4). In the non-deterministic case we shall show that the subspace $\mathscr{N}_t = \mathscr{M}^{\perp}_{-\infty} \cap \mathscr{M}_t$ is the past and present of the process $(h_t, -\infty < t < \infty)$, where $h_t = U_t(h'_0)$, $h'_0$ being the 0th normalized innovation of the discrete $f'_n$-process (§ 5). We shall then show (§ 6) that the $h_t$-process is weakly Markovian[1] with covariance $e^{-|t|}$ for lag $t$, and that if

$$\xi_t = T_t(h'_0), \qquad \text{where} \quad T_t = \frac{1}{\sqrt{2}}\left\{U_t - I + \int_0^t U_s ds\right\}, \tag{1.1}$$

the process $(\xi_t, \ -\infty < t < \infty)$ has stationary, orthogonal increments such that $|\xi_b - \xi_a|^2 = |b - a|$. These increments are the "differential innovations" of our $f_t$-process; for we shall show (6.6) that the set of all convergent stochastic integrals $\int_{-\infty}^{t} c(s)d\xi_s$, $c \in L_2(-\infty, t)$, is identical with the subspace $\mathscr{N}_t$ mentioned above. Since

$$\mathscr{M}_t = \mathscr{N}_t + \mathscr{M}_{-\infty}, \quad \mathscr{N}_t \perp \mathscr{M}_{-\infty},$$

it follows at once that $f_t = u_t + v_t$, where the $u_t$ form a one-sided moving

Received March, 21, 1962. Preliminary attempts in the direction of the research here presented were made by the first-named author some years ago in collaboration with Mr. P. D. Kapadia of Bombay, India, whose help we gratefully acknowledge. Much of the work was done recently at Indiana University, that of the first named author being supported in part by the Office of Naval Research.

[1] In this paper the term "weakly" has the same meaning as Doob's expression "in the wide sense" [5, p. 95].

average process, and the $v_t$ a deterministic one:

$$u_t = \int_0^\infty c(s) d_s \xi_{t-s}, \quad v_t = \text{projection of } f_t \text{ on } \mathscr{M}_{-\infty} .$$

We thus get the Wold decomposition cf. 6.7 below.

In justification of this new approach we may mention its simplicity and coherence. With the time-domain analysis so completed, one can develop the spectral theory in an equally coherent way. One can also deal conveniently with the extension to vector-valued processes. In comparison, an approach in which spectral considerations are brought to bear on time-domain questions or vice versa seems cumbersome and roundabout. But quite apart from this, our approach is essentially more general than one based on the spectral resolution of the group $(U_t, -\infty < t < \infty)$ and is more suggestive of further research, although it does not yield any really new results on univariate stationary processes. As prediction theory has advanced, its connection with the theory of shift-invariant subspaces of the Hardy class $H_2$ initiated by Beurling [2] has been noticed; see especially Helson and Lowdenslager [10] and Lax [14]. Recently Halmos [8] has brought to light a result, which shows that underlying both theories is a semi-group of isometries on a Hilbert space (cf. also [15]). (In the case under discussion, this semi-group comprises the (isometric) restrictions of the unitary operators $U_t^*$ to the subspace $M_0$.) One of us [16] has found that our approach based on use of the infinitesimal generator $iH$ and of the operator $T_t$ defined in (1.1) extends to general continuous parameter semi-groups of isometries to yield valuable results concerning their structure. But since in general these isometries will be non-normal, the generator $H$ will not be self-adjoint and the usual spectral considerations will fail; cf. Cooper [3]. Thus *it seems worthwhile to try to dispense with spectral tools in the analysis of time-domain problems.*

Hanner [9] was the first to make a purely time-domain analysis in the continuous parameter case. By an ingenious construction he proved the existence of differential innovations and derived the Wold decomposition. His approach, somewhat *ad hoc* in nature, has not been pursued in the literature, and its points of contact with the earlier work of Cooper [3] have gone unnoticed. Our approach differs from that of Hanner and Cooper in the transition we make to the discrete parameter case by means of the infinitesimal generator and the Cayley transform.

It is reasonably clear that our approach will work in the case of processes for which the differential innovations can be had by Hanner's method. As an instance we cite the study of continuous parameter random distributions due to K. Ito, Gelfand, and Balagangadharan [12, 7, 1]. It is also possible that our ideas may apply to some of the non-stationary processes studied recently by Cramer [4, 4′, 4″].

**2. The infinitesimal generator and Cayley transform.** Let $(U_t, -\infty < t < \infty)$ be a *strongly continuous group of unitary operators acting on a complex Hilbert space* $\mathfrak{X}$; i.e. let

$$(2.1)\quad \begin{cases} \text{(a)} & U_t \text{ be a unitary operator on } \mathfrak{X} \text{ onto } \mathfrak{X},\ -\infty < t < \infty. \\ \text{(b)} & U_s U_t = U_{s+t} = U_t U_s,\ -\infty < s, t < \infty. \\ \text{(c)} & U_{t+h} \to U_t (\text{strongly})^2 \text{ on } \mathfrak{X} \text{ as } h \to 0,\ -\infty < t < \infty. \end{cases}$$

It is known [17, p. 385] that the group has an *infinitesimal generator*

$$(2.2)\qquad iH = \lim_{h\to 0} \frac{1}{h}\{U_t - I\} \qquad \text{on } \mathscr{D},$$

where $H$ is a self-adjoint operator with domain $\mathscr{D}$, and $\mathscr{D}$ is a linear manifold everywhere dense in $\mathfrak{X}$. Also, cf. [19, p. 142 and 6, p. 622]

$$(2.3)\quad \begin{cases} \text{(a)} & H + iI \text{ is one-to-one on } \mathscr{D} \text{ onto } \mathfrak{X}, \\ \text{(b)} & (H+iI)^{-1} = \dfrac{1}{i}\displaystyle\int_0^\infty e^{-t} U_t dt \text{ is bounded and one-to-one} \\ & \text{on } \mathfrak{X} \text{ onto } \mathscr{D}, \text{ and } |(H+iI)^{-1}|_B \leqq 1^{(3)}. \end{cases}$$

Now let $V$ be the *Cayley transform* of $H$:

$$(2.4)\qquad V = c(H) = (H - iI)(H + iI)^{-1} \text{ on } \mathfrak{X}.$$

Then [19, p. 304]

$$(2.5)\quad \begin{cases} \text{(a)} & V \text{ is unitary on } \mathfrak{X} \text{ onto } \mathfrak{X}, \\ \text{(b)} & I - V = 2i(H+iI)^{-1} \text{ is one-to-one on } \mathfrak{X} \text{ onto } \mathscr{D}, \\ \text{(c)} & H = i(I+V)(I-V)^{-1} \text{ on } \mathscr{D}, \\ \text{(d)} & U_t V^n = V^n U_t \text{ on } \mathfrak{X},\ -\infty < n, t < \infty,\ n = \text{integer}. \end{cases}$$

In this section we shall establish the relationship between $U_t$ and $V^n$ for arbitrary $t$ and $n$ on which will hinge the subsequent development.

The $U_t$ are expressible in terms of $H$ by the Hille-Yosida exponential formula, cf. [17, p. 403],

$$(2.6)\quad \begin{cases} U_t = \lim\limits_{n\to\infty} \exp(tiHJ_n),\ (\text{strong})^2, & t \geqq 0 \\ J_n = \left(I - \dfrac{1}{h} iH\right)^{-1}. \end{cases}$$

One sees trivially that $J_n$ is a bounded operator and that so therefore is $iHJ_n = n(J_n - I)$. Hence the term $\exp(tiHJ_n)$ in (2.6) is definable

---

[2] It is to be understood in the sequel that all operator-limits are in the strong sense.

[3] $|T|_B$ refers to the Banach norm of the operator $T$.

by the usual power-series. We now assert two lemmas:

2.7 LEMMA. (*Expression of $U_t$ in terms of $V^k$*).

$$U_{\pm t} = e^{-t}I + \lim_{n\to\infty} \sum_{k=1}^{\infty} \frac{1}{k!}\left(\frac{-nt}{n+1}\right)^k \{(I + A_{\pm n})^k - I\} , \qquad t \geqq 0 ,$$

*where*

$$A_{\pm n} = \frac{2n}{n+1} \sum_{j=1}^{\infty} \left(\frac{n-1}{n+1}\right)^{j-1} V^{\pm j} , \qquad n \geqq 0.$$

*Proof.* Let $t \geqq 0$. Then by (2.6)

(1)
$$U_t = \lim_{n\to\infty} \sum_{k=0}^{\infty} \frac{t^k}{k!}(iHJ_n)^k .$$

Using (2.5)(c) we can express the R.H.S. of (1) in terms of $V$:

$$\begin{aligned} iHJ_n &= -(I + V)(I - V)^{-1}\left\{I + \frac{1}{n}(I + V)(I - V)^{-1}\right\}^{-1} \\ &= -\frac{n}{n+1}(I + A_n) \end{aligned}$$

after some simplification. Thus

$$(iHJ_n)^k = \left(-\frac{n}{n+1}\right)^k I + \left(-\frac{n}{n+1}\right)^k \{(I + A_n)^k - I\} , \qquad k \geqq 0 .$$

Hence from (1)

(2)
$$U_t = \lim_{n\to\infty} \exp\left(\frac{-nt}{n+1}\right) I + \lim_{n\to\infty} \sum_{k=1}^{\infty} \frac{1}{k!}\left(\frac{-nt}{n+1}\right)^k \{(I + A_n)^{-1} - I\} .$$

Since the first term on the R.H.S. is $e^{-t}I$, we have the desired expression for $U_t$, $t \geqq 0$.

To obtain the expression for $U_{-t}$, $t \geqq 0$, we note that $U_{-t} = U_t^*$, $V^* = V^{-1}$ and so $A_n^* = A_{-n}$, $n \geqq 0$. Thus, taking adjoints on both sides of (2), we get the desired result.

2.8 LEMMA. (*Expression of $V^n$ in terms of $U^t$*).

$$V^{\pm n} = I + 2\int_0^{\infty} L_n'(2t)e^{-t}U_{\pm t}dt , \qquad n \geqq 0$$

*where*

$$L_n(t) = \sum_{k=0}^{n} \frac{(-1)^k}{k!} \binom{n}{k} t^k , \qquad n \geqq 0, \text{ (}n\text{th } \textit{Laguerre polynomial}\text{)}.$$

*Proof.* The result obviously holds for $n = 0$. We establish it for $n > 0$ by induction. For $n = 1$, the desired result reduces to the equality

$$V = I - 2\int_0^\infty e^{-t} U_t dt \ , \tag{1}$$

the correctness of which is clear from (2.5)(b) and (2.5)(b). Next, assuming the result for $n$, we find using (1) that

$$V^{n+1} = I + 2\int_0^\infty \{L'_n(2t) - 1\} e^{-t} U_t dt - 4\int_0^\infty \int_0^\infty L'_n(2t) e^{-(s+t)} U_{s+t} ds dt \ .$$

Putting $\sigma = s + t$, and using Dirichlet's formulae we find that

$$4\int_0^\infty \int_0^\infty L'_n(2t) e^{-(s+t)} U_{s+t} ds dt = 2\int_0^\infty \{L_n(2\sigma) - 1\} e^{-\sigma} U_\sigma d\sigma \ .$$

Hence

$$V^{n+1} = I + 2\int_0^\infty \{L'_n(2t) - L_n(2t)\} e^{-t} U_t dt \ .$$

Since the Laguerre polynomials satisfy the recurrence relation $L_n = L'_n - L'_{n+1}$, cf. [18, p. 299, (10)] we get

$$V^{n+1} = I + 2\int_0^\infty L'_{n+1}(2t) e^{-t} U_t dt \ ,$$

as desired. The result thus holds for all $n \geqq 1$. Its validity for $n \leqq -1$ follows on taking adjoints and noting that $V^{-n} = (V^n)^*$ and $U_{-t} = U_t^*$.

We shall denote by $\mathfrak{S}(X_\lambda)_{\lambda \in \Lambda}$ the (closed) subspace spanned by the subsets $X_\lambda$ of $\mathfrak{X}$, for $\lambda \in \Lambda$. We now assert the following lemma:

2.9 Lemma. *For any $X \subseteqq \mathfrak{X}$, we have*

(a) $$\mathfrak{S}\{V^{-n}(X)\}_{n \geqq 0} = \mathfrak{S}\{U_{-t}(X)\}_{t \geqq 0}$$

(b) $$\mathfrak{S}\{V^{n}(X)\}_{n \geqq 0} = \mathfrak{S}\{U_{t}(X)\}_{t \geqq 0} \ .$$

*Proof* (a). Lemma 2.8 asserts that for $n \geqq 0$ $V^{-n}$ is the strong limit of a linear combination of the $U_{-t}$ for $t \geqq 0$. It follows that for any $X \subseteqq \mathfrak{X}$,

$$V^{-n}(X) \subseteqq \mathfrak{S}\{U_{-t}(X)\}_{t \geqq 0} \ ,$$

whence

$$\mathfrak{S}\{V^{-n}(X)\}_{n \geqq 0} \subseteqq \mathfrak{S}\{U_{-t}(X)\}_{t \geqq 0} \ .$$

On the other hand, Lemma 2.7 asserts that for $t \geqq 0$, $U_{-t}$ is the strong limit of a linear combination of the $V^{-n}$, for $n \geqq 0$, so that

$$U_{-t}(X) \subseteqq \mathfrak{S}\{V^{-n}(X)\}_{n \geqq 0} .$$

From this follows the inclusion reverse to (1), thereby yielding (a).

(b) can be derived similarly from Lemmas 2.7, 2.8 taking $V^n$, $U_t$, instead of $V^{-n}$, $U_{-t}$, with $n, t \geqq 0$.

3. **Weakly stationary stochastic processes.** In this section we shall recall the basic notions and results on weakly stationary stochastic processes.

By a *weakly stationary stochastic process* (S.P.) is meant a function $f$ on $(-\infty, \infty)$ to a complex Hilbert space $X$ such that the inner product

$$(f_s, f_t) = \gamma_{s-t} \tag{3.1}$$

depends only on the difference $s - t$ and not on $s$ and $t$ separately. The complex-valued function $\gamma$ on $(-\infty, \infty)$ is called the *covariance function* of the S.P. It is convenient to denote the values of $f$ and $\gamma$ at $t$ by $f_t$ and $\gamma_t$ rather than by $f(t)$ and $\gamma(t)$, and to denote the S.P. itself by $(f_t, -\infty < t < \infty)$ rather than by $f$.

We shall be especially interested in the subspaces

$$\begin{cases} \mathscr{M}_t = \mathfrak{S}(f_s)_{s \leqq t}, & -\infty < t < \infty \\ \mathscr{M}_{-\infty} = \bigcap_{-\infty < t < \infty} \mathscr{M}_t , & \mathscr{M}_\infty = \mathfrak{S}(f_s)_{s < \infty} . \end{cases} \tag{3.2}$$

We shall call $\mathscr{M}_t$ *the past and present of* $f_t$, $\mathscr{M}_{-\infty}$ *the remote past of the* S.P., *and* $\mathscr{M}_\infty$ *the space spanned by the* S.P. Obviously

$$\begin{cases} \mathscr{M}_{-\infty} \subseteqq \mathscr{M}_s \subseteqq \mathscr{M}_t \subseteqq \mathscr{M}_\infty , & -\infty < s < t < \infty \\ \mathscr{M}_{-\infty} = \bigcap_{t \leqq 0} \mathscr{M}_t . \end{cases} \tag{3.3}$$

It is known, cf. Karhunen [13, p. 55], that if $(f_t, -\infty < t < \infty)$ is a weakly stationary S.P., then there exists a group of unitary operators $U_t$ on $\mathfrak{X}$, $-\infty < t < \infty$, such that

$$f_{s+t} = U_t(f_s) , \qquad -\infty < s, t < \infty . \tag{3.4}$$

The operators $U_t$ are uniquely determined on the subspace $\mathscr{M}_\infty$ but not on $X$. We shall call $(U_t, -\infty < t < \infty)$ *the shift group of the* S.P. $(f_t, -\infty < t < \infty)$. It follows easily, cf. Hanner [9, p.162], that

$$U_t(\mathscr{M}_s) = \mathscr{M}_{s+t}, U_t(\mathscr{M}_{-\infty}) = \mathscr{M}_{-\infty}, U_t(\mathscr{M}_\infty) = \mathscr{M}_\infty, -\infty < s, t < \infty. \tag{3.5}$$

We call a S.P. $(f_t, -\infty < t < \infty)$ *mean-continuous*, if the function $f$ is continuous on $(-\infty, \infty)$ with respect to the metric induced by the norm of the Hilbert space $\mathfrak{X}$. From the stationarity condition (3.1) we readily infer the following:

3.6 LEMMA. *For a weakly stationary* S.P. $(f_t, -\infty < t < \infty)$ *with covariance function* $\gamma$ *mean-continuity is equivalent to each of the conditions*:

(i) $f$ *is continuous at* 0,

(ii) $\gamma$ *is continuous at* 0,

(iii) $\gamma$ *is continuous on* $(-\infty, \infty)$,

(iv) *the shift group* $(U_t, -\infty < t < \infty)$ *is strongly continuous on* $\mathscr{M}_\infty$.

The following result is known:

3.7 LEMMA. *If the* S.P. *is mean-continuous, then*

(a) $\mathscr{M}_\infty$ *is a separable subspace of* $\mathscr{X}$,

(b) $\mathscr{M}_{t-} = \mathscr{M}_t = \mathscr{M}_{t+}$, $-\infty < t < \infty$, *where* $\mathscr{M}_{t-} = clos. \bigcup_{s<t} \mathscr{M}_s$, $\mathscr{M}_{t+} = \bigcap_{s>t} \mathscr{M}_s$.

**4. The associated discrete parameter process.** Let $(f_t, -\infty < t < \infty)$ be a weakly stationary, *mean-continuous* S.P. with shift group $(U_t, -\infty < t < \infty)$. Let $V$ be the Cayley transform of $H$, where $iH$ is the infinitesimal generator of the shift group, cf. (2.2), (2.4). Let

$$f'_n = V^n(f_0) \,. \tag{4.1}$$

Then the bisequence $(f'_n)_{-\infty}^{\infty}$ is a discrete-parameter, weakly stationary S.P. with shift operator $V$. We shall call it *the discrete* S.P. *associated with* $(f_t, -\infty < t < \infty)$.

We shall denote the past and present of $f'_n$, the remote past, and the subspace spanned by the S.P. $(f'_n)_{-\infty}^{\infty}$ by $\mathscr{M}'_n$, $\mathscr{M}'_{-\infty}$, and $\mathscr{M}'_\infty$, respectively; thus

$$\mathscr{M}'_n = \mathfrak{S}(f'_k)_{k=-\infty}^{n}, \qquad \mathscr{M}'_{-\infty} = \bigcap_{n=\infty}^{\infty} \mathscr{M}'_n \,, \qquad \mathscr{M}'_\infty = \mathfrak{S}(f'_k)_{k=-\infty}^{\infty} \,. \tag{4.2}$$

It follows that

$$\begin{cases} \mathscr{M}'_{-\infty} \subseteqq \mathscr{M}'_m \subseteqq \mathscr{M}'_n \subseteqq \mathscr{M}'_\infty, \; -\infty < m < n < \infty \\ \mathscr{M}'_{-\infty} = \bigcap\limits_{n=-\infty}^{\circ} \mathscr{M}'_n \,. \end{cases} \tag{4.3}$$

As far as we know the associated discrete parameter S.P. $(f'_n)_{-\infty}^{\infty}$ has been defined in the literature, not by (4.1), but as the process whose spectral distribution is the Cayley transform (in the complex plane) of the spectral distribution of the given continuous parameter process, cf. e.g. Doob [5, p. 583]. It can be shown that the two definitions are equivalent. But as indicated in § 1 there are advantages in adopting a purely time-domain and spectral-free definition. For instance, in the light of Lemma 2.9 we can assert the following theorem, which reveals the close relationship between the two processes. Variants of parts (a), (b) of this theorem are know, cf. e.g. Doob [5, p. 583–84]; part (c) is new as for as we know.

4.4 THEOREM. (a) $\mathscr{M}_0 = \mathscr{M}'_0$, (b) $\mathscr{M}_\infty = \mathscr{M}'_\infty$, (c) $\mathscr{M}_{-\infty} = \mathscr{M}'_{-\infty}$.

*Proof*. (a) Take $X = \{f_0\}$ in 2.9(a). We then get

$$\mathscr{M}'_0 = \mathfrak{S}(V^{-n}(f_0))_{n \geqq 0} = \mathfrak{S}(U_{-t}(f_0))_{t \geqq 0} = \mathscr{M}_0 .$$

(b) Now take $X = \{f_0\}$ in 2.9(b). We then get $\mathfrak{S}(V^n(f_0))_{n \geqq 0} = \mathfrak{S}(U_t(f_0))_{t \geqq 0}$. Hence,

$$\begin{aligned} \mathscr{M}'_\infty &= \text{clos.}\,\{\mathscr{M}'_0 + \mathfrak{S}(V^n(f_0))_{n \geqq 0}\} \\ &= \text{clos.}\,\{\mathscr{M}_0 + \mathfrak{S}(U_t(f_0))_{t \geqq 0}\} \qquad \text{(by (a))} \\ &= \mathscr{M}_\infty . \end{aligned}$$

(c) Take $X = \mathscr{M}_{-\infty}$ in 2.9(b). Then using (3.5) we get

$$V^k(\mathscr{M}_{-\infty}) \subseteqq \mathfrak{S}(V^n(\mathscr{M}_{-\infty}))_{n \geqq 0} = \mathfrak{S}(U_t(\mathscr{M}_{-\infty}))_{t \geqq 0} = \mathscr{M}_{-\infty}, \quad k \geqq 0 .$$

Applying $V^{-k}$ to both sides, and using (a),

$$\mathscr{M}_{-\infty} \subseteqq V^{-k}(\mathscr{M}_{-\infty}) \subseteqq V^{-k}(\mathscr{M}_0) = V^{-k}(\mathscr{M}'_0) = \mathscr{M}'_{-k}, \quad k \geqq 0 .$$

Hence, cf. (4.3),

$$\mathscr{M}_{-\infty} \subseteqq \bigcap_{k=0}^{\infty} \mathscr{M}'_{-k} = \mathscr{M}'_{-\infty} . \tag{1}$$

Next taking $X = \mathscr{M}'_{-\infty}$ in 2.9(b), we get

$$U_s(\mathscr{M}'_{-\infty}) \subseteqq \mathfrak{S}(U_t(\mathscr{M}'_{-\infty}))_{t \geqq 0} = \mathfrak{S}(V^n(\mathscr{M}'_{-\infty}))_{n \geqq 0} = \mathscr{M}'_{-\infty}, \quad s \geqq 0.$$

Proceeding as before, we derive the inclusion relation reverse to that in (1). Thus (c).

## 5. Non-deterministic S.P. Pre-Wold decomposition.

We shall say that a S.P. $(f_t, -\infty < t < \infty)$ is *deterministic*, if and only if $\mathscr{M}_{-\infty} = \mathscr{M}_\infty$; otherwise *non-deterministic*. From the stationarity condition (3.1) we infer the following lemma, cf. Hanner [9, p. 163]:

5.1 LEMMA. *For a weakly stationary* S.P. *the following conditions are equivalent*:

(i) *the* S.P. *is deterministic*

(ii) $\mathscr{M}_s = \mathscr{M}_t$ *for all* $s, t$, $-\infty < s, t < \infty$

(iii) $\mathscr{M}_s = \mathscr{M}_t$ *for some* $s, t \quad -\infty < s < t < \infty$
(iv) $f_t \in \mathscr{M}_s$ *for some* $s, t, \quad -\infty < s < t < \infty$ .

Let the S.P. be non-deterministic. Then by 5.1 (iii) for any $t$ and any $s < t$, $\mathscr{M}_{-\infty} \subseteqq \mathscr{M}_s \subset \mathscr{M}_t$. Hence

$$\mathscr{N}_t = \mathscr{M}_{-\infty}^{\perp} \cap \mathscr{M}_t \neq \{0\}, \quad -\infty < t < \infty ,$$

and we get the decomposition

$$\mathscr{M}_t = \mathscr{M}_{-\infty} + \mathscr{N}_t, \quad \mathscr{M}_{-\infty} \perp \mathscr{N}_t \neq \{0\}, \quad -\infty < t < \infty . \tag{5.2}$$

Moreover from (3.5)

$$U_t(\mathscr{N}_s) = \mathscr{N}_{s+t}, \ -\infty < s, t < \infty . \tag{5.3}$$

If in the preceding paragraphs of this section we interpret $s, t$ as integers rather than as real numbers, we get the definition and properties of non-deterministic processes in the discrete parameter case. But in the discrete case, additional results are readily available. We recall some of these in the next paragraph.

Let $(f'_n)_{-\infty}^{\infty}$ be any weakly stationary, non-deterministic S.P. with shift operator $V$. Denote by $(f'_n \mid \mathscr{M}'_{n-1})$ the orthogonal projection of $f'_n$ on the subspace $\mathscr{M}'_{n-1}$, cf. (4.2). Then

$$g'_n = f'_n - (f'_n \mid \mathscr{M}'_{n-1}) \neq 0, \quad -\infty < n < \infty . \tag{5.4}$$

The vectors $g'_n$ and $h'_n = g'_n / |g'_n|$ are called the $n$th *innovation* and *normalized innovation vectors*, respectively, of the process $(f'_n)_{-\infty}^{\infty}$. It is easily seen that

$$(h'_m, h'_n) = \delta_{mn}, \quad h'_{m+n} = V^m(h'_n), \quad -\infty < m, n < \infty , \tag{5.5}$$

so that $(h'_n)_{-\infty}^{\infty}$ is an orthonormal S.P. with the same shift operator $V$ as $(f'_n)_{-\infty}^{\infty}$. It is an important fact that in the discrete analogue of (5.2), viz.

$$\mathscr{M}'_n = \mathscr{M}'_{-\infty} + \mathscr{N}'_n, \quad \mathscr{M}'_{-\infty} \perp \mathscr{N}'_n \neq \{0\} , \tag{5.6}$$

the subspace $N'_n$ is the past and present of $h'_n$:

$$\mathscr{N}'_n = \mathfrak{S}(h'_k)_{k=-\infty}^{n} = \mathfrak{S}(V^k(h'_0))_{k=-\infty}^{n} . \tag{5.7}$$

The relations (5.6), (5.7) constitute the *Wold decomposition* of $M'_n$. From this decomposition follows at once the canonical decomposition of $f'_n$ into a one-sided moving-average part and a deterministic part:

$$\begin{cases} f'_n = u'_n + v'_n, & -\infty < n < \infty \\ u'_n = (f'_n \mid \mathscr{N}'_{-\infty}) = \sum_{k=0}^{\infty} c_k h'_{n-k}, \quad c_0 = |g'_0|, & \sum_{k=0}^{\infty} |c_k|^2 < \infty \\ v'_n = (f'_n \mid \mathscr{M}'_n) \\ (u'_n)_{-\infty}^{\infty},\ (v'_n)_{-\infty}^{\infty} \text{ have the same shift operator } V \text{ as } (f'_n)_{-\infty}^{\infty}. \end{cases} \tag{5.8}$$

To revert to the continuous parameter case, let $(f_t, -\infty < t < \infty)$ be a weakly stationary, mean-continuous, non-deterministic S.P. with shift group $(U_t, -\infty < t < \infty)$. It is clear from the equalities in 3.7(b) that attempts to define "innovation vectors" $g_t$ for this process by an equation analogous to (5.4) will fail. Indeed, since there is no atomic time unit in the continuous parameter case, all that we may expect our $f_t$-process to possess are "differential innovations."

Now let $(f'_n)_{-\infty}^{\infty}$ be the discrete S.P. associated with $(f_t, -\infty < t < \infty)$. Since the latter process is non-deterministic, it follows from Theorem 4.4 that so is the former. Let $h'_0$ be its 0th normalized innovation vector, and let

$$h_t = U_t(h'_0), \quad -\infty < t < \infty . \tag{5.9}$$

The resulting process $(h_t, -\infty < t < \infty)$ plays an important role in the theory. In § 6 we shall show that it is weakly Markovian, and explain how the differential innovations of the $f_t$-process can be had from it.[4] Here we shall show that the subspaces $\mathscr{N}_t$ of (5.2) are its past and present subspaces:

5.10 THEOREM. (*Pre-Wold Decomposition*) *Let* $(f_t, -\infty < t < \infty)$ *be a weakly stationary, mean-continuous, non-deterministic* S.P. *with shift group* $(U_t, -\infty < t < \infty)$, *so that cf.* (5.2)

$$\mathscr{M}_t = \mathscr{M}_{-\infty} + \mathscr{N}_t, \quad \mathscr{M}_{-\infty} \perp \mathscr{N}_t ,$$

*Then* $\mathscr{N}_t = \mathfrak{S}(h_s)_{s \leqq t}$ *is the past and present of* $h_t$, $-\infty < t < \infty$.

*Proof*. By Theorem 4.4, $\mathscr{M}'_0 = \mathscr{M}_0$, $\mathscr{M}'_{-\infty} = \mathscr{M}_{-\infty}$. Hence, taking $t = 0 = n$ in (5.2), (5.6) we see that $\mathscr{N}'_0 = \mathscr{N}_0$. But taking $X = \{h'_0\}$ in 2.9(a), where $h'_0$ is the 0th normalized innovation of the associated discrete process, we find on using (5.7) that

$$\mathscr{N}_0 = \mathscr{N}'_0 = \mathfrak{S}(V^k(h'_0))_{k=-\infty}^{0} = \mathfrak{S}(U_s(h'_0))_{s \leqq 0} . \tag{5.11}$$

Hence by (5.3) and (5.9)

$$\mathscr{N}_t = U_t(\mathscr{N}_0) = \mathfrak{S}(U_s(h'_0))_{s \leqq t} = \mathfrak{S}(h_s)_{s \leqq t} .$$

## 6. Differential innovations and the Wold decomposition.

Let

[4] The physical significance of the $h_t$-process has been indicated by Wiener and Wintner [20]. When $\mathfrak{X}$ is the class of $L_2$-functions on a probability space $(\Omega, \mathscr{B}, P)$, and $t$ is the time, $h_t$ provides the weak (or wide sense) version of "random time", i.e. time as measured by a perfect clook which is subjected to Brownian fluctuations. More precisely, if

$$y_t(\omega) = \exp[i\lambda\{t + ax_t(\omega)\}], \qquad \omega \in \Omega$$

where $(x_t, -\infty < t < \infty)$ is the Brownian movement S.P., and $\lambda$, $a$ are constants such that $a\lambda = \sqrt{2}$, then the $y_t$- and $h_t$-processes have the same wide sense properties.

$(f_t, -\infty < t \in \infty)$ be a weakly stationary, mean-continuous, non-deterministic S.P. with shift group $(U_t, -\infty < t < \infty)$, and let $h'_0$ be the 0th normalized innovation vector of the associated discrete process $(f'_n)_{-\infty}^{\infty}$. In the next lemma we study the S.P. $(h_t, -\infty < t < \infty)$ defined by (5.9), the present and past subspaces $\mathscr{N}_t$ of which have been mentioned in the Pre-Wold decomposition 5.10.

6.1 LEMMA. (a) *The $h_t$-process is weakly (or wide sense) Markovian; more fully,*

$$(h_t \mid \mathscr{N}_s) = e^{s-t}h_s, \quad -\infty < s < t < \infty$$

*depends only on the terminal vector $h_s$ of $\mathscr{N}_s = \mathfrak{S}(h_\sigma)_{\sigma \leqq s}$.*

(b) *Its convariance function $\gamma$ is given by $\gamma_t = e^{-|t|}$, $-\infty < t < \infty$.*

*Proof.* (a) Let $t \geqq 0$. Then by 2.7

$$h_t = e^{-t}h'_0 + \lim_{n\to\infty} \sum_{k=1}^{n} \frac{1}{k!}\left(\frac{-nt}{n+1}\right)^k \{(I + A_n)^k - I\}(h'_0)$$

where

$$A_n(h'_0) = \frac{2n}{n+1} \sum_{j=1}^{\infty} \left(\frac{n-1}{n+1}\right)^{j-1} V^j(h'_0)\ .$$

Since by (5.5) the $h'_j = V^j(h'_0)$ constitute an orthonormal process, we see that

$$h_t = e^{-t}h'_0 + \eta_t, \quad \text{where} \quad \eta_t \perp h'_0, h'_{-1}, \cdots, t \geqq 0\ .$$

It follows from (5.10) that $\eta_t \perp \mathscr{N}_0 = \mathfrak{S}(h_s)_{s \leqq 0}$. Hence

$$e^{-t}h'_0 = (h_t \mid \mathscr{N}_0),\ t \geqq 0.$$

On applying $U_s$ to both sides we get

$$(1) \qquad e^{-t}h_s = (h_{s+t} \mid \mathscr{N}_s)\ , \quad -\infty < s < \infty\ , \quad t \geqq 0\ .$$

This reduces to the desired relation on changing the index.

(b) From (1) it follows at once that

$$\gamma_t = (h_{s+t}, h_s) = e^{-t}\ , \quad t \geqq 0.$$

This in turn entails that $\gamma_t = e^{-|t|}$, $-\infty < t < \infty$.

We shall now study the $\xi_t$-process mentioned in (1.1). By definition

$$(6.2) \qquad \xi_t = \frac{1}{\sqrt{2}}\left\{h_t - h_0 + \int_0^t h_s ds\right\}\ , \quad -\infty < t < \infty\ .$$

It follows at once that $\xi_0 = 0$ and

$$\text{(6.3)} \qquad \xi_b - \xi_a = \frac{1}{\sqrt{2}}\left\{h_b - h_a + \int_a^b h_s ds\right\}, \quad -\infty < a, b < \infty .$$

6.4 THEOREM. (a) *The $\xi_t$-process has increments which are stationary under the group $(U_t, -\infty < t < \infty)$, i.e.*

$$U_t(\xi_b - \xi_a) = \xi_{b+t} - \xi_{a+t}, \quad -\infty < a, b, t < \infty .$$

(b) *The $\xi_t$-process has orthogonal increments, i.e.*

$$\xi_b - \xi_a \perp \xi_d - \xi_c, \quad \textit{if} \quad -\infty < a < b \leqq c < d < \infty .$$

(c) $|\xi_b - \xi_a|^2 = |b - a|, \quad -\infty < a, b < \infty .$

(d) $(\xi_b - \xi_a, \xi_d - \xi_c) = |\xi_b - \xi_c|^2 = b - c$, *if* $-\infty < a < c < b < d < \infty$ .

*Proof.* (a) follows at once from (6.3) since $U_t(h_s) = h_{s+t}$.

(b) Let $a < b \leqq c < d$. Then from (6.3) and 6.1(b),

$$\begin{aligned} 2(\xi_b - \xi_a, \xi_d - \xi_c) &= \left(h_b - h_a + \int_a^b h_s ds, \quad h_d - h_c + \int_c^d h_t dt\right) \\ &= \left\{e^{b-d} - e^{b-c} + \int_c^d e^{b-t} dt\right\} - \left\{e^{a-d} - e^{a-c} + \int_c^d e^{a-t} dt\right\} \\ &\quad + \int_a^b \left\{e^{s-d} - e^{s-c} + \int_c^d e^{s-t} dt\right\} ds . \end{aligned}$$

Since the expression in each { } on the R.H.S. is zero, the result follows.

(c) First let $0 = a < b$. Then from (6.3) and 6.1(b)

$$\begin{aligned} 2|\xi_b - \xi_0|^2 &= \left(h_b - h_0 + \int_0^b h_s ds, \quad h_b - h_0 + \int_0^b h_t dt\right) \\ &= \left\{1 - e^{-b} + \int_0^b e^{t-b} dt\right\} - \left\{e^{-b} - 1 + \int_0^b e^{-t} dt\right\} \\ &\quad + \int_0^b \left\{e^{s-b} - e^{-s} + \int_0^b e^{-|s-t|} dt\right\} ds \\ &= 2(1 - e^{-b}) + 0 + \int_0^b \int_0^b e^{-|s-t|} dt ds . \end{aligned}$$

Since the last integral equals

$$\int_0^b \left\{\int_0^s e^{t-s} dt + \int_s^b e^{s-t} dt\right\} ds = 2b + 2(e^{-b} - 1) ,$$

it follows that $|\xi_b - \xi_0|^2 = b$.

Next, let $-\infty < a < b < \infty$. Then by (a) $\xi_b - \xi_a = U_a(\xi_{b-a} - \xi_0)$, $b - a > 0$, and so

$$|\xi_b - \xi_a|^2 = |\xi_{b-a} - \xi_0|^2 = b - a .$$

(c) is a simple consequence of (a), (b), the vertification of which we leave to the reader.

In view of the last theorem, the stochastic integral $\int_{-\infty}^{\infty} c(s)d\xi_s$ will exist for any complex-valued function $c \in L_2(-\infty, \infty)$, cf. Doob [5, Ch. IX, § 2]. In the next lemma we shall show that the vector $h_t$ is expressible in terms of the $\xi_s$ by means of such an integral. In effect we shall invert the relation expressed in (6.2):

6.5 LEMMA. (*Inversion formula*)

$$h_t = \sqrt{2}\left\{\xi_t - \int_{-\infty}^{t} e^{s-t}\xi_s ds\right\} = \sqrt{2}\int_{-\infty}^{t} e^{s-t}d\xi_s\ , \quad -\infty < t < \infty\ .$$

*Proof*. Since $h_t = U_t(h_0')$ and $U$ (as a function of $t$) is strongly continuous on $(-\infty, \infty)$, it follows that the vector-valued function $h$ *is* continuous for $t \in (-\infty, \infty)$, and therefore by (6.2) so is the function $\xi$. Hence the Riemann integral $\int_a^b e^{s-t}\xi_s ds$ exist for $-\infty < a < b \leqq t$. Moreover, since

$$\left|\int_a^b e^{s-t}\xi_s ds\right| \leqq \int e^{s-t}\,|\,\xi_s\,|\,ds = \int_a^b e^{s-t}\sqrt{s}\cdot ds \text{ or } \sqrt{s}\cdot ds\ ,$$

the infinite integral $\int_{-\infty}^{t} e^{s-t}\xi_s ds$ converges.

Now consider the case $t = 0$. We have from (6.2)

$$\begin{aligned}\sqrt{2}\int_{-\infty}^{0} e^{s}(\xi_0 - \xi_s)ds &= -\int_{-\infty}^{0} e^{s}\left\{h_s - h_0 + \int_0^s h_\sigma d\sigma\right\}ds \\ &= -\int_{-\infty}^{0} e^{s}h_s ds + h_0 + \int_{-\infty}^{0}\int_s^0 e^{s}h_\sigma d\sigma ds\ .\end{aligned}$$

Now by Dirichlet's formula the last integral equals

$$\int_{-\infty}^{0}\int_{-\infty}^{0} e^{s}h_\sigma ds d\sigma = \int_{-\infty}^{0}\left\{\int_{-\infty}^{0} e^{s}ds\right\}h_\sigma d\sigma = \int_{-\infty}^{0} e^{\sigma}h_\sigma d\sigma\ .$$

Hence

$$(1) \qquad \sqrt{2}\int_{-\infty}^{0} e^{s}(\xi_0 - \xi_s)ds = h_0\ .$$

Since for any real $t$, $U_t(\xi_0 - \xi_s) = \xi_t - \xi_{s+t}$, $U_t(h_0) = h_t$, we get the first equality in the lemma by applying $U_t$ to both sides of (1) and then changing variables.

The second equality follows on integrating by parts:

$$\int_{-\infty}^{t} e^{s-t}d\xi_s = [e^{s-t}\xi_s]_{s\to-\infty}^{s=t} - \int_{-\infty}^{t}\xi_s d_s(e^{s-t}) = \xi_t - \int_{-\infty}^{t} e^{s-t}\xi_s ds\ .$$

The use of integration by parts is justified as follows. In the first place, for $-\infty < a < t < \infty$ we have

$$(2) \qquad \int_a^t e^{s-t} d\xi_s = [e^{s-t}\xi_s]_{s=a}^{s=t} - \int_a^t \xi_s d_s(e^{s-t}) .$$

This follows from the fact that for a continuous integrand the stochastic integral is a Riemann-Stieltjes integral (with vector-valued integrator $\xi_s$) and that for the latter, integration by parts is valid, cf. [5, p. 429 (2.6)] and [11, p. 63 (3.31)]]. Next, the last integral in (2) is obviously equal to $\int_a^t \xi_s e^{s-t} ds$. Finally, since both $\int_{-\infty}^t e^{s-t} d\xi_s$, $\int_{-\infty}^t \xi_s e^{s-t} ds$ are known to exist, we can let $a \to -\infty$ in (2), cf. [5, p. 428 (2.4)].

