ALMOST PERIODIC FUNCTIONS IN A GROUP. I*

BY

J. v. NEUMANN

INTRODUCTION

1. The object of the present paper is to extend H. Bohr's famous theory of almost periodic functions [4, I]† to arbitrary groups, and to show that it gives just the maximum range over which the fundamental results of Frobenius-Schur representation theory [21; 22; 30] and its extensions by Peter and Weyl [32] hold. We shall see in particular that all bounded linear representations of a group are equivalent to unitary representations and belong to this class. Another point of importance is that we free ourselves completely from all topological assumptions (such as continuity, etc.) by the use of a definition of almost periodicity due to Bochner [2]. Thus we find that the general theory, which applies to every group $\mathfrak{G}$ whatsoever, is completely free from topological assumptions, but all of its results (for example, all series expansions) have a property of closure; if applied to functions which are continuous in a certain topology, they will lead only to functions of the same kind. It is remarkable that we find in the classical case of Bohr new almost periodic functions in addition to the known ones; even the elementary functions $f(a) = e^{2\pi\lambda a i}$ can be generalized (this connects with results of Ursell [28]). On the other hand, in some groups (for example in all semi-simple Lie groups) almost periodicity automatically implies continuity (this will be proved with the aid of a theorem of van der Waerden [29]).

2. The principal difficulty in building up a general theory of almost periodic functions lies in finding a generalization of the Bohr integral mean

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(x)dx$$

if the real numbers x and T are replaced by the elements of an arbitrary group $\mathfrak{G}$ which need not be even topological; also, the function $f(x)$ may be discontinuous. We meet this difficulty by finding an entirely new definition (cf. Definitions 4 and 5) which may be proved to be fit for the role of a "mean" under all conditions. The direct discussion of our mean is very simple and is given in Part I.

---

† The numbers in brackets refer to the bibliography at the end of the paper.

446   J. v. NEUMANN   [July

This mean is an extension of an integral in compact groups previously defined by the author [19]. It is defined by means entirely different from those employed in Haar's integral [11] with which it coincides for compact groups, but from which it differs widely for non-compact groups, first, because for such groups it is an integral-mean and not an integral; second, because it is free from topological restrictions, while Haar's integral applies only to locally compact and separable groups; third, because it is defined for almost periodic functions while Haar's integral is defined for measurable functions, and in general neither of these two classes contains the other.

3. The content of Parts I–V is as follows: Part I gives our general theory of the mean. Part II applies this theory (by using the powerful method of Weyl [31]) to prove the fundamental theorems of the Bohr theory, Parseval's formula and the approximation theorem. As we have to combine the devices contained in two papers of Weyl [30; 31] we find it advisable to give the proofs in full, even though the repetition is often almost literal. Part III repeats the main results of the Frobenius-Schur and Peter-Weyl theory of representations, and connects them with the theory of almost periodic functions. It provides a basis for the statement that the present general theory of almost periodic functions is the widest range over which this theory of representations holds without any loss of strength. Part IV connects our theory with topological and other restrictive conditions. By investigating the details of eight examples we illustrate the principal types of combinations of these notions which are likely to occur. Finally, we discuss the question as to how many almost periodic functions exist in a given group. Part V is entirely devoted to the proof that the maximal amount exists in Abelian groups (subject, however, to certain topological restrictions). Here the integral of Haar is used in combination with certain theorems of the author on operators and functions of operators [17]. The extension of some results of Haar on countably infinite Abelian groups [10] is of great importance for these investigations.

4. It is probable that most of the further developments of the Bohr theory will also apply to our general theory. Among these developments are finer convergence theorems, summability theorems, and Stepanoff's generalizations (where some topological restrictions will be necessary, as the Haar integral must be applied). In this connection it may be of interest to point out a needed generalization of an important notion of the Bohr theory, namely, the fact that the product of two elementary almost periodic functions is a function of the same kind: $e^{2\pi\lambda a i}e^{2\pi\mu a i}=e^{2\pi(\lambda+\mu)ai}$. This is unchanged for Abelian groups and leads to the important character-group; but in non-Abelian groups the corresponding situation is that the direct product of two irreducible representations (the elements $D_{\rho\sigma}(a;\mathfrak{C})$ of which are the analogues

of $e^{2\tau\lambda ai}$, cf. Definitions 11 and 12, and Theorems 24 and 28) is a sum of a finite number of irreducible representations, that is, there is the so-called composition formula

$$(*) \qquad D_{\rho\sigma}(a;\mathfrak{C})D_{\tau v}(a;\mathfrak{D}) = \sum_{\mathfrak{E},\xi,\eta}\Gamma_{\xi,\eta,\mathfrak{E}}(\rho,\sigma;\mathfrak{C}\mid\tau,v;\mathfrak{D})D_{\xi\eta}(a;\mathfrak{E}).$$

Another important notion in Bohr's theory is the independence of the expansion functions $e^{2\tau\lambda ai}$, $e^{2\tau\mu ai}$, $\cdots$ (that is, the linear independence of their exponents with integral coefficients), since almost periodic functions with such expansions possess particularly simple convergence properties. The corresponding requirement in our general theory is probably that the right-hand member of (*) should contain no term originating from the representation $D(a;\mathfrak{C})\equiv1$ if the left-hand member is any product of powers of $D(a;\mathfrak{C})$, $D(a;\mathfrak{D})$, $\cdots$.

<div align="center">I. EXISTENCE OF THE MEAN, GENERAL PROPERTIES</div>

5. Let $\mathfrak{G}$ be a group, that is, a set in which the operations ab and a^{-1} are defined and satisfy the group postulates. While $\mathfrak{G}$ may be topological* this property is not needed in Parts I–III and we do not yet make this assumption concerning $\mathfrak{G}$. Elements of $\mathfrak{G}$ will be denoted by $a, b, c, x, y, z, \cdots$, real or complex numbers by $m, n, u, v, \alpha, \beta, \xi, \eta, \cdots$, and functions defined in $\mathfrak{G}$ with complex numbers as values by $f(x), g(x), \cdots$.

For such functions $f(x)$ and $g(x)$ we define distance† by

$$D(f,g) = \text{l.u.b.}_x|f(x)-g(x)|.$$

A set $\mathfrak{M}$ of such functions is called conditionally compact (c.c.) if every sequence $f_1, f_2, \cdots$ extracted from it contains a subsequence $f_{n_1}, f_{n_2}, \cdots$ such that $D(f_{n_\mu}, f_{n_\nu})\to0$ as $\mu, \nu\to\infty$ (that is, a "fundamental" subsequence [13, p. 107]); this means that there exists a function f (not necessarily belonging to $\mathfrak{M}$) such that $D(f_{n_\mu}, f)\to0$ as $\mu\to\infty$.

We now extend Bohr's notion of almost periodic functions [4; 2, §5] to all $f(x)$ in $\mathfrak{G}$, but we prefer to generalize the definition given by S. Bochner [2], as it allows us to rid ourselves completely of topological conditions on $f(x)$ (continuity, etc.).

---

* That is, a topological set in the sense of Hausdorff [13, pp. 226–230]. One may take his topological system based on the notion of a neighborhood by means of Axioms 1, 2, 3 (or A, B, C) and one of the "separation" Axioms 4–8, such as 5. Furthermore, certain continuity assumptions have to be made concerning ab and a^{-1}. In Parts I–III we shall need no topology at all, in Part IV we must assume that ab is continuous in a for fixed b and in b for fixed a, and in Part V we must assume that ab is continuous in (a, b) and that a^{-1} is continuous in a.

† We shall consider only bounded functions. l.u.b.$_x$ denotes the least upper bound for all x's in $\mathfrak{G}$.

DEFINITION 1. *A function $f(x)$ in $\mathfrak{G}$ (with complex values) is called right almost periodic (r.a.p.) if the set R_f of all functions $f(xa)$ (x is the variable, a is a parameter running over $\mathfrak{G}$) is c.c.; it is called left almost periodic (l.a.p.) if the set L_f of all functions $f(ax)$ is c.c.; it is called almost periodic (a.p.) if it is r.a.p. and l.a.p.*

The equivalence of this definition to the obvious generalization of the Bohr definition is shown in the usual way if $f(x)$ is continuous; similarly, the uniform continuity of $f(x)$ follows in this case. But as we do not wish now to assume any topology in $\mathfrak{G}$, we shall not go into the details of this matter. On the other hand, the following theorems are of major importance:

THEOREM 1. *Each of the three notions r.a.p., l.a.p. and a.p. is invariant under the following operations: $f(xa)$, $f(ax)$, $\overline{f(x)}$, $\alpha f(x)$ (α any complex number), $f(x) \pm g(x)$, $f(x)g(x)$, and the operation of passing from $f_1(x)$, $f_2(x)$, $\cdots$ to $f(x)$ if $f_n(x)$ converges uniformly to $f(x)$ as $n \to \infty$. Passing from $f(x)$ to $f(x^{-1})$ interchanges r.a.p. and l.a.p. and leaves a.p. invariant.*

The statement concerning $f(x^{-1})$ is obvious. In the other cases we need to consider only r.a.p., as l.a.p. results, for example, by replacing ab by ba when defining $\mathfrak{G}$, and a.p. results by combining r.a.p. and l.a.p. That $f(xa)$ is r.a.p. is seen by replacing a_1, a_2, $\cdots$ (Definition 1) by a_1a, a_2a, $\cdots$; that $f(ax)$ is r.a.p. results from replacing x by ax; the situation concerning $\overline{f(x)}$ and $\alpha f(x)$ is obvious; the r.a.p. of $f(x) \pm g(x)$ and of $f(x)g(x)$ is proved by applying Definition 1 first to $f(x)$ and a_1, a_2, $\cdots$, and then to $g(x)$ and the subsequence which has been selected. An obvious and simple application of the diagonal process shows the invariance of r.a.p. under the operation of passing from $f_n(x)$ to $f(x)$.

THEOREM 2. *Every r.a.p. or l.a.p. function $f(x)$ is bounded.*

Again it is sufficient to consider r.a.p. If $f(x)$ were not bounded, we could select a sequence a_1, a_2, $\cdots$ such that $|f(a_n)| \to \infty$ as $n \to \infty$, and then no subsequence of $f(xa_1)$, $f(xa_2)$, $\cdots$ could have a finite limit at $x = 1$.

DEFINITION 2. *If $\mathfrak{M}$ is a set of functions in $\mathfrak{G}$, we call the set of all functions $\alpha_1 f_1(x) + \cdots + \alpha_n f_n(x)$ ($n = 1, 2, \cdots$; $\alpha_1, \cdots, \alpha_n$ non-negative real numbers such that $\alpha_1 + \cdots + \alpha_n = 1$; $f_1, \cdots, f_n$ any elements of $\mathfrak{M}$) the convex of $\mathfrak{M}$ and denote it by $\mathrm{Co}(\mathfrak{M})$.*

6. We prove

THEOREM 3. *If either of the sets $\mathfrak{M}$ and $\mathrm{Co}(\mathfrak{M})$ is c.c., the other is also c.c.*

If $\mathrm{Co}(\mathfrak{M})$ is c.c., its subset $\mathfrak{M}$ is c.c. Conversely, suppose that $\mathfrak{M}$ is c.c. The c.c. property of a set $\mathfrak{N}$ is equivalent to the following condition: for every

$\epsilon > 0$ there exists a finite number of functions $\bar{f}_1, \cdots, \bar{f}_m$ of $\mathfrak{N}$ $(m = m(\epsilon))$ such that, for each $f \subset \mathfrak{N}$, some $D(f, \bar{f}_\mu) \leqq \epsilon$, $\mu = 1, \cdots, m$ [13, pp. 108–109]. Now if an $\epsilon > 0$ is given, choose the functions $\bar{f}_1, \cdots, \bar{f}_m$ for $\mathfrak{M}$ and ϵ, put

$$\max_\mu \, \text{l.u.b.} \, |\bar{f}_\mu(x)| = C,$$

and select an integer $N \geqq Cm\epsilon^{-1}$. Then $\beta_1\bar{f}_1 + \cdots + \beta_m\bar{f}_m$ (where $\beta_1, \cdots, \beta_m$ are non-negative rational numbers with denominators N such that $\beta_1 + \cdots + \beta_m = 1$) can be written as a finite sequence $\bar{g}_1, \cdots, \bar{g}_M$ and have the property described above for $\text{Co}(\mathfrak{M})$ and 2ϵ.

DEFINITION 3. *If $f(x)$ is a real bounded function in $\mathfrak{G}$, we call*

$$\text{l.u.b.}_{x,y} |f(x) - f(y)| \quad (x \text{ and } y \text{ vary independently over } \mathfrak{G})$$

the oscillation of $f(x)$ and denote it by $\text{Osc}_x f(x)$. If $f(x)$ is not a constant, $\text{Osc}_x f(x) > 0$; otherwise, $\text{Osc}_x f(x)$ is zero.

THEOREM 4. *For every real $g \subset \text{Co}R_f$ we have $\text{Osc}_x g(x) \leqq \text{Osc}_x f(x)$. If the relation $\text{Osc}_x g(x) < \text{Osc}_x f(x)$ never occurs, and if $f(x)$ is l.a.p., $f(x)$ is necessarily a constant.*

The first statement is obvious. Suppose that the assumptions of the second statement are valid. Let $a_1', \cdots, a_n'$ be any elements of $\mathfrak{G}$; then

$$\frac{f(xa_1') + \cdots + f(xa_n')}{n} \subset \text{Co } R_f,$$

and thus

$$\text{Osc}_x \frac{f(xa_1') + \cdots + f(xa_n')}{n} = \text{Osc}_x f(x).$$

This implies that

$$\text{l.u.b.}_x \frac{f(xa_1') + \cdots + f(xa_n')}{n} = \text{l.u.b.}_x f(x).$$

Put $\text{l.u.b.}_x f(x) = C$; then for every $\epsilon > 0$ there exists an x' such that

$$\frac{f(x'a_1') + \cdots + f(x'a_n')}{n} \geqq C - \epsilon.$$

As all $f(x'a_\nu') \leqq C$, they all must also be $\geqq C - n\epsilon$.

Now choose a $\delta > 0$ and find a finite number of elements of L_f such that each element of L_f has a distance $\leqq \delta$ from one of them (cf. the proof of The-

orem 3). That is, find a finite number of elements $a_1, \cdots, a_n$ of $\mathfrak{G}$ such that for every a of $\mathfrak{G}$ there exists a $\mu = 1, 2, \cdots, n$ for which $|f(a_\mu x) - f(ax)| \leqq \delta$ identically. Now choose a b of $\mathfrak{G}$ and repeat the argument just described in the case where $\epsilon = \delta/n$, $a_1' = a_1^{-1}b$, $\cdots$, $a_n' = a_n^{-1}b$. Thus an x' exists for which all $f(x'a_\nu^{-1}b) \geqq C - \delta$, $\nu = 1, 2, \cdots, n$, and therefore, for a properly chosen $\mu = 1, 2, \cdots, n$, all $f(a_\mu a_\nu^{-1}b) \geqq C - 2\delta$. If $\nu = \mu$, then $f(b) \geqq C - 2\delta$.

On the other hand, $f(b) \leqq C$ and, as δ was arbitrary, it follows that $f(b) = C$. Finally, $f(x)$ is constant since b was arbitrary. This completes the proof of the theorem.

THEOREM 5. *If $f(x)$ is a.p., there exists a constant A toward which a certain sequence extracted from $\mathrm{Co}R_f$ converges uniformly.*

Since $f(x)$ is r.a.p., R_f and $\mathrm{Co}R_f$ are c.c. Denote real and imaginary parts by $\mathfrak{R}$ and $\mathfrak{J}$ respectively, consider the non-negative numbers $\mathrm{Osc}_x \mathfrak{R}g(x) + \mathrm{Osc}_x \mathfrak{J}g(x)$, $g \subset \mathrm{Co}R_f$ and call their greatest lower bound ω. We can extract a sequence $g_1(x), g_2(x), \cdots$ from $\mathrm{Co}R_f$ such that $\mathrm{Osc}_x \mathfrak{R}g_n(x) + \mathrm{Osc}_x \mathfrak{J}g_n(x) \to \omega$ as $n \to \infty$, and from this a subsequence $g_{n_1}(x), g_{n_2}(x), \cdots$, which converges uniformly to a function $g(x)$. Hence $\mathrm{Osc}_x \mathfrak{R}g(x) + \mathrm{Osc}_x \mathfrak{J}g(x) = \omega$. It is obvious that, $f(x)$ being l.a.p., every element $f(xa)$ of R_f is l.a.p. Therefore every element of $\mathrm{Co}R_f$ is l.a.p., and the uniform limit $g(x)$ as well as the real functions $\mathfrak{R}g(x)$ and $\mathfrak{J}g(x)$ are l.a.p. If we show that $\mathrm{Osc}_x \mathfrak{R}g(x) = \mathrm{Osc}_x \mathfrak{J}g(x) = 0$, we have $\mathfrak{R}g(x) = \text{constant}$, $\mathfrak{J}g(x) = \text{constant}$, that is, $g(x) = \text{constant}$, which proves our statement.

Suppose that $\mathrm{Osc}_x \mathfrak{R}g(x) > 0$. Then Theorem 4 shows that an $h \subset \mathrm{Co}R\mathfrak{R}_g$ exists such that $\mathrm{Osc}_x h(x) < \mathrm{Osc}_x \mathfrak{R}g(x)$. Here $h(x) = \alpha_1 \mathfrak{R}g(xa_1) + \cdots + \alpha_n \mathfrak{R}g(xa_n)$ $(\alpha_1, \cdots, \alpha_n$ each $\geqq 0$, $\alpha_1 + \cdots + \alpha_n = 1)$. Putting $k(x) = \alpha_1 g(xa_1) + \cdots + \alpha_n g(xa_n)$, we have $h(x) = \mathfrak{R}k(x)$, so that $\mathrm{Osc}_x \mathfrak{R}k(x) < \mathrm{Osc}_x \mathfrak{R}g(x)$. But it is obvious that $\mathrm{Osc}_x \mathfrak{J}k(x) \leqq \mathrm{Osc}_x \mathfrak{J}g(x)$. Therefore $\mathrm{Osc}_x \mathfrak{R}k(x) + \mathrm{Osc}_x \mathfrak{J}k(x) < \omega$. Now $g(x)$ can be uniformly approximated by functions $l \subset \mathrm{Co}R_f$, that is, $l(x) = \beta_1 f(xb_1) + \cdots + \beta_m f(xb_m)$ $(\beta_1, \cdots, \beta_m$ each $\geqq 0$, $\beta_1 + \cdots + \beta_m = 1)$. Hence $k(x)$ can be uniformly approximated by functions $q(x) = \alpha_1 \beta_1 f(xa_1 b_1) + \alpha_1 \beta_2 f(xa_1 b_2) + \cdots + \alpha_n \beta_m f(xa_n b_m)$, that is, by functions $q \subset \mathrm{Co}R_f$. Since $\mathrm{Osc}_x \mathfrak{R}k(x) + \mathrm{Osc}_x \mathfrak{J}k(x) < \omega$, the relation that $\mathrm{Osc}_x \mathfrak{R}q(x) + \mathrm{Osc}_x \mathfrak{J}q(x) < \omega$ results. This contradicts the definition of ω. Similarly $\mathrm{Osc}_x \mathfrak{J}g(x) > 0$ is disproved.

REMARK. *If a finite number of a.p. functions $f_1(x), \cdots, f_t(x)$ are given, it is possible to find a set of constants $A_1, \cdots, A_t$, toward which t sequences extracted from $\mathrm{Co}R_{f_1}, \cdots, \mathrm{Co}R_{f_t}$ respectively, with the same $\alpha_1, \cdots, \alpha_n$, $a_1, \cdots, a_n$, converge uniformly (that is, sequences of the form $\alpha_1^{(\nu)} f_1(xa_1^{(\nu)})$*

$+ \cdots + \alpha_{n_\nu}{}^{(\nu)} f_1(x a_{n_\nu}{}^{(\nu)}), \cdots, \alpha_{1_t}{}^{(\nu)} f(x a_1{}^{(\nu}) + \cdots + \alpha_{n_\nu}{}^{(\nu)} f_t(x a_{n_\nu}{}^{(\nu)}),$ *where* $\nu \to \infty, \alpha_1{}^{(\nu)} \geq 0, \cdots, \alpha_{n_\nu}{}^{(\nu)} \geq 0,$ *and* $\alpha_1{}^{(\nu)} + \cdots + \alpha_{n_\nu}{}^{(\nu)} = 1).$

The argument which proved Theorem 5 may be repeated here if we use $\mathrm{Osc}_x \, \Re f_1(x) + \mathrm{Osc}_x \, \Im f_1(x) + \cdots + \mathrm{Osc}_x \, \Re f_t(x) + \mathrm{Osc}_x \, \Im f_t(x)$ instead of $\mathrm{Osc}_x \, \Re f(x) + \mathrm{Osc}_x \, \Im f(x).$

DEFINITION 4. *A real number* A *which may be uniformly approximated by functions from* $\mathrm{Co}R_f$ *or* $\mathrm{Co}L_f$, *that is, a number* A *such that, for every* $\epsilon > 0$, *there exists a number* $n = 1, 2, \cdots$, *numbers* $\alpha_1, \cdots, \alpha_n$ *each* ≥ 0 *with* $\alpha_1 + \cdots + \alpha_n = 1$, *and elements* $a_1, \cdots, a_n$ *of* $\mathfrak{G}$ *such that the condition* $\left| \alpha_1 f(x a_1) + \cdots + \alpha_n f(x a_n) - A \right| \leq \epsilon$ *or* $\left| \alpha_1 f(a_1 x) + \cdots + \alpha_n f(a_n x) - A \right| \leq \epsilon$ *holds throughout* $\mathfrak{G}$, *is called a right-mean or a left-mean of* $f(x)$ *respectively.*

THEOREM 6. *If* $f(x)$ *is a.p., it has exactly one right-mean, exactly one left-mean, and these means are equal.*

The existence of a right-mean has been proved by Theorem 5. If we change the multiplication law ab in $\mathfrak{G}$ to ba, all notions remain unchanged except for the interchange of "right" and "left." Thus a left-mean must exist.

