LECTURES IN THEORETICAL PHYSICS 12. Boulder Colo. 1969.

Volume XII-C

MATHEMATICAL METHODS IN FIELD THEORY AND COMPLEX ANALYTIC VARIETIES

Edited by

KALYANA T. MAHANTHAPPA
and
WESLEY E. BRITTIN

Department of Physics and Astrophysics
University of Colorado

GORDON AND BREACH SCIENCE PUBLISHERS
NEW YORK • LONDON • PARIS

Library of Congress Catalog Card Number: 59-13034

Published by GORDON AND BREACH, Science Publishers, Inc.
440 Park Avenue South, New York, N.Y. 10016

Editorial office for Great Britain:
Gordon and Breach, Science Publishers Ltd.
12 Bloomsbury Way
London W.C.1

Editorial office for France:
Gordon & Breach
7-9 rue Emile Dubois
Paris 14e

Printed in the United States of America

THE LECTURERS

H. J. Borchers
Universität Göttingen, Göttingen, Germany

M. Guenin
University of Geneva, Geneva, Switzerland

R. C. Gunning
Princeton University, Princeton, New Jersey

P. M. Mathews
University of Madras, Madras, India

Eric H. Roffman
State University of New York at Stony Brook, Stony Brook, New York

I. T. Todorov
Institute for Advanced Study, Princeton, New Jersey

Walter Wyss
University of Colorado, Boulder, Colorado

PREFACE

Volume XII-C of the Lectures in Theoretical Physics contains the proceedings of the session on Mathematical Methods in Field Theory and Complex Analytic Varieties which was held simultaneously with the session on High Energy Collisions of Elementary Particles during the second part of the Twelfth Boulder Summer Institute for Theoretical Physics. It contains the text of all lectures. The text of discussions during the session have not been included.

The Institute was supported in part by a grant from the National Science Foundation and in part by the University of Colorado.

The editors thank the lecturers for their cooperation in the preparation of this volume and Mrs. Marion Higa for her expert typing of the manuscript.

Editors,

Wesley E. Brittin
Kalyana T. Mahanthappa

Department of Physics and
Astrophysics
University of Colorado

CONTENTS

I. Axiomatic Field Theory

ALGEBRAIC METHODS IN THEORETICAL PHYSICS†

Walter Wyss
Department of Mathematics and
Department of Physics and Astrophysics
University of Colorado
Boulder, Colorado

Contents

Introduction

The goal of theoretical physics is to describe the empirical world by means of mathematical structures. Such a structure has to simulate the preparations (states) and observations of systems under investigation.

We want to discuss algebraic structures, where the observables correspond to hermitian elements of a topological *-algebra and where the states are continuous positive linear functionals over this algebra. These lectures will deal with some generalities of this set-up and some examples that are used in quantum field theory and in statistical mechanics.

I. Topological *-algebras

The reason for the introduction of this setup will be touched upon by Profs. H. J. Borchers and M. Guenin.

---

†Presented at the INSTITUTE FOR THEORETICAL PHYSICS, University of Colorado, Summer 1969.

A. Definitions and Some Properties

Definition I.1. A topological *-algebra $\mathfrak{A}$ is a Hausdorff locally convex topological vector space over $\mathbb{C}$ with two additional operations, namely a multiplication (generally noncommutative) and an involution. These operations satisfy for $f, g, h \in \mathfrak{A}$, $\lambda \in \mathbb{C}$

a) 1) $f(gh) = (fg)h$
   2) $f(g+h) = fg + fh$, $(f+g)h = fh + gh$
   3) $f(\lambda g) = (\lambda f)g = \lambda(fg)$
   4) the product is separately continuous

b) 1) $(f^*)^* = f$
   2) $(f+g)^* = f^* + g^*$
   3) $(\lambda f)^* = \bar{\lambda} f^*$
   4) $(fg)^* = g^* f^*$

Remarks:

a) Since $\mathfrak{A}$ has to be a Hausdorff locally convex topological vector space, its topology is given by a set of seminorms $\{p_\alpha\}$ with the properties

1) $p_\alpha(f) \geq 0$
2) $p_\alpha(\lambda f) = |\lambda| \, p_\alpha(f)$
3) $p_\alpha(f+g) \leq p_\alpha(f) + p_\alpha(g)$

A seminorm is called a norm if $p_\alpha(f) = 0 \Leftrightarrow f = 0$.

b) In our discussions $\mathfrak{A}$ is always assumed to have an identity $\mathbf{1}$ for the multiplication.

c) Generally one does not require the involution to be continuous.

d) Notice that the set $\mathfrak{A}_0 = \{f \in \mathfrak{A};\ f = f^*\}$, i.e. the set of all hermitian elements is a real vector subspace of $\mathfrak{A}$ and closed if the involution is continuous.

Lemma I.1. Let $\mathfrak{A}$ be a *-algebra. Every element $f \in \mathfrak{A}$ can be written in the form

$$f = f_1 + if_2, \quad f_1, f_2 \in \mathfrak{A}_0,$$

i.e. $\mathfrak{A} = \mathfrak{A}_0 + i\mathfrak{A}_0$.

Proof. With

$$f_1 = \frac{1}{2}(f + f^*), \qquad f_2 = \frac{1}{2i}(f - f^*)$$

we have the desired representation. ∎

Definition I.2. Let $\mathfrak{A}$ be a *-algebra.

1) A subset $S \subset \mathfrak{A}$ is called self adjoint if it is stable under the involution.

2) Let M be a subset of $\mathfrak{A}$. The commutant M' of M is defined by $M' = \{f \in \mathfrak{A};\ fg = gf \ \forall\ g \in M\}$.

3) $\mathfrak{Z}(\mathfrak{A}) = \mathfrak{A} \cap \mathfrak{A}'$ is called the center of $\mathfrak{A}$ and is an abelian subalgebra of $\mathfrak{A}$.

Lemma I.2.

1) If $M \subset \mathfrak{A}$ is self adjoint, then M' is a *-subalgebra of $\mathfrak{A}$ and closed if the involution is continuous.

2) $M \subset M''$.

3) $M_1 \subset M_2 \Rightarrow M_2' \subset M_1'$.

Proof.

1) for $f, g \in M'$, $\lambda \in \mathbb{C}$ we have trivially
   a) $\lambda f + g \in M'$
   b) $f^* \in M'$
   c) $fg \in M'$

Furthermore let $M_f = \{g \in \mathfrak{A};\ gf = fg\}$. M_f is closed due to the separate continuity of the product. Hence $M' = \bigcap_{f \in M} M_f$ is closed too. The continuity of the involution thus turns M' into a closed *-subalgebra of $\mathfrak{A}$.

2) Let $f \in M$. Then every $g \in M'$ satisfies $gh = hg$, $\forall\ h \in M$, especially $gf = fg$. This means that $f \in M''$.

3) Suppose that $M_1 \subset M_2$. Then $M_2' = \{f \in \mathfrak{A};\ fg = gf, \forall\ g \in M_2\}$. Hence $M_2' \subset M_1'$. ■

Definition I.3. Let $\mathfrak{A}$ be a *-algebra and I a nontrivial subspace of $\mathfrak{A}$. I is called

1) a left ideal if $\mathfrak{A}I \subset I$

2) a right ideal if $I\mathfrak{A} \subset I$

3) a twosided ideal if $\mathfrak{A}I\mathfrak{A} \subset I$

4) a maximal ideal if it is not properly contained in any other nontrivial ideal of the same kind

5) a minimal ideal if it does not properly contain any other nontrivial ideal of the same kind.

Lemma I.3. Let $\mathfrak{A}$ be a *-algebra. Then

1) The identity is never contained in any proper ideal.

2) An element $f \in \mathfrak{A}$ has an inverse f^{-1} iff f is not contained in any ideal.

3) The involution maps the set of left ideals one-to-one onto the set of right ideals. Thus a self adjoint ideal (*-ideal) has to be two-sided.

Proof.

1) Let I be a left ideal. If $\mathbf{1} \in I$ then $\mathfrak{A} \subset I$ which means that I is not proper. The same is true for the other kinds of ideals.

2) a) Assume $\nexists f^{-1}$. Then the left ideal $I_f = \mathfrak{A}f \neq \mathfrak{A}$ and $f \in I_f$.

b) Now let $f \in I$. If f would have an inverse f^{-1}, then $f^{-1}f \in I$, i.e. $\mathbf{1} \in I$. But this is impossible.

3) Let I be a left ideal, i.e. $\mathfrak{A}I \subset I$. Then I* satisfies $I^*\mathfrak{A} \subset I^*$ and hence is a right ideal. Since the *-map is an involution we get the desired result.

Definition I.4. The Jacobson radical of a *-algebra $\mathfrak{A}$ is defined to be

$$\mathfrak{R}(\mathfrak{A}) = \{f \in \mathfrak{A};\ \exists (1 + gf)^{-1},\ \forall g \in \mathfrak{A}\} \ .$$

An algebra whose radical is trivial is called semisimple. An algebra with $\mathfrak{R}(\mathfrak{A}) = \mathfrak{A}$ is called radical.

Lemma I.4. If $\mathfrak{A}$ is not radical then $\mathfrak{R}(\mathfrak{A})$ is equal to the intersection of all maximal right (left) ideals and thus is two-sided.

Proof.

a) First suppose that f belongs to all maximal left ideals and not to $\mathfrak{R}(\mathfrak{A})$, i.e. $\nexists (1 + gf)^{-1}$ for some $g \in \mathfrak{A}$. Then $1 + gf \in I$, where I is some maximal left ideal. But $f \in I$ and this leads to the conclusion that $\mathbf{1} \in I$; this is impossible. Hence $f \in \mathfrak{R}(\mathfrak{A})$.

b) Now let $f \in \mathfrak{R}(\mathfrak{A})$ and suppose that there is a maximal left ideal I such that $f \notin I$. Then $I + \mathfrak{A}f$ is a left ideal that contains I. Due to the maximality of I we have $I + \mathfrak{A}f = \mathfrak{A}$, especially we get $1 = h - gf$, $h \in I$, $g \in \mathfrak{A}$ or $h = 1 + gf$. This contradicts the assumption that $f \in \mathfrak{R}(\mathfrak{A})$. ■

Definition I.5. Let $\mathfrak{A}$ be a *-algebra and

$$\pi = \{f \in \mathfrak{A};\ f = \sum \lambda_i f_i^* f_i,\ \lambda_i \geq 0,\ f_i \in \mathfrak{A}\} \ ,$$

where the sum extends over finitely many terms only. The closure (relative to the underlying topology) $K = \overline{\pi}$ is called the positive cone of $\mathfrak{A}$. K turns $\mathfrak{A}$ into an ordered topological vector space by

$$f \geq 0 \Leftrightarrow f \in K.$$

Lemma I.5. $\mathfrak{A}_0 = K - K$, i.e. K is generating for $\mathfrak{A}_0$.

Proof. Any $f \in \mathfrak{A}_0$ can be written as

$$f = \tfrac{1}{4}(1 + f)(1 + f) - \tfrac{1}{4}(1 - f)(1 - f) \qquad \blacksquare$$

There are two kinds of cones that will play an important role later.

Definition I.6.

1) A subset of $\mathfrak{A}_0$ is called full if

$$A = (A + K) \cap (A - K) .$$

2) The cone K is called normal in $\mathfrak{A}_0$ if there is a neighborhood basis of zero consisting of full sets.

3) The cone K is called a strict B-cone if $\mathfrak{B}_K = \{B \cap K - B \cap K;$ B bounded$\}$ is a fundamental system of bounded sets, i.e. every bounded set is contained in a suitable member of $\mathfrak{B}_K$.

Definition I.7. Let $\mathfrak{A}$, $\mathfrak{B}$ be two *-algebras.

1) A linear map

$$\tau : \mathfrak{A} \to \mathfrak{B}$$

is called a homomorphism if

$$\tau(fg) = \tau(f)\,\tau(g)$$

τ is said to be a *-homomorphism if in addition

$$\tau(f^*) = (\tau(f))^*$$

2) The set

$$\operatorname{Ker} \tau = \{f \in \mathfrak{A};\ \tau(f) = 0\}$$

is called the kernel of the homomorphism τ. If $\operatorname{Ker} \tau = 0$, then τ is called an isomorphism.

Lemma I.6. The *-ideals of a *-algebra are in one-to-one correspondence with the *-homomorphisms.

Proof. The kernel of a *-homomorphism is clearly a *-ideal. Let now I be a *-ideal. The canonical map

$$\tau: \mathfrak{A} \to \mathfrak{A}/I$$

is a *-homomorphism with Ker $\tau = I$. ■

Definition I.8. Let $\mathfrak{A}$ be a *-algebra. A *-representation of $\mathfrak{A}$ is a pair (H,τ), where τ is a *-homomorphism into the linear operators $\mathfrak{L}(H)$ of a topological innner product space H, i.e.

$$\tau: \mathfrak{A} \to \mathfrak{L}(H)$$

such that

$$(\tau(f^*)x, y) = (x, \tau(f)y) \quad \forall\, x,y \in H.$$

Definition I.9. A representation is called

1) faithful, if it is one-to-one.

2) algebraically irreducible, if there are no proper τ-invariant subspaces in H.

3) topologically irreducible, if there are no proper closed τ-invariant subspaces in H.

Definition I.10. Let $\mathfrak{B}$ be a subalgebra of the algebra of linear operators $\mathfrak{L}(E)$ on a topological vector space E, and for $x \in E$ let

$$E_x = \{y \in E;\ y = Ax,\ \forall\, A \in \mathfrak{B}\}\ .$$

E_x is clearly an invariant subspace under $\mathfrak{B}$.

1) If there is a $x \in E$ such that $E_x = E$, then $\mathfrak{B}$ is said to be algebraically cyclic and x is called an algebraically cyclic vector.

2) If there is a $x \in E$ such that $E_x = E$, then $\mathfrak{B}$ is said to be topologically cyclic and x is called a topologically cyclic vector.

3) A representation τ of an algebra $\mathfrak{A}$ is said to have one of the above properties if $\tau(\mathfrak{A})$ has that property.

Lemma I.7. $\mathfrak{B}$ is algebraically (topologically) irreducible iff every nonzero vector of E is algebraically (topologically) cyclic.

Proof. a) For any $x \in E$, E_x is an invariant subspace of E and due to the algebraic irreducibility has to coincide with E.

b) Suppose that every $x \in E$ is algebraically cyclic. Any invariant subspace $M \subset E$ contains some E_x and thus $M = E$. ■

Definition I.11. Let I be a left ideal of an algebra $\mathfrak{A}$. The representation

$$\tau: \mathfrak{A} \to \mathfrak{L}(\mathfrak{A}/I),$$

given by left multiplication, is called the left regular representation. Its kernel is given by

$$\operatorname{Ker} \tau = \{f \in \mathfrak{A};\ f\mathfrak{A} \subset I\},$$

i.e. Ker τ is the biggest two-sided ideal contained in I.

Lemma I.8. The left regular representation

$$\tau: \mathfrak{A} \to \mathfrak{L}(\mathfrak{A}/I)$$

will be algebraically irreducible iff I is maximal.

Proof. We show that there is a one-to-one correspondence between invariant subspaces of $\mathfrak{A}/I$ and left ideals containing I.

Let ε be the canonical map

$$\varepsilon: \mathfrak{A} \to \mathfrak{A}/I.$$

Then

$$\tau(f)\ \varepsilon(g) = \varepsilon(fg).$$

For a left ideal $J \supset I$, we find that $\varepsilon(J)$ is a τ-invariant subspace of $\mathfrak{A}/I$.

If M is any invariant subspace of $\mathfrak{A}/I$, then $J = \{f \in \mathfrak{A};\ \varepsilon(f) \in M\}$ is a left ideal containing I. One sees immediately that the above correspondence is one-to-one. ■

Lemma I.9. Let τ be an algebraically cyclic representation of an algebra $\mathfrak{A}$. Then there is a left ideal $I \subset \mathfrak{A}$ such that the left regular representation is algebraically equivalent to τ.

Proof. Let $\tau: \mathfrak{A} \to h(E)$ be algebraically cyclic and $x \in E$ an algebraically cyclic vector, i.e. $\tau(\mathfrak{A})x = E$. The subspace

$$I = \{f \in \mathfrak{A};\ \tau(f)x = 0\}$$

is a left ideal. With

$$\varepsilon : \mathfrak{A} \to \mathfrak{A}/I$$

being the canonical map, let

$$A: \mathfrak{A}/I \to E$$

be given by

$$A\,\varepsilon(f) = \tau(f)\,x.$$

A is one-to-one and due to the cyclicity onto. Hence with

$$\pi: \mathfrak{A} \to \mathfrak{L}(\mathfrak{A}/I)$$

being the left regular representation, we get

$$A\,\pi(f)\,\varepsilon(g) = A\,\varepsilon(fg) = \tau(fg)x = \tau(f)\,\tau(g)\,x$$

$$= \tau(f)\,A\,\varepsilon(g)\ , \qquad \forall\, \varepsilon(g) \in \mathfrak{A}/I,$$

or

$$A\,\pi(f) = \tau(f)\,A, \qquad \forall\, f \in \mathfrak{A}\ ,$$

which is the desired algebraic equivalence between π and τ. ■

Definition I.12. A two-sided ideal is called primitive if it is the biggest two-sided ideal contained in a maximal left ideal.

Lemma I.10.

1) An ideal is primitive if and only if it is the kernel of an algebraically irreducible representation.

2) The radical $\mathfrak{R}(\mathfrak{A})$ is equal to the intersection of all primitive ideals, i.e. the intersection of the kernels of all algebraically irreducible representations.

Proof.

1) a) Lemma I.9 and I.8 say that the kernel of an algebraically irreducible representation is primitive.

b) If the kernel is primitive, then there is a maximal left ideal I containing it. The corresponding left regular representation is then algebraically irreducible and has the above kernel.

2) We know that $\mathfrak{R}(\mathfrak{A})$ is equal to the intersections of all maximal left ideals. Hence the intersection of all primitive ideals is contained in $\mathfrak{R}(\mathfrak{A})$. On the other hand let $f \notin \{\cap\, I,\ I \text{ primitive}\}$.

Then there is an algebraically irreducible representation τ on a vector space E with $\tau(f) \neq 0$. Choose $x \in E$ with $\tau(f)x \neq 0$ and look at $I = \{g \in \mathfrak{A}; \tau(g)x = 0\}$. I is a maximal left ideal and $f \notin I$. Hence $f \notin \mathfrak{R}(\mathfrak{A})$. ■

Definition I.13. Let $\mathfrak{A}$ be a topological *-algebra.

1) The *-radical $\mathfrak{R}^*$ is defined to be the intersection of the kernels of all topological *-representations on a Hilbert space.

2) If $\mathfrak{R}^*(\mathfrak{A}) = 0$, then $\mathfrak{A}$ is called *-semisimple.

B. States and Representations

Let $\mathfrak{A}$ be a topological *-algebra with identity, K the positive cone of $\mathfrak{A}$ and $\mathfrak{A}'$ the topological dual of $\mathfrak{A}$, i.e. all continuous linear functionals on $\mathfrak{A}$.

Definition I.14. An element $T \in \mathfrak{A}'$ is called

1) hermitian if $T(f^*) = \overline{T(f)}$

2) positive if $T(f) \geq 0 \quad \forall f \in K$

Lemma I.11. If $T \in \mathfrak{A}'$ is positive, then it is hermitian and satisfies

$$|T(f^*g)|^2 \leq T(f^*f)\, T(g^*g),$$

the so-called Cauchy-Schwartz inequality.

Proof. Let $h = \lambda f + \mu g$. Then $T(h^*h) \geq 0 \quad \forall \lambda, \mu$ and $\forall g, f \in \mathfrak{A}$, or $T(h^*h) = |\lambda|^2 T(f^*f) + \bar{\lambda}\mu\, T(f^*g) + \lambda\bar{\mu}\, T(g^*f) + |\mu|^2\, T(g^*g)$. Hence $\bar{\lambda}\mu\, T(f^*g) + \lambda\bar{\mu}\, T(g^*f)$ is real. For $\lambda = \mu = 1$, resp. $\lambda = 1$, $\mu = i$, we get that $T(f^*g) = \overline{T(g^*f)}$. Furthermore let $\lambda = \bar{\lambda}$ and $\mu = T(g^*f)$. This leads to the positivity condition $|T(f^*g)|^2 \leq T(f^*f)\, T(g^*g)$. ■

Since we are mostly interested in positive linear functionals, it suffices to look at the hermitian elements $\mathfrak{A}_0$ of our *-algebra $\mathfrak{A}$. We assume the involution to be continuous and hence $\mathfrak{A}_0$ is a closed subspace of $\mathfrak{A}$. The topological dual $\mathfrak{A}_0'$ of $\mathfrak{A}_0$ is then isomorphic to the subspace of hermitian functionals on $\mathfrak{A}$.

Definition I.15. $K' = \{T \in \mathfrak{A}_0';\ T(f) \geq 0 \quad \forall f \in K\}$ is called the dual cone of K. This gives rise to an ordering of linear functionals by T_1, $T_2 \in \mathfrak{A}_0'$, $T_1 > T_2 \Leftrightarrow T_1 - T_2 \geq 0$. $\mathfrak{A}_0'$ thus is an ordered topological vector space.

Definition I.16. An element $T \in K'$ is called a state, provided $T(1) = 1$. Denote the set of states by σ.

Lemma I.12.

1) σ is a base for K', i.e. σ is convex and each $T \in K'$ has the representation $T = \lambda \cdot W$, $\lambda > 0$, $W \in \sigma$.

2) K' is a proper cone, i.e. $T \in K'$, $-T \in K'$ implies $T = 0$.

Proof.

1) If $T(1) \neq 0$ then we always can normalize T. In the case $T(1) = 0$ we find from the Cauchy-Schwartz inequality $|T(f)|^2 \leq T(1)\, T(f^*f)$ that $T(f) = 0 \ \forall f \in \mathfrak{U}$, or $T = 0$.

2) If $T \in K \cap (-K)$, $T \neq 0$, then $T(f^*f) = 0 \ \forall f \in \mathfrak{U}$. Hence $T(f) = 0 \ \forall f \in \mathfrak{U}$, or $T = 0$. ■

Definition I.17. Let $T \in \sigma$. Whenever $T = \lambda_1 T_1 + \lambda_2 T_2$ with T_1, $T_2 \in \sigma$, $\lambda_1 > 0$, $\lambda_2 > 0$, and $\lambda_1 + \lambda_2 = 0$, implies $T_1 = T_2 = T$, then T is called extremal in σ.

Definition I.18. To each $T \in K'$ we associate an ideal $I(T)$ in $\mathfrak{U}$ by

$$I(T) = \{f \in \mathfrak{U};\ T(f^*f) = 0\}.$$

$I(T)$ is called the left kernel of T.

Lemma I.13. $I(T)$ is a closed left ideal.

Proof. First it is a left ideal since with $f \in I(T)$ we have

$$|T((gf)^*(gf))|^2 \leq T(f^*f)\, T(f^*g^*gg^*gf) = 0,$$

i.e. $gf \in I(T)$, $\forall g \in \mathfrak{U}$. To show that $I(T)$ is closed we use the representation

$$I(T) = \{f \in \mathfrak{U};\ T(gf) = 0,\ \forall g \in \mathfrak{U}\}$$

$$= \bigcap_{g \in \mathfrak{U}} \{f \in \mathfrak{U};\ T(gf) = 0\}$$

Since T is continuous we know that $\{f \in \mathfrak{U};\ T(gf) = 0\}$ is closed for a fixed g.

Lemma I.14. If $T \in \sigma$ and $I(T)$ be maximal, then T is extremal.

Proof. Suppose T is not extremal. Then $T = \lambda_1 T_1 + \lambda_2 T_2$, $\lambda_1, \lambda_2 > 0$, $T_1 \neq T_2$ and hence $I(T) = I(T_1) \cap I(T_2)$, which says that $I(T)$ is not maximal. ■

Lemma I.15. Each element $T \in K'$ gives rise to a *-representation of $\mathfrak{A}$ (Gelfand-Segal construction).

Proof. Let $T \in K'$ and $I(T)$ the corresponding left kernel. Then $H(T) = \mathfrak{A}/I(T)$ is a Hausdorff locally convex topological vector space under the quotient topology. We then have the exact sequence

$$0 \to I(T) \to \mathfrak{A} \xrightarrow{\varepsilon} H(T) \to 0$$

$H(T)$ is also an inner product space by

$$(\varepsilon(f), \varepsilon(g)) = T(f^*g)$$

and carries a *-representation π of $\mathfrak{A}$ by

$$\pi(f)\, \varepsilon(g) = \varepsilon(fg).$$

With respect to the initial topology of $H(T)$ this representation is algebraically and hence topologically cyclic, with $G(1)$ being a cyclic vector. With respect to the inner product topology on $H(T)$, π is only topologically irreducible and not necessarily algebraic irreducible. ■

Remark:

1) There are various locally convex topologies on $\mathfrak{A}_0$ and $\mathfrak{A}_0'$, varying from the weak topology to the Mackey topology.

2) In the context of topological *-algebras there are two important questions:

a) Are positive functionals continuous?

b) Is the dual cone K' generating, i.e. has every functional $L \in \mathfrak{A}_0'$ the representation $L = T_1 - T_2$, with $T_1, T_2 \in K'$. When is this representation unique?

II. Examples

We now want to investigate some special topological *-algebras that are often used in theoretical physics. Our emphasis will be on some characteristic properties; for their proofs we frequently refer to the literature. The algebras under discussion are assumed to have an identity.

A. C*-algebras[1),2)]

Definition II.1. A topological *-algebra $\mathfrak{A}$ is called a C*-algebra provided

1) $\mathfrak{A}$ is a Banach space
2) For $A, B \in \mathfrak{A}$, the norm $\|\cdot\|$ satisfies

$$\|AB\| \leq \|A\| \, \|B\|$$

$$\|A^*A\| = \|A\|^2 \qquad \text{(C*-condition)}$$

Lemma II.1. $\mathfrak{B}(\mathfrak{H})$, the bounded operator on a Hilbert space, is a C*-algebra.

Proof. We know that $\mathfrak{B}(\mathfrak{H})$ is a Banach space and that

$$\|AB\| \leq \|A\| \, \|B\|; \quad \|A\| = \sup_{\|x\|=1} \|Ax\| \, , \quad x \in \mathfrak{H} \, .$$

We only have to show the C*-condition. Let $A \in \mathfrak{B}(\mathfrak{H})$; then

$$\|Ax\|^2 = (Ax, Ax) = (A^*Ax, x) \leq \|A^*A\| \, \|x\|^2 \, , \quad \forall \, x \in \mathfrak{H}.$$

Hence $\|A\|^2 \leq \|A^*A\| \leq \|A^*\| \, \|A\|$ and thus $\|A\| = \|A^*\|$ by symmetry. Therefore $\|A^*A\| = \|A\|^2$.

Properties II.1. Let $\mathfrak{A}$ be a complex C*-algebra. Then
1) the closure of a proper ideal is proper
2) $\mathfrak{A}$ is semisimple, i.e. $\mathfrak{R} = 0$
3) $\mathfrak{A}$ is *-semisimple, i.e. $\mathfrak{R}^* = 0$.
4) $\mathfrak{A}$ is isometrically *-isomorphic to a norm closed *-subalgebra of some $\mathfrak{B}(\mathfrak{H})$.
5) every topologically irreducible representation of $\mathfrak{A}$ is algebraically irreducible.

Let us now look at the hermitian part $\mathfrak{A}_0$ of a C*-algebra; this is a Banach space. Its dual $\mathfrak{A}_0'$ is also a Banach space in the strong topology $\beta(\mathfrak{A}_0', \mathfrak{A}_0)$ (norm topology).

Lemma II.2.
1) The positive cone $K \subset \mathfrak{A}_0$ is a normal strict B-cone.
2) The dual cone $K' \subset \mathfrak{A}_0'$ is a normal strict B-cone.

Proof.[3]

1) K is a strict B-cone because it is generating $\mathfrak{A}_0$. For the normality of K we have to show that for $A, B \in K$ we have $\|A + B\| \geq \|A\|$. Due to the fact that $\mathfrak{A}$ is a C*-algebra, we have an isometrically *-isomorphism π of $\mathfrak{A}$ onto a *-subalgebra of $\mathfrak{B}(\mathfrak{H})$. Hence $(x, \pi(A+B)x) - (x, \pi(A)x) = (x, \pi(B)x) \geq 0$, $\forall \, x \in \mathfrak{H}$, and thus

$$\|\pi(A+B)\| \geq \|\pi(A)\|$$

or

$$\|A+B\| \geq \|A\| \ .$$

2) This follows from the duality theorem between normal and strict B-cones. ■

These nice properties of K and K′ lead now to the following statement:

Lemma II.3.

1) Every positive functional on $\mathfrak{A}_0$ is continuous.

2) Every $T \in \mathfrak{A}_0'$ can be written as $T = T_1 - T_2$, where $T_1, T_2 \in K'$.

Proof. 1) This follows from the fact that $\mathfrak{A}_0$ is a Banach space and K generating. One could also use the fact that K has nonempty interior.

2) Since K′ is a strict B-cone in $\mathfrak{A}_0'$, it has to be generating. ■

Remark: Haag and Kastler[4] emphasized the importance of abstract C*-algebras because of the physical equivalence of all faithful representations.

B. The Field Algebra (Borchers Algebra)

A topological *-algebra of quite a different flavor than a C*-algebra is the so-called field algebra; it plays a fundamental role in Wightman's theory of quantized fields.

Definition II.2.[5] Let $\mathcal{S}_0 = \mathbb{C}$ and $\mathcal{S}_n = \mathcal{S}(\mathbb{R}^{4n})$, the Laurent Schwartz test function space. The field algebra is the topological direct sum

$$\mathfrak{A} = \bigoplus_{n=0}^{\infty} \mathcal{S}_n \ ,$$

equipped with the product

$$f = \{f_0, f_1, \ldots\}, \quad g = \{g_0, g_1, \ldots\}$$

$$(fg)_n(x_1, \ldots, x_n) = \sum_{k=0}^{n} f_k(x_1, \ldots, x_k) g_{n-k}(x_{k+1}, \ldots, x_n) \ ,$$

and the involution

$$(f^*)_n(x_1,\ldots,x_n) = \overline{f_n(x_n,\ldots,x_1)}.$$

The element

$$\mathbb{1} = \{1,0,\ldots\}$$

is the identity.

Properties II.2.[6] The field algebra $\mathfrak{A}$ has the following properties:

1) $\mathfrak{A}$ is a nuclear *-algebra, bornological and LF.

2) $\mathfrak{A}$ is semisimple, i.e. $\mathfrak{R} = 0$.

3) $\mathfrak{A}$ has no divisors of zero and only those elements that are a multiple of the identity have an inverse.

4) $\mathfrak{A}$ has no minimal ideals.

5) $\mathbb{1}$ and 0 are the only idempotents, i.e. satisfying $f^2 = f$.

6) the positive cone K has no interior points.

There are many other properties of $\mathfrak{A}$, some of them are the same as for C*-algebras, especially concerning the positive cone.

C. Von Neumann Algebras, Classification of Factors[7]

Definition II.3. Let $\mathfrak{A}$ be a *-subalgebra of some $\mathfrak{B}(\mathcal{H})$. $\mathfrak{A}$ is called a von Neumann algebra, provided $\mathfrak{A} = \mathfrak{A}''$. We restrict ourselves to separable Hilbert spaces.

Lemma II.4. A von Neumann algebra $\mathfrak{A}$ is a weakly closed *-subalgebra of $\mathfrak{B}(\mathcal{H})$; any weakly closed *-subalgebra of $\mathfrak{B}(\mathcal{H})$, containing the identity, is a von Neumann algebra. Since a weakly closed set is also strongly closed, a von Neumann algebra is also a C*-algebra; a C*-algebra however is not necessarily weakly closed.

Proof. That a von Neumann algebra is weakly closed is an immediate consequence of Lemma I.2, because the product is separately weakly continuous and the involution is weakly continuous. For the rest of the proof we refer to the literature. ∎

Definition II.4.

a) An hermitian idempotent $P \in \mathfrak{B}(\mathcal{H})$ is called a projection, i.e. $P^* = P$, $P^2 = P$.

b) Let $\mathcal{H}_1$, $\mathcal{H}_2$ be two Hilbert spaces. A map $V: \mathcal{H}_1 \to \mathcal{H}_2$ is called a partial isometry with initial domain M and final domain N,

if V maps M isometrically onto N and $M^{\perp}$, the orthogonal complement of M, into zero. $V^*V = P_M$ is then the initial domain projection and $VV^* = P_N$ is the final domain projection.

Definition II.5.

a) Two closed subspaces M, N of an Hilbert space $\mathfrak{H}$ are called equivalent, if there is a partial isometry U with initial domain M and final domain N; we write $M \sim N$.

b) We introduce the preorder amongst closed subspaces of $\mathfrak{H}$ by $M < N \Leftrightarrow$ M is equivalent to a subspace of N.

Definition II.6. Let $\mathfrak{A}$ be a von Neumann algebra on $\mathfrak{H}$.

a) A subspace $M \subset \mathfrak{H}$ is said to belong to $\mathfrak{A}$, if M is invariant under all unitary operators in $\mathfrak{A}'$; we write $M \eta \mathfrak{A}$.

b) A linear operator A (not necessarily bounded) is said to belong to $\mathfrak{A}$ if A commutes with all unitary operators in $\mathfrak{A}'$; we write $A \eta \mathfrak{A}$.

Lemma II.5.

a) If $A \eta \mathfrak{A}$ and A is bounded, then $A \in \mathfrak{A}$.

b) If $M \eta \mathfrak{A}$ and M is closed, then the projection P_M onto M belongs to $\mathfrak{A}$, i.e. $P_M \in \mathfrak{A}$.

Proof.

a) $A \in (\mathfrak{A}'(U))' = \mathfrak{A}'' = \mathfrak{A}$.

b) $M \eta \mathfrak{A}$ says that $UM \subset M$ for all unitary operators U in $\mathfrak{A}'$. Since $U^*M \subset M$ we have $M \subset UM$ and hence $M = UM$ and $M^{\perp} = UM^{\perp}$. That means that every unitary $U \in \mathfrak{A}'$ is reduced by M, i.e. $UP_M = P_M U$, $\forall\, U \in \mathfrak{A}'$. Hence $P_M \eta \mathfrak{A}$ and since P_M is bounded, we have $P_M \in \mathfrak{A}$. ■

The equivalence relation and preorder amongst closed subspaces belonging to $\mathfrak{A}$ can be carried over to an equivalence relation and preorder amongst projections belonging to $\mathfrak{A}$.

Definition II.7.

a) Let $\mathfrak{A}$ be a von Neumann algebra. Two projections $P_1, P_2 \in \mathfrak{A}$ are said to be equivalent if $P_1\mathfrak{H} \sim P_2\mathfrak{H}$; we write $P_1 \sim P_2$.

b) A preorder is given by $P_1 < P_2 \Leftrightarrow P_1 \sim P$ and $P\mathfrak{H} \subset P_2\mathfrak{H}$.

Definition II.8. A von Neumann algebra $\mathfrak{A}$ is called a factor if the center $\mathfrak{Z}(\mathfrak{A}) = \mathbb{C}$.

Since every von Neumann algebra $\mathfrak{A}$ has a direct integral decomposition into factors,[7] we will restrict ourselves from now on to factors.

Properties II.3.[1,7] Let $\mathfrak{A}$ be a factor. Then the following holds:

a) From $A \in \mathfrak{A}$, $A' \in \mathfrak{A}'$ and $AA' = 0$ it follows that either $A = 0$ or $A' = 0$.

b) Let $M \eta \mathfrak{A}$ and $N \eta \mathfrak{A}$, then either $M < N$ or $N < M$. Equivalently, if $P_M, P_N \in \mathfrak{A}$ then either $P_M < P_N$ or $P_N < P_M$.

c) If M and N belong to $\mathfrak{A}$ and $M < N$, $N < M$, then $M \sim N$. Equivalently, if $P_M, P_N \in \mathfrak{A}$, and $P_M < P_N$, $P_N < P_M$, then $P_M \sim P_N$.

Definition II.9. Let $\mathfrak{A}$ be a factor and M a closed subspace of $\mathfrak{K}$, belonging to $\mathfrak{A}$. Then

a_1) M is infinite $\Leftrightarrow$ M is equivalent to a nontrivial subspace of M.

a_2) A projection $P \in \mathfrak{A}$ is infinite $\Leftrightarrow$ $P\mathfrak{K}$ is infinite.

b_1) M is finite $\Leftrightarrow$ M is not infinite.

b_2) A projection $P \in \mathfrak{A}$ is finite $\Leftrightarrow$ $P\mathfrak{K}$ is finite.

c_1) M is called minimal if $M \neq 0$ and $N < M$ implies $N = 0$ or $N = M$.

c_2) A projection $P \in \mathfrak{A}$ is called minimal if $P \neq 0$ and $P_1 < P$ implies $P_1 = 0$ or $P_1 = P$.

Properties II.4.[1,8] On the set of closed subspaces belonging to a factor $\mathfrak{A}$ there exists a real valued function D, called the relative dimension, such that

(i) $D(M) = 0$ if $M = 0$ and $D(M) > 0$ if $M \neq 0$

(ii) $D(M) = \infty$ if M is infinite

(iii) $M \sim N \Rightarrow D(M) = D(N)$

(iv) $M \perp N \Rightarrow D(M + N) = D(M) + D(N)$

(v) $M < N$, M finite $\Rightarrow D(M) < D(N)$

Von Neumann-Murray Classification of Factors:[1,8]

For each type we give the range Δ of the relative dimension D.

All projections are infinite (purely infinite)	Some projections are infinite, some are finite (infinite)		All projections are finite (finite)	
	No minimal finite projections (continuous)	Has minimal finite projections (discrete)	No minimal finite projections (continuous)	Has minimal finite projections (discrete)
III	II_∞	I_∞	II_1	I_n
$\Delta = \{0, \infty\}$	$\Delta = [0, \infty]$	$\Delta = \{0, 1, 2, .. \infty\}$	$\Delta = [0, 1]$	$\Delta = \{0, 1, .., n\}$

Remark: In the von Neumann-Murray classification of factors it is known that in the classes II_1 and III there are nonisomorphic factors. This led Araki and Woods[9,10] to a detailed study of factors constructed as infinite tensor products of finite type I factors (ITPFI-factors) on separable Hilbert spaces. For completeness reasons we will give a summary of the Araki-Woods classification of ITPFI-factors; the details can be found in Refs. 9 and 10.

Araki-Woods Classification of ITPFI-factors[9,10]

To every ITPFI-factor $\mathfrak{A}$ one associates a subset $r_\infty(\mathfrak{A})$ of the non-negative real numbers; $r_\infty(\mathfrak{A})$ is called the asymptotic ratio set. There are the following standard sets for $r_\infty(\mathfrak{A})$:

$r_\infty(\mathfrak{A})$	Von Neumann-Murray classification	Isomorphisms
ϕ	I_n, $n=0,1,\ldots,<\infty$	one isomorphic class for each n
$\{0\}$	I_∞	all are isomorphic
$\{1\}$	hyperfinite type II_1	all are isomorphic
$\{0, x^n; n=0,\pm1,\pm2,..\}$ $0<x<1$	III	one and only one isomorphic class for each x
$\{0,1\}$	contains $I_\infty \otimes$ hyperfinite II_1, and III	non denumerably many type III isomorphic classes
$[0,\infty)$	III	only one isomorphic class

Remark.[9] The factor $\mathfrak{A}$ describing CCR for a nonrelativistic free Bose gas at finite temperature, finite density and no macroscopic occupation of the ground state is of type $r_\infty(\mathfrak{A}) = [0,\infty)$.

Also the factor $\mathfrak{A}$ describing CAR for a nonrelativistic free Fermi gas at finite temperature and finite density is of type $r_\infty(\mathfrak{A}) = [0,\infty)$.

Acknowledgment

I would like to thank Profs. W. E. Brittin and K. T. Mahanthappa for the warm hospitality at the Institute for Theoretical Physics in Boulder.

References

1. M. A. Naimark, Normed Rings (Nordhoff, Groningen, 1964).
2. C. E. Rickart, General Theory of Banach Algebras (Van Nostrand, Princeton, New Jersey, 1960).
3. H. H. Schaefer, Topological Vector Spaces (Macmillan, New York, 1966).
4. R. Haag and D. Kastler, J. Math. Phys. 5, 848 (1964).
5. H. J. Borchers, Nuovo Cimento 24, 214 (1962).
6. W. Wyss, in Lectures in Theoretical Physics, Vol. XID (Gordon and Breach, New York, 1969), p. 533.
7. J. Dixmier, Les Algèbres d'Operateurs dans l'Espace Hilbertien (Gauthier-Villars, Paris, 1969).
8. F. J. Murray and J. von Neumann, Ann. of Math 37, 116 (1936).
9. H. Araki and E. J. Woods, Publ. RIMS, Kyoto University, Ser. A, 3, 51 (1968).
10. H. Araki, Publ. RIMS, Kyoto University, Ser. A, 4, 585 (1969).

INTRODUCTION TO ALGEBRAIC TECHNIQUES IN QUANTUM FIELD THEORY AND STATISTICAL MECHANICS†

Marcel Guenin
Institute for Theoretical Physics
University of Geneva
Geneva, Switzerland

Table of Contents

†Presented at the INSTITUTE FOR THEORETICAL PHYSICS, University of Colorado, Summer 1969.

Section 0. Introduction

The use of modern mathematical techniques is, to say the least, not very fashionable among theoretical physicists. There are probably many reasons for the present state of things, and I may state two: first, the mathematical apparatus is often difficult and secondly, most of the physicists don't see the slightest advantage to this kind of game.

The aim of these lectures is mainly to bring the students to the point where they can at least understand the language and some of the physical ideas which can be formulated only in that mathematical language. The field chosen here is not one where it is apparent from the first sight which advantages can be obtained from these new approaches. Nevertheless the advantages are present and important enough so that researchers in that field are willing to spend years of hard work in order to gain some understanding. We expect the progresses to be slow and extend over many years before they lead toward results to be compared with experiments, but one should not forget the immense difficulties of the task. Field theory, for instance, dates back to 1928 with Heisenberg and Pauli. Only this year, in 1969, do we have the proof (Jaffe, Glimm, etc.) of the existence of a nontrivial field theory satisfying Wightman's axiom, and it will be certainly a long time before we can apply our results to practical purposes. We are, however, deeply convinced that there is no easy way out, and that any progress toward the construction of nontrivial theories will have to borrow from the results obtained by hard analysis. These statements are not to be considered as criticisms against the phenomenological approaches; we need results in these directions too, and good phenomenology and theory should be at the end complementary to each other.

Modern research in statistical mechanics and quantum field theory makes very extensive use of advanced functional analysis, and one may wonder why it is so that we are going to spend a lot of time on algebraic techniques. The reason is that first, functional analysis is studied by undergraduates to a greater extent than algebraic techniques, and secondly, that most of the problems we encounter in functional analysis arise after the algebraic formulation of the problems. As an example, we may quote the work of Glimm and Jaffe, where a very substantial part is devoted to the proof that a certain operator is self-adjoint. This is of course functional analysis, but the reason that we need this proof can only be understood in the algebraic formulation.

If I have chosen to speak both of field theory and of statistical mechanics, that is because of the great similarity of the problems

arising in the two approaches; in fact, they both may be characterized as being systems with an infinite number of degrees of freedom.

For the reader who wants more details, he can consult the list of references given at the end. As far as lecture notes or textbooks are concerned, I may quote my lectures here in Boulder in 1966 (algebraic methods in QFT), my lectures on algebraic methods in statistical mechanics (Springer 1969, in French) and the book of D. Ruelle on statistical mechanics (Benjamin 1969). The overlap between my lectures and Ruelle's book is not very great, but Ruelle's book represents of course a much more definitive treatment of the subject it covers.

I further want to point out that the subjects treated in these lectures were meant to make a whole with the lectures of Profs. Borchers and Wyss. Most of the repetitions are intentional, and should provide the reader with different points of view.

I finally want to express my gratitude to Professor W.E. Brittin for his kind invitation to spend some time in Boulder and my thanks to Professors H. Borchers and W. Wyss for many discussions.

Section I. Algebras

The purpose of this section is to give in a condensed form the principal definitions and the most elementary theorems which are needed for an understanding of the physical part.

For those who want to study further, we strongly recommend the two books of Dixmier, which are the basic reference works in that field. It may also be said that the material has been restricted because of Prof. Wyss' lectures, and that the overlap present between the two sets of lectures has been done on purpose.†

I.A. Generalities on Algebras

I.A.1. Definition: A set $\mathfrak{A}$ of elements $\{a,b,c,\ldots\}$ is an associative algebra over the real or complex (we shall simply say algebra, since we are going to consider only associative algebras) if:

(i) $\mathfrak{A}$ is a vector space

(ii) an operator of multiplication is defined in $\mathfrak{A}$ and satisfies

(α) $\alpha(ab) = (\alpha a)b$
(β) $a(\alpha b) = \alpha(ab)$ } (bilinear multiplication law)

(γ) $a(bc) = (ab)c$ (associativity)

(δ) $a(b+c) = ab + ac$
(ϵ) $(a+b)c = ac + bc$ } (distributivity)

†It should also be apparent that the content of this section does not vary considerably from other sets of lecture notes given by the same author, for instance in Boulder 1966 or in "Méthodes Algébriques en Mécanique Statistique," Springer 1969.

$\forall\, a,b,c \in \mathfrak{A}$, $\forall\, \alpha \in \mathbb{R}$ or $\mathbb{C}$. N.B. Many authors call such a structure "ring." Properly speaking, however, a ring does not contain a multiplication by scalars.

It is a commonly accepted abuse of language to call an element of an algebra "operator."

I.A.2. Definition: Two elements a and $b \in \mathfrak{A}$, $\mathfrak{A}$ an algebra, are said to commute if $ab = ba$. An algebra is said to be abelian (or commutative) if all its elements commute pairwise.

I.A.3. Definition: The center $\mathfrak{z}$ of an algebra $\mathfrak{A}$ is the set of all those elements of $\mathfrak{A}$ which commute with all elements of $\mathfrak{A}$:

$$\mathfrak{z}(\mathfrak{A}) = \{c \mid c \in \mathfrak{A},\ ac = ca\ \forall\, a \in \mathfrak{A}\}$$

$\mathfrak{z}$ is clearly a subalgebra of $\mathfrak{A}$.

I.A.4. Examples:

(i) In the algebra $\mathbb{M}_2(\mathbb{C})$ of all 2 by 2 matrices,

$$\mathfrak{z}(\mathbb{M}_2) = \left\{ \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \,\middle|\, \alpha \in \mathbb{C} \right\}$$

(ii) In the algebra of the complex matrices of the form

	n	m
n	A	0
m	0	B

$$\mathfrak{z} = \left\{ \begin{array}{c|cc} & n & m \\ \hline n & \begin{matrix} \alpha & 0 \\ 0 & \alpha \end{matrix} & 0 \\ \hline m & 0 & \begin{matrix} \beta & 0 \\ 0 & \beta \end{matrix} \end{array} \,\middle|\, \alpha, \beta \in \mathbb{C} \right\}$$

where A (resp. B) represent all n by n (resp. m by m) matrices.

I.A.5. Definition: Identity:

e is an identity (or unit) element of the algebra $\mathfrak{A}$, if $ae = ea = a$, $\forall a \in \mathfrak{A}$

g is a left identity element of the algebra $\mathfrak{A}$, if $ga = a$, $\forall a \in \mathfrak{A}$

d is a right identity element of the algebra $\mathfrak{A}$, if $ad = a$, $\forall a \in \mathfrak{A}$

I.A.6. Example: In the algebra $\mathfrak{A}$ of all 2 by 2 matrices of the form $\begin{pmatrix} \alpha & \alpha \\ \beta & \beta \end{pmatrix}$; $\alpha, \beta \in \mathbb{C}$, we do not have an identity element, nor a left

identity, but we have an infinity of right identities, namely all matrices of the form $\begin{pmatrix} \alpha & \alpha \\ 1-\alpha & 1-\alpha \end{pmatrix}$, $\alpha \in \mathbb{C}$.

I.A.7. Proposition:

(i) The identity element of an algebra $\mathfrak{A}$ is unique

(ii) If an algebra $\mathfrak{A}$ possesses both a right identity d and a left identity g, it possesses an identity element.

Proof: (i) let e and e′ be two identity elements, then $ee' = e = e'$, which means they are equal.

(ii) $gd = g = d$, put $e = g = d$.

I.A.8. Theorem: Any algebra $\mathfrak{A}$, without identity, can be identified with a sub-algebra of an algebra $\mathfrak{A}_1$ with an identity element.

Proof: (i) $\mathfrak{A}_1$ may be realized† as the set of all pairs (α, a) $\alpha \in \mathbb{C}$, $a \in \mathfrak{A}$, and the operations are defined as follows:

$$(\alpha, a) + (\beta, b) = (\alpha + \beta, a + b)$$

$$\beta(\alpha, a) = (\beta\alpha, \beta a)$$

$$(\alpha, a)(\beta, b) = (\alpha\beta, \beta a + \alpha b + ab)$$

(ii) $\mathfrak{A}$ may be identified to the subalgebra of $\mathfrak{A}_1$ consisting of the elements of the form $(0, a)$.

(iii) The identity element of $\mathfrak{A}_1$ is $(1, 0)$.

Note that $\mathfrak{A}_1$ is abelian if, and only if, $\mathfrak{A}$ is abelian.

N.B. For most of the physical applications which we shall consider, the algebras which we shall use shall possess an identity element. The origin of this fact is a deep one, and can be found, for instance, in the proposition calculus of Jauch and Piron. The identity element corresponds to the tautologic proposition asserting that the system exists.

---

†Another possible realization is to consider $\mathfrak{A}_1$ being the set of matrices of the form

$$\begin{pmatrix} \alpha & a \\ 0 & \alpha + a \end{pmatrix} \qquad \alpha \in \mathbb{C}, a \in \mathfrak{A}$$

with the usual matrix multiplication rule. The identity element of $\mathfrak{A}_1$ is then $\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and $\mathfrak{A}$ may be identified to the set of all matrices of the form $\begin{pmatrix} 0 & a \\ 0 & a \end{pmatrix}$, $a \in \mathfrak{A}$.

From a mathematical point of view, one has to remark that the addition of an identity element may change the topological structure of the algebra. It is, in fact, the analogous problem to the compactification of a set (for instance, adding the point at ∞ to $\mathbb{R}$).

I.A.9. Examples:

1) Let $\mathfrak{X}$ be a topological, locally compact space. Let $f(x)$ be a complex continuous function on $\mathfrak{X}$. It is said to vanish at infinity, if, $\forall\, \varepsilon > 0$, the set of all points $x \in \mathfrak{X}$, such that $|f(x)| > \varepsilon$, is contained in a compact of $\mathfrak{X}$. The set $\mathfrak{A}(\mathfrak{X})$ of all complex continuous functions on $\mathfrak{X}$, vanishing at ∞, is an algebra.

This algebra possesses an identity element, if, and only if, $\mathfrak{X}$ is compact.

2) Let $\mathfrak{H}$ be a Hilbert space, $\mathfrak{B}(\mathfrak{H})$, the algebra of all bounded operators on $\mathfrak{H}$. As a particular case, for $\mathfrak{H}$ of finite dimension n, $\mathfrak{B}(\mathfrak{H})$ may be identified with the algebra $\mathbb{M}_n(\mathbb{C})$ of all n by n matrices with complex coefficients.

3) Let $\mathfrak{D}$ (respectively $\mathcal{S}$) the space of Schwartz test functions on $\mathbb{R}^1$.

The set of operators of the form

$$\sum_{k=0}^{n} p_k(x)\, \frac{d^k}{dx^k}\,, \quad n < \infty$$

$p_k(x)$ being real polynomials, is an algebra. The identity element is $D^0 = 1$.

I.A.10. Definition: A subset $\mathfrak{I}$ of an algebra $\mathfrak{A}$ is called a left ideal [resp. a right ideal] if

(i) $\mathfrak{I}$ is a vector subspace of $\mathfrak{A}$

(ii) $x \in \mathfrak{I}$ and $a \in \mathfrak{A}$ implies that $ax \in \mathfrak{I}$ (resp. $xa \in \mathfrak{I}$), what we symbolically write as

$$\mathfrak{A}\mathfrak{I} \subset \mathfrak{I} \quad [\text{resp. } \mathfrak{I}\mathfrak{A} \subset \mathfrak{I}]$$

If $\mathfrak{I}$ is both a right and a left ideal, it is called a two-sided ideal.

Remarks: 1) Any algebra contains the ideals $\mathfrak{A}$ and $\{0\}$.

2) Any ideal is a subalgebra.

3) An ideal which is different from $\mathfrak{A}$ is said to be proper.

4) A proper ideal of $\mathfrak{A}$ is said to be minimal if it is different from $\{0\}$ and does not contain properly any ideal of the same type other than $\{0\}$. It is called maximal if it is not properly contained in an ideal of the same type other than $\mathfrak{A}$.

5) A proper ideal cannot contain the identity element.

I.A.11. Definition: An algebra which does not possess any proper two-sided ideal other than $\{0\}$ is said to be simple.

I.A.12. Examples:

1) The algebra of all n by n matrices, $\mathbb{M}_n(\mathbb{C})$, $n < \infty$, is simple.

2) In $\mathbb{M}_2(\mathbb{C})$,

$$\left\{\begin{pmatrix} 0 & a \\ 0 & b \end{pmatrix}\right\} = \mathfrak{I}_g \text{ is a left ideal}$$

and

$$\left\{\begin{pmatrix} 0 & 0 \\ a & b \end{pmatrix}\right\} = \mathfrak{I}_d \text{ is a right ideal}$$

3) In the algebra of all 2 by 2 matrices of the form $\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}$, the set

$$\mathfrak{I} = \left\{\begin{pmatrix} 0 & \alpha \\ 0 & 0 \end{pmatrix}\right\}$$

is a two-sided ideal.

4) In the algebra $\mathfrak{B}(\mathfrak{H})$ of all bounded operators on a Hilbert space, the set $\mathfrak{F}$ of all operators of finite rank is a two-sided ideal.

5) Again in $\mathfrak{B}(\mathfrak{H})$, the set $\mathfrak{C}$ of all compact operators is a two-sided ideal. Further, $\mathfrak{F}$ is minimal and $\mathfrak{C}$ maximal, $\mathfrak{F} \subset \mathfrak{C}$.

I.A.13. Definition: An element $i \in \mathfrak{A}$ is said to be idempotent if $ii = i$.

I.A.14. Example: in $\mathbb{M}_2(\mathbb{C})$, $i = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$ is an idempotent.

I.A.15. Definition: Let $\mathfrak{A}$ be an algebra over $\mathbb{C}$. An involution in $\mathfrak{A}$ is a mapping of $\mathfrak{A}$ onto itself: $a \mapsto a^*$, such that

(i) $(a^*)^* = a$

(ii) $(a + b)^* = a^* + b^*$

(iii) $(\lambda a)^* = \bar{\lambda} a^*$, $\lambda \in \mathbb{C}$, $\bar{\lambda}$ complex conjugate of λ

(iv) $(ab)^* = b^* a^*$

$\forall\, a, b \in \mathfrak{A}$.

An algebra with an involution is said to be a *-algebra. Naimark calls it a symmetric algebra.

a^* is called the adjoint of a.

a is called self-adjoint if $a^* = a$, normal if $aa^* = a^*a$.

A subset $\mathfrak{S} \subset \mathfrak{A}$ is said to be self-adjoint if $\mathfrak{S}^* = \mathfrak{S}$ ($\mathfrak{S}^* = \{a^* | a \in \mathfrak{S}\}$).

Remark: A self-adjoint ideal in a *-algebra is always two-sided.

I.A.16. Definition: A mapping $\beta: \mathfrak{A}_1 \xrightarrow{\text{into}} \mathfrak{A}_2$ is called a homomorphism (more exactly, an algebra homomorphism) if

(i) $\beta(\lambda a) = \lambda \beta(a)$

(ii) $\beta(a + b) = \beta(a) + \beta(b)$

(iii) $\beta(a \cdot b) = \beta(a) \cdot \beta(b)$

It is a *-homomorphism, if $\mathfrak{A}_1$ and $\mathfrak{A}_2$ are *-algebras and if

(iv) $\beta(a^*) = \beta(a)^*$

Remarks: The inverse image of the $\{0\}$ element of $\mathfrak{A}_2$ in $\mathfrak{A}_1$ is called the kernel of the homomorphism β. The kernel $\mathfrak{I}_\beta$ of a homomorphism β is a two-sided ideal: indeed, if $\beta(a) = \beta(b) = 0$, then $\beta(\lambda a) = \beta(a + b) = 0$, and $\beta(ac) = \beta(a)\beta(c) = 0$, $\beta(ca) = 0 \ \forall\, c \in \mathfrak{A}_1$.

If the kernel consists only of the $\{0\}$ element of $\mathfrak{A}_1$, the mapping is said to be faithful. A faithful homomorphism of $\mathfrak{A}_1$ onto $\mathfrak{A}_2$ is called an isomorphism. A faithful homomorphism of $\mathfrak{A}_1$ onto $\mathfrak{A}_1$ is called an automorphism. A representation of an algebra is a homomorphism of the algebra into the algebra of the linear operators on a certain vector space.

In these lectures, we shall only consider *-homomorphisms, *-representations, etc., and thus we shall omit the *-.

I.A.17. Example: Any algebra $\mathfrak{A}$ has a representation in the algebra of linear operators defined on $\mathfrak{A}$ (considered as a vector space). Indeed, to each element $a \in \mathfrak{A}$, one associates the mapping A_a of the vector space $\mathfrak{A}$ by defining $A_a x = ax$, $\forall\, x \in \mathfrak{A}$. One easily verifies that A_a is a representation of $\mathfrak{A}$. Furthermore, this representation is faithful if $\mathfrak{A}$ possesses an identity element. This representation is called the left regular representation.

I.A.18. Definition: A function $\mathfrak{A} \to \mathbf{R}$, $a \mapsto \|a\|$, $a \in \mathfrak{A}$ is called a prenorm on $\mathfrak{A}$ if

(i) $\|a\| \geq 0$

(ii) $\|a + b\| \leq \|a\| + \|b\|$

(iii) $\|\alpha a\| = |\alpha| \cdot \|a\|$

(iv) If $\mathfrak{A}$ possesses an identity element e, $\|e\| = 1$. This function is called a norm, if i) is replaced by

(i′) $\|a\| \geq 0$, $\|a\| = 0 \Leftrightarrow a = 0$.

An algebra on which a norm is defined is said to be a normed algebra.

I.A.19. Definition: A normed algebra, which is complete under its norm (i.e. is a Banach space) is called a Banach algebra.

If $\mathfrak{A}$ is a normed *-algebra, the condition $\|a\cdot b\| \leq \|a\|\cdot\|b\|$ implies that $\|a^*a\| \leq \|a^*\|\cdot\|a\|$. If we impose the condition $\|a^*a\| = \|a\|^2$, we get that $\|a\| = \|a^*\|$, but the reverse need not be true. A Banach algebra satisfying $\|a^*\| = \|a\|$ (but not necessarily $\|a^*a\| = \|a\|^2$, $\forall\, a \in \mathfrak{A}$) is called normed symmetric by Naimark, and a B*-algebra by Dixmier.

A Banach *-algebra, satisfying $\|a^*a\| = \|a\|^2$ is called a B*-algebra by Rickart, and a C*-algebra by Dixmier (it is also sometimes called an abstract C*-algebra).

Note the confusion existing in the literature about these denominations! In order to avoid any conflict, we shall call C*-algebra an algebra satisfying $\|a^*a\| = \|a\|^2$. These algebras are essentially the only ones which we shall encounter in these lectures.

I.B. Von Neumann Algebras

We introduce here a particular class of C*-algebras which enjoy remarkable properties. They often are used in physical applications and in the theory of C*-algebras themselves.

I.B.1. Let $\mathfrak{H}$ be a complex Hilbert space (not necessarily separable), and denote by $\mathfrak{B}(\mathfrak{H})$ the algebra of all bounded operators on $\mathfrak{H}$. Clearly, $\mathfrak{B}(\mathfrak{H})$ is a C*-algebra.

Let $\mathfrak{M}$ be an arbitrary subset of $\mathfrak{B}(\mathfrak{H})$. We denote by $\mathfrak{M}'$ the set of all elements of $\mathfrak{B}(\mathfrak{H})$ which commute with all elements of $\mathfrak{M}$:

$$\mathfrak{M}' = \{a' \mid a' \in \mathfrak{B}(\mathfrak{H}),\ a'a = aa',\ \forall\, a \in \mathfrak{M}\}$$

$\mathfrak{M}'$ is called the commutant of $\mathfrak{M}$. One easily sees that the commutant of any subset of $\mathfrak{B}(\mathfrak{H})$ is always an algebra, and that this algebra always contains an identity element.

In a similar way, one defines the double commutant, or bicommutant $\mathfrak{M}''$ of $\mathfrak{M}$, $\mathfrak{M}'' = (\mathfrak{M}')'$.

It is immediately clear that $\mathfrak{M}'' \supset \mathfrak{M}$. If $\mathfrak{M}_1 \subset \mathfrak{M}_2$ then also $\mathfrak{M}_1' \supset \mathfrak{M}_2'$. From these two relations follows that

$$\mathfrak{M}' = \mathfrak{M}''' = \mathfrak{M}^{(5)} = \dots$$

$$\mathfrak{M}'' = \mathfrak{M}^{(4)} = \mathfrak{M}^{(6)} = \dots$$

Finally, if $\mathfrak{M} = \mathfrak{M}^*$, then $\mathfrak{M}'$ is a *-algebra.

I.B.2. Definition: A von Neumann algebra $\mathfrak{A}$ on $\mathfrak{H}$, is a *-subalgebra $\mathfrak{A}$ of $\mathfrak{B}(\mathfrak{H})$ such that $\mathfrak{A} = \mathfrak{A}''$.
Example: The commutant $\mathfrak{M}'$ of any self-adjoint subset $\mathfrak{M} = \mathfrak{M}^*$ of $\mathfrak{B}(\mathfrak{H})$ is always a von Neumann algebra.
Remark: Von Neumann algebras are also called rings of operators, or W*-algebras. Sakai defines a W*-algebra as being a C*-algebra which is the dual of some Banach space. He then shows that such an algebra has a faithful representation as a von Neumann algebra on some Hilbert space. We shall therefore reserve the term W*-algebra for an algebra which is isomorphic to a von Neumann algebra, and when we speak of a particular von Neumann algebra, it will always be understood that it carries with it the particular Hilbert space on which it is defined. This is necessary because, as we shall see, an isomorphism of a von Neumann algebra into a $\mathfrak{B}(\mathfrak{H}')$ is not necessarily a von Neumann algebra.

Clearly, a von Neumann algebra is also a C*-algebra.
Example: Any C*-algebra on a finite dimensional Hilbert space is also a von Neumann algebra.

I.B.3. Proposition: The intersection of an arbitrary family of von Neumann algebras is a von Neumann algebra.

Proof: Let $\mathfrak{A} = \bigcap_{i\in I} \mathfrak{A}_i$. To say that $x \in \mathfrak{A}$ amounts to saying that x commutes with the elements of $\mathfrak{A}_i'$, thus with $\bigcup_{i\in I} \mathfrak{A}_i'$ and hence with $(\bigcup \mathfrak{A}_i')''$.

Corollary: The center $\mathfrak{Z}$ of a von Neumann algebra is

$$\mathfrak{Z} = \mathfrak{A} \cap \mathfrak{A}'$$

and is thus an abelian von Neumann algebra.

If $\mathfrak{A}$ is abelian, then $\mathfrak{A} = \mathfrak{A}'$ and $\mathfrak{Z} = \mathfrak{A}$.

I.B.4. Definition: A von Neumann algebra, the center of which contains only the scalar multiples of the identity, is called a factor.
Example: $\mathfrak{B}(\mathfrak{H})$ is a factor, indeed $\mathfrak{B}(\mathfrak{H})' = \{\lambda I\} = \mathfrak{Z}$.

I.B.5. Definition: Let $\mathfrak{A}$ be a *-subalgebra of $\mathfrak{B}(\mathfrak{H})$. A vector $x \in \mathfrak{H}$ is said to be cyclic (in French: "totalisateur") with respect to $\mathfrak{A}$, if the set $\{ax \mid a \in \mathfrak{A}\}$ is dense in $\mathfrak{H}$.

$x \in \mathfrak{H}$ is said to be a separating vector for $\mathfrak{A}$, if the conditions $a \in \mathfrak{A}$, $ax = 0$, imply $a = 0$.

In an evident way, this definition can be extended to subsets σ of $\mathfrak{H}$.

Proposition: Let σ be a subset of $\mathfrak{H}$, $\mathfrak{A}$ a *-subalgebra of $\mathfrak{B}(\mathfrak{H})$. Let further p be the projector on the closure of the set $\mathfrak{A}\sigma = \{ax \mid a \in \mathfrak{A}, x \in \sigma\}$. Then $p \in \mathfrak{A}'$, and is the smallest projection of $\mathfrak{A}'$ having σ contained in its range.

Proof: It is clear, that for any $t \in \mathfrak{A}$, $t\mathfrak{A}\sigma \subset \mathfrak{A}\sigma$, and therefore, by continuity (t being bounded), $tp\mathfrak{H} \subset p\mathfrak{H}$. But this last relation means exactly that $ptp = tp$. From this follows however, using that also $pt^*p = t^*p$

$$pt = (t^*p^*)^* = (pt^*p)^* = ptp = tp$$

and thus, $p \in \mathfrak{A}'$.

Let now $\overline{\mathfrak{A}\sigma} = p\mathfrak{H}$, and suppose $\exists\, p' \in \mathfrak{A}'$, $p'\mathfrak{H} \supset \mathfrak{A}\sigma$. We have that

$$p'\mathfrak{A}\sigma = \mathfrak{A}p'\sigma = \mathfrak{A}\sigma,$$

by hypothesis, but this means $p'p = p$, QED.

Theorem: Let $\mathfrak{A}$ be a von Neumann algebra. x is cyclic for $\mathfrak{A}$ if, and only if, x is separating for $\mathfrak{A}'$.

Proof: "$\Leftarrow$" Let p be the projector on the closure of $\mathfrak{A}x$. From the preceding proposition, we have that $p \in \mathfrak{A}'$. Because $I \in \mathfrak{A}'$, $(I - p)x = 0$. But by hypothesis, x is separating for $\mathfrak{A}'$, and hence $I - p = 0$, $p = I$, or $p\mathfrak{H} = \mathfrak{H}$, which means that the closure of $\mathfrak{A}x$ is $\mathfrak{H}$, or, in other words, that x is cyclic for $\mathfrak{A}$.

"$\Rightarrow$" The conditions $a' \in \mathfrak{A}'$, $a'x = 0$, imply that $a'ax = aa'x = 0$, $\forall\, a \in \mathfrak{A}$. But that means that $a' = 0$, since it is a bounded operator which vanishes on a dense set of vectors, namely $\mathfrak{A}x$. Thus x is separating for $\mathfrak{A}'$.

I.B.6. Theorem: Let $\mathfrak{A}$ be a C*-algebra. Any element $a \in \mathfrak{A}$ is a linear combination of two self-adjoint operators, or of four unitary operators, belonging to the algebra.

Proof: By construction:

Self-adjoint: put $a_1 = \frac{1}{2}(a+a^*)$, $a_2 = \frac{i}{2}(a^*-a)$, $a = a_1 + ia_2$

Unitary: a being bounded, one can choose λ_1 and λ_2 such that $a_i = \lambda_i a_i'$ with $\|a_i'\| < 1$, $i = 1,2$. Put now $u_i = a_i' - (1 - a_i'^2)^{\frac{1}{2}}$ u_i is clearly unitary, since $a_i^* = a_i$. It remains to show that $u_i \in \mathfrak{A}$, which amounts to proving that $(1 - a_i'^2)^{\frac{1}{2}} \in \mathfrak{A}$. This is true, since the Taylor series

$$1 - \frac{1}{2}a_i'^2 - \frac{1}{2\cdot 4}a_i'^4 - \frac{1\cdot 3}{2\cdot 4\cdot 6}a_i'^6 - \frac{1\cdot 3\cdot 5}{2\cdot 4\cdot 6\cdot 8}a_i'^8 - \ldots = (1 - a_i'^2)^{\frac{1}{2}}$$

converges in norm, because $\|a_i'\| < 1$, and $(1 - a_i^2)^{\frac{1}{2}} \in \mathfrak{A}$ because the algebra is complete in norm.

Put then

$$a = \tfrac{1}{4}\{\lambda_1(u_1 + u_1^*) + i\lambda_2(u_2 + u_2^*)\}.$$

I.B.7. Definition: An operator is said to be closed if $tx_i \to y$ and $x_i \to x$ imply $tx = y$.

Theorem: If t is closed, one may write $t = wk$, where w is a partially isometric operator, and k is self-adjoint and non-negative. This decomposition is unique, and is called the polar decomposition.

Definition: A closed operator t is said to be affiliated to a von Neumann algebra $\mathfrak{A}$ ($t\eta\mathfrak{A}$) if it commutes with all operators of $\mathfrak{A}'$, i.e. if $\forall\, b' \in \mathfrak{A}'$, $b't \subset tb'$.

Proposition: Let t be a closed operator and $t = wk$ its polar decomposition. Then $t\eta\mathfrak{A}$ if and only if all spectral projectors of k belong to $\mathfrak{A}$ and $w \in \mathfrak{A}$.

Proof: It is enough to show that in the case of a unitary operator $u' \in \mathfrak{A}'$, $u't \subset tu'$, $u'^{-1} \subset tu'^{-1}$ imply that $u'hu'^{-1} = h$ and $u'wu'^{-1} = w$.

We have that $u'tu'^{-1} \subset t \subset u'tu'^{-1}$ by hypothesis; from this follows that $u'tu'^{-1} = t = wh = u'wu'^{-1}u'hu'^{-1}$ and the proposition follows from the unicity of the polar decomposition, the inverse being trivial.

Corollary: Let $t\eta\mathfrak{A}$, $t = t^*$. Then for all continuous real functions f, $f(t)\eta\mathfrak{A}$.

I.C. Topological Considerations

I.C.1. We start by defining five different topologies on $\mathfrak{B}(\mathfrak{H})$.

Uniform topology: It is the topology induced by the norm of the operators. A basis of neighbourhoods of an element $t \in \mathfrak{B}(\mathfrak{H})$ is given by

$$U_\varepsilon(t) = \{a \mid a \in \mathfrak{B}(\mathfrak{H}),\ \|t - a\| < \varepsilon\}$$

Strong topology: A basis of neighbourhoods, indexed by ε and σ is induced by

$$U_{\varepsilon,\sigma}(t) = \{a \mid a \in \mathfrak{B}(\mathfrak{H}),\ \|(t - a)x\| < \varepsilon\ \ \forall\, x \in \sigma\}$$

where σ is a finite subset of $\mathfrak{H}$.

Weak topology:

$$u_{\varepsilon,\sigma_1,\sigma_2}(t) = \{a \mid a \in \mathfrak{B}(\mathfrak{H}), |(x,(t-a)y| < \varepsilon \quad \forall x \in \sigma_1 \quad \forall y \in \sigma_2\}$$

where σ_1 and σ_2 are finite subsets of $\mathfrak{H}$.

Ultrastrong topology: (sometimes called strongest topology)

$$u_{\varepsilon,\{x_i\}}(t) = \{a \mid a \in \mathfrak{B}(\mathfrak{H}), \sum_{i=1}^{\infty} \|(a-t)x_i\|^2 < \varepsilon\}$$

where x_i is an arbitrary sequence of vectors of $\mathfrak{H}$ such that $\sum_{i=1}^{\infty} |x_i|^2 < \infty$ (thus a filtering set $\{t_i\}$ converges towards t in the ultrastrong topology if, $\forall$ sequences $\{x_k\}$ such that $\Sigma|x_i|^2 < \infty$ $\{\Sigma_k t_i x_k\}$ converges towards $\Sigma_k t x_k$).

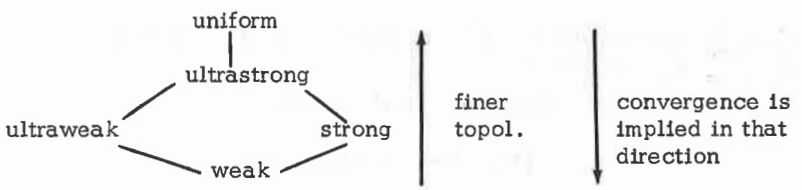
Ultraweak topology: We define

$$u_{\varepsilon,\{x_i\},\{y_i\}}(t) = \{a \mid a \in \mathfrak{B}(\mathfrak{H}), \sum_{k=1}^{\infty} |(x_i, (a-t)y_i)| < \varepsilon\}$$

where $\{x_i\}$, $\{y_i\}$ are arbitrary sequences of vectors of $\mathfrak{H}$ such that $\Sigma|x_i|^2 < \infty$ $\Sigma|y_i|^2 < \infty$.

I.C.2. We are not going to show that all these topologies are actually distinct if $\mathfrak{H}$ is infinite dimensional. It is clear for the uniform, strong and weak topologies. For the ultrastrong topology, one can show that it does coincide with the strong one bounded sets of $\mathfrak{B}(\mathfrak{H})$ (Dixmier 1957, p. 36) and similarly of the ultra weak and the weak topologies. One can construct an example (Dixmier, 1957, p. 48) of a set in $\mathfrak{B}(\mathfrak{H})$ whose strong and ultrastrong closure are not identical.

In fact, it is practically evident that we must have the following diagram



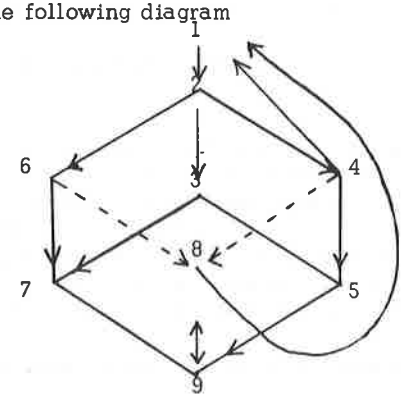
I.C.3. If we now consider a subalgebra of $\mathfrak{B}(\mathfrak{H})$ with identity element, we already know that it is a C*-algebra if it is closed in the uniform topology. If we want to close it under other topologies, it is, a priori, conceivable that different topologies will lead to different algebras. Under these circumstances, the famous density theorem of von Neumann is truly remarkable. In order to state and prove it, we will still need to define what is a "unit ball."

Definition: Let $\mathfrak{A}$ be a normed algebra, the unit ball $\mathfrak{A}_1$ is the set of all those operators of $\mathfrak{A}$ the norm of which does not exceed one.

Theorem: (density theorem of von Neumann) Let $\mathfrak{A}$ be a *-subalgebra of $\mathfrak{B}(\mathfrak{H})$, with identity element. The following nine conditions are equivalent

1) $\mathfrak{A} = \mathfrak{A}''$ (i.e. $\mathfrak{A}$ is a von Neumann algebra)
2) $\mathfrak{A}$ is weakly closed
3) $\mathfrak{A}_1$ is weakly closed
4) $\mathfrak{A}$ is strongly closed
5) $\mathfrak{A}_1$ is strongly closed
6) $\mathfrak{A}$ is ultraweakly closed
7) $\mathfrak{A}_1$ is ultraweakly closed
8) $\mathfrak{A}$ is ultrastrongly closed
9) $\mathfrak{A}_1$ is ultrastrongly closed

Proof: The sequence of the elements of the proof can be seen from the following diagram



→ used in the proof (4 → 1) shall be used in proving 8-1

--→ other possible proof path

(1 → 2): hypothesis: $\mathfrak{A} = \mathfrak{A}''$. To prove: Let $\{a_i \in \mathfrak{A}\}$ and $\underset{i \to \infty}{w - \lim} a_i = a$, then $a \in \mathfrak{A}$.

Proof: Let $a' \in \mathfrak{A}'$, then $a'a_i = a_i a'$, that is

$$|(x_1, (aa' - a'a)x_2)| < 2\varepsilon$$

because

$$|(x_1, (aa' - a'a)x_2)| = |(x_1, (a - a_i)a' + a'(a_i - a)x_2)| \leq$$

$$\leq |(x_1, (a - a_i)a'x_2)| + |(x_1, a'(a - a_i)x_2| < 2\epsilon$$

for a certain $i \geq i_0$ $(2 \to 4)$, $(2 \to 6)$, $(3 \to 7)$, $(5 \to 9)$, $(7 \to 9)$, and also $(3 \dashrightarrow 5)$, $(4 \dashrightarrow 8)$, $(6 \dashrightarrow 8)$ are true because of the partial ordering of the topologies, as seen above. $(2 \to 3)$, $(4 \to 5)$, $(8 \to 9)$, and also $(6 \dashrightarrow 7)$ are trivial by definition of the unit ball.

(4 → 1): Hypothesis: For any sequence $\{a_i\} \in \mathfrak{A}$, if $\{a_i\}$ converges strongly towards a, then $a \in \mathfrak{A}$. One has to show in any strong neighbourhood of each element $a'' \in \mathfrak{A}''$, there is an element of $\mathfrak{A}$.
Proof: Let $x \in \mathfrak{H}$ and p be the projector on the closure of $\mathfrak{A}x$. As shown in a theorem above, $p \in \mathfrak{A}'$; that means, however: $a''x = a''px = pa''x$; $a''x$ is contained in the closure of the set $\mathfrak{A}x$, $\forall x$, which is saying exactly that it is a strong limit of elements of $\mathfrak{A}$.

(8 → 1): Hypothesis: $\mathfrak{A}$ is ultrastrongly closed. Given $\epsilon > 0$, $\{x_i\}$ such that

$$\sum_{i=1}^{\infty} |x_i|^2 < \infty, \; x_i \in \mathfrak{H}, \; t \in \mathfrak{A}'' \quad ,$$

we want to show that there exists $s \in \mathfrak{A}$ such that

$$\sum_{i=1}^{\infty} |(s - t)x_i|^2 < \epsilon$$

and for that, we shall reduce the proof to the case (4-1).

Let $\mathfrak{K}$ be a family of mutually orthogonal Hilbert spaces $\mathfrak{K}_i$, and each $\mathfrak{K}_i$ being isomorphic to $\mathfrak{H}$, the isomorphism being given by $U_i\mathfrak{H} = \mathfrak{K}_i$. Let $y_i = U_i x_i$, it follows that

$$\sum_{i=1}^{\infty} |y_i|^2 < \infty$$

Let us write

$$y \text{ for } \{y_i\} \text{ and } |y|^2 = \sum_{i=1}^{\infty} |y_i|^2$$

which makes a Hilbert space out of $\mathfrak{K}$. For any $b \in \mathfrak{B}(\mathfrak{H})$ we define a corresponding $\tilde{b}$ acting on $\mathfrak{K}$ by

$$\tilde{b}y = \{U_i \, b \, U_i^* \, y_i\}$$

If a runs over $\mathfrak{A}$, the corresponding $\tilde{a}$ form a *-algebra $\tilde{\mathfrak{A}}$ on $\mathfrak{K}$.
If $t \in \mathfrak{A}''$, it is clear that $\tilde{t} \in \tilde{\mathfrak{A}}''$. We can now apply the argument used in (4 → 1), indeed there exists $\tilde{s} \in \tilde{\mathfrak{A}}$ such that $|(\tilde{s} - \tilde{t})y|^2 < \varepsilon$, which means

$$|(\tilde{s} - \tilde{t})y|^2 = \sum_{i=1}^{\infty} |(s - t)x_i|^2 < \varepsilon$$

(9 → 8): This is a general property of Banach spaces; see for instance Bourbaki, Espaces vectoriels topologiques, Chap. V, Section 5.1, cor. 1 of prop. 3.

I.D. Linear Functionals

From the standard formulation of Quantum Mechanics, we know that one assigns a number to each pair consisting of an observable and of a state, namely the expectation value of the observable in the considered state. One can therefore consider a state as being simply a mapping of the set of observables into the set of complex numbers, that is, a functional. We shall restrict our attention to linear functionals, because the linearity of these functionals is the mathematical expression of the principle of superposition.

I.D.1. Definition: A linear functional on an algebra $\mathfrak{A}$ is a mapping f of $\mathfrak{A}$ into $\mathbb{C}$ such that

$$f(\alpha a + \beta b) = \alpha f(a) + \beta f(b) \quad \forall \alpha, \beta \in \mathbb{C} \quad \forall a, b \in \mathfrak{A}$$

Note that we do not use here the algebraic structure, only the vector space structure.
Definition: A linear functional f on a Banach algebra $\mathfrak{A}$ is said to be uniformly continuous if, for any sequence $\{a_n\}$, $a_n \in \mathfrak{A}$, converging uniformly towards an element $a \in \mathfrak{A}$, $f(a_n)$ converges towards $f(a)$. (We shall say simply "continuous" in what follows.)

I.D.2. Definition: A linear functional on a *-algebra $\mathfrak{A}$ is said to be positive, if $f(t^*t) \geq 0$, $\forall t \in \mathfrak{A}$.

We want to deduce some useful properties of positive linear functionals. Let $\mathfrak{A}$ be a *-algebra, f a positive linear functional,

$a, b \in \mathfrak{A}$, $\alpha, \beta \in C$. Further, put $t = \alpha a + \beta b$. By definition, we have that

$$0 \leq f(t^*t) = |\alpha|^2 f(a^*a) + \bar{\alpha}\beta f(a^*b) + \alpha\bar{\beta} f(b^*a) + |\beta|^2 f(b^*b)$$

$f(a^*a)$, $f(b^*b)$ and $f(t^*t)$ being real, it follows that

$$\bar{\alpha}\beta f(a^*b) + \alpha\bar{\beta} f(b^*a)$$

is real too.

Putting $\alpha = \beta = 1$, it follows that

$$\text{Im } f(a^*b) + \text{Im } f(b^*a) = 0$$

Putting now $\alpha = 1$, $\beta = i$, we get that

$$\text{Re } f(a^*b) - \text{Re } f(b^*a) = 0$$

from which it follows that

$$f(a^*b) = \overline{f(b^*a)}$$

and in particular, for an algebra with identity

$$f(a^*) = \overline{f(a)}$$

If we now put $\beta = f(b^*a)$, and take α real, we get

$$\alpha^2 f(a^*a) + 2\alpha|\beta|^2 + |\beta|^2 f(b^*b) \geq 0$$

and if we now consider this expression as a quadratic form in α, we know that it is positive if and only if $|\beta|^4 - |\beta|^2 f(b^*b)\, f(a^*a) \leq 0$.

From this follows, however, in the case $\beta \neq 0$,

$$|\beta|^2 \leq f(a^*a)\, f(b^*b)$$

or

$$|(f(a^*b))|^2 = f(a^*b)\, f(b^*a) \leq f(a^*a)\, f(b^*b)$$

(Inequality of Schwartz, Cauchy, Bunyakovsky, etc.)

The inequality is of course also valid for $\beta = 0$.

I.D.3. Definition: Let $\mathfrak{A}$ be a Banach algebra. We call dual of $\mathfrak{A}$ the set $\mathfrak{A}^*$ of all linear functionals on $\mathfrak{A}$, continuous for the topology defined by the norm of $\mathfrak{A}$.

Proposition: $\mathfrak{A}^*$ is a Banach space with the norm $\|f\| = \sup_{\|a\| \leq 1} |f(a)|$.

Proof: The linearity is trivial; we only have to verify that the axioms of the norm are satisfied and that $\mathfrak{A}^*$ is complete:

$$\|f\| = 0 \Leftrightarrow \sup_{\|a\| = 1} |f(a)| = 0 \Leftrightarrow |f(a)| = 0 \quad \forall\, a \in \mathfrak{A}$$

$$\Leftrightarrow f = 0$$

$$\|\lambda f\| = \sup_{\|a\| = 1} |\lambda f(a)| = \sup |\lambda|\, |f(a)| = |\lambda|\, \|f\|$$

$$\|f_1 + f_2\| = \sup_{\|a\| = 1} |(f_1 + f_2)(a)| \leq \sup \{|f_1(a)| + |f_2(a)|\}$$

$$\leq \|f_1\| + \|f_2\|$$

$\mathfrak{A}^*$ is complete: If

$$\|f_n - f_m\| < \varepsilon \quad \forall\, n, m > N(\varepsilon)$$

then

$$\|f_n - f_m\| = \sup_{\|a\| = 1} |(f_n - f_m)(a)| \leq \varepsilon\, \|a\|$$

that is, for a fixed a, $\{f_n(a)\}$ converges. One therefore defines

$$f = \lim_{n \to \infty} f_n$$

so that

$$f(a) = \lim_{n \to \infty} f_n(a) \quad \forall\, a \in \mathfrak{A}$$

By continuity, f is linear and one easily sees that it is continuous; it therefore belongs to $\mathfrak{A}^*$. Q.E.D.