The formulae (6.2) and 6.5 together entail the following important result:

6.6 LEMMA. *For any real $t$, the past and present subspace $\mathscr{N}_t$ of $h_t$ is the set of all (convergent) stochastic integrals $\int_{-\infty}^t c(s)d\xi_s$, with complex-valued functions $c \in L_2(-\infty, t)$, i.e. $\mathscr{N}_t = \mathfrak{S}(\xi_\sigma - \xi_\tau)_{\sigma,\tau \leqq t}$.*

*Proof.* Denote by $\mathscr{N}_t^{(\xi)}$ the set of all such stochastic integrals. Let $-\infty < \tau \leqq t < \infty$. Then by 6.5

$$h_\tau = \sqrt{2}\int_{-\infty}^{\tau} e^{s-\tau} d\xi_s = \int_{-\infty}^{\tau} c(s) d\xi_s ,$$

where $c(s) = \sqrt{2}e^{s-\tau}$ on $(-\infty, \tau]$ and $c(s) = 0$ on $(\tau, t]$. Since $c \in L_2(-\infty, t]$, it follows that $h_\tau \in \mathscr{N}_t^{(\xi)}$. Hence $\mathscr{N}_t = \mathfrak{S}(h_\tau)_{\tau \leqq t} \subseteqq \mathscr{N}_t^{(\xi)}$.

To prove the reverse inclusion, let

$$g = \int_{-\infty}^t c(s) d\xi_s, \text{ where } c \in L_2(-\infty, t] .$$

Suppose first that $c$ is a step-function:

$$c(s) = \sum_{k=1}^n c_k \chi_{J_k}(s)$$

$\chi_{J_k}$ being the indicator function of the interval $J_k = [a_k, b_k] \subseteqq (-\infty, t]$. Then by definition (cf. Doob [5, p. 427 (2.1)])[5]

$$g = \sum_{k=1}^n c_k (\xi_{b_k} - \xi_{a_k}) .$$

From (6.3) it is clear that $g \in \mathscr{N}_t$. Next suppose $c \in L_2(-\infty, t]$, and $c = \lim_{n\to\infty} c^{(n)}$, where $c^{(n)}$ is a step-function. Then by definition

[5] We note that from 6.4(c) it follows that $\xi_{t-} = \xi_t = \xi_{t+}$, $-\infty < t < \infty$.

$$g = \lim_{n\to\infty} \int_{-\infty}^{t} c^{(n)}(s)d\xi_s \in \mathcal{N}_t ,$$

since $\mathcal{N}_t$ is closed. Thus $\mathcal{N}_t^{(\xi)} \subseteqq \mathcal{N}_t$.

We may sum up the main results established so far as follows:

6.7 THEOREM. (*Wold Decomposition I*) *Let* $(f_t, -\infty < t < \infty)$ *be a weakly stationary, mean-continuous, non-deterministic* S.P. *with shift group* $(U_t, -\infty < t < \infty)$. *Let* $h_0'$ *be the* 0*th normalized innovation of the associated discrete process, and let*

$$h_t = U_t(h_0') , \quad \xi_t = h_t - h_0 + \int_0^t h_s ds , \quad -\infty < t < \infty .$$

*Then* (a) $\mathcal{M}_t = \mathcal{N}_t + \mathcal{M}_{-\infty}$, $\mathcal{N}_t \perp \mathcal{M}_{-\infty}$, $-\infty < t < \infty$, *where* $\mathcal{N}_t = \mathfrak{S}(h_s)_{s\leqq t}$ *is the past and present of* $h_t$;

(b) *the* $\xi_t$*-process has stationary, orthogonal increments such that* $|\xi_t - \xi_s|^2 = |t-s|$; *moreover,* $\mathcal{N}_t = \mathfrak{S}(\xi_\sigma - \xi_\tau)_{\sigma,\tau\leqq t}$, *i.e.* $\mathcal{N}_t$ *is the set of all stochastic integrals* $\int_{-\infty}^{t} c(s)d\xi_s$ *with* $c \in L_2(-\infty, t]$.

6.8 UNIQUENSES THEOREM. *Let* $(\eta_t, -\infty < t < \infty)$ *be any process with the following properties*:

( i ) *it has orthogonal increments such that*

$$|\eta_b - \eta_a|^2 = |b-a| , \quad -\infty < a, b < \infty , \quad \textit{and} \quad \eta_0 = 0$$

(ii) $U_t(\eta_b - \eta_a) = \eta_{b+t} - \eta_{a+t}$, $-\infty < a, b, t < \infty$

(iii) $\mathfrak{S}(\eta_\sigma - \eta_\tau)_{\sigma,\tau\leqq 0} = \mathcal{M}_{-\infty}^{\perp} \cap \mathcal{M}_0$.

*Then* $\eta_t = e^{i\alpha}\xi_t$, *where* $\xi_t$ *is as in* 6.7, *and* $\alpha$ *is some real number.*

*Proof.* Our proof of this result is essentially that given by Hanner [9, p. 175–176]. Since our treatments and notations differ, we may indicate the main steps. We first show that

$$\mathfrak{S}(\eta_\sigma - \eta_\tau)_{\sigma,\tau\leqq b} = \mathcal{N}_b , \qquad \mathfrak{S}(\eta_\sigma - \eta_\tau)_{a\leqq\sigma,\tau\leqq b} = \mathcal{N}_a^{\perp} \cap \mathcal{N}_b ,$$

where $\mathcal{N}_b$ is an in 6.7(a). It follows from 6.7(b) that $\xi_b - \xi_a = \int_a^b f_{a,b}(s)d\eta_s$. By piecing together the functions $f_{n,n+1}$, $-\infty < n < \infty$, we can define a function $f$ on $(-\infty, \infty)$ such that

$$\xi_b - \xi_a = \int_a^b f(s)d\eta_s , \qquad a < b .$$

Using the fact that $\xi_b - \xi_a = U_h(\xi_{b-h} - \xi_{a-h})$, we can show that $f$ is essentially constant-valued on $(-\infty, \infty)$. From this the desired result is immediate.

An immediate corollary of Theorem 6.7 is the cannonical decomposition of the vector $f_t$ itself:

6.9 COROLLARY. (*Wold Decomposition II*) *With the hypothesis of Theorem* 6.7 *we have*

(a) $f_t = u_t + v_t, \quad u_t = (f_t \mid \mathscr{N}_t), \quad v_t = (f_t \mid \mathscr{M}_{-\infty})$;

(b) *the* $u_t$*-process in* (a) *is a one-sided moving average, i.e.*

$$u_t = \int_0^\infty c(s) d_s \xi_{t-s}\,, \quad -\infty < t < \infty, \quad \textit{where} \quad c \in L_2[0, \infty)\,.$$

*and* $\mathfrak{S}(u_s)_{s \leq t} = \mathscr{N}_t, \; -\infty < t < \infty$;

(c) *the* $v_t$*-process is deterministic, and* $\mathfrak{S}(v_s)_{s \leq t} = \mathscr{M}_{-\infty}$, *for* $-\infty < t < \infty$.

**7. Purely non-deterministic stochastic processes.** We call a weakly stationary S.P. *purely non-deterministic*, if and only if $\mathscr{M}_{-\infty} = \{0\}$. For completeness we state here the anologue of a theorem given by Kolmogorov for discrete parameter processes:

7.1 THEOREM. *For any weakly stationary, mean-continuous stochastic process* $(f_t, \; -\infty < t < \infty)$ *the following conditions are equivalent*:

( i ) $(f_t, \; -\infty < t < \infty)$ *is purely non-deterministic*;

(ii) $(f_t, \; -\infty < t < \infty)$ *is a one-sided moving average*:

$$f_t = \int_0^\infty c(s) d_s \xi_{t-s}\,, \quad c \in L_2[0, \infty]\,,$$

$(\xi_s, \; -\infty < s < \infty)$ *being a process with stationary and orthogonal increments such that* $| \xi_b - \xi_a |^2 = | b - a |$;

(iii) $\lim_{t \to \infty} (f_0 \mid \mathscr{M}_{-t}) = 0$.

*Proof.* The proof runs parallel to that in the discrete case and is omitted.

It follows from Corollary 6.9 and Theorem 7.1 that every weakly stationary, mean-continuous, non-deterministic S.P. $(f_t, \; -\infty < t < \infty)$ can be decomposed in the form $f_t = u_t + v_t$, where the $u_t$-process is purely non-deterministic, the $v_t$-process is deterministic, and all three processes have the same shift group $(U_t, \; -\infty < t < \infty)$. We shall refer to the $u_t$-and $v_t$-processes as the *purely non-deterministic part* and the *deterministic part* of the $f_t$-process. With an obvious notation, we have

$$\mathscr{M}_t = \mathscr{M}_t^{(u)} + \mathscr{M}_t^{(v)}\,, \qquad \mathscr{M}_\infty^{(u)} \perp \mathscr{M}_\infty^{(v)}$$
$$\mathscr{M}_t^{(u)} = \mathscr{N}_t\,, \qquad \mathscr{M}_t^{(v)} = \mathscr{M}_{-\infty}\,.$$

Now let $(u'_n)_{-\infty}^{\infty}$, $(v'_n)_{-\infty}^{\infty}$ be the purely non-deterministic and deterministic parts of the discrete process $(f'_n)_{-\infty}^{\infty}$ associated with $(f_t, -\infty < t < \infty)$. Then by 6.9(a), 4.4(c), and (5.8)

$$v_0 = (f_0 \mid \mathscr{M}_{-\infty}) = (f'_0 \mid \mathscr{M}'_{-\infty}) = v'_0 ,$$

and therefore

$$u_0 = f_0 - v_0 = f'_0 - v'_0 = u'_0 .$$

Moreover, the shift operator $V$ of the $u'_n$-, $v'_n$-processes is the Cayley transform of $H$, where $iH$ is the infinitesimal generator of the shift group $(U_t, -\infty < t < \infty)$ of the $u_t$-, $v_t$-processes. We can thus assert the following:

7.2 COROLLARY. *If $(f'_n)^{\infty}$ is the discrete process associated with the weakly stationary, mean-continuous, non-deterministic S.P. $(f_t, -\infty < t < \infty)$, then the purely non-deterministic and deterministic parts of $(f'_n)_{-\infty}^{\infty}$ are the discrete processes associated with the deterministic and purely non-deterministic parts of $(f_t, -\infty < t < \infty)$.*

## REFERENCES


1. K. Balagangadharan, *The prediction theory of stationary random distributions*, Memoirs of the College of Sci., University of Kyoto, series A, **33** (1960), 243-256.
2. A. Beurling, *On two problems concerning linear transformations in Hilbert space*, Acta Mathematica, **81** (1949), 239-255.
3. J. L. B. Cooper, *One-parameter semi-groups of isometric operators in Hilbert space*, Ann. Math., **48** (1947), 827-842.
4. H. Cramer, *On the linear prediction problem for certain stochastic processes*, Ark. for Mat., **4** (1959), 45-54.

4′. ———, *On some classes of non-stationary stochastic processes*, Proc. Fourth Berkeley Symposium of Mathematical Statistics and Probability, Berkeley (1960), 57-78.

4″. ———, *On the structure of purely non-deterministic stochastic processes*, Ark. for Mat., **4** (1961), 249-266.

5. J. L. Doob, *Stochastic Processes*, New York, 1953.
6. N. Dunford and J. Schwarz, *Linear Operators*, Part I, New York, 1958.
7. I. M. Gelfand, *Stationary random processes* (*Russian*), C. R. (Doklady), Acad. Sci. U.R.S.S., **100** (1955), 853-856, **208** (1961), 102-112.
8. P. Halmos, *Shifts of Hilbert spaces*, Crelle's Journal, **208** (1961), 102-112.
9. O. Hanner, *Deterministic and non-deterministic stationary random processes*, Ark. for Mat., **1** (1950), 161-177.
10. H. Helson and D. Lowdenslager, *Prediction theory and Fourier series in several variables*, Acta Mathematica, **99** (1958), 165-202.
11. E. Hille and R. S. Phillips, *Functional Analysis and Semigroups*, Providence, 1957.
12. K. Ito, *Stationary random distributions*, Memoirs of the College of Sci., Univ. of Kyoto, Series A, **28** (1953), 209-223.
13. K. Karhunen, *Über lineare Methoden in der Warhrscheinlickeitsrechnung*, Am. Acad. Sci. Fennicae, A, I, **37** (1947).
14. P. Lax, *Translation invariant subspaces*, Acta Mathematica, **101** (1959), 163-178.
15. P. Masani, *Shift invariant spaces and prediction theory*, Acta Mathematica, **107** (1962), 275-290.

16. ———, *On isometric flows on Hilbert space*. Bull. Amer. Math. Soc., **68** (1962), 624-632.

17. F. Riesz and B. Sz.-Nagy, *Functional Analysis*, Translated from 2nd French Ed., New York, 1959.

18. G. Sansone, *Orthogonal Polynomials*, New York, 1959.

19. M. H. Stone, *Linear Transformations in Hilbert Space*, New York, 1932.

20. N. Wiener and A. Wintner, *Random time*, Nature, **181** (1958), 561-562.

INDIANA UNIVERSITY

# NON-EXISTENCE OF ALMOST-COMPLEX STRUCTURES ON QUATERNIONIC PROJECTIVE SPACES

W. S. MASSEY

**1. Introduction.** In [3] F. Hirzebruch proved that the $n$-dimensional quaternionic projective space (which we denote by $P_n(\boldsymbol{H})$) does not admit any almost structure in case $n \neq 2$ or 3. According to Hirzebruch's lecture at the 1958 International Congress [5], Milnor has since proved that $P_2(\boldsymbol{H})$ and $P_3(\boldsymbol{H})$ do not admit almost complex structures. At the time of this writing, Milnor's proof has not yet been published.

It is the purpose of this note to give a short proof of this theorem of Milnor's making use of the theory of the ring $K(X)$ of complex vector bundles over a space $X$ due to Atiyah and Hirzebruch together with certain facts that are readily available in the literature. From the brief description given in Hirzebruch's lecture (loc. cit.) it seems that our method is quite different from Milnor's. Our method may be applicable in other cases to prove the existence or nonexistence of almost complex structures on a manifold.

**2. Summary of some known facts.** We will make use of the following results:

(a) The cohomology ring $H^*(P_n(\boldsymbol{H}), Z)$ is a truncated polynomial ring generated by a 4-dimensional cohomology class $u$ and subject to the single relation $u^{n+1} = 0$.

(b) Let $\tau_n$ denote the tangent bundle to $P_n(\boldsymbol{H})$. The total Pontrjagin class of $\tau_n$ is given by the formula

$$p(\tau_n) = (1 + u)^{2n+2}(1 + 4u)^{-1} \tag{1}$$

for appropriate choice of the generator $u$ (Borel and Hirzebruch [2], 15.5 or Hirzebruch [3]).

(c) We will use the following notation: If $\xi$ is a real $n$-plane bundle, then $\xi \otimes \boldsymbol{C}$ denotes its complexification, while if $\xi$ is a complex $n$-plane bundle, then $\xi_R$ denotes the real $2n$-plane bundle obtained by "restriction of coefficients" to the reals. Also, $\xi^*$ denotes the complex conjugate bundle. We then have the following relation for any complex vector bundle $\xi$:

$$\xi_R \otimes \boldsymbol{C} = \xi + \xi^* \quad \text{(Whitney sum)}\,. \tag{2}$$

(see Hirzebruch, [4], p. 68, proof of Theorem 4.5.1). Moreover, for any

Received January 10, 1962. During the preparation of this note the author was partially supported by N. S. F. Grant G 18995.

real vector bundle $\eta$,

$$p_i(\eta) = (-1)^i c_{2i}(\eta \otimes C) , \tag{3}$$

(Hirzebruch [4], p. 67).

(d) We will also need to make use of the properties of the ring $K(X)$ as summarized for example in § 2 of Atiyah and Todd [1]. We will use their notation and results without any further comment.

**3. The ring $K(P_n(H))$.** By proposition 2.3 of [1] the Chern character

$$ch: K(P_n(H)) \to H^*(P_n(H), Q)$$

is a monomorphism.

LEMMA 1. *The image of this homomorphism, denoted by $chP_n(H)$, is the subring of $H^*(P_n(H), Q)$ generated by*

$$2 \cosh \sqrt{v} = 2(1 + v/2! + v^2/4! + \cdots + v^n/(2n)!)$$

*where $v$ is an appropriately chosen generator of*

$$H^4(P_n(H), Z) \subset H^4(H), Q) .$$

*Proof.* Consider the well-known principal fibre bundle $\pi_n: S^{4n+3} \to P_n(H)$ with group $Sp(1)$; let $\eta_n$ denote the associated bundle with fibre a quaternionic vector space of dimension 1. We assert that the total symplectic Pontrjagin class of this bundle is

$$e(\eta_n) = 1 + e_1(\eta_n) = 1 + v$$

where $v$ is an appropriately chosen generator of $H^4(P_n(H), Z)$. This follows from the fact that $\pi_n: S^{4n+3} \to P_n(H)$ is a universal bundle for the group $Sp(1)$ (up to the dimension $4n + 2$), and that the integral cohomology ring of the classifying space for $Sp(1)$ is a polynomial ring generated by the symplectic Pontrjagin class $e_1$ (see Borel and Hirzebruch, [2], § 9.6).

Let $\xi_n$ denote the complex vector bundle obtained from $\eta_n$ by "restriction of coefficients" to the complex numbers; the associated principal bundle is a $U(2)$—bundle which is the extension of $\pi_n: S^{4n+3} \to P_n(H)$ under the standard inclusion $Sp(1) \subset U(2)$. By § 9.6 of [2],

$$e_1(\eta_n) = -c_2(\xi_n) ,$$

hence

$$c(\xi_n) = 1 - v .$$

The Chern character (see [2], § 9.1) of $\xi_n$ is

$$ch(\xi_n) = e^{y_1} + e^{y_2}$$

where

$$(1 + y_1)(1 + y_2) = c(\xi_n) = 1 - v\ .$$

Hence

$$y_1 + y_2 = 0\ , \quad y_1 y_2 = -v\ .$$

From this we conclude that

$$\begin{aligned} y_1 &= -y_2,\, v = y_1^2\ , \\ y_1 &= \sqrt{v}\,,\, y_2 = -\sqrt{v}\ , \\ ch(\xi_n) &= \exp(v) + \exp(-v) = 2\cosh\sqrt{v}\ . \end{aligned}$$

Hence $ch(P_n(\boldsymbol{H}))$ contains the subring generated by 2 cosh $\sqrt{v}$; it remains to show that it is exactly equal to this subring. This is done by induction on $n$ exactly as in the proof of Proposition 3.1 of [**1**]. The details may be left to the reader.

Note that $chP_n(\boldsymbol{H})$ may be equivalently described as the subring of $H^*(P_n(\boldsymbol{H}), \boldsymbol{Q})$ generated by

$$w = 2\cosh\sqrt{v} - 2 = v + 2v^2/4! + \cdots + 2v^n/(2n)!\ .$$

For many purposes this description of $ch\ P_n(H)$ is more convenient; note that $w^{n+1} = 0$, and $\{1, w, w^2, \cdots, w^n\}$ is a basis of $ch\ P_n(\boldsymbol{H})$ over the integers.

Lemma 1 and equation (1) above are both stated in terms of "appropriately" chosen generators, $v$ and $u$ respectively, of the infinite cyclic group $H^4(P_n(\mathrm{H}), \mathrm{Z})$. Therefore $u = \pm v$. *We assert that* $u = +v$. To prove this, it obviously suffices to show that $ch(\tau_n \otimes \mathrm{C})$ belongs to the subring of $H^*(P_n(\mathrm{H}), \mathrm{Q})$ generated by 2 cosh $\sqrt{u}$, but that it does not belong to the subring of $H^*(P_n(\boldsymbol{H}), \boldsymbol{Q})$ generated by 2 cosh $\sqrt{-u}$. This we will now do by an essentially straightforward, but rather lengthy, computation.

LEMMA 2. *The Chern character of* $\tau_n \otimes \boldsymbol{C}$ *is given by*

$$ch(\tau_n \otimes \boldsymbol{C}) = (4n + 4)\cosh\sqrt{u} - 4\cosh^2\sqrt{u}\ .$$

*Proof.* It follows from equations (1) and (3) that the total Chern class of $\tau_n \otimes \boldsymbol{C}$ is given by

$$c(\tau_n \otimes \boldsymbol{C}) = (1 - u)^{2n+2}(1 - 4u)^{-1}\ .$$

To compute the Chern character of $\tau_n \otimes \boldsymbol{C}$, we may proceed as follows: Write the total Chern class as a formal product

$$c(\tau_n \otimes \boldsymbol{C}) = \prod_{i=1}^{4n}(1 + x_i)\ ,$$

where the $x_i$'s have degree 2. Then

$$ch(\tau_n \otimes C) = \sum_{i=1}^{4n} \exp(x_i) ,$$

To actually carry out the computation, take the logarithm of both sides of the equation

$$\prod (1 + x_i) = (1 - u)^{2n+2}(1 - 4u)^{-1} ,$$

and use the MacLaurin series expansion of $\log(1 + z)$ and $\log(1 - z)$. The result is

$$\sum_{k>0} (-1)^{k+1} s_k/k = -\sum_{k>0} (2n + 2 - 4^k) u^k/k , \tag{4}$$

where

$$s_k = \sum_{i=1}^{4n} x_i^k .$$

Since each $x_i$ is of degree 2, while $u$ is of degree 4, we conclude from equation (4) that

$$\begin{aligned} s_k &= 0 \qquad \text{for } k \text{ odd} , \\ s_{2k} &= 2(2n + 2 - 4^k)u^k . \end{aligned}$$

Therefore

$$\begin{aligned} ch(\tau_n \otimes C) &= \sum_{i=1}^{4n} \exp(x_i) = \sum_{i=1}^{4n} \sum_k x_i^k/k ! \\ &= 4n + \sum_{k>0} s_k/k ! \\ &= 4n + \sum_{k>0} 2(2n + 2 - 4^k)u^k/(2k) ! \\ &= (4n + 4) \cosh \sqrt{u} - 2 \cosh \sqrt{4u} - 2 \\ &= (4n + 4) \cosh \sqrt{u} - 4 \cosh^2 \sqrt{u} , \end{aligned}$$

as was to be proved.

It is obvious from this formula that $ch(\tau_n \otimes C)$ belongs to the subring of $H^*(P_n(\boldsymbol{H}), \boldsymbol{Q})$ generated by $2 \cosh \sqrt{u}$; we must now prove that $ch(\tau_n \otimes C)$ does *not* belong to the subring generated by

$$2 \cosh \sqrt{-u} = 2 \cos \sqrt{u} .$$

Assume the contrary; then there exist integers $a_0, a_1, \cdots, a_n$ such that

$$ch(\tau_n \otimes C) = \sum_{k=0}^{n} a_k (2 \cos \sqrt{u} - 2)^k ,$$

that is,

$$\begin{aligned} &2\sum_{k=1}^{n}(2n+2-4^k)u^k/(2k)\,! \\ &\quad = \sum_{k=1}^{n} a_k(-u+2u^2/4!\,-2u^3/6!+\cdots\pm 2u^n/(2n)!)^k\,. \end{aligned}$$

If we compare coefficients of $u$, $u^2$, $u^3$, and $u^4$ in this equation, we obtain

$$\begin{aligned} a_1 &= -2n+2\,, \\ a_2 &= (n-4)/3\,, \\ a_3 &= (7-n)/18\,, \\ a_4 &= (5n-47)/504\,. \end{aligned}$$

For $n = 2$ or 3, $a_2$ is not an integer; for $n < 7$, $a_3$ is not an integer; and for any value of $n$, it is impossible that both $a_3$ and $a_4$ are integers. For, if $a_3$ is an integer, then

$$n \equiv 7 \bmod 18 \text{ or } 5n \equiv 35 \bmod 18\,,$$

while if $a_4$ is an integer, then

$$5n \equiv 47 \bmod 18$$

which is a contradiction.

This completes the proof that $u = +v$.

**4. Proof of the theorem.** Assume $\tau_n$ admits an almost complex structure $\theta_n$; we will show that this leads to a contradiction.

$\theta_n$ is a complex $2n$-plane bundle over $P_n(\boldsymbol{H})$ such that $\tau_n = \theta_{n\boldsymbol{R}}$. Then by equation (2)

$$\tau_n \otimes \boldsymbol{C} = \theta_{n\mathrm{R}} \otimes \boldsymbol{C} = \theta_n + \theta_n^*\,.$$

Next, recall that

$$ch_i(\theta_n^*) = (-1)^i\, ch_i(\theta_n)$$

where $ch_i$ denotes the component of $ch$ of degree $2i$. However, since the base space of the bundle $\theta_n$ is $P_n(\boldsymbol{H})$,

$$ch(\theta_n^*) = ch(\theta_n)\,.$$

Therefore

$$ch(\tau_n \otimes \mathrm{C}) = 2ch(\theta_n)\,.$$

It follows from Lemma 2 that

$$\begin{aligned} ch(\theta_n) &= (2n+2)\cosh\sqrt{u} - 2\cosh^2\sqrt{u} \\ &= -\tfrac{1}{2}(2\cosh\sqrt{u}-2)^2 + (n-1)(2\cosh\sqrt{u}-2) + 2n\,. \end{aligned}$$

This is the desired contradiction, since the Chern character of any complex vector bundle is an *integral* linear combination of the powers of $w = (2 \cosh \sqrt{u} - 2)$.

REMARK. We have actually proved a slightly stronger theorem, in that we have shown that for any integer $n > 1$, $P_n(\boldsymbol{H})$ does not admit a "generalized almost complex structure" as defined by Hirzebruch in his lecture [5]. As Hirzebruch remarks, this can be proved easily by induction on $n$, once the case $n = 2$ is taken care of. However, the above computations of $ch(P_n(\boldsymbol{H}))$ and $ch(\tau_n \otimes \boldsymbol{C})$ may be of some independent interest.

## BIBLIOGRAPHY

1. M. F. Atiyah and J. A. Todd, *On complex Stiefel manifolds,* Proc. Cambridge Phil. Soc., **56** (1960), 343-353.
2. A. Borel and F. Hirzebruch, *Characteristic classes and homogeneous spaces,* I. Amer. J. Math., **80** (1958), 458-538.
3. F. Hirzebruch, *Über die quaternionalen projektiven Räume* S.—B. Math. Nat. Kl. Bayer. Akad. Wiss. 1953, (1954), 301-312.
4. ———, *Neue Topologische Methoden in der Algebraischen Geometrie,* Ergeb. der Math. u. Grenzgebiete (N. F.) 9, Berlin, 1956.
5. ———, *Komplexe Manigfaltigkeiten,* Proc. Int. Congress. Math., 14-21 August 1958, Edinburgh, pp. 119-136.

# A THEOREM ON THE ACTION OF SO(3)

D. MONTGOMERY AND C. T. YANG

**1. Introduction.** We shall use notions given in [1]. Let $G$ be a compact Lie group acting on a locally compact Hausdorff space $X$. We denote by $F(G, X)$ the set of stationary points of $G$ in $X$, that is, $F(G, X) = \{x \in X \mid Gx = x\}$. If $G$ is a cyclic group generated by $g \in G$, $F(G, X)$ is also written $F(g, X)$.

Whenever $x \in X$, we call $Gx = \{gx \mid g \in G\}$ the *orbit* of $x$ and $G_x = \{g \in G \mid gx = x\}$ the *isotropy group* at $x$. By a *principal orbit* we mean an orbit $Gx$ such that $G_x$ is minimal. By an *exceptional orbit* we mean an orbit of maximal dimension which is not a principal orbit. By a *singular orbit* we mean an orbit not of maximal dimension. Denote by $U$ the union of all the principal orbits, by $D$ the union of all the exceptional orbits and by $B$ the union of all the singular orbits. Then $U$, $D$ and $B$ are all $G$-invariant and they are mutually disjoint. Moreover, $X = U \cup D \cup B$ and both $B$ and $D \cup B$ are closed in $X$.

Denote by $X^*$ the orbit space $X/G$ and by $\pi$ the natural projection of $X$ onto $X^*$. Whenever $A \subset X$, $A^*$ denotes the image $\pi A$. If $X$ is a connected cohomology $n$-manifold over $Z$ [1; p. 9], where $Z$ denotes the ring of integers, then the following results are known.

(1.1) *$U^*$ is connected* [1; p. 122] *so that whenever $x, y \in U$, $G_x$ and $G_y$ are conjugate.*

(1.2) $\dim_Z B^* \leqq \dim_Z U^* - 1$ *so that if $r$ is the dimension of principal orbits and $B_k$ is the union of all the $k$-dimensional singular orbits ($k < r$), then* $\dim_Z B_k \leqq n - r + k - 1$ [1; p. 118]. *Hence* $\dim_Z B \leqq n - 2$.

Denote by $E^{n+1}$ the euclidean $(n+1)$-space, by $S^n$ the unit $n$-sphere in $E^{n+1}$ and by SO(3) the rotation group of $E^3$. In this note $G$ is to be SO(3) and $X$ is to be a compact cohomology $n$-manifold over $Z$ with $H^*(X; Z) = H^*(S^n; Z)$.

Let us first observe the following examples.

1. Let $G = \mathrm{SO}(3)$ act trivially on $X = S^1$. (Here we have $n = 1$.)
2. Let $G = \mathrm{SO}(3)$ act on $E^{n+1} = E^5 \times E^{n-4}$ $(n \geqq 4)$ by the definition

$$g(x, y) = (gx, y)\ ,$$

where the action of $G$ on $E^5$ is an irreducible orthogonal action. Then $G$ acts on $X = S^n$ and in this action, the 2-dimensional orbits are all

Received December 20, 1961. The second named author is supported in part by the U. S. Army Research Office.

projective planes, $F(G, X)$ is an $(n-5)$-sphere and for every $x \in U$, $G_x$ is a dihedral group of order 4.

3. Let $G = \mathrm{SO}(3)$ act on $E^{n+1} = E^3 \times E^3 \times E^{n-5} (n \geqq 5)$ by the definition

$$g(x, y, z) = (gx, gy, z) ,$$

where the action on $E^3$ is the familiar one. Then $G$ acts on $X = S^n$ and in this action, the 2-dimensional orbits are all 2-spheres, $F(G, X)$ is an $(n-6)$-sphere and for every $x \in U$, $G_x$ is the identity group.

In all three examples, $D = \phi$ and $\dim B = n - 2$. The orbit space $X^*$ is $X$ itself in the first example and it is a closed $(n-3)$-cell with boundary $B^*$ in the other two examples.

The purpose of this note is to prove that if $X$ is a compact cohomology $n$-manifold over $Z$ with $H^*(X; Z) = H^*(S^n; Z)$, then every action of $G = \mathrm{SO}(3)$ on $X$ with $\dim_Z B = n - 2$ strongly resembles one of these examples. In fact, we shall prove the following:

THEOREM. *Let $X$ be a compact cohomology $n$-manifold over $Z$ with $H^*(X; Z) = H^*(S^n; Z)$ and let $G = \mathrm{SO}(3)$ act on $X$ with $\dim_Z B = n - 2$. Then $D = \phi$ and one of the following occurs.*

1. *$n = 1$ and $G$ acts trivially on $X$.*

2. *$n \geqq 4$ and for every $x \in U$, $G_x$ is a dihedral group of order 4. Moreover, the 2-dimensional ordits are all projective planes and $F(G, X)$ is a compact cohomology $(n-5)$-manifold over $Z_2$ with $H^*(F(G, X); Z_2) = H^*(S^{n-5}; Z_2)$, where $Z_2$ denotes the prime field of characteristic 2.*

3. *$n \geqq 5$ and for every $x \in U$, $G_x$ is the identity group. Moreover, the 2-dimensional orbits are all 2-spheres and $F(G, X)$ is a compact cohomology $(n-6)$-manifold over $Z_2$ with $H^*(F(G, X); Z_2) = H^*(S^{n-6}; Z_2)$.*

*In the last two cases, $B^*$ is a compact cohomology $(n-4)$-manifold over $Z$ with $H^*(B^*; Z) = H^*(S^{n-4}; Z)$ and $X^*$ is a compact Hausdorff space which is cohomologically trivial over $Z$ and such that $X^* - B^*$ is a cohomology $(n-3)$-manifold over $Z$.*

The proof of this theorem is given in the next three sections.

2. **The set $D$.** Let $X$ be a connected cohomology $n$-manifold over $Z$ and let $G = \mathrm{SO}(3)$ act on $X$ with $\dim_Z B = n - 2$. If $G$ acts trivially on $X$, it is clear that $n = 1$ and that $D = \phi$. Hence we shall assume that the action of $G$ on $X$ is nontrivial.

Since $G$ is a 3-dimensional simple group which has no 2-dimensional subgroup, it follows that

(2.1) *$G$ acts effectively on $X$ and no orbit is 1-dimensional.*

(2.2) *Principal orbits are 3-dimensional so that for every $x \in U \cup D$, $G_x$ is finite.*

By (2.1), principal orbits are either 2-dimensional or 3-dimensional. If principal orbits are 2-dimensional, then $B = F(G, X)$ so that, by (1.2), $\dim_Z B < n - 2$, contrary to our assumption.

(2.3) *Denote by $B_2$ the union of all the 2-dimensional orbits. Then $\dim_Z B_2 = n - 2$ so that $B_2 \neq \phi$ and $n \geqq 4$. Moreover, whenever $Gz$ is a 2-dimensional orbit, $G_z$ is either a circle group or the normalizer of a circle group and accordingly $Gz$ is either a 2-sphere or a projective plane.*

By (2.2), $n = \dim_Z X \geqq \dim_Z U \geqq 3$. We infer that $B_2 \neq \phi$ so that $n - 2 = \dim_Z B_2 \geqq 2$. Hence $n \geqq 4$.

(2.4) *Let $x \in U$. Whenever $y \in D$, there is a $g \in G$ such that $G_x$ is a normal subgroup of $G_{gy}$.*

Let $S$ be a connected slice at $y$ [1; p. 105]. Then $S$ is a connected cohomology $(n - 3)$-manifold over $Z$ and $G_y$ acts on $S$. As seen in [7], $S$ is also a connected cohomology $(n - 3)$-manifold over $Z_p$ for every prime $p$, where $Z_p$ denotes the prime field of characteristic $p$.

Let $x' \in S \cap U$. We claim that $G_{x'}$ is a normal subgroup of $G_y$. Since $G_y$ is a finite group (see (2.2)) and $G_{x'}$ is a subgroup of $G_y$, there exists a neighborhood $N$ of the identity in $G$ such that $N^{-1}G_{x'}N \cap G_y = G_{x'}$. Let $V$ be a neighborhood of $x'$ such that whenever $x'' \in V$, $hG_{x''}h^{-1} \subset G_{x'}$ for some $h \in N$. (For the existence of $V$, see [4; p. 216].) Then for every $x'' \in V \cap S$, $G_{x''} \subset N^{-1}G_{x'}N \cap G_y = G_{x'}$ so that $G_{x''} = G_{x'}$. Therefore $G_{x'}$ leaves every point of $V \cap S$ fixed. Since $S$ is a connected cohomology $(n - 3)$-manifold over $Z_p$ for every prime $p$, it follows from Newman's theorem [6] that $G_{x'}$ leaves every point of $S$ fixed. Hence $G_{x'} = \{g \in G_y \mid gx'' = x'' \text{ for all } x'' \in S\}$, which is clearly a normal subgroup of $G_y$. By (1.1), $G_x$ and $G_{x'}$ are conjugate so that our assertion follows.

(2.5) *Let $x \in U$. Whenever $Gz$ is 2-dimensional, there is a $g \in G$ such that $G_x \subset G_{gz}$. Hence $G_x$ is either cyclic or dihedral and it is cyclic if there is a 2-dimensional orbit which is a 2-sphere.*

For the rest of this section, we assume that

$$H_c^*(X; Z) = H^*(S^n; Z)\ .$$

Under this assumption, $H_c^0(X; Z) = H^0(S^n; Z) = Z$. Hence $X$ is compact.

(2.6) *Let $T$ be a circle group in $G$. Then $F(T, X)$ is a compact cohomology $(n - 4)$-manifold over $Z$ with $H^*(F(T, X); Z) = H^*(S^{n-4}; Z)$.*

Since $F(T, X)$ intersects every singular orbit at one or two points, $\dim_Z F(T, X) = \dim_Z B^* = n - 4$. Hence our assertion follows [1; Chapters IV and V].

(2.7) *Let $g \in G$ be of order $p^\alpha$, where $p$ is a prime and $\alpha$ is a positive integer. If $g \in G_x$ for some $x \in U \cup D$, then $F(g, X)$ is a compact cohomology $(n-2)$-manifold over $Z_p$ with $H^*(F(g, X); Z_p) = H^*(S^{n-2}; Z_p)$. Hence $F(g, X)$ intersects every principal orbit.*

It is known that $X$ is also a compact cohomology $n$-manifold over $Z_p$ with $H^*(X; Z_p) = H^*(S^n; Z_p)$. Since $G$ is connected, $g$ preserves the orientation of $X$. It follows that for some $r < n$ of the same parity, $F(g, X)$ is a compact cohomology $r$-manifold over $Z_p$ with $H^*(F(g, X); Z_p) = H^*(S^r; Z_p)$ [1; Chapters IV and V].

Let $T$ be the circle group in $G$ containing $g$. By (2.6), $F(g, X) \cap B = F(T, X)$ is a compact cohomology $(n - 4)$-manifold over $Z_p$. Since, by hypothesis, there exists a point of $U \cup D$ contained in $F(g, X)$, $F(g, X) \cap B$ is properly contained in $F(g, X)$ so that $r = n - 2$. Hence $F(g, X)$ is a compact cohomology $(n - 2)$-manifold over $Z_p$ with $H^*(F(g, X); Z_p) = H^*(S^{n-2}; Z_p)$.

Since $\dim_Z D^* < n - 3$ [1; p. 121] and since $F(g, X)$ intersects every exceptional orbit at a set of dimension $\leqq 1$, it follows that $\dim_{Z_p}(F(g, X) \cap D) \leqq \dim_Z(F(g, X) \cap D) < n - 2$. But we have $\dim_{Z_p} F(g, X) = n - 2$ and $\dim_{Z_p}(F(g, X) \cap B) = n - 4$. Therefore $F(g, X) \cap U \neq \phi$. Hence, by (1.1), $F(g, X)$ intersects every principal orbit.

(2.8) *Let $x \in U$ and $y \in D$. Let $p$ be a prime and let $\alpha$ be a positive integer. If $G_y$ has an element of order $p^\alpha$, so does $G_x$.*

Let $g \in G_y$ be of order $p^\alpha$. By (2.7), $F(g, X) \cap Gx \neq \phi$ so that for some $h \in G$, $hx \in F(g, X)$. Hence $h^{-1}gh$ is an element of $G_x$ of order $p^\alpha$.

(2.9) $D = \phi$.

Suppose that $D \neq \phi$. Let $x \in U$ and $y \in D$ be such that $G_x$ is a proper normal subgroup of $G_y$ (see (2.4)). We first claim that $G_y$ is dihedral.

It is well known that a finite subgroup of SO(3) is either cyclic or dihedral or tetrahedral or octahedral or icosahedral. If $G_y$ is cyclic, so is $G_x$. Let the order of $G_y$ be $p_1^{s_1} \cdots p_k^{s_k}$, where $p_1, \cdots p_k$ are distinct primes and $s_1, \cdots, s_k$ are positive integers. Then for every $i = 1, \cdots, k$, $G_y$ contains an element of order $p_i^{s_i}$ so that, by (2.8), $G_x$ also contains an element of order $p_i^{s_i}$. Hence $G_x$ is of order $\geqq p_1^{s_1} \cdots p_k^{s_k}$ and consequently $G_x = G_y$, contrary to the fact that $G_x$ is a proper subgroup of $G_y$. If $G_y$ is either tetrahedral or octahedral or icosahedral, then by (2.8), $G_x$ contains a subgroup of order 2 and a subgroup of order 3. In case $G_x$ is octahedral, it also contains a subgroup of order 4. Hence $G_x$, as a normal subgroup of $G_y$, is equal to $G_y$, contrary to our hypothesis. This proves that $G_y$ is dihedral.