Now let A be a right-mean, let B be a left-mean, and let ϵ be > 0. Choose $\alpha_1, \cdots, \alpha_n, \beta_1, \cdots, \beta_m, a_1, \cdots, a_n, b_1, \cdots, b_m$ such that

$$\left| \alpha_1 f(x a_1) + \cdots + \alpha_n f(x a_n) - A \right| \leq \epsilon,$$
$$\left| \beta_1 f(b_1 x) + \cdots + \beta_m f(b_m x) - B \right| \leq \epsilon.$$

If we replace x in the first equation by $b_1 x, \cdots, b_m x$ in succession, and add, we obtain

$$\left| \alpha_1 \beta_1 f(b_1 x a_1) + \alpha_1 \beta_2 f(b_2 x a_1) + \cdots + \alpha_n \beta_m f(b_m x a_n) - A \right| \leq \epsilon.$$

Similarly, if we replace x in the second equation by $x a_1, \cdots, x a_n$ in succession, we obtain

$$\left| \alpha_1 \beta_1 f(b_1 x a_1) + \alpha_1 \beta_2 f(b_2 x a_1) + \cdots + \alpha_n \beta_m f(b_m x a_n) - B \right| \leq \epsilon.$$

Therefore $\left| A - B \right| \leq 2\epsilon$ and, as ϵ may be arbitrarily small, $A = B$.

DEFINITION 5. *If* $f(x)$ *is a.p., we call the common value of its uniquely determined right- and left-means the mean of* $f(x)$, *and denote it by* $M_x f(x)$.*

We now state the most important properties of the mean.

---

* Definitions 3–5 and the argument of Theorems 3–6 are in very close analogy to the author's construction of the Haar-Lebesgue measure in compact groups [19]. It is noteworthy that for non-compact groups, where Haar proved by his method the existence of an integral [11], our method leads to an integral-mean.

THEOREM 7. *If $f(x)$ and $g(x)$ are a.p. functions, all the functions $f(xa)$, $f(ax)$, $f(x^{-1})$, $\overline{f(x)}$, $\alpha f(x)$, $f(x) \pm g(x)$ (α a complex number, a an element of $\mathfrak{G}$) are a.p. (cf. Theorem 1). Furthermore, we have the following:*

(1) $M_x[\alpha f(x)] = \alpha M_x f(x)$.

(2) $M_x[f(x) \pm g(x)] = M_x f(x) \pm M_x g(x)$.

(3) $M_x 1 = 1$.

(4) *If $f(x)$ is real and ≥ 0 throughout $\mathfrak{G}$, then $M_x f(x) \geq 0$; and if, in addition, $f(x) \not\equiv 0$, then $M_x f(x) > 0$.*

(5) $|M_x[f(x)]| \leq M_x[|f(x)|]$.

(6) $M_x[\overline{f(x)}] = \overline{M_x[f(x)]}$.

(7) $M_x f(xa) = M_x f(x)$.

(8) $M_x f(ax) = M_x f(x)$.

(9) $M_x f(x^{-1}) = M_x f(x)$.

The equations (1), (3), (5), (6) and the first half of (4) are obvious; as every left-mean of $f(x)$ is a left-mean of $f(xa)$ and as every right-mean of $f(x)$ is a right-mean of $f(ax)$, (7) and (8) are valid; as every right-mean of $f(x)$ is a left-mean of $f(x^{-1})$, (9) is true. Thus, only (2) and the second half of (4) remain unproved.

In order to prove (2), put $M_x f(x) = A$, $M_x g(x) = B$, let ϵ be > 0, and choose $\alpha_1, \cdots, \alpha_n$ ($\alpha_1, \cdots, \alpha_n$ each ≥ 0, $\alpha_1 + \cdots + \alpha_n = 1$) and $a_1, \cdots, a_n$ such that

$$|\alpha_1 f(xa_1) + \cdots + \alpha_n f(xa_n) - A| \leq \epsilon.$$

Now $\alpha_1 g(xa_1) + \cdots + \alpha_n g(xa_n)$ obviously has the same left-mean as $g(x)$, i.e., B. Therefore we can choose $\beta_1, \cdots, \beta_m$ ($\beta_1, \cdots, \beta_m$ each ≥ 0, $\beta_1 + \cdots + \beta_m = 1$) and $b_1, \cdots, b_m$ such that

$$\alpha_1 \beta_1 g(xb_1 a_1) + \alpha_1 \beta_2 g(xb_2 a_1) + \cdots + \alpha_n \beta_m g(xb_m a_n) - B| \leq \epsilon.$$

If we replace x in the first inequality by $xb_1, \cdots, xb_m$ in succession, and add, we obtain

$$|\alpha_1 \beta_1 f(xb_1 a_1) + \alpha_1 \beta_2 f(xb_2 a_1) + \cdots + \alpha_n \beta_m f(xb_m a_n) - A| \leq \epsilon.$$

Denote nm by p; $\alpha_1 \beta_1, \alpha_1 \beta_2, \cdots, \alpha_n \beta_m$ by $\gamma_1, \cdots, \gamma_p$ ($\gamma_1, \cdots, \gamma_p$ each ≥ 0, $\gamma_1 + \cdots + \gamma_p = 1$); $b_1 a_1, b_2 a_1, \cdots, b_m a_n$ by $c_1, \cdots, c_p$; we get, by adding and substracting our inequalities,

$$|\gamma_1(f(xc_1) \pm g(xc_1)) + \cdots + \gamma_p(f(xc_p) \pm g(xc_p)) - (A \pm B)| \leq 2\epsilon.$$

As ϵ may be arbitrarily small, this shows that $M_x[f(x) \pm g(x)] = A \pm B$. (An-

other way to prove (2) would be to apply the Remark following Theorem 5 to $f(x)$ and $g(x)$.)

In order to prove the second half of (4), assume $f(x) \geqq 0$ everywhere and $f(x_0) > 0$ for one particular x_0. For any $\epsilon > 0$ a finite number of elements of R_f exist such that each element of R_f has a distance $\leqq \epsilon$ from one of them (cf. the proof of Theorem 3). Hence there is a finite number of the elements $a_1, \cdots, a_n$ such that, for every a, there exists a $\mu = 1, 2, \cdots, n$ such that $|f(xa_\mu) - f(xa)| \leqq \epsilon$ identically. Now take $\epsilon = f(x_0)/2$. The substitution $x = x_0 a_\mu^{-1}$ shows that $f(x_0 a_\mu^{-1} a) \geqq f(x_0)/2$. Hence, for each a it follows that $f(x_0 a_\nu^{-1} a) \geqq 0$ for every $\nu = 1, \cdots, n$, but that $f(x_0 a_\nu^{-1} a) \geqq f(x_0)/2$ for at least one ν. Thus $f(x_0 a_1^{-1} a) + \cdots + f(x_0 a_n^{-1} a) \geqq f(x_0)/2$, that is, the function $g(y) = f(x_0 a_1^{-1} y) + \cdots + f(x_0 a_n^{-1} y) - f(x_0)/2$ is always $\geqq 0$. Hence the first half of (4) leads to the result that $M_y g(y) \geqq 0$, (2), (3), (6), (7), and (8) show that $M_y g(y) = n M_y f(y) - f(x_0)/2$, and it follows that

$$M_y f(y) \geqq \frac{f(x_0)}{2n} > 0.$$

THEOREM 8. *The formal properties* (1)–(9) *determine* $M_x f(x)$ *uniquely; in fact,* (1)–(3), *the first half of* (4), *and* (7) *or* (8) *are sufficient.*

It is sufficient to consider (1)–(3), the first half of (4), and (7), as (8) may be obtained by replacing ab in $\mathfrak{G}$ by ba. So assume that a functional $M_x' f(x)$, defined for all a.p. $f(x)$ and satisfying (1)–(3), the first half of (4), and (7), is given.

For every $\epsilon > 0$ we can choose $\alpha_1, \cdots, \alpha_n, a_1, \cdots, a_n$ ($\alpha_1, \cdots, \alpha_n$ each $\geqq 0$, $\alpha_1 + \cdots + \alpha_n = 1$) such that $|\alpha_1 f(xa_1) + \cdots + \alpha_n f(xa_n) - M_x f(x)| \leqq \epsilon$, or if $f(x)$ is real,

$$M_x f(x) - \epsilon \leqq \alpha_1 f(xa_1) + \cdots + \alpha_n f(xa_n) \leqq M_x f(x) + \epsilon.$$

Then (1)–(3), the first half of (4), and (7) show that $M_x f(x) - \epsilon \leqq M_x' f(x) \leqq M_x f(x) + \epsilon$, and as ϵ was arbitrary, $M_x' f(x) = M_x f(x)$. Property (1) with $\alpha = i$ shows that this holds also for pure imaginary $f(x)$, and property (2) shows that it holds for every $f(x)$.

Theorems 6–8 show that, for a.p. functions $f(x)$, there is exactly one way to define a notion $M_x f(x)$ possessing the essential formal properties of a mean. Our $M_x f(x)$ is the equivalent of the well known integral mean

$$\lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(x) dx$$

in Bohr's theory, when $\mathfrak{G}$ is the addition group of all real numbers. But even

in this case the form of our definition is essentially different from the usual one (for example, it does not use continuity), and it gives a new approach to the problem.

REMARK.* The notion of the mean can be modified in the following manner. Consider the doubled group $\mathfrak{G}\mathfrak{G}'$, that is, the set of all pairs $[a, a']$. This set is a group by virtue of the definitions $[a, a'] [b, b'] = [ab, b'a']$ and $[a, a']^{-1} = [a^{-1}, a'^{-1}]$. (This is similar to the construction in the next paragraph except that here we use $b'a'$ while there we use $a'b'$.) The argument in the proof of Theorem 9 below shows that if $f(x)$ is a.p. in $\mathfrak{G}$, then $f_0([x, x']) = f(xx')$ is a.p. in $\mathfrak{G}\mathfrak{G}'$. By Theorem 5 it then follows that there exists a constant A such that, for every $\epsilon > 0$, there exists a number $n = 1, 2, \cdots$, numbers $\alpha_1, \cdots, \alpha_n$ each ≥ 0 with $\alpha_1 + \cdots + \alpha_n = 1$, and elements $a_1, \cdots, a_n$ and $b_1, \cdots, b_n$ of $\mathfrak{G}$ such that the condition

$$\left| \alpha_1 f_0([x, y][a_1, b_1]) + \cdots + \alpha_n f_0([x, y][a_n, b_n]) - A \right| \leq \epsilon$$

holds for all x and y in $\mathfrak{G}$. If we write $c_1 = a_1 b_1, \cdots, c_n = a_n b_n$, this condition assumes the form

$$\left| \alpha_1 f(xc_1 y) + \cdots + \alpha_n f(xc_n y) - A \right| \leq \epsilon.$$

This mean is even easier to handle than our right- and left-means (which are special cases of it). This is due to the following fact: choose two arbitrary sets of numbers $\beta_1, \cdots, \beta_k$ and $\gamma_1, \cdots, \gamma_l$, each ≥ 0, with $\beta_1 + \cdots + \beta_k = 1$ and $\gamma_1 + \cdots + \gamma_l = 1$, and two arbitrary sets of elements $a_1, \cdots, a_k$ and $b_1, \cdots, b_l$ of $\mathfrak{G}$. In our last inequality replace x and y by xa_κ and $b_\lambda y$, multiply by $\beta_\kappa \gamma_\lambda$, and add over all $\kappa = 1, \cdots, k; \lambda = 1, \cdots, l$. Then we obtain an inequality of the same type except that there are knl terms instead of n terms and $\beta_\kappa \alpha_\nu \gamma_\lambda$ and $a_\kappa c_\nu b_\lambda$ appear in place of α_ν and a_ν. This shows that the conditions

$$\left| \alpha_1 f(xc_1 y) + \cdots + \alpha_m f(xc_m y) - A \right| \leq \epsilon,$$
$$\left| \alpha_1' g(xc_1' y) + \cdots + \alpha_n' g(xc_n' y) - B \right| \leq \epsilon$$

imply the conditions

$$\left| \alpha_1'' f(xc_1'' y) + \cdots + \alpha_{mn}'' f(xc_{mn}'' y) - A \right| \leq \epsilon,$$
$$\left| \alpha_1'' g(xc_1'' y) + \cdots + \alpha_{mn}'' g(xc_{mn}'' y) - B \right| \leq \epsilon$$

if α_ρ'' and c_ρ'' are $\alpha_\mu \alpha_\nu'$ and $c_\mu c_\nu'$ (in some order). This gives the uniqueness, the extension to complex $f(x)$, and the additivity of our new mean at once. Of course this mean coincides with our former means.

---

* Added February 4, 1934.

7. The applications to be made in the next chapter necessitate our proving some facts concerning double means. We therefore pass to this subject.

The group $\mathfrak{G}$ can be "doubled," that is, we can consider the set $\mathfrak{G}\mathfrak{G}$ of all pairs $[a, a']$, which by the definitions $[a, a'] [b, b'] = [ab, a'b']$, $[a, a']^{-1} = [a^{-1}, a'^{-1}]$ becomes a group, and we will denote functions in it by $f(x, x')$ instead of by $f([x, x'])$. All our notions apply to $\mathfrak{G}\mathfrak{G}$: we have a.p. functions $f(x, x')$ in $\mathfrak{G}\mathfrak{G}$, and a mean $M_{x,x'}f(x, x')$.

THEOREM 9. *If $f(x)$ is a.p. in $\mathfrak{G}$, the eight functions $f(xx')$, $f(x'x)$, $f(xx'^{-1})$, $\cdots$, $f(x'^{-1}x^{-1})$ are all a.p. in $\mathfrak{G}\mathfrak{G}$.*

Interchange of x and x', of ab in $\mathfrak{G}$ with ba, and of $f(x)$ with $f(x^{-1})$ reduces our task to discussing $f(xx')$ and $f(xx'^{-1})$ alone. Their a.p. character in $\mathfrak{G}\mathfrak{G}$ means that the sets of functions $f(axa'x')$, $f(xax'a')$, $f(axx'^{-1}a'^{-1})$, $f(xaa'^{-1}x'^{-1})$ in $\mathfrak{G}\mathfrak{G}$ are c.c. or else that the sets of functions of one or of two variables, $f(axby)$, $f(xayb)$, $f(axb)$, $f(xay)$ in $\mathfrak{G}$ are c.c. The third case arises from the first by setting $y=1$, the fourth from the second by setting $b=1$, and the second from the first by interchanging a and x with b and y, and ab in $\mathfrak{G}$ with ba. So we need to discuss only $f(axby)$.

Choose an $\epsilon>0$. As $f(x)$ is r.a.p., there is a finite number of elements $y_1, \cdots, y_n$, such that for each y there is a $\nu=1, \cdots, n$ for which $|f(zy) - f(zy_\nu)| \leq \epsilon$ identically. As each $f(xy_\nu)$ is r.a.p., the set of all functions $f(xby_\nu)$ is c.c. for every $\nu=1, \cdots, n$, and therefore even the set of all "vector-functions" with n components $[f(xby_1), \cdots, f(xby_n)]$ is c.c. Therefore a finite number of elements $b_1, \cdots, b_m$ exist such that for each b there is a $\mu=1, 2, \cdots, n$ for which $|f(zby_\nu) - f(zb_\mu y_\nu)| \leq \epsilon$ for all z and for every $\nu=1, 2, \cdots, n$. Our two inequalities together give the result that
$$|f(zby) - f(zb_\mu y)| \leq 3\epsilon.$$
Finally, $f(x)$ is l.a.p., so that there exists a finite set of elements $a_1, \cdots, a_l$ such that for every a there is a $\lambda=1, \cdots, l$ for which $|f(au) - f(a_\lambda u)| \leq \epsilon$ identically. This, together with our last inequality, implies that $|f(axby) - f(a_\lambda xb_\mu y)| \leq 5\epsilon$ identically.

As this holds for every $\epsilon>0$, the c.c. character of the set of functions $f(axby)$ is proved [13, pp. 108–109].

THEOREM 10. *If $f(x, x')$ is a.p. in $\mathfrak{G}\mathfrak{G}$, it is also a.p. in $\mathfrak{G}$ as a function of x or as a function of x'. Thus we can form $M_x f(x, x')$ and $M_{x'}f(x, x')$ which are a.p. in $\mathfrak{G}$ as a function of x' and as a function of x, respectively. Thus we can form $M_{x'}[M_x f(x, x')]$ and $M_x[M_{x'}f(x, x')]$. These expressions are both equal to $M_{xx'}f(x, x')$.*

The first statement is obvious. In the second and third statements it is sufficient to consider $M_x f(x, x')$ and $M_{x'}[M_x f(x, x')]$, as interchange of x

with x' and of $f(x, x')$ with $f(x', x)$ leads to the rest of the theorem.

Consider a sequence a_1, a_2, $\cdots$ of elements of $\mathfrak{G}$. As $f(x, x')$ is r.a.p., the sequence $f(x, x'a_1)$, $f(x, x'a_2)$, $\cdots$ contains a uniformly convergent subsequence $f(x, x'a_{n_1})$, $f(x, x'a_{n_2})$, $\cdots$ such that, for every $\epsilon > 0$ and almost all μ and ν, $\left| f(x, x'a_{n_\mu}) - f(x, x'a_{n_\nu}) \right| \leq \epsilon$. This implies that $\left| M_x f(x, x'a_{n_\mu}) - M_x f(x, x'a_{n_\nu}) \right| \leq \epsilon$. Thus the set of functions of x', $M_x f(x, x'a)$, is c.c. Therefore $M_x f(x, x')$ is r.a.p. and interchange of ab in $\mathfrak{G}$ with ba shows that it is l.a.p. Hence it is a.p.

Now it is obvious that $M'_{xx'} f(x, x') = M_{x'}[M_x f(x, x')]$ has Properties (1)–(4) and (7) enumerated in Theorem 7 if we look at it as an $[x, x']$-mean. Therefore we may conclude from Theorem 8 that it is $M_{xx'} f(x, x')$.

Theorems 9 and 10 may be extended by iterating them m times to functions of 2^m variables; by choosing $2^m \geq n$ and taking the functions constant in the last $2^m - n$ variables these theorems may be extended to functions of n variables.

II. Application of the method of Weyl and E. Schmidt
Proof of the fundamental theorems

8. The results of Part I enable us to apply the method of Weyl to the proof of the fundamental theorems of Bohr's theory of a.p. functions (in the addition group of real numbers) and to the discussion of the linear-orthogonal representations of continuous groups.* The present part, II, contains a proof of "Parseval's formula" (equivalent to Theorem 15), which runs exactly along the lines of Weyl's proof. It also contains the proof of the "approximation theorem" (equivalent to Theorem 18) where a different device, due to N. Wiener, has to be used because of the difficulties of constructing in our general case an a.p. function with the required properties (cf. [31, pp. 348–349], and our Theorem 17). The next part, III, contains an interpretation and application of these theorems connecting the theories of a.p. functions and of representations. In this, Weyl's method is of fundamental importance.

DEFINITION 6. *If $f(x)$ and $g(x)$ are a.p., we set*

$$h(x) = M_y[f(xy^{-1})g(y)] = M_y[f(y)g(y^{-1}x)] = f \times g.$$

We observe that the two expressions for $h(x)$ are equal by Theorem 7 and Properties (7) and (9), after making the substitution $y^{-1}x$ for y, and that $h(x)$ is a.p. by Theorems 9 and 10.

---

* Cf. H. Weyl [31], H. Weyl and F. Peter [32]. The operational methods used there are partly based on the thesis of E. Schmidt [20].