I.D.4. Theorem: Let $\mathfrak{A}$ be a C*-algebra with identity element e. Any positive linear functional on $\mathfrak{A}$ is bounded and continuous with respect to the norm of $\mathfrak{A}$), i.e.

$$|f(a)| \leq f(e)\,\|a\|\ ,\quad \|f\| = f(e)$$

Proof: Suppose first $a = a^*$, $\|a\| \leq 1$. The Taylor expansion of $(1 - a)^{\frac{1}{2}}$ converges in norm in $\mathfrak{A}$ towards an element $b \in \mathfrak{A}$, since $\mathfrak{A}$ is complete in norm. From the definition of a C*-algebra follows that $\|a\| = \|a^*\|$ and thus b will be self-adjoint too. By construction $b^*b = b^2 = e - a$; f being a positive functional, we get that $f(e - a) = f(b^*b) \geq 0$, and by linearity, $f(e) \geq f(a)$. We could repeat the same argument for $-a$, and thus get that $f(-a) \leq f(e)$, and therefore $|f(a)| \leq f(e)$. If $a^* = a$ and $\|a\| \geq 1$, one can repeat the argument for a, putting $a = a/(\|a\|)$, and use linearity at the end.

If now a is arbitrary, we first consider a^*a which is self-adjoint, we thus get $f(a^*a) \leq f(e)\cdot\|a^*a\| = f(e)\cdot\|a\|^2$. On the other hand, by the Schwarz inequality,

$$|f(a)|^2 = |f(ea)|^2 \leq f(e)\,f(a^*a) \leq f(e)^2\,\|a\|^2$$

$$\Rightarrow |f(a)| \leq f(e)\,\|a\| \quad \forall\, a \in \mathfrak{A}$$

$$\|f\| = \sup_{\|a\| = 1} |f(a)| \leq f(e)$$

but the equality is attained for $a = e$, thus $\|f\| = f(e)$.

I.D.5. Definition: Let f_1, f_2 be positive linear functionals on a *-algebra. f_1 is said to majorize f_2 if $f_1 - f_2$ is a positive functional.

I.D.6. Theorem: Let $\mathfrak{A}$ be a C*-algebra with identity e, f a linear functional. The condition $\|f\| = f(e)$ implies that f is positive.

The condition is therefore both necessary and sufficient (from the preceding theorem).

Proof: Put $f = f_1 + if_2$, where f_1 and f_2 are hermitian. We suppose that $f(e) = 1$, with $f_2(e) = 0$. Let $h \in \mathfrak{A}$, $h = h^*$. Put $u = \lambda e - ih$, where $\lambda \in \mathbf{R}$, we get

$$\|u\|^2 = \|\lambda^2 e + h^2\| \leq \lambda^2 + \|h\|^2$$

and on the other hand

$$|f(u)|^2 - |\lambda - if_1(h) + f_2(h)|^2 = \lambda^2 + 2\lambda f_2(h) + f_2(h)^2 + f_1(h)^2$$

thus

$$|f(u)|^2 - 2\lambda f_2(h) - \lambda^2 = f_1(h)^2 + f_2(h)^2 \geq 0 ,$$

adding

$$\|h\|^2 + \lambda^2 \geq \|u\|^2$$

we get

$$\|u\|^2 \leq |f(u)|^2 - 2\lambda f_2(h) + \|h\|^2 \leq \|u\|^2 - 2\lambda f_2(h) + \|h\|^2$$

and hence

$$2\lambda f_2(h) \leq \|h\|^2 \quad \forall \lambda \in \mathbb{R}$$

which means $f_2(h) = 0$, thus f is hermitian.

We write $x \geq 0$ if $\exists\ t \in \mathfrak{A}$ such that $x = t^*t$, and in an analogous way, $x \leq y$ if $y - x \geq 0$. Let h be given $0 \leq h \leq e$ so that $0 \leq e - h \leq e$ and therefore $\|e - h\| \leq 1$. Suppose $f(h) < 0$, that is, that f is not positive. We get

$$1 = f(e) = f(e - h) + f(h) < f(e - h) \leq \|e - h\| \leq 1$$

which is a contradiction.

I.D.7. Proposition: Let $\mathfrak{A}_1$ be a subalgebra of a normed algebra $\mathfrak{A}$ and f_1 a continuous linear functional on $\mathfrak{A}_1$. Then, there exists a continuous linear functional f on $\mathfrak{A}$ such that $f_1(a) = f(a)\ \forall\ a \in \mathfrak{A}_1$. If f_1 is bounded on $\mathfrak{A}_1$, with norm $\|f_1\|$, then f can be chosen such that $\|f\| = \|f_1\|$. If $\mathfrak{A}_1$ and $\mathfrak{A}$ are *-algebras and f_1 is positive, f can also be chosen positive.

Proof: This is the Hahn-Banach theorem: for a proof, see Naimark, Ch. 1, p.16, cor. 1, p. 17; or Day, Normed Linear Spaces, Springer 1958, Ch. 1, p.9; Banach, Op. linéaires, Warsaw 1932, Section 2, Ch. 1.

I.D.8. Definition: A positive linear functional f on a *-algebra is said to be pure if any positive linear functional f majorized by f is of the form $f = \lambda f$, $1 \geq \lambda \geq 0$.

This definition is equivalent to

$$f = \alpha_1 f_1 + \alpha_2 f_2 \Rightarrow f_1 = f_2 = f \quad \text{if } \alpha_1 \alpha_2 \neq 0 \\ \alpha_1 + \alpha_2 = 1$$

or

$$\alpha_1 \alpha_2 = 0$$

Therefore, on a normed *-algebra $\mathfrak{A}$, f is extremal in the set of all positive linear forms, of norm ≤ 1 on $\mathfrak{A}$.

I.D.9. Definition: Let $\mathfrak{A}$ be a Banach *-algebra, $\mathfrak{A}^*$ the set of all continuous linear functionals on $\mathfrak{A}$. Any fixed $a \in \mathfrak{A}$ defines a linear functional on $\mathfrak{A}^*$ (the f(a), this may be identified with an element $f_o \in \mathfrak{A}^{**}$). The uniform topology on $\mathfrak{A}^*$ is the topology given by the basis of neighbourhoods of an element $f_o \in \mathfrak{A}^*$

$$\mathfrak{A}_\varepsilon(f_o) = \{f \mid f \in \mathfrak{A}^*, \sup_{\substack{\|a\|=1 \\ a \in \mathfrak{A}}} |f(a) - f_o(a)| < \varepsilon\}$$

We define as W*-topology of $\mathfrak{A}^*$ (or $\mathfrak{A}$-topology of $\mathfrak{A}^*$) the weakest topology in which any element of $\mathfrak{A}$ is continuous as a linear functional on $\mathfrak{A}^*$.

This topology is Hausdorff, and may be characterized by the set of neighbourhoods

$$\mathfrak{N}_\varepsilon(f_o, \sigma) = \{f \mid f \in \mathfrak{A}^*, |f(a) - f_o(a)| < \varepsilon, \forall a \in \sigma\},$$

where σ is a finite subset of $\mathfrak{A}$.

I.D.10. Theorem: (Alaoglu-Bourbaki): Let $\mathfrak{N}$ be a Banach space, $\mathfrak{N}^*$ the dual space of the bounded, continuous linear functionals on $\mathfrak{N}$. Let further $\mathfrak{N}_1$ be the unit ball of $\mathfrak{N}^*$. $\mathfrak{N}_1^*$ is clearly closed in the norm topology. $\mathfrak{N}_1^*$ is compact in the W*-topology.
Proof: See Rickard, discussion at the bottom of p. 222, or Dunford and Schwartz, theorem 2, p. 424.
Corollary: The set $\mathfrak{P}_\mathfrak{A}$ of all positive linear functionals of norm 1 (which we shall call states) on a C*-algebra $\mathfrak{A}$, is closed and compact in the W*-topology.

The importance for us of this theorem and its corollary arises from the theorem of Markov-Kakutani, which we shall state below.

I.D.11. Definition: Let G be a topological group, $\mathfrak{A}$ a C*-algebra and τ_g: $a \to \tau_g a$, $g \in G$, $a \in \mathfrak{A}$, a representation of G in the group of the automorphism of $\mathfrak{A}$. A linear functional $f \in \mathfrak{A}^*$ is said to be invariant under G, if $f(a) = f(\tau_g a)$, $\forall g \in G$, $\forall a \in \mathfrak{A}$.
Theorem: Let $\mathfrak{A}$ be a C*-algebra, F an abelian family of automorphisms of $\mathfrak{A}$. There exists a state f on $\mathfrak{A}$, such that f is invariant under all elements of F.

Proof: Let us first note that the automorphisms τ of $\mathfrak{A}$ induce automorphisms τ^* of $\mathfrak{A}^*$, by the condition

$$(\tau_g^* f)(a) = f(\tau_g a), \ \forall a \in \mathfrak{A} .$$

This automorphism is clearly also an automorphism of the set $\mathfrak{P}_{\mathfrak{A}}$ of all states. We know from the corollary to the theorem of Alaoglu-Bourbaki that $\mathfrak{P}_{\mathfrak{A}}$ is closed and compact in the W*-topology; we also know that each automorphism of $\mathfrak{A}$ is also continuous under this topology. But these properties are exactly the hypothesis of the Theorem (Markov-Kakutani): Let $\mathfrak{K}$ be a compact, convex subset of a linear topological space $\mathfrak{X}$, F an abelian family of continuous linear mappings of $\mathfrak{K}$ onto itself. Then there exists a point $p \in \mathfrak{K}$ such that $fp = p$, $\forall f \in F$.
Proof: See Dunford and Schwartz, theorem 10.6, p. 456.

I.D.12. Definition: A positive linear functional f is said to be pure, or a pure state, if

a) $\|f\| = 1$

b) $f = \alpha_1 f_1 + \alpha_2 f_2$, $\alpha_i \geq 0$, f_i pos. lin. funct. $\|f_i\| = 1$ implies $f_1 = f_2 = f$ and $\alpha_1 + \alpha_2 = 1$, if $\alpha_1 \cdot \alpha_2 \neq 0$, or $\alpha_1 \cdot \alpha_2 = 0$, $\alpha_1 + \alpha_2 = 1$.

In other words, f is pure if and only if it is extremal in the set $\mathfrak{P}_{\mathfrak{A}}$ of all positive linear functionals, of norm ≤ 1 on $\mathfrak{A}$.

We finally shall need the theorem of Krein-Mil'man in the following form:
Theorem: (Krein-Mil'man): If $\mathfrak{A}$ is a Banach algebra, $\mathfrak{P}_{\mathfrak{A}}$ is the closed convex hull spanned by the pure states on $\mathfrak{A}$.

I.E. Topological Properties of Linear Functionals

We have introduced so far two kinds of topologies for the linear functionals, namely the W*-topology and the uniform one. In the particular case of the *-subalgebra of $\mathfrak{B}(\mathfrak{H})$, we can also use the various topologies defined previously on $\mathfrak{B}(\mathfrak{H})$. We shall therefore speak of strongly, ultrastrongly, weakly and ultraweakly continuous linear functionals on any *-subalgebra $\mathfrak{A}$ of $\mathfrak{B}(\mathfrak{H})$.

Furthermore, on a Hilbert space, there exists the particular class of linear functionals generated by vectors. We shall call them ω-forms:

$$\omega: \mathfrak{A} \to \mathbb{C}$$

$$\omega_{x,y}(a) \equiv (x, ay)$$

$$\omega_x(a) \equiv (x, ax)$$

It is clear that any ω-form is weakly continuous, and that ω_x is always a positive linear functional:

$$\omega_x(a^*a) = (x, a^*ax) = (ax, ax) = \|ax\|^2 \geq 0$$

The theorem which we are going to state below should show the connections between the various topologies introduced earlier and the ω-forms.

I.E.1. Proposition: Let $\mathfrak{A}$ be a *-subalgebra of $\mathfrak{B}(\mathfrak{H})$, with identity element, f a positive linear functional, majorized by ω_x. Then there exists $t' \in \mathfrak{A}'$, such that $f = \omega_{t'x}$.
Proof: For $s, t \in \mathfrak{A}$, we have that

$$|f(s^*t)|^2 \leq f(s^*s)\, f(t^*t) \leq |sx|^2 \cdot |tx|^2$$

If we define as a new scalar product on $\mathfrak{A}x$,

$$(sx, tx)' = f(s^*t)$$

we get a positive, sesquilinear, continuous functional defined on $\mathfrak{A}x$, which is also obviously continuous. Using Riesz theorem, we get that there exists an operator $t_o = t_o^*$, positive, on $\overline{\mathfrak{A}x}$, and such that

$$f(s^*t) = (t_o sx, tx)$$

If now $r, s, t \in \mathfrak{A}$, we get that

$$(t_o tsx, rx) = f((ts)^*r) = f(s^*(t^*r)) = (sx, t^*rx)' = (t_o sx, t^*rx)$$

$$= (tt_o sx, rx)$$

and therefore $t_o t = tt_o$ on $\overline{\mathfrak{A}x}$.

Let now p be the projector on $\overline{\mathfrak{A}x}$ $p \in \mathfrak{A}'$. Thus $t_o p$ is hermitian, positive and belongs to $\mathfrak{A}'$. Put $t'^2 = t_o p$, $t' > 0$, we then have that $\forall\, t \in \mathfrak{A}$.

$$f(t) = (t_o x, tx) = (t'^2 x, tx) = (t'x, tt'x) = \omega_{t'x}(t) \qquad \text{Q.E.D.}$$

I.E.2. Proposition: If $\omega_{x,y}$ is a positive linear functional on a *-algebra $\mathfrak{A}$ then there exists a $z \in \mathfrak{H}$, such that $\omega_{x,y} = \omega_z$ on $\mathfrak{A}$.
Proof:

$$4\omega_{x,y}(t) = 2\omega_{x,y}(t) + \overline{2\omega_{x,y}(t^*)}$$

$$= 2(x,ty) + 2(y,tx)$$

$$= (x+y,\ t(x+y)) - (x-y,\ t(x-y))$$

$$\leq (x+y,\ t(x+y))$$

thus $4\omega_{x,y} \leq \omega_{x+y}$ and the proposition follows from the preceding one.

I.E.3. Definition: Let $\mathfrak{A}$ be a von Neumann algebra, $\mathfrak{A}^+$ the set of all positive operators of $\mathfrak{A}$. A positive linear functional f on $\mathfrak{A}$ is said to be normal, if, for any increasing filtering set $\mathfrak{F} \subset \mathfrak{A}^+$ of upper bound $t \in \mathfrak{A}^+$, $f(t)$ is the upper bound of $f(\mathfrak{F})$.

Definition: Let $\mathfrak{A}$ be a von Neumann algebra, $\{p_i\}_{i\in I}$ an arbitrary family of mutually orthogonal projectors of $\mathfrak{A}$. A positive linear functional on $\mathfrak{A}$ is said to be completely additive, if

$$f\left(\sum_{i\in I} p_i\right) = \sum_{i\in I} f(p_i) \ .$$

Remark: In general the notion of complete additivity is stronger than that of σ-additivity (as used in the theory of probability or in measure theory, for instance) since σ-additivity applies only to countable families. It is clear that as long as we stay in separable Hilbert spaces the two notions are equivalent.

I.E.4. Theorem: Let $\mathfrak{A}$ be a von Neumann algebra, f a positive linear functional on $\mathfrak{A}$; the following five conditions are equivalent:

(i) f is normal

(ii) f is completely additive

(iii) $f = \sum_{i=1}^{\infty} \omega_{x_i}$, $\sum_{i=1}^{\infty} |x_i|^2 < \infty$

(iv) f is ultraweakly continuous

(v) f is ultrastrongly continuous

and the following three conditions are equivalent among themselves

(vi) f is weakly continuous

(vii) f is strongly continuous

(viii) $f = \sum_{i=1}^{N} \omega_{x_i}$, $N < \infty$

Proof: See Dixmier, loc. cit.:

(i) ⇔ (iii) ⇔ (v)	I.4.2	theorem 1, p. 54
(i) ⇔ (ii)	I.4	exercise 9, p. 65
(iv) ⇔ (v) ⇔ (iii)	I.3.3	theorem 1, p. 40
(vi) ⇔ (vii) ⇔ (viii)	I.3.3	theorem 1, p. 40

I.E.5. Remarks:

1) Let us note that in the standard formulation of quantum mechanics, states are represented by density matrices, and hence are of the form (iii) of the preceding theorem and thus normal states.

2) In the same way as we did define the continuity of linear functionals with respect to the various topologies introduced, we can speak of the continuity of the different kinds of morphisms we can consider. We have as an example the following

Definition: Let $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be von Neumann algebras, and φ a homomorphism of $\mathfrak{A}_1$ into $\mathfrak{A}_2$. φ is said to be a normal mapping, if, for any filtering increasing subset $\mathfrak{F} \subset \mathfrak{A}^+$ of upper bound $t \in \mathfrak{A}^+$, $\varphi(\mathfrak{F})$ has $\varphi(t)$ as an upper bound.

Theorem: The isomorphism of two von Neumann algebras is normal.

Proof: See Dixmier, loc. cit., I.4.3. cor. 1, p. 57 and theorem 2, p. 56.

Theorem: Let $\mathfrak{A}_1$ and $\mathfrak{A}_2$ be von Neumann algebras, φ a normal mapping of $\mathfrak{A}_1$ into $\mathfrak{A}_2$, such that $\varphi(e_{\mathfrak{A}_1}) = e_{\mathfrak{A}_2}$. Then $\varphi(\mathfrak{A}_1)$ is a von Neumann algebra and φ is continuous for the ultrastrong and the ultraweak topology. That means that the restriction of φ to bounded parts of $\mathfrak{A}_1$ is weakly and strongly continuous.

Proof: See Dixmier, I.4.3 loc. cit., cor. 2, p. 57.

I.F. The Construction of Gelfand, Naimark and Segal

I.F.1. The Gelfand-Naimark-Segal construction provides a representation into $\mathfrak{B}(\mathfrak{H})$ of every C*-algebra $\mathfrak{A}$ with an identity for each positive linear functional defined on $\mathfrak{A}$. The basic idea is to consider $\mathfrak{A}$ as a vector space, and the left regular representation on it. One then defines a scalar product and a norm on it using the given linear functional, and then makes the quotient space with the elements of vanishing norm. This quotient space once completed is then a Hilbert space on which a representation of $\mathfrak{A}$ is defined.

I.F.2. Let therefore φ be a positive linear functional on a C*-algebra $\mathfrak{A}$ with unit element e. We define an hermitian form on $\mathfrak{A}$ (considered now as a vector space) by

$$(x,y) = f(x^*y), \quad \text{for } x,y \in \mathfrak{A}$$

The set $\mathfrak{I}_f = \{x \mid x \in \mathfrak{A},\ (x,x) = 0\}$ is a left ideal of $\mathfrak{A}$. Indeed, suppose $y \in \mathfrak{A}$, we get

$$(yx,\ yx)^2 = f((yx)^*\ yx)^2 = f((x^*y^*y)x)^2$$

$$\leq f((x^*y^*y)^*x^*y^*y)\ f(x^*x) = 0$$

using Schwarz inequality. This means that $yx \in \mathfrak{I}_f$. We still have to verify that $\mathfrak{I}_f$ is a vector space: Suppose $x_1, x_2 \in \mathfrak{I}_f$, then

$$f((x_1+x_2)^*(x_1+x_2)) = f(x_1^*x_1) + f(x_1^*x_2) + f(x_2^*x_1) + f(x_2^*x_2) = 0$$

using Schwarz inequality for the second and third terms.

Call $\mathfrak{H}'$ the quotient space $\mathfrak{A}/\mathfrak{I}_f$, that is, the set of all classes of elements of $\mathfrak{A}$ equivalent modulo elements of $\mathfrak{I}_f$. In each of the classes $\xi, \eta, \ldots$, we pick out a "representative" element x, $y, \ldots$, and we define $(\eta,\xi) = (y,x) = f(y^*x)$. We have to show that this definition is consistent, that is, that the value of (ξ,η) does not depend upon the particular choice of a "representative." Suppose $x_1, x \in \xi$

$$|(y,(x-x_1))|^2 = |f(y^*(x-x_1))|^2 \leq f(yy^*)\cdot f((x-x_1)^*(x-x_1)) = 0$$

as $(x-x_1) \in \mathfrak{I}_f$.

It is clear that $(\xi,\eta) = \overline{(\eta,\xi)}$ so that $(\xi,\xi) \geq 0$ and that the form is sesquilinear. Now suppose $(\xi,\xi) = 0$; this means $f(x^*x) = 0$ for $x \in \xi$, thus $x \in \mathfrak{I}_f$ and $\xi = 0$. Therefore, the form (ξ,η) makes $\mathfrak{H}'$ a prehilbertian space--its completion by respect to this scalar product we shall call $\mathfrak{H}$.

I.F.3. We want now to construct a representation of $\mathfrak{A}$ into $\mathfrak{B}(\mathfrak{H})$. Take an arbitrary class $\xi = \{x\}$ with "representative" x, and let $a \in \mathfrak{A}$; denote by $\eta = \{ax\}$ the class with "representative" ax; the class η does not depend upon the particular choice of x as representative: suppose $x' \in \xi$ then $a(x-x') \in \mathfrak{I}_f$ since $\mathfrak{I}_f$ is a left ideal, thus $ax' \in \eta$. Putting $A_a\xi = \{ax\}$, we define a linear operator A_a on $\mathfrak{H}'$, and thus a mapping $a \to A_a$ of $\mathfrak{A}$ into the set of all linear operators on $\mathfrak{H}'$. We want to show that A_a is bounded.

$$|A_a\xi|^2 = (A_a\xi, A_a\xi) = (ax,ax) = f(x^*a^*ax).$$

We define the linear functional $f_1(a)=f(x^*ax)$. Clearly f_1 is positive: $f_1(a^*a) = f(x^*a^*ax) = f((ax)^*ax) \geq 0$; therefore, $|f_1(b)| \leq f(e)\|b\|$ and, putting $b = a^*a$, $f_1(a^*a) \leq f_1(e)\cdot\|a\|^2$, replacing f_1 by the definition we get $|f(x^*a^*ax)| \leq f(x^*x)\cdot\|a\|^2 = (\xi,\xi)\cdot\|a\|^2$, and thus

$$|A_a\xi|^2 \leq \|a\|^2\cdot|\xi|^2 \quad .$$

Therefore A_a is bounded on $\mathfrak{H}'$ and its norm is smaller or equal to the norm of a, $\|A_a\| \leq \|a\|$. A_a being a linear operator uniformly bounded on a dense subset $\mathfrak{H}'$ of $\mathfrak{H}$ can be extended to a bounded linear operator defined on the whole of $\mathfrak{H}$.

We have now to verify that the mapping $a \to A_a$, $b \to A_b$ etc. is a representation of $\mathfrak{A}$:

$$A_{\alpha a+\beta b}\xi = \{(\alpha a+\beta b)x\} = \alpha\{ax\}+\beta\{bx\} = \alpha A_a\xi + \beta A_b\xi$$

$$A_{ab}\xi = \{abx\} = A_a\{bx\} = A_aA_b\{x\} = A_aA_b\xi$$

and that this representation is a *-homomorphism $A_a^* = A_{a^*}$:

$$(\eta, A_a\xi) = (y,ax) = f(y^*ax) = f((a^*y)^*x) = (a^*y,x) = (A_{a^*}\eta,\xi) \ .$$

I.F.4. Let ξ_0 be the class of all elements of $\mathfrak{A}$ equal to e modulo $\mathfrak{I}_f$. Then $A_a\xi_0$ is the class containing the element a; therefore, the set of all elements of the form $A_a\xi_0$ coincides with the set of all classes, that is, with $\mathfrak{H}'$, but $\mathfrak{H}'$ is dense in $\mathfrak{H}$ and this just means that ξ_0 is a cyclic vector for the representation of $\mathfrak{A}$. (N.B. A representation with a cyclic vector is called a cyclic representation.)

I.F.5. $$f(a) = f(e^*ae) = (e,ae) = (\xi_0, A_a\xi_0).$$

Therefore, in the representation we have just constructed, f is an ω-form, and hence, weakly continuous and normal with respect to the weak closure of the representation of $\mathfrak{A}$ in $\mathfrak{H}$.

We can imagine that we started with $\mathfrak{A}$ defined on some Hilbert space $\mathfrak{H}_1$, and that f was not a sum of ω-forms in $\mathfrak{H}_1$. We arrive, therefore, at the conclusion that for a linear form on a C*-algebra the property of being expressible by ω-forms (or by density matrices) is a representation dependent one--it is true in some and not in others. This observation will be of the utmost importance for the physical applications.

I.F.6. The relation $\|A_a\| \leq \|a\|$ implies that the homomorphism $a \to A_a$ is norm continuous (this is in fact true for all representations of a Banach *-algebra with unit). But, in general, even if $\mathfrak{A}$ is a von Neumann algebra, the mapping will not be continuous in a weaker topology, and the image under an homomorphism of a von Neumann algebra need not be again a von Neumann algebra. However, as a special case, we have the

I.F.7. Proposition: Let $\mathfrak{A}$ be a von Neumann algebra, f a positive linear normal functional on $\mathfrak{A}$, φ the corresponding Gelfand homomorphism. Then φ is ultraweakly and ultrastrongly continuous and $\varphi(\mathfrak{A})$ is again a von Neumann algebra. The restriction of φ to bounded parts of $\mathfrak{A}$ is strongly and weakly continuous.
Proof: Combine th. 2, p. 56, prop. 1, p. 57 of Dixmier, loc. cit.

I.F.8. Definition: A representation $\mathfrak{A}_1$ of an algebra $\mathfrak{A}$ into $\mathfrak{B}(\mathfrak{H})$ is said to be irreducible if $\mathfrak{A}_1' = \{\lambda I\}$. This is equivalent to saying that $\mathfrak{A}_1$ does not leave invariant any closed subspace of $\mathfrak{H}$ other than $\{0\}$ on $\mathfrak{H}$.

I.F.9. Proposition: Let $\mathfrak{A}$ be a Banach *-algebra, f a positive linear functional on $\mathfrak{A}$, $\mathfrak{A}_1$ the associated Gelfand representation in $\mathfrak{B}(\mathfrak{H})$. Then $\mathfrak{A}_1$ is irreducible if and only if f is pure on $\mathfrak{A}$.
Proof: See Naimark, Normed Rings, p. 255 and ff.

I.G. Classification of Factors

I.G.1. Definition: Let $\mathfrak{A}$ be a von Neumann algebra. Two projectors p_1 and p_2 of $\mathfrak{A}$ are said to be equivalent (with respect to $\mathfrak{A}$) if there exists a partially isometric operator $v \in \mathfrak{A}$ such that $vv^* = p_2$, $v^*v = p_1$. One then writes $p_1 \sim p_2$. One easily sees that $\sim$ satisfies the properties of an equivalence relation.

I.G.2. Definition: A projector p of a von Neumann algebra is said to be infinite (with respect to $\mathfrak{A}$) if it is equivalent to a projector p_1 strictly smaller than p, $p \sim p_1 \lneq p$. A projector which is not infinite is finite.

I.G.3. Definition: p_1 is said to majorize p_2, $p_2 \propto p_1$ for $p_1, p_2 \in \mathfrak{A}$ if there exists $p \in \mathfrak{A}$, $p \sim p_2$, $p \leq p_1$.

I.G.4. Proposition: $p_1 \propto p_2$ and $p_2 \propto p_1$ imply $p_1 \sim p_2$.
Proof: Dixmier, loc. cit., p. 226, prop. 1.

I.G.5. Theorem: Let $\mathfrak{A}$ be a factor, p_1, $p_2 \in \mathfrak{A}$. Then either $p_1 \propto p_2$ or $p_2 \propto p_1$.
Proof: Dixmier, loc. cit., cor. 1, p. 228.

I.G.6. Definition: Let $\mathfrak{A}$ be a factor, thus it is said to be

(i) of type III_∞, or purely infinite if every projector of $\mathfrak{A}$ is infinite. A factor which is not purely infinite is said to be semi-finite.

(ii) A semi-finite factor which contains minimal projectors is said to be of type I or discrete. A factor which is not type I is also called continuous.

(iii) A semi-finite factor which does not contain minimal projectors is said to be of type II.

(iv) A factor is said to be infinite if it contains at least one infinite projector, and finite if it contains only finite projectors. A finite discrete factor is said to be of type I_n, an infinite discrete factor of type I_∞. A finite continuous factor is said to be of type II_1, and an infinite continuous factor is of type II_∞, if it is semi-infinite, and III_∞ otherwise.

I.G.7. Definition: A von Neumann algebra is said to be of type I (resp. II, III) if it is the direct integral of factors of type I (resp. II, III). It is finite if all its projectors are finite, otherwise infinite.

For more details on the classification of von Neumann algebras, consult the lectures by W. Wyss.

Section II. Statistical Mechanics

II.A. Generalities

II.A.1.

It is not possible, in the relatively small number of hours at our disposal, to pretend giving here a complete survey of all the applications of algebraic methods to statistical mechanics. We shall try, however, to give some of the leading ideas and important results in this relatively new field.

We want to warn the specialists of statistical mechanics that what we are going to present here will appear rather strange to them, but one should not forget that nearly all researchers in that field are quantum field theorists, and that many of the techniques and aims have been borrowed from that field.

Among the important results which we shall mention, there is the generalisation of ergodic theory to quantum systems, the notion of asymptotic abelianes, the analytical behavior of the thermodynamical functions, the time evolution, and so on.

II.A.2. As is also developed in the lectures by H. Borchers,[6] any quantum system can be abstracted in the following way:

Observables: Self-adjoint operators of a C*-algebra $\mathfrak{A}$

States: Positive linear functionals on $\mathfrak{A}$, of norm 1, $\mathfrak{P}_{\mathfrak{A}}$

In classical mechanics, we have as interpretations:

Observables: Continuous functions on phase space (vanishing at infinity if the latter is not compact)

States: Positive measures on phase space.

We thus have apparently two different interpretations for classical and quantum mechanics. This is happily not true and can be seen in the following way:

Let us start with an example, that of the ideal gas in a box. Let thus $\mathbb{R}^{\nu}$ be the phase space of that system, and, for instance, if we take the microcanonical ensemble, we get as admissible phase space a compact $\mathfrak{R}$ of the energy "surface" between E and E + dE. This set $\mathfrak{R}$ is compact because the energy is finite and the system is in a box.

The observables generate in an evident way a *-algebra $\mathfrak{A}_0$ of the continuous functions defined on $\mathfrak{R}$. One can make a C*-algebra out of it by taking the supremum norm. A state is defined by the probability of the system to be in the different points of $\mathfrak{R}$, which amounts to the definition of a positive measure ρ on $\mathfrak{R}$ and of the definition of the expectation value of $f \in \mathfrak{A}_0$ in that state; it is given by $\int_{\mathfrak{R}} f\rho$. But we know from the general mathematical theory that the positive measures on $\mathfrak{R}$ form exactly the positive cone of the dual of $\mathfrak{A}_0$. That means for us that we can interpret the classical mechanics exactly in the same way as quantum mechanics, by taking the self-adjoint elements of a C*-algebra as observables, and the positive linear functionals on that algebra as the states.

Note, however, that the essential characterization of the classical system is the fact that the C*-algebra is abelian.

Once we have made this abstraction for a classical system, one may wonder whether the usual interpretation of the observable by continuous functions and of the states by measures is always possible. That this is true is given by the

Theorem (Gelfand): Let $\mathfrak{A}$ be a C*-algebra, abelian. Then there exists a Hausdorff space $\mathfrak{X}$, locally compact such that $\mathfrak{A}$ is isomorphic to the algebra $C_0(\mathfrak{X})$ of continuous functions on $\mathfrak{X}$ decreasing at ∞. It is compact if, and only if $\mathfrak{A}$ possesses an identity element.

Proof: All the proofs are nearly the same; they differ only by the realization of the space $\mathfrak{X}$ as

- space of maximal ideals: Naimark § 4
- space of characters: Dixmier C*, th. 1.4.1

- space of pure lin. functionals: Kadison, Cargèse lectures, theorem 2.2.1.

What remains general in the algebraic approach is the fact that one is not specifying the manifold chosen as having the admissible position of phase space, nor the particular topology on it. This is of course of great importance in the thermodynamical limit, whenever the phase space is becoming infinite dimensional.

II.A.3. If we now are looking at the states on such a C*-algebra, we immediately observe that there exists a fantastic collection of them.

In the particular case of the ideal gas, for instance, we have that the states used in statistical mechanics are measures equivalent to the Lebesgues one. The δ-functions, on the other hand, are giving the analytical mechanics. In order to select a particular description, we are therefore going to impose further restrictions on the set of admissible states. The first (and most important perhaps) class of restrictions which we shall mention is bound to the notion of invariance.

II.B. Invariant States

II.B.1. Definitions: Let G be a topological group and $\tau_g : a \to \tau_g a$, $g \in E$, $a \in \mathfrak{A}$ a representation of G in the group of automorphisms of $\mathfrak{A}$. A linear functional $f \in \mathfrak{A}^*$ is said to be invariant under G (or G-invariant) if $f(a) = f(\tau_g a)$, $\forall\, a \in \mathfrak{A}$, $\forall\, g \in G$.

For abelian C*-algebras $\mathfrak{A}$, the notion of invariance can be translated in terms of the associated $C_o(\mathfrak{X})$. To the automorphisms corresponds the homeomorphisms of $\mathfrak{X}$. To the invariant states on $\mathfrak{A}$ one associates the (finite) Radon measures on $\mathfrak{X}$. We are again in the same set up as we started from, up to the difference that the measure on $\mathfrak{X}$ is bonded.

Another special case is the one where G is the group of inner automorphisms of a von Neumann algebra $\mathfrak{A}$. Then τ_g is trivial and the invariant positive functionals are finite traces on $\mathfrak{A}$.

II.B.2. The theorem on the existence of at least one invariant state (p. 44) is not always very useful because we don't know anything about the properties of this invariant state. It is possible to obtain much more powerful results, however, and for this we refer to the lectures by H. Borchers at the present school.

II.C. Quasilocal Algebras

II.C.1. Until now, we have not imposed any restrictions on the C*-algebra we were considering. There is, however, a crucial property, which will be of utmost importance in everything which follows. This is the local nature of all interactions. This property is ensured in quantum field theory by the locality axiom, and in statistical mechanics it should be reflected in the fast decrease of the correlation functions for great space separation. In the case of statistical mechanics, however, the situation is slightly more complicated than in quantum field theory, and this is because long range correlations do indeed appear. We, therefore, have to be careful by not excluding interesting cases, and we shall see the right description is one in which the algebra, the states and the invariance group together possess certain properties. We start with a rather elementary, but very important, example, the case of quasilocal algebras.

II.C.2. Let $a^*(x)$ and $a(x)$ be the usual (second quantized) creation and annihilation operators defined on a Fock space. More exactly, we shall consider the smeared out operators, symbolically written as

$$a^*(f) = \int f(x)\, a^*(x)dx; \quad a(f) = \int f(x)a(x)dx$$

where f is in $\mathcal{L}^2(\mathbb{R}^s)$.

For operators satisfying the canonical commutation rules, $a^*(f)$ and $a(f)$ are unbounded closed operators. In this case, one writes down the polar decomposition, and, instead of $a^*(f)$ or $a(f)$ itself, one considers the partial isometry and the spectral family of the hermitian part.

In the case of operators satisfying the canonical anticommutation relations, one has bounded operators on the Fock space, but we shall consider only operators which are products of the same number of a^* than of a. This is because we only want to retain observables.

There are now many different C*-algebras which may be associated with these operators, and there is no choice which would be the best under all circumstances. In statistical mechanics, however, the restriction of a well defined local density imposes restrictions which will enable us to make a "standard" choice.

II.C.3. Let $\mathfrak{D}_{\mathcal{O}}$ be the space of infinitely many times differentiable continuous functions of support contained in $\mathcal{O}$. We shall denote by $\mathfrak{B}(\mathcal{O})$ the C*-algebra generated by all operators

$$\{a^*(f),\ a(f) \mid f \in \mathfrak{D}_{\mathcal{O}}\}$$

and $\mathcal{A}(\mathcal{O})$ the von Neumann algebra generated by the same operators, thus

$$\mathcal{A}(\mathcal{O}) = \mathfrak{B}(\mathcal{O})''$$

Note that in the case of the commutation relation, $\mathcal{A}(\mathcal{O})$ is a factor of type I_∞.

We shall further define

$$\mathfrak{B} = \overline{\bigcup_{\mathcal{O} \text{ bounded}} \mathfrak{B}(\mathcal{O})}$$

as the algebra of quasilocal operators and

$$\mathcal{N} = \bigcup_{\mathcal{O} \text{ bounded}} \mathcal{A}(\mathcal{O}) \text{ resp. } \mathcal{A} = \overline{\bigcup_{\mathcal{O} \text{ bounded}} \mathcal{A}(\mathcal{O})}$$

We notice that the local algebras which we are defining here are not the same as those which we would define for a field theory, because in the latter case, we would generate the algebra using $\varphi(f)$, $\dot{\varphi}(f)$. The reason for that is that in the first case, there exists a local "number of particles operator." This is not the case in quantum field theory, the "number of particle operators" not being a local one.

II.C.4. In the case of the Fock representation, we know that there exists an operator "number of particles." Therefore we also know that for every bounded region $\mathcal{O}$, there exists an (unbounded in the case of the CCR) operator $N(\mathcal{O})$, affiliated to $\mathcal{A}(\mathcal{O})$. There is an inverse to this proposition, and it is given by the

Theorem: A representation of the canonical commutation or anticommutation relation by linear operators on a Hilbert space possesses an operator "number of particles" if, and only ,f it is unitarily equivalent to the Fock representation, or to a direct sum of representations unitarily equivalent to the Fock one, or, in other words, is quasi-equivalent to the Fock representation.

Proof: See Garding and Wightman, Proc. Nat. Acad. Sci. U.S. 40, 617, 622 (1954); Dell'Antonio, Doplicher, Ruelle, C.M.P. 2, 223 (1966).

II.C.5. For evident physical reasons, we are, in statistical mechanics, only interested in such representations π (faithful, and we shall see that this last property is always satisfied in this model, $\mathcal{A}$ being simple) of $\mathfrak{B}$ (or $\mathcal{A}$), such that there exists an operator "number of particles" for each $\pi(\mathfrak{B}(\mathcal{O}))$. But we know that this is true only if $\pi(\mathfrak{B}(\mathcal{O}))$ is quasiequivalent to the Fock representation, but this means, of course, that π can always be extended to $\mathcal{A}(\mathcal{O})$ (and hence to $\mathcal{A}$), and that the mapping of $\mathcal{A}(\mathcal{O})$ onto $\pi(\mathcal{A}(\mathcal{O}))$ will be normal. It is, therefore, permitted to work with the weakly closed $\mathcal{A}(\mathcal{O})$ instead of the $\mathfrak{B}(\mathcal{O})$.

II.C.6. The standard procedure, in statistical mechanics, for constructing states in the thermodynamical limit, is to define first local states, and then take the limit. This procedure also applies in our case; we have,however, a very severe restriction to impose on the limiting state f, and this is that f has to be normal on $\mathcal{A}(\mathcal{O})$. Otherwise, whenever we shall make the Gelfand construction with f, we shall not have a representation of $\mathcal{A}(\mathcal{O})$ quasiequivalent to $\mathcal{A}(\mathcal{O})$.

Fortunately enough, this criterion is simple to satisfy in our case, and this because we have shown that we could take the algebras $\mathcal{A}(\mathcal{O})$ as von Neumann algebras.

Let, therefore, B be a box, of volume V(B), centered at the origin, and $\mathcal{A}(B)$ the corresponding local von Neumann algebra. Let further f_B be a state on $\mathcal{A}(B)$ such that $f_B(N_B) < \infty$, where N_B is the operator "number of particles" in B.

Let now $\{B_i\}$ be an increasing sequence of boxes, such that

$$\lim_{i \to \infty} f_{B_i}(N_{B_i})(V(B_i))^{-1}$$

exists and is finite. We suppose, further, that f_{B_i} converges weakly on each $\mathcal{A}(\mathcal{O})$; that is, for i, j sufficiently large, and B_i, $B_j \supset \mathcal{O}$, and for any $a \in \mathcal{A}(\mathcal{O})$,

$$\lim_{ij \to \infty} |f_{B_i}(a) - f_{B_j}(a)| = 0$$

This hypothesis can be physically interpreted by saying that the physical observables of the infinite system should have as expectation values the limit of the expectation values for the observables in a finite system. In order that the limit has any meaning at all, it is of course also necessary to show that the limit is independent of the choice of the sequences of boxes (or other volume) taken. This has, of course, to be checked in every particular case, as we shall also see later.

Under this hypothesis, it should be clear that f_{B_i} converges toward a linear functional defined on the whole of $\mathcal{A}$:

Proof: It is evident that the limit is a functional on $\mathcal{N} = \bigcup_{\mathcal{O}\ \mathrm{bd.}} \mathcal{A}(\mathcal{O})$. Let now $b \in \mathcal{A}$, $\varepsilon > 0$. There exists $\mathcal{O}$ bounded, $\mathcal{O} \subset \mathbb{R}^3$ and an $a \in \mathcal{A}(\mathcal{O})$ such that $\|b-a\| < \varepsilon$. Let now $\hat{f}_i$ be a normal extension (arbitrary) of $f_i \equiv f_{B_i}$ to $\mathcal{A}''$. We have that

$$|\hat{f}_i(b) - \hat{f}_j(b)| < 2\varepsilon + |\hat{f}_i(a) - \hat{f}_j(a)| = 2\varepsilon + |f_i(a) - f_j(a)| < 3\varepsilon$$

Furthermore, the limit is not depending upon the extension f chosen: more exactly, let

$$\hat{f} = \lim_{i \to \infty} \hat{f}_i,\ \hat{g} = \lim_{i \to \infty} \hat{g}_i,\ \hat{f}_i = \hat{g}_i \text{ on } \mathcal{A}(\mathcal{O}), \text{ for } \mathcal{O} \text{ bounded}.$$

Then, for $b \in \mathcal{A}$, $\varepsilon > 0$, there again exists $a \in \mathcal{A}(\mathcal{O})$, $\mathcal{O}$ bounded, with $\|a-b\| < \varepsilon$. For i sufficiently large, we have that

$$|\hat{f}(b) - \hat{g}(b)| \leq 2\varepsilon + |\hat{f}(a) - \hat{g}(a)| \leq 4\varepsilon + |\hat{f}_i(a) - \hat{g}_i(a)|\ .$$

Let now i be such that $B_i \supset \mathcal{O}$, then $\hat{f}_i(a) = \hat{g}_i(a)$, $\forall\ a \in \mathcal{A}(\mathcal{O})$, thus $\hat{f} = \hat{g}$ on $\mathcal{A}$.

II.C.7. The next question to answer is of course whether we have any guarantee that the limiting state will be in any way an acceptable one, and, as we have already seen, one condition is to have a well defined local number of particles. That this is automatically true is due to a very extraordinary property of von Neumann algebras, namely the

Theorem: Let $\mathfrak{A}$ be a von Neumann algebra, f_i a sequence of positive linear normal functionals on $\mathfrak{A}$, converging weakly on $\mathfrak{A}$ to f. (i.e. $\forall\ a \in \mathfrak{A}$, $|f_i(a) - f(a)| \xrightarrow[i \to \infty]{} 0$) Then f is normal. If $\mathfrak{A}$ is a type I factor, the convergence is uniform.

Proof: See Sakai, Proc. Japan Acad. 33, 439 (1957); Dell'Antonio, Comm. Pure And Appl. Math. 20, 413 (1967).
In our case we therefore only need to recall that $\mathcal{A}(\mathcal{O})$ is type I.

Note that the above mentioned theorem only is valid for von Neumann algebras. It is in general false on C*-algebras.

Once one has constructed a particular state having the properties desired, one then makes the Gelfand construction. The fact that the limiting state will be normal on the $\mathcal{A}(\mathcal{O})$ implies that $\pi_f(\mathcal{A}(\mathcal{O}))$ is again a von Neumann algebra, and of the same type (but no longer necessarily a factor).

II.C.8. If we now consider the commutation properties of the algebras $\mathcal{A}(\mathcal{O})$, we first remark that these algebras are not local in the sense that $\mathcal{O}_1 \cap \mathcal{O}_2 = \{\phi\} \not\Longrightarrow \mathcal{A}(\mathcal{O}_1) \subset \mathcal{A}(\mathcal{O}_2)'$. We know, however, that the commutator $[a^*(x), a(y)]$ decreases like an exponential. This means that for great distances of their support, the observables essentially commute. This property is an extremely important one, and we shall study it in detail in the next section, but before that, we want quickly to examine what would happen if we had strict locality. We have the following theorem, due to Misra (H.P.A. 38, 189 (1965)).

II.C.9. Theorem: Let be given a mapping $\mathcal{O} \to \mathcal{A}(\mathcal{O})$ of the open bounded sets $\mathcal{O} \in \mathbb{R}^s$ into the set of all von Neumann algebras, such that

1) $\mathcal{A}(\mathcal{O})$ is a factor
2) $\mathcal{O}_1 \subset \mathcal{O}_2 \Rightarrow \mathcal{A}(\mathcal{O}_1) \subset \mathcal{A}(\mathcal{O}_2)$
3) $\mathcal{O}_1 \cap \mathcal{O}_2 = \phi \Rightarrow \mathcal{A}(\mathcal{O}_1) \subset \mathcal{A}(\mathcal{O}_2)'$
4) There exists a representation of a nontrivial noncompact group of translations (discrete or continuous) in $\mathrm{Aut}(\mathcal{A})$.
5) $\mathfrak{O}$ is separable.

Then the C*-algebra $\mathcal{A} = \overline{\bigcup_{\mathcal{O} \text{ bounded}} \mathcal{A}(\mathcal{O})}$ is simple.

Proof: We shall first show that $\mathcal{N} = \bigcup_{\mathcal{O} \text{ bounded}} \mathcal{A}(\mathcal{O})$ is simple.

Suppose $\mathfrak{J}$ is a two-sided ideal of $\mathcal{N}$, then there exists at least one $\mathcal{O}$ such that $\mathfrak{J} \cap \mathcal{A}(\mathcal{O}) \neq \phi$. Put $\mathfrak{J} \cap \mathcal{A}(\mathcal{O}) = \mathfrak{J}_{\mathcal{O}}$, $\mathfrak{J}_{\mathcal{O}}$ is clearly a two-sided ideal of $\mathcal{A}(\mathcal{O})$. But $\mathcal{A}(\mathcal{O})$ is supposed to be a factor, so it cannot be finite, since finite factors are simple. We may further suppose that there exists a nontrivial projector $p \in \mathfrak{J}_{\mathcal{O}}$ (Naimark IV.22.1). Take $\mathcal{A}(\mathcal{O}_1)$, $\mathcal{O}_1 \cap \mathcal{O} = \phi$, and $\mathcal{A}(\mathcal{O}_1)$ infinite, this certainly exists because of the invariance under translation. We claim, now, that p is an infinite projector with respect to the algebra $\mathcal{A}(\mathcal{O}_2)$, $\mathcal{O}_2 = \mathcal{O} \cup \mathcal{O}_1$.

Suppose it were not the case, then the reduced algebra $\mathcal{A}(\mathfrak{G}_2)_p$ would be finite but this algebra contains as subalgebra the algebra $\mathcal{A}(\mathfrak{G}_1)_p$. But as $p \in \mathcal{A}(\mathfrak{G}_1)'$, this implies that $\mathcal{A}(\mathfrak{G}_1)$ is isomorphic to $\mathcal{A}(\mathfrak{G}_1)$, which we supposed infinite, and as a finite algebra cannot have an infinite subalgebra, we conclude that p must be infinite with respect to $\mathcal{A}(\mathfrak{G}_2)$.

This in turn implies that there exists an isometric operator $w \in \mathcal{A}(\mathfrak{G}_2)$ such that $ww^* = 1$ and $w^*w = p$ (Naimark, VII 35.5. prop. VI). From this follows, however, that $wpw^* = ww^*ww^* = 1 \in \mathfrak{Z}(\mathfrak{G}_2)$. Thus $\mathfrak{Z}(\mathfrak{G}_2)$ is trivial and hence $\mathcal{N}$ is simple.

Let us now show that $\mathcal{A} = \overline{\mathcal{N}}$ is also simple. It is sufficient for that to show that all representations of $\mathcal{A}$ are faithful.

Suppose now that π is a representation of $\mathcal{A}$, $t \in \mathcal{A}$, and $\pi(t) = 0$. Let t_n be a sequence of elements of $\mathcal{N}$ such that t_n converges in norm toward t. We have that

$$\|\pi(t_n - t)\| = \|\pi(t_n) - \pi(t)\| = \|\pi(t_n)\| \leq \|t_n - t\|$$

But t_n belongs to a certain $\mathcal{A}(\mathfrak{G}_n)$, and as $\mathcal{N}$ is simple, $\pi(\mathcal{A}(\mathfrak{G}_2))$ is faithful, and thus preserves, the norm since $\mathcal{A}(\mathfrak{G}_n)$ is a C*-algebra, thus $\|\pi(t_n)\| = \|t_n\| \Rightarrow \|t_n\| \leq \|t_n - t\|$, thus $\|t_n\| \xrightarrow[n \to \infty]{} 0$, and hence $t = 0$. Therefore all representations of $\mathcal{A}$ are faithful and $\mathcal{A}$ is therefore simple. Q.E.D.

II.D. Asymptotically Abelian Systems

II.D.1. Definition: A pair $\{\mathfrak{A}, \alpha\}$ consisting of a C*-algebra $\mathfrak{A}$ and an homomorphism $g \to \alpha_g$ of a locally compact, noncompact topological group G into the group of the automorphisms of $\mathfrak{A}$, is called an asymptotically abelian system, if to any $\epsilon > 0$, $a, b \in \mathfrak{A}$, corresponds a compact $K \subset G$, such that $g \notin K$ implies

$$\|[a, \alpha_g(b)]\| < \epsilon$$

Such a pair is called weakly asymptotically abelian, if $\forall\ \epsilon > 0$, $a, b \in \mathfrak{A}$ and any state f on $\mathfrak{A}$, there exists a $K \subset G$, compact, and such that $g \notin K$ implies

$$|f([a, \alpha_g(b)])| < \epsilon \quad .$$

In order to simplify notations, and because it is anyway the most important case, we shall take as model for G the group T of translations x.

We shall further define:

II.D.2. Definition: Let f be a state on $\mathfrak{A}$. f is said to be strongly clustering (fortement essaimant, in French) if $\forall a, b \in \mathfrak{A}$,

$$\lim_{|x| \to \infty} |f(\tau_x(a)b) - f(\tau_x(a))\, f(b)| = 0$$

Let g be a function on $\mathbb{R}^3$, we define as the mean M(g) of g over $\mathbb{R}^3$,

$$M(g) = \lim_{L \to \infty} \frac{1}{4\pi L^3} \int_{|x| \leq L} g(x) d^3x$$

whenever it exists. This mean is the (C,1) Cesaro mean (cf. Hille and Phillips, Functional Analysis and Semi-Groups) and is also called the Wiener mean.

We shall say that a state is weakly clustering if

$$M(f(\tau_x(a)b) - f(\tau_x a)\, f(b)) = 0$$

II.D.3. It is clear that if a function g is continuous and tends to zero at infinity, M(g) = 0. We conclude therefore that any strongly clustering state is also weakly clustering.

If f is T-invariant, we have that

$$M(f(\tau_x a)) = M(f(a)) = f(a)$$

and we can write the property of weakly clustering as

$$M(f(\tau_x(a)b)) = f(a)\, f(b)$$

II.D.4. Definition: Let $\{\mathfrak{A}, \alpha\}$ be an asymptotically abelian system. The weakly clustering G-invariant states are called E-states, or ergodic states.

This class of states is truly remarkable; that they represent a generalisation of the classical notion of ergodic states should become clear after some results, especially the fact that ergodic states are extremal among the set of invariant states.

II.D.5. Theorem: Let $\{\mathfrak{A}, \alpha\}$ be an asymptotically abelian system, f a G-invariant state on $\mathfrak{A}$, and $\pi_f(\mathfrak{A})$ the Gelfand construction induced by f on $\mathfrak{H}_f$, $x_f \in \mathfrak{H}_f$ the cyclic vector such that $(x_f, \pi_f(a)x_f) = f(a)$, $\forall a \in \mathfrak{A}$, and $u_f(\alpha)$ the representation of G on $\mathfrak{H}_f$. Then the following properties are equivalent:

1) f is weakly clustering (and hence a E-state)
2) x_f is the only vector of $\mathfrak{H}_f$ invariant under $u_f(\alpha)$
3) Let $\mathfrak{R} = \{\pi_f(\mathfrak{A}) \cup u_f(G)\}''$, then $\mathfrak{R}' = \{\lambda I\}$

4) $\mathfrak{R}$ (as defined under 3)) is a factor

5) f is an extremal element of the convex set of all G-invariant states.

Proof: See Doplicher, Kadison, Kastler, Robinson, C.M.P. 6, 101 (1967); Lanford and Ruelle, JMP 8, 1460 (1967).

II.D.6. The fact that the E-states are extremal in the convex set of all G-invariant states is of course the most important result. It is perhaps useful at this point to recall an essential result of classical ergodic theory:

Theorem: Let K be a compact, metrizable space (phase space) and G a group of automorphisms of K (time evolution). Then a G-invariant measure on K may be decomposed in a unique way into extremal G-invariant measures (ergodic measures).

References and Proof: R. Phelps, Lectures on Choquet's Theorem, Van Nostrand (1966); K. Jacobs, Jahresber. der Deutschen Math. Vereins 67, 143 (1965).

We should like to have an extension of this classical result to the quantum case, that is, we want to show that any G-invariant state may be uniquely decomposed in E-states. As in classical theory, one lets correspond the pure thermodynamical phases to the E-states.

Let therefore $\mathfrak{P}(\mathfrak{A})$ be the set of states on $\mathfrak{A}$, μ a measure on $\mathfrak{B}(\mathfrak{A})$, normalized to $\int d\mu = 1$. To each μ corresponds a state f_μ by

$$f_\mu(a) = \int f(a)\, d\mu(f)$$

In the converse case, where f is given, we would like to have that f may be written as

$$f(a) = \int f_\xi(a)\, d\mu(f_\xi) \quad , \quad \int d\mu = 1$$

where the states f_ξ are G-invariant and extremal, thus E-states, and are parametrized by the intensive quantities of the system, like temperature, mean density, and so on.

II.D.7. Let f_μ be as above. One could imagine that $f_{\mu_1} = f_{\mu_2}$ may happen even if $\mu_1 \neq \mu_2$. In such a case the decomposition of f_μ would not be unique; it is also conceivable that the decomposition may not exist at all. We shall therefore develop some criterions.

We first define the measure δ_f, by

$$f(a) = \int f_\xi(a)\, d(\delta_f(\xi))$$

and we introduce a partial ordering among the set of measures on $\mathfrak{P}(\mathfrak{A})$:

$$\mu_1 < \mu_2 \Leftrightarrow \mu_1(\rho) < \mu_2(\rho) \quad \forall \rho \ ,$$

where ρ are convex continuous functionals on $\mathfrak{P}(\mathfrak{A})$, $\mu(\rho)$ being defined by

$$\mu(\rho) = \int \rho(f)\, d\mu(f)$$

and ρ convex meaning

$$\rho(\lambda f_1 + (1-\lambda) f_2) \leq \lambda \rho(f_1) + (1-\lambda)\, \rho(f_2) \ , \qquad 0 \leq \lambda \leq 1$$

We now define a particular linear convex (continuous) functional

$$\rho_a(f) \equiv f(a)$$

Verify that ρ_a is convex:

$$\begin{aligned} \rho_a(\lambda f_1 + (1-\lambda) f_2) &= (\lambda f_1 + (1-\lambda) f_2)(a) \\ &= \lambda f_1(a) + (1-\lambda) f_2(a) \\ &= \lambda \rho_a(f_1) + (1-\lambda) \rho_a(f_2) \end{aligned}$$

As we have the equality sign, we may also conclude that both ρ_a and $-\rho_a$ are convex, and that $\rho_a(f)$ is real if $a = a^*$. We are now ready to see under which (sufficient) conditions, we may have that

$$f_{\mu_1} = f_{\mu_2} \quad \text{and} \quad \mu_1 \neq \mu_2$$

Theorem: The measures related by the partial ordering $<$ induce the same state.

Proof: It is sufficient to remark that $\mu_1 < \mu_2$ implies

$$\mu_1(\rho_a) \leq \mu_2(\rho_a)$$

and also

$$\mu_1(-\rho_a) \leq \mu_2(-\rho_a)$$

or

$$-\mu_1(\rho_a) \leq -\mu_2(\rho_a)$$

and thus

$$\mu_1(\rho_a) = \mu_2(\rho_a)$$

that is,

$$\int f(a)d\mu_1(t) = \int f(a)d\mu_2(f) \qquad \text{QED.}$$

Theorem: If f is not extremal, there exists $\mu > \delta_f$, $\mu \neq \delta_f$.
Proof: Let now

$$f = \lambda f_1 + (1-\lambda)f_2, \quad f_1 \neq f_2, \quad \lambda \in (0,1)$$

Put $\mu = \lambda\delta_{f_1} + (1-\lambda)\delta_{f_2}$. We are going to show that $\delta_f < \mu$. Indeed,

$\delta_f(\rho) = \rho(f)$, and since, for a convex ρ,

$$\rho(f) \leq \lambda(f_1) + (1-\lambda)(f_2)$$

we can write

$$\delta_f(\rho) \leq \mu(\rho), \text{ thus } \delta_f < \mu.$$

As $\delta_f \neq \delta_g$, if $f \neq g$, we have that

$$\delta_f \neq \mu \qquad \text{QED.}$$

Corollary: If f is an E-state, δ_f is maximal with respect to the relation $<$.
In the case where f is not extremal, we have to prove the existence of a maximal measure $\mu > \delta_f$, and then show that its support contains only E-states.
The first statement is a consequence of the
Theorem: (Lanford and Ruelle, loc. cit.) Let $\mathfrak{A}$ be a C*-algebra with an identity, asymptotically abelian and f a G-invariant state on $\mathfrak{A}$. Let further $\pi_f(\mathfrak{A})$ be the Gelfand representation induced by f on $\mathfrak{H}_f$. Then there exists a unique maximal measure μ_f on the set of all G-invariant states, such that

$$\mu_f > \delta_f,$$

i.e., $\forall a \in \mathfrak{A}$, $\int g(a)d\mu_f(g) = \int g(a)d\delta_f(g) = f(a)$.

(This measure μ_f is given by

$$\mu_f(\rho_{a_1}\rho_{a_2}\cdots\rho_{a_n}) = (x_f, \pi_f(a_1)p_f\pi_f(a_2)p_f\cdots p_f\pi_f(a_n)x_f)$$

where $x_f \in \mathfrak{H}_f$ is cyclic for $\mathfrak{A}$ and where $f(a) = (x_f, \pi_f(a)x_f)$ $\forall a \in \mathfrak{A}$, p_f being the projector on the subspace of $\mathfrak{H}_f$ invariant under $u_f(G)$.)

It remains to show that the support of μ_f is contained in the set of E-states. Unfortunately, we don't have a completely general proof. For instance, it is easy to prove that

Theorem: If $\mathfrak{A}$ is further separable (as Banach space), then μ_f is supported by E-states.

This theorem does not satisfy us completely, since the algebras of interest in physics are rarely separable (if $\mathfrak{H}$ is of infinite dimension, $\mathfrak{L}(\mathfrak{H})$ is not separable). It is at this point that one may use the fact that, in most applications, $\mathfrak{A}$ may be realized as the norm closure of the union of $\mathcal{A}(\mathfrak{G})$, the $\mathcal{A}(\mathfrak{G})$ being of type I, and the fact that we want to have states with well defined local number of particles. One can thus prove the

Theorem: Let $\mathfrak{A}$ be a C*-algebra with identity element, asymptotically G-abelian, f a G-invariant state on $\mathfrak{A}$. Suppose further, that there exists a countable family $\{\mathfrak{A}_i\}$ of sub-C*-algebras of $\mathfrak{A}$, such that this set be dense in $\mathfrak{A}$. Let $\mathfrak{I}_i$ be a two-sided ideal of $\mathfrak{A}_i$, complete in norm and separable (as Banach space). Finally, suppose that the restriction of f to each $\mathfrak{I}_i$ be of norm 1. Then

1) $\mathfrak{H}_f$ is separable

2) If μ is a positive measure on the set of G-invariant states on $\mathfrak{A}$ such that $\mu > \delta_f$, μ being normalized to 1 on that set, then μ is maximal on that set if, and only if it has its support on E-states.

For the proof, see Lanford and Ruelle, loc. cit., and Ruelle, C.M.P. 3, 133 (1966).

We remark that this theorem applies to the quasilocal algebra $\mathcal{A}$ generated by the $\mathcal{A}(\mathfrak{G})$. The two-sided ideal being in that case of the form $\mathfrak{C} \otimes I$, since all $\mathcal{A}(\mathfrak{G}_i)$ are of the form $\mathfrak{L}(\mathfrak{H}) \otimes I$, $\mathfrak{C}$ being the two-sided ideal of compact operators. That f has norm 1 on $\mathfrak{C}$ implies that f will be normal on $\mathcal{A}(\mathfrak{G})$.

It is possible to formulate different sets of conditions; in particular, we have the

Theorem: Let f be a G-invariant state on a G-asymptotically abelian C*-algebra $\mathfrak{A}$, and let $\mathfrak{H}_f$ be the Hilbert space of the Gelfand construction induced by f.

Suppose that $\mathfrak{H}_f$ is separable. Then there exists a unique decomposition of f into E-states:

$$f(a) = \int g_\xi(a)\, d\mu(g_\xi) \quad ,$$

where the g_ξ are E-states.

Proof: See Kastler and Robinson, C.M.P. 3, 151 (1966), th. 2.

This theorem is therefore an answer, even if only partial, to the question asked above. The two last theorems are in fact not essentially different, the hypothesis of the first one implying the separability of $\mathfrak{H}_f$.

As far as the physical interpretation is concerned, it may be useful to develop some more properties of the E-states. Indeed, it is possible to show (cf. Kastler and Robinson, loc. cit.) that for an E-state f, and the corresponding $\mathfrak{H}_f$,

$$M\Big[(y, \{\pi_f(\tau_x a) - (x_f, \pi_f(a)x_f)\} z) \Big] = 0$$

$\forall\, a \in \mathfrak{A}$, $\forall\, y, z \in \mathfrak{H}_f$.

If one puts $y = z$, we get explicitly

$$\lim_{V\to\infty} \frac{1}{V} \int_V dx (y, \pi_f(\tau_x a) y) = (x_f, \pi_f(a) x_f)(y, y)$$

This means that in the Hilbert space of the Gelfand construction induced by f, all vector states describe states with at most local variations from the equilibrium, so that the mean of the measurements over all space is independent of the particular vector chosen.

It is further possible to show that

$$M\Big[\| \{\pi_f(\tau_x a) = (x_f, \pi_f(a)x_f)\} y \| \Big] = 0, \quad \forall\, y \in \mathfrak{H}_f, \ \forall\, a \in \mathfrak{A}$$

But this last statement exactly amounts to saying that the fluctuations of the spatial means of quasilocal observables vanish, which is a characteristic of system at the equilibrium point with a single thermodynamical phase.

II.E. Classification of Ergodic States

In the preceding section, we did discuss the role of the E-states and their relations to the equilibrium. There is, however, another way of looking at the problem, and that is to look at the spectrum of the infinitesimal generators of the translation in the representation induced by a given E-state.

II.E.1. We shall consider the following objects: Let T be the group of translations in R^S, $\mathfrak{A}$ a C*-algebra, T-asymptotically abelian, and

f a T-invariant state on $\mathfrak{A}$. Let further π_f be the Gelfand homomorphism defined by f, $\mathfrak{H}_f$ the corresponding Hilbert space, $u_f(x)$ a unitary representation of T on $\mathfrak{H}_f$, x_f the cyclic vector of $\mathfrak{H}_f$ (by respect to π_f) and invariant under $u_f(x)$. We already know that in the particular case where f is an E-state, x_f is the only vector of $\mathfrak{H}_f$ invariant under the $u_f(x)$.

We have the following theorem:

Theorem: We may write

$$u_f(x) = \sum_{p_n \in S_D} E(p_n)\, e^{ip_n x} + \int_{S_c} e^{ipx}\, dE(p)$$

where the sum corresponds to the discrete spectrum and the integral to the continuous one. The discrete spectrum is a subgroup of T, that is, if $p,q \in S_D$, then $p+q$ and $-p \in S_D$. Further, the projection $E(p_n)$ are one-dimensional.

Proof: See Kastler and Robinson, C.M.P. 3, 151 (1966).

This theorem enables us to give the following definitions

II.E.2. Definition: Let f be an E-state, x_f the T-invariant vector of $\mathfrak{H}_f$, cyclic under $\pi_f(\mathfrak{A})$, $u_f(x)$ the unitary representation of T on $\mathfrak{H}_f$. Then

i) f is called an E_I-state, if x_f is the only eigenvector of u_f (that is, S_D contains only the origin)

ii) f is called an E_{II}-state, if the spectrum S_D of u_f spans a a m-dimensional subspace of R^n, and if there is a nonvanishing minimal distance between any two points of S_D.

iii) f is called an E_{III}-state if it is not a E_I-state nor an E_{II}^m-state.

Remark that the condition ii) means that S_D has no accumulation point, thus for an E_{III}-state, S_D has an accumulation point.

As will become clearer after some examples, the idea would be to interpret E_I-states as corresponding to fluid phases, and E_{II}-states as corresponding to a lower symmetry, for instance a crystal state. There is no intuitive interpretation of the E_{III}-states as of now, but one can construct examples.

II.E.3. In order to discuss E_{III}-states, we are going to introduce a decomposition of these states into states invariant only under a subgroup of T, which we shall denote by T_L (by analogy to the discrete translations on a lattice). T_L is defined as being the subgroup which leaves the eigenvector of the representation of T invariant. It is clear that the E_{III}-state is invariant under T_L, but it is not extremal with

respect to T_L. We introduce the group $T/T_L = k$ of the translations modulo T_L. Let us call $\tilde{T}$ the dual group of T ("impulsion space"), S_D appears as a discrete subgroup of $\tilde{T}$, and the group T_L may be interpreted as the annihilator of S_D. In other words, T_L is isomorphic to the dual of S_D. S_D being discrete, k is compact.

II.E.4. Example: Take the particular case $T = R^2$. The E_{II}-states possible are E_{II}^1-states and E_{II}^2-states.

Consider first an E_{II}^1-state. S_D consists then in equally spaced points on a line, the equality of the distances following from the additivity of the eigenvalues. If we choose cartesian coordinates (x_1, x_2) in R^2, and (p_1, p_2) in $\tilde{T}$, we can suppose, without loss of generality, that $S_D = \{(aN, 0)\}$, where N runs over the integers and a is a fixed length. The associated subgroup T_{L1}^1 is the group of discrete translation in the direction x_1, where M runs over the integers.

Therefore T_L is the group of translations whose component in the direction x_1 is an integer multiple of $2\pi/a$, and the component in the direction x_2 is arbitrary.

The quotient group k consists thus in the translations in the direction x_1 modulo $2\pi/a$. It is therefore isomorphic to the group of rotations, modulo 2π of a circle of radius $1/a$. A function g over T_L can be considered as a sequence of functions $g_M(x_2)$, and the integral with respect to the Haar measure on T_L is given by

$$\sum_{M=-\infty}^{+\infty} \int_{-\infty}^{+\infty} g_M(x_2)\, dx_2$$

If we now consider an E_{II}^2-state, T_L consists of the translations of the form $(2\pi M_1/a_1, 2\pi M_2/a_2)$, where a_1 and a_2 are minimal fixed distances of S_D, and where M_1 and M_2 are integers. In this case, k is isomorphic to the product of two groups describing each the rotation on a circle. A function g on T_L is then given by

$$\sum_{M_1=-\infty}^{+\infty} \sum_{M_2=-\infty}^{+\infty} g_{M_1 M_2}$$

In general, we have the

II.E.5. Theorem: Let f be an E_{II}-state with respect to T. There exists a unique decomposition of f into states f_ξ, invariant and weakly clustering with respect to T_L (that is, E-states with respect to T_L).

This decomposition may be written using the Haar measure $dm(\xi)$ on the quotient group $K = T/T_L$,

$$f = \int_K f_\xi \, dm(\xi)$$

so that the states f are mapped onto each other by k, that is,

$$f_{\xi+\eta}(a) = f_\xi(\tau_y a)$$

where y is one representant of the class $\eta \in k$.

Furthermore, the f_ξ are E_I-states with respect to T_L. (These are sometimes called L-states.)

II.E.6. The generalisation of these results, obtained up to now for the group of translations, to a noncompact, locally compact group G is possible.

Let, therefore, G be a C*-algebra, weakly G-asymptotically abelian, f a G-invariant state on $\mathfrak{A}$, $\pi_f(\mathfrak{A})$ the Gelfand representation induced by f in $\mathfrak{H}_f$. Let further $u_f(G)$ be the unitary (strongly continous) representation of G on $\mathfrak{H}_f$. Let also e_f be the smallest projector in $\mathfrak{H}_f$, whose range contains all finite dimensional subspace invariant under $u_f(G)$. e_f is certainly nonzero, since $x_f \in e_f \mathfrak{H}_f$. Let finally N_f be the kernel of the representation of G on $e_f\mathfrak{H}_f$, and $\mathcal{G}_f = G/N_f$.

Definition:

i) f is an E_I-state, if $\mathcal{G}_f$ is reduced to the identity element.

ii) f is an E_{II}-state, if it is not an E_I-state, and if $\mathcal{G}_f$ is compact.

iii) f is an E_{III}-state, if $\mathcal{G}_f$ is not compact.

The connection with the definition given above is easy to make.

The decomposition of E_{II}-states is then expressed as follows

Theorem: If $\mathfrak{A}$ is G-asymptotically abelian, with an identity element, G being s-continuously represented ($g \to \alpha_g$) in the group of the automorphisms of $\mathfrak{A}$, then an E-state f (with respect to G) on $\mathfrak{A}$ is an E_{II}-state (with respect to G) if and only if there exists a maximal subgroup $H \subset G$, such that G/H is compact, and there exists a unique state f_o, which is an E_I-state with respect to H and such that

$$f(a) = \int_{G/H} f_o(\alpha_g' a) \, d\mu(g)$$

where $\mu(g)$ is the Haar measure on G/H, normalized to 1, and $\alpha \to \alpha'$ is the canonical homomorphism of G onto G/H.

In the example considered above, we had that

$$G = R^3 , \quad H = Z^1 \ (\text{or } Z^3), \quad d\mu(g) = dx, \quad x \in [0,1]$$

If f is an E_I-state with respect to H, then evidently $f = f_0$. In all other cases, $f_0 \neq f$, and f_0 is not invariant under G. We therefore conclude that an E_{II}-state may uniquely be represented as a mean over states possessing a lower symmetry.

II.F. Quantum Spin Systems

II.F.1. Let us consider a lattice Z^ν, and associate to each point $x \in Z^\nu$ an Hilbert space $\mathfrak{H}_x$. We shall suppose that all these $\mathfrak{H}_x$ are isomorphic to an $\mathfrak{H}$ of finite dimension, N + 1.

Remark that we can understand that to each point x, an "occupation number" n_x is associated, $n_x = 0, 1, \ldots, N$ and one can then think of $\frac{1}{2}(n_x - N)$ as the value of the component of a spin at x, hence the name often given of quantum spin systems, or, sometimes, lattice gas.

Let now $\mathfrak{G}$ be a finite set of points x_ξ. $\mathfrak{G} = \{x_\xi\}$. To $\mathfrak{G}$ we associate the direct product space $\mathfrak{H}_\mathfrak{G}$ by

$$\mathfrak{H}_\mathfrak{G} = \bigotimes_{x_\xi \in \mathfrak{G}} \mathfrak{H}_{x_\xi}$$

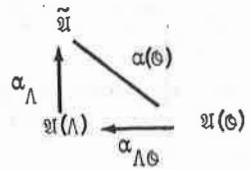
A translation by $a \in Z^\nu$ is described by $\tau_a \mathfrak{H}_x = \mathfrak{H}_{a+x}$ and $\tau_a \mathfrak{H}_\mathfrak{G} = \mathfrak{H}_{\mathfrak{G}+a} \equiv \bigotimes_{x \in \mathfrak{G}} \mathfrak{H}_{x+a}$.

To each finite set of points $\mathfrak{G} \subset Z^\nu$, we let correspond the algebra $\mathfrak{A}(\mathfrak{G}) \equiv \mathfrak{L}(\mathfrak{H}_\mathfrak{G})$.

If $\mathfrak{G} \subset \Lambda$, then $\mathfrak{H}_\Lambda$ is isomorphic to $\mathfrak{H}_\mathfrak{G} \otimes \mathfrak{H}_{\Lambda/\mathfrak{G}}$, and we define the the canonical imbedding $\alpha_{\Lambda\mathfrak{G}}$ of $\mathfrak{A}(\mathfrak{G})$ in $\mathfrak{A}(\Lambda)$ by

$$\alpha_{\Lambda\mathfrak{G}} A = A \otimes 1_{\Lambda/\mathfrak{G}} , \quad \forall A \in \mathfrak{A}(\mathfrak{G})$$

This last application may be used to define the relation between the quasilocal algebra, and the algebras $\mathfrak{A}(\Lambda)$, in defining α_Λ by the commutative diagram



that means that we can consider $\tilde{\mathfrak{A}}$ as being $\tilde{\mathfrak{A}} = \bigcup_{\Lambda} \alpha_{\Lambda}\, \mathfrak{A}(\Lambda)$, which is a normed algebra, with norm $\|\alpha_{\Lambda}A\| = \|A\|$ and involution $(\alpha_{\Lambda}A)^* = \alpha_{\Lambda}A^*$.

Let $\mathfrak{A}$ be the norm-completion of $\tilde{\mathfrak{A}}$. $\mathfrak{A}$ is a C*-algebra. The natural representation of the translations is given by

$$\tau_a : \mathfrak{A}(\Lambda) \to \mathfrak{A}(\Lambda + a), \quad \tau_a a = W_a\, A\, W_a^{-1},$$

with $W_a = V_{x+a}\, V_x^{-1}$, and V_x being the isomorphism between $\mathfrak{H}$ and $\mathfrak{H}_x$:

$$V_x \mathfrak{H} = \mathfrak{H}_x$$

We shall always write $\mathfrak{A}(\mathfrak{G}) \subset \mathfrak{A}$ in what follows, but this has to be interpreted as meaning $\alpha_{\mathfrak{G}}\mathfrak{A}(\mathfrak{G}) \subset \mathfrak{A}$.

Note that if $\mathfrak{G}_1$ and $\mathfrak{G}_2$ are disjoint, $\mathfrak{A}(\mathfrak{G}_1)$ and $\mathfrak{A}(\mathfrak{G}_2)$ commute (as subalgebras of $\mathfrak{A}$) as they are factors of type I, we see that the general discussion given above applies to this system; in particular $\mathfrak{A}$ is simple, T-asymptotically abelian.

II.F.2. We want to introduce now some forms of interaction, and we shall do this by introducing very general n-body potentials.

The interaction Φ is a function on the finite subsets of Z^{ν}, which satisfies

i) $\Phi(\mathfrak{G})$ is a self-adjoint element of $\mathfrak{A}(\mathfrak{G})$.

ii) $\Phi(\phi) = 0$ $\quad \phi$ = empty set.

iii) $\Phi(\mathfrak{G} + a) = \tau_a \Phi(\mathfrak{G})$

iv) $|\Phi| = \sum_{\substack{\mathfrak{G} \text{ finite} \\ o \in \mathfrak{G}}} \|\Phi(\mathfrak{G})\| \cdot N(\mathfrak{G})^{-1} < \infty$

with $N(\mathfrak{G})$ = number of points of $\mathfrak{G}$.

We introduce further the Hamiltonian $H_{\Lambda} = \sum_{\mathfrak{G} \subset \Lambda} \Phi(\mathfrak{G})$, Λ finite $\subset Z^{\nu}$. One would hope that, although $\lim_{\Lambda \to \infty} H_{\Lambda}$ doesn't exist in any sense,

$$\lim_{\Lambda \to \infty} e^{iH_{\Lambda}t}\, A\, e^{-iH_{\Lambda}t}, \quad A \in \mathfrak{A},$$

would make sense, and we would denote it by $\bar{\tau}_t A$. Unfortunately, iv) is too weak for proving that, so that we shall make a somehow stronger hypothesis:

$$\text{iv}') \quad \|\Phi\| = \sum_{o \in \mathfrak{G}} \|\Phi(\mathfrak{G})\| \exp\,(N(\mathfrak{G}) - 1) < \infty$$

and we shall finally denote by $\mathfrak{B}$ the Banach space of the functions Φ satisfying i), ii), iii), iv').

II.F.3. As an example, it is perhaps useful to look at the Ising model in two dimensions.

We have $\mathbb{Z}^3$, the spin component takes two values, thus $\mathfrak{H}_x$ is 2-dimensional. We can therefore describe the elements of $\mathfrak{H}_x$ by "spin up" as (1,0) and "spin down" as (0,1).

In the Ising model, the interaction energy (and the potential) is described, up to a positive factor, by $\mu_1 \cdot \mu_2$ between the nearest neighbours (μ takes the value +1 for spin up, and -1 for spin down).

$\Phi(\mathfrak{G})$ is therefore defined as

$$\Phi(\mathfrak{G}) = 0 \text{ if } \begin{cases} \text{i)} & N(\mathfrak{G}) \geq 3 \text{ or } N(\mathfrak{G}) = 1 \\ \text{ii)} & N(\mathfrak{G}) = 2 \text{ and the distance between } x_1 \text{ and } x_2 \in \mathfrak{G} \\ & \text{is larger than one lattice unit.} \end{cases}$$

in the remaining cases, we can write explicitly

$$\Phi(\mathfrak{G})[(1,0) \otimes (1,0)] = (1,0) \otimes (1,0)$$

$$\Phi(\mathfrak{G})[(1,0) \otimes (0,1)] = -(1,0) \otimes (0,1)$$

$$\Phi(\mathfrak{G})[(0,1) \otimes (1,0)] = -(0,1) \otimes (1,0)$$

$$\Phi(\mathfrak{G})[(0,1) \otimes (0,1)] = (0,1) \otimes (0,1)$$

which could be more elegantly written in a matrix form with

$$\Phi(\mathfrak{G}) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

The properties i), ii), iii) and iv') are easy to prove.

II.F.4. The natural definition for a "local" Hamiltonian H_Λ, corresponding to a finite region would be to put

$$H_\Lambda = \sum_{\mathfrak{G} \subset \Lambda} \Phi(\mathfrak{G})$$

Note that there is no "free" part. This is an essential feature of the lattice gas, and in fact what makes it both within reach of our techniques and of little physical physical content!

It is clear that the restrictions we imposed on the $\Phi(\mathfrak{G})$ are sufficient to ensure the existence of H_Λ as an operator on $\mathfrak{H}_\Lambda$. It is also clear that the limit $\Lambda \to \infty$ (that is tending to cover the whole lattice, what we always suppose to be done in a uniform way) does not exist.

There exists, however, other quantities of interest than the Hamiltonian itself, and, for instance, the first question we could ask is whether the dynamics, which supposedly should be generated by the Hamiltonian, still exist. For that, we want to prove that

$$\tau_t A = \lim_{\Lambda \to \infty} e^{iH_\Lambda t} A\, e^{-iH_\Lambda t}$$

exists, for all $A \in \mathfrak{A}$. If we were to start with an $A \in \mathfrak{A}(\mathfrak{G})$ for some finite $\mathfrak{G}$, and expanding $e^{iH_\Lambda t} A\, e^{-iH_\Lambda t}$ in series, we would get the expansion

$$\tau_t A = \lim_{\Lambda \to \infty} \sum_{n=0}^{\infty} \frac{(it)^n}{n!} \Omega_n(H_\Lambda, A)$$

where

$$\Omega_n(H_\Lambda, A) = [H_\Lambda, \Omega_{n-1}(H_\Lambda, A)], \quad \Omega_0(H_\Lambda, A) = A$$

Because of the local properties of the Hamiltonian, we observe that if Λ gets large, $\Omega_n(H_\Lambda, A)$ only depends upon Λ via N-body forces, where N is nearly $N(\Lambda)$. But these forces are necessarily very weak, because of our condition iv'). We therefore expect that, term by term, our perturbation expansion no longer depends upon Λ, for Λ sufficiently large. In quantitative terms, we have the

II.F.5. Lemma: Let $\Phi \in \mathfrak{B}$, $A \in \mathfrak{A}(\Lambda_1)$, $\Lambda_1 \subset \Lambda$. Then the following estimate holds:

$$\|\Omega_n(H_\Lambda, A)\| \leq \|A\| \cdot \exp(N(\Lambda_1)) \cdot n! (2\, \|\Phi\|)^n$$

Proof: We start from the fact that we can write

$$\Omega_n(H_\Lambda, A) = \sum_{\mathfrak{S}_1 \subset \Lambda} \cdots \sum_{\mathfrak{S}_n \subset \Lambda} [\Phi(\mathfrak{S}_n), [\ldots, [\Phi(\mathfrak{S}_1), A]\ldots]]$$

Because of "locality," the commutators vanish for regions of disjoint supports, we can therefore extend the sums to infinite ones by adding vanishing terms:

If $\mathfrak{S}_1 \cap \Lambda_1 = \Phi \Rightarrow [\Phi(\mathfrak{S}_1), A] = 0$

If $\mathfrak{S}_2 \cap (\mathfrak{S}_1 \cup \Lambda_1) = \phi \Rightarrow [\Phi(\mathfrak{S}_2), [\Phi(\mathfrak{S}_1), A]] = 0$, etc., and we write

$$\sum_{\mathfrak{S}_i \subset \Lambda} = \sum_{\mathfrak{S}_i} = \sum_{k_i=1}^{\infty} \sum_{\mathfrak{S}:\ N(\mathfrak{S})=k_i}$$

We therefore get

$$\|\Omega_n(H_\Lambda, A)\| \leq \|A\|\, 2^n \sum_{k_1 \ldots k_n} \left\{ \prod_{i=1}^{n} (N(\Lambda_1)+k_1+\ldots+k_{i-1}) \times \sum_{\substack{\mathfrak{S}_i \ni 0 \\ N(\mathfrak{S}_i)=k_i+1}} \|\Phi(\mathfrak{S}_i)\| \right\}$$

Now

$$\prod_{i=1}^{n} (N(\Lambda_1)+k_1+\ldots+k_{i-1}) \leq (N(\Lambda_1)+k_1+\ldots+k_n)^n$$

$$\leq n!\, \exp(N(\Lambda_1)+k_1+\ldots+k_n)$$

$$\leq n!\, \exp(N(\Lambda_1)) \prod_{i=1}^{n} \exp(k_i)$$

And thus

$$\|\Omega_n(H_\Lambda, A)\| \leq 2^n \|A\| n! \exp(N(\Lambda_1)) \sum_{k_1 \ldots k_n} \prod_{i=1}^{n} \sum_{\substack{\mathfrak{G}_i \ni 0 \\ N(\mathfrak{G}_i)=k_i+1}} e^{k_i} \|\Phi(\mathfrak{G}_i)\|$$

$$\leq 2^n \|A\| n! \exp(N(\Lambda_1)) \cdot \|\Phi\|^n \qquad \text{QED}$$

II.F.6. Theorem:

$$\bar{\tau}_t(A) = \lim_{\Lambda \to \infty} e^{iH_\Lambda t} A\, e^{-iH_\Lambda t}$$

exists $\forall\ A \in \mathfrak{A}(\mathfrak{G})$, $\mathfrak{G}$ finite, and $|t|$ sufficiently small, in the norm topology, and the convergence is uniform for small $|t|$. It defines uniquely a 1-parameter group of automorphisms on the whole of $\mathfrak{A}$.

Proof: (Sketch!)

