MC SYLLABUS 38.1

REPRESENTATIONS OF LOCALLY COMPACT GROUPS WITH APPLICATIONS

PART I

T.H. KOORNWINDER (ed.)

MATHEMATISCH CENTRUM AMSTERDAM 1979

AMS(MOS)subject classification scheme (1970):22D10,22E70,81A78,22E15,28-01,
28A70,20C15,22B05,22D25,22D30,
43A05

ISBN 90 6196 161 0

CONTENTS

PART I

PART II

INTRODUCTION

These two volumes contain the notes of a colloquium which was held at the Mathematical Centre during the academic year 1977/78. The colloquium was organized by the department of Applied Mathematics of the Mathematical Centre and it was directed by G. van Dijk at the Mathematical Institute of the University of Leiden, E.A. de Kerf at the Institute of Theoretical Physics of the University of Amsterdam and T.H. Koornwinder at the Mathematical Centre. The lectures were given by several people from the institutes mentioned above.

The colloquium had a dual purpose: to present the basic results on unitary representations of locally compact groups, with special emphasis on induced representations, and to give some applications of this theory to physics. A rather unusual arrangement of the lectures has been chosen. After a short survey of the mathematical results in Chapter I, applications to physics are presented in Chapters II and III. In particular, the representation theory of the Poincaré group from the physicist's point of view in Chapter III serves as a starting point for developing all the machinery which is needed for a rigorous mathematical treatment of the representations of this group. Thus, in Chapters IV through XI one meets Lie groups, the theory of measure and integration, general representation theory, representations of locally compact abelian groups, induced representations, the imprimitivity theorem, and representations of semidirect products. At the end of Chapter XI the synthesis with Chapter III is made. The syllabus concludes with two chapters on representations of semi-simple Lie groups, respectively in the compact case and the non-compact case (mainly $SL(2,\mathbb{R})$).

To be honest, the intended interaction and integration between the mathematical and the physical approach did not succeed as well as the organizers wished, but we still think that the effort has been worthwhile and we hope that the resulting notes will be useful for both mathematicians and phycisists.

Most chapters can be read independently from each other, the main exceptions being: Chapters VII → VIII and Chapters IX → X → XI. Knowledge of advanced mathematical or physical topics is not required for any of the chapters.

Each chapter is divided into sections, which are sometimes divided into subsections. The detailed contents are given at the beginning of each chapter. Formula numbers have the form (m.n), while definitions, lemmas, theorems, examples, etc., are consecutively numbered as m.n. In both cases m refers to the section number. Subsections do not affect the above numbering systems. When referring to formulas, etc., in other chapters we place the Roman chapter numbers in front of the Arabic formula number.

Finally I would like to thank all authors for their contribution and for their pleasant cooperation.

T.H. Koornwinder, editor.

I

GROUP REPRESENTATIONS: A FIRST INTRODUCTION AND SURVEY

G. VAN DIJK

Math. Inst., RU Leiden

CONTENTS

1. FINITE GROUPS

Let G be a finite group with n elements. Denote by $\ell^2(G)$ the complex vector space of all complex-valued functions on G provided with the scalar product $(f,g) = \sum_{x\in G} f(x)\overline{g(x)}$. Then $\dim \ell^2(G) = n$. G acts on $\ell^2(G)$ by left translation λ:

$$\lambda(x)f(y) = f(x^{-1}y) \qquad (x,y \in G,\ f \in \ell^2(G)).$$

λ is called the left regular representation of G.
We shall denote by $\hat{G}$ the set of equivalence classes of finite-dimensional irreducible representations of G on complex vector spaces. The following facts are well-known:

(1.1) Any representation π of G on a finite-dimensional complex vector space V can be made unitary by a suitable choice of the scalar product on V.

In particular $\hat{G}$ has a complete set of *unitary* representatives.

(1.2) Any (finite-dimensional) representation of G can be decomposed into a direct sum of irreducible representations of G and this decomposition is unique (modulo permutation and equivalence of the factors).

(1.3) Any irreducible representation π is equivalent to a subrepresentation of the left regular representation λ of G; actually π occurs $d(\pi)$ times in λ, where $d(\pi)$ = degree of π.

So the construction of a complete set of representatives of $\hat{G}$ remains. This problem is completely solved for special groups G only. The main tool is the inducing construction of Frobenius (see SERRE [7,§7]). Actually one mostly gives the *characters* of the irreducible representations. The construction of a model for the corresponding representations seems to be much more difficult and requires for instance in the case of the finite simple linear groups non-trivial results from algebraic geometry (cf. DELIGNE-LUSZTIG [3]).

2. COMPACT GROUPS

As already observed by Schur, much of the theory for finite groups goes through for compact groups, using integration on the group with respect to an invariant positive measure μ. Thus, if f is a continuous function on the group G, then

$$\int_G f(ax)d\mu(x) = \int_G f(x)d\mu(x)$$

for all $a \in G$. Such a measure exists on every compact G and is unique up to a scalar factor. It was first constructed by Haar. Actually, one should call μ left-invariant, but it is not difficult to show that it is also right-invariant for compact groups G.

Let G be a compact group, V a finite-dimensional complex vector space. A representation π of G on V is a *continuous* homomorphism of G into GL(V). Let us show that (1.1) remains true for compact G. Choose a scalar product (,) on V and put

$$\langle v,w\rangle = \int_G (\pi(x)v,\pi(x)w)d\mu(x) \qquad (v,w \in V).$$

Then $\langle\ ,\ \rangle$ is a G-invariant scalar product on V. The proof of (1.2) is similar to the finite case. The left regular representation λ of G on $L^2(G,\mu)$ is defined by

$$\lambda(x)f(y) = f(x^{-1}y) \qquad (f \in L^2(G,\mu);\quad x,y \in G).$$

This representation, which is infinite-dimensional if G is infinite, can be decomposed as the Hilbert-direct sum of irreducible finite-dimensional representations of G, each occuring a number of times equal to its degree. Moreover, each $\pi \in \hat{G}$ has a representative in this decomposition. This is the famous Peter-Weyl theorem.

Observe that we are naturally lead to consider *infinite-dimensional representations* of G. The regular representation λ is an infinite-dimensional unitary representation of G, if G is infinite.

Let $G = \mathbb{T}$, where $\mathbb{T}$ is the group of complex numbers of modulus 1. The irreducible unitary representations of $\mathbb{T}$ are the one-dimensional representations $z \mapsto z^n$ $(z \in \mathbb{T},\ n \in \mathbb{Z})$. So the Peter-Weyl theorem reduces to the well-known Parseval theorem in the theory of Fourier series:
$f(z) \sim \sum_{n=-\infty}^{\infty} c_n z^n$ (L^2-convergence),

$$\|f\|_2^2 = \frac{1}{2\pi}\int_0^{2\pi} |f(e^{i\phi})|^2\, d\phi = \sum_{n=-\infty}^{\infty} |c_n|^2 .$$

As to the construction of $\hat{G}$, one again specializes G, because the problem is too difficult for general compact groups G. We already considered $G = \mathbb{T}$. Similar observations hold for general abelian compact groups G. An important class is formed by the connected semisimple compact Lie groups, e.g. SO(n), SU(n), Sp(n). Hermann Weyl has given the characters of the irreducible (finite-dimensional) representations for this class of groups, in his famous *character formula*. The construction of models for the corresponding irreducible representations was performed much later by Borel-Weil-Bott, using cohomology of holomorphic vector bundels (WALLACH [8,Ch.6]).

3. LOCALLY COMPACT GROUPS

There are several good reasons to study unitary representations. This is natural because of the connection between Fourier analysis and unitary group representations. On the other hand, the work of Wigner has stressed the significance of unitary representations in physics. However, for many (especially in physics) relevant groups, all non-trivial unitary prepresentations are infinite-dimensional. Let us consider the groups $SL(2,\mathbb{R})$ and $SL(2,\mathbb{C})$, the latter being a two-fold covering of the proper Lorentz group $L_+^\uparrow$.

PROPOSITION 3.1. *Let* π *be a continuous unitary representation of* $SL(2,\mathbb{R})$ *(resp.* $SL(2,\mathbb{C})$*) on a* n*-dimensional complex Hilbert space* V. *Then* $\pi(x) = 1$ *for all* $x \in SL(2,\mathbb{R})$ *(resp.* $SL(2,\mathbb{C})$*).*

PROOF. Let $\mathfrak{g} = sl(2,\mathbb{R})$ (resp. $sl(2,\mathbb{C})$) be the Lie algebra of $SL(2,\mathbb{R})$ (resp. $SL(2,\mathbb{C})$). $\mathfrak{g}$ is simple (as a real Lie algebra). Denote by $d\pi$ the corresponding representation of $\mathfrak{g}$ on V. Then $d\pi = 0$ (so $\pi(x) = 1$ for all x) or $d\pi$ is injective. In the latter case, $d\pi$ transforms the Killing form of $\mathfrak{g}$ into the Killing form of $d\pi(\mathfrak{g})$:

$$\kappa_{d\pi(\mathfrak{g})}(d\pi(X),\, d\pi(Y)) = \kappa_{\mathfrak{g}}(X,Y) \quad (X,Y \in \mathfrak{g}).$$

Observe that $d\pi(X)$ and $d\pi(Y)$ are skew-hermitan in $\mathfrak{gl}(d\pi(\mathfrak{g}))$.

Therefore the signature of $\kappa_{d\pi(\mathfrak{g})}$ is (0,3) (resp. (0,6)), which differs from the signature of $\kappa_{\mathfrak{g}}$, being (2.1) (resp. (4,2)). This yields a contradiction. Thus $\pi(x) = 1$ for all $x \in SL(2,\mathbb{R})$ (resp. $SL(2,\mathbb{C})$). □

So we are brought to consider from now on also infinite-dimensional unitary representations on Hilbert spaces. The proper class of groups here is the class of locally compact groups. This class includes $SL(2,\mathbb{R})$, $SL(2,\mathbb{C})$ and $L_+^\uparrow$. Moreover, every locally compact group admits a (up to a scalar factor unique) left (and right) Haar measure, which is required already for the formulation of (1.3) for this class of groups.

DEFINITION 3.2. Let G be a group, whose underlying space is a topological space. G is said to be a *locally compact group* if

(i) G is a locally compact topological space;

(ii) the maps $(x,y) \mapsto xy$ from $G \times G$ into G and $x \mapsto x^{-1}$ from G to G, are continuous.

For technical reasons, G is assumed to be *second countable*: there is a countable basis for its topology.

Some examples:

(3.1) $G = \mathbb{R}$; Haar measure = Lebesgue measure, both left and right invariant.

(3.2) $G = \mathbb{T}$; $\int_G f(x)d\mu(x) = \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\phi})d\phi$ defines a bi-invariant measure.

(3.3) $G = \left\{\begin{pmatrix} \lambda & \mu \\ 0 & \lambda^{-1}\end{pmatrix} : \lambda \neq 0,\ \mu \in \mathbb{R}\right\}$. Left Haar measure $d_\ell g = \lambda^{-2} d\lambda d\mu$, right Haar measure $d_r g = d\lambda d\mu$.

A locally compact group G admitting a bi-invariant positive measure, is called *unimodular*.

4. INFINITE-DIMENSIONAL GROUP REPRESENTATIONS

Let G be a locally compact second countable group and H a separable complex Hilbert space. We denote $L(H)$ the algebra of bounded linear operators on H. A *representation* of G on H is a map $\pi: G \to L(H)$ such that

(i) $\pi(xy) = \pi(x)\pi(y)$ for all $x \in G$;

(ii) $\pi(e) = I$, where e is the identity element in G and I is the identity operator on H;

(iii) π is (strongly) continuous, that is, for each $v \in H$, the map $x \mapsto \pi(x)v$ is continuous from G to H.

If H_0 is a subspace of H, we say that H_0 is *invariant* under π if $\pi(x)H_0 \subset H_0$ for all $x \in G$. A representation π of G on H is *irreducible* if the only closed subspaces of H which are invariant under π are $\{0\}$ and

H. We remark that the concept of irreducibility is related to *closed* subspaces. Of course, if dim H is finite, "closed" can be removed from the definition.

A representation π of G on H is called a *unitary* representation if $\pi(x)$ is a unitary operator for all $x \in G$.

We say that two representations π_1, π_2 on H_1, H_2 respectively are *unitarily equivalent* if there exists a unitary operator A from H_1 onto H_2 such that $A\pi_1(x) = \pi_2(x)A$ for all $x \in G$. Clearly, unitary equivalence is an equivalence relation on the collection of all irreducible unitary representations of G. The set of equivalence classes of irreducible unitary representations of G is called the (unitary) *dual* of G and is denoted by $\hat{G}$.

Before presenting a few examples, we state the basic irreducibility criterion for unitary representations of a locally compact group. It is known as the continuous (or infinite-dimensional) version of Schur's Lemma.

<u>CRITERION 4.1</u>. *Let π be a unitary representation of a locally compact group G on H. Then π is irreducible if and only if every $A \in L(H)$, which satisfies the condition*

$$A\pi(x) = \pi(x)A$$

for all $x \in G$, has the form $A = c\cdot I$, where $c \in \mathbb{C}$.

As an immediate corollary, we have the fact that every irreducible unitary representation of a locally compact *abelian* group is one-dimensional: On the one hand, given a *character* χ of G, that is a continuous homomorphism from G to $\mathbb{T}$, we can define a unitary representation π of G on $\mathbb{C}$ by

$$\pi(x)z = \chi(x)z \qquad (x \in G,\ z \in \mathbb{C}).$$

Obviously, the representation is irreducible. On the other hand, suppose that π is an irreducible unitary representation of a locally compact abelian group G on H and fix $y \in G$. Then $\pi(y)\pi(x) = \pi(yx) = \pi(xy) = \pi(x)\pi(y)$ for all $x \in G$. Hence, by (4.1), $\pi(y) = \chi(y)\cdot I$ where $\chi(y) \in \mathbb{C}$, $|\chi(y)| = 1$. Since π is irreducible, H must be one-dimensional. Moreover, the map $y \mapsto \chi(y)$ $(y \in G)$ defines a character of G. So $\hat{G}$ may be identified in this case with the character group of G. So all irreducible unitary representations of $G = \mathbb{R}$ are one-dimensional and given by the characters $\chi_y(x) = e^{2\pi ixy}$ $(x \in \mathbb{R},\ y \in \mathbb{R})$. $\hat{G}$ is naturally isomorphic with $\mathbb{R}$ in this case.

Coming back to compact groups, it is not difficult to prove, applying the theory of compact operators on Hilbert space, that any irreducible unitary representation of a compact group is finite-dimensional.

Let G be a locally compact group and assume G unimodular. Let $H = L^2(G)$. For $f \in H$ and $x \in G$, we define

$$\lambda(x)f(y) = f(x^{-1}y) \qquad (y \in G).$$

Then λ is a unitary representation of G, the *left regular representation* of G. The *right regular representation* ρ of G is defined by

$$\rho(x)f(y) = f(yx) \qquad (x,y \in G,\ f \in H).$$

The unitary representations λ and ρ are unitarily equivalent via the unitary operator $A: H \to H$ given by $Af(x) = f(x^{-1})$ $(f \in H,\ x \in G)$. The representations λ and ρ are finite-dimensional if and only if G is a finite group.

5. DECOMPOSITION OF REPRESENTATIONS

In this section G is a *unimodular* locally compact group. So, in particular, $L^2(G)$ is defined unambiguously.

Let λ be the left regular representation of $\mathbb{R}$ on $L^2(\mathbb{R})$. Then λ does not contain any irreducible subrepresentation of $\mathbb{R}$. Indeed, suppose there is $f \in L^2(\mathbb{R})$, $f \neq 0$, $z \in \mathbb{R}$ such that

$$\lambda(x)f(y) = e^{2\pi ixz}\, f(y) \qquad (x,y \in \mathbb{R}).$$

We can choose $k \in C_c(\mathbb{R})$ such that $\hat{k}(z) = \int_{-\infty}^{\infty} k(x)\, e^{2\pi ixz}\, dx \neq 0$. Then

$$\int_{-\infty}^{\infty} f(y-x)k(x)dx = \int_{-\infty}^{\infty} e^{2\pi ixz}\, k(x)dx \cdot f(y).$$

So we can choose f continuous. Without loss of generality we may assume $f(0) = 1$. Therefore $f(y) = e^{-2\pi iyz}$ $(y \in \mathbb{R})$, which contradicts the fact that f should be in $L^2(\mathbb{R})$.

Actually one can prove the following:

THEOREM 5.1. *Let π be an irreducible unitary representation of G on H. The following three conditions are equivalent:*

(i) *there are $v,w \in H - \{0\}$ such that the function $x \mapsto (\pi(x)v,w)$ is in $L^2(G)$;*

(ii) *for all $v,w \in H$, the function $x \mapsto (\pi(x)v,w)$ is in $L^2(G)$;*

(iii) *π is equivalent to a subrepresentation of the left regular representation λ on $L^2(G)$.*

A nice proof of this theorem is given by BOREL [2, Théorème 5.15]. Such representations π are called *square-integrable*. In case $G = SL(2,\mathbb{R})$, they correspond exactly to the "discrete series" of BARGMANN [1].

Square-integrable representations behave in a sense similar to irreducible unitary representations of finite and compact groups. In particular, their coefficients satisfy orthogonality relations:

THEOREM 5.2. *Let* π,π' *be square-integrable representations of* G *on* H, H' *respectively.*

(i) *There exists a number* $d(\pi) > 0$, *called the formal degree of* π, *depending only on the normalization of the Haar measure on* G, *such that*

$$\int_G (\pi(x)v_1,w_1)\overline{(\pi(x)v_2,w_2)}dx = d(\pi)^{-1}(v_1,v_2)\overline{(w_1,w_2)}$$

for all $v_1,v_2,w_1,w_2 \in H$.

(ii) *If* π *is not equivalent to* π', *then*

$$\int_G (\pi(x)v,w)\overline{(\pi'(x)v',w')}dx = 0$$

for all $v,w \in H$, $v',w' \in H'$.

Obviously there is no hope that (1.2) could be proved for infinite-dimensional unitary representations of general locally compact G. However $G = \mathbb{R}$ provides us with a good substitute. Indeed, each $f \in L^2(\mathbb{R})$ can be written in the form $f(x) \sim \int_{-\infty}^{\infty} \hat{f}(y)\, e^{-2\pi ixy}\, dy$ (L^2-convergence), which is nothing but the Plancherel theorem from Fourier analysis. So λ is actually written as a *direct integral* of irreducible unitary representations, instead of a direct sum. By generalizing this concept, we get a good substitute of (1.2) for general locally compact G. But another problem comes up. It may happen that the decomposition is not unique. This has lead to the observation that G should be of type I, a concept coming from the theory of operator algebras of Murray and Von Neumann.

Now assume G of type I. Abelian groups, compact groups, connected semisimple Lie groups, nilpotent Lie groups are all type I. Let λ be the left regular representation of G on $L^2(G)$. There exists a unique positive measure ν on $\hat{G}$, such that

$$\lambda = \int_{\hat{G}} \pi d\nu(\pi)$$

is the direct integral decomposition of λ. ν is called the *Plancherel measure* of G. Now (1.3) suggests the following question: does every element of $\hat{G}$ lie in the support of the Plancherel measure? This is in general not the case. For $SL(2,\mathbb{R})$ and $SL(2,\mathbb{C})$ the "*complementary series*" does not enter in the decomposition of λ. In general λ decomposes into a continuous part and a discrete part. The discrete part is exactly the "discrete series" of G (that is, the square-integrable representations), whose members π enter with mass equal to the formal degree $d(\pi)$. (This is a very simplified picture of a complicated theory.) For finite and compact groups the decomposition is completely discrete and formal degree = degree. Moreover, (1.3) holds for these groups, as we observed previously.

6. CONSTRUCTION OF $\hat{G}$

This is the most difficult part of the program. Undoubtly the main technique to construct unitary representations of G is by inducing simple representations from smaller subgroups. This technique, which is the generalization of Frobenius' method for finite groups to general locally compact groups, is due to MACKEY [5]. It will be explained in the course of the colloquium. To prove that the induced representations obtained, are actually irreducible is by no means trivial and requires frequently hard analytic tools. If we restrict to $SL(n,\mathbb{R})$ and $SL(n,\mathbb{C})$, $\hat{G}$ is completely determined only if n = 2,3. Each connected Lie group is, at least infinitesimally, a semidirect product of a connected reductive Lie group and a connected nilpotent Lie group. So the program could be as follows:

(a) Construction of the irreducible unitary representations of a semidirect product of locally compact groups.

(b) Construction of $\hat{G}$ for G reductive (that is G/Z semisimple, where Z is the center of G).

(c) Construction of $\hat{G}$ for G a connected nilpotent Lie group G.

We intend to consider some special cases in the colloquium, in particular $SL(2,\mathbb{R})$.

Finally, let us remark that it is possible to define the notion of a *character* of an irreducible unitary representation of G. Actually, HARISH-CHANDRA [4] has constructed the characters of the discrete series of a connected semisimple Lie group G, with finite center. These are distributions on G, that is continuous linear forms on $C_c^\infty(G)$. Later on, SCHMID [6] has constructed models of the corresponding representations in the same spirit as Borel-Weil-Bott for compact groups.

LITERATURE

[1] BARGMANN, V., *Irreducible unitary representations of the Lorentz group,* Annals of Math. (2) 48 (1947), 568-640.

[2] BOREL, A., *Représentations de groupes localement compacts,* Lecture Notes in Mathematics 276, Springer-Verlag, Berlin, 1972.

[3] DELIGNE, P. & G. LUSZTIG, *Representations of reductive groups over finite fields,* Annals of Math. (2) 103 (1976), 103-161.

[4] HARISH-CHANDRA, *Discrete series for semisimple Lie groups* II, Acta Math. 116 (1966), 1-111.

[5] MACKEY, G.W., *The theory of unitary group representations,* The University of Chicago Press, Chicago, 1976.

[6] SCHMID, W., *On the characters of the discrete series,* Inventiones Math. 30 (1975), 47-144.

[7] SERRE, J-P., *Représentations linéaires des groupes finis,* Hermann, Paris, 1967.

[8] WALLACH, N.R., *Harmonic analysis on homogeneous spaces,* Marcel Dekker, New York, 1973.

II

QUANTUM MECHANICS AND SYMMETRY

G.G.A. BAUERLE

Inst. v. Theor. Fysica, Univ. van Amsterdam

CONTENTS

In this chapter we review the principles of quantum mechanics. We discuss the concept of symmetry for a quantum mechanical system and conclude with a discussion of symmetry groups and their representations. In this chapter we have no pretensions with regard to full mathematical rigour.

1. QUANTUM MECHANICS

1.1. Introduction

In the description of a physical system the *observables* and the *states* of the physical system are fundamental concepts. These observables and states are represented by mathematical entities. Examples of physical systems are:

EXAMPLE 1.1. A non-relativistic particle with mass m in a given external force field, with potential energy $V(\vec{r})$.

EXAMPLE 1.2. An electromagnetic field in a cavity.

The state of the system of Example 1.1 at a certain time t_0 is described in classical mechanics for instance by a pair of vectors $(\vec{r}(t_0), \vec{p}(t_0))$; $\vec{r}(t_0)$ is the position and $\vec{p}(t_0)$ the linear momentum of the particle at time t_0. Examples of observables (measurable quantities) are in this case: the position $\vec{r}$, the linear momentum $\vec{p}$, the angular momentum $\vec{L} = \vec{r} \wedge \vec{p}$ and the total energy $H = \frac{1}{2}\vec{p}^2/m + V(\vec{r})$. The evolution in time is determined by Newton's equations.

In Maxwell's theory the state of the system of Example 1.2 is described by a pair of vector fields $(\vec{E}(\vec{x},t), \vec{B}(\vec{x},t))$; $\vec{E}(\vec{x},t)$ is the electric field and $\vec{B}(\vec{x},t)$ is the magnetic induction. Examples of observables are the electric field, the magnetic induction, the energy density and the Poynting vector. The evolution in time of this system is determined by Maxwell's equations. Classical mechanics and Maxwell's theory of electromagnetism are examples of non-quantal theories.

We formulate now some of the general principles of quantum mechanics. These general principles are supposed to be valid for the quantum mechanical description of an arbitrary physical system. The first two postulates indicate how the states and the observables of the system are represented in a quantum mechanical theory. This is followed by a postulate about the physical interpretation and finally some remarks are made about the description of the evolution in time of the physical system.

1.2. General principles of quantum mechanics

We formulate the postulates about the states and the observables in the following way.

POSTULATE 1.3 (states). *A (pure) state of a physical system S is represented by a vector ($\neq 0$) in a separable Hilbert space $\mathcal{H}$ over the field of the complex numbers $\mathbb{C}$. The vectors $\psi \in \mathcal{H}(\psi \neq 0)$ and $\lambda\psi(\lambda \in \mathbb{C}, \lambda \neq 0)$ represent the same state of the system.*

It follows immediately that the state of a system can always be represented by a unit vector. But even then there is still the following arbitrariness. The unit vectors ψ and $e^{i\alpha}\psi(\alpha \in \mathbb{R})$ represent the same state of the system, $e^{i\alpha}(\alpha \in \mathbb{R})$ is called a *phase factor.*

For simple physical systems there also corresponds to every vector of $\mathcal{H}$ a physically realizable state of the system. This is however not true in in general. If there exists a superselection rule (BOGOLIUBOV [4], WICK [14]) for the system S then there are vectors in $\mathcal{H}$ which do not correspond to physically realizable states. For example if ψ_n is a vector describing a state of the system with total electrical charge ne, then the vector $\psi_0 + \psi_1$ does not correspond to a physically realizable state of the system.

The inner product of two state vectors $\psi, \chi \in \mathcal{H}$ is denoted by $\langle\psi,\chi\rangle$ and by definition it is anti-linear in ψ (usual convention in quantum mechanics). The norm of the vectors in $\mathcal{H}$ is positive definite. So we exclude in our discussion quantum mechanical formalisms with a vector space with indefinite metric (NAGY [12]) like e.g. the Gupta-Bleuler formalism of quantum electrodynamics.

POSTULATE 1.4 (observables). *An observable $\mathcal{A}$ of the physical system S is represented by a self-adjoint operator A with domain $\mathcal{D}_A \subset \mathcal{H}$, such that $\mathcal{D}_A$ is dense in $\mathcal{H}(\bar{\mathcal{D}}_A = \mathcal{H})$.*

In particular the total energy of the system S, being an observable, is represented by a self-adjoint operator, denoted by H and called the Hamiltonian. Self-adjoint operators have a spectral decomposition (resolution of the identity). For a self-adjoint (hermitean) operator on a finite-dimensional inner product space this is quite elementary and there the following theorem holds.

THEOREM 1.5 (HALMOS [8]). *To every self-adjoint linear transformation* A *on a finite-dimensional inner product space* V *there correspond real numbers* $\lambda_1,\ldots,\lambda_r$ $(0 < r \le \dim V,\ \lambda_1 < \lambda_2 \ldots < \lambda_r)$ *and self-adjoint projection operators* $P_1,\ldots,P_r$ *such that*

(i) $P_iP_j = 0$ *for* $i \neq j$ *and* $P_i \neq 0$ *for all* $i = 1,\ldots,r$;

(ii) $\sum_{j=1}^{r} P_j = I$ *(the identity operator on* V*)*;

(iii) $A = \sum_{j=1}^{r} \lambda_j P_j$ *(spectral form of* A*)*.

We define the projection operators

$$E_i = \sum_{j=1}^{i} P_j \ (i = 1,\ldots r) \text{ and } E_0 = 0$$

and these have the properties

(a) $E_i \le E_j$ for $\lambda_i \le \lambda_j$ (We recall that if E and E' are bounded self-adjoint operators on an inner product space V, then $E \le E'$ if $\langle\psi,E\psi\rangle \le \langle\psi,E'\psi\rangle$ for all $\psi \in V$);

(b) $E_r = I$;

(c) $A = \sum_{j=1}^{r} \lambda_j (E_j - E_{j-1})$.

A non-zero vector in the subspace $P_i H$ of H is an eigenvector of A with eigenvalue λ_i. For a self-adjoint operator in a Hilbert space the following classical theorem, which is rather analogous to Theorem 1.5, holds.

THEOREM 1.6 (HELMBERG [9]). *For every self-adjoint operator* A *on a Hilbert space* H *there exists a family of projections* $\{E(\lambda) \mid \lambda \in \mathbb{R}\}$, *called the spectral family of* A, *with the following properties:*

(i) $E(\lambda) \le E(\lambda')$ *for* $\lambda \le \lambda'$;

(ii) $(s)\lim_{\lambda\to-\infty} E(\lambda) = 0$ *and* $(s)\lim_{\lambda\to+\infty} E(\lambda) = I$

((s) lim *means the strong limit, i.e., if* A *and* A_n (n = 1,2,...) *are bounded operators on* $\mathcal{H}$, (s) lim A_n = A *means* $\lim_{n\to\infty} A_n f = Af$ *for all* $f \in \mathcal{H}$);

(iii) $\quad$ $(s)\lim_{\mu\downarrow\lambda} E(\mu) = E(\lambda)$ *for all* $\lambda \in \mathbb{R}$;

(iv) $\quad$ $\psi \in \mathcal{D}_A \subset \mathcal{H}$ *iff* $\int_{-\infty}^{\infty} \lambda^2 \, d\|E(\lambda)\psi\|^2 < \infty$;

(v) $\quad$ $A\psi = \int_{-\infty}^{\infty} \lambda dE(\lambda)\psi$ *for all* $\psi \in \mathcal{D}_A$.

In textbooks on quantum mechanics the postulate about the physical interpretation is often formulated by means of the eigenvectors and eigenvalues of the self-adjoint operator representing an observable. This is correct if the state vector space is a finite-dimensional inner product space or if the eigenvectors of the observable form a basis of the Hilbert space $\mathcal{H}$. In the general case this procedure does not apply except if we extend the Hilbert space $\mathcal{H}$. The use of eigenvectors can be avoided by means of the spectral family of the self-adjoint operator A of the observable $\mathcal{A}$.

POSTULATE 1.7 (interpretation). *If a measurement of the observable* $\mathcal{A}$, *represented by the self-adjoint operator* A *with spectral family* E(λ), *is made and the system is in the state represented by the unit vector* $\psi \in \mathcal{H}$ *then the probability* $W_\psi(a \in \Delta)$ *of finding the value* a *of the observable* $\mathcal{A}$ *in the interval* $\Delta = (\alpha_1, \alpha_2]$ *is given by*

(1.1) $$W_\psi(a \in \Delta) = \langle\psi, E_A(\Delta)\psi\rangle$$

with $E_A(\Delta) = E(\alpha_2) - E(\alpha_1)$.

We now state two immediate consequences of Postulate 1.7.

REMARK 1.8. Let A be a self-adjoint operator with a discrete and non-degenerate spectrum of eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n, \ldots$ and spectral decomposition $A = \sum_k \lambda_k P_k$. The probability $W_\psi(a \in \Delta)$ is then given by

$$W_\psi(a \in \Delta) = \sum_{\substack{n \\ \lambda_n \in \Delta}} \langle\psi, P_n\psi\rangle = \sum_{\substack{n \\ \lambda_n \in \Delta}} \langle P_n\psi, P_n\psi\rangle$$

where $\|\psi\| = 1$.
So $W_\psi(a \in \Delta) = 0$ if $\lambda_n \notin \Delta$ for all n.
The probability W_n that the result of the measurement will be the eigenvalue λ_n is given by

(1.2) $$W_n = \langle P_n\psi, P_n\psi\rangle = |\langle\phi_n,\psi\rangle|^2,$$

where ϕ_n is a unit eigenvector of A with eigenvalue λ_n ($A\phi_n = \lambda_n\phi_n$). W_n is called the *transition probability*. We stress, that transition probabilities are independent of the phase factors of the state vectors, because if $\hat{\psi} = e^{i\alpha}\psi$ and $\hat{\phi} = e^{i\beta}\phi$ then

$$|\langle\hat{\phi},\hat{\psi}\rangle|^2 = |\langle\phi,\psi\rangle|^2.$$

The same remark applies to (1.1), i.e. $W_\psi = W_{\hat{\psi}}$.

REMARK 1.9. We define $W(a') = W_\psi(a \in \Delta)$ where $\Delta = (-\infty, a']$. If the system is in the state represented by the unit vector $\psi \in H$, the *mean value (expectation value)* $\langle A\rangle_\psi$ of the observable A is defined by

$$\langle A\rangle_\psi = \int_{-\infty}^{\infty} a' \, d\, W(a').$$

From this follows that $\langle A\rangle_\psi = \langle\psi, \int_{-\infty}^{\infty} a' \, d\, E(a')\psi\rangle$.
Hence we get the following simple and very important formula for the expectation value:

(1.3) $$\langle A\rangle_\psi = \langle\psi, A\psi\rangle.$$

Postulate 1.7 gives the connection between measurable quantities (like transition probabilities and expectation values) and the concepts introduced in the Postulates 1.3 and 1.4 (description of a state by means of a vector and an observable by a self-adjoint operator). So far for the kinematical aspects.

The dynamical law governing the evolution in time can be expressed in various ways. Let $\psi(t)$ denote the state vector of the system at time t and let $A(t)$ denote the operator describing the observable A at time t. If $V(t)$ is an arbitrary time-dependent unitary operator on H and we define $\psi'(t) = V(t)\psi(t)$ and $A'(t) = V(t)A(t)V^{-1}(t)$, then the expectation value can be written as

(1.4) $$\langle A\rangle_\psi = \langle\psi,A\psi\rangle = \langle\psi',A'\psi'\rangle = \langle A'\rangle_{\psi'}.$$

This transformation of the state vectors and operators leaves the expectation values and transition probabilities invariant. So the description with primed vectors and primed operators is physically equivalent to the description with unprimed vectors and unprimed operators, because it gives the same values to the directly observable quantities such as expectation values and transition probabilities. The time-dependence of the expectation values $\langle A\rangle_\psi$ and $\langle A'\rangle_{\psi'}$ is the same (they are equal), but the time-dependence of ψ and ψ' (and also of A and A') is in general different. Hence the dynamical laws in the primed and unprimed description are different too. In the description of the dynamics, that is the evolution in time of the state-vectors and the self-adjoint operators of physical quantities, there are two extreme alternatives. They are called the *Schrödinger picture and the Heisenberg picture* (MESSIAH [11]).

In the Schrödinger picture, the state vectors are time-dependent. This time-dependence is described by the Schrödinger equation. Operators representing observables like position, momentum and angular momentum are time-independent.

In the Heisenberg picture the state vectors are independent of time. But the observables have a time - dependence described by the Heisenberg equation. Although in this chapter there is no further discussion of the dynamics of the system, we state for definiteness that we work in the Heisenberg picture.

1.3. Bra and ket notation of Dirac

Dirac denotes a vector $\psi \in H$ as $|\psi\rangle$, so

$$|\psi\rangle = \psi \in H,$$

$|\psi\rangle$ is called a *ket vector* or *ket* and ψ is the label of the ket. The continuous linear functionals χ: $H \to \mathbb{C}$ forming the dual space H^* of H are denoted by $\langle\chi| = \chi$ so that $\chi|\psi\rangle \equiv \langle\chi|\psi\rangle \in \mathbb{C}$, $\langle\chi|$ is called a *bra vector* or *bra*. Eigenkets and eigenbras are simply labelled by the eigenvalue, e.g.

(1.5) $A|a_i> = a_i|a_i>$,

(1.6) $<a_i|A = <a_i|a_i$.

If A is a self-adjoint operator with a discrete and non-degenerate spectrum, then the orthonormality relation reads

(1.7) $<a_i|a_j> = \delta_{ij}$.

The expression $|u><v|$ is defined as an operator on H by

(1.8) $(|u><v|)|\psi> = |u>(<v|\psi>) \in H$

for all $|\psi> \in H$.
The spectral form of the operator A can now be written as

(1.9) $A = \sum_i a_i|a_i><a_i|$

and the resolution of the identity reads

(1.10) $I = \sum_i |a_i><a_i|$.

The Dirac notation is usually very adequate for deriving certain results by means of formal manipulations. We illustrate this in the case of formula (1.9). From

$|\psi> = \sum_i c_i|a_i>$ and (1.7) follows $c_i = <a_i|\psi>$

and substituting the latter result in the former formula gives

$$|\psi> = \sum_i (<a_i|\psi>)|a_i>.$$

Hence

$$A|\psi> = \sum_i (<a_i|\psi>)A|a_i> = \sum_i (<a_i|\psi>)a_i|a_i>$$

$$= \sum_i a_i|a_i>(<a_i|\psi>) = (\sum_i a_i|a_i><a_i|)|\psi>$$

for all $|\psi\rangle \in H$ and this completes the formal verification of (1.9).

In the case that some eigenvalues are degenerate, one needs more labels to distinguish the eigenvectors.

In DIRAC [6] it is shown that in the case of a continuous spectrum this formalism remains formally applicable. Let B be a self-adjoint operator with a continuous non-degenerate spectrum and let the eigenvalue equation read

$$(1.11) \qquad B|b\rangle = b|b\rangle,$$

The orthonormality relation now becomes

$$(1.12) \qquad \langle b|b'\rangle = \delta(b-b')$$

and $|b\rangle$ is of course no longer a vector from H. Relations (1.9) and (1.10) can be written as

$$(1.13) \qquad B = \int b|b\rangle\langle b|db$$

and

$$(1.14) \qquad I = \int |b\rangle\langle b|db.$$

For comparison with Theorem 1.6 take $E(\lambda) = \int_{-\infty}^{\lambda} |b\rangle\langle b|db$.

For a self-adjoint operator with a partly discrete and partly continuous and eventually degenerate spectrum the obvious modifications have to be made.

A mathematically sound foundation of the Dirac formalism is obtained by means of a rigged Hilbert space (BOGOLIUBOV [4, Ch. 1.5, 1.6 and 4.1], GELFAND [7] and BÖHM [5]).

2. SYMMETRY

2.1. Symmetry transformations

The vectors $\psi \in H(\psi \neq 0)$ and $\lambda\psi \in H(\lambda \in \mathbb{C}, \lambda \neq 0)$ describe the same state of the physical system (Postulate 1.3). Even if we represent the state of the system by the unit vector $\tilde{\psi} = \|\psi\|^{-1}\psi$ ($\|\psi\| = <\psi,\psi>^{\frac{1}{2}}$) there is no uniqueness because the unit vectors $\tilde{\psi}$ and $e^{i\alpha}\tilde{\psi}$ ($\alpha \in \mathbb{R}$) represent the same state. We define the *ray* $\underline{\psi}$ as the set of vectors

$$\underline{\psi} = \{e^{i\alpha}\psi \,|\, \alpha \in \mathbb{R}\}.$$

The set $\tilde{\underline{\psi}}$ with $\|\tilde{\psi}\| = 1$ is called a unit ray.

A (pure) state of a system can be represented by a unit ray. In the following we disregard a minor complication which arises if there exists a (commutative) super-selection rule (BOGOLIUBOV [4], WICK [14]) and we simply assume that every vector of H represents a physically realizable state. Then there is a one-to-one correspondence between the states of the system and the unit rays in the Hilbert space H.

Below we give the definition of a symmetry transformation (or simply a symmetry) of a physical system. An example of a symmetry transformation is the translation over a certain distance of a physical system in a homogeneous (external) environment. Another example is that of a (fixed) rotation of a physical system in an isotropic environment. A transformation of a physical system is expressed by specifying the transformation of the states of the system. So in quantum mechanics a transformation of a physical system gives rise to a bijective transformation of the unit rays.

$$(2.1) \qquad \underline{T}: \underline{\psi} \subset H \to \underline{\psi}' = \underline{T\psi} \subset H'.$$

$\underline{T}$ is called a *ray transformation.*

In many instances H' is equal to H. An example where H' and H are unequal occurs in *charge conjugation,* this is the transformation of the states of the system where every particle of the system is replaced by its antiparticle and vice versa.

Intuitively a symmetry transformation is a transformation of the system in which "nothing" changes. By this we mean in particular that every transition probability between an arbitrary pair of states in the original situation is equal to the transition probability between the corresponding transformed states of the transformed system. We give the following

DEFINITION 2.1. For all rays $\underline{\phi}, \underline{\psi} \subset H$ and $\phi \in \underline{\phi}$, $\psi \in \underline{\psi}$ the *ray product* of $\underline{\phi}$ and $\underline{\psi}$ is

(2.2) $\underline{\phi} \cdot \underline{\psi} = |\langle\phi,\psi\rangle|$.

DEFINITION 2.2. A ray transformation $\underline{T}$ is called a *symmetry transformation* if

(2.3) $\underline{\phi} \cdot \underline{\psi} = (\underline{T}\ \underline{\phi}) \cdot (\underline{T}\ \underline{\psi})$

for all $\underline{\phi},\underline{\psi} \subset H$.

So symmetry transformations are transformations which preserve the transition probabilities. Symmetry transformations are produced in a simple way by unitary and by anti-unitary transformations. We recall that

DEFINITION 2.3. A transformation U: $H \to H'$ is called *unitary* if the domain of U is H and the range of U is H' and

$$\langle U\ \phi, U\ \psi\rangle = \langle\phi,\psi\rangle \quad \text{for all } \phi,\psi \in H.$$

REMARK 2.4. It can be proved (ACHIESER [1, §35, 36]), that the inverse U^{-1} exists and U^{-1} is unitary and U is linear.

DEFINITION 2.5. A transformation A: $H \to H'$ is called *anti-unitary* if the domain of A is H and the range of A is H' and if

$$\langle A\ \phi, A\ \psi\rangle = \langle\phi,\psi\rangle^* = \langle\psi,\phi\rangle \text{ for all } \phi,\psi \in H.$$

REMARK 2.6. It can be proved analogously that A^{-1} exists and that A^{-1} is anti-unitary and that furthermore A is *anti-linear,* i.e.

$$A(\lambda\psi + \mu\phi) = \lambda^* A\psi + \mu^* A\phi \text{ for all } \phi,\psi \in H \text{ and all } \lambda,\mu \in \mathbb{C}.$$

If U: $H \to H'$ is a unitary transformation then

$$\underline{U}\colon \underline{\psi} = \{e^{i\alpha}\psi \mid \alpha \in \mathbb{R}\} \to \underline{\psi}' = \{e^{i\beta}U\psi \mid \beta \in \mathbb{R}\}$$

is a symmetry transformation, because

$$\underline{\phi}' \cdot \underline{\psi}' = |\langle e^{i\alpha}\ U\phi, e^{i\beta}\ U\psi\rangle| = |\langle\phi,\psi\rangle| = \underline{\phi} \cdot \underline{\psi}\ .$$

Analogously, if A: $H \rightarrow H'$ is an anti-unitary transformation, then

$$\underline{A}: \underline{\psi} = \{e^{i\alpha}\psi | \alpha \in \mathbb{R}\} \rightarrow \underline{\psi}' = \{e^{i\beta} A\psi | \beta \in \mathbb{R}\}$$

is a symmetry transformation, because

$$\underline{\phi}' \cdot \underline{\psi}' = |\langle e^{i\alpha} A\phi, e^{i\beta} A\psi\rangle| = |\langle\psi,\phi\rangle| = \underline{\phi} \cdot \underline{\psi} .$$

In these two cases it is possible to replace the ray transformation by a transformation of the vectors.