REMARK. *$f \times g$ can be uniformly approximated by functions of the form $\gamma_1 f(xc_1) + \cdots + \gamma_n f(xc_n)$ ($\gamma_1, \cdots, \gamma_n$ are complex numbers), that is, for every $\epsilon > 0$ there exist numbers $\gamma_1, \cdots, \gamma_n$ and elements $c_1, \cdots, c_n$ such that*

$$\left| f \times g(x) - \gamma_1 f(xc_1) - \cdots - \gamma_n f(xc_n) \right| \leq \epsilon$$

holds throughout $\mathfrak{G}$.

$g(x)$ is a.p. and therefore bounded (Theorem 2); suppose $g(x) \leq C$. Now choose a $\delta > 0$. According to our remark in the proof of Theorem 3, it is possible to find a finite number of elements $b_1, \cdots, b_k$ of $\mathfrak{G}$ such that to every x there exists a $\kappa = 1, \cdots, k$ for which $\left| f(xz) - f(b_\kappa z) \right| \leq \delta$ holds identically. Now consider the a.p. y-functions $f(b_\kappa y^{-1})g(y)$, $\kappa = 1, \cdots, k$, and apply to them the Remark following Theorem 5 (with $\epsilon/2$): if $\epsilon > 0$, there exists a set of real numbers $\alpha_1, \cdots, \alpha_n$ ($\alpha_1, \cdots, \alpha_n$ each ≥ 0, $\alpha_1 + \cdots + \alpha_n = 1$) and a set of elements $a_1, \cdots, a_n$ of $\mathfrak{G}$ such that

$$\left| \alpha_1 f(b_\kappa a_1^{-1} y^{-1}) g(ya_1) + \cdots + \alpha_n f(b_\kappa a_n^{-1} y^{-1}) g(ya_n) - M_y[f(b_\kappa y^{-1})g(y)] \right| \leq \frac{\epsilon}{2}$$

holds identically for every $\kappa = 1, \cdots, k$. For every x there exists a κ such that $\left| f(xz) - f(b_\kappa z) \right| \leq \delta$ and therefore such that

$$\left| f(xu^{-1})g(u) - f(b_\kappa u^{-1})g(u) \right| \leq C\delta \quad \text{and}$$
$$\left| M_u[f(xu^{-1})g(u)] - M_u[f(b_\kappa u^{-1})g(u)] \right| \leq C\delta.$$

Hence we have the result that

$$\left| \alpha_1 f(xa_1^{-1} y^{-1})g(ya_1) + \cdots + \alpha_n f(xa_n^{-1} y^{-1})g(ya_n) - M_y[f(xy^{-1})g(y)] \right|$$

$$\leq 2C\delta + \frac{\epsilon}{2}.$$

Our statement is proved if we put $\delta = \epsilon/(4C)$ and $y = 1$, and substitute $\alpha_1 g(a_1), \cdots, \alpha_n g(a_n)$ for $\gamma_1, \cdots, \gamma_n$ and $a_1^{-1}, \cdots, a_n^{-1}$ for $c_1, \cdots, c_n$.

THEOREM 11. *The "multiplication" $f \times g$ is distributive (linear) in both factors, associative, and if $\mathfrak{G}$ is Abelian, commutative.*

The theorem is obvious except for associativity. Our second form for $h = f \times g$ gives

$$(f \times g) \times k(x) = M_z[M_y[f(y)g(y^{-1}z)]k(z^{-1}x)]$$
$$= M_z[M_y[f(y)g(y^{-1}z)k(z^{-1}x)]],$$
$$f \times (g \times k)(x) = M_y[f(y)M_z[g(y^{-1}z)k(z^{-1}x)]]$$
$$= M_y[M_z[f(y)g(y^{-1}z)k(z^{-1}x)]],$$

and these expressions are equal by Theorems 9 and 10.

DEFINITION 7. *If $f(x)$ is a.p., we denote $f \times f \times \cdots \times f$ (n factors) by f^n, and $\overline{f(x^{-1})}$ by $f'(x)$. Furthermore, we define*

$$Nf = \{M_x[|f(x)|^2]\}^{1/2}.$$

THEOREM 12. *Let $f(x)$ and $g(x)$ be a.p. functions. The following formulas hold:*

(1) *If $f \neq 0$, $Nf > 0$.*
(2) $N[\alpha f] = |\alpha| Nf$, $N[f \pm g] \leq Nf + Ng$, $N[f \times g] \leq (Nf)(Ng)$.
(3) $ff'(1) = f'f(1) = (Nf)^2$.
(4) $|M_x f(x)| \leq Nf$.
(5) $|f \times g(x)| \leq (Nf)(Ng)$.
(6) $M_x[|f(x)||g(x)|] \leq (Nf)(Ng)$.

Statements (1), the first part of (2), and (3) are obvious. The second part of (2), after being squared, means that

$$M_x[|f(x) \pm g(x)|^2] \leq M_x[|f(x)|^2] + M_x[|g(x)|^2] + 2(Nf)(Ng),$$
$$|M_x[\Re(f(x)\overline{g(x)})]| \leq (Nf)(Ng).$$

This obviously follows from (6). The third part of (2) again follows from (5) by squaring and applying M_x. (4) follows from (5) by putting $g(x) \equiv 1$, since $f \times 1(x) = M_y[f(y)]$.

Hence we need to prove only (5) and (6). Since

$$|f(y)g(y^{-1}x)| \leq \tfrac{1}{2}|f(y)|^2 + \tfrac{1}{2}|g(y^{-1}x)|^2$$

it follows that

$$|M_y[f(y)g(y^{-1}x)]| \leq \tfrac{1}{2}M_y[|f(y)|^2] + \tfrac{1}{2}M_y[|g(y^{-1}x)|^2]$$
$$\leq \tfrac{1}{2}M_y[|f(y)|^2] + \tfrac{1}{2}M_y[|g(y)|^2],$$
$$|f \times g(x)| \leq \tfrac{1}{2}(Nf)^2 + \tfrac{1}{2}(Ng)^2.$$

Starting from

$$|f(x)||g(x)| \leq \tfrac{1}{2}|f(x)|^2 + \tfrac{1}{2}|g(y)|^2$$

we obtain similarly

$$M_x[|f(x)||g(x)|] \leq \tfrac{1}{2}(Nf)^2 + \tfrac{1}{2}(Ng)^2.$$

If we replace f and g by γf and g/γ (γ real and >0) we see that $|f \times g(x)|$ and $M_x[|f(x)||g(x)|]$ do not exceed

$$\frac{\gamma^2}{2}(Nf)^2 + \frac{1}{2\gamma^2}(Ng)^2.$$

The greatest lower bound of this expression is $(Nf)(Ng)$. This completes the proof of (5) and (6).

THEOREM 13. *Let $f(x)$ be an a.p. function $\not\equiv 0$. Put*

$$\Gamma_n = N[f \times f' \times \cdots]^2 = N[f' \times f \times \cdots]^2 \qquad (n \text{ factors})$$
$$= (f \times f')^n(1) = (f' \times f)^n(1).$$

Then

$$\Gamma_n > 0, \qquad (\Gamma_n)^2 \leqq \Gamma_{n-1}\Gamma_{n+1}, \qquad \Gamma_{m+n} \leqq \Gamma_m\Gamma_n.$$

First we prove that the four expressions above for Γ_n are equal. Indeed the first and third expressions for Γ_n are equal, and so are the second and the fourth, since $(g \times h)' = h' \times g'$. The equality of the third and fourth expressions follows from

$$(f \times f')^n = f \times (f' \times f \times \cdots \times f'), \quad (f' \times f)^n = (f' \times f \times \cdots \times f') \times f,$$

and from

$$f \times g(1) = g \times f(1) = M_y[f(y)g(y^{-1})].$$

By (5) of Theorem 12, $|f \times g(1)| \leqq (Nf)(Ng)$. If we replace here f and g by $f \times f' \times \cdots$ with $n-1$ and $n+1$ factors respectively, we obtain $\Gamma_n^2 \leqq \Gamma_{n-1}\Gamma_{n+1}$. And if we replace f and g in (2) of Theorem 12 by $f \times f' \times \cdots$ or $f' \times f \times \cdots$ with m and n factors respectively, we obtain $\Gamma_{m+n} \leqq \Gamma_m\Gamma_n$. That $\Gamma_n \geqq 0$ is obvious; but the condition $\Gamma_n = 0$ would imply that $\Gamma_{n-1} = 0$ (because $\Gamma_{n-1}^2 = \Gamma_{n-2}\Gamma_n$ provided that $n \geqq 3$), so that $\Gamma_2 = 0$. This means that $N[f \times f'] = 0$, $f \times f'(x) \equiv 0$, hence $N[f]^2 = f \times f'(1) = 0$ (that is, $\Gamma_1 = 0$) and $f \equiv 0$, contrary to our assumption. Thus we have $\Gamma_n > 0$.

THEOREM 14. *Let $f(x)$ be as before, and define $\Gamma_1, \Gamma_2, \cdots$ as before. Then as $n \to \infty$,*

$$\frac{\Gamma_{n+1}}{\Gamma_n} \to \gamma, \qquad \frac{\Gamma_n}{\gamma^n} \to \kappa, \qquad \frac{(f \times f')^n(x)}{\gamma^n} \to \phi(x) \qquad (uniformly).$$

Furthermore, $0 < \gamma \leqq \Gamma_1$, $1 \leqq \kappa$, $\phi(x)$ is a.p., and $\phi' = \phi$, $\phi \times \phi = \phi$, $f \times f' \times \phi = \phi \times f \times f' = \gamma\phi$, $\phi(1) = \kappa$.

The formulas of Theorem 13 imply that

$$0 < \frac{\Gamma_2}{\Gamma_1} \leqq \frac{\Gamma_3}{\Gamma_2} \leqq \cdots \leqq \Gamma_1,$$

and therefore Γ_{n+1}/Γ_n has a limit γ as $n \to \infty$, $0 < \gamma \leqq \Gamma_1$. Furthermore,

$$\frac{\Gamma_{n+1}}{\Gamma_n} \leqq \gamma, \qquad \frac{\Gamma_n}{\gamma^n} \geqq \frac{\Gamma_{n+1}}{\gamma^{n+1}},$$

that is,

$$\frac{\Gamma_1}{\gamma} \geqq \frac{\Gamma_2}{\gamma^2} \geqq \cdots > 0,$$

and therefore Γ_n/γ^n has a limit κ as $n \to \infty$, $\kappa \geqq 0$. Finally we have $\Gamma_{m+n} \leqq \Gamma_m \Gamma_n$, that is,

$$\Gamma_n \geqq \frac{\Gamma_{m+n}}{\Gamma_m} = \frac{\Gamma_{m+1}}{\Gamma_m} \frac{\Gamma_{m+2}}{\Gamma_{m+1}} \cdots \frac{\Gamma_{m+n}}{\Gamma_{m+n-1}} \geqq \left(\frac{\Gamma_{m+1}}{\Gamma_m} \right)^n.$$

The limiting process $m \to \infty$ shows that $\Gamma_n \geqq \gamma^n$ and $\Gamma_n/\gamma^n \geqq 1$, and then the limiting process $n \to \infty$ shows that $\kappa \geqq 1$.

By (4) and (2) of Theorem 12,

$$\left| \frac{(f \times f')^n(x)}{\gamma^n} - \frac{(f \times f')^m(x)}{\gamma^m} \right|^2 = \left| f \times \left(\frac{(f' \times f)^{n-1}}{\gamma^n} - \frac{(f' \times f)^{m-1}}{\gamma^m} \right) \times f'(x) \right|^2$$

$$\leqq (Nf)^2 \left(N \left[\frac{(f' \times f)^{n-1}}{\gamma^n} - \frac{(f' \times f)^{m-1}}{\gamma^m} \right] \right)^2 (Nf')^2$$

$$= \Gamma_1^2 \left(\frac{(f' \times f)^{n-1}}{\gamma^n} - \frac{(f' \times f)^{m-1}}{\gamma^m} \right)^2 (1)$$

$$= \Gamma_1^2 \left(\frac{(f' \times f)^{2n-2}(1)}{\gamma^{2n}} - 2 \frac{(f' \times f)^{m+n-2}(1)}{\gamma^{m+n}} + \frac{(f' \times f)^{2m-2}(1)}{\gamma^{2m}} \right)$$

$$= \Gamma_1^2 \left(\frac{\Gamma_{2n-2}}{\gamma^{2n}} - 2 \frac{\Gamma_{m+n-2}}{\gamma^{m+n}} + \frac{\Gamma_{2m-2}}{\gamma^{2m}} \right).$$

As m and $n \to \infty$ the last expression converges to 0; thus the first expression converges uniformly to 0, that is, as $n \to \infty$, $(f \times f')^n(x)/\gamma^n$ converges uniformly to a limiting function $\phi(x)$. As the functions $(f \times f')^n(x)/\gamma^n$ are a.p., $\phi(x)$ is also a.p.

The relations

$$\left(\frac{(f \times f')^n}{\gamma^n} \right)' = \frac{(f \times f')^n}{\gamma^n}, \qquad \left(\frac{(f \times f')^n}{\gamma^n} \right)^2 = \frac{(f \times f')^{2n}}{\gamma^{2n}},$$

$$f \times f' \times \frac{(f \times f')^n}{\gamma^n} = \frac{(f \times f')^n}{\gamma^n} \times f \times f' = \gamma \frac{(f \times f')^{n+1}}{\gamma^{n+1}}$$

show that, when n becomes infinite (the convergences involved all being uni-

form), $\phi' = \phi$, $\phi \times \phi = \phi$, $f \times f' \times \phi = \phi \times f \times f' = \gamma \phi$. Finally,

$$\frac{\Gamma_n}{\gamma^n} = \frac{(f \times f')^n(1)}{\gamma^n} \to \phi(1)$$

and therefore $\phi(1) = \kappa$.

THEOREM 15. *Let $f(x)$ be as before. Then there is a (finite or infinite) sequence of real numbers γ_1, γ_2, $\cdots$ and a sequence of a.p. functions $\phi_1(x)$, $\phi_2(x)$, $\cdots$, all $\not\equiv 0$, such that $\gamma_1 > \gamma_2 > \cdots > 0$, $\phi_n' = \phi_n$, $\phi_n \times \phi_n = \phi_n$, $\phi_n(1) \geqq 1$, $\phi_m \times \phi_n = 0$ $(m \neq n)$, $f \times f' \times \phi_n = \phi_n \times f \times f' = \gamma_n \phi_n$, and $\gamma_1\phi_1(x) + \gamma_2\phi_2(x) + \cdots$ converges uniformly to $f \times f'(x)$.*

Apply Theorem 14, and put there $\gamma = \gamma_1$, $\phi(x) = \phi_1(x)$. $\phi_1(1) = \kappa \geqq 1$ proves that $\phi_1(x) \neq 0$. Now put $f^* = f - \phi_1 \times f$. Then $f^*(x)$ is a.p. and

$$f^* \times f'^* = (f - \phi_1 \times f) \times (f' - f' \times \phi_1)$$
$$= f \times f' - \phi_1 \times f \times f' - f \times f' \times \phi_1 + \phi_1 \times f \times f' \times \phi_1$$
$$= f \times f' - \gamma_1\phi_1 - \gamma_1\phi_1 + \gamma_1\phi_1 \times \phi_1 = f \times f' - \gamma_1\phi_1.$$

If $f^* \equiv 0$, this shows that $f \times f' = \gamma_1\phi_1$, that is, the Theorem holds for a sequence consisting of one element. Assume that $f^* \not\equiv 0$.

Then $f \times f' \times \phi_1 = \phi_1 \times f \times f' = \gamma_1\phi_1$ implies that $(f^* \times f'^*)^n = (f \times f' - \gamma_1\phi_1)^n = (f \times f')^n - \gamma^n\phi_1$, and thus

$$\frac{(f^* \times f'^*)^n(x)}{\gamma_1^n} = \frac{(f \times f')^n(x)}{\gamma_1^n} - \phi_1(x) \to 0 \text{ as } n \to \infty.$$

Therefore if we form $\gamma = \gamma_2$ and $\phi_2(x)$ of Theorem 14, for which we have

$$\frac{(f^* \times f'^*)^n(x)}{\gamma_2^n} \to \phi_2(x) \neq 0,$$

then it must be the case that $\gamma_1 > \gamma_2$.

By repeating this process with $f^{**} = f^* - \phi_2 \times f^*$, $f^{***} = f^{**} - \phi_3 \times f^{**}$, $\cdots$ we finally find a sequence of real numbers γ_1, γ_2, $\cdots$ and two sequences of a.p. functions $\phi_1(x)$, $\phi_2(x)$, $\cdots$ and $f(x)$, $f^*(x)$, $\cdots$ with the following properties:

$$\gamma_1 > \gamma_2 > \cdots > 0, \quad \phi_n' = \phi_n, \quad \phi_n \times \phi_n = \phi_n, \quad \phi_n(1) \geqq 1,$$
$$f^{(n-1)} \times f^{(n-1)'} \times \phi_n = \phi_n \times f^{(n-1)} \times f^{(n-1)'} = \gamma_n\phi_n, \quad f^{(n)} = f^{(n-1)} - \phi_n \times f^{(n)},$$

these sequences ending when an $f^{(n)}$ becomes $\equiv 0$, otherwise never ending.

These rules again imply the relations $f^{(n)} \times f^{(n)'} = f^{(n-1)} \times f^{(n-1)'} - \gamma_n\phi_n$. By adding these relations for all $n = 1$, $\cdots$, p, we obtain

(*) $\gamma_1\phi_1 + \cdots + \gamma_p\phi_p = f \times f' - f^{(p)} \times f^{(p)'}$.

We now wish to prove that $\phi_m \times \phi_n = 0$ for $m \neq n$. Application of (') shows that it is sufficient to consider $m > n$, that is, it is sufficient to prove that $\phi_{n+k+1} \times \phi_n = 0$ for $k = 0, 1, 2, \cdots$. Consider the equation $f^{(n+k)} \times f^{(n+k)\prime} \times \phi_n = 0$. For $k = 0$, the condition

$$f^{(n)} \times f^{(n)\prime} \times \phi_n = f^{(n-1)} \times f^{(n-1)\prime} \times \phi_n - \gamma_n \phi_n \times \phi_n = \gamma_n \phi_n - \gamma_n \phi_n = 0$$

obtains. If it holds for a certain $k = 0, 1, 2, \cdots$, we have

$$\phi_{n+k+1} \times f^{(n+k)} \times f^{(n+k)\prime} \times \phi_n = \begin{cases} \phi_{n+k+1} \times \left(f^{(n+k)} \times f^{(n+k)\prime} \times \phi_n \right) = 0, \\ \\ \left(\phi_{n+k+1} \times f^{(n+k)} \times f^{(n+k)\prime} \right) \times \phi_n = \gamma_{n+k+1} \phi_{n+k+1} \\ \hspace{9cm} \times \phi_n, \end{cases}$$

and thus $\phi_{n+k+1} \times \phi_n = 0$. This gives

$$f^{(n+k+1)} \times f^{(n+k+1)\prime} \times \phi_n = \left(f^{(n+k)} \times f^{(n+k)\prime} - \gamma_{n+k+1} \phi_{n+k+1} \right) \times \phi_n = 0,$$

that is, our equation holds for $k+1$. Therefore it holds for every $k = 0, 1, 2, \cdots$, and with it, its consequence $\phi_{n+k+1} \times \phi_n = 0$.

Application of $\phi_n \times \cdots$ or $\cdots \times \phi_n$ to (*) with $p = n - 1$ gives

$$\phi_n \times f \times f' = \phi_n \times f^{(n-1)} \times f^{(n-1)\prime} = \gamma_n \phi_n, \quad f \times f' \times \phi_n = f^{(n-1)} \times f^{(n-1)\prime} \times \phi_n = \gamma_n \phi_n.$$

The only thing remaining to be proved is the uniform convergence of $\gamma_1 \phi_1(x) + \cdots + \gamma_n \phi_n(x)$ to $f \times f'(x)$ as $n \to \infty$, or, according to (*), the uniform convergence of $f^{(n)} \times f^{(n)}(x)$ to 0. Now (*) implies, by (3) and (5) of Theorem 12, that $\gamma_1 \phi_1(1) + \cdots + \gamma_n \phi_n(1) = (Nf)^2 - (Nf^{(n)})^2 \leq (Nf)^2$, and since $|\phi_n \times \phi_n(x)| \leq \phi_n \times \phi_n(1)$, that is, $|\phi_n(x)| \leq \phi_n(1)$, (*) implies the uniform convergence of $\gamma_1 \phi_1(x) + \gamma_2 \phi_2(x) + \cdots$ to $g(x)$ as $n \to \infty$, where $g(x)$ must be a.p. Hence $f^{(n)} \times f^{(n)\prime}(x) \to f \times f(x) - g(x)$ uniformly. Moreover, the above mentioned convergence implies that $\gamma_n \to 0$ (because $\phi_n(1) \geq 1$). On the other hand, we have $\Gamma_2 / \Gamma_1 < \gamma$, $\Gamma_2 < \gamma \Gamma_1$, $(N[f \times f'])^2 < \gamma (Nf)^2$. If we replace f and γ by $f^{(n)}$ and γ_n we obtain $(N[f^{(n)} \times f^{(n)\prime}])^2 < \gamma_n (Nf^{(n)})^2 \leq \gamma_n (Nf)^2$. Thus

$$N[f^{(n)} \times f^{(n)\prime}] \to 0,$$

implying that $N[f \times f' - g] = 0$ and $g = f \times f'$.