1) In the perturbative expansion of $\bar{\tau}_t A$, each term converges (by lemma II.F.5) uniformly in Λ. Because of the estimate, given in lemma II.F.5, the series converges for $|t|$ small, $|t| < (2\|\Phi\|)^{-1}$.

2) One extends $\bar{\tau}_t$ to the whole of $\mathfrak{A}$ by continuity

3) If $|t_1|$, $|t_2|$ and $|t_1+t_2| < (2\|\Phi\|)^{-1}$, we then have the group property

$$\bar{\tau}_{t_1+t_2} A = \bar{\tau}_{t_1}\, \bar{\tau}_{t_2}\, A$$

using the uniformity of the convergence. One can then show that the automorphism of $\mathfrak{A}$ generated in such a way may be extended to all real t, with the same group property.

II.F.7. The standard way of defining an equilibrium state in statistical mechanics is to take the limit of a sequence of Gibbs grand canonical ensembles corresponding to increasing volumes:

$$\rho^\Phi(A) \equiv \lim_{\Lambda \to \infty} \langle A \rangle = \lim_{\Lambda \to \infty} \frac{\mathrm{Tr}_\Lambda [e^{-H_\Lambda} A]}{\mathrm{Tr}_\Lambda\, e^{-H_\Lambda}}$$

where Tr_Λ means the trace in the space $\mathfrak{H}_\Lambda$.

It is possible to show the following

Theorem: 1) $\rho^\Phi(A)$ exists on $\mathfrak{A}$

2) $\rho^{\Phi}(\bar{\tau}_t A) = \rho^{\Phi}(A)$, $\forall A \in \mathfrak{A}$

3) If $A, B \in \mathfrak{A}$, $\exists$ F, a bounded continuous function, defined on $\{z \mid z \in \mathbf{C},\ 0 \leq \mathrm{Im}\ z \leq \beta\}$, analytic in the strip $0 < \mathrm{Im}\ z < \beta$, and such that

$$\rho^{\Phi}(B\,\bar{\tau}_t A) = F(t) \quad \text{and} \quad \rho^{\Phi}(\bar{\tau}_t(A)B) = F(t + i\beta)$$

This last property is the Kubo-Martin-Schwinger boundary condition.

II.F.8. As far as the thermodynamical quantities are concerned, one can construct

$$\mathfrak{Z}(\mathfrak{G},\Phi) = \mathrm{Tr}_{\mathfrak{G}}\left[e^{-H_{\mathfrak{G}}}\right]$$

$$F(\mathfrak{G},\Phi) = \log \mathfrak{Z}(\mathfrak{G},\Phi)$$

$$P(\mathfrak{G},\Phi) = N(\mathfrak{G})^{-1} F(\mathfrak{G},\Phi)$$

One has the following.

Theorem: For $\Phi \in \mathfrak{B}$, the limit

$$P(\Phi) \equiv \lim_{N(\mathfrak{G})\to\infty} N(\mathfrak{G})^{-1} F(\mathfrak{G},\Phi) = \lim_{N(\mathfrak{G})\to\infty} P(\mathfrak{G},\Phi)$$

exists, and the function $\Phi \to P(\Phi)$ is convex and continuous on the Banach space $\mathfrak{B}$,

$$|P(\Phi) - P(\psi)| \leq \|\Phi - \psi\|$$

Proof: See Robinson CMP 7, 337 (1968).

If we now put $\beta = \frac{1}{kT}$ and define $\Xi_{\mathfrak{G}}(\beta,\Phi) = \mathfrak{Z}(\mathfrak{G},\beta\Phi)$ then Ξ can be interpreted as the partition function corresponding to the set $\mathfrak{G}$ and the temperature $T = (\beta k)^{-1}$. The free energy is then

$$p(\beta,\Phi) \equiv \frac{1}{\beta} P(\beta,\Phi) = \frac{1}{\beta} \lim_{N(\mathfrak{G})\to\infty} N(\mathfrak{G})^{-1} \log \Xi_{\mathfrak{G}}(\beta,\Phi)$$

II.F.9. The few points we discussed here have to be taken as a small sample of what can actually be done with these techniques. We hope, however, that this illustration will be sufficient to induce many to read the original literature.

Section III. Constructive Quantum Field Theory

Roughly speaking, constructive quantum field theory represents an attempt to use the knowledge gained in working in the different "axiomatic" approaches, for a head on attack of the basic problem of quantum field theory.

The starting point is certainly the 1964 Cargese lectures of Wightman; then came the thesis of Lanford and Jaffe, and work by the present author, all at Princeton. The decisive steps were made by Glimm and Jaffe in a series of fundamental papers, where they use another stream of ideas, coming mainly from Friedrichs (Perturbation of Spectra in Hilbert Spaces) and results of Nelson, Segal, Federbusch and others.

The present situation is probably best exposed, together with a lot of new results, in the beautiful 1968-69 Paris lectures of Hepp, from which we shall borrow quite a lot.

In the few remaining lectures, we plan to give a very short introduction to this subject, putting the accent on the main underlying ideas rather than on techniques. One should be aware, however, that the technicalities involved in this approach are by no means trivial, and that they make use of a great number of subtle tricks and refined mathematical analysis, and that no real understanding of the subject can be gained without mastering them.

III.A. Fock Space Techniques

III.A.1. In what follows, $\underline{x} \in \mathbb{R}^s$, $x \in \mathbb{R}^{s+1}$, s being the number of space dimensions $(x,y) = x^0y^0 - \underline{x}\cdot\underline{y}$, $\hbar = c = 1$.

$$\mu(\underline{k}) = +(m^2 + \underline{k}^2)^{\frac{1}{2}}$$

$\mathfrak{F}$ = Fock space;, $\mathfrak{F} = \bigoplus_{n=0}^{\infty} \mathfrak{F}_n$,

$$\varphi \in \mathfrak{F} \Leftrightarrow \varphi = \{\varphi_n | \varphi_n \in \mathfrak{F}_n, \quad \sum_{n=0}^{\infty} \|\varphi_n\|^2 < \infty\}$$

$\varphi_n(\underline{k}_1, \ldots, \underline{k}_n)$, symmetrical, $(\varphi_n, \varphi_n) = \|\varphi_n\|^2 = \prod_{i=1}^{n} \int \frac{d\underline{k}_i}{2\mu_i} |\varphi_n(\underline{k}_1 \ldots \underline{k}_n|^2$

$\mathfrak{F}_0 = \mathbb{C}$, $\mathfrak{F}_1 = \mathfrak{L}^2(d\Omega)$, $d\Omega(\underline{k}) = \frac{d^s\underline{k}}{2\mu(\underline{k})}$

One has the canonical representation of the commutation relations (CCR) in $\mathfrak{F}$:

$$[a(\underline{k}),\ a(\underline{\ell})] = 0$$

$$[a(\underline{k}),\ a^*(\underline{\ell})] = \delta(\underline{k}-\underline{\ell})$$

if we put for the annihilation operator $a(\underline{k})$ and the creation operator $a^*(\underline{k})$:

$$(a(\underline{k})\,\varphi_n)(\underline{k}_1,\ldots,\underline{k}_n) = \left(\frac{n}{2\mu(\underline{k})}\right)^{\frac{1}{2}} \varphi_n(\underline{k}_1,\ldots,\underline{k}_{n-1},\underline{k})$$

or, if one prefers, for $\mu^{\frac{1}{2}} f \in \mathfrak{L}^2(d\Omega) \Leftrightarrow f \in \mathfrak{L}^2(\mathbb{R}^s)$

$$(a(f)\,\varphi_n)(k_1,\ldots,k_{n-1}) = \sqrt{n}\int \frac{d\underline{k}}{\sqrt{2\mu(\underline{k})}} f(\underline{k})\,\varphi_n(\underline{k}_1,\ldots,\underline{k}_{n-1},\underline{k})$$

and, similarly

$$(a^*(\underline{k})\,\varphi_n)(\underline{k}_1,\ldots,\underline{k}_{n+1}) = \left(\frac{2\mu(\underline{k})}{n+1}\right)^{\frac{1}{2}} \sum_{j=1}^{n} \delta(\underline{k}-\underline{k}_j)\,\varphi_n(\underline{k}_1,\ldots,\hat{\underline{k}}_j,\ldots\underline{k}_n)$$

From these definitions, follows that

$$\varphi_n = (n!)^{-\frac{1}{2}} \int d\underline{k}_1\ldots d\underline{k}_n\,\varphi_n(\underline{k}_1\ldots\underline{k}_n)a^*(\underline{k}_1)\ldots a^*(\underline{k}_n)\varphi_0$$

and, for $f \in \mathfrak{L}^2(\mathbb{R}^s)$

$$a(f) \in \mathfrak{L}(\mathfrak{F}_n,\ \mathfrak{F}_{n-1}); \qquad a^*(f) \in \mathfrak{L}(\mathfrak{F}_n,\ \mathfrak{F}_{n+1})$$

with

$$\left.\begin{aligned} \|a(f)\|_{n,n-1} &\le \sqrt{n}\ \|f\|_2 \\ \|a^*(f)\|_{n,n+1} &\le \sqrt{n+1}\ \|f\|_2 \end{aligned}\right\} \qquad (A.1)$$

III.A.2. We define on $\mathfrak{F}$ the (unbounded) operator "number of particles" N by

$$N\varphi_n = n\varphi_n$$

It follows from the above bounds (A.1) that for $f \in \mathfrak{L}^2(\mathbb{R}^s)$

$$a^{\#}(f)(N+1)^{-\frac{1}{2}} \in \mathcal{L}(\mathfrak{F}, \mathfrak{F})$$

with

$$\|a^{\#}(f)(N+1)^{-\frac{1}{2}}\| \leq \|f\|_2$$

and where $a^{\#}$ stands for either a^* or a. More generally, we have the
Lemma: For $w \in \mathcal{L}^2(\mathbb{R}^{sm})$, $1 \leq i \leq m$

$$W = \int dk\, w(k_1, \ldots, k_m) a^*(k_1) \ldots a^*(k_i) a(k_{i+1}) \ldots a(k_m) \quad (A.2)$$

is defined on $D(N^{m/2})$ and

$$\|W(N+1)^{-m/2}\| \leq \|w\|_2$$

and we leave the proof as an exercise.

III.A.3. W is called a Wick monomial, with numerical kernel $w(k_1, \ldots, k_m)$ and operator kernel

$$w(k_1, \ldots k_m) a^*(k_1) \ldots a^*(k_i) a(k_{i+1}) \ldots a(k_m) \equiv w(k) : \prod_j a^{\#}(k_j) :$$

The annihilation operators always stand to the right of creation operators, and the order of the creation with respect to the annihilation operators among themselves is immaterial.

If $w \in \mathcal{S}'(\mathbb{R}^{sn})$, W formally defined by the expression (A.2) has a meaning as a mapping

$$W: \mathcal{S} \to \mathcal{S}'$$

where

$$\mathcal{S} = \bigoplus_{n=0}^{\infty} \mathcal{S}(\mathbb{R}^{sn}) \subset \mathfrak{F} \subset \mathcal{S}' = \prod_{n=0}^{\infty} \mathcal{S}'(\mathbb{R}^{sn})$$

because of the nuclear theorem (or as a sesquilinear form on $\mathcal{S} \times \mathcal{S}$).

The domain D(W) is defined by

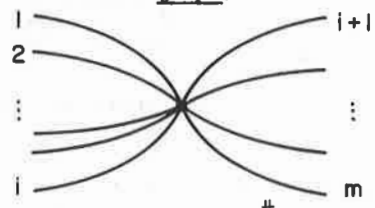
$$D(W) = \left\{ \varphi \,|\, \varphi \in \mathfrak{F}, \sum_{n=0}^{\infty} \|(W\varphi)_n\|^2 < \infty \right\}$$

Lemma: If $i = m$ in (A.2),

$$D(W) \neq \{0\} \Leftrightarrow \mathcal{S}'(\mathbb{R}^{sm}) \ni w$$

The proof is again left as an exercise.

III.A.4. For the operator kernel of W, and by an abuse of language, for W itself, we introduce a graph:



where the line "j" lies to the left if $a^{\#}(\underline{k}_j)$ equals $a^*(\underline{k}_j)$ and to the right, if $a^{\#}(\underline{k}_j)$ equals $a(\underline{k}_j)$.

III.A.5. The formal product of two Wick monomials may always be expressed as a sum of Wick monomials (Wick polynomial) as follows from the well-known Wick Theorem

Theorem: If $V = \int d\underline{k}\, v(\underline{k})\, a^*(\underline{k}_1)\ldots a^*(\underline{k}_i)a(\underline{k}_{i+1})\ldots a(\underline{k}_m)$

$$W = \int d\underline{k}\, w(\underline{k})\, a^*(\underline{\ell}_1)\ldots a^*(\underline{\ell}_j)a(\underline{\ell}_{j+1})\ldots a(\underline{\ell}_n)$$

Then

$$VW = \sum_{CS} \int d\underline{k}\, d\underline{\ell}\, v(\underline{k})\, w(\underline{\ell}) \times$$

$$:a^*(\underline{k}_1)\ldots a^*(\underline{k}_i)a(\underline{k}_{i+1})\ldots a(\underline{k}_m)a^*(\underline{\ell}_1)\ldots a^*(\underline{\ell}_j)a(\underline{\ell}_{j+1})\ldots a(\underline{\ell}_n):$$

$$= \sum_{\alpha \in CS} : \underbrace{VW}_{\alpha} :$$

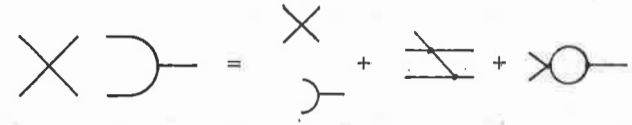
where $\sum_{CS}$ extends over all "contraction systems" between pairs:

$$(a(\underline{k}_{r_1}),\, a^*(\underline{\ell}_{s_1})) = \delta(\underline{k}_{r_1} - \underline{\ell}_{s_1}),\ldots,(a(\underline{k}_{r_t}), a^*(\underline{\ell}_{s_t})) = \delta(\underline{k}_{r_t} - \underline{\ell}_{s_t})$$

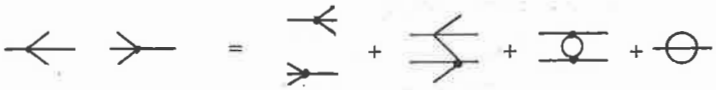
$0 \leq t \leq \min\{m-i, j\}$.

The proof of this theorem is well known and will not be repeated here. It is obtained in using the CCR.

It is useful to represent graphically the theorem of Wick. One gets for instance



or



and so on. Note that the exterior lines must always be to the right or to the left.

III.A.6. After integration over the δ functions (which is well defined if the multiplication of the kernels is), there are two ways of representing the kernel of $:\overbrace{VW}^{\alpha}:$:

$$a)\ a^*(\underline{k}_1)\ldots\overbrace{a^*(\underline{\ell}_{s_1})}\ldots a^*(\underline{\ell}_j)a(\underline{k}_{i+1})\ldots\overbrace{a(\underline{k}_{r_1})}\ldots a(\underline{\ell}_n)$$

$$\cdot v(\underline{k}_1\ldots\underline{k}_{r_1}\ldots\underline{k}_m)\ w(\underline{\ell}_1\ldots\underline{k}_{r_1}\ldots\underline{\ell}_n)$$

or

$$b)\ a^*(\underline{k}_1)\ldots\overbrace{a^*(\underline{\ell}_{s_1})}\ldots a^*(\underline{\ell}_j)a(\underline{k}_{i+1})\ldots\overbrace{a(\underline{k}_{r_1})}\ldots a(\underline{\ell}_n)$$

$$\cdot\int d\underline{k}_{r_1}\ldots v(\underline{k}_1\ldots\underline{k}_{r_1}\ldots\underline{k}_m)\ w(\underline{\ell}_1\ldots\underline{k}_{r_1}\ldots\underline{\ell}_n)$$

that is, before or after integration on the variables of the contracted lines. It should be clear that the representation a) is more suitable for introducing counterterms whereas the case b) is not an operator kernel.

III.A.7. Iteratively, one defines as Wick polynomial, the product $\prod_{j=1}^{n} W_j$; $\sum_{CS}$ extend over all compatible contractions $\{\alpha_{ij}\}$

$$\prod_{j=1}^{n} W_j = \sum_{\alpha\in CS} :\ \underbrace{W_1\ldots\underbrace{W_i\ldots\underbrace{W_n}_{\alpha_{in}}}_{\alpha_{1i}}}_{\alpha_{1n}}\ :$$

The numerical kernel of each term is deduced from $w_1 \otimes \ldots \otimes w_n$ by identification of certain variables, determined by $\{\alpha_{ij}\}$.

Each contraction system gives rise to a graph G, of vertices $V_1 \ldots V_n$, and ℓ lines $L_1 \ldots L_\ell$ ($\ell = \ell(G)$).

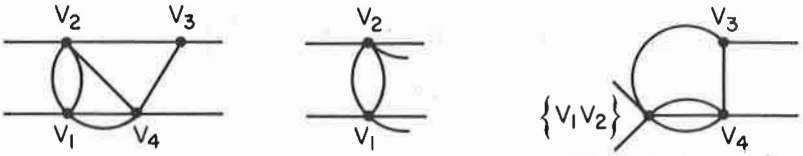
A subgraph $H \subset G$ is defined by a subset $\{V_1', \ldots, V_m'\} \subset \{V_1, \ldots, V_n\}$ with all lines of $\{L_1, \ldots L_\ell\}$ which are attached to its vertices.

One says that 2 subgraphs H_1 and H_2 are disjoint, if they don't have any vertex in common (but they may have common lines).

A line is said to be an interior line if it joins two vertices. In the reduced operator kernel, one has to integrate over the corresponding variable. Otherwise, it is called an exterior line.

The quotient of a graph G by a subgraph $H \subset G$ is defined by the identification of all vertices of H and the contractions of all interior lines of H.

Example:



Note that the topological equivalence class of graphs G is smaller than for Feynman graphs, the exterior lines having to go to the right or to the left, if they represent annihilators or creators.

Definition: A graph is said to be connected if all its vertices are connected by internal lines. It is said to be strongly connected, or one-particle irreducible, if the suppression of any line cannot make it non-connected.

III.B. Local Algebras Generated by Free Fields

III.B.1. We define as a local scalar neutral field on $\mathfrak{F}$ the operator valued distribution

$$\phi(\underline{x},t) = \frac{1}{(2\pi)^{s/2}} \int \frac{d\underline{k}}{(2\mu(\underline{k}))^{\frac{1}{2}}} \left\{ e^{i\underline{k}\underline{x}} \; a(\underline{k}) e^{-i\mu(\underline{k})t} + a^*(-\underline{k}) e^{i\mu(\underline{k})t} \right\}$$

We define as local algebra, $\mathfrak{R}(\mathcal{O})$ corresponding to an open region $\mathcal{O}$ of the s+1 dimensional Minkowsky space, the von Neumann algebra generated by all $\phi(f)$, $f \in \mathfrak{D}$, and the support of f being contained in $\mathcal{O}$.

$$\mathfrak{R}(\mathfrak{G}) = \{\phi(f) \,|\, f \in \mathfrak{D}(\mathbb{R}^{s+1}),\ \mathrm{Supp}\, f \subset \mathfrak{G}\}''$$

III.B.2. Let $\underline{\mathfrak{G}}$ be some open region of $\mathfrak{S}_t$, where $\mathfrak{S}_t$ is the hyperplane $x^0 = t$ (one could take a more general spacelike surface, but this does not bring anything essentially new). We shall suppose that the boundary of $\underline{\mathfrak{G}}$ is formed by piecewise many times continuously differentiable surfaces. Let $\mathfrak{G}$ be the double cone spanned by $\underline{\mathfrak{G}}$, that is,

$$\mathfrak{G} = \{x \,|\, (x-y, x-y) < 0,\ \forall y \in \mathfrak{S}_t,\ y \notin \underline{\mathfrak{G}}\} \equiv \underline{\mathfrak{G}}''$$

Let now $\mathfrak{R}(\underline{\mathfrak{G}}) = \{\phi(f), \phi(g) \,|\, f, g \in \mathfrak{D}(\mathbb{R}^s),\ \mathrm{Supp}\, f, g \subset \underline{\mathfrak{G}}\}''$ we have the following theorems due to Araki:

Theorem:

1) $\mathfrak{R}(\bigcup_i \mathfrak{G}_i) = \left\{ \bigcup_i \mathfrak{R}(\mathfrak{G}_i) \right\}''$

2) $\mathfrak{R}(\underline{\mathfrak{G}}) = \mathfrak{R}(\mathfrak{G})$, where $\mathfrak{G} = \underline{\mathfrak{G}}''$

3) $$\mathfrak{R}(\mathfrak{G}) = \bigcap_{\substack{i \\ \underline{\mathfrak{G}}_i \supset \underline{\mathfrak{G}}}} \mathfrak{R}(\underline{\mathfrak{G}}_i) = \bigcap_{\substack{i \\ \mathfrak{G}_i \supset \mathfrak{G}}} \mathfrak{R}(\mathfrak{G}_i) = \left\{ \bigcup_{\substack{i \\ \mathfrak{G}_i \subset \mathfrak{G}}} \mathfrak{R}(\mathfrak{G}_i) \right\}''$$

III.B.3. Theorem (Araki): $\underline{\mathfrak{G}}$, $\mathfrak{G}$ as before, $\mathfrak{G} = \mathfrak{G}''$; call

$$\mathfrak{G}' = \{x \,|\, (x-y, x-y) < 0,\ \forall y \in \mathfrak{G}\};\quad \mathfrak{G}' = \underline{\mathfrak{G}}'$$

Then $\mathfrak{R}(\mathfrak{G})' = \mathfrak{R}(\mathfrak{G}')$.

This is the famous "duality" theorem. Note that it is not true for arbitrary regions, but the condition $\mathfrak{G} = \mathfrak{G}''$ is in any case sufficient. In what follows, we shall call regions of the form $\mathfrak{G} = \mathfrak{G}''$, bounded by piecewise differentiable manifold, diamonds, and consider only such regions.

III.B.4. Theorem (Araki): $\mathfrak{G}$ as before, such that $\mathfrak{G}'$ is not empty. Then $\mathfrak{R}(\mathfrak{G})$ is a factor of type III.

III.B.5. It is also clear that the standards axioms of local algebras are satisfied, like locality, covariance, existence of the vacuum, etc.

III.C. Time Evolution of an Interacting System

III.C.1. As is well known, the requirements of translation invariance and of a nontrivial dynamics, that is the existence of a vacuum

polarization, exclude the possibility of having the total Hamiltonian defined on the Fock space $\mathfrak{H}$. This is Haag's theorem, and it can be stated in many different ways; one can, for instance, replace the translation invariance by Poincaré invariance, and then drop the assumption of a nontrivial vacuum polarization, and still get the result that the theory must be equivalent to a free field if the Hamiltonian is supposed to exist.

If we put ourselves in the Heisenberg picture, the time evolution is supposed to be given by

$$A_H(t) = e^{iHt} A(o)\, e^{-iHt}$$

where H is the total Hamiltonian. We know that in nontrivial cases, this Hamiltonian cannot exist, but this is not the really relevant question, because the (formal) solution to the field equations is $A_H(t)$, and it may be that $A_H(t)$ exists, without H existing; we have already seen such examples in statistical mechanics.

III.C.2. The easiest intuitive way of convincing ourselves that this might work is to use a perturbative expansion (Schwinger-Dyson): Suppose

$$H = H_o + H_I \text{ , put } H_I(t) = e^{iH_o t} H_I e^{-iH_o t}$$

$$\text{and } A(t) = e^{iH_o t} A(o)\, e^{-iH_o t}$$

$$A_H(t) = e^{iHt} A(o)\, e^{-iHt} = A(t) + i \int_o^t dt_1 \left[H_I(t-t_1),\, A(t)\right]$$

$$+ i^2 \int_o^t dt_1 \int_o^{t_1} dt_2 \left[H_I(t-t_1),\, \left[H_I(t-t_2), A(t)\right]\right]$$

$$+ \ldots \qquad \text{(C.1)}$$

Suppose now that A(o) is contained in $\mathfrak{R}(\underline{\mathcal{O}})$, where $\underline{\mathcal{O}}$ is some finite region of space, at time t = 0, contained in a sphere of radius a. Suppose further that H_I may be written in the form $H_I = \int :h_I:(\underline{x}) d^3\underline{x}$, where $:h_I:(\underline{x})$ is some local Wick polynomial in the (free) fields. H_I does not exist but it may happen that we are in the fortunate position of proving that H_{If} exists, where $H_{If} = \int :h_I:(\underline{x}) f(\underline{x}) d^3\underline{x}$ where $f(\underline{x}) \in \mathcal{D}(\mathbb{R}^3)$. We can take as a particular choice of f, the function g, $g \in D$, $g(\underline{x}) = 1$, for $|\underline{x}| \leq 1$, $g(\underline{x}) = 0$ for $|\underline{x}| \geq 2$, $g(\underline{x}) \geq 0$. We define then

$$H_{In} = \int :h_I:(\underline{x})\, g\left(\frac{\underline{x}}{n}\right) d^s\underline{x}$$

Consider, now, the first nontrivial term of the expansion (C.1) $:h_I:(x)$ being a Wick polynomial in the free fields, $H_{In}(t-t_1)$ has its support on the hyperplane $x^o = t - t_1$; $A(t)$ having its support on the hyperplane $x^o = t$. But, because of the local nature of the fields, the commutator of two fields vanishes outside the light cone in the variable of the difference of the arguments. Therefore, the only portion of $H_{In}(t - t_1)$ to contribute to the commutator is the part of the integrand having its support in the backward cone spanned by the support of A. That is, the first term of the expansion becomes independent of n, provided that $n > t + a$.

The same argument can be applied to all terms of the perturbative expansion, with the same result. The conclusion is therefore, that at least in perturbation theory, the locality of the interaction ensures the existence of the time evolution of local observables (or fields).

This intuitive argument can be transformed into a mathematically rigourous one, as we shall see, but we first need a result of the theory of semi-group.

III.C.3. Theorem: (Trotter Product Formula) Suppose H_o, H_{In}, $H_o + H_{In} = H_n$ are self-adjoint operators on a Hilbert space $\mathfrak{H}$. Then, in the strong topology of $\mathfrak{H}$, we have that

$$e^{iH_n t} = s - \lim_{r \to \infty} \left(e^{iH_o \frac{t}{r}}\, e^{iH_{In} \frac{t}{r}} \right)^r$$

This theorem is valid under much more general circumstances, but, at least until now, this supplementary generality has not been useful for our type of problems.

III.C.4. Theorem: (Guenin, Segal, Glimm, Jaffe) Suppose H_o, H_{In}, $H_n = H_o + H_{In}$, self-adjoint operators on $\mathfrak{H}$. $\mathfrak{R}(\mathfrak{G})$, a local system of von Neumann algebras, satisfying the properties

1) $\mathfrak{R}(\mathfrak{G}_1) \subset \mathfrak{R}(\mathfrak{G}_2)'$ if $\mathfrak{G}_1 \subset \mathfrak{G}_2{}'$;

2) $e^{iH_o t} \mathfrak{R}(\mathfrak{G})\, e^{-iH_o t} \subset \mathfrak{R}(\mathfrak{G}_t)$

3) $\mathfrak{R}(\bigcap_i \mathfrak{G}_i) = \bigcap_i \mathfrak{R}(\mathfrak{G}_i)$

Suppose further that H_{If} is self-adjoint for a sufficiently large family of elements of $\mathcal{S}(\mathbb{R}^s)$, and that H_{If_1} commutes with H_{If_2},

f_1 and f_2 belonging to $\mathcal{D}(\mathbb{R}^s)$, and, further, that $H_{If_1} \eta \mathcal{R}(\mathcal{O})'$ if the support of f_1 is contained in $\mathcal{O}'$. Then,

$$s - \lim_{n \to \infty} \left(e^{iH_n t} A e^{-iH_n t} \right) = \alpha_t[A]$$

exists, $\forall A \in \mathfrak{A}(\mathcal{O})$, and defines a one parameter group of isomorphisms of $\mathfrak{A}$. The resulting system of local rings is again local.

Proof: Using the Trotter product formula, we immediately get that

$$\alpha_t^n[A]\psi = \lim_{j \to \infty} \left(e^{iH_o \frac{t}{j}} e^{iH_{In} \frac{t}{j}} \right)^j A \left(e^{-iH_{In} \frac{t}{j}} e^{-iH_o \frac{t}{j}} \right)^j \psi, \quad \forall \psi \in \mathcal{D}$$

Let $A \in \mathcal{R}(\mathcal{O})$, where $\mathcal{O}$ is the diamond spanned by a sphere in $\mathbb{R}^s$ of radius a, and $\mathcal{O}_\varepsilon$ be the diamond spanned by a sphere in $\mathbb{R}^s$ of radius $a + \varepsilon$.

Let now $\varepsilon > 0$, split $g(x)$ in such a way that

$$g(x) = g_1(x) + g_2(x)$$

$g_1(x) \in \mathcal{D}$, $g_1(x) \geq 0$, and supp $g_1(\frac{x}{n}) \subset \underline{\mathcal{O}}_\varepsilon$ together with supp $g_2(\frac{x}{n}) \cap \underline{\mathcal{O}}_{\varepsilon/2}$ is empty. This is clearly always possible. We can then write

$$H_{I,n}(g) = H_{I,n}(g_1) + H_{I,n}(g_2) \quad .$$

By hypothesis, $H_{I,n}(g_1)$ and $H_{I,n}(g_2)$ commute, so that

$$\exp\left(iH_{I,n}(g)\frac{t}{j}\right) = \exp\left(iH_{I,n}(g_1)\frac{t}{j}\right) \exp\left(iH_{I,n}(g_2)\frac{t}{j}\right)$$

By hypothesis again, $H_{I,n}(g_2)$ commutes with $\mathfrak{A}(\mathcal{O}_{\varepsilon/2})$, and $\exp(iH_{I,n}(g_1)\frac{t}{j}) \in \mathfrak{A}(\mathcal{O}_\varepsilon)$.

Therefore, as $A \in \mathfrak{A}(\mathcal{O})$,

$$A_1(t) = \exp(iH_o \tfrac{t}{j}) \exp(iH_{I,n}(g)\tfrac{t}{j}) A \exp(-iH_{I,n}(g)\tfrac{t}{j}) \exp(-iH_o \tfrac{t}{j})$$

$$= \exp(iH_o \tfrac{t}{j}) \exp(iH_{I,n}(g_1)\tfrac{t}{j}) A \exp(-iH_{I,n}(g_1)\tfrac{t}{j}) \exp(-iH_o \tfrac{t}{j})$$

Or

$$H_{I,n}(g_1) \eta\, \mathfrak{A}(\mathcal{O}_\varepsilon),\ \exp\left(iH_{I,n}(g_1)\frac{t}{j}\right) A \exp\left(-iH_{I,n}(g_1)\frac{t}{j}\right) \in \mathfrak{A}(\mathcal{O}_\varepsilon)$$

and hence, $A_I(t) \in \mathfrak{A}(\mathcal{O}_{\varepsilon+\frac{t}{j}})$, using the property 2) of the hypothesis.

We can repeat the procedure, step by step, and after j steps, we conclude that

$$A_j(t) = \left[\exp\left(iH_o\frac{t}{j}\right)\exp\left(iH_{I,n}(g)\frac{t}{j}\right)\right]^j A \left[\exp\left(-iH_{I,n}(g)\frac{t}{j}\right)\exp\left(-iH_o\frac{t}{j}\right)\right]^j$$

depends on $g(\frac{x}{n})$ only in the region $\mathcal{O}_{t+j\varepsilon}$, and that $A_j(t) \in \mathfrak{A}(\mathcal{O}_{t+j\varepsilon})$.

We now make the important remark that $A_j(t)$ does not depend upon ε. This can be seen in taking a different ε, and thus different g_1, say g_1'. Now the difference between the mapping $A \to A_j(t)$ defined by g_2 and $A \to A_j'(t)$ vanishes, because the difference between g_1 and g_1' has its support outside $\mathcal{O}$, and thus is going to play exactly the same role as g_2 above. We conclude from that, that

$$A_j(t) \in \mathfrak{A}(\mathcal{O}_{t+j\varepsilon}) \quad , \quad \forall\, \varepsilon$$

and thus, $A_j(t) \in \mathfrak{A}(\mathcal{O}_t)$, using property 3) of the hypothesis. If we now take the strong limit, $s\text{-}\lim_{j\to\infty} A_j(t)$, we get that it will also belong to $\mathfrak{A}(\mathcal{O}_t)$, since the latter is strongly closed, being a von Neumann algebra.

Hence $\alpha_t^n[A] \in \mathfrak{A}(\mathcal{O}_t)$, from which the locality follows trivially.

Furthermore, $\alpha_t^n[A]$ only depends upon $g(\frac{x}{n})$ in the region $\mathcal{O}_t$. As $g(x) = 1$ for $|x| \leq 1$, it is sufficient to take $n > a + t$ in order to get $\alpha_t^n[A]$ independent of g and n. We therefore conclude that the spatial cut-off has been removed. Q. E. D.

III.D. Some Theorems of Friedrichs and the Γ Operation

Before we discuss the hierarchy of interactions, and before we can discuss the general setting of the problems, we need to introduce more of the general properties of Wick polynomials, and to formulate the key theorems of Friedrichs.

III.D.1. Definition: $W_1 \angle : W_2 \ldots W_n :$ is the sum of all those terms of the Wick expansion of $W_1 : W_2 \ldots W_n :$, which correspond to connected graphs.

We shall put, in a formal way:

$$W \; \angle \; :1: = W$$

III.D.2. Lemma (Friedrichs): As a formal power series,

$$V:e^{W}: = :(V \angle :e^{W}:)(:e^{W}:):$$

where

$$:e^{W}: = \sum_{n=0}^{\infty} \frac{1}{n!} \; :W^{n}: \quad .$$

Proof: If we consider a product of the form $V:W^{n}:$, it can be written as a sum of terms in which V is contracted with $:W^{n}:$ in all possible ways, that is contracted m times, $0 \leq m \leq n$ (of course, if there are only r annihilators in V, we have the supplementary relation $m \leq r$). (Note also that we mean contraction of the graphs; they may be much more than m pairs of contractions as far as lines are concerned.) The contribution corresponding to m factors contracted is

$$\binom{n}{m} : (V \angle :W^{n}:) \; W^{n-m}:$$

since there are $\binom{n}{m}$ possibilities of choosing m factors out of n. (Note that $V \angle :W^{m}: = 0$ if $m > n$, so that we do not need to remember this restriction any more.)

We therefore have that

$$V:W^{n}: = \sum_{m=0}^{n} \binom{n}{m} :(V \angle : W^{n}:) \; W^{n-m}:$$

and hence

$$V:e^{W}: = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{1}{n!} \binom{n}{m} : (V \angle :W^{m}:) \; W^{n-m}:$$

$$= \sum_{k,j=0}^{\infty} \frac{1}{k!} \frac{1}{j!} \; :(V \angle :W^{j}:) \; :W^{k}:$$

Q.E.D.

III.D.3. If $H_{o} = \int d\underline{k} \, \mu(\underline{k}) \, a^{*}(\underline{k}) a(\underline{k}) = \int d\underline{k} d\underline{\ell} \, \mu(\underline{k}) \, \delta(\underline{k}-\underline{\ell}) a^{*}(\underline{k}) a(\underline{\ell})$ is the self-adjoint operator representing the free Hamiltonians, we shall define

$$F_\tau = \int d\underline{k}[\mu(\underline{k})]^\tau \, a^*(\underline{k})\, a(\underline{k})$$

from which follows that

$$F_o = N, \qquad F_1 = H_o$$

If we define

$$\begin{aligned} \text{ad } F_\tau(W) &\equiv [F_\tau, W] \\ &= d\underline{k}\, w(\underline{k}) \left(\sum_{j=1}^{i} \mu(\underline{k}_j)^\tau - \sum_{j=i+1}^{m} \mu(\underline{k}_j)^\tau \right) \\ &\quad \cdot a^*(\underline{k}_1) \ldots a^*(\underline{k}_i) a(\underline{k}_{i+1}) \ldots a(\underline{k}_m) \end{aligned}$$

For W which contains only creation operators, one defines a Γ-operation which is the inverse of ad H_o = ad F_1. Friedrichs' definition of the Γ-operation is the same in the general case; it is, however, not the most convenient one. Indeed, with Friedrichs' Γ' operation, we would have that

$$\Gamma'(W) = \int d\underline{k}\, w(\underline{k}) \left[\sum_{j=1}^{i} \mu(\underline{k}_j) - \sum_{j=i+1}^{m} (\mu(\underline{k}_j) \right]^{-1} a^*(\underline{k}_1) \ldots$$

$$\ldots a^*(\underline{k}_i) a(\underline{k}_{i+1}) \ldots a(\underline{k}_n) \quad .$$

The trouble with that definition of Friedrichs is that $\left[\sum_{j=1}^{i} \mu_j - \sum_{j=i+1}^{m} \mu_j \right]$ can vanish, and therefore make the kernel of $\Gamma' W$ singular. We have already plenty enough singularities; what we need is a regularizating operation which would have similar properties. The idea of Glimm was to define

$$\Gamma(W) = \int d\underline{k}\, w(\underline{k}) \left[\sum_{j=1}^{i} \mu(\underline{k}_j) \right]^{-1} a^*(\underline{k}_1) \ldots a^*(\underline{k}_i) a(\underline{k}_{i+1}) \ldots a(\underline{k}_n)$$

Γ is certainly regularizing, since

$$D(\Gamma(W)) \supset D(W)$$

The Γ-operation of Glimm leaves invariant all W which consist only of annihilation operators, and it is identical with the Γ' of Friedrichs for W which consist only of creation operators. Therefore, for the most singular terms, the effect of Γ and Γ' will be the same.

Of course, there is another way of making the kernel not singular. Since $\left[\sum_{j=1}^{i} \mu_j - \sum_{j=i+1}^{m} \mu_j\right]$ is always real, we can add a small imaginary part.

We are thus led to define, following again Friedrichs,

$$\Gamma_{\pm}(W) = \int d\underline{k}\, w(\underline{k}) \left[\sum_{j=1}^{i} \mu_j - \sum_{j=i+1}^{m} \mu_j \pm i\varepsilon\right]^{-1} a^*(\underline{k}_1)\ldots a^*(\underline{k}_i) a(\underline{k}_{i+1})\ldots a(\underline{k}_n)$$

and we have that

$$\left[H_o, \Gamma_{\pm}(W)\right] = W$$

Let us now define, as a formal expression,

$$T_-(W) =: \exp \sum_{n=1}^{\infty} (-1)^n \left(\Gamma_-(W\ldots\Gamma_-(W)\ldots)\ldots\right)_L :$$

and

$$T_+(W) =: \exp \sum_{n=1}^{\infty} \left(\Gamma_+(W\ldots\Gamma_+(W))\ldots\right)_L :$$

where $(\ldots)_L$ means the sum over all connected graphs, with at least one external line.

Then one shows that (still as formal expressions)

$$(H_o + W + E(W))\, T_{\pm}(W) = T_{\pm}(W)\, H_o$$

and

$$T_{\pm}(W)^*\, T_{\pm}(W) = Z(W)^{-1}\, 1 \ ,$$

where $Z(W)$ is a wave function renormalization, and $E(W)$ an energy shift. We see that these operators $T_\pm$ of Friedrichs are the exact analogue of the Møller wave operators. An interesting formula of Friedrichs is the following:

$$T_- = :\exp(-\Gamma_-(Q_-)):$$

where Q_- is the solution to the equation

$$Q_- = W \swarrow :\exp-\Gamma_-(Q_-): - (\varphi_o, W \swarrow \exp-\Gamma_-(Q_-): \varphi_o)$$

III.E. Dressing Transformations and Classification of the Interactions

III.E.1. The formulas of the preceding section have clearly only a formal character. There is, however, a qualitative feature of the greatest importance, which we want to retain. We note that due to the intertwining property of $T_\pm$, $T_\pm$ formally maps elements in the domain of H_o into elements of $\overset{\mp}{H}_o + W$, that is the total Hamiltonian. More generally, we shall call dressing transformation a linear mapping of some dense domain of the Fock space $\mathfrak{F}$ into the domain of the total Hamiltonian, which may lie in another space. What interests us is the fact that a dressing transformation can cancel some of the singularities appearing in the formal expression for the Hamiltonian. It is this property which we want to retain, and to which one can give a rigorous and quantitative meaning. But for that, it is still beyond our technical possibilities to give an exact meaning to the expressions of Friedrichs. It was Glimm who realized that his Γ-operation permitted more accessible formulas, which would still lead to nontrivial dressing transformations. We have the

III.E.2. Lemma (Glimm): As a formal series, in the case $\Gamma(W) \neq W$,

$$H_o\Gamma(W) = W + :\Gamma(W)H_o:$$

$$H_o:e^{\Gamma(W)}: = :W(:e^{\Gamma(W)}:): + :H_o(:e^{\Gamma(W)}:): = :(W+H_o)(:e^{\Gamma(W)}:):$$

Proof: We evidently have, using Wick's theorem, that

$$H_o\Gamma(W) = :H_o\Gamma(W): + \underbrace{:H_o\Gamma(W):}_{1}$$

and

$$H_o : e^{\Gamma(W)}) : = :(H_o \angle :e^{\Gamma(W)}:)(:e^{\Gamma(W)}:):$$

but

$$\underbrace{H_o \quad \Gamma(W)}_{1} = W \qquad \text{and} \qquad H_o \angle :e^{\Gamma(W)}: = H_o + W$$

Q.E.D.

III.E.3. In order to make something rigorous out of all this, one has first to introduce suitable cut-offs. We shall define

$$\phi_\sigma(x,t) = \frac{1}{(2\pi)^{s/2}} \int_{|\underline{k}|<\sigma} \frac{d^s \underline{k}}{(2\mu(\underline{k}))^{1/2}} e^{i\underline{k}\cdot\underline{x}} \left\{a^*(-\underline{k})e^{i\mu(\underline{k})t} + a(\underline{k})\, e^{-i\mu(\underline{k})t}\right\}.$$

A cut-off interaction Hamiltonian $V_\sigma(g)$ will be defined as

$$V_\sigma(g) = \int d^s\underline{x}\, V_\sigma(\underline{x},0)\, g(\underline{x})$$

where $g(\underline{x}) \in \mathcal{S}(\mathbb{R}^s)$ is defined as previously, and $V_\sigma(\underline{x},0)$ is some local Wick polynomial, in which the fields ϕ have been replaced by ϕ_σ.

Clearly, the numerical kernel of $V_\sigma(g)$ will always be $\mathcal{L}^2$, if $\sigma < \infty$, $g \in \mathcal{S}$, so that $V_\sigma(g)$ and $H_o + V_\sigma(g)$ are operators on the Fock space $\mathfrak{F}$.

One defines then $\Gamma(V_\sigma)$, $T_\pm(V_\sigma)$, etc., which are obviously well defined for $\sigma < \infty$, and we are interested in taking the limit $\sigma \to \infty$.

III.E.4. Let $V_\sigma{}^c(g)$ be the pure creation part of $V_\sigma(g)$, and $V_\sigma{}^a(g) = V_\sigma(g) - V_\sigma{}^c(g)$. For the ultraviolet limit $\sigma \to \infty$, the following classification of interactions has been proposed by Glimm.

Type A: $\|V_\infty{}^c(g)\,\varphi_o\| < \infty$

Type B: $\|V_\infty{}^c(g)\,\varphi_o\| = \infty$, $\|\Gamma_\pm(V_\infty{}^c(g))\,\varphi_o\| < \infty$

Type C: $\|\Gamma_\pm(V_\infty{}^c(g))\,\varphi_o\| = \infty$, $\|\Gamma_\pm(V_\infty{}^a(g)\,\varphi\| < \infty$, $\forall \varphi \in \mathcal{S}$,

where $\varphi_o \in \mathfrak{F}$ is the Fock vacuum, $\mathcal{S}$ is the subset of states in $\mathfrak{F}$ with finitely many particles and n-particle wave function in $\mathcal{S}(\mathbb{R}^{ns})$.

In the language of power counting, A-models have no ultraviolet divergencies, B-models only logarithmic divergencies, C-models only logarithmic divergencies in linked graphs of the resolvent series.

If $V_\sigma(g)$ is not of type A, B, C, the corresponding theory can still be superrenormalizable, type D; simply renormalizable, type E; or nonrenormalizable, type F, in the conventional sense, as for instance described by Bogoliubov and Shirkov. This latter classification is due to Hepp.

III.E.5. Examples:

Type A: $:\phi_2{}^n:$ (upper index = Wick power; lower index = s + 1)

Type B: $:(\phi^n\bar{\psi}\psi)_2:$ $:\phi_3{}^2:$, $:\phi_3{}^3:$, $:\phi_4{}^2:$.

Type C: $:\phi_3{}^4:$, $:\phi_4{}^3:$, $:(\bar{\psi}\psi)_{n\geq 3}:$, $:\phi^2_{n\geq 5}:$

Type D: $:(\phi\bar{\psi}\psi)_3:$, $:\phi_5{}^3:$

Type E: $:(\bar{\psi}\psi)_2{}^2:$, $:(\phi\bar{\psi}\psi)_4:$, $:\phi_4{}^4:$

Type F: $(\bar{\psi}\psi)_4{}^2$

III.F. Some Results

III.F.1. Using the formal properties of $T_\pm(V_\sigma(g))$ (cf. preceding section)

$$(H_o + V_\sigma(g) + E(V_\sigma(g)))\ T_\pm(V_\sigma(g)) = T_\pm(V_\sigma(g))\ H_o$$

$$T_\pm(V_\sigma(g))^*\ T_\pm(V_\sigma(g)) = Z(V_\sigma(g))^{-1}\ \mathbb{1}$$

one expects that for theories of the type A, $H_o + V_\infty(g)$ is well defined on $\mathcal{D} \subset \mathcal{H}$.

From standard perturbation theory, we know that for theories of the type B or C, we will have to introduce counterterms. One therefore expects that, for theories of the type B, there exist operators of the form $R_\sigma(g) \subseteq R_\sigma(g)^*$, of order ≥ 2 in g, such that $\forall\, \varphi \in \mathcal{D}$

$$\| T_\pm(V_\sigma(g) + R_\sigma(g))\, \varphi \|^2 , \quad \sigma \leq \infty$$

stays finite. Using then the formal intertwining properties of $T_\pm$, one hopes to approximate $H_\infty(g)$ on $T_\pm(V_\infty(g) + R_\infty(g))\mathcal{D}$ in the strong topology. $(H_\sigma(g) = H_o + V_\sigma(g) + R_\sigma(g))$.

For type C theories, we have again counterterms $R_\sigma(g) \subseteq R_\sigma(g)^*$ of order ≥ 2 in g, and we expect that $\forall\, \varphi \in \mathcal{D}$

$$\|T_{\pm}(V_{\sigma}(g) + R_{\sigma}(g))\varphi\|^2 \cdot Z(V_{\sigma}(g) + R_{\sigma}(g))$$

stays finite for $\sigma \leq \infty$. In this case, we can no longer hope that $H_{\sigma}(g) = H_0 + V_{\sigma}(g) + R_{\sigma}(g)$ will have a strong limit on $T_{\pm}\mathfrak{N}$, but one hopes to approximate $H_{\infty}(g)$ as an operator on a new Hilbert space in the weak topology.

III.F.2. We know that in order to prove the existence of a dynamic, we have to show that H_0, $V_{\infty}(g) + R_{\infty}(g)$ and $H_{\infty}(g)$ are self adjoint operators on some space, and that the local algebras are also defined in that space and enjoy locality properties with respect to the interaction Hamiltonian. Except in the case $:\phi_2{}^4:$, this program has not yet been fulfilled, but we can state the following

Theorem: Let V be a real scalar local Wick polynomial in free massive fields.

A) If V is of type A, then $H_0 + V_{\infty}(g)$ is a real and symmetric operator on the dense domain $D(H_0) \cap D(V_{\infty}(g))$ $(\neq \{0\})$.

B) If V is of type B, then there exists a family of invertible mappings $T_{\rho\sigma}(g)\colon \mathfrak{N} \to \mathfrak{H}$ $(0 \leq \sigma \leq \infty,\ \rho \in \mathbb{Z}_+)$ and $R_{\sigma}(g) \subseteq R_{\sigma}(g)^*$ of order ≥ 2 in g, such that $\forall \varphi \in \mathfrak{N}$,

$$\operatorname*{s-lim}_{\sigma \to \infty} T_{\rho\sigma}(g)\varphi = T_{\rho\infty}(g)\varphi$$

$$\operatorname*{s-lim}_{\sigma \to \infty} (H_0 + V_{\sigma}(g) + R_{\sigma}(g))T_{\rho\sigma}(g)\varphi = H_{\infty}(g)T_{\rho\infty}(g)\varphi$$

$$\operatorname*{s-lim}_{\rho \to \infty} T_{\rho\sigma}(g)\varphi = \varphi$$

$H_{\infty}(g)$ is real and symmetric on the dense domain $\bigcup_{\rho<\infty} T_{\rho\infty}\mathfrak{N} \subset \mathfrak{H}$ which is disjoint from $D(H_0)$.

C) If V is of type C, then there exist invertible mappings

$$T_{\sigma}(g)\colon \mathfrak{N} \to \mathfrak{H},\ 0 \leq \sigma < \infty$$

$$T_{\infty}(g)\colon \mathfrak{N} \to \mathcal{K}(g)$$

where $T_{\infty}(g)\mathfrak{N}$ is dense in a new Hilbert space $\mathcal{K}(g)$. There exist $R_{\sigma}(g) \subseteq R_{\sigma}(g)^*$ of order ≥ 2 in g, and a wave function renormalization $Z_{\sigma}(g)$ such that $\forall \varphi, \psi \in \mathfrak{N}$

$$\lim_{\sigma \to \infty} Z_\sigma(g)(T_\sigma(g)\varphi,\ T_\sigma(g)\psi) = \langle T_\infty(g)\varphi,\ T_\infty(g)\psi\rangle$$

and

$$\lim_{\sigma \to \infty} Z_\sigma(g)(T_\sigma(g)\varphi,\ (H_o + V_\sigma(g) + R_\sigma(g))\ T_\sigma(g)\psi) =$$

$$= \langle T_\infty(g)\varphi,\ H_\infty(g)\ T_\infty(g)\psi\rangle$$

$H_\infty(g)$ is real and symmetric on $T_\infty(g)\mathcal{D}$, and $\langle .,. \rangle$ is the scalar product in $\mathcal{H}(g)$.

III.F.3. In the particular case of a $:\phi_2{}^4:$ interaction, much more powerful results are known (Glimm and Jaffe).

The essential steps are the following: Put

$$H_n = H_o + V_\infty\left(g\left(\frac{x}{n}\right)\right)$$

where

$$V_n = V_\infty\left(g\left(\frac{x}{n}\right)\right) = s - \lim_{\sigma \to \infty} V_\sigma\left(g\left(\frac{x}{n}\right)\right) ,$$

$$V_\sigma\left(g\left(\frac{x}{n}\right)\right) = \lambda \sum_{j=0}^{4} \binom{4}{j} \int a_\sigma^*(x)^j\, a_\sigma(\underline{x})^{4-j} g\left(\frac{x}{n}\right) d\underline{x}$$

the limit existing on the Fock space because this interaction is of type A.

Then

1) As a bilinear form on $\mathcal{D} \times \mathcal{D}$,

$$H_n \geq E_n, \qquad |E_n| < \infty$$

that is, H_n (renormalized by a constant E_n) is positive. E_n is dependent upon n, via $g\left(\frac{x}{n}\right)$, because $V_\infty(1)$ would also be a well defined bilinear form, but unbounded below. One chooses E_n to be equal to the lower bound of H_n.

2) As a bilinear form over $\mathcal{D} \times \mathcal{D}$, $\forall\, a > 1$, $\exists\, b = b(n,a,\sigma)$ such that

$$H_o{}^2 + V_{\sigma,n}{}^2 \leq a(H_{\sigma,n} + b)^2$$

This estimate is valid only for real, non-negative coupling constants λ. It allows us to prove that the resolvent of the self-adjoint operator $H_{0,n}$ converges in norm to the resolvent of a self-adjoint operator H_n. One further shows that H_n is essentially self-adjoint on $\mathcal{D}$. As V_n is also essentially self-adjoint on the same domain, and as the locality conditions are satisfied, one can apply the theorem of Section III.C.4. to prove the existence of local Heisenberg field.

III.F.4. In order to prove the existence of a vacuum state (cf. Borchers' lectures), one has to introduce a periodic box of volume V and the corresponding Fock space $\mathfrak{F}_V$. One then shows the estimate

3) $H_{n,V}\Big|_{\mathfrak{F}_V^\perp} \geq m \geq 0$

4) One then shows, using estimate 2) that $H_{n,V}\Big|_{\mathfrak{F}_V} = \hat{H}_{n,V}$ has a compact resolvent, hence it has a state of lowest energy, a vacuum vector $\varphi_{on,V}$:

$$\hat{H}_{n,V}\varphi_{on,V} = 0, \quad \|\varphi_{on,V}\| = 1$$

5) For some sequence of volumes V_j, $\underset{V_j \to \infty}{s - \lim} \varphi_{on,V_j} = \varphi_{on}$ exists and is a vacuum for H_n (renormalized).

6) φ_{on} is unique.

7) The limit $\omega(A) = \lim_{n \to \infty} (\varphi_{on}, A\,\varphi_{on})$ exists $\forall A \in \mathfrak{A}$ and defines, via the Gelfand construction, a representation with positive energy (cf. Borchers' lectures), where the time evolution is represented by a strongly continuous family of unitary operators.

8) This representation is locally Fock.

III.F.5. The fact that the theory $:\phi_2{}^4:$ is Lorentz invariant has been announced by J. Cannon and A. Jaffe.

There are still a certain number of questions left open:

- uniqueness of ω
- $\exists$ of n-point functions of the field operators
- gap in the spectrum of H
- 1-particle states
- cluster property
- asymptotic nature of the perturbation expansion

That means that we are not close to being out of jobs, especially if one considers that a $:\phi_2{}^4:$ theory is the simplest possible nontrivial model!

III.G. Example of the Quadratic Interaction

It is clear that everybody can solve a quadratic interaction. Nevertheless, this interaction has attracted people again and again, and for many different reasons. We have seen that in a 5-dimensional space-time, a $:\phi_5{}^2:$ interaction does belong to type C. As the big advantage of quadratic interactions is to be exactly soluble, we are going to be able to compare the exact dressing transformations with the approximate ones.

III.G.1. Let us put

$$V_{\sigma,n} = \frac{\delta m^2}{2} \int :\phi^2_{s+1}(\underline{x})\, g(\tfrac{x}{n})\, d^s\underline{x} \ , \qquad H_{\sigma,n} = H_o + V_{\sigma,n} \ .$$

This "interaction" thus corresponds to a mass renormalization. It is not difficult to show that

$$\phi_r(f,t) = \underset{\substack{\sigma\to\infty \\ n\to\infty}}{s\text{-}\lim}\ e^{iH_{\sigma,n}t}\ \phi(f,o)\ e^{-iH_{\sigma,n}t}$$

exists on $\mathcal{S}$, and that one can write explicitly

$$\phi_r(f,t) = \frac{1}{(2\pi)^{s/2}} \int \frac{d^s\underline{k}}{[2\Omega(k)]^{\frac{1}{2}}}\, e^{i\underline{k}\cdot\underline{x}} \left\{ a_r(\underline{k})\, e^{-i\Omega t} + a_r{}^*(-\underline{k}) e^{i\Omega t} \right\} f(\underline{x})\, d^s\underline{x}$$

with $\Omega(\underline{k})^2 = \mu^2 + \delta m^2$ and

$$a_r(\underline{k}) = \tfrac{1}{2}\left(\frac{\Omega}{\mu}\right)^{\frac{1}{2}} \left\{ \frac{\Omega+\mu}{\Omega}\, a(\underline{k}) + \frac{\Omega-\mu}{\Omega}\, a^*(-\underline{k}) \right\} \ .$$

The solution is thus a Bogoliubov-Valatin transformation, which we may also write as

$$\phi_r(f,t) = \frac{1}{(2\pi)^{s/2}} \int \frac{d^s\underline{k}}{[2\Omega]^{\frac{1}{2}}}\, e^{i\underline{k}\cdot\underline{x}} \left\{ a(\underline{k}) \left[\cos\Omega t + i\frac{\mu}{\Omega} \sin\Omega t \right] + a^*(-\underline{k}) \left[\cos\Omega t - i\frac{\mu}{\Omega} \sin\Omega t \right] \right\} f(\underline{x})\, d^s\underline{x}$$

In this form, it is evident that the solution is an entire analytic function in δm^2. This implies, in particular, that the Schwinger-Dyson perturbative expansion for ϕ_r will have an infinite radius of convergence, which can also be verified directly.

III.G.2. On the other hand, $\lim_{\substack{\sigma \to \infty \\ n \to \infty}}$ of $H_{\sigma,n}$ does not exist in Fock space; one has to perform an infinite renormalization, and change the Hilbert space for each δm^2.

The exact dressing transformation is given by

$$T_\sigma = \exp\left[\int_{|\underline{k}|<\sigma} d\underline{k}\ \mathrm{tgh}\left(\tfrac{1}{2}\frac{\Omega-\mu}{\Omega+\mu}\right)\left\{a^*(\underline{k})a^*(-\underline{k}) - a(\underline{k})a(-\underline{k})\right\}\right]$$

This dressing transformation is highly singular in the limit $\sigma \to \infty$, nevertheless, it is possible to show that

$$\lim_{\sigma\to\infty} \frac{(\varphi_1, T_\sigma^*\, A\, T_\sigma \varphi_2)}{(\varphi_a, T_\sigma^*\, T_\sigma\, \varphi_o)} \quad , \quad \forall \varphi_1, \varphi_2 \in \mathfrak{F} \ ,$$

exists $\forall A \in \mathfrak{A}$ and thus defines a new Hilbert space, $\mathcal{H}_{m^2+\delta m^2}$.

Note that $\mathcal{H}_\infty$ is well defined as a Hilbert space, this new representation of the commutation relations is no longer quasiequivalent to a Fock representation however.

Glimm's dressing transformation, in that particular case, would be much less singular, indeed, we would put the system in a box, and thus define

$$V_\sigma = \sum_{\underline{k}\in\Delta(\sigma)} V_{\underline{k}} \ , \quad V_{\underline{k}} = \frac{1}{2\mu}\left\{a^*(\underline{k})a^*(-\underline{k}) + a^*(\underline{k})a(\underline{k}) + a^*(-\underline{k})a(-\underline{k}) + a(\underline{k})a(-\underline{k})\right\}$$

and define

$$T_{\underline{k}\,\delta m^2}\varphi \equiv \exp -\lambda\, \Gamma(V_{\underline{k}})\varphi = \operatorname*{s-lim}_{n\to\infty} \sum_{m=0}^{n} \frac{(-\delta m^2 \Gamma(V_{\underline{k}}))^m}{m!}\ \varphi$$

which clearly exists.

With the same notations,

$$H_{\underline{k},\delta m^2} = \mu\left\{a^*(\underline{k})a(\underline{k}) + a^*(-\underline{k})a(-\underline{k})\right\} + \delta m^2\ V_{\underline{k}}$$

$$Z_{\underline{k},\delta m^2} = ||T_{\underline{k},\delta m^2}\varphi_o||^{-2} = 1 - \frac{\delta m^2}{16\mu^4}$$

$$E_{\underline{k},\delta m^2} = \underbrace{\delta m^2 V_{\underline{k}}}_{2}\ \Gamma(V_{\underline{k}}) = \frac{\delta m^2}{8\mu^3}$$

and one obtains on $\mathfrak{D}$

$$(H_{\underline{k},\delta m^2} + E_{\underline{k},\delta m^2})T_{\underline{k},\delta m^2}\varphi = T_{\underline{k},\delta m^2}\ G_{\underline{k},\delta m^2}\varphi$$

and

$$||\sum_{\underline{k}\in\Delta(\infty)} G_{\underline{k},\delta m^2}\ \varphi||^2 < \infty$$

One then shows that

$$Z_{\sigma,\delta m^2}^{\frac{1}{2}}\ T_{\sigma,\delta m^2} \equiv \prod_{\underline{k}\in\Delta(\sigma)} \mathfrak{Z}_{\underline{k},\delta m^2}^{\frac{1}{2}}\ T_{\underline{k},\delta m^2}$$

provides a good domain for

$$H_{\sigma,\delta m^2} + E_{\sigma,\delta m^2} = \sum_{\underline{k}\in\Delta(\sigma)} \left(H_{\underline{k},\delta m^2} + E_{\underline{k},\delta m^2}\right)$$

and finally, that, $\forall\varphi, \psi \in \mathfrak{D}$

$$\lim_{\sigma\to\infty} Z_{\lambda\sigma}(T_{\sigma,\delta m^2}\varphi,\ T_{\sigma,\delta m^2}\psi) = \langle T_{\infty,\delta m^2}\varphi,\ T_{\infty,\delta m^2}\psi\rangle$$

and defines the same Hilbert space as the exact transformation.

References

Section I:
J. Dixmier, Les algèbres d'opérateurs dans l'espace Hilbertien, 2e edition (Gauthier-Villars, Paris, 1969).
J. Dixmier, Les C*-algèbres (Gauthier-Villars, Paris, 1964).
M. Naimark, Normed Rings (Nordhoff, Groningen, 1964).
C. Rickart, General Theory of Banach Algebras (Van Nostrand, Princeton, 1960).

Section II:
D. Ruelle, Statistical Mechanics (Benjamin, New York, 1969).
J. P. Eckmann and M. Guenin, Mèthodes Algébriques en Mécanique Statistique, Springer Math. tracts No. 81, 1969.

Section III:
R. Jost, editor, Local Quantum Theory, Proceedings of the International School of Physics "Enrico Fermi" in Varenna (Academic Press, 1969). Lectures by Glimm, Jaffe, Araki and others.
K. Hepp, Théorie de la renormalisation, Springer tracts in physics No. 2, 1969.

PHYSICAL OBSERVABLES AND SYMMETRY GROUPS†

H. J. Borchers
Institut für Theoretische Physik
Universität Göttingen
Göttingen, Germany

I. Introduction

I want to give a sequence of lectures on the algebraic approach to quantum field theory. Recently this has become quite a big field, and it would hardly be possible to cover everything which is known about this subject in a course of two semesters. Therefore I have the choice of picking a special subject. This subject will be some aspects in connection with symmetry groups. I am picking these subjects since there has been some progress on these topics in recent years, and some things we will talk about are not published yet.

However, before going into our main subject, let us start with the discussion of some fundamentals; in particular we should answer the question, "Why do we use C*-algebras in physics?"

All physics is based on the fact that the set of objects we are dealing with can be split into two classes.

1) The first class we will call states, denoted by $\mathcal{S}$. These are the objects we want to analyze.

2) The second class is called observables, denoted by $\mathfrak{G}$. These are the devices with which we study the objects of the first class.

The splitting into the two classes is not permanent, but if we consider a certain theory we must have given such a splitting, which is fixed for this particular theory. No one can prevent us from making a theory of certain families of instruments. In such a case this family of observables will be the states which have to be investigated.

The observables we apply to states in order to get a number. This process will be called a measurement.

3) A measurement $\mathfrak{m}$ is a process which assigns to every pair $x \in \mathfrak{G}$ and $\varphi \in \mathcal{S}$ a real number $\langle x, \varphi \rangle \in R$

$$\mathfrak{m}: \mathfrak{G} \times \mathcal{S} \to R$$

---

†Presented at the INSTITUTE FOR THEORETICAL PHYSICS, University of Colorado, Summer 1969.

We have to impose some restrictions on this process of measurement. We would not call two things different if we cannot distinguish them by some measurement. This means
a) $\mathfrak{G}$ separates states, i.e., $\varphi_1, \varphi_2 \in \mathcal{S}$ and $\varphi_1 \neq \varphi_2$ then exists $x \in \mathfrak{G}$ such that

$$\langle x, \varphi_1 \rangle \neq \langle x, \varphi_2 \rangle$$

b) $\mathcal{S}$ separates the observables i.e. x_1, x_2 and $x_1 \neq x_2$ then exists $\varphi \in \mathcal{S}$ such that

$$\langle x_1, \varphi \rangle = \langle x_2, \varphi \rangle .$$

This setup is still very general and it is the aim of theoretical physics to find some additional mathematical structure which puts some order in the set of numbers given by measurements.

This process of finding a theory will in general not be a unique one. This process is governed by fashions and in particular by our knowledge of mathematics.

The modern development started off with the quantum theory of finite many degrees of freedom. This is a well developed and very successful theory. The theories now in the focus of theoretical interest are attempts to construct quantum theories of infinite many degrees of freedom. Since this new theory shall be a generalization of the usual quantum theory, it should cover the special situation which we have in quantum theory.

As everyone knows, there are two different aspects of the usual quantum theory, namely the Schrödinger picture and the Heisenberg picture. In the first one the emphasis is on the states, while in the Heisenberg picture the states are more in the background and the important part is the algebraic structure of the commutation relations. Both pictures contain, of course, the same information. This is even true in the case where you identify the states with the vectors in the representation Hilbert space. That this is so is a consequence of von Neumann's uniqueness theorem, which insures that all representations of the canonical commutation relation for a finite number of degrees of freedom are quasi-equivalent.

If one tries to generalize quantum theory from finitely many to an infinite number of degrees of freedom you have to make your choice, namely what part you should keep and what part you should change. The attitude towards this question has not always been the same. The elder versions of quantum field theory have always put the emphasis on the Hilbert space, which is also true for the so-called axiomatic

field theory, e.g. the L.S.Z. formalism[1] and the Wightman field theory.[2] I. E. Segal[3] was one who realized that one should look at the algebraic structure as the main ingredient of the theory. Every Hilbert space approach of the theory will then come out as a special realization of the algebraic approach.

The main ideas of this algebraic approach are the following: One assumes that one can consider the observables as the self-adjoint elements of a C*-algebra, now denoted by $\mathfrak{A}$. Then the physical states should be identified with positive normalized linear functionals on this C*-algebra $\mathfrak{A}$. These objects are called again states in the mathematical literature. If one wants to deal with a special physical theory one can impose structural conditions on the C*-algebra $\mathfrak{A}$ which restricts the enormous sets of mathematical objects to those which have reasonable physical properties.

As one knows from mathematics, most C*-algebras will have an uncountable number of different representations. At first sight this seems to be in contradiction to the assumption that C*-algebras could be useful for physics. But that these two statements do not exclude each other will be seen in the next section.

II. On Physical Equivalence of Representations

In this section we follow closely a paper by R. Haag and D. Kastler.[4] Let us assume we describe the observables as the self-adjoint elements of a C*-algebra $\mathfrak{A}$. What one usually does in such a case is to look for a representation π of this algebra, i.e. for a realization of the abstract algebra by bounded operators acting on a Hilbert space $\mathcal{H}$. If we do this, the question immediately arises, does every representation describe a different physics? Only if this is not the case can we say that the algebraic aspect can serve as a possible tool for describing physics.

When we have a representation π in a Hilbert space $\mathcal{H}$ we will identify the vectors $\xi \in \mathcal{H}$ as physical states and the expectation values $(\xi, \pi(x)\xi)$, where x is an observable, with the measurement of this state. More generally, if you want to consider a mixture given by a density matrix $R \geq 0$ and $\operatorname{tr} R = 1$ then $\operatorname{tr} R \cdot \pi(x)$ is the expectation value of x in this mixture.

Since, however, experimentalists can make only a finite number of measurements, and these also only with finite accuracy, we give the

II.1. Definition: Two representations π_1 and π_2 on the Hilbert spaces $\mathcal{H}_1$ and $\mathcal{H}_2$ are called physically equivalent, if for any finite number of observables $x_1, x_2, \ldots, x_N$ and density matrices $R_1, \ldots, R_N \in \mathfrak{B}(\mathcal{H}_1)$

and every $\epsilon > 0$, there exists density matrices $R_1', \ldots R_N' \in \mathfrak{B}(\mathcal{H}_2)$ such that

$$|\operatorname{tr} R_i \pi_1(x_i) - \operatorname{tr} R_i' \pi_2(x_i)| < \epsilon, \quad i = 1,2,\ldots,N.$$

Moreover, if $R_i = R_k$ then $R_i' = R_k'$ and vice versa.

The problem of finding different mathematical conditions for physical equivalence has been solved by J. M. G. Fell.[5]

Before stating the result we need some notation.

II.2. Definitions: Let $\mathfrak{A}$ be a C*-algebra. We denote by

1) $\mathfrak{A}^*$ the dual space of $\mathfrak{A}$ and by $\mathfrak{A}^{**}$ the second dual of $\mathfrak{A}$.

2) $E(\mathfrak{A})$ the set of states on $\mathfrak{A}$, i.e. the set of $\omega \in \mathfrak{A}^*$ such that (i) $\omega(x) \geq 0 \ \forall\, x \geq 0, x \in \mathfrak{A}$ and (ii) $||\omega|| = 1$.

3) Let π be a representation of $\mathfrak{A}$. Then we denote by $E(\pi)$ the set of vectorstates associated with π i.e.

$$E(\pi) = \left\{\omega_\xi;\ \xi \in \mathcal{H}_\pi,\ \omega_\xi(x) = (\xi, \pi(x)\xi),\ ||\xi|| = 1\right\}.$$

4) Co $E(\pi)$ the convex set generated by $E(\pi)$.

5) $\overline{\text{Co } E(\pi)}^{n,w}$ the closure of Co $E(\pi)$ in the norm-, resp. W*-topology $\mathfrak{A}^*$. (Remark: The W*-topology is the weak topology induced by the elements $x \in \mathfrak{A}$ on $\mathfrak{A}^*$. It is the same as $\sigma(\mathfrak{A}^*, \mathfrak{A})$ in the terminology of Bourbaki.)

With these notations we get the following result:

II.3. Theorem: Let $\mathfrak{A}$ be a C*-algebra and π_1 and π_2 be two representations of $\mathfrak{A}$. Then the following statements are equivalent.

1) π_1 and π_2 are physically equivalent.

2) $\overline{\text{Co } E(\pi_1)}^{w} = \overline{\text{Co } E(\pi_2)}^{w}$.

3) Ker π_1 = Ker π_2 where Ker π denotes the kernel of the representation π, i.e. the set of all $x \in \mathfrak{A}$ which are represented by the zero operator.

Proof:

$1 \Rightarrow 2$: The definition of physical equivalence just means that every weak neighborhood of a point Co $E(\pi_1)$ contains a point in Co $E(\pi_2)$ hence $\overline{\text{Co } E(\pi_1)}^{w} \supset$ Co $E(\pi_2)$. Since the assumptions are symmetric, 2 follows.

$2 \Rightarrow 3$: Let x be in the kernel of π_1. Then x annihilates every state in $E(\pi_1)$ and hence also in Co $E(\pi_1)$ and by continuity also in

$\overline{\text{Co } E(\pi_1)}^w = \overline{\text{Co } E(\pi_2)}^w$. This implies for any $\xi \in \mathcal{H}_{\pi_2}$, that $(\xi, \pi_2(x)\xi) = 0$, which then implies $(\xi_1, \pi_2(x)\xi_2) = 0 \ \forall \ \xi_1, \xi_2 \in \mathcal{H}_{\pi_2}$, hence $x \in \text{Ker } \pi_2$. Since the statements are symmetric, it follows that $\text{Ker } \pi_1 = \text{Ker } \pi_2$.

$3 \Rightarrow 1$: Since $\text{Ker } \pi_1 = \text{Ker } \pi_2 \Rightarrow \text{Ker } \pi_1^{\perp} = \text{Ker } \pi_2^{\perp}$. We have $(\xi_1, \pi_1(x)\xi_2) = 0$ for all pairs $\xi_1, \xi_2 \in \mathcal{H}_{\pi_1}$ only if $x \in \text{Ker } \pi_1$. This implies that the linear combinations of these matrix elements are weakly dense in $\text{Ker } \pi_1^{\perp}$. Hence every matrix element of π_2 can be approximated by matrix elements of π_1. Since every matrix element is a linear combination of expectation values we have that every expectation value of π_2 can be approximated by expectation values of π_1. Since the assumption was symmetric we get physical equivalence.

This result tells us that the only thing which counts for physics is the kernel of the representation, or to say it in another way, the algebra. However, one should not believe that the existence of different representations with the same kernel is something which is useless for physics. On the contrary, if you investigate a particular situation you should always adapt the representation to the particular situation. For example, if you describe the scattering of two particles, you look for a representation having a two particle incoming state. You could, of course, describe the same situation in a different representation, namely with a state, which describes besides the two incoming particles some other things, which are far enough away, such that it practically does not disturb the two particle scattering.

III. Observables and Symmetry Groups

Every physical problem is connected with the action of a group, sometimes as a symmetry group of the problem, very often as the group of time development, and not seldom as the combination of both.

We now have to translate this concept of symmetry groups into the language of observables and states. Such a symmetry means that we have associated to every element $g \in G$ two mappings, one, denoted by α_g, acting on the observables and another, denoted by α_g' acting on the states, such that for all $x \in \mathfrak{O}$ and all $\varphi \in \mathcal{S}$ we have

$$\langle \alpha_g x, \alpha_g' \varphi \rangle = \langle x, \varphi \rangle \quad .$$

If we want to describe our observables by a C*-algebra $\mathfrak{A}$ then α_g shall map positive operators onto positive operators and furthermore commuting elements onto commuting elements. The easiest way to realize such a mapping is by assuming that α_g is an automorphism of $\mathfrak{A}$.

III.1. Definition: Let $\mathfrak{A}$ be a C*-algebra. A mapping $\alpha: \mathfrak{A} \to \mathfrak{A}$ is called an automorphism if it fulfills the properties:

(i) $\alpha(\lambda x + y) = \lambda\, \alpha(x) + \alpha(y)$; $x, y \in \mathfrak{A}$, $\lambda \in \mathbb{C}$

(ii) $\alpha(xy) = \alpha(x)\, \alpha(y)$

(iii) $\alpha(x^*) = (\alpha(x))^*$
$\alpha(1) = 1$

(iv) α is a bijection with $||\alpha(x)|| = ||x||$.

The group of automorphisms acting on $\mathfrak{A}$ is denoted by Aut $\mathfrak{A}$.

If we now have a group G acting as symmetry group on our physical system, then, according to our preceding discussion, to every $g \in G$ we will have an automorphism $\alpha_g \in$ Aut $\mathfrak{A}$. Although it does not follow from our preceding discussion that the α_g form a representation of the group G we will restrict our discussion to the situation where

$$\alpha: G \to \text{Aut } \mathfrak{A}$$

is a representation of the group G, i.e. $\alpha_{g_1}\, \alpha_{g_2} = \alpha_{g_1 g_2}$ and $\alpha_1 =$ identity automorphism. In the following we will not deal with arbitrary groups but with locally compact groups. This covers all groups which are of physical interest, the Lie groups and the discrete groups.

Having a topological group one can, of course, define continuity properties of the representation α_g of the group.

III.2. Definitions: We say the automorphisms α_g, $g \in G$ act

a) norm continuous if for every $\epsilon > 0$ exists a neighborhood $\mathfrak{U} \subset G$ of the identity such that $||\alpha_g x - x|| \leq \epsilon ||x||$ for $g \in \mathfrak{U}$;

b) strongly continuous if for every $x \in \mathfrak{A}$ and every $\epsilon > 0$ there exists a neighborhood $\mathfrak{U} \subset G$ of the identity such that $\|\alpha_g x = x\| < \epsilon$ for $g \in \mathfrak{U}$

c) π-weakly continuous if π is a representation of $\mathfrak{A}$ and for every pair of vectors $\xi_1, \xi_2 \in \mathfrak{H}_\pi$ and $x \in \mathfrak{A}$ the function $(\xi_1, \pi(\alpha_g x)\xi_2)$ is a continuous function on G.

The case where the group acts norm continuous has been treated by R. Kadison and J. Ringrose[6] using the result of S. Sakai on derivations.[7] However, it is also known that norm continuous groups of automorphisms are not useful in physics because of too much analyticity.[8]

We will deal with the strongly continuous case. It is known from examples in field theory and statistical mechanics that this situation occurs in physics. Since for the application of physics it is much more natural to work with special representations which are adapted to

the special situation rather than with the abstract C*-algebra, our main interest will be to look for representations in which the automorphisms are implemented by unitary operators.

Speaking about representations π of the C*-algebra $\mathfrak{A}$ we will always mean non-degenerated representations. If not stated otherwise G will be a locally compact group and we assume that we have a representation

$$\alpha: G \to \mathrm{Aut}\,\mathfrak{A}$$

of this group.

III.3. Definitions: A representation π of $\mathfrak{A}$ is called

a) covariant if there exists a strongly continuous unitary representation $\rho: G \to \mathfrak{B}(\mathcal{H}_\pi)$ (all bounded operators on $\mathcal{H}_\pi$) such that $\pi(\alpha_g x) = \rho(g)\,\pi(x)\rho(g^{-1})$ for all $x \in \mathfrak{A}$.

b) quasicovariant if π is quasiequivalent to a covariant representation π'.

c) covariant extendible if π is unitary equivalent to a subrepresentation π_1 of a covariant representation π_2.

If one tries to give conditions for the covariance of a representation in the general situation it becomes clear that one has to cope with cohomology and with multiplicity problems. In order that we do not have to deal with these problems we will restrict ourselves to the case of finding conditions under which a representation is covariant extendible respectively under which it is a quasicovariant representation. These problems can be solved and have simple answers.

In the following it is always assumed that G acts as a strongly continuous group of automorphisms.

IV. States Which Are Continuous Under the Action of the Groups

We will denote by $E(\mathfrak{A})$ the set of states on $\mathfrak{A}$ and by $P(\mathfrak{A})$ the set of pure states of $\mathfrak{A}$.

IV.1. Definition: We say α_g' acts strongly continuous on a state $\omega \in E$ if $\|\alpha_g'\omega - \omega\| \to 0$ for $g \to 1$ and denote by E_C the set of states such that α_g' acts strongly continuous on them.

The importance of this set follows from

IV.2. Lemma: Let π be a covariant representation then every vector state of π belongs to E_C.

Proof: Let $\rho: G \to \mathfrak{B}(\mathcal{H}_\pi)$ be the strongly continuous unitary representation of G implementing the automorphisms α_g and $\xi \in \mathcal{H}_\pi$ then

$\alpha_g' \omega_\xi(x) = \omega_\xi(\alpha_g x) = (\rho^{-1}(g)\xi, \pi(x)\rho^{-1}(g)\xi)$. From this follows $|\alpha_g' \omega_\xi(x) - \omega_\xi(x)| = |((\rho^{-1}(g) - 1)\xi, \pi(x)\rho^{-1}(g)\xi) + (\xi, \pi(x)(\rho^{-1}(g)-1)\xi)|$ $\leq 2||x||\,||(\rho^{-1}(g) - 1)\xi|| \cdot ||\xi|| \to 0$ for $g \to 1$ since ρ is a strongly continuous representation. But this implies $||\alpha_g'\omega - \omega|| \to 0$ for $g \to 1$.

This Lemma tells us that states giving rise to covariant extendible representations must lie in E_c. This makes it worthwhile to study E_c more closely. Its properties are collected in

IV.3. Theorem: E_c has the following properties:

1. E_c is convex.
2. E_c is norm closed.
3. E_c is invariant under α_g'.
4. $\omega \in E_c$ and $x \in \mathfrak{A}$ such that $\omega(x^*x) = 1$ then $\omega_x \in E_c$ where $\omega_x(y) = \omega(x^*yx)$.
5. $\omega \in E_c$ and $\omega = \lambda\omega_1 + (1 - \lambda)\omega_2$ with $0 < \lambda < 1$ and $\omega_1, \omega_2 \in E$ then it follows that $\omega_1, \omega_2 \in E_c$.
6. $\omega \in E$, then there exists a unique decomposition $\omega = \lambda\omega_1 + (1 - \lambda)\omega_2$ with $0 \leq \lambda \leq 1$, $\omega_1 \in E_c$ and $\omega_2 \in E$ such that ω_2 does not majorise any state belonging to E_c.
7. There exists a family $E_c^\beta \subset E_c$, indexed by a semiordered set, such that
   - α) $E_c^\beta \subset E_c^\gamma$ for $\beta < \gamma$
   - β) E_c^β is convex and weakly closed
   - γ) $\bigcup_\beta E_c^\beta = E_c$

The rest of this section is devoted to the proof of this theorem.

Proof:

1. Let $\omega_1, \omega_2 \in E_c$ and $0 \leq \lambda \leq 1$ then $||\alpha_g'(\lambda\omega_1 + (1 - \lambda)\omega_2) - (\lambda\omega_1 + (1 - \lambda)\omega_2)|| = ||\lambda(\alpha_g'\omega_1 - \omega_1) + (1 - \lambda)(\alpha_g'\omega_2 - \omega_2)|| \leq \lambda||\alpha_g'\omega_1 - \omega_1|| + (1 - \lambda)||\alpha_g'\omega_2 - \omega_2|| \to 0$ for $g \to 1$.

2. In order to prove the second statement we remark first that the transposed mapping of an automorphism is a norm preserving map of the dual space $\mathfrak{A}^*$ of $\mathfrak{A}$. Let now ω be a limit point in norm of E_c then there exists a sequence $\omega_n \in E_c$ such that $||\omega_n - \omega|| \to 0$ for $n \to \infty$. Hence also $||\alpha_g'\omega_n - \alpha_g'\omega|| \to 0$. Let now $\varepsilon > 0$ then there exists an n such that $||\omega_n - \omega|| \leq \frac{\varepsilon}{3}$ and a $\mathfrak{u}$ such that $||\alpha_g'\omega_n - \omega_n|| \leq \frac{\varepsilon}{3}$ for $g \in \mathfrak{u}$. Hence $||\alpha_g'\omega - \omega|| \leq ||\alpha_g'(\omega - \omega_n)|| + ||\alpha_g'\omega_n - \omega_n|| + ||\omega_n - \omega|| \leq \varepsilon$ for $g \in \mathfrak{u}$. This implies α_g' acts strongly continuous on ω and hence $\omega \in E_c$.

3. This follows immediately from the group property $||\alpha_g'\alpha_h\omega - \alpha_h'\omega|| = ||\alpha_{h^{-1}gh}'\omega - \omega|| \to 0$ for $g \to 0$.

4. This follows from the relation $\alpha_g'\omega_x(y) - \omega_x(y) = \omega(x^*\alpha_g(y)x) - \omega(x^*yx) = (\omega(x^*\alpha_g(y)x) - \omega(\alpha_{g^{-1}}(x^*)y\,\alpha_{g^{-1}}(x))) + (\omega(\alpha_{g^{-1}}(x^*)y\,\alpha_{g^{-1}}(x)) - \omega(\alpha_{g^{-1}}(x^*)yx)) + (\omega(\alpha_g^{-1}(x^*)yx) - \omega(x^*yx))$. This implies $||\alpha_g'\omega_x - \omega_x|| \leq ||\alpha_g'\omega - \omega||\,||x||^2 + 2\,||\alpha_g^{-1}x - x||\cdot||x|| \to 0$ for $g \to 1$.

5. From 4 and 2 follows that with any state $\omega \in E_c$ also every vector state belonging to π_ω is an element of E_c (Ref. 2, 2.4.8). Now every state majorised by ω is a vector state (Ref. 9, 2.5.1). This implies 5.

6. We have seen that with any state $\omega \in E_c$ also every vector state belonging to π_ω is an element of E_c. Since E_c is convex and norm closed we see that all vector states of direct sums of representations, whose vector states are elements of E_c, belong also to E_c. We will collect these results in a proposition. To this end we first give a

IV.4. Definition: We say a representation π is affiliated to E_c if every vector state belonging to π is an element of E_c. We write $\pi\eta E_c$.

IV.5. Proposition:

1. Let $\pi\eta E_c$ and π_1 be a subrepresentation of π then $\pi_1\eta E_c$.
2. Assume $\pi_i\eta E_c$, $i \in I$ then $\sum_{i\in I} \oplus\, \pi_i\eta E_c$
3. Let $\pi\eta E_c$ and π_1 quasiequivalent to π then $\pi_1\eta E_c$.