Now the order of $G_y$ is even. It follows from (2.7) that whenever $g \in G$ is of order 2, $F(g, X)$ is a compact cohomology $(n-2)$-manifold over $Z_2$ with $H^*(F(g, X); Z_2) = H^*(S^{n-2}; Z_2)$. Let $H$ be a dihedral subgroup of $G$ of order 4. By Borel's theorem [1; p. 175], $F(H, X)$ is a compact cohomology $(n-3)$-manifold over $Z_2$ with $H^*(F(H, X); Z_2) = H^*(S^{n-3}; Z_2)$. Since $\dim_{Z_2}(F(H, X) \cap (D \cup B)) \leqq \dim_Z (F(H, X) \cap (D \cup B)) < n-3$, it follows that $F(H, X) \cap U$ is not null. Hence we may assume that $H \subset G_x \subset G_y$.

Let $T$ be the circle group in $G$ such that its normalizer contains $G_y$. Then $H \cap T \subset G_x \cap T \subset G_y \cap T$ so that $G_y \cap T$ is a cyclic group and $G_x \cap T$ is a proper subgroup of $G_y \cap T$ of even order. Let the order of $G_y \cap T$ be $2^{s_0} p_1^{s_1} \cdots p_k^{s_k}$, where $p_1, \cdots, p_k$ are distinct odd primes and $s_0, s_1, \cdots, s_k$ are positive integers. By (2.8), there are $k+1$ elements $g_0, g_1, \cdots, g_k$ of $G_x$ of order $2^{s_0}, p_1^{s_1}, \cdots, p_k^{s_k}$ respectively. Since $p_1, \cdots, p_k$ are odd, $g_1 \cdots, g_k$ are in $G_x \cap T$. Therefore no element of $G_x \cap T$ is of order $2^{s_0}$. But this implies that $s_0 > 1$ so that $g_0 \in G_x \cap T$. Hence we have arrived at a contradiction.

3. Case that the 2-dimensional orbits are all projective planes.

Let $X$ be a compact cohomology $n$-manifold over $Z$ with $H^*(X; Z) = H^*(S^n; Z)$ and let $G = \mathrm{SO}(3)$ act *nontrivially* on $X$ with $\dim_Z B = n-2$. Throughout this section, we assume that for some $x \in U$, $G_x$ is of even order.

(3.1) *Let $H$ be a dihedral subgroup of $G$ of order 4 and let $M$ be the normalizer of $H$ that is the octahedral group containing $H$. Then $F(H, X)$ is a compact cohomology $(n-3)$-manifold over $Z_2$ with $H^*(F(H, X); Z_2) = H^*(S^{n-3}; Z_2)$ and $K = M/H$ is isomorphic to the symmetric group of three elements and acts on $F(H, X)$. Moreover, the natural map of $F(H, X)/K$ into $X^*$ is onto.*

By (2.7), for every $g \in G$ of order 2, $F(g, X)$ is a compact cohomology $(n-2)$-manifold over $Z_2$ with $H^*(F(g, X); Z_2) = H^*(S^{n-2}; Z_2)$. It follows from Borel's theorem [1; p. 175] that $F(H, X)$ is a compact cohomology $(n-3)$-manifold over $Z_2$ with $H^*(F(H, X); Z_2) = H^*(S^{n-3}; Z_2)$.

Clearly $K = M/H$ is isomorphic to the symmetric group of three elements and the action of $M$ on $F(H, X)$ induces an action of $K$ on $F(H, X)$. Moreover, there is a natural map $f: F(H, X)/K \to X^*$.

Let $z \in F(H, X) \cap B$. If $Gz = z$, then $F(H, X) \cap Gz = z$. If $Gz$ is 2-dimensional, then $G_z$ contains $H$ so that by (2.3) it is the normalizer of a circle group. Therefore any two isomorphic dihedral subgroups of

$G_z$ are conjugate in $G_z$. Let $g$ be an element of $G$ with $gz \in F(H, X)$. It is clear that $g^{-1}Hg \subset g^{-1}G_{gz}g = G_z$ so that for some $h \in G_z$, $h^{-1}g^{-1}Hgh = H$ or $gh \in M$. Hence $gz = ghz \in Mz$. This proves that $F(H, X) \cap Gz \subset Mz$.

From these results it follows that $F(H, X)$ intersects every singular orbit at a finite set. [This and one or two facts mentioned below can be seen by examining the standard action of SO(3) on $S^2$ or on $P^2$ (viewed as the acts of lines through the region in $E^3$).] Therefore, by (1.2), $\dim_Z (F(H, X) \cap B) \leqq \dim_Z B^* < n - 3$. As a consequence of this result and that $D = \phi$ (see (2.9)), we have $F(H, X) \cap U \neq \phi$. Hence $F(H, X)$ intersects every principal orbit and consequently it intersects every orbit. This proves that the natural map $f: F(H, X)/K \to X^*$ is onto.

(3.2) *Every 2-dimensional orbit is a projective plane and intersects $F(H, X)$ at exactly three points.*

Let $Gz$ be a 2-dimensional orbit. By (3.1), $F(H, X)$ intersects $Gz$ so that we may assume that $z \in F(H, X)$. Since $G_z$ contains $H$, it follows from (2.3) that $G_z$ is the normalizer of a circle group. Hence $Gz$ is a projective plane.

In the proof of (3.1) we have shown that $F(H, X) \cap Gz \subset Mz$. But it is clear that $Mz \subset F(H, X) \cap Gz$. Hence

$$F(H, X) \cap Gz = Mz = M/(M \cap G_z) .$$

Since $M$ is of order 24 and $M \cap G_z$ is of order 8, it follows that $F(H, X) \cap Gz$ contains exactly three points.

(3.3) *$B^*$ is a compact cohomology $(n-4)$-manifold over $Z$ with $H^*(B^*; Z) = H^*(S^{n-4}; Z)$.*

Let $T$ be a circle group in $G$. It is clear that $F(T, X) \subset B$. Since, by (2.1) and (3.2), every singular orbit is either a point or a projective plane, it follows that $F(T, X)$ intersects every singular orbit at exactly one point. Therefore the natural projection $\pi$ maps $F(T, X)$ homeomorphically onto $B^*$ and hence our assertion follows from (2.6).

(3.4) *Let $Y = F(H, X) - F(G, X)$. Then $\bar{Y} = F(H, X)$ and every point of $Y$ has a neighborhood $V$ in $Y$ which is a cohomology $(n-3)$-manifold over $Z$ and such that the isotropy group is constant on $V - B$.*

Let $T$ be a circle group whose normalizer $N$ contains $H$. Then $F(H, X) \supset F(N, X) = F(T, X) \supset F(G, X)$. Since $F(H, X)$ is a compact $(n-3)$-manifold over $Z_2$ (see (3.1)) and since $F(T, X)$ is a compact $(n-4)$-manifold over $Z_2$ (see (2.6)), it follows that the closure of $F(H, X) - F(T, X)$ is $F(H, X)$. Hence $\bar{Y} = F(H, X)$.

Let $x \in Y \cap U$ and let $S$ be a slice at $x$. Then $S$ is a cohomology $(n-3)$-manifold over $Z$. Moreover, $G_y = G_x$ for all $y \in S$ so that $S \subset Y$. Since both $S$ and $Y$ are cohomology $(n-3)$-manifolds over $Z_2$, it follows that $S$ is open in $Y$. Hence our assertion follows by taking $S$ as $V$.

Let $z \in Y \cap B$ and let $S$ be a slice at $z$. Then $S$ is a cohomology $(n-2)$-manifold over $Z$ and $G_z$ is the normalizer of a circle group $T$ acting on $S$. Whenever $x \in S \cap U$, $G_x \cap T$ is a finite cyclic group in $T$ and the index of $G_x \cap T$ in $G_x$ is 2 because $G_x$ in a dihedral subgroup of $G_z$. Since the order of $G_x$ is independent of $x \in S \cap U$, so is the order of $G_x \cap T$. Hence $G_x \cap T$ is independent of $x \in S \cap U$ so that for $x \in F(H, S) \cap U$.

$$G_xS = H(G_x \cap T)S = HS = S$$

and

$$F(G_x, S) = F(G_x/(G_x \cap T), S) = F(H/(H \cap T), S) = F(H, S) .$$

Let $Q$ be a neighborhood of the identity of $G$ such that $Q^{-1}TQ \cap G_z = T$. If $gy \in F(H, X)$ with $g \in Q$ and $y \in S$, then $g^{-1}Hg \subset g^{-1}G_{gy}g = G_y \subset G_z$ so that $g^{-1}(H \cap T)g \subset Q^{-1}TQ \cap G_z = T$. Therefore $g^{-1}Tg = T$ or $g \in G_z$. Hence $gy \in G_zy \subset S$. This proves that $F(H, S) = F(H, X) \cap S = F(H, X) \cap QS$ is open in $F(H, X)$ so that it is a cohomology $(n-3)$-manifold over $Z_2$.

Since $S$ is a cohomology $(n-2)$-manifold over $Z$ with

$$F(H/(H \cap T), S) = F(H, S) ,$$

it follows that $F(H, S)$ is also a cohomology $(n-3)$-manifold over $Z$. (If $Z_2$ acts on a cohomology $m$ manifold over $Z$ with $F(Z_2)$ being a cohomology $(m-1)$-manifold over $Z_2$, then $F(Z_2)$ is also a cohomology $(m-1)$-manifold over $Z$.) That $G_x$ is constant on $F(H, S) \cap U$ is a direct consequence of the fact that $F(G_x, S) = F(H, S)$ for all $x \in F(H, S) \cap U$.

(3.5) *$Y$ is a connected cohomology $(n-3)$-manifold over $Z$ and the isotropy group is constant on $Y - B$.*

By (3.4), $Y$ is a cohomology $(n-3)$-manifold over $Z$. Let $T$ be a circle group in $G$ whose normalizer $N$ contains $H$. Then $F(H, X) \supset F(N, X) = F(T, X) \supset F(G, X)$. From (2.6) and (3.1), it is easily seen that $F(H, X) - F(T, X)$ has exactly two components with $F(T, X)$ as their common boundary. By (2.3), there exists a point $z$ of $F(T, X)$ such that $Gz$ is a projective plane so that $z \in F(T, X) - F(G, X)$. Hence $Y$ is connected.

Let $x \in Y \cap U$. Then $F(G_x, X) \cap Y$ is clearly closed in $Y$. But, by (3.4), it is also open in $Y$. Hence, by the connectedness of $Y$, $F(G_x, X) \cap Y = Y$.

(3.6) *Whenever* $x \in F(H, X) \cap U$, $G_x = H$. *Hence for every* $x \in U$, $G_x$ *is a dihedral group of order* 4.

Let $x$ be a point of $F(H, X) \cap U$. Since $H \subset G_x$, $F(H, X) \supset F(G_x, X)$. But, by (3.4) and (3.5), $F(H, X) \subset F(G_x, X)$. Hence $F(H, X) = F(G_x, X)$.

It is clear that $G' = \{g \in G \mid gF(H, X) = F(H, X)\}$ is a closed subgroup of $G$ containing $M$. Since $F(H, X) = F(G_x, X)$, $G_x$ is a normal subgroup of $G'$ so that $G'$ is contained in the normalizer of $G_x$. But, by (2.5), $G_x$ is dihedral and $H$ is the only dihedral group whose normalizer contains $M$. It follows that $G_x = H$. Hence, by (1.1), the isotropy group at any point of $U$ is a dihedral group of order 4.

(3.7) *Whenever* $x \in F(H, X)$, $F(H, X) \cap Gx = Kx$ *which contains one point or three points or six points according as* $Gx$ *is* 0-*dimensional or* 2-*dimensional or* 3-*dimensional.*

If $Gx$ is 0-dimensional, it is clear that $F(H, X) \cap Gx = x = Kx$. If $Gx$ is 2-dimensional, we have shown in the proof of (3.2) that $F(H, X) \cap Gx = Mx = Kx$ which contains exactly three points.

Now let $Gx$ be 3-dimensional. If $g$ is an element of $G$ with $gx \in F(H, X)$, then, by (3.6), $gHg^{-1} = gG_xg^{-1} = G_{gx} = H$ so that $g \in M$. Therefore $F(H, X) \cap Gx \subset Mx$. But it is obvious that $Mx \subset F(H, X) \cap Gx$. Hence

$$F(H, X) \cap Gx = Mx = Kx$$

which clearly contains six points.

From this result, it is easily seen that the natural map $f$: $F(H, X)/K \to X^*$ is a homeomorphism onto.

(3.8) *Whenever* $a \in K$ *is of order* 2, *we abbreviate* $F(a, F(H, X))$ *by* $F(a)$. *Then* $F(a) \subset B$ *and* $F(a)$ *is a compact cohomology* $(n-4)$-*manifold over* $Z$ *with* $H^*(F(a); Z) = H^*(S^{n-4}; Z)$. *Moreover,* $F(H, X) - F(a)$ *contains exactly two components* $V$ *and* $V'$ *with* $aV = V'$.

Whenever $x \in F(H, X) \cap U$, $G_x = H$ (see (3.6)) so that $x \notin F(a)$. Hence $F(a) \subset B$. Let $a = a'H$ with $a'$ being of order 4 and let $T$ be the circle group containing $a'$. Then $F(a) = F(T, X)$ and hence the first part follows from (2.6). Now $F(H, X)$ is a compact cohomology $(n-3)$-manifold over $Z_2$ with $H^*(F(H, X); Z_2) = H^*(S^{n-3}; Z_2)$ and $F(a) = F(a, F(H, X))$ is a compact cohomology $(n-4)$-manifold over $Z_2$. The second part follows.

(3.9) $F(H, X) - B$ *contains exactly six components and whenever* $P$ *is a component of* $F(H, X) - B$, $KP = F(H, X) - B$ *and the natural* *projection $\pi$ maps $P$ homeomorphically onto $U^*$.*

Let $P$ be a component of $F(H, X) - B$. Since the isotropy group is constant on $P$ (see (3.5)), the natural projection $\pi$ defines a local homeomorphism $\pi'$: $P \to U^*$. By (3.7), for every $x^* \in U^*$, $\pi'^{-1}x^*$ contains no more than six points. We infer that $\pi'$ is closed so that $\pi' P$ is both open and closed in $U^*$. Hence, by the connectedness of $U^*$, $\pi' P = U^*$.

Let $Q$ be a second component of $F(H, X) - B$ and let $y \in Q$. Then there is a point $x \in P$ such that $\pi x = \pi y$. Therefore, by (3.7), for some $k \in K$, $y = kx$ so that $Q = kP$. Hence $KP = F(H, X) - B$.

Let $x \in P$. By (3.8), $x$ and $ax$ belong to different components of $F(H, X) - F(a) \supset F(H, X) - B$. Therefore $aP$ is a component of $F(H, X) - B$ different from $P$. Similarly, $bP$ and $cP$ are components of $F(H, X) - B$ different from $P$.

If $aP, bP$ and $cP$ are not distinct, say $bP = cP$, then $\{k \in K \mid kP = P\}$ is of order 3 so that $P$ and $aP = bP = cP$ are the only two components of $F(H, X) - B$. Now $F(H, Z) - B = F(H, Z) - (F(a) \cup F(b) \cup F(c))$ and $F(a)$, $F(b)$, $F(c)$ are manifold over $Z$ of dimension one less than the dimension of $F(H)$. Hence $F(H, X) \cap B = F(a) \cap F(b) \cap F(c) = F(G, X)$. This is impossible, because the intersection of $F(H, X)$ and a 2-dimensional orbit is contained in $B$ but not contained in $F(G, X)$. From this result it follows that $P$, $aP$, $bP$, $cP$ are distinct components of $F(H, X) - B$. Hence $P$, $aP$, $bP$, $cP$, $bcP$, $cbP$ are all the distinct components of $F(H, X) - B$.

Now it is clear that for every $x^* \in U^*$, $\pi'^{-1}x^*$ contains exactly one point. Hence $\pi'$ is a homeomorphism.

(3.10) *Let $P$ be a component of $F(H, X) - B$. Then the map of $G/H \times P$ onto $U$ defined by $(gH, x) \to gx$ is a homeomorphsim onto. Hence $U$ is homeomorphic to the topological product of a principal orbit and $U^*$.*

This is an immediate consequence of (3.5) and (3.9).

(3.11) *The closure of $F(a) - F(G, X)$ is equal to $F(a)$. Hence* $\dim_{Z_2} F(G, X) \leqq \dim_Z F(G, X) \leqq n - 5$.

Suppose that the closure of $F(a) - F(G, X)$ is not equal to $F(a)$. Then there is a point $z$ of $F(G, X)$ and a neighborhood $A$ of $z$ such that $A \cap F(a) = A \cap F(G, X)$. Since $A \cap F(G, X) \subset F(b)$ and since, by (3.8), both $A \cap F(G, X)$ and $F(b)$ are cohomology $(n-4)$-manifolds over $Z$, $A \cap F(G, X)$ is open in $F(b)$ so that we may assume that $A \cap F(G, X) = A \cap F(b)$. Similarly, we may assume that $A \cap F(G, X) = A \cap F(c)$. Hence $A \cap F(G, X) = A \cap F(H, X) \cap B$. By (3.1) and (3.8), we may also assume that $KA = A$ and $A \cap (F(H, X) - F(a))$ contains exactly two components $Q$ and $Q'$. Now both $Q$ and $Q'$ are contained in $F(H, X) - B$ and $aQ = bQ = Q'$ Therefore $abQ = Q$ so that $ab$ maps the component of $F(H, X) - B$ containing $Q$ into itself, contrary to (3.9).

Since, by (3.8), $F(a)$ is a cohomology $(n-4)$-manifold over $Z$ and since $F(G, X)$ is nowhere dense in $F(a)$, it follows that $\dim_{Z_2} F(G, X) \leqq \dim_Z F(G, X) \leqq n-5$.

(3.12) *If $n = 4$, then $F(G, X)$ is null.*

This is a direct consequence of (3.11).

(3.13) *Let $T$ be a circle group in $G$, let $N$ be the normalizer of $T$ and let $A$ be an orbit. If $A$ is a projective plane, then $A/T$ is an arc and $N/T$ acts trivially on $A/T$ so that $F(N/T, A/T) = A/T = A/N$. If $A$ is 3-dimensional, then $A/T$ is a 2-sphere and $A/N$ is a closed 2-cell so that $F(N/T, A/T)$ is a circle.*

If $A$ is a projective plane, it is clear that $A/T$ is an arc and $N/T$ acts trivially on $A/T$. Therefore $A/N = A/T = F(N/T, A/T)$.

Now let $A$ be 3-dimensional. By (3.6), we may let $A = G/H = \{gH \mid g \in G\}$. Therefore $A/T$ is the double coset space $(G/H)/T$ and $(G/T)/H$ are homeomorphic. Since $G/T$ is a 2-sphere and since every element of $H$ preserves the orientation of $G/T$, it follows that $(G/T)/H$ is a 2-sphere. Hence $A/T$ is a 2-sphere.

As seen in [3], the double coset space $(G/N)/H$ is a closed 2-cell. Since $A/N$ may be regarded as the double coset space $(G/H)/N$ which is homeomorphic to $(G/N)/H$, we infer that $A/N$ is a closed 2-cell.

From these results, it follows that $f(N/T, A/T)$ is a circle.

(3.14) *$X^*$ is cohomological trivial over $Z$.*

Let $N$ be the normalizer of a circle group $T$ in $G$. Then $N/T$ is a cyclic group of order 2 which acts on $X/T$ with $(X/T)/(N/T) = X^*$. Since, by (2.6), $H^*(F(T, X); Z) = H^*(S^{n-4}; Z)$, it follows that $H(X/T; Z) = H^*(S^{n-1}; Z)$ [1; p. 65].

By (3.13), $F(N/T, B/T) = B/T$ and for every singular orbit $A$, $A/T$ is either a single point or an arc. It follows from the Vietoris map theorem that $H^*(B/T; Z) = H^*(B^*; Z) = H^*(S^{n-4}; Z)$ (see (3.3)). By (3.10) and (3.13), $F(N/T, U/T)$ is homeomorphic to the topological product of a circle and $U^*$ so that $H^{n-2}(F(N/T, U/T); Z) \neq 0$. Therefore $H^*(F(N/T, X/T); Z) = H^*(S^{n-2}; Z)$. Hence $H^*(X/N; Z) = 0$. By (3.13), for every orbit $A$, $A/N$ is either a single point or an arc or a closed 2-cell. It follows from the Vietoris map theorem that $H^*(X^*; Z) = H^*(X/N; Z) = 0$.

(3.15)
$$H_c^k(U^*; Z_2) = \begin{cases} Z_2 & \textit{for } k = n-3\,; \\ 0 & \textit{otherwise}\,. \end{cases}$$

This follows from (3.3), (3.14) and the cohomology sequence of $(X^*, B^*)$.

(3.16)
$$H_c^k(U; Z_2) = \begin{cases} Z_2 & \textit{for } k = n-3,\ n\,; \\ Z_2 \oplus Z_2 & \textit{for } k = n-2,\ n-1\,; \\ 0 & \textit{otherwise}\,. \end{cases}$$

Since for a principal orbit $A$, we have

$$H^k(A; Z_2) = \begin{cases} Z_2 & \text{for } k = 0, 3\,; \\ Z_2 \oplus Z_2 & \text{for } k = 1, 2\,; \\ 0 & \text{otherwise}\,, \end{cases}$$

our assertion follows from (3.10) and (3.15).

As a consequence of (3.16) and the cohomology sequence of $(X, B)$, we have

(3.17)
$$H^k(B; Z_2) = \begin{cases} Z_2 & \textit{for } k = 0,\ n-4\,; \\ Z_2 \oplus Z_2 & \textit{for } k = n-3,\ n-2\,; \\ 0 & \textit{otherwise}\,. \end{cases}$$

(3.18) *Let $T$ be a circle group in $G$ and let $n \geqq 5$. Then*

$$H_c^k(F(T, X) - F(G, X); Z_2) = \begin{cases} \widetilde{H}^{k-1}(F(G, X); Z_2) \ (\textit{the reduced group}) & \\ \qquad\qquad\qquad\qquad \textit{for } k = 1\,; & \\ H^{k-1}(F(G, X); Z_2) \oplus Z_2 & \textit{for } k = n-4\,; \\ H^{k-1}(F(G, X); Z_2) & \textit{otherwise}\,. \end{cases}$$

This follows from (2.6) and the cohomology sequence of $(F(T, X), F(G, X))$.

(3.19) *Let $n > 5$. Then*

$$H_c^k(B - F(G, X); Z_2) \begin{cases} H^k(B; Z_2) & \textit{for } k > n-4\,; \\ H^k(B; Z_2) \oplus H^{k-1}(F(G, X); Z_2) & \\ & \textit{for } k = n-4\,; \\ H^{k-1}(F(G, X); Z_2) & \\ & \textit{for } k = 2, \cdots, n-5\,; \\ \widetilde{H}^{k-1}(F(G, X); Z_2) & \textit{for } k = 1\,. \end{cases}$$

This follows from the cohomology sequence of $(B, F(G, X))$.

(3.20) *$B - F(G, X)$ is homeomorphic to the topological product of a projective plane and $F(T, X) - F(G, X)$. Hence*

$$\begin{aligned} &H_c^k(B - F(G, X); Z_2) \\ &\quad = H_c^k(F(T, X) - F(G, X); Z_2) \oplus H_c^{k-1}(F(T, X) - F(G, X); Z_2) \\ &\qquad\qquad\qquad \oplus H_c^{k-2}(F(T, X) - F(G, X); Z_2) \,. \end{aligned}$$

The first part follows from the that $F(T, X) - F(G, X)$ is a cross-section of the transformation group $(G, B - F(G, X))$ on which the isotropy group is constant. The second part follows from the first part and the fact that if $A$ is a projective plane, then

$$H^k(A; Z_2) = \begin{cases} Z_2 & \text{for } k = 0, 1, 2; \\ 0 & \text{otherwise.} \end{cases}$$

(3.21) $\dim_{Z_2} F(G, X) = n - 5$. *If* $n = 4$, *then* $B$ *contains exactly two projective planes. If* $n = 5$, *then* $F(G, X)$ *contains exactly two points. If* $n > 5$, *then* $H^{n-5}(F(G, X); Z_2) = Z_2$ *so that* $F(G, X)$ *is not null.*

Setting $k = n - 2$ in (3.20), we have, by (2.6) and (3.17),

$$Z_2 \oplus Z_2 = H_c^{n-4}(F(T, X) - F(G, X); Z_2) \,.$$

If $n = 4$, then, by (3.12), $H^\circ(F(T, X); Z_2) = Z_2 \oplus Z_2$ so that $F(T, X)$ contains exactly two points. Hence $B$ contains exactly two projective planes.

If $n = 5$, then $H_c^1(F(T, X) - F(G, X); Z_2) = \tilde{H}^\circ(F(G, X); Z_2) \oplus H^1(F(T, X); Z_2)$ so that $\tilde{H}^\circ(F(G, X); Z_2) = Z_2$. Hence $F(G, X)$ contains exactly two points.

If $n > 5$, it follows from (3.18) that $H^{n-5}(F(G, X); Z_2) = Z_2$. Hence $F(G, X)$ is not null.

(3.22) $H^*(F(G, X); Z_2) = H^*(S^{n-5}; Z_2)$.

For $n = 4$ and 5, the result has been shown in (3.12) and (3.21). For $n > 5$, our assertion follows from (3.18), (3.19), (3.20) and (3.21).

(3.23) $F(G, X)$ *is a compact cohomology* $(n - 5)$*-manifold over* $Z_2$.

To prove (3.23), we have only to localize the preceding computations. Details are omitted.

REMARK. There is no difficulty to use $Z$ in place of $Z_2$ in these computations. However, the computations over $Z$ will not strengthen our final results (3.22) and (3.23).

4. Case that the 2-dimensional orbits are all 2-spheres.

Let $X$ be a compact cohomology $n$-manifold over $Z$ with $H^*(X; Z) = H^*(S^n; Z)$ and let $G = \mathrm{SO}(3)$ act nontrivially on $X$ with $\dim_Z B = n - 2$.

Throughout this section, we assume that for some $x \in U$, $G_x$ is of odd order.

(4.1) *Let $H$ be a dihedral subgroup of $G$ of order 4. Then $F(H, X)$ is a compact cohomology $(n-6)$-manifold over $Z_2$ with $H^*(F(H, X); Z_2) = H^*(S^{n-6}; Z_2)$. Hence $n \geqq 5$.*

Let $g \in G$ be of order 2 and let $T$ be the circle group in $G$ containing $g$. Since for some $x \in U$, $G_x$ is of odd order, $F(g, X) \subset B$ so that $F(g, X) = F(T, X)$ is a compact cohomology $(n-4)$-manifold over $Z_2$ with $H^*(F(g, X); Z_2) = H^*(S^{n-4}; Z_2)$. By Borel's theorem [1; p. 175], $F(H, X)$ is a compact cohomology $(n-6)$-manifold over $Z_2$ with $H^*(F(H, X); Z_2) = H^*(S^{n-6}; Z_2)$. From this result it follows that $n-6 \geqq -1$. Hence $n \geqq 5$.

(4.2) *The 2-dimensional orbit are all 2-spheres.*

Suppose that this assertion is false. Then there is, by (2.3), a projective plane $Gz$. Denote by $T$ the identity component of $G_z$ and by $H$ a dihedral subgroup of $G_z$ of order 4. Let $S$ be a connected slice at $z$. Then $S$ is a cohomology $(n-2)$-manifold over $Z$ and $G_z$ acts on $S$. Moreover, $F(T, S) = F(T, X) \cap S$ is open in $F(T, X)$ so that it is a cohomology $(n-4)$-manifold over $Z$. Hence we may let $S$ be so chosen that $F(T, S)$ is connected and that both $S$ and $F(T, S)$ are orientable.

Since $T$ is a circle group and since $\dim_Z S - \dim_Z F(T, S) = 2$, it follows that $S/T$ is a connected cohomology $(n-3)$-manifold over $Z$ with boundary $F(T, S)$ [1; p. 196]. Hence we have a connected cohomology $(n-3)$-manifold $Y$ over $Z$ obtained by doubling $S/T$ on $F(T, S)$ [1; p. 196]. Since $S$ is orientable, so is $S/T - F(T, S)$. It follows from the connectedness of $F(T, S)$ that $Y$ is orientable.

It is clear that $K = G_z/T$ is a cyclic group of order 2 which acts on $S/T$ with $KF(T, S) = F(T, S)$. Since $F(K, F(T, S)) = F(H, S)$ is a cohomology $(n-6)$-manifold over $Z_2$, we infer from the dimensional parity that $K$ preserves the orientation of $F(T, S)$ [1; p. 79].

The action of $K$ on $S/T$ defines a natural action of $K$ on $Y$ which also preserves the orientation of $Y$. Hence $\dim_{Z_2} F(K, Y) > n-6$ so that for some $y^* = Ty \in S/T - F(T, S)$, $Ky^* = y^*$. But this implies that $G_z y = Ty$ so that $y$ is a point of $D$, contrary to (2.9). Hence (4.2) is proved.

(4.3) *$F(G, X)$ is a compact cohomology $(n-6)$-manifold over $Z_2$ with $H^*(F(G, X); Z_2) = H^*(S^{n-6}; Z_2)$.*

By (4.2), $F(G, X) = F(H, X)$. Hence our assertion follows from (4.1).

(4.4) *Whenever $x \in U$, $G_x$ is the identity group.*

If $X$ is strongly paracompact, the result can be found in [5]. But an unpublished result of Yang shows that it is true in general.

(4.5) *$B^*$ is a compact cohomology $(n-4)$-manifold over $Z$ with $H^*(B^*; Z) = H^*(S^{n-4}; Z)$.*

*Proof.* Let $T$ be a circle group in $G$ and $N$ its normalizer. Then $F(T, X)$ is a compact cohomology $(n-4)$-manifold over $Z$ with $H^*(F(T, X); Z) = H^*(S^{n-4}; Z)$ and $N/T$ is a cyclic group of order 2 acting on $F(T, X)$ with $F(T, X)/(N/T) = B^*$. Therefore $H^*(B^*; Z)$ is finitely generated [1; p. 44]. If $H$ is a dihedral subgroup of $N$ of order 4, it is easily seen that $F(N/T, F(T, X)) = F(H, X)$ so that $F(N/T, F(T, X))$ is a compact cohomology $(n-6)$-manifold over $Z_2$ with $H^*(F(N/T, F(T, X)); Z_2) = H^*(S^{n-6}; Z_2)$. Hence, by the dimensional parity theorem, $N/T$ preserves the orientation of $F(T, X)$.

By [1; pp. 63–64],

$$H^*(B^*; Z_2) = H^*(F(T, X)/(N/T); Z_2) = H^*(S^{n-4}; Z_2)\ .$$

We now use the following diagram from [1; p. 45]

$$\begin{tikzcd} \cdots \arrow[r] & H^k(B^*; Z) \arrow[r, "2"] \arrow[dr, "\pi^*"] & H^k(B^*; Z) \arrow[r, "q"] & H^k(B^*; Z_2) \arrow[r] & \cdots \\ & & H^k(F(T, X); Z) \arrow[u, "\mu"] & & \end{tikzcd}$$

in which the horizontal sequence is exact and the triangle is commutative. For $k \neq 0,\ n-4$, we have $H^k(B^*; Z_2) = 0$ and $H^k(F(T, X); Z) = 0$; hence $H^k(B^*; Z) = 0$. For $k = 0$, we have $H^0(B^*; Z) = Z$, because $B^*$ is clearly connected. For $k = n-4$, $H^{n-4}(B^*; Z)$ is a finitely generated group with $H^{n-4}(B^*; Z) \otimes Z_2 = H^{n-4}(B^*; Z_2) = Z_2$. It follows from the universal coefficient theorem that there is a finite subgroup $K$ of $H^{n-4}(B^*; Z)$ of odd order such that $H^{n-4}(B^*; Z)/K$ is $Z$ or $Z_2$. Since $K = 2K = \mu\pi^*K = 0$, $H^{n-4}(B^*; Z) = Z$ or $Z_2$. But $H^{n-4}(B^*; Z) \neq Z_2$, because $N/T$ preserves the orientation of $F(T, X)$. Hence $H^{n-4}(B^*; Z) = Z$.

By localizing this result, we can show that $B^*$ is a cohomology $(n-4)$-manifold over $Z$ near every point of $F(G, X)$. (This result is also shown in [2].) Since the projection of $F(T, X) - F(G, X)$ onto $B^* - F(G, X)$ is a local homeomorphism, $B^*$ is a cohomology $(n-4)$-manifold over $Z$ near every point of $B^* - F(G, X)$. Hence $B^*$ is a compact cohomology $(n-4)$-manifold over $Z$.

(4.6) *Let $T$ be a circle group in $G$ and let $N$ be the normalizer of $T$. Then $H^*(B/N; Z) = H^*(S^{n-4}; Z)$.*

Let $A$ be a singular orbit. If $A$ is a single point, so is $A/N$. If $A$

is a 2-sphere, we may let $A = G/T$. Therefore $A/N = (G/T)/N$ is homeomorphic to $(G/N)/T$ which is known to be a closed 2-cell [3]. Hence $A/N$ is a closed 2-cell.

Since, by (2.1) and (4.2), every singular orbit is either a single point or a 2-sphere, it follows from Vietoris map theorem that $H^*(B/N; Z) = H^*(B^*; Z)$. Hence our assertion follows from (4.5).

(4.7) $$H^k(X/N; Z) = \begin{cases} Z & \textit{for } k = 0 \text{ ;} \\ Z_2 & \textit{for } k = n-1 \text{ ;} \\ 0 & \textit{otherwise.} \end{cases}$$

Since $H^*(F(T, X); Z) = H^*(S^{n-4}; Z)$, it follows that $H^*(X/T; Z) = H^*(S^{n-1}; Z)$. Now $N/T$ is a cyclic group of order 2 acting on $X/T$ with $(X/T)/(N/T) = X/N$.

Let $A$ be an orbit. If $A$ is 3-dimensional, then, by (4.4), $A/T$ is a 2-sphere and $N/T$ acts freely on $A/T$. If $A$ is a 2-sphere, then $A/T$ is an arc and $F(N/T, A/T)$ is a single point. If $A$ is a point, then $F(N/T, A/T) = A/T = A$. Hence $F(N/T, X/T)$ is homeomorphic to $B^*$ so that, by (4.5), $H^*(F(N/T, X/T); Z_2)$.

As in the proof of (4.5), we can show that

(4.8) $$H_c^k(U/N; Z) = \begin{cases} Z & \textit{for } k = n-3 \text{ ,} \\ Z_2 & \textit{for } k = n-1 \text{ ,} \\ 0 & \textit{otherwise.} \end{cases}$$

(4.9) *There is an exact sequence*

$$\cdots \to H_c^{k-3}(U^*; Z_2) \to H_c^k(U^*; Z) \to H_c^k(U/N; Z) \to H_c^{k-2}(U^*; Z_2) \to \cdots .$$

By (4.4), $G$ acts freely on $U$. Hence we have the desired exact sequence as seen in [3].

(4.10) $$H_c^k(U^*; Z) = \begin{cases} Z & \textit{for } k = n-3 \text{ ,} \\ 0 & \textit{otherwise.} \end{cases}$$

Since $\dim_Z U^* = n-3$, we have

$$H_c^k(U^*; Z) = 0 \quad \text{for } k > n-3 .$$

It follows from (4.9) and (4.8) that $H_c^{n-3}(U^*; Z_2) = H_c^{n-1}(U/N; Z) = Z_2$. From (4.9), it is easily seen that $H_c^{n-3}(U^*; Z) = Z \oplus I$, where $I = im(H_c^{n-6}(U^*; Z_2) \to H_c^{n-3}(U^*; Z))$ so that every element of $I$ different from 0 is of order 2. By the universal coefficient theorem,

$$\begin{aligned} Z_2 = H_c^{n-3}(U^*; Z_2) &= H_c^{n-3}(U^*; Z) \otimes Z_2 \oplus \mathrm{Tor}(H^{n-2}(U^*; Z), Z_2) \\ &= Z_2 \oplus I . \end{aligned}$$

Hence $I = 0$, proving that

$$H_c^{n-3}(U^*; Z) = Z \ .$$

If $k < n-3$, then by (4.8) and (4.9), $H_c^k(U^*; Z) = H_c^{k-3}(U^*; Z_2)$. Hence for $k < n-3$,

$$H_c^k(U^*; Z) = 0 \ .$$

(4.11) *$X^*$ is cohomologically trivial over $Z$.*

This is an easy consequence of (4.5), (4.10) and the cohomology sequence of $(X^*, B^*)$.

## REFERENCES

1. A. Borel *et al.*, *Seminar on transformation groups,* Annals of Math., Studies, No. 46, Princeton University Press, 1960.
2. G. E. Bredon, *On the structure of orbit spaces of generalized manifolds,* (to appear).
3. P. E. Conner and E. E. Floyd, *A note on the action of SO*(3), Proc. Amer. Math. Soc., **10** (1959), 616-620.
4. D. Montgomery and L. Zippin, Topological transformation groups, Interscience Publishers, Inc., 1955.
5. D. Montgomery and H. Samelson, *On the action of SO*(3) *on* $S^n$, Pacific J. Math., **12** (1962), 649-659.
6. P. A. Smith, *Transformations of finite period III, Newman's theorem,* Ann. of Math. (2), **42** (1941), 446-458.
7. C. T. Yang, *Transformation groups on a homological manifold,* Trans. Amer. Math. Soc., **87** (1958), 261-283.

INSTITUTE FOR ADVANCED STUDY
UNIVERSITY OF PENNSYLVANIA

# A NOTE ON ABELIAN GROUP EXTENSIONS

R. J. NUNKE

In Exercise 21 page 248 of his book Abelian Groups L. Fuchs asks for a proof of the following

THEOREM. *If $A$ is a torsion-free and $C$ a torsion group, then* $\operatorname{Ext}(A, C)$ *is either* 0 *or contains an element of infinite order.*

Unfortunately the hint given with the exercise leads only to the conclusion that every countable subgroup of $A$ is free. Professor Fuchs has informed me that he meant to assume $A$ countable. The purpose of this note is to prove this theorem.

LEMMA. *If $C_1, C_2, \cdots$ is a sequence of abelian groups, $\Pi C_i$ their direct product and $\Sigma C_i$ their direct sum, then* $\operatorname{Ext}(A, \Pi C_i/\Sigma C_i) = 0$ *for all torsion-free groups $A$.*

*Proof.* A special case of this lemma with all the $C_i = Z$ the group of integers is a consequence of Theorem 1 of [1]. The proof of the special case given in [4] makes no use of the fact that $C_i = Z$. This proof will be sketched here. It is enough to prove the case in which $A$ is the rational numbers. Since $\operatorname{Ext}(A, \Pi C_i/\Sigma C_i)$ is a homomorphic image of $\operatorname{Ext}(A, \Pi C_i)$ we must show that each extension $0 \to \Pi C_i \to E \to A \to 0$ splits over $\Pi C_i/\Sigma C_i$, i.e., that there is a map $f\colon E \to \Pi C_i/\Sigma C_i$ whose restriction to $\Pi C_i$ is the canonical projection. With $A$ the rationals we choose elements $e^1, e^2, \cdots$ in $E$ such that $e^n$ maps onto $1/n!$ modulo $\Pi C_i$. Then $E$ is generated by $\Pi C_i$ and the $e$'s with relations

$$e^n = (n+1)e^{n+1} + c^n \qquad\qquad n = 1, 2, \cdots$$

where $c^n \in \Pi C_i$. We choose $b^n \in \Sigma C_i$ such that the first $n$ coordinates of $c^n + b^n$ are 0 and put

$$x^n = \textstyle\sum_{k \geqq n}(k!/n!)(c^k + b^k) .$$

Then

$$x^n = (n+1)x^{n+1} + c^n + b^n$$

and we can define $f$ to be the projection on $\Pi C_i$ and by $f(e^n) = x^n + \Sigma C_i$.