DEFINITION 2.7. A transformation T: $H \rightarrow H'$ is called *compatible* with the ray transformation $\underline{T}$

$$\text{if } \underline{T\psi} = \underline{T}\,\underline{\psi} \text{ for all } \psi \in H.$$

The possibility of replacing in general a symmetry transformation by a compatible (anti-) unitary transformation is a consequence of the theorem of Wigner (WIGNER [16], BARGMANN [3]), which will be discussed in the next subsection.

2.2. Wigner's theorem

THEOREM 2.8 (Wigner). *If H and H' are Hilbert spaces and*

$$\underline{T}: \underline{\psi} \subset H \rightarrow \underline{\psi}' \subset H'$$

is a symmetry transformation, then there exists a transformation T: $H \rightarrow H'$ *which is compatible with* $\underline{T}$ *and such that* T *is either unitary or anti-unitary if* $\dim H \geq 2$. *If* $\dim H = 1$ *there exists a unitary transformation* T_1: $H \rightarrow H'$ *and an anti-unitary transformation* T_2: $H \rightarrow H'$ *which are both compatible with* $\underline{T}$.

In section 2.3 we give the proof of Wigner's theorem à la Bargmann and an explicit discussion of the case with $\dim H = 2$.

DEFINITION 2.9. A transformation T: $H \rightarrow H'$ is called *additive* if $T(\psi + \phi) = T\psi + T\phi$ for all $\psi, \phi \in H$.

THEOREM 2.10 (uniqueness). *Let* T_1 *and* T_2 *be additive transformations from* H *onto* H' *which are both compatible with the ray transformation* $\underline{T}$ *and let* $\dim H \geq 2$, *then* $T_2 = e^{i\alpha}T_1$ $(\alpha \in \mathbb{R})$.

The proof of this theorem is also given in the next section.

2.3. Bargmann's proof of Wigner's theorem

In this section we prove Theorem 2.8. We first discuss the trivial case $\dim H = 1$. In this case there exists only one unit ray $\underline{e} \subset H$ and $\underline{T}$ is determined by $\underline{T}\,\underline{e} = \underline{e}'$ ($\underline{e}'$ unit ray in H'). Let $e \in \underline{e}$ and $e' \in \underline{e}'$, then T_1 defined by $T_1(\alpha e) = \alpha e'$ is a unitary mapping compatible with $\underline{T}$ and T_2 defined by $T_2(\alpha e) = \alpha^* e'$ is an anti-unitary mapping compatible with $\underline{T}$.

Hereafter in this section we assume $\dim H \geq 2$.

DEFINITION 2.11. The mapping Δ is defined by

$$(2.4) \qquad \Delta(\underline{\psi}_1,\underline{\psi}_2,\underline{\psi}_3) = \langle\psi_1,\psi_2\rangle\langle\psi_2,\psi_3\rangle\langle\psi_3,\psi_1\rangle$$

for all $\psi_1,\psi_2,\psi_3 \in H$.

REMARK 2.12. The right hand side of (2.4) is indeed independent of the choice of the representatives $\psi_i \in \underline{\psi}_i$.

REMARK 2.13. If T is either unitary or anti-unitary and compatible with $\underline{T}$, then

$$(2.5) \qquad \Delta(\underline{T\psi}_1,\underline{T\psi}_2,\underline{T\psi}_3) = \chi(\Delta(\underline{\psi}_1,\underline{\psi}_2,\underline{\psi}_3))$$

with $\chi(\lambda) = \begin{cases} \lambda & \text{if T is unitary} \\ \lambda^* & \text{if T is anti-unitary} \end{cases}$

and $\lambda \in \mathbb{C}$.

For $\dim H \geq 2$ Δ is not always real.

EXAMPLE 2.14. Let e and f be an orthonormal set of vectors in H and $e_1 = e$, $e_2 = (e - f)/\sqrt{2}$ and $e_3 = [e + f(1 - i)]/\sqrt{3}$, then $\Delta(\underline{e}_1,\underline{e}_2,\underline{e}_3) = \frac{i}{6}$.

CONCLUSION 2.15. *If* $\dim H \geq 2$, *Remark 2.13 and Example 2.14 imply that if* T *is unitary and compatible with* $\underline{T}$ *then there does not exist an anti-unitary transformation compatible with* $\underline{T}$, *and conversely if* T *is an anti-unitary transformation and compatible with* $\underline{T}$ *then there does not exist a unitary transformation compatible with* $\underline{T}$.

LEMMA 2.16. *Let* $\{\underline{f}_\alpha\}$ $(\alpha = 1,\dots,k)$ *be an orthonormal set of rays, i.e.* $\underline{f}_\alpha \cdot \underline{f}_\beta = \delta_{\alpha\beta}$, *and let* $\underline{f}'_\alpha = \underline{T}\,\underline{f}_\alpha$, *with* $\underline{T}$ *a symmetry transformation. Let* $f_\alpha \in \underline{f}_\alpha$, $f'_\alpha \in \underline{f}'_\alpha$ *and* $\psi = \sum_{\alpha=1}^{k} C_\alpha f_\alpha$ *then for any* $\psi' \in \underline{T\psi}$

$$\psi' = \sum_{\alpha=1}^{k} C'_\alpha f'_\alpha, \qquad |C'_\alpha| = |C_\alpha|.$$

PROOF. Because of $\underline{f}_\alpha \cdot \underline{f}_\beta = \delta_{\alpha\beta} = \underline{f}'_\alpha \cdot \underline{f}'_\beta$ we get $\langle f_\alpha, f_\beta\rangle = \delta_{\alpha\beta} = \langle f'_\alpha, f'_\beta\rangle$. Furthermore $\|\psi'\| = \|\psi\|$ and $|\langle f_\alpha,\psi\rangle| = |\langle f'_\alpha,\psi'\rangle|$, thus $\|\psi' - \sum_{\alpha=1}^{k} \langle f'_\alpha,\psi'\rangle f'_\alpha\|^2 = \|\psi'\|^2 - \sum_{\alpha=1}^{k} |\langle f'_\alpha,\psi'\rangle|^2 = \|\psi\|^2 - \sum_{\alpha=1}^{k} |\langle f_\alpha,\psi\rangle|^2 = \|\psi - \sum_{\alpha=1}^{k} \langle f_\alpha,\psi\rangle f_\alpha\|^2 = 0$

and we conclude that $\psi' - \sum_{\alpha=1}^{k} \langle f'_\alpha,\psi'\rangle f'_\alpha = 0$. □

The transformation T compatible with $\underline{T}$ will be constructed in four steps, labelled a,b,c and d.

Step a. Choose a unit ray $\underline{e} \subset H$ and let $\underline{e}' = \underline{T}\,\underline{e}$. Let $e \in \underline{e}$ and $e' \in \underline{e}'$ then we define

$$T\,e = e'.$$

Let P be the set of vectors in H orthogonal to e and P' the set of vectors in H' orthogonal to e'. Every vector $\psi \in H$ has a unique decomposition

(2.6) $\qquad \psi = \alpha e + \phi, \qquad \phi \in P.$

Step b. We now define T on the vectors ψ of (2.6) with $\alpha = 1$ and $\phi \neq 0$ i.e. on vectors $\psi = e + \phi$ $(\phi \neq 0,\ \phi \in P)$. Because $\psi = e + \|\phi\|\hat\phi$ with $\hat\phi = \phi/\|\phi\|$, Lemma (2.16) implies for any $\psi' \in \underline{T\psi}$

$$\psi' = c_0 e' + c_1\hat\phi' \quad \text{with} \quad |c_0| = 1,\ |c_1| = \|\phi\| \text{ and } \hat\phi' \in \underline{T}\,\underline{\hat\phi} \subset P'.$$

Hence $\underline{T}\,\underline{\psi}$ contains the uniquely determined vector

$$c_0^{-1}\,\psi' = e' + c_0^{-1}\,c_1\hat{\phi}' \quad (c_0^{-1}\,c_1\hat{\phi}' \in P')$$

and we define the mapping

$$S\colon \phi \in P \to c_0^{-1}\,c_1\hat{\phi}' \in P'.$$

Finally we define T on the subset $\{e+\phi \mid \phi \neq 0, \phi \in P\}$ as follows

$$T(e+\phi) = e' + S\phi. \tag{2.7}$$

Step c. We define T on the subspace P by

$$T\phi = S\phi = T(e+\phi) - Te \quad (\phi \in P). \tag{2.8}$$

In the following lemma properties of the mapping S are given.

LEMMA 2.17. *The mapping* $S\colon P \to P'$ *has the following properties:*

$$S(\phi_1 + \phi_2) = S\phi_1 + S\phi_2, \tag{2.9}$$

$$S(\lambda\phi) = \chi(\lambda)S(\phi) \tag{2.10}$$

and

$$\langle S\phi_1, S\phi_2\rangle = \chi(\langle\phi_1,\phi_2\rangle), \tag{2.11}$$

where $\phi, \phi_1, \phi_2 \in P$, $\lambda \in \mathbb{C}$ *and* $\chi(\lambda)$ *is equal to either* λ *or* λ^* *for all* $\lambda \in \mathbb{C}$.

PROOF. Note that

$$|\langle S\phi_1, S\phi_2\rangle|^2 = |\langle\phi_1,\phi_2\rangle|^2 \tag{2.12}$$

and

$$|\langle e' + S\phi_1, e' + S\phi_2\rangle|^2 = |\langle e+\phi_1, e+\phi_2\rangle|^2$$

or

(2.13) $$|1+\langle S\phi_1,S\phi_2\rangle|^2 = |1+\langle\phi_1,\phi_2\rangle|^2$$

for $\phi_1,\phi_2 \in P$. Since $|1+z|^2 = 1+|z|^2+2\,\mathrm{Re}\,z$ $(z \in \mathbb{C})$, formulae (2.12) and (2.13) imply

(2.14) $$\mathrm{Re}\langle S\phi_1,S\phi_2\rangle = \mathrm{Re}\langle\phi_1,\phi_2\rangle.$$

It follows from (2.12) and (2.14) that

(2.15) $$\langle S\phi_1,S\phi_2\rangle = \langle\phi_1,\phi_2\rangle \quad \text{or} \quad \langle\phi_1,\phi_2\rangle^*,$$

so that in particular

(2.16) $$\langle S\phi,S\phi\rangle = \langle\phi,\phi\rangle.$$

The definition of the mapping S implies for each non-zero $\phi \in P$

(2.17) $$S(\lambda\phi) = \chi_\phi(\lambda)S\phi, \ \lambda \in \mathbb{C}, \ \chi_\phi(\lambda) \in \mathbb{C}.$$

We will prove that

(2.18) $$\chi_\phi(\lambda) = \lambda \quad \text{or} \quad \lambda^*$$

for all non-zero $\phi \in P$ and all $\lambda \in \mathbb{C}$. In fact, (2.17) implies

$$\langle S\phi,S(\lambda\phi)\rangle = \chi_\phi(\lambda)\langle S\phi,S\phi\rangle$$

and because of (2.15) and (2.16) it follows that

$$\langle\phi,\lambda\phi\rangle \quad \text{or} \quad \langle\phi,\lambda\phi\rangle^* = \chi_\phi(\lambda)\langle\phi,\phi\rangle.$$

This proves (2.18).

Now we prove that either $\chi_\phi(\lambda) = \lambda$ for all $\lambda \in \mathbb{C}$ or $\chi_\phi(\lambda) = \lambda^*$ for all $\lambda \in \mathbb{C}$. Let $\lambda,\mu \in \mathbb{C}$ and $\lambda,\mu \notin \mathbb{R}$. Suppose $\chi_\phi(\lambda) = \lambda$ and $\chi_\phi(\mu) = \mu^*$, then for $\|\phi\| = 1$

$$\lambda\mu = \langle S(\mu\phi),S(\lambda\phi)\rangle = \langle\mu\phi,\lambda\phi\rangle \quad \text{or} \quad \langle\mu\phi,\lambda\phi\rangle^*.$$

It follows that $\lambda\mu = \mu^*\lambda$ or $\mu\lambda^*$ so that μ or λ is real. This contradicts the hypothesis.

This implies

$$(2.19) \qquad \chi_\phi(\alpha\lambda) = \chi_\phi(\alpha)\chi_\phi(\lambda) \quad \text{for all} \quad \alpha,\lambda \in \mathbb{C};$$

$$(2.20) \qquad \chi_\phi(\lambda+\mu) = \chi_\phi(\lambda) + \chi_\phi(\mu) \quad \text{for all} \quad \lambda,\mu \in \mathbb{C};$$

$$(2.21) \qquad \chi_\phi(\lambda)^* = \chi_\phi(\lambda^*) \quad \text{for all} \quad \lambda \in \mathbb{C}.$$

We next prove

$$(2.22) \qquad \chi_{\alpha\phi}(\lambda) = \chi_\phi(\lambda) \quad \text{for all} \quad \alpha,\lambda \in \mathbb{C}.$$

(2.17) implies

$$S(\lambda(\alpha\phi)) = \chi_{\alpha\phi}(\lambda)S(\alpha\phi) = \chi_{\alpha\phi}(\lambda)\chi_\phi(\alpha)S\phi = \chi_\phi(\alpha\lambda)S\phi.$$

Hence $\chi_{\alpha\phi}(\lambda)\chi_\phi(\alpha) = \chi_\phi(\alpha\lambda) = \chi_\phi(\alpha)\chi_\phi(\lambda)$ because of (2.19) and this proves (2.22).

Now we prove (2.9). If ϕ_1 and ϕ_2 are linearly dependent, then there exists a vector $\phi \in P$ such that $\phi_1 = c_1\phi$ and $\phi_2 = c_2\phi$. In this case (2.20) and (2.22) imply

$$S(\phi_1+\phi_2) = S((c_1+c_2)\phi) = \chi_\phi(c_1+c_2)S\phi = \chi_\phi(c_1)S\phi + \chi_\phi(c_2)S\phi =$$
$$= S\phi_1 + S\phi_2.$$

We next consider the case where $<\phi_1,\phi_2> = 0$ and $\phi_1,\phi_2 \neq 0$. Then (2.15) implies $<S\phi_1,S\phi_2> = 0$ and (2.16) reads $<S\phi_i,S\phi_i> = <\phi_i,\phi_i>$ $(i = 1,2)$. Lemma 2.16 implies

$$S(\phi_1+\phi_2) = a_1S\phi_1 + a_2S\phi_2, \quad |a_i| = 1 \ (i = 1,2)$$

and it follows that

$$<S\phi_i,S(\phi_1+\phi_2)> = a_i<S\phi_i,S\phi_i> = a_i<\phi_i,\phi_i> =$$

$$<\phi_i,\phi_1+\phi_2> \text{ or } <\phi_i,\phi_1+\phi_2>^* = <\phi_i,\phi_i> \ (i = 1,2).$$

Hence $a_i = 1$ $(i = 1,2)$ and this proves (2.9) in case ϕ_1 and ϕ_2 are orthogonal. By combining the results of both cases we obtain (2.9) for arbitrary $\phi_1, \phi_2 \in P$.

Next we prove that the function $\chi_\phi(\lambda)$ does not depend on ϕ. Let $\phi, \psi \in P$ and non-zero. If ϕ and ψ are linearly dependent then (2.22) implies

$$(2.23) \qquad \chi_\phi = \chi_\psi$$

So suppose ϕ and ψ linearly independent, then $S\phi$ and $S\psi$ are linearly independent (this is proved analogously to Lemma 2.18 because of (2.15) and (2.16)) and for $\lambda \in \mathbb{C}$ we have

$$S(\lambda(\phi+\psi)) = \chi_{\phi+\psi}(\lambda)S(\phi+\psi) = \chi_{\phi+\psi}(\lambda)S\phi + \chi_{\phi+\psi}(\lambda)S\psi =$$

$$= S(\lambda\phi + \lambda\psi) = S(\lambda\phi) + S(\lambda\psi) = \chi_\phi(\lambda)S\phi + \chi_\psi(\lambda)S\psi,$$

hence

$$(2.24) \qquad \chi_\phi(\lambda) = \chi_\psi(\lambda) = \chi_{\phi+\psi}(\lambda).$$

By combining the results of both cases (i.e. (2.23) and (2.24)) we obtain

$$(2.25) \qquad \chi_\phi(\lambda) = \chi_\psi(\lambda) \quad \text{for all non-zero} \quad \phi, \psi \in P$$

and we define

$$(2.26) \qquad \chi(\lambda) = \chi_\phi(\lambda), \quad \phi \in P, \quad \phi \neq 0, \quad \lambda \in \mathbb{C}.$$

Evidently $\chi(\lambda)$ equals either λ or λ^* for all $\lambda \in \mathbb{C}$. (2.17) and (2.26) imply (2.10).

Next we verify (2.11) of Lemma 2.17. If $\langle\phi_1,\phi_2\rangle = 0$ then $\langle S\phi_1, S\phi_2\rangle = 0$ and (2.11) is trivially satisfied. So it is sufficient to prove (2.11) for the case that ϕ_1 and ϕ_2 are linearly dependent, i.e. $\phi_1 = c_1\phi, \phi_2 = c_2\phi$ where $\|\phi\| = 1$. Hence $\langle S\phi_1, S\phi_2\rangle = \chi^*(c_1)\chi(c_2)\langle S\phi, S\phi\rangle = \chi(c_1^* c_2) = \chi(\langle c_1\phi, c_2\phi\rangle) = \chi(\langle\phi_1,\phi_2\rangle)$. □

<u>Step d</u>. Finally we define T on the subset of vectors $\psi = \alpha e + \phi$, $\alpha \neq 0,1$ and $\phi \in P$. Let $\psi' = \alpha^{-1}\psi = e + \alpha^{-1}\phi$, then $\alpha^{-1}\phi \in P$ and $T\psi'$ is defined in step c by (2.7). We define for $\alpha \neq 0,1$

(2.27) $\quad T\psi = \chi(\alpha)T\psi'.$

Hence

$$T(\alpha e+\phi) = \chi(\alpha)T(e+\alpha^{-1}\phi) = \chi(\alpha)(e'+S(\alpha^{-1}\phi)) =$$

$$= \chi(\alpha)e' + \chi(\alpha)\chi(\alpha^{-1})S\phi = \chi(\alpha)e' + S\phi.$$

So

(2.28) $\quad T(\alpha e+\phi) = \chi(\alpha)e' + S\phi$

and this formula is also valid for $\alpha = 0,1$.

The transformation T is now defined on $H(\dim H \geq 2)$ and compatible with $\underline{T}$ and Lemma 2.17 implies that T is either unitary or anti-unitary, because

$$<T(\alpha_1 e+\phi_1),T(\alpha_2 e+\phi_2)> = <\chi(\alpha_1)e'+S\phi_1,\chi(\alpha_2)e'+S\phi_2> =$$

$$= \chi(\alpha_1)^*\chi(\alpha_2) + <S\phi_1,S\phi_2> = \chi(\alpha_1^*\alpha_2) + \chi(<\phi_1,\phi_2>) =$$

$$= \chi(\alpha_1^*\alpha_2 + <\phi_1,\phi_2>) = \chi(<\alpha_1 e,\alpha_2 e> + <\phi_1,\phi_2>) = \chi(<\alpha_1 e+\phi_1,\alpha_2 e+\phi_2>).$$

This concludes the proof of Theorem 2.8.

In the following we prove the uniqueness Theorem 2.10.

LEMMA 2.18. *If* $T: H \to H'$ *is compatible with the symmetry transformation* $\underline{T}$ *and* ψ_1, ψ_2 *are linearly independent then* $T\psi_1$, $T\psi_2$ *are linearly independent.*

PROOF. ψ_1, ψ_2 are linearly independent iff

$$G(\psi_1,\psi_2) = <\psi_1,\psi_1><\psi_2,\psi_2> - |<\psi_1,\psi_2>|^2 > 0.$$

But $G(T\psi_1,T\psi_2) = G(\psi_1,\psi_2)$. So the linear independence of ψ_1, ψ_2 implies the linear independence of $T\psi_1$, $T\psi_2$. □

LEMMA 2.19. *If* $T_1: H \to H'$ *and* $T_2: H \to H'$ *are both additive and compatible with the symmetry transformation* $\underline{T}$ *then* $T_1 0 = T_2 0 = 0$. *For all* $\psi \neq 0$ *define* $\tau(\psi) \in \mathbb{C}$ *by*

(2.29) $\quad T_2\psi = \tau(\psi)T_1\psi.$

If ψ_1 and ψ_2 are linearly independent, then $\tau(\psi_1) = \tau(\psi_2)$.

PROOF. The first part of the lemma is trivial. Let $\psi = \psi_1 + \psi_2$, then

$$\tau(\psi_1)T_1\psi_1 + \tau(\psi_2)T_1\psi_2 = T_2\psi_1 + T_2\psi_2 = T_2(\psi_1+\psi_2) = T_2\psi =$$

$$= \tau(\psi)T_1\psi = \tau(\psi)T_1\psi_1 + \tau(\psi)T_1\psi_2$$

and Lemma 2.18 asserts that $T_1\psi_1$ and $T_1\psi_2$ are linearly independent. It follows that $\tau(\psi_1) = \tau(\psi_2) = \tau(\psi)$. □

PROOF OF THEOREM 2.10. Let $\psi_0 \neq 0$ be a fixed vector in H and let $\tau(\psi_0) = \theta$, then $|\theta| = 1$. We show that $\tau(\psi) = \theta$ for all $\psi \neq 0$, $\psi \in H$. If ψ_0 and ψ are linearly independent Lemma 2.19 implies $\tau(\psi) = \tau(\psi_0) = \theta$. If ψ_0 and ψ are linearly dependent, then $\psi = \mu\,\psi_0$, $\mu \neq 0$. Choose $\phi \in H$ and linear independent of ψ_0. Hence ϕ is linearly independent of ψ. Then by Lemma 2.19 it follows that $\tau(\phi) = \tau(\psi_0)$ and $\tau(\phi) = \tau(\psi)$ so that $\tau(\phi) = \tau(\psi_0) = \theta$. □

We conclude this section by considering the case with $H = H'$ and dim $H = 2$ explicitly, because of its connection with the theory of spinors.

A unit vector in $\mathbb{C}^2$ can be represented by

$$\psi = e^{i\alpha}\begin{pmatrix}\cos\frac{\theta}{2}\\ e^{i\phi}\sin\frac{\theta}{2}\end{pmatrix}, \quad (0 \le \theta \le \pi;\ 0 \le \phi,\ \alpha < 2\pi).$$

The *Pauli matrices* are defined by

$$\sigma^1 = \begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix},\ \sigma^2 = \begin{pmatrix}0 & -i\\ i & 0\end{pmatrix},\ \sigma^3 = \begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}.$$

We define a one-to-one mapping M: $\underline{\psi} \to \vec{n}$ between the rays $\underline{\psi} \subset \mathbb{C}^2$ and the vectors $\vec{n}$ in $\mathbb{R}^3$ by $n^i = \langle\psi,\sigma^i\psi\rangle$ $(i = 1,2,3$ and $\psi \in \underline{\psi})$. If $\underline{\psi}$ is a unit ray, then $\vec{n} = (\sin\theta\ \cos\phi,\ \sin\theta\ \sin\phi,\ \cos\theta)$ is a unit vector, (θ,ϕ) are the polar angles. M maps the unit rays one-to-one onto the unit sphere in $\mathbb{R}^3$.

A symmetry transformation in ray space gives via the mapping M an isometric transformation with fixed origin in $\mathbb{R}^3$ and vice versa. This is easily verified, because

$$\delta_{\alpha\gamma}\delta_{\beta\delta} = \tfrac{1}{2}\,\vec{\sigma}_{\alpha\beta}\cdot\vec{\sigma}_{\delta\gamma} + \tfrac{1}{2}\,\delta_{\alpha\beta}\,\delta_{\delta\gamma}$$

implies

$$(\vec{n}_1 - \vec{n}_2)^2 = (\underline{\psi}_1^2 + \underline{\psi}_2^2)^2 - 4(\underline{\psi}_1\cdot\underline{\psi}_2)^2$$

with

$$\vec{n}_k = \langle\psi_k, \vec{\sigma}\psi_k\rangle \quad (k = 1,2).$$

An isometry with fixed origin in $\mathbb{R}^3$ can be a proper rotation R ($\det R = +1$) or an improper rotation R' ($\det R' = -1$). With a proper rotation in $\mathbb{R}^3$ corresponds a unitary transformation in $\mathbb{C}^2$ and this unitary transformation is compatible with the given symmetry transformation. An improper rotation can be written as a product of a rotation and a reflection or as a product of a rotation and the inversion. We consider the reflection

$$(n^1, n^2, n^3) \to (n^1, -n^2, n^3) \quad \text{or} \quad (\theta,\phi) \to (\theta, 2\pi - \phi)$$

with the (1,3)-plane as mirror plane. This reflection in $\mathbb{R}^3$ corresponds to the transformation of complex conjugation K in $\mathbb{C}$,

$$K:\ \psi = \begin{pmatrix} a \\ b \end{pmatrix} \to \psi^* = \begin{pmatrix} a^* \\ b^* \end{pmatrix}.$$

Thus with an improper rotation in $\mathbb{R}^3$ corresponds an anti-unitary transformation in $\mathbb{C}^2$, because the product of K and a unitary transformation is an anti-unitary transformation.

2.4. Symmetry groups

From now on we consider symmetry transformations with $H = H'$. The product of two symmetry transformations is a symmetry transformation. The identity ray transformation E: $\underline{\psi} \to \underline{\psi}$ is a symmetry transformation. And because symmetry transformations are one-to-one, the inverse of a symmetry transformation exists and is a symmetry transformation.

Thus a set of symmetry transformations closed under multiplication and taking the inverse forms a group; this is called a *symmetry group*. Also the corresponding abstract group is sometimes called the symmetry group. Symmetry groups play a very important role in physics. We mention the important roles of the rotation group and the permutation group in atomic physics and that of the Poincaré group (inhomogeneous Lorentz group) in relativistic physics. In elementary particle physics we come across the so-called internal symmetries, for instance the isospin group SU(2) and SU(3). In the following a symmetry group will always be supposed to be a Lie group.

Let G be a symmetry group, then to every element $g \in G$ there corresponds an (anti-) unitary transformation U(g) on H which is compatible with the ray transformation $\underline{U}(g)$. If $g_1 g_2 = g_3$ where $g_1, g_2, g_3 \in G$ then

$$\underline{U}(g_1)\ \underline{U}(g_2) = \underline{U}(g_3). \tag{2.30}$$

The ray transformations give a realization of the group G. Now the uniqueness theorem says that U(g) is determined by $\underline{U}(g)$ apart from an arbitrary phase factor. So (2.30) implies

$$U(g_1)\ U(g_2) = \omega(g_1,g_2)U(g_1g_2), \tag{2.31}$$

where $\omega(g_1,g_2)$ is a phase factor.

The operators U(g) satisfying (2.31) form a *projective representation* or a *ray representation*. We choose U always such that, for the unit element $e \in G$, $U(e) = I$ holds. The U's one encounters are very often unitary and this is due to the following. If for $g \in G$ there exists an element $h \in G$ such that $g = h^2$, then $U(g) = \omega(h,h)^{-1}\ U(h)^2$ and U(g) is unitary because the product of two unitary operators is unitary and also the product of two anti-unitary operators is unitary.

If we define $\xi(g_1,g_2)$ by

$$\omega(g_1,g_2) = e^{i\xi(g_1,g_2)}$$

so that

$$U(g_1)U(g_2) = e^{i\xi(g_1,g_2)}U(g_1g_2)$$

then $\xi(e,e) = 0$ and

$$\xi(g_1,g_2) + \xi(g_1g_2,g_3) = \xi(g_1,g_2g_3) + \xi(g_2,g_3) \pmod{2\pi}$$

because of the associativity of group multiplications. If we shift the phases of the operators U(g) by

(2.32) $$U(g) \to \hat{U}(g) = e^{i\zeta(g)}\, U(g),$$

then

(.2.33) $$\xi(g_1,g_2) \to \hat{\xi}(g_1,g_2) = \xi(g_1,g_2) + \zeta(g_1) + \zeta(g_2) - \zeta(g_1g_2),$$

where $\hat{\xi}$ is defined by

$$\hat{U}(g_1)\hat{U}(g_2) = e^{i\hat{\xi}(g_1,g_2)}\, \hat{U}(g_1g_2).$$

Thus the phase factors in the multiplication rule of the projective representations are not uniquely determined, because of (2.33).

The freedom contained in (2.33) can sometimes be used to simplify the phase factors in (2.31). For the projective representations of the group of proper rotations SO(3), one can make shifts of the phases of the unitary operators U(g) such that $\omega = \pm 1$. For the universal covering group of SO(3), that is SU(2), it is even possible to choose $\omega = 1$. If SO(3) is the symmetry group of a system, then SU(2) is called its quantum mechanical symmetry group.

In the following chapter the focus will be on the Poincaré group.

DEFINITION 2.20. The linear transformation Λ: $x^\mu \to \hat{x}^\mu = \Lambda^\mu{}_\nu x^\nu$ (μ,ν = 0,1,2,3 and summation convention) is called a *Lorentz transformation* if $\eta_{\mu\nu}\hat{x}^\mu\hat{x}^\nu = \eta_{\mu\nu}x^\mu x^\nu$, where $\eta_{\mu\nu} = \mathrm{diag}(1,-1,-1,-1)$. Λ is called a *proper* Lorentz transformation if $\Lambda^o{}_o > 0$ and det $\Lambda = +1$. The *proper Lorentz group* $L_+^\uparrow$ is the pair $(L,\cdot)$ where L is the set of proper Lorentz transformations and the multiplication $(\cdot)$ is defined by $(\Lambda_1{}^\mu{}_\nu \Lambda_2{}^\nu{}_\sigma) = (\Lambda_1\cdot\Lambda_2)^\mu{}_\sigma$.

DEFINITION 2.21. The *proper Poincaré group* $P_+^\uparrow$ is the pair $(\{(a,\Lambda)\},\cdot)$, where $a \in \mathbb{R}^4$, $\Lambda \in L_+^\uparrow$ and the multiplication $(\cdot)$ is defined by $(a_1,\Lambda_1)\cdot(a_2,\Lambda_2) = (a_1+\Lambda_1 a_2, \Lambda_1\Lambda_2)$.

For the projective representations of the Poincaré group we have with regard to the phase factors a situation analoguous to the case of the proper rotation group (WIGNER [15], BARGMANN [2]).

We now introduce the universal covering group of the proper Poincaré group $P_+^\uparrow$. We recall the homomorphism of the unimodular group SL(2,$\mathbb{C}$) and the proper Lorentz group $L_+^\uparrow$:

$$A \in SL(2,\mathbb{C}) \to \Lambda = \Lambda(A) \in L_+^\uparrow \quad ,$$

where $\Lambda(A)^\mu{}_\nu = \frac{1}{2}\,\mathrm{Tr}(\sigma^\mu A \sigma^\nu A^\dagger)\,(\mu,\nu=0,1,2,3)$, $\sigma^1,\sigma^2,\sigma^3$ are the Pauli matrices defined in section 2.3, σ^0 is the 2x2 unit matrix and $A^\dagger$ is the adjoint of A.

DEFINITION 2.22. The *inhomogeneous* SL(2,$\mathbb{C}$) (also denoted by iSL(2,$\mathbb{C}$)) is the group $(\{(a,A)\},\cdot)$ where $a \in \mathbb{R}^4$, $A \in SL(2,\mathbb{C})$ and the multiplication is defined by

$$(a_1,A_1)\cdot(a_2,A_2) = (a_1 + \Lambda(A_1)a_2, A_1A_2).$$

REMARK 2.23. The inhomogeneous SL(2,$\mathbb{C}$) is the universal covering group of the proper Poincaré group $P_+^\uparrow$.

DEFINITION 2.24. Let $\underline{U}(g)$ be a realization of a symmetry group G by means of ray transformations, then $g \to \underline{U}(g)$ is called *continuous* if $g \to \underline{\psi}.(\underline{U}(g)\underline{\psi}) \in \mathbb{R}$ is continuous for all $\psi \in \mathcal{H}$. The realization $g \to \underline{U}(g)$ is called *unitary* if for all $g \in G$ the ray transformation $\underline{U}(g)$ is a symmetry transformation.

THEOREM 2.25 (WIGNER [15]). *If* $\underline{U}(g)$ *is a continuous unitary realization of a symmetry group* G, *then there exists a neighborhood* N_0 *of the identity* $e \in G$ *and there exist unitary operators* U(g) *compatible with* $\underline{U}(g)$ *for* $g \in N_0$ *such that* U(g) *is strongly continuous on* N_0, *i.e., for all* $\varepsilon > 0$ *and for all* $\psi \in \mathcal{H}$ *there exists a neighborhood* $N_\varepsilon(g) \subset N_0$ *such that* $\|U(g')\psi - U(g)\psi\| < \varepsilon$ *for all* $g' \in N_\varepsilon(g)$.

The following theorem is crucial for the relativistic invariance of a quantum mechanical system. We write $\underline{U}(a,\Lambda) = \underline{U}(g)$ and $U(a,\Lambda) = U(g)$ for $g \in P_+^\uparrow$, and $U(a,A) = U(g)$ for $g \in iSL(2,\mathbb{C})$.

THEOREM 2.26 (WIGNER [15], BARGMANN [2]). *If* $\underline{U}(a,\Lambda)$ *is a continuous unitary realization of the proper Poincaré group* $P_+^\uparrow$ *by ray transformations, then there exists a strongly continuous unitary representation* U(a,A) *of* iSL(2,$\mathbb{C}$), *such that* U(a,A) *is compatible with* $\underline{U}(a,\Lambda(A))$ *for all* $(a,A) \in iSL(2,C)$.

In the next chapter it will become clear that it is the covering group of $P_+^\uparrow$, which plays an important role in relativistic quantum physics. BARGMANN [2] discusses the situation for other groups, see also VARADARAJAN [13, Ch. X] and further references therein. Not every projective representation can be reduced to a representation as is shown by the projective representations of the Galilei group (LEVY-LEBLOND [10] and references therein).

The author would like to thank T.H. Koornwinder for simplifying the proof of lemma 2.17.

LITERATURE

[1] ACHIESER, N.I. & I.M. GLASMANN, *Theorie der linearen Operatoren im Hilbert-Raum,* Akademie-Verlag, Berlin, 1954.

[2] BARGMANN, V., *On unitary ray representations of continuous groups,* Annals of Math. (2) 59 (1954), 1-46.

[3] BARGMANN, V., *Note on Wigner's theorem on symmetry operations,* J. Mathematical Phys. 5 (1964), 862-868.

[4] BOGOLIUBOV, N.N., A.A. LOGUNOV & I.T. TODOROV, *Introduction to axiomatic quantum field theory,* Benjamin, Reading (Mass.), 1975.

[5] BÖHM, A., *Rigged Hilbert space and mathematical description of physical systems,* in "Lectures in Theoretical Physics, Vol. 9A" (Proceedings Ninth Boulder Summer Institute for Theoretical Physics, 1966) (Brittin, Barut and Guenin, eds.), pp. 255-317, Gordon and Breach, New York, 1967.

[6] DIRAC, P.A.M., *The principles of quantum mechanics,* Oxford University Press, Oxford, 1947.

[7] GELFAND, I.M. & N.Ya. VILENKIN, *Generalized functions,* Vol. 4, Academic Press, New York, 1964.

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

[9] HELMBERG, G., *Introduction to spectral theory in Hilbert space,* North-Holland Publishing Company, Amsterdam, 1969.

[10] LEVY-LEBLOND, J., *Galilei group and nonrelativistic quantum mechanics,* J. Mathematical Phys. 4 (1963), 776-788.

[11] MESSIAH, A., *Quantum mechanics,* Vol. I, North-Holland Publishing Company, Amsterdam, 1961.

[12] NAGY, K.L., *State vector spaces with indefinite metric in quantum field theory,* Noordhoff, Groningen, 1966.

[13] VARADARAJAN, V.S., *Geometry of quantum theory,* Vol. II, Van Nostrand, New York, 1970.

[14] WICK, G.C., A.S. WIGHTMAN & E.P. WIGNER, *The intrinsic parity of elementary particles,* Phys. Rev. (2) 88 (1952), 101-105.

[15] WIGNER, E.P., *On unitary representations of the inhomogeneous Lorentz group,* Annals of Math. (2) 40 (1939), 149-204.

[16] WIGNER, E.P., *Group theory and its applications to atomic spectra,* Academic Press, New York, 1959.

III

UNITARY IRREDUCIBLE REPRESENTATIONS OF THE CONTINOUS POINCARE GROUP $P_+^\uparrow$

E.A. DE KERF

Inst.v. Theor. Fysica, Univ. van Amsterdam

CONTENTS

1. INTRODUCTION

The *Poincaré group* P is the group of inhomogeneous linear transformations on the four-dimensional space-time manifold which preserve a certain quadratic form. If we denote space-time coordinates with respect to some Lorentz frame (inertial frame) by $x^\mu = (x^0, x^1, x^2, x^3)$ with $x^0 = ct$ then the invariant form reads

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu$$

with $\eta = \mathrm{diag}(1,-1,-1,-1)$. (Throughout this chapter the summation convention is used.) The linear transformations are the mappings $x \to \hat{x}$ with

$$\hat{x}^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu .$$

Λ is a non-singular real 4×4 matrix and $a^\mu = (a^0, a^1, a^2, a^3)$ an arbitrary set of four real numbers. Invariance under these mappings means

$$ds^2 = \eta_{\mu\nu} dx^\mu dx^\nu = \eta_{\mu\nu} d\hat{x}^\mu d\hat{x}^\nu$$

and leads to a restriction on the matrices Λ:

$$\eta_{\mu\nu} \Lambda^\mu{}_\rho \Lambda^\nu{}_\kappa = \eta_{\rho\kappa} .$$

From this it follows that $(\det \Lambda)^2 = 1$ and $(\Lambda^0{}_0)^2 \geq 1$, so one has either $\det \Lambda = 1$ or $\det \Lambda = -1$ and $\Lambda^0{}_0 \geq 1$ or $\Lambda^0{}_0 \leq -1$. We will denote the transformations in P by (a,Λ). Transformations of type (a,I) are called *translations,* transformations $(0,\Lambda)$ are called *Lorentz transformations*. The composition law of the group reads

$$(a_2,\Lambda_2)(a_1,\Lambda_1) = (a_2 + \Lambda_2 a_1, \Lambda_2 \Lambda_1),$$

the identity element is $(0,I)$. The inverse of (a,Λ) is the group element $(a,\Lambda)^{-1} = (-\Lambda^{-1}a, \Lambda^{-1})$. Both the translations $T = \{(a,I)\}$ and the Lorentz transformations $L = \{(0,\Lambda)\}$ are subgroups. The abelian translation subgroup T is an invariant subgroup:

$$(b,\Lambda)(a,I)(b,\Lambda)^{-1} = (\Lambda a,I).$$

The Poincaré group is the semidirect product of T and L:

$$P = T \circledS L .$$

The Lorentz transformations are automorphisms on T.

In this chapter we will restrict ourselves to the connected component of the identity in P, the so-called *continuous Poincaré group* $P_+^\uparrow$. This is the group of transformations (a,Λ) with Λ satisfying $\det \Lambda = 1$ and $\Lambda^0{}_0 \geq 1$. Lorentz transformations having these properties constitute the subgroup $L_+^\uparrow$ of the full Lorentz group L. The interest in the group $P_+^\uparrow$ for physics stems from the fact that this group is one of the best established symmetry groups in nature. (Special relativity!) In classical physics this means that physical laws must be covariant with respect to this group. The Maxwell equations of electrodynamics provide a beautiful example of such covariant laws of physics. The implications of a symmetry for quantum physics have already been discussed in Chapter II. To each space-time transformation $(a,\Lambda) \in P_+^\uparrow$ there corresponds a transformation $\underline{T}(a,\Lambda)$ on the states $\{\underline{\phi}\}$ of a physical system. If $\underline{\phi}$ is a state then $\hat{\underline{\phi}} = \underline{T}(a,\Lambda)\underline{\phi}$ is a state. The mapping of states is such that the transition probability $(\underline{\chi}\cdot\underline{\phi})^2 = |\langle\chi|\phi\rangle|^2$ (i.e. the probability that a measurement yields $\underline{\chi}$ if the system is originally in the state $\underline{\phi}$) is invariant under this mapping:

$$\hat{\underline{\chi}}\cdot\hat{\underline{\phi}} = \underline{T}(a,\Lambda)\underline{\chi}\cdot\underline{T}(a,\Lambda)\underline{\phi} = \underline{\chi}\cdot\underline{\phi}.$$

Moreover the group property is maintained:

$$\underline{T}(a_2,\Lambda_2)\underline{T}(a_1,\Lambda_1)\underline{\phi} = \underline{T}(a_2 + \Lambda_2 a_1,\Lambda_2\Lambda_1)\underline{\phi},\ \forall\underline{\phi}.$$

Using Wigner's theorem (Theorem II.2.8) we can study the mapping of state vectors. To a transformation $\underline{T}(a,\Lambda)$ corresponds a transformation $U(a,\Lambda)$ which is compatible with $\underline{T}(a,\Lambda)$, either unitary or anti-unitary and determined up to a phase factor. The operators $U(a,\Lambda)$ obey the rule

$$U(a_2,\Lambda_2)U(a_1,\Lambda_1) = \omega(a_2,\Lambda_2;a_1,\Lambda_1)U(a_2 + \Lambda_2 a_1,\Lambda_2\Lambda_1),$$

thus giving a projective representation of $P_+^\uparrow$. It has been shown that the operators $U(a,\Lambda)$ are unitary and that the phase factors can be chosen in such a way that one has either

$$U(a_2,\Lambda_2)U(a_1,\Lambda_1) = U(a_2 + \Lambda_2 a_1, \Lambda_2\Lambda_1)$$

or

$$U(a_2,\Lambda_2)U(a_1,\Lambda_1) = \pm\, U(a_2 + \Lambda_2 a_1, \Lambda_2\Lambda_1),$$

cf. Theorem II.2.26. In the latter case one speaks of a double-valued representation. The double-valuedness can be done away with if one considers instead of the Lorentz group $L_+^\uparrow$ the universal covering group $SL(2,\mathbb{C})$ of $L_+^\uparrow$. There is a homomorphic mapping from $SL(2,\mathbb{C})$ onto $L_+^\uparrow$. Denoting an element of $SL(2,\mathbb{C})$ by A this homomorphism is such that the elements A and -A are mapped onto the same element $\Lambda(A) = \Lambda(-A)$ in $L_+^\uparrow$. In our analysis we will take the inhomogeneous $SL(2,\mathbb{C})$ group. Group multiplication is then written as

$$(a_2,A_2)(a_1,A_1) = (a_2 + \Lambda(A_2)a_1, A_2A_1).$$

For the representation of this group one has

$$U(a_2,A_2)\ U(a_1,A_1) = U(a_2 + \Lambda(A_2)a_1, A_2A_1).$$

We will determine the irreducible unitary representations of the covering group which is sometimes called the *quantum mechanical Poincaré group*. It will turn out that there are representations characterized by two real numbers m and s. The number m is the mass of a closed system and s is the intrinsic angular momentum (spin). Throughout our discussion the emphasis will be on the physical aspects. Mathematical rigour will be left to the mathematicians.

From the mathematical point of view the representation theory of the Lie group $P_+^\uparrow$ is an interesting example of the representation theory for semidirect products with abelian normal subgroup. The method used in obtaining the irreducible representations is called the *method of induced representations*. This means that the representations of the full group are obtained from representations of certain subgroups.