Proof: The first statement follows immediately from the definition of affiliated representations. The second statement follows from the remarks given just before the definition IV.4. There remains the third statement. Since π and π_1 are quasiequivalent there exists a representation π_2 quasiequivalent to π such that π and π_1 can be identified with subrepresentations of π_2 (Ref. 10, I §4 COROLLAIRE). Taking into account the result of statement 1 we need only to prove that π_2 is affiliated to E_c. Now, according to Ref. 9, 5.3.1, we can choose for π_2 a multiple of π. Hence $\pi_2\eta E_c$ by the second statement.

We now turn back to the proof of the theorem. Let us denote by $\pi = \sum_{\omega\in E_c} \oplus \pi_\omega$ which is by Proposition IV.5 affiliated to E_c. If we

denote by π_u the universal representation then there exists a unique central projection P of the weak closure of π_u such that π is quasi-equivalent to $P\pi_u$ and therefore $P\pi_u\eta E_c$ by Proposition IV.5. On the other hand if ω is a vector state of $(1-P)\pi_u$ then α_g' does not act strongly continuous on ω by construction of P. By definition of the universal representation π_u every state ω is a vector state ω_ξ. Now $\lambda\omega_1 = \omega_{P\xi}$, $(1-\lambda)\omega_2 = \omega_{(1-P)\xi}$ with $\lambda = ||P\xi||$ gives the decomposition of ω_ξ such that $\omega_1 \in E_c$ and ω_2 does not majorise any state belonging to E_c. The last statement follows from the fact that every state majorised by ω_2 is a vector state of $(1-P)\pi_u$.

To prove the last part of Theorem II.3 we remark first that for any $\omega \in E_c$ the expression $||\alpha_g'\omega - \omega||$ defines a continuous non-negative function on G which vanishes at the identity. We call $I_1(G)$ the maximal ideal of bounded continuous functions vanishing at the identity, $I_1^+(G)$ is a semi ordered set.

IV.6. Definition. For every function $\beta \in I_1^+(G)$ we define E_c^β as the set of $\omega \in E_c$ such that $||\alpha_g'\omega - \omega|| \leq \beta(g)$.

It follows from the definition of E_c^β that $E_c^\beta \subset E_c^\gamma$ for $\beta < \gamma$. Furthermore, every ω belongs to all E_c^β with $\beta(g) \geq ||\alpha_g'\omega - \omega||$ hence $\bigcup_\beta E_c^\beta = E_c$. Since the set of functions which are smaller than a given one is a convex set it follows that E_c^β is a convex set. Let now ω_α be a weakly convergent net in E_c^β then we have for every $x \in \mathfrak{A}$ the relation $|\omega_\alpha(\alpha_g x - x)| \leq ||x|| \cdot \beta(g)$ hence this relation holds also for the limit point ω, this means $\omega \in E_c^\beta$.

IV.7. Remarks.

1. We want to emphasize that the extremal points of E_c^β are not pure states in general. This can easily be seen from the example of continuous functions on G vanishing at infinity having the translations as a strongly continuous group of automorphisms.

2. From the fact that α_g acts strongly continuous on $\mathfrak{A}$ it follows that E_c is not empty. For $f \in \mathfrak{L}_1(G)$ with $f \geq 0$ and $\int f(g)\alpha_g = 1$ we define $x \to x(f) = \int f(g)\alpha_g x \alpha_g$, a linear order preserving map from $\mathfrak{A} \to \mathfrak{A}$. This implies for any $\omega \in E$ that we have a state $\omega_f(x) = \omega(x(f))$. Now $|(\alpha_g'\omega_f - \omega)(x)| = |\omega(x(f_g) - x(f))| = |\omega(x(f_g - f))| \leq ||x||\,||f_g - f||_{\mathfrak{L}_1} \to 0$ for $g \to 1$. This implies $\omega_f \in E_c(f_g(h) = f(g^{-1}h))$.

V. Covariance Algebras

This section follows closely the paper of Doplicher, Kastler, and Robinson,[11] and of D. Testard.[14]

Let G be a locally compact group, dg a left invariant Haar measure on G and $\Delta(g)$ the module function.

V.1. Definitions:

1. $\mathfrak{A}_1{}^G$ denotes the set of functions $F: G \to \mathfrak{A}$, i.e. $F(g) \in \mathfrak{A}$, such that $F(g)$ is a measurable function on G. $\mathfrak{A}_1{}^G$ is a Banach space with the norm $||F||_1 = \int ||F(g)|| dg$.

2. On $\mathfrak{A}_1{}^G$ we define an involution by $(F^{\#})(g) := \alpha_g F(g^{-1})^* \Delta(g^{-1})$ The star on the right hand side of the equation denotes the involution in the algebra $\mathfrak{A}$.

3. For every two elements $F, G \in \mathfrak{A}_1{}^G$ we define a product by the equation $(F \bigstar G)(g) = \int dg'\ F(g')\alpha_{g'}\ G(g'^{-1} g)$.

4. We define a homomorphism H from the algebra $\mathfrak{A}$ to the set of bounded linear operators acting on $\mathfrak{A}_1{}^G$ by $(H(x)F)(g) = xF(g)$ $x \in \mathfrak{A}$, $F \in \mathfrak{A}_1{}^G$ and

5. A representation θ of G in the set of bounded linear operators acting on $\mathfrak{A}_1{}^G$ by $(\theta(g')F)(g) = \alpha_{g'}\ F(g'^{-1} g)$.

V.2. Proposition: Under the involution defined in 2 and the product defined in 3 $\mathfrak{A}_1{}^G$ becomes a B*-algebra. The map H(x) has norm $||x||$ and $\theta(g)$ is a representation of G of norm one and acts strongly continuous on $\mathfrak{A}_1{}^G$. $\theta(g)$ is not an automorphism of $\mathfrak{A}_1{}^G$.

Proof: Since α_g is a left invariant measure we get $(F \bigstar (G \bigstar H))(g) = \int dg_1 dg_2\ F(g_1)\ \alpha_{g_1}\{G(g_2)\alpha_{g_2}\ H(g_2^{-1} g_1^{-1}\ g)\}$ and with $g_1 \cdot g_2 = g_2'$ this becomes $= \int dg_1\ dg_2'\ F(g_1)\ \alpha_{g_1}\{G(g_1^{-1} g_2')\}\ \alpha_{g_2'} H(g_2'^{-1} g) = ((F \bigstar G) \bigstar H)(g)$. Next $||F \bigstar G||_1 = \int dg\, dg_1 ||F(g_1)\alpha_{g_1} G(g_1^{-1})|| \leq \int dg\, dg_1 ||F(g_1)||\ ||G(g_1^{-1} g|| > \{\int dg\, ||G(g)||\} \cdot \{\int dg_1\ ||F(g_1)||\} = ||F||_1\ ||G||_1$. This proves that $\mathfrak{A}_1{}^G$ is an algebra. Next $((F^{\#})^{\#})(g) = \alpha_g (F^{\#})(g^{-1})^* \Delta(g^{-1}) = \alpha_g\{\alpha_g{}^{-1} F(g)^* \Delta(g)\}^* \Delta(g^{-1}) = F(g)$; and $(F \bigstar G)^{\#}(g) = \alpha_g (F \bigstar G)(g)^* \Delta(g^{-1}) = \alpha_g\{\int dg_1\ F(g_1)\alpha_{g_1}\ G(g_1^{-1}\ g^{-1})\}^* \Delta(g^{-1}) = \int dg_1\ \{\alpha_{gg_1}\ G(g_1^{-1}\ g^{-1})\}^* \{\alpha_g\ F(g_1)\}^* \Delta(g^{-1})$ with $g \cdot g_1 = g_2$ and $\Delta(g^{-1}) = \Delta(g_2^{-1})\Delta(g_2 g^{-1})$ we get $= \int dg_2 \{\alpha_{g_2} G(g_2^{-1})\ \Delta(g_2^{-1})\}^* \alpha_{g_2}\{\alpha_{g_2{}^{-1} g}\ F(g^{-1} g_2)\Delta(g^{-1} g_2)\}^* = \int dg_2\ G^{\#}(g_2)\ \alpha_{g_2}\ F^{\#}(g_2^{-1} g) = (G^{\#} \bigstar F^{\#})(g)$.

This proves $\mathfrak{A}_1{}^G$ is a B*-algebra.

Now $||H(x)F||_1 = \int dg\ ||xF(g)|| \leq ||x|| \int dg\ ||F(g)|| = ||x|| \cdot ||F||_1$ and $||\theta(g)F||_1 = \int dg_1\ ||\alpha_g\ F(g^{-1} g_1)|| = \int dg_1\ ||F(g_1)|| = ||F||_1$.

It remains to show that $\theta(g)$ acts strongly continuous.

$||\theta(g)F - F||_1 = \int dg_1 ||\alpha_g F(g^{-1}g_1) - F(g)|| \leq \int dg_1 ||\alpha_g F(g^{-1}g_1) - F(g^{-1}g_1)|| + \int ||F(g^{-1}g_1) - F(g_1)|| dg_1 \to 0$ for $g \to 1$. This is first clear for continuous functions $F(g)$ but by continuity it follows from this for arbitrary $F \in \mathfrak{A}_1^G$. This proves the proposition.

For the later applications of the covariance algebras we need some further properties of the operators $H(x)$ and the representation $\theta(g)$ which we will list in the following

V.3. Proposition: For every $x \in \mathfrak{A}$ and every $g \in G$ the operators $H(x)$ and $\theta(g)$ defined by

$$(H(x)F)(g) = x\,F(g)$$
$$(\theta(g')F)(g) = \alpha_{g'}\,F(g'^{-1}g)$$

have the following properties:

a) $\theta(e)F = F$

$\theta(g_1)\theta(g_2)F = \theta(g_1 g_2)F$

b) $||\theta(g)F||_1 = ||F||_1$

$\lim_{g \to g_0} ||\theta(g)F - \theta(g_0)F||_1 = 0$

c) $F^{\#} \bigstar \theta(g)G = (\theta(g^{-1})F)^{\#} \bigstar G$

$(\theta(g)F)^{\#} \bigstar \theta(g)G = F^{\#} \bigstar G$

d) $(\theta(g)F) \bigstar G = \theta(g)(F \bigstar G)$

e) $H(x_1)H(x_2)F = H(x_1 x_2)F$

$H(1)F = F$

f) $||H(x)F||_1 \leq ||x||\,||F||_1$

g) $F^{\#} \bigstar H(x)G = (H(x^*)F)^{\#} \bigstar G$

$(H(x)F)^{\#} \bigstar H(x)G = F^{\#} \bigstar H(x^*x)G$

h) $(H(x)F) \bigstar G = H(x)(F \bigstar G)$

i) Let $\eta \in E(\mathfrak{A}_1^G)$ and $F \in \mathfrak{A}_1^G$ fixed; then $\omega(x) \equiv \eta(F^{\#} \bigstar H(x)F)$ is a positive linear functional on $\mathfrak{A}$ with $||\omega|| = \eta(F^{\#} \bigstar F)$

k) $H(\alpha_g x)F = \theta(g)H(x)\theta(g^{-1})F$

Proof: a) and b) have already been proved in the last proposition. Because of a) the second line of c) is implied by a) and the first line. So we have to prove the first line:

$$(F^{\#} \bigstar \theta(g_1)G)(g) = \int dg_2\, F^{\#}(g_2)\, \alpha_{g_2} \alpha_{g_1}\, G(g_1^{-1} g_2^{-1} g)$$

$$= \int dg_2 \Delta(g_2^{-1}) \{\alpha_{g_2} F(g_2^{-1})^*\} \alpha_{g_2 g_1}\, G(g_1^{-1} g_2^{-1} g)$$

On the other hand

$$((\theta(g_1^{-1})F)^{\#} \bigstar G)(g) = \int dg_2\, (\theta(g_1^{-1})F)^{\#}(g_2)\, \alpha_{g_2}\, G(g_2^{-1} g)$$

$$= \int dg_2\, \alpha_{g_2} \{\theta(g_1^{-1})F\}^* (g_2^{-1}) \Delta(g_2^{-1})\, \alpha_{g_2}\, G(g_2^{-1} g)$$

$$= \int dg_2\, \alpha_{g_2} \{\alpha_{g_1^{-1}}\, F(g_1 g_2^{-1})\}^* \Delta(g_2^{-1}) \alpha_{g_2}\, G(g_2^{-1} g)$$

Since now $\alpha_{g_2} \Delta(g_2^{-1})$ is a right invariant measure we get by setting $g_3 = g_2 g_1^{-1}$

$$= \int dg_3 \Delta(g_3^{-1})\, \alpha_{g_3} \{F(g_3^{-1})\}^* \alpha_{g_3 g_1}\, G(g_1^{-1} g_3^{-1} g)$$

This implies statement c).

Now statement d):

$$((\theta(g_1)F) \bigstar G)(g) = \int dg_2\, (\theta(g_1)F)(g_2) \alpha_{g_2}\, G(g_2^{-1} g)$$

$$= \int dg_2\, \alpha_{g_1} \{F(g_1^{-1} g_2)\}\, \alpha_{g_2}\, G(g_2^{-1} g)$$

$$= \alpha_{g_1} \int dg_2\, F(g_1^{-1} g_2)\, \alpha_{g_1^{-1} g_2}\, G(g_2^{-1} g)$$

and with $g_3 = g_1^{-1} g_2$ this becomes

$$\alpha_{g_1} \int dg_3\, F(g_3)\, \alpha_{g_3}\, G(g_3^{-1} g_1^{-1} g) = (\theta(g_1)(F \bigstar G))(g)\ .$$

Statements e) and f) again have been treated in the previous proposition. Proving now g):

$$(F^{\#} \bigstar H(x)G)(g) = \int dg_1\, F^{\#}(g_1)\, \alpha_{g_1} \{x\, G(g_1^{-1} g)\}$$

$$= \int dg_1\, \alpha_{g_1} \{F(g_1^{-1})\} \Delta(g_1^{-1}) \alpha_{g_1} \{x\} \alpha_{g_1} \{G(g_1^{-1} g)\}$$

$$= \int dg_1\, \alpha_{g_1} \{x^* F(g_1^{-1})\}^* \Delta(g_1^{-1})\, \alpha_{g} \{G(g_1^{-1} g)\}$$

$$= \int dg_1\, (x^* F)^{\#}(g_1)\, \alpha_{g_1}\, G(g_1^{-1} g)$$

$$= ((H(x^*)F)^{\#} \bigstar G)(g)\ .$$

The second line of g) is a simple consequence of the first line and statement e).

h) $((H(x)F) \bigstar G)(g) = \int \alpha_{g_1} \times F(g_1)\, \alpha_{g_1}\, G(g_1^{-1} g)$

$= x \int \alpha_{g_1}\, F(g_1)\, \alpha_{g_1}\, G(g_1^{-1} g) = (H(x)(F \bigstar G))(g)$

i) $\eta \in E(\mathfrak{A}_1^G)$ and $F \in \mathfrak{A}_1^G$ then clearly $\omega(x) = \eta(F^{\#} \bigstar H(x)F)$ defines a linear functional on $\mathfrak{A}$. It has the properties:

$$\omega(x^*x) = \eta(F^{\#} \bigstar H(x^*x)F) = \eta((H(x)F)^{\#} \bigstar H(x)F) \geq 0 .$$

This implies ω is a positive linear functional on $\mathfrak{A}$. Now

$$||\omega|| = \omega(1) = \eta(F^{\#} \bigstar F) \quad .$$

It remains to show the property

k) We have

$$(H(\alpha_{g_1} x)F)(g) = (\alpha_{g_1} x)F(g) = \alpha_{g_1}\{x\alpha_{g_1}^{-1} F(g)\}$$

$$= \theta(g_1)(x\alpha_{g_1}^{-1} F(g_1 g)) = \theta(g_1)H(x)(\alpha_{g_1}^{-1} F(g_1 g))$$

$$= \theta(g_1)H(x)\theta(g_1^{-1})F(g) \quad .$$

This proves the proposition.

We do not want to list all properties of these covariance algebras since we need them only as a tool for our further investigations. The importance of these algebras is given in the following

V.4. Theorem: There is a one to one correspondence between non-degenerate covariant representations $(\pi, \rho(g))$ of the algebra $\mathfrak{A}$ and non-degenerate representations $\hat{\pi}$ of $\mathfrak{A}_1^G$ given by

$$\hat{\pi}(F) = \int \pi(F(g))\ \rho(g)\ dg \qquad F \in \mathfrak{A}_1^G$$

and

$$\pi(x) \cdot \hat{\pi}(F) = \hat{\pi}(H(x)F) \qquad x \in \mathfrak{A}$$

$$\rho(g) \cdot \hat{\pi}(F) = \hat{\pi}(\theta(g)F) \qquad g \in G \quad .$$

If we assume that $1 \in \mathfrak{A}$ then we have also

$$\pi(x) = \text{str.lim}_{fn \to \delta} \hat{\pi}(x \cdot fn), \quad fn \in \mathfrak{L}_1(G)$$

$$\rho(g) = \text{str.lim}_{fn \to \delta} \hat{\pi}(\theta(g) fn)$$

where δ is the Dirac measure of the identity of the group and fn converges to δ in the W*-topology in the dual of the continuous functions on G.

Proof: Assume first we have a covariant representation (π, ρ) then define $\hat{\pi}(F) = \int \pi(F(g)\rho(g)\, dg$ then we get

$$\hat{\pi}(F)\,\hat{\pi}(G) = \int \pi(F(g)\rho(g)\,\pi(G(g_1)\rho(g_1)\, dg\, dg_1$$

$$= \int \pi(F(g)\rho(g)\,\pi(G(g^{-1}g_1))\rho(g^{-1}g_1) dg\, dg_1$$

$$= \int \pi(F(g))\,\pi(\alpha_g\, G(g^{-1}g_1)\,\rho(g_1)\, dg\, dg_1$$

$$= \int \pi\{F(g)\alpha_g\, G(g^{-1}g_1)\}\,\rho(g_1)\, dg\, dg_1$$

$$= \int \pi\{(F \bigstar G)(g_1)\}\,\rho(g_1) dg_1 = \hat{\pi}(F \bigstar G).$$

Next

$$\hat{\pi}(F^{\#}) = \int \pi(F^{\#}(g))\,\rho(g)\, dg$$

$$= \int \pi(\alpha_g F(g^{-1})^*)\,\rho(g)\,\Delta(g^{-1}) dg$$

$$= \int \rho(g)\,\pi(F(g^{-1})^*)\,\Delta(g^{-1})\, dg$$

$$= \{\int \pi(F(g^{-1}))\,\rho(g^{-1})\,\Delta(g^{-1}) dg\}^*$$

$$= \{\int \pi(F(g))\,\rho(g)\, dg\}^* = \hat{\pi}(F)^* \quad .$$

Hence $\hat{\pi}$ defines a representation of $\mathfrak{A}_1^G$. Assume now the representation $\hat{\pi}$ of $\mathfrak{A}_1^G$ is given then define $\pi(x)\,\hat{\pi}(F) = \hat{\pi}(H(x)F)$. We have

$$\pi(x_1)\,\pi(x_2)\,\hat{\pi}(F) = \hat{\pi}(H(x_1)\,H(x_2)F) = \hat{\pi}(H(x_1x_2)F) = \pi(x_1x_2)\,\hat{\pi}(F) \quad .$$

Now

$$\hat{\pi}(F)\,\pi^*(x)\,\hat{\pi}(G) = (\hat{\pi}(G^{\#})\pi(x)\hat{\pi}(F^{\#}))^* = (\hat{\pi}(G^{\#})\hat{\pi}(H(x)F^{\#}))^*$$

$$= \hat{\pi}((G^{\#} \mathbin{\hat{\star}} H(x)F^{\#})^{\#}) = \hat{\pi}(((H(x^*)G)^{\#} \mathbin{\hat{\star}} F^{\#})^{\#})$$

$$= \hat{\pi}(F \mathbin{\hat{\star}} H(x^*)G) = \hat{\pi}(F)\,\hat{\pi}(H(x^*)G) = \hat{\pi}(F)\,\pi(x^*)\,\hat{\pi}(G)\ .$$

In the same manner follows $\rho(g)\,\hat{\pi}(F) = \hat{\pi}(\theta(g)F)$ with the properties $\rho(g_1)\,\rho(g_2) = \rho(g_1 g_2)$, $\rho^*(g) = \rho(g^{-1})$ $\rho(g)\pi(x)\rho(g^{-1}) = \pi(\alpha_g\, x)$. The strong continuity of $\rho(g)$ follows directly from the strong continuity of $\theta(g)$.

Finally

$$\hat{\pi}(f_n x)\cdot\hat{\pi}(F)\xi = \hat{\pi}(x\cdot f_n \mathbin{\hat{\star}} F)\xi = \pi(x)\hat{\pi}(f_n \mathbin{\hat{\star}} F)\xi \to \pi(x)\,\hat{\pi}(F)\xi\ .$$

This means $\hat{\pi}(f_n x) \to \pi(x)$ on a dense set of vectors and since $\hat{\pi}(f_n x)$ is bounded by $||x||\cdot||f||_1$ it converges on every vector. The same argument holds for $\hat{\pi}(\theta(g)f_n)$. This proves the theorem.

VI. Characterization of Vector States of Covariant Representations

In this section we want to prove that every state belonging to E_c is a vector state of a covariant representation. The main result of this section will be

VI.1. Theorem: The following statements are equivalent

1. π is covariant extendible
2. $\pi\ \eta\ E_c$ (see IV.4 for the definition of affiliation)
3. π is the direct sum of cyclic representations such that the states ω_{ξ_i}, $\{\xi_i\}_{i\in I}$ the cyclic vectors, belong to E_c.

Proof of the first part:

1. $\Rightarrow$ 2. follows from Lemma IV.2.
2. $\Rightarrow$ 3. Since every representation is direct sum of cyclic representations (Ref. 9, 2.2.7).
3. $\Rightarrow$ 2. ω_{ξ_i} belongs to E_c. Hence $\pi_{\omega_{\xi_i}}\ \eta\ E_c$ by Theorem IV.3, 4 and 2. Hence

$$\pi = \sum_{i\in I} \oplus \pi_{\omega_{\xi_i}}\ \eta\ E_c$$

by Proposition IV.5.

For proving the implication 3 ⇒ 1 we will use the covariance algebras. By means of this correspondence between representations of $\mathfrak{A}_1^G$ and covariant representations of $\mathfrak{A}$ we have a natural map φ from the states of $\mathfrak{A}_1^G$ into the states of $\mathfrak{A}$. Since for a given covariant representation there might exist several different group representations implementing the automorphisms α_g, the map φ will in general be many to one. But we have

VI.2. Theorem: The image of the mapping φ from $E(\mathfrak{A}_1^G)$ into $E(\mathfrak{A})$ is onto $E_c(\mathfrak{A})$, i.e. $\varphi\, E(\mathfrak{A}_1^G) = E_c(\mathfrak{A})$.

Proof: From Lemma IV.2 follows $\varphi\, E(\mathfrak{A}_1^G) \subset E_c(\mathfrak{A})$. It remains to show the converse inclusion. Let $J \in \mathfrak{A}_1^G$ such that $J(g)$ is a continuous function with compact support. $(J^{\#} \bigstar F \bigstar J)(g)$ is a continuous function in g (Ref. 11, Theorem 2). In particular $F \to (J^{\#} \bigstar F \bigstar J)(1)$ defines a linear positive map Ψ_J from $\mathfrak{A}_1^G$ into $\mathfrak{A}$. Hence the transposed map Ψ_J' sends positive linear forms of $\mathfrak{A}$ into positive linear forms of $\mathfrak{A}_1^G$. This map is given by $(\Psi_J'\, \omega)F = \omega((J^{\#} \bigstar F \bigstar J)(1))$. (Ref. 11, Lemma 5). The rest of the proof of Theorem VI will be given in the following two lemmas.

VI.3. Lemma: Let $\omega \in E_c(\mathfrak{A})$ then there exists a sequence $J_n \in \mathfrak{A}_1^G$ continuous and with compact support such that $\varphi \circ \Psi_{J_n}' \omega$ converge in norm to ω.

Proof: We have

$$(J^{\#} \bigstar F \bigstar J)(1) = \int \alpha_g(J^{\#}(g^{-1}))\alpha_g(F(g^{-1}h))\alpha_h(J(h^{-1}))\Delta(g^{-1})dgdh \ .$$

From this follows

$$\Psi_J'\omega(F) = \int dgdh \ \ \omega(\alpha_g(J^{\#}(g^{-1})\alpha_g(F(g^{-1}h))\alpha_h(J(h^{-1}))) \ \Delta(g^{-1}) \ .$$

Now construct $\varphi \circ \Psi_J'\omega(x)$ by replacing $F(g)$ by $x \cdot \delta(g)$ where $\delta(g)$ denotes the Dirac measure at the identity of G. Hence we get

$$\varphi \circ \Psi_J'\, \omega(x) = \int dg\ \omega(\alpha_g\, J^{\#}(g^{-1})x\, J(g^{-1}))\ \Delta(g^{-1}) \qquad .$$

Since we have assumed that $1 \in \mathfrak{A}$ we can put $J_n(g) = 1\, f(g)$, $f(g)$ continuous function with compact support on G. Thus we get

$$\varphi \circ \Psi_{J_n}'\ \omega(x) = \int dg\ |f(g^{-1})|^2\ \alpha_g'\, \omega(x)\ \Delta(g^{-1}) \ .$$

Since α'_g acts strongly continuous we can choose $f_n(g)$ such that $|f_n(g^{-1})|^2 \Delta(g^{-1})$ converges to $\delta(g)$ in the dual space of the continuous functions on G. Hence $\varphi o \Psi'_{f_n} \omega$ converges in norm to ω.

This Lemma tells us that $\varphi\, E(\mathfrak{A}_1 G)$ is norm dense in $E_c(\mathfrak{A})$. Therefore the proof of Theorem VI.2 is established by the following

VI.4. Lemma: The image of $E(\mathfrak{A}_1 G)$ under the map φ is closed in norm.

Proof: Let ω be a limit point of $\varphi\, E(\mathfrak{A}_1^G)$ then there exists a sequence $\omega_n \in \varphi\, E(\mathfrak{A}_1 G)$ such that $||\omega_n - \omega|| \to 0$.

Let $\hat{\omega}_n \in E(\mathfrak{A}_1^G)$ such that $\varphi\hat{\omega}_n = \omega_n$. Denote by $\hat{\pi} = \sum_n \oplus \hat{\pi}_{\hat{\omega}_n}$ the representation of $\mathfrak{A}_1^G$ induced by $\{\hat{\omega}_n\}$. By the correspondence theorem exists a covariant representation π of $\mathfrak{A}$ such that $\varphi\hat{\omega}_\xi = \omega_\xi$ for every $\xi \in \mathcal{H}_{\hat{\pi}}$. Since the vector states are norm closed (R. V. Kadison, Ref. 12) there exists a vector $\eta \in \mathcal{H}_{\hat{\pi}}$ with $\omega = \omega_\eta = \varphi\,\hat{\omega}_\eta$.

After the proof of Theorem VI.2 we can complete the proof of Theorem VI.1. Let π be a representation of $\mathfrak{A}$ with $\pi_\eta\, E_c$ then it is the direct sum of cyclic representation. $\pi = \sum_{i \in I} \oplus \pi_{\omega_i}$. Since ω_i belongs to E_c there exists $\hat{\omega}_i \in E(\mathfrak{A}_1^G)$ such that $\varphi\hat{\omega}_i = \omega_i$. Now $\sum_{i \in I} \oplus \hat{\pi}_{\hat{\omega}_i}$ is a representation $\hat{\pi}$ of $\mathfrak{A}_1^G$. Let π_1 be corresponding covariant representation of $\mathfrak{A}$ then one checks easily that π is a subrepresentation of π_1. This proves Theorem VI.1.

After having established necessary and sufficient conditions for the existence of covariant extension of a given representation we will ask next for the class of quasicovariant representations.

Before we go into the details of the discussions we have to recall the condition for quasiequivalence in terms of states. Let π_1 and π_2 be two quasiequivalent representations then every normal state on π_1 defines a normal state on π_2 and vice versa. Hence π_1 and π_2 have the same sets of normal states. Now the set of normal states of a representation π is norm closed and coincides with the convex closure of its vector states.

VI.5. Definition: Let π be a representation then we denote

1. $E(\pi)$ the set of all vector states of π
2. $\overline{Co}^n\, E(\pi)$ the norm closed convex hull of $E(\pi)$.

Since for any state $\omega \in \overline{Co}^n E(\pi)$ the representation π_ω is quasiequivalent to a subrepresentation of π, it follows that $\sum_{\omega \in \overline{Co}^n E(\pi)} \oplus \pi_\omega$ is

quasiequivalent to π. But this implies that the two representations π_1 and π_2 are quasiequivalent if $\overline{Co}^n E(\pi_1) = \overline{Co}^n E(\pi_2)$.

After this discussion of quasiequivalent representations we can state the result.

VI.6. Theorem: Let π be a representation of $\mathfrak{A}$ then π is quasicovariant if and only if

α) $\pi \eta E_C$ and

β) $\alpha'_g \overline{Co}^n E(\pi) = \overline{Co}^n E(\pi)$ for all $g \in G$.

Proof: Let π be a quasicovariant representation then there exists a covariant representation π_1 which is quasiequivalent to π. Hence $E(\pi) \subset E_C$ and since E_C is convex and norm closed it follows that $E_C \supset \overline{C_o}^n E(\pi_1) = \overline{C_o}^n E(\pi)$ but this implies that $\pi \eta E_C$. Since π_1 is a covariant representation it follows that $\alpha'_g E(\pi_1) = E(\pi_1)$ and hence also the convex closure of $E(\pi)$ is invariant under the action of α'_g.

Let now π fulfill conditions α) and β) then by Theorem VI.1 there exists a covariant representation π_1 such that π is a subrepresentation of π_1. Let now P be the central carrier of π (in the center of the weak closure of π_1) then $P\pi_1$ is quasiequivalent to π. It remains to show that $P\pi_1$ is a covariant representation. To this end let us denote by $\rho(g)$ the strongly continuous unitary representation of G on $\mathfrak{H}_{\pi_1}$ implementing the automorphisms. Let $\xi \in \mathfrak{H}_{\pi_1}$ such that $P\xi = \xi$; then by condition β $\rho(g)\xi$ defines a normal state of $P\pi_1$ and since the representations $P\pi_1$ and $(1 - P)\pi_1$ are disjoint it follows that $P\rho(g)\xi = \rho(g)\xi$. This implies P commutes with $\rho(g)$ and hence $P\pi_1$ is a covariant representation. This proves the Theorem.

Remarks:

1. We refrain from investing conditions under which a representation is covariant. The necessary condition that $E(\pi)$ is invariant under α'_g is not sufficient for solving the problem as we will see in an example. As far as I have looked into this problem the invariance of $E(\pi)$ will probably be sufficient in the cases where π'' is purely infinite and in the case where π'' is finite only when the coupling is smaller or equal to one, this means if π' is small compared to π''.

Example: Let $\mathfrak{A}$ be the continuous functions on the real line R vanishing at infinity with the translations as automorphism groups. Let π_o be the natural representation in $\mathfrak{L}_2(R)$ and P be a projection in π_o'' such that $P \neq 0$ and $1 - P \neq 0$. Let $\mathfrak{K}$ be any Hilbert space of at least two dimensions. Define $\mathfrak{K} = P\mathfrak{L}_2(R) \otimes \mathfrak{K} \oplus (1 - P)\mathfrak{L}_2(R)$ and $\pi(x) = P\pi_o(x) \otimes 1 \oplus (1 - P)\pi_o(x)$. One sees immediately $E(\pi) = E(\pi_o)$ and

hence $E(\pi)$ is invariant under the action of α'_g. However α_g is not unitary implementable since multiplicity is a unitary invariant.

2. From the way we have constructed states on the covariance algebra $\mathfrak{A}_1{}^G$ in the proof of Theorem VI.2 it follows that there exists a faithful covariant representation of $\mathfrak{A}$ (compare also Ref. 11, Lemma 5). From this we will see that a two-sided ideal in $\mathfrak{A}$ is kernel of a covariant representation if and only if it is invariant.

The content of the Sections IV and VI follows a paper by H. J. Borchers.[13]

VII. On An Algebra Related to the Covariance Algebra

During the next four sections G will always denote an n-parametric abelian group, i.e. the vector group of R^n. We denote by $\hat{G}$ the character group of G which is isomorphic to R^n the dual space of R^n. Therefore, it makes sense to speak about a cone in $\hat{G}$. In the following V^+ will denote a fixed cone in $\hat{G}$ which is closed, convex, has interior points and has its apex at the origin, such that its dual cone has also interior points.

We will say that a strongly continuous unitary representation $\rho: G \to B(\mathcal{H})$ on a Hilbert space $\mathcal{H}$ fulfills the spectrum condition if the spectrum of ρ is contained in V^+.

VII.1. Definition: A representation π of $\mathfrak{A}$ will be called positive, $\pi > 0$, if there exists a strongly continuous unitary representation $\rho: G \to B(\mathcal{H}_\pi)$ such that (i) ρ fulfills the spectrum condition and (ii) ρ implements the automorphisms, i.e. for all $x \in \mathfrak{A}$ and $g \in G$ we have $\rho(g)\,\pi(x)\,\rho(g^{-1}) = \pi(\alpha_g x)$.

Let us denote by $C_0(\hat{G})$ the C*-algebra of continuous functions on $\hat{G}$, tending to zero at infinity. Let now $\mathfrak{A}$ be a concrete C*-algebra acting on a Hilbert space $\mathcal{H}$ and $\rho: G \to B(\mathcal{H})$ a strongly continuous unitary representation of G implementing the automorphisms α_g which by assumptions shall act strongly continuous on $\mathfrak{A}$.

VII.2. Definition: We denote by

(i) $(\mathfrak{A}, G)$ the C*-algebra generated by $\{\mathfrak{A}, \rho(f);\ f \in C_0(\hat{G})\}$

(ii) $I(\mathfrak{A}, G)$ the smallest norm closed two sided ideal containing all elements of the form $\{x\rho(f),\ \rho(f)x;\ x \in \mathfrak{A},\ f \in C_0(\hat{G})\}$

(iii) $(\mathfrak{A}, G)_0$ the smallest sub C*-algebra of $(\mathfrak{A}, G)$ containing all operators of the form $\rho(f_1)x\rho(f_2)$ $x \in \mathfrak{A}$ $f_1, f_2 \in C_0(\hat{G})$.

As a consequence of the continuity property and the spectrum condition we have the important

VII.3. Lemma: The two sets $I(\mathfrak{A}, G)$ and $(\mathfrak{A}, G)_0$ coincide.

Proof: By definition we have $(\mathfrak{A},G)_0 \subset I(\mathfrak{A},G)$. In order to show the converse inclusion it suffices that elements of the form $x\rho(f')$ and $\rho(f)x$ contained in $(\mathfrak{A},G)_0$. And since $(\mathfrak{A},G)_0$ is a C*-algebra it will do to consider only elements of the form $x\rho(f)$.

Now let $f \in C_0(\hat{G})$ have compact support, then $\rho(g)\rho(f) = \rho(e^{ig\hat{g}}f(\hat{g}))$ is an entire analytic function in g. On the other hand $\rho(g)$ has an analytic continuation into the tube $\{g+ig';\ g' \in \hat{V}^+\}$ where $\hat{V}^+$ denotes the dual cone of V^+. Furthermore $\rho(g+ig') \in \rho(C_0(\hat{G}))$ for all $g' \in \hat{V}^+$. Then for any $x \in \mathfrak{A}$, $\rho(g+ig')\, x\rho(-g-ig')\rho(f) \in (\mathfrak{A},G)_0$, $g' \in \hat{V}^+$. Now $\rho(g+ig')\, x\rho(-g-ig')\rho(f)$ is continuous in norm for $g' \in \hat{V}^+$ and also for $g' = 0$. Hence $\rho(g+ig')\, x\rho(-g-ig')\rho(f)$ tends in norm to $\rho(g)\, x\rho(-g)\rho(f)$ for $g' \to 0$. This implies $x\rho(f) \in (\mathfrak{A},G)_0$ for all $f \in C_0(\hat{G})$ with compact support. Since these functions are norm dense in $C_0(\hat{G})$ it follows that $x\rho(f) \in (\mathfrak{A},G)_0$ for all $f \in C_0(\hat{G})$. This proves the lemma.

This last lemma is a preparation for the following statement, which in turn will be the main tool for the first part of the investigation of positive representations.

VII.4. Proposition: Every representation π of $(\mathfrak{A},G)$ whose restriction to $I(\mathfrak{A},G)$ is not degenerated defines a positive representation of $\mathfrak{A}$.

Proof: From the preceding Lemma we have $I(\mathfrak{A},G) = (\mathfrak{A},G)_0$ which implies that $\pi(\rho(C_0(\hat{G}))$ is not degenerated. This in turn implies that we have a strongly continuous unitary representation $\theta(g)$ of G such that for every $f \in C_0(\hat{G})$ the relation $\theta(f) = \pi(\rho(f))$ holds (Ref. 1, §13). Since the spectrum of $\rho(g)$ is contained in V^+, the spectrum of $\theta(g)$ is contained in V^+ too. To finish the proof it remains to show that $\theta(g)$ implements the automorphism α_g. For $x \in \mathfrak{A}$ and $f_1, f_2 \in C_0(\hat{G})$ we have

$$\pi(\rho(f_1))\,\theta(g)\,\pi(x)\,\theta^{-1}(g)\,\pi(\rho(f_2))$$

$$= \pi(\rho(f_1 e^{ig\hat{g}}))\,\pi(x)\,\pi(\rho(e^{-ig\hat{g}}f_2)) = \pi(\rho(f_1 e^{ig\hat{g}})x\rho(e^{-ig\hat{g}}f_2))$$

$$= \pi(\rho(f_1)\alpha_g x\rho(f_2)) = \pi(\rho(f_1))\,\pi(\alpha_g x)\,\pi(\rho(f_2))$$

and since $\pi(\rho(C_0(\hat{G})))$ is not degenerated it follows that

$$\theta(g)\,\pi(x)\,\theta^{-1}(g) = \pi(\alpha_g x)\ .$$

For the use of this proposition we will remind the reader that every representation of a two-sided ideal can be extended to a representation of the whole algebra without enlarging the Hilbert space (Ref. 9, Proposition 2.10.4). This last proposition is not quite what we will use later. Therefore we state

VII.5. Corollary: Let Δ be a non-empty bounded set in the spectrum of $\rho(g)$ and $f \in C_0(\hat{G})$ a real function with the properties $f(\hat{g}) = 1$ for $\hat{g} \in \Delta$ and $f(\hat{g}) < 1$ for $\hat{g} \notin \Delta$. Assume that ω is a state on $(\mathfrak{A},G)$ with the property $\omega(\rho(f)) = 1$. Then π_ω restricted to $\mathfrak{A}$ is a positive representation.

Proof: It follows from the properties of ω that ω restricted to $I(\mathfrak{A},G)$ is not zero. On the other hand since $\rho(f) \leq 1$ and $\omega(\rho(f) = 1$ it follows that the cyclic vector ξ_0 is eigenvector of $\pi(\rho(f))$ with eigenvalue 1. Since the projection P onto the essential subspace of $\pi(I(\mathfrak{A},G))$ commutes with $\pi(\mathfrak{A},G)$ and has the property $P\xi_0 = \xi_0$ it follows that $P = 1$. Thus $\pi(I(\mathfrak{A},G))$ is not degenerated and the result follows from Proposition VII.4.

VIII. States Fulfilling the Spectrum Condition

During this section we will make the simplifying assumption that the algebra $\mathfrak{A}$ contains the identity. All results we will obtain can easily be translated to the case where $\mathfrak{A}$ does not contain an identity. We denote by $E(\mathfrak{A})$ the set of states on $\mathfrak{A}$ and by $P(\mathfrak{A})$ the pure states which are the extremal points of $E(\mathfrak{A})$. Furthermore we will define:

VIII.1. Definition: $E^+(\mathfrak{A})$ denotes the set of states $\omega \in E(\mathfrak{A})$ such that $\pi_\omega > 0$. The main goal of this section will be to prove the following

VIII.2. Theorem: E^+ has the following properties

1. E^+ is convex.
2. $\omega \in E^+$, $x \in \mathfrak{A} \Rightarrow \omega_x \in E^+$ with $\omega_x(y) = \frac{\omega(x^*yx)}{\omega(x^*x)}$ $y \in \mathfrak{A}$.
3. $\omega \in E^+$ and $\omega = \lambda\omega_1 + (1-\lambda)\omega_2$ with $0 < \lambda < 1$ and $\omega_1, \omega_2 \in E \Rightarrow \omega_1, \omega_2 \in E^+$.
4. E^+ is norm closed.
5. Denote by α'_g the transposed automorphism. Then E^+ is invariant under α'_g and α'_g acts strongly continuous on E^+.
6. Let $\omega \in E$ then there exists a unique decomposition $\omega = \lambda\omega_1 + (1-\lambda)\omega_2$ with $0 \leq \lambda \leq 1$, $\omega_1 \in E^+$, $\omega_2 \in E$ and there exists no state $\omega_3 < \omega_2$ with $\omega_3 \in E^+$ (see Ref. 1, 2.5 for the order relation).

7. There exists a family $E_n^+ \subset E^+$ such that
   - α) $E_{n+}^+ \subset E_{n+1}^+$
   - β) E_n^+ is convex and weakly closed and invariant under α'_g
   - γ) $C_o(E_n^+ \cap P)$ is weakly dense in E_n^+
   - δ) $\bigcup_n E_n^+$ is norm dense in E^+.

The rest of this section will be devoted to the proof of this theorem. For proving the first six statements may we remind the reader of a result due to the author.[15]

VIII.3. Theorem: Let π be any positive representation of $\mathfrak{A}$ then there exists a strongly continuous unitary representation $\rho: G \to B(\mathfrak{H}_\pi)$ such that

- α) $\rho(g)$ implements the automorphism α_g
- β) $\rho(g)$ fulfills the spectrum condition
- γ) $\rho(g)$ belongs to the weak closure of $\pi(\mathfrak{A})$.

From this theorem we immediately have the

VIII.4. Corollary:

a) The direct sum of positive representations is a positive representation, i.e. let $\pi_{i \in I} > 0$ then $\pi = \sum_{i \in I} \oplus \pi_i > 0$.

b) Any subrepresentation of a positive representation is a positive representation.

c) Any representation quasiequivalent to a positive representation is a positive representation.

Proof: Theorem VIII.3 permits us to choose the representation $\rho(G)$ in the weak closure of $\pi(\mathfrak{A})$. This implies b) and c). Now the direct sum of continuous representations is again continuous. Hence $\sum_{i \in I} \oplus \rho_i$ implements the automorphism for $\sum_{i \in I} \oplus \pi_i$. Since every ρ_i fulfills the spectrum condition we have for all $f \in C_o(\hat{G})$, vanishing on V^+, $\rho_i(f) = 0$. Hence $\sum_{i \in I} \oplus \rho_i(f) = 0$ for these functions. This implies $\sum_{i \in I} \oplus \rho_i$ fulfills the spectrum condition.

Proof of Theorem VIII.2 first part:

1. Let $\omega_1, \omega_2 \in E^+$ and $0 \leq \lambda \leq 1$ then $\pi_{\lambda\omega_1 + (1-\lambda)\omega_2}$ is unitary equivalent to a subrepresentation of $\pi_{\omega_1} + \pi_{\omega_2}$. Hence $\lambda\omega_1 + (1-\lambda)\omega_2$

$\in E^+$ by corollary VIII.4.

2. Let $\omega \in E^+$, $x \in \mathfrak{A}$ then ω_x gives rise to a subrepresentation of π_ω. Hence $\omega_x \in E^+$.

3. Let $\omega \in E^+$ and $0 < \lambda < 1$ and $\omega = \lambda\omega_1 + (1 - \lambda)\omega_2$. Then $\omega_1 < \omega$, $\omega_2 < \omega$ and ω_1 as well as ω_2 define subrepresentations of π_ω (Ref. 2, 2.5.1). Hence $\omega_1, \omega_2 \in E^+$.

4. Let ω be a limit point in norm of E^+. Then there exists a sequence $\omega_n \in E^+$ with $||\omega_n - \omega|| \xrightarrow[n\to\infty]{} 0$. Let $\pi = \sum_n \oplus \pi_{\omega_n}$ then $\pi > 0$ and $\omega_\eta = \omega_\xi$ with $\xi \in \mathfrak{H}_\pi$. Now the set of vector states of π is norm closed (R. V. Kadison, Ref. 12). Therefore ω is some vector state ω_η, $\eta \in \mathfrak{H}_\pi$, and ω_η defines a subrepresentation of π. Since $\pi > 0$, it follows that $\omega \in E^+$.

5. Let $\omega \in E^+$ and $\rho: G \to B(\mathfrak{H}_\pi)$ a strongly continuous unitary representation implementing α_g. Then $\alpha'_g\omega(x) = \omega(\alpha_g x) = (\xi, \pi(\alpha_g x)\xi) = (\rho^{-1}(g)\xi, \pi(x)\rho^{-1}(g)\xi)$. This means that $\alpha'_g\omega$ is a vector state of π_ω which implies $\alpha'_g\omega \in E^+$.

Now

$$|(\rho^{-1}(g)\xi, \pi(x)\rho^{-1}(g)\xi) - (\xi, \pi(x)\xi)| =$$

$$|((\rho^{-1}(g) - 1)\xi, \pi(x)\rho^{-1}(g)\xi) + (\xi, \pi(x)(\rho^{-1}(g) - 1)\xi)| \leq$$

$$2||x||\,||(\rho^{-1}(g) - 1)\xi||\,||\xi|| \xrightarrow[g\to 0]{} 0$$

since ρ is a strongly continuous representation. Hence $||\alpha'_g\omega - \omega|| \xrightarrow[g\to 0]{} 0$.

6. Denote by π_u the universal representation and by $\pi = \sum_{\omega \in E^+} \oplus \pi_\omega$. π is a positive representation. Now there exists a projection belonging to the center of π_u such that $P\pi_u$ is quasiequivalent to π. Therefore, every state $\omega \in E^+$ is a vector state ω_ξ of π_u with $P_\xi = \xi$. Let now $\omega \in E$, then η such that $\omega = \omega_\eta$. Assume $P^u_\eta \neq 0$ and $(1 - P)\eta \neq 0$ then

$$\omega_1 = \omega_{\frac{P\eta}{||P\eta||}} \in E^+ \quad \text{and} \quad \omega_2 = \omega_{\frac{(1-P)\eta}{||(1-P)\eta||}} \notin E^+$$

and

$$\omega = ||P\eta||\omega_1 + ||(1-P)\eta||\omega_2 \quad .$$

If $\omega_3 < \omega_2$ then ω_3 is a vector state of $(1 - P)\pi_u$ since P belongs to the center of π_u'' and hence $\omega_3 \notin E^+$.

To prove the last statement we will use the results of Section VII. To this end we identify the algebra $\mathfrak{A}$ of Section VII with the representation $\sum_{\omega \in E^+} \oplus \pi_\omega = \pi$ and $\rho(G)$ with the strongly continuous unitary representation of the group G which fulfills the spectrum condition, which exists in $\mathfrak{H}_\pi$ since $\pi > 0$.

Let $\underline{a}$ be an interior point of V^+ and denote by Δ_n the sets $V^+ \cap (n \cdot \underline{a} + V^-)$ in $\hat{G}$. We define now E_n^+ the set of states on $\mathfrak{A}$ which have an extension to $(\mathfrak{A}, G)$ such that $\omega(\rho(f)) = 1 \; \forall \; f \in C_0(\hat{G})$ with $f(\hat{g}) = 1$, $\hat{g} \in \Delta_n$ and $\omega(\rho(f)) = 0 \; \forall \; f \in C_0(\hat{G})$ vanishing in Δ_n. It is clear (VII.4) that the representation π_ω of $(\mathfrak{A}, G)$ defined by $\omega \in E_n^+$ is not degenerated on $I(\mathfrak{A}, G)$ and we have therefore by Proposition VII.3 $E_n^+ \subset E^+$. The convexity of E_n^+ is clear. We now want to show that E_n^+ is weakly closed. To this end let $f \in C_0(\hat{G})$ be a real function $0 \leq f \leq 1$, $f(\hat{g}) = 1$ for $\hat{g} \in \Delta_n$, $f(\hat{g}) < 1$ for $\hat{g} \notin \Delta_n$, and ω a weak limit point of E_n^+. By this we mean that ω is a limit point in the weak topology defined by the elements $x \in \mathfrak{A}$. We have to show that ω has an extension to the algebra $(\mathfrak{A}, G)$ with the desired properties. Consider first the linear set $\pi(\mathfrak{A}) + \lambda\rho(f)$, $\lambda \in C$. On this set we define an extension of ω by the equation $\omega'(\pi(x) + \lambda\rho(f)) = \omega(x) + \lambda$.

We show now that this extension takes non-negative values on non-negative operators. There are two cases, the first one $\rho(f) \in \pi(\mathfrak{A})$. In this case nothing has to be proved, since $\omega(\rho(f)) = 1$ for all $\omega \in E_n^+$ and hence also for the limit points. In the second case $\rho(f)$ does not belong to $\pi(\mathfrak{A})$. Then $\pi(x) + \lambda\rho(f) \geq 0$ implies $\pi(x)^* = \pi(x)$ and $\bar{\lambda} = \lambda$ and ω' is non-negative on such elements if

$$\begin{array}{cc} \sup \omega(x) \leq 1 \leq \inf \omega(x) & . \\ \pi(x) \leq \rho(f) \qquad \pi(x) \geq \rho(f) & \end{array}$$

This, however, is true for every element $\omega \in E_n^+$ by construction of this set, and hence also for the limit point ω. This proves that ω' is a positive linear functional and has therefore by the extension theorem (Ref. 2, 2.10.1) an extension ω'' to a state of $(\mathfrak{A}, G)$ with $\omega''(x) = \omega(x)$ for all $x \in \mathfrak{A}$ and $\omega''(\rho(f)) = 1$. Since $||\rho(f)|| = 1$ and $\xi > 0$ follows that the cyclic vector ξ_0 of the representation $\pi_{\omega''}$ is eigenvector of $\pi_{\omega''}(\rho(f^2))$ with eigenvalue 1. This implies together with the special choice of f that $\omega''(\rho(h)) = 1$ for all h which are one on Δ_n and $\omega''(\rho(h)) = 0$ for all h which vanish on Δ_n. But this implies $\omega \in E_n^+$ and hence E_n^+ is weakly closed.

Next let ω be an extremal point of E_n^+ and assume that there exists a decomposition $\omega = \lambda\omega_1 + (1 - \lambda)\omega_2$, $0 \leq \lambda \leq 1$ then by VIII.2.3

$\omega_1, \omega_2 \in E^+$. Assume $\lambda > 0$, then there exists an operator $T \in \pi_\omega(\mathfrak{A})'$ with $\omega_1 = \omega_{T\xi_o}$ where ξ_o is the cyclic vector for π_ω. Let now P_n be the spectral projection carried by Δ_n and which belongs to $\pi_\omega(\mathfrak{A})''$. Then we have by definition of ω the relation $P_n\xi_o = \xi_o$ and hence $P_nT\xi_o = TP_n\xi_o = T\xi_o$ which means that $\omega_1 \in E_n^+$. Since the same argument holds for ω_2 we have $\lambda = 0$ or 1 which proves that every extremal point of E_n^+ is a pure state. From the fact that the projections P_n tend to one for $n \to \infty$ it follows that every $\omega \in E^+$ can be approximated in norm by elements in E_n^+ because every state is a vector state of π. This completes the proof of the Theorem.

There is one immediate consequence

VIII.5. Corollary: Assume $E^+ \neq \phi$ then E^+ contains at least one invariant state

Proof: Since $E^+ \neq \phi$ there exists a n such that $E_n^+ \neq \phi$. Applying to this set the Markov-Kakutani fix point theorem (Ref. 16, Part I, V.10.6) we get the desired result.

IX. Some Conditions for Positive Representations

In this section we want to present the second tool for handling positive representations. These are generalizations of techniques developed earlier by G. F. Dell'Antonio[17] and myself.[15] Before proving something we will state the result, but we need some notation for this.

Let a be an interior point of V^+ then

IX.1. Definition: $\mathfrak{L}_a$ denotes the smallest left ideal in $\mathfrak{A}$ generated by the element of the form $\int f(g)\alpha_g x\, dg$ with $x \in \mathfrak{A}$, $f \in \mathfrak{L}_1(G)$ and $\tilde{f}(\hat{g})=0$ for $\hat{g} \in -a + V^+$. Here $\tilde{f}$ denotes the Fourier transform of f.

With this definition we can formulate the result as follows

IX.2. Theorem: Let π be a representation of $\mathfrak{A}$ and denote by $\mathfrak{H}_a \subset \mathfrak{H}_\pi$ the subspace of vectors $\xi \in \mathfrak{H}_\pi$ such that $\pi(y)\xi = 0$ for all $y \in \mathfrak{L}_a$. Then the representation π is positive if and only if $\bigcup_{a \in V^+} \mathfrak{H}_a$ is dense in $\mathfrak{H}_\pi$.

One sees immediately that this condition is fulfilled for positive representations. The proof that this condition is also sufficient will be the content of this section.

We start with

IX.3. Proposition: Let π be a representation fulfilling the condition of Theorem IX.2 then Ker π is invariant under the automorphisms α_g.

Proof: Let P_a be the projection onto $\mathcal{K}_a$ then it follows from the construction of P_a that $P_a\pi(\alpha_g x)P_a$ is an entire analytic function in g of some exponential type. One sees this by the following argument. Let $f \in \mathcal{L}_1(G)$ with $\hat{f}(\hat{g}) = 0$ for $\hat{g}$ in $-a+V^+$. Then $\int P_a\pi(\alpha_g x)P_a f(g)dg = 0$ by definition of P_a. Hence the Fourier transform has support in $-a+V^+$. Now let $\hat{f}(\hat{g}) = 0$ for $\hat{g} \in a+V^-(V^- = -V^+)$ then $\int P_a\pi(\alpha_g x)P_a f(g)dg = 0$ since the adjoint $\int P_a\pi(\alpha_g x^*)P_a \bar{f}(g)\,dg = 0$. This implies $P_a\pi(\alpha_g x)P_a$ has support in $\{-a+V^+\} \cap \{a+V^-\}$ which is compact. Let now $x \geq 0$ and $x \in \text{Ker}\,\pi$ then $P_a\pi(\alpha_g x)P_a$ has a zero of second order at $g = 0$. Assume, by induction, that $P_a\pi(\alpha_g x)P_a$ has for any $x \in \text{Ker}\,\pi$ a zero of order w at the origin. We want to show that it has a zero of order 2n. To this end consider the expression

$$P_a\{\pi(\alpha_g x) + \pi(\alpha_{g+h} x) - \pi(\alpha_g|x-\alpha_h x|)\}P_a$$

The element in the bracket is an element of Ker π if $g = 0$ and if $g + h = 0$. Hence we have for sufficiently small g and h

$$||P_a\{\pi(\alpha_g x) + \pi(\alpha_{g+h} x) - \pi(\alpha_g|x-\alpha_h x|)\}P_a|| \leq C||g||^n\,||g+h||^n$$

or

$$||P_a\{\pi(\alpha_g x) + \pi(\alpha_{g+h} x)\}P_a|| \leq C||g||^n\,||g+h||^n + ||x-\alpha_h x||$$

and since $||x-\alpha_h x|| \xrightarrow[h\to 0]{} 0$ we get $2||P_a\pi(\alpha_g x)P_a|| \leq C||g||^{2n}$. This means we have a zero of arbitrary high order and since $P_a\pi(\alpha_g x)P_a$ is analytic in g it follows that $P_a\pi(\alpha_g x)P_a = 0\ \forall\ g$. But since $P_a \to 1$ we have $\pi(\alpha_g x) = 0\ \forall\ g$ which means that Ker π is invariant.

IX.4. In order to prove the implementability of the automorphisms we have to construct some algebras such that α_g acts norm continuous so that we can apply the result of Sakai and Kadison. To do this we will investigate the algebra generated by the projections $\{P_a\}$ and $\pi(\mathfrak{A})$. Since the projections P_a are defined by the invariant left ideals it is natural to extend the automorphisms α_g^π defined by $\alpha_g^\pi\,\pi(x) = \pi(\alpha_g x)$ such that $\alpha_g^\pi P_a = P_a$. Doing this there remains one question, namely whether all relations inside the algebra defined by $\pi(\mathfrak{A})$ and $\{P_a\}$ remain invariant under α_g^π. In order to prove this we have to make a detour such that we reduce the problem to the application of Proposition IX.3.

We will denote by $\mathfrak{F}(\pi(\mathfrak{A}), P_a)$ the free algebra generated by $\pi(\mathfrak{A})$ and the set $\{P_a\}$. On $\mathfrak{F}(\pi, P_a)$ we will define the automorphisms

$\hat{\alpha}_g$ by $\hat{\alpha}_g \pi(x) = \pi(\alpha_g x)$ and $\hat{\alpha}_g P_a = P_a$. Furthermore we will have a family of representations $\theta_g : \mathfrak{F} \to B(\mathcal{H}_\pi)$ defined by $\theta_g \pi(x) = \pi(\alpha_g x) \in B(\mathcal{H}_\pi)$ $\theta_g P_a = P_a \in B(\mathcal{H}_\pi)$. The kernel of the representation θ_g will be just all relations among the operators $\pi(\alpha_g x)$ and P_a when they act on $\mathcal{H}_\pi$. With the help of this family of representations we can define a semi norm on $\mathfrak{F}$ as follows

$$N(y) = \sup_g ||\theta_g(y)|| \qquad y \in \mathfrak{F}$$

IX.5. Lemma: The semi norm N on $\mathfrak{F}$ has the following properties

a) $N(\lambda y) = |\lambda| N(y) \qquad y \in \mathfrak{F}, \ \lambda \in C$
b) $N(y^*) = N(y)$
c) $N(y_1 + y_2) \leq N(y_1) + N(y_2)$
d) $N(y_1 y_2) \leq N(y_1) N(y_2)$
e) $N(y^* y) = N(y)^2$

Proof: a) and b) are an immediate consequence of the definition

c)

$$N(y_1 + y_2) = \sup_g ||\theta_g(y_1) + \theta_g(y_2)|| \leq \sup_g \{||\theta_g(y_1)|| + ||\theta_g(y_2)||\}$$

$$\leq \sup_{g_1} ||\theta_{g_1}(y_1)|| + \sup_{g_2} ||\theta_{g_2}(y_2)|| = N(g_1) + N(g_2)$$

d)

$$N(y_1 y_2) = \sup_g ||\theta_g(y_1 y_2)|| \leq \sup_g ||\theta_g(y_1)|| \, ||\theta_g(y_2)||$$

$$\leq \{\sup_{g_1} ||\theta_{g_1}(y_1)||\}\{\sup_{g_2} ||\theta_{g_2}(y_2)||\} = N(y_1) N(y_2)$$

e) From d) and b) follows

$$N(y^* y) \leq N(y)^2 \ .$$

Let now g_o such that

$$||\theta_{g_o}(y)|| \geq N(y) - \varepsilon$$

then follows

$$N(y)^2 \leq \{ ||\theta_{g_o}(y)|| + \epsilon \}^2 = ||\theta_{g_o}(y^*y)|| + 2\epsilon ||\theta_{g_o}(y)|| + \epsilon^2$$

$$\leq N(y^*y) + 2\epsilon\, N(y) + \epsilon^2 \ .$$

Since ϵ is arbitrary it follows that $N(y)^2 \leq N(y^*y)$. This proves the lemma.

Let now $I = \{ y \in \mathfrak{F};\ N(y^*y) = 0 \}$. These are, of course, all relations among the operators $\pi(x)$ and P_a which are preserved by the mapping α_g^{π}. I is a two-sided ideal in $\mathfrak{F}$ and N defines on $\mathfrak{F}/I$ a norm such that Lemma IX.5 guarantees that the completion of $\mathfrak{F}/I$ is a C*-algebra which we will denote by $\overline{\mathfrak{F}/I}$.

IX.6. Lemma: The mapping $\hat{\alpha}_g$ defines a strongly contiuous group of automorphisms on $\overline{\mathfrak{F}/I}$.

Proof: It is clear that $\hat{\alpha}_g$ defines automorphisms, since by definition of $\mathfrak{F}$, $\hat{\alpha}_g$ was an automorphism of $\mathfrak{F}$ and the two-sided ideal I is by definition invariant under $\hat{\alpha}_g$. Now the continuity property is also trivial

$$N(\hat{\alpha}_g P_a - P_a) = N(P_a - P_a) = 0$$

$$N(\hat{\alpha}_g \pi(x) - \pi(x)) = N(\pi(\alpha_g x - x) = \sup_h ||\pi(\alpha_h(\alpha_g x - x))||$$

$$= ||\pi(\alpha_g x - x)|| \to 0 \qquad \text{for } g \to 0 \ .$$

Since these elements generate the whole algebra $\mathfrak{F}$ we have strong continuity on $\overline{\mathfrak{F}/I}$.

Now we are ready to use Proposition IX.3.

IX.7. Proposition: All relations among the operators $\pi(x)$ and $\{P_a\}$ in $B(\mathfrak{H}_\pi)$ are invariant under the automorphisms α_g^{π}.

Proof: We may consider the C*-algebra generated by the concrete operators $\pi(x)$ and $\{P_a\}$ as the representation θ_o of $\overline{\mathfrak{F}/I}$. If we can show that this representation fulfills the conditions of Theorem IX.2 then it follows from the Proposition IX.3 that the kernel of this representation is invariant, which means that all relations of $\mathfrak{F}$ are invariant under $\hat{\alpha}_g$.

By construction of P_a we have $\pi(\mathfrak{L}_a)P_a = 0$ hence this holds also for the weak closure $\pi(\mathfrak{L}_a)^-$. Thus it remains to show that for every $f \in \mathfrak{L}_1(G)$ with $\tilde{f}(\hat{g}) = 0$ for $\hat{g} \in -a+V^+$ expressions of the form

$$\int f(g)\pi(\alpha_g x_1)\, P_{a_1}\pi(\alpha_g x_2)P_{a_2}\ldots\pi(\alpha_g x_n)dg =$$

$$\int f(g)\theta_o\{\hat{\alpha}_g\pi(x_1)P_{a_1}\pi(x_2)P_{a_2}\ldots\pi(x_n)\}\,dg$$

belong to $\pi(\mathfrak{L}_o)^-$, since by linearity and norm continuity such a relation would stay true for all elements $y \in \mathfrak{F}/I$. Now an element of the form $\pi(x_1)P_{a_1}\pi(x_2)P_{a_2}\ldots\pi(x_n)$ belongs to the weak closure of $\pi(\mathfrak{A})$ and can therefore, according to Kaplansky's density theorem, be approximated by a bounded family of operators $\pi(x_\beta) \in \pi(\mathfrak{A})$. Due to the fact that $(1 - P_a)$ can be approximated by the approximate right identity of $\mathfrak{L}_a$ we can find for every bounded countable set $\{g_\nu\}$ a bounded family of elements $\pi(x_\beta)$ such that $\pi(\alpha_{g_\nu} x_\beta)$ converges to $\pi(\alpha_{g_\nu} x_1)P_{a_1}\ldots \pi(\alpha_{g_\nu} x_n)$.

But for any $x \in \mathfrak{A}$, $P_a\pi(\alpha_g x)P_a$ is an entire function of fixed exponential type. Therefore $P_a\pi(\alpha_g x_\beta)P_a$ is a bounded family of entire functions converging on a properly chosen countable set $\{g_\nu\}$. This implies the convergence as entire functions by Vitalie's theorem.[18] This, however, implies

$$P_a\int f(g)\pi(\alpha_g x_\beta)dg\, P_a \xrightarrow{\text{strongly}} P_a\int f(g)\theta_o\{\hat{\alpha}_g\pi(x_1)P_{a_1}\ldots\pi(x_n)\}\,dg\, P_a$$

and since $P_a \xrightarrow[a\to\infty]{} 1$ we have

$$\int f(g)\pi(\alpha_g x_\beta)dg \xrightarrow{\text{strongly}} \int f(g)\,\theta_o\{\hat{\alpha}_g\,\pi(x_1)P_{a_1}\ldots\pi(x_n)\}\,dg$$

which gives the desired result.

Having established that the representation $\theta_o(\overline{\mathfrak{F}/I})$ fulfills the condition of theorem IX.2 we are able to give the

Proof of Theorem IX.2. Let us denote by $(\pi(\mathfrak{A}), P_a)$ the C*-algebra generated by the operators $\pi(x)$, $x \in \mathfrak{A}$ and $\{P_a\}$. On this algebra we can define a strongly continuous group of automorphisms $\hat{\alpha}_g$: $\hat{\alpha}_g\pi(x) = \pi(\alpha_g x)$, $\hat{\alpha}_g P_a = P_a$ (consequence of Propositions IX.3, IX.6 and Lemma IX.5). Moreover we have that P_a belongs to $\pi(\mathfrak{L}_a)^-$ which implies that for any $y \in (\pi(\mathfrak{A}), P_a)$ and a fixed $\hat{\alpha}_g P_a y P_a$ is an entire function of fixed exponential type. Hence $\hat{\alpha}_g$ acts norm continuous on such elements, moreover such elements form a sub C*-algebra of $(\pi(\mathfrak{A}), P_b)$ and hence $\hat{\alpha}_g$ restricted to this sub algebra is unitary implementable with unitaries belonging to the weak closure of this sub algebra. The spectrum of this representation is bounded and can easily be adjusted to be contained in $V^+ \cap \{a+V^-\}$. Let us call this representation $\rho_a(g)$ and denote by F_a the central carrier of P_a in $\pi(\mathfrak{A})''$.

We define a representation $\hat{\rho}_a(g)$ of G in $F_a\mathcal{H}_\pi$ by the relation

$$\hat{\rho}_a(g)\, y\, P_a\xi = (\hat{\alpha}_g y)\, \rho_a(g)\, P_a\xi; \quad y \in (\pi(\mathfrak{A}), P_b), \ \xi \in \mathcal{H}_\pi$$

It can easily be checked that $\hat{\rho}_a(g)$ is strongly continuous, unitary and implements the automorphism $\hat{\alpha}_g$. It can be checked that the spectrum of $\hat{\rho}_a(g)$ is contained in V^+. To this end we write $\rho_a(g) = \int_{\Delta a} e^{ig\hat{g}} E(d\hat{g})$ with $\Delta a = V^+ \cap \{a + V^-\}$, on the other hand for any $b \in \Delta a$ we have that $\int_{\Delta b} E(d\hat{g})$ is annihilated by $\pi(\mathfrak{L}_b)$. If $f \in \mathfrak{L}_1(G)$ with $\tilde{f}(\hat{g}) = 0$ for $\hat{g} \in V^+$ we have

$$\int f(g)\hat{\rho}_a(g)\, y\, P_a\xi\, dg = \int f(g)(\hat{\alpha}_g y)\, \rho_a(g) P_a\xi\, dg$$

$$= \iint_{\Delta a} f(g) e^{ig\hat{g}} (\hat{\alpha}_g y)\, dg\, E(d\hat{g}) P_a\xi = 0$$

since $\int f(g) e^{ig\hat{g}} \hat{\alpha}_g y\, dg$ belongs to $\pi(\mathfrak{L}_{\hat{g}})$. Since now the l.u.b. of the P_a is 1 also the l.u.b. of the F_a is 1. This allows us to construct a strongly continuous unitary representation $\rho(g) \in \pi(\mathfrak{A})$ which implements the automorphisms α_g. (In the construction of $\rho(g)$ we have omitted several technical details which can be looked up in Refs. 15 and 17.)

X. On Kernels of Positive Representations

We are now prepared for proceeding in the general discussion of positive representations. But we are interested in characterising the two-sided ideals which may appear as kernels of positive representations. We start again with the formulation of the result but we still need a

X.1. Definition: We define I_a as the maximal two-sided ideal contained in the left ideal $\mathfrak{L}_a$ (Definition IX.1).

With this notation we get the following result:

X.2. Theorem: Denote by $I_o = \bigcap_{\omega \in E^+} \operatorname{Ker} \pi_\omega$ then

a) $I_o = \bigcap_a I_a$

b) Let I be a two-sided ideal then there exists a positive representation π with $\operatorname{Ker}\pi \subset I$ if and only if $I_o \subset I$.

As a preparation for the proof of this theorem we show first

X.3. Proposition: Let $\omega \in E$ be a state, assume there exists an $a \in V_o^+$ such that $\omega(\mathfrak{L}_a) = 0$ then ω belongs to E^+.

Proof: Since $\omega(\mathfrak{L}_a) = 0$ it follows that for the cyclic vector $\xi_o \in \mathcal{H}_\omega$ $\pi_\omega(\mathfrak{L}_a)\xi_o = 0$. Hence $P_a \neq 0$. Let now $f \in \mathfrak{L}_1(G)$ and supp $\tilde{f}(\hat{g}) \subset \{-b+V^+\} \cap \{b+V^-\}$; $b \in V_o^+$ then $\pi(\mathfrak{L}_{a+b})\pi(x(f))\xi_o = 0$ hence $P_{a+b} \neq 0$. Let $P = \text{l.u.b.}_{b \in V^+} P_b$. Then for every $f \in \mathfrak{L}_1(G)$ we find $P\pi(x(f))\xi_o = \pi(x(f))\xi_o$ and since α_g acts strongly continuous $P\pi(x)\xi_o = \pi(x)\xi_o \; \forall \, x \in \mathfrak{U}$. But this implies $P = 1$ and the result follows from Theorem IV.2.

Proof of Theorem X.2:

For any state ω annihilating $\mathfrak{L}_a$ we have $\pi_\omega > 0$ by Proposition X.3. But these states characterize $\mathfrak{L}_a$ (Ref. 9, 2.9.5). Hence $I_a = \text{Ker} \sum_{\omega(\mathfrak{L}_a)=0} \oplus \pi_\omega$. From this follows $I_o \subset \cap I_a$. Let now Δ_a denote the set $V^+ \cap \{a+V^-\}$. Then every state $\omega \in E_{\Delta_a}^+$ annihilates $\mathfrak{L}_a$. Hence $I_a \subset \bigcap_{\omega \in E_{\Delta_a}^+} \text{Ker}\, \pi_\omega$. From this follows

$$\bigcap_a I_a \subset \bigcap_a \bigcap_{\omega \in E_{\Delta_a}^+} \text{Ker}\, \pi_\omega = \bigcap_{\omega \in E^+} \text{Ker}\, \pi_\omega = I_o$$

where the last equality follows from Theorem III.2. Hence $I_o = \bigcap_a I_a$ and statement a) is proved.

Let now Λ_a denote the smallest closed left ideal containing I and $\mathfrak{L}_a$ and J_a the maximal two-sided ideal contained in Λ_a. From the proposition X.3 follows the existence of a positive representation π_a with $\text{Ker}\, \pi_a = J_a$. From this follows $\text{Ker} \sum_a \oplus \pi_a = \bigcap_a J_a$. It remains to show that $\bigcap_a J_a = I$. To do so we prove first

X.4. Lemma: $J_a = I + I_a$.

Proof: Since $\mathfrak{L}_a$ is a closed left ideal, it is generated by its positive elements $\mathfrak{L}_a \cap \mathfrak{U}^+$ (Ref. 9, 2.9.3). Denote by C_a the C*-algebra generated by $\mathfrak{L}_a \cap \mathfrak{U}^+$. We have $C_a \subset \mathfrak{L}_a$ since $\mathfrak{L}_a \cap \mathfrak{U}^+$ is invariant under taking the adjoint, i.e. $x_1, \ldots x_n \in \mathfrak{L}_a \cap \mathfrak{U}^+$ implies $x_1 x_2 \ldots x_n \in \mathfrak{L}_a$ and $(x_1 x_2 \ldots x_n)^* = x_n x_{n-1} \ldots x_1 \in \mathfrak{L}_a$. Since $C_a \supset \mathfrak{L}_a \cap \mathfrak{U}^+$ also C_a generates $\mathfrak{L}_a$ as a left ideal. Now $C_a + I$ is norm closed (Ref. 9, 1.8.4) and generates the left ideal $\Lambda_a = \mathfrak{L}_a + I$. Let now $z \in \Lambda_a \cap \mathfrak{U}^+$

then also $z^{\frac{1}{2}} \in \Lambda_a \cap \mathfrak{A}^+$ and it exists a sequence $y_n \in \mathfrak{A}$, $x_n \in C_a + I$ with norm $\lim_{n\to\infty} y_n x_n = z^{\frac{1}{2}}$. Hence norm $\lim_{n\to\infty} x_n^* y_n^* y_n x_n = z$. Since $x_n \in C_a + I$ we can write $x_n = x_n^1 + x_n^2$ with $x_n^1 \in C_a$, $x_n^2 \in I$. Therefore

$$x_n^* y_n^* y_n x_n = x_n^1 * y_n^* y_n x_n^1 + (x_n^2 * y_n^* y_n x_n^1 + x_n^1 * y_n^* y_n x_n^2 + x_n^2 * y_n^* y_n x_n^2) \in C_a + I$$

since

$$x_n^1 * y_n^* y_n x_n^1 \in \mathfrak{L}_a \cap \mathfrak{A}^+ \subset C_a$$

and the rest belongs to I since it is a two-sided ideal. This implies $\Lambda_a \cap \mathfrak{A}^+ \subset I + C_a$ and in particular $J_a \cap \mathfrak{A}^+ \subset I + C_a$ and since $I + C_a$ is a C*-algebra $J_a \subset I + C_a$. Assume $x_1 + x_2 \in C_a + I$ and $y(x_1+x_2)z \in C_a + I$ for all y, $z \in \mathfrak{A}$ then it follows that $yx_1z \in C_a + I$ since $x_3 \in I$ with $x_1 \in C_a$. Putting first $z = 1$ it follows $yx_1 \in \mathfrak{L}_a$, hence $yx_1 \in C_a$. This implies $yx_1 \in \mathfrak{L}_a{}^*$ or $yx_1z \in \mathfrak{L}_a{}^*$ and this means that $yx_1z \in C_a$ for all $y,z \in \mathfrak{A}$. Hence $yx_1z \in I_a$ and $J_a \subset I + I_a$. Since the converse inclusion is trivial we have $J_a = I + I_a$.

From this lemma follows immediately $\bigcap_a J_a = \bigcap_a I + I_a = I$ since $\bigcap_a I_a \subset I$, which proves Theorem X.2.

For the application in physics we are interested in conditions which guarantee the existence of one faithful positive representation. We are now able to formulate such conditions.

X.5. Theorem: The following statements are equivalent:

a) There exists a positive representation π with $\operatorname{Ker}\pi = 0$.

b) For every two-sided ideal I there exists a representation $\pi > 0$ with $\operatorname{Ker}\pi = I$.

c) Denote by $P^+_{rim}\,\mathfrak{A}$ the set of kernels of representations $\pi > 0$ which are also irreducible. For every two-sided ideal $I \subset \mathfrak{A}$ there exists a set $T \subset P^+_{rim}\,\mathfrak{A}$ with $I = \bigcap_{J \in T} J$.

d) E^+ is weakly dense in E.

e) $\bigcap_{a \in V^+} \mathfrak{L}_a = 0$.

Remarks: It is not known whether $P^+_{rim}\,\mathfrak{A}$ coincides with $P_{rim}\,\mathfrak{A}$ or not, i.e. given a primitive ideal I does there exist an irreducible positive representation π with $\operatorname{Ker}\pi = I$?

The statement e) answers the question raised by the paper of S. Doplicher.[17]

Proof: a) and d) are equivalent by the theorem of Fell.[5] a) and b) are equivalent by statement b) of theorem X.2. If c) holds then b) follows trivially. If b) holds then $I = \bigcap_a I_a + I$ and $\overline{\mathfrak{L}_a + I}$ is the intersection of the maximal left ideals which contain $\mathfrak{L}_a + I$. A state annihilating such maximal left ideal gives rise to an irreducible positive representation. Hence $T_a \subset P^+_{rim}\, \mathfrak{A}$ such that $I + I_a = \bigcap_{J \in T_a} J$.

But this implies $I = \bigcap_{J \in \bigcup_a T_a} J$, which proves c).

If now e) holds then $\bigcap_a I_a = 0$, and a) follows by Theorem X.2.

If a) holds then there exists an isometric positive representation. Let now Δ_a be the set $V^+ \cap \{a+V^-\}$ and P_a the spectral projection associated to Δ_a of the representation $\rho(g)$. Then $\mathfrak{L}_a \subset \{x \in \mathfrak{A};\ \pi(x)P_a = 0\} = M_a$. But since $\{P_a\}$ is a resolution of the identity it follows that $\bigcap_a \pi(M_a) = 0$ and since π is faithful $\bigcap_a M_a = 0$; but this implies $\bigcap_a \mathfrak{L}_a = 0$.

References

1. H. Lehmann, K. Symanzik and W. Zimmermann, Nuovo Cimento 1, 205 (1955).
2. A. S. Wightman, Phys. Rev. 101, 860 (1956).
3. I. E. Segal, Ann. of Math. 48, 930 (1947).
4. R. Haag and D. Kastler, J. Math. Phys. 5, 848 (1964).
5. J. M. G. Fell, Trans. Amer. Math. Soc. 94, 365 (1960).
6. R. V. Kadison and J. Ringrose, Comm. Math. Phys. 4, 32 (1967).
7. S. Sakai, Ann. Math. 83, 273 (1966).
8. R. V. Kadison, Comm. Math. Phys. 4, 258 (1967).
9. J. Dixmier, Les C*-algèbres et leurs représentations (Gauthier-Villars, Paris, 1964).
10. J. Dixmier, Les algèbres d'opérateurs dans l'espace Hilbertien (Gauthier-Villars, Paris, 1957).
11. S. Doplicher, D. Kastler, and D. W. Robinson, Comm. Math. Phys. 3, 1 (1966).
12. R. V. Kadison, Trans. Am. Math. Soc. 103, 304 (1962).
13. H. J. Borchers, Comm. Math. Phys. 14, 305 (1969).
14. D. Testard, Ann. Inst. Henri Poincaré, Section A VI, 267 (1967).
15. H. J. Borchers, Comm. Math. Phys. 2, 49 (1966).

16. N. Dunford and J. T. Schwartz, Linear Operators (Interscience, New York, 1958).
17. G. F. Dell'Antonio, Comm. Math. Phys. 2, 384 (1966).
18. e.g. E. C. Titchmarsh, The Theory of Functions (Oxford University Press, 1939).
19. S. Doplicher, Comm. Math. Phys. 1, 1 (1964).

II. Wave Equations

WAVE EQUATIONS FOR PARTICLES OF ARBITRARY SPIN: RELATIVISTIC INVARIANCE, AND QUANTIZATION†

P. M. Mathews
University of Madras
Madras, India

Table of Contents

I. Introduction

The formulation, by Dirac, of the spectacularly successful relativistic wave equation for spin -$\frac{1}{2}$ particles, prompted an intensive search for other wave equations which could (hopefully) serve as a basis for a relativistically invariant description of particles of various spins. The higher-spin equations which resulted from this activity[1] have taken many forms, their bewildering variety being a reflection of the very considerable freedom in the (subjective) choice of reasonable criteria to be met, from among many possibilities which are not all

†Presented at the INSTITUTE FOR THEORETICAL PHYSICS, University of Colorado, Summer 1969.

mutually compatible. Each choice brings its own advantages and disadvantages, and the fact that efforts to find an approach which is satisfactory in all respects have spanned a period of some four decades (and are still continuing) testifies to the inherent difficulty of the task. Nevertheless, work during the last few years has been instrumental in providing deep insights into several aspects of the problem, and it is with some of this work that this series of lectures will be primarily concerned.

We begin with a brief recapitulation of the familiar notions regarding the role of relativity in quantum mechanics, and then proceed to a presentation of the essential details concerning the structure of the group constituted by the coordinate transformations of special relativity (as defined conveniently through the Lie algebra, i.e. the commutation relations of infinitesimal generators of the group) and concerning its irreducible representations. We have to have this information on hand for ready reference, since we will be making extensive use of it in a discussion of the assumptions regarding transformation properties of wave functions, and later, in the derivation of relativistic Schrödinger equations for arbitrary spin. The discrete transformations, space inversion, time reversal and charge conjugation (which do not find a place in the Lie algebra) will have an important role in our discussions, and a careful definition of what we mean by them (independently of any wave equation) is needed and will be given.

The major part of these lectures will be devoted to relativistic wave equations in the Schrödinger form

$$i\frac{\partial \psi(\underset{\sim}{x},t)}{\partial t} = H\,\psi(\underset{\sim}{x},t)\ . \qquad (I.1)$$

The reasons behind the choice of this form will be clear after the discussion of Sec. II. The operator H must be so determined as to ensure invariance of (I.1) under relativistic transformations. It can be shown that in the c-number theory, where ψ is a numerical-valued function, there are two possibilities for H for particles of any given spin a and mass m. If invariance under boosts transverse to the momentum direction (which leads to very complicated equations) is not imposed, the possibilities swell to 4 (infinite) classes characterized by specific types of realizations of the discrete transformations (Sec. IV). But when ψ is promoted to the role of a quantized field operator, and then required to behave in conformity with the principle of microcausality, the situation changes completely. This requirement picks out, for each spin, a particular realization of the discrete operations and determines H completely, and insists that the "statistics" be in conformity with the familiar spin-statistics relation.[4] Though it is

true that all derivations of the spin-statistics theorem have the condition of causality as their basic ingredient, the full extent of the power of this condition has not been manifest in earlier treatments. But what is perhaps the most fascinating aspect of the approach to be presented is the possibility of seeing explicitly and precisely what happens as the invariance requirements are imposed one by one--or dropped one at a time.

The primary aim of these lectures will thus be to try to throw some light on the role of the different constraints which go into the setting up of a relativistic theory of free higher-spin fields. About the difficult problem of interaction of these fields we will have nothing to say, though it is hoped that the insight gained into the structure of higher-spin wave equations would help towards an eventual solution of the interacting case.

II. Relativistic Quantum Mechanics--General Considerations

A. The Transformations of Special Relativity

Observers O and O′ in different inertial frames ascribe to an event A different sets of space-time coordinates

$$x_A = \{x_A{}^{\mu}\} \quad \text{and} \quad x_A{}' = \{x_A'^{\mu}\} \quad , (\mu = 0,1,2,3)$$

such that the space-time separation between two events A and B, measured by

$$(\Delta x)^2 \equiv g_{\mu\nu}(\Delta x)^{\mu}(\Delta x)^{\nu}, \qquad \Delta x \equiv x_A^{\mu} - x_B^{\mu} \quad , \tag{II.1}$$

is independent of the frame of reference. We employ the usual convention of summation over repeated indices. The nonvanishing elements $g_{\mu\nu}$ of the metric tensor are taken to be

$$g_{00} = -g_{11} = -g_{22} = -g_{33} = 1. \tag{II.2}$$

The index $_0$ refers to the time-coordinate.

Linear transformations $x^{\mu} \to x'^{\mu}$ which preserve the form (II.1) comprise the following:

(1) The proper homogeneous Lorentz transformations (hereafter simply called Lorentz transformations):

$$x^{\mu} \to x'^{\mu} = \Lambda^{\mu}{}_{\nu}\, x^{\nu} \tag{II.3}$$

where the $\Lambda^{\mu}{}_{\nu}$ are real and satisfy

$$g_{\mu\nu}\, \Lambda^{\mu}{}_{\rho}\, \Lambda^{\nu}{}_{\sigma} = g_{\rho\sigma} \quad , \tag{II.4a}$$

$$\Lambda^0{}_0 > 0 \text{ and } \det \Lambda = +1 \ . \qquad \text{(II.4b)}$$

These are induced either by transformations of the rotation group (rotations of the space coordinate axes with respect to a fixed origin), in which case

$$\Lambda^0{}_0 = 1 \text{ and } \Lambda^0{}_i = \Lambda^i{}_0 = 0, \ (i = 1,2,3) \qquad \text{(II.5a)}$$

and

$$\tilde{\Lambda}\Lambda = 1 \ ; \qquad \text{(II.5b)}$$

or by going from one to another of two reference frames in uniform relative motion (i.e. by a "boost"), the space axes in the two frames being taken parallel. In this case

$$\Lambda = \tilde{\Lambda} \ . \qquad \text{(II.6)}$$

All the transformations in (II.3) are induced by a sequence of rotations and boosts, and form the proper homogeneous Lorentz group (or simply, the Lorentz group), of which the rotation group is a subgroup. We will denote the Lorentz group by $\mathfrak{L}$.