PROPOSITION. If $C$ is the direct sum of infinitely many copies of

Received April 25, 1962. This work was supported by the National Science Foundation research grant NSF-G 11098.

$D$ and if $A$ is torsion-free with $\operatorname{Ext}(A, D) \neq 0$, then $\operatorname{Ext}(A, C)$ has an element of infinite order.

*Proof.* Since $D$ is a direct summand of $C$ we have $\operatorname{Ext}(A, C) \neq 0$. The sequence

$$\operatorname{Ext}(A, \Sigma C) \to \operatorname{Ext}(A, \Pi C) \to \operatorname{Ext}(A, \Pi C/\Sigma C) \to 0$$

is exact where $\Sigma C$ is the direct sum and $\Pi C$ the direct product of countably many copies of $C$. By the lemma $\operatorname{Ext}(A, \Pi C/\Sigma C) = 0$ so that the left-most map in the sequence is an epimorphism. Since $A$ is torsion-free $\operatorname{Ext}(A, C)$ is divisible and hence has elements of arbitrarily large finite order if it has nonzero elements of finite order at all. Hence $\operatorname{Ext}(A, \Pi C) \cong \Pi \operatorname{Ext}(A, C)$ has an element of infinite order. It follows that $\operatorname{Ext}(A, \Sigma C)$ also has an element of infinite order. Since $C$ is the direct sum of infinitely many copies of $D$ we have $\Sigma C \cong C$ so that $\operatorname{Ext}(A, \Sigma C) \cong \operatorname{Ext}(A, C)$ proving the proposition.

Now to prove the theorem we suppose that $A$ is torsion-free, $C$ is torsion and that $\operatorname{Ext}(A, C)$ is a nonzero torsion group. Then $\operatorname{Ext}(A, C)$ has a nonzero $p$-primary component for some prime $p$. Since $C = C' \oplus E$ where $C'$ is the $p$-primary component of $C$ and $E$ is the sum of the other primary components we have

$$\operatorname{Ext}(A, C) = \operatorname{Ext}(A, C') \oplus \operatorname{Ext}(A, E)\ .$$

Multiplication by $p$ is an automorphism of $E$, hence also an automorphism of $\operatorname{Ext}(A, E)$. It follows that $\operatorname{Ext}(A, C')$ is a nonzero torsion group. Hence in proving the theorem we may assume that $C$ is $p$-primary.

In [3] it was shown that, for $A$ torsion-free and $C$ $p$-primary,

$$\operatorname{Ext}(A, C) \cong \operatorname{Ext}(A, M)$$

where $M$ is a direct sum of copies of $\Sigma Z/p^n Z$, the number of copies being equal to the final rank of $C$. If $C$ has bounded order, then $\operatorname{Ext}(A, C) = 0$ for all torsion-free groups $A$. Otherwise the final rank of $C$ is infinite. This last case is the one to be considered. Then $M$ is the direct sum of countably many copies of itself and the proposition shows that $\operatorname{Ext}(A, M)$ is either 0 or has an element of infinite order.

The referee has pointed out that a stronger form of the lemma in this paper has been proved by A. Hulanicki (Bull. Acad. Pol. Sci. Ser. Sci. Math. Astr. Phys., 10 (1962), 77–80.) He showed that each element of infinite height in $\Pi C_i/\Sigma C_i$ is in the maximal divisible subgroup, hence this group is algebraically compact.

## References

1. S. Balcerzyk, *On factor groups of some subgroups of a complete direct sum of infinite cyclic groups*, Bull. Acad. Polon. Sci. Ser. Sci. Math. Astr. Phys., **7** (1959), 141–142.
2. L. Fuchs, *Abelian Groups*, Budapest, Publishing house of the Hungarian Academy of Sciences, 1958.
3. R. J. Nunke, *On the extensions of a torsion module*, Pacific J. Math., **10** (1960), 597–606.
4. ———, *Slender groups*, Acta Sci. Math. Szeged, **23** (1962), 67–73.

University of Washington

# A COMPLETE SET OF UNITARY INVARIANTS FOR OPERATORS GENERATING FINITE $W^*$-ALGEBRAS OF TYPE I

CARL PEARCY

**1. Introduction.** The principal object of this paper is to give a complete set of unitary invariants for a certain class of operators on Hilbert space. The operators considered are exactly those operators which generate a $W^*$-algebra which is finite of type $I$ in the terminology of Kaplansky [6]. Such an operator is a direct sum of homogeneous $n$-normal operators, and a homogeneous $n$-normal operator can be regarded as a continuous function from a totally disconnected topological space to the full ring of $n \times n$ complex matrices. Thus it was conjectured by Kaplansky that if one could find a suitable set of invariants for complex matrices, one could also solve the unitary equivalence problem for homogeneous $n$-normal operators, and Brown's solution [2] of the problem in the case $n = 2$ strengthened this belief. A complete set of unitary invariants for $n \times n$ matrices was furnished by Specht. In [10] he showed that there is a collection of traces attached to every $n \times n$ complex matrix such that two matrices are unitarily equivalent if and only if the corresponding traces in this collection are equal. A generalization of the trace of a matrix to $n$-normal operators is given by Diximier in [3], and it was thus natural to suppose that the generalized Specht invariants would serve for homogeneous $n$-normal operators. (See page 20, [7].)

Unfortunately, the Specht invariants have the unpleasant feature that they are infinite in number, and for $n$ fixed it seemed likely that some finite subset would serve. Herein it is shown (Theorems 1 and 2) that there is always a subset of less than $4^{n^2}$ traces which is a complete set of unitary invariants for $n \times n$ complex matrices. Furthermore, the same invariants form a set of orthogonal invariants for $n \times n$ real matrices. (One observes that Specht's proof does not generalize to the real case, due to the failure there of Burnside's theorem.)

The (local) unitary equivalence problem for homogeneous $n$-normal operators generating the same $W^*$-algebra is then considered, and it is shown that the same finite number of Dixmier traces is a (local) complete set of unitary invariants for such operators (Theorem 3). Finally the question of global unitary equivalence for operators which generate a finite $W^*$-algebra of type $I$ is considered, and a global complete set of

Received September 28, 1961, and in revised form December 20, 1961. This paper is essentially a part of the thesis submitted in partial fulfillment of the requirements for the Ph. D. degree at Rice University.

unitary invariants is determined (Theorem 5). In particular, to each such operator $A$ is attached a countable collection of mutually commuting normal operators $N_i(A)$. Then $A$ is unitarily equivalent to $B$ if and only if there is a unitary isomorphism $\varphi$ between the respective Hilbert spaces which satisfies $\varphi N_i(A)\varphi^{-1} = N_i(B)$ for all $i$.

The author wishes to express his appreciation to Professor Arlen Brown for his encouragement and patient criticism during the preparation of this paper.

## 2. $n \times n$ matrices.

We first obtain the result for $n \times n$ matrices, or what is the same thing, for operators on an $n$-dimensional (complex) Hilbert space. The reader is reminded that a ring of operators, or $W^*$-algebra, is a self-adjoint algebra of operators closed in the weak operator topology acting on a Hilbert space. $W^*$-algebras are not assumed to contain the identity operator.

Throughout this paper $W$ will denote the free multiplicative semigroup generated by the two free variables $x$ and $y$. Words in this collection are denoted by $w(x, y)$, and the collection of all words $w(x, y)$ with the property that the sum of the exponents appearing in $w(x, y)$ does not exceed $n$ is denoted by $W(n)$. Also, if $A$ is an operator, the notation $W_A(n)$ denotes the collection of all operators $w(A, A^*)$ with $w(x, y) \in W(n)$.

LEMMA 2.1. *If $A$ is an operator on an $n$-dimensional Hilbert space, and $d$ is a positive integer such that every operator in $W_A(d+1)$ is a linear combination of operators in $W_A(d)$, then $W_A(d)$ spans the $*$-algebra $V$ generated by $A$.*

*Proof.* Clearly, $V$ consists of all linear combinations of words $w(A, A^*)$ where $w(x, y) \in W$. If $W_A(d)$ does not span $V$, then there is a word $w(A, A^*)$ which is independent of $W_A(d)$ with the property that the sum of the exponents in $w(x, y)$ is a minimum. A contradiction is easily reached by factoring $A$ (or $A^*$) out of $w(A, A^*)$ and writing the other factor as a linear combination of operators in $W_A(d)$.

LEMMA 2.2. *If $A$ and $V$ are as before, then $V$ is spanned by the collection of operators $W_A(n^2)$.*

*Proof.* This follows from Lemma 2.1 and the fact that $V$ can contain at most $n^2$ linearly independent operators.

We introduce the notation $\sigma(A)$ for the trace of an operator $A$ acting on a finite dimensional space.

THEOREM 1. *If $A$ and $B$ are operators on an $n$-dimensional*

*Hilbert space* $\mathscr{H}$, *and* $\sigma[w(A, A^*)] = \sigma[w(B, B^*)]$ *for every word* $w(x, y) \in W(2n^2)$, *then* $A$ *and* $B$ *are unitarily equivalent.*

*Proof.* Let $R(A)$ and $R(B)$ be the *-algebras generated by $A$ and $B$ respectively. If $A^* = \lambda A$ for some scalar $\lambda$, then it is easy to see that $B^* = \lambda B$, so that $A$ and $B$ are normal, and the traces assumed equal are more than enough to guarantee the unitary equivalence of $A$ and $B$. Thus, we can assume that $A$ and $A^*$ are linearly independent, and it results from the preceding lemmas that there is a basis $\beta(A) = \{w_i(A, A^*)/w_i(x, y) \in \tau\}$ of $R(A)$ such that $\tau \subset W(n^2 - 1)$, $w_1(x, y) = x$ and $w_2(x, y) = y$. It follows easily from the hypothesis and the fact that $\sigma(CC^*) = 0$ implies $C = 0$ for arbitrary $C$ that $\beta(B) = \{w_i(B, B^*)/w_i(x, y) \in \tau\}$ is a basis for $R(B)$. To complete the proof, it suffices to show that if $w_j(x, y)$ is any word in $\tau$ and $w_j(A, A^*)A = \sum_i \alpha_{ij} w_i(A, A^*)$, $w_j(A, A^*)A^* = \sum_i \gamma_{ij} w_i(A, A^*)$, then $w_j(B, B^*)B = \sum_i (\alpha_{ij} w_i(B, B^*)$ and $w_j(B, B^*)B^* = \sum_i \gamma_{ij} w_i(B, B^*)$. For if this is so, then it is clear that if any word $w(A, A^*)$ is formed by multiplications of appropriate powers of $A$ by appropriate powers of $A^*$, and the corresponding word $w(B, B^*)$ is formed similarly, we will obtain $w(A, A^*) = \sum_i \delta_i w_i(A, A^*)$ and $w(B, B^*) = \sum_i \delta_i w_i(B, B^*)$. This implies that $\sigma[w(A, A^*)] = \sigma[w(B, B^*)]$, and the result will follow from the original theorem of Specht. Thus let $w_j(x, y) \in \tau$ and consider $L = w_j(B, B^*)B - \sum_i \alpha_{ij} w_i(B, B^*)$ and $N = w_j(B, B^*)B - \sum_i \gamma_{ij} w_i(B, B^*)$. Since $LL^*$ and $NN^*$ are linear combinations of words each of which is in $W_B(2n^2)$, it follows from the hypothesis that $\sigma(LL^*) = \sigma(NN^*) = 0$, so that $L = N = 0$, and the proof is complete.

It is easy to see that some of the equalities $\sigma[w(A, A^*)] = \sigma[w(B, B^*)]$, $w(x, y) \in W(2n^2)$, follow from others as a result of properties of the trace function, and thus there are smaller sets of invariants than the set indicated by Theorem 1. For example, it suffices to assume equality for words of the form $x^i$ and $x^i y^j x^k \cdots y^t$ in view of the identity $\sigma(A^*) = [\sigma(A)]^*$ and the fact that the trace of any commutator is zero. Detailed consideration of the case $n = 3$ indicated (see § 5) that it is probably not worthwhile to pursue the question of how many words can thus be dispensed with, so we content ourselves with the observation that there are more distinct sequences of positive integers each having the property that the sum of its terms is at most $2n^2$ than there are traces needed.

THEOREM 2. *There is a complete set of unitary invariants for* $n \times n$ *complex matrices containing fewer than* $4^{n^2}$ *elements.*

*Proof.* By induction, the number of distinct sequences of positive integers each having the property that the sum of its terms is a given positive integer $k$ is $2^{k-1}$, and one sums the resulting geometric series.

The following corollary extends the above result to real matrices.

COROLLARY. *Any collection of traces which is a complete set of unitary invariants for $n \times n$ complex matrices is also a complete set of orthogonal invariants for $n \times n$ real matrices, and therefore there is a complete set of orthogonal invariants containing fewer than $4^{n^2}$ elements.*

*Proof.* We can assume $A$ and $B$ are real $n \times n$ matrices with $UAU^* = B$, $U$ complex unitary. Let $U = R + iJ$ where $R$ and $J$ are real matrices. Then $RA = BR$, $JA = BJ$, $RA^* = B^*R$, $JA^* = B^*J$, and one can choose a real $\lambda$ such that $S = R + \lambda J$ is nonsingular. It follows that $SAS^{-1} = B$ and $SA^*S^{-1} = B^*$, and the usual construction yields an orthogonal real matrix $V$ such that $VAV^* = B$.

3. **Homogeneous $n$-normal operators.** Terms such as abelian projections, equivalence of projections, and homogeneity of projections are taken as defined in [6]. A $W^*$-algebra $R$ is *$n$-normal* if it satisfies the identity

$$(*) \qquad \sum \pm X_1 X_2 \cdots X_{2n} = 0$$

where the sum is taken over all permutations on $2n$ objects, and the sign is determined by the parity of the permutation. An $n$-normal algebra is *homogeneous $n$-normal* (also called type $I_n$) if the unit is homogeneous of order $n$, and an operator is (homogeneous) $n$-normal if the $W^*$-algebra it generates is (homogeneous) $n$-normal.

The imposition of (*) on an algebra $R$ restricts the number of nonzero, orthogonal, equivalent projections in $R$ to a maximum of $n$, and since every direct summand of an $n$-normal algebra contains an abelian projection [2], it follows easily that any $n$-normal algebra is a direct sum of algebras of type $I_k$ where $k \leqq n$. Kaplansky [5] and Brown [2] gave a structure theory for these algebras, and according to [2], if $R$ is a homogeneous $n$-normal algebra, then $R$ is unitarily isomorphic to the algebra of all $n \times n$ matrices with entries from an abelian $W^*$-algebra $Z'$ containing 1. By applying the representation theorem for abelian $C^*$-algebras to $Z'$, one obtains that $Z'$ is $C^*$-isomorphic to the $C^*$-algebra $C(\mathfrak{X})$ of all continuous complex-valued functions on a compact Hausdorff space $\mathfrak{X}$. Now $Z'$ is weakly closed, and it has been shown that this gives $\mathfrak{X}$ the additional properties that the closure of every open set is open, and the compact open sets form a base for the topology [11]. It results that $R$ is $C^*$-isomorphic to the $C^*$-algebra $M_n(\mathfrak{X})$ of all continuous functions from $\mathfrak{X}$ to the full ring $M_n$ of $n \times n$ complex matrices, where $\|A(\cdot)\| = \sup_{t \in \mathfrak{X}} \|A(t)\|$. If $A = (A_{ij}) \in R$, then $A$ corresponds to the function $A(\cdot) \in M_n(\mathfrak{X})$ whose value at $t \in \mathfrak{X}$ is $A(t) = (a_{ij}(t))$, where $a_{ij}(\cdot)$ is the function in $C(\mathfrak{X})$ corresponding to $A_{ij}$ in $Z'$. See [2] for details.

It will be useful from here on to have a notation for a diagonal matrix which has the same entry $E$ in every position on the main diagonal and zeros elsewhere. We hereby adopt the notation $\mathrm{Diag}(E)$ for this matrix whose size will always be clear from context.

Dixmier [3] has demonstrated the existence of a unique center valued trace-like function (called by him *"l'application $\natural$ canonique"*) defined on finite $W^*$-algebras. This function, which we denote by $D(\cdot)$, is linear, a unitary invariant, constant on the center of the algebra, preserves the $^*$-operation, and has the property that if $A_\lambda$ is a net of uniformly bounded operators converging weakly to $A$, then $D(A_\lambda)$ converges weakly to $D(A)$. For more information concerning this function, see [3].

Our intention is to use operators of the form $D[w(A, A^*)]$ as unitary invariants for operators $A$ generating finite $W^*$-algebras of type $I$. To this end, let $R$ be a homogeneous $n$-normal $W^*$-algebra. Then as mentioned, we can take $R$ to be the $W^*$-algebra of all $n \times n$ matrices over an abelian algebra $Z'$, and any $A \in R$ has the form $A = (A_{ij})$. Thus one can define a mapping $A \rightarrow \mathrm{Diag}(1/n \sum_i A_{ii})$ from $R$ to the center of $R$, and it is not hard to see that this mapping has all of the afore mentioned properties of $D(\cdot)$, and in addition is globally weakly and uniformly continuous. From these considerations and from Theorem 3, page 267, [3], it follows that $D(A) = \mathrm{Diag}(1/n \sum_i A_{ii})$. The usefulness of this fact is that under the $C^*$-isomorphism between $R$ and $M_n(\mathfrak{X})$, any operator $D[w(A, A^*)] \in R$ corresponds to the function $\mathrm{Diag}(1/n\, \sigma[w(A(\cdot), A^*(\cdot)])$ in $M_n(\mathfrak{X})$.

We solve the local unitary equivalence problem first in the simplest case where the $n$-normal operators $A$ and $B$ under consideration are both in the homogeneous $W^*$-algebra $R$ and where $A$ is assumed to generate $R$. We begin by supposing that $D[w(A, A^*)] = D[w(B, B^*)]$ for $w(x, y) \in C$, where $C$ is any collection of words $w(x, y)$ furnishing a complete set of unitary invariants for $n \times n$ matrices. Then it follows that $A(t)$ is unitarily equivalent to $B(t)$ for each $t \in \mathfrak{X}$, and as a result $D[w(A, A^*)] = D[w(B, B^*)]$ for all words $w(x, y) \in W$. At this point we make two observations. The first is that the problem of finding a unitary operator in $R$ satisfying $UAU^* = B$ is equivalent to being able to choose the unitary matrix $U(t)$ implementing the equivalence of $A(t)$ and $B(t)$ in a continuous fashion. The second follows: consider the mapping $\phi: p(A, A^*) \leftrightarrow p(B, B^*)$ between the algebraic $^*$-algebras generated by $A$ and $B$. It is clear, since $\phi$ was shown above to be trace preserving, that $\phi$ is in fact a norm-preserving $^*$-algebra isomorphism, and as such can be extended to a $C^*$-isomorphism between the $C^*$-algebras generated by $A$ and $B$. Thus the question of whether $A$ is unitarily equivalent to $B$ is exactly the question of whether the isomorphism $\phi$

is implemented by a unitary operator. To answer this question, it is useful to consider the problem locally in $M_n(\mathfrak{X})$, where the appropriate $AW^*$-algebras are more accessible. The following lemmas lead to the result.

LEMMA 3.1. *Suppose $A$ generates the homogeneous $n$-normal $W^*$-algebra $R$, and as such corresponds to the function $A(\cdot) \in M_n(\mathfrak{X})$. If $\mathscr{U}$ is any compact open subset of $\mathfrak{X}$, then there is a $t \in \mathscr{U}$ such that $A(t)$ generates the full*-algebra $M_n$ of $n \times n$ complex matrices.*

*Proof.* Suppose there is some compact open $\mathscr{U}$ such that for each $t \in \mathscr{U}$, the *-algebra of matrices generated by $A(t)$ [which is, of course, a direct sum of factors] is not the full algebra $M_n$. Let (**) be the polynomial identity obtained from (*) by replacing $n$ by $n-1$. It follows from the facts about polynomial identities in [1] that for any $t \in \mathscr{U}$, the *-algebra of matrices generated by $A(t)$ satisfies (**). Now the characteristic function of $\mathscr{U}$ corresponds to a projection $E'$ in $Z'$, and thus the operator $E = \mathrm{Diag}\,(E')$ is a central projection in $R$. What we have just proved is that the algebraic *-algebra generated by $EA$ satisfies (**). It follows by continuity that $ER$ satisfies (**) also, which is impossible because $ER$ is homogeneous $n$-normal with $R$ and thus contains $n^2$ matrix units which cannot satisfy (**).

The next lemma uses the fact that if $\mathscr{U}$ is any compact open subset of $\mathfrak{X}$, then the algebra $M_n(\mathscr{U})$ of continuous functions from $\mathscr{U}$ to $M_n$, considered as a normed algebra with sup norm, is a $C^*$-algebra and, in fact, an $AW^*$-algebra.

LEMMA 3.2. *Suppose that $A(\cdot)$ and $B(\cdot)$ are elements of $M_n(\mathfrak{X})$ such that for every word $w(x, y) \in W$ and for every $t \in \mathfrak{X}$, $\sigma[w(A(t), A^*(t))] = \sigma[w(B(t), B^*(t))]$.* Suppose further that $s \in \mathfrak{X}$ is such that $A(s)$ generates $M_n$. Then there is a compact open set $\mathscr{U}$ containing $s$ and a unitary element $V(\cdot) \in M_n(\mathscr{U})$ such that for each $t \in \mathscr{U}$, $B(t) = V(t)A(t)V^*(t)$.

*Proof.* Since $A(s)$ generates $M_n$, there are $n^2$ words $w_i(x, y)$ such that the matrices $w_i(A(s), A^*(s))$ are linearly independent, and we can take $w_1(A(s), A^*(s)) = A(s)$ and $w_2(A(s), A^*(s)) = A^*(s)$. Since $A(\cdot)$ can be regarded as a matrix with continuous functions as entries, there is a compact open set $\mathscr{U}$ containing $s$ such that for $t \in \mathscr{U}$, the $n^2$ matrices $w_i(A(t), A^*(t))$ remain linearly independent. Thus for each $t \in \mathscr{U}$, one obtains, just as in the proof of Theorem 1, that the $n^2$ matrices $w_i(B(t), B^*(t))$ are linearly independent. Furthermore, if

$$w_i(A(t), A^*(t))w_j(A(t), A^*(t)) = \sum_k d_{ij}^k(t)w_k(A(t), A^*(t)) ,$$

then the same equation holds with $A$ everywhere replaced by $B$. Now any element $T(\cdot) \in M_n(\mathscr{U})$ is such that $T(t) = \sum_{i=1}^{n^2} c_i(t)w_i(A(t), A^*(t))$ for $t \in \mathscr{U}$, where the $c_i(\cdot)$ are uniquely determined continuous complex-valued functions on $\mathscr{U}$. This is the crucial fact, for it allows us to define the mapping

$$\phi\colon \sum c_i(\cdot)w_i(A(\cdot), A^*(\cdot)) \rightarrow \sum c_i(\cdot)w_i(B(\cdot), B^*(\cdot))$$

of $M_n(\mathscr{U})$ onto itself. Using the facts mentioned above, it is not hard to see that $\phi$ is in fact a $^*$-algebra automorphism of $M_n(\mathscr{U})$ which leaves the center elementwise fixed. It follows from Theorem 3, [5] that there is a unitary element $V(\cdot) \in M_n(\mathscr{U})$ implementing $\phi$, and since $\phi$ maps $A(\cdot)$ to $B(\cdot)$ we have the desired result. (It is perhaps worth remarking that instead of using Kaplansky's theorem above, the desired unitary element $V(\cdot)$ could have been constructed via a construction from standard algebra.)

THEOREM 3. *Suppose $A$ is a homogeneous $n$-normal operator generating the $W^*$-algebra $R$, and suppose $B$ is any operator in $R$. Suppose also that $C$ is any collection of words $w(x, y)$ with the property that the associated traces form a complete set of unitary invariants for $n \times n$ complex matrices. Finally, suppose that $D[w(A, A^*)] = D[w(B, B^*)]$ for each $w(x, y) \in C$. Then there is a unitary element $U \in R$ such that $UAU^* = B$.*

*Proof.* Consider collections of nonzero, orthogonal, central projections $E_\lambda$ in $R$ for which there exists some unitary operator $V_\lambda$ in $R$ satisfying $BE_\lambda = V_\lambda A V_\lambda^* E_\lambda$. By Zorn one obtains a maximal collection $\{E_\lambda\}$. Let $F = \sup_\lambda \{E_\lambda\} = \sum_\lambda E_\lambda$. To show that $F$ is the unit of $R$, suppose not. Then the central projection $1 - F$ is nonzero, and thus is of the form $\operatorname{Diag}(E')$ where $E'$ is a projection in $Z'$. Now $E'$ corresponds to the characteristic function of a compact open subset $\mathscr{U}_1$ of $\mathfrak{X}$, and by Lemmas 3.1 and 3.2 we can drop down to a compact open subset $\mathscr{U}$ of $\mathscr{U}_1$ such that there is a unitary $V(\cdot) \in M_n(\mathscr{U})$ with $B(t) = V(t)A(t)V^*(t)$ for every $t \in \mathscr{U}$. Then of course $V(\cdot)$ can be extended to a unitary element $V(\cdot) \in M_n(\mathfrak{X})$, and if $E$ is the central projection in $R$ corresponding to the set $\mathscr{U}$, we have $BE = VAV^*E$. This contradicts the maximality of the collection $\{E_\lambda\}$, and thus $\sum_\lambda E_\lambda = 1$. If $U$ is defined as $\sum_\lambda E_\lambda \cdot V_\lambda$, it is an easy matter to verify that $U$ is a unitary operator in $R$ and that $UAU^* = B$.

We can remove the restriction in Theorem 3 that $A$ generates a homogeneous algebra, provided we maintain the requirement that $A$ generates a $W^*$-algebra of type $I$, finite. For it is known that any such algebra $R$ is a direct sum $\sum_{i \in I} \oplus R_i$ where $I$ is some (perhaps infinite) subset of the positive integers and each $R_i$ is homogeneous $i$-normal, and it is easy to see that the Dixmier trace $D(\cdot)$ on $R$ is the direct sum of the functions $D_i(\cdot)$ defined as previously on the homogeneous summands $R_i$. Thus we get immediately.

THEOREM 4. *If $A$ generates the finite $W^*$-algebra $R$ of type $I$, $B$ is any operator in $R$, and $D[w(A, A^*)] = D[w(B, B^*)]$ for each $w(x, y) \in W$, then there is a unitary operator $U \in R$ such that $UAU^* = B$.*

**4. Global unitary equivalence.** We now shift our attention from the question of local unitary equivalence to the question of global unitary equivalence. In other words, if $A$ and $B$ are operators on the Hilbert spaces $\mathscr{H}$ and $\mathscr{K}$ respectively, and each generates a finite $W^*$-algebra of type $I$, we wish to set forth necessary and sufficient conditions for the existence of a unitary isomorphism $\varphi$ mapping $\mathscr{H}$ onto $\mathscr{K}$ and satisfying $\varphi A \varphi^{-1} = B$. Suppose $A$ and $B$ generate the $W^*$-algebras $R(A)$ and $R(B)$ respectively. Let $D_a(\cdot)$ be the Dixmier trace defined on the algebra $R(A)$, and similarly let $D_b(\cdot)$ be the trace on the algebra $R(B)$. In order to eventually arrive at a complete set of unitary invariants for $A$ and $B$, we must set forth conditions which will ensure that the algebras $R(A)$ and $R(B)$ are unitarily equivalent, and the following lemma begins this program.

LEMMA 4.1. *If $A$ and $B$ are as above, and if there is a unitary isomorphism $\varphi$ mapping $\mathscr{H}$ onto $\mathscr{K}$ such that $\varphi D_a[w(A, A^*)]\varphi^{-1} = D_b[w(B, B^*)]$ for each $w(x, y) \in W$, then $\varphi Z(A)\varphi^{-1} = Z(B)$, where $Z(A)$ and $Z(B)$ are the centers of $R(A)$ and $R(B)$ respectively.*

*Proof.* It clearly suffices to demonstrate that the $W^*$-algebra $Z(\Gamma)$ which the collection $\Gamma = \{D_a[w(A, A^*)]/w(x, y) \in W\}$ generates is $Z(A)$. By the fundmental density theorem, $Z(\Gamma)$ is the ultraweak ("*ultrafaible*") closure of the algebraic $*$-algebra generated by $\Gamma$, and $R(A)$ is the ultraweak closure of the algebraic $*$-algebra generated by $A$. If $K \in Z(A)$, then there is a net of polynomials $p_\lambda(A, A^*)$ converging ultraweakly to $K$, and $D_a[p_\lambda(A, A^*)]$ converges ultraweakly to $D_a(K) = K$.

One conjectures that the number of traces required in the previous lemma can be reduced somewhat if it is assumed that $R(A)$ and $R(B)$ are both $n$-normal for some $n$. (This is equivalent to supposing that there exists a positive integer $n$ such that neither $R(A)$ nor $R(B)$ has a nontrivial $i$-homogeneous summand with $i > n$.) The following lemma affirms this conjecture.

LEMMA 4.2. *If $R(A)$ and $R(B)$ are both $n$-normal $W^*$-algebras, $C$ is any collection of words $w(x, y)$ furnishing a complete set of unitary* *invariants for $n \times n$ complex matrices, and* $\varphi D_a[w(A, A^*)]\varphi^{-1} = D_b[w(B, B^*)]$ *for* $w(x, y) \in C$, *then* $\varphi Z(A)\varphi^{-1} = Z(B)$.

*Proof.* As before, it suffices to demonstrate that the $W^*$-algebra which the collection $\Omega = \{D_a[w(A, A^*)]/w(x, y) \in C\}$ generates is $Z(A)$. On the other hand, we know from Lemma 4.1 that the collection $\Gamma = \{D_a[w(A, A^*)]/w(x, y) \in W\}$ generates $Z(A)$. Thus it suffices to show that $\Omega$ and $\Gamma$ generate the same $W^*$-algebra, or even less, the same $C^*$-algebra. Now $R(A)$ is a direct sum of homogeneous algebras $R_i$, and thus is $C^*$-isomorphic to an algebra of the form $\sum_{i\in I} \oplus M_i(\mathfrak{X}_i)$ where $I$ is a subset of the first $n$ positive integers. Consider the compact Hausdorff space $\mathfrak{X} = \bigcup_{i\in I} \mathfrak{X}_i$, defined by agreeing that a set $\mathscr{U}$ is open in $\mathfrak{X}$ if and only if $\mathscr{U} \cap \mathfrak{X}_i$ is open in $\mathfrak{X}_i$ for each $i \in I$. Clearly, $Z(A)$ is $C^*$-isomorphic to the $C^*$-algebra $F$ of all continuous complex-valued functions on $\mathfrak{X}$. For any $w(x, y) \in W$, let $f_w \in F$ be the element corresponding to $D_a[w(A, A^*)]$ in $Z(A)$. If $A$ corresponds to $\sum_{i\in I} \oplus A_i(\cdot)$, then it is easy to see that $f_w/\mathfrak{X}_i$ (the restriction of $f_w$ to $\mathfrak{X}_i$) is equal to $1/i\, \sigma[w(A_i(\cdot), A_i^*(\cdot))]$. We want to prove that $\Omega_1 = \{f_w \in F/w(x, y) \in C\}$ and $\Gamma_1 = \{f_w \in F/w(x, y) \in W\}$ generate the same closed subalgebra of $F$. Define $g_w/\mathfrak{X}_i = i \cdot f_w/\mathfrak{X}_i$ for $i \in I$. Then $\Omega_1$ and $\Gamma_1$ generate the same closed subalgebra of $F$ if and only if $\Omega_2 = \{g_w/w(x, y) \in C\}$ and $\Gamma_2 = \{g_w/w(x, y) \in W\}$ do also. We apply the Stone-Weierstrass Theorem to prove that $\Omega_2$ and $\Gamma_2$ do indeed generate the same $C^*$-subalgebra of $F$, and thus complete the argument. Suppose $t_1, t_2 \in \mathfrak{X}$, and suppose $g_w(t_1) = g_w(t_2)$ for each $w(x, y) \in C$. Say $t_1 \in \mathfrak{X}_i$ and $t_2 \in \mathfrak{X}_j$. Then the matrices $A_i(t_1)$ and $A_j(t_2)$ can both be made into $n \times n$ matrices by forming direct sums with appropriate sized zero matrices, and one sees by virtue of the hypothesis on $C$ that the resulting $n \times n$ matrices are unitarily equivalent. Thus $g_w(t_1) = g_w(t_2)$ for all $w(x, y) \in W$, and it remains only to show that if $t \in \mathfrak{X}_i$ is such that $g_w(t) = 0$ for all $w(x, y) \in C$, then $g_w(t) = 0$ for all $w(x, y) \in W$. This is immediate, however, since then $A_i(t)$ is unitarily equivalent to the zero matrix and thus is equal to zero.

One knows (Theorem 3, [2]) that two $n$-homogeneous $W^*$-algebras whose centers are unitarily isomorphic are then themselves unitarily isomorphic, and the next lemma gives conditions under which the homogeneous summands of two finite $W^*$-algebras of type $I$ can be aligned.

LEMMA 4.3. *Suppose $A$ generates the $n$-homogeneous $W^*$-algebra $R(A)$ with center $Z$, and suppose $B$ generates the $m$-homogeneous $W^*$-algebra $R(B)$ whose center is also $Z$. Suppose also that* $D_a[w(A, A^*)] = D_b[w(B, B^*)]$ *for each* $w(x, y) \in W(\max[2n^2, 2m^2])$. *Then* $m = n$.

*Proof.* We can regard $R(A)$ and $R(B)$ as matrix algebras over the common center $Z$, and if $\mathfrak{X}$ is taken to be the maximal ideal space of $Z$, we obtain $C^*$-isomorphisms of $Z$ onto $C(\mathfrak{X})$, $R(A)$ onto $M_n(\mathfrak{X})$, and $R(B)$ onto $M_m(\mathfrak{X})$. If $A \leftrightarrow A(\cdot)$ and $B \leftrightarrow B(\cdot)$ under these isomorphisms, then as usual $D_a[w(A, A^*)]$ and $D_b[w(B, B^*)]$ correspond respectively to the $n \times n$ matrix Diag $(1/n\,\sigma[w(A(\cdot), A^*(\cdot))])$ and the $m \times m$ matrix Diag $(1/m\sigma[w(B(\cdot), B^*(\cdot))])$. It follows from the hypothesis and the isomorphism between $C(\mathfrak{X})$ and $Z$ that $m/n\sigma[w(A(t),A^*(t))]=\sigma[w(B(t),B^*(t))]$ for each $t \in \mathfrak{X}$ and for each $w(x, y) \in W(\max[2n^2, 2m^2])$. By Lemma 3.1 we can choose a point $s \in \mathfrak{X}$ such that $A(s)$ generates $M_n$, and thus find $n^2$ words $w_i(x, y) \in W(n^2)$ such that the matrices $w_i(A(s), A^*(s))$ are linearly independent. Proceeding just as in the proof of Theorem 1, one concludes that the $n^2$ matrices $w_i(B(s), B^*(s))$ are linearly independent, and thus $m \geqq n$. The result follows by symmetry.

We are now in a position to prove the central result of the paper.

THEOREM 5. *Suppose $A$ is an operator acting on the Hilbert space $\mathscr{H}$ and generating the finite $W^*$-algebra $R(A)$ of type I. Let $D_a(\cdot)$ be the Dixmier trace defined on the algebra $R(A)$. Then $A$ is unitarily equivalent to an operator $B$ acting on the Hilbert space $\mathscr{K}$ if and only if*

*(1) $B$ generates a $W^*$-algebra $R(B)$ which is finite of type I, and*

*(2) there is a unitary isomorphism $\varphi$ of the Hilbert space $\mathscr{H}$ onto the Hilbert space $\mathscr{K}$ satisfying $\varphi D_a[w(A, A^*)]\varphi^{-1} = D_b[w(B, B^*)]$ for each $w(x, y) \in W$, where $D_b(\cdot)$ is the Dixmier trace on the algebra $R(B)$.*

*Proof.* If there is a unitary isomorphism $\varphi$ of $\mathscr{H}$ onto $\mathscr{K}$ satisfying $\varphi A\varphi^{-1} = B$, then $\varphi R(A)\varphi^{-1} = R(B)$, and $\varphi Z(A)\varphi^{-1} = Z(B)$, where $Z(A)$ and $Z(B)$ are the centers of the respective algebras $R(A)$ and $R(B)$. That $\varphi D_a[w(A, A^*)]\varphi^{-1} = D_b[w(B, B^*)]$ for each $w(x, y) \in W$ follows easily from the uniqueness of the Dixmier trace (Theorem 3, page 267, [3]). Going the other way, suppose $B$ generates the finite $W^*$-algebra $R(B)$ of type $I$, and suppose there is a unitary isomorphism $\varphi$ of $\mathscr{H}$ onto $\mathscr{K}$ such that $\varphi D_a[w(A, A^*)]\varphi^{-1} = D_b[w(B, B^*)]$ for $w(x, y) \in W$. Let $A_1$ be the operator $\varphi A\varphi^{-1}$ acting on $\mathscr{K}$, and suppose $A_1$ generates the $W^*$-algebra $R(A_1)$ with center $Z(A_1)$ and Dixmier trace $D_{a_1}(\cdot)$. Then another uniqueness argument shows that $D_b[w(B, B^*)] = D_{a_1}[w(A_1, A_1^*)]$ for $w(x, y) \in W$, and from Lemma 4.1 we obtain $Z(B) = Z(A_1)$. Write $R(A_1) = \sum_{i\in I} \oplus R_i$ and $R(B) = \sum_{j\in J} \oplus T_j$ where $R_i$ and $T_i$ are homogeneous $i$-normal algebras, and $I$ and $J$ are subsets of the positive integers. (It is convenient to regard the above direct sums as internal in this situation, and we do so.) If $E_i$ is the unit of the algebra $R_i$, then at least $E_i$ is a central projection in $Z(B)$, and we show that $E_i$

is in fact the unit of $T_i$ (thus proving $I \subset J$). Write $E_i = \sum_{j \in J} \oplus F_j$, where each $F_j$ is a central projection in $T_j$. If $F_j$ is nonzero, then the algebra $F_j R(B) = F_j T_j$ is $j$-homogeneous, and since $F_j \leqq E_i$, $F_j R(A_1) = F_j R_i$ is $i$-homogeneous. It is easy to see that Lemma 4.3 is applicable to the operators $F_j A_1$ and $F_j B$, and it results that $j = i$ and hence $E_i = F_i$. Thus $E_i$ is dominated by the unit of the algebra $T_i$, and from symmetry considerations one can conclude that $E_i$ is the unit of the algebra $T_i$ and that $I = J$. In other words, for $i \in I$, the homogeneous algebras $R_i$ and $T_i$ have the common center $E_i Z(B)$. If Theorem 3, [2], is now applied for each $i \in I$, there results a unitary operator $V$ such that $VR(A_1)V^* = R(B)$ and $V$ commutes with $Z(A_1)$. Consider $A_2 = VA_1V^*$ which clearly generates the $W^*$-algebra $R(B)$. This fact and another uniqueness argument yield $D_b[w(A_2, A_2^*)] = D_b[w(B, B^*)]$ for $w(x, y) \in W$, and it follows from Theorem 4 that there is a unitary operator $Y \in R(B)$ satisfying $YA_2Y^* = B$. Thus $(YV\varphi)A(YV\varphi)^{-1} = B$, and the argument is complete.

As was the case in Lemma 4.2, if it is known that the operators $A$ and $B$ of Theorem 5 each generate an $n$-normal $W^*$-algebra, it is possible to get by with assumptions on fewer traces.

THEOREM 6. *If the $W^*$-algebras $R(A)$ and $R(B)$ of Theorem* 5 *are $n$-normal, and $\varphi D_a[w(A, A^*)]\varphi^{-1} = D_b[w(B, B^*)]$ for $w(x, y) \in W(2n^2)$, then $A$ and $B$ are unitarily equivalent.*

The proof is similar to that of Theorem 5 and is omitted.