9. Having reached the final result of the E. Schmidt-Weyl theory, we now pass to the approximation theorems. But as we have already mentioned, we are now giving only their proofs and shall discuss their real meaning in the next part.

DEFINITION 8. *An a.p. function $\phi(x)$ such that $\phi' = \phi$ and $\phi \times \phi = \phi$ is called a unit. Two units $\phi(x)$ and $\psi(x)$ such that $\phi \times \psi = 0$ are called orthogonal.*

THEOREM 16†. *For every a.p. function $f(x)$ and every $\epsilon > 0$ there exists a unit $\phi(x)$ such that $N[f - \phi \times f] \leqq \epsilon$.*

If $f \equiv 0$, then $\phi \equiv 0$. Hence we may assume that $f \not\equiv 0$. Then apply Theorem 13. $\psi_n = \phi_1 + \cdots + \phi_n$ is a unit (because $\phi_n' = \phi_n$, $\phi_n \times \phi_n = \phi_n$, $\phi_m \times \phi_n = 0$ for $m \neq n$), and we have

$$(N[f - \psi_n \times f])^2 = \left(f - \sum_{\nu=1}^{n} \phi_\nu \times f\right) \times \left(f' - \sum_{\nu=1}^{n} f' \times \phi_\nu\right) \quad (1)$$

$$= \left(f \times f - \sum_{\nu=1}^{n} \phi_\nu \times f \times f - \sum_{\nu=1}^{n} f \times f' \times \phi_\nu + \sum_{\mu,\nu=1}^{n} \phi_\mu \times f \times f' \times \phi_\nu\right) \quad (1)$$

$$= \left(f \times f - \sum_{\nu=1}^{n} \gamma_\nu \phi_\nu - \sum_{\nu=1}^{n} \gamma_\nu \phi_\nu + \sum_{\mu,\nu=1}^{n} \gamma_\nu \phi_\mu \times \phi_\nu\right) \quad (1)$$

$$= \left(f \times f - \sum_{\nu=1}^{n} \gamma_\nu \phi_\nu\right) \quad (1) \to 0 \quad \text{as } n \to \infty .$$

Thus $\phi = \psi_n$, for a sufficiently large n, yields the desired result.

THEOREM 17. *For every a.p. function $f(x)$ and every $\epsilon > 0$ there exists an a.p. function $g(x)$ such that $|f(x) - g \times f(x)| \leqq \epsilon$ for every x.*

Following N. Wiener, consider the "translation function" of $f(x)$,

$$e(x) = \text{l.u.b.}_y | f(x^{-1}y) - f(y)|,$$

which was introduced by S. Bochner [2]. As $f(x)$ is a.p., it is easily seen that $e(x)$ is also a.p. Furthermore, $e(x) \geqq 0$, $e(1) = 0$. Now define the function

$$F(u) = \begin{cases} 1 - \dfrac{u}{\epsilon}, & 0 \leqq u \leqq \epsilon, \\ \\ 0, & u \geqq \epsilon. \end{cases}$$

As $F(u)$ is continuous, $\phi(x) = F(e(x))$ is a.p. It is obvious that $\phi(x) \geqq 0$, $\phi(1) = 1$; and if l.u.b.$_y|f(x^{-1}y) - f(y)| > \epsilon$, then $\phi(x) = 0$. Thus $\phi(x) \neq 0$ implies that $|f(x^{-1}y) - f(y)| \leqq \epsilon$, and therefore, always $|\phi(x) (f(x^{-1}y) - f(y))| \leqq \epsilon \phi(x)$. Consequently, $|M_x[\phi(x) (f(x^{-1}y) - f(y))]| \leqq \epsilon M_x \phi(x)$. Now $M_x[\phi(x) (f(x^{-1}y) - f(y))] = \phi \times f(y) - f(y) M_x \phi(x)$, and therefore the function

---

† Cf. Theorems 28 and 29, where the results of Theorems 16 and 18 will be interpreted and applied.

$$g(y) = \frac{\phi(y)}{M_x \phi(x)}$$

meets the requirements.

THEOREM 18. *For every a.p. function $f(x)$ and every $\epsilon > 0$ there exists an a.p. function $g(x)$ and a unit $\phi(x)$ such that $\left| f(x) - \phi \times g \times f(x) \right| \leq \epsilon$ for every x.*[†]

Choose the function $g(x)$ of Theorem 17 corresponding to $f(x)$ and $\epsilon/2$, and choose the function $\phi(x)$ of Theorem 16 corresponding to $g(x)$ and $\epsilon/(2Nf)$. Then $\left| f(x) - g \times f(x) \right| \leq \epsilon/2$, $\left| g \times f(x) - \phi \times g \times f(x) \right| \leq N[g - \phi \times g]Nf \leq \epsilon/2$, and therefore $\left| f(x) - \phi \times g \times f(x) \right| \leq \epsilon$.

III. THEORY OF LINEAR REPRESENTATIONS OF $\mathfrak{G}$

10. We define the representations in the usual manner:

DEFINITION 9. *If to every $a \subset \mathfrak{G}$ there corresponds a matrix*

$$D(a) = \{ D_{\rho\sigma}(a) \} \qquad (\rho, \sigma = 1, \cdots, s)$$

of degree s such that $D(1) = 1$, (the unit matrix of degree s), $D(ab) = D(a) \cdot D(b)$, then we call $D(a)$ a representation of $\mathfrak{G}$. (No continuity is assumed.) Two representations $D(a)$ and $D'(a)$ are called equivalent if they are of the same degree s, and if a fixed matrix $U = \{ U_{\rho\sigma} \}$, $\rho, \sigma = 1, \cdots, s$, exists which transforms one representation into the other: $U^{-1}D(a)U = D'(a)$. A representation $D(a)$ is called reducible (completely reducible) if it is equivalent to a representation $D'(a)$ such that $D'_{\rho\sigma}(a) = 0$ identically (in a) whenever $\rho \leq t$, $\sigma > t$ ($\rho \leq t$, $\sigma > t$ or $\rho > t$, $\sigma \leq t$)[‡], for a fixed value of t, $1 \leq t \leq s - 1$. Representations without these properties are irreducible (completely irreducible).

For finite groups $\mathfrak{G}$ Frobenius and Schur gave a complete theory of all representations [21, 22]; for continuous groups $\mathfrak{G}$ close analogues of their results were established by Schur for the rotation group in three dimensions, and in much broader generality by Weyl for all compact Lie-groups [30]. These results were extended to all compact groups $\mathfrak{G}$ by Haar [11, pp. 166–169] with the help of his notion of "right-invariant" Lebesgue measure in groups. We shall push the extension further to all groups $\mathfrak{G}$, but in order to do this it is natural and necessary to restrict the domain of representations of $\mathfrak{G}$ by means of

---

[†] Cf. footnote to Theorem 16.

[‡] These are the fundamental notions of the Frobenius-Schur theory of group representations [21, 22, 30].

THEOREM 19. *The following conditions on a representation $D(a)$ of $\mathfrak{G}$ are equivalent to each other*:

A. $D(a)$ *is equivalent to a unitary*† *representation.*

B. *All elements $D_{\rho\sigma}(a)$ of $D(a)$ are bounded.*

C. *All elements $D_{\rho\sigma}(a)$ of $D(a)$ are a.p.*

If $D'(a)$ is unitary, then, by the footnote just cited in the case where $\rho = \sigma$,

$$\sum_{\tau=1}^{s} | D_{\rho\tau}'(a) |^2 = 1, \qquad | D_{\rho\tau}'(a) | \le 1,$$

and all $D_{\rho\tau}'(a)$ are bounded. Therefore the elements $D_{\rho\tau}(a)$ of any $D(a)$ equivalent to $D'(a)$ must also be bounded. Thus A implies B.

If all $D_{\rho\sigma}(a)$ are bounded, every sequence $D_{\rho\sigma}(a_n)$, $n = 1, 2, \cdots$, contains a subsequence which converges for all ρ, $\sigma = 1, \cdots, s$. And then, since $D(xa_n) = D(x)D(a_n)$ and $D(a_n x) = D(a_n)D(x)$, the representations $D(xa_n)$ and $D(a_n x)$, and hence all $D_{\rho\sigma}(xa_n)$ and $D_{\rho\sigma}(a_n x)$, converge uniformly. Thus all $D_{\rho\sigma}(a)$ are a.p., that is, B implies C.

If all $D_{\rho\sigma}(a)$ are a.p., so are the expressions $\sum_{\tau=1}^{s} D_{\rho\tau}(a)\overline{D_{\sigma\tau}(a)}$, and we can form

$$A_{\rho\sigma} = M_x\left[\sum_{\tau=1}^{s} D_{\rho\tau}(x)\overline{D_{\sigma\tau}(x)} \right].$$

Now

$$\sum_{\tau=1}^{s} D_{\rho\tau}(a)\overline{D_{\sigma\tau}(a)} = \overline{\sum_{\tau=1}^{s} D_{\sigma\tau}(a)\overline{D_{\rho\tau}(a)}}$$

and for every system $\xi_1, \cdots, \xi_s$ which is not identically zero,

$$\sum_{\rho,\sigma=1}^{s}\left[\sum_{\tau=1}^{s} D_{\rho\tau}(a)\overline{D_{\sigma\tau}(a)} \right] \xi_\rho\bar{\xi}_\sigma = \sum_{\tau=1}^{s}\left| \sum_{\rho=1}^{s} D_{\rho\tau}(a)\xi_\rho \right|^2 > 0,$$

so that

$$A_{\rho\sigma} = \overline{A}_{\sigma\rho} \quad \text{and} \quad \sum_{\rho,\sigma=1}^{s} A_{\rho\sigma}\xi_\rho\bar{\xi}_\sigma > 0.$$

Therefore the matrix $A = \{A_{\rho\sigma}\}$ is Hermitian and positive definite. Hence

---

† That is, to a representation $D'(a)$ in which all matrices $D'(a) = \{D_{\rho\sigma}'(a)\}$ $(\rho, \sigma = 1, \cdots, s)$ are unitary. A matrix $U = \{U_{\rho\sigma}\}$ is called unitary if its adjoint $U^* = \{\overline{U}_{\sigma\rho}\}$ is reciprocal to it, that is, if $UU^* = U^*U = 1_s$, or more explicitly,

$$\sum_{\tau=1}^{s} U_{\rho\tau}\overline{U}_{\sigma\tau} = \sum_{\tau=1}^{s} U_{\tau\rho}\overline{U}_{\tau\sigma} = \delta_{\rho\sigma} = \begin{cases} 1, & \rho = \sigma, \\ 0, & \rho \ne \sigma. \end{cases}$$

there exists a matrix $X = \{X_{\rho\sigma}\}$ such that $A = XX^*$. On the other hand,

$$
\begin{aligned}
A_{\rho\sigma} &= M_x \left[\sum_{\tau=1}^{s} D_{\rho\tau}(x)\overline{D_{\sigma\tau}(x)} \right] = M_x \left[\sum_{\tau=1}^{s} D_{\rho\tau}(ax)\overline{D_{\sigma\tau}(ax)} \right] \\
&= M_x \left[\sum_{\rho',\sigma'=1}^{s} \sum_{\tau=1}^{s} D_{\rho\rho'}(a)D_{\rho'\tau}(x)\overline{D_{\sigma\sigma'}(a)}\,\overline{D_{\sigma'\tau}(x)} \right] \\
&= \sum_{\rho',\sigma'=1}^{s} D_{\rho\rho'}(a)\overline{D_{\sigma\sigma'}(a)} M_x \left[\sum_{\tau=1}^{s} D_{\rho'\tau}(x)\overline{D_{\sigma'\tau}(x)} \right] \\
&= \sum_{\rho',\sigma'=1}^{s} D_{\rho\rho'}(a)\overline{D_{\sigma\sigma'}(a)} A_{\rho'\sigma'},
\end{aligned}
$$

that is, $A = D(a)AD(a)^*$, or $XX^* = D(a)XX^*D(a)^*$, $X^{-1}D(a)XX^*D(a)^*X^{*-1} = 1$, $(X^{-1}D(a)X)(X^{-1}D(a)X)^* = 1$. In other words, the equivalent representation $X^{-1}D(a)X = D'(a)$ is unitary. Thus C implies A.

Our three statements together prove the equivalence of A, B, and C.

DEFINITION 10. *We call normal the representations satisfying one of the equivalent conditions of Theorem* 19.

11. The fundamental theorems of the theory of orthogonal representations may now be proved in the classical way [21, 22, 30, 32].

THEOREM 20. *Let $D(a)$ and $E(a)$ be completely irreducible normal representations of degrees s and t respectively, and let A be a rectangular matrix with s rows and t columns. If $D(a)A \equiv AE(a)$ for every a, then either $A = 0$ or $s = t$ and $\det A \neq 0$, the latter alternative of course implying the equivalence of $D(a)$ and $E(a)$. If $D(a) = E(a)$, then $A = \alpha 1$ (α being a complex number).*

In all these statements (except the last) $D(a)$ and $E(a)$ may be replaced by two equivalent representations. Therefore we may assume them to be unitary. Even then further transformations by unitary matrices X and Y are possible. They carry A into $A' = X^{-1}AY$. Now by such transformations we can obtain $A' = \{A_{\rho\sigma'}\}$, $\rho = 1, \cdots, s$; $\sigma = 1, \cdots, t$, such that

$$
A_{\rho\sigma'} = \begin{cases} c_\rho, \text{ for } \rho = \sigma = 1, \cdots, r, \text{ all } c_\rho > 0, \\ 0, \text{ for all other } \rho \text{ and } \sigma, \end{cases}
$$

where r is the rank of A' and $r \leqq s$, $r \leqq t$. Therefore we may assume that A itself has this form.

Under these conditions the relation $D(a)A = AE(a)$ implies that $D_{\rho\sigma}(a) = 0$ for $\rho > r$ and $\sigma \leqq r$, and that $E_{\rho\sigma}(a) = 0$ for $\rho \leqq r$ and $\sigma > r$. Since A^* also has the form we assumed for A, $D(a)^* = D(a)^{-1} = D(a^{-1})$, $E(a)^* = E(a)^{-1} = E(a^{-1})$,

we get, by applying * and replacing a by a^{-1}, $A^*D(a) = E(a)A^*$, so that $D_{\rho\sigma}(a) = 0$ for $\rho \leq r$ and $\sigma > r$, and $E_{\rho\sigma}(a) = 0$ for $\rho > r$ and $\sigma \leq r$. Thus the complete irreducibility of $D(a)$ requires that r be 0 or s, and the complete irreducibility of $E(a)$ requires that r be 0 or t. Hence either $r = 0$, in which case $A = 0$, or $r = s = t$, in which case $\det A \neq 0$.

If $D(a) = E(a)$, every $A - \alpha 1$ has the same property as A. If α is a root of the characteristic equation of A, we have $\det [A - \alpha 1] = 0$, so that our alternative requires that $A - \alpha 1 = 0$, and $A = \alpha 1$.

THEOREM 21. *Let $D(a)$ and $E(a)$ be completely irreducible normal representations of degrees s and t respectively. If they are inequivalent, we have*

$$D_{\rho\sigma} \times E_{\tau\upsilon}(x) \equiv 0.$$

Considering $D(a)$ alone, we have

$$D_{\rho\sigma} \times D_{\tau\upsilon}(x) \equiv \begin{cases} \dfrac{1}{s} D_{\rho\upsilon}(x) \ for \ \sigma = \tau, \\[2mm] 0 \quad for \ \sigma \neq \tau. \end{cases}$$

Form the (rectangular) matrix

$$A = \{A_{\tau\sigma}\}, \quad A_{\tau\sigma} = D_{\rho\sigma} \dot\times E_{\tau\upsilon}(x)$$

for a given choice of ρ, υ, and x. Then

$$\sum_{\tau'=1}^{t} E_{\tau\tau'}(a)A_{\tau'\sigma} = M_y \left[\sum_{\tau'=1}^{t} D_{\rho\sigma}(xy^{-1})E_{\tau\tau'}(a)E_{\tau'\upsilon}(y) \right]$$

$$= M_y[D_{\rho\sigma}(xy^{-1})E_{\tau\upsilon}(ay)] = M_z[D_{\rho\sigma}(z)E_{\tau\upsilon}(az^{-1}x)],$$

$$\sum_{\sigma'=1}^{s} A_{\tau\sigma'}D_{\sigma'\sigma}(a) = M_y \left[\sum_{\sigma'=1}^{s} D_{\rho\sigma'}(y)D_{\sigma'\sigma}(a)E_{\tau\upsilon}(y^{-1}x) \right]$$

$$= M_y[D_{\rho\sigma}(ya)E_{\tau\upsilon}(y^{-1}x)] = M_z[D_{\rho\sigma}(z)E_{\tau\upsilon}(az^{-1}x)]$$

(the variable y being replaced by $z = xy^{-1}$ and $z = ya$ respectively), and therefore

$$E(a)A = AD(a).$$

Thus we can apply Theorem 20. If $D(a)$ and $E(a)$ are inequivalent, it results that $A = 0$, and if $D(a) = E(a)$, $A = \alpha_{\rho\upsilon}(x)1$, ($\alpha_{\rho\upsilon}(x)$ being a complex number). This implies (1) that $A_{\tau\sigma} = 0$, that is, $D_{\rho\sigma} \times E_{\tau\upsilon}(x) = 0$ if $D(a)$ and $E(a)$ are inequivalent or if $D(a) = E(a)$ and $\sigma \neq \tau$; and (2) that $A_{\tau\sigma} = \alpha_{\rho\upsilon}(x)$, which is independent of σ, if $D(a) = E(a)$ and $\sigma = \tau$. Hence all the statements of our Theorem are proved if we show that $\alpha_{\rho\upsilon}(x) = (1/s)D_{\rho\upsilon}(x)$. This follows immediately from

$$s\alpha_{\rho v}(x) = \sum_{\sigma=1}^{s} A_{\sigma\sigma} = M_y\left[\sum_{\sigma=1}^{s} D_{\rho\sigma}(xy^{-1})D_{\sigma v}(y)\right]$$

$$= M_y D_{\rho v}(xy^{-1}y) = M_y D_{\rho v}(x) = D_{\rho v}(x).$$

THEOREM 22. *For normal representations reducibility and complete reducibility are equivalent, so that irreducibility and complete irreducibility are also equivalent.*

That complete reducibility implies reducibility is obvious. Assume now that $D(a)$ is reducible without being completely reducible. As we can replace $D(a)$ by any equivalent representation, we may assume $D(a)$ to be in the form described in Definition 9. Then there would be a pair of indices ρ and σ such that $D_{\rho\sigma}(x)\equiv 0$. By Theorem 21, this relation implies that $D_{\rho\sigma}\times D_{\sigma\rho}\equiv D_{\rho\rho}\equiv 0$, in spite of the fact that $D_{\rho\rho}(1)=1$. Thus $D(a)$ must be completely reducible.

12. Our next task is to formulate the connection between the units of Definition 8 and representations. This is accomplished by means of

THEOREM 23. *For every unit $\phi(x)$ there exist a number of inequivalent irreducible unitary representations $D^{(1)}(a), \cdots, D^{(u)}(a)$ of degrees $s_1, \cdots, s_u$ respectively such that*

$$\phi(x) = \sum_{\omega=1}^{u} s_\omega \sum_{\rho,\sigma=1}^{s_\omega} \alpha_{\rho\sigma}^{(\omega)} D_{\rho\sigma}^{(\omega)}(x).$$

Here every matrix $\alpha^{(\omega)} = \{\alpha_{\rho\sigma}^{(\omega)}\}$ is idempotent, that is, $\alpha^{(\omega)*}=\alpha^{(\omega)}$ (cf. footnote on page 465), $(\alpha^{(\omega)})^2=\alpha^{(\omega)}$.† Conversely, every $\phi(x)$ which is formed in this way (where $D^{(\omega)}(a)$ and $\alpha^{(\omega)}$ satisfy our conditions) is a unit.