(ii) The inhomgeneous Lorentz transformations

$$x^\mu \to x'^\mu = x^\mu - a^\mu \qquad \text{(II.7)}$$

representing a translation of the origin of space-time coordinates by a four-vector a^μ. The transformations (II.3) and (II.7) together form the Poincarè group or inhomogeneous Lorentz group. We will denote this group by $\mathcal{P}$, and any generic element by L. All transformations L of $\mathcal{P}$ are continuous with the identity, i.e. there exists a continuous sequence of transformations of $\mathcal{P}$ which links any element L to the identity element. (The latter relates two reference frames which coincide.)

(iii) The improper transformations: space inversion

$$\Lambda^i{}_j = -\delta^i{}_j, \quad \Lambda^0{}_0 = 1, \quad \Lambda^0{}_j = \Lambda^j{}_0 = 0 \qquad \text{(II.8)}$$

and time reversal

$$\Lambda^i{}_j = \delta^i{}_j, \quad \Lambda^0{}_0 = -1, \quad \Lambda^0{}_j = \Lambda^j{}_0 = 0 \ . \qquad \text{(II.9)}$$

These are not continuous with the identity and are therefore referred to as discrete transformations. The group obtained by adjoining these to the Poincarè group will be referred to in the sequel as the extended

Poincarè group. The term relativistic transformation will be used to denote generally any transformation belonging to the extended Poincarè group.

Besides these, we will have to deal with another discrete transformation, called charge conjugation, which can be defined only in relation to representations of the group considered above, and will be formally introduced later.

B. Relativistically Invariant Description of Quantum Mechanical States

The meaning of the requirement of relativistic invariance on a quantum mechanical system can be stated as follows:

(i) All possible states of a quantum system, as described by an observer in some inertial frame of reference, at some fixed but arbitrary time in his own frame, constitute a linear vector space. This vector space should remain unchanged with time of observation in a given reference frame, and should be common to all observers in the same or different inertial reference frames.

(ii) An observer in a given frame describes the time evolution of the system by a trajectory running through the tips of the state vectors into which a given state transforms itself with the passage of time. The specific forms of the possible trajectories are determined by the dynamics of the system. For relativistic invariance, the set of all possible trajectories should be independent of the observer. Two different observers watching the evolution of a given system would describe it by two different trajectories, both of which, however, belong to this common set. Thus one trajectory gets mapped on to another under a relativistic transformation; but it must be kept in mind that one point on a trajectory (the state at one instant of time) does not get mapped into one point on the other, since a fixed time in one reference frame does not correspond to a single instant of time for all observers in a relatively moving frame.

(iii) With every pair of states of a quantum system at a fixed time in any reference frame is associated a unitary inner product. To ensure that this inner product has a relativistically invariant meaning, it is necessary to demand that if two bundles of trajectories (representing the evolution of various states of a system as seen from two reference frames) are "cut," each at a fixed time in its own frame, then the two states obtained from the corresponding members of the two bundles should be related by a linear transformation which is unitary with respect to the inner product and depends only on the relation between the two frames, and not on which particular trajectory is considered. It must be emphasized that this requirement goes beyond mere identity of the set of all states as seen from different reference

frames as demanded in (i), or correspondence between whole trajectories as seen from different frames, required in (ii).

To see in more concrete terms what these statements mean, consider a single particle described by a coordinate space wave function $\psi(\underset{\sim}{x},t)$ whose time dependence gives a "trajectory" in the space of wave functions. The above requirements are then (i) that the set of all possible wave functions $\psi(\underset{\sim}{x})$ at fixed time be independent of the reference frame used, and (ii) that if a particular form of the function $\psi(\underset{\sim}{x},t)$ describes a possible time evolution (trajectory) in one reference frame, it gives also a possible time dependence in any other frame ($\underset{\sim}{x}$,t) being the space and time coordinates in the frame concerned). This is usually ensured by defining the ψ's as solutions of a wave equation which has the same form in all frames, i.e., a relativistically invariant wave equation. Of course, different observers O and O′, watching the state of a particle evolve, would describe it by different functions $\psi(\underset{\sim}{x},t)$ and $\psi'(\underset{\sim}{x},t)$, but in view of the mixing of space and time coordinates in boost transformations, ψ at a fixed time does not go into ψ' at a fixed time; rather, the whole trajectory (the wave function at all times) in any frame goes into the determination of each point on the trajectory (fixed time wave function) as seen from another moving reference frame. The assertion (iii) is now that in spite of the above, a unitary transformation U(L) depending only on the relativistic transformation L relating the reference frames exists such that

$$\psi'_{t'}(\underset{\sim}{x}) = U(L)\ \psi_t(\underset{\sim}{x}) \qquad \text{(II.10)}$$

where t, t′ are written as subscripts to emphasize that ψ, ψ' are taken at (arbitrary) fixed times in their respective frames. The necessity for the existence of a Hermitian operator H which acting on fixed-time wave functions $\psi(\underset{\sim}{x})$ gives $i\,\partial\psi(\underset{\sim}{x},t)/\partial t$ is a consequence of the above requirement (in the special case of time translations).

III. The Realization of Relativistic Transformations On Wave Functions

The operators U(L) relating fixed time wave functions in two reference frames as in (II.10) must form a unitary ray representation

$$U(L_2)\ U(L_1) = \omega(L_2L_1)\ U(L_2L_1)$$

of the transformation group, in order that the effect of a succession of relativistic transformations on the wave function be consistent with the group property of the transformations themselves. (As is well known, the possibility of having the unimodular factor ω in (III.1) arises from the arbitrary phase that can be associated with any given

state, and as shown by Wigner[5] many years ago, is reducible to ±1 in any unitary irreducible representation of the Poincarè group.)

We now review briefly the group structure of the relativistic transformations, and then present basic information regarding the irreducible representations, which will be used in the discussion of (II.10) in the sequel.

A. Structure of the Poincarè Group $\mathcal{P}$

If we consider an infinitesimal relativistic transformation characterised by an infinitesimal real parameter ϵ, the corresponding element $\hat{L}$ of the abstract† Poincarè group differs infinitesimally from the identity (denoted by $\hat{1}$):

$$\hat{L} = \hat{1} + i\,\epsilon\,\hat{G} \quad . \tag{III.1}$$

$\hat{G}$ is called the generator of the transformation, and generates the whole family of elements $\exp(i\,\epsilon\,\hat{G})$ of $\mathcal{P}$ when ϵ is given arbitrary real values. The Poincarè group $\mathcal{P}$ has 10 independent generators:

(i) $\hat{J} = (\hat{J}_1, \hat{J}_2, \hat{J}_3)$ which generate rotations about the three space axes, and are characterised by the commutation relations

$$[\hat{J}_i, \hat{J}_j] = i\,\epsilon_{ijk}\,\hat{J}_k \quad . \tag{III.2}$$

These generate the rotation subgroup of $\mathcal{P}$, with elements $\exp(i\theta\underline{n}\cdot\underline{J})$ corresponding to rotations through angle θ in the positive sense about the unit vector $\underline{n}$.

(ii) $\underline{\hat{K}} = (\hat{K}_1, \hat{K}_2, \hat{K}_3)$, which generate boosts in the directions of the space axes, and satisfy

$$[\hat{J}_i, \hat{K}_j] = i\,\epsilon_{ijk}\,\hat{K}_k \tag{III.3}$$

and

$$[\hat{K}_i, \hat{K}_j] = -i\,\epsilon_{ijk}\,\hat{K}_k \quad . \tag{III.4}$$

† An abstract group is defined solely by the "multiplication table" of the group elements, or in the case of Lie groups like the Poincarè group, by the commutation relations of the generators of the group, which defines the Lie algebra associated with the group. The elements of the abstract group (or Lie algebra) are indicated here by symbols with "hats." The elements in any (matrix or other operator) realization of the group will be denoted by symbols without hats.

The $\hat{J}_i$ and $\hat{K}_i$ together generate the homogeneous Lorentz subgroup $\mathcal{L}$ of $\mathcal{P}$, with elements of the form $\exp\{i\,\theta\,\underset{\sim}{n}\cdot\hat{\underset{\sim}{J}} + i\,\theta'\underset{\sim}{n}'\cdot\hat{\underset{\sim}{K}}\}$.

(iii) $\hat{P}_0$ and $\hat{\underset{\sim}{P}} = (\hat{P}_1, \hat{P}_2, \hat{P}_3)$, which generate translations of the origin of coordinates along the time axis and the three space axes respectively, without change of orientation of the axes. They satisfy

$$[\hat{P}_i, \hat{P}_j] = 0 \tag{III.5}$$

$$[\hat{J}_i, \hat{P}_j] = i\,\epsilon_{ijk}\,\hat{P}_k \tag{III.6}$$

$$[\hat{K}_i, \hat{P}_j] = i\,\delta_{ij}\,\hat{P}_0 \tag{III.7}$$

$$[\hat{P}_i, \hat{P}_0] = 0 \tag{III.8}$$

$$[\hat{J}_i, \hat{P}_0] = 0 \tag{III.9}$$

$$[\hat{K}_i, \hat{P}_0] = i\,\hat{P}_i \tag{III.10}$$

Eqs. (III.2) through (III.10) define the Lie algebra of the ten generators $\hat{J}_i$, $\hat{K}_i$, $\hat{P}_0$, $\hat{P}_i$ of $\mathcal{P}$. The same commutation relations will be obeyed, by definition, by the representatives of these generators (denoted by the same symbols but without hats) in any realization of the Lie algebra.

It is pertinent to note here that by complex conjugation of the representatives of the group elements in any representation, one gets another representation whose generators are obtained from the original ones by the replacement

$$G \to -G^* \quad . \tag{III.11}$$

That the Lie algebra remains unchanged under this mapping is easily verified. In the case of the rotation group, each representation $D(\underset{\sim}{s})$ (corresponding to spin s, $s = 0, \frac{1}{2}, 1, \ldots$), is known to be equivalent to its complex conjugate;[6] i.e., there exists a unitary matrix ζ which transforms the generators $\underset{\sim}{s}$ (the three-vector of spin-s angular momentum matrices) into $-\underset{\sim}{s}^*$:

$$\zeta^{-1}\,\underset{\sim}{s}\,\zeta = -\underset{\sim}{s}^* \quad , \tag{III.12a}$$

$$\zeta\zeta^\dagger = 1, \quad \zeta\zeta^* = (-1)^{2s}, \quad \tilde{\zeta} = (-1)^{2s}\zeta \quad , \tag{III.12b}$$

where the tilde sign denotes transposition. In the special case when the spin is $s = \frac{1}{2}$ we have the familiar representation $\underset{\sim}{s} = \frac{1}{2}\underset{\sim}{\sigma}$ for the

angular momentum operator, in terms of the Pauli matrices σ_1, σ_2, σ_3. The matrix ζ here is $i\sigma_2$. We will need to make use of the result (III.12) rather frequently.

B. Irreducible Representations of the Lorentz Group $\mathcal{L}$

(i) Finite Dimensional Representations

It follows from (III.2)-(III.4) that the combinations

$$\underset{\sim}{\hat{M}} = \tfrac{1}{2}(\underset{\sim}{\hat{J}} + i\underset{\sim}{\hat{K}}) \quad \text{and} \quad \underset{\sim}{\hat{N}} = \tfrac{1}{2}(\underset{\sim}{\hat{J}} - i\underset{\sim}{\hat{K}}) \tag{III.13}$$

of the generators of $\mathcal{L}$ obey

$$[\hat{M}_i, \hat{M}_j] = i\varepsilon_{ijk}\hat{M}_k, \; [\hat{N}_i, \hat{N}_j] = i\varepsilon_{ijk}\hat{N}_k, \; [\hat{M}_i, \hat{N}_j] = 0 \tag{III.14}$$

so that they behave like generators of two independent rotation groups. Thus $\underset{\sim}{\hat{M}}^2$ and $\underset{\sim}{\hat{N}}^2$ are Casimir operators of $\mathcal{L}$, i.e. they commute with all generators, and are therefore given, in any irreducible representation, by the unit matrix multiplied by their respective eigenvalues. By choosing a specific pair of eigenvalues familiar from angular momentum theory, namely $m(m+1)$ of $\underset{\sim}{M}^2$ and $n(n+1)$ of $\underset{\sim}{N}^2$ (where m and n are, independently, non-negative integers or half-integers), we get an irreducible representation of $\mathcal{L}$, denoted by $D(m,n)$. It is of dimension $(2m+1)(2n+1)$. The representation matrices $\underset{\sim}{M}$ and $\underset{\sim}{N}$ here are Hermitian, and lead to a Hermitian

$$\underset{\sim}{J} = \underset{\sim}{M} + \underset{\sim}{N} \tag{III.15a}$$

and an anti-Hermitian

$$\underset{\sim}{K} = -i(\underset{\sim}{M} - \underset{\sim}{N}) \quad . \tag{III.15b}$$

The representation $D(m,n)$ is thus non-unitary, as it should be, since $\mathcal{L}$ is a non-compact group and non-compact groups cannot have any non-trivial finite-dimensional unitary representations.

When complex conjugation of the above representation is performed, $\underset{\sim}{J} \to -\underset{\sim}{J}^*$ and $\underset{\sim}{K} \to -\underset{\sim}{K}^*$, so that $\underset{\sim}{M} \to -\underset{\sim}{N}^*$ which is equivalent to $\underset{\sim}{N}$ in view of the remarks at the end of the last subsection. Thus the result of complex conjugation is effectively the interchange $\underset{\sim}{M} \leftrightarrow \underset{\sim}{N}$, so that $D(m,n) \leftrightarrow D(n,m)$.

It is to be noted that each irreducible representation $D(m,n)$ of $\mathcal{L}$ is in general reducible with respect to the rotation subgroup. This is clear from Eq. (III.15a) which shows the true angular momentum vector (rotation generator) $\underset{\sim}{J}$ to be the sum of two angular momentum-like vectors $\underset{\sim}{M}$ and $\underset{\sim}{N}$ with definite magnitudes m and n. Consequently

this representation must contain all the spin values $(m+n)$, $(m+n-1)$, $\ldots$, $|m-n|$, each value occurring once. More formally stated, $D(m,n)$ reduces with respect to the rotation subgroup of $\mathfrak{L}$ into a direct sum of irreducible representations $D(s)$ of the rotation group,

$$D(m,n) \sim D(m+n) \oplus D(m+n-1) \oplus \ldots \oplus D(|m-n|) \quad . \tag{III.16}$$

It is only the representations of the types $D(s,0)$ and $D(0,s)$ which have a unique spin content s. For this reason these are especially interesting, and we employ these exclusively in our development of the theory for arbtrary spin s in later sections. For future convenience we make a brief mention here of the salient properties of these representations.

In the representation $D(s,0)$, $\underset{\sim}{N} \equiv \frac{1}{2}(\underset{\sim}{J} - i\underset{\sim}{K}) = 0$, so that $\underset{\sim}{K} = -i\underset{\sim}{J}$. Further, since it is $(2s+1)$-dimensional, and irreducible with respect to the rotation subgroup as mentioned above, $\underset{\sim}{J}$ must be given by the vector $\underset{\sim}{s}$ of $(2s+1)$-dimensional angular momentum matrices. Thus, for

$$D(s,0): \qquad \underset{\sim}{J} = \underset{\sim}{s}, \ \underset{\sim}{K} = -i\underset{\sim}{s} \quad . \tag{III.17}$$

The complex conjugate representation, $D(0,s)$, would then be characterised (according to the prescription (III.11)) by the generators

$$\underset{\sim}{J} = -\underset{\sim}{s}^* \quad \text{and} \quad \underset{\sim}{K} = -i\underset{\sim}{s}^* \quad , \tag{III.18}$$

but one could equally well use an equivalent set of generators, obtained by similarity transformation of (III.18) by the matrix ζ defined in (III.12), wherein one has

$$D(0,s): \qquad \underset{\sim}{J} = \underset{\sim}{s} \qquad \text{and} \qquad \underset{\sim}{K} = i\underset{\sim}{s} \quad . \tag{III.19}$$

The transformation of $(2s+1)$-component spin functions under finite transformations in the representations generated by (III.17), (III.18) and (III.19) are, respectively,

$$\chi_u \to [D(L)]_u{}^v \chi_v \tag{III.17a}$$

$$\eta_{\dot{u}} \to [D^*(L)]_{\dot{u}}{}^{\dot{v}} \eta_{\dot{v}} \tag{III.18a}$$

$$\varphi^{\dot{u}} \to [D^{\dagger -1}(L)]^{\dot{u}}{}_{\dot{v}} \varphi^{\dot{v}} \tag{III.19a}$$

where the indices u, v run from 1 to $(2s+1)$. The dotting on the indices in (III.18a) serves as a reminder that the transformation matrix to be used is the complex conjugate of $D(L) \equiv D^{(s,0)}(L)$. And the

raising (or lowering) of any index, accomplished by operating with the the matrix ζ of (III.12), for example

$$\varphi^{\dot{u}} = \zeta^{\dot{u}\dot{v}}\eta_{\dot{v}} \quad \text{or} \quad \eta_{\dot{u}} = (\zeta^{-1})_{\dot{u}\dot{v}}\,\varphi^{\dot{v}} = (-)^{2s}\,\zeta^{*}_{\dot{u}\dot{v}}\,\varphi^{\dot{v}}$$

is to be accompanied by combined transposition and inversion of the representation matrices. The reader is invited to verify that (III.18) and (III.19) are generators of (III.18a) and (III.19a), and that both of them correspond to D(0,s) within equivalence. It is a consequence of (III.18a) and (III.19a) that

$$\eta_{\dot{u}}\,\varphi^{\dot{u}} \text{ (sum over } \dot{u}) \qquad \text{(III.20)}$$

is invariant, and a similar invariant can be defined in terms of the undotted indices too. The properties (III.12b) of the index-raising and -lowering operator ζ show that $\eta_{\dot{u}}\,\varphi^{\dot{u}} = (-1)^{2s}\eta^{u}\varphi_{\dot{u}}$.

The above notation is a generalization of the usual spinor notation, applied to two-component spinors[7] χ_A, φ^A etc. (A = 1,2) which transform according to $D(\frac{1}{2},0)$ and $D(0,\frac{1}{2})$ respectively. What we have called χ_u (u = 1,2,...,2s+1) is equivalent to a spinor of rank (2s) totally symmetric in (2s) indices, each of which ranges from 1 to 2.

$$\chi_u \sim \chi_{A_1 A_2 \ldots A_{2s}} \quad . \qquad \text{(III.21a)}$$

Similarly

$$\varphi^{\dot{u}} \sim \varphi^{\dot{A}_1 \dot{A}_2 \ldots \dot{A}_{2s}} \quad . \qquad \text{(III.21b)}$$

More generally, a quantity transforming according to the representation D(m,n) may be denoted equivalently by

$$\psi_u{}^{\dot{v}} \sim \psi^{\dot{B}_1 \dot{B}_2 \ldots \dot{B}_{2n}}_{A_1 A_2 \ldots A_{2m}} \quad , \qquad \text{(III.21c)}$$

with u = 1,2,...,2m+1 and v = 1,2,...2n+1. The indices u and $\dot{v}$ transform according to (III.17a) and (III.19a) respectively:

$$\psi_u{}^{\dot{v}} \to [D^{(m,o)}(L)]_u{}^{u'}\,[D^{(n,o)\dagger -1}(L)]^{\dot{v}}{}_{\dot{v}'}\,\psi_{u'}{}^{\dot{v}'} \qquad \text{(III.22a)}$$

The infinitesimal generators may then be written as

$$\underline{J} = \underline{s}^{(m)} \times 1 + 1 \times \underline{s}^{(n)} \qquad \text{(III.22b)}$$

$$\underset{\sim}{K} = -i[\underset{\sim}{s}^{(m)} \times 1 - 1 \times \underset{\sim}{s}^{(n)}] \qquad \text{(III.22c)}$$

where $\underset{\sim}{s}^{(m)}$ and $\underset{\sim}{s}^{(n)}$ are spin m and spin n angular momentum matrices which act on the indices u and $\dot{v}$ respectively. The four-dimensional representation $D(\frac{1}{2},\frac{1}{2})$ is of special interest because it is equivalent to the transformation of four-vectors, and every conventional (manifestly covariant) relativistic wave equation, linear in the space-time differential operators, involves a vector operator coupled to the vector $\partial/\partial x^{\mu}$. The reader may verify that the four linear combinations $V^{\mu} = g^{\mu\dot{A}B}V_{\dot{A}B}$ of the components of a spinor $V_{\dot{A}B}$ (which transforms according to $D(\frac{1}{2},\frac{1}{2})$ as indicated by one dotted and one undotted index) transform like a four-vector. The coefficients $g^{\mu\dot{A}B}$ are conveniently defined as the elements of matrices $g^0 = 1$ and $g^i = \sigma_i$ $(i = 1,2,3)$ where the σ_i are the Pauli matrices.

(ii) Infinite Dimensional Representations

As already mentioned, the above finite-dimensional representations of $\mathfrak{L}$ are not unitary. We shall not make explicit use of the infinite-dimensional representations (unitary or otherwise), but for completeness we include here the classification of such representations too. They are usually labelled, not by the eigenvalues of $\underset{\sim}{M}^2$ and $\underset{\sim}{N}^2$, but by two numbers k_o and c defined in terms of the eigenvalues

$$k_o^2 + c^2 - 1 \quad \text{of} \quad \underset{\sim}{I}^2 - \underset{\sim}{K}^2 \equiv 2(\underset{\sim}{M}^2 + \underset{\sim}{N}^2) , \qquad \text{(III.23a)}$$

and

$$-ik_o c \quad \text{of} \quad \underset{\sim}{I}'\underset{\sim}{K} \equiv \underset{\sim}{M}^2 - \underset{\sim}{N}^2 . \qquad \text{(III.23b)}$$

It is known[8] that with every pair (k_o, c), where

$$k_o = 0, \tfrac{1}{2}, 1, \ldots \qquad \text{(III.23c)}$$

and

$$c = \text{a complex number},$$

is associated an irreducible representation of $\mathfrak{L}$ which is unitary if (a) c is pure imaginary (Principal series representations), or if (b) $k_o = 0$ and c is any real number such that $0 \leq c \leq 1$ (Supplementary series representations).

All other representations are non-unitary, and in particular, the finite-dimensional ones already discussed are recovered if $|c| - k_o$ is

a positive integer. The relation to the labels m,n used earlier for this case is

$$k_o = |m - n| \quad \text{and} \quad |c| = m+n+1 \quad . \tag{III.24}$$

Quite generally, each representation reduces under the rotation subgroup into irreducible representations corresponding to spins $s = k_o$, $k_o+1, \ldots$, there being an upper limit $s_{max} = |c|-1$ only if (III.24) is satisfied.

C. Wave Functions and Their Transformations

It is customary to assume that the wave function $\psi(\underset{\sim}{x},t)$ is locally covariant† in the sense that if the wave function at a given space-time point, labelled by coordinates $\underset{\sim}{x},t$ and $\underset{\sim}{x}',t'$ by different observers, is seen by these observers to be $\psi(\underset{\sim}{x},t)$ and $\psi'(\underset{\sim}{x}',t')$ respectively, then

$$\psi'(\underset{\sim}{x}',t') = S(L)\ \psi(\underset{\sim}{x},t) \tag{III.25}$$

where S(L) is a numerical matrix independent of coordinates or differential operators. Local covariance might seem to be a self-evident requirement, but there are formalisms which do not possess this. The transformation of the components of the wave function in Wigner's unitary irreducible representations,[5] for example, is given by a momentum-dependent matrix in the momentum space, and is non-local (in the above sense) in configuration space. But we leave this point for later discussion and confine ourselves to (III.25) for the time being. The S(L) are then matrices forming one of the irreducible representations discussed in Sec. III.B., or a direct sum of a number of these. The inhomogeneous transformations (translations) are supposed to affect only the space-time coordinates, not the spin indices. By rewriting (III.25) as $\psi'(Lx) = S(L)\ \psi(x)$, or

$$\psi'(x) = S(L)\ \psi(L^{-1}x) \quad , \tag{III.26}$$

and taking an infinitesimal transformation $L = 1 + i\epsilon F$, $S(L) = 1 + i\epsilon G^{(s)}$, we find that under such a transformation,

†The term "manifest covariance" has been used with different meanings in the literature. Here we reserve it to describe the explicitly covariant appearance of an equation (i.e. of the differential matrix operator acting on the wave function in the equation).

$$\psi(x) \to \psi'(x) = [1 + i\epsilon(G^{(o)} + G^{(s)})]\,\psi(x) \quad . \tag{III.27}$$

Here $G^{(s)}$ is the "spin" part of the generator of infinitesimal transformations on ψ, and is given by the appropriate linear combination (depending on the actual Lorentz transformation performed) of the generators (III.22b,c) if ψ transforms according to $D(m,n)$. For the "orbital" part $G^{(o)}$, which reflects the effect of the change in argument $x \to L^{-1}x$ on the functional form of ψ, we have the familiar forms†

$$P_o^{(o)} = -i\frac{\partial}{\partial t} \equiv p_o \quad , \tag{III.28a}$$

$$\underset{\sim}{P}^{(o)} = -i\,\underset{\sim}{\nabla} \equiv \underset{\sim}{p} \quad , \tag{III.28b}$$

$$\underset{\sim}{J}^{(o)} = \underset{\sim}{x} \times \underset{\sim}{p} \quad , \tag{III.28c}$$

$$\underset{\sim}{K}^{o} = t\underset{\sim}{p} + \underset{\sim}{x}p_o \quad . \tag{III.28d}$$

These generators (as well as the $G^{(s)}$) satisfy Eqs. (III.5)-(III.10).

It might seem, from the fact that the "orbital" and "spin" parts of the generators are simply additive, that what we have here is a simple coupling of two representations, in the same sense as the addition of orbital and spin angular momenta in the theory of the rotation group. This would in fact be true, if we were content with having representations of $\mathcal{P}$ defined over functions ψ of space and time (or, in the language employed in Sec. II, representations which map whole trajectories into one another). But the quantum mechanical aspect of relativistic wave equations calls for a representation defined over wave functions at fixed time, in order that an invariant scalar product between fixed time states can be defined. Eqs. (III.28) do not provide such a representation because the operator p_o in (III.28a) and (III.28d) are clearly undefinable on fixed-time functions. One is then compelled to introduce an operator (to be denoted by -H), which is defined over functions $\psi(\underset{\sim}{x},t)$ at fixed t (i.e., operates on the spin index and on $\underset{\sim}{x}$, but not on t) and produces the same effect on them as p_o (which acts on the time variable). In other words, one has to have a wave equation

†For example, $K_1^{(o)}$ would be obtained from the fact that $L^{-1}x$ in this case is $(x^o+\beta\gamma x^1, x^1+\beta\gamma x^o, x^2, x^3)$, so that to first order in small $\beta(=v/c)$, $\psi(L^{-1}x) = \psi(x) + \beta\left(x^1\frac{\partial}{\partial x^o} + \beta x^o\frac{\partial}{\partial x^1}\right)\psi(x)$.

$$p_o\psi = -H\psi \quad \text{or} \quad i\frac{\partial\psi}{\partial t} = H\psi \quad . \tag{III.29}$$

The operator H, which we will hereafter refer to as the Hamiltonian for obvious reasons, must be such that on substituting $-H$ for P_o everywhere, the Eqs. (III.5)-(III.10) remain valid, i.e., $-H$ must behave like P_o with respect to group properties in order that (III.29) be relativistically invariant. The important point now is that an operator H which has these properties will, quite generally, be a matrix-differential operator, i.e. it acts both on the space and spin variables. Therefore the generators of $\mathcal{P}$ on the wave functions no longer contain a pure "orbital" part independent of the "spin" part. Consider, as an example, the generators in the interesting case when the transformation of the wave function at a given space-time point is according to the reducible representation $D(o,s) \oplus D(s,o)$. The wave function is then $2(2s+1)$-dimensional, and we have

$$P_o = -i\frac{\partial}{\partial t_o} \to -H \tag{III.30a}$$

$$\underset{\sim}{P} = -i\underset{\sim}{\nabla} \equiv \underset{\sim}{p} \tag{III.30b}$$

$$\underset{\sim}{I} = \underset{\sim}{x} \times \underset{\sim}{p} + \underset{\sim}{S} \tag{III.30c}$$

$$\underset{\sim}{K} = t\underset{\sim}{p} + \underset{\sim}{x}p_o + i\underset{\sim}{\lambda} \to t\underset{\sim}{p} - \underset{\sim}{x}H + i\underset{\sim}{\lambda} \tag{III.30d}$$

where

$$\underset{\sim}{S} = \begin{pmatrix} \underset{\sim}{s} & o \\ o & \underset{\sim}{s} \end{pmatrix} , \qquad \underset{\sim}{\lambda} = \begin{pmatrix} \underset{\sim}{s} & o \\ o & -\underset{\sim}{s} \end{pmatrix} \equiv \rho_3 \underset{\sim}{S} \quad . \tag{III.31}$$

The matrix ρ_3 is one of the Pauli matrices

$$\rho_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} , \qquad \rho_2 = \begin{pmatrix} o & -i \\ i & o \end{pmatrix} , \qquad \rho_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} , \tag{III.32}$$

whose elements are to be thought of as $(2s+1)$-dimensional matrices. The pure spin parts $\underset{\sim}{S}$ and $i\underset{\sim}{\lambda}$ of these generators are obtained by putting the $(2s+1)$-dimensional matrices (III.19) of $D(o,s)$ and (III.17) of $D(s,o)$ together to form the block-diagonal matrix of dimension $2(2s+1)$ for the reducible representation $D(o,s) \oplus D(s,o)$. We shall later on determine the operator H for arbitrary spin s in such a way that at least when invariance under the discrete operations T, C, P also are required, H cannot be merely a function of the differential

operator $\underset{\sim}{p}$, but must contain matrices constructed from $\underset{\sim}{s}$, $\underset{\sim}{\lambda}$, etc. Actually this fact can already be seen from the well-known example of the Dirac Hamiltonian (the Dirac wave function being something which transforms according to $D(0,\frac{1}{2}) + D(\frac{1}{2},0)$). In the notation used here it is

$$H_{\text{Dirac}} = 2\underset{\sim}{\lambda}\cdot\underset{\sim}{p} + \rho_1 m \quad . \tag{III.33}$$

To summarise the argument of the preceding paragraph, introduction of the notion of a scalar product, essential for quantum mechanical interpretation of ψ, necessitates the definition of a Hamiltonian operator.† The explicit construction of an invariant scalar product is another matter. The scalar product between two wave functions φ, ψ is not, in general, given by $\int\varphi^\dagger\psi d^3x$ because in order that this quantity be invariant under relativistic transformations, one must have, for example, $\int\varphi^\dagger\psi d^3x = \int[1 + i\varepsilon G)\varphi]^\dagger[(1 + i\varepsilon G)\psi]d^3x$, (III.34), i.e. the generators G must be Hermitian in the ordinary sense. But this is not necessarily so. Indeed, in view of the non-Hermitian term $\underset{\sim}{x}H$ and the anti-Hermitian term $i\underset{\sim}{\lambda}$ in (III.30d), it is evident prima facie that $\underset{\sim}{K}$ cannot be Hermitian in the example considered there, unless exceptional circumstances prevail. The spin $\frac{1}{2}$ case is exceptional, and by introducing the Dirac Hamiltonian (III.33) in (III.30d) one verifies that $\underset{\sim}{K}$ is Hermitian. In the general case, however, one has to define the scalar product as[9]

$$(\varphi,\psi) = \int\varphi^\dagger M\psi d^3x \tag{III.35}$$

where M is a "metric" operator in the space of wave functions,‡ and is to be determined in such a way that (III.35) is relativistically invariant, i.e.,

$$\int\varphi^\dagger M\psi d^3x = \int[(1 + i\varepsilon G)\varphi]^\dagger M[(1 + i\varepsilon G)]d^3x \quad , \tag{III.36}$$

or

$$MG = G^\dagger M \quad . \tag{III.37}$$

---

†It is not implied that for any specified local transformation property of ψ, an equation in the form (III.29) should exist. To take the example of spinless particles, the wave function ψ obeying the Klein-Gordon equation and its derivatives $\partial_\mu\psi$ must be taken together to get an equation of this form.

‡We don't have a quantized field yet, and the metric M here must be clearly distinguished from metrics (definite or otherwise) on the space of states of a quantized field.

Of course one must also have

$$M^{\dagger} = M \qquad \text{(III.38)}$$

in order that (φ,φ) be real. By (III.37), the actual expression for M will depend on the expressions for the generators G, in particular, on that for H. We must therefore postpone further consideration of M till after we have gone into the problem of determination of H.

There is one question, however, which may be legitimately raised at this point. Why are we stuck with a metric M (which is, in general, a matrix-cum-differential operator), while Foldy,[10] for instance, is able to define a simple scalar product $\int \varphi_c^* \psi_c d^3x$ between wave functions φ_c, ψ_c in his "canonical" representation? The answer is that we are insisting on something which Foldy is not, namely, local covariance of the wave function in the sense of Eq. (III.25). The "spin" part of the boost generator, i.e. the part which acts on the index of the multi-component wave function ψ--the counterpart of the term $i\underset{\sim}{\lambda}$ in (III.30d)--contains, in Foldy's case, terms like $\rho_1 (\underset{\sim}{s} \times \underset{\sim}{p})/(m + E)$, (where $E = +\sqrt{p^2 + m^2}$), which are nonlocal in the configuration space. Actually it is this canonical representation, wherein the transformation of the wave function is not locally covariant, that emerges directly from an analysis of the unitary irreducible representations of $\mathcal{P}$ according to the method of Wigner. The canonical representation is related to the representation (III.30) by a similarity transformation† which is momentum-dependent, and is not, in general, unitarity. The simplest way of seeing this is by noting that if the Hermitian metric operator M in (III.35) is positive definite, then it can be decomposed in the form $M = R^{\dagger}R$, and the scalar product is then

$$(\varphi,\psi) = \int \varphi^{\dagger} M \psi d^3x = \int \varphi_c^{\dagger} \psi_c d^3x \qquad \text{(III.39)}$$

where

$$\psi_c = R\psi, \quad \varphi_c = R\varphi \quad . \qquad \text{(III.40)}$$

Thus in terms of the transformed functions ψ_c, φ_c, the scalar product is of the simple type, without any metric operator.

D. The Discrete Transformations

We have already introduced the discrete transformations of space and time inversion. It is obvious that performance of either of

†This is really a generalized Foldy-Wouthuysen transformation, for an explicit determination of which, see Ref. 9 (in the case of half-integral spin).

these operations twice in succession leads to the identity transformation $x^\mu \to x^\mu$. Further, the performance of both time and space inversion, in either order, takes $x^\mu \to -x^\mu$. Thus, for the abstract operators $\hat{P}$ of space inversion and $\hat{T}$ of time inversion, it is true that $\hat{P}^2 = \hat{T}^2 = 1$, and $\hat{P}\hat{T} = \hat{T}\hat{P}$. Naturally, $\hat{P}^2$ and $\hat{T}^2$ should take every state of a quantum system into itself. However the wave function representing the state need not be mapped identically into itself--it is quite sufficient if the wave function is taken into a constant multiple of itself, where the multiplying factor is a "phase factor" of unit modulus. This is because multiplication of a wave function by a phase factor does not alter the physical state which it represents. As a consequence, the operators P, T representing the effect of $\hat{P}$, $\hat{T}$ on the wave function need satisfy only the weaker conditions

$$P^2 \sim 1, \quad T^2 \sim 1, \quad PT \sim TP \quad , \tag{III.41}$$

where the sign $\sim$ means equality to within a factor of unit modulus. Our insistence on leaving the door open for these arbitrary phase factors may appear a little too legalistic at this stage; however the wisdom of not curtailing this freedom will become evident when we try to second-quantize the wave equation we derive. On the other hand, in commutation relations involving P or T and the generators of the connected group $\mathcal{P}$, no free phase factors remain, for reasons of continuity with the identity transformation. The relations are

$$PP_0 = P_0P, \quad P\underset{\sim}{P} = -\underset{\sim}{P}P, \quad P\underset{\sim}{J} = \underset{\sim}{J}P, \quad P\underset{\sim}{K} = -\underset{\sim}{K}P \quad , \tag{III.42}$$

$$TP_0 = P_0T, \quad T\underset{\sim}{P} = -\underset{\sim}{P}T, \quad T\underset{\sim}{J} = -\underset{\sim}{J}T, \quad T\underset{\sim}{K} = \underset{\sim}{K}T \quad . \tag{III.43}$$

These equations reproduce the familiar behaviour of $\underset{\sim}{P}$, $\underset{\sim}{J}$ etc. (considered as observables: momentum, angular momentum etc., rather than merely as generators of the group) under space and time inversion. It is pertinent to observe here that the correct transformation properties (III.43) under T are obtained only if T is chosen to be an antilinear operator,[10] i.e. T must consist of complex conjugation together with a linear operation. To illustrate this point, consider the operator $(1 + i\epsilon P_3)$ for a displacement through an infinitesimal distance ϵ along the z-axis, followed or preceded by a time reversal. Since the space and time directions are completely independent, one evidently has $(1 + i\epsilon P_3)T = T(1 + i\epsilon P_3)$. If T were a linear operator, this equation would immediately lead to $P_3T = TP_3$, which states that the third component of the momentum remains unchanged under time reversal. But we know that momentum does change sign under T. To ensure this property we have to make T antilinear, in which case

$T + i\varepsilon P_3 T = (1 + i\varepsilon P_3)T = T(1 + i\varepsilon P_3 T) = T + T(i\varepsilon P_3) = T - i\varepsilon T P_3$, since the sign of i is changed on pulling it through an antilinear T. The correct relation $TP_3 = -P_3 T$ is thus obtained. The same can be said of the other generators.

The effect of P and T on wave functions $\psi(\underline{x},t)$ will thus have to be defined by

$$P\psi(\underline{x},t) = \sigma\psi(-\underline{x},t) \tag{III.44}$$

and

$$T\psi(\underline{x},t) = \tau\psi^*(\underline{x},-t) \tag{III.45}$$

where σ, τ are purely numerical matrices, and the antilinearity of T is reflected in the complex conjugation of ψ in (III.45).

Consider now the operation C, called charge conjugation, about which we have so far been silent. Unlike T and P, C is not a relativistic transformation. It is not an element of the extended Poincaré group: since C leaves the coordinates x^μ unchanged, it adds nothing to the group structure. Then what is C? It is an operation which maps a representation of $\mathcal{P}$ on to its complex conjugate, and is defined as an antilinear operator such that

$$CP_0 = -P_0 C,\ C\underline{P} = -\underline{P}C,\ C\underline{J} = -\underline{J}C,\ C\underline{K} = -\underline{K}C. \tag{III.46}$$

For the same reasons as in the case of P and T, it must satisfy

$$C^2 \sim 1,\ CP \sim PC,\ CT \sim TC\ . \tag{III.47}$$

Its effect on the wave function is given by

$$C\psi(\underline{x},t) = \kappa\psi^*(\underline{x},t) \tag{III.48}$$

where κ is a matrix with purely numerical elements.

The matrices σ, τ, κ are so far undetermined, but are strongly restricted by the conditions (III.41)-(III.48). Considering parity, for instance, it follows from the use of Eqs. (III.30) and (III.44) that the first two of the Eqs. (III.42) are trivially satisfied, while the last one, $P\underline{K}\psi(\underline{x},t) = -\underline{K}P\psi(\underline{x},t)$, requires

$$\sigma(K\psi)(-\underline{x},t) = -K\sigma\psi(-\underline{x},t)$$

or

$$\sigma(-t\underline{p} - \underline{x}p_0 + i\underline{\lambda})\,\psi(-\underline{x},t) = (-t\underline{p} - \underline{x}p_0 - i\underline{\lambda})\,\sigma\psi(-\underline{x},t)\ .$$

To ensure this, we must have

$$\sigma\underset{\sim}{\lambda} = -\underset{\sim}{\lambda}\sigma \ , \quad \text{or} \quad \sigma\rho_3\underset{\sim}{S} = -\rho_3\underset{\sim}{S}\sigma \tag{III.49a}$$

Similarly, $P\underset{\sim}{I} = \underset{\sim}{I}P$ demands

$$\sigma\underset{\sim}{S} = \underset{\sim}{S}\sigma \tag{III.49b}$$

Eqs. (III.49) determine σ to be a linear combination of the Pauli matrices ρ_1 and ρ_2. We will develop the theory with the specific choice

$$\sigma = \rho_1 \ . \tag{III.50}$$

An over-all phase factor which is permitted by $P^2 \sim 1$ turns out to be not important, and mixing in of a term proportional to ρ_2 in σ would be equivalent to combining (III.50) with a transformation which changes the relative phase of the $D(0,\frac{1}{2})$ and $D(\frac{1}{2},0)$ parts of the wave function. The effect of this can be studied independently, but we will restrict ourselves in the following to the identification (III.50).

As for the time reversal operation, again the first two of Eqs. (III.43) are satisfied trivially--keeping in mind the antilinearity of T --by virtue of the explicit forms (III.30a,b) for P_0 and $\underset{\sim}{P}$. The requirement $T\underset{\sim}{I} = -\underset{\sim}{I}T$ leads to

$$\tau[(\underset{\sim}{x} \times -i\underset{\sim}{\nabla} + \underset{\sim}{S})\,\psi(\underset{\sim}{x},-t)]^* = -[\underset{\sim}{x} \times -i\underset{\sim}{\nabla} + \underset{\sim}{S}]\tau\psi^*(\underset{\sim}{x},-t) \ ,$$

or

$$\tau\underset{\sim}{S}^* = -\underset{\sim}{S}\tau \ . \tag{III.51a}$$

Similarly, from $T\underset{\sim}{K} = \underset{\sim}{K}T$, one finds

$$T(i\underset{\sim}{\lambda}) = (i\underset{\sim}{\lambda})T, \text{ or } T\underset{\sim}{\lambda} = -\underset{\sim}{\lambda}T \ , \tag{III.51b}$$

i.e.,

$$\tau(i\underset{\sim}{\lambda})^* = (i\underset{\sim}{\lambda})\tau, \text{ or } \tau\rho_3\underset{\sim}{S}^* = -\rho_3\underset{\sim}{S}\tau \ . \tag{III.51c}$$

To satisfy the two Eqs. (III.51a,c) simultaneously, τ must be of the form

$$\tau = \begin{pmatrix} \tau' & 0 \\ 0 & \tau'' \end{pmatrix} \tag{III.52}$$

where τ', τ'' are matrices of dimension $(2s + 1)$ which satisfy

$$\tau' \underset{\sim}{s}^* = -\underset{\sim}{s}\tau' \text{ and } \tau'' \underset{\sim}{s}^* = -\underset{\sim}{s}\tau'' \quad . \tag{III.53}$$

These equations fix τ' and τ'' to be multiples of the matrix ζ defined in Eqs. (III.12), and the requirement $T^2 \sim 1$, which reduces to $\tau\tau^* \sim 1$, restricts these multiples to be unimodular phase factors. Thus, apart from an overall phase factor which we will ignore,

$$\tau = \begin{pmatrix} \zeta & 0 \\ 0 & \zeta e^{i\theta_\tau} \end{pmatrix} \quad , e^{i\theta_\tau} = \pm 1 \quad . \tag{III.54}$$

The constraint $e^{i\theta_\tau} = \pm 1$ comes from the condition $PT \sim TP$.

An entirely analogous treatment of the charge conjugation operator C shows that

$$\kappa = \begin{pmatrix} 0 & \zeta \\ \zeta e^{i\theta_c} & 0 \end{pmatrix} \quad , e^{i\theta_c} = \pm 1 \quad . \tag{III.55}$$

The two ambiguous phases in (III.54) and (III.55) are uncorrelated, and we thus have four kinds of possibilities. In what follows, the commutation rules

$$TP = e^{i\theta_\tau} PT \text{ and } CP = e^{i\theta_c} PC \tag{III.56}$$

of P with T and C will play a crucial role. It may also be noted that

$$T^2 = (-1)^{2s} \tag{III.57}$$

and

$$C^2 = e^{i\theta_c} (-1)^{2s} \quad . \tag{III.58}$$

Finally, we observe that the off-diagonal matrices σ and κ (Eqs. (III.50) and (III.55)) link the $D(0,s)$ and $D(s,0)$ parts of the wave function. In general, if parity and/or charge conjugation operation is to be defined on a locally covariant wave function, then every irreducible representation of $\mathfrak{L}$ contained in the wave function must be accompanied by its complex conjugate irreducible representation.

IV. Derivation of Wave Equation Invariant Under P and Under the Discrete Operations

We are now in a position to construct an H such that Eq. (III.29) is invariant under the above operations. The wave function ψ will be assumed to transform according to $D(0,s) + D(s,0)$ since this would assure a unique spin s, while at the same time permitting definition of the discrete operations T, C, P. Since ψ has no redundant components, there will be no supplementary conditions in the theory. This is certainly an advantage, but the price to be paid for it is that the equation cannot be of the first order in all derivatives and manifestly covariant (except for $s = \frac{1}{2}$). The reason is that the matrix four-vector which is needed for coupling to the four-vector ∂_μ in such equations transforms† as $D(\frac{1}{2},\frac{1}{2})$, and therefore any wave function on which it operates must contain, along with any part transforming as $D(m,n)$, another part transforming according to at least one of the representations $D(m \pm \frac{1}{2}, n \pm \frac{1}{2})$. This is obviously not satisfied for $D(0,s) \oplus D(s,0)$ unless $s = \frac{1}{2}$.

But let us insist that we will be content with having covariance, even if not manifest. Then what is required for invariance of the Eq. (III.29) is clearly that H, as a function of the "dynamical variables" involved in the problem, viz., $\underline{p}$, $\underline{s}$, ρ_1, ρ_2, ρ_3, should have identically the same commutation rules with the generators P_0, $\underline{P}$, $\underline{J}$, K of P and with T, C, P, as $i\frac{\partial}{\partial t} = -P_0$ has. Specifically, the requirements, obtained by replacing P_0 by $-H$ in Eqs. (III.8), (III.9), (III.10), (III.42), (III.43) and (III.46), are

$$[P_i, H] = 0 \quad , \tag{IV.1}$$

$$[J_i, H] = 0 \quad , \tag{IV.2}$$

$$[K_i, H] = -iP_i \quad , \tag{IV.3}$$

$$PH = HP \quad , \tag{IV.4}$$

$$TH = HT \quad , \tag{IV.5}$$

$$CH = -HC \quad . \tag{IV.6}$$

†Recall the statement at the end of Sec. III.B.(i) that the four-vector transformation law is equivalent to $D(\frac{1}{2},\frac{1}{2})$.

Further, in the absence of any supplementary condition† on the wave function, all components of ψ must satisfy the Klein-Gordon equation corresponding to a specified mass m, and this implies

$$H^2 = p^2 + m^2 \equiv E^2 \quad . \tag{IV.7}$$

Our task is now to find an H which is a solution of Eqs. (IV.1)-(IV.7).

We observe first of all that Eq. (IV.1), together with $[P_o, H] = 0$, requires H to be free of $\underset{\sim}{x}$ and t. Eqs. (IV.2) and (IV.4) demand that H be a true scalar with respect to rotations and space inversion. Now it is easy to verify with the aid of Eqs. (III.49) that the only true scalar that can be constructed from the available variables are

$$\underset{\sim}{\lambda}\cdot\underset{\sim}{p} \quad \text{and} \quad \rho_1(\underset{\sim}{\lambda}\cdot\underset{\sim}{p}) \quad . \tag{IV.8}$$

Therefore H must be a function of these only. Now, any function of $\underset{\sim}{\lambda}\cdot\underset{\sim}{p}$ is reducible to a polynomial of degree 2s with coefficients depending on $p = |\underset{\sim}{p}|$. This is because, by virtue of the definition (III.31) of $\underset{\sim}{\lambda}$, $\underset{\sim}{\lambda}\cdot\underset{\sim}{p}$ is a matrix with (2s+1) distinct eigenvalues νp ($\nu = -s, -s+1, \ldots, s-1, s$) -- each eigenvalue occurring twice--so that it satisfies a characteristic equation of degree (2s+1). Consequently H can be written as a polynomial in $(\underset{\sim}{\lambda}\cdot\underset{\sim}{p})$ plus ρ_1 times another polynomial in $(\underset{\sim}{\lambda}\cdot\underset{\sim}{p})$.

At this point it is convenient to impose the Klein-Gordon condition (IV.7) before worrying about the remaining invariance conditions. The process of squaring H which would apparently lead to very complicated expressions, can actually be carried out very simply and elegantly by exploiting once again the matrix nature of $(\underset{\sim}{\lambda}\cdot\underset{\sim}{p})$. All we need to do is to make use of the well known spectral expansion theorem for any diagonalizable matrix A, which states that if Λ_ν are projection operators to the eigenvalues ν of A, then

$$\Lambda_\mu \Lambda_\nu = \Lambda_\mu \, \delta_{\mu\nu} \tag{IV.9}$$

---

†When the wave equation has redundant components, (IV.7) cannot be insisted upon. The simplest example is the Hamiltonian of the spin-1 particle as described by the Kemmer equation,[11] which satisfies $H^3 = E^2 H$. But those components of the wave function which correspond to the zero eigenvalue of H allowed by this equation are constrained to vanish on account of a supplementary condition, which cuts down the independent components in the wave function from 10 to the necessary 6.

and

$$f(A) = \sum_{\nu} f(\nu)\, \Lambda_{\nu} \quad . \tag{IV.10}$$

In the present context we take for A the matrix $\lambda_p \equiv \underline{\lambda}\cdot\underline{p}/p$ (where $p = |\underline{p}|$) so that its eigenvalues are $\nu = -s, -s+1, \ldots, s-1, s$. The explicit form of the projection operator is then

$$\Lambda_{\nu} = \prod_{\mu \neq \nu} \left(\frac{\lambda_p - \mu}{\nu - \mu} \right) , \quad \lambda_p = \frac{\underline{\lambda}\cdot\underline{p}}{p} \quad . \tag{IV.11}$$

It is clear that Λ_{ν} contains both odd and even powers of λ_p, a feature which would create some inconvenience when we try to impose T and C invariance, if H were expressed directly in terms of the Λ_{ν}. To forestall this difficulty, we define

$$B_{\nu} = \Lambda_{\nu} + \Lambda_{-\nu} \quad \text{except } B_0 = \Lambda_0 \tag{IV.12a}$$

and

$$C_{\nu} = \Lambda_{\nu} - \Lambda_{-\nu} \quad , \qquad (\nu \geq 0) \quad ; \tag{IV.12b}$$

which would make B_{ν} an even function of λ_p and C_{ν} an odd function. Eqs. (IV.9) and (IV.12) lead to

$$B_{\mu}B_{\nu} = C_{\mu}C_{\nu} = B_{\mu}\delta_{\mu\nu}, \quad B_{\mu}C_{\nu} = C_{\mu}\delta_{\mu\nu} \quad . \tag{IV.13a}$$

Also, as a special case of (IV.10),

$$1 = \sum_{\nu=-s}^{s} \Lambda_{\nu} = \sum_{\nu \geq 0} B_{\nu} \quad . \tag{IV.13b}$$

Now instead of using (IV.10) to write H as a linear combination of the Λ_{ν}, one can equally well write H in terms of B_{ν} and C_{ν}. We thus have, quite generally,

$$H = \left(\sum b_{\nu} B_{\nu} + \sum c_{\nu} C_{\nu} \right) + \rho_1 \left(\sum b_{\nu}' B_{\nu} + \sum c_{\nu}' C_{\nu} \right) \quad , \tag{IV.14}$$

where the coefficients b_{ν}, c_{ν}, b_{ν}', c_{ν}' may be functions of p. The Klein-Gordon condition can now be trivially imposed in terms of these

coefficients by using Eqs. (IV.12)-(IV.14). One finds the requirement to be

$$b_\nu^2 + c_\nu^2 + b_\nu'^2 - c_\nu'^2 = E^2 \qquad \text{(IV.15)}$$

for each $\nu \geq 0$.

Our remaining task is to use the time reversal and charge conjugation invariance conditions (IV.5) and (IV.6) and the boost invariance condition (IV.3) to determine the coefficients in (IV.14) as functions of p. Consider first time reversal. Since $T\underset{\sim}{\lambda} = -\underset{\sim}{\lambda}T$ and $T\underset{\sim}{P} = -\underset{\sim}{P}T$ from Eqs. (III.43) and (III.51b), we have $T(\underset{\sim}{\lambda}\cdot\underset{\sim}{p}) = (\underset{\sim}{\lambda}\cdot\underset{\sim}{p})T$, so that T leaves the projection operators B_ν and C_ν in (IV.14) unchanged. But it takes each coefficient to its complex conjugate, and further, on account of (III.66), $T\rho_1 = e^{i\theta_T}\rho_1 T$. Therefore

$$TH = \left\{\left(\sum b_\nu^* B_\nu + \sum c_\nu^* C_\nu\right) + e^{i\theta_T}\rho_1\left(\sum b_\nu'^* B_\nu + \sum c_\nu'^* C_\nu\right)\right\} T,$$

and condition (IV.5) which requires this to be equal to HT--with H as in (IV.14)--yields

$$b_\nu^* = b_\nu \ , \quad c_\nu^* = c_\nu \ , \quad e^{i\theta_T} b_\nu'^* = b_\nu' \ , \quad e^{i\theta_T} c_\nu'^* = c_\nu' . \qquad \text{(IV.16)}$$

These are reality conditions on the coefficients, whose exact nature depends on whether the phase $e^{i\theta_T}$ is taken to be +1 or -1. Since we have no means of discriminating between these at the moment, we must keep both possibilities open.

An exactly similar analysis can be made of the requirement of C-invariance. The conditions obtained for T and C invariance are summarised in Table I. It will be observed from the table that if we require both T and C invariance, the b_ν will have to vanish, as well as either the b_ν' or the c_ν'. The various forms which result for the possible Hamiltonians are listed below. For convenience of future comparison we give side by side also two possible forms of the metric operator in each case. Derivation of the metric operators will be given later.

Case (i): $\underline{TP = PT,\ CP = -PC}$.

$$H = \sum c_\nu C_\nu + \rho_1 \sum b_\nu' B_\nu \qquad \text{(IV.17)}$$

with c_ν, b_ν' real, and

$$c_\nu^2 + b_\nu'^2 = E^2 \equiv \underset{\sim}{p}^2 + m^2 \ . \qquad \text{(IV.18)}$$

Table I

Conditions On H, Eq. (IV.14), For T and C Invariance*

	b_ν	c_ν	b_ν'	c_ν'
T(+)	R	R	R	R
T(-)	R	R	I	I
C(+)	I	R	I	R
C(-)	I	R	R	I
T and C	0	R	R if T(+), C(-) I if T(-), C(+) 0 otherwise	R if T(+), C(+) I if T(-), C(-) 0 otherwise

* R = Real, I = Imaginary, T(+) means TP = +PT, C(-) means CP = -CP, etc.

Possible metric operators are

$$M_1 = m^{-1} \sum b_\nu' B_\nu \quad , \tag{IV.19}$$

$$M_2 = (Em)^{-1}\Big[\sum b_\nu' c_\nu C_\nu + \rho_1 \sum b_\nu'^2 B_\nu\Big] \quad . \tag{IV.20}$$

Case (ii): TP = -PT, CP = PC

H is of the same form as in case (i), except that b_ν' is now pure imaginary. The metric operators are

$$M_1 = (iEm)^{-1}\Big[\sum c_\nu b_\nu' B_\nu - i\rho_1 \sum b_\nu'^2 C_\nu\Big] \quad , \tag{IV.21}$$

$$M_2 = (im)^{-1} \sum b_\nu' C_\nu \quad . \tag{IV.22}$$

Case (iii): TP = PT, CP = PC

$$H = \sum c_\nu C_\nu + \rho_1 \sum c_\nu' C_\nu \tag{IV.23}$$

with c_ν, c_ν' real, and

$$c_\nu^2 - c_\nu'^2 = E^2 \quad . \tag{IV.24}$$

In this case

$$M_1 = (Em)^{-1}\left[\sum c_\nu c_\nu' B_\nu - \rho_1 \sum c_\nu'^2 B_\nu\right], \qquad \text{(IV.25)}$$

$$M_2 = m^{-1}\sum c_\nu' C_\nu \quad . \qquad \text{(IV.26)}$$

Case (iv): TP = -PT, CP = -PC

H has the same form as in case (iii), except that c_ν' is now imaginary. As for the metric operators,

$$M_1 = (im)^{-1}\sum c_\nu' B_\nu \quad , \qquad \text{(IV.27)}$$

$$M_2 = (iEm)^{-1}\left[\sum c_\nu c_\nu' C_\nu + i\rho_1 \sum c_\nu'^2 C_\nu\right]. \qquad \text{(IV.28)}$$

Let us now turn to the boost invariance condition, which is the only one left to be considered. It can be rewritten, using (III.30d), as

$$[H,\ t\underset{\sim}{p} - \underset{\sim}{x}H + i\underset{\sim}{\lambda}] = i\underset{\sim}{p} \quad . \qquad \text{(IV.29)}$$

An alternative form, obtained by writing $i\underset{\sim}{p} = \frac{1}{2}[\underset{\sim}{x}, E^2] = \frac{1}{2}(xH^2 - H^2x)$ in (IV.17), is[†]

$$\left[H,\ \{-i[\underset{\sim}{x},H] - 2\underset{\sim}{\lambda}\}\right] = 0 \quad . \qquad \text{(IV.30)}$$

Explicit evaluation of the left hand member in (IV.29) or (IV.30) is difficult in the general case, though it can be done. However it turns out to be sufficient for our purposes, and much simpler, to consider the weaker condition obtained by scalar multiplication of (IV.30) on the left by $\underset{\sim}{p}$, namely

$$\left[H,\ -i\underset{\sim}{p}\cdot[\underset{\sim}{x},H]\right] = 2[H,\ \underset{\sim}{\lambda}\cdot\underset{\sim}{p}] \quad . \qquad \text{(IV.31)}$$

The commutator $-i[\underset{\sim}{x},H]$ in this equation is the gradient of H in $\underset{\sim}{p}$-space, and the scalar product with $\underset{\sim}{p}$ limits the differentiation to one with respect to the "radial" variable p. The restrictions imposed by

---

[†]Eq. (IV.30) shows that the velocity operator, $-i[\underset{\sim}{x},H]$, is not a constant of the motion, since its commutator with H is $[H,\ 2\underset{\sim}{\lambda}] \neq 0$ for any nonzero spin. It is interesting that the existence of Zitterbewegung appears as a direct consequence of the requirement of invariance under boosts.

(IV.31) thus take the form of ordinary differential equations for the coefficients c_ν and b_ν' or c_ν' occurring in H. The derivation of these differential equations is given in the Appendix. As shown there, the equation for c_ν is the same in all four cases, and is given by

$$E^2 \frac{dc_\nu}{dp} = c_\nu p + 2\nu(E^2 - c_\nu^2) \quad . \tag{IV.32}$$

This being a first order differential equation, its solution will depend on one arbitrary constant ℓ_ν, which we define by

$$\lim_{p \to 0} c_\nu = m\ell_\nu \quad . \tag{IV.33}$$

It is easy to show that the solution is

$$c_\nu = E \cdot \frac{(E+p)^{4\nu}(1+\ell_\nu) - m^{4\nu}(1-\ell_\nu)}{(E+p)^{4\nu}(1+\ell_\nu) + m^{4\nu}(1-\ell_\nu)} \quad . \tag{IV.34}$$

It follows then from Eq. (IV.18) that b_ν', occurring in cases (i) and (ii), is

$$b_\nu' = \frac{2E(E+p)^{2\nu} m^{2\nu}(1-\ell_\nu^2)^{\frac{1}{2}}}{(E+p)^{4\nu}(1+\ell_\nu) + m^{4\nu}(1-\ell_\nu)} \quad ; \tag{IV.35}$$

and from Eq. (IV.24), that c_ν', which shows up in cases (iii) and (iv), is

$$c_\nu' = \frac{2E(E+p)^{2\nu} m^{2\nu}(\ell_\nu^2-1)^{\frac{1}{2}}}{(E+p)^{4\nu}(1+\ell_\nu) + m^{4\nu}(1-\ell_\nu)} \quad . \tag{IV.36}$$

These expressions may be cast into simpler form by the substitution

$$E = m \cosh\theta \qquad p = m \sinh\theta \quad , \tag{IV.37}$$

together with the replacement of the parameters ℓ_ν also by hyperbolic functions which however have to be chosen in such a way that the reality conditions on b_ν', c_ν' (Table I) are ensured.

Case (i): b_ν' real.

Here, from (IV.39), $\ell_\nu^2 < 1$, and we set

$$\frac{1}{\sqrt{1-\ell_\nu^2}} = \cosh\eta_\nu \quad , \quad \frac{\ell_\nu}{\sqrt{1-\ell_\nu^2}} = \sinh\eta_\nu \quad . \tag{IV.38}$$

Then (IV.34) and (IV.35) become

$$c_\nu' = E \tanh (2\nu\theta + \eta_\nu) \qquad \text{(IV.39a)}$$

$$b_\nu' = E \operatorname{sech} (2\nu\theta + \eta_\nu) \qquad \text{(IV.39b)}$$

Case (ii): b_ν' imaginary, $\ell_\nu^2 > 1$.
Defining

$$\frac{\ell_\nu}{\sqrt{\ell_\nu^2 - 1}} = \cosh \eta_\nu \quad \text{and} \quad \frac{1}{\sqrt{\ell_\nu^2 - 1}} = \sinh \eta_\nu \,, \qquad \text{(IV.40)}$$

we get

$$c_\nu = E \coth (2\nu\theta + \eta_\nu) \qquad \text{(IV.41a)}$$

$$b_\nu' = iE \operatorname{cosech} (2\nu\theta + \eta_\nu) \,. \qquad \text{(IV.41b)}$$

Case (iii): c_ν' real, $\ell_\nu^2 > 1$.
Define η_ν through (IV.40). Then we obtain

$$c_\nu = E \coth (2\nu\theta + \eta_\nu) \qquad \text{(IV.42a)}$$

$$c_\nu' = E \operatorname{cosech} (2\nu\theta + \eta_\nu) \,. \qquad \text{(IV.42b)}$$

Case (iv): c_ν' imaginary, $\ell_\nu^2 < 1$.
Defining η_ν through (IV.38), we find

$$c_\nu = E \tanh (2\nu\theta + \eta_\nu) \qquad \text{(IV.43a)}$$

$$c_\nu' = iE \operatorname{sech} (2\nu\theta + \eta_\nu) \,. \qquad \text{(IV.43b)}$$

Substitution of these expressions for the coefficients in H into Eqs. (IV.17) and (IV.23) finally yields four classes of Hamiltonians, each class being characterized by specific commutation relations of P with T and C, and containing an infinite number of members by virtue of the completely arbitrary parameters η_ν which occur in them. If we applied the boost condition (IV.30) in full instead of its "longitudinal" component (IV.31) alone, cases (ii) and (iv) would be eliminated entirely, and in the other two cases only $\eta_\nu = 0$ would be permitted. The proof of this statement is quite complicated and will not be given here. We shall however continue to work with the infinite variety of Hamiltonians determined above, in order to demonstrate all the better the power of the constraints imposed by the requirement of quantizability.

V. The Invariant Scalar Product, and Orthonormality of Solutions of the Wave Equation

Let us now turn to the question of the invariant scalar product, which is such a vital ingredient of relativistic quantum mechanics. We have already noted that the scalar product must be of the form (III.35), with a Hermitian metric M satisfying (III.37), to be determined. We will assume that the scalar product is invariant under space-inversion (MP = PM). Its behaviour under T and C will have to be considered separately.

As we have noted earlier, the generators $\underset{\sim}{P}$ and $\underset{\sim}{J}$, Eqs. (III.30b) and (III.30c), are Hermitian, and when introduced in Eq. (III.37), lead to $M\underset{\sim}{P} = \underset{\sim}{P}M$ and $M\underset{\sim}{J} = \underset{\sim}{J}M$. These equations, together with MP = PM (space-inversion invariance), state that M should be a true scalar operator, independent of space coordinates (and of time too, of course). As in the case of the Hamiltonian, this means that M must be a polynomial in $\underset{\sim}{\lambda}\cdot\underset{\sim}{p}$, plus ρ_1 times another such polynomial, with the additional restriction that this whole expression be Hermitian. The boost invariance condition $MK = K^{\dagger}M$ becomes

$$M(t\underset{\sim}{p} + \underset{\sim}{x}H + i\underset{\sim}{\lambda}) = (t\underset{\sim}{p} + H\underset{\sim}{x} - i\underset{\sim}{\lambda})M, \qquad (V.1)$$

and the consequence of this requirement can be worked out, after taking a scalar product of (V.1) with $\underset{\sim}{p}$, exactly as we determined the consequences of (IV.30) for H. We get again a set of first order differential equations for the coefficients in the expansion of M in terms of the projection operators B_ν and C_ν. Solution of these equations leads, in each of our four cases, to a linear combination (with arbitrary constant coefficients) of two parts, one of which is invariant under charge conjugation, while the other changes sign. Clearly, each part by itself is admissible as a metric operator, the first of these being positive definite and the other indefinite. We denote these by M_1 and M_2 respectively, and their expressions in the four cases are given* in Sec. IV. It may be verified in each case that

$$M_2 = M_1(H/E) \quad . \qquad (V.2)$$

*These forms presuppose a convenient choice of arbitrary constants arising in the integration of the differential equations. Unlike in the case of H, the constants here appear as factors multiplying the terms corresponding to different ν's in the expressions for M, and are therefore simply normalization constants which can be chosen to normalize the different solutions of our wave equation according to our convenience.

The overall factors of i in some of the M's are included to ensure Hermiticity, and the factors of m^{-1} make the metric operator dimensionless.

We have thus a choice of two metric operators for each of the Hamiltonians, and we have no a priori reason to choose one in preference to the other. The search for the invariant scalar product has only added to the freedom which existed in the choice of the Hamiltonian itself! However that may be, when a scalar product is defined with a particular metric, all observables have to be Hermitian with respect to the metric, i.e., if A is an operator representing an observable, A must satisfy

$$MA = A^{\dagger}M \tag{V.3}$$

in order that its expectation value,

$$\langle A\rangle = (\varphi, A\varphi) = \int \varphi^{\dagger} M A \varphi \, d^3x \tag{V.4}$$

be real. It is immaterial whether or not A is Hermitian in the ordinary sense. As far as observables like energy, related to the generators of the Poincarè group are concerned, this property is already ensured by the invariance condition (III.37). It is interesting to note that the coordinate variable $\underset{\sim}{x}$ does not satisfy this requirement,† and is therefore not an observable.

Let us consider now the plane wave solutions of our wave equation, and their normalization with respect to the metric M. Observe first of all that H is expressed completely in terms of the operator $\lambda_p \equiv \rho_3 (\underset{\sim}{s}\cdot\underset{\sim}{p}/p)$ which commutes with the helicity $(\underset{\sim}{s}\cdot\underset{\sim}{p}/p)$. Thus plane wave solutions of H can be labelled by the value of the momentum‡ $\underset{\sim}{q}$, sign of the energy ϵ, and helicity h. Let $\psi^{\epsilon}_{\underset{\sim}{q}h}(\underset{\sim}{x},t)$ be such a solution. Clearly,

$$\psi^{\epsilon}_{\underset{\sim}{q}h}(\underset{\sim}{x},t) = u^{\epsilon}_{\underset{\sim}{q}h}\cdot(2\pi)^{-\frac{3}{2}} \exp[i(\underset{\sim}{q}\cdot\underset{\sim}{x} - \epsilon\omega t)], \tag{V.5a}$$

---

†The only exception is the case of the Dirac particle. In the general case one can define an "observable position" $\underset{\sim}{X} = R\,\underset{\sim}{x}\,R^{-1}$. This transformation is really a generalised Foldy-Wouthuysen transformations. Ref. 9 gives details regarding this for the case of half-integral spins.

‡We use $\underset{\sim}{q}$ and $\omega = +(\underset{\sim}{q}^2 + m^2)^{\frac{1}{2}}$ for momentum and magnitude of the energy, to emphasize that they are numerical values rather than operators $\underset{\sim}{p} = -i\underset{\sim}{\nabla}$ and $E = (\underset{\sim}{p}^2 + m^2)^{\frac{1}{2}}$.

$$\underset{\sim}{p}\,\psi^{\epsilon}_{\underset{\sim}{q}h} = \underset{\sim}{q}\,\psi^{\epsilon}_{\underset{\sim}{q}h};\quad H\,\psi^{\epsilon}_{\underset{\sim}{q}h} = \epsilon\omega\,\psi^{\epsilon}_{\underset{\sim}{q}h};\quad \frac{\underset{\sim}{S}\cdot\underset{\sim}{p}}{p}\,\psi^{\epsilon}_{\underset{\sim}{q}h} = h\,\psi^{\epsilon}_{\underset{\sim}{q}h} \tag{V.5b}$$

where $u^{\epsilon}_{\underset{\sim}{q}h}$ is a $(2s+1)$-component spinor. By considering the equality

$$\int \psi^{\epsilon\,\dagger}_{\underset{\sim}{q}h}(\underset{\sim}{x},t)\, MG\psi^{\epsilon'}_{\underset{\sim}{q}',h'}(\underset{\sim}{x},t)\, d^3x$$

$$= \int [G\,\psi^{\epsilon}_{\underset{\sim}{q}h}(\underset{\sim}{x},t)]^{\dagger}\, M\,\psi^{\epsilon'}_{\underset{\sim}{q}',h'}(\underset{\sim}{x},t)\, d^3x \quad , \tag{V.6}$$

wherein G is taken successively as $\underset{\sim}{p}$, H and $(\underset{\sim}{S}\cdot\underset{\sim}{p}/p)$, and the condition (III.37) is used, one can easily verify by virtue of (V.5b) that the plane wave solutions can be orthonormalized in the following way:

$$\int \psi^{\epsilon\,\dagger}_{\underset{\sim}{q}h}(\underset{\sim}{x},t)\, M\,\psi^{\epsilon'}_{\underset{\sim}{q}',h'}(\underset{\sim}{x},t)\, d^3x = \eta\,\delta_{\epsilon\epsilon'}\,\delta_{hh'}\,\delta(\underset{\sim}{q}-\underset{\sim}{q}') \quad . \tag{V.7}$$

The normalization is according to

$$\eta = 1 \quad \text{if} \quad M = M_1 \tag{V.8a}$$

$$\eta = \epsilon \quad \text{if} \quad M = M_2 \quad . \tag{V.8b}$$

The fact that in (V.8b), η is positive for positive energy states and negative for negative energy states reflects the indefinite character of the metric M_2. The orthonormality properties of the spinors $u^{\epsilon}_{\underset{\sim}{q}h}$ are obtained from (V.7) and (V.5a):

$$u^{\epsilon\,\dagger}_{\underset{\sim}{q}h}\, M(\underset{\sim}{q})\,\eta^{-1}\, u^{\epsilon'}_{\underset{\sim}{q}h'} = \delta_{\epsilon\epsilon'}\,\delta_{hh'} \quad . \tag{V.9}$$

This equation can be used to obtain an expression for a certain "sum over states" which we will encounter while carrying out the second quantization. Multiply (V.9) on the left by $u^{\epsilon}_{\underset{\sim}{q}h}$ and sum over h. We get

$$\left(\sum_h u^{\epsilon}_{\underset{\sim}{q}h}\, u^{\epsilon\,\dagger}_{\underset{\sim}{q}h}\, M(\underset{\sim}{q})\,\eta^{-1}\right) u^{\epsilon'}_{\underset{\sim}{q}h'} = \delta_{\epsilon\epsilon'}\, u^{\epsilon}_{\underset{\sim}{q}h'} \quad . \tag{V.10}$$

Thus the bracketed operator in (V.10) is a projection operator which leaves spinors corresponding to momentum $\underset{\sim}{q}$ and energy $\epsilon\omega$ unchanged while it annihilates states with the wrong sign of the energy. Therefore

$$\sum_h u^{\epsilon}_{\underline{q}h}\, u^{\epsilon\,\dagger}_{\underline{q}h}\, M(\underline{q})\, \eta^{-1} = \tfrac{1}{2}\left[1 + \frac{H(\underline{q})}{\epsilon\omega}\right] \quad . \tag{V.11}$$

With the use of (V.5a), this equation leads to the following identity involving the space-time dependent wave functions $\psi^{\epsilon}_{\underline{q}h}$:

$$S^{\epsilon} \equiv \sum_h \int d^3q\, \psi^{\epsilon}_{\underline{q}h}(\underline{x},t)\, \psi^{\epsilon\,\dagger}_{\underline{q}h}(\underline{y},\tau)$$

$$= (2\pi)^{-3} \int d^3q\, \eta \cdot \tfrac{1}{2}\left[1 + \frac{H(\underline{q})}{\epsilon\omega}\right] M^{-1}(\underline{q})\ \exp[i\underline{q}\cdot(\underline{x}-\underline{y}) - i\epsilon\omega(t-\tau)]$$

$$= \frac{-i\epsilon}{(2\pi)^3} \int \frac{d^3q}{2\omega}\left[i\eta\{\epsilon\omega + H(\underline{q})\}\, M^{-1}(\underline{q})\right] \exp[i\underline{q}\cdot(\underline{x}-\underline{y}) - i\epsilon\omega(t-\tau)] \tag{V.12}$$

It should be noted that the right hand side of (V.12), like the left hand side, is actually independent of which of the two metric operators M_1 or M_2 is used, because $\eta M^{-1}(\underline{q}) = M_1{}^{-1}(\underline{q}) = \epsilon M_2{}^{-1}(\underline{q})$. We will need to make explicit use of this expression in the next section.

VI. Second Quantization

We have now carried the c-number theory of arbitrary spin fields as far as is possible without making extra assumptions beyond invariance under the Poincaré group and the discrete transformations.* The next question is whether the theory is quantizable, and if so, what new restrictions or new features emerge from the quantization procedure. We shall attempt to carry out the second quantization by expanding the field ψ in terms of the plane wave solutions $\psi^{\epsilon}_{\underline{q}h}$ of the wave equation and subjecting the expansion coefficients to quantum conditions.[12),4)] Thus, we write

$$\psi(\underline{x},t) = \sum_h \int d^3q \left\{\psi^{+}_{\underline{q}h}(\underline{x},t)\, a(\underline{q},h) + \psi^{-}_{-\underline{q},h}(\underline{x},t)\, b^{*}(\underline{q},h)\right\} \tag{VI.1}$$

and suppose that

$$[a(\underline{q},h),\ a^{*}(\underline{q}',h')]_{\pm} = [b(\underline{q},h),\ b^{*}(\underline{q}',h')]_{\pm}$$

$$= \delta_{hh'}\, \delta(\underline{q}-\underline{q}') \ , \tag{VI.2}$$

with all other anticommutators/commutators vanishing. (The plus/minus sign on the brackets in (VI.2) indicate anticommutators/commutators, corresponding to Fermi/Bose statistics. We do not consider

*It may be recalled, however (last paragraph of Sec. IV), that invariance under transverse boosts has been held in abeyance.

the possibility of having more general parastatistics.) On top of all the ambiguities we have noted so far, we now have one more: whether to use the + or - sign in (VI.2) in quantizing the theory for a specific spin. So matters seem to be going from bad to worse! However this trend gets reversed dramatically if we impose one very natural condition, that of microcausality, which we take in the form

$$[\psi_\alpha(\underline{x},t),\ \psi^*_\beta(\underline{y},\tau)]_\pm = 0 \qquad \text{(VI.3)}$$

for space-like separations between the points $(\underline{x},t)$ and $(\underline{y},\tau)$, i.e., for $(t-\tau)^2 - (\underline{x}-\underline{y})^2 < 0$. The condition (VI.3) would ensure commutation at space-like separations, of certain bilinear functions of the field which represent supposedly observable densities (like energy and momentum densities) if such densities are local functions of the field; this will be verified to be the case for the fields we consider.

We shall show:

(a) that microcausality cannot be achieved in any theory using Hamiltonians of cases (ii) and (iv);

(b) that case (i) is consistent with microcausality if and only if the spin is half-integral, the quantization is carried out using Fermi statistics (i.e. plus sign in (VI.2)), and all the parameters η_ν in H, entering through the coefficients c_ν and $b_\nu{}'$ as determined in (VI.39), vanish; and

(c) that case (iii) is consistent with microcausality if and only if the spin is integral, the minus sign in (VI.2), (Bose statistics), is employed, and all the parameters η_ν in H, entering through the coefficients c_ν, $c_\nu{}'$ as given in (VI.42), vanish.

To substantiate the above statements, which testify to the truly astonishing power of the microcausality condition, let us evaluate the anticommutator/commutator in (VI.3) using (VI.1) and (VI.2). It is a trivial matter to see that

$$\left[\psi_\alpha(\underline{x},t),\ \psi^*_\beta(\underline{y},\tau)\right]_\pm = \sum_h \int d^3q\ \psi^+_{\underline{q}h,\alpha}(\underline{x},t)\ \psi^{+*}_{\underline{q}h,\beta}(\underline{y},\tau)$$

$$\pm \sum_h \int d^3q\ \psi^-_{\underline{q}h,\alpha}(\underline{x},t)\ \psi^{-*}_{\underline{q}h,\beta}(\underline{y},\tau)$$

$$\equiv S^+ \pm S^- \qquad \text{(VI.4)}$$

where the explicit form of the sums S^+ and S^- has been already determined in (V.12). This form suggests that we try to express $S^\epsilon(\underline{x},t)$ in terms of the familiar Lorentz-invariant functions

$$\Delta^{\epsilon}(\underline{x},t) = \frac{-i\epsilon}{(2\pi)^3}\int \frac{d^3\underline{q}}{2\omega}\; e^{i\epsilon(\underline{q}\cdot\underline{x} - \omega t)}\ , \qquad \text{(VI.5)}$$

since it is known that

$$\Delta = \Delta^{+} + \Delta^{-} \qquad \text{(VI.6)}$$

has the causality property we are looking for (i.e. vanishing outside the light cone). For, we can then hope to pull out the factor

$$[\,i\eta\,\{\epsilon\omega + H(\underline{q})\}\; M^{-1}(\underline{q})\,] \qquad \text{(VI.7)}$$

from the integrand in S^{ϵ}, Eq. (V.12), by replacing $\underline{q}$ by the differential operator $\underline{p} = -i\underline{\nabla}$ and $\epsilon\omega$ by $i\frac{\partial}{\partial t} = -p_0$, and thus express S^{ϵ} as Δ^{ϵ} acted on by an operator which is a function of $\underline{p}$ and p_0. To see what kind of an operator emerges out of this procedure, we must first make the q and ω dependence of the hyperbolic functions contained in M^{-1} and H in (VI.7) manifest by expanding these functions in powers of $\cosh\theta = (\omega/m)$ and $\sinh\theta = (q/m)$. When this is done, if (VI.7) turns out to be of the form $f(\underline{q},\epsilon\omega)$, without any ε-dependence other than through the combination εω, then its role in place of the square-bracketed factor in (V.12) is to make S^{ϵ} equal to $f(\underline{p},-p_0)\Delta^{\epsilon}$, so that the sum $S^{+} + S^{-} = f(\underline{p},-p_0)\Delta$ would be causal, provided the operator f is local (i.e. is a polynomial in the differential operators $\underline{p}$, p_0). In this case, quantization must be done according to Fermion commutation rules to ensure causality. On the other hand if (VI.7) reduces to the form $\epsilon f(\underline{q},\epsilon\omega)$ then it is the difference $S^{+} - S^{-} = f(\underline{p},-p_0)\Delta$ which is local, and Boson commutation rules would then be required to ensure causality. If neither of these happens, then one cannot get causality with either type of commutation rule, and such a case would have to be considered unphysical.

Let us now substitute in (VI.7) each of the four types of Hamiltonians with the associated metric operators, and see what happens.

Case (i). Introducing the expressions (IV.7) and (IV.9) for H and M, with (IV.39), into (VI.7), and multiplying out with the help of (IV.12), we find that (VI.7) reduces to

$$\mathrm{im}\Big[\sum \epsilon\cosh(2\nu\theta + \eta_\nu)\cdot B_\nu + \sum \sinh(2\nu\theta + \eta_\nu)\cdot C_\nu + \rho_1\Big]\,. \qquad \text{(VI.8)}$$

Now use the expansions

$$\sinh 2\nu\theta = \sum_{n\ \mathrm{odd}} \alpha_n \sinh^n\theta = \sum_{n\ \mathrm{odd}} \alpha_n (q/m)^n \qquad \text{(VI.9a)}$$

$$\cosh 2\nu\theta = \sum_{n \text{ odd}} \beta_n \cosh^n \theta = \sum_{n \text{ odd}} \beta_n (\omega/m)^n \qquad \text{(VI.9b)}$$

for 2ν odd (half-integer spin), and

$$\sinh 2\nu\theta = \sum_{n \text{ odd}} \alpha_n' \sinh\theta \cosh^n\theta = \sum_{n \text{ odd}} \alpha_n' (q/m)(\omega/m)^n , \qquad \text{(VI.10a)}$$

$$\cosh 2\nu\theta = \sum_{n \text{ even}} \beta_n' \cosh^n \theta = \sum_{n \text{ even}} \beta_n' (\omega/m)^n \qquad \text{(VI.10b)}$$

for 2ν even (integer spin), where α_n, β_n, α_n', β_n' are constants. Inspection of Eqs. (VI.8)-(VI.10) shows that if and only if all η_ν were set equal to zero and 2ν is odd (i.e. the spin is half-integral), (VI.8) reduces to the form $f(\underline{q}, \epsilon\omega)$. Under these conditions, the anticommutator $[\psi_\alpha, \psi_\beta^*]_+$ would reduce to $f(\underline{p}, -p_o)\Delta$, which is causal because $f(\underline{p}, -p_o)$ is a local operator. The locality property is verified by identifying (VI.8), subject to $\eta_\nu = 0$, as

$$f(q, \epsilon\omega) = \mathrm{im}\left[\epsilon\omega \cdot \omega^{-1} \cosh 2\lambda_q\theta + \sinh 2\lambda_q\theta + \rho_1 \right] \qquad \text{(VI.11)}$$

with the aid of (IV.10) and (IV.12) as applied to the hyperbolic functions of matrices in this last expression, and checking that $\omega^{-1}\cosh 2\lambda_q\theta$ and $\sinh 2\lambda_q\theta$ are polynomials in $\underline{\lambda}\cdot\underline{q}$. For this last step, explicit expansions given in the Appendix A to Ref. 13 may be employed. The assertion (b) on page 172 is thus proved. The proof of assertion (c) regarding case (iii) follows in an entirely analogous fashion, the only difference being that when the H and M for this case, Eqs. (IV.23), (IV.26) and (IV.42), are introduced into (VI.7), it reduces (for $\eta_\nu = 0$) to the form $\epsilon f(\underline{q}, \epsilon\omega)$, leading to Bose statistics since it is the commutator $[\psi_\alpha, \psi_\beta^*]_- = S^+ - S^-$ that is now causal. In cases (ii) and (iv), one of the terms in (VI.7) is $\sum C_\nu$ (apart from irrelevant factors) and it is easy to check that it contains inverse powers of q which make it impossible to arrive at a <u>local</u> operator $f(\underline{p}, -p_o)$. Therefore we cannot have microcausality in these cases, confirming the assertion (a) on page 172.

Thus we see that quantization consistent with microcausality leads to a theory with the standard spin-statistics connection which is unique for any given spin--except in one respect. We still do not know whether there is anything to choose between the two possible metrics M_1 and M_2 in each case. Even this remaining ambiguity disappears[4] when the familiar role of the total energy and momentum

operators of the quantized field as translation generators on the field is taken seriously, requiring for instance that

$$[\mathcal{H}, \psi] = -i\frac{\partial \psi}{\partial t} \tag{VI.12}$$

where the field Hamiltonian $\mathcal{H}$ is obviously to be defined as

$$\mathcal{H} = \int \psi^{\dagger} M H \psi \, d^3x \quad . \tag{VI.13}$$

Evaluation of (VI.13) using the field expansion (VI.1) and the properties (V.5b), (V.7) and (V.8) of the plane waves $\psi^{e}_{\underline{g}h}$ leads to

$$\mathcal{H} = \sum_{h} \int d^3q \, \omega \, [a^{*}(\underline{q},h)\, a(\underline{q},h) \mp b(\underline{q},h)\, b^{*}(\underline{q},h)] \tag{VI.14}$$

where the upper (minus) sign in (VI.14) results if the choice $M = M_1$ of the metric is made, and the lower (plus) sign if we take $M = M_2$. It is a simple matter to verify now that with this expression for H, (VI.12) is satisfied only if

(a) the minus sign in (VI.14) is used (i.e. $M = M_1$, positive definite charge density) with fermion commutation rules, and

(b) the plus sign (i.e. $M = M_2$, indefinite charge density) is used with boson commutation rules.