## 5. Remarks.

(1) It is of interest to ask how near the upper bound $4^{n^2}$ obtained in § 2 is to the least upper bound on the number of traces required to form a complete set of unitary invariants for $n \times n$ matrices. In this connection, it is well known that for $n = 2$ the collection $\{\sigma(A), \sigma(A^2), \sigma(AA^*)\}$ is a complete set of invariants, and also the author has shown [9] that for $n = 3$, a collection of nine traces suffices. Thus it would appear that the estimate $4^{n^2}$ is not very good, but it is thought that to obtain any substantial improvement, a completely new approach will be necessary.

(2) (*Added in proof*) I wish to acknowledge my indebtedness to Don Deckard for pointing out a slight simplification in my original proof of Theorem 1 which enabled me to reduce the number of traces needed from $16^{n^2}$ to $4^{n^2}$.

(3) Whether the sets of invariants provided by Theorem 5 and 6 are satisfactory is, of course, open to question. We present the following facts in support of their reasonableness:

(a) Two normal operators are unitarily equivalent if and only if their associated spectral measures are, and thus a solution of this simpler problem requires the simultaneous unitary equivalence of the corresponding elements in two infinite families of commuting projections.

(b) Brown [2] has given a complete set of unitary invariants for homogeneous binormal operators which requires the simultaneous unitary equivalence of four commuting normal functions of the operators. Furthermore, he shows by example that one cannot do away with the simultaneity of this unitary equivalence.

## REFERENCES


1. A. S. Amitsur and J. Levitski, *Minimal identities for algebras,* Proc. Amer. Math. Soc., **1** (1950), 449-463.
2. A. Brown, *The unitary equivalence of binormal operators,* Amer. J. Math., **76** (1954), 414-434.
3. J. Dixmier, *Les algebres d'operateurs dans l'espace Hilbertien,* Paris, 1957.
4. P. R. Halmos, *Introduction to Hilbert space and the theory of spectral multiplicity,* New York, 1943.
5. I. Kaplansky, *Algebras of type I,* Annals of Math., **56** (1952) 460-472.
6. ———, *Projections in Banach algebras,* Annals of Math., **53** (1951), 235-249.
7. I. Kaplansky and others, *Some aspects of analysis and probability,* New York, 1958.
8. F. J. Murray and J. v. Neumann, *On rings of operators,* Annals of Math., **37** (1936), 116-229.
9. C. Pearcy, *A complete set of unitary invariants for* $3 \times 3$ *complex matrices,* Trans. Amer. Math. Soc., **104** (1962), 425-429.
10. W. Specht, *Zur Theorie der Matrizen II,* Jahresbericht der Deutschen Mathematiker Vereinigung, **50** (1940), 19-23.
11. M. H. Stone, *Boundedness properties in function lattices,* Canadian J. Math., **1** (1949), 176-186.


RICE UNIVERSITY

# INTEGRAL CLOSURE OF RINGS OF SOLUTIONS OF LINEAR DIFFERENTIAL EQUATIONS

EDWARD C. POSNER

Let $K$ be an ordiary differential field of characteristic zero with field of constants $C$. Let $R$ be a differential subring of $K$ containing $C$ and having quotient field $K$. A differential subring $V$ of an extension differential field $M$ of $K$ is called a fundamental differential ring (over $R$) if $V$ contains $R$ and if, for each $v$ in $V$, there exist $v_2, \cdots, v_n$ in $V$, $n$ depending on $v$, such that $v, v_2, \cdots, v_n$ form a fundamental system of solutions of a homogeneous linear differential equation with coefficients in $K$. Throughout this paper, $\{\cdots\}$ denotes differential ring adjunction, $<\cdots>$ differential field adjunction.

THEOREM 1. *Let $K, C, R, M, V$ be as above. Then $V$ is a fundamental differential ring over $R$ if and only if $V = R\{v_{\alpha i}, \alpha \in A, 1 \leqq i \leqq n_\alpha\}$, $A$ an indexing set, where for each $\alpha$ in $A$, $v_{\alpha 1}, v_{\alpha 2}, \cdots, v_{\alpha n_\alpha}$ form a fundamental system of solutions of a homogeneous linear differential equation over $K$.*

*Proof.* If $V$ is a fundamental differential ring over $R$, we may let $A = V$; the interest attaches to the converse. It amounts to proving that every differential polynomial with coefficients in $R$ in the $v_{\alpha i}$ is one element of a fundamental system of solutions of a homogeneous linear differential equation over $K$, all the elements of which system of solutions belong to $V$. By use of induction, we may reduce the problem to consideration of the four differential polynomials $s'$, $s+t$, $st$, and $rs$, $r \in K$. We treat the polynomials $s'$ and $s+t$; the polynomials $st$ and $rs$ are treated in a like manner.

Let $s^{(n)} + a_{n-1}s^{(n-1)} + \cdots + a_0 s = 0$, $a_i \in K$, $0 \leqq i \leqq n-1$. (There is no loss of generality in supposing that the leading coefficient of this differential equation is 1.) If $a_0 = 0$, then $s'$ already satisfies a homogeneous linear differential equation (of order $n-1$) over $K$; if $a_0 \neq 0$, we differentiate the expression

$$\left(\left(\frac{1}{a_0}\right)s^{(n)} + \left(\frac{a_{n-1}}{a_0}\right)s^{(n-1)} + \cdots + \left(\frac{a_1}{a_0}\right)s' + s\right)$$

to obtain a homogeneous linear differential equation of order $n$ in $s'$

Received May 23, 1960, and in revised form February 8, 1962. Supported by NASA Contract NASw-6 between the Jet Propulsion Laboratory of the California Institute of Technology and the National Aeronautics and Space Administration. I am indebted to the referee for suggesting valuable improvements incorporated into this paper.

with coefficients in $K$.

To prove the result for $s+t$, let $s, t$ be in $V$ with $s+t \neq 0$; let $s, s_2, \cdots, s_n$ be $n$ elements of $V$ forming a fundamental system of solutions of a homogeneous linear differential equation over $K$, and the same for $t, t_2, \cdots, t_m$. Let $s_1 = s, t_1 = t$. Let $u_1 = s_1 + t_1$ and choose $u_2, u_3, \cdots, u_r$ from among $s_1, s_2, \cdots, s_n; t_1, t_2, \cdots, t_m$ such that $u_1, u_2, \cdots, u_r$ form a basis over the constants for the vector space spanned over the constants by $s_1, s_2, \cdots, s_n; t_1, t_2, \cdots, t_m$. Let $W(z_1, z_2, \cdots, z_p)$ denote the wronskian of the $p$ elements $z_1, z_2, \cdots, z_p$. Consider the linear differential operator of order $r$, $\mathscr{L}(y) = W(y, u_1, \cdots, u_r)/W(u_1, \cdots, u_r)$. (Since $u_1, \cdots, u_r$ are linearly independent over constants, their wronskian is nonzero.) $\mathscr{L}(u_r) = 0$, $1 \leqq \lambda \leqq r$, and $\mathscr{L} \neq 0$ since the coefficient of $y^{(r)}$ is $1 = W(u_1, \cdots, u_r)/W(u_1, \cdots, u_r)$. We shall prove that all the coefficients of $\mathscr{L}$ are in $K$; $\mathscr{L}(y) = 0$ will then be the sought-after differential equation.

Let $\sigma$ be a differential isomorphism of $K\langle s_1, s_2, \cdots, s_n; t_1, t_2, \cdots, t_m\rangle$ over $K$; then $\sigma(s_\mu) = \sum_{i=1}^{n} c_{\mu i} s_i$, $1 \leqq \mu \leqq n$ and $\sigma(t_\nu) \sum_{j=1}^{m} d_{\nu j} t_j$, $1 \leqq \nu \leqq m$, where the $c_{\mu i}$ and $d_{\nu j}$ are constants. This is true because $s_1, s_2, \cdots, s_n$ span over constants the vector space of solutions of the homogeneous linear differential equation over $K$ satisfied by $s_1$; similarly for $t_1, \cdots, t_m$. These two sets of equations taken together imply $\sigma(u_\lambda) = \sum_{k=1}^{r} e_{\lambda k} u_k$, $1 \leqq \lambda \leqq r$, $e_{\lambda k}$ constants, for each $\sigma(u_\lambda)$ is in the vector space spanned over the constants by $s_1, \cdots, s_n; t_1, \cdots, t_m$.

This implies that $W(y, \sigma u_1, \cdots, \sigma u_r) = (\det(e_{\lambda k}))W(y, u_1, \cdots, u_r)$, and similarly $W(\sigma u_1, \cdots, \sigma u_r) = (\det(e_{\lambda k}))W(u_1, \cdots, u_r)$. Therefore the coefficients $a_p$, $0 \leqq p \leqq r$, of $\mathscr{L}(y)$ are invariant under $\sigma$, for all differential isomorphisms $\sigma$ of $K\langle s_1, \cdots, s_n; t_1, \cdots, t_m\rangle$ over $K$. By Theorem 2.6, pg. 16 of [1], $a_p$ is in $K$, as required. This proves the theorem.

The above theorem has the following immediate consequence.

COROLLARY. *If $M$ is a universal differential field extension of $K$* ([2], Sec. 5, esp. pg. 771, Theorem), *the set $V$ of all elements of $M$ satisfying a homogeneous linear differential equation over $K$ forms a fundamental differential ring.*

The following lemma isolates the key property of fundamental differential rings that will be used to prove integral closure. An element $w$ in an extension differential field of $K$ is called a wronskian over $K$ if $w \neq 0$ and $w'/w$ belongs to $K$.

LEMMA. *Let $V$ be a fundamental differential ring over $R$. Then any nonzero differential ideal $I$ of $V$ contains a wronskian over $K$.*

*Proof.* Let $u_1$ be a nonzero element of the differential ideal $I$ of $V$, and let $u_2, u_3, \cdots, u_n$ be $n-1$ elements of $V$ such that $u_1, u_2, \cdots, u_n$ form a fundamental system of solutions of a homogeneous linear differential equation over $K$. Then $W(u_1, u_2, \cdots, u_n)$ is a nonzero element of $I$: it is nonzero since $u_1, u_2, \cdots, u_n$ are linearly independent over constants; it belongs to $I$ because each term in the expansion of the determinant defining $W(u_1, \cdots, u_n)$ contains a derivative of $u_1$ as a factor. Since $W(u_1, \cdots, u_n)$ is a wronskian over $K$, the proof is complete.

DEFINITION. A differential ring is called differentiably simple if it has no differential ideals other than zero and itself.

THEOREM 2. *Let $R$ be differentiably simple (in particular, $R = K$), and for every wronskian $w$ over $K$ belonging to $V$, let there exist a nonzero $h$ in $R$ such that $h/w$ is in $V$. Then $V$ too is differentiably simple. (When $R = K$, the assumption is that $V$ contains the inverse of every wronskian over $K$ which belongs to $V$.)*

*Proof.* Let $I$ be a nonzero differential ideal of $V$. To prove that $I = V$, let $w$ be a wronskian over $K$ in $I$; such exist by the lemma. Now by hypothesis, there is a nonzero $h$ in $R$ with $h/w$ in $V$. Thus $w \cdot h/w = h$ is in $I$, so that $I \cap R$ is not the zero ideal of $R$. Since $I \cap R$ is a differential ideal of $R$ and $R$ is differentiably simple, $I \cap R = R$, so that $1 \in I \cap R$, and $1 \in I$. Thus $I = V$ as required.

The next theorem is a sort of converse to the previous theorem. (Here $V$ need not be a fundamental differential ring over $R$; $V$ can be any differential subring of $M$ containing $R$.)

THEOREM 3. *Let $V$, but not necessarily $R$, be differentiably simple, and let $w$ be a wronskian over $K$ belonging to $V$. Then there is a nonzero $h$ in $R$ such that $h/w$ is in $V$. (Thus if $R = K$, $1/w$ is in $V$.)*

*Proof.* Since $K$ is the quotient field of $R$, there exist $b, c$ in $R$, with $c \neq 0$, such that $w' = (b/c)w$. Let $I$ denote the set of elements of $V$ of the form $vc^{-p}w$, $p$ a nonnegative integer, $v$ an element of $V$. $I$ can readily be shown to be an ideal of $V$; we shall prove that $I$ is closed under differentiation. If $vc^{-p}w \in I$, then $(vc^{-p}w)' = v'c^{-p}w - pvc^{-p-1}c'w + vc^{-p}w' = (v'c)c^{-p-1}w - (pvc')c^{-p-1}w + (bv)c^{-p-1}w = (v'c - pvc' + bv)c^{-p-1}w$ is an element of $V$ and hence of $I$. Thus $I$ is a differential ideal of $V$, and is nonzero since $w$ is in $I$. Since $V$ is differentiably simple, $I = V$, and $1 \in I$. Thus $1 = vc^{-p}w$ for some $v \in V$, $p \geqq 0$. Then, if $c^p = h$, we have $h/w = v \in V$, with $h$ an element of $R$. This proves the theorem.

The following theorem with $K = C$ generalizes a consequence of a result of Ritt ([4], Sec. 1, pg. 681) to the effect that if $C$ is the field of complex numbers, the ring $C\ [e^{\lambda x}$, all complex $\lambda]$ is integrally closed in its quotient field. In fact, Theorem 4 also implies that $C\ [x, e^{\lambda x}]$ is is integrally closed in its quotient field.

THEOREM 4. *Let $K$ be a differential field of characteristic zero with field of constants $C$. Let $K$ be differential algebraic over $C$. Let $V$ be a fundamental differential ring over $K$ which contains the inverse of every wronskian over $K$ in it. Then $V$ is integrally closed in its quotient field* (*it is differentiably simple by Theorem* 2).

*Proof.* Let $u$ be an element of the quotient field $M$ of $V$ integral over $V$: that is, there exist elements $v_i$ in $V$, $1 \leqq i \leqq n$, such that $u^n + \sum_{i=1}^{n} v_i u^{n-i} = 0$, and there exist $v_{n+1}, v_{n+2}$ in $V$ with $u = v_{n+1}/v_{n+2}$. Let $v_i$ be a solution of a homogeneous linear differential equation $\mathscr{L}_i(y) = 0, 1 \leqq i \leqq n+2$, where $\mathscr{L}_i(y) = \sum_{j=0}^{n_i} b_{ij} y^{(j)}, 1 \leqq i \leqq n+2, 0 \leqq j \leqq n_i; b_{in_i} = 1, 1 \leqq i \leqq n+2$. Furthermore let $v_{ik}, 1 \leqq k \leqq n_i$, be for each $i$ a fundamntal system of solutions of $\mathscr{L}_i(y) = 0$, with $v_{i1} = v_i$. Let $Y$ be a differential indeterminate, and, for each $i, j$, let $P_{ij}(Y) \in C\{Y\}$ be a differential polynomial of lowest order $r_{ij}$ say satisfied by $b_{ij}$ over $C$ and such that the degree of $P_{ij}$ in $Y^{(r_{ij})}$ is as small as possible among these differential polynomials of order $r_{ij}$. Define the separant $S_{ij}$ of $P_{ij}$ as the (partial) derivative of $P_{ij}$ with respect to $Y^{(r_{ij})}$. One verifies, using the minimal property of the $P_{ij}$, that $S_{ij}(b_{ij})$ is nonzero. Then $b_{ij}^{(r_{ij}+1)}$ is $S_{ij}^{-1}(b_{ij})$ multiplied by a differential polynomial over $C$ in $b_{ij}$ of order at most $r_{ij}$. This implies that $C\{b_{ij}\} = C[b_{ij}^{(p)}, 0 \leqq p \leqq r_{ij}]$, all $i, j$. (This argument is well known.)

Now define $\bar{V} = C\{b_{ij}, S_{ij}^{-1}(v_{ij}), v_{ik}$, all $1 \leqq i \leqq n+2, 0 \leqq j \leqq n_i, 1 \leqq k \leqq n_i\}$; observe $\bar{V} \subset V$. Since $\mathscr{L}_i$ has leading coefficient 1 and $\mathscr{L}_i(v_{ik}) = 0, 1 \leqq i \leqq n+2, 1 \leqq k \leqq n_i$, and because of the above property of each $C\{b_{ij}\}$, one concludes that $\bar{V} = C[b_{ij}^{(p)}, S_{ij}^{-1}(b_{ij}), v_{ik}^{(q)}$, all $1 \leqq i \leqq n+2, 0 \leqq j \leqq n_i, 1 \leqq k \leqq n_i\ 0 \leqq p \leqq r_{ij}, 0 \leqq q \leqq n_i - 1]$. This is what we were after: we have proved that $\bar{V}$ is finitely generated as an ordinary ring over $C$. We can now apply Theorem 2 of [3] to conclude that the integral closure $\bar{O}$ of $\bar{V}$ in its quotient field $\bar{M}$ is in fact a differential subring of $\bar{M}$. But $u$ is in $\bar{O}$; if we can prove that $\bar{O}$ is contained in $\bar{V}$, the proof will be completed.

So consider the ideal $\bar{I}$ of $\bar{V}$ consisting of all $h$ in $\bar{V}$ such that $h\bar{O} \subset \bar{V}$. By [5], pg. 267, Theorem 9, $\bar{I}$ is nonzero; a fortiori, the ideal $I$ of $V$ consisting of those $h$ in $V$ with $h\bar{O} \subset V$ is also nonzero, since it contains $\bar{I}$. We assert that $I$ is a differential ideal of $V$: let $\omega \in \bar{O}$; then $h\omega \in V, (h\omega)' = h'\omega + h\omega' \in V$. Since $\bar{O}$ is closed under differentiation by [3], pg. 1393, lemma, $\omega' \in \bar{O}$, so that, since $h \in I$,

$h\omega' \in V$. Thus $h'\omega$ is in $V$ if $\omega$ is in $\bar{O}$ and $h$ is in $I$. In other words, $I$ is a differential ideal of $V$. Since $V$ is differentiably simple by Theorem 2, and $I$ is nonzero, we conclude that $I = V$. Therefore $1 \in I$. This implies that $\bar{O} = 1 \cdot \bar{O}$ is contained in $V$, as promised. This completes the proof of Theorem 4.

(The above theorem could be strengthened by use of the following unproved result: a differentiably simple ring of characteristic zero is integrally closed in its quotient field. This result would generalize Theorem 1 of [3].)

Theorem 4 has the following corollary.

COROLLARY. *Let $K$ be a differential field of characterististic zero with field of constants $C$. Let $K$ be differential algebraic over $C$. Let $M$ be a universal differential field extension of $K$. Let $V$ be the subset of $M$ comprising those elements of $M$ satisfying a homogeneous linear differential equation over $K$. Then $V$ is integrally closed in its quotient field.*

*Proof.* That $V$ is a fundamental differential ring over $K$ follows from the corollary to Theorem 1. To prove $V$ integrally closed in its quotient field, we shall prove that $V$ contains the inverse of every wronskian over $K$ in it, and then apply Theorem 4.

Now if $w$ is a wronskian over $K$ in $V$, then $w \neq 0$ and $w' = kw$, $k \in K$. Then $(1/w)' = (-1/w^2)\cdot w' = (-1/w^2)\cdot kw = -k\cdot(1/w)$. So $1/w$ satisfies a (first order) homogeneous linear differential equation over $K$; by the definition of $V$, $(1/w)$ belongs to $V$, as required for the application of Theorem 4.

REMARK. Let $V_1 = V$ and $V_{n+1}$, $n \geqq 1$, be the differential subring of $M$ consisting of those elements of $M$ satisfying a homegeneous linear differential equation with coefficients in $V_n$. Then $V_{n+1}$ contains $L_n$ (thus $\bigcup_{n=1}^{\infty} V_n = V_\infty$ is a field), for if $f(\neq 0)$ is in $V_n$, then $(1/f)' = -f'/f\cdot 1/f$. Thus $1/f$ satisfies a first order homogeneous linear differential equation with coefficients in $L_n$ and so is in $V_{n+1}$. Since $V_{n+1}$ contains $V_n$, and now the inverse of every nonzero element in $V_n$, $V_{n+1}$ contains $L_n$. But each $L_n$ is differential algebraic over $C$, and $M$ is still a universal differential extension of $L_n$. The above corollary thus implies that each $V_n$ is integrally closed in its quotient field $L_n$, $n \geqq 1$.

## BIBLIOGRAPHY

1. I. Kaplansky, *An Introduction to Differential Algebra*, Paris, Hermann, 1957.
2. E. R. Kolchin, *Galois theory of differential fields*, Amer. J. of Math., **75** (1953), 753-824.

3. E. C. Posner, *Integral closure of differential rings,* Pacific J. Math., **10** (1960), 1393-1396.
4. J. F. Ritt, *On the zeros af exponential polynomials,* Trans. Amer. Math. Soc., **52** (1928), 680-686.
5. O. Zariski, and P. Samuel, *Commutative Algebra,* Vol. I, Princeton, van Nostrand, 1958.

CALIFORNIA INSTITUTE OF TECHNOLOGY

# ASYMPTOTICS III: STATIONARY PHASE FOR TWO PARAMETERS WITH AN APPLICATION TO BESSEL FUNCTIONS

D. SATHER

**1. Introduction.** The method of stationary phase has long been a valuable analytical tool for investigating the asymptotic behavior as $p \to \infty$ of integrals of the form

$$I(p) = \int_0^a Q(t) \exp(ipF(t))dt .$$

As a natural generalization of the method of stationary phase involving one parameter we will investigate the asymptotic behavior of an integral of the form

$$I(h, k) = \int_0^a t^{\gamma-1} q(t) \exp[i(ht^\lambda f(t) + kt^\nu g(t))]dt$$

where $h$ and $k$ tend to infinity independently.

It will be shown that under certain restrictions between the real numbers $\lambda$, $\nu$ and $\gamma$ that the asymptotic form of $I(h, k)$ is determined by the behavior of the ratio $kh^{-\nu/\lambda}$ as $h, k \to \infty$ and by the character of $f$ and $g$ in a neighborhood of $t = 0$. For example, if $\gamma < \nu < \lambda$, $\gamma > 0$, $f(0) > 0$, $g(0) > 0$ and $kh^{-\nu/\lambda} \to \infty$ then

$$I(h, k) \backsim \frac{q(0)\Gamma\left(\frac{\gamma}{\nu}\right) \exp\left(\frac{i\pi\gamma}{2\nu}\right)}{\nu[kg(0)]^{\gamma/\nu}} .$$

As an immediate application of our results we will determine the asymptotic behavior of the Bessel function $J_\nu(x)$ in Watson's transition region, i.e. when $\nu$, $x$ and $|\nu - x|$ are large and $\nu/x$ is nearly equal to 1. In particular, we will obtain a simple rigorous proof of Nicholson's formulas under the restriction that $0 < \limsup \; x^{-1/3}|\nu - x| < \infty$.

**2. General assumptions.** Throughout the paper we shall use $A \backsim B$ to mean $\lim A/B = 1$, and all limits will mean the limit as $h$ and $k$ tend to infinity. A similar remark applies to order symbols.

We shall consider $I(h, k)$ under the following general assumptions:

Received January 3, 1962. This paper was written at the University of Minnesota in part under Contract Nonr 710(16), sponsored by the Office of Naval Research, and in part under a National Science Foundation Fellowship. The author wishes to express his appreciation to Professor W. Fulks for suggesting the problem and for giving valuable aid in its solution.

(i) $k = o(h)$,
(ii) $\lambda > 0$, $\nu > 0$, and $\gamma > 0$,
(iii) $q(0) \neq 0$ and $g(0) \neq 0$,
(iv) $f, g$ and $q$ are real valued functions such that $f \in C^2$, $g \in C^2$ and $q \in C$ on $[0, a]$,
(v) $\lambda f(t) + tf'(t) > 0$ on $[0, a]$.

For convenience we shall consider here only the case $f(0) > 0$. If $f(0) < 0$ and $-f$ satisfies certain obvious conditions one obtains analogous results with $-i$ and $-g$ replacing $i$ and $g$, respectively.

**3. Preliminary lemmas.** We shall first establish the following lemmas.

LEMMA I. *Consider* $I(p) = \int_0^b \omega(t)\Psi(t)\exp(ip\Phi(t))dt$. *Suppose $d$ is a nonnegative constant and $p, \alpha$ and $\mu$ are functions of $h$ and $k$ such that $p \to \infty$, $\mu \to 0$ and $\alpha$ is bounded as $h, k \to \infty$.*

(i) $\Phi(t) = t^r\phi(\alpha t)$, $\Psi(t) = t^{s-1}\psi(\mu(t+d))$, *the functions $\phi$ and $\psi$ are real with $\psi(0) \neq 0$, $r > 0$, $0 < s < r$, $\phi(\alpha t) > 0$ for $0 \leqq t \leqq c'$, $c' > 0$ and $\phi \in C^{n+2}$ and $\psi \in C^n$ for $0 \leqq t \leqq c'$ where $m$ and $n$ are the least integers such that $mr > 1$ and $n \geqq m(r-s)+1$, respectively.*

(ii) *$b$ is a constant such that $0 < b < c'$ and $bMK/m_0 r < 1$ where $M = \text{maximum}_{0 \leqq t \leqq c'} |\phi'(t)|$, $m_0 = \text{minimum}_{0 \leqq t \leqq c'} \phi(t)$ and $K \geqq \alpha$ when $h, k$ are sufficiently large.*

(iii) *$\omega = u + iv$ is a complex valued function such that $u(0) = 1$, $v(0) = 0$ and $u, v \in C^n$ for $0 \leqq t \leqq c'$. Then*

$$I(p) \backsim \frac{\psi(0)\Gamma\left(\frac{s}{r}\right)\exp\left(\frac{i\pi s}{2r}\right)}{r[p\phi(0)]^{s/r}}.$$

*Proof.* We may set $v \equiv 0$ since it will be seen that the contribution from $v$ to $I(p)$ is negligible because $v(0) = 0$. Let $x = t[\phi(\alpha t)]^{1/r}$. Since $x'(t) > 0$ for $0 \leqq t \leqq b$ and $x \in C^{n+2}$ there exists a unique inverse function, say $t(x)$, such that $t \in C^{n+2}$ for $0 \leqq x \leqq b[\phi(\alpha b)]^{1/r} = a$, $t(0) = 0$ and $t'(0) = a_1 = [\phi(0)]^{-1/r}$. Hence we may write $t(x) = a_1 x + a_2 x^2 + \cdots + a_{n-1}x^{n-1} + A(x)x^n$ where $A \in C^2$ and $a_l$ is bounded as $h, k \to \infty$ for $2 \leqq l \leqq n-1$. We may assume that $c'$ is sufficiently small such that if $t(x) = a_1 x(1 + w(x))$ then $|w(x)| < 1$ for $0 \leqq x \leqq a$. This implies that

$$(t(x))^{s-1} = a_1^{s-1}(1 + b_1 x + \cdots + b_{n-2}x^{n-2} + z(x)x^{n-1})x^{s-1}$$

where $z \in C$ and $b_l$ is independent of $x$ for $1 \leqq l \leqq n-2$. If we now expand $\psi$ and $\omega$ about $t = 0$ and substitute $t(x)$, and let $B(x) = \omega(t(x))(t(x))^{s-1}\psi(\mu(t(x)+d))$ we have

$$B(x)\frac{dt(x)}{dx} = a_1^s u(\mu d)x^{s-1} + c_0 x^s + \cdots + c_{n-3}x^{s+n-3} + D(x)x^{n+s-2}$$

where $h$ and $k$ are sufficiently large such that $\mu d < b$, $D \in C$ and $c_l$ is bounded as $h, k \to \infty$ and independent of $x$ for $0 \leqq l \leqq n-3$. Therefore,

$$I(p) = a_1^s \psi(\mu d) \int_0^a x^{s-1} \exp(ipx^r) dx + J(p) + \sum_{l=0}^{n-3} c_l \int_0^a x^{s+l} \exp(ipx^r) dx$$

where $J(p) = \int_0^a D(x) x^{n+s-2} \exp(ipx^r) dx$. Since $\int_0^\infty e^{it} t^{\beta-1} dt = \exp(i\pi\beta/2)\Gamma(\beta)$ for $0 < \beta < 1$ and $r \leqq n + s - 1 < r + 1$ when $r > 1$ we have

$$\begin{aligned} I(p) &= \frac{a_1^s \psi(\mu d) \exp\left(\frac{i\pi s}{2r}\right) \Gamma\left(\frac{s}{r}\right)}{r p^{s/r}} + J(p) + o(p^{-s/r}) \\ &= \frac{a_1^s \psi(0) \exp\left(\frac{i\pi s}{2r}\right) \Gamma\left(\frac{s}{r}\right)}{r p^{s/r}} + J(p) + o(p^{-s/r}) . \end{aligned}$$

Finally an integration by parts yields $J(p) = 0(1/p)$ since $n-(r+1-s) \geqq 0$ by the choice of $n$ and $D \in C$. This completes the proof of Lemma I for the case $r > 1$. For $0 < r \leqq 1$ one makes the change of variable $t = x^m$ and the desired result follows from the case $r > 1$.

LEMMA II. *Suppose that in addition to the assumptions of Lemma I that $r$ is an even integer, $s = 1$, $\phi(\alpha t) > 0$ for $-c' \leqq t \leqq c'$, $b$ satisfies the same conditions as in Lemma I except that $M$ and $m_0$ are now determined for $-c' \leqq t \leqq c'$, and $\omega$, $\psi$ and $\phi$ are now in their respective differentiability classes given in Lemma I for $-c' \leqq t \leqq c'$. Then*

$$\int_{-b}^{b} \omega(t) \Psi(t) \exp(ip\Phi(t)) dt \backsim \frac{2\psi(0)\Gamma\left(\frac{1}{r}\right)\exp\left(\frac{i\pi}{2r}\right)}{r[p\phi(0)]^{1/r}} .$$

*The proof follows immediately from Lemma I.*

We will introduce the following functions which will be used throughout the remainder of the paper:

$$F(t) = t^\lambda f(t), \; G(t) = t^\nu g(t) \text{ and } Q(t) = t^{\gamma-1} q(t).$$

LEMMA III. *Under the general assumptions on $F$, $G$ and $Q$ we have for each arbitrarily small but fixed positive constant $c < a$ that*

$$L(h, k) = \int_c^a Q(t) \exp[i(hF(t) + kG(t))] dt = 0(1/h) .$$

*Proof*. Let $H(t) = F(t) + (k/h)G(t)$. Then $H'(t) > 0$ for $c \leqq t \leqq a$ and $h, k$ sufficiently large since $\lambda f(t) + tf'(t) > 0$ by hypothesis and

$k = o(h)$. Hence an integration by parts implies $L(h, k) = 0(1/h)$.

This completes the necessary lemmas and the main results of the paper will now be presented.

**4. The asymptotic evaluation of $I(h, k)$.** We shall first consider the case where $kh^{-\nu/\lambda} \to 0$ so that $I(h, k)$ is almost completely determined by the character of $hf$ at the origin.

THEOREM I. *Suppose that*

1. $f \in C^{n+2}$ and $q \in C^n$ for $0 \leqq t \leqq c$, $c > 0$, *where $m$ and $n$ are the least integers such that* $m\lambda > 1$ and $n \geqq m(\lambda - \gamma) + 1$, *respectively,*

2. *if $0 \leqq t \leqq c$ then $t^\nu g(\beta t) = b_0 + b_1 t + \cdots + b_{n-2}t^{n-2} + B(t)t^{n-1}$ where $B \in C$ and $b_l$ is bounded as $\beta \to 0$ for $0 \leqq l \leqq n-2$,*

3. *$k^\lambda = o(h^\nu)$ and $\gamma < \lambda$. Then*

$$I(h, k) \backsim \frac{q(0)\Gamma\left(\frac{\gamma}{\lambda}\right)\exp\left(\frac{i\pi\gamma}{2\lambda}\right)}{\lambda[hf(0)]^{\gamma/\lambda}} .$$

*Proof of Theorem I.* For $c$ as given we have

$$I(h,k) = \int_0^c + \int_c^a = I'(h, k) + 0(1/h)$$

by Lemma III. Let $t = xk^{-1/\nu}$, $\tilde{f}(x) = f(xk^{-1/\nu})$, $\tilde{g}(x) = g(xk^{-1/\nu})$, $\tilde{Q}(x) = Q(xk^{-1/\nu})$ and $p = hk^{-\lambda/\nu}$. For any $b$ such that $0 < b < c$ we have

$$I'(h, k) = k^{-1/\nu}\int_0^b [\tilde{Q}(x)\exp(i\tilde{g}(x)x^\nu)]\exp(ip\tilde{f}(x)x^\lambda)dx$$

$$+ k^{-1/\nu}\int_b^{ck^{1/\nu}} = I''(h, k) + J(h, k), \text{ respectively.}$$

Set $\mu = k^{-1/\nu}$, $\psi = q$, $\phi = f$, $\lambda = r$, $\gamma = s$, $\omega(x) = \exp(ix^\nu\tilde{g}(x))$ and note that $f(0) > 0$ implies that $\tilde{f}(x) > 0$ for $0 \leqq x \leqq c'$, $c' > 0$, so that $b$ may be chosen to satisfy the requirements of Lemma I. Hence by Lemma I

$$I''(h, k) \backsim \frac{q(0)\Gamma(\gamma/\lambda)\exp(i\pi\gamma/2\lambda)}{\lambda[hf(0)]^{\gamma/\lambda}} .$$

Therefore to complete the proof of Theorem I it is sufficient to show that $h^{\gamma/\lambda}J(h, k) = o(1)$. Let $d = bk^{-1/\nu}$, $H(t) = F(t) = +(k/h)G(t)$ and $P(t) = \lambda f(t) + tf'(t) + kt^{\nu-\lambda}/h[\nu g(t) + g'(t)t]$. Note that $P(d) \to \lambda f(0) = 2B > 0$ as $h, k \to \infty$ since $k^\lambda = 0(h^\nu)$ and $P(t)$ is continuous for $0 < d \leqq t \leqq a$. We may assume that $c$ is such that for $h, k$ sufficiently large, $P(t) \geqq B$ for the entire closed interval $d \leqq t \leqq c$. This implies $H'(t) \geqq Bt^{\lambda-1} > 0$ for $0 < d \leqq t \leqq c$ and hence we can integrate $J(h, k)$ by parts as follows;

$$J(h, k) = \int_d^c Q(t) \exp(ihH(t))dt = \frac{Q(c) \exp(ihH(c))}{ihH'(c)} - \frac{Q(d) \exp(ihH(d))}{ihH'(d)}$$
$$- \frac{1}{ih} \int_d^c \frac{Q'(t) \exp(ihH(t))dt}{H'(t)} + \frac{1}{ih} \int_d^c \frac{Q(t)H''(t) \exp(ihH(t))dt}{(H'(t))^2}$$
$$= 0(1/h) + A + J'(h, k), \text{ respectively.}$$

Using the estimates $H'(t) \geqq Bt^{\lambda-1}$ and $|H''(t)| \leqq Kt^{\lambda-2}$ for some $K$ we see immediately that $J = 0(k^{(\lambda-\gamma)\nu}/h)$. Since $k^\lambda = o(h^\nu)$ this implies $h^{\gamma/\lambda}J(h, k) = o(1)$ which completes the proof of Theorem I.

We state the following corollary to Theorem I which may apply when $t^\nu g(\beta t)$ does not have the required smoothness at the origin but $f, g$ and $q$ are highly differentiable on $[0, c]$, $c > 0$.

*Corollary. Suppose that* $\nu + \gamma > \lambda$ *and*

1. $f \in C^{n+2}$, $g \in C^n$ and $q \in C^n$ for $0 \leqq t \leqq c$, $c > 0$ where $m$ and $n$ are the least integers such that $m(\nu+\gamma-\lambda) \geqq 2$, $m\lambda > 1$ and $n \geqq m(\lambda-\gamma)+1$,
2. $k^\lambda = o(h^\nu)$ and $\gamma < \lambda$. Then

$$I(h, k) \backsim \frac{q(0)\Gamma\left(\frac{\gamma}{\lambda}\right) \exp\left(\frac{i\pi\gamma}{2\lambda}\right)}{\lambda[hf(0)]^{\gamma/\lambda}} .$$

*Proof.* Note that $m\nu \geqq m(\lambda - \gamma) + 2 > n$ by the definition of $n$ and hence $x^{m\nu} \in C^n$. The change of variable $t = x^m$ and the use of Theorem I completes the proof.

We shall next consider the case where the behavior of $kg$ at the origin becomes a significant factor in the asymptotic evaluation of $I(h, k)$.

THEOREM II. *Suppose that*

1. $q \in C^n$ *and* $g \in C^{n+2}$ *for* $0 \leqq t \leqq c$, $c > 0$, *where* $m$ *and* $n$ *are the least integers such that* $m\nu > 1$ *and* $n \geqq m(\nu - \gamma) + 1$, *respectively,*
2. *if* $0 \leqq t \leqq c$ *then* $t^\lambda f(\beta t) = b_0 + b_1 t + \cdots + b_{n-2}t^{n-2} + B(t)t^{n-1}$ *where* $B \in C$ *and* $b_l$ *is bounded as* $\beta \to 0$ *for* $0 \leqq l \leqq n - 2$,
3. $g(0) > 0$, $h^\nu = o(k^\lambda)$ *and* $\gamma < \nu < \lambda$. *Then*

$$I(h, k) \backsim \frac{q(0)\Gamma\left(\frac{\gamma}{\nu}\right) \exp\left(\frac{i\pi\gamma}{2\nu}\right)}{\nu[kg(0)]^{\gamma/\nu}} .$$

*Proof of Theorem II.* The proof of Theorem II follows from the proof of Theorem I with the roles of $f$ and $g$, $\lambda$ and $\nu$, $h$ and $k$ intercharged.

COROLLARY. *Suppose that*

1. $f \in C^n$, $q \in C^n$ *and* $g \in C^{n+2}$ *for* $0 \leqq t \leqq c$, $c > 0$, *where* $m$ *and* $n$ *are the least integers such that* $m(\lambda + \gamma - \nu) \geqq 2$, $m\nu > 1$ *and* $n \geqq m(\nu - \gamma) + 1$,
2. $g(0) > 0$, $h^\nu = o(k^\lambda)$ *and* $\gamma < \nu < \lambda$. *Then*

$$I(h, k) \backsim \frac{q(0)\Gamma\left(\dfrac{\gamma}{\nu}\right)\exp\left(\dfrac{i\pi\gamma}{2\nu}\right)}{\nu[kg(0)]^{\gamma/\nu}} .$$

When $kh^{-\nu/\lambda} \to \infty$ and $g(0) < 0$ the character of both $F$ and $G$ in a neighborhood of $t = 0$ becomes important since for $h$ and $k$ sufficiently large they determine uniquely in some $(0, c_0)$ a number $\tau$ such that $hF'(\tau) + kG'(\tau) = 0$ and in terms of which the asymptotic form of $I(h, k)$ may be expressed.

THEOREM III. *Suppose that* $g(0) < 0$, $\nu < \lambda$, $\gamma < \lambda$, $h^\nu = o(k^\lambda)$, $f \in C^6$, $g \in C^6$ *and* $q \in C^2$ *for* $0 \leqq t \leqq c$, $c > 0$, *and hypothesis* 1 *and* 2 *Theorem* II *are satisfied when* $\nu \geqq \gamma$.