By a suitable choice of $D^{(\omega)}(a)$ we can give the matrices $\alpha^{(\omega)}$ the form

$$\alpha_{\rho\sigma}^{(\omega)} = \begin{cases} 1 \text{ for } \rho = \sigma \leq s_\omega', \text{ with some } s_\omega' = 1, \cdots, s_\omega, \\ 0 \text{ for all other } \rho \text{ and } \sigma. \end{cases}$$

Consider the (a.p.) solutions $f(x)$ of the equation $\phi\times f=f$. Assume that it is possible to find s solutions $g_1, \cdots, g_s$ among them which satisfy the conditions

$$g_\mu \times g_\nu'(1) = M_x[g_\mu(x)\overline{g_\nu(x)}] = \begin{cases} 1 \text{ for } \mu = \nu, \\ 0 \text{ for } \mu \neq \nu. \end{cases}$$

Put

---

† That is,

$$\alpha_{\rho\sigma}^{(\omega)} = \overline{\alpha_{\sigma\rho}^{(\omega)}}, \qquad \sum_{\sigma=1}^{s_\omega}\alpha_{\rho\sigma}^{(\omega)} \alpha_{\sigma\tau}^{(\omega)} = \alpha_{\rho\tau}^{(\omega)}.$$

$$\psi(x, y) = \phi(xy^{-1}) - \sum_{\mu=1}^{s} g_\mu(x)\overline{g_\mu(y)}.$$

Then

$$M_{x,y}\big[|\psi(x, y)|^2\big] = M_{x,y}\big[|\phi(xy^{-1})|^2\big] - \sum_{\mu=1}^{s} M_{x,y}\big[\phi(xy^{-1})\overline{g_\mu(x)}g_\mu(y)\big]$$

$$- \sum_{\mu=1}^{s} M_{x,y}\big[\overline{\phi(xy^{-1})}\,g_\mu(x)\,\overline{g_\mu(y)}\big] + \sum_{\mu,\nu=1}^{s} M_{x,y}\big[g_\mu(x)\overline{g_\mu(y)}\,\overline{g_\nu(x)}g_\nu(y)\big].$$

By Theorems 9 and 10, the orthogonality properties of g_μ, and the relations $\phi \times g_\mu = g_\mu$, $g_\mu' \times \phi = g_\mu'$, this turns out to be equal to

$$M_x\big[|\phi(x)|^2\big] - s - s + s = (N\phi)^2 - s.$$

Therefore we have $(N\phi)^2 - s \geq 0$, so that $s \leq (N\phi)^2$. Thus the possible numbers s are bounded, and they have a maximal value. Assume that s is this maximal value and choose $g_1, \cdots, g_s$ accordingly. If a solution $f(x)$ of $\phi \times f = f$ is such that $f \times g_\mu' = 0$ for all $\mu = 1, \cdots, s$, then necessarily $f \equiv 0$, for otherwise we could put $g_{s+1} = f/(Nf)$, implying that $Ng_{s+1} = 1$, that is, $g_{s+1} \times g'_{+1}(1) = 1$, and $g_{s+1} \times g_\mu'(1) = 0$ as well as $g_\mu \times g'_{s+1}(1) = 0$ for $\mu = 1, \cdots, s$, so that

$$g_\mu \times g_\nu'(1) = \begin{cases} 1 \text{ for } \mu = \nu, \\ 0 \text{ for } \mu \neq \nu, \end{cases} \qquad \mu, \nu = 1, \cdots, s+1,$$

contradicting our assumption that s is maximal.

If we define $f(x)$ to be $\psi(x, a)$, we find by a simple computation that $\phi \times f = f$ and $f \times g_\mu'(1) = 0$. Therefore $f(x) \equiv 0$, $\psi(x, a) \equiv 0$, and, as a was arbitrary, $\psi(x, y) \equiv 0$, that is,

(‡) $$\phi(xy^{-1}) \equiv \sum_{\mu=1}^{s} g_\mu(x)\overline{g_\mu(y)}.$$

As (‡) implies that

$$\phi \times f(x) = M_y\big[\phi(xy^{-1})f(y)\big] = \sum_{\mu=1}^{s} M_y\big[\overline{g_\mu(y)}f(y)\big]g_\mu(x),$$

every solution of $\phi \times f = f$ has the form $\sum_{\mu=1}^{s} \alpha_\mu g_\mu$ (α_μ being complex numbers). It is obvious that $f(xa)$ is a solution along with $f(x)$, so that $g_\mu(xa)$ is a solution, and we can write

$$g_\sigma(xa) = \sum_{\rho=1}^{s} D_{\rho\sigma}(a)g_\rho(x).$$

The orthogonality properties of g_μ determine the coefficients $D_{\rho\sigma}(a)$ uniquely:

$$D_{\rho\sigma}(a) = M_x[g_\sigma(xa)\overline{g_\rho(x)}] = g'_\rho \times g_\sigma(a).$$

Hence $D(1) = 1_s$, and if we put $D(a) = \{D_{\rho\sigma}(a)\}$, $\rho, \sigma = 1, \cdots, s$, we obviously have $D(ab) = D(a)D(b)$, that is, $D(a)$ is a representation. As ($\ddagger$) holds if we replace x and y by xa and ya, that is, if we replace $g_\mu(x)$ and $g_\mu(y)$ by $g_\mu(xa)$ and $g_\mu(ya)$, the transformations $D(a)$ must be unitary.

All this implies that

$$\phi(x) = \sum_{\sigma=1}^{s} g_\sigma(x)\overline{g_\sigma(1)} = \sum_{\rho,\sigma=1}^{s} D_{\rho\sigma}(x)g_\rho(1)\overline{g_\sigma(1)},$$

so that $\phi(x)$ is a linear aggregate of all $D_{\rho\sigma}(x)$. Now (cf. Theorem 22) $D(a)$ can be transformed into an equivalent $D'(a)$ which consists of a certain number of irreducible representations $D^{(1)}(a), \cdots, D^{(v)}(a)$ (of degrees $s_1, \cdots, s_v$ respectively, where $s_1 + \cdots + s_v = s$) which succeed each other along the main diagonal, and zeros in all other places. Therefore $\phi(x)$ is also a linear aggregate of the elements $D_{\rho\sigma}'(x)$, that is, of the elements $D^{(\omega)}_{\rho\sigma}(x)$. Now if some representation $D^{(\omega)}(x)$ is equivalent to another representation $D^{(\tau)}(x)$, the elements $D^{(\omega)}_{\rho\sigma}(x)$ are linear aggregates of elements of $D^{(\tau)}(x)$. Therefore, in expressing $\phi(x)$ as a linear aggregate of all elements $D^{(\omega)}_{\rho\sigma}(x)$, it is sufficient to keep only one member of each class of equivalent representations $D^{(\omega)}(x)$. Those representations thus kept may be labeled $D^{(1)}(x), \cdots, D^{(u)}(x)$, $u \leq v$. So we finally have the result that $D^{(1)}(a), \cdots, D^{(u)}(a)$ (of degrees $s_1, \cdots, s_u$ respectively) is a set of inequivalent irreducible unitary representations, and $\phi(x)$ is a linear aggregate of the elements $D^{(\omega)}_{\rho\sigma}(x)$, that is, we can write $\phi(x)$ in the form

$$\phi(x) = \sum_{\omega=1}^{u} s_\omega \sum_{\rho,\sigma=1}^{s_\omega} \alpha^{(\omega)}_{\rho\sigma} D^{(\omega)}_{\rho\sigma}(x).$$

If by means of this equation we now determine the meaning of $\phi' = \phi$ and $\phi \times \phi = \phi$, remembering that

$$(D^{(\omega)}(x))^* = D^{(\omega)}(x)^{-1} = D^{(\omega)}(x^{-1}),$$

$$\overline{D^{(\omega)}_{\sigma\rho}(x)} = D^{(\omega)}_{\rho\sigma}(x^{-1}), \quad \overline{D^{(\omega)}_{\rho\sigma}(x^{-1})} = D^{(\omega)}_{\sigma\rho}(x),$$

that is, $D^{(\omega)\prime}_{\sigma\rho}(x) = D^{(\omega)}_{\sigma\rho}(x)$, and

$$D_{\rho\sigma}^{(\omega)} \times D_{\tau v}^{(\chi)} = \begin{cases} \dfrac{1}{s_\omega} D_{\rho v}^{(\omega)} \text{ if } \omega = \chi \text{ and } \sigma = \tau, \\ \\ 0 \quad \text{ for all other } \omega, \chi, \rho, \sigma, \tau, v, \end{cases}$$

we obtain exactly the conditions in our Theorem. Furthermore it is clear that every matrix $\alpha^{(\omega)} = \{\alpha_{\rho\sigma}^{(\omega)}\}$, being idempotent, can be transformed into the form given at the end of our Theorem. And the inverse transformations of the representations $D^{(\omega)}(x)$, which carry them into equivalent representations, bring about just these $\alpha^{(\omega)}$-transformations.

13. We choose a system of "representants" for the inequivalent irreducible (normal or orthogonal) representations of $\mathfrak{G}$:

DEFINITION 11. *Let I be the set of all irreducible normal representations of $\mathfrak{G}$. Call each subset $\mathfrak{C}$ of I which consists of all the elements of I equivalent to one of its elements a class. It is obvious that every element of I belongs to exactly one class. Call the set of all classes C. Each element $\mathfrak{C}$ of C contains unitary representations (since every normal representation is equivalent to a unitary representation). Select one unitary representation from each $\mathfrak{C}$ of C, call it the representant of $\mathfrak{C}$, denote it by $D(a; \mathfrak{C})$†, and denote its degree by $s(\mathfrak{C})$.*

THEOREM 24. *The (a.p.) functions $D_{\rho\sigma}(x; \mathfrak{C})$, $\mathfrak{C}$ in C, ρ and $\sigma = 1, \cdots, s(\mathfrak{C})$, have the property that*

$$D_{\rho\sigma}(\mathfrak{C}) \times D_{\tau v}(\mathfrak{D}) = \begin{cases} \dfrac{1}{s(\mathfrak{C})} D_{\rho v}(\mathfrak{C}) \text{ for } \mathfrak{C} = \mathfrak{D} \text{ and } \sigma = \tau, \\ \\ 0 \quad \text{ for all other } \mathfrak{C}, \mathfrak{D}, \rho, \sigma, \tau, v. \end{cases}$$

This implies that

$$M_x[D_{\rho\sigma}(x; \mathfrak{C})\overline{D_{v\tau}(x; \mathfrak{D})}] = \begin{cases} \dfrac{1}{s(\mathfrak{C})} \text{ for } \mathfrak{C} = \mathfrak{D}, \rho = v, \sigma = \tau, \\ \\ 0 \text{ for all other } \mathfrak{C}, \mathfrak{D}, \rho, \sigma, \tau, v. \end{cases}$$

The first formula has been proved in Theorem 21. The second formula follows from it if we put the variable equal to 1 and remember that $D_{\tau v}(\mathfrak{D})' = \overline{D_{v\tau}(\mathfrak{D})}$, $D(1, \mathfrak{D}) = 1$.

Thus the functions $s(\mathfrak{C})^{1/2} D_{\rho\sigma}(x; \mathfrak{C})$ form a "normalized orthogonal" system. This is the basis for the formulation of the usual expansion theorems. The key theorems of the theory will be proved as Theorems 28 and 29.

---

† This does not imply an essential use of the "axiom of choice" because it would in most cases be possible to characterize a $D(a; \mathfrak{C})$ in $\mathfrak{C}$ in a unique way. To abbreviate we shall also use the notation $D(\mathfrak{C})$ omitting the argument a.

DEFINITION 12. *If $f(x)$ is an a.p. function, the complex numbers*

$$\tilde{\alpha}_{\rho\sigma}(\mathfrak{C}) = f \times D_{\rho\sigma}(1;\mathfrak{C})' = M_x[f(x)\overline{D_{\rho\sigma}(x;\mathfrak{C})}]$$

are called its expansion coefficients. The matrices $\tilde{\alpha}(\mathfrak{C}) = \{\tilde{\alpha}_{\rho\sigma}(\mathfrak{C})\}$ are called the expansion matrices.

THEOREM 25. *If f and g have the expansion matrices $\tilde{\alpha}(\mathfrak{C})$ and $\tilde{\beta}(\mathfrak{C})$ ($\mathfrak{C}$ running over C), $f \pm g$, θf (θ any complex number), f' and $f \times g$ have the expansion matrices $\tilde{\alpha}(\mathfrak{C}) \pm \tilde{\beta}(\mathfrak{C})$, $\theta\tilde{\alpha}(\mathfrak{C})$, $\tilde{\alpha}(\mathfrak{C})^*$, and $\tilde{\alpha}(\mathfrak{C})\tilde{\beta}(\mathfrak{C})$ respectively. A unit ϕ is characterized by the fact that all its expansion matrices are idempotent, so that only a finite number of them can be $\neq 0$.*

The statements concerning $f \pm g$ and θf are obvious; as to f' it is sufficient to remark that

$$M_x[f'(x)\overline{D_{\rho\sigma}(x;\mathfrak{C})}] = M_x[\overline{f(x^{-1})}D_{\rho\sigma}(x^{-1};\mathfrak{C})']$$

$$= M_x[\overline{f(x)}D_{\sigma\rho}(x;\mathfrak{C})'] = \overline{M_x[f(x)\overline{D_{\sigma\rho}(x;\mathfrak{C})}]} = \overline{\tilde{\alpha}_{\sigma\rho}(\mathfrak{C})}.$$

The following computation proves our statement with regard to $f \times g$ (cf. Theorems 7 and 8):

$$M_x[f \times g(x)\overline{D_{\rho\sigma}(x;\mathfrak{C})}] = M_{x,y}[f(y)g(y^{-1}x)\overline{D_{\rho\sigma}(x;\mathfrak{C})}]$$

$$= M_{y,z}[f(y)g(z)\overline{D_{\rho\sigma}(yz;\mathfrak{C})}]$$

$$= \sum_{\tau=1}^{s(\mathfrak{C})} M_{y,z}[f(y)g(z)\overline{D_{\rho\tau}(y;\mathfrak{C})}\,\overline{D_{\tau\sigma}(z;\mathfrak{C})}]$$

$$= \sum_{\tau=1}^{s(\mathfrak{C})} M_y[f(y)\overline{D_{\rho\tau}(y;\mathfrak{C})}]M_z[g(z)\overline{D_{\tau\sigma}(z;\mathfrak{C})}]$$

$$= \sum_{\tau=1}^{s(\mathfrak{C})} \tilde{\alpha}_{\rho\tau}(\mathfrak{C})\tilde{\beta}_{\tau\sigma}(\mathfrak{C}).$$

This discussion shows that the idempotence of all expansion matrices $\tilde{\alpha}(\mathfrak{C})$ is characteristic of units ϕ, but it follows from Theorem 21 that all those matrices which have a $D(a;\mathfrak{C})$ not identical to a $D^{(\omega)}(a)$ for some $\omega = 1, \cdots, u$ mush vanish, so that only a finite number are different from zero. (We here use the fact that if two functions have the same expansion matrices, they coincide; that is, if all expansion matrices of a function f vanish, then $f(x) \equiv 0$. This follows from Theorem 28 (the proof of which does not depend upon Theorem 25) by putting $f(x) \equiv g(x)$.)

THEOREM 26. *If f has the expansion matrices $\tilde{\alpha}(\mathfrak{C}) = \{\tilde{\alpha}_{\rho\sigma}(\mathfrak{C})\}$ and if $\mathfrak{C}_1, \cdots, \mathfrak{C}_n$ are elements of C, then of all linear aggregates g of the elements $D_{\rho\sigma}(\mathfrak{C}_\omega)$, $\omega = 1, \cdots, n$; $\rho, \sigma = 1, \cdots, s(\mathfrak{C}_\omega)$, which can be written in the form*

$$g(x) = \sum_{\omega=1}^{n} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} \alpha_{\rho\sigma}(\mathfrak{C}_\omega) D_{\rho\sigma}(x;\mathfrak{C}_\omega)$$

that one which minimizes the expression $(N[f-g])^2$ is characterized by the property that $\alpha_{\rho\sigma}(\mathfrak{C}_\omega) = \tilde{\alpha}_{\rho\sigma}(\mathfrak{C}_\omega)$. The minimum value in question is

$$(Nf)^2 - \sum_{\omega=1}^{n} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} |\tilde{\alpha}_{\rho\sigma}(\mathfrak{C}_\omega)|^2.$$

The proof is contained in the well known computation

$$(N[f-g])^2 = M_x[|f(x)-g(x)|]^2$$

$$= M_x[|f(x)|^2] - M_x[f(x)\overline{g(x)}] - M_x[\overline{f(x)}g(x)] + M_x[|g(x)|^2]$$

$$= (Nf)^2 - \sum_{\omega=1}^{n} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} \overline{\alpha_{\rho\sigma}(\mathfrak{C}_\omega)} M_x[f(x)\overline{D_{\rho\sigma}(x;\mathfrak{C}_\omega)}]$$

$$- \sum_{\omega=1}^{n} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} \alpha_{\rho\sigma}(\mathfrak{C}_\omega) M_x[\overline{f(x)}D_{\rho\sigma}(x;\mathfrak{C}_\omega)]$$

$$+ \sum_{\omega,\chi=1}^{n} s(\mathfrak{C}_\omega)s(\mathfrak{C}_\chi) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} \sum_{\tau,\upsilon=1}^{s(\mathfrak{C}_\chi)} \alpha_{\rho\sigma}(\mathfrak{C}_\omega)\overline{\alpha_{\tau\upsilon}(\mathfrak{C}_\chi)} M_x[D_{\rho\sigma}(x;\mathfrak{C}_\omega)\overline{D_{\tau\upsilon}(x;\mathfrak{C}_\chi)}]$$

$$= (Nf)^2 - \sum_{\omega=1}^{n} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} \tilde{\alpha}_{\rho\sigma}(\mathfrak{C}_\omega)\overline{\alpha_{\rho\sigma}(\mathfrak{C}_\omega)}$$

$$- \sum_{\omega=1}^{n} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} \overline{\tilde{\alpha}_{\rho\sigma}(\mathfrak{C}_\omega)}\alpha_{\rho\sigma}(\mathfrak{C}_\omega) + \sum_{\omega=1}^{n} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} |\alpha_{\rho\sigma}(\mathfrak{C}_\omega)|^2$$

$$= \left\{ (Nf)^2 - \sum_{\omega=1}^{n} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} |\tilde{\alpha}_{\rho\sigma}(\mathfrak{C}_\omega)|^2 \right\}$$

$$+ \sum_{\omega=1}^{n} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} |\alpha_{\rho\sigma}(\mathfrak{C}_\omega) - \tilde{\alpha}_{\rho\sigma}(\mathfrak{C}_\omega)|^2.$$

THEOREM 27. *(Bessel's inequality.) If f has the expansion matrices $(\tilde{\alpha}\mathfrak{C}) = \{\tilde{\alpha}_{\rho\sigma}(\mathfrak{C})\}$, the number of those $\mathfrak{C}$ for which $\tilde{\alpha}(\mathfrak{C}) \neq 0$ is at most countably infinite, that is, it is possible to arrange them in a finite or infinite sequence $\mathfrak{C}_1, \mathfrak{C}_2, \cdots$. Then we have*

$$(Nf)^2 \geqq \sum_{\omega} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} |\tilde{\alpha}_{\rho\sigma}(\mathfrak{C}_\omega)|^2.$$

Since, for all other $\mathfrak{C}$'s, $\tilde{\alpha}_{\rho\sigma}(\mathfrak{C}) = 0$, we can instead write

$$(Nf)^2 \geqq \sum_{\mathfrak{C}} s(\mathfrak{C}) \sum_{\rho,\sigma=1}^{s(\mathfrak{C})} |\tilde{\alpha}_{\rho\sigma}(\mathfrak{C})|^2.$$

The number of $\mathfrak{C}$'s for which

$$s(\mathfrak{C}) \sum_{\rho,\sigma=1}^{s(\mathfrak{C})} | \tilde{\alpha}_{\rho\sigma}(\mathfrak{C}) |^2 \geq \epsilon, \quad \epsilon > 0,$$

is, because of the last statement of Theorem 26, certainly finite and $\leq (Nf)^2/\epsilon$. Putting $\epsilon = 1, \frac{1}{2}, \frac{1}{3}, \cdots$ successively, we see that, with at most countably infinitely many exceptional $\mathfrak{C}$'s, it is always the case that

$$s(\mathfrak{C}) \sum_{\rho,\sigma=1}^{s(\mathfrak{C})} | \tilde{\alpha}_{\rho\sigma}(\mathfrak{C}) |^2 = 0,$$

that is, $\tilde{\alpha}_{\rho\sigma}(\mathfrak{C}) = 0$. This proves the first statement of our Theorem.

For every n the last statement of Theorem 26 shows that

$$\sum_{\omega=1}^{n} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} | \tilde{\alpha}_{\rho\sigma}(\mathfrak{C}_\omega) |^2 \leq (Nf)^2.$$

Hence the sum

$$\sum_{\omega} s(\mathfrak{C}_\omega) \sum_{\rho,\sigma=1}^{s(\mathfrak{C}_\omega)} | \tilde{\alpha}_{\rho\sigma}(\mathfrak{C}_\omega) |^2$$

is convergent (if the number of terms is infinite) and $\leq (Nf)^2$.