A remarkable aspect of this result is that the connection established is between the positive definiteness or indefiniteness of the charge density and the statistics of the particles (instead of the spin of the particles as in Pauli's classical proof).

Finally, it should be observed that in the cases where we have found quantization consistent with microcausality to be possible, the quantities like charge, momentum and energy are expressible as integrals of local functions of the field (provided time derivatives are admitted), even though we did not introduce locality as a separate requirement. To see this, consider first the total charge or number operator, given by $\int \psi^{\dagger} M \psi d^3x$. Here M is to be taken as M_1 for fermion fields and M_2 for boson fields, as indicated above. In case (i), which we showed to be appropriate for half-integral spin fermion fields,

$$H = E\sum \tanh 2\nu\theta \cdot C_{\nu} + \rho_1 E \sum \operatorname{sech} 2\nu\theta \cdot B_{\nu}$$

$$= E(\tanh 2\lambda_p\theta + \rho_1 \operatorname{sech} 2\lambda_p\theta) \tag{VI.15}$$

and

$$M_1 = (E/m) \sum \text{sech } 2\nu\theta \cdot B_\nu = (E/m) \text{ sech } 2\lambda_p \theta , \qquad \text{(VI.16)}$$

in view of (IV.17), (IV.19), (IV.39) and the fact that $\eta_\nu = 0$, proved in the last section. It follows from (VI.15) and (VI.16) that

$$M_1 = \frac{1}{2m} (\rho_1 H + H \rho_1) \quad . \qquad \text{(VI.17)}$$

We therefore have

$$\int \psi^\dagger M_1 \psi d^3x = \frac{1}{2m} \int \left[\psi^\dagger \rho_1 \, i \frac{\partial \psi}{\partial t} - i \frac{\partial \psi^\dagger}{\partial t} \rho_1 \psi \right] d^3x \qquad \text{(VI.18)}$$

on using the Hermiticity of H and the wave equation. In case (iii), appropriate for integral spin boson fields, one has to take the metric to be M_2. It is found that

$$M_2 = \frac{1}{2m} (\rho_1 H + H^\dagger \rho_1) \quad . \qquad \text{(VI.19)}$$

(Note that H is not Hermitian in this case, though $MH = H^\dagger M$.) One again obtains

$$\int \psi^\dagger M_2 \psi d^3x = \frac{1}{2m} \int \left[\psi^\dagger \rho_1 \, i \frac{\partial \psi}{\partial t} - i \frac{\partial \psi^\dagger}{\partial t} \rho_1 \psi \right] d^3x \quad . \qquad \text{(VI.20)}$$

It is trivial to verify with the aid of (VI.18) and (VI.20) that momentum and energy densities are also local functions of ψ.

VII. Discussion

The theory we have presented above is very different in appearance from theories of manifestly covariant wave equations, many different forms of which have been proposed and studied extensively for over thirty years from the field theoretic, matrix algebraic and group theoretic points of view, and discussed at considerable length, for instance in the books by Corson and Umezawa.[1] The arbitrary spin equations due to Dirac,[14] Fierz and Pauli,[15] Bargmann and Wigner[16] and many others, are expressible in the form

$$(i\beta^\mu \partial_\mu + m)\psi = 0 \qquad \text{(VII.1)}$$

which has been investigated with great generality by Bhabha[17] and Harish-Chandra.[18] All equations of this form, with the sole exception of the Dirac equation for spin $\frac{1}{2}$, involve wave functions with a nonunique spin content. The reason, as has been briefly mentioned already, is that a set of matrices β^μ transforming like a vector, which

is equivalent to $D(\frac{1}{2},\frac{1}{2})$, cannot be defined over wave functions transforming irreducibly, say as $D(m,n)$, under $\mathfrak{L}$. $D(\frac{1}{2},\frac{1}{2})$ must necessarily connect $D(m,n)$ with one or more of the representations $D(m \pm \frac{1}{2}, n \pm \frac{1}{2})$, and it is only with $D(\frac{1}{2},0) + D(0,\frac{1}{2})$ that one can meet this requirement and yet maintain a unique spin value, $\frac{1}{2}$. In all other cases the equation must describe a multi-spin (and multi-mass) system such as the case discussed by Bhabha,[17] or else one must make β^0 in (VII.1) singular in such a way that time derivatives of only those components of ψ which belong to the correct spin appear with nonzero coefficients (i.e. have equations of motion), the other components being then merely subject to supplementary conditions (free of time derivatives) which serve to eliminate them. The conditions necessary for ensuring that particles described by (VII.1) have unique mass and spin have been found under very general conditions by Harish-Chandra.[18] A clear and succinct account of these, with explicit construction of the matrices β^μ for spins $\frac{1}{2}$, $\frac{3}{2}$, $\frac{5}{2}$, is given in a recent paper by Capri.[19]

While the theory of manifestly covariant wave equations is extremely elegant, the presence of supplementary conditions creates difficulties, notably when the introduction of interactions is attempted.[20] More pertinent from our point of view is the fact that invariance under proper transformations and under all or at least some of the discrete transformations come packaged together and cannot be separated if covariance is to be manifest. This can be seen already from the case of the Dirac equation (spin $\frac{1}{2}$) which is actually invariant under T, C and P though none of these was asked for while deriving the equation. Therefore if the effects of various invariance conditions are to be separately analysed, the assumption of manifest covariance must be given up.

The first attempt to treat this question systematically was in a very interesting paper by Foldy,[10] who used a "canonical" form of the wave equation, which is not manifestly covariant and involves a wave function which does not possess local covariance. Foldy drew attention to the existence of ambiguities associated with various possible commutation rules for discrete symmetry operators, but did not carry the analysis much further. Consideration of this question within the very restricted freedom allowed by manifest covariance is also contained in a paper by Pursey[21] which deals with a general classification of relativistic wave equations through an anlysis of the relations that must exist between the different Lorentz-irreducible parts $D(m,n)$ of a wave function when a unique spin and mass content is imposed by expressing each of them in terms of the relevant unitary irreducible representation of the Poincarè group. In both the above papers, only the c-number theory is considered. Equations without manifest covariance, of the type corresponding to our case (i) with

$\eta_\nu = 0$, were first obtained (with explicit determination of H for a few low-spin values) by Weaver, Hammer and Good,[22] by starting from a special form assumed for the Hamiltonian of a particle at rest. Actually none of our general classes of Hamiltonians except this particular type, has a definite limit as $\underline{p} \to 0$. And of course the Hamiltonians H in all higher-spin cases are nonlocal operators in configuration space. Despite these apparently disturbing features, the second-quantizable theories do have observables expressed locally in terms of the field functions, as we have shown in the last section. The wave functions in these particular cases coincide, as is to be expected, with those obtained by Weinberg,[13] who takes ψ to be a quantized field from the beginning. There are interesting questions which remain, and are under investigation: for example whether, if one or more of the discrete symmetries are "broken," one would get more general types of quantizable theories. For the present we will conclude by noting that our wave equation (invariant under T, C and P separately), in the quantizable cases determined in the last section, is really the essence of all manifestly covariant wave equations: for this equation is what would result if, starting from any manifestly covariant equation which involves various irreducible representations D(m,n), all except the D(0,s) and D(s,0) parts were expressed in terms of these two by making use of the supplementary conditions.

Appendix: Implementation of the Boost-Invariance Condition

To illustrate how the coefficients c_ν and b_ν' or c_ν' in H are determined by the boost-invariance condition in the form (IV.31), we exhibit here the details of the calculation in case (i). It is convenient to use for this purpose two different ways of writing H:

$$H = \sum c_\nu C_\nu = \rho_1 \sum b_\nu' B_\nu \tag{A.1}$$

or

$$H = \sum_{\ell \text{ odd}} f_\ell (\underline{\lambda}\cdot\underline{p})^\ell + \rho_1 \sum_{\ell \text{ even}} g_\ell (\underline{\lambda}\cdot\underline{p})^\ell , \tag{A.2}$$

where ℓ ranges from 0 to 2s and the f_ℓ are nonzero only for odd values of ℓ and g_ℓ for even values. The relation between the coefficients in the above two equations can be easily obtained from the spectral representation of $(\underline{\lambda}\cdot\underline{p})$:

$$(\lambda_p)^\ell = (\underline{\lambda}\cdot\underline{p}/p)^\ell = \sum_{\nu=-s}^{s} \nu^\ell \Lambda_\nu \tag{A.3}$$

so that

$$(\underset{\sim}{\lambda}\cdot\underset{\sim}{p})^{\ell} = \sum_{\nu=-s}^{s} (\nu p)^{\ell} \Lambda_{\nu} = \sum_{\nu\geq 0} (\nu p)^{\ell} B_{\nu} \qquad (\ell \text{ even}) \tag{A.4a}$$

$$= \sum_{\nu\geq 0} (\nu p)^{\ell} C_{\nu} \qquad (\ell \text{ odd}) \tag{A.4b}$$

Introducing (A.4) into (A.2) and comparing with (A.1), we find that

$$c_{\nu} = \sum_{\ell} (\nu p)^{\ell} f_{\ell}, \quad b_{\nu}' = \sum_{\ell} (\nu p)^{\ell} g_{\ell} . \tag{A.5}$$

Now, using the fact that $-i[\underset{\sim}{x},H]$ is just the gradient of H in $\underset{\sim}{p}$-space, it is easy to verify that

$$-i\underset{\sim}{p}\cdot[\underset{\sim}{x},H] = \sum_{\ell} \left(\frac{df_{\ell}}{dp} p + f_{\ell}\, \ell\right) (\underset{\sim}{\lambda}\cdot\underset{\sim}{p})^{\ell}$$

$$+ \rho_1 \sum_{\ell} \left(\frac{dg_{\ell}}{dp} p + g_{\ell}\, \ell\right) (\underset{\sim}{\lambda}\cdot\underset{\sim}{p})^{\ell} . \tag{A.6}$$

By using (A.4) and (A.5) we can reduce this to

$$-i\underset{\sim}{p}\cdot[\underset{\sim}{x},H] = \sum p \frac{dc_{\nu}}{dp} \cdot C_{\nu} + \rho_1 \sum p \frac{db_{\nu}'}{dp} \cdot B_{\nu} . \tag{A.7}$$

Substituting this in the left hand side of (A.1) and using the representation (A.4b) for $\underset{\sim}{\lambda}\cdot\underset{\sim}{p}$ on the right hand side, we obtain, after evaluating the commutators,

$$\frac{dc_{\nu}}{dp} b_{\nu}' - c_{\nu} \frac{db_{\nu}'}{dp} = 2\nu\, b_{\nu}' . \tag{A.8}$$

But we already have the relation

$$c_{\nu}^{2} + b_{\nu}'^{2} = E^{2} \tag{A.9}$$

from which it follows that

$$c_\nu \frac{dc_\nu}{dp} + b_\nu' \frac{db_\nu'}{dp} = p \ . \qquad (A.10)$$

Solution of the simultaneous equations (A.8) and (A.10) for the derivatives of c_ν and b_ν' yields

$$\frac{dc_\nu}{dp} = \frac{c_\nu p + 2\nu b_\nu'^2}{E^2} = \frac{c_\nu p + 2\nu(E^2 - c_\nu^2)}{E^2} \ , \qquad (A.11)$$

$$\frac{db_\nu'}{dp} = \frac{-2\nu c_\nu b_\nu' + p b_\nu'}{E^2} \ . \qquad (A.12)$$

From (A.11) we can obtain a simple equation for (c_ν/E), which can be immediately solved. With the use of the initial condition (IV.33) one gets the solution to be (IV.34), and then b_ν' (IV.35) is deduced using (A.9).

It is a striking characteristic of the coefficients c_ν, b_ν', c_ν', Eqs. (IV.34), (IV.35), (IV.36), that they are independent of the spin of the particle.

References

1. A fairly exhaustive treatment of manifestly covariant wave equations may be found in:
   E. M. Corson, Introduction to Tensors, Spinors and Relativistic Wave Equations (Hafner Publishing Co., New York, 1953), and H. Umezawa, Quantum Field Theory (North-Holland, Amsterdam, 1956). See also Y. Takahashi in Lectures in Theoretical Physics, Vol. XA, ed. A. O. Barut and W. E. Brittin (Gordon and Breach, New York, 1968). This volume contains also a presentation of a group-theoretic approach to quantum fields by P. T. Matthews.
2. P. M. Mathews, Phys. Rev. 143, 978 (1966).
3. P. M. Mathews and S. Ramakrishnan, Nuovo Cimento 50, 339 (1967).
4. P. M. Mathews, J. Math. Phys. Sci. (Madras) 1, 197 (1967).
5. E. P. Wigner, Ann. of Math. 40, 149 (1939).
6. See, for example, U. Fano and G. Racah, Irreducible Tensorial Sets (Academic Press Inc., New York, 1959), Appendix C.
7. See, for example, E. M. Corson, Ref. 1.
8. See, for example, M. A. Naimark, Linear Representations of the Lorentz Group (McMillan, New York, 1964).
9. P. M. Mathews, Phys. Rev. 143, 985 (1966).
10. L. L. Foldy, Phys. Rev. 102, 568 (1956).
11. N. Kemmer, Proc. Roy. Soc. (London) A 173, 91 (1939).

12. P. M. Mathews, Phys. Rev. 155, 1415 (1967).
13. S. Weinberg, Phys. Rev. 133, B1318 (1964).
14. P. A. M. Dirac, Proc. Roy. Soc. (London) A155, 447 (1936).
15. M. Fierz, Helv. Phys. Acta 12, 3 (1939); M. Fierz and W. Pauli, Helv. Phys. Acta 12, 297 (1939) and Proc. Roy. Soc. (London) A 173, 211 (1939).
16. V. Bargmann and E. P. Wigner, Proc. Natl. Acad. Sci. U. S. 34, 211 (1948).
17. H. J. Bhabha, Revs. Mod. Phys. 17, 200 (1945); 21, 451 (1949).
18. Harish-Chandra, Phys. Rev. 71, 793 (1947); Proc. Roy. Soc. (London) 192 A, 195 (1947).
19. A. Z. Capri, Phys. Rev. 178, 2427 (1969).
20. K. Johnson and E. C. G. Sudarshan, Ann. of Phys. (New York) 13, 126 (1961).
21. D. L. Pursey, Ann. of Phys. (New York) 32, 157 (1965).
22. D. L. Weaver, C. L. Hammer and R. H. Good, Jr., Phys. Rev. 135, B241 (1964).

ON THE THREE-DIMENSIONAL FORMULATION OF THE RELATIVISTIC TWO-BODY PROBLEM†

I. T. Todorov‡
Institute for Advanced Study
Princeton, New Jersey

Table of Contents

---

†Presented at the INSTITUTE FOR THEORETICAL PHYSICS, University of Colorado, Summer 1969.

‡On leave from Joint Institute for Nuclear Research, Dubna, USSR, and from Physical Institute of the Bulgarian Academy of Sciences, Sofia, Bulgaria.

Summary

The main purpose of the present notes is to discuss in some detail a recent three-dimensional "quasipotential" approach to the relativistic two-body problem developed by C. Itzykson, V. Kadyshevsky and the author. The quasipotential equation is derived from conventional local Hamiltonian formalism on the basis of Kadyshevsky's diagram technique for the "non-covariant" perturbation expansion. The quasipotential is defined as a series of irreducible graphs, analogous to the series for the kernel of the Bethe-Salpeter equation. The whole approach is related to the 4-dimensional B.-S. formalism as the old fashioned off-energy-shell perturbation theory is related to the off-mass-shell covariant technique of Feynman and Dyson. The non-uniqueness of the off-energy-shell extrapolation of the scattering amplitude is used to obtain a simpler "local" version of the quasipotential equation. It is shown that for the scalar Coulomb potential $V(p,q) = \frac{\alpha}{(p-q)^2}$ the latter equation is equivalent to a simple infinite-component wave equation (similar, but not identical, to the equations considered by Nambu, Fronsdal and Barut). The energy eigenvalues are calculated in this case and are found to be SO(4)-degenerate (just as in the non-relativistic Coulomb problem). In contrast with the Wick-Cutkosky model they do not depend on any additional quantum member (other than the principal quantum number n).

As an introduction the Lippmann-Schwinger and the Bethe-Salpeter equations are briefly reviewed.

I. Introduction: Linear Off-Shell Equations for the Scattering Amplitude

Two methods have been applied for the determination of the relativistic elastic scattering amplitude:

(i) Linear off-mass-shell or off-energy-shell equations (the Bethe-Salpeter and the quasipotential equation).

(ii) Nonlinear on-shell equations based on unitarity and dispersion relations.

The second method has never been exactly formulated. All closed sets of equations treated up to now make, for example, the two-particle approximation in the unitarity condition. The complete formulation is expected to involve an infinite set of equations for the multiparticle scattering amplitudes. However, the analytic structure of the multi-particle amplitudes is far from being understood.

We shall deal in these lectures with the first approach only. It is not uniquely defined because of the non-uniqueness of the off-shell extrapolation of the scattering amplitude. We shall start with a brief

review of non-relativistic Lippmann-Schwinger equation and relativistic Bethe-Salpeter equation.

A. Schrödinger Wave-Function and Non-Relativistic Off-Energy-Shell Scattering Amplitude[1),2)]

The non-relativistic scattering amplitude has been first defined by the asymptotic expression of the Schrödinger wave function satisfying certain boundary conditions. Thus, its off-shell extrapolation is dictated by the Schrödinger equation. One simplifying feature of the non-relativistic two-body problem is that the center-of-mass motion can be trivially extracted reducing thus the problem to the study of scattering and bound states of a single particle in an external field. In the time-independent formulation the scattering of an (effective) particle of reduced mass m_R, initial momentum $\underline{q}$, and energy E is determined by the Schrödinger equation

$$\left(-\frac{1}{2m_R}\Delta + v(\underline{x}) - E\right)\psi_E(\underline{q},\underline{x}) = 0 \qquad (I.1)$$

with the boundary condition

$$\psi_E(\underline{q},\underline{x}) = e^{i\underline{q}\underline{x}} + \varphi_E(\underline{x})$$

where

$$\lim_{r\to\infty}\varphi_E(\underline{x}) = 0, \quad \lim_{r\to\infty}\left[\frac{\partial}{\partial r}\varphi_E(\underline{x}) - i\sqrt{2m_R E}\,\varphi_E(x)\right] = 0, \quad r = |\underline{x}| \,. \qquad (I.2)$$

Eqs. (I.1) and (I.2) can be combined in a single integral equation

$$\psi_E(\underline{q},\underline{x}) = e^{i\underline{q}\underline{x}} + \int G_E(\underline{x}-\underline{y})v(\underline{y})\psi_E(\underline{q},\underline{y})d^3\underline{y} \qquad (I.3)$$

where the "Green function" G_E is given by

$$G_E(\underline{x}-\underline{y}) = -2m_R\frac{\exp\{i\sqrt{2m_R E}\,|\underline{x}-\underline{y}|\}}{4\pi|\underline{x}-\underline{y}|} = \frac{-2m_R}{(2\pi)^3}\int\frac{e^{-i\underline{q}(\underline{x}-\underline{y})}}{\underline{q}^2 - 2m_R E - i0}\,d^3\underline{q} \,. \qquad (I.4)$$

We define the off-shell scattering amplitude T as the Fourier transform of the product $-v\psi_E$:

$$T_E(\underline{p},\underline{q}) = -\frac{1}{4\pi}\int e^{-i\underline{p}\underline{x}}v(x)\psi_E(\underline{q},\underline{x})\,d^3\underline{x}$$

$$= -\frac{1}{4\pi}\int e^{-i\underline{p}\underline{x}}\left(\frac{1}{2m_R}\Delta + E\right)\psi_E(\underline{q},\underline{x})\,d^3x \qquad (I.5)$$

($\underline{p}$ is the final momentum of the reduced particle). Let the potential V in momentum space be

$$V(\underline{p}-\underline{q}) = \frac{1}{4\pi} \int v(\underline{x})\, e^{i(\underline{p}-\underline{q})\underline{x}}\, d^3\underline{x} \ . \tag{I.6}$$

[We have chosen the normalization in such a way that the Yukawa potential $\frac{1}{r} e^{-\mu r}$ goes into $(\mu^2+(\underline{p}-\underline{q})^2)^{-1}$.] Let us multiply both sides of Eq. (I.3) by $-\frac{1}{4\pi} v(\underline{x})$ and take the Fourier transform. Using (I.4)-(I.6) we obtain the Lippmann-Schwinger equation for the off-shell amplitude:

$$T_E(\underline{p},\underline{q}) + V(\underline{p}-\underline{q}) + \frac{m_R}{\pi^2} \int V(\underline{p}-\underline{k})\, \frac{T_E(\underline{k},\underline{q})}{k^2 - 2m_R E - i0}\, d^3\underline{k} = 0 \ . \tag{I.7}$$

The energy shell is defined by

$$\underline{p}^2 = \underline{q}^2 = 2m_R E \ . \tag{I.8}$$

[We note that on the surface (I.8) the denominator of the integrand in (I.7) vanishes.] It can be proved[1] that if $v(\underline{x})$ decreases at infinity fast enough so that

$$\int |v(\underline{x})|\, d^3x < \infty \tag{I.9}$$

then $\psi_E(\underline{q},\underline{x})$ has the following asymptotic behaviour:

$$\psi_E(\underline{q},\underline{x}) \underset{(r \to \infty)}{\approx} e^{i\underline{q}\underline{x}} + \frac{e^{iqr}}{r}\, \underline{T}_E(q\underline{n}, \underline{q}) \tag{I.10}$$

where $\underline{T}$ is the on-shell amplitude, $q = \sqrt{2m_R E}$, and the three-dimensional unit vector $\underline{n}$ is defined by $r\underline{n} = \underline{x}$. [Eq. (I.10) is often taken as a definition of the scattering amplitude.]

B. Bethe-Salpeter Equation[3)-7)]

Consider the problem of two relativistic interacting particles associated with the local Heisenberg fields $\underline{\psi}_1(x)$ and $\underline{\psi}_2(x)$. All information (both about the scattering and the bound states problem) is contained in the Green function $\tilde{G}$ defined as the vacuum expectation value of the time-ordered product of four fields:

$$\tilde{G}(x_1, x_2; y_1, y_2) = \langle 0 | T(\underline{\psi}_1(x_1)\underline{\psi}_2(x_2)\underline{\psi}_1^*(y_1)\underline{\psi}_2^*(y_2)) | 0 \rangle \ . \tag{I.11}$$

Let $\Delta'_{Fa}(p)$, $a = 1,2$, be the complete (or "modified") Feynman propagators (i.e., the two-particle Green functions in momentum space) and let $I(p_1, p_2; q_1, q_2)\delta(p_1+p_2-q_1-q_2)$ be the sum of all $\psi_1+\psi_2$-

irreducible Feynman graphs without radiative corrections on the external lines. [We say that the diagram D of the $\psi_1+\psi_2$-elastic scattering amplitude is $\psi_1+\psi_2$-reducible if it can be decomposed by cutting a ψ_1 and a ψ_2 lines in two graphs D′ and D″ of the same process such that D′ contains both incoming lines of D (with momenta q_1, q_2) and D″ contains both outgoing lines of D (with momenta p_1, p_2). If this is not possible the diagram D is called $(\psi_1+\psi_2)$-irreducible.] Then, the Bethe-Salpeter (B.-S.) equation for the Green function G (defined through the Fourier transform of (I.11) by factoring out the 4-dimensional δ-function which exhibits energy-momentum conservation) has the form

$$\delta(p_1+p_2-q_1-q_2)\, G(p_1,p_2;q_1,q_2) = \Delta'_{F1}(p_1)\,\Delta'_{F2}(p_2)\{\delta(p_1-q_1)\,\delta(p_2-q_2)$$

$$-\frac{i}{(2\pi)^2}\,\delta(p_1+p_2-q_1-q_2)\int I(p_1,p_2;k_1,k_2)\,\delta(k_1+k_2-q_1-q_2)$$

$$\times G(k_1,k_2;q_1,q_2)\,d^4k_1\,d^4k_2\ . \qquad \text{(I.12)}$$

Equation (I.12) can be rewritten in terms of the off-mass-shell scattering amplitude in a form similar to the Lippmann-Schwinger equation (I.7). To do this we introduce the following set of independent variables: the center-of-mass momentum

$$P = p_1+p_2 = q_1+q_2 \qquad \text{(I.13)}$$

and the relative momenta p and q defined by

$$p = \mu_2 p_1 - \mu_1 p_2,\ q = \mu_2 q_1 - \mu_1 q_2,\ \mu_a = \frac{m_a}{m_1+m_2}\ ,\ a = 1,2,$$

$$(p_1 = \mu_1 P+p,\ p_2 = \mu_2 P - p)\ . \qquad \text{(I.14)}$$

The off-mass-shell amplitude $T_p(p,q)$, equal to the sum of all connected Feynman diagrams without radiative corrections on the external lines, is related to the Green function G by

$$\Delta'_{F1}(p_1)\,\Delta'_{F2}(p_2)\,T_P(p,q)\,\Delta'_{F1}(q_1)\,\Delta'_{F2}(q_2) =$$

$$= (2\pi)^3[G(p_1p_2,q_1q_2)-\Delta'_{F1}(p_1)\Delta'_{F2}(p_2)\delta(p-q)] \qquad \text{(I.15)}$$

Multiplying both sides of Eq. (I.12) by the product of the inverse two-particle Green functions and using (I.15) we obtain the following equation for T:

$$T_P(p,q) = I_P(p,q) - \frac{i}{(2\pi)^2}\int I_P(p,k)\Delta'_{F1}(\mu_1 P+k)\Delta'_{F2}(\mu_2 P-k)T_P(k,q)d^4k. \tag{I.16}$$

The corresponding homogeneous equation for the B.-S. wave function

$$\varphi_{P,\alpha}(p) = \int \langle 0|T(\underline{\psi}_1(x)\underline{\psi}_2(y))|P,\alpha\rangle \exp\{i[(\mu_1 P+p)x+(\mu_2 P-y)y]\}d^4(x-y)$$
$$= \int \langle 0|T(\underline{\psi}_1(\tfrac{x}{2})\underline{\psi}_2(-\tfrac{x}{2}))|P,\alpha\rangle\, e^{ipx}\, d^4x \tag{I.17}$$

($|P,\alpha\rangle$ being a bound state of total momentum P and other quantum numbers α) is

$$\left[\Delta'_{F1}(\mu_1 P+p)\,\Delta'_{F2}(\mu_2 P-p)\right]^{-1}\varphi_P(p) = \frac{-i}{(2\pi)^2}\int I_P(p,k)\,\varphi_P(k)\,d^4k. \tag{I.18}$$

The wave function φ is simply related to the residue at the pole of the Green function for $P^2 = M_B^2$ where M_B is the mass of the bound state (see Ref. 8). We mention that if Eq. (I.18) is extrapolated to the two-particle mass-shell: $P = q_1+q_2$, $q_a^2 = m_a^2$, $a = 1,2$, then the substitution

$$\varphi_P(p) = (2\pi)^2 i\,\delta(p-q) + \Delta'_{F1}(\mu_1 P+p)\,\Delta'_{F2}(\mu_2 P-p)\,T_P(p,q) \tag{I.19}$$

would reduce it to Eq. (I.16) for the off-shell amplitude T.

Both the Green functions Δ'_{Fa} and the kernel I_P of Eq. (I.18) are defined by their perturbation expansions. Consider as an example the simplest case of two complex scalar fields ψ_1 and ψ_2 interacting via a neutral scalar field φ of mass μ with an interaction Lagrangian

$$L(x) = g{:}(\psi_1^*(x)\psi_1(x) + \psi_2^*(x)\psi_2(x))\,\varphi(x){:} \tag{I.20}$$

[g has the dimension of mass (i.e., of inverse length) in this model]. Then, in lowest order in perturbation theory (for both Δ'_{Fa} and I_P) Eq. (I.18) reads

$$[m_1^2-(\mu_1 P+p)^2][m_2^2-(\mu_2 P-p)^2]\varphi_P(p) = \frac{ig^2}{(2\pi)^2}\int\frac{1}{\mu^2-(p-k)^2-i0}\,\varphi_P(k)d^4k. \tag{I.21}$$

This is the so called ladder approximation in the B.-S. equation. (The iterative solution of the corresponding approximate equation for the off-shell amplitude gives the sum of all ladder graphs.)

Comparison of Eq. (I.18) or (I.21) with the non-relativistic Schrödinger equation displays a number of undesirable features of the B.-S. equation (see Ref. 5). First of all, it involves a fourth

coordinate, the relative energy $p_o(k_o)$ (or the relative time in the original B.-S. formulation[3]) which does not have a clear physical meaning; its presence makes obscure the non-relativistic limit of the B.-S. equation and leads to extra (unphysical) solutions (see Sec. I.C.). The operator in the left-hand side of (I.21) is a fourth-order polynomial in p (i.e., a 4th order differential operator in coordinate space). This is another source for extra solutions of the B.-S. equation (at least in the limit $g \to 0$). The strong singularity of the kernel of Eq. (I.16) does not permit to apply standard mathematical tools to this equation. There is no positive definite scalar product in the space of wave functions. [About the normalization of the solutions of (I.18) (or (I.21)) see Refs. 8 and 9.]

C. Stability Conditions and Wick Rotation.[5] A Solvable Model: The Scalar Coulomb Problem[5),6),10),11)]

Wick[5] observed that an important additional information about the B.-S. wave function can be obtained from the stability condition for particles 1 and 2:

$$\langle 0|\underline{\psi}_a(x)|q_n n\rangle = 0 \text{ for } q_n^2 < m_a^2 \quad a = 1,2. \tag{I.22}$$

The B.-S. wave function (I.17) may be written as

$$\varphi_{P\alpha}(p) = \int \Big\{ \theta(x_o) \sum_{\substack{n \\ q_n^2 \geq m_1^2}} \langle 0|\underline{\psi}_1(\tfrac{x}{2})|q_n n\rangle\langle q_n n|\underline{\psi}_2(-\tfrac{x}{2})|P\alpha\rangle$$

$$+\theta(-x_o) \sum_{\substack{n \\ q_n^2 \geq m_2^2}} \langle 0|\underline{\psi}_2(-\tfrac{x}{2})|q_n n\rangle\langle q_n n|\underline{\psi}_1(\tfrac{x}{2})|P\alpha\rangle \Big\} e^{ipx} d^4x = \varphi_P^+(p) + \varphi_P^-(p). \tag{I.23}$$

(As usually, the sum over intermediate states $\sum_n$ incorporates integration over the continuous variables.) In the rest frame of the bound state, for

$$P = (2E, \underline{0}) = (m_1 + m_2 - B, \underline{0}) \quad (B > 0) \tag{I.24}$$

we have

$$\varphi_P^+(p) = (2\pi)^3 \sum_{\substack{n \\ E_{1n} \geq E_{1P}}} \frac{\langle 0|\psi_1(0)|q_{1n}n\rangle\langle q_{1n}n|\psi_2(0)|P\alpha\rangle}{P_o - E_{1n} + E + i0}\ \delta(\underline{q}_{1n} - \underline{p}), \tag{I.25}$$

$$\varphi_P^-(p) = (2\pi)^3 \sum_{\substack{n \\ E_{2n} \geq E_{2P}}} \frac{\langle 0|\psi_2(0)|q_{2n}n\rangle\langle q_{2n}n|\psi_1(0)|P\alpha\rangle}{P_o + E_{2n} - E - i0}\ \delta(\underline{q}_{2n} + \underline{p}) \qquad (E_{ap} = \sqrt{m_n^2 + \underline{p}^2}) \tag{I.26}$$

We see from these expressions that $\varphi_P^{\pm}(p)$ are analytic functions of p_o with cuts for

$$p_o \geq E_{1p} - E = \sqrt{m_1^2 + \underline{p}^2} - \tfrac{1}{2}(m_1 + m_2 - B) = E_+ \quad ,$$

and

$$p_o \leq E - E_{2p} = \tfrac{1}{2}(m_1 + m_2 - B) - \sqrt{m_2^2 + \underline{p}^2} = E_- \quad , \tag{I.27}$$

respectively. Hence, their sum $\varphi_{P,\alpha}(p)$ is analytic in the p_o plane with the two cuts (I.27). We mention that for non-zero binding energy (i.e., for $B > 0$) there is always a real interval of analyticity, since

$$E_+ - E_- = E_{1p} + E_{2p} - m_1 - m_2 + B > 0. \tag{I.28}$$

For the sake of simplicity, we shall only treat the equal-mass case

$$m_1 = m_2 = m,\ E_+ = -E_- = E_p - m + \tfrac{1}{2}B \quad (E_p = \sqrt{m^2 + \underline{p}^2}) \tag{I.29}$$

in what follows. Assuming that the wave function of the bound-state $\varphi_P(p)$ goes to zero for $p_o \to \infty$ (in any direction) faster than $\frac{1}{p_o}$ and using the established analyticity of φ we can rotate the line of integration in the p_o plane in (I.21) on an angle $\frac{\pi}{2}$ in the positive direction; this amounts to replacing p_o by ip_4 and k_o by ik_4. Thus, we obtain

$$[(m^2 + \vec{p}^2 - E^2)^2 + 4E^2 p_4^2]\ \varphi_E(p) = \frac{g^2}{(2\pi)^2} \int \frac{1}{\mu^2 + (\vec{p} - \vec{k})^2}\ \varphi_P(\vec{k})\, d^4\vec{k} \tag{I.30}$$

where $p^2 = \underline{p}^2 + p_4^2 = p_1^2 + p_2^2 + p_3^2 + p_4^2$. In the unphysical point $E = 0$ ("the point of maximal binding") this equation has a manifest O(4) symmetry which has been extensively exploited in recent years (for a bibliography see Ref. 7).

It is much less obvious that Eq. (I.30) possesses a "hidden" O(4) symmetry for any eigenvalue of E in the case of zero-mass exchange ($\mu = 0$) in which the bound-state problem has been solved exactly.[5),6)] A standard way to see this is to extend the method of Fock[12)] (used originally to display the O(4) symmetry of the non-relativistic Coulomb problem), and to perform the following stereographic projection of the 4-dimensional momentum space onto the unit sphere in 5 dimension:

$$\vec{\xi} = \frac{2p_5}{\vec{p}^2 + p_5^2}\vec{p} \quad , \qquad \xi_5 = \frac{p_5^2 - \vec{p}^2}{p_5^2 + \vec{p}^2} \quad , \qquad p_5 = \sqrt{m^2 - E^2} \quad . \tag{I.31}$$

In these variables Eq. (I.30) (for $\mu = 0$) becomes[13)*]

$$[m^2 - (1+\xi_4^2)E^2]\psi_E(\xi) = \frac{g^2}{(2\pi)^3}\int \frac{\psi_E(\eta)\delta(\eta^2+\eta_5^2-1)}{1-\vec{\xi}\vec{\eta}-\xi_5\eta_5} \tag{I.32}$$

where

$$\psi_E(\xi) = \frac{1}{(1+\xi_5)^3}\,\varphi_E(p) \quad . \tag{I.33}$$

Equation (I.32) obviously allows an O(4) invariance group in the (1,2,3,5)-space. It is reduced to an ordinary second order differential equation in ξ_4 (see Refs. 6 and 10; we will skip the details). The equation in ξ_4 will obviously introduce a new quantum number, say k, in addition to the usually encountered quantum numbers n, ℓ, ζ $(=\ell_z)$ which span the O(4) representation space. In particular, the O(4)-degenerate energy levels E_{kn} will depend on this extra quantum number which has no non-relativistic analogue and no clear physical meaning; it corresponds to the relative energy (or the relative time). That seems to be a serious defect of the B.-S. equation. The unphysical variable comes in not only in some intermediate step--in the equation--but also in the observable result: in the energy eigenvalues. Besides, this feature is certainly not peculiar to the Wick-Cutkosky model only, since it originates from a simple counting of variables (and their conjugate quantum numbers). If the situation

---

*In order to obtain (I.32) from (I.30) one makes use of the following identities:

$$\vec{p} = \frac{2p_5}{1+\xi_5}\vec{\xi}, \quad \vec{p}^2 = 4(m^2-E^2)\frac{1-\xi_5}{1+\xi_5} \quad , \quad (\vec{p}-\vec{k})^2 = 8(m^2-E^2)\frac{1-\vec{\xi}\vec{\eta}-\xi_5\eta_5}{(1+\xi_5)(1+\eta_5)}$$

$$d^4k = 16\,\frac{(m^2-E^2)^2}{(1+\eta_5)^4}\,\frac{d^4\vec{\eta}}{|\eta_5|}$$

would remain unchanged for the B.-S. equation for the positronium (or the hydrogen atom), where the energy levels are measured with a high precision, we would find out that nothing in nature corresponds to the quantum number k.

Is there any way out of this difficulty (other than just throwing away part of the solutions on the basis that they do not have a non-relativistic limit)? Before drawing far going conclusions we should look more carefully at Eq. (I.32). It is not a standard equation for the energy eigenvalues. For fixed E it could be considered as a canonical equation for the eigenvalues of the dimensionless coupling

$$\lambda = \frac{g^2}{2m^2} \tag{I.34}$$

This would lead to a set of eigenvalues

$$\lambda_{kn} = f_{kn}\left(\frac{E}{m}\right) .$$

In particular, if E = 0, Eq. (I.32) becomes O(5) symmetric and the eigenvalues λ°_{kn} depend only on the sum k + n:

$$\lambda^{\circ}_{kn} = f_{kn}(0) = (k+n)(k+n+1), \quad k = 0,1,2,\ldots, \quad n = 1,2,\ldots \tag{I.35}$$

Only if the equation $\lambda = f_{kn}\left(\frac{E}{m}\right)$ can be solved with respect to E for all λ may we consider $E_{kn}(\lambda)$ as defined for any k, n and λ. However we cannot assume that all values of the coupling constant are admissible in Eq. (I.32). If we postulate (as it is customary)that the bound state energy should be non-negative, then the coupling constant λ cannot exceed 2. Indeed, according to (I.35), 2 is the lowest value of λ for E = 0; hence, for $\lambda > 2$ the lowest energy eigenvalue would be negative, since $E_{o_1}(\lambda)$ is a decreasing function of λ. On the other hand, the bound state eigenvalues E_{kn} should not exceed the elastic scattering threshold m. According to the Cutkosky analysis this excludes the eigenvalues E_{kn} with $k > 0$ if $\lambda < \frac{1}{4}$. In the admissible interval $\frac{1}{4} \leq \lambda \leq 2$, however, some of the extra solutions (with $k \geq 1$) do in fact appear. The question arises whether the condition $E_{kn} < m = E_{threshold}$ will be sufficient to exclude the positive k's in a more realistic situation, say for the positronium. I would be glad if some of you give me the answer to this question. In any case, the simplest and most straightforward calculation of the fine-structure of the hydrogen (and of the positronium) spectrum are performed within the framework of the quasipotential equation which we are going to consider next.

For a further analysis of the Wick-Cutkosky model in connection with relativistic symmetries and infinite-component wave equations we refer to Refs. 10 and 11.

II. Relativistic Quasipotential Equation[14)-20)]

A. Review of Non-Covariant Perturbation Theory[21)]

Our aim now is to derive a (3-dimensional!) relativistic analogue of the Lippmann-Schwinger equation on the basis of the conventional Hamiltonian formalism. First of all, we derive a graph technique for the old fashioned ("non-covariant") perturbation theory.

We start with the equation for the operator-valued function $R(\varkappa_1, \varkappa_2)$:

$$R(\varkappa_1, \varkappa_2) + \tilde{H}(\varkappa_1 - \varkappa_2) + \frac{1}{2\pi} \int \tilde{H}(\varkappa_1 - \varkappa) \frac{1}{\varkappa - i0} R(\varkappa, \varkappa_2)\, d\varkappa = 0 \qquad \text{(II.1)}$$

where $\tilde{H}(\varkappa)$ is the Fourier transform of the interaction Hamiltonian $H(\tau)$:

$$H(\tau) = \int_{x_0=\tau} H(x)\, d^3\underline{x}, \quad \tilde{H}(\varkappa) = \int H(\tau) e^{-i\varkappa\tau} d\tau \ . \qquad \text{(II.2)}$$

The direction of the time axis will be specified later. The operator R is related to the scattering operator $S = S(\infty, -\infty)$ by

$$S(\tau, -\infty) = 1 + \frac{1}{2\pi} \int R(\varkappa, 0) \frac{e^{i\varkappa\tau}}{\varkappa - i0}\, d\varkappa \qquad \text{(II.3)}$$

or

$$S = S(\infty, -\infty) = 1 + iR(0,0). \qquad \text{(II.4)}$$

[We have used the identity

$$\frac{1}{2\pi i} \lim_{\tau \to \pm\infty} \frac{e^{i\varkappa\tau}}{\varkappa - i0} = \begin{cases} \delta(x) & (\text{for } \tau \to +\infty) \\ 0 & (\text{for } \tau \to -\infty) \ . \end{cases}]$$

A diagram technique was developed by Kadyshevsky[21),16)] for the calculation of the matrix elements of R. In the case of a theory of spinless particles it can be summarized in the following way. To any ordinary Feynman diagram of N vertices we let correspond a set of N! new graphs having the same picture as the original one with all possible numeration of the vertices $1, \ldots, N$. Every internal line is orientated toward the vertex with smaller number. Furthermore we let a spurion (dotted) line enter the vertex 1, connect 1 with 2, 2 with 3

and so on (always oriented toward the vertex with larger number), and finally go out of the vertex N. [For instance, to the second order Feynman graph of Fig. 1a, correspond the two diagrams of Fig. 1b and 1c.] The conservation law in each vertex of the new diagrams takes into account the energies of the dotted lines. For instance, to vertex 1 of the graph shown in Fig. 1b corresponds the factor

$$-\frac{g}{\sqrt{2\pi}}\,\delta(q_1+k-p_1+(\varkappa_1-\varkappa)n)$$

where n is a 4-dimensional unit vector in the direction of the time axis. To a solid line of mass μ and momentum k we make correspond the "on-mass-shell propagator"

$$\delta^{+}_{\mu}(k)=\theta(k_o)\,\delta(k^2-\mu^2)\quad . \tag{II.5}$$

To an internal dotted line with "energy" $\varkappa$ we make correspond the propagator

$$\frac{1}{2\pi}\,\frac{1}{\varkappa-i0}\quad . \tag{II.6}$$

These rules give rise to the old-fashioned ("non-covariant") perturbation expansion for the scattering amplitude. If we start with a local interaction Hamiltonian H(x) the on-shell amplitude (for $\varkappa_1=\varkappa_2=0$) does not depend on the choice of the direction of the time axis. For instance, the contribution from the diagrams on Fig. 1b, c is

$$\frac{1}{(2\pi)^3}\,\delta(p_1+p_2-q_1-q_2+(\varkappa_2-\varkappa_1)n)\;T^{(2)}$$

where

$$T^{(2)}=\tfrac{1}{2}g^2\left(\frac{1}{\omega_{p_1-q_1}}\,\frac{1}{\varkappa_1+q_1^0-p_1^0+\omega_{p_1-q_1}-i0}+\frac{1}{\omega_{p_2-q_2}}\,\frac{1}{\varkappa_1+q_2^0-p_2^0+\omega_{p_2-q_2}-i0}\right) \tag{II.7}$$

$$\omega_k=\sqrt{\mu^2+\underline{k}^2}\quad . \tag{II.8}$$

For $\varkappa_1=\varkappa_2=0$, $q_1-p_1=p_2-q_2$ Eq. (II.7) reduces to the covariant Feynman rule for the on-shell amplitude T:

$$T^{(2)}=\frac{g^2}{\omega_{p_1-q_1}-(p_1^0-q_1^0)^2-i0}=\frac{g^2}{\mu^2-(p_1-q_1)^2-i0} \tag{II.9}$$

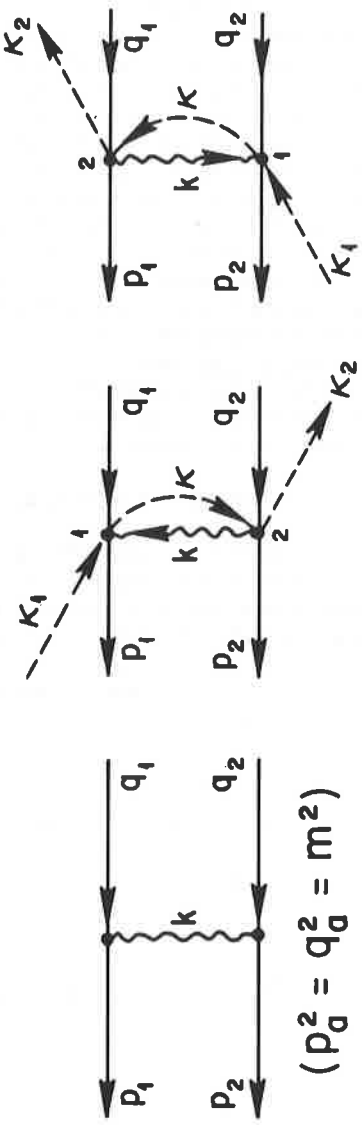

K_2
q_1
q_2
K
2
k
1
p_1
p_2
K_1
1c
q_1
q_2
K
K_2
1
2
k
K_1
p_1
p_2
1b
q_1
q_2
k
p_1
p_2
$(p_a^2 = q_a^2 = m^2)$
1a

FIG. 1

All ultraviolet divergences in higher order diagrams can be reduced to divergences in the integration over the variable $\varkappa$, and renormalization can be carried out in a way similar to the subtraction procedure in dispersion integrals (for more details see Ref. 21).

In general, the Kadyshevsky series (or, if you prefer, the old-fashioned perturbation expansion) is nothing else but a rearrangement of the familiar series of Feynman graphs. However, the physical picture corresponding to individual diagrams in the two approaches is different: While in a Feynman graph the 4-momentum on each internal line is off-mass-shell but the energy-momentum is conserved in each vertex, in our case each internal 4-momentum is on its mass-shell but the conservation law in each vertex is violated along the n-axis (though the overall energy momentum is conserved provided that $\varkappa_1 = \varkappa_2 = 0$). An undeniable formal advantage of the Feynman rules is their compactness: one Feynman graph with N vertices corresponds to N! different graphs in the Kadyshevsky formalism.

B. Two-Particle Quasipotential Equation

We consider the model (I.20) of interaction of two complex scalar fields ψ_1 and ψ_2 of mass m and a real scalar field φ of mass μ. The interaction Hamiltonian for this model is

$$H(x) = -L(x) = -g(:\psi_1^*(x)\psi_1(x)\varphi(x): + :\psi_2^*(x)\psi_2(x)\varphi(x):) \ . \qquad (II.10)$$

Our aim will be to write an equation for the (off-shell) elastic scattering amplitude of two equally charged particles ψ_1 and ψ_2. [It is known that contrary to the electromagnetic interaction via a vector field A_μ, the scalar interaction (II.10) of two such particles is attractive.]

First of all, following Ref. 1, we remark that Eq. (II.1) may be written in a more compact symbolic form

$$R + \tilde{H} + \tilde{H}G_0R = 0 \ . \qquad (II.11)$$

If we introduce the quasi-particle states $|\varkappa)$ normalized by

$$(\varkappa_1|\varkappa_2) = \delta(\varkappa_1 - \varkappa_2) \qquad (II.12)$$

and put

$$(\varkappa_1|R|\varkappa_2) = R(\varkappa_1,\varkappa_2), \ (\varkappa_1|\tilde{H}|\varkappa_2) = \tilde{H}(\varkappa_1-\varkappa_2), \ (\varkappa_1|G_0|\varkappa_2) = \frac{\delta(\varkappa_1-\varkappa_2)}{2\pi(\varkappa_1-i0)} \qquad (II.13)$$

and define the "matrix" multiplication as an integral over $\varkappa$, then Eq. (II.1) is "obtained" from (II.11) by taking the matrix element between $\varkappa_1$ and $\varkappa_2$. Iterating once Eq. (II.11) we get

$$R = -\tilde{H} + \tilde{H}G_o\tilde{H} + \tilde{H}G_o\tilde{H}G_oR \quad . \tag{II.14}$$

We take the matrix element of both sides of (II.14) between two $\psi_1+\psi_2$-particle states $\langle p_1 p_2|$ and $|q_1 q_2\rangle$. (We use the covariant normalization

$$\langle p|q\rangle = 2E_p\delta(\underline{p}-\underline{q}), \quad E_p = \sqrt{m^2+\underline{p}^2} \tag{II.15}$$

for one-particle states of momenta p and q and mass m.) Observing that for the interaction Hamiltonian (II.10)

$$\langle p_1 p_2|\tilde{H}|q_1 q_2\rangle = 0 \tag{II.16}$$

and separating the contribution of the intermediate $\psi_1+\psi_2$-particle state we obtain

$$\begin{aligned}\langle p_1 p_2|R|q_1 q_2\rangle = &\langle p_1 p_2|\tilde{H}G_o\tilde{H}|q_1 q_2\rangle \\ &+\int\langle p_1 p_2|\tilde{H}G_o\tilde{H}|k_1 k_2\rangle G_o\langle k_1 k_2|R|q_1 q_2\rangle (dk_1)(dk_2) \\ &+\sum_{n>2}\int d\sigma_n(k_1,\ldots,k_n)\langle p_1 p_2|\tilde{H}G_o\tilde{H}|k_1\ldots k_n\rangle \\ &\times G_o\langle k_1\ldots k_n|R|q_1 q_2\rangle \end{aligned} \tag{II.17}$$

where $(dk) = \delta_m^+(k)d^4k$ and $d\sigma_n(k_1,\ldots k_n)$ is the corresponding invariant measure for the n-particle intermediate state. [We mention that the matrix elements of the type $\langle p_1 p_2|R|q_1 q_2\rangle$ are still considered as operators in the space of the spurions energies $\varkappa$.] Our aim is to define a kernel K which incorporates the contribution from the n-particle states ($n > 2$) in order to obtain a linear equation for the two-particle amplitude. Let Π_2 be the projection operator on the subspace of two different ψ-particles ψ_1 and ψ_2. We define the kernel K by

$$\begin{aligned}K &= \tilde{H}G_o\tilde{H}\frac{1}{1-G_o(1-\Pi_2)\tilde{H}G_o\tilde{H}} \\ &= \tilde{H}G_o\tilde{H} + \tilde{H}G_o\tilde{H}G_o(1-\Pi_2)\tilde{H}G_o\tilde{H} + \ldots \quad .\end{aligned} \tag{II.18}$$

Then Eq. (II.17) can be rewritten in the form

$$\langle p_1 p_2 | R | q_1 q_2 \rangle = \langle p_1 p_2 | K | q_1 q_2 \rangle + \int \langle p_1 p_2 | K | k_1 k_2 \rangle G_0 \langle k_1 k_2 | R | q_1 q_2 \rangle (dk_1)(dk_2) \tag{II.19}$$

[We have taken into account that Π_2 commutes with G_0 so that $G_0(1-\Pi_2) = (1-\Pi_2)G_0(1-\Pi_2)$.]

In what follows it will be useful to introduce the complete "Green function" $G(k_1, k_2)$ which includes all possible radiative corrections on the lines with on-mass-shell momenta k_1 and k_2. For instance, in second order in perturbation theory,

$$(\varkappa | G^{(2)}(k_1, k_2) | \varkappa') = \frac{1}{2\pi}\delta(\varkappa-\varkappa')\left\{\frac{1}{\varkappa - i0} + \left(\frac{g}{2\pi}\right)^2 \left[\frac{F(\varkappa n, k_1)}{2k_{20}} + \frac{F(\varkappa n, k_2)}{2k_{10}}\right]\right\}$$

$$= \delta(\varkappa-\varkappa')\, G^{(2)}_{\varkappa}(k_1, k_2) \quad . \tag{II.20}$$

where

$$F(\varkappa n, k) = \int_{x_0(k)}^{\infty} dx\, f(x^2 - \underline{k}^2)\, \frac{(x+\varkappa)(k_0^2+x^2) + 2xk_0^2}{(x^2-k_0^2)^2[(x+\varkappa-i0)^2 - k_0^2]} \quad ,$$

$$x_0(k) = [(m+\mu)^2 + \underline{k}^2]^{\frac{1}{2}} \quad , \quad k_0 = (nk), \quad k^2 = k_0^2 - \underline{k}^2 = m^2 \tag{II.20a}$$

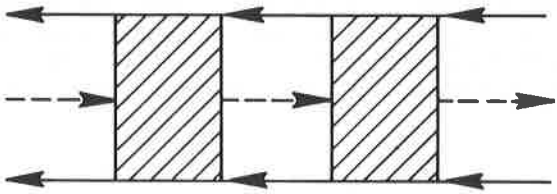
and f is defined by the phase space integral:

$$\frac{1}{\pi}\int \delta_m^+(k-q)\, \delta_\mu^+(q)\, d^4q = \theta(k_0)\, \theta(k^2-(m+\mu)^2)\, f(s) \quad ,$$

$$f(z) = \frac{1}{8z}\, [z^2 - 2(m^2+\mu^2)z + (m^2-\mu^2)^2]^{\frac{1}{2}} \quad .$$

Let us define the connected scattering amplitude R^C as the sum of all "solid-line-connected" graphs (i.e. graphs which remain connected when the dotted lines carrying momenta $\varkappa n$ are removed). Then an equation identical to (II.19) holds true for R^C with G_0 replaced by G and $\langle p_1 p_2 | K | q_1 q_2 \rangle$ replaced by the sum $\langle p_1 p_2 | K^C | k_1 k_2 \rangle$ of all (two-particle) irreducible diagrams, defined in the following way. A (connected)diagram of the $\psi_1\psi_2$-scattering amplitude is called irreducible if it cannot be split in two parts which are linked by a ψ_1 and a ψ_2-lines oriented in the direction of the external lines (say, from right to left) and one dotted line oriented in the opposite

direction (see Fig. 2) and if it does not contain radiative corrections on the external lines.



Reducible Graph

FIG. 2

In order to take the energy-momentum conservation explicitly into account, we put

$$\langle p_1 p_2 | (\varkappa_1 | R^C | \varkappa_2) | q_1 q_2 \rangle = \delta(p_1+p_2-q_1-q_2-(\varkappa_1-\varkappa_2)n) T_{\varkappa_1\varkappa_2}(p_1 p_2; q_1 q_2),$$

$$\langle p_1 p_2 | (\varkappa_1 | K^C | \varkappa_2) | q_1 q_2 \rangle = -\delta(p_1+p_2-q_1-q_2-(\varkappa_1-\varkappa_2)n) V_{\varkappa_1\varkappa_2}(p_1 p_2; q_1 q_2). \quad (II.21)$$

This allows us to rewrite Eq. (II.19) in the form

$$T_{\varkappa_1\varkappa_2}(p_1 p_2; q_1 q_2) + V_{\varkappa_1\varkappa_2}(p_1 p_2; q_1 q_2) +$$

$$+ \int \ldots \int \delta(p_1+p_2-k_1-k_2-(\varkappa_1-\varkappa_2)n)\, V_{\varkappa_1\varkappa}(p_1 p_2; k_1 k_2)$$

$$\times G_{\varkappa}(k_1 k_2) T_{\varkappa\varkappa_2}(k_1 k_2; q_1 q_2)\ (dk_1)(dk_2)\, d_{\varkappa} = 0\ , \quad (II.22)$$

where

$$G_{\varkappa}(k_1, k_2) = \frac{1}{2\pi(\varkappa - i0)} + \left(\frac{g}{2\pi}\right)^2 R_{\varkappa}(k_1, k_2; g^2)\ , \quad (II.23)$$

$R_{\varkappa}$ being regular for $\varkappa = 0$ (the first member of the series $R_{\varkappa}$ is given by (II.20)). The "potential" V is given by the sum of all strictly irreducible graphs defined by the condition that they cannot be represented in the form of Fig. 2. It can be shown that in the non-relativistic limit $V_{\varkappa_1\varkappa_2}$ goes into the non-relativistic potential. We shall see this later in Sec. IV.A for the special case of scalar Coulomb interaction.

C. Center of Mass Variables. Equation for the Wave Function

In what follows we shall treat Eq. (II.22) in the center of mass frame (assuming that the unit vector n, which defines the time axis, is collinear to p_1+p_2). [If we were interested in the t-channel behavior of the scattering amplitude (i.e. for time-like p_1-q_1) it would have been advantageous to choose n along p_1-q_1.] In this frame we put

$$\underline{p}_1 = -\underline{p}_2 = \underline{p},\ \underline{q}_1 = -\underline{q}_2 = \underline{q},\ \underline{n} = 0;$$

$$p_1^0 = p_2^0 = p^0,\ q_1^0 = q_2^0 = q^0;$$

$$|p_0| = E_p = \sqrt{m^2+\underline{p}^2},\ p_0 - \tfrac{1}{2}\varkappa_1 = q_0 - \tfrac{1}{2}\varkappa_2 \equiv E. \qquad \text{(II.24)}$$

$$T_{\varkappa_1\varkappa_2}(p_1p_2;\, q_1q_2) = T_E(p,q),\ V_{\varkappa_1\varkappa_2}(p_1p_2;\, q_1q_2) = V_E(p,q),$$

$$G_\varkappa(k_1,k_2) = G_2{}_{(k_0-E)}(k_0,\underline{k},\ k_0,-\underline{k}) \equiv 2k_0\, G_E(k). \qquad \text{(II.25)}$$

In these variables Eq. (II.21) can be written in the form

$$T_E(p,q) + V_E(p,q) + \int V_E(p,k)G_E(k)T_E(k,q)(dk) = 0. \qquad \text{(II.26)}$$

The corresponding equation for the complete two-particle Green function

$$\mathcal{G}_E(p;q) = G_E(p)(p_0+q_0)\delta(\underline{p}-\underline{q}) + G_E(p)T_E(p,q)G_E(q) \qquad \text{(II.27)}$$

is

$$\mathcal{G}_E(p,q) + G_E(p)\int V_E(p,k)\mathcal{G}_E(k,q)(dk) = (p_0+q_0)\delta(\underline{p}-\underline{q})G_E(p). \qquad \text{(II.28)}$$

Let there exist a r-fold degenerate ($r \geq 1$) bound state of mass 2B in the $\psi_1\psi_2$-system. Assume that in analogy with the Bethe-Salpeter equation and with the non-relativistic Lippmann-Schwinger equation the Green function $\mathcal{G}_E(p,q)$ has a simple pole for $E = B$ and in the neighbourhood of this pole can be written in the form†

$$\mathcal{G}_E(p,q) = N_B \sum_{a=1}^{r} \frac{\varphi_{Ba}(p)\,\bar{\varphi}_{Ba}(q)}{B - E - i0} + \text{regular terms for } E \sim B, \qquad \text{(II.29)}$$

† The form of the singular term in (II.29) is consistent with the non-covariant perturbation rules described in Sec. II.A if we assume the existence of r particles of mass 2B coupled to ψ_1 and ψ_2.

where $\varphi_{Ba}(p)$ will be interpreted as the wave function of the bound state of mass 2B and other quantum numbers specified by a and N_B is a normalization factor. Inserting (II.29) in Eq. (II.28) and comparing the residues at the pole E = B we obtain

$$\sum_{a=1}^{r} [\varphi_{Ba}(p) + G_B(p) \int V_B(p,k)\varphi_{Ba}(k)(dk)] \bar{\varphi}_{Ba}(q) = 0$$

Since $\bar{\varphi}_{Ba}(q)$ are linearly independent this implies the following homogeneous equation for each of the wave functions $\varphi_B(p)$:

$$G_B^{-1}(p)\,\varphi_B(p) + \int V_B(p,k)\varphi_B(k)(dk) = 0. \qquad \text{(II.30)}$$

In order to obtain the normalization condition for the wave function we apply to both sides of Eq. (II.28) the integral operator

$$\int \mathcal{G}_E(p,p')\, G_E^{-1}(p')\cdot(dp') \quad .$$

This leads to the following nonlinear equation for $\mathcal{G}_E$:

$$\int \mathcal{G}_E(p,k)\, G_E^{-1}(k)\mathcal{G}_E(k,q)(dk) +$$

$$+ \int\int \mathcal{G}_E(p,k_1)\, V_E(k_1,k_2)\mathcal{G}_E(k_2,q)(dk_1)(dk_2) = \mathcal{G}_E(p,q). \qquad \text{(II.31)}$$

Inserting (II.29) in (II.31) and comparing the residues at the pole E = B in both sides we obtain the following orthonormalization condition (cf. Refs. 30, 31)

$$N_B \int\int \bar{\varphi}_{Ba}(k_1)\left[-\frac{\partial}{\partial B}(G_B^{-1}(k_1)2E_{k_1}\,\delta(\underline{k}_1-\underline{k}_2) + V_B(k_1,k_2)\right]$$

$$\times \varphi_{Bb}(k_2)(dk_1)(dk_2) = \delta_{ab}\,. \qquad \text{(II.32)}$$

Consider the special case when $V_E(p,q)$ does not depend on E and G_μ is replaced by the first term in the expansion (II.23), so that, according to (II.25)

$$G_E^{-1}(k) \approx 8\pi E_k(E_k - E)\,. \qquad \text{(II.33)}$$

Choosing the normalization factor $N_B = \frac{1}{4\pi}$ we reduce Eq. (II.32) in this case to the normalization condition for the nonrelativistic Schrödinger wave function

$$\int \bar{\varphi}_{Ba}(\underline{k})\, \varphi_{Bb}(\underline{k})\, d^3\underline{k} = \delta_{ab}. \qquad \text{(II.34)}$$

We stress that Eq. (II.30) does not suffer from the diseases of the 4-dimensional Bethe-Salpeter equation discussed at the end of Sec. I.B. In contrast to the Bethe-Salpeter equation the three-dimensional Eq. (II.30) admits an unambiguous non-relativistic limit. However, we pay a certain price for the nice features of the quasipotential equation. If we replace the potential V_E by its second order approximation then the known analytic properties of the scattering amplitude will be distorted by the iterative solution of Eq. (II.26) (which is not the case for the corresponding Bethe-Salpeter equation).

III. Simplified Version of the Quasipotential Equation Consistent with Elastic Unitarity

A. Non-Uniqueness of the Off-Shell Extrapolation of the Scattering Amplitude and of the Corresponding Quasipotential Equation

In spite of the attractive general properties of the quasipotential equation (II.24) (or (II.26)) discussed at the end of the previous section, it has one defect: it is too complicated to provide exactly soluble problems in any reasonable approximation. Indeed, already in lowest order in perturbation theory the potential $V_E^{(2)}$ evaluated from the second order graphs at Fig. 1 has the "non-local"* form

$$V_E^{(2)}(p,q) = \frac{g^2}{\omega_{p-q}(2E-p_0-q_0-\omega_{p-q}+i0)} \qquad \text{(III.1)}$$

where $\omega_{p-q} = \sqrt{\mu^2+(\underline{p}-\underline{q})^2}$ and the corresponding quasipotential equation cannot be solved exactly even in the limit of zero mass exchange ($\mu = 0$).

However, it is known, that one can write different three-dimensional (quasipotential) equations which give rise to the same perturbation expansion for the on shell amplitude. For instance, the original quasipotential equation of Logunov and Tavkhelidze[14]**

$$T_{E_q}(\underline{p},\underline{q}) + V_{E_q}(\underline{p},\underline{q}) + \frac{1}{4\pi}\int V_{E_q}(\underline{p},\underline{k}) \frac{T_{E_q}(\underline{k},\underline{q})}{E_k^2-(E_q+i0)^2} \frac{d^3\underline{k}}{2E_k} = 0 \qquad \text{(III.2)}$$

*In analogy with the non-relativistic Schrödinger equation we call a potential $V(p,q)$ "local" if it depends on the difference $p-q$ only.

**We have changed the sign convention for V adopted in Ref. 14. Our choice fits the non-relativistic limit for the potential.

differs from our Eq. (II.24) both in the Green's function (i.e. the energy denominator) and the potential [their choice of the second order off-shell amplitude and potential being

$$T^{(2)}_{E_q}(\underline{p},\underline{q}) = -V^{(2)}_{E_q}(\underline{p},\underline{q}) = \frac{g^2}{\mu^2+(\underline{p}-\underline{q})^2}\ ] \ .$$

Both Eqs. (II.24) and (III.2) belong to a large family of linear equations of the type

$$T + V + VGT = 0 \qquad \text{(III.3)}$$

which have the following property in common: for real V, in the physical region, they lead automatically (at least formally) to the elastic unitarity condition.

To describe the whole class of equations of the type (III.3) with this property we write the solution of (III.3) as

$$T = -\frac{1}{1+VG}V = -V\frac{1}{1+GV} \ . \qquad \text{(III.4)}$$

If the potential is Hermitian, $V = V^*$, then the discontinuity of T in the s-channel is given by

$$T - T^* = \frac{-1}{1+VG}V + \frac{1}{1+VG^*}V = T(G - G^*)\,T^* \ . \qquad \text{(III.5)}$$

In order to make Eq. (III.5) identical with the elastic unitarity condition

$$\underline{T}(p,q) - \underline{T}^*(p,q) = \frac{i}{4E}\int \underline{T}(p,k)\,\underline{T}^*(k,q)\delta(E_k^2 - E^2)\,d^3\underline{k} \qquad \text{(III.6)}$$

(where, for the on shell amplitude $T(p,q)$, $p_o = q_o = E$) we have to specify accordingly the discontinuity of the Green function. It is readily verified that for both Green functions G_E (II.25) and

$$G'_E(E_k) = \frac{1}{4\pi[E_k^2 - (E+i0)^2]}$$

(corresponding to Eqs. (II.24) and (III.2), respectively) the discontinuity is the same:

$$G_E - G_E^* = G'_E - G'^*_E = \frac{1}{4E}\delta(E_k - E) \qquad \text{(III.7)}$$

and it leads to (III.6) (we use in both cases the invariant volume element $\frac{d^3k}{2E_k}$ on the upper hyperboloid).

We will exploit the freedom of the off-shell extrapolation of the scattering amplitude in order to write a simpler equation consistent with (III.7) (i.e. with the elastic unitarity condition). The potential in any such modified quasipotential equation is calculated from a definite off-shell extrapolation of the perturbation expansion of the amplitude T (see Ref. 14). [We require, for instance that in the lowest order $V^{(2)} = -T^{(2)}$, where $T^{(2)}$ coincides on shell with (II.9); this latter requirement is violated in the quasipotential equation proposed in Ref. 18.]

B. Discussion of a Simplified Version of the Quasipotential Equation

We will consider the following model equation of the type (III.3)

$$T_E(p,q) + V_E(p,q) + \frac{1}{8\pi E}\int V_E(p,q)\,\frac{\epsilon(k_o)}{k_o - E - i0}\,T_E(k,q)\,\delta(k^2-m^2)\,d^4k = 0.$$

$$(\epsilon(k_o) = \operatorname{sgn} k_o = \theta(k_o) - \theta(-k_o)) \qquad \text{(III.8)}$$

and the corresponding homogeneous equation

$$(E - p_o)\,\varphi_E(p) = \frac{1}{8\pi E}\int V_E(p,k)\,\varphi_E(k)\,\epsilon(k_o)\,\delta(k^2-m^2)\,d^4k = 0. \qquad \text{(III.9)}$$

It is readily checked that the Green function

$$G_E(k) = \frac{\epsilon(k_o)}{8\pi E(k_o - E - i0)}$$

corresponding to these equations fulfills the elastic unitarity condition (III.7). This choice of G_E is among the simplest possibilities (consistent with (III.7)) since the operator in the left-hand side of (III.9) is a first degree polynomial in p_o. Besides we will restrict ourselves to the second order approximation in the potential choosing it as the "local" energy independent extrapolation

$$V_E(p,q) = \frac{g^2}{(p-q)^2 - \mu^2 + i0} \qquad \text{(III.10)}$$

of $-T^{(2)}$ (II.9). An important feature of Eqs. (III.8), (III.9) is that they involve integration over the two-sheeted hyperboloid $k^2 = m^2$. (We mention that $T_E(k,q)$ could be interpreted for $k_o < 0$ as the amplitude of a process with four incoming particles, which is possible off

energy shell.) We will see in Sec. IV that in the case of scalar Coulomb potential (i.e. for $\mu = 0$) the presence of the lower sheet of the hyperboloid $k^2 = m^2$ in the domain of integration in (III.9) is essential in order to ensure the O(4) symmetry of the bound state problem.

A more complicated model, with Green's function

$$G_E(k) = \frac{1}{8\pi E_k (k_o - E - i0)} \tag{III.11}$$

(and also involving integration over a two-sheeted hyperboloid) was considered in Ref. 22. It leads to the same O(4) degeneracy of the energy levels.

As some justification of Eq. (III.8) we observe that the exact expression for the fourth order box diagram (Fig. 3) after integration over the internal energy k_o in the center-of-mass frame can be written in the form:

$$T_{box}(p,q) = \frac{1}{8\pi E} \int V_E(p,k) \left[\frac{1}{k_o - E - i0} + \epsilon(k_o) \frac{\omega^3_{p-k} + \omega^3_{q-k} + \omega_{p-k}\omega_{q-k} - (E-k_o)^2}{\omega_{p-k}\omega_{q-k}(\omega_{p-k} + \omega_{q-k})} \right] V_E(k,q)\delta(k^2 - m^2)d^4k \tag{III.12}$$

where V_E is given by (III.11). We see that this expression contains the second iteration of Eq. (III.9) plus a term which is regular in the physical region and, hence, does not contribute to the imaginary part of T.

The comparison between (III.12) and the second iteration of (III.9), i.e. the integral

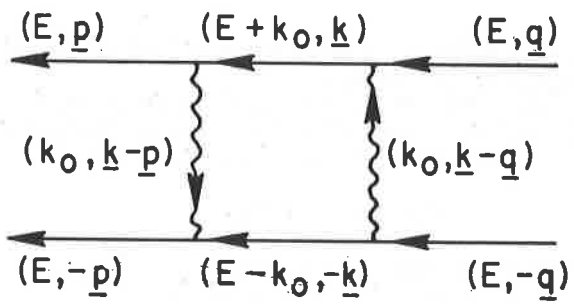


FIG. 3

$$T_2 = \frac{g^4}{8\pi E} \int \frac{1}{(p-k)^2 - \mu^2} \frac{\varepsilon(k_0)}{k_0 - E - i0} \frac{1}{(k-q)^2 - \mu^2} \delta(k^2 - m^2)\, d^4k \qquad \text{(III.13)}$$

(for $p_0 = q_0 = E$) may give us a feeling of the discrepancy of the fourth order calculated from the Bethe-Salpeter equation in the ladder approximation and the quasipotential equation (III.9). It is possible to evaluate explicitly both (III.12) and (III.13) for the case of forward scattering. The result for the box diagram is

$$T_{box}(p,p) = \frac{g^4}{4m^2\mu^2} F(E) \qquad \text{(III.14)}$$

where

$$F(E) = \frac{i\frac{\pi}{2} \operatorname{th}\alpha + \theta_0 \operatorname{ctg}\theta_0 - \alpha \operatorname{th}\alpha}{\operatorname{ch}2\alpha + \cos 2\theta_0} \qquad \text{for } E \geq m \qquad \text{(III.15a)}$$

with

$$\cos\theta_0 = \frac{\mu}{2m}, \quad \operatorname{ch}\alpha = \frac{E}{m}, \quad \alpha > 0;$$

$$F(E) = \frac{\theta \operatorname{ctg}\theta - \theta_0 \operatorname{ctg}\theta_0}{\cos 2\theta - \cos 2\theta_0} \quad \text{for } E^2 = m^2 \sin^2\theta < m^2 . \qquad \text{(III.15b)}$$

Eq. (III.15a) may be considered as analytic continuation of (III.15b) for complex θ ($\theta = \frac{\pi}{2} + i\alpha$). With the same notation the forward contribution from (III.13) for $E \geq m$ is

$$T_2 = \frac{g^4}{4m^2\mu^2} \frac{\frac{\pi}{2}\left(i\operatorname{th}\alpha + \operatorname{ctg}\theta_0\right)}{\operatorname{ch}2\alpha + \cos 2\theta_0} . \qquad \text{(III.16)}$$

The imaginary parts of (III.15a) and (III.16) coincide as they should.

IV. The Scalar Coulomb Problem: Solution and Relation to Infinite-Component Wave Equations[22),20)]

A. O(4) Symmetry of the Scalar Coulomb Problem in Fock Variables[22)]

In this section we shall exhibit the O(4) symmetry and solve Eq. (III.10) with the attractive scalar Coulomb potential

$$V_E(p,k) = \frac{2\alpha m^2}{\pi(p-k)^2} \qquad \text{(IV.1)}$$

(i.e. with the potential (III.11) for $\mu = 0$, $g^2 = 2\frac{\alpha m^2}{\pi}$). To do this we shall use the Fock variables $p \leftrightarrow u$, $k \leftrightarrow v$ where

$$u_j = \frac{\sqrt{1-\omega^2}}{p_0 - \omega m} p_j, \quad j = 1,2,3, \quad u_4 = \frac{m - p_0 \omega}{p_0 - \omega m},$$

$$p_0 = m \frac{1+\omega u_4}{u_4 + \omega}, \quad p_j = m \frac{\sqrt{1-\omega^2}}{u_4+\omega} u_j, \quad \omega = \frac{E}{m} < 1 \tag{IV.2}$$

(cf. the treatment of Wick-Cutkosky model in Sec. I.C). Eqs. (IV.2) define a map of the two-sheeted hyperboloid $p^2 = m^2$ onto the Euclidean unit sphere in 4 dimension: $\vec{u}^2 = \underline{u}^2 + u_4^2 = 1$. If we restrict p to vary on the upper sheet of the hyperboloid, then $\vec{u}$ would cover only the part of the unit sphere for which $-\omega \le u_4 \le 1$. (It is clear at this point that our subsequent results would not be true had we restricted the integration in the right-hand side of Eq. (III.10) to the upper hyperboloid.) The Jacobian of the transformation (IV.2) is

$$\left| \frac{D(p_0, p_1, p_2, p_3)}{D(u_1, u_2, u_3, u_4)} \right| = m^4 \left(\frac{\sqrt{1-\omega^2}}{|u_4 + \omega|} \right)^6 . \tag{IV.3}$$

The singularity at $u_4 = -\omega$ corresponds to $|p_0| \to \infty$. It is compensated in the integrand of (III.10) by the assumed decrease of $\varphi_E(k)$ at infinity. Using (IV.3) and the identities

$$(p-k)^2 = m^2 \frac{(\omega^2-1)(\vec{u}-\vec{v})^2}{(u_4+\omega)(v_4+\omega)}, \quad \epsilon(k_0)(v_4+\omega) = |v_4+\omega|,$$

$$p_0 - E = m \frac{1-\omega^2}{u_4+\omega}, \quad k^2 - m^2 = m^2(1-\omega^2) \frac{1-\vec{v}^2}{(v_4+\omega)^2} , \tag{IV.4}$$

we transform (III.10) into

$$\psi_\omega(u) = \frac{\alpha}{4\pi^2 \omega \sqrt{1-\omega^2}} \int_{\vec{v}^2=1} \frac{\psi_\omega(\vec{v})}{(\vec{u}-\vec{v})^2} d\Omega_{\vec{v}} \tag{IV.5}$$

where

$$\psi_\omega(\vec{u}) = \frac{\varphi_E(p)}{(u_4+\omega)^2} , \tag{IV.6}$$

and $d\Omega_{\vec{v}}$ is the volume element on the unit sphere in four dimension. Eq. (IV.5) coincides with the non-relativistic Schrödinger equation for the Coulomb problem in Fock variables if we replace $\omega\sqrt{1-\omega^2}$ in the right-hand side of $\sqrt{-\frac{E_{NR}}{2\mu}}$, where μ is the reduced mass. Just as well as in the non-relativistic case this equation is manifestly O(4)-invariant. A complete set of solutions of the eigenvalue problem

(IV.5) is given by the spherical harmonics in four dimension $Y_{n\ell\zeta}(\vec{u})$ which satisfy

$$Y_{n\ell\zeta}(\vec{u}) = \frac{n}{4\pi^2} \int_{\vec{v}^2} \frac{1}{1-\vec{u}\vec{v}} \, Y_{n\ell\zeta}(\vec{v}) \, d\Omega_{\vec{v}}, \quad n = 1,2,\ldots, \tag{IV.7}$$

$$[\underline{L}^2 - \ell(\ell+1)]Y_{n\ell\zeta} = 0, \quad (L_3 - \zeta)Y_{n\ell\zeta} = 0, \tag{IV.8}$$

where $\underline{L}$ is the angular momentum operator in three dimension, $\ell = 0, 1, \ldots, n-1$, $-\ell \le \zeta \le \ell$ (see e.g. Ref. 23). Comparing (IV.5) with (IV.7) we see that the eigenvalues of ω are determined by the equation

$$\frac{\alpha}{\omega_n \sqrt{1-\omega_n^2}} = 2n \, . \tag{IV.9}$$

This leads to

$$\omega_n = \sqrt{\tfrac{1}{2} + \tfrac{1}{2}\sqrt{1 - \left(\frac{\alpha}{n}\right)^2}} \, . \tag{IV.10}$$

(We have excluded the second root of (IV.9) by the requirement that $\omega_n \to 1$ for $\alpha \to 0$.) The binding energy $B = 2m(1 - \omega_n)$ goes to the correct non-relativistic expression

$$-E_{NR} = \frac{\mu\alpha^2}{2n^2} \quad \left(\mu = \frac{m}{2}\right) \tag{IV.11}$$

for $\alpha \to 0$. The O(4)-degeneracy of Eq. (IV.5) (just as well as the degeneracy of the Wick-Cutkosky equation (I.32)) displays the main qualitative distinction between the scalar "Coulomb" interaction and the real electromagnetic interaction (via a 4-vector potential) which necessarily leads to a fine splitting of the relativistic energy levels with respect to the total angular momentum.

B. Algebraization of the Scalar Coulomb Problem[24),25),26),22)]

Now we will establish a one-to-one correspondence between the quasipotential equation

$$(E - p_o)\varphi_E(p) = \frac{g^2}{8\pi E} \int \frac{1}{(p-k)^2} \, \varphi_E(k) \, \varepsilon(k_o) \, \delta(k^2 - m^2) \, d^4k \tag{IV.12}$$

and an infinite-component wave equation written in terms of the generators of the zero helicity representation of the conformal group SO(4,2). A similar algebraization has been carried out for the Bethe-Salpeter equation (for the same case of scalar Coulomb interaction) in Ref. 10.

We will make use of the well known degenerate representation of SO(4,2) which can be realized on the set of homogeneous functions on the upper light cone $u_o = \sqrt{u_1^2 + u_2^2 + u_3^2 + u_4^2}$ of degree of homogeneity -2 or -1 (see e.g. Ref. 25). It is equivalent to the zero helicity representation[27] of the conformal group [for an explicit demonstration of the equivalence of the two representations see Ref. 22a (Appendix)]. This representation can be realized equivalently on the space $\mathcal{H}_1$ of functions defined on the double sheeted hyperboloid $p^2=1$, equipped with the scalar product

$$(\varphi,\psi) = \frac{1}{\pi^4}\iint \overline{\varphi(p)}\,\frac{-1}{(p-q)^2}\,\psi(q)\,\delta(p^2-1)\delta(q^2-1)d^4p\,d^4q. \qquad (IV.13)$$

The (homogeneous) Lorentz group acts in $\mathcal{H}_1$ as a group of argument transformations:

$$[U(\Lambda)\psi](q) = \psi(\Lambda^{-1}q) \ .$$

The generators Γ_μ and Γ_5 of SO(4,2) (i.e. the representatives of the Dirac γ-matrices $\frac{i}{2}\gamma_\mu$ and $\frac{i}{2}\gamma_5$ in this infinite-dimensional unitary representation) are defined by the following non-local operators:

$$\left[\Gamma_\mu\varphi\right](p) = -\frac{2}{\pi^2}\int \frac{q_\mu}{[(p-q)^2]^2}\,\varphi(q)\,\varepsilon(q_o)\,\delta(q^2-1)\,d^4q \qquad (IV.14)$$

$$\left[\Gamma_5\varphi\right](p) = -\frac{2}{\pi^2}\int \frac{1}{[(p-q)^2]^2}\,\varphi(q)\,\varepsilon(q_o)\,\delta(q^2-1)\,d^4q \ . \qquad (IV.15)$$

Comparing (IV.14) with (IV.15) we see that

$$p_\mu\,\varphi(p) = \left[\frac{1}{\Gamma_5}\,\Gamma_\mu\,\varphi\right](p) \ . \qquad (IV.16)$$

It can be verified directly (see Ref. 22) that the operators $p_\mu = \frac{1}{\Gamma_5}\Gamma_\mu$ commute between themselves and that $p_\mu p^\mu = 1$. It is also not difficult to check that the inverse of the operator (IV.15) is given by

$$\left[\frac{1}{\Gamma_5}\varphi\right](p) = \frac{1}{2\pi^2}\int \frac{-1}{(p-q)^2}\,\varphi(q)\,\varepsilon(q_o)\delta(q^2-1)d^4q \ . \qquad (IV.17)$$

Changing p and q in (IV.16), (IV.17) to p/m, q/m and inserting in Eq. (IV.12) we get the following "algebraic form" of the quasipotential equation:

$$\left(\frac{m}{\Gamma_5}\Gamma_0 - E\right)\varphi_E = \frac{\alpha m^2}{2E\Gamma_5}\varphi_E \qquad \text{(IV.18)}$$

where $\alpha = \frac{\pi}{2m^3} g^2$. The discrete spectrum corresponding to Eq. (IV.18) can be found by multiplying both sides by Γ_5 from the left and performing a rotation in the (0,5)-plane (cf. Ref. 24). The result is

$$\left(E\sqrt{m^2 - E^2}\,\Gamma_0 - \frac{\alpha}{2}m^2\right)\varphi_E = 0 \ . \qquad \text{(IV.19)}$$

Finally we recall that the eigenvalues of Γ_0 in the given representation are all positive integers (see e.g. Ref. 22) and find

$$E_n^2 = \frac{m^2}{2}\left(1 + \sqrt{1 - \frac{\alpha^2}{n^2}}\right)$$

in agreement with (IV.10). Eq. (IV.18) admits also a continuous spectrum corresponding to the two-particle scattering states.

In conclusion we would like to make the following remarks.

1) The preceding argument gives a simple prescription for the "algebraization" of the (free) 4-momentum:

$$p_\mu = \frac{m}{\Gamma_5}\Gamma_\mu \qquad \text{(IV.20)}$$

(see (IV.16)). This prescription is independent of the interaction under consideration.

2) The simple algebraization of the potential based on Eq. (IV.17) is peculiar to the case of zero mass exchange. The potential (III.11) with $\mu > 0$ leads already to considerable complications (see Sec. III.2 of Ref. 22). The reason is that the kernel in the scalar product (IV.13) in $\mathcal{K}_1$ is closely related to the Coulomb potential. If, onthe other hand, we adapt the scalar product in our representation space to the potential for $\mu > 0$ the simplicity of the free Hamiltonian will be distorted. However, we can use Eqs. (IV.15), (IV.17) and (IV.20) to solve the inverse problem: given ad hoc an infinite-component wave equation in the representation space of the ladder representation of U(2,2) (see Refs. 24, 25, 26) to reconstruct an equivalent integral equation in momentum space.

3) The potential in the right-hand side of (IV.18) (with a minus sign) will coincide with the non-relativistic attractive Coulomb potential in coordinate space $\left(-\frac{\alpha}{r}\right)$ if we identify r with $\frac{1}{m}\Gamma_5$. This observation is not accidental. It has been argued in Ref. 17 that in general for spin 0 particles the relativistic generalization of r is given by $r^2 = \frac{1}{m^2}(\underline{N}^2 - \underline{L}^2)$ where $\underline{L}$ and $\underline{N}$ are the generators of the homogeneous Lorentz group. In our case $\underline{N}^2 - \underline{L}^2 = \Gamma_5^2$.

4) As mentioned before, the simplest evaluation of the Lamb shift corrections to the hydrogen or positronium energy levels (or of the hyperfine splitting of the hydrogen levels due to the nucleon form-factor[28),29)]) has been done on the basis of the Logunov-Tavkhelidze quasipotential equation. It would be interesting to carry out this more realistic calculations on the basis of the quasipotential equation (III.10) and of the algebraic technique developed here.

Acknowledgments

The author would like to thank Dr. C. Kaysen for his hospitality at the Institute for Advanced Study where the main part of these lectures was prepared. It is a pleasure to thank also Professors W. E. Brittin and K. T. Mahanthappa for their hospitality at the Institute for Theoretical Physics at the University of Colorado.