So far for the introduction. This chapter is organized as follows. In section 2 we describe in some detail the groups $P_+^\uparrow$ and $L_+^\uparrow$. In section 3 we discuss SL(2,ℂ) and the homomorphism onto $L_+^\uparrow$. Section 4 is devoted to the construction of the unitary irreducible representations of the quantum mechanical Poincaré group. In section 5 we summarize the results and we make a digression to relativistic wave equations. Section 6 contains some information on the rotation group, the group SU(2) and on the representations of these groups. In section 7 we discuss the concept of Rigged Hilbert space. In such a space the eigenvalue problem of Hermitian operators with a continuous spectrum can be formulated properly.

2. THE GROUPS $L_+^\uparrow$ AND $P_+^\uparrow$

The coordinates of an event x in space-time are given with respect to an inertial frame K by the real numbers

$$x^\mu = (x^0,x^1,x^2,x^3) = (x^0,\vec{x}), \qquad \text{where } x^0 = ct.$$

The distance between two events x and y is the quadratic form

$$(2.1) \qquad (x-y)^2 = (x^0-y^0)^2 - (\vec{x}-\vec{y})^2 = \eta_{\mu\nu}(x-y)^\mu(x-y)^\nu$$

with

$$(\eta)_{\mu\nu} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

The *Poincaré group* P is the group of real linear transformations $x \to \hat{x}$ of the form

$$(2.2) \qquad \hat{x}^\mu = \Lambda^\mu{}_\nu x^\nu + a^\mu$$

under which (2.1) is invariant, i.e.

$$(2.3) \qquad ds^2 = \eta_{\mu\nu}dx^\mu dx^\nu = \eta_{\mu\nu}d\hat{x}^\mu d\hat{x}^\nu .$$

From (2.2) and (2.3) one obtains a condition on the matrices Λ:

(2.4) $$\eta_{\mu\nu}\Lambda^{\mu}{}_{\kappa}\Lambda^{\nu}{}_{\lambda} = \eta_{\kappa\lambda} \quad .$$

Taking $\kappa = \lambda = 0$ one finds from (2.4)

$$(\Lambda^{0}{}_{0})^{2} = 1 + \sum_{i=1}^{3} (\Lambda^{i}{}_{0})^{2} \quad ,$$

which shows that $\Lambda^{0}{}_{0}$ is either ≥ 1 or ≤ -1. Writing the left hand side of (2.4) as matrix multiplication

(2.5) $$(\Lambda^{T})^{\kappa}{}_{\mu}\eta_{\mu\nu}\Lambda^{\nu}{}_{\lambda} = \eta_{\kappa\lambda},$$

we obtain $(\det \Lambda)^{2} = 1$.

The set of matrices satisfying (2.4) is called the *Lorentz group* L. The elements of the Poincaré group P (inhomogeneous Lorentz group) are denoted by (a,Λ). The multiplication rule reads

(2.6) $$(a_2,\Lambda_2)(a_1,\Lambda_1) = (a_2 + \Lambda_2 a_1, \Lambda_2\Lambda_1).$$

It has already been stated in the introduction that P is the semi-direct product of the abelian subgroup $T = \{(a,I)\}$ of space-time translations and the Lorentz group L: $P = T \circledS L$. T is an invariant subgroup, i.e.

(2.7) $$(b,\Lambda)^{-1}(a,I)(b,\Lambda) = (\Lambda^{-1}a,I).$$

From (2.7) we obtain the useful formula

(2.8) $$(a,\Lambda) = (a,I)(0,\Lambda) = (0,\Lambda)(\Lambda^{-1}a,I).$$

The Lorentz group L contains the subgroup

$$L_{+}^{\uparrow} = \{\Lambda,\ \Lambda \in L \mid \det \Lambda = 1,\ \Lambda^{0}{}_{0} \geq 1\} \quad ,$$

which is called the *proper orthochronous Lorentz group*.

The definition of vectors and tensors with respect to the Lorentz group L is straightforward. A *contravariant vector* V is a four-component object $V^{\mu} = (V^0,V^1,V^2,V^3)$ with the transformation law

$$\hat{V}^\mu = \Lambda^\mu{}_\nu V^\nu \quad . \tag{2.9}$$

The invariant tensor $(\eta)_{\mu\nu}$ can be used to define the *covariant* components of V:

$$V_\mu = \eta_{\mu\nu} V^\nu \quad . \tag{2.10}$$

The *invariant "length"* of V is

$$(V)^2 = (V^0)^2 - \vec{V}^2 = \eta_{\mu\nu} V^\mu V^\nu = \eta_{\mu\nu} \hat{V}^\mu \hat{V}^\nu = V_\mu V^\mu \quad .$$

A vector will be called *time-like* if $(V)^2 > 0$, *light-like* if $(V)^2 = 0$ and *space-like* if $(V)^2 < 0$.

From now on we will restrict ourselves to the group $L_+^\uparrow$ and the corresponding inhomogeneous group $P_+^\uparrow = \{(a,\Lambda) \mid \Lambda \in L_+^\uparrow\}$. Transformations from $L_+^\uparrow$ have the property that vectors V with $(V)^2 > 0$ and $V^0 > 0$ are transformed into vectors $\hat{V}$ with $\hat{V}^0 > 0$ and $(\hat{V})^2 > 0$. We will now give some examples of Lorentz transformations from $L_+^\uparrow$.

<u>EXAMPLE 2.1</u>. *Rotations*.

Taking $\Lambda^0{}_0 = 1$ we obtain from (2.4) that Λ must have the form

$$\Lambda = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & R & \\ 0 & & & \end{pmatrix}$$

with the matrix R satisfying $R^T R = I$. So $\hat{x}^0 = x^0$ and $\hat{x}^i = R^i{}_j x^j$. These transformations are isomorphic with 3-dimensional rotations. Specifying a rotation by the rotational axis $\vec{n}(\vec{n}^2 = 1)$ and the angle θ we have

$$\hat{\vec{x}} = \vec{x} + (\vec{n} \wedge \vec{x})\sin\theta + \vec{n} \wedge (\vec{n} \wedge \vec{x})(1-\cos\theta), \qquad \hat{x}^0 = x^0. \tag{2.11}$$

See also section 6.4. We will sometimes use the notation

$$\hat{x}^\mu = R(\vec{n},\theta)^\mu{}_\nu\, x^\nu , \tag{2.12}$$

where it is understood that $R^0{}_0 = 1$ and $R^0{}_i = R^i{}_0 = 0$. The rotations form a subgroup of $L_+^\uparrow$. A rotation $(\vec{n},\Theta)$ can be represented by a point with coordinates $(n_1\Theta, n_2\Theta, n_3\Theta)$. The parameter space is a ball with radius π. Identifying points $\vec{n}\pi$ and $-\vec{n}\pi$ on the surface we have a one-to-one correspondence between rotations and points at the ball. This shows that the rotation group is a compact group.

EXAMPLE 2.2. *Special Lorentz transformations.*

From the passive point of view special Lorentz transformations give the relation between the coordinates x^μ of a space-time event with respect to an inertial frame K and the coordinates $\hat{x}^\mu$ of the same event with respect to an inertial frame $\hat{K}$ which moves with respect to K with constant velocity $\vec{v}$. The spatial axes of $\hat{K}$ are parallel to those of K (no rotation). If we take $\hat{K}$ and K such that there origins coincide at $x^0 = 0$, the special Lorentz transformation reads

$$\text{(2.13)} \qquad \begin{cases} \hat{x}^0 = \gamma x^0 - \gamma\vec{\beta}\cdot\vec{x}, \\ \hat{\vec{x}} = -\gamma\vec{\beta}x^0 + \dfrac{\gamma^2}{\gamma+1}\vec{\beta}(\vec{\beta}\cdot\vec{x}) + \vec{x}, \end{cases}$$

with $\vec{\beta} = \dfrac{\vec{v}}{c}$, $\gamma = (1 - \vec{\beta}^2)^{-\frac{1}{2}}$, $|\vec{\beta}| < 1$ and c the speed of light in vacuum. We will adopt the active point of view. The special Lorentz transformation is then the mapping $x \to \hat{x}$ with

$$\text{(2.14)} \qquad \begin{cases} \hat{x}^0 = \gamma x^0 + \gamma\vec{\beta}\cdot\vec{x}, \\ \hat{\vec{x}} = \vec{x} + \gamma\vec{\beta}x^0 + \dfrac{\gamma^2}{\gamma+1}\vec{\beta}(\vec{\beta}\cdot\vec{x}). \end{cases}$$

A special Lorentz transformation is completely determined by the three components of $\vec{\beta}$. Instead of these parameters one can use a unit vector $\vec{m} = \vec{\beta}/|\vec{\beta}|$ and a parameter $\chi \geq 0$ defined by $\gamma = \cosh\chi$, $\gamma|\vec{\beta}| = \sinh\chi$. In this parametrization (2.14) reads

$$\text{(2.15)} \qquad \begin{cases} \hat{x}^0 = x^0\cosh\chi + \vec{m}\cdot\vec{x}\sinh\chi, \\ \hat{\vec{x}} = \vec{x} + x^0\vec{m}\sinh\chi + \vec{m}(\vec{m}\cdot\vec{x})(\cosh\chi - 1). \end{cases}$$

The special Lorentz transformations do not form a subgroup. The product of two special transformations is in general not a special transformation. Special Lorentz transformations may be represented by points $\vec{m}\chi$ with $\chi \geq 0$

but otherwise arbitrary. This shows that the Lorentz group $L_+^\uparrow$ is a non-compact group.

It can be shown that any Lorentz transformation Λ can be written as the product of a rotation and a special Lorentz transformation. A Lorentz transformation can therefore be characterized completely by the six parameters $(\vec{n},\theta;\vec{m},\chi)$.

The tangent space of $L_+^\uparrow$ is easily obtained from (2.11) and (2.15). Taking $\vec{n}_1 = (1,0,0)$, $\vec{n}_2 = (0,1,0)$, $\vec{n}_3 = (0,0,1)$ we get by differentiation with respect to Θ and putting $\Theta = 0$ the tangent vectors

$$I^1 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \equiv I^{23}, I^2 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \equiv I^{31} ,$$

(2.16)

$$I^3 = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} = I^{12} .$$

From (2.15) one finds by putting $\vec{m}_1 = (1,0,0)$, $\vec{m}_2 = (0,1,0)$ and $\vec{m}_3 = (0,0,1)$ and by differentiation with respect to χ:

$$I^{01} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} , I^{02} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} ,$$

(2.17)

$$I^{03} = \begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} .$$

The six matrices form a basis for the Lie algebra of $L_+^\uparrow$. An arbitrary element F of the algebra can be written in the form

(2.18) $$F = \frac{1}{2}\,\omega_{\mu\nu} I^{\mu\nu} ,$$

where we have defined $I^{\mu\nu} = -I^{\nu\mu}$ so that $\omega_{\mu\nu}$ can be taken anti-symmetric: $\omega_{\mu\nu} = -\omega_{\nu\mu}$. The matrices $I^{\mu\nu}$ are explicitly given by

$$(2.19) \qquad (I^{\mu\nu})^{\kappa}{}_{\lambda} = \eta^{\mu\kappa}\eta^{\nu}{}_{\lambda} - \eta^{\nu\kappa}\eta^{\mu}{}_{\lambda},$$

where $\eta^{\mu\kappa} = \eta_{\mu\kappa}$ and $\eta^{\mu}{}_{\lambda} = \delta^{\mu}{}_{\lambda}$. The exponential mapping gives for the group elements

$$(2.20) \qquad \Lambda = e^{\frac{1}{2}\omega_{\mu\nu}I^{\mu\nu}} .$$

One also uses a slightly different convention. With the definition

$$(2.21) \qquad K^{\mu\nu} = i\ I^{\mu\nu},$$

a group element takes the form

$$(2.22) \qquad \Lambda = e^{-\frac{1}{2}i\omega_{\mu\nu}K^{\mu\nu}} .$$

The matrices $K^{\mu\nu}$ have the commutation relations

$$[K^{\mu\nu},K^{\kappa\lambda}] = i(\eta^{\mu\lambda}K^{\nu\kappa} + \eta^{\nu\kappa}K^{\mu\lambda} - \eta^{\mu\kappa}K^{\nu\lambda} - \eta^{\nu\lambda}K^{\mu\kappa}) .$$

The hermitian matrices $J^1 = K^{23}$, $J^2 = K^{31}$, $J^3 = K^{12}$ are called the generators of rotations. The anti-hermitian matrices $K^1 = K^{01}$, $K^2 = K^{02}$ and $K^3 = K^{03}$ are the generators of special Lorentz transformations. Expressed in $\vec{J}$ and $\vec{K}$ the commutation relations read

$$(2.23) \qquad [J^k,J^{\ell}] = i\ \varepsilon^{klm}J^m,$$

$$(2.24) \qquad [J^k,K^{\ell}] = i\ \varepsilon^{klm}K^m,$$

$$(2.25) \qquad [K^k,K^{\ell}] = -i\ \varepsilon^{klm}J^m.$$

The complexification of the Lie algebra of $L_+^{\uparrow}$ can be decomposed into the sum of two 3-dimensional algebras. To this end one defines

$$(2.26) \qquad \vec{A} = \frac{1}{2}(\vec{J} + i\vec{K}),\ \vec{B} = \frac{1}{2}(\vec{J} - i\vec{K}).$$

The matrices A^1,A^2,A^3 and B^1,B^2,B^3 have the same commutation relations as J^1,J^2,J^3 while the commutator of an A and a B is zero. This decomposition facilitates the search for the finite-dimensional representations of the Lorentz group $L_+^\uparrow$.

3. THE GROUP SL(2,$\mathbb{C}$) AND THE HOMOMORPHISM ONTO $L_+^\uparrow$

The group SL(2,$\mathbb{C}$) is the group of 2×2 complex matrices A with $\det A = 1$. We show that this group can be mapped homomorphically onto $L_+^\uparrow$. The mapping is two-to-one. The elements A and -A have the same image $\Lambda(A) = \Lambda(-A)$. To obtain an explicit form of $\Lambda(A)$ we associate with a 4-vector $x^\mu = (x^0,x^1,x^2,x^3)$ the hermitian matrix

$$(3.1)\qquad X = \begin{pmatrix} x^0 - x^3 & -x^1 + ix^2 \\ -x^1 - ix^2 & x^0 + x^3 \end{pmatrix} = x^0\sigma^0 - \vec{x}\cdot\vec{\sigma} = x^\mu\sigma_\mu ,$$

where σ^0 is the unit matrix and $\sigma^1,\sigma^2,\sigma^3$ are the three Pauli matrices, cf. Section 6.4. The matrix X satisfies

$$(3.2)\qquad X^\dagger = X,\ \det X = (x^0)^2 - (\vec{x})^2 = x^2,\quad \mathrm{Tr}\, X = 2x^0.$$

As any 2×2 hermitian matrix can be written in the form (3.1) we have a one-to-one correspondence between 2×2 hermitian matrices and elements of $\mathbb{R}^4$.

Let $A \in SL(2,\mathbb{C})$ then $A^\dagger \in SL(2,\mathbb{C})$ and we can consider the following transformation of X:

$$(3.3)\qquad \hat{X} = A\, X\, A^\dagger.$$

The matrix $\hat{X}$ satisfies

$$(3.4)\qquad \hat{X}^\dagger = \hat{X} \quad\text{and}\quad \det\hat{X} = \det X.$$

Because of the hermiticity we can write $\hat{X}$ in the form (3.1) :

$$(3.5)\qquad \hat{X} = \hat{x}^\mu\sigma_\mu = A(x^\nu\sigma_\nu)A^\dagger.$$

As $\det\hat{X} = \det X$ we have

(3.6) $$\hat{x}^2 = x^2.$$

The transformation (3.3) gives a linear mapping $x \to \hat{x}$ with conservation of the Lorentz length. So we conclude that this mapping is a Lorentz transformation

(3.7) $$\hat{x}^\mu = \Lambda(A)^\mu{}_\nu x^\nu.$$

It is readily verified that the mapping $A \to \Lambda(A)$ is a homomorphism. To obtain the explicit form of $\Lambda(A)$ we need some properties of the matrices σ. The set $\sigma^\mu = (\sigma^0,\sigma^1,\sigma^2,\sigma^3)$ contains the unit matrix and the Pauli matrices. The set σ_μ is defined as $\sigma_\mu = \eta_{\mu\nu}\sigma^\nu = (\sigma^0,-\sigma^1,-\sigma^2,-\sigma^3)$. We will also use the notation $\tilde{\sigma}_\mu = \sigma^\mu$ and $\tilde{\sigma}^\mu = \sigma_\mu$. It is now easy to see that

(3.8) $$\mathrm{Tr}(\tilde{\sigma}^\mu\sigma_\rho) = 2\delta^\mu{}_\rho.$$

Considering again (3.5)

$$\hat{x}^\mu\sigma_\mu = A(x^\nu\sigma_\nu)A^\dagger = (A\sigma_\nu A^\dagger)x^\nu,$$

we obtain using (3.8)

$$\hat{x}^\mu = \frac{1}{2}\,\mathrm{Tr}(\tilde{\sigma}^\mu A\sigma_\nu A^\dagger)x^\nu,$$

which yields

(3.9) $$\Lambda(A)^\mu{}_\nu = \frac{1}{2}\,\mathrm{Tr}(\tilde{\sigma}^\mu A\sigma_\nu A^\dagger).$$

The kernel of the homomorphism is obtained by looking for those A which yield $\Lambda(A)^\mu{}_\nu = \delta^\mu{}_\nu$. The kernel consists of the matrices I and $-I \in SL(2,\mathbb{C})$. We have therefore a two-to-one mapping from $SL(2,\mathbb{C})$ to L.

We must now show that we have a homomorphism onto $L^\uparrow_+$. A first indication is directly obtained from (3.9). Calculating $\Lambda(A)^0{}_0$ one finds

(3.10) $$\Lambda(A)^0{}_0 = \frac{1}{2}\,\mathrm{Tr}(A\,A^\dagger) > 0.$$

We will proceed in another direction and use the so-called *polar decomposition*. Any element $A \in SL(2,\mathbb{C})$ can uniquely be written as the product of a unitary matrix U and a positive definite hermitian matrix H:

$$A = UH. \tag{3.11}$$

The proof of (3.11) is by construction. Define the matrix H by

$$H^2 = A^{\dagger}A. \tag{3.12}$$

The matrix $A^{\dagger}A$ is positive definite and hermitian. Both eigenvalues of $A^{\dagger}A$ are real and positive. We can therefore define the hermitian matrix H uniquely by taking the positive roots:

$$H = (A^{\dagger}A)^{\frac{1}{2}}, \ \det H = 1. \tag{3.13}$$

The matrix U defined by $U = A\,H^{-1}$ is then a unitary matrix with det U = 1:

$$U^{\dagger} = H^{-1}A^{\dagger}; \ U^{\dagger}U = H^{-1}A^{\dagger}AH^{-1} = I. \tag{3.14}$$

The unitary unimodular matrices U in SL(2,$\mathbb{C}$) constitute the subgroup SU(2). These matrices can be parametrized by a unit vector $\vec{n}$ and an angle $\Theta(0 \le \Theta < 2\pi)$:

$$U(\vec{n},\Theta) = I \cos \frac{1}{2}\Theta - i\,\vec{n}\cdot\vec{\sigma} \sin \frac{1}{2}\Theta = \exp[-i\,\Theta\vec{n}\cdot\vec{\sigma}/2]. \tag{3.15}$$

For details we refer to Section 6.

Substituting (3.15) in (3.9) we obtain the Lorentz transformations corresponding to the unitary subgroup. A direct calculation reveals that $\Lambda(U) = \Lambda(\vec{n},\Theta)$ is just the rotation given by formula (2.12). So the unitary subgroup is mapped onto the subgroup of rotations of $L_+^{\uparrow}$.

The subset of hermitian matrices H is not a subgroup. A matrix H with the aforementioned properties can be written in the form

$$H = h_0\sigma^0 + h_1\sigma^1 + h_2\sigma^2 + h_3\sigma^3 \tag{3.16}$$

with h_0, h_1, h_2, h_3 real and

$$\det H = h_0^2 - h_1^2 - h_2^2 - h_3^2 = 1, \quad \mathrm{Tr}\,H = 2h_0 > 0. \tag{3.17}$$

Instead of the parameters h we can use a unit vector $\vec{m} = \vec{h}/|\vec{h}|$ and a parameter $\chi \geqq 0$ defined by

(3.18) $$h_0 = \cosh \chi/2, \quad |\vec{h}| = \sinh \chi/2.$$

With this parametrization (3.16) becomes

(3.19) $$H = \sigma^0 \cosh \chi/2 + \vec{m}\cdot\vec{\sigma} \sinh \chi/2$$

$$= \exp[\chi\vec{m}\cdot\vec{\sigma}/2] = \exp[-i\chi\vec{m}\cdot i\vec{\sigma}/2].$$

Substitution of (3.19) in (3.9) yields after some calculation the Lorentz transformation $\Lambda(H) = \Lambda(\vec{m},\chi)$ corresponding to $H(\vec{m},\chi)$. The result is precisely the matrix for a special Lorentz transformation given in formula (2.15). From the way of parametrization it is clear that one gets all of $L_+^\uparrow$.

We give some examples of special Lorentz transformation which will be useful in the analysis of the irreducible representations of the inhomogeneous SL(2,ℂ) group. From relativistic mechanics we know that the energy E and the momentum $\vec{p}$ of a particle with mass m are the components of a 4-vector $p^\mu = (p^0,\vec{p})$ with $p^0 = E/c$. The invariant length of this vector is $(p)^2 = (p^0)^2 - (\vec{p})^2 = m^2c^2$. For a particle at rest with respect to some Lorentz frame the momentum $\vec{p}$ is zero and $p = (mc,\vec{0})$. We will denote this particular vector by $\underset{0}{p}$. We can now ask for the special Lorentz transformation which transforms $\underset{0}{p}$ in p. The answer is most easily given in terms of SL(2,ℂ). We represent $\underset{0}{p} = (mc,\vec{0})$ and $p^\mu = (p^0,\vec{p})$ by the matrices

$$\underset{0}{P} = \begin{pmatrix} mc & 0 \\ 0 & mc \end{pmatrix} = \underset{0}{p}^\mu \sigma_\mu, \qquad P = \begin{pmatrix} p^0-p^3 & -p^1+ip^2 \\ -p^1-ip^2 & p^0+p^3 \end{pmatrix} = p^\mu \sigma_\mu$$

and look for the hermitian matrix H which satisfies

(3.20) $$H \underset{0}{P} H = P.$$

Solving (3.20) for H gives

$$H(p) = \frac{mc\sigma^0 + p^\mu \sigma_\mu}{[2\,mc\,(p^0 + mc)]^{\frac{1}{2}}} \tag{3.21}$$

The inverse matrix $H^{-1}(p)$ transforms the 4-momentum p^μ satisfying $(p)^2 = m^2c^2$ and $p^0 > 0$ into $\underset{0}{p} = (mc, o)$. This shows that any 4-momentum on the hyperboloid $(p)^2 = (p^0)^2 - (\mathbf{p})^2 = m^2c^2\,(p^0 > 0)$ can be mapped into the point $\underset{0}{p}$. In fact for any two points p and q on the hyperboloid there exists a matrix A in SL(2,$\mathbb{C}$) which connects these two points. An example of such a matrix is given by

$$A_{q\leftarrow p} = H(q)H^{-1}(p). \tag{3.22}$$

We call such a hyperboloid an *orbit*. The group SL(2,$\mathbb{C}$) (and therefore $L_+^\uparrow$) acts transitively on this orbit.

Another example which shows the advantage of working in SL(2,$\mathbb{C}$) instead of in $L_+^\uparrow$ is the solution of the following problem: determine the subgroup of $L_+^\uparrow$ which leaves invariant an arbitrary but fixed p^μ on the hyperboloid $(p)^2 = m^2c^2(p^0 > 0)$. In SL(2,$\mathbb{C}$) this means that we must determine the subgroup of matrices A_p which satisfy

$$A_p\, p^\mu \sigma_\mu A_p^\dagger = p^\mu \sigma_\mu. \tag{3.23}$$

This subgroup is called the *little group* (or stabilizer) belonging to p. To determine the structure of this subgroup we remark that the little groups belonging to different points on the same orbit are isomorphic. Let q and p $(q \neq p)$ be on the same orbit and let $\{A_q\}$ be the little group of q:

$$A_q q^\mu \sigma_\mu A_q^\dagger = q^\mu \sigma_\mu. \tag{3.24}$$

Using the transitivity of SL(2,$\mathbb{C}$) on the orbit we can establish a one-to-one relationship between $\{A_p\}$ and $\{A_q\}$:

$$A_{q\leftarrow p} A_p A^{-1}_{q\leftarrow p} = A_q. \tag{3.25}$$

Formula (3.25) gives to each A_p a uniquely determined A_q. The matrix $A_{q\leftarrow p}$ is not unique. Instead of $A_{q\leftarrow p}$ one can take $A_q\, A_{q\leftarrow p}$ where A_q is any element of the little group of q. This shows that it is in fact the right coset $\{A_q\}.\, A_{q\leftarrow p}$ which does the job. This isomorphism between the little groups allows us to take the most simple vector on the orbit, i.e. $\underset{0}{p}$, to study the structure of the little group. The matrices $A_{\underset{0}{p}}$ must satisfy

$$(3.26) \qquad A_{\underset{0}{p}} \begin{pmatrix} mc & 0 \\ 0 & mc \end{pmatrix} A^{\dagger}_{\underset{0}{p}} = \begin{pmatrix} mc & 0 \\ 0 & mc \end{pmatrix} .$$

From (3.26) one obtains $A_p A_p^{\dagger} = I$. The little group of $\underset{0}{p}$ is therefore the unitary subgroup SU(2).

For light-like vectors things are more complicated. A light-like vector $k^{\mu} = (k^0,\vec{k})$ satisfies $(k)^2 = 0$, $k^0 > 0$. It is not hard to show that any light-like vector can be obtained from $\underset{0}{k} = (1,0,0,1)$ by performing a special Lorentz transformation in the z-direction which transforms $\underset{0}{k}$ into $\hat{k} = (k^0,0,0,k^0)$ and a rotation which transforms $\hat{k}$ into k. To determine the structure of the little group of a light-like vector we can therefore study the matrices $A_{\underset{0}{k}} \in SL(2,\mathbb{C})$ which satisfy

$$(3.27) \qquad A_{\underset{0}{k}} \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} A^{\dagger}_{\underset{0}{k}} = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} .$$

We will not go into the details but merely state the result. The little group of a light-like vector is isomorphic to the group E(2), the euclidean group in the 2-dimensional euclidean space.

As a final example we consider the little group of the four-vector $p \equiv 0$. This is of course all of $SL(2,\mathbb{C})$.

4. UNITARY IRREDUCIBLE REPRESENTATIONS OF THE QUANTUM MECHANICAL POINCARÉ GROUP

4.1. Energy-momentum eigenstates

The statement that the inhomogeneous $SL(2,\mathbb{C})$ group is a group of symmetry transformations for a quantum system means that to each element (a,A) there corresponds a unitary or anti-unitary operator $\underline{\underline{U}}(a,A)$ on the Hilbert space H of quantum states. The operators $\underline{\underline{U}}(a,A)$ satisfy *)

$$(4.1) \qquad \underline{\underline{U}}(a_2,A_2)\underline{\underline{U}}(a_1,A_1) = \underline{\underline{U}}(a_2 + \Lambda(A_2)a_1, A_2A_1).$$

*) From now on the operators U on the Hilbert space will be denoted by $\underline{\underline{U}}$.

We show that the operators $\underline{\underline{U}}$ are unitary operators. This follows from the fact that any transformation (a,A) can be written as the square of a transformation (b,B) ($B \in SL(2,\mathbb{C})$):

$$(4.2) \qquad (a,A) = (b,B)(b,B).$$

For the corresponding operators we have

$$(4.3) \qquad \underline{\underline{U}}(a,A) = \underline{\underline{U}}(b,B)\underline{\underline{U}}(b,B).$$

Now $\underline{\underline{U}}(b,B)$ is either a unitary or an anti-unitary operator. In both cases the product, i.e. $\underline{\underline{U}}(a,A)$, is a unitary operator.

The problem is now to determine all irreducible sets of unitary operators satisfying (4.1). To solve this problem we consider first the operators of the abelian subgroup of space-time translations $\{(a,I)\}$:

$$(4.4) \qquad \underline{\underline{U}}(a,I)\underline{\underline{U}}(b,I) = \underline{\underline{U}}(b,I)\underline{\underline{U}}(a,I) = \underline{\underline{U}}(a+b,I).$$

Notice that the symbol a stands for $a^\mu = (a^0,a^1,a^2,a^3)$. The unitary operators $U(a,I)$ can be represented by **)

$$(4.5) \qquad \underline{\underline{U}}(a,I) = e^{i\, a_\mu P^\mu}$$

with

$$(4.6) \qquad P^\mu = -i \left.\frac{\partial \underline{\underline{U}}(a,I)}{\partial a_\mu}\right|_{a=0}.$$

The (generally unbounded) hermitian operators $P^\mu = (P^0,\vec{P})$, being the generators of translations, are interpreted as the operators of *momentum* ($\vec{P}$) and *energy* (P^0). As the translation group is an abelian group these operators commute:

$$(4.7) \qquad [P^\mu,P^\nu] = 0.$$

**) We should have written $e^{ia_\mu P^\mu/\hbar}$, where $\hbar$ is Planck's constant divided by 2π. We will omit all factors of $\hbar$ and c.

Next we exploit the fact that the translation subgroup is an invariant subgroup. From (2.7) we obtain

$$\underline{\underline{U}}^{-1}(b,A)\underline{\underline{U}}(a,I)\underline{\underline{U}}(b,A) = \underline{\underline{U}}(\Lambda^{-1}(A)a,I). \tag{4.8}$$

So

$$\underline{\underline{U}}^{-1}(b,A)e^{ia_\mu P^\mu}\underline{\underline{U}}(b,A) = e^{i(\Lambda^{-1}(A)a)_\mu P^\mu} = e^{ia_\mu(\Lambda(A)P)^\mu}. \tag{4.9}$$

Differentiation of (4.9) with respect to a_μ and taking $a_\mu = 0$ afterwards, gives

$$\underline{\underline{U}}^{-1}(b,A)P^\mu\underline{\underline{U}}(b,A) = \Lambda(A)^\mu{}_\nu P^\nu. \tag{4.10}$$

The operators (P^0,P^1,P^2,P^3) transform as a 4-vector under the transformations of the group.

The operator

$$(P)^2 = \eta_{\mu\nu}P^\mu P^\nu \tag{4.11}$$

is an invariant operator:

$$\underline{\underline{U}}^{-1}(b,A)\eta_{\mu\nu}P^\mu P^\nu\underline{\underline{U}}(b,A) = \eta_{\mu\nu}\Lambda(A)^\mu{}_\rho\Lambda(A)^\nu{}_\kappa P^\rho P^\kappa = \eta_{\rho\kappa}P^\rho P^\kappa. \tag{4.12}$$

Here we have used (2.4). The operator $(P)^2$ commutes with $\{\underline{\underline{U}}(b,A)\}$. Using the extension of Schur's lemma to the case of unbounded linear operators we conclude that $(P)^2$ is a multiple of the unit operator in an irreducible representation:

$$(P)^2 = m^2 I. \tag{4.13}$$

As $(P)^2$ is hermitian, m^2 is real.

We will restrict ourselves to the cases $m \geq 0$. Using the properties of the operators P^μ we can now start the construction of a basis for an irreducible representation. We take as a basis the common eigenvectors $|p;\alpha\rangle$ of the commuting operators P^μ:

$$P^\mu|p;\alpha\rangle = p^\mu|p;\alpha\rangle. \tag{4.14}$$

The eigenvalues p^μ are real. The label α accounts for a possible degeneracy of energy-momentum states and will be determined later. Notice that we could have used a more complete labeling to indicate the irreducibility of the basis. Instead of $|p;\alpha\rangle$ we should have written $|m;p;\alpha\rangle$. From (4.13) we obtain

$$(4.15)\qquad (P)^2|p;\alpha\rangle = \eta_{\mu\nu}p^\mu p^\nu|p;\alpha\rangle = m^2|p;\alpha\rangle.$$

This restricts the eigenvalues p^μ to the set satisfying

$$(4.16)\qquad (p)^2 = m^2, \qquad p^0 = \pm\sqrt{\vec{p}^{\,2}+m^2}\,.$$

Again we make a restriction. We consider only the case $p^0 = \sqrt{\vec{p}^{\,2}+m^2}$. The reason is clear. Formula (4.16) is precisely the relativistic relation between energy and momentum for a system with *restmass* m. With the plus sign for p^0 the energy and momentum of a basis state $|p;\alpha\rangle$ are completely specified by giving the momentum $\vec{p}$, we shall however stick to the notation $|p;\alpha\rangle$. The continuum states $|p;\alpha\rangle$ will be normalized according to*)

$$(4.17)\qquad \langle p';\alpha'|p;\alpha\rangle = (2\pi)^3\, 2p^0\delta(\vec{p}-\vec{p}\,')\delta_{\alpha\alpha'}\quad .$$

We now investigate the transformation properties of these states. Under a translation (a,I) things are simple:

$$(4.18)\qquad \underline{\underline{U}}(a,I)|p;\alpha\rangle = e^{ia_\mu P^\mu}|p;\alpha\rangle = e^{ia_\mu p^\mu}|p;\alpha\rangle\ .$$

To obtain the effect of a Lorentz transformation we use (4.10) with b = 0:

$$(4.19)\qquad P^\mu\underline{\underline{U}}(0,A)|p;\alpha\rangle = \underline{\underline{U}}(0,A)\Lambda(A)^\mu{}_\nu P^\nu|p;\alpha\rangle = \Lambda(A)^\mu{}_\nu p^\nu\underline{\underline{U}}(0,A)|p;\alpha\rangle.$$

Equation (4.19) shows that the transformed state is again an eigenstate of P^μ, the eigenvalues being $\Lambda(A)^\mu{}_\nu p^\nu$, i.e. the Lorentz transformed of p^μ.

---

*) The continuum states $\{|p,\alpha\rangle\}$ are not in Hilbert space. The eigenvalue problem for P^μ can be solved in a Rigged Hilbert space. This concept is discussed in section 7.

Invoking the transitivity of $SL(2,\mathbb{C})$ on the orbit $p^2 = m^2 (p^0 > 0)$ this means that all the energy-momentum eigenstates $|p;\alpha\rangle$ can be reached from the set of states $|p_0;\alpha\rangle (p_0 = (m,0))$. From (4.19) we conclude that

$$\underline{\underline{U}}(0,A)|p;\alpha\rangle = C_{\beta\alpha}(\Lambda(A)p,A)|\Lambda(A)p;\beta\rangle. \tag{4.20}$$

For the matrices $C(\Lambda(A)p,A)$ we obtain from

$$\underline{\underline{U}}(0,A_2)\underline{\underline{U}}(0,A_1) = \underline{\underline{U}}(0,A_2A_1) \tag{4.21}$$

the multiplication rule

$$C(p,A_2)C(\Lambda^{-1}(A_2)p,A_1) = C(p,A_2A_1). \tag{4.22}$$

We are interested in solutions of (4.22) for which $C(p,A)$ is a *finite-dimensional* unitary matrix for $A \in SL(2,\mathbb{C})$. Notice that the dependence on the momentum p is essential. Without the p-dependence the problem would not have a solution satisfying our requirements because $SL(2,\mathbb{C})$ does not have finite-dimensional unitary representations except the trivial one (cf. Prop. I.3.1).

Due to the p-dependence the matrices $C(p,A)$ do not yield a representation of $SL(2,\mathbb{C})$. Considering however the subspace $\{|p;\alpha\rangle | p \text{ fixed}\} \equiv H(p)$ we see that the matrices $C(p,A)$ constitute a representation of the little group belonging to p. We denote this little group by $L(p)$ and the elements by A_p. For $A_p \in L(p)$ we have

$$C(p,(A_2)_p)C(p,(A_1)_p) = C(p,(A_2)_p(A_1)_p). \tag{4.23}$$

The problem is now reduced to finding the finite-dimensional irreducible unitary representations of the little group. Once we have found a representation of the little group we can construct a representation of the full group.

In section 3 we have already discussed the notion of the little group. There are three cases to be considered:

(a) $(p)^2 = m^2 \quad (m > 0, p^0 > 0);$

(b) $(p)^2 = 0 \quad (p^0 > 0);$

(c) $p = 0.$

4.2. The little group belonging to a time-like vector (case (a))

4.2.1. Representations of the little group

For time-like 4-momentum the little group is the subgroup SU(2). The unitary irreducible representations are well-known. We will denote these representations by $D^{(s)}(U)$, $U \in SU(2)$. The label s can take the values $0, \frac{1}{2}, 1, \frac{3}{2} \ldots$. The matrices $D^{(s)}(U)$ form a representation in a $(2s+1)$-dimensional vector space. Taking the standard 4-vector $\underset{0}{p} = (m, \vec{0})$ we have in $H(\underset{0}{p})$

$$(4.24) \qquad \underline{\underline{U}}(0,U)\,|\underset{0}{p};\sigma\rangle^{(s)} = D^{(s)}_{\sigma'\sigma}(U)\,|\underset{0}{p};\sigma'\rangle^{(s)} \qquad (\sigma,\sigma' = s, s-1, \ldots, -s).$$

The number s has a physical interpretation. With the standard momentum $\underset{0}{p}$ there are in a representation with label s precisely $(2s+1)$ states $|\underset{0}{p};\sigma\rangle$. Under a rotation these states transform according to the unitary irreducible representation $D^{(s)}$. The basis states $|\underset{0}{p};\sigma\rangle^{(s)}$ are labelled with the eigenvalues of the operators $(\vec{\Sigma}^{(s)})^2$ and $(\Sigma^{(s)})^3$. The states $|\underset{0}{p};\sigma\rangle$ are the (2s+1) spin states of a particle with mass $m > 0$ and spin s. The spin operators are $s = \hbar\vec{\Sigma}$. See section 6):

$$(4.25) \qquad \begin{cases} S^3\,|\underset{0}{p};\sigma\rangle^{(s)} = \hbar\sigma\,|\underset{0}{p};\sigma\rangle^{(s)} & (\sigma = s, s-1, \ldots, -s), \\ \vec{S}^2\,|\underset{0}{p};\sigma\rangle^{(s)} = \hbar^2 s(s+1)\,|\underset{0}{p};\sigma\rangle^{(s)}. \end{cases}$$

We will omit the label s on the state vectors.

4.2.2. Inducing the representation

We will start from

$$(4.24) \qquad \underline{\underline{U}}(0,U)\,|\underset{0}{p};\sigma\rangle = D^{(s)}_{\sigma'\sigma}(U)\,|\underset{0}{p};\sigma'\rangle$$

to construct the operator $\underline{\underline{U}}(0,A)$, $A \in SL(2,\mathbb{C})$. First we define a basis in $H(p)$ coupled to the basis in $H(\underset{0}{p})$:

$$(4.26) \qquad |p;\sigma\rangle = \underline{\underline{U}}(0,H(p))\,|\underset{0}{p};\sigma\rangle.$$

$H(p)$ is the matrix (3.21) which transforms $\underset{0}{p}$ into p. Secondly we notice that to any matrix $A \in SL(2,\mathbb{C})$ there is a uniquely defined matrix in SU(2), i.e. in the little group $L(p_0)$. Consider the matrix A_{p_0} defined as

(4.27) $$A_{p_0} \equiv H^{-1}(\Lambda(A)p)A\,H(p).$$

This matrix satisfies (3.26). The matrix A_{p_0} defined by (4.27) or rather the matrix $\Lambda(A_{p_0})$ in $L_+^\uparrow$ is called a *Wigner rotation*. It clearly depends on both p and A. We will denote this matrix by U(p,A). So

$$U(p,A) = H^{-1}(\Lambda(A)p)A\,H(p)$$

or

(4.28) $$A = H(\Lambda(A)p)U(p,A)H^{-1}(p).$$

Using (4.28) we can obtain the unitary operator $\underline{\underline{U}}(0,A)$:

(4.29) $$\begin{aligned}\underline{\underline{U}}(0,A)|p;\sigma\rangle &= \underline{\underline{U}}(0,A)\underline{\underline{U}}(0,H(p))|p_0;\sigma\rangle = \\ &= \underline{\underline{U}}(0,H(\Lambda(A)p))\underline{\underline{U}}(0,U(p,A))|p_0;\sigma\rangle = \\ &= D^{(s)}_{\sigma'\sigma}(U(p,A))|\Lambda(A)p;\sigma'\rangle = \\ &= D^{(s)}_{\sigma'\sigma}(H^{-1}(\Lambda(A)p)A\,H(p))|\Lambda(A)p;\sigma'\rangle\ .\end{aligned}$$

Using (4.28) one can verify that (4.29) defines indeed a representation:

(4.30) $$\underline{\underline{U}}(0,A_2)\underline{\underline{U}}(0,A_1)|p,\sigma\rangle = D^{(s)}_{\sigma'\sigma}(U(p,A_2A_1))|\Lambda(A_2A_1)p,\sigma'\rangle.$$

As we know already the action of a translation (4.18), we can now give the result of $\underline{\underline{U}}(a,A)$ on our basis:

(4.31) $$\underline{\underline{U}}(a,A)|p;\sigma\rangle = e^{ia_\mu(\Lambda(A)p)^\mu}\,D^{(s)}_{\sigma'\sigma}(U(p,A))|\Lambda(A)p;\sigma'|$$

with

$$U(p,A) = H^{-1}(\Lambda(A)p)A\,H(p).$$

4.2.3. Normalized states

Up to now we have been working with states $|p;\sigma\rangle$ with the normalization

$$\langle p;\sigma|p';\sigma'\rangle = (2\pi)^3 2p^0 \delta(\vec{p}-\vec{p}')\delta_{\sigma\sigma'} \quad . \tag{4.32}$$

The completeness relation for these states is

$$\int \sum_{\sigma=-s}^{s} |p;\sigma\rangle \frac{d^3p}{(2\pi)^3 2p^0} \langle p;\sigma| = I. \tag{4.33}$$

We now consider states $|\phi\rangle$ which are normalized to one:

$$\langle\phi|\phi\rangle = 1. \tag{4.34}$$

With respect to the basis $\{|p;\sigma\rangle\}$ a state $|\phi\rangle$ is represented by the probability amplitude $\phi_\sigma(p)$ defined as follows:

$$|\phi\rangle = \int \sum_\sigma |p;\sigma\rangle \frac{d^3p}{(2\pi)^3 2p^0} \langle p;\sigma|\phi\rangle = \int \sum_\sigma |p;\sigma\rangle \phi_\sigma(p) \frac{d^3p}{(2\pi)^3 2p^0} \quad . \tag{4.35}$$

The normalization of $|\phi\rangle$ is then given by

$$\langle\phi|\phi\rangle = \int \sum_\sigma |\phi_\sigma(p)|^2 \frac{d^3p}{(2\pi)^3 2p^0} \quad . \tag{4.36}$$

The integration is on the upper part of the hyperboloid $p^2 = m^2$, $p^0 = \sqrt{\vec{p}^2+m^2}$. The measure $\frac{d^3p}{p^0}$ is a Lorentz invariant measure.