THEOREM 28. (*Parseval's equation.*) *If f and g have the expansion matrices* $\tilde{\alpha}(\mathfrak{C}) = \{\tilde{\alpha}_{\rho\sigma}(\mathfrak{C})\}$ *and* $\tilde{\beta}(\mathfrak{C}) = \{\tilde{\beta}_{\rho\sigma}(\mathfrak{C})\}$, *then*

$$M_x[f(x)\overline{g(x)}] = \sum_{\mathfrak{C}} s(\mathfrak{C}) \sum_{\rho,\sigma=1}^{s(\mathfrak{C})} \tilde{\alpha}_{\rho\sigma}(\mathfrak{C})\overline{\tilde{\beta}_{\rho\sigma}(\mathfrak{C})},$$

where the series $\sum_{\mathfrak{C}}$ *contains at most countably infinitely many terms* $\neq 0$, *and is absolutely convergent (if infinite at all).*

If this is proved for $f = g$, we obtain the real part of the statement by replacing our f by $(f+g)/2$ and $(f-g)/2$ and subtracting. Replacing f and g by if and g gives the imaginary part and completes the proof. Hence we may assume that $f = g$, that is, we must show that

$$(Nf)^2 - \sum_{\mathfrak{C}} s(\mathfrak{C}) \sum_{\rho,\sigma=1}^{s(\mathfrak{C})} | \tilde{\alpha}_{\rho\sigma}(\mathfrak{C}) |^2 = 0.$$

If we take all finite subsets of C for the $\mathfrak{C}_1, \cdots, \mathfrak{C}_n$ in Theorem 26, we see that the left side of the above equation is the greatest lower bound of all $(N[f-g])^2$ if g is any linear aggregate of any finite number of elements $D_{\rho\sigma}(\mathfrak{C})$. Hence we need to show merely that it can be made $\leq \epsilon^2$ for any $\epsilon > 0$. This is accomplished by choosing the unit $\phi(x)$ according to Theorem 16, be-

cause $\phi \times f$ is a linear aggregate of a finite number of elements $D_{\rho\sigma}(\mathfrak{C})$. By Theorem 23 we need to prove this only for the elements $D_{\tau v}(\mathfrak{D}) \times f$ and it follows that

$$D_{\tau v}(\mathfrak{D}) \times f(x) = M_y[D_{\tau v}(xy^{-1}; \mathfrak{D})f(y)]$$

$$= \sum_{\lambda=1}^{s(\mathfrak{D})} M_y[D_{\tau\lambda}(x; \mathfrak{D})D_{\lambda v}(y^{-1}; \mathfrak{D})f(y)]$$

$$= \sum_{\lambda=1}^{s(\mathfrak{D})} M_y[D_{\lambda v}(y^{-1}; \mathfrak{D})f(y)]D_{\tau\lambda}(x; \mathfrak{D}).$$

THEOREM 29. (*Approximation Theorem.*) *For every a.p. function $f(x)$ and every $\epsilon > 0$ there exists a linear aggregate $h(x)$ of a finite number of elements $D_{\rho\sigma}(x; \mathfrak{C})$ (which can be limited to such elements for which the expansion matrix $\tilde{\alpha}(\mathfrak{C})$ of f is $\neq 0$) such that $|f(x) - h(x)| \leq \epsilon$ for every x.*

By Theorem 18 we may put $h = \phi \times f \times g$, so that we need to prove merely that $\phi \times f \times g$ is a linear aggregate of the desired kind. By Theorem 23 we may consider $D_{\tau v}(\mathfrak{D}) \times f \times g$. The last formula of the preceding proof shows that this is a linear aggregate of a finite number of elements $D_{\tau\lambda}(\mathfrak{D})$. This formula gives for the coefficient of $D_{\tau\lambda}(\mathfrak{D})$ in $D_{\rho\sigma}(\mathfrak{D}) \times f \times g$ (replace its f by $f \times g$)

$$M_y[D_{\lambda v}(y^{-1}; \mathfrak{D})f \times g(y)] = M_y[\overline{D_{\lambda v}(y; \mathfrak{D})}'f \times g(y)]$$

$$= M_y[f \times g(y)\overline{D_{v\lambda}(y; \mathfrak{D})}].$$

Thus it is the expansion coefficient of $D_{v\lambda}(\mathfrak{D})$ in $f \times g$, and this is equal to zero if the expansion matrix of $D_{v\lambda}(\mathfrak{D})$ in f is zero (cf. the statement of Theorem 25 concerning $f \times g$).

THEOREM 30. *Each a.p. function is the limit of a uniformly convergent sequence of functions each of which is a linear aggregate of a finite number of elements $D_{\rho\sigma}(\mathfrak{C})$, and conversely.*

The statement follows from Theorem 29 by putting $\epsilon = 1, \frac{1}{2}, \frac{1}{3}, \cdots$ in succession. The converse statement is a consequence of the a.p. character of all elements $D_{\rho\sigma}(\mathfrak{C})$.

IV. ALMOST PERIODICITY AND CLOSED FAMILIES OF FUNCTIONS

14. Parts I–III give a fairly complete theory of a.p. functions in an arbitrary group $\mathfrak{G}$, absolutely free from the customary restriction of continuity. We now introduce restrictions of this type, but in a more general manner, by considering certain families of functions.

DEFINITION 13. *A set S of functions f(x) (defined in $\mathfrak{G}$ with complex numbers as values) is called a closed family (cl.f.) if it has the following properties:*

(1) *If f(x) is in S, every f(xa) is in S.*

(2) *If f(x) is in S, every f(ax) is in S.*

(3) *If f(x) is in S, every αf(x) is in S.*

(4) *If f(x) and g(x) are in S, f(x) ± g(x) is in S.*

(5) *If $f_1(x)$, $f_2(x)$, $\cdots$ are in S and if $f_n(x)$ converges uniformly to f(x) as $n \rightarrow \infty$, then f(x) is in S.*

THEOREM 31. *If S is a cl.f. and contains either f or g, then it contains $f \times g$; if $D(a) = \{D_{\rho\sigma}(a)\}$ is an irreducible normal representation, S contains every $D_{\rho\sigma}$ if it contains one $D_{\rho\sigma}$; if the system of representative irreducible normal representations $D(a; \mathfrak{C})$ is given (cf. Definition 11), and if S contains f, then S contains all elements $D_{\rho\sigma}(\mathfrak{C})$ of every $D(\mathfrak{C})$ whose expansion matrix $\tilde{\alpha}(\mathfrak{C})$ (cf. Definition 12) is $\neq 0$.*

Therefore Theorems 28 (Parseval's equation), 29 (Approximation Theorem), and 30 remain true if we restrict ourselves throughout to functions in S.

If f belongs to S, $f \times g$ belongs to S by the Remark following Definition 6. The case where g belongs to S can be reduced to the case where f belongs to S by replacing ab by ba in $\mathfrak{G}$. If a $D_{\rho\sigma}$ belongs to S, every $D_{\rho'\sigma'}$ belongs to it, since, by Theorem 21, $s^2 D_{\rho'\rho} \times D_{\rho\sigma} \times D_{\sigma\sigma'} = D_{\rho'\sigma'}$. Finally,

$$M_y[f(y)D_{\sigma'\sigma}(y^{-1}x; \mathfrak{C})] = M_y\left[f(y)\sum_{\tau=1}^{s} D_{\sigma'\tau}(y^{-1}; \mathfrak{C})D_{\tau\sigma}(x; \mathfrak{C})\right]$$

$$= \sum_{\tau=1}^{s} M_y[f(y)D_{\sigma'\tau}(y^{-1}; \mathfrak{C})]D_{\tau\sigma}(x; \mathfrak{C}) = \sum_{\tau=1}^{s} M_y[f(y)\overline{D_{\tau\sigma'}(y; \mathfrak{C})}]D_{\tau\sigma}(x; \mathfrak{C}),$$

that is,

$$f \times D_{\sigma'\sigma}(\mathfrak{C}) = \sum_{\tau=1}^{s} \tilde{\alpha}_{\tau\sigma'}(\mathfrak{C})D_{\tau\sigma}(\mathfrak{C}),$$

$$D_{\rho\rho'}(\mathfrak{C}) \times f \times D_{\sigma'\sigma}(\mathfrak{C}) = \frac{1}{s} \tilde{\alpha}_{\rho'\sigma'}(\mathfrak{C})D_{\rho\sigma}(\mathfrak{C}).$$

Hence if $\tilde{\alpha}(\mathfrak{C}) \neq 0$ for a given $\mathfrak{C}$, that is, if any $\tilde{\alpha}_{\rho'\sigma'}(\mathfrak{C}) \neq 0$, and if f belongs to S, then $f \times D_{\sigma'\sigma}(\mathfrak{C})$, $D_{\rho\rho'}(\mathfrak{C}) \times f \times D_{\sigma'\sigma}(\mathfrak{C})$, and $D_{\rho\sigma}(\mathfrak{C})$ in turn belong to S.

If we keep these facts in mind we see that the proofs of Theorems 28, 29, and 30 still hold in S.

DEFINITION 14. *If a topology T is given in $\mathfrak{G}$† we denote the set of all T-continuous functions by $[T]$.*

THEOREM 32. *If a topology T is given in $\mathfrak{G}$† in which ab is continuous in a for a fixed b, and in b for a fixed a, then $[T]$ is a cl.f.*

The statement is obvious.

The a.p. functions of a cl.f. S are determined (in the manner described in Theorem 30; cf. the last statement of Theorem 31) by the elements $D_{\rho\sigma}(\mathfrak{C})$ belonging to it. This greatly facilitates the determination of all a.p. functions of a given cl.f. S.

15. We shall discuss some examples in detail.

EXAMPLE 1. Let $\mathfrak{G} = \mathfrak{G}_{rat}$ be the set of all rational numbers with addition as the rule of composition. As the group is Abelian, all irreducible representations are of degree 1, $D(a; \mathfrak{C}) = \{D_{\rho\sigma}(a; \mathfrak{C})\}$, $\rho, \sigma = 1$, so that we have a single element $D_{11}(a; \mathfrak{C}) = \phi(a; \mathfrak{C})$. The fact that this is a unitary representation is expressed by the relation

(*) $$\phi(a)\phi(b) = \phi(a + b), \qquad |\phi(a)| = 1.$$

Every rational number can be written in the form $a = m/n!$ ($m = 0, \pm 1, \pm 2, \cdots ; n = 1, 2, \cdots$). Now put

$$\phi\left(\frac{1}{n!}\right) = e^{2\pi\lambda_n i}, \qquad 0 \leq \lambda_n < 1.$$

Then

(**) $$\phi(a) = \phi\left(\frac{m}{n!}\right) = e^{2\pi\lambda_n m i}.$$

Since

$$\frac{n+1}{(n+1)!} = \frac{1}{n!},$$

it follows that

($^{*}_{**}$) $$(n + 1)\lambda_{n+1} \equiv \lambda_n \pmod 1 \qquad (n = 1, 2, \cdots)$$

and, on the other hand, it is clear that ($^{*}_{**}$) makes (**) a definition prescribing a unique value for $\phi(a)$ which satisfies (*). So the general solution is (**), with the further condition ($^{*}_{**}$). An alternative way of writing ($^{*}_{**}$) is

($^{**}_{**}$) $$\lambda_{n+1} \equiv \frac{\lambda_n + p_n}{n+1}, \qquad p_n = 0, 1, \cdots, n; \quad n = 1, 2, \cdots.$$

† See first footnote on page 447.

EXAMPLE 2. Take the same $\mathfrak{G} = \mathfrak{G}_{rat}$, but take its normal topology $T = T_0$ (distance $|a-b|$) and consider $S = [T_0]$. The question then is, for which λ_n's does the $\phi(a)$ of (∗∗) belong to $[T_0]$? That is, when is it T_0-continuous? It is obvious that this means that $n!\lambda_n$ is bounded, and as (∗∗) implies that $n!\lambda_n = \lambda_1 + 1!p_1 + \cdots + (n-1)!p_{n-1}$, it means that only a finite number of the p_m's are $\neq 0$. Thus $n!\lambda_n$ is ultimately constant, say λ, and we have

$$(\S) \qquad\qquad\qquad \phi(a) = e^{2\pi\lambda a i} \qquad\qquad\qquad (\lambda \text{ real}).\dagger$$

EXAMPLE 3. Take the same $\mathfrak{G} = \mathfrak{G}_{rat}$, but take its p-adic topology $T = T_p$ ($p = 2, 3, 5, \cdots$ a prime number; distance is then 2^{N_0}, where N_0 is the minimal exponent $N = 0, \pm 1, \pm 2, \cdots$ for which the least denominator of $p^N(a-b)$ is not divisible by p) and consider $S = [T_p]$. The question is, for which λ_n's is the $\phi(a)$ of (∗∗) T_p-continuous? In T_p, $p^\nu/n! \to 0$ as $\nu \to \infty$ ($n = 1, 2, \cdots$, but fixed), so that $\exp(2\pi\lambda_n p^\nu i) \to 1$, $\lambda_n p^\nu \to 0$ (mod 1), which implies, of course, that there is a $\nu = \nu_n$ for which $\lambda_n p^\nu$ is an integer. This can be expressed in the following manner: there is a ν for which $\lambda_1 p^\nu$ is an integer, and p_n in (∗∗) is divisible by the greatest divisor of $n+1$ which is prime to p. On the other hand, it is not difficult to see that this condition is sufficient.

EXAMPLE 4. Let $\mathfrak{G} = \mathfrak{G}_{real}$ be the set of all real numbers with addition as the rule of composition. Again we first determine all a.p. functions, that is, all irreducible unitary representations. This again means solving (∗), but now with a and b running over all real numbers. Equation (∗) can be solved by the following procedure:

Choose a rational linear basis of the set of real numbers, that is, a set B such that for every real number a the equation $a = \alpha_1\xi_1 + \cdots + \alpha_n\xi_n$ ($n = 1, 2, \cdots; \alpha_1, \cdots, \alpha_n$ rational numbers, all $\neq 0$; $\xi_1, \cdots, \xi_n$ different elements of B) has exactly one solution [12, pp. 459–462]. For every ξ of B we can define the quantity

$$\Gamma^{(a)}(\xi) = \begin{cases} \alpha_m & \text{if } \xi = \text{some } \xi_m, \\ 0 & \text{if } \xi \neq \text{each } \xi_m, \end{cases} \qquad m = 1, \cdots, n;$$

then $\Gamma^{(a)}(\xi)$ is always rational, we have

$$a = \sum_{\xi \text{ in } B} \Gamma^{(a)}(\xi)\xi,$$

where only a finite number of terms are $\neq 0$, and thus $\Gamma^{(a+b)}(\xi) = \Gamma^{(a)}(\xi) + \Gamma^{(b)}(\xi)$. From this it follows at once that every solution $\phi(a)$ of (∗) for real a's is of the form

---

† Thus there exist discontinuous a.p. functions of a rational variable. This fact was proved by Ursell [28, Second Note].

(†) $$\phi(a) = \prod_{\xi \text{ in } B} \phi_\xi(\Gamma^{(a)}(\xi)),$$

where each $\phi_\xi(c)$ is a solution of (*) for rational c's, and thus only a finite number of factors are $\neq 1$. Conversely, it is obvious that every $\phi(a)$ in (†) is a solution. Therefore the general solution is given by (†) if, for every ξ of B, we choose a $\phi_\xi(c)$ from (**) and $\binom{**}{**}$ with $\lambda_n = \lambda_{\xi,n}$ and $p_n = p_{\xi,n}$ dependent on ξ.

EXAMPLE 5. Take the same $\mathfrak{G} = \mathfrak{G}_{\text{real}}$, but consider the set of all Lebesgue-measurable functions, $S = S_m$, which is obviously a cl.f. The question is, which functions $\phi(a)$ of (†) are Lebesgue-measurable? As they are solutions of $\phi(a)\phi(b) = \phi(a+b)$ and $|\phi(a)| = 1$, we can infer from their measurability that they must be of the form

(§) $$\phi(a) = e^{2\pi\lambda a i}$$ (λ real).‡

EXAMPLE 6. Take the same $\mathfrak{G} = \mathfrak{G}_{\text{real}}$, but take its normal topology $T = T_0$ (distance $|a-b|$) and consider $S = [T_0]$. The question is, which functions $\phi(a)$ of (†) are T_0-continuous? As every T_0-continuous function is measurable, all such functions must be of the form (§); and as all functions (§) are continuous, this again gives the general solution.‖

EXAMPLE 7. Take the same $\mathfrak{G} = \mathfrak{G}_{\text{real}}$, but in it take a new topology $T = T(\lambda_1, \cdots, \lambda_k)$, where the only relation $n_1\lambda_1 + \cdots + n_k\lambda_k = 0$, with $n_1, \cdots, n_k = 0, \pm 1, \pm 2, \cdots$, shall be $n_1 = \cdots = n_k = 0$; distance is defined by¶

$$\left[\left| e^{2\pi\lambda_1 a i} - e^{2\pi\lambda_1 b i} \right|^2 + \cdots + \left| e^{2\pi\lambda_k a i} - e^{2\pi\lambda_k b i} \right|^2\right]^{1/2}$$
$$= 2\left[\sin^2 \pi\lambda_1(a-b) + \cdots + \sin^2 \pi\lambda_k(a-b)\right]^{1/2}.$$

Hence the condition $a_n \to a$ in $T(\lambda_1, \cdots, \lambda_k)$ as $n \to \infty$ means that $a_n \to a$ with

---

‡ This is analogous to a result of Fréchet [9] who discussed $f(a)+f(b)=f(a+b)$. Cf. also Sierpinski [23] and Banach [1]. The simplest way to prove our statement is this:

Put $\psi_\epsilon(a) = \int_a^{a+\epsilon} \phi(x)dx$. Then $\psi_\epsilon(a)$ is continuous in a and satisfies $\psi_\epsilon(a)\phi(b) = \psi_\epsilon(a+b)$. If we had $\psi_\epsilon(a) = 0$ for every ϵ, then, as $(\partial/\partial\epsilon)\psi_\epsilon(a)$ is equal to $\phi(a+\epsilon)$ except over a set of measure zero, it would lead to a function $\phi(a+\epsilon) = 0$, except over a set of measure zero, which contradicts the condition $|\phi(a+\epsilon)| = 1$. Thus we can find ϵ_0 and a_0 such that $\psi_{\epsilon_0}(a_0) \neq 0$, and then our equation shows that $\phi(b) = \psi_{\epsilon_0}(a_0+b)/\psi_{\epsilon_0}(a_0)$, that is, $\phi(b)$ is continuous. Then $\psi_{\epsilon_0}(a)$ is differentiable, so that (by our last equation) $\phi(b)$ is also. Now we differentiate $\phi(a)\phi(b) = \phi(a+b)$ and get $\phi'(a)\phi(b) = \phi'(a+b)$, that is, $\phi'(b) = \beta\phi(b)$ when $a = 0$. This means that $\phi(a) = \alpha e^{\beta a}$, and our original conditions make $\alpha = 1$, $\beta = 2\pi\lambda i$, λ real.

‖ Thus there exist discontinuous a.p. functions of a real variable, but they are all non-measurable. These facts have also been proved by Ursell [28, First Note].

¶ For $k = 1$ this is not only a new topology in $\mathfrak{G}_{\text{real}}$, but this also implies an identification of elements congruent mod $1/\lambda_1$. After this identification it is the normal topology. For $k > 1$ it implies no identifications, but it is a new topology.

respect to mod $1/\lambda_1, \cdots,$ and mod $1/\lambda_k$ simultaneously. Therefore $\mathfrak{G}_{\text{real}}$ is compact when metrically completed in this topology and every uniformly $T(\lambda_1, \cdots, \lambda_k)$-continuous function is a.p. (cf. Theorem 36). The question is, which functions $\phi(a)$ of (†) are uniformly $T(\lambda_1, \cdots, \lambda_k)$-continuous? As $T(\lambda_1, \cdots, \lambda_k)$-continuity implies T_0-continuity, they must have the form (§), that is, $\phi(a) = e^{2\pi\lambda a i}$. Now the condition $a_n \to a$ with respect to mod $1/\lambda_1, \cdots,$ and mod $1/\lambda_k$ should imply that $\phi(a_n) \to \phi(a)$ so that $e^{2\pi\lambda a_n i} \to e^{2\pi\lambda a i}$. This is the case if and only if

$$\lambda = n_1\lambda_1 + \cdots + n_k\lambda_k, \qquad n_1, \cdots, n_k = 0, \pm 1, \pm 2, \cdots .^*$$

EXAMPLE 8. Let $\mathfrak{G}$ be a semi-simple Lie group.† The determination of all a.p. functions again means the determination of all irreducible unitary representations (which now of course need not be of degree 1). But such a representation is always continuous in the normal topology T_0 of $\mathfrak{G}$.‡ Therefore all a.p. functions in this $\mathfrak{G}$ are automatically T_0-continuous, in contrast with Examples 1, 2, and 4, 6 (cf. footnotes † and ‖ on pages 478 and 479 respectively). Thus there is no need to discuss $S = [T_0]$ separately.