A. *If* $\nu < 2\gamma$ *then*

$$I(h, k) \backsim \frac{\sqrt{2}\, q(0)\Gamma\left(\dfrac{1}{2}\right)\exp\left(\dfrac{i\pi}{4}\right)}{(\lambda - \nu)^{1/2}} \left[\frac{(\lambda h f(0))^{\nu - 2\gamma}}{(-\nu k g(0))^{\lambda - 2\gamma}}\right]^{1/2(\lambda - \nu)}$$
$$\times \exp\left[i(hF(\tau) + kG(\tau))\right] .$$

B. *If* $\nu = 2\gamma$ *then*

$$I(h, k) \backsim \frac{q(0)\Gamma\left(\dfrac{1}{2}\right)\exp\left(\dfrac{i\pi}{4}\right)}{(-\nu k g(0))^{1/2}} \left\{\frac{\sqrt{2}\exp\left[i(hF(\tau) + kG(\tau))\right]}{(\lambda - \nu)^{1/2}} - i\nu^{-1/2}\right\} .$$

C. *If* $\nu > 2\gamma$ *then*

$$I(h, k) \backsim \frac{q(0)\Gamma\left(\dfrac{\gamma}{\nu}\right)\exp\left(\dfrac{\gamma\pi}{2\nu i}\right)}{\nu[-kg(0)]^{\gamma/\nu}} .$$

*Proof of Theorem III.* We may assume that $c$ is such that $G'(t) < 0$ and $f(t) > 0$ for $0 < t \leqq c$. For $0 < t \leqq c$ let $D(t) = F'(t)/-G'(t)$ with $D(0) = 0$. Then $D'(t) = t^{\lambda+\nu-\xi}/(G'(t))^2[\nu\lambda f(t)g(t)(\nu - \lambda) + tE(t)]$ for $0 < t \leqq c$ where $E$ is continuous on $[0, c]$. Hence there exists $c_0$ such that $0 < c_0 < c$, $D'(t) > 0$ for $0 < t \leqq c_0$, $D(c_0) > 0$ and for $h$ and $k$ sufficiently large $k/h < D(c_0)$. This implies that there exists a unique $\tau \in (0, c_0)$ such that $D(\tau) = k/h$ which is equivalent to $hF'(\tau) + kG'(\tau) = 0$. Moreover from the definition of $D$ we have

$$\tau = \left(\frac{-\nu k g(0)}{\lambda h f(0)}\right)^{1/\lambda - \nu} (1 + o(1))$$

which implies that $\tau^{\lambda-\nu} = o(k/h) = o(1)$.

If we now let $H(t) = F(t) + (k/h)G(t)$ and expand $h(H(t) - H(\tau))$ about $t = \tau$ we have using the integral form of the remainder

$$h(H(t) - H(\tau)) = h\int_\tau^t (t - y)F''(y)dy + k\int_\tau^t (t - y)G''(y)dy$$
$$= hR(t, \tau) + kS(t, \tau), \text{ respectively.}$$

We may further assume that $c_0$ is so small that $f, f', f'', g, g'$ and $g''$ are of constant sign for $0 \leqq t \leqq c_0$. If we apply the mean value theorem for integrals and substitute $t = \tau(x + 1)$ we have for $-1 < x < 1$ that

$$T(x, \tau) = R(\tau(x + 1), \tau) = \frac{\tau^\lambda x^2}{2}[\lambda(\lambda - 1)f(\tau_0(x))\alpha_0(x)$$
$$+ \tau(\lambda + 1)\alpha_1(x)f'(\tau_1(x)) + \tau^2 f''(\tau_2(x))\alpha_2(x)] = \frac{\tau^\lambda x^2}{2} P_1(x)$$

where $\alpha_0, \alpha_1, \alpha_2 \in C^\infty$, $\alpha_0(0) = \alpha_2(0) = \alpha_1(0) = 1$, $P_1 \in C^4$ and $P_1(0) = \lambda(\lambda - 1)f(0) + o(1)$. Similarly

$$W(x, \tau) = S(\tau(x + 1), \tau) = \tfrac{1}{2}\tau^\nu x^2 P_2(x)$$

where $P_2 \in C^4$ and $P_2(0) = \nu(\nu - 1)g(0) + o(1)$. Let $d_0 = (c_0/\tau) - 1$, $I'(h, k) = \exp(-ihH(\tau))I(h, k)$ and choose $b$ such that $\tau(b + 1) < c_0$ and $0 < b < 1$.

$$I'(h, k) = \int_0^{c_0} + 0\left(\frac{1}{h}\right) = I''(h, k) + 0\left(\frac{1}{h}\right).$$

$$I''(h, k) = \tau\int_{-1}^{-b} + \tau\int_{-b}^{b} + \tau\int_b^{d_0} Q(\tau(x + 1))\exp[i(hT(x, \tau) + kW(x, \tau))]dx$$
$$= L(h, k) + I'''(h, k) + J(h, k), \text{ respectively.}$$

Let

$\mu = \tau$, $p = (1/2)h\tau^\lambda$, $\psi = q$, $\omega(x) = (1 + x)^{\gamma-1}$ and $\phi(x) = P_1(x) + (k\tau^{\nu-\lambda}/h)P_2(x)$.

Then $\phi(0) = \lambda(\lambda - \nu)f(0) + o(1)$ implies for $h, k$ sufficiently large that $\phi(x) > 0$ for $-c' \leqq x \leqq c'$, $c' > 0$. Hence $b$ may be chosen small enough that the conditions on $b$ in Lemma II are satisfied. Therefore

$$I'''(h, k) \backsim \frac{\sqrt{2}\,q(0)\Gamma\left(\frac{1}{2}\right)\exp\left(\frac{i\pi}{4}\right)}{(\lambda - \nu)^{1/2}}\left[\frac{(\lambda hf(0))^{\nu-2\gamma}}{(-\nu kg(0))^{\lambda-2\gamma}}\right]^{1/2(\lambda-\nu)}.$$

The contribution of $L(h, k)$ to $I(h, k)$ may be determined by considering

$$L'(h, k) = \int_0^{\tau(1-b)} Q(t)\exp(ihH(t))dt .$$

We note that the uniqueness of $\tau$ in $[\varepsilon, c_0]$, $\varepsilon > 0$, implies that $H'(t) \neq 0$ on $\varepsilon \leqq t \leqq \tau(1-b)$ for every $\varepsilon > 0$. In fact, there exists a number $K > 0$ which is independent of $\varepsilon$ and for which we have $|H'(t)| \geqq K(k/h)t^{\nu-1}$ for $\varepsilon \leqq t \leqq \tau(1-b)$.

(i) For $\nu < \gamma$ the usual integration by parts together with the above inequality for $H'(t)$ yields that $L'(h,k) = o(1/k)$. Hence $L'(h,k) = o(h^{\nu-2\gamma}/k^{\lambda-2\gamma})^{1/2(\lambda-\nu)}$ since $h^\nu = o(k^\lambda)$.

(ii) For $\nu > \gamma$ we rewrite $L'(h,k)$ as

$$L'(h,k) = \int_0^{\tau(1-b)} Q(t)\exp\{-i[kt^\nu(-g(t)) + ht^\lambda(-f(t))]\}dt$$

and apply Theorem II with $-g$ playing the role of $f$. Hence

$$L'(h,k) \backsim \frac{q(0)\Gamma\left(\frac{\gamma}{\nu}\right)\exp\left(\frac{-i\pi\gamma}{2\nu}\right)}{\nu[-kg(0)]^{\gamma/\nu}} .$$

(iii) Finally for $\nu = \gamma$ a closer examination of the proof of Lemma I together with the change of variable $t = xh^{-1/\lambda}$ implies for $p' = k/h^{\nu/\lambda}$ that $L'(h,k) = h^{-\gamma/\lambda}0(1/p') = 0(1/k)$.

The given relation $h^\nu = o(k^\lambda)$ and the calculation

$$k^{\gamma/\nu}0\left(\left(\frac{h^{\nu-2\gamma}}{k^{\lambda-2\gamma}}\right)^{1/2(\lambda-\nu)}\right) = 0\left(\left(\frac{h}{k^{\lambda/\nu}}\right)^{\nu-2\gamma/2(\lambda-\nu)}\right)$$

then imply that $I'''(h,k) = o(k^{-\gamma/\nu})$ if $\nu > 2\gamma$ and $L'(h,k) = o((h^{\nu-2\gamma}/k^{\lambda-2\gamma})^{1/2(\lambda-\nu)})$ if $\gamma \leqq \nu < 2\gamma$. When $\nu = 2\gamma$ we note that both $L'(h,k)$ and $I'''(h,k)$ are of the same order so that both terms contribute to $I(h,k)$.

To complete the proof of Theorem III we need only show that $J(h,k)$ is negligible compared to $I'''(h,k)$. For $P(t)$ defined as in the proof of Theorem I and $d = \tau(b+1)$ we have

$$P(d) = \lambda f(0)[1-(b+1)^{\nu-\lambda}](1+o(1)) .$$

Then $P(d) > 0$ for $h$ and $k$ sufficiently large and hence proceeding exactly as in the proof of Theorem I we obtain $H'(t) \geqq Bt^{\lambda-1} > 0$ for $0 < d \leqq t \leqq c_0$ and $2B = \lambda f(0)[1-(1+b)^{\nu-\lambda}]$. We now write

$$J(h,k) = \int_d^{c_0} Q(t)\exp(ihH(t))dt$$

and integrate by parts as in Theorem I to obtain $J(h,k) = 0((h^{\nu-\gamma}/k^{\lambda-\gamma})^{1/\lambda-\nu})$. Hence $h^\nu = o(k^\lambda)$ implies that

$$\left(\frac{k^{\lambda-2\gamma}}{h^{\nu-2\gamma}}\right)^{1/2(\lambda-\nu)} J(h,k) = 0\left(\left(\frac{h^\nu}{k^\lambda}\right)^{1/2(\lambda-\nu)}\right) = o(1) .$$

To obtain the value of $\exp[i(hF(\tau)+kG(\tau))]$ in a more explicit form we need to know more about the exact relation between $h$ and $k$. For example we shall state the following corollary under more stringent assumptions.

COROLLARY. *If in addition to the above assumptions in Theorem* III *we have* $k^{\lambda+1}=o(h^{\nu+1})$ *then*

$$\exp[i(hF(\tau)+kG(\tau))]\backsim\exp\left\{\frac{i(\nu-\lambda)}{\lambda\nu}\left[\frac{(-\nu kg(0))^{\lambda}}{(\lambda hf(0))^{\nu}}\right]^{1/\lambda-\nu}\right\}.$$

*Proof of the Corollary to Theorem III.* We will use the same notation as in the proof of Theorem III. If we expand $hH(\tau)$ about the origin and substitute for $\tau$ we have

$$\begin{aligned}\exp(\,(ihH(\tau))&=\exp\left\{i\left[hf(0)\left(\frac{-\nu kg(0)}{\lambda hf(0)}\right)^{\lambda/\lambda-\nu}\right.\right.\\&\quad\left.\left.+\,kg(0)\left(\frac{-\nu kg(0)}{\lambda hf(0)}\right)^{\nu/\lambda-\nu}\right]\left[1+0\left(\left(\frac{k}{h}\right)^{1/\lambda-\nu}\right)\right]\right\}\\&=\exp\left\{\frac{i(\nu-\lambda)}{\lambda\nu}\left[\frac{(-\nu kg(0))^{\lambda}}{(\lambda hf(0))^{\nu}}\right]^{1/\lambda-\nu}+0\left(\left(\frac{k^{\lambda+1}}{h^{\nu+1}}\right)^{1/\lambda-\nu}\right)\right\}.\end{aligned}$$

Hence if $k^{\lambda+1}=o(h^{\nu+1})$ the Corollary is established.

Finally, we shall consider the case where lim sup $kh^{-\nu/\lambda}$ is bounded away from both 0 and $\infty$.

THEOREM IV. *Suppose that* $\gamma<\lambda$, $\nu<\lambda$ *and* $0<\limsup p<\infty$ *where* $p=kh^{-\nu/\lambda}$. *Then*

$$\begin{aligned}I(h,k)&\backsim q(0)h^{-\gamma/\lambda}\int_0^\infty x^{\gamma-1}\exp[i(f(0)x^{\lambda}+pg(0)x^{\nu})]dx\\&=q(0)\left(\frac{k}{h}\right)^{\gamma/\lambda-\nu}\int_0^\infty x^{\gamma-1}\exp[ip^{\lambda/\lambda-\nu}(f(0)x^{\lambda}+g(0)x^{\nu})]dx\,.\end{aligned}$$

*Proof of Theorem IV.* We will consider only values of $c>0$ such that (i) $\nu g(t)+tg'(t)$ is of constant sign for $0\leqq t\leqq c$ and (ii) for each $\varepsilon>0$ we have $|q(t)-q(0)|<\varepsilon$, $|f(t)-f(0)|<\varepsilon$ and $|g(t)-g(0)|<\varepsilon$ for $0\leqq t\leqq c$. Set $H(t)=F(t)+(k/h)G(t)$ and $I'(h,k)=\int_0^c$ as usual. Let $m=\text{minimum}_{0\leqq t\leqq a}\lambda f(t)+tf'(t)>0$, $\omega=\limsup p$ and $M=\text{maximum}_{0\leqq t\leqq a}(1,|f^{(l)}|,|q|,|q'|,|g^{(l)}|,|\nu g(t)+tg'(t)|)$ for $l=0,1,2$. Consider a number $b>1$ chosen such that $b>N=(4M\omega/m)^{1/(\lambda-\nu)}$. If $d=bh^{-1/\lambda}<c$ then for $0<d\leqq t\leqq c$ we have for $g(0)<o$

$$H'(t)\geqq t^{\lambda-1}\left(m-\frac{kM}{hd^{\lambda-\nu}}\right)\geqq t^{\lambda-1}\left(m-\frac{mk}{4\omega h^{\nu/\lambda}}\right)\geqq\frac{1}{2}mt^{\lambda-1}$$

since $2\omega > kh^{-\nu/\lambda}$ for $h$ and $k$ sufficiently large. Hence $H'(t) \geqq (1/2)mt^{\lambda-1} > 0$ for $0 < d \leqq t \leqq c$. Let $t = xh^{-1/\lambda}$, $\tilde{q}(x) = q(xh^{-1/\lambda})$, $\tilde{f}(x) = f(xh^{-1/\lambda})$ and $\tilde{g}(x) = g(xh^{-1/\lambda})$. Then

$$h^{\gamma/\lambda}I'(h, k) = \int_0^{ch^{1/\lambda}} x^{\gamma-1}\tilde{q}(x) \exp[i(\tilde{f}(x)x^\lambda + p\tilde{g}(x)x^\nu)]dx$$
$$= \int_0^b + \int_b^{ch^{1/\lambda}} = I''(h, k) + J(h, k), \text{ respectively.}$$

We will first estimate $J(h, k)$ in terms of the number $b$. Since $H'(t) \geqq (1/2)mt^{\lambda-1} > 0$ for $0 < d \leqq t \leqq c$ we may integrate $J(h, k)$ by parts as follows:

$$J(h, k) = h^{\gamma/\lambda}\left\{\frac{Q(c)\exp(ihH(c))}{ihH'(c)} - \frac{Q(d)\exp(ihH(d))}{ihH'(d)}\right.$$
$$- \frac{1}{ih}\int_d^c \frac{Q'(t)\exp(ihH(t))dt}{H'(t)} + \frac{1}{ih}\int_d^c \frac{Q(t)F''(t)\exp(ihH(t))dt}{[H'(t)]^2}$$
$$\left. + \frac{k}{ih^2}\int_d^c \frac{Q(t)G''(t)\exp(ihH(t))dt}{[H'(t)]^2}\right\}$$
$$= 0(h^{\gamma-\lambda/\lambda}) + A + J'(h, k) + J''(h, k) + J'''(h, k), \text{ respectively.}$$

Hence $|A| \leqq 2M/mb^{\lambda-\gamma} = Bb^{\gamma-\lambda}$,

$$|J'(h, k)| \leqq \frac{2Mh^{\gamma/\lambda}}{mh}\int_d^c t^{\gamma-\lambda-1}dt < \frac{2M}{m(\lambda-\gamma)b^{\lambda-\gamma}} = B'b^{\gamma-\lambda},$$

$$|J''(h, k)| \leqq \frac{4M^2}{m^2(\lambda-\gamma)b^{\lambda-\gamma}} = B''b^{\gamma-\lambda}, \text{ and}$$

$$|J'''(h, k)| \leqq \frac{4M^2kh^{-\nu/\lambda}}{m^2(2\lambda-\gamma-\nu)b^{2\lambda-\gamma-\nu}} \leqq \frac{8M^2\omega}{m^2(2\lambda-\gamma-\nu)b^{\lambda-\gamma}} = B'''b^{\gamma-\lambda}.$$

Define

$$h^{\gamma/\lambda}I_0(h, k) = \int_0^\infty x^{\gamma-1}q(0)\exp[i(f(0)x^\lambda + pg(0)x^\nu)]dx = \int_0^b + R(b).$$

Then there exists a number $K$ which is independent of $h, k$ and $\varepsilon$ and for which $|J(h, k)| \leqq Kb^{\gamma-\lambda}$ and $|R(b)| \leqq Kb^{\gamma-\lambda}$. Consider

$$(^*)h^{\gamma/\lambda}(I_0 - I) = \int_0^b x^{\gamma-1}(q(0) - \tilde{q}(x))\exp[i(\tilde{f}(x)x^\lambda + p\tilde{g}(x)x^\nu)]dx$$
$$+ q(0)\int_0^b x^{\gamma-1}(1 - P(x))\exp[i(f(0)x^\lambda + pg(0)x^\nu)]dx$$
$$+ R(b) + 0(h^{\gamma-\lambda/\lambda}) - J(h, k)$$
$$= L(h, k) + L'(h, k) + R(b) + 0(h^{\gamma-\lambda/\lambda}) - J(h, k),$$
respectively,

where $P(x) = \exp\{i[\tilde{f}(x) - f(0))x^\lambda + p(\tilde{g}(x) - g(0))x^\nu]\}$. By the choice of $c$ for each $\varepsilon > 0$ we have $|L(h,k)| \leqq M\varepsilon b^{\lambda+\gamma}$, $|L'(h,k)| < 2M\varepsilon b^{\lambda+\gamma} + 2M\varepsilon\omega b^{\nu+\gamma}$. If we take lim sup of both sides of (*) as $h, k \to \infty$ we obtain

$$0 \leqq \limsup h^{\gamma/\lambda} |I_0 - I| \leqq 3M\varepsilon b^{\lambda+\gamma} + 2M\varepsilon b^{\nu+\gamma} + 2Kb^{\gamma-\lambda}$$

which is true for $\varepsilon > 0$ and $b > N$. Since $\limsup h^{\gamma/\lambda}|I_0 - I|$ is independent of both $\varepsilon$ and $b$ we first let $\varepsilon \to 0$ and then $b \to \infty$. Hence $h^{\gamma/\lambda}(I_0 - I) = o(1)$ which implies that $I(h,k) \backsim I_0(h,k)$. To obtain the alternate form of $I_0(h,k)$ we let $x = h^{1/\lambda}(k/h)^{1/(\lambda-\nu)}t$.

**5. Discussion of the suggested application.** Consider for $x > 0$ Schlafli's generalization of Bessel's integral:

$$\begin{aligned} J_\nu(x) &= \frac{1}{\pi}\int_0^\pi \cos(\nu t - x\sin t)dt - \frac{\sin\nu\pi}{\pi}\int_0^\infty \exp[-\nu t - x \sin h\, t]dt \\ &= \frac{1}{\pi} R\int_0^\pi \exp[i(\nu t - x\sin t)]dt + 0\left(\frac{1}{\nu}\right). \end{aligned}$$

Let $F(t) = t - \sin t$ and $|G(t)| = t$. We rewrite $F(t)$ as $F(t) = (1/6)t^3 \cos(r(t))$ and let $h = x$, $k = |\nu - x|$, $q(t) \equiv 1$ and $f(t) = 1\backslash 6\cos(r(t))$. It follows that the condition $3f(t) + tf'(t) > 0$ for $0 \leqq t \leqq \pi$ is satisfied since $F'(t) = 1 - \cos t > 0$ for $0 < t \leqq \pi$.

We note that our Theorem I and III yield the dominant terms of some well known complete asymptotic expansions for $J_\nu(x)$ with $\tau=$Arc-cos $\nu/x$ in Theorem III[1]. For the case $0 < \limsup x^{-1/3}|\nu - x| < \infty$ we have by Theorem IV with $p = x^{-1/3}|(\nu - x)x^{-1/3}|$ that

$$J_\nu(x) \backsim \frac{1}{\pi x^{1/3}}\int_0^\infty \cos\left(\frac{1}{6}t^3 + pt\right)dt$$

where the expression on the right is one of Airy's integrals[2], whose evaluation for $p > 0$ and $p < 0$ yields precisely Nicholson's formulas when $\nu$ is an integer[3].

[1] See W. Magnus and F. Oberhettinger, "Formeln und Satze fur die Speziellen Funktionen der Mathematischen Physik," Springer-Verlag, Berlin, 1948, pp. 33-34. Our theorems I and III give results which are equivalent to the dominant terms of the expansions ($b_3$) and ($b_1$), respectively.

[2] See, for example, G. N. Watson, "Theory of Bessel Functions," Cambridge, 1944, pp. 188-190.

[3] See G. N. Watson, op. cit., pp. 248-249.

INSTITUTE OF TECHNOLOGY
UNIVERSITY OF MINNESOTA

# BOUNDS OF ANALYTIC FUNCTIONS OF TWO COMPLEX VARIABLES IN DOMAINS WITH THE BERGMAN-SHILOV BOUNDARY

J. ŚLADKOWSKA

**Introduction.** From the function-theoretic point of view, the three-dimensional boundary of a domain (in the space of two complex variables) does not play a role analogous to the boundary curve in the theory of one variable. In order to be able to use methods similar to those in one variable, Bergman introduces analytic polyhedra, i.e., domains bounded by finitely many segments of analytic hypersurfaces.[1] On the three-dimensional boundary of an analytic polyhedron lies a two-dimensional manifold which, from the function-theoretic point of view, plays a role similar to that of the boundary curve. In studying the value distribution of holomorphic and meromorphic functions in an analytic polyhedron, we can distinguish with Bergman two types of problems:

(1) derivation of bounds for a function in terms of values on the (two-dimensional) distinguished boundary (the so-called Bergman-Šilov boundary),

(2) studies of the relations between the value distribution on the complementary part of the boundary and in the interior of the domain. While studies of problems of type (1) proceed along the lines similar to those in the case of one variable (through repeated use of the Cauchy and Poisson-Jensen formula, etc.), the investigation of problems of type (2) has a different character. Bergman and Charzyński considered the case of functions $f(z_1, z_2)$ which belong to a normal family in every lamina. For instance, they assume $f(z_1, z_2)$ to be a Schlicht function in every lamina. In this case it is possible to obtain bounds for $|f|$ in terms of its maximum along a *one-dimensional* boundary manifold. In the present paper, the investigation of problems of type (2) is continued, and we assume that the function $f$ in every lamina is mean multivalent of order $p$ (see § 1 for details). The order $p = p(\lambda)$ is a function of the parameter $\lambda$[2]; $p(\lambda)$ is square-integrable.

Let $\mathfrak{S}^2$ be a segment of an analytic surface $\mathfrak{S}_0^2$ which intersects the polyhedron. We obtain bounds for $|f(z_1, z_2)|$, $(z_1, z_2) \in \mathfrak{S}^2$, in terms of

(a) the minimum and the maximum of $|f|$ on the one-dimensional manifold mentioned before,

(b) a quantity connected with $p(\lambda)$,

Received May 16, 1962. This work was done under NSF Grant 10375.

[1] An analytic hypersurface is a one-parameter family of analytic surfaces called a laminas.

[2] The laminas of a segment of an analytic hypersurface depend on a parameter $\lambda$.

(c) certain constants which depend only upon the domain and the surface $\mathfrak{G}_0^2$.

0. **Definitions and notations.** We shall consider an arbitrary bounded domain $\mathfrak{B}$ lying in the space of two complex variables $z_1, z_2, z_k = x_k + iy_k, k = 1, 2$. We assume that the boundary $\mathfrak{b}^3$ of this domain consists of finitely many segments

$$\mathfrak{e}_k^3, \qquad k = 1, \cdots, n, \tag{0.1}$$

of analytic hypersurfaces. Every such segment is given by a parametric representation of the form

$$z_1 = h_{1k}(Z_k, \lambda_k), \quad z_2 = h_{2k}(Z_k, \lambda_k), \tag{0.2}$$

where $h_{1k}(Z_k, \lambda_k)$ and $h_{2k}(Z_k, \lambda_k)$ are continuously differentiable functions of $Z_k, \lambda_k$ in the set $\{(Z_k, \lambda_k): |Z_k| \leqq 1, 0 \leqq \lambda_k \leqq 2\pi\}$. For a fixed $k$ and $\lambda_k$ the corresponding set of points (0.2) will be called a lamina of $\mathfrak{e}_k^3$ and designated $\mathfrak{J}_k^2(\lambda_k)$. We assume that

$$\mathfrak{J}_k^2(\lambda_k') \cap \mathfrak{J}_k^2(\lambda_k'') = 0 \quad \text{if } \lambda_k' \neq \lambda_k'', \tag{0.3}$$

and that for fixed $\lambda_k$

$$(h_{1k}(Z_k', \lambda_k), h_{2k}(Z_k', \lambda_k)) \neq (h_{1k}(Z_k'', \lambda_k), h_{2k}(Z_k'' \lambda_k)). \tag{0.4}$$

The set $\mathfrak{F}^2$ of points (0.2) corresponding to the values $|Z_k|=1, k=1, \cdots, n$, constitutes the so-called Bergman-Šilov boundary surface of $\mathfrak{B}$ on which the maximum principle holds for functions regular in $\overline{\mathfrak{B}}$ (see [1]). We shall also assume that for every $|Z_k^{(0)}| < 1, \lambda_k^{(0)}, k = 1, \cdots, n$, and for sufficiently small $\sigma > 0$, the set of points (0.2) which correspond to the values

$$|Z_k - Z_k^{(0)}| < \sigma, \quad |\lambda_k - \lambda_k^{(0)}| < \sigma$$

of the parameters contain all the points of $b^3$ lying sufficiently near the point

$$z_1^{(0)} = h_{1k}(Z_k^{(0)}, \lambda_k^{(0)}), \quad z_k^{(0)} = h_{2k}(Z_k^{(0)}, \lambda_k^{(0)}).$$

The set of points of four-dimensional space of the form

$$z_1 = g_1(\zeta), z_2 = g_2(\zeta), \zeta \in \overline{\mathfrak{D}}, \tag{0.5}$$

where $\mathfrak{D}$ is a domain in the $\zeta$-plane, and the expressions on the right-hand sides of (0.5) are holomorphic functions of $\zeta$ in $\mathfrak{D}$ and continuous in $\overline{\mathfrak{D}}$, is called an analytic surface.

The set of points which corresponds to the values $\zeta \in \partial(\mathfrak{D})$[3] will be

[3] $\partial(\mathfrak{D})$ = boundary of $\mathfrak{D}$

called the boundary of the surface.

The complement of $\mathfrak{B}$ with respect to the whole space will be called $\mathfrak{P}_0$.

1. **Bounds for the function $f(z_1, z_2)$ on the analytic surface.** Let $\mathfrak{B}$ be a domain described in § 0 and let $\mathfrak{S}_0^2$ denote an analytic surface of the form (0.5). We assume that $\mathfrak{S}_0^2$ has common points with $\mathfrak{B}$ and its whole boundary lies in $\mathfrak{P}_0$. Further, let the intersection $\mathfrak{S}_0^2$ with $\mathfrak{B}$ satisfy the following conditions:

1°. The intersection is a segment

$$\mathfrak{S}^2 = \mathfrak{S}_0^2 \cap \mathfrak{B} = \{(z_1, z_2): z_1 = g_1(\zeta), z_2 = g_2(\zeta), |\zeta| < 1\} .$$

Here $g_1(\zeta)$, $g_2(\zeta)$ are analytic functions which are regular in $|\zeta| < 1$ and continuous in $|\zeta| \leqq 1$.

2°. The boundary curve $\mathfrak{g}^1$ of $\mathfrak{S}^2$ is the intersection $\overline{\mathfrak{S}}_0^2$ with $\mathfrak{b}^3$. We assume that

$$g^1 = \{(z_1, z_2): z_1 = g_1(e^{i\varphi}), z_2 = g_2(e^{i\varphi}), 0 \leqq \varphi \leqq 2\pi\}$$

can be divided into $J$ parts

$$\begin{aligned} \mathfrak{g}_j^1 &= \{(z_1, z_2): z_1 = g_1(e^{i\varphi}), z_2 = g_2(e^{i\varphi}), \varphi_j \leqq \varphi \leqq \varphi_{j+1}\} , \\ j &= 1, \cdots, J, \varphi_1 < \varphi_2 < \cdots < \varphi_{J+1} = \varphi_1 + 2\pi , \end{aligned}$$

so that $\mathfrak{g}_j^1 \in \mathfrak{e}_{k_j}^3$, $k_{j_1} \neq k_{j_2}$ for $j_1 \neq j_2$ and only the points

$$(g_1(e^{i\varphi_j}), g_2(e^{i\varphi_j})), j = 1, \cdots, J ,$$

belong to $\mathfrak{F}^2$.

3°. Every point of $\mathfrak{g}_j^1$ lies in a certain lamina, say

$$\mathfrak{F}_{k_j}^2(\lambda_{k_j}) = \{(z_1, z_2): z_1 = h_{1k_j}(Z_{k_j}, \lambda_{k_j}), z_2 = h_{2k_j}(Z_{k_j}, \lambda_{k_j})\} .$$

Hence, by (0.3) and (0.4), functions $\lambda_{k_j} = \lambda_{k_j}(\varphi)$ and $Z_{k_j} = Z_{k_j}(\varphi)$, $\varphi_j \leqq \varphi \leqq \varphi_{j+1}$, exist such that

$$\begin{aligned} \mathfrak{g}_j^1 = \{(z_1, z_2): z_1 = h_{1k_j}(Z_{k_j}(\varphi), \lambda_{k_j}(\varphi)), z_2 = h_{2k_j}(Z_{k_j}(\varphi), \lambda_{k_j}(\varphi)), \\ \varphi_j \leqq \varphi \leqq \varphi_{j+1}\} . \end{aligned}$$

We assume that $\lambda_{k_j}(\varphi)$, $Z_{k_j}(\varphi)$, $j = 1, \cdots, J$, are continuous and that $\lambda_{k_j}(\varphi)$ are also monotone in the intervals $\langle\varphi_j, \varphi_{j+1}\rangle$. Therefore, the derivatives $\lambda'_{k_j}(\varphi)$ exist almost everywhere.

4°. Since $\lambda_{k_j}(\varphi)$ are monotone in $\langle\varphi_j, \varphi_{j+1}\rangle$, there exist inverse functions $\varphi_j(\lambda_{k_j})$ in the intervals $\langle\alpha_j, \beta_j\rangle = \lambda_{k_j}(\langle\varphi_j, \varphi_{j+1}\rangle)$. The derivatives

$\varphi_j'(\lambda_{k_j})$ also exist in $\langle\alpha_j, \beta_j\rangle$ almost everywhere. We shall assume that

$$|\varphi_j'(\lambda_{k_j})| \leqq Q\,, \qquad j = 1, \cdots, J\,.$$

5°. The intersection $\mathfrak{S}_0^2$ is such that the expressions $1 - |Z_{k_j}(\varphi)|$ go to zero no faster than some positive power of $\varphi - \varphi_j$ or $\varphi - \varphi_{j+1}$ if $\varphi \to \varphi_j +$ or $\varphi \to \varphi_{j+1} -$, respectively.

The hypotheses 1°, 2°, and 3° are the same as hypotheses 1, 2, and 3 in [4], p. 188. Instead of hypothesis 6, [4] we have the weaker hypothesis 5°.[4]

We define now a family of functions in a domain $\mathfrak{B}$. The function $f(z_1, z_2)$ defined in $\overline{\mathfrak{B}}$ will be called the function of the family $\mathscr{F}_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$, $P > 0$, if it satisfies the following conditions:

1°°. $f(z_1, z_2)$ is regular in the set $\mathfrak{B}_1 = \overline{\mathfrak{B}} \backslash \mathfrak{F}^2$ continuous in $\mathfrak{B}_2 = \mathfrak{B}_1 \cup \mathfrak{g}^1 \cap \mathfrak{F}^2$.

2°°. $f(z_1, z_2) \neq 0$ in $\mathfrak{B}_2$.

3°°. On almost every lamina $\mathfrak{F}_{k_j}^2(\lambda_{k_j})$, $\alpha_j \leqq \lambda_{k_j} \leqq \beta_j$, the function $f(z_1, z_2) = f(h_{1k_j}(Z_{k_j}, \lambda_{k_j}), h_{2k_j}(Z_{k_j}, \lambda_{k_j}))$ considered as a function of one variable $Z_{k_j}$ in the circle $|Z_{k_j}| < 1$ is a mean multivalent function of the order $p_j(\lambda_{k_j})$ in the sense of Biernacki, see [5], [7].[5]

4°°. The functions $p_j(\lambda_{k_j})$ may grow to infinity, but in such a way that they are square-integrable in $\langle\alpha_j, \beta_j\rangle$.

5°°. $$\sum_{j=1}^{J}\left(\frac{1}{2\pi}\int_{\alpha_j}^{\beta_j} p_j^2(\lambda_{k_j})d\lambda_{k_j}\right)^{1/2} \leqq JP\,.$$ [6]

Definition. *Every $f(z_1, z_2)$ which belongs to $\mathscr{F}_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$ will be called a mean multivalent function of the order $P$ with respect to $\mathfrak{S}_0^2$.*

We set

$$l = \min\,[1, \min_{\substack{\alpha_j \leqq \lambda_{k_j} \leqq \beta_j \\ j=1,\cdots,J}} |f(h_{1k_j}(0, \lambda_{k_j}), h_{2k_j}(0, \lambda_{k_j}))|]\,, \tag{1.1}$$

$$L = \max\,[1, \max_{\substack{\alpha_j \leqq \lambda_{k_j} \leqq \beta_j \\ j=1,\cdots,J.}} |f(h_{1k_j}(0, \lambda_{k_j}), h_{2k_j}(0, \lambda_{k_j}))|]\,. \tag{1.1'}$$

[4] From hypothesis 6 it follows that $1-|Z_{k_j}(\varphi)|$ must go to zero no faster than $1/\log|\varphi-\varphi_j|$.

[5] A function $f(z)$ regular in $|z| < 1$ is called mean multivalent of order $p$ in the sence of Biernacki if

$$p(R) = \frac{1}{2\pi}\int_0^{2\pi} n(Re^{i\theta})d\theta \leqq p$$

for every positive number $R$. Here $n(Re^{i\theta})$ designate the number of $Re^{i\theta}$ — points of $f(z)$ in $|z| < 1$.

[6] The integrals here are in the sense of Lebesgue.

THEOREM 1. *For every* $\varepsilon > 0$ *there exists* $r_0$, $0 < r_0 < 1$, *so that at every point of* $G^2$, *say at* $z_1^0 = g_1(\zeta_0)$, $z_2^0 = g_2(\zeta_0)$, *the function* $f(z_1, z_2) \in \mathscr{F}_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$ *satisfies the inequality*

$$(*) \quad \left(e^{-\varepsilon} l\left(\frac{1-r}{1+r}\right)^{2JPQ}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)} \leqq |f(z_1^0, z_2^0)| \leqq \left(e^{\varepsilon} L\left(\frac{1+r}{1-r}\right)^{2JPQ}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)}$$

*for every* $r \geqq r_0$.

*Proof.* First, we prove the last inequality in (*). Let $\varepsilon > 0$. By hypothesis 5° there exist positive numbers $a_j$ and $b_j$ such that

$$\lim_{\varphi \to \varphi_j +} \frac{1 - |Z_{k_j}(\varphi)|}{(\varphi - \varphi_j)^{a_j}} \text{ and } \lim_{\varphi \to \overline{\varphi_{j+1}}} \frac{1 - |Z_{k_j}(\varphi)|}{(\varphi_{j+1} - \varphi)^{b_j}}$$

are different from zero.[7] Hence, there are positive numbers, say $A_j$ and $B_j$, and a positive number $\eta'$ such that

$$(1.2) \qquad 1 - |Z_{k_j}(\varphi)| > A_j(\varphi - \varphi_j)^{a_j}$$

and

$$(1.2') \qquad 1 - |Z_{k_j}(\varphi)| > B_j(\varphi_{j+1} - \varphi)^{b_j}$$

for $0 < \varphi - \varphi_j < \eta'$ or $0 < \varphi_{j+1} - \varphi < \eta'$, respectively. Further, since the functions

$$\omega_{1j}(x) = x \log^2 \frac{2}{A_j x^{a_j}} + 2a_j x \log \frac{2}{A_j x^{a_j}} + 2a_j^2 x$$

and

$$\omega_{2j}(x) = x \log^2 \frac{2}{B_j x^{b_j}} + 2b_j x \log \frac{2}{B_j x^{b_j}} + 2b_j^2 x$$

go to zero for $x \to 0+$, there exists an $\eta'' > 0$ such that

$$(1.3) \qquad \omega_{1j}(x) < \frac{\pi\varepsilon^2}{8QJ^2P^2}, \quad \omega_{2j}(x) = \frac{\pi\varepsilon^2}{8QJ^2P^2}$$

for $0 < x < \eta''$. If we set now

$$\eta = \min(\eta', \eta''),$$

then the inequalities (1.2), (1.2') and (1.3), respectively, are satisfied for $0 < \varphi - \varphi_j < \eta$, $0 < \varphi_{j+1} - \varphi < \eta$ and $0 < x < \eta$.

Let $(z_1^0, z_2^0) \in \mathfrak{S}^2$. Then $z_1^0 = g_1(\zeta_0)$, $z_2^0 = g_2(\zeta_0)$. We consider now the function $f(z_1, z_2)$ in the segment $\overline{\mathfrak{S}}^2$, i.e., $f(g_1(\zeta), g_2(\zeta))$ in $|\zeta| \leqq 1$. Considered as a function of $\zeta$ it is regular in $|\zeta| < 1$ and continuous in

[7] $a_j, b_j$ may be infinite.

$|\zeta| \leqq 1$. For $\zeta = e^{i\varphi}$ it has the following bounds

$$|f(g_1(e^{i\varphi}), g_2(e^{i\varphi}))| \leqq L\Big(\frac{1 + |Z_{k_j}(\varphi)|}{1 - |Z_{k_j}(\varphi)|}\Big)^{2p_j(\lambda_{k_j}(\varphi))} \quad \text{for } \varphi \in (\varphi_j, \varphi_{j+1}), \qquad j = 1, \cdots, J. \tag{1.4}$$

This is a consequence of the fact that $f(z_1, z_2)$ assumes at the point $(g_1(e^{i\varphi}), g_2(e^{i\varphi}))$ the values of the multivalent functions $f(h_{1k_j}(Z_{k_j}, \lambda_{k_j}), h_{2k_j}(Z_{k_j}, \lambda_{k_j}))$ (of order $p_j(\lambda_{k_j}(\varphi))$) at the point $Z_{k_j}(\varphi)$, see [7], p. 116. We divide the line $\mathfrak{g}_j^1$ into two parts $\check{\mathfrak{g}}_j^1$ and $\check{\check{\mathfrak{g}}}_j^1$ as follows

$$\check{\mathfrak{g}}_j^1 = \{(z_1, z_2): z_1 = g_1(e^{i\varphi}), z_2 = g_2(e^{i\varphi}), \varphi \in \langle \varphi_j + \eta, \varphi_{j+1} - \eta\rangle\},$$
$$\check{\check{\mathfrak{g}}}_j^1 = \{(z_1, z_2): z_1 = g_1(e^{i\varphi}), z_2 = g_2(e^{i\varphi}), \varphi \in \langle \varphi_j, \varphi_j + \eta) \cup (\varphi_{j+1} - \eta, \varphi_{j+1}\rangle\}.$$

It is easy to see that $r_0$ exists such that for every point $(g_1(e^{i\varphi}), g_2(e^{i\varphi}))$ the inequality $|Z_{k_j}(\varphi)| \leqq r_0$ holds (this follows from the continuity of the functions $Z_{k_j}(\varphi)$). Therefore, for these points, the inequalities (1.4) give the following bounds

$$|f(g_1(e^{i\varphi}), g_2(e^{i\varphi}))| \leqq L\Big(\frac{1 + r}{1 - r}\Big)^{2p_j(\lambda_{k_j}(\varphi))} \tag{1.5}$$

for every $r \geqq r_0$ and for $\varphi \in \langle \varphi_j + \eta, \varphi_{j+1} - \eta\rangle$, $j = 1, \cdots, J$. On the complementary part of $\mathfrak{g}^1$[8] we have the inequalities

$$|f(g_1(e^{i\varphi}), g_2(e^{i\varphi}))| \leqq L\Big(\frac{2}{A_j(\varphi - \varphi_j)^{\alpha_j}}\Big)^{2p_j(\lambda_{k_j}(\varphi))} \quad \text{for } \varphi \in (\varphi_j, \varphi_j + \eta), \tag{1.6}$$

and

$$|f(g_1(e^{i\varphi}), g_2(e^{i\varphi}))| \leqq L\Big(\frac{2}{B_j(\varphi_{j+1} - \varphi)^{b_j}}\Big)^{2p_j(\lambda_{k_j}(\varphi))} \quad \text{for } \varphi \in (\varphi_{j+1} - \eta, \varphi_{j+1}). \tag{1.6'}$$

This follows from (1.4), (1.2) and (1.2′).