The transformation properties of $\phi_\sigma(p)$ follow from (4.31). We consider the state

$$|\hat{\phi}\rangle = \underline{\underline{U}}(a,A)|\phi\rangle \quad . \tag{4.37}$$

One derives

$$\langle p;\sigma|\hat{\phi}\rangle \equiv \hat{\phi}_\sigma(p) = e^{ia_\mu p^\mu} D^{(s)}_{\sigma\sigma'}(H^{-1}(p)A\, H(\Lambda(A^{-1})p)\phi_{\sigma'}(\Lambda^{-1}(A)p) = \tag{4.38}$$

$$= (\underline{\underline{U}}(a,A)\phi)_\sigma(p) \quad .$$

This is our final form for the representation characterized by $[m,s]$, $m > 0$, $s = 0, \frac{1}{2}, 1, \ldots$.

4.3. The little group belonging to a light-like vector (case (b))

The method of constructing the representations is exactly the same as for case (a). There is however a difficulty which has to do with the structure of the little group. We discuss this point first. Referring to section 3 we consider the little group for the standard 4-momentum $\underset{0}{k} = (1,0,0,1)$. The matrices satisfying (3.27) have the general form

$$A_{\underset{0}{k}} = \begin{pmatrix} \alpha & 0 \\ \beta & \alpha^* \end{pmatrix}, \quad |\alpha|^2 = 1, \tag{4.39}$$

with α and β complex. Putting $\alpha = e^{-\frac{1}{2}i\phi}$ and $\beta = e^{-\frac{1}{2}i\phi}z$ with $z = \xi^1 + i\xi^2$ we have

$$A_{\underset{0}{k}} = \begin{pmatrix} e^{-\frac{1}{2}i\phi} & 0 \\ e^{-\frac{1}{2}i\phi}z & e^{\frac{1}{2}i\phi} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ z & 1 \end{pmatrix}\begin{pmatrix} e^{-\frac{1}{2}i\phi} & 0 \\ 0 & e^{\frac{1}{2}i\phi} \end{pmatrix}. \tag{4.40}$$

We denote these matrices by (z,ϕ). The multiplication rule for these matrices reads

$$(z_2,\phi_2)(z_1,\phi_1) = (z_2 + e^{i\phi}z_1, \phi_2 + \phi_1). \tag{4.41}$$

The matrices $\{(z,0)\}$ constitute an invariant abelian subgroup:

$$(w,\phi)(z,0)(w,\phi)^{-1} = (w + e^{i\phi}z,\phi)(-e^{i\phi}w,-\phi) = (e^{i\phi}z,0). \tag{4.42}$$

The little group for the vector $\underset{0}{k} = (1,0,0,1)$ is the semidirect product of the 2-parameter abelian subgroup of matrices $\{(z,0)\}$ and the one-parameter subgroup $\{(0,\phi)\}$. It is not difficult to show that this little group is isomorphic with the Euclidean group E(2) of the plane.

To construct unitary irreducible representations of E(2) is a problem similar to the construction of the representation of $P_+^\uparrow$. This leads to infinite-dimensional representations of the little group itself. In such representations there would be an infinite number of degenerate states for each 4-momentum and one would have a continuous number of spin states instead of a finite number. There are however representations which do have a discrete spin. We will discuss briefly how these representations are obtained.

Consider the Lie algebra of $L(\underset{0}{k})$. The matrices

(4.43) $$\frac{1}{2}\sigma^3 = \frac{1}{2}\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}, \quad t^1 = \begin{pmatrix}0 & 0\\ i & 0\end{pmatrix}, \quad t^2 = \begin{pmatrix}0 & 0\\ -1 & 0\end{pmatrix}$$

are the tangent vectors at the identity. The commutation relations are given by

(4.44) $$[t^1,t^2] = 0, \quad [\frac{\sigma^3}{2},t^1] = it^2, \quad [\frac{\sigma^3}{2},t^2] = -it^1,$$

and a group element in a neighbourhood of the identity is written as

(4.45) $$(z,\phi) = I - i\phi\frac{\sigma^3}{2} - i\xi^1 t^1 - i\xi^2 t^2.$$

In a unitary representation of $L(k)$ the generators (4.43) are represented by hermitian operators J^3, T^1 and T^2 respectively with commutation relations similar to (4.44). A basis for a representation can be constructed by taking the eigenvectors of the commuting operators T^1 and T^2:

(4.46) $$T^i|\ell^i;\eta> = \ell^i|\ell^i;\eta> \qquad (i = 1,2).$$

The label η accounts for a possible degeneracy of states with a given $\vec{\ell} = (\ell^1,\ell^2)$. The operator $\vec{T}^2 = (T^1)^2 + (T^2)^2$ commutes with T^1, T^2 and J^3, i.e., $\vec{T}^2$ is a Casimir operator. Invoking Schur's lemma we conclude that $\vec{\ell}^2 = r^2 > 0$ in an irreducible representation. The effect of the operations generated by J^3 is to rotate $\vec{\ell}$ along the circle $\vec{\ell}^2 = r^2$. This is analogous to the transformation of p into Λp along the hyperbola $p^2 = m^2$. One sees that the representation of $L(\underset{0}{k})$ is in general of infinite dimension. There are however finite-dimensional representations of $L(\underset{0}{k})$. They are obtained by taking $r = 0$. This means that the operators T^1 and T^2 are represented by the null-operator. In such representations we have

(4.47) $$\underline{\underline{U}}(z,\phi)\,|\underset{0}{k};\vec{0},\eta> = e^{-i\phi J^3}|\underset{0}{k};\vec{0},\eta> \quad .$$

Irreducible representations are now easily constructed. We have to take the irreducible representations of the abelian subgroup $\{(0,\phi)\}$ of $L(\underset{0}{k})$. These represenstations are one-dimensional and can be characterized by the eigenvalues of the hermitian operator J^3:

(4.48) $$J^3|\underset{0}{k};[\lambda]> = \lambda|\underset{0}{k};[\lambda]> .$$

We have omitted the label $\vec{0}$ (cf. (4.47)). A restriction on λ is obtained in the following way. In $L(\underset{0}{k})$ we have for $\phi = 2\pi$:

(4.49) $$(0,2\pi)(0,2\pi) = (0,4\pi) = (0,0).$$

So $e^{-i2\pi\lambda} \cdot e^{-i2\pi\lambda} = 1$. This gives

(4.50) $$\lambda = 0, \pm\frac{1}{2}, \pm 1, \pm\frac{3}{2}, \ldots \quad .$$

The number λ characterizing an irreducible representation of $L(\underset{0}{k})$ is called the *helicity*. The helicity is the projection of the angular momentum along the direction of motion:

(4.51) $$J^3|\underset{0}{k},[\lambda]\rangle = \frac{\vec{J}\cdot\vec{P}}{|\vec{P}|}|\underset{0}{k},[\lambda]\rangle = \lambda|\underset{0}{k},[\lambda]\rangle \; .$$

Examples of massless particles with definite helicity are the neutrino ($\lambda = \frac{1}{2}$) and the anti-neutrino ($\lambda = -\frac{1}{2}$). Photons occur in two helicity states $\lambda = 1$ and $\lambda = -1$.

From the representation of the little group $L(\underset{0}{k})$ we can construct the representation of the full group. We start from

(4.52) $$\underline{\underline{U}}(0,A_{\underset{0}{k}})|\underset{0}{k},[\lambda]\rangle = e^{-i\phi\lambda}|\underset{0}{k},[\lambda]\rangle.$$

To obtain the operator $\underline{\underline{U}}(0,A)$ acting on states with momentum k we define

(4.53) $$|k;[\lambda]\rangle \equiv \underline{\underline{U}}(0,U(\vec{k}))\underline{\underline{U}}(0,H(\hat{k}))|\underset{0}{k},[\lambda]\rangle = \underline{\underline{U}}(0,A_{k\leftarrow \underset{0}{k}})|\underset{0}{k},[\lambda]\rangle.$$

Here $H(\hat{k})$ is the positive definite hermitian matrix in $SL(2,\mathbb{C})$ which transforms $\underset{0}{k} = (1,0,0,1)$ into $\hat{k} = (k^0,0,0,k^3)$, $(k^3 = k^0)$, and $U(\vec{k})$ is a unitary matrix which rotates $\hat{k}$ into $k = (k^0,\vec{k})$:

(4.54) $$H(\hat{k})\begin{pmatrix}0 & 0\\ 0 & 2\end{pmatrix}H(\hat{k}) = \begin{pmatrix}0 & 0\\ 0 & 2k^0\end{pmatrix}, \qquad H(\hat{k}) = \begin{pmatrix}\frac{1}{\sqrt{k^0}} & 0\\ 0 & \sqrt{k^0}\end{pmatrix},$$

(4.55) $$U(\vec{k})\begin{pmatrix}0 & 0\\ 0 & 2k^0\end{pmatrix}U^\dagger(\vec{k}) = k^\mu\sigma_\mu \; .$$

There is of course some arbitrariness in the choice of $U(\vec{k})$. The usual choice is the matrix which rotates the unit vector $\vec{1}_3$ into $\vec{k}/|\vec{k}|$ with rotation axis $\vec{n} = \vec{1}_3 \wedge \vec{k}/|\vec{k}|$.

With this convention everything is fixed. The operator $\underline{\underline{U}}(0,A)$ can be obtained by realizing that $A \in SL(2,\mathbb{C})$ can be related to a little group element A_{k_0}:

$$A^{-1}_{\Lambda(A)k\leftarrow k_0} \, A \, A_{k\leftarrow k_0} = A_{k_0} \in L(k_0). \tag{4.56}$$

Using (4.56) we have

$$\begin{aligned} \underline{\underline{U}}(0,A)\,|k,[\lambda]\rangle &= \underline{\underline{U}}(0,A)\underline{\underline{U}}(0,A_{k\leftarrow k_0})\,|k_0,[\lambda]\rangle \\ &= \underline{\underline{U}}(0,A_{\Lambda(A)k\leftarrow k_0})\underline{\underline{U}}(0,A_{k_0})\,|k_0,[\lambda]\rangle \\ &= e^{-i\phi(p,A)\lambda}\,|\Lambda(A)k,[\lambda]\rangle \ . \end{aligned} \tag{4.57}$$

The angle $\phi(k,A)$ depends on k and A can be obtained from (4.56):

$$\cos[\tfrac{1}{2}\,\phi(k,A)] = \tfrac{1}{2}\,\mathrm{Tr}(A^{-1}_{\Lambda(A)k\leftarrow k_0} \, A \, A_{k\leftarrow k_0}). \tag{4.58}$$

The transformation properties of normalized states in a representation $[0,\lambda]$ are given by

$$\hat{\phi}(k,[\lambda]) = e^{i\phi(k,A^{-1})}\,\phi(\Lambda^{-1}(A)k,\lambda). \tag{4.59}$$

This ends our discussion of case (b). The final case (c), i.e. the representations for $p \equiv 0$, are easily dealt with. The little group is $SL(2,\mathbb{C})$ itself. Being a non-compact Lie group $SL(2,\mathbb{C})$ has no finite-dimensional unitary representations except the trivial representation. We have therefore

$$\underline{\underline{U}}(a,A)\,|0\rangle = |0\rangle. \tag{4.60}$$

The state $|0\rangle$ is the invariant vacuum state. It carries neither energy nor momentum.

5. SUMMARY

We have constructed those unitary irreducible representations of the inhomogeneous SL(2,ℂ) group which can be labelled with two quantum numbers. That the representations are irreducible is made plausible by the construction. We have built the representations from an irreducible representation of the little groups. The representations discussed in the foregoing do have a physical interpretation.

A representation [m,s] gives the kinematical properties of a free particle with mass m and spin s. For a particle having momentum $\vec{p}$ there are (2s + 1) mutually orthogonal states $|p;\sigma\rangle$. Under a rotation these states transform as

$$\underline{\underline{U}}(0,R)\,|p,\sigma\rangle = D^{(s)}_{\sigma'\sigma}(R)\;|R\,p,\sigma'\rangle.$$

The transformation properties under translations and Lorentz transformations is given by (4.31).

A representations $[0,\lambda]$ gives the kinematical states of a massless system. For a given momentum there is only one state $|p,\lambda\rangle$.

These representations can be considered as the starting point for the construction of *covariant space-time operator fields*. To obtain these operator fields one has to through a number of steps. We discuss briefly one of these steps.

Instead of the canonical basis $\{|p,\sigma\rangle\}$ one can introduce the so-called *spinor basis*

(5.1) $$|p,\sigma\rangle^{(s,0)} = D^{(s,0)}_{\sigma'\sigma}\,(H^{-1}(p))\,|p,\sigma'\rangle.$$

The matrices $D^{(s,0)}(A)$ constitute a (2s + 1)-dimensional representation of SL(2,ℂ). These matrices are not unitary. The representation $D^{(s,0)}(A)$ has the property that it coincides with $D^{(s)}(U)$ for the matrices of the unitary subgroup:

(5.2) $$D^{(s,0)}(U) = D^{(s)}(U), \quad U \in SU(2).$$

Using this relationship the transformation properties of the spinor basis are easily found from (4.31):

(5.3) $$\underline{\underline{U}}(0,A)\,|p,\sigma\rangle^{(s,0)} = D^{(s,0)}_{\sigma'\sigma}(H^{-1}(p))\underline{\underline{U}}(0,A)\,|p,\sigma'\rangle$$

$$= D^{(s,0)}_{\sigma'\sigma}(H^{-1}(p))\,D^{(s)}_{\sigma''\sigma'}(H^{-1}(\Lambda(A)p)A\,H(p))\,|\Lambda(A)p,\sigma''\rangle$$

$$= D^{(s,0)}_{\sigma''\sigma}(H^{-1}(\Lambda(A)p)A\,H(p)H^{-1}(p))\,|\Lambda(A)p,\sigma''\rangle$$

$$= D^{(s,0)}_{\rho\sigma}(A)\,|\Lambda(A)p,\rho\rangle^{(s,0)}.$$

This shows that the spin-label transforms according to a finite-dimensional representation of SL(2,$\mathbb{C}$). The complicated Wigner rotation has disappeared from the transformation law. The wave function $\phi^{(s,0)}_\sigma(p) \equiv {}^{(s,0)}\langle p;\sigma|\phi\rangle$ transforms according to

(5.4) $$\hat{\phi}^{(s,0)}_\sigma(p) = D^{(s,0)}_{\sigma\rho}(A)\phi^{(s,0)}_\rho(\Lambda(A^{-1})p).$$

Instead of the representation $D^{(s,0)}$ one can take $D^{(0,s)}$ which is not equivalent to $D^{(s,0)}$ but which shares the property of being equal to $D^{(s)}$ for the unitary subgroup:

(5.5) $$D^{(0,s)}(U) = D^{(s)}(U).$$

Defining

(5.6) $$|p,\sigma\rangle^{(0,s)} \equiv D^{(0,s)}_{\sigma'\sigma}(H^{-1}(p))\,|p,\sigma'\rangle$$

one obtains for the wave functions $\phi^{(0,s)}_\sigma(p) \equiv {}^{(0,s)}\langle p,\sigma|\phi\rangle$ the transformation behaviour

(5.7) $$\hat{\phi}^{(0,s)}_\sigma(p) = D^{(0,s)}_{\sigma\rho}(A)\phi^{(0,s)}_\rho(\Lambda(A^{-1})p).$$

It is interesting to notice that the spinor bases $\{|p,\sigma\rangle^{(s,0)}\}$ and $\{|p,\sigma\rangle^{(0,s)}\}$ being defined from the canonical basis are not independent. One finds from (5.1) and (5.6)

(5.8) $$|p,\sigma\rangle^{(s,0)} = (D^{(0,s)}(H(p)))^2_{\lambda\sigma}\,|p,\lambda\rangle^{(0,s)}.$$

For the wave functions one obtains

$$(5.9) \qquad \phi_\sigma^{(s,0)}(p) = (D^{(0,s)}(H(p)))^2_{\sigma\rho}\phi_\rho^{(0,s)}(p) \quad ;$$

$$\phi_\sigma^{(0,s)}(p) = (D^{(s,0)}(H(p)))^2_{\sigma\rho}\phi_\rho^{(s,0)}(p) \quad .$$

Defining now the so-called *doubled wave function*

$$(5.10) \qquad \phi(p) = \begin{pmatrix} \phi^{(s,0)}(p) \\ \phi^{(0,s)}(p) \end{pmatrix}$$

one has the transformation law

$$(5.11) \qquad \hat{\phi}(p) = \begin{pmatrix} D^{(s,0)}(A) & 0 \\ 0 & D^{(0,s)}(A) \end{pmatrix} \phi(\Lambda(A^{-1})p),$$

this clearly is a reducible representation. The relationship (5.9) can now be put in the form

$$(5.12) \qquad \Omega(p)\phi(p) = 0$$

with

$$(5.13) \qquad \Omega(p) = \begin{pmatrix} I & -(D^{(0,s)}(H(p)))^2 \\ -(D^{(s,0)}(H(p)))^2 & I \end{pmatrix} \quad .$$

Defining the Fourier transform of $\phi(p)$ by

$$(5.14) \qquad \phi(x) = (2\pi)^{-3/2}\int \frac{d^3p}{2p^0}\,\phi(p)e^{-ipx}$$

one arrives at the relativistic wave equation for a particle with spin s:

$$(5.15) \qquad \Omega(-i\partial^\mu)\phi(x) = 0.$$

We have discussed this matter to show the importance of the finite-dimensional representations of the Lorentz group $L_+^\uparrow$. To obtain wave equations one has to work with reducible representations. The wave equation restricts the redundancy in the number of components of the wave function in such a way that there are essentially $(2s+1)$ independent components for a particle with spin s.

6. REPRESENTATIONS OF THE GROUP SU(2)

For the construction of the unitary representations of the Poincaré group for massive systems (see section 4.2) one needs the representations of the group SU(2). In this section we review briefly some properties of the groups SU(2) and SO(3) and the representations of SU(2).

6.1. The group SU(2)

The group SU(2) is the group of unitary 2x2 matrices U with det U = 1. A convenient and physically useful parametrization of the matrices U is obtained in the following way. Let U be given by

$$U = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}. \tag{6.1}$$

From $U^{\dagger} = U^{-1}$ (unitarity) and det U = 1 one obtains

$$U = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix} \quad \text{with } |\alpha|^2 + |\beta|^2 = 1. \tag{6.2}$$

We now put $\alpha = a_0 - ia_3$, $\beta = -ia_1 - a_2$ with a_0, a_1, a_2, a_3 real numbers. Expressed in these parameters we have

$$U = \begin{pmatrix} \alpha & \beta \\ -\beta^* & \alpha^* \end{pmatrix} = a_0\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - ia_1\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} - ia_2\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} - ia_3\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \tag{6.3}$$

$$= a_0 I - i\vec{a}.\vec{\sigma}.$$

The matrices $\sigma_1, \sigma_2, \sigma_3$ introduced in (6.3) are the *Pauli matrices*

$$\sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_2 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{6.4}$$

while

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

From det U = 1 we obtain

(6.5) $$a_0^2 + a_1^2 + a_2^2 + a_3^2 = a_0^2 + \vec{a}^2 = 1.$$

This shows that SU(2) is in 1-1 correspondence with the unit sphere in $\mathbb{R}^4$. The restriction on the parameters can be made explicit by putting

(6.6) $$\begin{cases} a_0 = \cos \phi/2, \\ |\vec{a}| = \sin \phi/2, \qquad 0 \leq \phi < 2\pi, \end{cases}$$

and the matrices of SU(2) take the form

(6.7) $$U(\vec{n},\phi) = I \cos \phi/2 - i\, \vec{n}.\vec{\sigma} \sin \phi/2 = e^{-i\phi\vec{n}.\vec{\sigma}/2} \quad \text{with } \vec{n} = \vec{a}/|\vec{a}|.$$

The matrices are determined by three independent real parameters, an angle ϕ and a unit vector $\vec{n}$. For later use we list some properties of the Pauli matrices:

(6.8)

(a) $[\sigma_j,\sigma_k] \equiv \sigma_j\sigma_k - \sigma_k\sigma_j = 2i\varepsilon_{jk\ell}\sigma_\ell \qquad (j,k,\ell = 1,2,3),$

(b) $\{\sigma_j,\sigma_k\} \equiv \sigma_j\sigma_k + \sigma_k\sigma_j = 2\delta_{jk}.I,$

(c) $\sigma_j\sigma_k = \delta_{jk} + i\varepsilon_{jk\ell}\sigma_\ell,$

(d) $\mathrm{Tr}\ \sigma_j = 0; \quad \sigma_j = \sigma_j^\dagger = \sigma_j^{-1}, \quad \det \sigma_j = -1,$

(e) $\mathrm{Tr}\ \sigma_j\sigma_k = 2\delta_{jk},$

(f) $(\vec{a}.\vec{\sigma})(\vec{b}.\vec{\sigma}) = (\vec{a}.\vec{b}) + i(\vec{a}\wedge\vec{b}).\vec{\sigma}.$

The symbol $\varepsilon_{jk\ell}$ is the 3-dimensional Levi-Civitā symbol.

6.2. Infinitesimal transformations

For infinitesimal ϕ the matrices $U(\vec{n},\phi)$ take the form

(6.9) $$U_{inf}(\vec{n},\phi) = I - i\, \vec{n}.\frac{\vec{\sigma}}{2}\phi + O(\phi^2).$$

The mapping $\phi \to U(\vec{n},\phi)$ is a path through I with tangent vector $-i\vec{n}.\frac{\vec{\sigma}}{2}$. The matrices $\sigma_1/2$, $\sigma_2/2$ and $\sigma_3/2$ are called the generators of the group. They satisfy the commutation relations

$$(6.10) \qquad [\frac{\sigma_j}{2}, \frac{\sigma_k}{2}] = i\varepsilon_{jk\ell}\frac{\sigma_\ell}{2}.$$

Apart from a factor $\hbar$ (Planck's constant h divided by 2π) we have here the commutation relations of the operators of angular momentum for a quantum mechanical system.

REMARK. In the literature one often uses another parametrization for SU(2): $U(\phi,\theta,\psi) = e^{-i\phi\sigma_3/2}\, e^{-i\theta\sigma_2/2}\, e^{-i\psi\sigma_3/2}$, where ϕ,θ,ψ are the Euler angles.

6.3. The group SO(3)

The group SO(3) is the rotation group in $\mathbb{R}^3$. Elements of SO(3) are real 3x3 orthogonal matrices R with det R = 1. A convenient parametrization for the matrices R is obtained from the following geometrical considerations. Let $R(\vec{n},\phi)$ denote a right-handed rotation around the unit vector $\vec{n}$ with rotation angle ϕ. The action of $R(\vec{n},\phi)$ is a mapping

$$\vec{x} \to \hat{\vec{x}} = R(\vec{n},\phi)\vec{x}.$$

We now put $\vec{x} = \vec{x}_{11} + \vec{x}_\perp$ with $\vec{x}_{11} = (\vec{n}.\vec{x})\vec{n}$ and $\vec{n}.\vec{x}_\perp = 0$. Decomposing in the same way $\hat{\vec{x}}$, i.e.

$$\hat{\vec{x}} = \hat{\vec{x}}_{11} + \hat{\vec{x}}_\perp,$$

we have $\hat{\vec{x}}_{11} = \vec{x}_{11}$. The vector $\hat{\vec{x}}_\perp$ is obtained from $\vec{x}_\perp$ by the rotation over ϕ in the plane through $\vec{x}_\perp$ and orthogonal to $\vec{n}$ (cf. Figure 1):

$$\hat{\vec{x}}_\perp = \frac{\vec{x}_\perp}{|\vec{x}_\perp|}|\vec{x}_\perp| \cos\phi + \frac{\vec{n}\wedge\vec{x}_\perp}{|\vec{n}\wedge\vec{x}_\perp|}|\vec{x}_\perp| \sin\phi =$$

$$= (\vec{x}-\vec{x}_{11}) \cos\phi + (\vec{n}\wedge\vec{x}) \sin\phi.$$

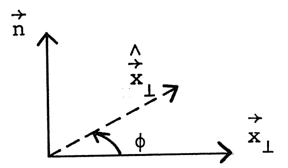

Figure 1.

The result is

$$(6.11)\qquad \hat{\vec{x}} = \vec{x}_{11} + \hat{\vec{x}}_{\perp} = \vec{n}(\vec{x}.\vec{n})+(\vec{x}-\vec{n}(\vec{x}.\vec{n}))\cos\phi + (\vec{n}\wedge\vec{x})\sin\phi =$$

$$= \vec{x} + (\vec{n}\wedge\vec{x})\sin\phi + \vec{n}\wedge(\vec{n}\wedge\vec{x})(1-\cos\phi) = R(\vec{n},\phi)\vec{x}.$$

Writing out (6.11) one finds that the matrices $R(\vec{n},\phi)$ can be written in the form

$$(6.12)\qquad R(\vec{n},\phi) = I + (\vec{n}.\vec{I})\sin\phi + (\vec{n}.\vec{I})^2(1-\cos\phi),$$

where we have introduced the matrices $\vec{I} = (I_1,I_2,I_3)$ given by

$$(6.13)\qquad I_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad I_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad I_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

The matrices are the tangent vectors at the unit element of the mapping $\phi \to R(\vec{n},\phi)$ with $\vec{n}$ = (1,0,0), (0,1,0) and (0,0,1), respectively. The matrices I_j satisfy the relations

$$[I_j,I_k] = \varepsilon_{jk\ell}I_\ell.$$

Defining the hermitian matrices $\vec{J} = i\,\vec{I}$ we have

$$(6.14)\qquad [J_j,J_k] = i\varepsilon_{jk\ell}J_\ell.$$

Comparing (6.10) and (6.14) one sees that the groups SU(2) and SO(3) have the same Lie algebra. It is not hard to show that the matrices $R(\vec{n},\phi)$ can be written in exponential form

$$(6.15)\qquad R(\vec{n},\phi) = e^{\phi\vec{n}.\vec{I}} = e^{-i\phi\vec{n}.\vec{J}}.$$

6.4. <u>The homomorphism from SU(2) onto SO(3)</u>

The groups SU(2) and SO(3) are related by a two-to-one homomorphism from SU(2) onto SO(3). This homomorphism can be made explicit by the following construction. Consider the matrices

$$H(\vec{x}) \equiv \begin{pmatrix} x_3 & x_1-ix_2 \\ x_1+ix_2 & -x_3 \end{pmatrix} = \vec{x}.\vec{\sigma}, \quad \vec{x} \in \mathbb{R}^3. \tag{6.16}$$

One readily obtains the following properties for $H(\vec{x})$:

$$\mathrm{Tr}\, H = 0, \quad H = H^\dagger, \quad \det H = -\vec{x}^2, \quad x_i = \tfrac{1}{2}\mathrm{Tr}(\sigma_i H(\vec{x})).$$

This is a one-one correspondence between traceless hermitian matrices H and points $\vec{x} \in \mathbb{R}^3$.

Consider next the unitary transformation

$$U(\vec{n},\phi)H(\vec{x})U^{-1}(\vec{n},\phi) \equiv H'(\vec{x}), \quad U(\vec{n},\phi) \in SU(2). \tag{6.17}$$

The matrix H' satisfies

$$\mathrm{Tr}\, H' = 0 \quad \text{and} \quad H' = H'^\dagger.$$

This means that there is a uniquely determined vector $\hat{\vec{x}} \in \mathbb{R}^3$ which is obtained from

$$H'(\vec{x}) = \hat{\vec{x}}.\vec{\sigma}. \tag{6.18}$$

From (6.17) and (6.18) we obtain moreover

$$\det H'(\vec{x}) = -\hat{\vec{x}}^2 = \det H(\vec{x}) = -\vec{x}^2. \tag{6.19}$$

It is clear that the mapping $\vec{x} \to \hat{\vec{x}}$ associated with (6.17) is linear. We therefore conclude from (6.17) and (6.19) that $\vec{x}$ and $\hat{\vec{x}}$ are related by an orthogonal transformation

$$\vec{x} \to \hat{\vec{x}} = R(U)\vec{x}. \tag{6.20}$$

The mapping $U \to R(U)$ is a homomorphism the kernel of which consists of the matrices I and -I. The mapping $U \to R(U)$ is therefore two-to-one; the elements U and -U are mapped into $R(U) = R(-U)$.

The matrices R(U) can be obtained from (6.17) by using the properties of the Pauli matrices. Written out (6.17) reads

$$Ux_i\sigma_i U^{-1} = x_i U\sigma_i U^{-1} = \hat{x}_i\sigma_i.$$

Multiplying by σ_j and taking the trace gives

(6.21) $$\hat{x}_j = \tfrac{1}{2}x_i \mathrm{Tr}(\sigma_i U\sigma_j U^{-1}) \equiv R_{ji}(U)x_i.$$

Substituting now for U the matrix $U(\vec{n},\phi)$ given by (6.7) one finds

(6.22) $$R(U(\vec{n},\phi)) = I + (\vec{n}.\vec{I})\sin\phi + (\vec{n}.\vec{I})^2(1-\cos\phi).$$

This is precisely the form of the rotation matrices given in (6.12). From the above considerations we conclude that the mapping $U \to R(U)$ is a two-to one homomorphism from SU(2) onto SO(3).

6.5. Representations of SO(3)

Starting from SO(3) one can consider the correspondence

$$R(\vec{n},\phi) \to \pm\, U(\vec{n},\phi).$$

In physical texts SU(2) is often called a double-valued representation of SO(3). Studying the properties of physical systems under rotations, one usually starts from the rotation group. To describe massive particles with half integer spin $s = \frac{1}{2}$ (e.g. electrons) one runs into the double-valued representation $R(U) \to \pm\, U$.

To describe higher spins one needs the representations of SU(2). The unitary irreducible representations of SU(2) are denoted by $D^{(s)}(U)$. They are completely characterized by the number s which can take values from $\{0,\frac{1}{2},1,\frac{3}{2},2,\ldots\}$. In the next subsection we construct these representations.

6.6. Representations of SU(2)

Let $U(\vec{n},\phi) \to D(U)$ be a unitary representation of SU(2). The generators Σ_1,Σ_2 and Σ_3 in the representation D(U) are hermitian matrices which satisfy the commutation relations of the group SU(2), i.e.

(6.23) $$[\Sigma_j,\Sigma_k] = i\varepsilon_{jk\ell}\Sigma_\ell.$$

The matrices $\vec{\Sigma} = (\Sigma_1,\Sigma_2,\Sigma_3)$ transform according to the adjoint representation

(6.24) $$D^{-1}(U)\ \Sigma_j\ D(U) = R_{jk}(U)\ \Sigma_k.$$

We construct the representation of SU(2) by solving (6.23) for all irreducible sets $\{\Sigma_1,\Sigma_2,\Sigma_3\}$ and using the exponential mapping

(6.25) $$D(U(n,\phi)) = e^{-i\phi(\vec{n}.\vec{\Sigma})}.$$

To obtain the solutions of (6.23) we use the so-called construction operator formalism.

First, we notice that the operator $\vec{\Sigma}^2 = \Sigma_1^2 + \Sigma_2^2 + \Sigma_3^2$ satisfies

(6.26) $$[\vec{\Sigma}^2,\Sigma_j] = 0, \qquad j = 1,2,3.$$

Invoking Schur's lemma we have in an irreducible representation:

(6.27) $$\vec{\Sigma}^2 = \alpha \qquad (\alpha \geq 0).$$

The number α is real because $\vec{\Sigma}^2$ is hermitian. Moreover, α is non-negative because $\vec{\Sigma}^2$ is a positive operator.

We now consider the commuting operators $\vec{\Sigma}^2$ and Σ_3. Denoting the common eigenvectors of these operators by $\{|\alpha,m\rangle\}$ we have

(6.28) $$\vec{\Sigma}^2|\alpha,m\rangle = \alpha|\alpha,m\rangle \qquad (\alpha \geq 0),$$

$$\Sigma_3|\alpha,m\rangle = m|\alpha,m\rangle \qquad (m \text{ real}).$$

We determine the eigenvalues α and m for irreducible representations. To this end we define operators Σ_+ and Σ_-:

(6.29) $$\Sigma_+ = \Sigma_1+i\Sigma_2, \qquad \Sigma_- = \Sigma_1 - i\Sigma_2.$$

Expressed in those operators the commutation relations (6.23) simplify:

(6.30) $$[\Sigma_+,\Sigma_-] = 2\Sigma_3, \quad [\Sigma_3,\Sigma_\pm] = \pm\Sigma_\pm, \quad [\vec{\Sigma}^2,\Sigma_\pm] = 0.$$

The operators Σ_+,Σ_- and Σ_3 constitute the so-called *Cartan basis*.

Starting from an eigenvector $|\alpha,m\rangle$ we obtain from (6.30) that $\Sigma_\pm|\alpha,m\rangle$ are also eigenvectors of $\vec{\Sigma}^2$ and Σ_3:

(6.31)
$$\begin{cases} \vec{\Sigma}^2\Sigma_\pm|\alpha,m\rangle = \Sigma_\pm\vec{\Sigma}^2|\alpha,m\rangle = \alpha\ \Sigma_\pm|\alpha,m\rangle, \\ \Sigma_3\Sigma_\pm|\alpha,m\rangle = (\Sigma_\pm\Sigma_3 \pm \Sigma_\pm|\alpha,m\rangle = (m\pm 1)\Sigma_\pm|\alpha,m\rangle. \end{cases}$$

Repeated application of Σ_+ to $|\alpha,m\rangle$ leads to a chain of eigenvectors of Σ_3 with eigenvalues m, m+1, m+2, m+3,... . Repeated application of Σ_- leads to a chain of eigenvectors with eigenvalues m, m−1, m−2, m−3,... . We are going to show that this chain must break and that it contains for arbitrary but fixed α only a finite number of elements.

Consider the norm of the eigenvector $\Sigma_+|\alpha,m\rangle$:

(6.32)
$$\langle\alpha,m|\Sigma_+^\dagger\Sigma_+|\alpha,m\rangle = \langle\alpha,m|\Sigma_-\Sigma_+|\alpha,m\rangle = \langle\alpha,m|\Sigma_1^2+\Sigma_2^2-\Sigma_3|\alpha,m\rangle =$$
$$= \langle\alpha,m|\vec{\Sigma}^2-\Sigma_3^2-\Sigma_3|\alpha,m\rangle = (\alpha-m^2-m)\ \langle\alpha,m|\alpha,m\rangle.$$

The left-hand side of this equation, being a norm, is non-negative. It may be zero if $\Sigma_+|\alpha,m\rangle = 0$. We obtain

(6.33)
$$\alpha - m^2 - m \geq 0.$$

Along the same lines one obtains from the norm of $\Sigma_-|\alpha,m\rangle$:

(6.34)
$$\langle\alpha,m|\Sigma_-^\dagger\Sigma_-|\alpha,m\rangle = \langle\alpha,m|\Sigma_+\Sigma_-|\alpha,m\rangle = (\alpha-m^2+m)\ \langle\alpha,m|\alpha,m\rangle,$$

so

(6.35)
$$\alpha - m^2 + m \geq 0.$$

From (6.33) and (6.35) one obtains for the eigenvalues m of Σ_3 the restriction

(6.36)
$$-\sqrt{\alpha} \leq m \leq \sqrt{\alpha}.$$

This shows that the chain of eigenvalues given above must break.

Before finishing this analysis we want to make a remark on the degeneracy of the eigenvalues α and m. From (6.32) and (6.34) one sees that both $\Sigma_+\Sigma_-|\alpha,m\rangle$ and $\Sigma_-\Sigma_+|\alpha,m\rangle$ are proportional to $|\alpha,m\rangle$. This means that if there is a degeneracy we will build up chains which do not mix. It is, therefore, justified to assume that the eigenvalues are non-degenerate. This means that

$$\Sigma_+|\alpha,m\rangle = \beta_+|\alpha,m+1\rangle,\ \Sigma_-|\alpha,m\rangle = \beta_-|\alpha,m-1\rangle, \tag{6.37}$$

where β_+ and β_- can be fixed by normalization.

Let us now finish the discussion on the chains. From (6.36) it follows that there is a maximum value of m which we call s. For this value we have $|\alpha,s\rangle \neq 0$ and $\Sigma_+|\alpha,s\rangle = 0$. From (6.32) we obtain for this eigenvalue:

$$\alpha - s^2 - s = 0. \tag{6.38}$$

As the values of m are also bounded below there must be a minimum value of m. Calling this value s' we have $|\alpha,s'\rangle \neq 0$, $\Sigma_-|\alpha,s'\rangle = 0$. From (6.33) we obtain

$$\alpha - s'^2 + s' = 0. \tag{6.39}$$

From (6.38) and (6.39) follows $s' = s + 1$ or $s' = -s$. The minimum value of m is therefore $s' = -s$. Starting from the highest eigenvalue s of Σ_3, we must arrive, applying Σ_-, at the lowest eigenvalue $-s$ in unit steps. This means that 2s is a non-negative integer and the number s can take values from the set $\{0,\frac{1}{2},1,\frac{3}{2},\ldots\}$.

For an arbitrary but fixed s from this set, we have a (2s+1)-dimensional space of vectors which are eigenvectors of both $\vec{\Sigma}^2$ and Σ_3. For these vectors α takes the value $\alpha = s(s+1)$. We denote these vectors by $\{|s,m\rangle\}$ and we have

$$\vec{\Sigma}^2|s,m\rangle = s(s+1)|s,m\rangle, \tag{6.40}$$

$$\Sigma_3|s,m\rangle = m|s,m\rangle \qquad (m = -s,-s+1,\ldots,+s).$$

6.7. Normalization and matrix elements

We adopt the convention that the vectors $\{|s,m\rangle\}$ are normalized to one. From (6.32), (6.34) and (6.37) we obtain for β_+ and β_-

$$|\beta_+|^2 = s(s+1)-m(m+1), \quad |\beta_-|^2 = s(s+1)-m(m-1).$$

The standard convention is to take β_+ and β_- real, this gives

(6.41) $$\Sigma_+|s,m\rangle = \sqrt{s(s+1)-m(m+1)}\;|s,m+1\rangle,$$

$$\Sigma_-|s,m\rangle = \sqrt{s(s+1)-m(m-1)}\;|s,m-1\rangle.$$

Using this, we arrive at the matrix elements of $\Sigma_+, \Sigma_-, \Sigma_3$ and $\vec{\Sigma}^2$ in a representation characterized by s:

(6.42) $$\begin{cases} \langle s,m'|\vec{\Sigma}^2|s,m\rangle = s(s+1)\delta_{m,m'}, \\ \langle s,m'|\Sigma_3|s,m\rangle = m\delta_{m,m'}, \\ \langle s,m'|\Sigma_+|s,m\rangle = \sqrt{s(s+1)-m(m+1)}\;\delta_{m',m+1}, \\ \langle s,m'|\Sigma_-|s,m\rangle = \sqrt{s(s+1)-m(m-1)}\;\delta_{m',\,m-1}. \end{cases}$$

The matrix elements of Σ_1 and Σ_2 may be obtained from those of Σ_- and Σ_+ using the definition (6.29). The representations of the Lie algebra obtained in the above manner are irreducible. This is clear from the construction of the (2s+1)-dimensional representation spaces.

From the representations of the algebra we arrive at the representations of the group by exponentiation. The matrices $D^{(s)}(U)$ defined by

(6.43) $$D^{(s)}(U) = e^{-i\phi\vec{n}.\vec{\Sigma}^{(s)}},$$

where $\vec{\Sigma}^{(s)}$ are the matrices $\vec{\Sigma}$ in a representation labelled by s, constitute a representation of the group SU(2). These representations are used in section 4.2.

7. RIGGED HILBERT SPACE

7.1. The problem of a mathematical foundation for generalized eigenvectors

In the quantum mechanical description of the physical properties of a system one uses the language of Hilbert space. The states of the system are represented by unit vectors (or rather unit rays) and the observables by hermitian operators (see Ch.II). The connection between the mathematical formalism and reality is comprimed in a number of postulates which contain the physical interpretation of mathematical quantities. We consider here the so-called Expansion Postulate which relates the results of measurements of an observable A and the eigenvalues and eigenvectors of the hermitian operator

$\underline{\underline{A}}$ corresponding to A. To explain the meaning of this postulate we assume for simplicity that $\underline{\underline{A}}$ is defined on all of the Hilbert space H and that $\underline{\underline{A}}$ has a non-degenerate discrete spectrum of eigenvalues. The eigenvalue equation reads

$$\underline{\underline{A}}|a_n\rangle = a_n|a_n\rangle \quad (n=0,1,2,\ldots),$$

$$\langle a_n|a_m\rangle = \delta_{nm}.$$

The eigenvectors constitute a complete orthonormal system in H and any state $|u\rangle$ can be expanded as

$$|u\rangle = \sum_0^\infty c_n|a_n\rangle \quad (\langle u|u\rangle=1)$$

with $c_n = \langle a_n|u\rangle$ and $\Sigma_0^\infty |c_n|^2 = 1$. The completeness of the system of eigenvectors is expressed as

$$\sum_0^\infty |a_n\rangle\langle a_n| = I$$

and the operator $\underline{\underline{A}}$ is given by

$$\underline{\underline{A}} = \sum_0^\infty a_n|a_n\rangle\langle a_n|.$$

The interpretation going with this is the following: Let the system be in the state represented by $|u\rangle$, a measurement of the observable A will then yield an eigenvalue of $\underline{\underline{A}}$. The quantity $|c_n|^2 = |\langle a_n|u\rangle|^2$ is the probability that the result is the eigenvalue a_n. Due to the measurement the state of the system is changed. If the outcome is a_n the system will be in the state $|a_n\rangle$ after the measurement. (Reduction of the state vector.)

All this is very nice as long as the operators have a discrete spectrum. There are however many examples for which the spectrum is partly discrete and partly continuous or even completely continuous. Examples of the latter case are the momentum operator $\underline{\underline{p}}$ and the position operator $\underline{\underline{x}}$. Such operators have no eigenvectors in Hilbert space. If one wants to maintain the above formalism one must extend the Hilbert space to a space in which the eigenvalue problem can be properly formulated for the continuous spectrum. In introductory courses on quantum mechanics one does not worry very much about these problems. One simply uses the formalism developed by DIRAC [5] to cope with the continuous spectrum. We illustrate this with the well-known example of the momentum operator.

Consider a system for which the space of states is $H = L^2(\mathbb{R}^3)$, i.e., the states are described by "functions" $u(\vec{x})$ which satisfy

$$<u|u> = \int_{\infty}^{\infty} u^*(\vec{x})u(\vec{x})d^3x < \infty.$$

The generators of translations of the system are the momentum operators $\underline{\vec{p}} = \frac{\hbar}{i}\nabla$. The eigenvalue equation

$$\underline{\vec{p}}u_{\vec{p}}(\vec{x}) = \vec{p}u_{\vec{p}}(\vec{x}) \qquad (\vec{p}=(p_1,p_2,p_3) \text{ real})$$

is solved by

$$u_{\vec{p}}(\vec{x}) = (2\pi\hbar)^{-3/2}e^{i\vec{p}.\vec{x}/\hbar}.$$

For these eigenfunctions we have

$$<u_{\vec{p}}|u_{\vec{p}}> = \int_{-\infty}^{\infty} (2\pi\hbar)^{-3}d^3x = \infty$$

and $u_{\vec{p}}(x)$ is not in H.