References

1. V. DeAlfaro and T. Regge, Potential Scattering (North-Holland Publishing Co., Amsterdam, 1965).
2. M. Goldberger, K. Watson, Collision Theory (J. Wiley and Sons, Inc., New York, London, Sydney, 1964).
3. E. E. Salpeter and H. A Bethe, Phys. Rev. 84, 1232 (1951).
4. M. Gell-Mann and F. Low, Phys. Rev. 84, 350 (1951).
5. G. C. Wick, Phys. Rev. 96, 1124 (1954).
6. R. E. Cutkosky, Phys. Rev. 96, 1135 (1954).
7. N. Nakanishi, Kyoto University preprint RIMS-51 (1969); a complete bibliography on the subject is given in this paper.
8. D. Lurié, A. J. Macfarlane and Y. Takahashi, Phys. Rev. 140, B1091 (1965).
9. C. H. Llewellyn Smith, Nuovo Cimento 60A, 348 (1969).
10. E. Kyriakopoulos, Phys. Rev. 174, 1846 (1968).
11. J. E. Mandula, California Institute of Technology preprint CALT-68-185 (1969).
12. V. A. Fock, Z. Physik 98, 145 (1936).
M. Lévy, Proc. Roy. Soc. (London) A204, 145 (1950).
13. R. Delbourgo, A. Salam and J. Strathdee, Nuovo Cimento 50A, 193 (1967).
14. A. A. Logunov and A. N. Tavkhelidze, Nuovo Cimento 29, 380 (1963).
15. R. Blankenbecler and R. Sugar, Phys. Rev. 142, 1051 (1966).
16. V. G. Kadyshevsky, Nucl. Phys. B6, 125 (1968).
V. G. Kadyshevsky and M. D. Mateev, Nuovo Cimento 55A, 275 (1968).
17. V. G. Kadyshevsky, R. M. Mir-Kasimov and N. B. Skachkov, Nuovo Cimento 55A, 233 (1968).

18. M. D. Mateev, R. M. Mir-Kasimov and M. Freeman, J.I.N.R., Dubna, preprint P2-4107 (1968).
19. V. G. Kadyshevsky, M. D. Mateev and R. M. Mir-Kasimov, J.I.N.R., Dubna, preprint E2-4030 (1968).
20. C. Itzykson, V. G. Kadyshevsky and I. T. Todorov, "Three-dimensional Formulation of the Relativistic Two-body Problem and Infinite-component Wave Equations" (to be published).
21. V. G. Kadyshevsky, Soviet Phys. JETP 19, 443 and 597 (1964).
V. G. Kadyshevsky, Soviet Phys. Doklady 10, 46 (1965).
22. a) C. Itzykson and I. T. Todorov, Proceedings of the Coral Gables Conference on Fundamental Interactions at High Energy, 1969.
b) I. T. Todorov, Proceedings of the Battelle-Seattle Rencontres in Mathematics and Physics, 1969.
23. M. Bander and C. Itzykson, Rev. Mod. Phys. 38, 330 and 346 (1966).
24. Y. Nambu, Progr. Theor. Phys. (Kyoto) Suppl. 37 and 38, 368 (1966).
Y. Nambu, Phys. Rev. 160, 1171 (1967).
25. C. Fronsdal, Phys. Rev. 156, 1665 (1967).
26. A. O. Barut and H. Kleinert, Phys. Rev. 157, 1180 (1967).
A. O. Barut and H. Kleinert, Phys. Rev. 160, 1149 (1967).
H. Kleinert, Fortschr. Phys. 16, 1 (1968).
27. G. Mack and I. Todorov, J. Math. Phys. 10, 2078 (1969).
28. R. N. Faustov, Nucl. Phys. 75, 669 (1966), and references to earlier work of Faustov quoted there.
29. H. Grotch and D. R. Yennie, Rev. Mod. Phys. 41, 350 (1969).
30. C. H. Llewellyn Smith, Nuovo Cimento 60A, 348 (1969);
V. A. Matveev, J.I.N.R., Dubna, preprint P2-432F (1968).
31. R. N. Faustov and A. A. Helashvili, J.I.N.R., Dubna, preprint P2-4345 (1969).

LOCALIZED SOLUTIONS OF RELATIVISTIC NONLINEAR DIFFERENTIAL EQUATIONS†

Eric H. Roffman
Department of Mathematics
State University of New York At Stony Brook
Stony Brook, New York 11790

Introduction

For linear equations such as the Schrödinger equation, the typical behavior of a wave packet is that it will spread and its maximum value will tend to zero as time passes.

Many physical phenomena have a different behavior, if friction is neglected. For example, think of drops on a window pane sliding downward under the force of gravity. They may fragment. They may collide with other drops on the way down. They may leave a trail of tiny drops behind them. But they do retain a coherent localized shape, more or less, all the way down.

This kind of behavior--localized units which interact without spreading--may be typical of certain solutions of a whole class of nonlinear partial differential equations. These lectures will summarize what is known about constructing such equations, and solving them.

First we shall treat some ordinary differential equations. This will illustrate the characteristics of nonlinearity. Furthermore, the equations we study--the Riccati equation, the elliptic equations, and the Van der Pol equation--will have features that we shall use when treating partial differential equations.

We shall then be concerned with properties of nonlinear wave equations of the form

$$\Box\varphi = m^2\varphi - \lambda\varphi(\varphi\bar{\varphi}) + \beta\varphi(\varphi\bar{\varphi})^2 \ ,$$

with positive definite energy.

---

†Presented at the INSTITUTE FOR THEORETICAL PHYSICS, University of Colorado, Summer 1969.

These are the simplest prototypes of the differential equations of quantum field theory. Quantum field theory considers nonlinear equations for linear operators. We are looking for solutions which are ordinary functions. (Well behaved functions, in fact, which go to zero exponentially for large $|\vec{x}|$.)

We shall describe stationary solutions, and the interaction process. We shall describe several kinds of behavior which an equation might, a priori, give rise to.

Then we shall show that the strongest possible property--the strong particle scattering property--is satisfied by at least one equation, the Korteweg-de Vreis equation. We shall report on the work that has been done on this and related equations.

I. Nonlinear Ordinary Differential Equations

In this lecture our main object is the study of the Riccati differential equation, the equations for elliptic functions, and the Van der Pol equation. Each of these equations has features we shall refer to again.

We shall begin with a very simple example which illustrates the differences between linear and nonlinear differential equations.

Consider the linear differential equation

$$\frac{df}{dt} = \lambda f \tag{I.1}$$

The solution is

$$f = f(0)\, e^{\lambda t} . \tag{I.2}$$

Note that the functions $f_a = ae^{\lambda t}$, $f_b = be^{\lambda t}$ and $f_{a+b} = (a+b)e^{\lambda t} = f_a + f_b$ all satisfy the equation. Also note that if $f(0)$ is finite then $f(t)$ is finite for all time.

On the other hand, consider

$$\frac{df}{dt} = \lambda f^2 \tag{I.3}$$

This has a solution obtained by writing

$$f^{-2}\, df = \lambda\, dt$$

$$\int_{f(t=t_o)}^{f(t)} f^{-2}\, df = \lambda \int_{t=t_o}^{t} dt$$

$$-f^{-1}(t) + f^{-1}(0) = \lambda (t - t_o) \tag{I.4}$$

Hence

$$f(t) = \frac{f(0)}{1 - \lambda f(0)(t-t_o)} \tag{I.5}$$

$$f_a = \frac{a}{1 - \lambda a(t-t_o)} \quad \text{and} \quad f_b = \frac{b}{1 - \lambda b(t-t_o)} \tag{I.5}$$

both satisfy the equation, and so does

$$f_{(a+b)} = \frac{a+b}{1 - \lambda (a+b)t} \ , \quad \text{but } f_{a+b} \neq f_a + f_b \ .$$

Furthermore, even if $f_a(t_o) = a$ is finite, $f_a(t)$ is not finite for all time, if λ is real. It becomes infinite in a finite time, namely

$$t = t_o + \frac{1}{\lambda a} \ .$$

Finally, a linear differential equation may have singularities, but they are independent of the boundary values. On the other hand, here we see that the position of the singularity depends on the value a of f at t_o. This is called a moving pole, because its position moves with the boundary value.

A. Riccati Equation

The first equation is called the Riccati equation, which was first considered in a special case by John Bernoulli in 1694, and first solved in a special case in 1701.

We will derive the equation from its solution,

$$y = \frac{a(x) + k\, b(x)}{c(x) + k\, d(x)} \ , \text{ where } bc \neq ad \ . \tag{I.6}$$

k is the constant of integration, and to obtain the differential equation we must eliminate k.

$$y' = \frac{a' + kb'}{c + kd} - \frac{(c' + kd')y}{c + kd} \ . \tag{I.7}$$

Solving for k

$$k = -\frac{a - yc}{b - yd} \quad \text{from (I.6)}$$

also

$$k = -\frac{(a' - y'c - c'y)}{b' - y'd - d'y} \quad \text{from (I.7)} \ .$$

So, recalling the assumption that $bc - ad \neq 0$,

$$y' + \frac{(a'd - ad' + bc' - b'c)}{(bc - ad)}\, y + \frac{(cd' - c'd)}{(bc - ad)}\, y^2 = \frac{a'b - ab'}{(bc - ad)} \quad . \tag{I.8}$$

This is often written as

$$y' + Qy + Ry^2 = P \quad . \tag{I.9}$$

Note that if $P = 0$ then $a = 0$, $b = 0$, or $a = \lambda b$. Also, setting $y = \frac{1}{v}$ we obtain $v' - Qv = R$ a linear equation with general solution

$$v = Ce^{\int Q dx} + e^{\int Q dx} \int^x R(t)\, e^{-\int Q dt}\, dt \quad . \tag{I.10}$$

It is an exercise for the reader to show the relationship between this solution and the previous one (Eq. (I.6)).

One of the most interesting properties of the Riccati equation is the following:

Theorem: To every Riccati equation corresponds a linear second order differential equation, and to every linear second order differential equation corresponds a family of Riccati equations.

Proof: Set

$$y = \frac{1}{R} (\log u)'$$

then Eq. (I.9) becomes

$$Ru'' - (R' - QR)u' - PR^2 u = 0 \quad . \tag{I.11}$$

Conversely, the equation

$$Au'' + Bu' + Cu = 0$$

becomes the Riccati equation

$$y' + [(\log R)' + B/A]y + Ry^2 = - C/AR \quad .$$

We can choose R so that

$$R = e^{-\int^x (B/A) dx}$$

giving

$$y' + Ry^2 = C/AR \quad .$$

B. Elliptic Functions

Consider the equation

$$y'' = a - b^2 y \quad . \tag{I.12}$$

The solution is

$$y = \frac{a}{b^2} + \alpha \sin(bx + \beta) \quad .$$

If we multiply by y' and integrate we get

$$\frac{(y')^2}{2} - \left[ay - \frac{b^2 y^2}{2}\right] = \text{const} \; .$$

The constant is evaluated at $x = -\beta/b$ and is $\frac{(\alpha b)^2}{2} - \frac{1}{2}\frac{a^2}{b^2}$ so

$$(y')^2 = \left[(\alpha b)^2 - \frac{a^2}{b^2}\right] + 2ay - b^2 y^2 \quad .$$

The simplest generalization of Eq. (I.12) is

$$y'' = \sum^{N} a_n y^n \tag{I.13}$$

or

$$(y')^2 = \sum^{N} \frac{a_n}{n+1} y^{n+1} + b \tag{I.14}$$

If $1 < N \leq 3$ the solution is an elliptic function.

The following transformations and definitions are standard. If the polynomial roots are ± 1 and $\pm 1/k$ the equation may be written

$$(y')^2 = (1 - y^2)(1 - k^2 y^2)$$

letting $y = \sin\theta$ we obtain

$$(\theta')^2 = (1 - k^2 \sin^2\theta) \; .$$

We define $F(x,k)$

$$F(x,k) = \int_0^x \frac{dx}{\sqrt{(1 - x^2)(1 - k^2 x^2)}}$$

or

$$F(\theta,k) = \int_0^{\theta} \frac{d\varphi}{\sqrt{1 - k^2 \sin^2\varphi}} \equiv u_k(\theta) \quad .$$

Define

$$sn(u,k) = \sin\theta$$

$$cn(u,k) = \cos\theta \quad .$$

The transformation from Eq. (I.13) to Eq. (I.14) will be particularly useful later.

C. Van der Pol Equation

The equation

$$\ddot{y} - \epsilon(1 - y^2)\dot{y} + ay = 0 \tag{I.15}$$

was discussed by B. Van der Pol in 1926 in connection with the oscillations in a triode amplifier circuit.

Since the elliptic functions were discussed extensively in the early 1800's, and Riccati's equation in the early 1700's, we have roughly a century passing between the study of each of these three nonlinear equations.

If we set $x = \dot{y}$ in Eq. (I.15) we have

$$\frac{dx}{dt} = \epsilon x + ay - \epsilon x y^2 \tag{I.16a}$$

$$\frac{dy}{dt} = x \quad . \tag{I.16b}$$

Note that if we set $x = y = 0$, the equation becomes

$$\frac{dx}{dt} = 0$$

$$\frac{dy}{dt} = 0 \quad .$$

Hence $x(t) = y(t) = 0$ is a constant solution. Near $x = y = 0$ we can diagonalize Eqs. (I.16).

Set

$$x = A\xi + B\eta \tag{I.17a}$$

$$y = C\xi + D\eta \tag{I.17b}$$

and look for a solution of the form

$$\frac{d\xi}{dt} = \lambda_1 \xi \qquad \frac{d\eta}{dt} = \lambda_2 \eta \tag{I.18}$$

near $x = y = 0$ (i.e. the ϵxy^2 term is neglected).

Substituting Eqs. (I.17) in Eqs. (I.16) and using (I.18) we obtain

$$A\lambda_1 \xi + B\lambda_2 \eta = \epsilon(A\xi + B\eta) + a(C\xi + D\eta)$$

$$C\lambda_1 \xi + D\lambda_2 \eta = A\xi + B\eta \quad .$$

So

$$\lambda_1 = A/C = \epsilon + a\ C/A$$

$$\lambda_2 = B/D = \epsilon + a\ D/B$$

or

$$(\lambda_1, \lambda_2) = (A/C,\ B/D) = \frac{\epsilon \pm \sqrt{\epsilon^2 - 4a}}{2} \quad .$$

So, near $x = y = 0$ the solution is

$$y = e^{\frac{1}{2}\epsilon t}\left(e^{\frac{t}{2}\sqrt{\epsilon^2 - 4a}} + e^{-\frac{t}{2}\sqrt{\epsilon^2 - 4a}}\right) \tag{I.19a}$$

$$x = \frac{dy}{dt} = e^{\frac{1}{2}\epsilon t}\left(\epsilon \cosh\left(\frac{t}{2}\sqrt{\epsilon^2 - 4a}\right) + \tfrac{1}{2}\sqrt{\epsilon^2 - 4a}\ \sinh\left(\frac{t}{2}\sqrt{\epsilon^2 - 4a}\right)\right). \tag{I.19b}$$

Near $x = y = 0$ we see that for $\epsilon > 0$, $a > 0$, x and y are both growing. If $\epsilon^2 < 4a$ the trajectory spirals because $\sqrt{\epsilon^2 - 4a}$ is imaginary.

In fact, when one studies the trajectory in detail one sees that one family of trajectories moves out from $x = y = 0$ to a certain limiting closed trajectory. A second family starts from large x and y and spirals down to the same limiting closed trajectory.

The diagram which shows the trajectories of solutions is called the phase diagram. It looks like this for the Van der Pol equation:

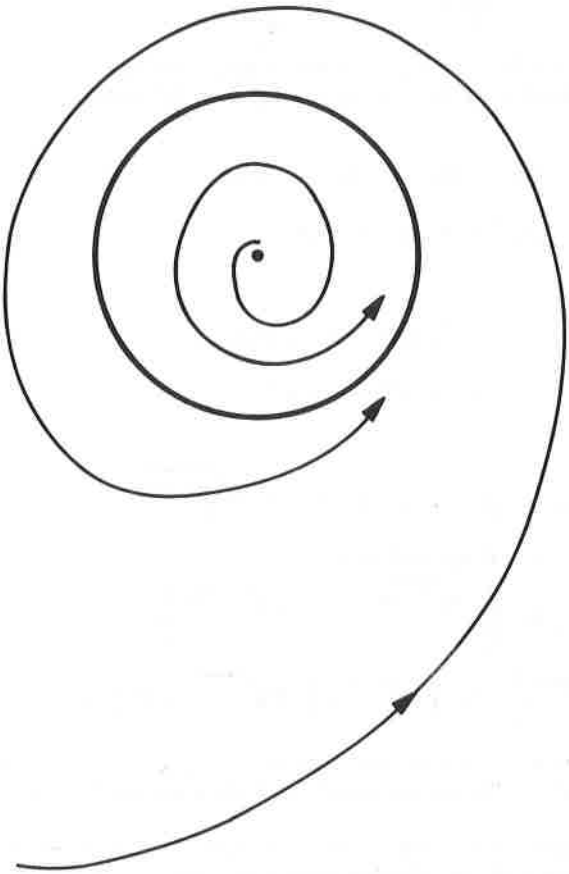


Fig. 1. Phase diagram with limit cycle for Van der Pol's equation.

Note that for any finite time we can start at any point on the x,y plane, and then for $t \to +\infty$ we will be asymptotic to the limit cycle. For $t \to -\infty$ there are only three places to start from--$x \to 0$, $y \to 0$; $x \to \infty$, $y \to \infty$; or someplace on the limit cycle. Thus, the boundary conditions that can be imposed at $t \to \pm\infty$ are somewhat different from the boundary conditions at finite time.

Another interesting feature of the Van der Pol equation is the way it approaches a linear equation as $\epsilon \to 0$. From Eq.(I.19), we see that as $\epsilon \to 0$, near $x = y = 0$ the solution approaches a circle. However, no matter how small ϵ may be, as long as it is greater than zero, the solution after spiraling long enough will eventually reach the one and only one limit cycle characteristic of the Van der Pol equation. The approach to the linear equation as $\epsilon \to 0$ is therefore not uniform.

II. Stationary and Elementary Solutions of Field Equations

We now begin the main subject of these lectures, the properties of certain self-interacting scalar fields.

Starting with an equation of the form

$$-\Box\varphi = m^2\varphi - \lambda\varphi(\bar{\varphi}\varphi) + \beta\varphi(\bar{\varphi}\varphi)^2 \qquad m^2, \lambda, \beta > 0 . \quad \text{(II.1)}$$

We want to know when such an equation has solutions of the form

$$\varphi(x,t) = e^{-i\omega t}\,\theta(\vec{x}) \quad . \qquad \text{(II.2)}$$

Such a solution is called stationary.

We shall always require positive definite energy for any solution of any equation we study.

A. Integral Invariants

We shall begin by recalling the derivation of the field equations and the energy-momentum tensor from the Lagrangian.

Consider the Lagrangian

$$L = \int \mathfrak{L}\, d'x = \int [\,(\dot{\bar{\varphi}}\dot{\varphi}) - \sum_i (\bar{\varphi}_i \varphi_i) - V(\bar{\varphi}\varphi)]\, dx \qquad \text{(II.3)}$$

dx denotes an integral over space and time.

The field equation is given by

$$\partial_o\left(\frac{\partial \mathcal{L}}{\partial \dot{\bar{\varphi}}}\right) - \frac{\partial \mathcal{L}}{\partial \bar{\varphi}} = 0$$

$$\ddot{\varphi} - \sum_i \varphi_{ii} + \varphi F(\varphi\bar{\varphi}) = 0 \qquad \text{(II.4)}$$

where $F(y) = V'(y)$

and $\dfrac{\partial(\varphi_i\bar{\varphi}_i)}{\partial\bar{\varphi}}$ is understood as $-\varphi_{ii}$.

The Lagrangian is invariant under two groups of transformations. First, the translations and rotations making up the Poincaré group leave the Lagrangian invariant. Second, phase changes of the form $\varphi \to e^{i\lambda}\varphi$ also leave it invariant.

From the translation invariance we obtain energy momentum conservation

$$\partial^\mu T_{\mu\nu} = 0 \qquad \text{(II.5)}$$

where

$$T_{\mu\nu} = \left(\frac{\partial \mathcal{L}}{\partial \varphi_\mu}\right)\bar{\varphi}_\nu + \frac{\partial \mathcal{L}}{\partial \varphi_\mu}\varphi_\nu - g_{\mu\nu}\mathcal{L} \quad . \qquad \text{(II.6)}$$

From Eqs. (II.5) and (II.6) we see that the energy E is conserved, where

$$E = \int \mathcal{E}\, d\vec{x} = \int T_{oo}\, d\vec{x} =$$

$$\int\left[\dot{\bar{\varphi}}\,\dot{\varphi} + \sum_i \bar{\varphi}_i\varphi_i + V(\varphi\varphi)\right] d\vec{x} \qquad \text{(II.7)}$$

$d\vec{x}$ denotes an integral over space.

Notice that the energy is positive definite if and only if the potential V is positive definite.

The momentum is also conserved

$$P_i = \int \mathcal{P}_i\, d\vec{x} = \int T_{oi}\, d\vec{x}$$

$$= 2\int \text{Real}\,(\dot{\bar{\varphi}}\,\varphi_i)\, d\vec{x} \quad . \qquad \text{(II.8)}$$

From invariance under transformations belonging to the homogeneous Lorentz group we may obtain the conservation of angular momentum and the "center of mass relation."

From the gauge invariance of the first kind

$$\varphi \to e^{i\lambda}\varphi$$

we obtain the conservation of "charge"

$$Q = 2\ \mathrm{Im} \int (\varphi\, \bar{\varphi})\ d\vec{x}. = \int \mathfrak{D}\ d\vec{x} \qquad \text{(II.9)}$$

B. Stationary Solutions

After this brief discussion of some important integral invariants of any solution of Eq. (II.4), let us return to the solutions (II.2) of (II.1). By substitution we obtain

$$\nabla^2\theta = \theta(m^2 - \omega^2 - \lambda\theta^2 + \beta\theta^4) \quad \text{for real } \theta\ . \qquad \text{(II.10)}$$

This is now an elliptic (time independent) differential equation. Compared to Eq. (1) it is trivial.

We shall seek solutions of (10) with the boundary conditions $\theta'(\vec{x} = 0) = 0$, $\theta(|\vec{x}| \to \infty) \to 0$, $\theta'(|\vec{x}| \to \infty) \to 0$, θ is everywhere smooth and finite. We shall also seek solutions symmetrical about $x = 0$.

We can write

$$\nabla^2\theta = \frac{d^2}{dr^2}\,\theta(r) + \frac{n-1}{r}\,\frac{d}{dr}\,\theta(r) = \theta(r)([m^2 - \omega^2] - \lambda[\theta(r)]^2 + \beta[\theta(r)]^4)$$

where r represents the coordinate in n space dimensions. Notice that in one space dimension the second term drops out.

Recalling the procedure we adopted for the elliptic functions, we multiply through by $\frac{d\theta}{dr}$ and integrate

$$\left[\left[\frac{d}{dr}\,\theta(r)\right]^2 + \omega^2\theta^2 - V(\theta(r))\right]_{r_2}^{r_1} = -\int_{r_2}^{r_1} \frac{n-1}{r}\left(\frac{d\theta}{dr}\right)^2 dr\ .$$

Using the boundary conditions at ∞ we obtain

$$(\theta'(r))^2 + \omega^2\theta^2(r) - V(\theta(r)) = -\int_{\infty}^{r} \frac{n-1}{r}\left(\frac{d\theta}{dr}\right)^2 dr\ . \qquad \text{(II.12)}$$

Notice that in one space dimension the right hand side is absent, making possible many simplifications.

Equation (II.12) has many important consequences:

Cor. 1: For positive definite potential (which we always require) a stationary solution has $\omega^2 > 0$.

Proof: From the boundary conditions, at the origin $\theta'(r) = 0$, hence Eq. (II.12) becomes

$$\omega^2 \theta^2(0) - V(\theta(0)) = \int_0^\infty \frac{n-1}{r} \left(\frac{d\theta}{dr}\right)^2 dr \ .$$

For positive definite V this is possible only if $\omega^2 > 0$.

Because of the absence of the second term in Eq.(II.11), the case of one space dimension can be solved quite explicitly. We shall now assume one space dimension.

Cor.2: In one space dimension, coming from a positive definite potential, there is a stationary solution of the form (II.2) for Eq. (II.1) if and only if

(a) the equation

$$\omega^2 y^2 = V(y)$$

has a solution $y_o > 0$ and

(b)

$$y_o(m^2 - \lambda y_o^2 + \beta y_o^4) = \frac{dV(y_o)}{dy} < 0.$$

Furthermore if (a) and (b) are true, then

$$\omega^2 = V(y_o)/y_o^2, \quad \theta(x=0) = y_o$$

and

$$\lim_{|x| \to \infty} \theta = c\, e^{-|x|\sqrt{m^2 - \omega^2}} \ .$$

Proof:

In one space dimension Eq. (II.12) becomes

$$(\theta'(r))^2 + \omega^2 \theta^2(r) = V(\theta(r)) \ . \qquad \text{(II.13)}$$

(Because θ is symmetric about $x = 0$, we use r for $|x|$.) Hence at the origin, for a localized solution,

$$\omega^2 \theta^2(r) = V(\theta(r)) \ . \qquad \text{(II.14)}$$

Conversely, if $\theta(0) = y_o$,

$$\omega^2 y_o^2 = V(y_o^2)$$

and $\theta'(0) = 0$ then Eq. (II.13) is satisfied. Since

$$\theta'' = \frac{dV(y_o)}{dy} < 0$$

the slope θ' becomes negative.

If we now integrate Eq. (II.13)

$$\theta' = \sqrt{V(\theta) - \omega^2 \theta^2} \tag{II.15}$$

we obtain a function which satisfies Eq. (II.13), and at r=0 is sloping downward. The function will continue to decrease until it is asymptotic to $\theta=0$. It cannot cross $\theta=0$ because at $\theta=0$ $\theta' \neq 0$ from Eq. (II.13), and it cannot turn up or stop decreasing because from Eq. (II.13) if $\theta'=0$ then $\omega^2 \theta^2 = V(\theta)$, but this equation, from the form of the potential, has no solutions other than $\theta=y_o=\theta(r=0)$ and $\theta=0$.

We have already seen that $\theta(r=0) = y_o$ and that ω^2 satisfies the equation

$$\omega^2 y_o^2 = V(y_o) \ .$$

If ω^2 were such that this equation had no solution, then we cannot carry out this procedure and there would be no stationary solution with frequency ω.

From Eq. (II.13) we see that not only does $\theta(r)$ go to zero for large r, it goes to zero like

$$\theta(r) \to e^{-\sqrt{m^2 - \omega^2}\, r} \tag{II.16}$$

Note that one can easily write generalizations of Cor. 2 for more general potentials.

By means of Eq. (II.12), and corollaries one and two, we have obtained quite a complete picture of when stationary localized solutions can exist and what they look like.

Now we also can use Eq. (II.12) to derive some of their integral properties. These are interesting because it turns out that the energy momentum tensor of a stationary solution has a rather special form.

Cor. 3: For a stationary solution $\varphi = e^{-i\omega t}\, \theta(x)$ in one space dimension

$$\mathcal{E} = 2V(\theta(x)) \tag{II.17}$$

$$\mathcal{P} = 0 \tag{II.18}$$

$$T_{ii} = 0 \tag{II.19}$$

$$\mathcal{L} = 2(\omega^2\theta^2(x) - V(\theta(x)) = -2(\theta'(x))^2 \tag{II.20}$$

$$\mathcal{Q} = 2\omega\,\theta^2(x) \tag{II.21}$$

Proof:

$\mathcal{P}$ comes out zero since $\dot{\theta}$ is imaginary. $\mathcal{Q}$ is immediate from the definition. For the rest, substitute

$$(\theta')^2 = V(\theta) - \omega^2\theta^2$$

into their respective definitions.

We see that the energy momentum tensor may be written

$$T_{\mu\nu} = 2\begin{pmatrix} V & 0 \\ 0 & 0 \end{pmatrix} \quad . \tag{II.22}$$

As an example, we shall work out in some detail the properties of the potential

$$V(a) = m^2 a - \lambda a^2 + \beta a^3$$

Here we have made some changes of notation. (For the potential of Eq. (II.1) we had $V(\varphi) = m^2\varphi\bar{\varphi} - \frac{\lambda}{2}(\varphi\bar{\varphi})^2 + (\beta/3)(\varphi\bar{\varphi})^3$.)

We write $V(a) = a\tilde{V}(a) = (m^2 - \lambda a + \beta a^2)a$. The minimum of $\tilde{V}$ occurs at $\lambda/2\beta = a$ at which

$$\tilde{V} = (m^2 - \lambda^2/4\beta)$$

hence V is positive definite if $m^2 > \lambda^2/4\beta$.

A stationary solution exists, according to Cor. 2, if $\omega^2 a = V(a)$. Therefore

$$m^2 > \omega^2 > m^2 - \frac{\lambda^2}{4\beta} \tag{II.23a}$$

and

$$a = \frac{\lambda}{2\beta} - \sqrt{\omega^2 - (m^2 - \frac{\lambda^2}{4\beta})} \tag{II.23b}$$

where

$$\theta(x=0) = \sqrt{a}$$

for a stationary solution of frequency ω.

Notice that as $\omega^2 \to m^2$, $a \to 0$, while for $\omega^2 \to (m^2 - \frac{\lambda^2}{4\beta})$, $a \to \frac{\lambda}{2\beta}$. So there is a continuous family of stationary solutions centered about $x = 0$ with maxima ranging from zero to $\sqrt{\lambda/2\beta}$.

It is not possible to have a stationary solution with $\omega^2 = m^2 - \frac{\lambda^2}{4\beta}$, $\theta(x=0) = \sqrt{\lambda/2\beta}$. For at $x = 0$, the equation for θ is

$$\theta'' = \theta(m^2 - \omega^2 - 2\lambda a + 3\beta a^2) \ .$$

Substitution gives

$$\theta'' = 0 \ .$$

Hence (II.23a) is a strict inequality.

C. Elementary Solutions

We have already noted that the Lagrangian for Eq. (II.1) is invariant under the transformations of the Poincaré group. Hence, if $\varphi(x_\mu)$ is any solution of Eq. (II.1), translations, rotations and Lorentz boosts may be applied to φ and they yield new solutions of Eq. (II.1).

$$\varphi(x_\mu + a_\mu)$$

is obtained after translation.

$$\varphi(x_\mu) \to \varphi(\Lambda_\mu{}^\nu x_\nu + a_\mu) \qquad \text{(II.24)}$$

is the most general transformation of φ by elements of the Poincaré group. If $\varphi(x_\mu)$ satisfies Eq. (II.1), then $\varphi(\Lambda_\mu{}^\nu x_\nu + a_\mu)$ also satisfies Eq. (II.1).

If a stationary solution $e^{-i\omega t}\theta(x)$ is transformed in this way we obtain a solution of the form

$$e^{-i\omega\Lambda_\mu{}^0(x^\mu - a^\mu)}\theta[\Lambda_i{}^\mu(x_\mu - a_\mu)] = \varphi(x,t) \ . \qquad \text{(II.25)}$$

Such a solution is called an elementary solution. It is centered at $x^\mu = a^\mu$ and is moving with velocity $v_i = \frac{\Lambda^0{}_i}{\Lambda^0{}_0}$, without any change in shape.

III. Dynamics and Scattering

A. Introduction and Formalism

We have been studying the localized, stationary solutions of field equations, and we have also noted that, as a consequence of Poincaré invariance, a whole family of elementary solutions, which can be centered anywhere and may be moving with arbitrary velocity ($0 \leq v < 1$), can be generated from the stationary solutions.

We now ask whether these stationary solutions can scatter. That is, in some sense can we have a solution of the wave equation

$$-\Box\varphi = \varphi(m^2 - \lambda(\varphi\bar{\varphi}) + \beta(\varphi\bar{\varphi})^2) \qquad \text{(III.1)}$$

which at $t = -\infty$ looks like two or more elementary solutions approaching each other. And if so, what does such a solution look like at $t = +\infty$.

Before going into this problem in more detail, let us recall the characteristic features of solving the linear Schrödinger equation.

If φ is a solution of

$$-i\partial_t\varphi = \nabla^2\varphi - V(x)\varphi$$

then we can expand φ in the form

$$\varphi = \int d\mu_E \varphi_E(x) e^{+iEt}$$

where $\varphi_E(x)$ is a solution of

$$E\varphi_E(x) = \nabla^2 \varphi_E - V(x)\varphi_E \quad .$$

The particular solution is determined essentially by boundary conditions.

We see that because we are dealing with a linear equation, for which the superposition principle is valid, we have been able to give the general time dependent solution by means of a superposition of functions each of which has trivial time dependence, and is determined by the solution of a time independent elliptic equation.

Unfortunately, these techniques do not apply to nonlinear equations.

Our first objective is to find suitable boundary conditions.

There is a striking resemblance between stationary solutions and the constituents $e^{iEt}\varphi_E(x)$ which made up a general solution of

the Schrödinger equation. This suggests that stationary and, more generally, elementary solutions are of particular interest. Of course, superpositions of elementary solutions are not themselves solutions of Eq. (III.1), because of the nonlinearity. However, as we shall see, to a limited extent, the sum of elementary solutions can be asymptotic to a solution.

We shall investigate the following sort of property: A partial differential equation will be said to have a particle interpretation if

(a) It has a family of elementary solutions

(b) For any collection $\{\varphi_i(x_\mu)\}_{i=1,\ldots,N}$ of elementary solutions which satisfy $\lim_{t\to\pm\infty} \varphi_i(x,t)\,\varphi_j(x,t) \to 0$ for all i,j $(i \neq j)$, there is a solution $\varphi_+(x_\mu)$ and a solution $\varphi_-(x_\mu)$ of Eq. (III.1) such that

$$\lim_{t\to+\infty} \varphi_+(x_\mu) = \sum_{i=1}^{N} \varphi_i(x_\mu)$$

$$\lim_{t\to-\infty} \varphi_-(x_\mu) = \sum_{i=1}^{N} \varphi_i(x_\mu)$$

We denote by Φ^+ the set of all such solutions φ_+ and by Φ^- the set of all such solutions φ_-.

(c) There exist solutions which are in both Φ^- and Φ^+. Such a solution is asymptotic to a sum of elementary solutions at $t = -\infty$ and, whatever happens at finite times, at $t = +\infty$ it again becomes asymptotic to a sum of elementary solutions. Denote such a set of functions by Φ. Clearly

$$\Phi \subset \Phi^+ \cap \Phi^- \tag{III.2}$$

Note that if $\varphi \in \Phi$, then

$$\lim_{t\to-\infty} \varphi = \sum_{i=1}^{N} \varphi_i(x_\mu)$$

and

$$\lim_{t\to+\infty} \varphi = \sum_{j=1}^{M} \varphi_j(x_\mu)$$

where the collection $\{\varphi_i\}_{i=1,\ldots N}$ is not necessarily the same as the collection $\{\varphi_j\}_{j=1,\ldots M}$.

(d) For any $\varphi_1 \in \Phi$, such that

$$\lim_{t \to +\infty} \varphi_1 \to \sum_{i=1}^{N} \varphi_i$$

there is a $\varphi_2 \in \Phi$, such that

$$\lim_{t \to -\infty} \varphi_2 \to \sum_{i=1}^{N} \varphi_i$$

and the converse is also true.

This says that any state which is the result of a scattering process can also give rise to a scattering process. And every initial state is also the result of a scattering process.

For any equation which satisfies the conditions a-d, a set Φ of (nontrivial) solutions which satisfies c and d will be called a particle space.

We have used the conditions at $t = \pm\infty$ to select out certain types of boundary conditions as particularly interesting. This is reminiscent of the Van der Pol equation for which the boundary conditions one might specify at $t = \pm\infty$ were different from those at finite time.

In addition to the condition of a particle interpretation, several other kinds of behavior are a priori possible for a nonlinear partial differential equation. Some are stronger, some weaker than conditions a-d.

First, we shall define the notion of the spectrum of a particle space.

Let Φ be a particle space. Then any $\varphi \in \Phi$ satisfies

$$\lim_{t \to -\infty} \varphi = \sum_{i=1}^{N} \varphi_i(x)$$

$$\lim_{t \to +\infty} \varphi = \sum_{j=1}^{M} \varphi_j(x)$$

where $\varphi_i(x)$ and $\varphi_j(x)$ are elementary solutions.

Each elementary solution φ_i has a certain energy E_i in its rest frame. We define the spectrum of φ, $\mathcal{E}_\varphi$, as the pair

$$\mathcal{E}_\varphi = (\{E_i\}_{i=1,N}, \{E_j\}_{j=1,m}) \ . \tag{III.3}$$

Then we define the spectrum of Φ, $\mathcal{E}_\Phi$, as the collection of the spectra of its elements

$$\mathcal{E}_\Phi = \{\mathcal{E}_\varphi | \varphi \in \Phi\} \ . \tag{III.4a}$$

Equivalently,

$$\mathcal{E}_\Phi = \left(\left\{ E_i \Big| \lim_{t \to -\infty} \varphi = \varphi_i + \sum_{k=1}^{N} \varphi_k, \varphi \in \Phi \right\} \ , \right.$$

$$\left. \left\{ E_j \Big| \lim_{t \to +\infty} \varphi = \varphi_j + \sum_{\ell=1}^{M} \varphi_\ell, \varphi \in \Phi \right\} \right) \tag{III.4b}$$

writing

$$\mathcal{E}_\Phi = (\mathcal{E}_+, \mathcal{E}_-) \ ,$$

we note that because of condition (d), $\mathcal{E}_+ = \mathcal{E}_-$, hence

$$\mathcal{E}_\Phi = (\mathcal{E}, \mathcal{E}) \ . \tag{III.5}$$

We can also define the spectra of Φ^+, Φ^-, and $\varphi_\pm \in \Phi^\pm$, i.e. if

$$\lim_{t \to \pm\infty} \varphi_\pm = \sum_{i=1}^{N} \varphi_i$$

then $\mathcal{E}_\varphi = \{E_i\}$ and $\mathcal{E}_\Phi = \{\mathcal{E}_\varphi\}$ respectively.

With the concept of a spectrum, we can make the following definition:

An equation has the strong particle scattering property if

(a) it has a particle interpretation

(b) it has a particle space Φ which satisfies

i) The spectrum of Φ is discrete and finite

ii) If $\varphi \in \Phi$,

$$\lim_{t \to -\infty} \varphi = \sum_{i=1}^{N} \varphi_i \quad ,$$

and φ_i is any elementary solution whose rest energy is in the spectrum of Φ_1 and such that $\lim_{t \to -\infty} \varphi_j(x)\, \varphi = 0$, then there is a solution $\varphi_1 \in \Phi$ such that

$$\lim_{t \to -\infty} \varphi_1 = \varphi_j + \sum_{i=1}^{N} \varphi_i = \lim_{t \to -\infty} (\varphi_j + \varphi)$$

Another possible behavior that the equation can have is the general scattering property:

Some (or all) solutions of the equation may be written asymptotically as the superposition of a sum of elementary solutions and a solution of the linearized equation. (For Eq. (III.1), the linearized equation is Eq. (III.1) with $\lambda = \beta = 0$.)

We would like to be able to prove that a particle interpretation holds for Eq. (III.1), or even that the strong particle scattering property holds. We have so far achieved only weaker results which we give below.

B. Analytical Results

Let $M_\varphi(t) = \sup_x |\varphi(x,t)|$. $M_\varphi(t)$ is the maximum value of the modulus of φ.

We shall say that φ attenuates if $M_\varphi(t) \to 0$, i.e., if for any $a > 0$, there is a t such that $M_\varphi(t) < a$.

To begin the investigation of scattering, let us note that a set of initial conditions--called Cauchy data--for Eq. (III.1) consists of giving the function at $t = 0$, and its time derivative, $\dot{\varphi}(x,t=0)$, at $t = 0$.

If $\psi(x,t)$ is a function, we shall say that ψ is Cauchy data for φ if we set

$$\varphi(x,t_o) = \psi(x,t_o)$$

$$\dot{\varphi}(x,t_o) = \psi(x,t_o)$$

We shall now prove the main result for the nonlinear Klein-Gordon equation:

Theorem: There exist nonattenuating solutions of nonlinear Klein-Gordon equations. In particular, if Eq. (III.1) has stationary solutions φ which satisfy

$$Q(\varphi_i) > \frac{1}{m} E(\varphi_i) \tag{III.6}$$

Then a solution $\varphi(x,t)$ with Cauchy data ψ does not attenuate, where

$$\psi = \sum_i \varphi_i([\Lambda_i]_\mu{}^\nu x_\nu + [a_i]_\mu) \tag{III.7}$$

with (Λ_i, a_i) a transformation in the Poincaré group such that Λ_i is close to one (i.e. $(\Lambda_i)^0{}_0 \approx 1$), while $(a_i)_\mu - (a_j)_\mu$ is large and space-like.

Lemma: Suppose

$$\Box\varphi = \varphi_{tt} - \Delta\varphi = -\varphi v(|\varphi|^2) \tag{III.8a}$$

where

$$\frac{dV(a)}{da}$$

and

$$V(a) = am^2 - \lambda a^2 + \beta a^3$$

$$m^2, \lambda, \beta > 0 \text{ and } V(a) > 0 \text{ if } a > 0 \tag{III.8b}$$

Then, if

$$M_\varphi(t) < \epsilon ,$$

$$\mathcal{Q}(x) \le \frac{1}{m} + \delta \; \mathcal{E}(x) \text{ where } \delta \to 0 \text{ as } \epsilon \to 0 \tag{III.9}$$

Proof: $V(a) \ge am^2 - \delta_1$, for $0 \le a \le \epsilon^2$ as $\epsilon \to 0$. δ_1 is simply the maximum of $|V(a) - am^2|$ in this range.

Also,

$$\mathcal{Q}(x) = 2 \operatorname{Im} (\dot{\varphi}(x)\,\bar{\varphi}(x)) \le \frac{1}{m}(|\dot{\varphi}|^2 + m^2|\varphi|^2) \le \frac{1}{m}\mathcal{E} - \frac{1}{m}\delta_3 + \frac{1}{m}\delta_1$$

where $\delta_3 = |\varphi'(x)|^2$. Then, certainly

$$\mathcal{Q}(x) \le (\frac{1}{m} + \delta)\,\mathcal{E} \quad \text{and } \delta \to 0 \text{ as } \epsilon \to 0.$$

Lemma: If, given any Cauchy data ψ for φ,

$$Q(\psi) > (\frac{1}{m} + \delta)\, E(\psi) \quad , \tag{III.10}$$

then there is an ϵ such that

$$M_{\varphi}(t) > \epsilon \quad \text{for all time} \quad .$$

Proof: Choose ϵ small enough so that for $\varphi(x) < \epsilon$

$$\mathcal{Q}(\varphi(x)) \leq (\frac{1}{m} + \delta)\, \mathcal{E}(\varphi(x)) \quad . \tag{III.11}$$

Suppose now that $M_{\varphi}(t_o) < \epsilon$, then (III.11) is true at each point, and therefore

$$Q(\varphi) \leq (\frac{1}{m} + \delta)\, E(\varphi) \quad . \tag{III.12}$$

However, this is impossible since E and Q are each conserved, and satisfy (III.10) at the initial time.

It now follows that the theorem is true if we can exhibit Cauchy data which satisfy (III.10).

One set of Cauchy data which suffices in one space dimension was given by P. D. Lax:

For any $a > 0$ such that $V(a) < m^2 a$, set

$$\varphi(x) = a \qquad 0 \leq |x| \leq R$$

$$\varphi(x) = a(R + 1 - |x|) \qquad R \leq |x| \leq R + 1$$

$$\varphi(x) = 0 \qquad R + 1 \leq |x|$$

$$\dot{\varphi}(x) = im\,\varphi(x) \quad . \tag{III.13}$$

For R sufficiently large, this satisfies Eq. (III.10) and the function $\varphi(x,t)$ which evolves from this Cauchy data does not attenuate.

Let us now turn to the nonattenuation of certain functions with Cauchy data of the form (III.7).

Recall that a stationary solution with frequency ω exists for a potential of the form (III.8b) if

$$m^2 > \omega^2 > m^2 - \lambda^2/4\beta$$

and that

$$\mathcal{E}(\varphi(x)) = 2\,V(\varphi(x)) = 2(m^2|\varphi|^2 - \lambda|\varphi|^4 + \beta|\varphi|^6)$$

$$Q = 2\omega|\varphi|^2$$

Now $m^2 - \lambda a + \beta a^2$ decreases monotonically from m^2 to its minimum.

For a stationary solution $e^{i\omega t}\theta(x)$, we have

$$\omega^2\theta^2(x=0) = V(\theta(0))$$

and $\theta(x)$ and $V(\theta(x))$ both decrease monotonically from $x = 0$. Hence

$$m^2\theta^2(x) > V(\theta(x)) > \omega^2\theta^2(x)$$

Set

$$\omega = m\delta \qquad \delta < 1$$

near

$$x = 0 \qquad \frac{V(\theta(x))}{\mathfrak{D}(\theta(x))} = m\delta < m$$

for

$$x \to \infty \qquad \frac{V(\theta(x))}{\mathfrak{D}(\theta(x))} = m/\delta > m,$$

however both V and $\mathfrak{D}$ are small.

Numerical calculations show that in fact it is possible to find stationary solutions of wave equations in one, two and three dimensions which satisfy Eq. (III.10). The argument above explains why this is plausible: near $x = 0$ where both the energy density and charge density are large the densities satisfy an inequality

$$\mathfrak{D} > \frac{1}{m}\,\mathcal{E}\,.$$

Although this condition is violated for large $|x|$, the densities there are small. Most of the contribution to the integrals for E and Q obviously comes from the region where the densities are large.

For an elementary solution, translation obviously does not change the energy. A Lorentz transformation varies the energy continuously. If E is the energy of a stationary solution, then $(\Lambda^0{}_0)E$ is the energy of a boosted solution. Charge is a scalar under Poincaré transformations.

Hence if a stationary solution $\varphi_i(x_\mu)$ satisfies (III.10), the elementary solution

$$\varphi_i(\Lambda_\mu{}^\nu x_\nu + a_\mu)$$

satisfies (III.10) providing $\Lambda_0{}^0$ is sufficiently close to 1.

Now let us consider sums of elementary solutions, of the form (III.7).

It is easy to see that if for each pair i,j, $(a_i)_\mu - (a_j)_\mu = R_\mu^{ij}$ is large and spacelike, then, when we compute the energy and charge we find

$$E(\psi) = \sum_i E(\varphi_i) + (\text{Remainder})_1$$

$$Q(\psi) = \sum_i Q(\varphi_i) + (\text{Remainder})_2$$

where the remainders consist of products $\varphi_i(x)\ \varphi_j(x)$ which go rapidly to zero as $R_\mu^{ij} R_{ij}{}^\mu \to -\infty$, because of the exponential decrease of stationary solutions.

It now follows directly that Eq. (III.10) is satisfied for the sum if it is satisfied for each φ_i, and if the remainders are sufficiently small. Hence a solution φ with Cauchy data ψ does not attenuate, as asserted.

We have now proved that some solutions which start out as superpositions of elementary solutions never attenuate. They have even a stronger property, in fact. For they have finite total energy, and there is necessarily a finite energy concentrated around the maxima of the function. So that we have in fact proved that "tangible" lumps of energy scatter into other "tangible" lumps of energy.

It would be very desirable to obtain exactly the form of the functions for $t = +\infty$.

We shall give another result which is consistent with, and even suggestive of the possibility that a particle interpretation, or something like it, holds for Eq. (III.1).

A crude variational calculation indicates that of all solutions with charge q, the stationary solution with charge q has the lowest energy.

$$E = \int [\dot{\bar{\varphi}}\,\dot{\varphi} - \bar{\varphi}\,\varphi_{ii} + V(\varphi\,\bar{\varphi})]\,dx$$

$$Q = \tfrac{1}{2}\int (\bar{\varphi}\,\dot{\varphi} - \dot{\bar{\varphi}}\,\varphi)\,dx$$

If we minimize E subject to the constraint $Q = q/2$, then using Lagrange multipliers, we minimize

$$E = \int \{[\dot{\bar{\varphi}}\,\dot{\varphi} - \bar{\varphi}\,\varphi_{ii} + V(\varphi\,\bar{\varphi})] + \lambda[\dot{\varphi}\,\bar{\varphi} - \dot{\bar{\varphi}}\,\varphi]\}dx - \lambda q$$

Varying with respect to $\dot{\bar{\varphi}}(x)$, we find

$$\dot{\varphi} - \lambda\varphi = 0 \qquad \text{(III.15a)}$$

from which we obtain

$$\varphi(x,t) = e^{\lambda t}\,\theta(x) \qquad \text{(III.15b)}$$

Varying with respect to $\bar{\varphi}$, we find

$$-\varphi_{ii} + \varphi\,\frac{\partial V(\varphi\bar{\varphi})}{\partial(\varphi\,\bar{\varphi})} + \lambda\dot{\varphi} = 0 \qquad \text{(III.16)}$$

Using Eq. (III.15), we see that Eq. (III.16) is

$$\Delta\theta = \theta\left(\frac{\partial V(\theta^2)}{\partial\theta^2} + \lambda^2\right)$$

which is precisely the equation for a stationary solution.

Of course, variation with respect to λ now requires the charge of $e^{\lambda t}\,\theta(x)$ to be exactly $q/2$, for we get, using (III.15),

$$\text{Im}\ \lambda\int(|\theta|^2)\,dx = q.$$

C. Numerical Results

Extensive numerical calculations have been performed to investigate the dynamical behavior of scattering solutions. We give a brief summary of the results.

a) We have noted that for a positive definite potential there are no real, stationary solutions. Nevertheless we have shown numerically that real functions $\theta(x,t)$ exist which dissipate their energy very slowly.

b) If two localized wave packets are made to move toward one another, they will interact, then separate, remaining more or less intact, with more or less their original shape, but with a possible change in velocity, phase, and perhaps a time delay in overall position.

c) One can plot such quantities as the position of maxima of the modulus of the function itself, or the position of the maxima of the energy density, to obtain a "scattering diagram."

d) It is obvious, though amusing in practice, to see the effects of Lorentz contraction, time dilation and the relativity of "simulaneity" on moving wave packets.

f) One can make motion pictures of the scattering process.

g) Interesting effects can be observed for different values of the parameters of a scattering process. In one case a "resonant" intermediate state was formed: During the interaction a single wave packet was formed from the two incoming packets, and it lasted a particularly long time before it separated into two outgoing packets.

IV. Other Equations

It might seem unlikely that any equation has a particle interpretation, and even more unlikely that one could establish it, given the difficulties of analyzing nonlinear partial differential equations. However, there is an equation, the Korteweg-de Vreis (K-de V) equation, which has been found to satisfy the strong particle scattering property (when restated with slightly different terminology). In addition, this equation can be solved completely, in the sense that, given a function at $t = 0$, $u(x)$, one can find a solution of the K-de V equation $u(x,t)$ which satisfies $u(x,0) = u(x)$. In fact, in some cases $u(x,t)$ can be written as an explicit rational fraction of sums of exponentials.

The K-de V equation is nonrelativistic. We will show, however, that even for relativistic equations some remarkable properties can hold.

A. Relativistic Equations

First, we ask the reader, as a challenge to his manipulative ability, to investigate the equation

$$\Box\varphi = e^{\varphi} \tag{IV.1}$$

The objective is to find the general solution of Eq. (IV.1). A general solution is a function

$$\phi_{A,B}(x,t)$$

which is a solution and which is known, given that

$$\phi(x,0) = A(x)$$

$$\dot{\phi}(x,0) = B(x) \quad .$$

We offer the following hints. Letting

$$\xi = \tfrac{1}{2}(x + t) \qquad \eta = \tfrac{1}{2}(x - t)$$

the equation becomes

$$\varphi_{\xi\eta} = e^{\varphi} \quad . \tag{IV.2}$$

Notice that this implies

$$(\varphi_{\xi\xi} - \tfrac{1}{2}\varphi_{\xi}^{2})_{\eta} = 0$$

which means that a solution of (IV.2) satisfies

$$\varphi_{\xi\xi} - \tfrac{1}{2}\varphi_{\xi}^{2} + f(\xi) = 0 \qquad \text{for arbitrary } f(\xi) \quad .$$

Making the Riccati transformation

$$\varphi_{\xi} = -2\frac{u_{\xi}}{u}$$

we obtain

$$-2\frac{u_{\xi\xi}}{u} + 2\left(\frac{u_{\xi}}{u}\right)^{2} - \tfrac{1}{2} - 2\left(\frac{u_{\xi}}{u}\right)^{2} + f(\xi) = 0$$

$$= -2u_{\xi\xi} + u(\xi)\, f(\xi)$$

which is a linear equation. The problem of obtaining the general solution $\phi_{AB}(x,t)$ is now left to the reader.

Before treating the K-de V equation, we will study another relativistic equation, the sin-Gordon equation

$$\Box\varphi = \sin\varphi \tag{IV.3a}$$

for real φ in one space dimension.

We again make the transformation

$$x \to \tfrac{1}{2}(x + t) \qquad t \to \tfrac{1}{2}(x - t) \quad ,$$

giving

$$\varphi_{xt} = \sin\varphi \tag{IV.3b}$$

We will show that Eq. (IV.3) has an infinite number of conservation laws. The method we use is a formalization, by David Wiley, of the method used to prove that the K-de V equation has an infinite number of conservation laws.

Let us consider a new function y which satisfies

$$y = \varphi + p(y_x, \epsilon) \tag{IV.4}$$

and the equation

$$y_{xt} = \sin(y)\, k(y_x, \epsilon) \tag{IV.5}$$

We suppose that y satisfies these two equations for every value of ϵ. We now take the xt derivatives of Eq. (IV.4), and apply (IV.5) and (IV.3).

$$y_{xt} = \varphi_{xt} + p'(y_x\, \epsilon)(y_{xt})_x + p''(y_x, \epsilon)\, y_{xt}\, y_{xx} \tag{IV.6a}$$

$$\sin y\, k(y_x, \epsilon) = \sin\varphi + p'(\sin y)_x\, k + p'\, \sin y\, k'(y_x, \epsilon) y_{xx}$$

$$+ p''\, \sin y\, k(y_x, \epsilon)\, y_{xx} \tag{IV.6b}$$

It is clear that this equation can be satisfied only if the coefficient of y_{xx} vanishes. That is

$$p'k' + p''k = 0 \tag{IV.7}$$

furthermore

$$\sin y\, k = \sin(y - p) + p'(\sin y)_x\, k \quad . \tag{IV.8}$$

These conditions are satisfied, with the additional conditions: p 0 as $\epsilon \to 0$; and $k \to 1$ as $\epsilon \to 0$, if for Eqs. (IV.3), (IV.4), and (IV.5) we have

$$\varphi_{xt} = \sin\varphi \tag{IV.9}$$

$$y_{xt} = \sin y \sqrt{1 - \epsilon^2 y_x^2} \tag{IV.10}$$

$$y = \varphi + \sin^{-1}(\epsilon y_x) \tag{IV.11}$$

Eq. (IV.10) is particularly interesting because it has a conservation law for each ϵ.

$$\left[\frac{(\sqrt{1-(\epsilon y_x)^2}-1)}{\epsilon^2}\right]_t = (\cos y)_x \qquad \text{(IV.12)}$$

Eq. (IV.12) is written in that form so that as $\epsilon \to 0$ it converges smoothly to

$$\tfrac{1}{2}(\varphi_x^2)_t = (\cos\varphi)_x \qquad \text{(IV.13)}$$

which is a conservation law for (IV.9)

Equations (IV.9) and (IV.10) are not equivalent under the mapping (IV.11).

Equation (IV.11) plus Eq. (IV.10) imply Eq. (IV.9):

$$\varphi_{xt} + \left[\frac{\epsilon y_{xt}}{\sqrt{1-(\epsilon y_x)^2}}\right]_x = \sin y\sqrt{1-(\epsilon y_x)^2}$$

$$\varphi_{xt} + \epsilon\cos y\, y_x = \sin y\sqrt{1-(\epsilon y_x)^2}$$

$$\varphi_{xt} = \sin(y - \sin^{-1}(\epsilon y_x)) = \sin\varphi$$

However, Eq. (IV.11) plus Eq. (IV.9) does not imply Eq. (IV.10), but rather a more general equation:

We obtain, first

$$y_{xt} - \left[\frac{\epsilon y_{xt}}{\sqrt{1-(\epsilon y_x)^2}}\right]_x = \sin(y - \sin^{-1}(\epsilon y_x))$$

$$= \sin y\sqrt{1-(\epsilon y_x)^2} - \epsilon y_x\cos y \,.$$

Now we cannot simplify the second term, so we obtain

$$\left(\sqrt{1-(\epsilon y_x)^2} - \epsilon\frac{\partial}{\partial x}\right)\left(\frac{y_{xt}}{\sqrt{1-(\epsilon y_x)^2}} - \sin y\right) = 0$$

which does not imply Eq. (IV.10), although it is satisfied if Eq. (IV.10) is true.

To obtain an infinite number of conservation laws, it is only necessary now to expand Eq. (IV.12) as a formal power series in ϵ, and

use Eq. (IV.11), also iterated formally in ϵ as a power series in φ and its derivatives. The first few conservation laws obtained in this way are

$$F_1 = \int \left(\frac{1}{2}\varphi_x^2\right) dx = \text{constant}$$

$$F_2 = \int \left(\frac{1}{2}\varphi_{xx}^2 + \frac{1}{8}\varphi_x^4\right) dx = \text{constant}$$

$$F_3 = \int \left(\frac{1}{2}\varphi_{xxx}^2 + \frac{5}{4}\varphi_x^2\,\varphi_{xx}^2 + \frac{1}{16}\varphi_x^6\right) dx = \text{constant}.$$

These conservation laws are expressed for Eq. (IV.3b). To obtain the conservation laws for the original equation (IV.3a) requires only a change of variables.

B. The Korteweg-de Vreis Equation

The K-de V equation is

$$u_t + uu_x + u_{xxx} = 0 \tag{IV.14}$$

This equation has been found to be relevant to many problems in plasma physics and fluid motion. Considering the interesting properties it has, it is noteworthy that the equation was studied for its physical properties rather than "invented" for its mathematical properties.

We shall show that the K-de V equation can be solved--that is, given $u(x)$ we can find a solution $u(x,t)$ such that $u(x) = u(x,t=0)$. Furthermore, it turns out that the solutions have the strong scattering property.

To begin with, we note that the function

$$\varphi_a(x,t) = 3a^2\ \mathrm{sech}^2\left[\frac{a}{2}x - \frac{a^3}{2}t\right] \tag{IV.15}$$

is a solution of (IV.14) which, obviously, does not change shape as it moves. The functions $\varphi_a(x,t)$ are the elementary solutions of the K-de V equation.

Historically, the theory of the K-de V equation was developed by first noticing numerically the stability of isolated or solitary wave packets. Then many conservation laws were discovered by trial and error manipulation of the equation. The fact was noted that another equation also had many conservation laws, namely:

$$v_t + v^2 v_x + v_{xxx} = 0 \tag{IV.16}$$

It was conjectured that Eqs. (IV.14) and (IV.16) both had an infinite number of conservation laws and that the two equations were related. This was proved, and then the general solution was found.

Let us relate u and v by the Riccati equation:

$$u = v^2 \pm i\sqrt{6}\, v_x \quad . \tag{IV.17}$$

Then substitution reveals

$$u_t + uu_x + u_{xxx} \equiv (2v \pm i\sqrt{6}\,\frac{\partial}{\partial x})(v_t + v^2 v_x + v_{xxx}) \; . \tag{IV.18}$$

We now apply the linearizing change of variables for the Riccati equation to v

$$v = i\sqrt{6}\,\frac{\psi_x}{\psi} \tag{IV.19}$$

and obtain for u, after translation of u by λ,

$$\psi_{xx} + \frac{1}{6}(u - \lambda)\psi = 0 \; . \tag{IV.20}$$

This is a remarkable equation, because it turns out that (IV.20) in conjunction with (IV.16), the K-de V equation, allows (IV.16) to be solved completely. It turns out to be convenient later to replace u by -6u, then the K-de V equation becomes

$$u_t - 6uu_x + u_{xxx} = 0 \tag{IV.14a}$$

and Eq. (IV.20) becomes

$$\psi_{xx} - u\,\psi = \lambda\psi \tag{IV.20a}$$

At each time t_o, Eq. (IV.20a) is a Sturm-Liouville equation for ψ, with potential $u(x,t_o)$.

It turns out that, by means of Gelfand-Levitan inverse scattering theory, u can be obtained from ψ. On the other hand, from the K-de V equation, given $u(x,0)$, we can find ψ at every time.

The Gelfand-Levitan equation is

$$K(x,y) + B(x+y) + \int_x^\infty B(y+z)\, K(x,z)dz = 0 \quad . \tag{IV.21}$$

B is a known function, evaluated from Eq. (IV.20a). K is solved for in Eq. (IV.21). Then

$$u(x) = -2\frac{d}{dx} K(x,x) \tag{IV.22}$$

The time is here a hidden parameter, (IV.21) and (IV.22) both being evaluated at each time. B is defined by

$$B(\xi) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} b(k)\, e^{ik\xi} dk + \sum_{n=1}^{N} C_n^2\, e^{-K_n\xi} \tag{IV.23}$$

where k, K_n, C_n and $b(k)$ are defined as follows:

$-K_n^2 = \lambda^{(n)}$, the nth discrete eigenvalue of Eq. (IV.20) (IV.24a)

$k^2 = \lambda_k$ represents the continuous spectrum . (IV.24b)

The C_n are normalization constants in the following sense. For a bound state wave function $\psi^{(n)}$ we require

$$\int_{-\infty}^{+\infty} \psi^{(n)2}\, dx = 1 \quad . \tag{IV.25a}$$

Then as $x \to +\infty$

$$\psi \to C_n\, e^{-K_n x} \quad . \tag{IV.25b}$$

For the continuous spectrum, since $u \to 0$ as $|x| \to \infty$, ψ is a linear combination of e^{+ikx} and e^{-ikx}. We set

$$\psi = e^{-ikx} + b(k)\, e^{+ikx} \qquad \text{as } x \to +\infty \tag{IV.26a}$$

$$\psi = a(k)\, e^{-ikx} \qquad x \to -\infty \tag{IV.26b}$$

$b(k)$ represents the reflection of a plane wave by the potential u, $a(k)$ represents the transmission, and

$$|a(k)|^2 + |b(k)|^2 = 1 \quad . \tag{IV.26c}$$

We now set out to evaluate these quantities.

There are two ways to prove the remarkable fact that the bound state eigenvalues $\lambda^{(n)}$ are constant if u evolves according to the K-de V equation.

First we write

$$u = \lambda + \frac{\psi_{xx}}{\psi}$$

substituting this into the K-deV equation yields

$$\lambda_t \psi^2 + [\psi R_x - \psi_x R]_x = 0 \qquad \text{(IV.27a)}$$

where

$$R = \psi_t + \psi_{xxx} - 3(u + \lambda)\,\psi_x \qquad \text{(IV.27b)}$$

Notice that we can now integrate (IV.27a) to give

$$0 = \int_{-\infty}^{+\infty} \lambda_t \psi^2 \, dx = \lambda_t \int \psi^2 \, dx = \lambda_t \quad . \qquad \text{(IV.28)}$$

Another way to prove that λ is constant reveals much about the structure of the K-de V equation.

Suppose that to some differential equation

$$u_t = K(u) \qquad \text{(IV.29)}$$

there is associated a linear operator L_u in some Hilbert space, i.e.,

$$u(t) \to L_{u(t)} \qquad L_{u(t)}: \; H \to H \quad . \qquad \text{(IV.30)}$$

The eigenvalues of $L_{u(t)}$ will remain constant as u evolves according to (IV.29), if the time evolution of L is effected by unitary operators, i.e., if

$$L_{u(t)} = U^{-1}(t)\, L_{u(o)}\, U(t) \qquad \text{(IV.31)}$$

where U(t) is a unitary operator for each t. Differentiating (IV.31) we obtain

$$\left(\frac{dL_u}{du}\right) u_t = \left(\frac{dL_u}{du}\right) K(u) = [L_{u(o)}, B] \qquad \text{(IV.32)}$$

where B is the generator of the unitary transformation U. If L is linear in u, as is the operator

$$L(u) = \frac{\partial^2}{\partial x^2} - u \qquad \text{(IV.33)}$$

in Eq. (IV.20a), then (IV.32) becomes

$$K(u) = c\,[L_u, B] \tag{IV.34}$$

with c a constant (equal to one for the operator in (IV.33)).

Conversely, given a generator B, and an operator L linear in u, evaluating the commutator in (IV.34) yields a function K(u) which generates a differential equation such that if

$$u_t = K(u),$$

then the eigenvalues of $L_{u(t)}$ remain constant in time.

Taking for L the linear operator $\frac{\partial^2}{\partial x^2} + \frac{1}{6}u$, and for B the operator

$$\frac{1}{12}B = \frac{\partial^3}{\partial x^3} + \frac{1}{4}u\frac{\partial}{\partial x} + \frac{1}{8}u_x\,,$$

we find

$$u_t = [L, B] = 12\left[\frac{\partial^2}{\partial x^2} + \frac{1}{6}u,\ \frac{\partial^3}{\partial x^3} + \frac{1}{4}u\frac{\partial}{\partial x} + \frac{1}{8}u_x\right]$$

$$= -uu_x - u_{xxx}\,,$$

which is the K-de V equation of Eq. (IV.14). This proves that the eigenvalues λ of L are constant, since if u(t) transforms by K-de V, then L transforms unitarily.

Using Eqs. (IV.28), one obtains the differential equation for the time development of ψ. Using the fact that ψ is a bound state solution, we find that

$$o = R = \psi_t + \psi_{xxx} - 3(u + \lambda)\,\psi_x \tag{IV.35}$$

is the equation of motion for ψ. For continuum solutions, one must be a little more detailed.

Using Eqs. (IV.24), (IV.25), and (IV.26) and noting that $u \to 0$ asymptotically, we obtain simple equations for a, b, and c with solutions

$$C_n(t) = C_n(0)\,e^{4K_n^3 t} \tag{IV.36a}$$

$$b(k,t) = b(k,o)e^{8ik^3 t} \tag{IV.36b}$$

$$a(k,t) = a(k,o) \tag{IV.36c}$$

The quantities $C_n(o)$, $b(k,o)$ and $a(k,o)$ are all obtained by solving Eq. (IV.20a) with the initial data $u(x) = u(x,o)$.

From this information a complete analysis of the properties of the K-de V equation can be made. One result is that $u(x,o)$ is a potential for which $b(k,o) = 0$ if and only if $u(x,t)_1$ as $t \to \pm\infty$ is composed only of solitons, i.e. a superposition of elementary solutions of the form (IV.15).

Moreover, if $b = 0$, then it follows from the Gelfand-Levitan equation that $u(x,t)$ is a rational fraction of exponentials.

As an example, consider the solitary wave

$$u(x,t) = -2 \operatorname{sech}^2 (x - 4t)$$

then

$$u(x,o) = -2 \operatorname{sech}^2 x \quad .$$

It turns out that the problem

$$-\nabla^2 \psi + (2 \operatorname{sech}^2 (x))\psi = \lambda\psi$$

can be solved exactly, and it has one eigenvalue.

$$K_1 = 1$$

$$C_1(0) = \sqrt{2}$$

$$b(k,0) = 0 \quad .$$

The solitary wave, as we noted, is reflectionless. The Gelfand-Levitan equation becomes

$$K(x,y) + 2\, e^{8t-(x+y)} + 2e^{8t-y} \int_x^\infty K(x,z)\, e^{-z} dz = 0 \ .$$

Let $K(x,y) = L(x)e^{-y}$ as an ansatz, then

$$L(x) + 2e^{8t-x} + 2e^{8t} L(x) \int_x^\infty e^{-2z}\, dz = 0$$

$$L(x) = \frac{-2\, e^{8t-x}}{1 + 2e^{8t}\, \frac{e^{-2x}}{2}}$$

$$K(x,x) = \frac{-2e^{-xx}}{e^{-8t} + e^{-2x}} = \frac{-2}{e^{2x-8t}+1}$$

$$u = -2\left[\frac{(-2)(2e^{2x-8t})}{(e^{2x-8t}+1)^2}\right] = -2\left(\frac{2}{e^{x-4t} + e^{-x+4t}}\right)^2$$

$$= -2 \operatorname{sech}^2 (x - 4t)$$

With the initial condition

$$u(x,0) = -6 \operatorname{sech}^2 x$$

there are two distinct eigenvalues

$$K_1 = 1, \quad K_2 = 2$$

and the exact solution turns out to be

$$u(x,t) = -12\left[\frac{3 + 4 \cosh (2x-8t) + \cosh(4x-64t)}{[3 \cosh (x-28t) + \cosh(3x-36t)]^2}\right].$$

This solution exhibits the following properties, which are typical of solutions of the K-de V equation

a) As $t \to +\infty$ or $-\infty$ u describes two solitons with velocities $v_1 = (2K_1)^2$ and $v_2 = (2K_2)^2$.

b) The two solitons as $t \to -\infty$ are positioned so that the faster is approaching the slower, and as $t \to +\infty$ so that the two solitons are diverging.

c) The solitons are displaced from the position they would have occupied if they had not interacted.

V. Conclusions and Summary

For the K-de V equation a strong particle scattering property holds. This equation is dissimilar to the equations of particle physics in the following respects: the solutions are real, the equation has high derivatives in x, and is nonrelativistic, with only one space dimension.

For the sin-Gordon equation, which is real and one dimensional but relativistic, we have proved the existence of an infinite number of conservation laws.

For the nonlinear Klein-Gordon equation, we have shown that solutions exist, which do not attenuate, in one, two, and three dimensions, but we have not proved either a particle interpretation, or an infinite number of conservation laws.

No equations are known which have a finite number of conservation laws and a particle interpretation.

If a strong particle scattering property held for a relativistic, realistic equation, it would give rise to a mathematical model for the relativistically invariant dynamics of classical extended particles.

References:

Part I was prepared from the book, Harold T. Davis, Introduction to Nonlinear Differential and Integral Equations (Dover, 1962). This is an excellent introduction to the theory of nonlinear equations. It is particularly well suited to the interests of physicists, because of the scope and the emphasis on obtaining explicit properties of the solutions.

Parts II and III on the nonlinear Klein-Gordon report on research, most of which has not been previously published. A description of the theorem that equations do not attenuate is given in: Eric H. Roffman, "Localized Solutions of Nonlinear Wave Equations," Bulletin of the American Mathematical Society 70, 76 (1970).

Some interesting work has been done on the Yang-Mills field by T. T. Wu and C. N. Yang. It has been shown that some localized solutions of the differential equations exist. One particularly noteworthy property of these solutions is that space-time and internal symmetry indices are intertwined--the solution does not factor into an isotopic vector and an isotopic spin independent function of space-time.

The system of equations consisting of the Dirac electron field coupled to the electromagnetic field has also been studied.

H. Wakano has showed that under certain approximations a localized solution exists. In unpublished work, I have shown that (a) under the same approximations as Wakano makes there are solutions with one, two, and three nodes; and probably one solution with n nodes, for each n. (b) The corrections to Wakano's equations can be calculated. They consist of terms with increasing angular asymmetry. There is an infinite series of such terms, and it is unlikely

that the series terminates. (c) There is a partial differential equation, which has a solution if and only if there is a stationary solution which has an identically vanishing vector potential. M. Wakano, Prog. Theor. Phys. 35, 1117 (1966).

Part IV: The first report on the interesting properties of the Korteweg-de Vreis equation appeared in N. J. Zabusky and M. D. Kruskal, Phys. Rev. Letters 15, 240 (1965). The method for solving the equation was given in C. S. Gardner, J. M. Green, M. D. Kruskal, and R. M. Miura, Phys. Rev. Letters 19, 1095 (1967). A series of articles on the K-de V equation is being published by Gardner, Green, Kruskal, Miura, and Su. Lax's method of attacking the K-de V equation generates many equations with similar properties. P. D. Lax, Comm. in Pure and Appl. Math. 21, 467 (1968). The method for finding an infinite number of solutions of the sin-Gordon equation, and integrating the exponential equation is given by Kruskal in unpublished papers, and the method we give here is due to D. Wiley, in another unpublished paper.

For the reader's convenience, we list below some references on the inverse scattering problem. N. Levinson, Phys. Rev. 89, 755 (1953); I. M. Gelfand and B. M. Levitan, A. M.S. Translations Ser. 2, 1, 253 (1955); I. Kay and H. E. Moses, N.C. Ser. X, 3, 276 (1956); I. Kay and H. E. Moses, J. Appl. Phys. 27, 1503 (1956).

III. Complex Analytic Varieties

COMPLEX ANALYTIC VARIETIES†

R. C. Gunning
Department of Mathematics
Princeton University
Princeton, New Jersey

I am rather hesitant to give a series of mathematical lectures to an audience consisting principally of physicists. It is not that physicists know too little mathematics, but on the contrary that they know too much mathematics, or more precisely that they know too well exactly which mathematical problems they want solved. They want to know the holomorphy envelope of a particular domain reeking with physical significance, or the nature of the singularities of a particular Feynman integral of phenomenal importance. Consequently they frequently feel a certain lack of rapport with mathematicians, who persist in backing off to look at more general problems or broader classes of problems, and who often prove quite useless as aides for solving a particular problem. These differing viewpoints seem part of the nature of things, or we would not have evolved separately into mathematicians and physicists over the past few centuries; and I shall not attempt to reverse this possibly irreversible consequence of the specialization forced upon us by the increase of knowledge but not of intellectual capacity.

My aim here is to give a survey of what might be called the local-geometrical aspect of the theory of functions of several complex variables. There has been a good deal of mathematical activity in this area in the recent and the not so recent past, leading to some quite deep and some quite surprising results. Some of the purely mathematical advances have been achieved by physicists, such as Federbush, Pham, and others, who were led to this work through the study of Feynman integrals. I cannot attempt to discuss the relevance of these results to physics, since I do not know enough physics to do so; and I shall not pretend that these lectures will be directly and immediately relevant to current work in physics. However it may be

†Presented at the INSTITUTE FOR THEORETICAL PHYSICS, University of Colorado, Summer 1969.

of some interest and even of some use to physicists to see a general survey of this field, avoiding for the most part detailed proofs or technical complications that do not shed much light on the present picture.

I. Complex Analytic Functions

To begin this survey, it may be in order to review briefly the elementary properties of analytic functions of several complex variables; to a great extent these properties are direct extensions of the familiar properties of analytic functions of a single complex variable.

Consider the set $\mathbb{C}^n$ of n-tuples of complex numbers, points of which will be denoted by $z = (z_1, \ldots, z_n)$ where $z_j \in \mathbb{C}$; this set will be identified as a topological space with the ordinary Euclidean space $\mathbb{R}^{2n}$ of dimension 2n. A complex-valued function f(z) in an open subset $U \subseteq \mathbb{C}^n$ is called complex analytic (or holomorphic) in U if in some open neighborhood of any point $a = (a_1, \ldots, a_n) \in U$ the values of the function are given by a convergent multiple power series

$$f(z_1, \ldots, z_n) = \sum_{\nu_1, \ldots \nu_n = 0}^{\infty} c_{\nu_1 \ldots \nu_n} (z_1 - a_1)^{\nu_1} \ldots (z_n - a_n)^{\nu_n} . \tag{I.1}$$

It is familiar from elementary analysis that such a series is absolutely and uniformly convergent in any suitably small polydisc

$$\Delta_a(\varepsilon) = \left\{ (z_1, \ldots, z_n) \in \mathbb{C}^n \;\middle|\; |z_j - a_j| < \varepsilon_j \text{ for } j = 1, \ldots, n \right\} \tag{I.2}$$

of center $a = (a_1, \ldots, a_n)$ and polyradius $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)$ with $\varepsilon_j > 0$; so the order in which the series is summed is quite immaterial. Note that when all the coordinates z_j except z_k are given the values a_j, the series expansion reduces to

$$f(a_1, \ldots, a_{k-1}, z_k, a_{k+1}, \ldots, a_n) = \sum_{\nu_k = 0}^{\infty} c_{0 \ldots 0 \nu_k 0 \ldots 0} (z_k - a_k)^{\nu_k} ,$$

so that $f(a_1, \ldots, a_{k-1}, z_k, a_{k+1}, \ldots, a_n)$ is a complex analytic function of the single complex variable z_k in the usual sense; that is to say, a complex analytic function of several complex variables is analytic in each variable separately. Consequently many of the familiar results from the theory of functions of a single complex variable can be applied quite directly. For instance, the Cauchy integral formula

can be applied to each variable separately, to obtain an iterated integral formula

$$f(z_1,\dots,z_n) = \frac{1}{(2\pi i)^n} \int_{|\zeta_1-a_1|=\epsilon_1} d\zeta_1 \cdots \int_{|\zeta_n-a_n|=\epsilon_n} d\zeta_n \frac{f(\zeta_1,\dots,\zeta_n)}{(\zeta_1-z_1)\cdots(\zeta_n-z_n)}$$

valid for all points $z \in \Delta_a(\epsilon)$ whenever $f(z)$ is complex analytic in an open neighborhood of the closure of the polydisc $\Delta_a(\epsilon)$. Since the function $f(z)$ is continuous, this iterated integral is equivalent to the multiple integral

$$f(z_1,\dots,z_n) = \frac{1}{(2\pi i)^n} \int_{|\zeta_j-a_j|=\epsilon_j} \frac{f(\zeta_1,\dots,\zeta_n)}{(\zeta_1-z_1)\cdots(\zeta_n-z_n)}\, d\zeta_1 \cdots d\zeta_n \,, \tag{I.3}$$

which is the several complex variable form of the Cauchy integral theorem. It should be pointed out that this integral is not over the full boundary of the polydisc $\Delta_a(\epsilon)$ when $n \geq 2$, so that this formula is not an exact analogue of the one variable Cauchy integral formula; indeed there is no exact analogue of the classical formula when $n \geq 2$, but rather there are a number of integral formulas embodying various aspects of the classical formula.

Note that in deriving the integral formula (I.3) it is not really necessary to require that the function $f(z)$ be a complex analytic function of n variables; it is sufficient merely to assume that $f(z)$ is analytic in each variable separately in a neighborhood of $\Delta_a(\epsilon)$, and is smooth enough (say continuous or Lebesgue integrable) that the above iterated integral is equivalent to a multiple integral. Having obtained formula (I.3), however, note also that

$$\frac{1}{(\zeta_1-z_1)\cdots(\zeta_n-z_n)} = \sum_{\nu_1,\dots,\nu_n=0}^{\infty} \frac{(z_1-a_1)^{\nu_1}\cdots(z_n-a_n)^{\nu_n}}{(\zeta_1-a_1)^{\nu_1+1}\cdots(\zeta_n-a_n)^{\nu_n+1}} \,,$$

where this series is absolutely and uniformly convergent for $|\zeta_n| = \epsilon_n$ and for any fixed point $z \in \Delta_a(\epsilon)$; and upon substituting this series in (I.3) and integrating term by term, there results a multiple power series expansion of the form (I.1). Consequently, whenever $f(z)$ is analytic in each variable separately in an open neighborhood of a polydisc $\Delta_a(\epsilon)$ and is smooth enough, it is an analytic function of n variables in $\Delta_a(\epsilon)$, and indeed has a series expansion (I.1) converging

throughout $\Delta_a(\epsilon)$. Actually the smoothness assumption is also unnecessary; any function analytic in each variable separately is complex analytic as a function of n variables. This result, first due to Hartogs, is surprisingly much harder to prove, though.

If f(z) is a complex-valued function of one complex variable $z = x + iy$ which is differentiable in the underlying real coordinates x,y, the classical Cauchy-Riemann criterion for f(z) to be complex analytic is that

$$\frac{\partial f}{\partial \bar{z}} = 0 \quad \text{where} \quad \frac{\partial}{\partial \bar{z}} = \tfrac{1}{2}\left(\frac{\partial}{\partial x} + i\,\frac{\partial}{\partial y}\right) \quad .$$

Consequently a function $f(z_1,\ldots,z_n)$ of n complex variables $z_j = x_j + iy_j$, which is differentiable in the underlying real coordinates x_j, y_j, is complex analytic if and only if

$$\frac{\partial f}{\partial \bar{z}_j} = 0 \quad \text{for } j = 1,\ldots,n \ .$$

One of the fundamental properties of analytic functions of one complex variable, an immediate consequence of the Cauchy-Riemann conditions, is that the composition of two analytic functions is again analytic. For several complex variables, the analogous condition involves a complex analytic mapping $G: \mathbb{C}^n \to \mathbb{C}^n$ defined by an n-tuple of complex analytic functions $g_j(z)$; such a mapping can be viewed as a coordinate change in $\mathbb{C}^n$, introducing new coordinates $w_j = g_j(z)$ for $j = 1,\ldots,n$. Then if $f(w_1,\ldots,w_n)$ is a complex analytic function, the composition $f(G(z)) = f(g_1(z),\ldots,g_n(z))$ is also complex analytic; for by the chain rule

$$\frac{\partial f(G(z))}{\partial \bar{z}_j} = \sum_{k=1}^{n} \left(\frac{\partial f}{\partial w_k}\,\frac{\partial g_k}{\partial \bar{z}_j} + \frac{\partial f}{\partial \bar{w}_k}\,\frac{\partial \bar{g}_k}{\partial \bar{z}_j}\right) = 0,$$

since the functions $f, g_1, \ldots, g_k$ are all analytic. This result is of course purely local, so the function f(w) and the mapping G(z) need only be defined locally.

The set of zeros of a non-constant analytic function of one complex variable consists of a discrete set of points in the domain of analyticity of the function. This is no longer true for functions of more than one complex variable; the coordinate function z_n is a complex analytic function in $\mathbb{C}^n$ whose zero locus is the linear subspace $z_n = 0$, which is equivalent to the full space $\mathbb{C}^{n-1}$ of n-1 complex variables. For a general analytic function of several complex

variables, the zero locus can be quite complicated indeed; the aim of the present lectures is to provide a survey of the properties of such sets. An easy preliminary property is that a function f(z) analytic in an open connected subset $U \subseteq \mathbb{C}^n$ vanishes on an open subset of U if and only if f(z) vanishes identically on U. To see this, let E be the interior of the set of points at which $f(z) = 0$, so that E is a non-empty open subset of U. To see that $E = U$, which is of course the desired result, it is only necessary to show that E has no boundary points inside U. Suppose that $a \in U$ lies on the boundary of E; and select a polydisc $\Delta_a(\epsilon) \subseteq U$. There is a point $b \in \Delta_a(\frac{1}{2}\epsilon) \cap E$, since a is a boundary point of E; and evidently $\Delta_b(\frac{1}{2}\epsilon) \subseteq U$. The function f(z) has a multiple power series expansion centered at b and converging throughout $\Delta_b(\frac{1}{2}\epsilon)$; but since $b \in E$, this series necessarily vanishes identically, hence the function f(z) is identically zero in $\Delta_b(\frac{1}{2}\epsilon)$. However $a \in \Delta_b(\frac{1}{2}\epsilon)$, hence a is an interior point of E, contradicting the assumption that a was a boundary point of E and thereby concluding the proof.

II. Analytic Hypersurfaces in Standard Form

A subset V of an open domain $U \subseteq \mathbb{C}^n$ is said to be an analytic hypersurface of U if in some open neighborhood of each point of U the subset V is the set of zeros of a complex analytic function. Such a subset is necessarily relatively closed in U; and, except for the trivial case in which the defining function vanishes identically, V is a proper subset of U containing no interior points. The aim of the present section is to derive a useful general description of an analytic hypersurface in an open neighborhood of a point on it; after changing coordinates if necessary, it can be assumed that the point of interest is the origin.