Applying now the Poisson formula to the function $\log|f(g_1(\zeta), g_2(\zeta))|$, which is harmonic in $|\zeta| < 1$ and continuous in $|\zeta| \leqq 1$, and using the inequalities (1.5), (1.6) and (1.6′) we obtain

$$\begin{aligned}
\log|f(z_1^0, z_2^0)| &= \frac{1}{2\pi}\int_0^{2\pi} \log|f(g_1(e^{i\varphi}), g_2(e^{i\varphi}))|\, re\frac{e^{i\varphi} + \zeta_0}{e^{i\varphi} - \zeta_0}d\varphi \\
&= \sum_{j=1}^{J}\Big[\frac{1}{2\pi}\int_{\varphi_j+\eta}^{\varphi_{j+1}-\eta} \log L\Big(\frac{1 + r}{1 - r}\Big)^{2p_j(\lambda_{k_j}(\varphi))} re\frac{e^{i\varphi} + \zeta_0}{e^{i\varphi} - \zeta_0}d\varphi \\
&\quad + \frac{1}{2\pi}\int_{\varphi_j}^{\varphi_{j+1}} \log L\Big(\frac{2}{A_j(\varphi - \varphi_j)^{\alpha_j}}\Big)^{2p_j(\lambda_{k_j}(\varphi))} re\frac{e^{i\varphi} + \zeta_0}{e^{i\varphi} - \zeta_0}d\varphi \\
&\quad + \frac{1}{2\pi}\int_{\varphi_{j+1}-\eta}^{\varphi_{j+1}} \log L\Big(\frac{2}{B_j(\varphi_{j+1} - \varphi)^{b_j}}\Big)^{2p_j(\lambda_{k_j}(\varphi))} re\frac{e^{i\varphi} + \zeta_0}{e^{i\varphi} - \zeta_0}d\varphi\Big]
\end{aligned}$$

[8] Except for the points $(g_1(e^{i\varphi_j}), g_2(e^{i\varphi_j}))$.

Interchanging the variables of integration and applying the Schwarz inequality to the last two integrals we get

$$\begin{aligned}\log|f(z_1^0, z_2^0)| &\leqq \frac{1+|\zeta_0|}{1-|\zeta_0|}\sum_{j=1}^{J}\Big[\frac{1}{2\pi}\int_{\varphi_j}^{\varphi_{j+1}}\log L\,d\varphi \\ &+ 2\log\Big(\frac{1+r}{1-r}\Big)\Big|\frac{1}{2\pi}\int_{\lambda_{k_j}(\varphi_j+\eta)}^{\lambda_{k_j}(\varphi_{j+1}-\eta)} p_j(\lambda_{k_j})\varphi_j'(\lambda_{k_j})d\lambda_{k_j}\Big| \\ &+ \frac{2}{2\pi}\sqrt{\Big|\int_{\lambda_{k_j}(\varphi)}^{\lambda_{k_j}(\varphi_j+\eta)} p_j^2(\lambda_{k_j})\varphi_j'(\lambda_{k_j})d\lambda_{k_j}\Big|}\sqrt{\int_{\varphi_j}^{\varphi_j+\eta}\log^2\frac{2}{A_j(\varphi-\varphi_j)^{a_j}}d\varphi} \\ &+ \frac{2}{2\pi}\sqrt{\Big|\int_{\lambda_{k_j}(\varphi_{j+1}-\eta)}^{\lambda_{k_j}(\varphi_{j+1})} p_j^2(\lambda_{k_j})\varphi_j'(\lambda_{k_j})d\lambda_{k_j}\Big|}\sqrt{\int_{\varphi_{j+1}-\eta}^{\varphi_{j+1}}\log^2\frac{2}{B_j(\varphi_{j+1}-\varphi)^{b_j}}d\varphi}\Big].\end{aligned}$$

Evaluating the integrals

$$\int_{\varphi_j}^{\varphi_j+\eta}\log^2\frac{2}{A_j(\varphi-\varphi_j)^{a_j}}d\varphi \quad\text{and}\quad \int_{\varphi_{j+1}-\eta}^{\varphi_{j+1}}\log^2\frac{2}{B_j(\varphi_{j+1}-\varphi)^{b_j}}d\varphi\ ,$$

using hypotheses $4^\circ$, $4^{\circ\circ}$ and $5^{\circ\circ}$[9] and inequalities (1.3), we have

$$\begin{aligned}\log|f(z_1^0, z_2^0)| &\leqq \frac{1+|\zeta_0|}{1+|\zeta_0|}\Big[\log L + Q\cdot 2\log\Big(\frac{1+r}{1-r}\Big)\sum_{j=1}^{J}\frac{1}{2\pi}\int_{\alpha_j}^{\beta_j}p_j(\lambda_{k_j})d\lambda_{k_j} \\ &\quad + \frac{2}{2\pi}\sqrt{Q}\sum_{j=1}^{J}\Big(\sqrt{\frac{1}{2\pi}\int_{\lambda_{k_j}(\varphi_j)}^{\lambda_{k_j}(\varphi_j+\eta)}p_j^2(\lambda_{k_j})d\lambda_{k_j}}\cdot\sqrt{\omega_{1j}(\eta)} \\ &\quad + \sqrt{\frac{1}{2\pi}\int_{\lambda_{k_j}(\varphi_{j+1}-\eta)}^{\lambda_{k_j}(\varphi_{j+1})}p_j^2(\lambda_{k_j})d\lambda_{k_j}}\cdot\sqrt{\omega_{2j}(\eta)}\Big) \\ &\leqq \frac{1+|\zeta_0|}{1-|\zeta_0|}\Big[\log L + 2QJP\log\Big(\frac{1+r}{1-r}\Big) \\ &\quad + \frac{2}{\sqrt{2\pi}}\sqrt{Q}\cdot\frac{\sqrt{\pi}\cdot\varepsilon}{2\sqrt{2}\sqrt{Q}\,JP}\cdot 2JP\Big] \\ &= \frac{1+|\zeta_0|}{1-|\zeta_0|}\Big(\log L\Big(\frac{1+r}{1-r}\Big)^{2PQJ} + \varepsilon\Big).\end{aligned}$$

Therefore, finally

$$|f(z_1^0, z_2^0)| \leqq \Big(e^{\varepsilon}L\Big(\frac{1+r}{1-r}\Big)^{2JPQ}\Big)^{(1+|\zeta_0|)/(1+|\zeta_0|)} \tag{1.7}$$

which is the first inequality of (*). We notice now that

[9] From $5^{\circ\circ}$ and the Schwarz inequality, we get

$$\begin{aligned}\sum_{j=1}^{J}\Big|\frac{1}{2\pi}\int_{\lambda_{k_j}(\varphi_j+\eta)}^{\lambda_{k_j}(\varphi_{j+1}-\eta)}p_j(\lambda_{k_j})d\lambda_{k_j}\Big| &\leqq \sum_{j=1}^{J}\frac{1}{2\pi}\int_{\alpha_j}^{\beta_j}p_j(\lambda_{k_j})d\lambda_{k_j} \\ &\leqq \sum_{j=1}^{J}\frac{1}{2\pi}\sqrt{\beta_j-\alpha_j}\sqrt{\int_{\alpha_j}^{\beta_j}p_j^2(\lambda_{k_j})d\lambda_{k_j}} \leqq \sum_{j=1}^{J}\Big(\frac{1}{2\pi}\int_{\alpha_j}^{\beta_j}p_j^2(\lambda_{k_j})d\lambda_{k_j}\Big)^{1/2} \leqq JP\,.\end{aligned}$$

$$\frac{1}{f(z_1, z_2)} \in \mathscr{F}_{\mathfrak{B}}(\mathfrak{S}_0^2, P) \quad \text{if} \quad f(z_1, z_2) \in \mathscr{F}_{\mathfrak{B}}(\mathfrak{S}_0^2, P)^{10} .$$

Moreover,

$$\begin{aligned}\max\Big(1, \max_{\alpha_j \leqq \lambda_{k_j} \leqq \beta_j} &\Big|\frac{1}{f(h_{1k_j}(0, \lambda_{k_j}), h_{2k_j}(0, \lambda_{k_j}))}\Big|\Big) \\ &= \max\Big(1, \frac{1}{\min\limits_{\alpha_j \leqq \lambda_{k_j} \leqq \beta_j} |f(h_{1k_j}(0, \lambda_{k_j}), h_{2k_j}(0, \lambda_{k_j}))|}\Big) \\ &\leqq \frac{1}{\min(1, \min\limits_{\alpha_j \leqq \lambda_{k_j} \leqq \beta_j} |f(h_{1k_j}(0, \lambda_{k_j}), h_{2k_j}(0, \lambda_{k_j}))|)} = \frac{1}{l} .\end{aligned}$$

Applying the inequality (1.7) to the function $1/f(z_1, z_2)$, we obtain the inequality

$$\text{(1.8)} \qquad \frac{1}{f(z_1, z_2)} \leqq \left(e^{\varepsilon} \frac{1}{l}\left(\frac{1+r}{1-r}\right)^{2JPQ}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)}$$

for $r \geqq r_0$; $r_0$ is here the same as in (1.7), because it is independent of the function. From (1.8) we have

$$\text{(1.9)} \qquad |f(z_1^0, z_2^0)| \geqq \left(e^{-\varepsilon} l\left(\frac{1+r}{1-r}\right)^{2JPQ}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)}$$

The inequalities (1.7) and (1.9) give the conclusion of the theorem.

REMARK 1. Modifying the definition of the family $\mathscr{F}_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$, we obtain somewhat simpler analogous results. Instead of hypothesis $4^{\circ\circ}$ we assume that the function $p_j(\lambda_{k_j}(\varphi))$ considered as a function of the variable $\varphi$ is square-integrable in the interval $\langle\varphi_j, \varphi_{j+1}\rangle$, and we replace condition $5^{\circ\circ}$ by the condition

$$\sum_{j=1}^{J} \left(\frac{1}{2\pi}\int_{\varphi_j}^{\varphi_{j+1}} p_j^2(\lambda_{k_j}(\varphi))d\varphi\right)^{1/2} \leqq JP .$$

The assumptions that $\lambda_{k_j}(\varphi)$ are continuous and monotonic and that $|\varphi'(\lambda_{k_j})| \leqq Q$ are now superfluous. The family of functions which satisfy these conditions will be called $\mathscr{F}^*_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$. For functions of that family we can prove

THEOREM 1′. *For every* $\varepsilon > 0$ *there exists* $r_0$, $0 < r_0 < 1$, *such that for every point* $(z_1^0, z_2^0) \in \mathfrak{S}^2$ *and for every* $f(z_1, z_2) \in \mathscr{F}^*_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$ *the inequality*

$$\text{(1.10)} \quad \left(l\left(\frac{1-r}{1+r}\right)^{2JP} e^{-\varepsilon}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)} \leqq |f(z_1^0, z_2^0)| \leqq \left(L\left(\frac{1+r}{1-r}\right)^{2JP} e^{\varepsilon}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)}$$

[10] If $f(z) \neq 0$ and is mean-multivalent of order $p$ in the sense of Biernacki, then $1/f(z)$ has the same property.

*holds for every* $r \geqq r_0$.

The number $r_0$ is chosen in the following way. Let $\varepsilon > 0$ be an arbitrary number. The number $\eta'$ is chosen in the same way as in the proof given above; $\eta''$ is such that

$$\sqrt{\omega_{1j}(x)} < \frac{\pi\varepsilon^2}{8J^2P^2} \quad \text{and} \quad \sqrt{\omega_{2j}(x)} < \frac{\pi\varepsilon^2}{8J^2P^2}$$

for $0 < x < \eta''$. We set $\eta = \min(\eta', \eta'')$ and for $\eta$ we choose a number $r_0$ as previously.

REMARK 2. If the surface $\mathfrak{S}_0^2$ intersects only one boundary segment, say $\mathfrak{e}_k^3$, and the line of intersection $\mathfrak{g}^1$ lies $\mathfrak{e}_{kr_0}^3$, where

$$\mathfrak{e}_{kr_0}^3 = \{(z_1, z_2): z_1 = h_{1k}(Z_k, \lambda_k), |Z_k| \leqq r_0\} ,$$

then

$$(1.11) \quad \left(l\left(\frac{1-\mathrm{r}}{1+r}\right)^{2PQ}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)} \leqq |f(z_1^0, z_2^0)| \leqq \left(L\left(\frac{1+r}{1-r}\right)^{2PQ}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)}$$

for every $r \geqq r_0$ in the case that $f(z_1, z_2) \in \mathscr{F}_{\mathfrak{B}}(\mathfrak{S}_2^2, P)$ and

$$(1.11') \quad \left(l\left(\frac{1-r}{1+r}\right)^{2P}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)} \leqq |f(z_1^0, z_2^0)| \leqq \left(L\left(\frac{1+r}{1-r}\right)^{2P}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)}$$

if $f(z_1, z_2) \in \mathscr{F}^*_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$.

Indeed, for every $\varphi$ for which $(g_1(e^{i\varphi}), g_2(e^{i\varphi})) \in e_{kr_0}^3$, the corresponding point $Z_k(\varphi)$ satisfies the inequality

$$(1.12) \qquad |Z_k(\varphi)| \leqq r_0 ,$$

and therefore

$$|f(g_1(e^{i\varphi}), g_2(e^{i\varphi}))| \leqq L\left(\frac{1+r}{1-r}\right)^{2p_k(\lambda_k(\varphi))}$$

for $\varphi \in \langle 0, 2\pi \rangle$. Applying, as previously, the Poisson formula and using the inequality (1.12), we obtain (1.11) in the first case and (1,11′) in the second.

REMARK 3. The result of Theorems 1 and 1′ can be obtained without requiring that $\mathfrak{B}$ is an analytic polyhedron. It is sufficient to assume that the part of the boundary which intersects by $\mathfrak{S}_0^2$ is a sum of the analytic hypersurfaces mentioned in § 0. Concerning the complementary part of the boundary no special hypotheses are needed.

REMARK 4. The lower and upper bounds of $|f|$ are expressed in terms of the minimum and the maximum of $|f|$ on the manifold

$$\bigcup_{j=1}^{J} \bigcup_{\lambda_{k_j} \in \langle \alpha_j, \beta_j \rangle} (h_{1k_j}(0, \lambda_{k_j}), h_{2k_j}(0, \lambda_{k_j})) .$$

We note that analogous bounds can be obtained in term of the minimum and the maximum of $|f|$ on a manifold

$$\bigcup_{j=1}^{J} \bigcup_{\lambda_{k_j} \in \langle \alpha_j, \beta_j \rangle} (h_{1k_j}(Z_{k_j}(\lambda_{k_j}), \lambda_{k_j}), h_{2k_j}(Z_{k_j}(\lambda_{k_j}), \lambda_{k_j})) ,$$

where $|Z_{k_j}(\lambda_{k_j})| < 1$ and $Z_{k_j}(\lambda_{k_j})$ are continuously differentiable functions of $\lambda_{k_j} \in \langle \alpha_j, \beta_j \rangle$. These new bounds are obtained by changing the parametric representations of $\mathrm{e}^3_{k_j}$, $j = 1, \cdots, J$, as follows:

$$\mathrm{e}^3_{k_j} = \{(z_1, z_2): z_1 = \tilde{h}_{1k_j}(\tilde{Z}_{k_j}, \lambda_{k_j}), z_2 = \tilde{h}_{2k_j}(\tilde{Z}_{k_j}, \lambda_{k_j})\} ,$$

$|\tilde{Z}_{k_j}| \leqq 1$, where

$$\tilde{h}_{\kappa k_j}(\tilde{Z}_{k_j}, \lambda_{k_j}) = h_{\kappa k_j}\left(\frac{\tilde{Z}_{k_j} + A_{k_j}(\lambda_{k_j})}{1 - A_{k_j}(\lambda_{k_j})\tilde{Z}_{k_j}}, \lambda_{k_j}\right) ,$$

Here

$$A_{k_j}(\lambda_{k_j}) = \begin{cases} Z_{k_j}(\alpha_j) & \text{for } \lambda_{k_j} \in \langle 0, \alpha_j) , \\ Z_{k_j}(\lambda_{k_j}) & \text{for } \lambda_{k_j} \in \langle \alpha_j, \beta_j \rangle , \\ Z_{k_j}(\beta_j) & \text{for } \lambda_{k_j} \in (\beta_j, 2\pi\rangle . \end{cases}$$

REMARK 5. Let $0 < R < 1$ and let

$$\mathfrak{S}^2_R = \{(z_1, z_2): z_1 = g_1(\zeta), z_2 = g_2(\zeta), |\zeta| < R\} .$$

Then for every $(z_1, z_2) \in \mathfrak{S}^2_R$ and for every $f(z_1, z_2) \in \mathscr{F}_{\mathfrak{B}}(\mathfrak{S}^2_0, P)$ the inequality

$$(**) \quad \left(e^{-\varepsilon} l\left(\frac{1 - r_0}{1 + r_0}\right)^{2JPQ}\right)^{(1+R)/(1-R)} \leqq |f(z_1, z_2)| \leqq \left(e^{\varepsilon} L\left(\frac{1 + r_0}{1 - r_0}\right)^{2JPQ}\right)^{(1+R)/(1-R)}$$

holds; here $r_0$ depends only upon $\mathfrak{B}$, $\mathfrak{S}^2_0$ and $\varepsilon$.

Let $\{\mathfrak{S}^2\}_{\mathfrak{r}_0}$ be the set of all segments $\mathfrak{S}^2$ of analytic surfaces $\mathfrak{S}^2_0$ for which the conditions 1°–5° are fulfilled and for which the set of the corresponding numbers $r_0$ has an upper bound smaller than $\mathfrak{r}_0 < 1$. We set

$$\mathfrak{S}_R = \bigcup_{\mathfrak{S}^2 \in \{\mathfrak{S}^2\}_{\mathfrak{r}_0}} \mathfrak{S}^2_R .$$

For every $(z_1, z_2) \in \mathfrak{S}_R$ the inequality

$$(***) \quad \left(e^{-\varepsilon} l\left(\frac{1 - r_0}{1 + r_0}\right)^{2JPQ}\right)^{(1+R)/(1-R)} \leqq |f(z_1, z_2)| \leqq \left(e^{\varepsilon} L\left(\frac{1 + r_0}{1 - r_0}\right)^{2JPQ}\right)^{(1+R)/(1-R)}$$

holds. Corresponding to $\{\mathfrak{S}^2\}_{\mathfrak{r}_0}$ we define a sequence of sets $\{\mathfrak{A}_n\}$ by induction as follows:

1. $\mathfrak{A}_1 = \mathfrak{S}_R$.

2. $\mathfrak{A}_{n+1}$ is the set of all points $(z_1, z_2) \in \mathfrak{B} \backslash (\overline{\mathfrak{A}}_1 \cup \cdots \cup \overline{\mathfrak{A}}_n)$ which belong to at least one of the analytic surfaces $\mathfrak{S}_n^2$ lying in $\mathfrak{B}$ and having its boundary in $\mathfrak{A}_1 \cup \cdots \cup \mathfrak{A}_n$. The sum of all the sets $\mathfrak{A}_n$ will be denoted by $\mathfrak{A}_R$ and called the associated domain corresponding to the set $\{\mathfrak{S}^2\}_{r_0}$ and to the number $R$. We can prove (similarly to [6], p. 33) that the inequality (***) holds in the full set $\mathfrak{A}_R$ and consequently also in its closure $\overline{\mathfrak{A}}_R$.

2. **The case of a bounded** $p(\lambda)$. If we replace the hypotheses $4^{\circ\circ}$ and $5^{\circ\circ}$ by the condition

$$(2.1) \qquad p_j(\lambda_{k_j}) \leqq P \quad \text{for } \lambda_{k_j} \in \langle \alpha_j, \beta_j \rangle, \, j = 1, \cdots, J \,,$$

the function $f$ which satisfies the hypotheses $1^{\circ\circ}$–$3^{\circ\circ}$ and the condition (2.1) belongs to the family $\mathscr{F}_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$ and even to the family $\mathscr{F}^*_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$[11]. For these functions the inequality (*) follows from Theorem 1′. However, repeating the proof of theorem 1′ and using the condition (2,1) yields a better result.

THEOREM 2. *For every $\varepsilon > 0$ there exists $r_0$, $0 < r_0 < 1$, such that for every point $(z_1^0, z_2^0) \in \mathfrak{S}^2$ and for every function $f \in \mathscr{F}^*_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$ the inequalities*

$$(2.2) \quad \left(l\Big(\frac{1-r}{1+r}\Big)^{2P} e^{-\varepsilon}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)} \leqq |f(z_1^0, z_2^0)| \leqq \left(L\Big(\frac{1+r}{1-r}\Big)e^{\varepsilon}\right)^{(1+|\zeta_0|)/(1-|\zeta_0|)}$$

*hold for every $r \geqq r_0$ if $f$ satisfies condition* (2.1).

The proof of Theorem 2 proceeds in a way analogous to that of Theorem 1′. Let $\eta = \min(\eta', \eta'')$, where $\eta'$ has the same meaning as in the proof on p. 8, and $\eta'' > 0$ is chosen in such a way that for $0 < x < \eta''$

$$\hat{\omega}_{1j}(x) = x \log \frac{2}{A_i} - a_j x \log x - a_j x < \frac{\pi}{2JP} \varepsilon \,,$$

$$\hat{\omega}_{2j}(x) = x \log \frac{2}{B_j} - b_j x \log x - b_j x < \frac{\pi}{2JP} \varepsilon$$

hold. We choose $r_0$ in the same way as before. If we assume, instead of hypothesis $5^\circ$, that $1 - |Z_{k_j}(\varphi)|$ goes to zero no faster than $(\varphi - \varphi_j)^{a_j}$ or $(\varphi_{j+1} - \varphi)^{b_j}$, where $0 < \alpha_j, b_j < 1/2P$, when $\varphi \to \varphi_j +$ or $\varphi \to \varphi_{j+1} -$, respectively (hypothesis $5^{\circ\prime}$), we can obtain a better inequality.

THEOREM 3. *For every sufficiently small $\varepsilon > 0$ there exists $r_0$, $0 < r_0 < 1$ such that for every point $(z_1^0, z_2^0) \in \mathfrak{S}^2$ and for every function $f \in \mathscr{F}^*_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$ the inequalities*

[11] Indeed, the functions $p_j(\lambda_{k_j}(\varphi))$, $\varphi \in \langle \varphi_j, \varphi_{j+k} \rangle$, being bounded, are square-integrable.

$$(2.3)\quad \frac{1-|\zeta_0|}{1+|\zeta_0|}\Big(l\Big(\frac{1-r}{1+r}\Big)^{2P}-\varepsilon\Big) \leqq |f(z_1^0, z_2^0)| \leqq \frac{1+|\zeta_0|}{1-|\zeta_0|}\Big(L\Big(\frac{1+r}{1-r}\Big)^{2P}+\varepsilon\Big)^{12}$$

*hold for every* $r \geqq r_0$, *if the function* $f$ *satisfies the condition* (2.1) *and if, instead of hypothesis* 5°, *hypothesis* 5°′ *is fulfilled.*

*Proof.* Let $\varepsilon > 0$ and $\varepsilon < 1/l$. It follows from hypothesis 5°′ that there exist numbers $A_j, B_j > 0$ and $\eta' > 0$ such that

$$(2.4)\quad 1-|Z_{k_j}(\varphi)| > A_j(\varphi-\varphi_j)^{a_j} \quad\text{and}\quad 1-|Z_{k_j}(\varphi)| > B_j(\varphi_{j+1}-\varphi)^{b_j}$$

for $0 < \varphi - \varphi_j < \eta'$ and $0 < \varphi_{j+1} - \varphi < \eta'$, respectively. Let $\eta'' > 0$ be a number such that, for $0 < x < \eta''$,

$$(2.5)\qquad \breve{\omega}_{1j}(x) = L\left(\frac{2}{A_j}\right)^{2P}\frac{x^{1-2Pa_j}}{1-2Pa_j} < \frac{\pi}{J}\varepsilon\,,$$

and

$$(2.5')\qquad \breve{\omega}_{2j}(x) = L\left(\frac{2}{B_j}\right)^{2P}\frac{x^{1-2Pb_j}}{1-2Pb_j} < \frac{\pi}{J}\varepsilon$$

hold. We set $\eta = \min(\eta', \eta'')$. There exists $r_0$, $0 < r_0 < 1$, such that

$$(2.6)\qquad |Z_{k_j}(\varphi)| \leqq r_0 \quad\text{for } \varphi \in (\varphi_j+\eta, \varphi_{j+1}-\eta),\ j = 1, \cdots, J\,.$$

Applying the Cauchy formula to the function $f(g_1(\zeta), g_2(\zeta))$ which is regular in $|\zeta| < 1$ and continuous in $|\zeta| \leqq 1$, dividing the interval of integration and using the inequalities (2.6), (1.4), (2.4), (2.5), and (2.5′), we obtain

$$\begin{aligned}(2.7)\quad |f(z_1^0, z_2^0)| &= |f(g_1(\zeta_0), g_2(\zeta_0))| = \left|\frac{1}{2\pi}\int_0^{2\pi} f(g_1(e^{i\varphi}), g_2(e^{i\varphi}))\, re\,\frac{e^{i\varphi}+\zeta_0}{e^{i\varphi}-\zeta_0}\,d\varphi\right| \\ &\leqq \frac{1+|\zeta_0|}{1-|\zeta_0|}\sum_{j=1}^{J}\Big[\frac{1}{2\pi}\int_{\varphi_j+\eta}^{\varphi_{j+1}-\eta} L\Big(\frac{1+r}{1-r}\Big)^{2P} d\varphi \\ &\quad + \frac{1}{2\pi}\int_{\varphi_j}^{\varphi_j+\eta} L\Big(\frac{2}{A_j(\varphi-\varphi_j)^{a_j}}\Big)^{2P} d\varphi \\ &\quad + \frac{1}{2\pi}\int_{\varphi_{j+1}-\eta}^{\varphi_{j+1}} L\Big(\frac{2}{B_j(\varphi_{j+1}-\varphi)^{b_j}}\Big) d\varphi\Big] \\ &\leqq \frac{1+|\zeta_0|}{1-|\zeta_0|}\Big[L\Big(\frac{1+r}{1-r}\Big)^{2P}+\varepsilon\Big]\,.\end{aligned}$$

If we apply the inequality obtained above to the function $1/f(z_1, z_2)$, which also belongs to $\mathscr{F}^*_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$ and for which the condition (2.1) holds, we have the inequality

[12] Here $l = \min|f(h_{1k_j}(0, \lambda_{k_j}), h_{2k_j}(0, \lambda_{k_j}))|$, $L = \max|f(h_{1k_j}(0, \lambda_{k_j}), h_{2k_j}(0, \lambda_{k_j}))|$, $l$ may be larger than 1 or $m$ smaller than 1.

$$\left|\frac{1}{f(z_1, z_2)}\right| \leqq \frac{1+|\zeta_0|}{1-|\zeta_0|}\left(\frac{1}{l}\left(\frac{1+r}{1-r}\right)^{2P} + \varepsilon\right)^{13}$$

for $r \geqq r_0$. Hence,

$$(2.8) \quad |f(z_1^0, z_2^0)| \geqq \frac{1-|\zeta_0|}{1+|\zeta_0|}\frac{1}{\frac{1}{l}\left(\frac{1+r}{1-r}\right)^{2P} + \varepsilon} > \frac{1-|\zeta_0|}{1+|\zeta_0|}\left(l\left(\frac{1-r}{1+r}\right)^{2P} - \varepsilon\right).$$

From (2.7) and (2.8), (2.3) follows.

REMARK 1. The inequality on the right hand side of (2.3) is obtained in the same way as an inequality obtained by Bergman (see [4] p. 190). Bergman assumes that the function $f$ omits the values 0 and 1 in every lamina and, instead of the inequality (1.5), he applies an inequality, which follows from the Schottky theorem.

The case when $\mathfrak{S}_0^2$ intersects $\mathfrak{b}^3$ along only one segment $\mathfrak{e}_k^3$ so that the line of the intersection $\mathfrak{g}^1$ lies in $\mathfrak{e}_{kr_0}^3$ is of special interest. This case is considered in remark 2 of § 1. We assume there that the function $f$ belongs to the family $\mathscr{F}_{\mathfrak{B}}^*(\mathfrak{S}_0^2, P)$. However, if we assume in addition that $p(\lambda_k) \leqq P$ (this means that $f(h_{1k}(Z_k, \lambda_k), h_{2k}(Z_k, \lambda_k))$ is mean multivalent of at most order $P$ in every lamina $\mathfrak{J}_k^2(\lambda_k)$ for which $\mathfrak{J}_k^2(\lambda_k) \cap \mathfrak{g}^1 \neq 0$) we obtain a better result, using, instead of the Poisson formula, the minimum and maximum principles (see [6], p. 31). This method yields the following theorem:

THEOREM 4. *If* $\mathfrak{g}^1 \subset \mathfrak{e}_{kr_0}^3$, $f \in \mathscr{F}_{\mathfrak{B}}^*(\mathfrak{S}_0^2, P)$ *and the additional condition* $p(\lambda_k) \leqq P$ *is satisfied on every lamina* $\mathfrak{J}_k^2(\lambda_k)$ *which is intersected by* $\mathfrak{g}^1$, *then for every* $(z_1^0, z_2^0) \in \mathfrak{S}^2$ *and for every* $r \geqq r_0$ *the inequality*

$$l\left(\frac{1-r}{1+r}\right)^{2P} \leqq |f(z_1^0, z_2^0)| \leqq L\left(\frac{1+r}{1-r}\right)^{2P}$$

*holds. Here,*

$$l = \min_{\lambda_k \in s_k} |f(h_{1k}(0, \lambda_k), h_{2k}(0, \lambda_k))|$$
$$L = \max_{\lambda_k \in s_k} |f(h_{1k}(0, \lambda_k), h_{2k}(0, \lambda_k))|$$

$s_k$ *designates the set of* $\lambda_k$ *for which* $\mathfrak{J}_k^2(\lambda_k) \cap \mathfrak{g}^1 \neq 0$.

REMARK 1. Bergman [2], [3], [4] obtained an upper bound for $|f|$ on an analytic surface, which intersects $\mathfrak{b}^3$ along a line lying in $\mathfrak{e}_{kr_0}^3$, under the assumption that $f$ is a univalent function in every lamina $\mathfrak{J}_k^2(\lambda_k)$.

---

13 $$\max_{\alpha_j \leqq \lambda_{k_j} \leqq \beta_j}\left|\frac{1}{f(h_{1k_j}(0, \lambda_{k_j}), h_{2k_j}(0, \lambda_{k_j}))}\right| = \frac{1}{\min\limits_{\alpha_j \leqq \lambda_{k_j} \leqq \beta_j} |f(h_{1k_j}(0, \lambda_{k_j}), h_{2k_j}(0, \lambda_{k_j}))|} = \frac{1}{l}.$$

The bound is expressed in terms of the maximum of $|f|$ and of the maximum of the absolute value of the derivate of $f(h_{1k}(Z_k, \lambda_k), h_{2k}(Z_k, \lambda_k))$ with respect $Z_k$ on a one-dimensional manifold lying on $\mathfrak{b}^3$.

**3. Example.** Let $\Phi$ be a univalent function in $|t| < 1$, continuous in $|t| \leqq 1$, $|\Phi(t)| < 1$ for $|t| \leqq 1$ and $t \neq \exp(i\lambda_1^0)$, $|\Phi[\exp(i\lambda_1^0)]| = 1$. In addition we assume that

$$\int_0^{2\pi} \frac{d\lambda_1}{(1 - |\Phi(e^{i\lambda_1})|)^4} \tag{3.1}$$

exists. Let

$$\mathfrak{B} = \{(z_1, z_2): z_1 = Z, z_2 = \Phi(t), |Z| < 1, |t| < 1\} .$$

$\mathfrak{B}$ is a domain which is obtained from the bicylinder $|Z| < 1, |t| < 1$ by pseudo-conformal mapping $z_1 = Z$, $z_2 = \Phi(t)$. Its three-dimensional boundary $\mathfrak{b}^3$ consists of two segments, say $\mathfrak{e}_1^3$ and $\mathfrak{e}_2^3$, of analytic hypersurfaces:

$$\mathfrak{e}_1^3 = \{(z_1, z_2): z_1 = Z, z_2 = \Phi(e^{i\lambda_1}), |Z| \leqq 1, 0 \leqq \lambda_1 \leqq 2\pi\}$$
$$\mathfrak{e}_2^3 = \{(z_1, z_2): z_1 = e^{i\lambda_2}, z_2 = \Phi(t), |t| \leqq 1, 0 \leqq \lambda_2 \leqq 2\pi\} .$$

$\mathfrak{B}$ is obviously an analytic polyhedron. The Bergman-Šilov boundary of $\mathfrak{B}$ is a two-dimensional manifold

$$\mathfrak{F}^2 = \{(z_1, z_2): z_1 = e^{i\lambda_1}, z_2 = \Phi(e^{i\lambda_2}), 0 \leqq \lambda_1 \leqq 2\pi, 0 \leqq \lambda_2 \leqq 2\pi\} .$$

Let $\mathfrak{G}_0^2$ be a plane

$$z_1 = r_0 e^{i\theta_0}, 0 < r_0 < 1 , \quad \theta_0 \text{ real number} .$$

The common part $\overline{\mathfrak{G}}^2 = \mathfrak{G}_0^2 \cap \overline{\mathfrak{B}}$ can be represented in the form

$$\overline{\mathfrak{G}}^2 = \{(z_1, z_2): z_1 = r_0 e^{i\theta_0}, z_2 = \Phi(\zeta), |\zeta| \leqq 1\} .$$

The intersection $\mathfrak{g}^1 = \mathfrak{G}_0^2 \cap \mathfrak{b}^3$ has the parametric representation

$$\mathfrak{g}^1 = \{(z_1, z_2): z_1 = r_0 e^{i\theta_0}, z_2 = \Phi(e^{i\varphi}), 0 \leqq \varphi \leqq 2\pi\} .$$

Here, $\varphi = \lambda_1$ and $(d\varphi)/(d\lambda_1) = 1$. $\mathfrak{G}_0^2$ intersects the segment $\mathfrak{e}_1^3$ only, and the line of the intersection $\mathfrak{g}^1$ lies in $\mathfrak{e}_{1r_0}^3$. We consider the function

$$f(z_1, z_2) = \exp\left(\frac{1}{(1 - z_1 z_2)^2} - 1\right) . \tag{3.2}$$

It is holomorphic in $\mathfrak{B}$; its singularities lie on the line

$$\mathfrak{S}^1 = \{(z_1, z_2): z_1 = e^{i\psi}, z_2 = \Phi[\exp(i\lambda_1^0)], 0 \leqq \psi \leqq 2\pi\} ,$$

which belongs to $\mathfrak{F}^2$. $f(z_1, z_2)$ is different from zero and holomorphic in the segment $\overline{\mathfrak{G}}^2$ ($\overline{\mathfrak{G}}^2$ has no common points with $\mathfrak{F}^2$). Now, we shall prove that on every lamina

$$\mathfrak{J}_1^2(\lambda_1) = \{(z_1, z_2): z_1 = Z, z_2 = \Phi(e^{i\lambda_1}), |Z| \leqq 1\}, \tag{3.3}$$

except on lamina $\mathfrak{J}_1^2(\lambda_1^0)$, the function (3.2) is mean multivalent of order

$$p(\lambda_1) = \frac{1}{\pi}\left(2 + \frac{1}{(1 - |\Phi(e^{i\lambda_1})|)^2}\right) + 1$$

in the sense of Biernacki.

Let $a$ be an arbitrary complex number such that

$$|a| \leqq e^{-2} \quad \text{or} \quad |a| \geqq 1. \tag{3.4}$$

We want to estimate the number of $a$-points of function (2.3) in lamina (3.3). This number is equal to the number of $a$-points of the function

$$\exp\left(\frac{1}{(1 - Z\Phi(e^{j\lambda_1}))^2} - 1\right) \tag{3.5}$$

in the circle $|Z| < 1$. We must estimate the number of roots of the equation

$$\exp\left(\frac{1}{(1 - Z\Phi(e^{i\lambda_1}))^2} - 1\right) = a \tag{3.6}$$

which lie in $|Z| < 1$. From (3.6) we have

$$Z = \frac{1}{\Phi(e^{i\lambda_1})}\left(1 - \frac{1}{\sqrt{1 + \log a}}\right).$$

As $|Z| < 1$,

$$\left|1 - \frac{1}{\sqrt{1 + \log \mathrm{a}}}\right| < |\Phi(e^{i\lambda_1})|.$$

Hence, by hypothesis (3.4)

$$\frac{1}{\sqrt{|1 + \log a|}} \geqq 1 - |\Phi(e^{i\lambda_1})|$$

and

$$|1 + \log a| \leqq \frac{1}{(1 - |\Phi(e^{i\lambda_1})|)^2}. \tag{3.7}$$

From (3.7) it follows that

$$|\arg a| \leqq \frac{1}{(1 - |\Phi(e^{i\lambda_1})|)^2}.$$

If we set $\arg a = \operatorname{Arg} a + 2k\pi$, $k = 0, \pm 1, \pm 2, \cdots$, where $|\operatorname{Arg} a| \leqq \pi$, then

$$2|k|\pi - |\operatorname{Arg} a| \leqq \frac{1}{(1-|\Phi(e^{i\lambda_1})|)^2},$$
$$(2|k|-1)\pi \leqq \frac{1}{(1-|\Phi(e^{i\lambda_1})|)^2}$$

and finally

$$|k| \leqq \frac{1}{\pi}\left(1+\frac{1}{2}\frac{1}{(1-|\Phi(e^{i\lambda_1})|)^2}\right).$$

The number $n(a, \lambda_1)$ of $a$-points of (3.2) in lamina (3.3) cannot exceed $2|k|+1$; this means that

$$p(|a|, \lambda_1) = \frac{1}{2\pi}\int_0^{2\pi} n(a, \lambda_1) d \arg a \leqq \frac{1}{\pi}\left(2+\frac{1}{(1-|\Phi(e^{i\lambda_1})|)^2}\right)+1$$

for $|a| \leqq e^{-2}$ or $|a| \geqq 1$. For numbers $a$ such that $e^{-2} \leqq |a| \leqq 1$, the corresponding number $p(|a|, \lambda_1)$ is $\leqq 1$. Hence,

$$p(\lambda_1) = \sup_{|a|} p(|a|, \lambda_1) \leqq \frac{1}{\pi}\left(2+\frac{1}{(1-|\Phi(e^{i\lambda_1})|)^2}\right)+1\,.$$

The function $p(\lambda_1)$ becomes infinite as $\lambda_1 \to \lambda_1^0$, but $\int_0^{2\pi} p^2(\lambda_1) d\lambda_1$ exists, as a consequence of (3.1), and

$$(3.8)\quad \frac{1}{2\pi}\int_0^{2\pi} p^2(\lambda_1) d\lambda_1 \leqq \frac{2}{\pi^3}+\frac{2}{\pi^2}+\frac{1}{2\pi} + \left(\frac{2}{\pi^3}+\frac{1}{\pi^2}\right)\int_0^{2\pi}\frac{d\lambda_1}{(1+|\Phi(e^{i\lambda_1})|)^2} + \frac{1}{2\pi^3}\int_0^{2\pi}\frac{d\lambda_1}{(1-|\Phi(e^{i\lambda_1})|)^4},$$

The function (3.2) belongs to the family $\mathfrak{F}_{\mathfrak{B}}(\mathfrak{S}_0^2, P)$, where $\mathfrak{B}$ and $\mathfrak{S}_0^2$ are the domains and the analytic surface described above. $P$ equals a square root of the right-hand side of (3.8). Here, $Q=1, J=1$,

$$l = \min\,(1, \min_{0\leqq\lambda_1\leqq 2\pi} |f(0, \Phi(e^{i\lambda_1})|) = 1$$
$$L = \max\,(1, \max_{0\leqq\lambda_1\leqq 2\pi} |f(0, \Phi(e^{i\lambda_1})|) = 1\,.$$

Applying Theorem 1 and remark 1 of § 1, we can say: for every $(z_1^0, z_2^0) = (r_0 e^{i\theta_0}, \Phi(\zeta_0))$, $|\zeta_0| < 1$, and for every $r \geqq r_0$ the inequalities

$$(3.9)\quad \left(\frac{1-r}{1+r}\right)^{2p(1+|\delta_0|)/(1-|\delta_0|)} \leqq \left|\exp\left(\frac{1}{(1-r_0 e^{i\theta_0}\Phi(\zeta_0))^2}-1\right)\right| \leqq \left(\frac{1+r}{1-r}\right)^{2p(1+|\zeta_0|)/(1-|\zeta_0|)}$$

hold. The inequality on the right-hand side of (3.9) gives a better estimate for $r_0$ and $|\zeta_0|$ sufficiently near to 1, then the inequality

$$\left|\exp\left(\frac{1}{(1-r_0 e^{i\theta_0}\Phi(\zeta_0))^2}-1\right)\right| \leqq \exp\left(\frac{1}{(1-r|\zeta_0|)^2}-1\right),^{14}$$

which we may obtain directly.