In a more abstract notation we write

$$\underline{\vec{p}}|\vec{p}> = \vec{p}|\vec{p}>$$

and the "eigenvectors" are normalized according to

$$<\vec{p}|\vec{p}'> = \delta^3(\vec{p}-\vec{p}').$$

The completeness of the system of eigenvectors is expressed by

$$I = \int |\vec{p}>d^3p<\vec{p}|$$

and the expansion of $|u>$ is

$$|u> = \int |\vec{p}><\vec{p}|u>d^3p.$$

The probability that a measurement of the momentum will yield a result $\vec{p}$ in $\Delta p_1.\Delta p_2.\Delta p_3$ around $\vec{p}$ is given by $|<\vec{p}|u>|^2\Delta p_1.\Delta p_2.\Delta p_3$. In the language of $L^2(\mathbb{R}^3)$ the expansion of $|u>$ is simply the Fourier expansion

$$u(x) = (2\pi\hbar)^{-3/2} \int_{-\infty}^{\infty} d^3p \; \tilde{u}(\vec{p}) e^{i\vec{p}.\vec{x}/\hbar}.$$

Notwithstanding the beauty of this formalism one might feel a bit uneasy about the mathematical aspects, e.g., one could ask in what kind of space the eigenvalue problem should be formulated to include the continuous spectrum. The answer to this question can be found in the literature, see GELFAND [6], BÖHM [3], ANTOINE [1], BOGOLIUBOV [2]. The space best suited is in fact a triple of spaces called *Gelfand Triple* or *Rigged Hilbert Space*:

$$\Omega \subset H \subset \Omega'.$$

Here Ω is a countably normed nuclear Hilbert space, H is a Hilbert space in which Ω is dense, and Ω' is the dual space of Ω.

In such a space the eigenvalue problem can be properly formulated using the concept of generalized eigenvectors. Generalized eigenvectors are elements of Ω', i.e. continuous linear functionals. We give the definition. Let $\underline{\underline{A}}$ be a hermitian operator on Ω. Then a generalized eigenvector F_a belonging to the eigenvalue a is an element of Ω' which satisfies

$$F_a(\underline{\underline{A}}u) = aF_a(u), \quad \forall u \in \Omega.$$

Using this concept one can prove that a hermitian operator $\underline{\underline{A}}$ has a complete set of generalized eigenvectors belonging to real eigenvalues. In the example of the momentum operator the generalized eigenvectors are the bras $\langle\vec{p}|$. In the following subsection we are going to discuss the various mathematical concepts involved in the definitions of Rigged Hilbert Space and we will give examples of such a space.

7.2. The countably normed Hilbert space Ω

We consider a linear vector space Ω over the complex numbers. Let there be given on Ω a countable set of inner products $\{(u,v)_n\}_{n=1}^{\infty}$ with the property $(u,u)_n \geq 0$ and $(u,u)_n = 0 \Leftrightarrow u = 0$. From these inner products one obtains a countable set of norms by defining

(7.1) $$\|u\|_n = \sqrt{(u,u)_n}.$$

We assume these norms to be ordered according to

(7.2) $$\|u\|_1 \leq \|u\|_2 \leq \|u\|_3 \;\ldots\; .$$

This can always be achieved by a redefinition of the inner products. If (7.2) does not hold one defines $((u,v))_n = \Sigma_{k=1}^{n} (u,v)_k$ and the norms obtained from these new inner products satisfy (7.2).

Using the norms (7.1) one can introduce on Ω a topology by defining a set of neighbourhoods $U_{n,\varepsilon}(0)$ of $u = 0$:

(7.3) $$U_{n,\varepsilon}(0) = \{u \in \Omega \mid \|u\|_n < \varepsilon\}.$$

A sequence (u_k) is said to *converge* to an element u (notation: $u_k \to u$) if $\forall U_{n,\varepsilon}(0)\ \exists N(n,\varepsilon)$ such that:

$$k > N(n,\varepsilon) \Rightarrow (u_k - u) \in U_{n,\varepsilon}(0), \text{ i.e. } \|u_k - u\|_n < \varepsilon.$$

A sequence (u_k) is a *Cauchy sequence* if $\forall U_{n,\varepsilon}(0)\ \exists N$ such that:

$$k,\ell > N(n,\varepsilon) \Rightarrow (u_k - u_\ell) \in U_{n,\varepsilon}(0).$$

The space Ω is said to be a *complete space* if all Cauchy sequences have a limit.

We need one more concept concerning the norms. The norms (7.1) must be pairwise compatible. Two norms say $p_1(u)$ and $p_2(u)$ defined on the same space Ω are called *compatible* if every sequence which is fundamental (Cauchy) in both norms and which converges in one of these norms converges also in the other norm.

DEFINITION 7.1. A *countably normed Hilbert space* Ω is a linear topological space in which the topology is given by a countable set of pairwise compatible norms (obtained from inner products) and which is complete.

REMARK 7.2. The norms are assumed to be ordered as in (7.2).

REMARK 7.3. The topology given by (7.3) is equivalent to the topology induced by the metric

$$d(u,v) \equiv \sum_{n=1}^{\infty} \frac{1}{2^n} \cdot \frac{\|u-v\|_n}{1+\|u-v\|_n} .$$

Using $d(u,v)$ as the distance between u and v the space Ω becomes a metric space and the completeness of Ω is in the sense of the metric.

REMARK 7.4. Ω is in general not a Hilbert space.

7.3. The Hilbert spaces Ω_n

Consider in a countably normed Hilbert space the inner product $(.,.)_n$ for arbitrary but fixed n. Relative to this inner product Ω is in general not a Hilbert space. One can however complete Ω with respect to the norm $\|.\|_n$ induced by $(.,.)_n$ and in doing so one obtaines a Hilbert space which will be called Ω_n. The space Ω is by definition dense in Ω_n. Performing this construction for all n one obtains a countable set of Hilbert spaces $\{\Omega_n\}_{n=1}^{\infty}$ which as a consequence of the ordering (7.2) are ordered according to

$$(7.4) \qquad \Omega_1 \supset \Omega_2 \supset \Omega_3 \;\ldots\ldots \supset \Omega.$$

From the completeness of Ω one can prove that Ω is the intersection of the spaces Ω_n

$$(7.5) \qquad \Omega = \bigcap_{n=1}^{\infty} \Omega_n.$$

From (7.5) one sees that Ω, being the intersection of normed spaces, is in general "smaller" than a normed space.

7.4. The dual spaces Ω'_n and Ω'

Consider the Hilbert space Ω_n. Let Ω'_n be its dual space, i.e. the linear space of bounded linear functionals F on Ω_n. The norm of a functional $F \in \Omega'_n$ is defined as

$$(7.6) \qquad p_n(F) = \sup_{\|u\|_n \leq 1} |F(u)|.$$

As is well-known from the Riesz-Fréchet theorem, the dual space Ω'_n is a Hilbert space isomorphic with Ω_n. We recall the contents of this theorem. First of all let u be a fixed element of Ω_n, then F_u defined by $F_u(v) = (u,v)_n$ is a bounded linear functional, i.e. $F_u \in \Omega'_n$. According to the theorem all bounded linear functionals on Ω_n are of this form, i.e., for each $F \in \Omega'_n$ there is a unique vector $u \in \Omega_n$ such that $F(v) = (u,v)_n$. The correspondence $F \leftrightarrow u$ is anti-linear. If $F \rightarrow u$ then $\alpha F \rightarrow \alpha^* u$. This follows from:

$$(\alpha F)(v) \equiv \alpha . F(v) = \alpha(u,v)_n = (\alpha^* u,v)_n.$$

The correspondence is isometric. If $F \rightarrow u$ then

$$p_n(F) = \sup_{\|v\|_n \le 1} |F(v)| = \sup_{\|v\|_n \le 1} |(u,v)|_n = \|u\|_n .$$

The inner product in Ω_n' is defined as $(F_1,F_2)_n = (u_2,u_1)_n$. This makes $(F_1,F_2)_n$ linear in the second argument, $(F_1,\alpha F_2)_n = (\alpha^* u_2,u_1)_n = \alpha(F_1,F_2)_n$. One usually identifies Ω_n' and Ω_n, but it is useful to keep in mind the anti-linearity of the mapping between Ω_n and Ω_n'.

We now consider the dual spaces Ω_n' for all n. This gives a sequence of Hilbert spaces ordered according to

$$\Omega_1' \subset \Omega_2' \subset \Omega_3' \ \dots \ . \tag{7.7}$$

The norms $\{p_n(F)\}_{n=1}^{\infty}$ are also ordered:

$$p_1(F) \ge p_2(F) \ge p_3(F) \ge \dots \ . \tag{7.8}$$

The dual space Ω' of Ω is the union

$$\Omega' = \bigcup_{n=1}^{\infty} \Omega_n'. \tag{7.9}$$

Indeed, a functional F on Ω is continuous iff F is continuous with respect to some norm $\| \ \|_m$.

7.5. Nuclear spaces

Consider the Hilbert spaces Ω_m and Ω_n with $\Omega_m \supset \Omega_n$ (n>m). The space Ω_n was obtained from Ω by completing Ω with respect to the norm $\| \ \|_n$. This means that the elements of Ω constitute an everywhere dense set in Ω_n. Ω_m was obtained in a similar way and Ω is also dense in Ω_m.

Consider now the injection of Ω_n into Ω_m, written as

$$\overset{(n)}{u} \mapsto \overset{(m)}{u} = \hat{T}_m^n \overset{(n)}{u},$$

where $\overset{(n)}{u} = \overset{(m)}{u} = u$ is an element $u \in \Omega$ but now considered as an element of Ω_n and Ω_m, respectively. The mapping $\hat{T}_m^n$ is a continuous linear mapping from a dense set in Ω_n onto a dense set in Ω_m. According to a well-knwon theorem such a mapping can be extended to a continuous linear operator T_m^n which maps Ω_n onto a dense subset in Ω_m by defining

$$T_m^n u = \lim_{k\to\infty} \hat{T}_m^n \overset{(n)}{u}_k, \text{ where } u = \lim_{k\to\infty} \overset{(n)}{u}_k, \ (\overset{(n)}{u}_k) \in \Omega.$$

We now introduce the concept of nuclearity of Ω.

DEFINITION 7.5. A countably normed Hilbert space $\Omega = \cap \Omega_n$ in which all Ω_n are separable is called *nuclear* if for any m there is an n,n > m, such that the mapping T_m^n from Ω_n into Ω_m takes the form

$$(7.10) \qquad T_m^n u = \sum_{k=1}^{\infty} \lambda_k (e_k, u)_n d_k \qquad (u \in \Omega_n),$$

where $\{e_k\}_{k=1}^{\infty}$ is an orthonormal system in Ω_n, $\{d_k\}_{k=1}^{\infty}$ an orthonormal system in Ω_m, $\lambda_k \geq 0$ and $\Sigma_{k=1}^{\infty} \lambda_k < \infty$.

An operator T_m^n which satisfies (7.10) is called a *nuclear operator*. One gets a bit of insight in this complicated expression if one takes $u = e_\ell$. This gives

$$T_m^n e_\ell = \lambda_\ell d_\ell,$$

so for T_m^n to be nuclear there must exist in Ω_n an orthonormal system of vectors $\{e_k\}$ and in Ω_m an orthonormal system of vectors $\{d_k\}$ such that e_k is mapped on $\lambda_k d_k$ with $\lambda_k \geq 0$ and $\Sigma \lambda_k < \infty$.

REMARK 7.6. Formula (7.10) can be rewritten in a slightly different form by using the connection between Ω_n and Ω_n'. To e_k there corresponds a functional $F_k \in \Omega_n'$ such that $F_k(u) = (e_k, u)_n$. We can thus write

$$(7.11) \qquad T_m^n u = \sum_{\lambda=1}^{\infty} \lambda_k F_k(u) d_k,$$

where $\{F_k\}_{k=1}^{\infty}$ is an orthonormal system of functionals in Ω_n'.

From now on we assume the countably normed Hilbert space Ω to be nuclear.

7.6. Rigged Hilbert space

We start from a countably normed nuclear Hilbert space Ω. Let there be given on Ω a scalar product (u,v) with $(u,u) \geq 0$, $(u,u) = 0 \Leftrightarrow u = 0$. For this scalar product we suppose that it satisfies apart from the usual properties the following requirements:
If $u_k \to u$ (in the topology of Ω) then

$$(7.12) \qquad \lim_{k\to\infty} (u_k, v) = (u,v), \qquad \forall v \in \Omega.$$

From $(u,v) = (v,u)^*$ we have also

(7.13) $\lim_{k\to\infty} (v,u_k) = (v,u)$.

These requirements make (u,v) a continuous linear functional in the second argument and a continuous anti-linear functional in the first argument, where continuity is with respect to the topology in Ω. An application of the Banach-Steinhaus theorem yields that (u,v) is jointly continuous in u and v with respect to the topology of $\Omega\times\Omega$. Such functionals are continuous (bounded) with respect to some norm $\|.\|_m$ (cf. the remark after (7.9)), i.e. there is a positive real number M and an index m such that

(7.14) $|(u,v)| \leq M\|u\|_m\|v\|_m$.

The space Ω can be completed relative to the norms $\|u\| = \sqrt{(u,u)}$ induced by the scalar product (.,.). In this way one obtains a Hilbert space H and Ω is dense in H. From the properties of Ω (nuclearity) one can prove that H is a seperable Hilbert space, i.e., there is a countable complete orthonormal system in H.

We consider the natural imbedding T of Ω in H. For $u \in \Omega$ we have $Tu = u \in H$. The operator T is a continuous linear operator. Identifying Ω with the subset $T\Omega$ in H we can write

(7.15) $\Omega \subset H$.

Considering the dual spaces we have H' and Ω'. The adjoint T' of T defined by

(7.16) $h'(Tu) = T'h'(u), \quad \forall u \in \Omega \ \& \ \forall h' \in H'$,

is a continuous mapping from H' into Ω'. From the anti-linear relationship between H and H' it follows that T' can be considered as an anti-linear mapping from H into Ω'. Identifying H with the image $T'H$ of H in Ω' we can write

(7.17) $H \subset \Omega'$.

Starting from Ω we are thus led to a triple of spaces Ω, H and Ω'. The above discussed mappings T and T' are symbolically represented by

(7.18) $\Omega \subset H \subset \Omega'$.

This set of spaces is the *Gelfand Triple* or *Rigged Hilbert Space*. One also encouters the name *enriched Hilbert space*.

We must next discuss a property of the operator T which is crucial in the analysis of the spectrum of self-adjoint linear operators.

7.7. Nuclearity of the mapping $\Omega \to T\Omega \subset H$

The mapping T from Ω into H is also continuous with respect to some norm $\|.\|_m$ on Ω_m. This follows from (7.14). For $u \in \Omega$ we have $Tu = u \in H$ and

$$(7.19) \qquad (Tu,Tu) = \|u\|^2 \le M\|u\|_m^2.$$

This means that T can be extended to a continuous mapping from Ω_m into H. Due to the ordering of norms (7.2), T is also continuous in the norm $\|.\|_n$ for $n > m$. We denote the extension of T for $n \ge m$ by T_n. These operators are continuous linear mappings from the spaces $\Omega_n (n \ge m)$ into H. Now choose $n > m$ such that T_m^n is nuclear, clearly $T_n = T_m T_m^n$. We now use a theorem which says that the product of a continuous operator and a nuclear operator is a nuclear operator. So T_n is a nuclear operator. This means that there exist orthonormal bases $\{e_k\}_{k=1}^{\infty}$ and $\{h_k\}_{k=1}^{\infty}$ in Ω_n and H, respectively, such that

$$(7.20) \qquad T_n u = \sum_{k=1}^{\infty} \lambda_k (e_k,u)_n h_k, \qquad \forall u \in \Omega_n,$$

where $\lambda_k \ge 0$ and $\Sigma\, \lambda_k < \infty$.

We can use (7.20) to give an explicit form for the operator T. For all $u \in \Omega$ we have $Tu = T_n U$ and

$$(7.21) \qquad Tu = \sum_{k=1}^{\infty} \lambda_k (e_k,u)_n h_k, \qquad \forall u \in \Omega.$$

Using again the Riesz-Fréchet theorem we can write $(e_k,u)_n = F_k(u)$ with $F_k \in \Omega_n'$.

The results obtained up to now can be summarized as follows: Let $\Omega \subset H \subset \Omega'$ be a Rigged Hilbert space and T the natural inbedding of Ω into H then T can be expressed as

$$(7.22) \qquad Tu = \sum_{k=1}^{\infty} \lambda_k F_k(u) h_k, \qquad \forall u \in \Omega.$$

This means that there is an index n and there are orthonormal systems $\{F_k\}_{k=1}^{\infty}$ and $\{h_k\}_{k=1}^{\infty}$ in Ω_n' and H, respectively, such that (7.22) holds with the λ_k's satisfying the by now well-known requirements.

Formula (7.22) enables one to prove that, once the Hilbert space H and the space Ω are realized as function spaces, i.e., elements of Ω are functions u(x), there exists to each x a linear functional F_x such that $F_x(u) = u(x)$. This will be the next topic.

7.8. Function spaces

We consider a Rigged Hilbert space $\Omega \subset H \subset \Omega'$. Let the Hilbert space H be realized as a space $L^2(X,S,\mu)$ where (X,S,μ) is some measure space. Elements are measurable "functions" u(x) defined up to a set $N \in S$ with $\mu(N) = 0$ and $\int_X |u(x)|^2 d\mu(x) < \infty$. The problem of constructing $\forall x \in X$ a continuous linear functional F_x on L^2 which assigns to $u \in H$ the value u(x) in the point x cannot be formulated properly in general. The reason is that, for a specific x, u(x) is not uniquely given for $u \in L^2(X)$, unless {x} has nonzero measure.

However suppose that we have a linear space V of functions on X such that each $u \in L^2(X)$ has one and only one representative in V and such that each function in V is a representative of some $u \in L^2(X)$. That is, V is another model of $L^2(X)$, consisting of functions rather than equivalence classes of functions. Now we may ask: does there exist a family $\{F_x | x \in X\}$ of continuous linear functionals on V such that $F_x(u) = u(x)$, $\forall u \in V$ and $x \in X$? This would imply that for any representative u of an element of $L^2(X)$ (not necessarily lying in V) we have $F_x(u) = u(x)$ a.e. $[\mu]$.

It can be shown for a wide collection of measure spaces, including $\mathbb{R}^n$ with Lebesgue measure, that such a collection of continuous linear functionals cannot exist. On a Rigged Hilbert space however such functionals do exist. We explain how such functionals are constructed. We assume that the Hilbert space H is again realized as a L^2-space. We can then use the imbedding T of Ω into H to give a realization of Ω as a space of equivalence classes of functions:

$$u \in \Omega \mapsto Tu = u \in H.$$

Take λ_k, h_k and F_k as in (7.22). For each k choose a function $x \to h_k(x)$ which is a representative of $h_k \in L^2(X)$. Now define $\forall x \in X$ a functional F_x by

$$F_x = \sum_{k=1}^{\infty} \lambda_k h_k(x) F_k. \tag{7.23}$$

The functionals $\{F_k\}_{k=1}^{\infty}$ consitute a complete orthonormal set of continuous linear functionals in Ω_n'. One can prove that the right hand side of (7.23) converges to a functional in Ω_n' for all x with the possible exception of a set of measure zero. On this set we take $F_x := 0$. In this way one obtains a continuous linear functional $\forall x \in X$. Applying F_x defined by (7.23) to $u \in \Omega$ we have

$$(7.24) \qquad F_x(u) = \sum_{k=1}^{\infty} \lambda_k F_k(u) h_k(x) \quad \text{a.e. } [\mu].$$

On the other hand we have

$$(7.22) \qquad Tu = u = \sum_{k=1}^{\infty} \lambda_k F_k(u) h_k,$$

where the right hand side converges in the L^2-norm. This implies that

$$(7.25) \qquad u(x) = \sum_{k=1}^{\infty} \lambda_k F_k(u) h_k(x)$$

converges a.e. $[\mu]$. From (7.24) and (7.25) we have

$$(7.26) \qquad F_x(u) = u(x) \quad \text{a.e. } [\mu].$$

As the function $u(x)$ assigned to $u \in \Omega$ is defined up to a set of μ-measure zero one can define $u(x)$ on this set such that (7.26) holds for all x. Assuming that this has been done we have for all $x \in X$ a bounded linear functional F_x which acting on $u \in \Omega$ gives the value of the function assigned to u in the point x.

This result can be extended to the case that the Hilbert space $\mathcal{H}$ is realized as a direct sum or a direct integral of Hilbert spaces. The latter concept is presented in §VIII.7. We will not go into this extension.

7.9. Operators on Rigged Hilbert space; the eigenvalue problem

Let A be a linear operator, defined on Ω, which satisfies

$$(7.27) \qquad (Au,v) = (u,Av), \quad \forall u \ \& \ v \in \Omega,$$

where $(.,.)$ is the inner product discussed in section 7.6. As Ω is dense in $\mathcal{H}$ in the sense of this inner product one can define the closure $\bar{A}$ of A: If a sequence $\{u_k\} \in \Omega$ converges to $u \in \mathcal{H}$, $u_k \to u$, and if $\{Au_k\} \subset \Omega$ converges to an element $v \in \mathcal{H}$, $Au_k \to v$, then $\bar{A}$ is defined by $v = \bar{A}u$. The operator A is said to be hermitian on the Rigged Hilbert space $\Omega \subset \mathcal{H} \subset \Omega'$ if the closure $\bar{A}$ is hermitian on $D_{\bar{A}} \cap \mathcal{H}$ with respect to the scalar product.

Let now A be hermitian on $\Omega \subset H \subset \Omega'$. A linear functional $F_a \in \Omega'$, $F_a \neq 0$, is said to be a *generalized eigenvector* belonging to the eigenvalue a of A if

$$(7.28) \qquad F_a(Au) = aF_a(u), \quad \forall u \in \Omega.$$

The set of generalized eigenvectors of A is said to be *complete* if $F_a(u) \equiv 0 \Rightarrow u = 0$. We now state the main theorem concerning the spectrum of hermitian operators.

THEOREM 7.7. *A self-adjoint (hermitian) operator* A *on a rigged Hilbert space possesses a complete set of generalized eigenvectors corresponding to real eigenvalues.*

The proof of the theorem can be found in [6]. An analogous theorem holds for unitary operators on a rigged Hilbert space. The eigenvalues are complex numbers of modulus one.

7.10. Examples of Rigged Hilbert spaces

To conclude this section on Rigged Hilbert spaces we give some examples. In the first two examples we start from a Hilbert space and we construct Rigged Hilbert spaces from it. The third example concerns the spaces S and S'.

EXAMPLE 7.8. Let H be a separable Hilbert space with orthonormal basis $\{e_n\}_{n=1}^{\infty}$. We define a countably normed Hilbert space Ω, such that $\Omega \subset H \subset \Omega'$ becomes a Gelfand triple. Let Ω_n (n=0,1,2,...) consists of all elements $u = \Sigma_{k=1}^{\infty} c_k e_k \in H$ such that $\Sigma_{k=1}^{\infty} k^{4n}|c_k|^2 < \infty$. Define on Ω_n an inner product

$$(u,v)_n = \sum_{k=1}^{\infty} k^{4n}(u,e_k)(e_k,v).$$

Clearly the vectors $\{k^{-2n}e_k\}_{k=1}^{\infty}$ form an orthonormal system in Ω_n. The norms $\|.\|_n$ obtained from the inner products $(.,.)_n$ are ordered as in (7.2) and the Hilbert spaces Ω_n satisfy $H = \Omega_0 \supset \Omega_1 \supset \Omega_2 \ldots$. Consider now the injection T_m^n from Ω_n into $\Omega_m (n > m)$ defined by $T_m^n u = u$, $u \in \Omega_n$. For the basis vectors in Ω_n we have $T_m^n(k^{-2n}e_k) = k^{2(m-n)}(k^{-2m}e_k)$ and $\Sigma_{k=1}^{\infty} k^{2(m-n)} < \infty$. Comparing this with the requirements discussed in section 7.5 we see that T_m^n is a nuclear operator. Finally define the space Ω by $\Omega = \bigcap_{n=0}^{\infty} \Omega_n$, then Ω is a countably normed Hilbert space and $\Omega \subset H \subset \Omega'$ is a Gelfand triple.

EXAMPLE 7.9. We consider the space $H = L^2(\mathbb{R} \pmod{2\pi})$ with inner product

$$(u,v) = \frac{1}{2\pi}\int_0^{2\pi} \overline{u(x)}\,v(x)\,dx.$$

The functions $x \to e^{ikx} = \chi_k(x)$, $k \in \mathbb{Z}$, constitue an orthonormal basis in H. Consider now the linear space Ω_n $(n=0,1,2,\ldots)$ consisting of all $u \in H$ for which the $(2n)^{th}$ derivative (in distributional sense) is in L^2. This is just the space of all u with Fourier series $u(x) = \Sigma_{k=-\infty}^{\infty} c_k e^{ikx}$ such that $\Sigma_{k=-\infty}^{\infty} k^{4n}|c_k|^2 < \infty$. On Ω_n we define an inner product as follows:

$$(u,v)_n = \frac{1}{2\pi}\int_0^{2\pi} \left\{\overline{u^{(2n)}(x)}\,v^{(2n)}(x)\,dx + \bar{u}(x)v(x)\right\}dx$$

$$= \sum_{k=-\infty}^{\infty} (k^{4n}+1)(u,\chi_k)(\chi_k,v).$$

The functions $x \to \frac{1}{(k^{4n}+1)^{\frac{1}{2}}}\, e^{ikx}$, $k \in \mathbb{Z}$ form an orthonormal system in Ω_n. By this construction we obtain again a set of Hilbert spaces ordered according to $H = \Omega_0 \supset \Omega_1 \supset \Omega_2 \ \ldots$. The injection $T_m^n\colon \Omega_n \to \Omega_m$ $(n > m)$ is a nuclear operator and $\Omega \subset H \subset \Omega'$ with $\Omega := \bigcap_{n=0}^{\infty} \Omega_n$ is a Gelfand triple.

EXAMPLE 7.10. The final example to be discussed is the triple of spaces $S(\mathbb{R}) \subset L^2(\mathbb{R}) \subset S'(\mathbb{R})$ (cf. BÖHM [3]). Here $S(\mathbb{R})$ is the space of C^∞ functions of fast decrease. The topology on S which makes S a countably normed Hilbert space is introduced by means of the inner products

$$(u,v)_n = \int_{-\infty}^{\infty} (1+x^2)^{2n}\left(\sum_{k=0}^{n} \overline{u^{(k)}(x)}\,v^{(k)}(x)\right)dx.$$

The inner product on S that leads to $H = L^2(\mathbb{R})$ is the usual one:

$$(u,v) = \int_{-\infty}^{\infty} \bar{u}(x)v(x)\,dx.$$

Denoting by S_n $(n=0,1,2,\ldots)$ the linear space obtained from S by imposing the inner product $(.,.)_n$ and by completion in the norm $\|.\|_n$ we have $H = S_0 \supset S_1 \supset S_2 \supset \ldots \supset S$ and $S = \bigcap_{n=0}^{\infty} S_n$. In $H = S_0$ we have the orthonormal basis given by the harmonic oscillator eigenfunctions $\{u_\ell(x) = N_\ell \exp(-\frac{1}{2}x^2)H_\ell(x)\}_{\ell=0}^{\infty}$ where H_ℓ $(\ell=0,1,2,\ldots)$ are the Hermite polynomials and N_ℓ are normalization constants. We will not go into the problem of the nuclearity of the injection $T_m^n\colon S_n \to S_m$ $(n > m)$. The triple $S(\mathbb{R}) \subset L^2(\mathbb{R}) \subset S'(\mathbb{R})$, where S' is the space of tempered distributions on S, is a Gelfand triple. The generator of translations $\underline{\underline{p}} = -i\hbar d/dx$ is a well-defined operator on S, moreover $\underline{\underline{p}}$ is hermitian with respect to the inner product (u,v) for u & $v \in S$. The eigenvalue problem for $\underline{\underline{p}}$ can be formulated. The generalized eigenfunctions $F_p(x) = (2\pi\hbar)^{-\frac{1}{2}} e^{ipx/\hbar}$ $(p \in \mathbb{R})$ are elements of S' and $F_p(\underline{\underline{p}}u) = pF_p(u)$, $\forall u \in S$. The action of F_p on S is explicitly given by

$$F_p(u) = (2\pi\hbar)^{-\frac{1}{2}} \int_{-\infty}^{\infty} e^{-ipx/\hbar} u(x)\,dx.$$

Using Dirac's notation: $F_p(u) = \langle p|u\rangle$. The set $\{F_p\}$ is complete: from $F_p(u) \equiv 0$ it follows that $u = 0$

Acknowledgement. The author wishes to thank T.H. Koornwinder for the stimulating discussions and his critical remarks, especially on the section of Rigged Hilbert Space. The examples 7.8 and 7.9 are due to him.

LITERATURE

On the Poincaré group: [2, ch.4,5,6], [7], [8].
On Quantum Mechanics: [5].
On Rigged Hilbert spaces: [1], [2, ch.1,4,8], [3], [4], [6].

[1] ANTOINE, J.P., *Dirac formations and symmetry problems in quantum mechanics*. II. *Symmetry problems*, J. Math. Phys. 10 (1969), 2276-2290.

[2] BOGOLIUBOV, N.N., A.A. LUGUNOV & I.T. TODOROV, *Introduction to axiomatic quantum field theory*, Benjamin, Reading (Mass.), 1975.

[3] BÖHM, A., *Rigged Hilbert space and mathematical description of physical systems*, in Boulder Lectures 1966, Gordon & Breach, New York, 1967.

[4] BÖHM, A., *The Rigged Hilbert space and quantum mechanics*, Lecture Notes in Physics, 78, Springer-Verlag, Berlin, 1978.

[5] DIRAC, P.A.M., *The principles of quantum mechanics*, Oxford University Press, Oxford, 1947.

[6] GELFAND, I.M., & N.J. WILENKIN, *Verallgemeinerte Funktionen, IV*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1964.

[7] WIGHTMAN, A.S., *L'Invariance dans la mécanique quantique relativiste,* in "Relations de dispersion et Particules élémentaires" (Ecole d'été de physique théorique, Les Houches, 1960), John Wiley & Sons Inc., New York, 1960.

[8] WIGNER, E.P., *Unitary representations of the inhomogeneous Lorentz group including reflections,* in Lectures Notes of the Istanbul Summer School 1962, Gordon & Breach, New York, 1964.

IV

LIE GROUPS AND LIE ALGEBRAS

T.H. KOORNWINDER

Mathematisch Centrum

CONTENTS

In this chapter we intend to give a rather detailed account of the basic facts on Lie groups, and on Lie algebras as far as they are related to Lie groups. It is divided in three sections. In the first section we develop the notions of an analytic manifold and of a Lie group, using the unitary group and other linear groups as motivating examples. We conclude this section with some topological concepts which are important in Lie group theory, for instance connectedness. The next section deals with the relationship between Lie groups and Lie algebras. In order to define the Lie algebra of a Lie group we first introduce tangent vectors and vector fields on an analytic manifold. The connection between Lie groups and Lie algebras can be made very close by means of the exponential mapping. This part of section 2 together with the two next subsections, which explore the connection between subgroups and subalgebras and between Lie group homomorphisms and Lie algebra homomorphisms, can be considered as the heart of this chapter. The two final subsections of section 2 deal with locally isomorphic Lie groups and with the adjoint representation. Section 3, concluding this chapter, shortly discusses the various kinds of Lie groups and Lie algebras: simple, semisimple, solvable and nilpotent.

We have tried to write this chapter in an informal way, motivating every new step and giving almost no formal proofs. However, in writing this chapter the author learnt that many of the relevant proofs can be sketched in a few lines, such that the reader can easily make things complete. For certain harder proofs we have given suitable references, mostly to CHEVALLEY [2] and VARADARAJAN [10]. To some extent the author was influenced by the rather elementary approach followed in MILLER [7, Chap. 5]. Of course, the theory of Lie groups cannot be given without using some concepts from topology, but we hope that a little knowledge of topology on $\mathbb{R}^n$ and some intuitive notion of general topological spaces will be sufficient for understanding this chapter.

1. LIE GROUPS

Informally stated, a Lie group is a group G on which a neighbourhood of the identity element e can be described by real coordinates $(x_1,x_2,\ldots,x_m)$ such that the mapping

$$g \to (x_1(g), x_2(g), \ldots, x_m(g))$$

identifies the neighbourhood of e with an open subset of $\mathbb{R}^m$, the coordinates $x_i(gh)$ $(i = 1,\ldots,m)$ are analytic functions of the coordinates $x_1(g)$, $\ldots, x_m(g)$, $x_1(h),\ldots,x_m(h)$, and the coordinates $x_i(g^{-1})$ $(i = 1,\ldots,m)$ are analytic functions of $x_1(g),\ldots,x_m(g)$, where g and h are in a sufficiently small neighbourhood of e. Furthermore, on a Lie group G there must exist local coordinates around any point $g \in G$ and if on some region two such local coordinate systems overlap then the transformation from the one to the other coordinate system must be analytic. If $(x_1(g),\ldots,x_m(g))$ are local coordinates around e then a standard method to introduce local coordinates around some $g_0 \in G$ is by the definition

$$(y_1(g_0g),\ldots,y_m(g_0g)) := (x_1(g),\ldots,x_m(g)),$$

where g is near e.

1.1. Linear groups

The prototype of a Lie group is the so-called *general linear group* $GL(n,\mathbb{C})$ which consists of all complex nonsingular n×n matrices. Note that $GL(n,\mathbb{C})$ is a subset of the set $M_n(\mathbb{C})$ consisting of all complex n×n matrices. By using the n^2 matrix entries A_{ij} of $A \in M_n(\mathbb{C})$ as coordinates we can identify $M_n(\mathbb{C})$ as a complex linear space with the n^2-dimensional complex vector space $\mathbb{C}^{n^2}$. The usual topology on $\mathbb{C}^{n^2}$ thus defines a topology on $M_n(\mathbb{C})$. Now $GL(n,\mathbb{C})$ is the set of all $A \in M_n(\mathbb{C})$ such that $\det A \neq 0$. Since det is a continuous complex-valued function on $M_n(\mathbb{C})$ it follows that $GL(n,\mathbb{C})$ is an open subset of $M_n(\mathbb{C})$.

If $A,B \in GL(n,\mathbb{C})$ then

$$(AB)_{ij} = \sum_{k=1}^{n} A_{ik} B_{kj}$$

and

$$(A^{-1})_{ij} = (-1)^{i+j} m_{ji}/\det A\,,$$

where m_{ji} is the $(j,i)^{th}$ minor of the matrix A. Remember that a mapping F from an open subset V of $\mathbb{C}^k$ (or $\mathbb{R}^k$) into an open subset W of $\mathbb{C}^\ell$ (or $\mathbb{R}^\ell$) is called *complex analytic* (or *real analytic*) if for each $z_0 \in V$ and for each $j = 1,\ldots,\ell$ the j^{th} complex (or real) coordinate $F_j(z)$ of $F(z)$ can be written as a power series in the complex (or real) variables $(z_1-z_{0,1}),\ldots,(z_k-z_{0,k})$ for z in some neighbourhood of z_0. So we conclude that the mapping $(A,B) \to AB$ is complex analytic from $GL(n,\mathbb{C}) \times GL(n,\mathbb{C})$ into $GL(n,\mathbb{C})$ and the mapping $A \to A^{-1}$ is complex analytic from $GL(n,\mathbb{C})$ into $GL(n,\mathbb{C})$. The complex vector space $\mathbb{C}^k$ can be identified with the real vector space $\mathbb{R}^{2k}$ by taking the real and imaginary parts of the complex coordinates as real coordinates. Thus a complex analytic mapping F from $V \subset \mathbb{C}^k$ into $W \subset \mathbb{C}^\ell$ (V and W open) can also be considered as a real analytic mapping from $V \subset \mathbb{R}^{2k}$ into $W \subset \mathbb{R}^{2\ell}$. In particular, the group operations on $GL(n,\mathbb{C})$ are also real analytic mappings.

By definition a *linear group* is a subgroup of $GL(n,\mathbb{C})$ for some n. Many of the groups encountered in physics (for instance SO(3), SU(2) and the Lorentz group L) are linear groups. Very often, these groups are *closed* (i.e. topologically closed) subgroups of some $GL(n,\mathbb{C})$. For instance, the unitary group $U(n) := \{T \in M_n(\mathbb{C}) \mid T^*T = I\}$, which is a subgroup of $GL(n,\mathbb{C})$, is a closed subset of $M_n(\mathbb{C})$ since it is the inverse image of I under the continuous mapping $T \to T^*T$ from $M_n(\mathbb{C})$ into itself. Hence U(n) is also closed in $GL(n,\mathbb{C})$. (In the definition of U(n) the symbol T^* denotes the *adjoint* of the matrix T, i.e. $(T^*)_{ij} := \bar{T}_{ji}$; it corresponds with $T^\dagger$ in the physical literature.) A list of some important linear groups is given in Table 1 in §3.4.

For linear groups there is a simple compactness criterium. Remember that the compact subsets of some vector space $\mathbb{R}^m$ are precisely the subsets which are both closed and bounded. This yields:

PROPOSITION 1.1. *A subgroup* G *of* $GL(n,\mathbb{C})$ *is compact if and only if* G *is both closed and bounded in* $M_n(\mathbb{C})$. *(Boundedness of* G *means that all matrix entries* T_{ij} *are bounded for* $T \in G$.)

In particular, note that U(n) is bounded in $M_n(\mathbb{C})$ and hence compact, since all columns of a matrix $T \in U(n)$ are unit vectors in $\mathbb{C}^n$. However, the Lorentz group L, which is a subgroup of $GL(4,\mathbb{C})$, contains the special Lorentz transformations given by (III. 2.15), which clearly form an unbounded subset of $M_4(\mathbb{C})$. This implies that L is noncompact.

We will now indicate how an analytic structure can be defined on a number of familiar linear groups such that the group operations become analytic. The main problem in doing so is that often one set of coordinates is not sufficient to cover the whole group. We have to work with a number of overlapping coordinate systems, i.e., we have to introduce the notion of an analytic manifold and of an analytic mapping from one analytic manifold to another. Then we can define a Lie group. The discussion of a certain class of linear groups in subsections 1.2 and 1.3 below will serve as a guiding example for the general definitions of analytic manifolds and Lie groups in subsection 1.4. In § 1.2 we will use the exponential mapping in order to define local coordinates around the identity element such that the group operations become locally analytic. In § 1.3 we will extend the local analytic structure to the whole group by means of left translations. Then the group operations become globally analytic.

1.2. Introducing a local Lie group structure by means of the exponential mapping

Let $A \in M_n(C)$ and define the exponential of A by the power series

$$e^A = \exp A := \sum_{j=0}^{\infty} \frac{1}{j!} A^j, \tag{1.1}$$

where absolute convergence holds for each of the n^2 matrix entries. Then:

$$e^0 = I \,;\, e^{A+B} = e^A e^B \text{ if } AB = BA; \ (e^A)^{-1} = e^{-A}.$$

Hence $e^A \in GL(n,\mathbb{C})$ for all $A \in M_n(\mathbb{C})$.

PROPOSITION 1.2 (cf. MILLER [7,§5.1]). *The mapping* exp *is a complex analytic mapping from* $M_n(\mathbb{C})$ *into* $GL(n,\mathbb{C})$.
There is an open neighbourhood V *of* 0 *in* $M_n(\mathbb{C})$ *such that the image set* exp(V) *is an open neighbourhood of* I *in* $GL(n,\mathbb{C})$, *the mapping* exp *is one-to-one from* V *onto* exp(V), *and both* exp: V → exp(V) *and the inverse mapping* $\exp^{-1}$: exp(V) → V *are complex analytic mappings.*

We will denote the mapping $\exp^{-1}$: exp(V) → V in the above theorem by log. The second part of Proposition 1.2 follows from the *inverse function theorem*: If F is a complex analytic mapping from an open subset U of $\mathbb{C}^k$ into $\mathbb{C}^k$ and if for some $z_0 \in U$ the determinant of the Jacobian matrix

$\left(\frac{\partial F_i(z)}{\partial z_j}\bigg|_{z=z_0}\right)$ is nonzero then there is an open neighbourhood $V \subset U$ of z_0 such that $F(V)$ is an open neighbourhood of $F(z_0)$ in $\mathbb{C}^k$, the mapping F is one-to-one from V onto $F(V)$ and the inverse mapping F^{-1}: $F(V) \to V$ is also complex analytic. A similar theorem holds for real analytic mappings. In this situation the mapping F: $V \to F(V)$ is called a (real or complex) *analytic diffeomorphism*.

For many familiar linear groups G it can be verified that the following assumption holds:

ASSUMPTION 1.3. G is a subgroup of $GL(n,\mathbb{C})$ and there is a real linear subspace L of $M_n(\mathbb{C})$ such that

$$\exp(L \cap V) = G \cap \exp(V), \tag{1.2}$$

where V is some open neighbourhood of 0 in $M_n(\mathbb{C})$ on which exp: $V \to \exp(V)$ is a complex analytic diffeomorphism. (L is called a real linear subspace of $M_n(\mathbb{C})$ if $A,B \in L$ and $\alpha,\beta \in \mathbb{R}$ imply that $\alpha A+\beta B \in L$.)

Formula (1.2) is illustrated by Figure 1.

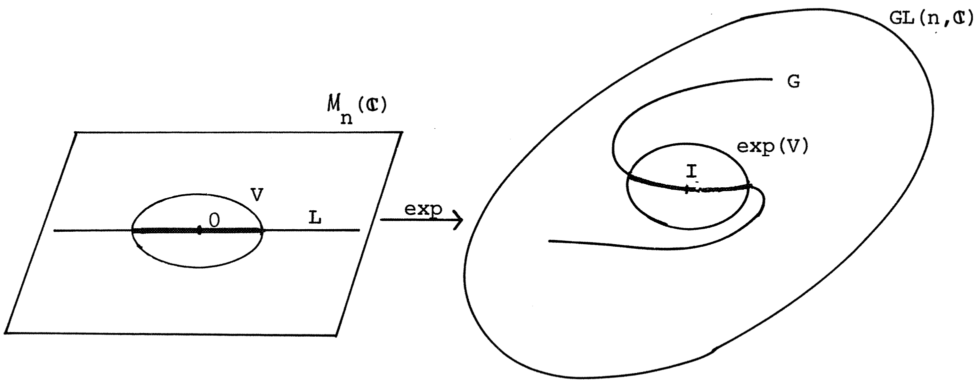

Figure 1.

If G, V and L satisfy Assumption 1.3 then we may choose $2n^2$ real linear coordinates x_i on $M_n(\mathbb{C})$ such that the linear subspace L consists of all $A \in M_n(\mathbb{C})$ for which the last $2n^2 - m$ coordinates are zero, where m is the dimension of L. Then the mapping log identifies the open neighbourhood exp(V) of I in GL(n,$\mathbb{C}$) with the open neighbourhood V of 0 in $M_n(\mathbb{C})$ and describes at the same time a real analytic coordinate transformation from the real and imaginary parts of the matrix entries of $T \in \exp(V)$ to the $2n^2$ coordinates x_i of log T. Furthermore, the mapping log: $\exp(V) \to V$ restricted to G identifies the open neighbourhood $G \cap \exp(V)$ of I in the group G with the open neighbourhood $L \cap V$ of 0 in the linear space L and introduces at the same time the first m coordinates $x_1, \dots, x_m$ of log T as coordinates for $T \in G \cap \exp(V)$. In the following we will not use these explicit coordinates x_i, but we will simply say that coordinates for $T \in G \cap \exp(V)$ are given by log T and that these coordinates take their values on the open subset $L \cap V$ of the linear space L.