Consider then a complex analytic function f(z) in a polydisc $\Delta(\epsilon) = \{(z_1, \ldots, z_n) \mid |z_j| < \epsilon_j\}$ centered at the origin in $\mathbb{C}^n$, and the hypersurface

$$V = \{z \in \Delta(\epsilon) \mid f(z) = 0\} .$$

Suppose that the function f(z) is regular in the variable z_n, in the sense that $f(0, \ldots, 0, z_n)$ considered as an analytic function of the single complex variable z_n does not vanish identically. This is not a serious restriction, since any analytic function can be made regular in this sense by a suitable linear change of coordinates in $\mathbb{C}^n$, provided that the function does not vanish identically; for select any point $b \in \Delta(\epsilon)$ for which $f(b) \neq 0$, and choose coordinates such that $b = (0, \ldots, 0, 1)$. Suppose further that the hypersurface V actually passes through the origin. The function $f(0, \ldots, 0, z_n)$ as a function

of the single complex variable z_n will then have a zero of some order $r > 0$ at the origin $z_n = 0$. The zeros of the function $f(0,\dots,0,z_n)$ are isolated, so by choosing the radius ϵ_n sufficiently small it can further be assumed that $f(0,\dots,0,z_n) \neq 0$ for $0 < |z_n| \leqq \epsilon_n$, with $f(0,\dots,0,z_n)$ being defined and analytic in an open neighborhood of this closed disc of radius ϵ_n centered at the origin $z_n = 0$. Thus there is a constant $\delta > 0$ such that $|f(0,\dots,0,z_n)| \geqq \delta$ whenever $|z_n| = \epsilon_n$. With these preparations out of the way, it is quite easy to give a rough general description of the hypersurface V.

Since the function $f(z)$ is continuous, after choosing the constants $\epsilon_1,\dots,\epsilon_{n-1}$ sufficiently small it can be assumed that

$$|f(z_1,\dots,z_n) - f(0,\dots,0,z_n)| < \delta \tag{II.1}$$

whenever

$$|z_1| < \epsilon_1,\dots,|z_{n-1}| < \epsilon_{n-1}, |z_n| = \epsilon_n.$$

Now consider the polydisc

$$\Delta'(\epsilon') = \left\{ z' = (z_1,\dots,z_{n-1}) \in \mathbb{C}^{n-1} \mid |z_j| < \epsilon_j \text{ for } j = 1,\dots,n-1 \right\} ;$$

and for each fixed point $z' \in \Delta'(\epsilon')$, consider the function $f(z_1,\dots,z_{n-1},z_n)$ as a function of the complex variable z_n alone. Since $f(0,\dots,0,z_n)$ has zeros of total order r in the disc $|z_n| < \epsilon_n$, and since the function $f(z_1,\dots,z_{n-1},z_n)$ satisfies (II.1), it follows from Rouché's theorem that $f(z_1,\dots,z_{n-1},z_n)$ has zeros of total order r in the disc $|z_n| < \epsilon_n$, and of course no zeros on the boundary of that disc. Label these r points $\varphi_1(z'),\dots,\varphi_r(z')$ in some order, noting that they depend on the choice of the point $z' \in \Delta'(\epsilon')$, and taking multiplicities into account by repeating a multiple zero the correct number of times. Of course the labeling is quite random, so the functions $\varphi_j(z')$ need not even be continuous when considered as functions of $z' \in \Delta'(\epsilon')$; however if this random element is eliminated by considering symmetric polynomials in the functions $\varphi_j(z')$, the resulting function is well behaved, indeed is analytic in $\Delta'(\epsilon')$. To see this, note that it follows from the usual Cauchy integral formula for functions of one complex variable that

$$\frac{1}{2\pi i} \int_{|\zeta_n| = \epsilon_n} \frac{\zeta_n^{\nu} \frac{\partial}{\partial \zeta_n} f(z_1,\dots,z_{n-1},\zeta_n)}{f(z_1,\dots,z_{n-1},\zeta_n)} \, d\zeta_n = \sum_{j=1}^{r} \varphi_j(z_1,\dots,z_{n-1})^{\nu} ,$$

while the above integral is evidently a complex analytic function in $\Delta'(\epsilon')$; and since any symmetric polynomial in the functions $\varphi_j(z')$ can be expressed as a polynomial in these power sums, any such expression is complex analytic in $\Delta'(\epsilon')$. In particular, introduce the function

$$p(z_1,\ldots,z_n) = \prod_{j=1}^{r} \left(z_n - \varphi_j(z_1,\ldots,z_{n-1})\right)$$

$$= z_n^r + c_1(z')\, z_n^{r-1} + \ldots + c_r(z') \quad ,$$

a polynomial of degree r in the variable z_n with coefficients which are the elementary symmetric functions of the $\varphi_j(z')$ and hence which are analytic in $\Delta'(\epsilon')$. This polynomial thus has leading coefficient 1, and the remaining coefficients are complex analytic functions in $\Delta'(\epsilon')$ which vanish at the origin; such a polynomial is called a Weierstrass polynomial at the origin in the variable z_n. By construction, for any fixed point $z' \in \Delta'(\epsilon')$ this Weierstrass polynomial has the same zeros in z_n as the original function f(z); consequently there always exists a Weierstrass polynomial defining the same analytic hypersurface as the given analytic function f(z).

Consider the special case in which the function f(z) is regular in the variable z_n at the origin and $f(0,\ldots,0,z_n)$ has a zero of order 1 at $z_n = 0$; this is equivalent to saying that

$$f(0,\ldots,0) = 0, \quad \frac{\partial f}{\partial z_n}(0,\ldots,0) \neq 0 \ . \qquad \text{(II.2)}$$

In this case the associated Weierstrass polynomial has the special form $p(z) = z_n - \varphi_1(z_1,\ldots,z_{n-1})$ where $\varphi_1(z')$ is analytic in $\Delta'(\epsilon')$; so the hypersurface V defined by either the function f(z) or the Weierstrass polynomial p(z) can be described very simply by the parametric equation $z_n = \varphi_1(z_1,\ldots,z_{n-1})$ for $z' \in \Delta'(\epsilon')$. In this case, it is further possible to introduce a new coordinate system $(w_1,\ldots,w_n)$ in some open neighborhood of the origin defined by

$$w_1 = z_1,\ldots,w_{n-1} = z_{n-1} \ , \quad w_n = z_n - \varphi_1(z_1,\ldots,z_{n-1}) \ ;$$

it is easily verified that the Jacobian of this change of variables is non-singular. For the new coordinate system, the hypersurface is merely the coordinate hyperplane $w_n = 0$. In general, a point on an analytic hypersurface V is called a regular point if it is possible to choose coordinates in an open neighborhood of a in $\mathbb{C}^n$ such that

locally V is a coordinate hyperplane; and a hypersurface V is called a complex manifold of complex dimension n - 1 if it is regular at each point. If a subvariety V is defined by an analytic function f(z) satisfying (II.5), it then follows that V is regular at the origin, that is, that the origin is a regular point of V; and of course conversely whenever V is regular at the origin, it can be defined by an analytic function f(z) satisfying (II.2) in some set of coordinates. It is obvious that this problem of the correct choice of a set of coordinates for verifying condition (II.2) can be avoided by noting that an analytic hypersurface V is regular at the origin in $\mathbb{C}^n$ if and only if it can be defined by an analytic function f(z) such that

$$f(0,\ldots,0) = 0 \qquad \frac{\partial f}{\partial z_j}(0,\ldots,0) \neq 0 \text{ for some } j. \quad \text{(II.3)}$$

The points of V which are not regular points are called singular points. (It should be noted that the preceding observations include a proof of the complex analytic form of the implicit function theorem; for if f(z) is an analytic function satisfying (II.2), the zero locus of f(z) is described parametrically by $z_n = \varphi_1(z_1,\ldots,z_{n-1})$, hence $f(z_1,\ldots,z_{n-1}, \varphi_1(z_1,\ldots,z_{n-1})) \equiv 0$, and $\varphi_1(0,\ldots,0) = 0$.)

Returning once more to the general case, the discriminant of the Weierstrass polynomial p(z) is the function defined by

$$\delta(z') = \prod_{\substack{\mu \neq \nu \\ \mu,\nu=1,\ldots,r}} \left(\varphi_\mu(z') - \varphi_\nu(z')\right) \quad ; \qquad \text{(II.4)}$$

this is also a symmetric polynomial in the $\varphi_j(z')$, hence it is a complex analytic function of z' in $\Delta'(\epsilon')$. It is a familiar algebraic result that this discriminant does not vanish identically if the polynomial p(z) has no multiple factors. (If it is possible to write the polynomial p(z) as a product of polynomials in z_n of lower degrees with coefficients analytic in $(z_1,\ldots,z_{n-1})$, say $p(z) = p_1(z)\ldots p_s(z)$, the polynomials $p_j(z)$ are called factors. If two factors are the same, the polynomial is said to have a multiple factor. It is obvious that any repeated factor can be dropped, and the resulting polynomial will define the same analytic hypersurface V.) Clearly there is no loss of generality in considering only polynomials p(z) with no multiple factors; hence it can be assumed that the discriminant $\delta(z')$ does not vanish identically in $\Delta'(\epsilon')$. The set

$$D = \{z' = (z_1,\ldots,z_{n-1}) \in \Delta'(\epsilon') \mid \delta(z') = 0\} \qquad \text{(II.5)}$$

is then a complex analytic hypersurface in the polydisc $\Delta'(\epsilon')$ in the space of n - 1 complex variables. Now consider the natural projection from the polydisc $\Delta(\epsilon) \subseteq \mathbb{C}^n$ onto the polydisc $\Delta'(\epsilon') \subseteq \mathbb{C}^{n-1}$ which takes a point $z = (z_1, \ldots, z_n) \in \Delta(\epsilon)$ to the point $z' = (z_1, \ldots, z_{n-1}) \in \Delta'(\epsilon')$. Under this projection the hypersurface $V \subset \Delta(\epsilon)$ is mapped onto the entire polydisc $\Delta'(\epsilon')$. For any fixed point $a' = (a_1, \ldots, a_{n-1}) \in \Delta'(\epsilon') - D$ there are exactly r distinct points in the hypersurface V mapping onto the point a', namely the points with coordinates $(a_1, \ldots, a_{n-1}, \varphi_j(a'))$ for $j = 1, \ldots, r$. Each of these points is a point of order one for the defining function, hence is a regular point of the hypersurface; and as noted in the preceding paragraph, it is possible to select complex analytic functions $\varphi_j(z')$ in some open neighborhood of $a' \in \Delta'(\epsilon')$ such that $\varphi_j(a')$ have the specified values and that in a neighborhood of the point $(a_1, \ldots, a_{n-1}, \varphi_j(a'))$ the hypersurface is described parametrically by the equation $z_n = \varphi_j(z')$. Thus over a small enough open neighborhood of a' the hypersurface consists of r sheets, each of which is mapped onto the neighborhood of a' homeomorphically by the projection; equivalently, the projection exhibits the points of V lying over $\Delta'(\epsilon') - D$ as an r-sheeted covering space of $\Delta'(\epsilon') - D$. It is evident that as $z' \in \Delta'(\epsilon') - D$ approaches a point of D, some of these distinct sheets approach coincidence; in particular as z' approaches the origin, all the sheets come together at the origin in $\mathbb{C}^n$.

The points of V lying over D will be called the branch points of this covering, and the set of branch points will be denoted by $B \subset V$; the set B is thus defined by the pair of analytic equations

$$B = \{ z \in \Delta(\epsilon) \mid f(z) = \delta(z) = 0 \} , \qquad \text{(II.5)}$$

where of course $\delta(z_1, \ldots, z_n) = \delta(z_1, \ldots, z_{n-1})$ is independent of the variable z_n. Clearly all the points of V - B are regular points of the hypersurface, so that V - B is a complex analytic submanifold of dimension n - 1 of the domain $(\Delta'(\epsilon') - D) \times (|z_n| < \epsilon_n) \subseteq \Delta(\epsilon) \subseteq \mathbb{C}^n$. It is possible that some further points of the set B may also be regular points of the hypersurface V, but appear as branch points only because of the particular choice of projection in this representation of the hypersurface; for the set B can be viewed as being defined by the equations $f(z) = \frac{\partial f(z)}{\partial z_n} = 0$, while the singular points of V are possibly the points given by the further equations $f(z) = \frac{\partial f(z)}{\partial z_1} = \ldots = \frac{\partial f(z)}{\partial z_n} = 0$. Examples of this will be given later.

Before discussing this situation further, a brief digression to establish another general property of analytic functions of several complex variables is in order. Recall from classical function theory

that a bounded analytic function in a punctured disc $0 < |z| < \epsilon$ in the complex plane can always be extended analytically to a function in the entire disc $|z| < \epsilon$; this is the Riemann removable singularities theorem. An equivalent formulation of course is that a function $f(z)$ bounded and analytic in the complement of an analytic subvariety of a disc extends analytically to a function in the entire disc; and this formulation extends to functions of several complex variables. To see this, consider the analytic hypersurface $V \subset \Delta(\epsilon)$ as described above; and suppose that $g(z_1, \ldots, z_n)$ is a bounded analytic function in the open set $\Delta(\epsilon) - V$. Note that the hypersurface V avoids an open neighborhood of any boundary point $(z_1, \ldots, z_n)$ with $|z_1| < \epsilon_1, \ldots, |z_{n-1}| < \epsilon_{n-1}$, $|z_n| = \epsilon_n$; so the function $g(z)$ remains analytic in the region $|z_1| < \epsilon_1, \ldots, |z_{n-1}| < \epsilon_{n-1}$, $\epsilon_n^* \leq |z_n| < \epsilon_n$ for some number ϵ_n^*. Consider the integral

$$h(z_1, \ldots, z_n) = \frac{1}{2\pi i} \int_{|\zeta_n| = \epsilon_n^*} \frac{g(z_1, \ldots, z_{n-1}, \zeta_n)}{\zeta_n - z_n} \, d\zeta_n ,$$

noting that this is analytic in the polydisc of radius $(\epsilon_1, \ldots, \epsilon_{n-1}, \epsilon_n^*)$; for the function $g(z_1, \ldots, z_{n-1}, \zeta_n)$ is analytic in $|z_1| < \epsilon_1, \ldots, |z_{n-1}| < \epsilon_{n-1}$ whenever $|\zeta_n| = \epsilon_n^*$. Now for any fixed point $(z_1, \ldots, z_{n-1})$, the function $g(z_1, \ldots, z_{n-1}, z_n)$ as a function of z_n alone is bounded and analytic in the complement of an analytic subvariety of the disc $|z_n| < \epsilon_n$, so by the classical theorem this function can be extended to an analytic function throughout the disc $|z_n| < \epsilon_n$; but then necessarily $g(z_1, \ldots, z_n) = h(z_1, \ldots, z_n)$, so that $h(z)$ provides the desired analytic extension.

An analytic hypersurface V in the form described above can also be viewed as the graph of a multiple valued analytic function of $n - 1$ complex variables. For consider the segment of V lying over an open neighborhood of a point $a' \in \Delta'(\epsilon') - D$ and parametrized by the equation $z_n = \varphi_1(z')$; this portion of V is just the graph of the analytic function $\varphi_1(z')$. Now the function $\varphi_1(z')$ can evidently be continued analytically along any path in $\Delta'(\epsilon') - D$ in a uniquely defined manner, and the graph is always a portion of the hypersurface V. After analytic continuation around a closed path in $\Delta'(\epsilon') - D$ beginning and ending at a', there results an analytic function in an open neighborhood of a' describing some portion of V, but not necessarily the same portion described by the function $\varphi_1(z')$; that is, after such an analytic continuation, the function $\varphi_1(z')$ may return to coincide with $\varphi_j(z)$ for some index $1 < j \leq r$. This function $\varphi_j(z')$ in turn can be continued along another closed path, and may return to coincide with yet

another $\varphi_k(z')$. There are only r possible effects of such continuation; and any continuation will have the effect of a permutation of the r function elements $\varphi_j(z')$, corresponding to a permutation of the sheets of the covering space V - B over $\Delta'(\varepsilon')$ - D. The set of all functions $\varphi_j(z')$ can be viewed thus as a single r-valued analytic function in $\Delta'(\varepsilon')$ - D, and V - B is the graph of this function.

Note that the space $\Delta'(\varepsilon')$ - D is connected, so that the functions $\varphi_j(z')$ can be continued from any point $a' \in \Delta'(\varepsilon')$ - D to any other point $b' \in \Delta'(\varepsilon')$ - D, and the full set V - B is described as above. (To see this, suppose contrariwise that $\Delta'(\varepsilon')$ - D is not connected. Consider the function $g(z')$ defined in $\Delta'(\varepsilon')$ - D by setting $g(z') = 0$ for all points z' in one connected component of $\Delta'(\varepsilon')$ - D and $g(z') = 1$ for all points z' in the remainder of $\Delta'(\varepsilon')$ - D. The resulting function is bounded and analytic in the complement of the analytic hypersurface $D \subset \Delta'(\varepsilon')$, so by the generalized Riemann removable singularities theorem extends to a complex analytic function throughout $\Delta'(\varepsilon')$; but this is clearly impossible, for the function $g(z')$ would have to vanish identically since it is zero in an open subset of $\Delta'(\varepsilon')$.) On the other hand, the space V - B need not be connected. Consider a single connected component V_1 of V - B, and suppose that over an open neighborhood of $a' \in \Delta'(\varepsilon')$ - D this component is parametrized by the functions $\varphi_1(z'), \ldots, \varphi_s(z')$ for $s \leq r$. It is evident that after any analytic continuation along a closed path in $\Delta'(\varepsilon')$ - D these functions will be merely permuted among themselves, as will be the remaining set of function elements $\varphi_{s+1}(z'), \ldots, \varphi_r(z')$; indeed, the functions $\varphi_1(z'), \ldots, \varphi_s(z')$ will be permuted transitively among themselves by the set of all possible such continuations, since V_1 is assumed to be connected. Now the function

$$p_1(z) = \prod_{j=1}^{s} (z_n - \varphi_j(z'))$$

$$= z_n^{\,s} + c_1(z_1, \ldots, z_{n-1})\, z_n^{\,s-1} + \ldots + c_s(z_1, \ldots, z_{n-1})$$

describes the set V_1, for points $z' \in \Delta'(\varepsilon')$ - D; but this is a polynomial in z_n with coefficients which are symmetric polynomials in $\varphi_1(z'), \ldots, \varphi_s(z')$, hence which are single-valued complex analytic functions in $\Delta'(\varepsilon')$ - D. These coefficients are also bounded, since $|z_n| < \varepsilon_n$; hence by the generalized Riemann removable singularities theorem they extend to analytic functions in all of $\Delta'(\varepsilon')$, and consequently the function $p_1(z)$ itself extends to a complex analytic function throughout $\Delta'(\varepsilon')$. Clearly the point set closure of V_1 in $\Delta(\varepsilon)$ is

the analytic hypersurface $\{ z \in \Delta(\epsilon) \mid p_1(z) = 0 \}$. An analytic hypersurface V in $\Delta(\epsilon)$ is said to be irreducible if it cannot be written as the union of two hypersurfaces properly contained in V. These observations show that V is irreducible precisely when V - B is connected, and that any hypersurface can be written as the union of finitely many irreducible hypersurfaces in a sufficiently small polydisc.

Finally a few words should be said about the set of all possible analytic functions defining a given hypersurface V. Suppose that V is represented as above in a polydisc $\Delta(\epsilon) \subseteq \mathbb{C}^n$, and that p(z) is the Weierstrass polynomial with no multiple factors defining that hypersurface; and consider an arbitrarily complex analytic function f(z) in an open neighborhood of this polydisc $\Delta(\epsilon)$. The Weierstrass division theorem asserts that this function f(z) can be written in a unique way in the form $f(z) = p(z)q(z) + r(z)$, where q(z), r(z) are analytic in $\Delta(\epsilon)$ and r(z) is a polynomial in z_n of order strictly less than the order r of the Weierstrass polynomial p(z). (To see that this is so, introduce the function

$$q(z) = \frac{1}{2\pi i} \int_{|\zeta_n| = \epsilon_n} \frac{f(z_1, \ldots, z_{n-1}, \zeta_n)}{p(z_1, \ldots, z_{n-1}, \zeta_n)} \frac{d\zeta_n}{\zeta_n - z_n} ,$$

noting that this is clearly a complex analytic function in $\Delta(\epsilon)$ since the denominator $p(z_1, \ldots, z_{n-1}, \zeta_n)$ is non-zero for $|z_1| < \epsilon_1, \ldots, |z_{n-1}| < \epsilon_{n-1}, |\zeta_n| = \epsilon_n$. Write this difference

$$r(z) = f(z) - p(z)\, q(z)$$

$$= \frac{1}{2\pi i} \int_{|\zeta_n| = \epsilon_n} \frac{f(z_1, \ldots, z_{n-1}, \zeta_n)}{\zeta_n - z_n} \, d\zeta_n$$

$$- \frac{1}{2\pi i} \int_{|\zeta_n| = \epsilon_n} \frac{p(z_1, \ldots, z_n)\, f(z_1, \ldots, z_{n-1}, \zeta_n)}{(\zeta_n - z_n)\, p(z_1, \ldots, z_{n-1}, \zeta_n)} \, d\zeta_n$$

$$= \frac{1}{2\pi i} \int_{|\zeta_n| = \epsilon_n} \frac{f(z_1, \ldots, z_{n-1}, \zeta_n)}{p(z_1, \ldots, z_{n-1}, \zeta_n)} \frac{[p(z_1, \ldots, z_{n-1}, \zeta_n) - p(z_1, \ldots, z_{n-1}, z_n)]}{\zeta_n - z_n} \, d\zeta_n ;$$

note that this function is analytic in $\Delta(\epsilon)$, and since the numerator in the integrand is divisible by $(\zeta_n - z_n)$ with quotient a polynomial in z_n of order strictly less than r, the entire integral r(z) is a polynomial in z_n of order strictly less than r. This proves the existence of such a representation for the function f(z), while the uniqueness is fairly evident directly.) Now suppose that the function f(z) vanishes on the hypersurface V; so that writing $f(z) - p(z)q(z) = r(z)$, the polynomial r(z) also vanishes on V. However for any fixed point $z' \in \Delta'(\epsilon') - D$ there are r points of V lying over z', hence there are r distinct zeros of the polynomial r(z), while the degree of this polynomial in z_n is strictly less than r; this can only happen if the polynomial is identically zero, so that the coefficients of the polynomial r(z) vanish throughout $\Delta'(\epsilon') - D$, and consequently $r(z) \equiv 0$. That is to say, whenever f(z) is analytic in $\Delta(\epsilon)$ and vanishes on V, then necessarily $f(z) = p(z)q(z)$ for some analytic function q(z) in $\Delta(\epsilon)$, where p(z) is the Weierstrass polynomial with no multiple factors defining the hypersurface V. In particular, if f(z) has the same order at the regular points of V as has the polynomial p(z), then $f(z) = p(z)q(z)$ where q(z) is complex analytic and non-vanishing in $\Delta(\epsilon)$.

It is clear from this that the set of singular points of the hypersurface V defined by the Weierstrass polynomial p(z) with no multiple factors is precisely the point set defined by the analytic equations

$$p(z) = \frac{\partial p(z)}{\partial z_1} = \ldots = \frac{\partial p(z)}{\partial z_n} = 0 \ ;$$

for the regular points are precisely those for which $f(z) = \partial f(z)/\partial z_j \neq 0$ for some function f(z) vanishing on V and some index j, and any such function f(z) must be of the form $f(z) = p(z)q(z)$. Further, if the hypersurface V is reducible, let $p_1(z), \ldots, p_t(z)$ be the Weierstrass polynomials defining the various components; any function vanishing on V is necessarily of the form $f(z) = q(z)p_1(z)\ldots p_t(z)$ for some analytic function q(z). Then in the obvious sense, the hypersurface V is irreducible precisely when the function f(z) defining the hypersurface is irreducible.

III. Geometry of Analytic Hypersurfaces

The simplest non-trivial analytic hypersurface is of course an irreducible analytic hypersurface in $\mathbb{C}^2$; and the local geometrical or topological properties of any such hypersurface can be described quite easily in terms of the standard form discussed in the preceding section. The natural projection from the polydisc $\Delta(\epsilon) \subseteq \mathbb{C}^2$ to the polydisc $\Delta'(\epsilon') \subseteq \mathbb{C}^1$ exhibits V - B as a connected r-sheeted covering of

$\Delta'(\varepsilon') - D$. The analytic subvariety D of the disc $\Delta'(\varepsilon')$ in the complex plane is a discrete set of points; and choosing $\varepsilon' = \varepsilon_1$ sufficiently small, it can be assumed that D is just the origin itself. Topologically there is a unique r-sheeted connected covering of the punctured disc; the covering space is again a punctured disc, and the covering projection wraps this r times around the base space. The entire space V is just this covering space, a punctured disc, with the set B added; and in this case B consists of the origin itself, so that topologically V is merely a disc.

This result can be derived another way, providing at the same time a useful additional analytic parametrization of the hypersurface V in $\mathbb{C}^2$. The mapping $t \in \mathbb{C} \to z_1 = t^r \in \mathbb{C}$ exhibits the punctured t-plane as an r-sheeted connected covering of the punctured z_1-plane, so must be topologically the same as the covering projection from $V - B$ to $\Delta'(\varepsilon') - D$; this suggests that the latter covering projection might be described in terms of this simple power mapping. To do so quite explicitly, consider one of the function elements $\varphi_1(z_1)$ describing one sheet of the hypersurface V over a neighborhood of the point $z_1 = 1$; this function element can be continued analytically along any path in the punctured disc $\Delta'(\varepsilon') - D$, and there results a multiple-valued analytic function in $\Delta'(\varepsilon') - D$ with graph $V - B$. Now $\theta(t) = \varphi_1(t^r)$ is clearly a well defined analytic function element in a neighborhood of $t = 1$, and it can be continued analytically along any path in a punctured disc in the t-plane. However when t varies in a closed path circling the origin once, $z_1 = t^r$ varies in a closed path in the z_1-plane circling the origin r times, and the continuation of $\varphi_1(z_1)$ along such a path evidently leads back to the original function $\varphi_1(z_1)$; consequently analytic continuation of $\theta(t)$ yields a single-valued complex analytic function in a punctured disc in the t-plane. Since $|\theta(t)| = |\varphi_1(t^r)| < \varepsilon_2$, it follows from the Riemann removable singularities theorem that $\theta(t)$ can be continued to an analytic function at the origin as well, and of course $\theta(0) = 0$. It is clear that the analytic hypersurface $V \subset \Delta(\varepsilon)$ can be described parametrically as the set of all points $(z_1, z_2) \in \Delta(\varepsilon)$ given by

$$z_1 = t^r, \qquad z_2 = \theta(t) \; ; \tag{III.1}$$

and the mapping $t \to (t^r, \theta(t))$ is a homeomorphism from a disc in the t-plane onto the entire hypersurface $V \subset \Delta(\varepsilon)$. Conversely any function $\theta(t)$ for which this mapping is a homeomorphism does describe a complex analytic hypersurface $V \subset \Delta(\varepsilon)$ in standard form.

Although the set V is just a disc topologically, it is imbedded in the polydisc $\Delta(\varepsilon)$ in a possibly rather complicated manner. The boundary of V, the set K at which V intersects the boundary of $\Delta(\varepsilon)$, is described parametrically by the equations

$$z_1 = \epsilon_1 e^{isr}, \qquad z_2 = \theta(\epsilon_1 e^{is}) \quad \text{for } 0 \leqq s \leqq 2\pi \ ;$$

the boundary of $\Delta(\epsilon)$ is topologically the three-dimensional sphere S^3, and $K = V \cap S^3$ is a smooth curve in S^3, that is to say, is a knot in S^3. For instance, when $r = 2$ and $\theta(t) = t^3$, this is the knot $z_1 = e^{2is}$, $z_2 = e^{3is}$, $0 \leqq s \leqq 2\pi$, lying on the torus $|z_1| = |z_2| = 1$ in the boundary S^3; this is a non-trivial knot in the three sphere, wrapping around the torus twice in one direction and three times in another. There has been a rather extensive study of the knots that arise this way and their relationship to the parametrizing functions t^r, $\theta(t)$, for the hypersurface V; indeed these knots have been taken as a basis for the classification of the singularities of hypersurfaces in $\mathbb{C}^2$. The details will not be given here, but the interested reader is referred to Refs. 8 and 9.

For analytic hypersurfaces in polydiscs in $\mathbb{C}^n$ for dimensions $n \geqq 2$ the situation is much more complicated, and relatively little is known about the topological properties of the singularities. Actually this aspect of several complex variables has only come under intensive investigation quite recently, and is presently a rather active field of research. These investigations require some further mathematical machinery, since the topological properties of higher dimensional spaces cannot be described readily in familiar and intuitive terms. Consequently the discussion here necessarily will be rather sketchy, and some of the concepts used may not be too familiar. A brief review of some of the topology required can be found in Ref. 10, and a more encyclopedic survey in Ref. 11.

Consider first the analytic hypersurface V of the entire space $\mathbb{C}^n$ defined as the set of zeros of a polynomial $f(z)$ of the form

$$f(z) = z_1^{a_1} + \ldots + z_n^{a_n} \quad \text{for integers } a_j > 0. \tag{III.2}$$

Note that $\partial f(z)/\partial z_j = a_j z_j^{a_j - 1}$ so that V is nonsingular at all of its points whenever $a_j = 1$ for some index j; and restricting attention to the interesting case in which $a_j > 1$ for all j, the hypersurface V has an isolated singularity at the origin while all other points are regular points. To consider this singularity, introduce the $(2n - 1)$ dimensional sphere

$$S^{2n-1} = \left\{ z \in \mathbb{C}^n \;\middle|\; |z_1|^2 + \ldots + |z_n|^2 = 1 \right\} \ ,$$

and let $K = V \cap S^{2n-1}$. It is easily verified that K is a real differentiable submanifold of dimension $2n - 3$ in the sphere S^{2n-1}; and since it is evident that V is topologically a cone over K, the topological nature of V at its singularity is described quite completely by the

topological nature of K. This is of course parallel to the situation in the special case n = 2 described above.

It seems easier to describe the topology of $S^{2n-1} - K$ than that of K itself. Note first of all that the mapping φ: $(S^{2n-1}-K) \times \mathbb{R}^+ \to \mathbb{C}^n - V$, which takes any point $(z_1, \ldots, z_n) \in S^{2n-1} - K$ and any positive real number $x \in \mathbb{R}^+$ into the point $\varphi(z_1, \ldots, z_n; x) = (x^{1/a_1} z_1, \ldots, x^{1/a_n} z_n)$ with $x^{1/a_j} > 0$, is a homeomorphism; so since $\mathbb{R}^+$ is really topologically trivial, it suffices to describe the set $\mathbb{C}^n - V$, the complement of the hypersurface V in $\mathbb{C}^n$. Now consider the mapping f: $\mathbb{C}^n \to \mathbb{C}$ defined by the polynomial function (III.2); the set $f^{-1}(0)$ is just the hypersurface V itself, while the set $f^{-1}(\zeta)$ for any $\zeta \in \mathbb{C} - O$ is the hypersurface V_ζ defined by

$$V_\zeta = \left\{ z \in \mathbb{C}^n \mid z_1^{a_1} + \ldots + z_n^{a_n} = \zeta \right\} .$$

It is clear that the hypersurface V_ζ is a regular complex analytic submanifold of $\mathbb{C}^n$ whenever $\zeta \neq 0$. It is also clear that these submanifolds are analytically equivalent; for given any complex number $\xi \in \mathbb{C} - O$, the mapping φ_ξ: $\mathbb{C}^n \to \mathbb{C}^n$ defined by $\varphi_\xi(z_1, \ldots, z_n) = (\xi^{1/a_1} z_1, \ldots, \xi^{1/a_n} z_n)$ is a complex analytic homeomorphism of $\mathbb{C}^n$ to itself taking V_ζ onto $V_{\xi\zeta}$. Indeed restricting ξ to the disc $\Delta_1(1)$ in which the roots ξ^{1/a_j} can be chosen as single-valued complex analytic functions, the mapping which takes any point $\xi \in \Delta_1(1)$ and any point $(z_1, \ldots, z_n) \in V_\zeta$ into the point $\varphi_\xi(\zeta) \in \mathbb{C}^n$ is obviously a complex analytic homeomorphism from the product manifold $\Delta_1(1) \times V_\zeta$ onto the inverse image under f of a disc $\Delta_\zeta(|\zeta|)$ around the point $\zeta \in \mathbb{C}$; that is to say, the restricted mapping f: $\mathbb{C}^n - V \to \mathbb{C} - O$ has the property that the inverse image of a sufficiently small disc $\Delta_\zeta(\varepsilon)$ around any point $\zeta \in \mathbb{C} - O$ is analytically equivalent to the product manifold $V_1 \times \Delta_\zeta(\varepsilon)$. (Since all the manifolds V_ζ are equivalent, it is really only necessary to consider one of them, say the manifold V_1 corresponding to $\zeta = 1$.) In the standard mathematical terminology, the mapping f: $\mathbb{C}^n - V \to \mathbb{C} - O$ is a complex analytic fibre bundle over $\mathbb{C} - O$ with fibre V_1.

Fibre bundles have been quite thoroughly studied, and it is known how the topology of the bundle can be calculated in terms of the topology of the fibre and the structure of the bundle. It can be shown that the fibre V_1 has the homotopy type of a set of $\mu = (a_1 - 1)(a_2 - 1)\ldots(a_n - 1)$ spheres S^{n-1} with a single point in common. (This calculation is not difficult, but it is too much of a digression to be considered in detail here. It is perhaps of interest to note that the calculation was first carried out by physicists, in connection with the study of Feynman integrals;[12),13)] another calculation is given in Ref. 16.) It can also be shown that the fibre bundle is definitely not a trivial bundle, that is to say, that the fibre bundle is not

topologically equivalent merely to the product space $(\mathbb{C} - O) \times V_1$. (To see this it suffices to consider the effect of translating the fibre V_1 around the origin once, using the local trivialization mapping φ_ξ. As ξ ranges once around the unit circle the corresponding fibres $\varphi_\xi V_1$ are smoothly deformed into one another; but at the end the fibre V_1 is not translated back to itself unchanged, but rather translated back to itself under the action of the transformation induced by the mapping $\chi\colon (z_1, \ldots, z_n) \to (e^{2\pi i/a_1} z_1, \ldots, e^{2\pi i/a_n} z_n)$. This transformation is not generally topologically trivial; indeed its effect on the homology of V_1 is calculated explicitly in Refs. 13 and 16, although again the details will not be given here. The final result is that on the homology group $H_{n-1}(V_1, \mathbb{Z}) = \mathbb{Z}$ the induced mapping χ_* is a linear transformation with eigenvalues $e^{2\pi i\left(\frac{\nu_1}{a_1}+\ldots+\frac{\nu_n}{a_n}\right)}$ for $0 < \nu_j < a_j$. The structure of the bundle is determined by this automorphism χ_*.) The homotopy groups of the bundle space $\mathbb{C}^n - V$ are then almost fully determined immediately by the exact homotopy sequence of the bundle.[11] For $n \geq 3$ it follows that $\pi_1(\mathbb{C}^n - V) = \mathbb{Z}$, $\pi_{n-1}(\mathbb{C}^n - V) = \mathbb{Z}^\mu$, and $\pi_j(\mathbb{C}^n - V) = 0$ otherwise; while for $n = 2$, it follows that $\pi_1(\mathbb{C}^n - V)$ is an extension of $\mathbb{Z}$ by a free group of rank μ, and $\pi_j(\mathbb{C}^n - V) = 0$ otherwise. The homology groups of the bundle space $\mathbb{C}^n - V$ can be calculated from the spectral sequence of the bundle, as in Ref. 14; for $n \geq 3$, it follows that $H_j(\mathbb{C}^n - V, \mathbf{Z}) = 0$ for $j \neq 0, 1, n-1, n$, and that $H_{n-1}(\mathbb{C}^n - V, \mathbf{Z}) = H_n(\mathbb{C}^n - V, \mathbf{Z}) = 0$ precisely when

$$\prod_{0<\nu_j<a_j} \left(1 - e^{2\pi i\left(\frac{\nu_1}{a_1}+\ldots+\frac{\nu_n}{a_n}\right)}\right) = \pm 1 . \qquad \text{(III.3)}$$

To return to the manifold K itself then, it follows from Poincaré duality that $H_j(K, \mathbf{Z}) = H^{2n-3-j}(K, \mathbf{Z})$, from Alexander duality that $H^{2n-3-j}(K, \mathbf{Z}) = H_{j+1}(S^{2n-1} - K, \mathbf{Z})$, and from earlier observations that $H_{j+1}(S^{2n-1} - K, \mathbf{Z}) = H_{j+1}(\mathbb{C}^n - V, \mathbf{Z})$. Consequently $H_j(K, \mathbf{Z}) = 0$ for $j \neq 0, n-2, n-1, 2n-3$; and $H_{n-2}(K, \mathbf{Z}) = H_{n-1}(K, \mathbf{Z}) = 0$ precisely when condition (III.3) is fulfilled. A separate calculation is necessary to show that for $n \geq 4$ the manifold K is simply-connected, as in Ref. 14. When $n \geq 4$ and the exponents a_j are such that (III.3) holds, the manifold K has dimension $2n - 3 \geq 5$ and has the homology of the $(2n-3)$-sphere; and it follows from the generalized Poincaré hypothesis that K is really homeomorphic to the sphere S^{2n-3}. In these cases, the hypersurface V is topologically a manifold at its singular point, quite parallel to the special case $n = 2$ described above. However when (III.3) does not hold, or when $n = 3$, the manifold K is not

a sphere, and V is then not topologically a manifold at its singular point; these exceptional cases have not been described in very much detail.

A very surprising recent observation[14),15),16)] is that even in those cases that K is homeomorphic to the sphere S^{2n-3}, the differentiable manifold K is not necessarily differentiably homeomorphic to the sphere S^{2n-3} with its usual differentiable structure. For instance, the 28 differentiable manifolds $K(k) = S^9 \cap V(k)$ where $V(k) = \{ z \in C^{10} \mid z_1^2 + z_2^2 + z_3^2 + z_4^3 + z_5^{6k-1} = 0 \}$, are the 28 different differentiable manifolds homeomorphic to the 7 sphere, for $k = 1, \ldots, 28$.

All of these considerations so far have been limited to hypersurfaces defined by polynomials f(z) of the form (III.2). Partial extensions have been made to general polynomial hypersurfaces with an isolated singularity, though without such detailed results.[16)] Very little is really known beyond this, although some results obtained with quite a different approach will be mentioned later.

IV. Analytic Functions on Hypersurfaces

If V is a complex hypersurface in a polydisc $\Delta(\epsilon)$ in $\mathbb{C}^n$, a complex valued function $\tilde{f}(z)$ defined only on the set V is called an analytic function on the hypersurface V if it is locally the restriction to V of an analytic function in $\mathbb{C}^n$, that is to say, if for each point $a \in V$ there is an analytic function $f_U(z)$ in an open neighborhood U of the point a in $\mathbb{C}^n$ such that $\tilde{f} \mid U \cap V = f_U \mid V$. Actually the function $\tilde{f}(z)$ is analytic on V if and only if it is the restriction to V of a function f(z) analytic throughout all of the polydisc $\Delta(\epsilon)$; the proof of this assertion is quite nontrivial (see Theorem VIII A 18 in Ref. 4), and will not be given here at all, but the result will be used to simplify the present discussion. Note that there are many different functions analytic in $\Delta(\epsilon)$ which restrict to the same function $\tilde{f}$ on V; any two differ by an analytic function in $\Delta(\epsilon)$ which vanishes on V. At the regular points of V this definition coincides with what would naturally be taken to be the definition of an analytic function on V; for in a neighborhood U in which local coordinates $w_1, \ldots, w_n$ can be chosen so that $V \cap U$ is the hyperplane $w_n = 0$, it is apparent that a function $\tilde{f}(w_1, \ldots, w_{n-1})$ is analytic on V precisely when it is a complex analytic function of the n - 1 complex variables $w_1, \ldots, w_{n-1}$. At the singular points the situation is certainly not so clear, but the usefulness of this notion will be more apparent later.

The set of analytic functions on V can be described algebraically in the following manner. Let $\mathcal{O}_n$ denote the set of all analytic functions of n complex variables in the polydisc $\Delta(\epsilon)$, and $\mathcal{O}_V$ denote the set of all analytic functions on the hypersurface V. It is clear that these are both rings, in the algebraic sense; and that restriction

is a ring homomorphism from $\mathcal{O}_n$ onto $\mathcal{O}_V$. The kernel of this homomorphism, the subset of all elements of $\mathcal{O}_n$ mapping to the zero element of $\mathcal{O}_V$, is just the ideal $\mathcal{I}(V)$ of all functions $f \in \mathcal{O}_n$ vanishing on V; this will be called merely the ideal of the hypersurface V for short. Thus algebraically $\mathcal{O}_V = \mathcal{O}_n/\mathcal{I}(V)$, the quotient (or residue class) ring of $\mathcal{O}_n$ modulo the ideal $\mathcal{I}(V)$. Now suppose that V is in the standard form, and is defined by a Weierstrass polynomial p of degree r in the variable z_n. The ideal $\mathcal{I}(V)$ is the principal ideal generated by $p(z)$, that is, $\mathcal{I}(V)$ consists of all elements $p(z)f(z)$ for $f(z) \in \mathcal{O}_n$. Letting $\mathcal{O}_{n-1}$ denote the subset of $\mathcal{O}_n$ consisting of all functions which are independent of the variable z_n, it is clear that $\mathcal{O}_{n-1} \cap \mathcal{I}(V) = 0$; there are no nontrivial multiples of $p(z)$ independent of z_n. This means that under the restriction mapping from $\mathcal{O}_n$ to $\mathcal{O}_V$ the subring $\mathcal{O}_{n-1}$ is mapped isomorphically to its image; thus the subring $\mathcal{O}_{n-1} \subset \mathcal{O}_n$ can be identified with its image in $\mathcal{O}_V$, hence it can be viewed as a subring $\mathcal{O}_{n-1} \subset \mathcal{O}_V$. Further let $\mathcal{O}_{n-1}[z_n]$ denote the ring of polynomials in z_n with coefficients in the ring $\mathcal{O}_{n-1}$. As a consequence of the Weierstrass division theorem, any function $f(z) \in \mathcal{O}_n$ can be written in the form $f(z) = p(z)\,q(z) + r(z)$ for some polynomial $r(z) \in \mathcal{O}_{n-1}[z_n]$; and since $p(z)\,q(z) \in \mathcal{I}(V)$, the functions $f(z)$ and $r(z)$ will have the same restriction to V. Therefore to define analytic functions on V it is sufficient to consider merely polynomials in $\mathcal{O}_{n-1}[z_n]$ so that $\mathcal{O}_V = \mathcal{O}_{n-1}[z_n]/\mathcal{I}(V) \cap \mathcal{O}_{n-1}[z_n]$. The function $z_n \in \mathcal{O}_n$ restricts to an analytic function $\tilde{z}_n$ on V, which satisfies the polynomial equation $p(\tilde{z}_n) = 0$. That is to say, $\mathcal{O}_V = \mathcal{O}_{n-1}[\tilde{z}_n]$ is an algebraic extension of the ring $\mathcal{O}_{n-1}$ by a single element $\tilde{z}_n$ which is the root of an algebraic equation with coefficients in $\mathcal{O}_{n-1}$; indeed, this polynomial equation has leading coefficient 1, so that $\tilde{z}_n$ is actually integral over $\mathcal{O}_{n-1}$ in the algebraic sense.

As noted above, at the regular points of V the analytic functions on V can be characterized intrinsically as the functions which are analytic in the local coordinates on V; but at the singular points of V this characterization is of course impossible. However, recalling that the singular points of V lie in the set of zeros of an analytic function on V, it naturally occurs to one to try an analogue of the generalized Riemann removable singularities theorem to handle the singular points. Although this does not really work, it leads to an additional useful construction. A complex valued function $\tilde{f}(z)$ defined only on the regular points of V is called a weakly analytic function on V if it is complex analytic on the regular points of V and locally bounded at all points of V; thus the restriction of such a function to some open neighborhood of any singular point of V is a bounded analytic function on the regular points of V contained within that neighborhood. As defined here, the weakly analytic functions on V may be a properly large

class of functions than the analytic functions on V. The analytic hypersurface V is said to be normal precisely when these two classes of functions coincide, that is to say, when every weakly analytic function on V is actually analytic on V; so the normal hypersurfaces are just those hypersurfaces for which the generalized Riemann removable singularities theorem holds for the analytic functions on the hypersurface.

Not all analytic hypersurfaces are normal. To see this, consider for example an analytic hypersurface V of a polydisc $\Delta(\epsilon)$ in $\mathbb{C}^2$; and recall that such a hypersurface can be described parametrically by $z_1 = t^r$, $z_2 = \theta(t)$ for some analytic function $\theta(t)$ with $\theta(0) = 0$. Clearly the singular locus is at most the origin $z_1 = z_2 = 0$, $(t = 0)$; and that point really is a singular point only when $r > 0$ and $\theta'(0) = 0$. Now the parametrizing coordinate t can be viewed as a complex-valued function on the hypersurface V, and it is evidently weakly analytic since outside of the origin it can be expressed as an analytic function of z_1, namely $t = z_1^{1/r}$. If this function is also analytic on V, it is the restriction to V of a complex analytic function $T(z_1, z_2)$ in some open neighborhood of the origin in $\mathbb{C}^2$; and so $t = T(t^r, \theta(t))$. It is obvious that such a relation can hold only when either $r = 1$ or $\theta'(0) \neq 0$; and therefore V is normal if and only if it is regular at all points. Thus a hypersurface of a polydisc in $\mathbb{C}^2$ can never be normal when it has singularities; but there are normal hypersurfaces of polydiscs in $\mathbb{C}^n$ for $n > 2$ which have singularities.

To examine this situation in somewhat more detail, consider a hypersurface V of a polydisc $\Delta(\epsilon) \subseteq \mathbb{C}^n$, and suppose that V is represented in standard form by a Weierstrass polynomial $p(z)$ of degree r in z_n. As usual, introduce the discriminant $\delta(z') = \delta(z_1, \ldots, z_{n-1})$ of the polynomial $p(z)$, and let $D = \{ z' \in \Delta'(\epsilon') \mid \delta(z') = 0 \}$. For any point $z' \in \Delta'(\epsilon') - D$, the points of V lying over an open neighborhood of z' form r separate sheets of regular points of V, described parametrically by the r equations $z_n = \varphi_j(z')$ for $j = 1, \ldots, r$. If $\tilde{f}(z)$ is a weakly analytic function on V, its values on these r separate sheets will be the r complex analytic functions $\tilde{f}(z_1, \ldots, z_{n-1}, \varphi_j(z_1, \ldots, z_{n-1}))$ of z'. The problem is to find an analytic function $f(z)$ such that the restriction of $f(z)$ to the hypersurface V is the given function $\tilde{f}(z)$. For any fixed point $z' = (z_1, \ldots, z_{n-1}) \in \Delta'(\epsilon') - D$, the values of the function $f(z_1, \ldots, z_{n-1}, z_n)$ are given at the r points $z_n = \varphi_j(z')$; the simplest way to construct a function $f(z)$ with these values is by means of polynomial interpolation. Thus suppose that $f(z)$ is taken to be in the form of a polynomial

$$f(z) = c_1(z')\, z_n^{r-1} + \ldots + c_{r-1}(z') z_n + c_r(z') \; ;$$

the conditions on the coefficients that this polynomial have the required values at the r points $z_n = \varphi_j(z')$ are just that

$$f(z', \varphi_j(z')) = c_1(z')\,\varphi_j(z')^{r-1} + \ldots + c_{r-1}(z')\,\varphi_j(z') + c_r(z') \quad ,$$

for $j = 1, \ldots, r$. These are r linear equations among the r unknown coefficients $c_k(z')$; so by Cramer's rule the solutions are given by

$$c_k(z') = \frac{\det(\varphi_j(z')^{r-1}, \ldots, \varphi_j(z')^{r-k+1}, f(z', \varphi_j(z')), \ldots, \varphi_j(z'), 1)}{\det(\varphi_j(z')^{r-1}, \ldots, \varphi_j(z'), 1)}$$

$$= \frac{N_k(z')}{D(z')} \quad ,$$

where a typical row is shown for each matrix. It is clear that interchanging any two of the functions $\varphi_j(z')$ has the effect of interchanging two rows of the matrix, hence changes the sign of both the numerator expression $N_k(z')$ and the denominator expression $D(z')$. Writing $c_k(z') = n_k(z')/d(z')$ where $n_k(z') = N_k(z')\,D(z')$ and $d(z') = D(z')^2$, it follows that $n(z')$ and $d(z')$ are both symmetric polynomials in the expressions $\varphi_j(z')$; and therefore, as proved earlier, both $n(z')$ and $d(z')$ are complex analytic functions in all of $\Delta'(\epsilon')$. Actually the denominator $d(z')$ is a well known expression, called the van der Monde determinant, which is identical with the discriminant $\delta(z')$ of the polynomial $p(z)$. The function

$$\delta(z')\,f(z) = n_1(z')\,z_n^{r-1} + \ldots + n_{r-1}(z')\,z_n + n_r(z')$$

is therefore an analytic function in all of $\Delta(\epsilon)$, even a polynomial in the variable z_n; and the restriction of this function to the hypersurface V is $\delta(z')\,\tilde{f}(z)$.

This shows that for any weakly analytic function $\tilde{f}(z)$ on the hypersurface V, the product $\delta(z)\tilde{f}(z)$ is an analytic function on V, where $\delta(z) = \delta(z')$ is the discriminant of the Weierstrass polynomial $p(z)$ defining the hypersurface; hence any weakly analytic function $f(z)$ is at least the restriction to V of a meromorphic function in $\mathbb{C}^n$, with a fixed denominator $\delta(z')$. Not all meromorphic functions in $\mathbb{C}^n$ restrict to weakly analytic functions on V, of course.

As a final remark, it can be shown that the weakly analytic functions form a finite dimensional module over the ring of analytic functions in the strict sense, at least when the polydisc $\Delta(\epsilon)$ is sufficiently small; that is to say, there are finitely many weakly analytic

functions $\tilde{f}_1(z), \ldots, \tilde{f}_m(z)$ on V such that any other weakly analytic function can be written in the form $\tilde{f}(z) = g_1(z)\,\tilde{f}_1(z) + \ldots + g_m(z)\,\tilde{f}_m(z)$ for some analytic functions $g_j(z)$ on V. This is rather more difficult, though, so the proof will not be given here.

V. Analytic Subvarieties in Standard Form

Generalizing the notion of an analytic hypersurface, a subset V of an open domain $U \subseteq \mathbb{C}^n$ is said to be an analytic subvariety of U if in some open neighborhood of each point of U the subset V is the set of common zeros of a finite number of complex analytic functions. A point a on a subvariety V is called a regular point if it is possible to choose coordinates $(w_1, \ldots, w_n)$ in an open neighborhood of z in $\mathbb{C}^n$ such that locally V is a linear subspace $\{w \mid w_{k+1} = \ldots = w_n = 0\}$, for some integer k called the complex dimension of V; a subvariety V is called a complex manifold of complex dimension k if it is regular and of complex dimension k at each point. The points of a subvariety which are not regular points are called the singular points. An analytic subvariety V of an open polydisc $\Delta(\epsilon) \subseteq \mathbb{C}^n$ can be determined as the set of common zeros of a finite number of functions analytic throughout $\Delta(\epsilon)$, provided the polydisc is sufficiently small. The set of all analytic functions in $\Delta(\epsilon)$ which vanish on V form an ideal $\mathcal{I}(V)$ in the ring $\mathcal{O}_n$ of all analytic functions in $\Delta(\epsilon)$, called for short the ideal of V. A subvariety V of $\Delta(\epsilon)$ is called irreducible if it cannot be written as a union $V = V_1 \cup V_2$ where V_j are analytic subvarieties of $\Delta(\epsilon)$ properly contained in V. It is easy to see that a subvariety V of $\Delta(\epsilon)$ is irreducible precisely when the ideal $\mathcal{I}(V)$ is a prime ideal in $\mathcal{O}_n$; recall that an ideal $\mathcal{I} \subset \mathcal{O}_n$ is said to be prime if whenever $f_1 f_2 \in \mathcal{I}$ at least one of the elements f_j is contained in $\mathcal{I}$.

An analytic subvariety V of an open polydisc $\Delta(\epsilon)$ in $\mathbb{C}^n$ can be represented in a standard form somewhat similar to the standard form for a hypersurface, as discussed in Sec. II. It is first necessary to choose a suitable system of coordinates in the ambient space $\mathbb{C}^n$. Selecting any nontrivial function $f_n(z) \in \mathcal{I}(V)$, after a suitable nonsingular linear change of coordinates in $\mathbb{C}^n$ it can be assumed that $f_n(z)$ is regular in the variable z_n; and then of course $f_n(z)$ can be replaced by a Weierstrass polynomial in z_n having the same zero locus, which polynomial will then be called $f_n(z)$. Then select a nontrivial function $f_{n-1}(z) \in \mathcal{I}(V)$ which depends only on the coordinates $z_1, \ldots, z_{n-1}$; after a change of coordinates involving only these variables, it can be assumed that $f_{n-1}(z)$ is a Weierstrass polynomial in z_{n-1} with coefficients analytic in $z_1, \ldots, z_{n-2}$. The process can then be continued until for some integer k the ideal $\mathcal{I}(V)$ contains no functions depending only on the variables $z_1, \ldots, z_k$; and there remains the sequence of Weierstrass polynomials $p_{k+1}, \ldots, p_n$ in $\mathcal{I}(V)$, noting

that the further coordinate changes leave these in the form of Weierstrass polynomials.

The further normalization is perhaps described most easily in terms of analytic functions on the subvariety V, and in a rather algebraic manner. As before let $\mathfrak{O}_n$ be the ring of all analytic functions in $\Delta(\epsilon)$; and for any $m \leqq n$ let $\mathfrak{O}_m$ denote the subring of those functions in $\mathfrak{O}_n$ which depend only on the variables $z_1, \ldots, z_m$. The ring of analytic functions on V is the residue class or quotient ring $\mathfrak{O}_V = \mathfrak{O}_n/\mathfrak{J}(V)$; and the image in $\mathfrak{O}_V$ of a function $f \in \mathfrak{O}_n$ will be denoted by $\tilde{f}$. Now since the polynomial $f_n(z) = z_n^r + a_1 z_n^{r-1} + \ldots + a_r$ is contained in $\mathfrak{J}(V)$, upon passing to the residue class ring it follows that $0 = \tilde{z}_n^r + \tilde{a}_1 \tilde{z}_n^{r-1} + \ldots + \tilde{a}_r$; therefore $\tilde{z}_n$ is an integral algebraic quantity over the subring $\tilde{\mathfrak{O}}_{n-1}$, the image of the subring $\mathfrak{O}_{n-1} \subset \mathfrak{O}_n$ in $\mathfrak{O}_V$. Furthermore, by the Weierstrass division theorem any function $f \in \mathfrak{O}_n$ can be written in the form $f = f_n g_n + h_n$ for some functions $g_n, h_n \in \mathfrak{O}_n$, where h_n is a polynomial in z_n; so in the residue class ring, $\tilde{f} = \tilde{h}_n \in \tilde{\mathfrak{O}}_{n-1}[\tilde{z}_n]$, the latter symbol denoting the set of polynomials in $\tilde{z}_n$ with coefficients from $\tilde{\mathfrak{O}}_{n-1}$. Altogether then, $\mathfrak{O}_V = \tilde{\mathfrak{O}}_{n-1}[\tilde{z}_n]$, where $\tilde{z}_n$ is an integral algebraic quantity over the ring $\tilde{\mathfrak{O}}_{n-1}$. Applying the same argument with the polynomial f_{n-1}, it follows that $\tilde{\mathfrak{O}}_{n-1} = \tilde{\mathfrak{O}}_{n-2}[\tilde{z}_{n-1}]$, where $\tilde{z}_{n-1}$ is an integral algebraic quantity over the ring $\tilde{\mathfrak{O}}_{n-2}$. This argument can be repeated, until finally $\tilde{\mathfrak{O}}_k = \mathfrak{O}_k$ since $\mathfrak{O}_k \cap \mathfrak{J}(V) = 0$. It is a known algebraic result that such a succession of integral algebraic extensions can be made simultaneously; this is the theorem of the transitivity of integral extensions.[17] Therefore it follows that $\mathfrak{O}_V = \mathfrak{O}_k[\tilde{z}_{k+1}, \ldots, \tilde{z}_n]$ where $\tilde{z}_j$ are integral algebraic quantities over $\mathfrak{O}_k$, that is to say, where the elements $\tilde{z}_j$ are roots of polynomial equations

$$0 = \tilde{z}_j^{r_j} + c_j \tilde{z}_j^{r_j-1} + \ldots + c_j^{(r_j-1)} \tilde{z}_j + c_j^{(r_j)}$$

with coefficients $c_j^{(m)} \in \mathfrak{O}_k$. The polynomials

$$p_j(z) = z_j^{r_j} + c_j z_j^{r_j-1} + \ldots + c_j^{(r_j-1)} z_j + c_j^{(r_j)}$$

are consequently analytic functions in $\mathfrak{O}_n$ which belong to the ideal $\mathfrak{J}(V)$.

Suppose further that the ideal $\mathfrak{J}(V)$ is really a prime ideal in $\mathfrak{O}_n$. It then follows that when the functions $p_j(z)$ are taken to be the polynomials of lowest degree of the required form in $\mathfrak{J}(V)$, they are uniquely determined Weierstrass polynomials. This set of n - k polynomials will be called the first set of canonical polynomials for the

ideal $\mathcal{J}(V)$. It also follows that the ring $\mathcal{O}_V$ is an integral domain, a ring with no divisors of zero, hence has a well defined quotient field $\mathcal{M}_V$, the field of meromorphic functions on V; and that $\mathcal{M}_V = \mathcal{M}_k[\tilde{z}_{k+1}, \ldots, \tilde{z}_n]$, an algebraic extension of the field $\mathcal{M}_k$ of meromorphic functions of k complex variables. Now if the coordinates $z_{k+1}, \ldots, z_n$ are suitably chosen, the field $\mathcal{M}_V$ will be generated by the image of a single coordinate, say z_{k+1}; this is the theorem of the primitive element.[17] Since $\mathcal{M}_V = \mathcal{M}_k[\tilde{z}_{k+1}]$, every meromorphic function on V can be written as a polynomial in $\tilde{z}_{k+1}$ with coefficients in $\mathcal{M}_k$, so the field $\mathcal{M}_V$ is described completely by the single polynomial equation p_{k+1}. Indeed, every analytic function $f \in \mathcal{O}_V$ can also be written as a polynomial in $\tilde{z}_{k+1}$, but again the coefficients are meromorphic functions in $\mathcal{M}_k$; for it is not necessarily true that $\mathcal{O}_V = \mathcal{O}_k[\tilde{z}_{k+1}]$. However for an analytic function $\tilde{f} \in \mathcal{O}_V$, the denominator of the coefficients of the representing polynomial can be taken to be a fixed element, the discriminant $\delta \in \mathcal{O}_k$ of the polynomial p_{k+1}; that is to say, for any element $\tilde{f} \in \mathcal{O}_V$, the product $\delta\tilde{f} \in \mathcal{O}_k[\tilde{z}_{k+1}]$. In particular, for each coordinate z_j for $j = k+2, \ldots, n$, there is a polynomial $q_j^* \in \mathcal{O}_k[\tilde{z}_{k+1}]$ such that $\delta\tilde{z}_j = q_j^*(\tilde{z}_{k+1})$; and consequently the complex analytic functions

$$q_j(z) = \delta(z_1, \ldots, z_k)\, z_j - q_j^*(z_1, \ldots, z_k; z_{k+1}) \quad ,$$

which are polynomials in z_{k+1} and z_j, belong to the ideal $\mathcal{J}(V)$. These will be called the second set of canonical polynomials for the ideal $\mathcal{J}(V)$. The further discussion of the standard form for general analytic subvarieties rests just on these two sets of canonical equations.

First, one formal algebraic property of these canonical polynomials should be noted. If f(z) is an arbitrary analytic function in $\Delta(\varepsilon)$, an application of the Weierstrass division theorem shows that it can be written in the form $f(z) = p_n(z)\, g_n(z) + f^{(n-1)}(z)$, where $g_n(z)$, $f^{(n-1)}(z)$ are analytic in $\Delta(\varepsilon)$ and $f^{(n-1)}(z)$ is a polynomial in z_n. Applying the division theorem again, each coefficient of the polynomial $f^{(n-1)}(z)$ can be divided by $p_{n-1}(z)$, leaving as remainder a polynomial in z_{n-1}; so that $f(z) = p_n(z)\, g_n(z) + p_{n-1}(z)\, g_{n-1}(z) + f^{(n-2)}(z)$, where all the functions are analytic in $\Delta(\varepsilon)$ and $f^{(n-2)}(z)$ is a polynomial in z_{n-1} and z_n. Repeating the argument, finally secure that

$$f(z) = p_{k+2}(z)\, g_{k+2}(z) + \ldots + p_n(z)\, g_n(z) + f^{(k+1)}(z) \quad , \qquad (V.1)$$

where all these functions are analytic in $\Delta(\varepsilon)$ and $f^{(k+1)}(z)$ is a polynomial in the variables $z_{k+2}, \ldots, z_n$. Now notice that for $j = k+2$,

$\ldots,n$ and for any positive integer ν, it follows that $\delta^{\nu} z_j^{\nu}$ is equal to a multiple of the canonical polynomial $q_j(z)$, plus a remainder which is an analytic function only of the variables $z_1,\ldots,z_{k+1}$, indeed, which is even a polynomial in z_{k+1}. Consequently (V.1) can be rewritten, for some integer N, as

$$\delta^N f(z) = \sum_{j=k+2}^{n} p_j(z)\, g_j(z) + \sum_{j=k+2}^{n} q_j(z)\, h_j(z) + f_1(z) \ , \qquad (V.2)$$

where all the functions are analytic in $\Delta(\epsilon)$ and the remainder $f_1(z)$ depends only on the variables $z_1,\ldots,z_{k+1}$. The function $f_1(z)$ can then be divided by the Weierstrass polynomial $p_{k+1}(z)$ so that

$$\delta^N f(z) = \sum_{j=k+1}^{n} p_j(z)\, g_j(z) + \sum_{j=k+2}^{n} q_j(z)\, h_j(z) + f_0(z) \ , \qquad (V.3)$$

where $f_0(z) \in \mathcal{O}_k[z_{k+1}]$ and the degree of $f_0(z)$ in the variable z_{k+1} is strictly less than the degree r of the polynomial $p_{k+1}(z)$. If the original element f(z) belongs to the ideal $\mathcal{J}(V)$, so does the remainder $f_0(z)$; but since $p_{k+1}(z)$ is the polynomial in $\mathcal{O}_k[z_{k+1}] \cap \mathcal{J}(V)$ of lowest degree, necessarily $f_0(z) = 0$. Notice further that if f(z) is from the beginning a polynomial in the variables $z_{k+1},\ldots,z_n$, the first step in the above construction is unnecessary; so that f(z) can be written in the form (V.3) with $g_{k+2}(z) = \ldots = g_n(z) = 0$. If f(z) is also an element of the ideal $\mathcal{J}(V)$, the remainder $f_0(z)$ is zero; so that whenever $f(z) \in \mathcal{J}(V) \cap \mathcal{O}_{k+1}[z_{k+2},\ldots,z_n]$, the product $\delta^N f(z)$ can be expressed as a linear combination of the canonical polynomials p_{k+1}, $q_{k+2},\ldots,q_n$. This holds in particular for the polynomials $p_{k+2},\ldots,p_n$, so after multiplying (V.3) by a further power of δ, these terms can be omitted from that formula. The final result then is that for any analytic function f(z) in $\Delta(\epsilon)$,

$$\delta^N f(z) = p_{k+1}(z)\, g(z) + q_{k+2}(z)\, h_{k+2}(z) + \ldots + q_n(z)\, h_n(z) + f_0(z) \ , \qquad (V.4)$$

for some integer N, where all these functions are analytic in $\Delta(\epsilon)$ and $f_0(z)$ depends only on the variables $z_1,\ldots,z_{k+1}$ and is a polynomial in z_{k+1} of degree strictly less than r, the degree of $p_{k+1}(z)$; moreover if $f(z) \in \mathcal{J}(V)$, necessarily $f_0(z) = 0$.

To describe the standard form for the irreducible subvariety V, introduce the subset $B \subset V$ defined by $B = \{z \in V \mid \delta(z) = 0\}$, where $\delta(z) = \delta(z_1,\ldots,z_k)$ is, as above, the discriminant of the polynomial $p_{k+1}(z)$.

Note that B is itself an analytic subvariety of $\Delta(\epsilon)$ contained in V. It follows immediately from (V.4) that the set V - B is determined completely by the canonical polynomials $p_{k+1}(z), q_{k+2}(z), \ldots, q_n(z)$; for any function $f(z) \in \mathcal{J}(V)$ has the property that $\delta^N f(z)$ is expressible as a linear combination of these canonical polynomials, so that provided $\delta(z) \neq 0$, the function f(z) vanishes at z if and only if these polynomials vanish at z. That is to say,

$$V - B = \left\{ z \in \Delta(\epsilon) \,\middle|\, \delta(z) \neq 0,\ p_{k+1}(z) = q_{k+2}(z) = \ldots = q_n(z) = 0 \right\}.$$

For any point $z = (z_1, \ldots, z_n) \in V$, necessarily $p_{k+1}(z) = p_{k+1}(z_1, \ldots, z_{k+1}) = 0$; so that the natural projection $\pi_{k+1}: \mathbb{C}^n \to \mathbb{C}^{k+1}$ defined by $\pi_{k+1}(z_1, \ldots, z_n) = (z_1, \ldots, z_{k+1})$ maps the subvariety $V \subset \mathbb{C}^n$ into the hypersurface $V_{k+1} \subset \mathbb{C}^{k+1}$ defined by

$$V_{k+1} = \left\{ (z_1, \ldots, z_{k+1}) \in \Delta(\epsilon) \,\middle|\, p_{k+1}(z_1, \ldots, z_{k+1}) = 0 \right\} .$$

This hypersurface is in the standard form described in Sec. II; the natural projection $\pi_k: \mathbb{C}^{k+1} \to \mathbb{C}^k$ exhibits V_{k+1} as an r-sheeted branched covering of a polydisc $\Delta'(\epsilon') \subseteq \mathbb{C}^k$, and the branch points in V_{k+1} form the subvariety

$$B_{k+1} = \left\{ (z_1, \ldots, z_{k+1}) \in V_{k+1} \,\middle|\, \delta(z_1, \ldots, z_k) = 0 \right\}$$

consisting of those points of V_{k+1} lying over the hypersurface

$$D = \left\{ (z_1, \ldots, z_k) \in \Delta'(\epsilon') \,\middle|\, \delta(z_1, \ldots z_k) = 0 \right\}$$

of the polydisc $\Delta'(\epsilon')$. The complement $V_{k+1} - B_{k+1}$ is a k-dimensional analytic submanifold of $(\Delta'(\epsilon') - D) \times \mathbb{C}$, given parametrically by the r analytic functions $z_{k+1} = \varphi_j(z_1, \ldots, z_k)$, $j = 1, \ldots, r$, over an open neighborhood of any point of $\Delta'(\epsilon') - D$. Since the coordinates $z_{k+2}, \ldots, z_n$ of a point $(z_1, \ldots, z_n) \in V - B$ are expressed analytically in terms of the coordinates $z_1, \ldots, z_{k+1}$ by the canonical equations $q_{k+2}(z) = \ldots = q_n(z) = 0$, it follows that the induced mapping $\pi_{k+1}: V \to V_{k+1}$ leads to a one-to-one mapping from V - B onto $V_{k+1} - B_{k+1}$. The complement V - B is a k-dimensional complex analytic submanifold of $(\Delta'(\epsilon') - D) \times \mathbb{C}^{n-k}$, an r-sheeted covering of $\Delta'(\epsilon') - D$ under the mapping $\pi_k \pi_{k+1}$; and over an open neighborhood of any point of $\Delta'(\epsilon') - D$ the r sheets of the manifold V - B are described parametrically by the equations.

$$z_{k+1} = \varphi_j(z_1, \ldots, z_k), \quad z_m = \frac{q_m^*(z_1, \ldots, z_k; \varphi_j(z_1, \ldots, z_k)}{\delta(z_1, \ldots, z_k)}$$

$$m = k+2, \ldots, n$$

for $j = 1, \ldots, r$. This situation can be perhaps best kept in mind by referring to the following diagram.

$$\begin{array}{lccccc} & V - B & \subseteq & V & \subset & \mathbb{C}^n \\ & \downarrow & & \downarrow & & \downarrow \\ \text{(homeomorphism)} & & & & & \pi_{k+1} \\ & V_{k+1} - B_{k+1} & \subseteq & V_{k+1} & \subset & \mathbb{C}^{k+1} \\ \text{(r-sheeted covering, branch locus } B_{k+1}\text{)} & \downarrow & & \downarrow & & \downarrow \\ & \Delta'(\varepsilon') - D & \subseteq & \Delta'(\varepsilon') & \subseteq & \mathbb{C}^k \end{array} \qquad \pi_k$$

The canonical equations do not serve to describe the points of V lying over the discriminant locus $D \subset \Delta'(\varepsilon')$. The equations $q_{k+2}(z) = \ldots = q_n(z) = 0$ are trivial whenever $z' \in D$; and although the equations $p_{k+1}(z) = \ldots = p_n(z) = 0$ show that there are at most finitely many points $z \in V$ lying over any point $z' \in \Delta'(\varepsilon')$ even over a point $z' \in D$, these equations generally describe an analytic subvariety properly larger than V. However, this problem can be finessed by observing that V is the point set closure of $V - B$ in $\Delta(\varepsilon)$, hence that it suffices merely to describe the set $V - B$; the proof of this is not altogether trivial, and will be omitted here.

Turning now to some consequences of this standard form for analytic subvarieties, it should be pointed out first of all that at no point in the entire preceding discussion was any use made of the condition that the prime ideal under consideration be precisely the ideal of all analytic functions vanishing on an irreducible subvariety. Beginning with an arbitrary prime ideal $\mathfrak{I} \subset \mathfrak{O}_n$, the canonical polynomials can be constructed as above, and the set of common zeros of all the functions in the ideal $\mathfrak{I}$ is an analytic subvariety in the standard form. The ideal of all analytic functions vanishing on V is now a prime ideal $\mathfrak{I}(V)$ which contains the original ideal $\mathfrak{I}$; actually these two ideals coincide, so any prime ideal is precisely the ideal of all functions vanishing on an irreducible subvariety. (To see this, consider any analytic function $f \in \mathfrak{I}(V)$. As before, after dividing through by the canonical polynomials p_j, q_j, it follows that $\delta^N f = f_0 + f_1$ where $f_0 \in \mathfrak{I}$ and f_1 is an analytic function depending only on the variables $z_1, \ldots, z_{k+1}$, indeed, a polynomial in z_{k+1} of degree strictly less than the degree of p_{k+1}. The remainder f_1 vanishes on V_{k+1};

but again this can only happen when f_1 vanishes identically, so that actually $\delta^N f \in \mathfrak{J}$. Since $\mathfrak{J}$ is a prime ideal and $\delta \notin \mathfrak{J}$, it follows further that $f \in \mathfrak{J}$, and the desired result is therewith demonstrated.) Knowing this, it is essentially a purely algebraic argument to derive next the Hilbert zero theorem: if $f_1(z), \ldots, f_m(z)$ are analytic functions in a polydisc $\Delta(\epsilon)$, if $V = \{ z \in \Delta(\epsilon) \mid f_1(z) = \ldots = f_m(z) = 0 \}$, and if $f(z)$ is another analytic function in $\Delta(\epsilon)$ which vanishes on V, then $f^N = g_1 f_1 + \ldots + g_m f_m$ for some integer N and some analytic functions $g_1, \ldots, g_m$ in $\Delta(\epsilon)$. The details can be found in Ref. 4 and elsewhere.

Next, it is worth commenting in some detail on the relations between the analytic subvariety $V \subset \mathbb{C}^n$ and the hypersurface $V_{k+1} \subset \mathbb{C}^{k+1}$ appearing in the above standard representation for V. This of course involves a comparison of analytic subvarieties of polydiscs in complex spaces of different dimensions; and such comparisons arise in many other contexts as well. In general, consider complex analytic subvarieties $V_1 \subset \Delta(\epsilon_1) \subseteq \mathbb{C}^{n_1}$ and $V_2 \subset \Delta(\epsilon_2) \subseteq \mathbb{C}^{n_2}$. A continuous mapping $f: V_1 \to V_2$ is said to be a complex analytic mapping between these two subvarieties if there is a complex analytic mapping $F: \Delta(\epsilon_1) \to \mathbb{C}^{n_2}$ such that the restriction of F to the subset V_1 is just f, or in symbols, such that $F \mid V_1 = f$; and these two subvarieties are said to be analytically equivalent if there are complex analytic mappings $f: V_1 \to V_2$ and $g: V_2 \to V_1$ such that the compositions fg and gf are the appropriate identity mappings. Note that this latter condition can be restated as the condition that there exist complex analytic mappings $F: \Delta(\epsilon_1) \to \mathbb{C}^{n_2}$ and $G: \Delta(\epsilon_2) \to \mathbb{C}^{n_1}$ such that $FG \mid V_2$ and $GF \mid V_1$ are both the identity mappings; this does not mean that FG and GF are themselves identity mappings, so $\Delta(\epsilon_1)$ and $\Delta(\epsilon_2)$ still may be polydiscs in complex spaces of different dimensions. This notion of equivalence thus allows one to speak of analytic subvarieties without reference to the spaces in which they are imbedded; an equivalence class is called an analytic variety, and a space which has locally the structure of an analytic variety is called an analytic space. As another approach to the same end, note that whenever $f: V_1 \to V_2$ is a complex analytic mapping and g is an analytic function on V_2, the composition gf is clearly an analytic function on V_1; and conversely, if $f: V_1 \to V_2$ is a continuous mapping such that gf is an analytic function on V_1 whenever g is an analytic function on V_2, then f is a complex analytic mapping. (To see this, let $z_1, \ldots, z_{n_1}$ be the coordinates in $\mathbb{C}^{n_1}$ and $w_1, \ldots, w_{n_2}$ be the coordinates in $\mathbb{C}^{n_2}$. Note that the restriction of w_j to the subvariety V_2 is a complex analytic function on V_2; therefore $w_j f$ is a complex analytic function on V_1, the restriction to V_1 of some complex analytic function F_j in $\Delta(\epsilon_1)$. The set of functions $F_1, \ldots, F_{n_2}$ define a complex analytic mapping $F: \Delta(\epsilon_1) \to \mathbb{C}^{n_2}$;

and clearly the restriction of this mapping to V_1 is precisely the given mapping f, since the restriction of the function F_j gives the w_j-coordinate of the image under f of a point on V_1.) Thus analytic mappings are precisely the continuous mappings which preserve complex analytic functions; and analytic equivalences are topological homeomorphisms which identify the rings of complex analytic functions.

In the representation in standard form for the subvariety V, the natural projection induces a complex analytic mapping from V onto V_{k+1}; but this mapping is not generally an analytic equivalence. For on the one hand, the mapping need not be a topological homeomorphism, since some of the points of the subset B may be collapsed upon projection to B_{k+1}; and on the other hand, even if it is a homeomorphism, the inverse map may not be analytic, since there may be analytic functions on V which do not induce analytic functions on V_{k+1}. However, since the restriction of the projection mapping is a complex analytic equivalence between the complex manifolds $V - B$ and $V_{k+1} - B_{k+1}$, it is evident that it induces a one-to-one correspondence between the weakly analytic functions on V and V_{k+1}; so that the varieties V and V_{k+1} can be viewed as being weakly equivalent. Actually the coordinates $z_{k+2},\ldots,z_n$ of a point $z \in V$ are weakly analytic functions of the point $(z_1,\ldots,z_{k+1}) \in V_{k+1}$, and hence they define a weakly analytic mapping from V_{k+1} back to V, exhibiting the weak analytic equivalence in yet another way; it should be noted that these functions are really only defined on the regular points of V_{k+1}, since an inverse mapping need not be a single-valued function on B_{k+1}. If the hypersurface V_{k+1} is normal, so that weakly and strongly analytic functions coincide, then of course the subvarieties V and V_{k+1} are necessarily analytically equivalent. For any hypersurface V_{k+1} it can be shown that there exists a normal analytic subvariety V such that V_{k+1} is just the projection of into $\mathbb{C}^{k+1}$, as in the standard representation; this subvariety is called the normalization of V_{k+1}, and is, roughly speaking, the analytic subvariety arising from V_{k+1} by making all the weakly analytic functions strictly analytic.

Finally, the integer k appearing prominently in the preceding discussion is called the (complex) dimension of the analytic subvariety V. This integer can be defined in terms of the above standard form, although of course it is then apparently dependent on the choice of coordinates in $\mathbb{C}^n$; however the dense open subset $V - B$ of V is a complex analytic manifold of dimension k, so that this can be taken as another definition of the complex dimension of the entire subvariety V, and it is then clear that this dimension is intrinsically defined. It is also clear that an analytic hypersurface of a polydisc in $\mathbb{C}^n$ is an analytic subvariety of dimension $n - 1$, and conversely; so that the analytic subvarieties of dimension $n - 1$ are precisely the analytic

subvarieties that can be defined as the set of zeros of a single analytic function. Somewhat more generally, if V_1 and V_2 are irreducible analytic subvarieties of a polydisc in $\mathbb{C}^n$ of dimensions k_1 and k_2, then each irreducible component of the intersection $V_1 \cap V_2$ has dimension $\geq k_1 + k_2 - n$; the proof is not difficult, and will not be given here, but details can be found in Ref. 1 or 7. (To motivate this formula to some extent, consider the case that V_1 and V_2 are coordinate planes in $\mathbb{C}^n$; then V_1 is defined by setting $n - k_1$ coordinates equal to zero, V_2 is defined by setting $n - k_2$ coordinates equal to zero, so $V_1 \cap V_2$ is defined by setting at most $2n - k_1 - k_2$ coordinates equal to zero and hence has dimension not less than $n - (2n - k_1 - k_2) = k_1 + k_2 - n$.) As a consequence, each irreducible component of the set of common zeros of $n - k$ analytic functions in a polydisc in $\mathbb{C}^n$ has dimension $\geq k$. It should be emphasized that the converse is not generally true; that is, an analytic subvariety consisting of one or more irreducible components of dimension k in a polydisc in $\mathbb{C}^n$ cannot necessarily be defined as the set of common zeros of $n - k$ analytic functions. The precise characterization of this special class of subvarieties, called geometrical complete intersections, is a rather difficult matter that is not yet completely understood.

VI. Miscellaneous Properties

Of course a great deal more is known about complex analytic varieties than it has been possible even to mention here; but the preceding lectures may have provided some sort of general picture of what these varieties are like and how they are handled, and also may have illuminated some pitfalls to be avoided when physically faced with an analytic variety. Before concluding, though, at least a few words must be said about some of the major topics not previously mentioned; and it must be repeated that the entire discussion has been limited to the purely local and essentially geometrical properties of complex analytic varieties.

In studying complex analytic functions with singularities, there are a number of theorems guaranteeing that certain classes of singularities are really removable; the Riemann removable singularities theorem is one example, and its usefulness has been apparent in the preceding discussion. Similar questions arise in studying complex analytic varieties. Suppose that W is an r-dimensional analytic subvariety of a polydisc Δ in $\mathbb{C}^n$, and that V is a k-dimensional analytic subvariety of the open subset $\Delta - W$ of that polydisc; the question arises whether the subvariety V can be continued through W to be an analytic subvariety of all of Δ. This subvariety V can always be so extended when $r < k$, and there are further conditions under which such an extension is possible when $r = k$; these questions are treated

in Refs. 18 and 19 for instance, where further references can also be found.

The tangent space to a manifold is a very useful auxiliary tool in studying the properties of that manifold, and it would of course be quite desirable to have an analogue in the case of a complex analytic variety with singularities. Actually there are a number of candidates for the tangent set to a complex analytic variety, with various properties and applications; some examples are mentioned in Ref. 4, and a much more extensive discussion can be found in Refs. 20 and 21.

Finally, perhaps the strongest possible tool for studying the singularities of complex analytic varieties and subvarieties is the resolution of singularities. For a one-dimensional analytic variety V with a singular point, as has been demonstrated in the preceding lectures, there exists a complex analytic manifold $\hat{V}$ admitting an analytic mapping $\pi: \hat{V} \to V$ which is a topological homeomorphism; and the properties of the singularity of V can be described directly in terms of this regular parametrization, by means of the properties of analytic functions and mappings on the complex manifold $\hat{V}$. In higher-dimensional cases, varieties with singularities are not necessarily topologically homeomorphic to manifolds; but nonetheless there is an analogue of the one-dimensional situation. If V is an irreducible analytic subvariety of a polydisc in $\mathbb{C}^n$, and if S is the singular locus of V, there exists a complex manifold $\hat{V}$ admitting an analytic mapping $\pi: \hat{V} \to V$ with the following properties: (i) the restriction of π is a complex analytic homeomorphism between the complex manifolds $\hat{V} - \pi^{-1}(S)$ and $V - S$; (ii) $\pi^{-1}(S)$ is an analytic subvariety of $\hat{V}$, and the restriction of π is a proper analytic mapping from $\pi^{-1}(S)$ onto S. (To say that π is proper is merely the assertion that the inverse image of a compact set is compact; generally the inverse image of a point of S will not even be a finite point set, but rather a compact complex analytic subvariety of $\hat{V}$.) The properties of the singularities of V can then be described in terms of the properties of the complex manifold $\hat{V}$ and this parametrization. For two-dimensional subvarieties, this process was described in Ref. 22; and some applications to the topological properties of the singularities were given in Ref. 23. Note that this case proved exceptional in the discussion in Sec. III; but a good deal of light has been shed on this problem by the resolution of singularities techniques, and the subject is now being very actively investigated. The three-dimensional analytic case was treated in Ref. 24, and the general case in Ref. 25; the general results are quite deep and involved, and it will be some time before the situation is really thoroughly understood.

References

Among the books and general surveys of various aspects of the theory of functions of several complex variables are the following:

1. S. Abhyankar, Local Analytic Geometry (Academic Press, New York, 1964).
2. H. Behnke and P. Thullen, Theorie der Funktionen mehrerer komplexer Veränderlichen (Chelsea, New York, reprint).
3. S. Bochner and W. T. Martin, Several Complex Variables (Princeton Univ. Press, Princeton, 1948).
4. R. C. Gunning and H. Rossi, Analytic Functions of Several Complex Variables (Prentice-Hall, Englewood Cliffs, N. J., 1965).
5. L. Hörmander, An Introduction to Complex Analysis in Several Variables (Van Nostrand, Princeton, 1966).
6. B. Malgrange, Lectures on the Theory of Functions of Several Complex Variables (Tata Institute, Bombay, 1960).
7. R. Narasimhan, Introduction to the Theory of Analytic Spaces (Springer Verlag, Berlin, 1966).

Additional books and papers referred to in the lectures are the following:

8. J. E. Reeve, Rend. Sem. Mat. Torino 14, 159 (1955).
9. O. Zariski, Algebraic Surfaces (Chelsea, New York, reprint).
10. M. J. Greenberg, Lectures on Algebraic Topology (W. A. Benjamin, New York, 1967).
11. E. Spanier, Algebraic Topology (McGraw-Hill, New York, 1967).
12. P. Federbush, J. Math. Phys. 6, 941 (1965).
13. F. Pham, Bull. Soc. Math. France 93, 333 (1965).
14. E. Brieskorn, Invent. Math. 2, 1 (1966).
15. F. Hirzebruch and K. H. Mayer, O(n)-Mannigfaltigkeiten, exotische Sphären, und Singularitäten (Springer, Berlin, 1968).
16. J. W. Milnor, Singular Points of Complex Hypersurfaces (Princeton Univ. Press, Princeton, 1968).
17. B. L. Van der Waerden, Modern Algebra, Vols. I and II (Frederick Ungar, New York, 1950).
18. R. Remmert and K. Stein, Math. Ann. 126, 263 (1953).
19. G. Stolzenberg, Volumes, Limits, and Extensions of Analytic Varieties (Springer, Berlin, 1966).
20. H. Whitney, Annals of Math. 81, 469 (1965).
21. H. Whitney, Local Properties of Analytic Varieties, Differential and Combinatorial Topology (Princeton Univ. Press, Princeton, 1965), pp. 205-244.
22. F. Hirzebruch, Math. Annalen 126, 1 (1953).
23. D. B. Mumford, Inst. Hautes Etudes Sci. Publ. Math. 9, 5 (1961).

24. N. Kuhlmann, Math. Annalen 151, 304 (1963); 154, 387 (1964).
25. H. Hironaka, Annals of Math. 79, 109 (1964).