## Bibliography


1. S. Bergman, *Über eine in gewissen Bereichen mit Maximumfläche gültige Integraldarstellung der Funktionen zweier komplexer Variabler*, I, II, Math. Z. **39** (1934), 76-94; 605-608.

2. ———, *Über eine Abschätzung von meromorphen Funktionen zweier komplexer Veränderlicher in Bereichen mit ausgezeichneter Randfläche*, Trav. Inst. Math. Tbilissi **1** (1937), 187-204.

3. ———, *Über das Varhalten der Funktionen von zwei komplexen Veränderlichen in Gebieten min einer ausgezeichneten Randfläche*, Compositio Mathematica **14** (1938), 107-123.

4. ———, *Bounds for analytic functions in domains with a distinguished boundary surface*, Math. Z., **63** (1955), 173-194.

5. M. Biernacki, *Sur les fonctions multivalentes d'ordre p*, C. R. Acad. Sci., Paris **204** (1936), 449-451.

6. Z. Charzyński, *Bounds for analytic functions of two complex variables*, Math. Z., **75** (1961), 29-35.

7. W. K. Hayman, *Multivalent functions*, Cambridge University Press, 1958.

8. D. C. Spencer, *On finitely mean valent functions*, II. Trans. Amer. Math. Soc., **48** (1940), 418-435.

9. ———, *On finitely mean valent functions*, Proc, London Math. Soc., (2), **47** (1941), 201-211.


STANFORD UNIVERSITY
UNIVERSITY OF LODZ

[14] $|\Phi(\zeta_0)| \leqq |\zeta_0|$ by Schwarz' lemma.

# HYPONORMAL OPERATORS

JOSEPH G. STAMPFLI

We say a bounded linear transformation $T$ on a Hilbert space $H$ is hyponormal if $\|Tx\| \geqq \|T^*x\|$ for all $x \varepsilon H$ or equivalently if $T^*T - TT^* \geqq 0$. The notion of hyponormality was introduced in [6], through under another name. In [8], Putnam studied properties of the operator $J_\theta = e^{i\theta}T + e^{-i\theta}T^*$, where $T$ is hyponormal. Lemmas 1 through 6 which appear below occur as exercises in [1], and will be quoted without proof. Henceforth the term operator will mean bounded linear transformation.

LEMMA 1. *Let $T$ be a hyponormal operator on the Hilbert space $H$, then $\|(T - zI)x\| \geqq \|(T^* - \bar{z}I)x\|$ for $x \varepsilon H$, i.e. $T - zI$ is hyponormal.*

LEMMA 2. *Let $T$ be hyponormal on $H$; then $Tx = zx$ implies $T^*x = \bar{z}x$.*

LEMMA 3. *Let $T$ be hyponormal on $H$ with $Tx_1 = z_1x_1$, $Tx_2 = z_2x_2$ and $z_1 \neq z_2$. Then $(x_1, x_2) = 0$.*

LEMMA 4. *If $T$ is hyponormal on $H$ and $M \subset H$ is invariant under $T$; then $T|_M$ is hyponormal.*

LEMMA 5. *Let $T$ be hyponormal on $H$, with $M \subset H$ invariant under $T$ and let $T|_M$ be normal. Then $M$ reduces $T$.*

LEMMA 6. *Let $T$ be hyponormal on $H$ and Let $M = \{x \varepsilon H : Tx = zx\}$. Then $M$ reduces $T$ and $T|_M$ is normal.*

THEOREM 1. *Let $T$ be hyponormal on $H$, then $\|T\| = R_{sp}(T)$* (the *spectral radius of $T$*).

*Proof.* For $x \varepsilon H$, $\|x\| = 1$ *we have*

$$\|Tx\|^2 = (Tx, Tx) = (T^*Tx, x) \leqq \|T^*Tx\| \leqq \|T^2x\|.$$

But then $\|T\|^2 \leqq \|T^2\| \leqq \|T\|^2$ which implies $\|T\|^2 = \|T^2\|$.

Now

$$\begin{aligned}\|T^nx\|^2 &= (T^nx, T^nx) = (T^*T^nx, T^{n-1}x) \\ &\leqq \|T^*T^nx\| \cdot \|T^{n-1}x\| \leqq \|T^{n-1}x\| \cdot \|T^{n-1}x\| .\end{aligned}$$

Received February 15, 1962.

Thus $\|T^n\|^2 \leqq \|T^{n-1}\|\cdot\|T^{n-1}\|$, and combining this with the equality above, a simple induction argument yields $\|T^n\| = \|T\|^n$ for $n = 1, 2, \cdots$. Since $R_{sp}(T) = \lim_{n\to\infty} \|T^n\|^{1/n} = \lim_{n\to\infty} \|T\|$ the proof is finished.

COROLLARY. *The only quasi-nilpotent hyponormal operator is the transformation which is identically zero.*

THEOREM 2. *Let $T$ be hyponormal on the Hilbert space $H$ and let $z_0$ be an isolated point in the spectrum of $T$. Then $z_0 \varepsilon \sigma_p(T)$, the point spectrum of $T$.*

*Proof.* By Lemma 1 we may assume $z_0 = 0$. Choose $R > 0$ sufficiently small that 0 is the only point of $\sigma(T)$ contained in or on the circle $|z| = R$. Define

$$E = \int_{|z|=R} (T - zI)^{-1} dz \ .$$

Then $E$ is a nonzero projection which commutes with $T$ (see [9]; projection as used here does not necessarily mean self-adjoint).

Thus $EH$ is invariant under $T$ and $T|_{EH}$ is hyponormal. Also

$$\sigma(T|_{EH}) = \sigma(T) \cap \{|z| < R\}$$

by, [9] p. 421, so $\sigma(T|_{EH}) = \{0\}$ .

From the last corollary we may conclude that $T|_{EH}$ is the zero transformation. In fact, it is now clear that $EH = \{x \varepsilon H: Tx = 0\}$ which implies $EH$ actually reduces $T$.

THEOREM 3. If $T$ is hyponormal on $H$ with a single limit point in its spectrum, then $T$ is normal.

*Proof.* We may assume by Lemma 1 that the limit point is 0. By Theorem 1 there exists $z_1 \varepsilon \sigma(T)$ such that $|z_1| = \|T\|$.

Let $M_1 = \{x \varepsilon H: Tx = z_1 x\}$; $M_1$ is not empty by theorem 2 and since $M_1$ reduces $T$, we conclude from theorem 2 that $T|_{M_1\perp}$ does not have $z_1$ in its spectrum. We note also by Lemma 6 that $T|_{M_1}$ is normal. We continue in this way, selecting points in $\sigma(T)$ ordered by absolute value, setting $M_i = \{x \varepsilon H: Tx = z_i x\}$.

Then $M_1 \oplus \cdots \oplus M_n$ reduces $T$ and $T|_{|M_1\oplus\cdots\oplus M_n}$ is normal. We observe that $T|_{|M_1\oplus\cdots\oplus M_n|\perp}$ is hyponormal with its spectral radius equal to its norm. Thus, since 0 is the only limit point of $\sigma(T)$, the normal operators $T|_{M_1\oplus\cdots\oplus M_n}$ must converge to $T$ in the uniform operator topology. Hence, $T$ is normal.

COROLLARY 1. *If $T$ is a hyponormal, completely continuous opera-*

*tor, then $T$ is normal.*

COROLLARY 2. *If $T$ is hyponormal* on *$H$ with only a finite number of limit points in its spectrum; then $T$ is normal.*

*Proof.* Let $z_1$ be a limit point of $\sigma(T)$ and choose a smooth simple closed curve $G$ which does not intersect $\sigma(T)$ and contains only the limit point $z_1$ in its interior. Now define

$$E_1 = \int_G (T - zI)^{-1} dz.$$

Then $T$ is invariant on $E_1H$ and

$$\sigma(T) \cap [\text{Interior } G] = \sigma(T|_{E_1H})$$

so $\sigma(T|_{E_1H})$ can have only one limit point.

We now apply theorem 3 to $T|E_1H$ to conclude that it is normal. Then by Lemma 5, $T$ is reduced by $E_1H$. We may thus turn our attention to $T$ on $(E_1 H)_\perp$ and continue this process until the limit points are exhausted.

Theorem 1 and the first corollary to Theorem 3 have been proved independently by both T. Ando and S. Berberian, and will soon appear.

The subsequent theorem generalizes the well-known result which equates similarity equivalence of normal operators with unitary equivalence. However, there is a strong restriction on the spectrum of the operator.

THEOREM 4. *If $T$ is a hyponormal scalar operator and $\sigma(T)$ has zero area, then $T$ is normal.*

*Proof.* Since $T$ is scalar, (see [2]), $T = QAQ^{-1}$ where $Q$ is positive self-adjoint and $A$ is normal. Let $A = \int zdE(z)$. For $\varepsilon > 0$, there exists a set of half-open, half-closed disjoint squares $\{R_i\}_{i=1}^k$ with each $R_i$ of dimension $1/n \times 1/n$ such that $k/n^2 = \text{area } (U_{i=1}^k R_i) < \varepsilon$ where

$\sigma(T) \subset (U_{i=1}^k R_i)$. Now for $x_i \varepsilon E(R_i)H$; $z_i$ the center of $R_i$

we have

$$\|(A - z_iI)x_i\| \leqq \left[\int_{R_i} |z - z_i|^2 d\, \|E(z)x_i\|^2\right]^{1/2} \leqq \frac{1}{n} \,.$$

Thus

$$\begin{aligned} \|(A - z_iI)Q^2x_i\| &= \|A^* - \bar{z}_iI)Q^2x_i\| = \|Q(T^* - \bar{z}_iI)Qx_i\| \\ &\leqq \|(T^* - \bar{z}_iI)Qx_i\| \cdot \|Q\| \leqq \|(T - z_iI)Qx_i\| \cdot \|Q\| \end{aligned}$$

$$= \| Q(A - z_iI)x_i \| \cdot \| Q \| \leqq \| Q \|^2 \cdot \frac{1}{n} \quad \text{and}$$

$$\| Q^2(A - z_iI)x_i \| \leqq \| Q \|^2 \cdot \frac{1}{n} .$$

Combining these we have

$$\| (AQ^2 - Q^2A)x_i \| \leqq 2 \| Q \|^2 \cdot \frac{1}{n} \ \text{for}\ x_i \varepsilon E(R_i)H.$$

Now $E(R_i)H$ is orthogonal to $E(R_j)H$ for $i \neq j$ so for $y \varepsilon H$, $\| y \| = 1$ we have

$$y = \textstyle\sum_{i=1}^k a_i x_i \ \text{where}\ \sum_{i=1}^k | a_i |^2 = 1 .$$

Thus

$$\begin{aligned} \| (AQ^2 - Q^2A)y \| &= \| \textstyle\sum_{i=1}^k a_i(AQ^2 - Q^2A)x_i \| \\ &\leqq \textstyle\sum_{i=1}^k | a_i | \, \| (AQ^2 - Q^2A)x_i \| \\ &\leqq \left\{ \textstyle\sum_{i=1}^k | a_i |^2 \sum_{i=1}^k \left( 2 \| Q \|^2 \cdot \frac{1}{n} \right)^2 \right\}^{1/2} = 2 \| Q \|^2 \cdot k^{1/2}/n \\ &\leqq 2 \| Q \|^2 \varepsilon^{1/2} \end{aligned}$$

implying that $AQ^2 = Q^2A$.

Noting that $Q$ is positive we may conclude from the spectral theorem that $AQ = QA$ and thus $T = A$ which completes the proof.

The author has been unable to decide whether the condition on the area of the spectrum in the last theorem may be omitted. He would conjecture that it cannot. There is a generalization of the theorem quoted above which states that if $A$ and $B$ are normal operators, $Q$ an arbitrary operator such that $AQ = QB$; then $A^*Q = QB^*$. This statement does not hold if $A$ is normal and $B$ semi-normal. To see this, let $H$ be a Hilbert space with the basis $\{\phi_i\}_{i=-\infty}^{\infty}$ and define $A\phi_i = \phi_{i+1}$, all $i$; $B\phi_i = \phi_{i+1}$, $i \geqq 0$ $B\phi_i = 0$, $i < 0$; and $Q = B$. Then it is clear that $AQ = QB$ but $A^*Q\phi_0 = \phi_0 \neq QB^*\phi_0 = 0$.

Before going on to the next theorem we must recall some results from the literature. Let $B$ be a normal operator of finite spectral multiplicity $n$, and let $T$ be an operator commuting with $B$. Then there is a finite measure $v(\cdot)$ defined on Borel sets of the complex plane and vanishing outside of $\sigma(B)$ and $n$ Borel sets $e_1, \cdots, e_n$ with $e_1$ the plane and $e_i \subset e_{i+1}$, such that if we define $v_i(e) = v(e \cap e_i)$ for such Borel set $e$; set $\hat{H} = \sum_{i=1}^n L_2(v_i)$ and define

$$\hat{B}f(s) = \hat{B}(f_1(s), \cdots, f_n(s)) = (sf_1(s), \cdots, sf_n(s)) = sf(s)$$

for $f(a)\varepsilon H$; then $B$ and $\hat{B}$ are unitarily equivalent. Also there exist measurable functions $a_{ij}(s)i, j = 1, \cdots, n$ such that if we define

$$Tf(a) = \begin{vmatrix} a_{1,1}(s) & \cdots & a_{1,n}(s) \\ \vdots & & \vdots \\ a_{n,1}(s) & \cdots & a_{n,n}(s) \end{vmatrix} \begin{vmatrix} f_1(s) \\ \vdots \\ f_n(s) \end{vmatrix}$$

for $f(s)\varepsilon\hat{H}$; then $T$ and $T$ are unitarily equivalent. The foregoing may be found in [3] Chapter X theorem 5.10, in [4], [7] and in its earliest form is due to von Neumann.

Using results of Gonshor (see [5] Theorem 3 and remarks in section 6) we may define $\hat{H}$ in such a way that

$$T = \begin{vmatrix} a_{11}(s) & a_{12}(s) & \cdots & a_{1n}(s) \\ 0 & a_{22}(s) & \cdots & a_{2n}(s) \\ \vdots & 0 & & \vdots \\ 0 & & & a_{nn}(s) \end{vmatrix}$$

or roughly speaking $\hat{T}$ has super diagonal form. In what follows we will identify $\hat{T}$ with $T$ and $\hat{B}$ with $B$.

THEOREM 5. *Let $T$ be a hyponormal operator with $T^n = B$ where $n$ is a positive integer and $B$ is a normal operator; then $T$ is normal.*

*Proof.* For $x_0\varepsilon H$, let $M = clm[B^iB^{*j}T^kx_0]$ (the closed linear manifold spanned by the iterates) for $k = 0, 1, \cdots, n-1$; $i, j = 0, 1, 2, \cdots$.

Then $M$ reduces $B$ and $B$ has spectral multiplicity $n$ on $M$ (we will assume $x_0$, $Tx_0$, $\cdots$, $T^{n-1}x_0$ are linearly independent). Also $M$ is invariant under $T$ since $TB^iB^{*j}T^kx_0 = B^iB^{*j}T^{k+1}x_0$ and invariance holds under closure. The Fuglede theorem is used in obtaining the last equality.

Let us now consider $T|_M$ which we may write as

$$\begin{vmatrix} a_{11}(s) & \cdots & a_{1n}(s) \\ 0 & \cdot & \\ & \cdot & \\ & & \cdot \\ & & a_{nn}(s) \end{vmatrix} .$$

Then for the vector $f_1 = (x_{\sigma(B)}, 0, \cdots, 0)$, where $x_{\sigma(B)}$ is the characteristic function of $\sigma(B)$,

we have

$$\| T|_M f_1 \|^2 = \int_{\sigma(B)} | a_{11}(s) |^2 dv_1(s)$$

and

$$\| (T|_M)^* f_1 \|^2 = \sum_{j=1}^n \int_{\sigma(B)} | a_{1j}(s) |^2 dv_j(s) .$$

But $\| T|_M f_1 \| \geqq \| (T|_M)^* f_1 \|$ since the restriction of a hyponormal operator to an invariant subspace is hyponormal. Thus $a_{1j}(s) = 0$ a.e. $(v_j)$ for $j = 2, 3, \cdots, n$. Continuing this argument, we find that $a_{ij}(s) = 0$ a.e. $(v_j)$ for $i \neq j$. We conclude from this that $T|_M$ is normal, or $\| T|_M y \| = \| (T|_M)^* y \|$ for $y \varepsilon M$. Therefore,

$$\| Tx_0 \| = \| T|_M x_0 \| = \| (T|_M)^* x_0 \| \leqq \| T^* x_0 \| .$$

But this with hyponormality implies that $\| Tx_0 \| = \| T^* x_0 \|$. Since $x_0$ is arbitrary $T$ must be normal.

COROLLARY. *If $T$ is hyponormal and commutes with a normal operator having finite spectral multiplicity; then $T$ is normal.*

EXAMPLE. Let $\{\varphi_i\} i = -\infty$ be a basis for the Hilbert space $H$ and define $T\varphi_i = a_i \varphi_{i+1}$. Then if $|a_i| \leqq |a_{i+1}|$, all $i$, $T$ is hyponormal. $T$ will be normal if and only if $|a_i| = |a_{i+1}|$ for all $i$. If $|a_k| = |a_{k+1}|$ for some fixed $k$ and $|a_j| \neq |a_k|$ for some $j > k$ then $T$ will not be subnormal. Another example of hyponormal operator which is not subnormal is given in [6].

## REFERENCE


1. S. K. Berberian, *Introduction to Hilbert Space*, New York, Oxford University Press 1961.
2. N. Dunford, *Spectral operators*, Pacific J. Math., **4** (1954), 321-354.
3. N. Dunford and J. T. Schwartz, *Linear Operators II*, New York, Interscience Publishers, to be published.
4. S. R. Foguel, *Normal operators of finite multiplicity*, Comm. Pure Appl. Math, **11** (1958), 297-313.
5. H. Gonshor, *Spectral theory for a class of non-normal operators*, Can. J. Math, **8** (1956), 449-461.
6. P. R. Halmos, *Normal dilations and extensions of operators*, Summa Bras. Math, **2** (1950), 124-134.
7. M. A. Naimark and S .V. Fomin, *Continuous direct sums of Hilbert spaces and some of their applications*, Amer. Math, Soc. Trans., Vol. 5, series 2, 35-65.
8. C. R. Putnam, *On semi-normal operators*, Pacific J. Math., **7** (1957), 1649-1652.
9. F. Riesz and B. Sz.-Nagy, *Functional Analysis*, New York, Frederick Ungar, 1955.


NEW YORK UNIVERSITY

# SOME EXTREMAL PROPERTIES OF LINEAR COMBINATIONS OF KERNELS ON RIEMANN SURFACES

GEORGES G. WEILL

**1. Introduction.** Let $\Gamma_a$ be the Hilbert space of analytic differentials of finite Dirichlet norm on an open Riemann surface. We shall consider analytic singularities which are finite linear combinations of elements of the type

$$s_j dz = \sum_{k=0}^{\infty} \frac{c_k^j dz}{(z - \xi_j)^{k+2}} + \frac{d^j dz}{z - \xi_j} .$$

Let

$$sdz = \sum_{j=1}^{N} s_j dz , \quad \sum_{j=1}^{N} d^j = 0 .$$

To a given singularity $sdz$ there correspond Bergman kernels

$$k_s(z, \zeta)dz \quad \text{and} \quad h_s(z, \zeta)dz$$

for the space $\Gamma_a$.

We now consider various subspaces $\Gamma_\alpha \subset \Gamma_a$, and show that linear combinations of the kernels for $\Gamma_\alpha$ of the form

$$h_s dz + \lambda k_s dz ,$$

where $\lambda$ is complex, extremalize an explicity given functional.

We proved in our thesis [2] that, for the space $\Gamma_{ae}$ of analytic exact differentials on a *planar* Riemann surface,

$$k_s dz = \frac{1}{2} \frac{\partial}{\partial z}(p_1 - p_0)dz$$
$$h_s dz = \frac{1}{2} \frac{\partial}{\partial z}(p_1 + p_0)dz$$

where $p_1$ and $p_0$ are Sario's principal functions with the corresponding singularities [1, Chapter III].

Here we show that the right hand sides still enjoy the same properties on an arbitrary Riemann surface, for the subspace $\Gamma_p \cap \Gamma_{ase}$, where $\Gamma_{ase} = \left\{adz \colon adz \in \Gamma_a, \int_\gamma adz = 0, \ \gamma \text{ any dividing cycle}\right\}$, and $\Gamma_p$ is generated over the complex numbers by $\{\Gamma_p\} = \{adz\colon adz = \partial p/\partial z, \ p$ a single-valued harmonic function on $W$, with finite Dirichlet integral.$\}$

Received August 21, 1961.

**2. Inner products and singular differentials.** We shall be concerned here with the Hilbert space $\Gamma_a$ of analytic differentials on a given Riemann surface $W$. The inner product of two analytic differentials $adz = \alpha dx + \beta dy$ and $a_1dz = \alpha_1 dx + \beta_1 dy$ is defined as:

$$(adz, a_1dz)_W = -i\int_W a\bar{a}_1 dz\overline{dz} = \int_W (\alpha\bar{\alpha}_1 + \beta\,\bar{\beta}_1)dxdy\ .$$

If we now consider differentials analytic on $W$, except for a singularity of the type $dz/(z-\zeta)^{m+2}$, $m \geqq 0$, we delete a disk $\delta$ of radius $r$ about $z = \zeta$ and define for differentials $bdz$ and $b_1dz$ analytic except for a singularity of the above type, the inner product

$$(bdz, b_1dz)_W = \lim_{r\to 0}\,(bdz, b_1dz)_{W-\delta}\ ,$$

which amounts to considering the Cauchy principal value for the inner product. In the case of a singularity $dz/(z-\zeta_1) - dz/(z-\zeta_2)$, we replace $\delta$ by disks about $z = \zeta_1$ and $z = \zeta_2$, plus a narrow strip along a cut joining $z = \zeta_1$ to $z = \zeta_2$ and define in the same fashion the inner product by a Cauchy limit.

The previous remarks may be extended to finite linear combinations of singularities of the type

$$s_jdz = \sum_{k=0}^{\infty}\frac{c_k^j dz}{(z-\zeta_j)^{k+2}} + \frac{d^j dz}{(z-\zeta_j)}\ ,$$

provided $\sum_{j=1}^N d^j = 0$.

**3. Extremal properties of the kernels.** Let $sdz = \sum_{j=1}^N s_jdz$ be a singularity differential and $k_sdz, h_sdz$ be the Bergman kernels correspond to that singularity. We shall consider linear combinations

$$(h_s + \lambda k_s)dz$$

which are normalized in the sense that they all exhibit the same singularity.

We recall that for $l(z)dz \in \Gamma_a$, the Bergman kernels corresponding to a singularity $sdz$, enjoy the following properties:

$$\text{for } sdz = \frac{dz}{(z-\zeta)^{m+2}}\ ,\ m \geqq 0 \qquad (ldz, k_sdz) = \frac{2\pi l^{(m)}(\zeta)}{(m+1)!}$$

$$(ldz, h_sdz) = 0$$

$$\text{for } sdz = \frac{dz}{z-\zeta_1}\ \frac{dz}{z-\zeta_2} \qquad (ldz, k_sdz) = -(ldz, h_sdz) = 2\pi\int_c ldz$$

where $c$ is a path from $\zeta_1$ to $\zeta_2$.

For $sdz = as_1dz + bs_2dz$, ($a, b$ constant),

$$k_s dz = ak_{s1}dz + bk_{s2}dz$$
$$h_s dz = ah_{s1}dz + bh_{s2}dz \,.$$

Such a linear property is a consequence of the uniqueness of the kernels. Notice that in particular: $(ldz, k_s dz) = \bar{a}(ldz, k_{s1}dz) + \bar{b}(ldz, k_{s2}dz)$. Let now $a_s dz$ be a differential, analytic except for the singularity $sdz$. We form

$$\begin{aligned}(1)\qquad &\| a_s dz - (h_s + \lambda k_s)dz \|^2 = \| a_s dz\|^2 - \| h_s dz \|^2 + | \lambda |^2 \| k_s dz\|^2 \\ &\quad + 2Re((h_s - a_s)dz, h_s dz) + 2Re\bar{\lambda}((h_s - a_s)dz, k_s dz) \,.\end{aligned}$$

Assume now that in a disk about $z = \zeta_j$

$$h_s dz = s_j dz + \sum_{k=0}^{\infty} b_k^j (z - \zeta_j)^k dz$$

$$a_s dz = s_j dz + \sum_{k=0}^{\infty} a_k^j (z - \zeta_j)^k dz \,.$$

We then compute:

$$2Re((h_s - a_s)dz, h_s dz) = -4\pi \sum_{j=1}^{N} Re\bar{d}^j \int_{c_j} (h_s - a_s)dz \,,$$

$$2Re\bar{\lambda}((h_s - a_s)dz, k_s dz) = 4\pi \sum_{j=1}^{N} Re\bar{\lambda}\left[\sum_{k=1}^{\infty} \frac{(b_k^j - a_k^j)\bar{c}_k^j}{k+1} + \bar{d}^j \int_{c_j} (h_s - a_s)dz\right.$$

using the linear property of the kernels, with respect to the coefficients of the singularity. We now write (1) in the following form:

$$\begin{aligned}\| a_s dz \|^2 &- 4\pi \sum_{j=1}^{N} Re\left[\sum_{k=0}^{\infty} \frac{\bar{\lambda} a_k^j \bar{c}_k^j}{k+1} + (\bar{\lambda} - 1)\bar{d}_j \int_{c_j} (a_s - s)dz\right] = \| h_s dz \|^2 \\ &- | \lambda |^2 \| k_s dz \|^2 - 4\pi \sum_{j=1}^{N} Re\left[\sum_{k=0}^{\infty} \frac{\bar{\lambda} b_k^j \bar{c}_k^j}{k+1} + (\bar{\lambda} - 1)\bar{d}^j \int (h_s - s)dz\right] \\ &+ \| a_s dz - (h_s + \lambda k_s)dz \|^2 \,.\end{aligned}$$

We can now study the value of the bracket in the functional, and prove that

$$\sum_{j=1}^{N}\left[\sum_{k=0}^{\infty} \frac{\bar{\lambda} b_k^j \bar{c}_k^j}{k+1} + \bar{\lambda}\bar{d}^j \int_{c_j} (h_s - s)dz\right] = 0 \,.$$

We shall summarize our results in a theorem:

THEOREM III A. *Let* $sdz = \sum_{j=1}^{N} s_j dz$ *where*

$$s_j dz = \sum_{k=0}^{\infty} \frac{c_k^j dz}{(z - \zeta_j)^{k+2}} + \frac{d^j dz}{(z - \zeta_j)}$$

*be an analytic singularity with* $\sum_{j=1}^{N} d^j = 0$.

Let $k_s dz, h_s dz$ be the Bergman kernels corresponding to $sdz$, and let $\lambda$ be a complex parameter.

Then the linear combination $(h_s + \lambda k_s)dz$ minimizes the functional:

$$\| a_s dz \|^2 - 4\pi \sum_{j=1}^{N} Re\Big[\sum_{k=0}^{\infty} \frac{\bar{\lambda} a_k^j \bar{c}_k^j}{k+1} + (\bar{\lambda} - 1)\bar{d}^j \int_{c_j} (a_s - s)dz\Big]$$

over the class of differentials $a_s dz$, analytic except for the singularity $sdz$. The minimum is

$$\| h_s dz \|^2 + 4\pi \sum_{j=1}^{N} Re\bar{d}^j \int_{c_j} (h_s - s)dz + |\lambda|^2 \| k_s dz \|^2 ,$$

and the deviation from the minimum is

$$\| a_s dz - (h_s + \lambda k_s)dz \|^2 .$$

*Proof.* $h_s dz + \lambda e^{i\theta} k_s dz$ for $\theta$ real is a competing function; therefore:

$$\| h_s dz \|^2 - |\lambda|^2 \| k_s dz \|^2 - 4\pi \sum_{j=1}^{N} Re\Big[\sum_{k=0}^{\infty} \frac{\bar{\lambda} \bar{c}_k^j b_k^j}{k+1} + (\bar{\lambda} - 1)\bar{d}^j \int_{c_j} (h_s - s)dz\Big]$$

$$\leqq \| h_s dz \|^2 - |\lambda|^2 \| k_s dz \|^2 - 4\pi \sum_{j=1}^{N} \Big[\sum_{k=1}^{\infty} \frac{\bar{\lambda} e^{-i\theta} \bar{c}_k^j \bar{b}_k^j}{k+1} + (\bar{\lambda} e^{-i\theta} - 1)\bar{d}^j \int_{cj} (h_s - s)dz\Big] .$$

It follows that

$$\sum_{j=1}^{N} Re\Big[\sum_{k=0}^{\infty} \frac{\bar{\lambda} \bar{c}_k^j b_k^j}{k+1} + \bar{\lambda} \bar{d}^j \int_{c_j} (h_s - s)dz\Big] \geqq \sum_{j=1}^{N} Re\Big\{e^{-i\theta}\Big[\sum_{k=0}^{\infty} \frac{\bar{\lambda} \bar{c}_k^j b_k^j}{k+1} + \bar{\lambda} d^j \int_{c_j} (h_s - s)dz\Big]\Big\}$$

which is only possible if the bracket is real. It cannot be real fer all $\lambda$ except if it is equal to zero.

**4. Particular cases-applications.** Assume now that $adz = (\partial p/\partial z)dz$, where $p$ is a single-valued harmonic function on $W$, except for a singularity $Re\, S(z) = \sum_{j=1}^{N} Re\, S_j(z)$, with

$$ReS_j(z) = d^j \log |z - \zeta_j| + Re\Big[\sum_{k=0}^{\infty} \frac{c_k^j}{(-k-1)(z-\zeta_j)^{k+1}}\Big] ,$$

*where $d^j$ is real.* The singularity of $(\partial p/\partial z)dz$ is then $sdz = \sum_{j=1}^{N} s_j dz$, with

$$s_j dz = \frac{d^j dz}{z - \zeta_j} + \sum_{k=0}^{\infty} \frac{c_k^j dz}{(z-\zeta_j)^{k+2}} .$$

Moreover if $p = Re\,\{S_j(z) + \sum_{k=0}^{\infty} A_k^j(z-\zeta_j)^k\}$ near $z = \zeta_j$ and

$$\frac{\partial p}{\partial z}\,dz = s_j dz + \sum_{k=0}^{\infty} a_k^j(z-\zeta_j)^k\,,$$

it follows that $A_{k+1}^j = a_k^j/k+1$ for $k \geqq 0$. We notice furthermore that $\|\,adz\,\|^2 = 2(B(p) - A(p))$, where $B(p) = \int_\beta p dp^*$ ($\beta$ the ideal boundary of $\omega$) and $A(p) = 2\pi \sum_{j=1}^{N} d^j \int_{c_j} (a_s - s)dz$. The functional to be minimized becomes:

$$2\Big[B(p) - 2\pi \sum_{j=1}^{N} Re\Big[\sum_{k=0}^{\infty} \frac{\bar{\lambda} a_k^j \bar{c}_k^j}{k+1} + \bar{\lambda} d^j \int_{c_j}(a_s - s)dz\Big]\Big]\,.$$

We notice that the differentials $adz = (\partial p/\partial z)dz$ with $p$ single valued harmonic function generate a subspace $\Gamma_p \subset \Gamma_a$. If $k_{sp}dz$ and $h_{sp}dz$ are the Bergman kernels for $\Gamma_p$, they correspond to two functions $K_s$ harmonic and $H_s$ harmonic except for the singularity $Re\,S(z)$ and such that:

$$k_{sp}dz = \frac{\partial K_s}{\partial z}\,dz$$

$$h_{sp}dz = \frac{\partial H_s}{\partial z}\,d_s\,.$$

We can write the value of the minimum as:

$$2B(H_{sp}) + |\,\lambda\,|^2\,\|\,k_{sp}dz\,\|^2\,.$$

We now shall prove the following theorem.

THEOREM IV A: *Let $(\partial_{p0}/\partial z)dz$ and $(\partial_{p1}/\partial z)dz$ be the analytic differentials with singularity sdz, corresponding to the principal functions $p_0$ and $p_1$. Then*

$$\tfrac{1}{2}\partial/dz(p_1 - p_0)dz = k_{sp}dz$$

$$\tfrac{1}{2}\partial/dz(p_1 + p_0)dz = h_{sp}dz\,,$$

*where $h_{sp}dz$ and $k_{sp}dz$ are the orthogonal and reproducing kernels for $\Gamma_p \cap \Gamma_{ase}$, corresponding to the singularity sdz.*

*Proof.* First, we know from the definition of $p_0$ and $p_1$, that $(\partial_{p0}/\partial z)dz$ and $(\partial_{p1}/\partial z)dz$ are elements of $\Gamma_p \cap \Gamma_{ase}$. Second, from (1. Chapter III. Theorem 9E where only the notation is different), $(\partial_{p0}/\partial z)dz$ minimizes the same functional as $h_{sp}dz - k_{sp}dz$ (which corresponds to $\lambda = -1$), and $(\partial_{p1}/\partial z)dz$ minimizes the same functional as $(h_{sp}dz + k_{sp})dz$, (which corresponds to $\lambda = 1$). The theorem follows.

We shall consider here a family of functions $P$ harmonic, except for a singularity of the type $ReS(z)$; the periods of $P$ vanish along all dividing cycles. It follows that the differentials $(\partial P/dz)dz$ are elements of $\Gamma_P \cap \Gamma_{ase}$, except for a singularity $s(z)dz$.

We shall call $H_s$ the function corresponding to $h_{sp}dz$, and $K_s$ the one corresponding to $k_{sp}dz$. The following results are consequences of the main Theorem.

THEOREM IV B: *Among all functions $P$ with singularity $1/(z-\zeta)$, $H_s + \lambda K_s$ minimizes the functional $B(P) - 2\pi Re\bar{\lambda}A_1$.*

THEOREM IV C: *Among all functions $P$ with singularity $\log|(z-\zeta_1)/(z-\zeta_2)|$, $H_s + \lambda K_s$ minimizes $B(P) - 2\pi Re\bar{\lambda}(A_0^1 - A_0^2)$.*

THEOREM IV D: *Among all functions $P$ with singularity $ReS(z)$, $H_s$ minimizes the functional $B(P)$.*

We shall now consider exact differentials, analytic except for some singularity $s(z)dz = \sum_{j=1}^N s_j(z)dz$, which may be written $f'(z)dz = df(z)$, where $f$ is a function analytic except for a singularity $S(z) = \sum_{j=1}^N s_j(z)$ such that $S'(z)dz = s(z)dz$; then $f = S_j(z) + \sum_{k=0}^\infty \alpha_k(z-\zeta_j)^k$ near $z = \zeta_j$. We proved [II] the existence of a non-zero reproducing kernel if $W \notin 0_{AD}$. We shall now find a sufficient condition for the existence of an orthogonal kernel. We recall that in the case of a planar Riemann surface

$$\Gamma_h = \Gamma_{he} + \Gamma_{he}^* \cap \Gamma_{ho}^* .$$

We shall consider here Riemann surfaces on which

$$\Gamma_h = \Gamma_{he} + \Gamma_{he}^* .$$

We call such surfaces type $W_E$. On a surface of type $W_E$

$$\Gamma_{h0} \cap \Gamma_{ho}^* = [\Gamma_{he} + \Gamma_{he}^*]^\perp = 0 .$$

We then get the following lemma:

LEMMA IV E: *On a surface of type $W_E$, given a singularity $s(z)dz = dz/(z-\zeta)^{m+2}$, $m \geqq 0$, there exists a differential analytic exact, except for the corresponding singularity.*

*Proof*. Let $\Theta$ be constructed as in [1, Chapter V. 18.19]. The differential $\Theta - i\Theta^*$ is square integrable and hence has the decomposition?.

$$\Theta - i\Theta^* = \omega_h + \omega_{eo} + \omega_{eo}^* = \omega_{he} + \omega_{he}^* + \omega_{eo} + \omega_{eo}^* .$$

It follows that

$$\eta = \Theta - \omega_{eo} - \omega_{he} = i\Theta^* + \omega_{he}^* + \omega_{eo}^*$$

is harmonic exact except for the singularity and so is $\eta^*$. We may write $\eta = \phi + \bar{\psi}$ where $\psi$ is analytic and $\phi$ is analytic except for the singularity. It follows that $\phi$ is the differential mentioned in the lemma; $\phi = dF_m$ where $F_m$ is an analytic function except for the singularity

$$\frac{-1}{(m+1)(z-\zeta)^{m+1}}$$

and from [2] there exists an orthogonal kernel $dH_m$ for $\Gamma_{ae}$ on $W_E$.

*Note.* An analogous proof works for differentials with $s(z)dz = dz/(z-\zeta_1) - dz/(z-\zeta_2)$; we have only to discard the periods about $z=\zeta_1$ and $z=\zeta_2$.

From the existence of orthogonal kernels for $\Gamma_{ae}$ we can state the following theorems; here $B(f) = \frac{1}{2}\int_\beta f d\bar{f}$; $H_s$ and $K_s$ are analytic functions whose differentials are respectively the orthogonal and reproducing kernels for $\Gamma_{ae}$, corresponding to the singularity.

THEOREM IV F: *Among all functions $f$ analytic except for a simple pale at $z=\zeta$ with expansion $f = c_1/(z-\zeta) + \alpha_1(z-\zeta) + \cdots$ in a neighborhood of $z=\zeta$, $H_s + \lambda K_s$ minimizes the functional $B(f) + 2\pi Re\bar{\lambda}\bar{c}_1\alpha_1$.*

THEOREM IV G: *Among all functions $f(z)$ analytic except for the singularity*

$$S(z) = \sum_{j=1}^{N}\sum_{k=0}^{\infty} \frac{c_k^j}{(z-\zeta_j)^{k+1}(-k-1)},$$

*the function $H_s$ minimizes $B(f)$.*

## BIBLIOGRAPHY

1. L. V. Ahlfors and L. Sario, *Riemann surfaces*, Princeton University Press, Princeton 1960.

2. G. G. Weill, *Reproducing kernels and orthogonal kernels for analytic differentials on Riemann surfaces*, Pacific J. Math., **12** (1962), 729–767.

Errata to the paper

*"A remark on the Nijenhuis tensor"*

BY E. T. KOBAYASHI

LEMMA 1.3 (page 965 in proof) is not correct as it is stated, the hypothesis is too weak. The hypothesis should be stated as: "If $\theta_1, \cdots, \theta_g$ are completely integrable distributions of dimensions $r_1, \cdots, r_g$ on $M$, *such that the sums* $\theta_i + \theta_j: x \to \theta_i(x) + \theta_j(x)$ *are also completely integrable* $(i, j = 1, \cdots, g; i \neq j)$, *and* such that

$$\theta_1(x) + \theta_2(x) + \cdots + \theta_g(x) = T_x \text{ (direct sum)}$$

for each $x \in M$" where the underline portion is what should be added to the old version.

**Pacific Journal of Mathematics**
Vol. 12, No. 4 April, 1962