Note that in Assumption 1.3 L is completely determined by G and V, since it is the real linear span of $\log(G \cap \exp(V))$. In fact, L only depends on G and it will turn out in § 2 that L is the Lie algebra of G.

Let us verify Assumption 1.3 for two examples. First let G be the so-called *special linear group* $SL(n,\mathbb{C}) := \{T \in M_n(\mathbb{C}) \mid \det T = 1\}$. This is a subgroup of GL(n,$\mathbb{C}$). Choose V as in Proposition 1.2 with the additional assumption that $|\mathrm{tr}\, A| < 2\pi$ if $A \in V$. Let L consist of all $A \in M_n(\mathbb{C})$ with trace zero. This is a complex linear subspace, which is denoted by $\mathfrak{sl}(n,\mathbb{C})$, cf. Table 2 in §3.4. Then (1.2) follows by the use of

$$\det(\exp A) = e^{\mathrm{tr}\, A} \tag{1.3}$$

(cf. MILLER [7, Cor. 5.3]). Since L is a complex linear space, the mapping log here defines complex coordinates locally around I on SL(n,$\mathbb{C}$).

As a second example let G = U(n), choose V as in Proposition 1.2, let $W \subset V$ be an open neighbourhood of 0 in $M_n(\mathbb{C})$ such that A^* and $-A \in V$ if $A \in W$. Let L consist of all $A \in M_n(\mathbb{C})$ which are *skew-hermitian*, i.e. $A + A^* = 0$. This is a real but not complex linear subspace of $M_n(\mathbb{C})$, which is denoted by $\mathfrak{u}(n)$. Now using $(\exp A)^* = \exp(A^*)$ and $(\exp A)^{-1} = \exp(-A)$ the reader can easily verify (1.2) with V replaced by W.

PROPOSITION 1.4. *Let* G, V *and* L *be as in Assumption* 1.3. *Then there is an open neighbourhood* $W \subset L \cap V$ *of* 0 *in* L *such that:*
(a) exp A exp B and $(\exp A)^{-1}$ *are in* exp(V) *if* $A, B \in W$.
(b) *the mapping* $(A,B) \to \log(\exp A \exp B)$ *is analytic from* $W \times W$ *into* $L \cap V$;
(c) *the mapping* $A \to \log((\exp A)^{-1})$ *is analytic from* W *into* $L \cap V$.

By "analytic" we usually mean "real analytic". However, if L is a complex linear subspace of $M_n(\mathbb{C})$ then the mappings in (b) and (c) are complex analytic. Note that we can always choose W such that $A \in W$ implies $-A \in W$. Then for $A \in W$ we have $(\exp A)^{-1} = \exp(-A) \in \exp(V)$ and $\log((\exp A)^{-1}) = -A$. Thus part (c) of the proposition is a triviality.

The proof of Proposition 1.4 is straightforward. We can interpret the result as follows. By means of log the local group structure of G is transferred to an open neighbourhood of 0 in the real vector space L. For sufficiently small $A, B \in L$ the product operation is defined by the mapping $(A,B) \to \log(\exp A \exp B)$ and it is analytic. With respect to this product operation the inverse of sufficiently small $A \in L$ is $-A$. The set W in Proposition 1.4 together with the local group structure just described is an example of a *local Lie group,* cf. MILLER [7, § 5.2].

Of course there are many ways to introduce local analytic coordinates on a given closed linear subgroup of $GL(n,\mathbb{C})$. However, the above method using log is very canonical. In fact, the coordinates thus obtained are called *canonical coordinates.*

1.3. Extension of the local Lie group structure to the whole group

In practice it happens quite often that we have G, V and L as in Assumption 1.3 such that $\exp(L \cap V)$ covers the whole group G except for a set of lower dimension. Then the local coordinates on G defined by log are usually sufficient for practical purposes. However, in order to describe the exceptional points of G in a neat way and in order to obtain a general theory, we want to cover the whole group G (which is supposed to satisfy Assumption 1.3) by open subsets on which local analytic coordinates are defined such that the group operations become analytic. Again there are many different ways to do this, but an easy, straightforward and canonical method is as follows. Let V and L be as in Assumption 1.3. Let $g_0 \in G$. Then we must choose local coordinates around g_0 such that the mapping $g \to g_0^{-1} g$ is analytic from the open neighbourhood $g_0 \exp(L \cap V)$ of

g_0 in G onto the open neighbourhood $\exp(L \cap V)$ of I in G. Now the analyticity of the above mapping has to be understood in terms of the coordinates we already had chosen around I. Hence the mapping $g \to \log(g_0^{-1}g)$ must be analytic from $g_0 \exp(L \cap V)$ onto the open neighbourhood $L \cap V$ of 0 in the linear space L, i.e. $\log(g_0^{-1}g)$ must depend analytically on the local coordinates of g chosen around g_0. Now what is easier than considering $\log(g_0^{-1}g)$ itself as the local coordinates for g around g_0? Rather quickly we now arrive at the following proposition.

Let G, V and L be as in Assumption 1.3. Let $W \subset V$ be an open neighbourhood of 0 in $M_n(\mathbb{C})$ and write for each $g_0 \in G$:

$$(1.4) \qquad U_{g_0} := g_0 \exp(L \cap W) = G \cap g_0 \exp(W),$$

which is an open neighbourhood of g_0 in G. Now the mapping ϕ_{g_0} defined by

$$(1.5) \qquad \phi_{g_0}(g) := \log(g_0^{-1}g), \qquad g \in U_{g_0},$$

is one-to-one from U_{g_0} onto the open neighbourhood $L \cap W$ of 0 in L and it defines local coordinates around g_0 on G.

<u>PROPOSITION 1.5</u>. *The above neighbourhood* W *of* 0 *in* $M_n(\mathbb{C})$ *can be chosen such that:*

(i) *If* $g_0, h_0 \in G$ *and* U_{g_0} *and* U_{h_0} *have nonempty intersection then the mapping*

$$\phi_{g_0}(g) \to \phi_{h_0}(g)$$

is analytic for g *restricted to this intersection.*

(ii) *If* $g_0, h_0 \in G$ *then the mapping*

$$(\phi_{g_0}(g), \phi_{h_0}(h)) \to \phi_{g_0h_0}(gh)$$

is analytic for (g,h) *in a certain neighbourhood of* (g_o, h_0).

(iii) *If* $g_0 \in G$ *then the mapping*

$$\phi_{g_0}(g) \to \phi_{g_0^{-1}}(g^{-1})$$

is analytic for g *in a certain neighbourhood of* g_0.

1.4. Analytic manifolds and Lie groups

The choice of the open sets U_{g_0} and the mapping ϕ_{g_0}, together with part (i) of Proposition 1.5., makes G into an analytic manifold. Let us now give the formal definition of an analytic manifold. For technical reasons we will work with Hausdorff spaces: A topological space M is called *Hausdorff* if any two distinct points of M have disjoint neighbourhoods. A *local chart* of *dimension* m on a topological Hausdorff space M is a pair (U,ϕ) such that U is a nonempty open subset of M and ϕ is a one-to-one mapping from U onto some open subset $\phi(U)$ of $\mathbb{R}^m$ which is a homeomorphism (i.e., both ϕ and the inverse mapping ϕ^{-1} are continuous mappings).

DEFINITION 1.6. A *real analytic manifold* of *dimension* m is a nonempty topological Hausdorff space M together with a collection of local charts (U_α,ϕ_α) of dimension m such that:

(i) The union of all U_α's equals M.

(ii) If U_α and U_β have nonempty intersection then the mapping $\phi_\alpha \circ \phi_\beta^{-1}$ is analytic from $\phi_\beta(U_\alpha \cap U_\beta)$ onto $\phi_\alpha(U_\alpha \cap U_\beta)$ (which are both open subsets of $\mathbb{R}^m$), cf. Figure 2.

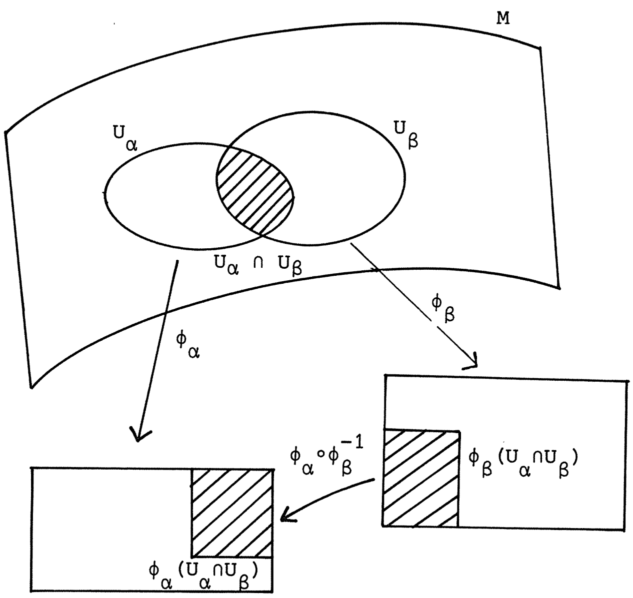

Figure 2.

A complex analytic manifold is defined in a similar way, with $\mathbb{R}^m$ replaced by $\mathbb{C}^m$ and the mappings $\phi_\alpha \circ \phi_\beta^{-1}$ being complex analytic. By the identification of $\mathbb{C}^m$ with $\mathbb{R}^{2m}$ any complex analytic manifold also becomes a real analytic manifold. For convenience we will restrict ourselves to the real case and we will skip the adjective "real".

A collection of local charts (U_α,ϕ_α) on M satisfying (i) and (ii) of Definition 1.6 is called an *atlas* for M. If (U_α,ϕ_α) is a local chart and $\phi_\alpha(p) = (x_1(p),\ldots,x_m(p))$ $(p \in U_\alpha)$ then the real-valued functions x_i $(i = 1,\ldots,m)$ on U_α are called *local coordinates* on M.

The geographic terminology introduced above immediately leads to the unit sphere S^2 in $\mathbb{R}^3$ as a standard example of a (two-dimensional) analytic manifold. Let $x \in S^2$ have cartesian coordinates (x_1,x_1,x_3) and define six local charts $(U_{k,j},\phi_{k,j})$ $(k = 1,2,3;\ j = 0,1)$ on S^2 by

$$U_{k,j} := \{x \in S^2 \mid (-1)^j x_k > 0\}$$

and

$$\phi_{k,j}(x) := \begin{cases} (x_2,x_3) & \text{if } k = 1 \text{ and } j = 0,1, \\ (x_1,x_3) & \text{if } k = 2 \text{ and } j = 0,1, \\ (x_1,x_2) & \text{if } k = 3 \text{ and } j = 0,1. \end{cases}$$

Then the $U_{k,j}$'s cover S^2 and condition (ii) is easily verified, for instance

$$\phi_{1,0}(U_{1,0} \cap U_{2,0}) = \{(y,z) \mid y > 0,\ y^2+z^2 < 1\}$$

and

$$\phi_{2,0} \circ \phi_{1,0}^{-1}(y,z) = (\sqrt{1-y^2-z^2},\ z).$$

Suppose that we have on M two atlases $\{(U_\alpha,\phi_\alpha)\}$ and $\{(V_\beta,\psi_\beta)\}$. Then by convention M is the same analytic manifold with respect to both atlases if for all α,β the mapping

$$\phi_\alpha \circ \psi_\beta^{-1} \colon \psi_\beta(V_\beta \cap U_\alpha) \to \phi_\alpha(V_\beta \cap U_\alpha)$$

and its inverse are both analytic, i.e., if the union of both atlases is again an atlas for M. In fact, we might extend the atlas $\{(U_\alpha,\phi_\alpha)\}$ in

Definition 1.6 to an atlas which is maximal under the conditions (i) and (ii). Such an atlas is called a *complete atlas*.

It is now clear that in Proposition 1.5 the local charts (U_{g_0},ϕ_{g_0}) $(g_0 \in G)$ make G into an analytic manifold. Parts (ii) and (iii) of this proposition make G into a Lie group. In order to give the formal definition of a Lie group we first have to define the notion of an analytic mapping from one analytic manifold to another.

DEFINITION 1.7. Let M and N be analytic manifolds. A continuous mapping $F: M \to N$ is called *analytic* if for each $p \in M$ there are local charts (U_α,ϕ_α) on M and (V_β,ψ_β) on N such that $p \in U_\alpha$, $F(p) \in V_\beta$ and the mapping

$$\psi_\beta \circ F \circ \phi_\alpha^{-1}: \phi_\alpha(U_\alpha) \to \psi_\beta(V_\beta)$$

is analytic in some neighbourhood of $\phi_\alpha(p)$.

It is very important to note that this definition does not depend on the particular choice of the local charts. Indeed, if $(U_{\alpha'},\phi_{\alpha'})$, $(V_{\beta'},\psi_{\beta'})$ is another pair of local charts such that $p \in U_{\alpha'}$, $F(p) \in V_{\beta'}$ then the mapping

$$\psi_{\beta'} \circ F \circ \phi_{\alpha'}^{-1} = (\psi_{\beta'} \circ \psi_\beta^{-1}) \circ (\psi_\beta \circ F \circ \phi_\alpha^{-1}) \circ (\phi_\alpha \circ \phi_{\alpha'}^{-1})$$

is analytic around $\phi_\alpha'(p)$ if $\psi_\beta \circ F \circ \phi_\alpha^{-1}$ is analytic around $\phi_\alpha(p)$.

If in Definition 1.7 we take $N = \mathbb{R}$, which is trivially an analytic manifold, then we obtain the definition of an analytic function on M.

If M and N are analytic manifolds with local charts (U_α,ϕ_α) and (V_β,ψ_β), respectively, then $M\times N$ becomes an analytic manifold with respect to the local charts $(W_{\alpha,\beta},\chi_{\alpha,\beta})$, where $W_{\alpha,\beta} = U_\alpha \times V_\beta$ and $\chi_{\alpha,\beta}(p,q) = (\phi_\alpha(p), \psi_\beta(q))$ $(p \in U_\alpha,\ q \in V_\beta)$. Now we are ready to define a Lie group:

DEFINITION 1.8. A *Lie group* is a group G which is also an analytic manifold such that the mapping $(g,h) \to gh$ is analytic from $G\times G$ to G and the mapping $g \to g^{-1}$ is analytic from G to G.

In fact, the last condition is redundant since it follows from an application of the implicit function theorem to the equation $gh = e$, which has solution $h = g^{-1}$. Observe that the conditions (ii) and (iii) of Proposition 1.5 make G into a Lie group.

1.5. Some topological properties of Lie groups

Since in a Lie group the group operations are analytic, they are certainly continuous. This makes any Lie group into a topological group.

DEFINITION 1.9. A *topological group* is a group G which is also a topological Hausdorff space such that the mapping $(x,y) \to xy$ is continuous from $G\times G$ to G and the mapping $x \to x^{-1}$ is continuous from G to G.

It is very satifactory to know that if a topological group G has an analytic structure which is compatible with the topology of G and which makes G into a Lie group, then this analytic structure is unique (cf. CHEVALLEY [2, Ch.4, §13, Theorem 3]).

We conclude section 1 with the discussion of some important aspects of Lie groups which only depend on the topological and group structure of G and hence can be formulated in terms of topological groups. In section 2 we will pass to the relationship between Lie groups and Lie algebras, where the analytic structure of a Lie group will be fully exploited.

A topological Hausdorff space is called *locally Euclidean* of dimension n if each point of the space has an open neighbourhood which is homeomorphic to some open subset of $\mathbb{R}^n$. A topological group G is already locally Euclidean if the identity element in G has such an open neighbourhood. Clearly any Lie group G is locally Euclidean. Just pick a local chart (U_α, ϕ_α) with $p \in U_\alpha$. The property of being locally Euclidean implies local compactness: a topological Hausdorff space is called *locally compact* if each point p of the space has a compact neighbourhood, i.e., p is contained in an open subset of the topological space for which the closure is compact. In particular, any Lie group is locally compact. Many results in the general representation theory of Lie groups have their natural setting in the context of locally compact topological groups. The restriction of local compactness is essential, since then the existence of a Haar measure is assured (cf. Ch. V).

Another important property which a topological space can have is *connectedness*. For locally Euclidean spaces this property coincides with arcwise connectedness. A topological space M is called *arcwise connected* if any pair of points $x_0, x_1 \in M$ can be connected with each other by means of a continuous curve, i.e., there exists a continuous mapping $t \to x(t)$ from the unit interval [0,1] into M such that $x(0) = x_0$ and $x(1) = x_1$. For a locally Euclidean space M the *component* (or connected component) of $x_0 \in M$ coincides with the *arcwise connected component* of x_0. This last concept is defined as the set of all $x \in M$ which can be connected with x_0 by means of a continuous curve. Components are always closed subsets. On a locally Euclidean space they are also open. The component of the neutral element e in a topological group G has a particularly nice form:

PROPOSITION 1.10. *Let* G *be a topological group. Then the component* G_0 *of* e *is a closed normal subgroup of* G. *If* G *is a Lie group then* G_0 *is also open in* G.

PROOF (in the case of a Lie group G).
Let $x_1, y_1 \in G_0$ and let the curves $t \to x(t)$ and $t \to y(t)$ connect e with x_1 and y_1, respectively. Let $g \in G$. Then the curves $t \to x(t)y(t)$, $t \to x(t)^{-1}$ and $t \to gx(t)g^{-1}$ connect e with x_1y_1, x_1^{-1} and gx_1g^{-1}, respectively. □

Any nonempty open subset N of an analytic manifold M with local charts (U_α, ϕ_α) becomes an analytic manifold of the same dimension as M with respect to the local charts $(U_\alpha \cap N, \phi_\alpha)$. Hence the component G_0 of e in a Lie group G is also a Lie group and G is a disjoint union of its components, which all have the form gG_0 $(g \in G)$ and which are both open and closed subsets of G. *It is usual to include in the definition of a Lie group* G *that it has only countably many components*. This is equivalent to the condition that G satisfies the second axiom of countability (see below).

Let us inspect connectedness for some special Lie groups. We already introduced the unitary group U(n). Let the *special unitary group* SU(n) consist of all $T \in U(n)$ with $\det T = 1$. The *orthogonal group* O(n) consists of all real $n \times n$ matrices T such that $T^t T = I$, where T^t is the *transpose* of T: $(T^t)_{ij} := T_{ji}$. The *special orthogonal group* SO(n) consists of all $T \in O(n)$ with $\det T = 1$. All these groups are subgroups of $GL(n,\mathbb{C})$. By the methods developed in subsections 1.2 and 1.3 it can easily be shown that they are Lie groups. It is known that SO(n), SU(n) and U(n) are connected (cf. CHEVALLEY [2, Ch. 2, § 5, Prop. 3]). This is particularly easy to see for the circle group $\mathbb{T} = SO(2) = U(1)$ and for SU(2), since this group is precisely the set of all matrices

$$\begin{pmatrix} a_0 - ia_3 & -ia_1 - a_2 \\ -ia_1 + a_2 & a_0 + ia_3 \end{pmatrix}, \quad a_0, a_1, a_2, a_3 \in \mathbb{R} \text{ and } a_0^2 + a_1^2 + a_2^2 + a_3^2 = 1.$$

Hence SU(2) is homeomorphic with the unit sphere S^3 in $\mathbb{R}^4$ and thus connected.

The continuous image of a connected topological space is again connected. Since det is a continuous mapping from O(n) onto the disconnected set $\{-1,1\}$, the group O(n) cannot be connected. It consists of two components: the normal subgroup SO(n) and the set of all $T \in O(n)$ with $\det T = -1$.

The noncompact groups $GL(n,\mathbb{C})$ and $SL(n,\mathbb{C})$ are also connected. This can be shown by the use of the *polar decomposition* which was given in § III.3 for the case of $SL(2,\mathbb{C})$: any $T \in SL(2,\mathbb{C})$ can be written as $T = UH$ with $U \in SU(2)$ and H a positive definite hermitian 2×2 matrix of determinant 1. After bringing H in diagonal form we obtain

$$T = U_1\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}U_2, \quad U_1,U_2 \in SU(2), \quad a > 0.$$

Since SU(2) is connected, it is now easy to connect any $T \in SL(2,\mathbb{C})$ with I by means of a continuous curve.

It was shown in § III.3 that the proper Lorentz group $L_+^\uparrow$ is the image of a continuous two-to-one homomorphism from $SL(2,\mathbb{C})$. Hence $L_+^\uparrow$ must be connected. This can also be seen from the fact that any proper Lorentz transformation can be written as the product of a rotation and a special Lorentz transformation (cf. § III.2).

The following Proposition 1.11 is quite useful. A subset V of a group G is called *symmetric* if $V^{-1} = V$. If V is a neighbourhood of e in a topological group G then $V \cap V^{-1}$ is a symmetric neighbourhood of e.

PROPOSITION 1.11 (cf. CHEVALLEY [2, Chap. 2, §4, Theor. 1]). *Let* V *be a symmetric neighbourhood of* e *in a connected topological group* G. *Then each* $g \in G$ *can be written as a product of finitely many elements of* V.

An application of this proposition is the result that any connected Lie group satisfies the second axiom of countability. By definition a topological space M satisfies the *second axiom of countability* if there is a countable collection $\{V_j\}$ of open subsets of M such that each open subset of M is a union of V_j's. This property implies separability: a topological space M is called *separable* if there exists a countable subset of M which is dense in M. It is well known that Euclidean spaces and their subsets have these properties. Now let G be a connected Lie group and apply Proposition 1.11 with V being homeomorphic with an open subset of some $\mathbb{R}^m$. Then it is not very difficult to prove that G satisfies the second axiom of countability. In the general representation theory of locally compact groups many technical complications are avoided by requiring that the groups satisfy the second axiom of countability.

A final topological concept of importance for us is the notion of simple connectedness. Consider an arcwise connected topological space M (for instance a connected Lie group). Then M is called *simply connected* if each closed continuous curve in M can be continuously deformed to one point. More formally stated this means that for each continuous map x: [0,1] → M such that x(0) = x(1) there are a point x_0 and a continuous map X: [0,1] × [0,1] → M such that

$$X(s,0) = x(s), \quad 0 \le s \le 1; \quad X(0,t) = X(1,t), \quad 0 \le t \le 1,$$

and

$$X(s,1) = x_0, \qquad 0 \le s \le 1.$$

All spheres S^n (n = 2,3,...) are simply connected, hence SU(2) is a simply connected Lie group, since it is homeomorphic with S^3. By the use of the two-to-one homomorphism from SU(2) onto SO(3) it is easily shown that SO(3) is not simply connected. A refinement of the polar decomposition for SL(2,ℂ) shows that the mapping $(U,A) \to U\,e^A$ is a homeomorphism from SU(2) × {all 2×2 hermitian matrices of trace zero} onto SL(2,ℂ). Hence SL(2,ℂ) is simply connected and the two-to-one homomorphism from SL(2,ℂ) onto $L_+^\uparrow$ shows that $L_+^\uparrow$ cannot be simply connected. An important application of simple connectedness will be given in the subsections 2.6 and 2.7.

2. THE RELATIONSHIP BETWEEN LIE GROUPS AND LIE ALGEBRAS

2.1. Lie algebras

A real (or complex) *Lie algebra* is a real (or complex) linear vector space L on which a bilinear mapping (X,Y) → [X,Y] from L×L into L is given such that

$$[X,Y] = -[Y,X] \tag{2.1}$$

and

(2.2) $[X,[Y,Z]] + [Y,[Z,X]] + [Z,[X,Y]] = 0.$

We call $[X,Y]$ the *commutator product* of X and Y, this is antisymmetric by (2.1). The identity (2.2) is called the *Jacobi identity*. Clearly, any complex Lie algebra is also a real Lie algebra.

The prototype of a (complex) Lie algebra is $M_n(\mathbb{C})$, where the commutator product is defined by

(2.3) $[X,Y] = XY - YX.$

A real (or complex) *subalgebra* of a real (or complex) Lie algebra L is a real (or complex) linear subspace M of L such that $[X,Y] \in M$ whenever $X,Y \in M$. Then M becomes a Lie algebra itself with respect to the commutator product of L. A *linear Lie algebra* is a real subalgebra of $M_n(\mathbb{C})$ for some n. The linear subspaces $s\ell(n,\mathbb{C})$ and $u(n)$ of $M_n(\mathbb{C})$ introduced in subsection 1.1 are examples of linear Lie algebras: Let $X,Y \in M_n(\mathbb{C})$. If $\mathrm{tr}X = \mathrm{tr}Y = 0$ then $\mathrm{tr}(XY-YX) = 0$. If $X^* = -X$, $Y^* = -Y$ then $(XY-YX)^* = Y^*X^* - X^*Y^* = YX - XY = -(XY-YX)$. By *Ado's theorem* (cf. VARADARAJAN [10, Theor.3. 17.7]) any finite-dimensional real Lie-algebra is isomorphic with some linear Lie algebra.

Lie algebras are an interesting object of study in their own right. However, they become of cardinal importance for us because of their relationship with Lie groups. In order to understand this relationship we have to introduce the concepts of tangent vectors and vector fields for an analytic manifold.

2.2. Tangent vectors and vector fields

Let M be an analytic manifold. In the case that M is an open subset of $\mathbb{R}^m$ the definition of a tangent vector to M at a point $p \in M$ is easily given: it is some vector a in $\mathbb{R}^m$ attached to the point p. In the case of general M we may pick a local chart (U,ϕ) on M such that $p \in U$ and we may define a tangent vector to M at p as some vector a in $\mathbb{R}^m$ attached to the point $\phi(p) \in \phi(U) \subset \mathbb{R}^m$. However, the vector a will depend on the choice of the local chart and we have to specify this dependence.

We will give three different but equivalent methods to define a tangent vector:

(a) By means of some equivalence class of differentiable curves starting at $p \in M$.

(b) A formal method where, for each local chart around $p \in M$, a vector $a \in \mathbb{R}^m$ is attached to p and where it is described how the vector a transforms under a change of the local chart.

(c) A tangent vector is considered as a linear functional acting on the space of C^∞-functions which are locally defined around $p \in M$.

It is important to know all these three approaches. Which of the three will be used in practice depends on the situation and on everybody's taste. Below we will discuss each of the three approaches.

Approach (a) (equivalence classes of curves). Consider the sphere S^2, which is an elementary example of an analytic manifold. In classical geometry the tangent plane to S^2 at some point $p \in S^2$ is the plane lying in $\mathbb{R}^3$ which touches to S^2 at p. A tangent vector to S^2 at p is a vector lying in the tangent plane to S^2 and having its origin in p. The advantage of this definition is that it has a geometrically clear meaning. The disadvantage is that it uses vectors lying outside the manifold. For general analytic manifolds we would like to have a more intrinsic definition of tangent vectors such that the geometric interpretation remains clear. In the example of S^2 note that for each tangent vector $a = (a_1,a_2,a_3)$ to S^2 at p there is a family of differentiable curves $t \to x(t)$ lying in S^2, passing through p ($x(0) = p$), and touching to the tangent vector a at p in the sense that $x'(0) = a$. Two differentiable curves $t \to x(t)$ and $t \to y(t)$ in S^2 through p are in the same family if and only if they touch each other in the sense that $x'(0) = y'(0)$. This suggests the definition formulated below.

Let M be an analytic manifold and $p \in M$. By a *curve in* M *starting at* p we will mean a continuous mapping $t \to x(t)$ from some real interval $[0,\delta]$, $\delta > 0$, into M such that $x(0) = p$ and the right derivative

$$\frac{d}{dt}\phi(x(t))\Big|_{t=0} := \lim_{t\downarrow 0} \frac{\phi(x(t)) - \phi(x(0))}{t}$$

exists for some local chart (and hence for all local charts) (U,ϕ) with $p \in U$. We will call two such curves $t \to x(t)$ and $t \to y(t)$ *equivalent* to each other if for some (and hence for all) local chart(s) (U,ϕ) with $p \in U$ we have

$$\frac{d}{dt}\,\phi(x(t))\Big|_{t=0} = \frac{d}{dt}\,\phi(y(t))\Big|_{t=0}.$$

DEFINITION 2.1(a). Let M be an analytic manifold and $p \in M$. A *tangent vector* to M at p is an equivalence class of curves in M starting at p.

Let us choose a local chart (U,ϕ) on M with $p \in U$, let $t \to x(t)$ be a curve in M starting at p and let us write by abuse of notation $x(t) = (x_1(t),\ldots,x_m(t)) \in \phi(U) \subset \mathbb{R}^m$ instead of $\phi(x(t))$, and $p = (p_1,\ldots,p_m) \in \phi(U) \subset \mathbb{R}^m$ instead of $\phi(p)$. Then

$$\text{(2.4)} \qquad x_i(t) = p_i + t\,x_i'(0) + o(t) \text{ for } t \downarrow 0,\ i = 1,\ldots,m,$$

and two curves $t \to x(t)$ and $t \to y(t)$ in M starting at p are equivalent if and only if $x_i'(0) = y_i'(0)$, $i = 1,\ldots,m$. Note that $x'(0) = (x_1'(0),\ldots,x_m'(0)) \in \mathbb{R}^m$ and that any vector $a \in \mathbb{R}^m$ can be obtained in this way, for instance by means of the curve $t \to p+ta$. We conclude that the tangent vectors to M at p are in one-to-one correspondence with the vectors in $\mathbb{R}^m$, where we associate with a representative $t \to x(t)$ of a tangent vector the vector $x'(0) \in \mathbb{R}^m$. However, this identification depends on the particular local chart we have chosen.

Approach (b) (the formal method). Let $t \to x(t)$ be a curve in M starting at p which represents some tangent vector to M at p according to Definition 2.1(a). Let $(x_1,\ldots,x_m)$ and $(\tilde{x}_1,\ldots,\tilde{x}_m)$ be two sets of local coordinates obtained from two local charts on M around p. Then $a := x'(0)$ and $\tilde{a} := \tilde{x}'(0)$ are the vectors in $\mathbb{R}^m$ corresponding to the tangent vector in the two coordinate systems, respectively. We have

$$\tilde{a}_i = \tilde{x}_i'(0) = \sum_{j=1}^{m} x_j'(0)\,\frac{\partial \tilde{x}_i}{\partial x_j}\Big|_{x=p} = \sum_{j=1}^{m} a_j\,\frac{\partial \tilde{x}_i}{\partial x_j}\Big|_{x=p}.$$

This means that the vector $a = (a_1,\ldots,a_m)$ transforms in a *contravariant* way under an analytic change of coordinates. (According to the notation conventions in tensor analysis we should have written a^i instead of a_i.) Thus we can give the following alternative to Definition 2.1(a):

DEFINITION 2.1(b). A *tangent vector* to M at p is an attachment of a vector $a \in \mathbb{R}^m$ to p for each local coordinate system $(x_1,\ldots,x_m)$ on M around p such that under an analytic coordinate transformation $(x_1,\dots,x_m) \to (\tilde{x}_1,\dots,\tilde{x}_m)$ the vector a transforms to $\tilde{a}$ by the rule

$$(2.5)\qquad \tilde{a}_i = \sum_{j=1}^{m} a_j \frac{\partial \tilde{x}_i}{\partial x_j}\Big|_{x=p}, \qquad i = 1,\dots,m.$$

Similarly, one might consider vectors $b \in \mathbb{R}^m$ transforming in a covariant way under an analytic change of coordinates, i.e.

$$\tilde{b}_i = \sum_{j=1}^{m} b_j \frac{\partial x_j}{\partial \tilde{x}_i}\Big|_{\tilde{x}=\tilde{p}}, \qquad i = 1,\dots,m.$$

Such vectors correspond to so-called differential 1-forms on M, but these objects will not be discussed in the present chapter.

Approach (c) (linear functionals). On a still higher level of abstraction than in Approach (a) we can describe tangent vectors as linear functionals on a certain linear function space $C^\infty(p)$. This function space consists of all real-valued functions f defined on some open neighbourhood of p in M such that for some (and hence for all) local chart(s) (U,ϕ) around p the function $f\circ\phi^{-1}$ is a C^∞-function on some open neighbourhood of $\phi(p)$ in $\mathbb{R}^m$. Furthermore, two functions in $C^\infty(p)$ are identified with each other if they coincide on some open neighbourhood of p in M. Thus $C^\infty(p)$ becomes a linear space.

Let $t \to x(t)$ be a curve in M starting at p, which represents a tangent vector according to Definition 2.1(a). Let

$$(2.6)\qquad Af := \frac{d}{dt} f(x(t))\Big|_{t=0}, \qquad f \in C^\infty(p).$$

Then A is a linear functional on $C^\infty(p)$ which satisfies

$$(2.7)\qquad A(fg) = (Af)g(p) + (Ag)f(p), \qquad f,g \in C^\infty(p).$$

Let us again choose a local chart on M around p and let $x'(0) = a \in \mathbb{R}^m$ in terms of these coordinates. Then (2.6) gives

$$(2.8)\qquad Af = \sum_{i=1}^{m} a_i \frac{\partial f(x)}{\partial x_i}\Big|_{x=p}.$$

Hence the linear functional A defined by (2.6) only depends on the equivalence class to which the curve $t \to x(t)$ belongs and the linear functionals A of the form (2.8) (for a fixed local chart) are in one-to-one correspondence with the tangent vectors to M at p. It is rather easy to show that for a given local chart around p any linear functional A on $C^\infty(p)$ satisfying (2.7) takes the form (2.8) (cf. HELGASON [4, Ch. 1, § 2.1]). So we have yet another alternative to Definition 2.1(a):

DEFINITION 2.1(c). A *tangent vector* to M at p is a linear functional on $C^\infty(p)$ satisfying (2.7).

The set of all tangent vectors to M at p is a linear space (cf. Def. 2.1(c)) which has the same dimension as M (cf. Def. 2.1(b)). It is called the *tangent space* to M at p and it is denoted by $T_p(M)$.

Next we consider the concept of a vector field:

DEFINITION 2.2. A *vector field* X on an analytic manifold M is an attachment to each $p \in M$ of a tangent vector to M at p.

Hence a vector field X is specified if for each $p \in M$ a tangent vector $X_p \in T_p(M)$ is given.

Let $p \in M$, $f \in C^\infty(p)$ and let X be a vector field on M. According to approach (c), for each q in some open neighbourhood of p the real number $X_q f$ is well-defined. Hence $q \to X_q f$ is a function on some open neighbourhood of p which we will denote by Xf:

$$(2.9) \qquad (Xf)(q) := X_q f.$$

For some local chart around p combination of (2.8) and (2.9) gives

$$(2.10) \qquad (Xf)(q) = \sum_{i=1}^{m} a_i(q) \left. \frac{\partial f(x)}{\partial x_i} \right|_{x=q}.$$

Hence, a vector field X takes locally, with respect to some local chart, the form of a linear differential operator with variable coefficients $a_i(q)$. In approach (b) the mapping $q \to (a_1(q),\ldots,a_m(q))$ describes the vector field with respect to the given coordinate system.

A function f defined on some open neighbourhood of $p \in M$ is called *analytic* at p if for each (and hence for all) local chart(s) (U,ϕ) around p the function $f\circ\phi^{-1}$ is analytic at $\phi(p) \in \mathbb{R}^m$. A vector field X is called *analytic* if for each $p \in M$ and for each function f which is analytic at p, the function Xf is analytic at p. It follows easily that this condition is equivalent to that fact that for some (and hence for all) local chart(s) (U,ϕ) the functions $q \to a_i(q)$ in (2.10) are analytic on $\phi(U) \subset \mathbb{R}^m$.

Let X, Y be analytic vector fields on M. For $f \in C^\infty(p)$ we define

$$[X,Y](f) := X(Yf) - Y(Xf). \tag{2.11}$$

Then $f \to [X,Y](f)$ is a well-defined linear mapping from $C^\infty(p)$ into itself, which sends analytic functions to analytic functions. Let us choose a local chart, let (2.10) hold for X and let a similar identity hold for Y with a_i replaced by b_i. Then an easy calculation shows that

$$\begin{aligned}([X,Y](f))(q) &= \\ &= \sum_{i=1}^{m} \left(\sum_{j=1}^{m} a_j(q) \left.\frac{\partial b_i(x)}{\partial x_j}\right|_{x=q} - \sum_{j=1}^{m} b_j(q) \left.\frac{\partial a_i(x)}{\partial x_j}\right|_{x=q} \right) \cdot \\ &\quad \cdot \left.\frac{\partial f(x)}{\partial x_i}\right|_{x=q} .\end{aligned} \tag{2.12}$$

Hence $[X,Y]$ is again an analytic vector field on M. It follows easily that the linear space of all analytic vector fields on M with the commutator product defined by (2.11) becomes a Lie algebra. However, this Lie algebra is usually infinite-dimensional.

An analytic mapping F from an analytic manifold M to an analytic manifold N induces linear mappings from tangent spaces to M into related tangent spaces to N. This is most easily seen by following approach (a). Let the curve $t \to x(t)$ in M starting at p represent some tangent vector to M at p. Then the mapping $t \to F(x(t))$ is continuous from $[0,\delta]$ into N such that $F(x(0)) = F(p)$ and the first right derivative of this mapping at $t = 0$ exists. Hence it represents some tangent vector to N at $F(p)$. Now choose local charts (U,ϕ) on M around p and (V,ψ) on N around $F(p)$. By abuse of notation we may consider F as a mapping from $\phi(U) \subset \mathbb{R}^m$ to $\psi(V) \subset \mathbb{R}^n$. Let

$$x_i(t) = p_i + t a_i + o(t) \quad \text{for } t \downarrow 0, \quad i = 1,\dots,m,$$

where $a \in \mathbb{R}^m$. Then

$$F_i(x(t)) = F_i(p) + t \sum_{j=1}^{m} \frac{\partial F_i(x)}{\partial x_j}\bigg|_{x=p} a_j + o(t)$$

for $t \downarrow 0$, $i = 1,\ldots,n$. Hence the tangent vector represented by the curve $t \to F(x(t))$ does not depend on the choice of the representative $t \to x(t)$ for the original tangent vector to M at p. Furthermore, if $b_i := \frac{d}{dt} F_i(x(t))\Big|_{t=0}$, $i = 1,\ldots,n$, then

$$b_i = \sum_{j=1}^{m} \frac{\partial F_i(x)}{\partial x_j}\bigg|_{x=p} a_j, \quad i = 1,\ldots,n, \tag{2.13}$$

i.e., the mapping $a \to b$ is linear and its matrix is the Jacobian matrix of the mapping F at p. Summarizing we have defined:

<u>DEFINITION 2.3</u>. Let M and N be analytic manifolds and let $F: M \to N$ be an analytic mapping. For each $p \in M$ the *differential* dF_p of the mapping F is a linear mapping from the tangent space $\mathcal{T}_p(M)$ into the tangent space $\mathcal{T}_{F(p)}(N)$ which can be defined in either of the two following ways (corresponding to the approaches (a) and (b) for tangent vectors):

(a) If the curve $t \to x(t)$ in M starting at p is a representative of the tangent vector $A \in \mathcal{T}_p(M)$ then the curve $t \to F(x(t))$ is a representative of the tangent vector $dF_p(A) \in \mathcal{T}_{F(p)}(N)$.

(b) If the vector $a \in \mathbb{R}^m$ is a tangent vector to M at p with respect to some local chart on M around p then the vector $b \in \mathbb{R}^n$ defined by (2.13) is the image of a under dF_p with respect to some local chart on N around F(p).

The reader may try to define dF_p in the spirit of approach (c).

Now suppose that F is a one-to-one analytic mapping from an analytic manifold M into itself such that the inverse mapping $F^{-1}: M \to M$ is also analytic, i.e., F is an *analytic diffeomorphism* of M. Then for all $p \in M$ the m×m Jacobian matrix $\left(\frac{\partial F_i(x)}{\partial x_j}\Big|_{x=p}\right)$ is non-singular, hence dF_p is a one-to-one linear mapping from $\mathcal{T}_p(M)$ onto $\mathcal{T}_{F(p)}(m)$. Now let X be an analytic vector field on M. Then for each $p \in M$ we have $dF_p(X_p) \in \mathcal{T}_{F(p)}(M)$. Hence all tangent vectors $dF_p(X_p)$, $p \in M$, together form a new vector field, denoted by dF(X). Then

$$(dF(X))_{F(p)} = dF_p(X_p),$$

i.e.

$$(2.14)\qquad (dF(X))_p = dF_{F^{-1}(p)}(X_{F^{-1}(p)}).$$

It can be easily proved that the vector field dF(X) is again analytic and that for two analytic vector fields X and Y we have

$$(2.15)\qquad dF([X,Y]) = [dF(X),dF(Y)].$$

2.3. The Lie algebra associated with a Lie group

Finally we are ready to define the Lie algebra associated with a given Lie group G. This may be done by following any of the three approaches (a), (b) or (c) of subsection 2.2. For reasons of elegance we will start here with approach (c), then we give an equivalent definition in the spirit of (b) and we conclude with approach (a). Shortly summarized the three definitions of the Lie algebra g of the Lie group G are as follows:

(c) g is the collection of all left invariant vector fields on G. This has a natural Lie algebra structure and it is in one-to-one linear correspondence with the tangent space $T_e(G)$.

(b) Let $(x_1,\dots,x_m)$ be local coordinates on G around e such that e has coordinates $(0,\dots,0)$. Consider the power series expansion of $(xy)_i$ $(i = 1,\dots,m)$ in terms of $x_1,\dots,x_m,y_1,\dots,y_m$. Then g is the linear space $\mathbb{R}^m$ with the structure constants of the Lie algebra obtained from the coefficients of the second degree terms in the above power series expansion. The transformation of these structure constants under an analytic change of coordinates on G around e can be explicitly given.

(a) The tangent space $T_e(G)$ is made into a Lie algebra g by defining the commutator product $C = [A,B]$ of $A,B \in T_e(G)$ in terms of a curve $t \to z(t)$ representing C which is constructed from curves $t \to x(t)$ and $t \to y(t)$ representing A and B, respectively.

In later subsections we will mostly use approach (a).

Approach (c) (left invariant vector fields). Let G be a Lie group. Then G is also an analytic manifold. For each $x \in G$ the mapping L_x defined by

$$(2.16)\qquad L_x g := x\,g, \qquad g \in G,$$

is a one-to-one analytic transformation from G onto G for which the inverse mapping $(L_x)^{-1} = L_{x^{-1}}$ is also analytic. Hence, if X is an analytic vector field on G and $x \in G$ then $(dL_x)(X)$ (as defined by (2.14)) is again an analytic vector field on G. Let us call an analytic vector field X on G *left invariant* if $(dL_x)(X) = X$ for each $x \in G$, in other words if

$$(2.17) \qquad (dL_x)(X_g) = X_{xg} \qquad \text{for all } x,g \in G.$$

By (2.15) the commutator product of two left invariant analytic vector fields is again a left invariant analytic vector field. Hence the left invariant analytic vector fields on G form a subalgebra $\mathfrak{g}$ of the Lie algebra of all analytic vector fields on G.

DEFINITION 2.4(c). *The Lie algebra $\mathfrak{g}$ associated with a Lie group* G is the linear space of all left invariant analytic vector fields with the commutator product defined by (2.11).

The above definition of $\mathfrak{g}$ is rather abstract and it is not yet clear at this stage whether $\mathfrak{g}$ contains any vector field different from zero. However, we prefer this definition to the two other variants below, since it is immediately clear from this definition that $\mathfrak{g}$ satisfies the axioms of a Lie algebra, in particular the Jacobi identity (2.2). In the other two approaches this requires some calculations and it is kind of a surprise that the Jacobi identity comes out.