16. Examples 1–8 sufficiently illustrate the various possibilities of combining a.p. functions with topology to make further comment unnecessary. We shall now investigate another phenomenon.

THEOREM 33. *Let $\mathfrak{G}$ be a group and S a cl.f. of functions in it. The following conditions on two elements a and b of $\mathfrak{G}$ are equivalent*:

A. $D(a; \mathfrak{C}) = D(b; \mathfrak{C})$ (*that is*, $D_{\rho\sigma}(a; \mathfrak{C}) = D_{\rho\sigma}(b; \mathfrak{C})$) *for every $\mathfrak{C}$ of C for which the elements $D_{\rho\sigma}(\mathfrak{C})$ belong to S.*
B. $D(a) = D(b)$ (*that is*, $D_{\rho\sigma}(a) = D_{\rho\sigma}(b)$) *for every normal representation for which the elements $D_{\rho\sigma}$ belong to S.*
C. $f(a) = f(b)$ *for every a.p. function in S.*

A is a special case of B, so that B implies A.

As every $D_{\rho\sigma}(x)$ is a.p. (Theorem 19 and Definition 10), B is a special case of C so that C implies B.

Finally, Theorems 30 and 31 show that A implies C.

Our three statements together prove the equivalence of A, B, and C.

DEFINITION 15. *We call two elements a and b of $\mathfrak{G}$ which satisfy one of the equivalent conditions of Theorem 33 S-coherent (if S is the set of all functions, we abbreviate this to coherent). We denote the set of those elements which are S-coherent (coherent) with 1 by $\mathfrak{G}^S$ ($\mathfrak{G}_0$).*

---

* That is, we are led to the Bohl-Esclangon [3] quasi-periodic functions with the basis $\lambda_1, \cdots, \lambda_n$. Cf. also H. Bohr [4, II, pp. 111–117].

† For a detailed discussion of this notion cf. E. Cartan [7].

‡ This is a most remarkable difference between the behavior of Abelian Lie groups (cf. Example 4) and semi-simple Lie groups. It was discovered by B. L. van der Waerden [29, p. 785].

THEOREM 34. $\mathfrak{G}^S$ *is an invariant subgroup of* $\mathfrak{G}$, *and if* $S = [T]$ *for a topology* T *of* $\mathfrak{G}$, *then* $\mathfrak{G}^S$ *is* T-*closed. Those elements of* $\mathfrak{G}$ *which are coherent with a given* a *form the coset of* $\mathfrak{G}_0^S$ *in* $\mathfrak{G}$ *belonging to* a.

Consider the condition B in Theorem 33 (either A or C could also be used). If a and b belong to $\mathfrak{G}_0^S$ we have $D(ab) = D(a)D(b) = 1$, $D(a^{-1}) = D(a)^{-1} = 1$, that is, a^{-1} and ab belong to it; if only a belongs to $\mathfrak{G}_0^S$, we have $D(b^{-1}ab) = D(b)^{-1}D(a)D(b) = D(b)^{-1}D(b) = 1$, that is, $b^{-1}ab$ also belongs to it. If $S = [T]$, every $D(a)$ is T-continuous, each set $D(a) = 1$ is T-closed, and so their common part $\mathfrak{G}_0^S$ is also. That a and b are coherent means that we always have $D(a) = D(b)$, $D(a^{-1}b) = D(a)^{-1}D(b) = 1$, that is, that $a^{-1}b$ belongs to $\mathfrak{G}_0^S$. Hence the elements b form exactly the coset of $\mathfrak{G}_0^S$ in $\mathfrak{G}$ belonging to a.

DEFINITION 16. *If* $\mathfrak{G}_0^S = 1$ ($\mathfrak{G}_0 = 1$) *we call* $\mathfrak{G}$ *and* S ($\mathfrak{G}$) *maximally a.p.; if* $\mathfrak{G}_0^S = \mathfrak{G}$ ($\mathfrak{G}_0 = \mathfrak{G}$) *we call* $\mathfrak{G}$ *and* S ($\mathfrak{G}$) *minimally a.p.*

These two cases are indeed the two extremes which can occur. If $\mathfrak{G}$ and S are minimally a.p., then for every a.p. $f(x)$ of S we always have $f(a) = f(1)$, that is, the constants are the only a.p. functions in S. And for every normal representation $D(a)$ with the elements $D_{\rho\sigma}(a)$ in S, it must be $D(a) = D(1) = 1$, so that if $D(a)$ is irreducible its degree must be 1. If, on the other hand, $\mathfrak{G}$ and S are maximally a.p., then there exists, for every pair a and b in $\mathfrak{G}$, $a \neq b$, a $\mathfrak{C}$ from C such that all $D_{\rho\sigma}(\mathfrak{C})$ are in S with $D(a; \mathfrak{C}) \neq D(b; \mathfrak{C})$, and an a.p. function $f(x)$ in S such that $f(a) \neq f(b)$. Even more is true:

THEOREM 35. *If* $f(x)g(x)$ *is in* S *whenever* $f(x)$ *and* $g(x)$ *are in* S, *and if* $\mathfrak{G}$ *and* S *are maximally a.p., then, for any finite set* $a_1, \cdots, a_n$ *of distinct elements of* $\mathfrak{G}$ *and any set of complex numbers* $\alpha_1, \cdots, \alpha_n$, *an a.p. function* $f(x)$ *exists in* S *with the prescribed values* $f(a_1) = \alpha_1, \cdots, f(a_n) = \alpha_n$.

If $a \neq b$, there is an a.p. function $g(x)$ in S such that $g(a) \neq g(b)$, so that

$$h(x) = \frac{g(x) - g(b)}{g(a) - g(b)}$$

is an a.p. function in S with $h(a) = 1$ and $h(b) = 0$. For every pair a and b ($a \neq b$), choose such a function $h(x)$ and denote it by $h(a,b;x)$. Then

$$f(x) = \sum_{\nu=1}^{n} \alpha_\nu \prod_{\mu=1, \mu \neq \nu}^{n} h(a_\nu, b_\mu; x)$$

has all the properties required.

There are also some other ways to characterize $\mathfrak{G}_0^S$, but we shall not discuss them here.

17. If $\mathfrak{G}$ and S are maximally a.p., we can introduce a topology by means of their a.p. functions. In this connection the following notions are of importance:

DEFINITION 17. *If $\mathfrak{G}$ and S are maximally a.p. we define a topology FS in $\mathfrak{G}$ by considering the following "neighborhoods" $\mathfrak{N}(a)$ of an element a of $\mathfrak{G}$†: Choose a finite number of a.p. functions $f_1, \cdots, f_n$ and an $\epsilon > 0$; then $\mathfrak{N}(a) = \mathfrak{N}(a; f_1, \cdots, f_n, \epsilon)$ is the set of all b's such that $|f_1(a) - f_1(b)| < \epsilon, \cdots, |f_n(a) - f_n(b)| < \epsilon$. (If S is the set of all functions we abbreviate this to F.)*

One sees at once that *FS* satisfies Hausdorff's Axioms (cf. first footnote on page 447).

DEFINITION 18. *If two topologies T_1 and T_2 for a set $\mathfrak{S}$ are given, T_1 is called weaker than T_2 if every T_1-neighborhood of an element a of $\mathfrak{S}$ contains a T_2-neighborhood of a.*

Obviously, every set which is closed or open in the T_1-sense, and every function which is continuous in the T_1-sense, has the same property in the T_2-sense. Thus for a group $\mathfrak{S} = \mathfrak{G}$, $[T_1]$ is a subset of $[T_2]$ (cf. Definition 14). On the other hand, it is obvious that if S_1 and S_2 are cl.f. and S_2 is a subset of S_1, then FS_1 is weaker than FS_2.

We intend to go more deeply into the theory of $[T]$ and *FS* on another occasion. At present let us merely remark that for every $\mathfrak{G}$ (even for a non-topologically given $\mathfrak{G}$) F is a topology determined by $\mathfrak{G}$ alone (if $\mathfrak{G}$ is maximally a.p.). Discussion of Examples 1 and 4 shows without much difficulty that $\mathfrak{G}_{rat}$ and $\mathfrak{G}_{real}$ are maximally a.p. (even with their cl.f. $[T_0]$ or $[T_p]$ and $[T_0]$ or $[T(\lambda_1, \cdots, \lambda_k)]$ respectively (for $k > 1$, cf. footnote* on page 480)) and that the F's are very "strong"; the condition $a_n \to a$ in F as $n \to \infty$ means that all a_n's, with a finite number of exceptions, are equal to a. On the other hand, Example 8 shows that, for a semi-simple Lie group $\mathfrak{G}$ (if it is maximally a.p.), $F = F[T_0]$. Theorem 36 shows that if $\mathfrak{G}$ is compact in T_0, $\mathfrak{G}$ and $[T_0]$ are maximally a.p. and $F[T_0] = T_0$. Thus, if $\mathfrak{G}$ is a semi-simple and compact Lie group, it is maximally a.p. and $F = T_0$.

18. The case where a group $\mathfrak{G}$ and a topology T have the properties that $\mathfrak{G}$ and $[T]$ are maximally a.p. and $F[T] = T$ is of particular importance.

THEOREM 36. *$\mathfrak{G}$ and $[T]$ are maximally a.p. and $F[T] = T$ in each of the two following cases (in case B, $F[T] = T$ should be understood to mean only that the condition $a_n \to a$ as $n \to \infty$ is equivalent to it in both senses):*

A. *$\mathfrak{G}$ is compact in T.*

---

† See first footnote on page 447.

B. $\mathfrak{G}$ *is locally compact† and separable† in* T, $\mathfrak{G}$ *is an Abelian group, and ab and* a^{-1} *are* T-*continuous in a and b, and a respectively.‡*

If $\mathfrak{G}$ is compact in T, every continuous $f(x)$ is a.p.: for $\mathfrak{G}$ being compact, $f(x)$ is uniformly continuous; if any sequence a_1, a_2, $\cdots$ is given, we can extract from it a subsequence a_{n_1}, a_{n_2}, $\cdots$ which converges to a limit a, and then we have the result that $f(xa_n) \to f(xa)$ and $f(a_{n_\nu}x) \to f(ax)$ uniformly as $\nu \to \infty$. Thus $[T]$ consists only of a.p. functions.

Now it is possible to define a distance $D(a, b)$ in $\mathfrak{G}$ which is equivalent to the topology T [27, 25]. $f(x) = D(a, x)$ belongs to $[T]$ and we have the result that $f(a) = 0 \neq f(b)$, proving that $\mathfrak{G}$ and $[T]$ are maximally a.p.; and the neighborhood $\mathfrak{N}(a; f, \epsilon)$ (cf. Definition 17) is the sphere with the center a and the radius ϵ, proving that T is weaker than $F[T]$. But $F[T]$ is obviously weaker than T, and therefore $F[T] = T$. Thus A is proved.

The proof of B will be given at the end of Part V.

Minimally a.p. groups likewise exist, for example, the group $\mathfrak{g}_{(n)}$ of all linear transformations of determinant 1 in the real euclidean space of n dimensions, $n = 2, 3, \cdots$. As it is a semi-simple Lie group, indeed even simple [6], all its bounded representations are continuous [29]; as it is a linear group, their continuity implies their differentiability [14, p. 37]. Hence we need only determine those irreducible representations of $\mathfrak{g}_{(n)}$ which arise from "infinitesimal representations," and see if there exist any bounded ones among them. Now these representations and their traces (characteristics) are known [70; 30, pp. 287, 300], and only the identity, $D(a) \equiv 1$, has a bounded trace, so no other representation can be bounded. Application of Theorem 33, criterion A, shows that $\mathfrak{g}_{(n)}$ is minimally a.p.

The group $\mathfrak{g}'$ of all transformations $y = ax + b$ (a and b real, $a > 0$) is neither minimally nor maximally a.p., as a simple discussion shows.

V. ABELIAN GROUPS

19. We assume throughout Part V that the assumptions of Theorem 36, case B (which we shall finally prove), hold; thus we assume that a group $\mathfrak{G}$ and a topology T are given, that $\mathfrak{G}$ is locally compact and separable in T and Abelian, and that ab, a^{-1} are T-continuous.

Under the above topological assumptions, A. Haar has shown the existence of a right-invariant Lebesgue integral [11, pp. 166–167]. Thus it is possible to define for complex-valued functions $f(x)$ defined in $\mathfrak{G}$ (i) a notion of measura-

---

† "Locally compact" means that each element a has a conditionally compact neighborhood [13, p. 107]; as a group $\mathfrak{G}$ is homogeneous it is sufficient to postulate this for the element 1. "Separable" means that there exists a countably infinite "equivalent system of neighborhoods"; if the topology is originated by a distance notion, one may postulate the existence of a countably infinite everywhere dense subset [13, p. 125, and p. 229, Axiom 10].

‡ See first footnote on page 447.

bility, (ii) a notion of summability, (iii) an integral $\int_{\mathfrak{G}} f(x)dx$. On the basis of (i) and (ii), moreover, it is possible to do this in such a manner that (i)–(iii) have all the formal properties of these notions as in the usual Lebesgue theory, and besides are invariant under the substitution of $f(xa)$ for $f(x)$.

We now consider all measurable functions $f(x)$ in $\mathfrak{G}$ for which $|f(x)|^2$ is summable, that is, $\int_{\mathfrak{G}} |f(x)|^2 dx$ is finite. These functions form a Hilbert space $\mathfrak{H}_{\mathfrak{G}}$ if we define the inner product (f, g) to be $\int_{\mathfrak{G}} f(x)\overline{g(x)}dx$†, provided that $\mathfrak{G}$ is infinite, which we will assume to be the case. (If it is finite, it is compact and falls under case A.) In $\mathfrak{H}_{\mathfrak{G}}$,

(#) $$O_a f(x) = f(xa)$$

defines a linear and unitary operation (that is, an operation which leaves (f, g) invariant), and it follows that

(##) $$O_a O_b = O_{ab}.$$

Now we use the Abelian character of $\mathfrak{G}$, by virtue of which (##) implies that O_a and O_b commute. As O_a is unitary, its adjoint† is $O_a^* = O_a^{-1}$ and, by (##), $O_a^* = O_{a^{-1}}$. Thus every O_b commutes with every O_a and O_a^*, and the set of all operators O_a has been called Abelian [16, p. 389]. Therefore a theorem proved by the author applies to this set: there exists a bounded Hermitian operator R such that every O_a is a function of R,

($\overset{\#}{\underset{\#}{\#}}$) $$O_a = \phi_a(R),$$

where $\phi_a(\lambda)$ is a complex-valued function of the variable λ.§ That the functions $\phi_a(\lambda)$ can be used for the discussion of the group $\mathfrak{G}$ has been noted by Haar and successfully applied to countably infinite Abelian groups [10, p. 131]; cf. also Wiener and Paley [33]. Theorem 37 will be an application of this idea in the full generality allowed by Haar's right-invariant Lebesgue integral. It must be remarked, however, that Haar's method of discussing countably infinite Abelian groups has been considerably simplified by Wiener and Paley [33], but that their simplification seems not to apply to our general case, and that we have to use Haar's original method.

THEOREM 37. *If $\mathfrak{G}$ and T fulfill the assumptions formulated at the beginning of this part (that is, if $\mathfrak{G}$ is locally compact, separable, and Abelian), there exists a function in two variables $\phi(a, \lambda)$ (a in $\mathfrak{G}$, λ real) with the following properties:*

---

† For the modern theory of Hilbert space cf. J. v. Neumann [15, pp. 63–70, 108–111]. Cf. further M. H. Stone [24, pp. 1–32].

§ The notion of a function of an operator is due originally to F. Riesz. More general forms have been given to it by J. v. Neumann [17, pp. 202–213] and M. H. Stone [24, pp. 221–241]. The theorem in question has been proved by J. v. Neumann [17, p. 214].

$\phi(a, \lambda)$ *is a Baire function in* (a, λ),* *and there exists a "resolution of the iden-* *tity"* $E(\lambda)$ *such that*

$$(\ddagger) \qquad\qquad (O_a f, g) = \int_{-\infty}^{\infty} \phi(a, \lambda)\, d(E(\lambda)f, g)\dagger$$

identically in a *and* $f(x)$ *and* $g(x)$.

For the function $\phi(a, \lambda) = \phi_a(\lambda)$ in ($\#\#$), the theorem mentioned in the footnote§ on page 484 leads to all our statements except for the Baire character of $\phi(a, \lambda)$ in (a, λ) (it would show the Baire character in λ, but we need it also in a).

We know that finite linear aggregates of functions $f_{\mathfrak{O}}$ where $\mathfrak{O}$ is a conditionally compact open set, therefore having finite measure, and

$$f_{\mathfrak{O}}(a) = \begin{cases} 1 \text{ for } a \text{ in } \mathfrak{O}, \\ 0 \text{ elsewhere}, \end{cases}$$

are everywhere dense in our functional space [15, p. 110]. From now on "everywhere dense" will be interpreted in the sense of the distance

$$D(f, g) = \|f - g\| = [(f - g, f - g)]^{1/2} = \left[\int_{\mathfrak{G}} |f(x) - g(x)|^2 dx \right]^{1/2}$$

but not in the sense of the distance l.u.b.$_x |f(x) - g(x)|$. Now, if $\mathfrak{O}_1$ is an open set the closure of which is part of $\mathfrak{O}$, we can find a continuous function§

$$f_{\mathfrak{O},\mathfrak{O}_1}(a) \begin{cases} = 1 \text{ for } a \text{ in } \mathfrak{O}, \\ = 0 \text{ for } a \text{ not in } \mathfrak{O}, \\ \geqq 0 \text{ and } \leqq 1 \text{ elsewhere}. \end{cases}$$

If we let $\mathfrak{O}_1$ converge to $\mathfrak{O}$, then $f_{\mathfrak{O},\mathfrak{O}_1}(a)$ converges everywhere to, and is majorized by, $f_{\mathfrak{O}}(a)$, so that $f_{\mathfrak{O}}(a)$ is its limit in the sense of the distance $\|f - g\|$. Therefore continuous functions f which are $\neq 0$ only in conditionally compact sets are everywhere dense in our functional space. Since $a_n \to a$ implies $xa_n \to xa$, we have $f(xa_n) \to f(xa)$ for these functions, and, by the second property,

$$\|O_{a_n} f - O_a f\| = \left[\int_{\mathfrak{G}} |f(xa_n) - f(xa)|^2 dx \right]^{1/2} \to 0.$$

Hence $a_n \to a$ implies $O_{a_n} f \to O_a f$ for an everywhere dense set of f's, but as all

---

* That is, it can be obtained from continuous functions in (a, λ) by successive limiting processes wherein the limit is always taken of everywhere convergent sequences.

† $\int_{-\infty}^{\infty}$ is a Lebesgue-Stieltjes integral over λ. For an explanation of the terminology used, see [15, p. 92] or [24, p. 174].

§ This is a problem of Fréchet, first solved by Hahn. Cf. [26, Anhang III, p. 290].

O_b's are unitary operators, and therefore uniformly continuous in f, the implication holds for every f. Consequently $O_a f$ is a continuous function in a for every fixed f.

A simple computation shows, after substituting $E(\mu)g$ in ($\ddagger$), at the end of Theorem 37 in place of g [cf. 17, p. 206],

$$(O_a f, E(\mu)g) = \int_{-\infty}^{\mu} \phi(a, \lambda) d(E(\lambda)f, g).$$

Now choose a complete normalized orthogonal system $f_1, f_2, \cdots$, put $f = g = f_n$, $n = 1, 2, \cdots$, multiply by 2^{-n}, and add. The infinite series thus obtained in the left- and right-hand members converge uniformly since $(O_a f, E(\mu)g)$ and $(E(\lambda)f, g)$ are both $\leq \|f\| \|g\|$ in absolute value. The result is

$$\sum_{n=1}^{\infty} 2^{-n}(O_a f_n, E(\mu)f_n) = \int_{-\infty}^{\mu} \phi(a, \lambda) d\left[\sum_{n=1}^{\infty} 2^{-n}(E(\lambda)f_n, f_n) \right].$$

$(O_a f_n, E(\mu)f_n)$ is continuous in a and continuous on the right in μ, $(E(\lambda)f_n, f_n)$ is continuous on the right in λ and monotonically increasing, and the same properties hold for the uniformly convergent sums

$$F(a, \mu) = \sum_{n=1}^{\infty} 2^{-n}(O_a f_n, E(\mu)f_n), \qquad G(\lambda) = \sum_{n=1}^{\infty} 2^{-n}(E(\lambda)f_n, f_n).$$

Thus $F(a, \mu)$ and $G(\lambda)$ are Baire functions, the latter is monotonically increasing, and

$$F(a, \mu) = \int_{-\infty}^{\mu} \phi(a, \lambda) dG(\lambda).$$

If we consider $G(\lambda)$ as the variable (instead of λ), then the well known theorem on the differentiability of integrals shows that†

$$\lim_{\delta, \epsilon \to 0^+} \frac{F(a, \mu + \delta) - F(a, \mu - \epsilon)}{G(\mu + \delta) - G(\mu - \epsilon)}$$

exists and equals $\phi(a, \mu)$ except, however, for a set of μ's dependent on a whose $\xi = G(\mu)$-image§ is a set (of real numbers) of Lebesgue measure zero. Now the function

$$\phi_1(a, \lambda) = \begin{cases} \displaystyle\lim_{\delta, \epsilon \to 0^+} \frac{F(a, \mu + \delta) - F(a, \mu - \epsilon)}{G(\mu + \delta) - G(\mu - \epsilon)} & \text{when this limit exists,} \\ 0 & \text{otherwise} \end{cases}$$

---

† Cf. [5, pp. 544–545]. Analogous results concerning "central derivatives" of F with respect to G are due to Daniell [8].