It is important to note that the mapping $X \to X_e$ establishes a one-to-one linear correspondence between the left invariant analytic vector fields X on G and the tangent vectors to G at e. In fact, if X is a left invariant analytic vector field on G then (2.17) yields that

$$(2.18) \qquad X_g = (dL_g)(X_e) \qquad \text{for all } g \in G.$$

Hence X_e determines X uniquely. Conversely, if $A \in T_e(G)$ then it is easily seen that

$$X_g := (dL_g)(A), \qquad g \in G,$$

defines a left invariant analytic vector field X such that $X_e = A$.

We conclude that the Lie algebra $\mathcal{g}$ of G has the same dimension as G. Furthermore, by the identification of $\mathcal{g}$ with $\mathcal{T}_e(G)$, the tangent space $\mathcal{T}_e(G)$ gets a Lie algebra structure which cannot be understood intrinsically from $\mathcal{T}_e(G)$ but for the definition of which we have to go to neighbouring tangent spaces.

Approach (b) (structure constants obtained from the power series expansion of the product). Let us consider how the just-mentioned Lie algebra structure of $\mathcal{T}_e(G)$ looks in terms of a local chart (U,ϕ) on G with $e \in U$. If $x \in U$ then let $\phi(x) \in \phi(U) \subset \mathbb{R}^m$ have coordinates $(x_1,\ldots,x_m)$. Let X be a left invariant analytic vector field on G. If $x \in U$ then write

$$X_x = (a_1(x),\ldots,a_m(x))$$

in terms of the local chart. (Here we have used definition 2.1(b) for the tangent vector $X_x \in \mathcal{T}_x(G)$.) Combination of (2.18) and (2.13) gives:

$$a_i(x) = \sum_{k=1}^{m} \left.\frac{\partial (xy)_i}{\partial y_k}\right|_{y=e} a_k(e), \qquad i = 1,\ldots,m.$$

If Y is a left invariant analytic vector field with $Y_x = (b_1(x),\ldots,b_m(x))$ $(x \in U)$ then $b_i(x)$ can be similarly expressed in terms of $b_1(e),\ldots,b_m(e)$. Let $Z := [X,Y]$ and write $Z_x = (c_1(x),\ldots,c_m(x))$ $(x \in U)$. Then it follows from (2.12) that

$$c_i = \sum_{j,k=1}^{m} a_j b_k \left\{ \left.\frac{\partial^2 (xy)_i}{\partial x_j \partial y_k}\right|_{\substack{x=e\\y=e}} - \left.\frac{\partial^2 (xy)_i}{\partial x_k \partial y_j}\right|_{\substack{x=e\\y=e}} \right\} .$$

Suppose that the local chart is chosen such that $\phi(e) = (0,0,\ldots,0)$. The function $(xy)_i$ is analytic in $x_1,\ldots,x_m,y_1,\ldots,y_m$ at $(0,\ldots,0,0,\ldots,0)$. One quickly sees that up to the second degree terms its power series expansion must have the form

$$(2.19) \qquad (xy)_i = x_i + y_i + \sum_{j,k=1}^{m} \gamma_{i;j,k}\, x_j y_k + \cdots .$$

We conclude that the coefficients c_i in (2.20) are equal to

$$c_i = \sum_{j,k=1}^{m} (\gamma_{i;j,k} - \gamma_{i;k,j}) a_j b_k.$$

DEFINITION 2.4(b). *The Lie algebra $\mathfrak{g}$ associated with a Lie group* G (of dimension m) is the linear space $\mathbb{R}^m$ on which a commutator product

$$[a,b]_i := \sum_{j,k=1}^{m} g_{i;j,k}\, a_j b_k \qquad (a,b \in \mathbb{R}^m) \tag{2.20}$$

is defined depending on the choice of local coordinates $(x_1, x_2, \ldots, x_m)$ around $e(=(0,\ldots,0))$ such that the *structure constants* $g_{i;j,k}$ equal

$$g_{i;j,k} := \gamma_{i;j,k} - \gamma_{i;k,j} \tag{2.21}$$

with the $\gamma_{i;j,k}$'s given by (2.19).

Since the equivalent Definition 2.4(a) of $\mathfrak{g}$ is coordinate free, the passage from $g_{i;j,k}$ to $\tilde{g}_{i;j,k}$ under an analytic change of local coordinates on G from $(x_1,\ldots,x_m)$ to $(\tilde{x}_1,\ldots,\tilde{x}_m)$ can be calculated from (2.20) and the transformation rule (2.5) for tangent vectors. The result is

$$\tilde{g}_{p;q,r} = \sum_{i,j,k=1}^{m} g_{i;j,k} \frac{\partial \tilde{x}_p}{\partial x_i} \frac{\partial x_j}{\partial \tilde{x}_q} \frac{\partial x_k}{\partial \tilde{x}_r}\;, \tag{2.22}$$

where all partial derivatives are taken at $0 \in \mathbb{R}^m$ corresponding with $e \in G$. It follows that $g_{i;j,k}$ transforms as a tensor, contravariant in its first index and covariant in the other two indices.

Approach (a) (the commutator product of two tangent vectors at e defined in terms of representing curves). Let $(x_1,\ldots,x_m)$ be local coordinates around $e \in G$ such that e has coordinates $(0,\ldots,0)$. It follows easily from (2.19) and (2.21) that for x, y in some neighbourhood of e we have

$$(x^{-1})_i = -x_i + \sum_{j,k=1}^{m} \gamma_{i;j,k}\, x_j x_k + \cdots\;, \tag{2.23}$$

$$(xyx^{-1})_i = y_i + \sum_{j,k=1}^{m} g_{i;j,k}\, x_j y_k + \cdots\;, \tag{2.24}$$

$$(xyx^{-1}y^{-1})_i = \sum_{j,k=1}^{m} g_{i;j,k}\, x_j y_k + \cdots\;. \tag{2.25}$$

This last expansion makes it possible to characterize the Lie algebra structure of $\mathcal{T}_e(G)$ with the use of Definition 2.1(a) for tangent vectors. Let $t \to x(t)$ and $t \to y(t)$ be curves in G starting at e which represent A and $B \in \mathcal{T}_e(G)$, respectively. Define a curve $t \to z(t)$ by

$$(2.26) \qquad z(t) := x(t^{\frac{1}{2}})y(t^{\frac{1}{2}})(x(t^{\frac{1}{2}}))^{-1}(y(t^{\frac{1}{2}}))^{-1}.$$

If $x_i(t) = ta_i + o(t)$ and $y_i(t) = tb_i + o(t)$ for $t \downarrow 0$ (in terms of the local coordinates) then it follows from (2.25) that

$$z_i(t) = t \sum_{j,k=1}^{m} g_{i;j,k}\, a_j b_k + o(t) \qquad \text{for } t \downarrow 0.$$

In view of (2.20) the following definition is now justified.

DEFINITION 2.4(a). *The Lie algebra* g *associated with a Lie group* G *is the linear space* $\mathcal{T}_e(G)$ *such that the commutator product* $[A,B]$ *of* $A,B \in \mathcal{T}_e(G)$ *has representing curve* $t \to z(t)$ *defined by* (2.26) *if* A *and* B *have representing curves* $t \to x(t)$ *and* $t \to y(t)$*, respectively.*

This concludes our description of the three approaches to the Lie algebra of G. Let us now examine the Lie algebra of the Lie group $GL(n,\mathbb{C})$. This group can be covered by one set of $2n^2$ real coordinates: the real and imaginary parts of the n^2 matrix entries. Usually we do not write these coordinates explicitly, but we preserve the matrix notation. Let $t \to T(t)$ be a curve in $GL(n,\mathbb{C})$ starting at I. Then

$$(2.27) \qquad T(t) = I + tA + o(t) \qquad \text{for } t \downarrow 0,$$

where $A \in M_n(\mathbb{C})$, and any $A \in M_n(\mathbb{C})$ can be obtained in this way. Hence the tangent vector represented by the curve $t \to T(t)$ can be identified with A in (2.27) and the tangent space $\mathcal{T}_I(GL(n,\mathbb{C}))$ can be identified with $M_n(\mathbb{C})$. Next we look for the left invariant analytic vector field X on $GL(n,\mathbb{C})$ for which X_I is represented by $t \to T(t)$. By (2.18) and Definition 2.3(a) we can represent $X_S \in \mathcal{T}_S(GL(n,\mathbb{C}))$ ($S \in GL(n,\mathbb{C})$) by the curve $t \to ST(t)$ and from (2.27) we have

$$(2.28) \qquad ST(t) = S + tSA + o(t) \qquad \text{for } t \downarrow 0.$$

Thus we have proved:

PROPOSITION 2.5. *Let* $A \in M_n(\mathbb{C})$. *Then the attachment of* $SA \in M_n(\mathbb{C})$ *to* $S \in GL(n,\mathbb{C})$ *defines a left invariant analytic vector field on* $GL(n,\mathbb{C})$ *(following approach (b)). All left invariant analytic vector fields on* $GL(n,\mathbb{C})$ *can be obtained in this way.*

If $M_n(\mathbb{C})$ is identified with $T_I(GL(n,\mathbb{C}))$ as above then a Lie algebra structure is induced on $M_n(\mathbb{C})$. In order to find the commutator product note that

$$(I+A)(I+B)(I+A)^{-1}(I+B)^{-1} = I + (AB-BA) + \ldots$$

for $A,B \in M_n(\mathbb{C})$ in some neighbourhood of 0. Hence

$$I + C(t) := (I+t^{\frac{1}{2}}A)(I+t^{\frac{1}{2}}B)(I+t^{\frac{1}{2}}A)^{-1}(I+t^{\frac{1}{2}}B)^{-1} =$$

$$= I + t(AB-BA) + o(t) \qquad \text{as } t \downarrow 0,$$

where $A,B \in M_n(\mathbb{C})$ arbitrarily. By Definition 2.4(a) the curve $t \to I + C(t)$ represents the commutator product of the tangent vectors at I represented by $t \to I + tA$ and $t \to I + tB$. Hence the induced Lie algebra structure on $M_n(\mathbb{C})$ is

$$[A,B] := AB - BA,$$

which coincides with (2.3).

2.4. The exponential mapping

The correspondence between a Lie group G and its Lie algebra $\mathfrak{g}$ can be made even closer by means of the exponential mapping. We will illustrate this concept first for the Lie group $GL(n,\mathbb{C})$ and its Lie algebra $M_n(\mathbb{C})$. Some important properties of the mapping $\exp: M_n(\mathbb{C}) \to GL(n,\mathbb{C})$ were formulated in Proposition 1.2. In order to extend the definition and properties of exp to arbitrary pairs $(\mathfrak{g},G)$ we have to find a definition which is more conceptual then the one given by means of the power series (1.1).

Clearly we know $\exp A$ for all $A \in M_n(\mathbb{C})$ if we know the mappings

$t \to \exp tA$ from $\mathbb{R}$ into $GL(n,\mathbb{C})$ for all $A \in M_n(\mathbb{C})$. These mappings can be characterized in a very conceptual way. Note that

(2.29) $\quad \exp tA = I + tA + o(t) \quad$ for $t \downarrow 0$,

(2.30) $\quad \exp(t_1 + t_2)A = \exp t_1 A \exp t_2 A$,

(2.31) $\quad \frac{d}{dt} \exp tA = \exp(tA)\cdot A$.

Now we have:

PROPOSITION 2.6. *Let* $A \in M_n(\mathbb{C})$, $T(t) = \exp tA$. *Then the mapping* $T: \mathbb{R} \to GL(n,\mathbb{C})$ *can be uniquely characterized in either of the two following ways:*

(a) T *is an analytic homomorphism from* $\mathbb{R}$ *(considered as additive group) into* $GL(n,\mathbb{C})$ *such that* $\left.\frac{dT(t)}{dt}\right|_{t=0} = A$.

(b) T *is an analytic solution of the differential equation* $\frac{dT}{dt} = TA$ *with initial value* $T(0) = I$.

Note that the condition $\left.\frac{dT(t)}{dt}\right|_{t=0} = A$ is equivalent to the statement that the curve $t \to T(t)$ is a representative of the tangent vector A to $GL(n,\mathbb{C})$ at I. Similarly, the fact that T is a solution of the differential equation $\frac{dT}{dt} = TA$ can also be formulated in the form: for each $s \in \mathbb{R}$ the curve $t \to T(t+s)$ is a representative of the tangent vector $T(s)A$ to $GL(n,\mathbb{C})$ at $T(s)$. Now we know from Proposition 2.5 that $X_T := TA$ defines a left invariant analytic vector field X on $GL(n,\mathbb{C})$ such that $X_I = A$. We can rephrase the first part of condition (b) of Proposition 2.6 in the form: T is an integral curve of the left invariant analytic vector field on G which is generated by A.

Let M be an arbitrary analytic manifold and let X be an analytic vector field on M. An analytic mapping $t \to x(t)$ from some real open interval I into M is called an *integral curve* of the vector field X if for each $s \in I$ the mapping $t \to x(s+t)$ is a representative of the tangent vector $X_{x(s)}$ to M at $x(s)$. If we have a local chart on M around $x(t_0)$, $t_0 \in I$, and if in terms of the local coordinates

$$X_x = \sum_{i=1}^{m} a_i(x) \frac{\partial}{\partial x_i}$$

then such an integral curve is a solution of the system

$$\frac{dx_i}{dt} = a_i(x_1,\ldots,x_m), \qquad i = 1,\ldots,m.$$

It is well-known from the theory of ordinary differential equations that for a given initial value $x(t_0)$ this system has a unique analytic solution on some open interval around t_0.

We can now formulate the analogue of Proposition 2.6 for a general Lie group G with Lie algebra $\mathfrak{g}$ which we will identify with $T_e(G)$.

PROPOSITION 2.7. *Let* G *be a Lie group. Let* X *be a left invariant analytic vector field on* G *and put* $X_e = A$. *Then there is a unique mapping* $t \to x_A(t)$ *from* $\mathbb{R}$ *into* G *such that the following two equivalent conditions are satisfied:*

(a) $t \to x_A(t)$ *is an analytic homomorphism from* $\mathbb{R}$ *into* G *and it is a representative of the tangent vector* A *to* G *at* e.

(b) $t \to x_A(t)$ *is an (analytic) integral curve of the left invariant analytic vector field* X *on* G *and* $x_A(0) = e$.

PROOF. See for instance CODDINGTON & LEVINSON [3, Ch.1] for the results on ordinary differential equations used below.

(i) Proof that (a) ⇒ (b): Let $t \to x(t)$ be an analytic homomorphism from $\mathbb{R}$ into G such that it is a representative of $A \in T_e(G)$. Let X be the left invariant analytic vector field on G with $X_e = A$. Then $x(t_0+t) = x(t_0)x(t)$. Hence the curve $t \to x(t_0+t)$ is a representative of the tangent vector to G at $x(t_0)$ given by

$$dL_{x(t_0)}A = X_{x(t_0)},$$

cf. Definition 2.3(a) and formulas (2.16) and (2.18). This proves that $t \to x(t)$ is an integral curve of the vector field X through e.

(ii) Proof that (b) ⇒ (a): Let $t \to x(t)$ be an analytic integral curve of the left invariant analytic vector field X on G such that $x(0) = e$. Fix $t_1 \in \mathbb{R}$. We will prove that $x(t_1+t) = x(t_1)x(t)$ by showing that both sides, considered as functions of t, are integral curves of X. In fact, the curve $\tau \to x(t_1+t+\tau)$ (t fixed) represents the tangent vector $X_{x(t_1+t)}$ and the curve $\tau \to x(t_1)x(t+\tau)$ represents the tangent vector

$$dL_{x(t_1)} X_{x(t)} = X_{x(t_1)x(t)},$$

cf. Definition 2.3(a) and formulas (2.16) and (2.17). Now it is generally true that two integral curves $t \to y(t)$ $(t \in I_1)$ and $t \to z(t)$ $(t \in I_2)$ of an analytic vector field X on an analytic manifold M for which $y(t_0) = z(t_0)$ for some $t_0 \in I_1 \cap I_2$, have the property that $y(t) = z(t)$ for all t in the intersection of the intervals I_1 and I_2. Otherwise, for some $t_1 \in I_1 \cap I_2$ we have $y(t_1) = z(t_1)$ but $y(t) \not\equiv z(t)$ for t in any neighbourhood of t_1. Choose local coordinates $(x_1,\ldots,x_m)$ around $y(t_1)$. Then, for $i = 1,\ldots,m$:

$$\frac{dy_i}{dt} = a_1(y_1,\ldots,y_m), \quad \frac{dz_i}{dt} = a_1(z_1,\ldots,z_m)$$

and $y_i(t_1) = z_i(t_1)$, where a_i is an analytic function. Hence the uniqueness property of solutions of systems of ordinary differential equations leads to a contradiction. We conclude that $x(t_1+t) = x(t_1)x(t)$ for all t.

(iii) Existence and uniqueness proof: By the above argument, if an integral curve x of the left invariant vector field X exists with $x(0) = e$ then it is unique. By choosing local coordinates around e we see that the existence of an analytic integral curve is assured on some interval $(-\delta,\delta)$ because of the existence theorem for solutions of systems of analytic ordinary differential equations. Suppose that this local solution x cannot be extended to $\mathbb{R}$ but to some maximal smaller interval (a,b). Let, for instance, $b < \infty$. Now define $y(t) := x(\frac{1}{2}b)x(t-\frac{1}{2}b)$, $a + \frac{1}{2}b < t < \frac{3}{2}b$. From the left invariance of X we conclude that y is an integral curve of X which coincides with x for $a + \frac{1}{2}b < t < b$, since $x(\frac{1}{2}b) = y(\frac{1}{2}b)$. Hence the integral curve $t \to y(t)$ gives us a continuation of x to the interval $(a,\frac{3}{2}b)$. This is a contradiction. □

It follows from (b) that the mapping $(A,t) \to x_A(t)$ is analytic from $\mathfrak{g} \times \mathbb{R}$ into G and that $x_A(t) = x_{tA}(1)$. Let us prove the first fact. Choose local coordinates around e. Then, locally around 0, $t \to x_A(t)$ is the solution of the system

$$\begin{cases} \dfrac{dx_i}{dt} = a_i(x_1,\ldots,x_m), & i = 1,\ldots,m, \\ x_i(0) = e_i, & \end{cases}$$

where $(a_1(x),\dots,a_m(x))$ denotes the tangent vector $X_x = dL_x(A)$ in terms of the local coordinates. It follows from (2.13) that

$$a_i(x) = \sum_{j=1}^{m} \left.\frac{\partial(xy)_i}{\partial y_j}\right|_{y=e} \alpha_j,$$

where $(\alpha_1,\dots,\alpha_m) := (a_1(e),\dots,a_m(e))$ denotes A in terms of the local coordinates. Thus $t \to x_A(t)$ is a solution of

$$\frac{dx_i}{dt} = \sum_{j=1}^{m} \left.\frac{\partial(xy)_i}{\partial y_j}\right|_{y=e} \alpha_j, \qquad i = 1,\dots,m. \tag{2.32}$$

Since the right hand side of (2.32) depends analytically on A and x, we conclude from the theory of systems of analytic ordinary differential equations that for each $A_0 \in \mathcal{T}_e(G)$ the functions $(x_A)_i(t)$ are analytic in A and t, where (A,t) is in some neighbourhood of $(A_0,0)$. Finally, the global analyticity follows from the fact that $x_A(t) = (x_A(n^{-1}t))^n$.

We now define the *exponential mapping*

$$\exp A := x_A(1) \qquad (A \in \mathfrak{g}).$$

Then $x_A(t) = \exp tA$ and exp is an analytic mapping from $\mathfrak{g}$ into G. Since exp has Jacobian determinant 1 for $A = 0$, we arrive at the following theorem.

THEOREM 2.8. *The mapping* exp *is analytic from* $\mathfrak{g}$ *into* G *and it maps* 0 *to* e. *It maps some open neighbourhood of* 0 *in* $\mathfrak{g}$ *one-to-one onto some open neighbourhood of* e *in* G *such that the inverse mapping is also analytic. For each* $A \in \mathfrak{g}$ *the mapping* $t \to \exp tA$ *is an analytic homomorphism from* $\mathbb{R}$ *into* G *and all analytic homomorphisms from* $\mathbb{R}$ *into* G *are obtained in this way.*

PROOF. Clearly $x_A(t) = \exp tA$ and $\exp: \mathfrak{g} \to G$ is analytic. Let us calculate the Jacobian of exp at 0. Choose local coordinates around e and let $E_1,\dots,E_m$ be basis vectors of $\mathcal{T}_e(G)$ corresponding to these coordinates. Then

$$\left.\frac{\partial(\exp(\alpha_1E_1+\dots+\alpha_mE_m))_i}{\partial\alpha_j}\right|_{A=0} = \left.\frac{\partial(\exp(\alpha_jE_j))_i}{\partial\alpha_j}\right|_{A=0} =$$

$$= \left.\frac{d(x_{E_j}(t))_i}{dt}\right|_{t=0} = \left.\frac{\partial y_i}{\partial y_j}\right|_{y=e} = \delta_{ij},$$

where we used (2.32). Hence exp: $\mathfrak{g} \to G$ has Jacobian determinant 1 at $A = 0$. An application of the inverse function theorem then yields Theorem 2.8. □

The subgroups $\{\exp tA | t \in \mathbb{R}\}$ of G with $A \in \mathfrak{g}$, $A \neq 0$, are called *one-parameter subgroups*. The inverse mapping log of exp defines a local chart on some open neighbourhood of e. The corresponding coordinates which take values on some open neighbourhood of 0 in $\mathfrak{g}$ are called *canonical coordinates*.

It follows from Theorem 2.8 that any Lie group G has a symmetric neighbourhood V of e in G such that each $g \in V$ can be written as $g = h^2$ for some $h \in V$. Combination of this fact with Proposition 1.11 enables us to prove the following general result, the physical relevance of which was discussed in § II.2.4.

PROPOSITION 2.9. *Let* U *be a mapping from a connected Lie group* G *into the set of all unitary and anti-unitary operators on some Hilbert space* $\mathcal{H}$ *such that*

$$U(g_1)U(g_2) = \omega(g_1,g_2)U(g_1g_2), \qquad g_1,g_2 \in G,$$

where ω *is some complex-valued function with* $|\omega(g_1,g_2)| = 1$. *Then* $U(g)$ *is a unitary operator for all* $g \in G$.

PROOF. Let V be as above. If $g \in V$ then $g = h^2$ for some $h \in V$ and $U(g) = \omega(h,h)^{-1}U(h)^2$ is unitary. Then by induction $U(g_1g_2 \ldots g_k)$ is unitary for $g_1,\ldots,g_k \in V$. Application of Proposition 1.10 completes the proof. □

In the next two subsections we will make important use of the exponential mapping when we establish the connection between Lie subgroups of G and subalgebras of $\mathfrak{g}$ and between Lie group homomorphisms and homomorphisms for the corresponding subalgebras.

2.5. Lie subgroups and subalgebras

In the beginning of this chapter we already met several closed subgroups of $GL(n,\mathbb{C})$ (like $U(n)$) which turned out to be Lie groups. In fact these groups are Lie subgroups of $GL(n,\mathbb{C})$, but we did not yet define Lie subgroups. First we have to define analytic submanifolds.

DEFINITION 2.10. A subset N of an analytic manifold M is called an *analytic submanifold* of M if N is an analytic manifold itself and if a collection of local charts on M and N, respectively, can be chosen which are related to each other in the following way: For each $p \in N$ there are local charts (U,ϕ) on M and (V,ψ) on N such that $p \in V \subset U$ and:

(a) $\phi(U) = \{(x_1,\dots,x_m) \in \mathbb{R}^m \mid |x_i| < 1 \text{ for } i = 1,\dots,m\}$, $\phi(p) = (0,\dots,0)$;

(b) $\phi(V) = \{(x_1,\dots,x_n,0,\dots,0) \in \mathbb{R}^m \mid |x_i| < 1 \text{ for } i = 1,\dots,n\}$;

(c) If $q \in V$ and $\phi(q) = (x_1,\dots,x_n,0,\dots,0)$ then $\psi(q) = (x_1,\dots,x_n)$.

The definition tells us that for each $p \in N$ there is a cubic coordinates neighbourhood U of p in M such that some slice of U through p is a cubic coordinate neighbourhood V of p in N and the local coordinates of N are obtained by projecting the local coordinates of M on this slice.

DEFINITION 2.11. A *Lie subgroup* H of a Lie group G is a subgroup H of G which is also an analytic submanifold of G, such that H becomes a Lie group with repect to this analytic structure.

It follows from Proposition 1.5 that subgroups G of $GL(n,\mathbb{C})$ which satisfy Assumption 1.3 are Lie subgroups of $GL(n,\mathbb{C})$. However, these subgroups are nicer than what can happen in the general case. Their topology induced by $GL(n,\mathbb{C})$ coincides with their topology as an analytic manifold. Generally, if N is an analytic submanifold of an analytic manifold M then the topology of N as analytic manifold is finer than the topology induced on N by M and the two topologies need not coincide. If they do, we say that N is *regularly imbedded* in M. For a Lie subgroup H of a Lie group G it can be proved that H is regularly imbedded in G if and only if H is a closed subgroup of G (cf. VARADARAJAN [10, Theorem 2.5.4]). It is very nice to know that any closed subgroup H of G has a unique analytic structure such that it becomes a regularly imbedded analytic submanifold of G and hence a Lie subgroup of G (cf. CHEVALLEY [2, Chap. 4, § 14]). An example of a Lie subgroup of $GL(2,\mathbb{C})$ which is not regularly imbedded is given by the group G of all matrices

$$\begin{pmatrix} e^{ita} & 0 \\ 0 & e^{itb} \end{pmatrix}, \qquad t \in \mathbb{R},$$

where a, b are real and nonzero and a/b is irrational. Here G is made into a Lie group by using $t \in \mathbb{R}$ as a global coordinate on G.

We would like to identify the Lie algebra h of a Lie subgroup H of a Lie group G with some subalgebra of the Lie algebra g of G. For doing this we have to know how the tangent space at p to some analytic submanifold N of an analytic manifold M can be identified with a linear subspace of the tangent space at p to M. Use related local charts (U,ϕ) and (V,ψ) around p on M and N, respectively, as described in Definition 2.10. If $t \to x(t)$ is the representative of some tangent vector to N at p such that with respect to the local chart (V,ψ) we have

$$x_i(t) = ta_i + o(t), \qquad i = 1,\dots,n, \quad t \downarrow 0,$$

then $t \to x(t)$ is also the representative of some tangent vector to M at p and with respect to the local chart (U,ϕ) we have

$$x_i(t) = \begin{cases} ta_i + o(t), & i = 1,\dots,n, \quad t \downarrow 0, \\ 0 \qquad\qquad , & i = n+1,\dots,m. \end{cases}$$

Hence, in terms of the local charts (U,ϕ) and (V,ψ) the tangent space $T_p(N)$ is indentified with a linear subspace of the tangent space $T_p(M)$ by the imbedding

$$(a_1,\dots,a_n) \to (a_1,\dots,a_n,0,\dots,0).$$

Now let H be a Lie subgroup of the Lie group G. Then the Lie algebra $h = T_e(H)$ can be identified with a linear subspace of the Lie algebra $g = T_e(G)$. We will show that this linear subspace is a subalgebra of g and that the Lie algebra structure of h is the same as the Lie algebra structure induced on h as a subalgebra of g. We follow approach (a) of subsection 2.3. Let $t \to x(t)$ and $t \to y(t)$ be representatives of tangent vectors $A,B \in T_e(H)$, respectively. Then these curves are also representatives of the corresponding tangent vectors in $T_e(G)$. Let the curve $t \to z(t)$ be defined by (2.26). Then $t \to z(t)$ represents the commutator product of A and B both in $T_e(H)$ and in $T_e(G)$. Hence, under the canonical identification of $T_e(H)$ with a linear subspace of $T_e(G)$ these two commutator products coincide.

Let G, H, g and h be as above. In subsection 2.4 we have defined exponential mappings $\exp_G$: $g \to G$ and $\exp_H$: $h \to H$. Of course, these two mappings are related to each other: $\exp_H$ is the restriction of $\exp_G$ to h, where h

is considered as a subalgebra of $\mathfrak{g}$. This is seen immediately, since for $A \in \mathfrak{h}$ both $t \to \exp_H tA$ and $t \to \exp_G tA$ are analytic homomorphisms from $\mathbb{R}$ into G which are also representatives of the tangent vector A. Then $\exp_H tA = \exp_G tA$ by Proposition 2.7. This result can be applied to groups G (including all familiar linear groups) which satisfy Assumption 1.3. The Lie algebra $\mathfrak{g}$ of such a group G is a subalgebra of $M_n(\mathbb{C})$ and $\exp_G: \mathfrak{g} \to G$ is the restriction of exp to $\mathfrak{g}$, where $\exp: M_n(\mathbb{C}) \to GL(n,\mathbb{C})$ is defined by (1.1). Now $\exp_G$ is an analytic diffeomorphism from some open neighbourhood of 0 in $\mathfrak{g}$ to some open neighbourhood of I in G by Theorem 2.8. Then it is clear from Assumption 1.3 that $\mathfrak{g}$ coincides with the linear subspace L introduced there.

As a second application of the fact that $\exp_H$ is the restriction of $\exp_G$ to $\mathfrak{h}$ we show that a connected Lie subgroup H of a Lie group G is uniquely determined by its Lie algebra $\mathfrak{h} \subset \mathfrak{g}$. For if H_1 and H_2 are two such subgroups with the same Lie algebra $\mathfrak{h}$ then by Theorem 2.8 they have a common open neighbourhood of e of the form $\exp(\mathfrak{h} \cap V)$, where V is some open neighbourhood of 0 in $\mathfrak{g}$. Now application of Proposition 1.11 shows that $H_1 = H_2$.

Next consider the following problem. Let G be a Lie group with Lie algebra $\mathfrak{g}$ and let $\mathfrak{h}$ be a subalgebra of $\mathfrak{g}$. Does there exist a Lie subgroup H of G which has $\mathfrak{h}$ as Lie algebra? The answer is positive and the component H_0 of e in H is uniquely determined by $\mathfrak{h}$. See CHEVALLEY [2, Ch. 4, §4] for the proof in the general case. Here we only consider the case that G = GL(n,$\mathbb{C}$). We will make use of the *Campbell-Baker-Hausdorff formula* which states that, for A, B in some open neighbourhood V of 0 in $M_n(\mathbb{C})$,

$$\log(\exp A \exp B) = A + B + \frac{1}{2}[A,B] + \dots , \tag{2.33}$$

where the n'th term is a finite linear combination of commutor products $[X_1,[X_2,[X_3,\dots,[X_{n-1},X_n]\dots]$, with $X_i = A$ or B for all $i = 1,\dots,n$, and where the series converges absolutely and uniformly for $A,B \in V$. See MILLER [7, Theorem 5.5] for a proof. Let $\mathfrak{h}$ be a subalgebra of $M_n(\mathbb{C})$. It follows from (2.33) that if $\exp A, \exp B \in \exp(V \cap \mathfrak{h})$ then $\exp A \exp B \in \exp(\mathfrak{h})$. Hence $\exp(V \cap \mathfrak{h})$ becomes a local Lie subgroup of GL(n,$\mathbb{C}$). Now let H be the subgroup of GL(n,$\mathbb{C}$) which is generated by $\exp(V \cap \mathfrak{h})$. Then it is not difficult to define an analytic structure on H such that H becomes a Lie subgroup of GL(n,$\mathbb{C}$). For the case of H satisfying Assumption 1.3 this procedure was discussed in §1.3.

Let us summarize the results of this subsection in the following theorem:

THEOREM 2.12. *Let* G *be a Lie group with Lie algebra* $\mathfrak{g}$. *There is a one-to-one correspondence between the connected Lie subgroups* H *of* G *and the subalgebras* $\mathfrak{h}$ *of* $\mathfrak{g}$ *such that* $\mathfrak{h}$ *is the Lie algebra of* H.

2.6. Lie group homomorphisms and Lie algebra homomorphisms

Let $\mathfrak{g}$ and $\mathfrak{h}$ be Lie algebras. A linear mapping $f: \mathfrak{g} \to \mathfrak{h}$ is called a *homomorphism* if $f([A,B]) = [f(A),f(B)]$ for all $A,B \in \mathfrak{g}$. Let $\mathfrak{g}$ and $\mathfrak{h}$ be the Lie algebras of the Lie groups G and H, respectively. We will relate the homomorphisms $f: \mathfrak{g} \to \mathfrak{h}$ to the analytic homomorphisms $F: G \to H$. (By an *analytic homomorphism* we mean an analytic mapping from G into H which is also a group homomorphism, i.e. $F(xy) = F(x)F(y)$ for all $x,y \in G$.) This relationship has a very important application to finite-dimensional representation theory: In order to find the analytic homomorphisms from G into $GL(n,\mathbb{C})$ it is sufficient to find the homomorphisms from $\mathfrak{g}$ into $M_n(\mathbb{C})$. This last problem is purely algebraic and, to a large extent, linear. Therefore its solution may be easier then the corresponding group problem. We earlier mentioned results, where the topological structure already implies the analytic structure, for instance: a closed subgroup of a Lie group is a Lie subgroup. There is a similar result for analytic homomorphisms: if G and H are Lie groups and $F: G \to H$ is a continuous homomorphism then F is analytic (cf. CHEVALLEY [2, Ch. 4, § 13, Prop. 1]).

Let G and H be Lie groups and let $F: G \to H$ be an analytic homomorphism. Then dF_e is a linear mapping from $T_e(G) = \mathfrak{g}$ into $T_e(H) = \mathfrak{h}$ (cf. Definition 2.3). If $t \to x(t)$ and $t \to y(t)$ are representatives of A and $B \in \mathfrak{g}$, respectively, then $t \to F(x(t))$ and $t \to F(y(t))$ are representatives of $dF_e(A)$ and $dF_e(B) \in \mathfrak{h}$, respectively. Now the mapping

$$t \to F(x(t^{\frac{1}{2}})y(t^{\frac{1}{2}})(x(t^{\frac{1}{2}}))^{-1}(y(t^{\frac{1}{2}}))^{-1}) =$$

$$= F(x(t^{\frac{1}{2}}))F((y(t^{\frac{1}{2}}))(F(x(t^{\frac{1}{2}})))^{-1}(F(y(t^{\frac{1}{2}})))^{-1}$$

is a representative of both $dF_e([A,B])$ and $[dF_e(A), dF_e(B)]$, hence these two elements of $\mathfrak{h}$ coincide and dF_e (which we will write henceforth as dF) is a homomorphism from $\mathfrak{g}$ into $\mathfrak{h}$.

If G is a connected Lie group then the analytic homomorphism $F:G \to H$ is uniquely determined by $dF = f: \mathfrak{g} \to \mathfrak{h}$. In fact, if $A \in \mathfrak{g}$ then $t \to F(\exp_G tA)$ is an analytic homomorphism from $\mathbb{R}$ into H and it is also a representative of $f(A) \in T_e(H)$. Hence in view of Proposition 2.7:

(2.34) $$F(\exp_G tA) = \exp_H(tf(A)).$$

Thus $F(g)$ is uniquely determined by f for g in some neighbourhood of e in G (cf. Theorem 2.8) and the global uniqueness of F follows from Proposition 1.11.

Next we consider the existence problem for F if f is given. Let G and H be Lie groups, G connected, $\mathfrak{g}$ and $\mathfrak{h}$ the corresponding Lie algebras and $f\colon \mathfrak{g} \to \mathfrak{h}$ an homomorphism. Does there exist an analytic homomorphism $F\colon G \to H$ such that $dF = f$? From (2.34) we know that if F exists then it must be defined in some neighbourhood of $e \in G$ by

(2.35) $$F(\exp_G A) = \exp_H(f(A)),$$

where A is in some open neighbourhood of $0 \in \mathfrak{g}$. This locally defined mapping is clearly analytic and it is also a local homomorphism. This last fact follows from the observation that $\{(A,f(A)) \mid A \in \mathfrak{g}\}$ is a subalgebra of the Lie algebra $\mathfrak{g} \oplus \mathfrak{h}$, for which the corresponding connected Lie subgroup K of the Lie group $G \times H$ (cf. Theorem 2.12) is generated by all elements $(\exp_G A, \exp_H f(A)) \in G \times H$, where $A \in \mathfrak{g}$. Each element in some open neighbourhood of e in this subgroup K can be written as $(\exp_G A, \exp_H f(A))$ for some unique A in a certain open neighbourhood of 0 in $\mathfrak{g}$. Hence, for A_1 and A_2 sufficiently small:

$$(\exp_G A_1 \exp_G A_2, \exp_H f(A_1) \exp_H f(A_2)) =$$
$$= (\exp_G A_3, \exp_H f(A_3)) \qquad \text{for some } A_3 \in \mathfrak{g}.$$

By the use of (2.35) it follows that

$$F(\exp_G A_1)F(\exp_G A_2) = F(\exp_G A_1 \exp_G A_2)$$

for A_1, A_2 sufficiently small. Thus F is a *local analytic homomorphism* from G to H, i.e. an analytic mapping F from some open neighbourhood of e in G into H such that $F(g_1g_2) = F(g_1)F(g_2)$ for g_1 and g_2 in a sufficiently small neighbourhood of e in G.

The global extension of F can be done by the use of Proposition 1.11. We know that $F(xy) = F(x)F(y)$ for x, y in some open symmetric neighbourhood V of e. A general element of G can be written as $x_1x_2\ldots x_k$, with all

$x_i \in V$. Then define

(2.36) $$F(x_1x_2\ldots x_k) := F(x_1)F(x_2),\ldots,F(x_k),$$

and F becomes an homomorphism, since

$$F(x_1x_2\ldots x_ky_1y_2\ldots y_\ell) = F(x_1)\ldots F(x_k)F(y_1)\ldots F(y_\ell) =$$
$$= F(x_1\ldots x_k)F(y_1\ldots y_\ell).$$

The only problem is that F may be multi-valued, since definition (2.36) may depend on the choice of the x_i's. This is a very serious fact which cannot be repaired. For instance, the two-to-one homomorphism from SU(2) onto SO(3) is locally a one-to-one homomorphism. Hence the inverse mapping is locally a one-to-one homomorphism from SO(3) to SU(2), for which the global extension is multi-valued: a one-to-two mapping from SO(3) onto SU(2).

If the Lie group is both connected and simply connected then the global extension of F to a one-valued homomorphism is always possible. In order to see this let $g \in G$ and connect g with e by a continuous curve $t \to x(t)$. Then the local definition of F around e together with the condition that F is a continuous homomorphism uniquely determine F on the curve as a function of t. In particular, F(g) is determined, but its definition may depend on the choice of the curve. However, since G is simply connected, any two continuous cruves connecting e with g can be continuously deformed into each other and this deformation may be done such that the two endpoints remain fixed. Hence the value of F(g) does not depend on the choice of the curve. Let us summarize our results in the following Theorem (see also CHEVALLEY [2, Chap. 4, §6]):

THEOREM 2.13. *Let* G *and* H *be connected Lie groups with Lie algebras* $\mathfrak{g}$ *and* $\mathfrak{h}$*, respectively. There is a one-to-one correspondence between the homomorphisms* $f: \mathfrak{g} \to \mathfrak{h}$ *and the local analytic homomorphisms* $F: G \to H$*, such that* $dF = f$*. If* G *is simply connected then each local analytic homomorphism* $F: G \to H$ *extends uniquely to a global analytic homomorphism* $F: G \to H$.

2.7. Locally isomorphic Lie groups

As an application of Theorem 2.13 let us consider to which extent a connected Lie group is determined by its Lie algebra. Two Lie algebras g and h are called *isomorphic* if there is a one-to-one homomorphism from g onto h. Two isomorphic Lie algebras are essentially the same and we may consider the family of all Lie algebras isomorphic to a given Lie algebra as an equivalence class. Now what about the family of all connected Lie groups which correspond to a given equivalence class of isomorphic Lie algebras? We already mentioned Ado's theorem that any Lie algebra is isomorphic to a linear Lie algebra. Then Theorem 2.12 shows that this linear Lie algebra corresponds to a connected Lie subgroup of $GL(n,\mathbb{C})$ for some n. Hence the family of connected Lie groups corresponding to an equivalence class of isomorphic Lie algebras is nonempty. Now let G and H be two connected Lie groups belonging to this family. Then their Lie algebras g, h are isomorphic, i.e. $f: g \to h$ and $f^{-1}: h \to g$ are both one-to-one homomorphisms onto. It follows from Theorem 2.13 that G and H are *locally isomorphic* Lie groups: the local analytic homomorphism $F: G \to H$ satisfying $dF = f$ is an analytic diffeomorphism from some open neighbourhood of e in G onto some open neighbourhood of e in H. Conversely, it follows from Theorem 2.13 that locally isomorphic Lie groups have isomorphic Lie algebras. Summarizing: there is a one-to-one correspondence between equivalence classes of locally isomorphic connected Lie groups and equivalence classes of isomorphic Lie algebras.

Each equivalence class of locally isomorphic connected Lie groups contains a simply connected Lie group, which is unique up to global analytic isomorphisms. Starting with an arbitrary connected Lie group G a simply connected Lie group $\tilde{G}$ being locally isomorphic to G can be constructed as follows. Let $\tilde{G}$ consist of all continuous curves in G connecting e with some $g \in G$ and consider two such curves as being equivalent if they have the same endpoint and can be continuously deformed into each other. If $t \to x(t)$ and $t \to y(t)$ are two such curves then define their product by the curve $t \to x(t)y(t)$. This definition of the product respects the above convention of equivalence. Thus $\tilde{G}$ becomes a group on which locally around e an analytic structure can be defined such that the mapping which associates with each curve $t \to x(t)$ its endpoint $x(1)$ is a local analytic isomorphism from $\tilde{G}$ to G. Finally the local analytic structure can be globally extended on $\tilde{G}$ such that $\tilde{G}$ becomes a Lie group, which turns out to be connected and simply connected. The above local isomorphism extends to an analytic homomorphism ω from $\tilde{G}$ onto G, which is not one-to-one except if G is simply connected itself. This homomorphism ω is *locally homeomorphic*, i.e. for each $g \in \tilde{G}$ there is an open neighbourhood V of g in $\tilde{G}$ such that $\omega(V)$ is an open neighbourhood of $\omega(g)$ in G and ω: $V \to \omega(V)$ is a homeomorphism. We call $\tilde{G}$ the *universal covering group* of G. It has the following remarkable property: If F: G $\to$ H is a local analytic homomorphism from G into a Lie group H then there is a unique global analytic homomorphism $\tilde{F}$: $\tilde{G} \to H$ such that $F \circ \omega = \tilde{F}$ locally around e. This is a corollary of Theorem 2.13. Thus we can make precise in what sense the local analytic homomorphism F: G $\to$ H corresponding to a given homomorphism f: $\mathfrak{g} \to \mathfrak{h}$ extends to a multivalued global homomorphism: Consider the one-valued global homomorphism $\tilde{F}$: $\tilde{G} \to H$ corresponding to f; if $g \in G$ then F(g) can take all values $F(\tilde{g})$, where $\tilde{g}$ is any element of $\omega^{-1}(g)$.