§ If $G(\mu)$ is discontinuous at $\mu = \mu_0$, the image of $\mu = \mu_0$ is supposed to be the whole jump-interval $G(\mu_0 - 0) \leq \xi \leq G(\mu_0 + 0)$.

is obviously a Baire function, and the set Σ of λ's for which $\phi_1(a, \lambda) \neq \phi(a, \lambda)$ has a $\xi = G(\mu)$-image of Lebesgue measure zero. Since $2^n G(\mu) - (E(\mu)f_n, f_n)$ is monotonically increasing, the $\xi = (E(\mu)f_n, f_n)$-image of Σ is also of Lebesgue measure† zero, and therefore every $\xi = (E(\mu)f, f)$-image is of Lebesgue measure zero [17, p. 213, the last remark of Part II].

Hence we may replace $\phi(a, \lambda)$ in (‡) by $\phi_1(a, \lambda)$, and this will not affect the validity of (‡) for $f = g$; now if we replace our f by $(f+g)/2$ and $(f-g)/2$ and subtract, we get the real part of the general (‡); if we replace f and g by if and g, we get its imaginary part, and prove it altogether. Thus $\phi_1(a, \lambda)$ meets all our requirements.

THEOREM 38. *Under the assumptions of Theorem 37, $\phi(a, \lambda)$ can even be chosen as a continuous function in a satisfying the equations*

$$\phi(ab, \lambda) = \phi(a, \lambda)\,\phi(b, \lambda), \qquad |\phi(a, \lambda)| = 1.$$

A simple computation shows [17, p. 206] that

$$(O_a O_b f, g) = \int_{-\infty}^{+\infty} \phi(a, \lambda)\,\phi(b, \lambda)\,d(E(\lambda)f, g),$$

$$(O_a^* O_a f, g) = \int_{-\infty}^{+\infty} |\phi(a, \lambda)|^2\,d(E(\lambda)f, g);$$

on the other hand,

$$(O_{ab}f, g) = \int_{-\infty}^{+\infty} \phi(ab, \lambda)\,d(E(\lambda)f, g), \qquad (f, g) = \int_{-\infty}^{+\infty} d(E(\lambda)f, g).$$

Now $O_a O_b = O_{ab}$, $O_a^* O_a = 1$; therefore the right sides of our equations are equal. An analogous computation shows that if we substitute $E(\mu)g$ for g [17, p. 206] and subtract, we get

$$\int_{-\infty}^{\mu} (\phi(ab, \lambda) - \phi(a, \lambda)\,\phi(b, \lambda))\,d(E(\lambda)f, g) = 0,$$

$$\int_{-\infty}^{\mu} (|\phi(a, \lambda)|^2 - 1)\,d(E(\lambda)f, g) = 0.$$

Putting $f = g$ shows that the equations of our Theorem hold except for a set of λ's (depending on the pair a and b and on a respectively), the $\xi = (E(\lambda)f, f)$-image of which has Lebesgue measure zero† (this condition holds for all f's simultaneously). Returning to the complete normalized orthogonal

---

† The Lebesgue measure of the $\xi = H(\mu)$-image of a set $\mathfrak{S}$ is $\int_{\mathfrak{S}} dH(\mu)$, and therefore, if it is 0 for $H(\mu)$, it will be 0 for every other function $K(\mu)$ for which $H(\mu) - K(\mu)$ is monotonically increasing [cf. 17, p. 198, rule d, and p. 199].

system $f_1, f_2, \cdots$ in the proof of Theorem 37, and to the corresponding

$$G(\lambda) = \sum_{n=1}^{\infty} 2^{-n}(E(\lambda)f_n, f_n),$$

we see that also the $\xi = G(\lambda)$-images have Lebesgue measure zero (this follows for each $\sum_{n=1}^{N} 2^{-n}(E(\lambda)f_n, f_n)$ from the integral formula in the footnote on page 487, and for

$$G(\lambda) = \sum_{n=1}^{\infty} 2^{-n} (E(\lambda)f_n, f_n)$$

from the fact that the difference is monotonically increasing, ≥ 0, and $\leq \sum_{n=-N+1}^{\infty} 2^{-n} = 1/2^N$.

For a-sets and b-sets (that is, subsets of $\mathfrak{G}$) we have a measure, namely, the Haar-Lebesgue measure. For λ-sets (that is, sets of real numbers) we shall consider the Lebesgue measure of the $\xi = G(\lambda)$-image (cf. the footnote§ on page 486), and call it the λ-measure. All these measures have the formal properties of the Lebesgue measure.* We can, by the analogue of the process which leads from linear to plane measure [cf. 18, p. 588] use these measures to define measures with similar properties for (a, b)-sets, (a, λ)-sets, and (a, b, λ)-sets. If we use these defining processes, the theorem of Fubini holds for all combinations of the variables a, b, λ because its proof [5, pp. 622–628] applies unchanged.

As we are dealing with Baire functions, the (a, b, λ)- and (a, λ)-sets for which $\phi(ab, \lambda) \neq \phi(a, \lambda)\phi(b, \lambda)$ and $|\phi(a, \lambda)| \neq 1$ are Borel sets and therefore measurable. Hence Fubini's theorem can be applied to them; as for fixed a, b and a respectively they give λ-sets of zero λ-measure, they are sets of zero (a, b, λ)-measure and (a, λ)-measure themselves. This again implies that if λ does not belong to a certain (fixed) set $\mathfrak{S}_1$ of zero λ-measure, and if a does not belong to a certain set $\mathfrak{S}_2^{(\lambda)}$ (depending on λ) of zero (Haar) measure, then we have the result that, if b does not belong to a certain set $\mathfrak{S}_3^{(\lambda, a)}$ (depending on λ and a) of zero (Haar) measure, then $\phi(ab, \lambda) = \phi(a, \lambda)\phi(b, \lambda)$, and, at any rate, $|\phi(a, \lambda)| = 1$.

Now choose a conditionally compact open set $\mathfrak{O}$. If we had, for a certain λ,

$$\int_{\mathfrak{O}} \phi(x, \lambda)dx = 0$$

for every $\mathfrak{O}$, this would imply that

$$\int_{\mathfrak{M}} \phi(x, \lambda)dx = 0$$

---

* By this we mean that they satisfy Carathéodory's postulates I–V [5, pp. 238–239, 258].

for every measurable set $\mathfrak{M}$, and thus $\phi(x, \lambda) = 0$ except for an x-set of λ-measure zero. This contradicts the fact that we have $|\phi(x, \lambda)| = 1$ except for an x-set of λ-measure zero. Therefore choose $\mathfrak{O}$ such that

$$\int_{\mathfrak{O}} \phi(x, \lambda) dx \neq 0,$$

and denote the set of all (ax)'s (x in $\mathfrak{O}$) by $a\mathfrak{O}$.

Assume λ in $\mathfrak{S}_1$ and a in $\mathfrak{S}_2^{(\lambda)}$. Then we have $\phi(ax, \lambda) = \phi(a, \lambda)\phi(x, \lambda)$ if x is not in $\mathfrak{S}_3^{(\lambda, a)}$ and this implies that

$$\phi(a, \lambda) \int_{\mathfrak{O}} \phi(x, \lambda) dx = \int_{\mathfrak{O}} \phi(ax, \lambda) dx = \int_{a\mathfrak{O}} \phi(y, \lambda) dy,$$

$$\phi(a, \lambda) = \frac{\int_{a\mathfrak{O}} \phi(x, \lambda) dx}{\int_{\mathfrak{O}} \phi(x, \lambda) dx}.$$

Now, by well known theorems on Lebesgue integrals, the numerator is continuous in a, the denominator is constant and $\neq 0$, and for this argument we need not restrict a to $\mathfrak{S}_2^{(\lambda)}$ (λ, of course, is in $\mathfrak{S}_1$). Hence we may define a continuous function $\phi_2(a, \lambda)$ by putting it equal to

$$\frac{\int_{a\mathfrak{O}} \phi(x, \lambda) dx}{\int_{\mathfrak{O}} \phi(x, \lambda) dx}$$

(λ in $\mathfrak{S}_1$); and we have $\phi_2(a, \lambda) = \phi(a, \lambda)$ if a is not in $\mathfrak{S}_2^{(\lambda)}$.

In this case, if b is not in $\mathfrak{S}_2^{(\lambda)}$ and ab is not in $\mathfrak{S}_2^{(\lambda)}$, we have $\phi_2(ab, \lambda) = \phi_2(a, \lambda)\phi_2(b, \lambda)$. But as we except only a b-set of zero (Haar) measure, this holds in an everywhere dense b-set, and thus, for reasons of continuity, for every b. So the above formula is true for every b, and $|\phi_2(a, \lambda)| = 1$ is true if a is not in $\mathfrak{S}_2^{(\lambda)}$. For the same continuity reasons, therefore, there are no a-exceptions at all. Thus (if λ is not in $\mathfrak{S}_1$) $\phi_2(a, \lambda)$ meets the requirements of our theorem if it can replace $\phi(a, \lambda)$.

By definition, $\phi_2(a, \lambda)$ is a Baire function in (a, λ). Hence the (a, λ)-set for which $\phi(a, \lambda) \neq \phi_2(a, \lambda)$ is a Borel set and therefore measurable. Hence Fubini's theorem applies to it, and since for a fixed λ, except in a set of zero λ-measure ($\mathfrak{S}_1$), it gives an a-set of zero (Haar) measure ($\mathfrak{S}_2^{(\lambda)}$), it is a set of zero (a, λ)-measure itself. This again implies that if a does not belong to a certain (fixed) set $\mathfrak{S}_1'$ of zero (Haar) measure, then $\phi(a, \lambda) = \phi_2(a, \lambda)$ provided λ does not belong to a certain set $\mathfrak{S}_2'^{(a)}$ (depending on a) of zero λ-measure. If we change $\phi_2(a, \lambda)$ for the λ's of $\mathfrak{S}_1$ into 1, we obtain its continuity in a and the equations of our Theorem for all λ's without exceptions, and the state-

ment just now proved still holds if we replace $\mathfrak{S}_2'^{(a)}$ by $\mathfrak{S}_2'^{(a)} + \mathfrak{S}_1$ which is also a set of zero λ-measure.

If a does not belong to $\mathfrak{S}_1'$, we have $\phi(a, \lambda) = \phi_2(a, \lambda)$ except for a λ-set with zero λ-measure, that is, with a $\xi = G(\lambda)$-image (cf. the footnote§ on page 486) of zero Lebesgue measure. This proves, as at the end of the proof of Theorem 37, that

$$(O_a f, g) = \int_{-\infty}^{+\infty} \phi_2(a, \lambda) \, d(E(\lambda)f, g)$$

identically in $f(x)$ and $g(x)$ if a does not belong to $\mathfrak{S}_1'$. Since $\mathfrak{S}_1'$ has zero (Haar) measure, the domain of validity in a is everywhere dense. But both sides are continuous functions of a: this was shown for the left side at the beginning of the proof of Theorem 37, and follows for the right side from the continuity of $\phi_2(a, \lambda)$ in a for all λ's. Hence our equation holds for all a's.

Thus $\phi_2(a, \lambda)$ meets all our requirements.

THEOREM 39. *If the assumptions of Theorems 37 and 38 are satisfied, and if ab and a^{-1} are continuous in (a, b) and in a respectively, then the condition $\phi(a, \lambda) = \phi(b, \lambda)$ for all λ's is equivalent to the condition $a = b$, and the condition $\phi(a_n, \lambda) \to \phi(a, \lambda)$ as $n \to \infty$ for all λ's is equivalent to the condition $a_n \to a$ as $n \to \infty$.*

The first statement follows from the second by putting $a_1 = a_2 = \cdots = b$. The necessity of the criterion in the second statement is obvious, as all the functions $\phi(a, \lambda)$ are continuous in a. So the only thing we need to prove is its sufficiency.

Therefore suppose that $\phi(a_n, \lambda) \to \phi(a, \lambda)$ as $n \to \infty$ for all λ's. Then ($\ddagger$) in Theorem 37 shows that $(O_{a_n}f, g) \to (O_a f, g)$ as $n \to \infty$ for any $f(x)$ and $g(x)$ of our functional space. Now let $\mathfrak{O}$ be a conditionally compact open set, and define

$$f_{\mathfrak{O}}(x) = \begin{cases} 1 & \text{for } x \text{ in } \mathfrak{O}, \\ 0 & \text{elsewhere.} \end{cases}$$

Put $f(x) = f_{\mathfrak{O}}(x)$ and $g(x) = f_{\mathfrak{O}}(ax)$. Then

$$(O_{a_n}f, g) = \int_{\mathfrak{G}} f(a_n x)\overline{f(ax)}dx,$$

$$(O_a f, g) = \int_{\mathfrak{G}} |f(ax)|^2 dx > 0,$$

and $(O_{a_n}f, g) \to (O_a f, g)$ implies that if n is sufficiently large, $(O_{a_n}f, g) \neq 0$ and therefore $f(a_n x)f(ax) \not\equiv 0$. Hence there is an x for which $a_n x$ and ax both belong

to $\mathfrak{O}$, and as $a_n a^{-1} = (a_n x)(ax)^{-1}$, aa_n^{-1} can be written in the form uv^{-1}, where u and v both belong to $\mathfrak{O}$. As every open set has conditionally compact subsets, this holds for every open $\mathfrak{O}$.

Now let $\mathfrak{N}$ be a neighborhood of a. Then we can find an open set $\mathfrak{O}$ for which every uv^{-1}, u and v in $\mathfrak{O}$, belongs to $\mathfrak{N}$. (Here is where the extra continuity assumptions are used.) Then our result shows that if n is sufficiently large, a_n belongs to $\mathfrak{N}$. This means that $a_n \to a$ as $n \to \infty$.

Theorems 37 and 38, combined with Theorem 19, show that each $\phi(a, \lambda)$, when considered as an a-function, is a.p. and belongs to $[T]$. Therefore Theorem 39 proves exactly the statements of Theorem 36, case B.

BIBLIOGRAPHY

1. S. Banach, *Sur l'équation fonctionnelle f(x+y)=f(x)+f(y)*, Fundamenta Mathematicae, vol. 1 (1920), pp. 123–124.
2. S. Bochner, *Beiträge zur Theorie der fastperiodischen Funktionen*, Mathematische Annalen, vol. 96 (1927), pp. 119–147.
3. P. Bohl, *Über die Darstellung von Funktionen einer Variabeln durch trigonometrische Reihen mit mehreren einer Variabeln proportionalen Argumenten*, Thesis, Dorpat, 1893.
4. H. Bohr, *Zur Theorie der fastperiodischen Funktionen*. I, Acta Mathematica, vol. 45 (1925), pp. 29–127. II, Ibidem, vol. 46 (1925), pp. 101–214.
5. C. Carathéodory, *Vorlesungen über reelle Funktionen*. 2d edition. Leipzig and Berlin, 1927.
6. E. Cartan, *Sur la structure des groupes de transformations finis et continus*, Thesis, Paris, 1894.
7. E. Cartan, *Les groupes projectifs qui ne laissent invariante aucune multiplicité plane*, Bulletin de la Société Mathématique de France, vol. 41 (1913), pp. 53–96.
8. P. J. Daniell, *Differentiation with respect to a function of bounded variation*, Transactions of the American Mathematical Society, vol. 19 (1918), pp. 353–362.
9. M. Fréchet, *Pri la funkcie ekvacio f(x+y)=f(x)+f(y)*, L'Enseignement Mathématique, vol. 15 (1913), pp. 390–393.
10. A. Haar, *Über unendliche kommutative Gruppen*, Mathematische Zeitschrift, vol. 33 (1931), pp. 129–159.
11. A. Haar, *Der Massbegriff in der Theorie der kontinuirlichen Gruppen*, Annals of Mathematics, (2), vol. 34 (1933), pp. 147–169.
12. G. Hamel, *Eine Basis aller Zahlen und die unstetigen Lösungen der Funktionalgleichung f(x+y)=f(x)+f(y)*, Mathematische Annalen, vol. 60 (1905), pp. 459–462.
13. F. Hausdorff, *Mengenlehre*. 2d edition. Berlin and Leipzig, 1927.
14. J. v. Neumann, *Über die analytischen Eigenschaften von Gruppen linearer Transformationen und ihrer Darstellungen*, Mathematische Zeitschrift, vol. 30 (1929), pp. 3–42.
15. J. v. Neumann, *Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren*, Mathematische Annalen, vol. 102 (1930), pp. 49–131.
16. J. v. Neumann. *Zur Algebra der Funktionaloperationen und Theorie der normalen Operatoren*, Mathematische Annalen, vol. 102 (1930), pp. 370–427.
17. J. v. Neumann, *Über Funktionen von Funktionaloperatoren*, Annals of Mathematics, (2), vol. 32 (1931), pp. 191–226.
18. J. v. Neumann, *Zur Operatorenmethode in der klassischen Mechanik*, Annals of Mathematics, (2), vol. 33 (1932), pp. 587–642.
19. J. v. Neumann, *Zum Haarschen Mass in topologischen Gruppen*, Compositio Mathematica, vol. 1 (1934), pp. 106–114.

20. E. Schmidt, *Zur Theorie der linearen und nichtlinearen Integralgleichungen*, Mathematische Annalen, vol. 63 (1907), pp. 433–476.

21. I. Schur, *Neue Begründung der Theorie der Gruppencharaktere*, Sitzungsberichte der Preussischen Akademie, Phys. Math. Kl., 1905, pp. 406–432.

22. I. Schur, *Neue Anwendungen der Integralrechnung auf Probleme der Invariantentheorie*, Sitzungsberichte der Preussischen Akademie, Phys. Math. Kl., 1924, pp. 183–208.

23. W. Sierpinski, *Sur l'équation fonctionnelle $f(x+y)=f(x)+f(y)$*, Fundamenta Mathematicae, vol. 1 (1920), pp. 125–129.

24. M. H. Stone, *Linear Transformations in Hilbert Space*, American Mathematical Society Colloquium Publications, vol. XV, 1932.

25. A. Tichonoff, *Über einen Metrisationssatz von P. Urysohn*, Mathematische Annalen, vol. 95 (1926), pp. 139–142.

26. P. Urysohn, *Über die Mächtigkeit der zusammenhängenden Mengen*, Mathematische Annalen, vol. 94 (1925), pp. 262–308.

27. P. Urysohn, *Zum Metrisationsproblem*, Mathematische Annalen, vol. 94 (1925), pp. 309–315.

28. H. D. Ursell, *Normality of almost periodic functions*, First Note, Journal of the London Mathematical Society, vol. 4 (1929), pp. 123–127. Second Note, Ibidem, vol. 5 (1930), pp. 47–50.

29. B. L. van der Waerden, *Stetigkeitssätze der halbeinfachen Lieschen Gruppen*, Mathematische Zeitschrift, vol. 36 (1933), pp. 780–786.

30. H. Weyl, *Theorie der Darstellung kontinuirlicher halbeinfacher Gruppen durch lineare Transformationen*. I, Mathematische Zeitschrift, vol. 23 (1925), pp. 271–309.

31. H. Weyl, *Integralgleichungen und fastperiodische Funktionen*, Mathematische Annalen, vol. 97 (1927), pp. 338–356.

32. H. Weyl and F. Peter, *Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuirlichen Gruppe*, Mathematische Annalen, vol. 97 (1927), pp. 737–755.

33. N. Wiener and R. E. A. C. Paley, *Characters of Abelian groups*, Proceedings of the National Academy of Sciences, vol. 19 (1933), pp. 253–257.

INSTITUTE FOR ADVANCED STUDY,
    PRINCETON, N. J.