Examples of pairs $(G,\tilde{G})$ are SO(3) with universal covering group SU(2) and $L_+^\uparrow$ with universal covering group SL(2,$\mathbb{C}$). Here the mapping ω is two-to-one. A much more elementary example is the circle group $\mathbb{T}$ with universal covering group $\mathbb{R}$. Here the mapping ω is countably infinite-to-one. The notion of the universal covering group is of particular importance in representation theory: Each finite-dimensional representation of the Lie algebra $\mathfrak{g}$ leads to a representation of a corresponding connected Lie group G, which may be multi-valued. However, the corresponding representation of the universal covering group $\tilde{G}$ is one-valued.

We conclude this subsection with a theorem which summarizes the results:

THEOREM 2.14. *There is a one-to-one correspondence between the equivalence classes of isomorphic Lie algebras and the equivalence classes of locally isomorphic connected Lie groups. Each equivalence class of the last kind contains a linear Lie group. It contains also a simply connected Lie group which is unique up to analytic isomorphisms. If* $\tilde{G}$ *and* G *are locally isomorphic connected Lie groups,* $\tilde{G}$ *simply connected, then any local analytic isomorphism* ω: $\tilde{G} \to G$ *extends to a one-valued global analytic homomorphism which is locally homeomorphic.*

See VARADARAJAN [10, §2.6] for the details of the proofs.

2.8. The adjoint representation

In the search for representations of a Lie group G with Lie algebra $\mathfrak{g}$, one particular representation of G and $\mathfrak{g}$ is automatically given, without any further construction. This is the so-called adjoint representation, for which the representation space is given by the Lie algebra $\mathfrak{g}$ itself. The adjoint representation is an important tool for the structure theory of Lie algebras (cf. § 3). It is also a nice application of the previous subsections.

Let us first compare with each other the *inner automorphism* $\alpha(x)$ $(x \in G)$ of a Lie group G and the *inner derivations* ad A $(A \in \mathfrak{g})$ of a Lie algebra $\mathfrak{g}$. These are defined by:

$$(2.37) \qquad (\alpha(x))(g) := x\,g\,x^{-1}, \qquad g \in G,$$

$$(2.38) \qquad (\mathrm{ad}\,A)(B) := [A,B], \qquad B \in \mathfrak{g}.$$

The inner automorphisms of G are special *Lie group automorphisms,* i.e. group isomorphisms F from G onto G such that the both F and F^{-1} are analytic. An application of the Jacobi identity (2.2) shows that

$$(2.39) \qquad (\mathrm{ad}\,A)([B,C]) = [(\mathrm{ad}\,A)(B),C] + [B,(\mathrm{ad}\,A)(C)], \qquad A,B,C \in \mathfrak{g}.$$

A *derivation* of a Lie algebra $\mathfrak{g}$ is a linear mapping D: $\mathfrak{g} \to \mathfrak{g}$ such that

$$(2.40) \qquad D([A,B]) = [D(A),B] + [A,D(B)].$$

Hence the inner derivations ad A of $\mathfrak{g}$ are special derivations of $\mathfrak{g}$. Note that a *normal subgroup* of G is a subgroup of G which is invariant under all inner automorphisms of G. Similarly, we define an *ideal* of the Lie algebra $\mathfrak{g}$ as a linear subspace of $\mathfrak{g}$ which is invariant under all inner derivations of $\mathfrak{g}$. In other words: the linear subspace $\mathfrak{n}$ of $\mathfrak{g}$ is an ideal if $[A,B] \in \mathfrak{n}$ for all $A \in \mathfrak{g}$, $B \in \mathfrak{n}$. Clearly, an ideal is also a subalgebra.

We can also consider $\alpha(x)$ and ad A as functions of x and A, respectively. Observe that the mapping $x \to \alpha(x)$ is a homomorphism from G into the group Aut(G) of all Lie group automorphisms of G:

$$(2.41) \qquad \alpha(xy) = \alpha(x)\circ\alpha(y), \qquad x,y \in G.$$

Another application of the Jacobi identity (2.2) gives:

$$(2.42) \qquad \mathrm{ad}[A,B] = \mathrm{ad}\,A \circ \mathrm{ad}\,B - \mathrm{ad}\,B \circ \mathrm{ad}\,A.$$

Hence the mapping $A \to \mathrm{ad}\,A$ is a homomorphism from the Lie algebra $\mathfrak{g}$ into the Lie algebra $L(g)$ which consists of all linear mappings from g into itself and where the commutator product is defined by (2.3). Generally, a representation of a Lie algebra g on a linear space V is a homomorphism from g into $L(V)$. Hence ad defines a special representation of g on g itself (considered as a linear space): the *adjoint representation* of g.

Inner automorphisms of a Lie group G and inner derivations of the corresponding Lie algebra g cannot be related to each other by a straightforward application of Theorem 2.13, since both the action of $\alpha(x)$ on G and the dependence of $\alpha(x)$ on x have to be transferred to the Lie algebra. We need an intermediate step, which will be given by the adjoint representation of G.

Fix $x \in G$. The Lie group automorphism $\alpha(x): G \to G$ induces a Lie algebra automorphism $d(\alpha(x)): g \to g$. (By an *automorphism of a Lie algebra* g we mean a one-to-one homomorphism from g onto itself.) Let us define

(2.43) $$\mathrm{Ad}\ x := d(\alpha(x)), \qquad x \in G.$$

Then

(2.44) $$(\mathrm{Ad}\ x)([A,B]) = [(\mathrm{Ad}\ x)(A),(\mathrm{Ad}\ x)(B)], \quad A,B \in g,$$

and, as a corollary of (2.41):

(2.45) $$\mathrm{Ad}(xy) = \mathrm{Ad}\ x \circ \mathrm{Ad}\ y, \qquad x,y \in G.$$

Hence the mapping $x \to \mathrm{Ad}\ x$ is a homomorphism from G into $GL(g)$, the group of all nonsingular linear transformations of g (considered as a linear space). The representation of G on g thus defined is called the *adjoint representation* of G.

It is easily shown that the mapping $x \to \mathrm{Ad}\ x$ from G into $GL(g)$ is analytic. It is sufficient to show that the mapping $x \to (\mathrm{Ad}\ x)(B)$ is analytic from G into g for each $B \in g$. Choose a local chart (U,ϕ) around e on G and pick a tangent vector $B \in T_e(G) = g$. Then the mapping

$$(x,t) \to \phi(x(\exp tB)x^{-1})$$

is analytic for $x \in G$, t in some neighbourhood of 0. Identifying $T_e(g)$ with $\mathbb{R}^m$ (cf. Definition 2.1(b)) we obtain:

(2.46) $$(\mathrm{Ad}\ x)(B) = \frac{d}{dt}\,\phi(x(\exp tB)x^{-1})\Big|_{t=0}.$$

The right hand side of (2.40) is clearly analytic in x.

If G is a linear Lie group, i.e. a Lie subgroup of $GL(n,\mathbb{C})$, then we can identify its Lie algebra $\mathfrak{g}$ with some subalgebra of $M_n(\mathbb{C})$, cf. § 2.5. In that case it follows from (2.46) that

(2.47) $$(\mathrm{Ad}\ T)(B) = TBT^{-1}, \qquad T \in G,\ B \in \mathfrak{g}.$$

The analytic homomorphism Ad: $G \to GL(\mathfrak{g})$ induces a Lie algebra homomorphism d(Ad): $\mathfrak{g} \to L(\mathfrak{g})$. We will show that

(2.48) $$d(\mathrm{Ad}) = \mathrm{ad}.$$

Take again a local chart (U,ϕ) around e on G. Then it follows from (2.45) that

(2.49) $$((d(\mathrm{Ad}))(A))(B) = \frac{\partial^2}{\partial s\,\partial t}\,\phi(\exp sA\ \exp tB\ (\exp sA)^{-1})\Big|_{\substack{s=0\\t=0}}.$$

By (2.24) and (2.20) the i-th coordinate of the right hand side of (2.49) equals

$$\sum_{j,k=1}^{m} g_{i;j,k}\, A_j B_k = [A,B]_i.$$

This completes the proof of (2.48).

The passage from $\alpha(x)$ to Ad x and from Ad to ad by means of (2.43) and (2.48) can be inverted by means of the exponential mapping. Application of (2.35) with F = Ad, f = ad and H = $GL(\mathfrak{g})$ gives

(2.50) $$\mathrm{Ad}(\exp_G A) = \exp_{GL(\mathfrak{g})}(\mathrm{ad}\ A), \qquad A \in \mathfrak{g}.$$

Another application of (2.35) with $F = \alpha(x)$ $(x \in G)$, f = Ad x, H = G gives

(2.51) $$x(\exp_G A)x^{-1} = \exp_G((\mathrm{Ad}\ x)(A)), \qquad x \in G,\ A \in \mathfrak{g}.$$

For linear Lie groups G formulas (2.50) and (2.51) take the form

$$(2.52)\qquad e^{A}Be^{-A} = e^{\mathrm{ad}\,A}(B), \qquad A,B \in \mathfrak{g},$$

$$(2.53)\qquad Te^{A}T^{-1} = e^{TAT^{-1}}, \qquad T \in G,\ A \in \mathfrak{g}.$$

where the exponential mapping is defined by the power series (1.1).

Let us summarize the preceding results in a theorem.

THEOREM 2.15. *Let* G *be a Lie group with Lie algebra* $\mathfrak{g}$. *Then there are unique mappings* α: G $\to$ Aut(G), Ad: G $\to$ GL($\mathfrak{g}$), ad: $\mathfrak{g} \to L(\mathfrak{g})$ *satisfying the following properties:*

(a) α: G $\to$ Aut(G) *is a group homomorphism;*
Ad x: $\mathfrak{g} \to \mathfrak{g}$ *is a Lie algebra automorphism for each* x $\in$ G;
Ad: G $\to$ GL($\mathfrak{g}$) *is an analytic group homomorphism;*
ad A: $\mathfrak{g} \to \mathfrak{g}$ *is a derivation for each* A $\in \mathfrak{g}$.
ad: $\mathfrak{g} \to L(\mathfrak{g})$ *is a Lie algebra homomorphism.*

(b) $(\alpha(x))(g) = xgx^{-1}$ $(x,g \in G)$ *and*
$(\mathrm{ad}\,A)(B) = [A,B]$ $(A,B \in \mathfrak{g})$.

(c) Ad x = d(α(x)) (x $\in$ G) *and* ad = d(Ad).

(d) $\mathrm{Ad}(\exp A) = e^{\mathrm{ad}\,A}$ $(A \in \mathfrak{g})$ *and*
$x(\exp A)x^{-1} = \exp((\mathrm{Ad}\,x)(A))$ $(x \in G,\ A \in \mathfrak{g})$.

In the beginning of this subsection we shortly discussed normal subgroups of G and ideals of $\mathfrak{g}$. Under the canonical relationship between connected Lie subgroups and subalgebras normal subgroups are related to ideals:

THEOREM 2.16. *Let* G *be a Lie group with Lie algebra* $\mathfrak{g}$. *There is a one-to-one correspondence between connected normal Lie subgroups* N *of* G *and ideals* $\mathfrak{n}$ *of* $\mathfrak{g}$ *such that* $\mathfrak{n}$ *is the Lie algebra of* N.

See CHEVALLEY [2, Ch. 4, § 11, Prop. 2]. For the proof use (2.43) and (2.48) in the one direction and (2.50) and (2.51) in the other direction.

For a simple application of the adjoint representation consider the construction of the two-to-one homomorphism from SU(2) onto SO(3). Formula (2.47) defines the representation Ad of SU(2) on its Lie algebra $\mathfrak{su}(2)$, which consists of all skew-hermitian 2×2 matrices with trace zero, i.e. all matrices

$$A = \begin{pmatrix} -ia_3 & -ia_1-a_2 \\ -ia_1+a_2 & ia_3 \end{pmatrix}, \qquad a_1,a_2,a_3 \in \mathbb{R}.$$

Then $\det A = a_1^2 + a_2^2 + a_3^2$ is invariant under all transformations Ad T ($T \in SU(2)$), hence Ad T can be identified with an orthogonal transformation of $\mathbb{R}^3$. Then it is left to prove that Ad: $SU(2) \to O(3)$ has image $SO(3)$ and is two-to-one. The last fact follows from the following general result:

PROPOSITION 2.17. *Let* G *be a connected Lie group with Lie algebra* $\mathfrak{g}$. *Then the kernel* $\{x \in G \mid \mathrm{Ad}\, x = I_{\mathfrak{g}}\}$ *of the homomorphism* Ad: $G \to GL(\mathfrak{g})$ *equals the center* $Z := \{x \in G \mid xy = yx$ *for all* $y \in G\}$ *of* G.

PROOF. If $x \in Z$ then $\alpha(x) = id_G$, hence $\mathrm{Ad}\, x = d(\alpha(x)) = I_{\mathfrak{g}}$. Conversely, if $\mathrm{Ad}\, x = I_{\mathfrak{g}}$ then $x(\exp A)x^{-1} = \exp A$ for all $A \in \mathfrak{g}$ (cf. (2.51)), hence $xyx^{-1} = y$ for all $y \in \mathfrak{g}$ by an application of Theorem 2.8 and Proposition 1.11. Thus $y \in Z$. □

It is an easy excercise to prove that the center of SU(2) consists of the 2×2 matrices I and -I. Hence the mapping Ad: $SU(2) \to O(3)$ is two-to-one.

Finally we will prove that $\mathrm{Ad}(SU(2)) = SO(3)$. Again we use a more general result. Note that the kernel $\{x \in G \mid F(x) = e\}$ of an analytic homomorphism F: $G \to H$ (G, H Lie groups) is a closed normal subgroup of G and that the kernel $\{A \in \mathfrak{g} \mid f(A) = 0\}$ of a homomorphism f: $\mathfrak{g} \to \mathfrak{h}$ ($\mathfrak{g}$, $\mathfrak{h}$ Lie algebras) is an ideal of $\mathfrak{g}$.

PROPOSITION 2.18. *Let* G *and* H *be Lie groups with Lie algebras* $\mathfrak{g}$ *and* $\mathfrak{h}$, *respectively. Let* F: $G \to H$ *be an analytic homomorphism with kernel* $N = \{x \in G \mid F(x) = e\}$. *Then* N *is a closed normal Lie subgroup of* G *and its Lie algebra* $\mathfrak{n}$ *is the kernel* $\{A \in \mathfrak{g} \mid dF(A) = 0\}$ *of* dF: $\mathfrak{g} \to \mathfrak{h}$.

PROOF. Choose an open neighbourhood V of 0 in g such that $\exp_G$: $V \to \exp_G(V)$ is an analytic diffeomorphism and $\exp_H$: $dF(V) \to H$ is one-to-one. Then it follows from the identity

$$F(\exp_G A) = \exp_H(dF(A)), \qquad A \in \mathfrak{g},$$

(cf. (2.35)) that $x \in N \cap \exp_G(V)$ if and only if $x = \exp_G A$ for some $A \in V$

with $dF(A) = 0$. This means that $N \cap \exp_G(V)$ is a local Lie subgroup of G with Lie algebra n equal to the kernel of dF. The local Lie group structure of N can be extended globally by standard methods (cf. § 1.3). □

COROLLARY 2.19. *Let* G *be a connected Lie group with Lie algebra* g. *Then the center* Z *of* G *is a closed normal Lie subgroup of* G *for which the Lie algebra* z *is equal to the center* $\{A \in g \mid [A,B] = 0$ *for all* $B \in g\}$ *of* g.

PROOF. Use Propositions 2.17 and 2.18 and formula (2.48). □

Since SU(2) has center {I,-I}, its Lie algebra $su(2)$ has center {0}. The center of $su(2)$ is the kernel of ad: $su(2) \to o(3)$. Hence ad is one-to-one and it is onto since the Lie algebras $su(2)$ and $o(3)$ have both dimension 3. If follows from Theorem 2.13 that SU(2) and O(3) are locally isomorphic Lie groups. In particular, Ad(SU(2)) contains a neighbourhood of I in O(3). An application of Proposition 1.11 shows that Ad(SU(2)) equals the component of I in O(3).

This component is certainly included in SO(3). In § 1.5 we showed that it equals SO(3) since SO(3) is the image of SU(2) under Ad. This would give a cicular reasoning here. However note that the connectedness of SO(3) also follows from the fact that each $T \in SO(3)$ can be written as

$$T = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{pmatrix} \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_2 & -\sin\theta_2 \\ 0 & \sin\theta_2 & \cos\theta_2 \end{pmatrix}$$

for some $\theta_1, \theta_2, \phi \in \mathbb{R}$, and thus can be connected with I by means of a continuous curve in SO(3).

The treatment of this example has become more lengthy than we intended. However, we used it as an excuse to formulate some propositions which are worthwhile to be known themselves. The reader should also be aware that anyone having a working knowledge of the general facts formulated until here can deal with the example Ad: $SU(2) \to SO(3)$ in a few lines.

3. SEMISIMPLE, SOLVABLE AND NILPOTENT LIE ALGEBRAS AND LIE GROUPS*)

3.1. Prototypes and definitions of the various kinds of Lie algebras

A trivial example of a Lie algebra is a *commutative Lie algebra*, i.e. an arbitrary linear space g with the commutator product defined by $[A,B] := 0$ for all $A,B \in g$. If G is a connected Lie group with Lie algebra g then G is a commutative group if and only if g is a commutative Lie algebra.

Next consider the Lie algebra $o(3)$ of SO(3). For a suitable basis $\{I_1,I_2,I_3\}$ of $o(3)$ we have

$$[I_1,I_2] = I_3,\ [I_2,I_3] = I_1,\ [I_3,I_1] = I_2. \tag{3.1}$$

This Lie algebra is noncommutative, but it is still of a very simple form. Let us examine the ideals of $o(3)$. Clearly, $\{0\}$ and $o(3)$ itself are ideals of $o(3)$. Suppose that $n \neq \{0\}$ is an ideal of $o(3)$. Without loss of generality we may assume that n contains an element $A = a_1I_1 + a_2I_2 + a_3I_3$ with $a_1 \neq 0$. Then $[A,I_2] = a_1I_3 - a_3I_1 \in n$, hence $a_1^{-1}[[A,I_2],I_1] = I_2 \in n$, and $I_1,I_3 \in n$ by (3.1), so $n = o(3)$. We conclude that $o(3)$ has no nontrivial ideals. A Lie algebra g is called *simple* if g is noncommutative and $\{0\}$ and g are the only ideals in g. Hence $o(3)$ is simple. In a similar way it can be shown that the Lie algebra of the proper Lorentz group $L_+^\uparrow$ is simple. (Use formulas (2.23), (2.24) and (2.25) in Chapter III).

Simple Lie algebras are a kind of building blocks of general Lie algebras. If $g_1,g_2,\ldots,g_k$ are Lie algebras then let their *direct sum* $g = g_1 \oplus g_2 \oplus \ldots \oplus g_k$ be the linear space consisting of all k-tuples $(A_1,A_2,\ldots,A_k)$ with $A_i \in g_i$ $(i = 1,\ldots,k)$, which becomes a Lie algebra with respect to the commutator product

$$[(A_1,\ldots,A_k),(B_1,\ldots,B_k)] := ([A_1,B_1],\ldots,[A_k,B_k]).$$

The subspace $\{(0,\ldots,0,A_i,0,\ldots,0) \mid A_i \in g_i\}$ of g is an ideal of g which can be identified with g_i. A Lie algebra is called *semisimple* if it is the direct sum of simple Lie algebras.

Conversely, let us start with an arbitrary Lie algebra g and let us try to decompose it as a direct sum of ideals of g. If g has no nontrivial

---

*) In this section all Lie algebras are supposed to be real and finite-dimensional.

ideals then we are ready. Otherwise, if $\mathcal{g}_1$ is a nontrivial ideal of $\mathcal{g}$ then we look for another complementary ideal $\mathcal{g}_2$ of $\mathcal{g}$, i.e. $\mathcal{g}$ is the linear span of $\mathcal{g}_1$ and $\mathcal{g}_2$ and $\mathcal{g}_1 \cap \mathcal{g}_2 = \{0\}$. Then $\mathcal{g} = \mathcal{g}_1 \oplus \mathcal{g}_2$. Let us suppose that such a complementary ideal exists. Then we can apply the same procedure to $\mathcal{g}_1$ and $\mathcal{g}_2$. If each time we can find complementary ideals then this procedure must come to a natural end as soon as a direct sum decomposition $\mathcal{g} = \mathcal{h}_1 \oplus \ldots \oplus \mathcal{h}_k$ is attained such that no $\mathcal{h}_i$ has nontrivial ideals. Then each $\mathcal{h}_i$ is either simple or it is a one-dimensional Lie algebra, which is necessarily commutative. A Lie algebra $\mathcal{g}$ is called *reductive* if $\mathcal{g}$ is a direct sum of simple and of commutative Lie algebras.

Unfortunately (or maybe fortunately), not every Lie algebra is reductive. Consider for instance the class $\mathcal{t}(n,\mathbb{R})$ of all real upper triangular n×n matrices, i.e. real n×n matrices of the form

$$\begin{pmatrix} a_{11} & & & * \\ & \cdot & & \\ & & \cdot & \\ 0 & & & a_{nn} \end{pmatrix}. \tag{3.2}$$

It is easily seen that $\mathcal{t}(n,\mathbb{R})$ is a real subalgebra of $M_n(\mathbb{C})$. The subclass $\mathcal{t}_0(n,\mathbb{R})$ consisting of all matrices of the form (3.2) with $a_{11} = a_{22} = \ldots = a_{nn} = 0$ is an ideal of $\mathcal{t}(n,\mathbb{R})$. This ideal has a very peculiar property: if $A,B \in \mathcal{t}(n,\mathbb{R})$ then $[A,B] \in \mathcal{t}_0(n,\mathbb{R})$ and $\mathcal{t}_0(n,\mathbb{R})$ is the linear span of all commutator products $[A,B]$ $(A,B \in \mathcal{t}(n,\mathbb{R}))$. In general, the *derived algebra* $\mathcal{g}'$ of a Lie algebra $\mathcal{g}$ is defined as the linear span of all commutator products $[A,B]$ $(A,B \in \mathcal{g})$. It follows by the use of Jacobi identity (2.2) that $\mathcal{g}'$ is an ideal of $\mathcal{g}$. Returning to our example suppose that $\mathcal{t}(n,\mathbb{R})$ has an ideal $\mathcal{n}$ which is complementary to $\mathcal{t}_0(n,\mathbb{R})$. Then $\mathcal{n}$ must possess an element of the form

$$A = \begin{pmatrix} 1 & & & * \\ & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{pmatrix}$$

Let B be the element of $\mathcal{t}_0(n,\mathbb{R})$ having $B_{1,2} = 1$ and all other B_{ij}'s zero. Let $C := [A,B]$. Then $C_{1,2} = 1$, but, on the other hand, $C = 0$ since $\mathcal{t}_0(n,\mathbb{R}) \cap \mathcal{n} = \{0\}$ by assumption. This is a contradiction. Hence $\mathcal{t}(n,\mathbb{R})$ is not reductive.

By recurrence let us define the p^{th} *derived algebra* $g^{(p)} := (g^{(p-1)})'$ of a Lie algebra g, where $g^{(1)} := g'$. Then $g \supset g' \supset g^{(2)} \supset \ldots \supset g^{(p)} \supset \ldots$ and all $g^{(p)}$'s are ideals of g. This sequence must terminate somewhere, i.e. $g^{(p)} = g^{(p+1)}$ for some p. If g is simple then clearly $g' = g$. (This also holds if g is semisimple). If g is commutative then $g' = \{0\}$ and $g^{(2)} = g'$. If $g = t(n,\mathbb{R})$ then $g^{(p)}$ $(2^{p-1} < n)$ consists of all real n×n matrices of the form

$$\begin{pmatrix} 0 & \cdots & 0 & a_{1,q+1} & & & * \\ & & & & \cdot & & \\ & & & & & \cdot & \\ & & & & & & \cdot \\ & & \bigcirc & & & & a_{n-q,n} \\ & & & & & & 0 \\ & & & & & & \vdots \\ & & & & & & 0 \end{pmatrix} \qquad \text{with } q = 2^{p-1}.$$

Hence $g^{(n)} = \{0\}$ and $g^{(n+1)} = g^{(n)}$. A Lie algebra g is called *solvable* if $g^{(p)} = \{0\}$ for some p. Hence semisimple Lie algebras are never solvable, commutative Lie algebras are always solvable and $t(n,\mathbb{R})$ is a nontrivial example of a solvable Lie algebra.

Even if we are only interested in semisimple Lie algebras and the corresponding Lie groups (like SO(3) and $L_+^\uparrow$) then we are still forced to pay some attention to solvable Lie algebras, since they naturally occur as solvable subalgebras of semisimple Lie algebras, if we analyze the structure of the last ones. For instance, the Lie algebra $s\ell(n,\mathbb{R})$ consisting of all real n×n matrices with trace zero is a semisimple Lie algebra and it is the direct sum of the semisimple subalgebra $o(n)$ consisting of all real n×n skew-symmetric matrices and the solvable subalgebra $st(n,\mathbb{R}) := \{A \in t(n,\mathbb{R}) \mid \text{tr } A = 0\}$. Here the direct sum is taken with respect to the linear structure, not with respect to the Lie algebra structure; $o(n)$ and $st(n,\mathbb{R})$ are subalgebras but not ideals of $s\ell(n,\mathbb{R})$. The solvable subalgebra $st(n,\mathbb{R})$, in its turn, can be written as the direct sum of the commutative subalgebra $sd(n,\mathbb{R})$ consisting of all real n×n diagonal matrices with trace zero and the nilpotent ideal $t_0(n,\mathbb{R})$ we already met. (See below for the definition of nilpotent.) Again this is a direct sum of linear spaces, not of Lie algebras; since $t_0(n,\mathbb{R})$ is an ideal, it is a semidirect sum of Lie algebras.

The resulting decomposition

$$(3.3)\qquad \mathfrak{sl}(n,\mathbb{R}) = \mathfrak{o}(n) + \mathfrak{sd}(n,\mathbb{R}) + \mathfrak{t}_0(n,\mathbb{R})$$

is an example of the *Iwasawa decomposition* $\mathfrak{g} = \mathfrak{k} + \mathfrak{a} + \mathfrak{n}$ of a semisimple Lie algebra of the noncompact type, where the subalgebras $\mathfrak{k}$, $\mathfrak{a}$, $\mathfrak{n}$ are semisimple, commutative and nilpotent, respectively.

We have yet to define a nilpotent Lie algebra, for which $\mathfrak{t}_0(n,\mathbb{R})$ is the prototype. Remember that a linear transformation A from a linear vector space V into itself is called nilpotent if $A^p = 0$ for some natural number p. We call a Lie algebra $\mathfrak{g}$ *nilpotent* if the linear mappings ad A: $\mathfrak{g} \to \mathfrak{g}$ are nilpotent for all $A \in \mathfrak{g}$. In the case of $\mathfrak{t}_0(n,\mathbb{R})$ it is easily verified that $(\text{ad } A)^{n-1}(B) = 0$ for all $A,B \in \mathfrak{t}_0(n,\mathbb{R})$. All commutative Lie algebras are nilpotent; all nilpotent Lie algebras are solvable (cf. VARADARAJAN [10, Theorem 3.7.2]). The Lie algebra $\mathfrak{st}(n,\mathbb{R})$ is solvable but not nilpotent. (Consider $(\text{ad } A)^n$ with A a diagonal matrix of trace zero.)

3.2. The Killing form; semisimple Lie algebras and Lie groups

The semisimplicity of a Lie algebra can also be characterized by the use of the Killing form. On an arbitrary Lie algebra $\mathfrak{g}$ the *Killing form* is a symmetric bilinear form κ on $\mathfrak{g}$ which is given by

$$(3.4)\qquad \kappa(A,B) := \text{tr}(\text{ad } A \circ \text{ad } B), \qquad A,B \in \mathfrak{g}.$$

Choose a basis $\{E_1,\ldots,E_m\}$ for $\mathfrak{g}$ and fix $A,B \in \mathfrak{g}$. We have

$$(3.5)\qquad (\text{ad } A \circ \text{ad } B)(E_j) = [A,[B,E_j]] = \sum_{i=1}^{m} c_{ij}E_i$$

for certain coefficients c_{ij}. Then

$$(3.6)\qquad \text{tr}(\text{ad } A \circ \text{ad } B) := \sum_{i=1}^{m} c_{ii},$$

independent of the choice of the basis.

The importance of the Killing form κ is illustrated by its nice behaviour under automorphisms σ and derivations D of the Lie algebra $\mathfrak{g}$:

$$(3.7)\qquad \kappa(\sigma A,\sigma B) = \kappa(A,B), \qquad A,B \in \mathfrak{g},$$

and

$$\kappa(DA,B) + \kappa(A,DB) = 0, \qquad A,B \in \mathfrak{g}. \tag{3.8}$$

For the proof of (3.7) we use (3.5):

$$(\mathrm{ad}\,\sigma A \circ \mathrm{ad}\,\sigma B)(\sigma E_j) = [\sigma A,[\sigma B,\sigma E_j]] = [\sigma A,\sigma[B,E_j]] =$$
$$= \sigma[A,[B,E_j]] = \sum_{i=1}^{m} c_{ij}\, \sigma E_j.$$

Since $\{\sigma E_1,\ldots,\sigma E_m\}$ is also a basis for $\mathfrak{g}$, it follows that

$$\mathrm{tr}(\mathrm{ad}\,\sigma A \circ \mathrm{ad}\,\sigma B) = \sum_{i=1}^{m} c_{ii}.$$

Combination with (3.6) yields (3.7).

We reduce (3.8) to (3.7) by observing that the exponential e^D of a derivation D of $\mathfrak{g}$ is an automorphism of $\mathfrak{g}$. In fact, let $A,B \in \mathfrak{g}$. Then

$$[e^D A, e^D B] = \sum_{i,j=0}^{\infty} \frac{1}{i!j!}[D^i A, D^j B] = \tag{3.9}$$
$$= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{i=0}^{n} \binom{n}{i} [D^i A, D^{n-i} B] =$$
$$= \sum_{n=0}^{\infty} \frac{1}{n!} D^n [A,B] = e^D [A,B],$$

where we iterated the derivation property (2.40) of D.

If D is a derivation of $\mathfrak{g}$ then tD ($t \in \mathbb{R}$) is also a derivation, hence e^{tD} is an automorphism and (3.7) yields

$$\kappa(e^{tD}A, e^{tD}B) = \kappa(A,B), \qquad A,B \in \mathfrak{g}.$$

Then (3.8) is obtained by differentiation of both sides of the last identity with respect to t for t = 0.

A special case of (3.8) is

$$\kappa((\mathrm{ad}\,C)(A),B) + \kappa(A,(\mathrm{ad}\,C)(B)) = 0, \qquad A,B,C \in \mathfrak{g}. \tag{3.10}$$

This identity can also be proved by a straightforward computation, without the use of (3.7). It follows from (3.9) that $e^{\mathrm{ad}\,C}$ $(C \in \mathfrak{g})$ is an automorphism of $\mathfrak{g}$. Hence a special case of (3.7) is:

$$\kappa(e^{\mathrm{ad}\,C}A, e^{\mathrm{ad}\,C}B) = \kappa(A,B), \qquad A,B,C \in \mathfrak{g}, \tag{3.11}$$

If $\mathfrak{g}$ is the Lie algebra of the Lie group G then Ad x $(x \in G)$ is an automorphism of $\mathfrak{g}$ and (3.7) implies that

$$\kappa((\mathrm{Ad}\,x)(A), (\mathrm{Ad}\,x)(B)) = \kappa(A,B), \qquad A,B \in \mathfrak{g},\ x \in G. \tag{3.12}$$

Let us calculate κ explicitly in some easy cases. For $\mathfrak{g} = \mathcal{O}(3)$ it follows from (3.1) that

$$\kappa(a_1I_1 + a_2I_2 + a_3I_3,\ b_1I_1 + b_2I_2 + b_3I_3) =$$

$$= -2(a_1b_1 + a_2b_2 + a_3b_3) \qquad \text{for real } a_i\text{'s and } b_i\text{'s.}$$

Note that κ is a negative-definite bilinear form on $\mathcal{O}(3)$: $\kappa(A,A) \le 0$ for all $A \in \mathcal{O}(3)$ and $\kappa(A,A) = 0$ implies $A = 0$.

In general, a real bilinear form λ on a real linear space V is called *positive-definite* or *negative-definite* if $\lambda(v,v) \ge 0$ or $\lambda(v,v) \le 0$, respectively, for all $v \in V$ and if $\lambda(v,v) = 0$ implies $v = 0$. In both cases the bilinear form is called *definite*. If $\lambda(v,v)$ attains both positive and negative values for $v \in V$ then λ is called *indefinite*. The bilinear form λ is called *nondegenerate* if for each $v \in V$ the following holds: if $\lambda(v,w) = 0$ for all $w \in V$ then $v = 0$. Clearly, if λ is definite then λ is nondegenerate.

Next we calculate the Killing form on the Lie algebra $\mathfrak{sl}(2,\mathbb{R})$ consisting of all real 2×2 matrices with trace 0. Take

$$J_1 = \begin{pmatrix} 0 & \frac12 \\ \frac12 & 0 \end{pmatrix}, \quad J_2 = \begin{pmatrix} 0 & \frac12 \\ -\frac12 & 0 \end{pmatrix}, \quad J_3 = \begin{pmatrix} \frac12 & 0 \\ 0 & -\frac12 \end{pmatrix}$$

as a basis for $\mathfrak{sl}(2,\mathbb{R})$. Then

$$[J_1,J_2] = -J_3, [J_2,J_3] = -J_1, [J_3,J_1] = J_2.$$

Hence

$$\kappa(a_1J_1 + a_2J_2 + a_3J_3, b_1J_1 + b_2J_2 + b_3J_3) =$$
$$= 2a_1b_1 - 2a_2b_2 + 2a_3b_3 \qquad \text{for real } a_i\text{'s and } b_i\text{'s.}$$

Note that the bilinear form κ is nondegenerate on $s\ell(2,\mathbb{R})$ but not definite.

There is the following criterium for semisimple Lie algebras:

THEOREM 3.1 (cf. VARADARAJAN [10, Theorem 3.9.2]). *A Lie algebra g is semisimple if and only if the Killing form on g is nondegenerate.*

A Lie group is *semisimple* if its Lie algebra is semisimple. A connected Lie group is called *solvable* or *nilpotent* if its Lie algebra is solvable or nilpotent, respectively.

The semisimple Lie algebras $o(3)$ and $s\ell(2,\mathbb{R})$ correspond to semisimple Lie groups SO(3) and SL(2,$\mathbb{R}$), respectively. In the first case, SO(3) is compact and the Killing form on $o(3)$ is negative-definite. In the second case, SL(2,$\mathbb{R}$) is noncompact and the Killing form on $s\ell(2,\mathbb{R})$ is indefinite. These are examples of a general phenomenon. If g is a semisimple Lie algebra then the adjoint representation of g is faithful, i.e. the homomorphism $A \to \text{ad } A$, $A \in g$, is one-to-one (cf. VARADARAJAN [10, p. 214]). Hence the subalgebra $\text{ad } g := \{\text{ad } A | A \in g\}$ of $L(g)$ is isomorphic to g. Let G be the *adjoint group* of g: the connected Lie subgroup of GL(g) corresponding to the subalgebra ad g of $L(g)$. This group is generated by the elements $e^{\text{ad } A}$, $A \in g$. It follows from (3.11) that the Killing form on g (which is nondegenerate) is invariant under G. It can be proved that G is a closed subgroup of GL(g) (cf. VARADARAJAN [10, Theorem 3.16.3]). Now suppose that κ is negative-definite on g. Then G is a closed subgroup of the group of transformations of g which are orthogonal with respect to the inner product $-\kappa$. Hence G is compact. Thus we know that if κ is negative-definite on g then there is at least one compact connected Lie group with Lie algebra isomorphic to g. A theorem of Weyl states that the universal covering group $\tilde{G}$ of a compact connected semisimple Lie group G is again compact (cf. VARADARAJAN [10, Theorem 4.11.6]). Since each connected Lie group with Lie algebra isomorphic to g can be obtained as the image under a continuous homomorphism from $\tilde{G}$ (cf. Theorem 2.14) we conclude that any connected Lie group G for which the Killing form on its Lie algebra is negative-definite, is compact.

A converse result also holds. We summarize everything in the following theorem:

<u>THEOREM 3.2</u> (cf. VARADARAJAN [10, Theorem 4.11.7 and Exercise 4.35(a)]).

(a) *A connected semisimple Lie group is compact if and only if the Killing form on its Lie algebra is definite. If the Killing form is definite then it is negative-definite.*

(b) *Any connected compact Lie group* G *is locally isomorphic with a direct product* G_1 *of a torus group* $\mathbb{T}^n = \mathbb{T}\times\mathbb{T}\times\ldots\times\mathbb{T}$ (n = 0,1,2,...) *and a finite number of connected, simply connected, compact Lie groups with simple Lie algebras such that* G_1 *covers* G, *i.e. the local analytic isomorphism from* G_1 *to* G *extends to a global analytic homomorphism from* G_1 *onto* G.

A semisimple Lie algebra is of the *compact* or of the *noncompact* type according to whether the Killing form is negative-definite or not, respectively.

The simple Lie algebras are completely classified. Up to isomorphisms, the class of simple Lie algebras of the compact type consists of the Lie algebras $\mathcal{o}(n)$ $(n \geq 7)$, $\mathcal{su}(n)$ $(n \geq 2)$, $\mathcal{sp}(n)$ $(n \geq 2)$ (i.e. the Lie algebra of the symplectic group Sp(n)), and five exceptional Lie algebras (cf. VARADARAJAN [10, §4.5]). Thus, in view of Theorem 3.2, the compact connected Lie groups are classified up to local isomorphisms.

3.3. <u>The Levi decomposition</u>

Let us finally consider the so-called Levi decomposition of a general Lie algebra as the (semidirect) sum of a solvable ideal and a semisimple subalgebra. This leads to the decomposition of a connected Lie group G as the semidirect product QM of a solvable normal Lie subgroup Q and a semisimple Lie subgroup M. Then G is completely determined by Q, M and the action via inner automorphisms of M on Q. Thus the representation theory for general connected Lie groups G may follow from a complete knowledge of the representations of semisimple and of solvable Lie groups. In this colloquium we will mainly restrict ourselves to the case that Q is abelian in the above decomposition G = QM. The Poincaré group $P_+^\uparrow$ provides a example of this case: here Q is the abelian Lie group $\mathbb{R}^4$ and M is the semisimple Lie group $L_+^\uparrow$.

In a Lie algebra g there exists a unique solvable ideal q which is maximal, i.e. if $q_1 \supset q$ is a solvable ideal in g then $q_1 = q$. This ideal is called the *radical* of g, denoted by rad g (see VARADARAJAN [10, p. 204]). Now g is solvable if and only if $g = \text{rad } g$; g is semisimple if and only if rad $g = 0$. (This last criterium is often used as a definition for semisimple Lie algebras.) In the intermediate situation we have:

THEOREM 3.3 (cf. VARADARAJAN [10, Theorem 3.14.1]). *In a Lie algebra g there is a subalgebra m such that $g = \text{rad } g + m$. Then m is semisimple.*

This decomposition of g is called a *Levi decomposition*. Since rad g is an ideal, any two subalgebra's m_1 and m_2 which may occur in the Levi decomposition are isomorphic. The Levi decomposition is used as a tool for the proof of Ado's theorem (every Lie algebra is a linear Lie algebra).

Next let G be a connected Lie group with Lie algebra g. Let $g = \text{rad } g + m$ be a Levi decomposition. Let Q and M be the connected Lie subgroups of G corresponding to the subalgebras rad g and m of g, respectively. Then Q is closed in G and it is a maximal solvable connected normal Lie subgroup of G, which uniquely exists. It is called the *radical* of the connected Lie group G. Then each $g \in G$ can be written as $g = qm$ for some $q \in Q$ and $m \in M$:

THEOREM 3.4 (cf. VARADARAJAN [10, Theorem 3.18.3]). *Let G be a connected Lie group. Then there are connected Lie subgroups Q and M of G such that Q is a normal subgroup which is solvable and closed in G, M is semisimple and for each $g \in G$ there are $q \in Q$ and $m \in M$ with $g = qm$. Furthermore, if G is also simply connected then there are Q and M with the additional properties that M is closed in G and the mapping $(q,m) \to qm$ is an analytic diffeomorphism from $Q \times M$ onto G.*

3.4. Tables of some important linear Lie groups and Lie algebras

Table 1 below contains the symbols and definitions of some important linear Lie groups and Table 2 gives similar information for the corresponding Lie algebras. Throughout, $\mathbb{F}$ denotes either $\mathbb{R}$ or $\mathbb{C}$.

name	definition
$GL(n,\mathbb{F})$	nonsingular n×n matrices over $\mathbb{F}$
$SL(n,\mathbb{F})$	$\{T \in GL(n,\mathbb{F}) \mid \det T = 1\}$
$U(n)$	$\{T \in GL(n,\mathbb{C}) \mid T^*T = I\}$
$SU(n)$	$\{T \in U(n) \mid \det T = 1\}$
$O(n)$	$\{T \in GL(n,\mathbb{R}) \mid T^tT = I\}$
$SO(n)$	$\{T \in O(n) \mid \det T = 1\}$
$T(n,\mathbb{F})$	$\{T \in GL(n,\mathbb{F}) \mid T_{ij} = 0 \text{ for } i > j\}$
$ST(n,\mathbb{F})$	$\{T \in T(n,\mathbb{F}) \mid \det T = 1\}$
$T_0(n,\mathbb{F})$	$\{T \in T(n,\mathbb{F}) \mid T_{11} = \dots = T_{nn} = 1\}$
$D(n,\mathbb{F})$	$\{T \in GL(n,\mathbb{F}) \mid T_{ij} = 0 \text{ for } i \neq j\}$
$SD(n,\mathbb{F})$	$\{T \in D(n,\mathbb{F}) \mid \det T = 1\}$

Table 1.

name	definition
$M_n(\mathbb{F})$	n×n matrices over $\mathbb{F}$
$\mathfrak{sl}(n,\mathbb{F})$	$\{A \in M_n(\mathbb{F}) \mid \operatorname{tr} A = 0\}$
$\mathfrak{u}(n)$	$\{A \in M_n(\mathbb{C}) \mid A^*+A = 0\}$
$\mathfrak{su}(n)$	$\{A \in \mathfrak{u}(n) \mid \operatorname{tr} A = 0\}$
$\mathfrak{o}(n) = \mathfrak{so}(n)$	$\{A \in M_n(\mathbb{R}) \mid A^t+A = 0\}$
$\mathfrak{t}(n,\mathbb{F})$	$\{A \in M_n(\mathbb{F}) \mid A_{ij} = 0 \text{ for } i > j\}$
$\mathfrak{st}(n,\mathbb{F})$	$\{A \in \mathfrak{t}(n,\mathbb{F}) \mid \operatorname{tr} A = 0\}$
$\mathfrak{t}_0(n,\mathbb{F})$	$\{A \in \mathfrak{t}(n,\mathbb{F}) \mid A_{11} = \dots = A_{nn} = 0\}$
$\mathfrak{d}(n,\mathbb{F})$	$\{A \in M_n(\mathbb{F}) \mid A_{ij} = 0 \text{ for } i \neq j\}$
$\mathfrak{sd}(n,\mathbb{F})$	$\{A \in \mathfrak{d}(n,\mathbb{F}) \mid \operatorname{tr} A = 0\}$

Table 2.

LITERATURE

In the list below MILLER [7, Ch.5] is a rather elementary introduction to Lie groups and Lie algebras; WARNER [11], SAGLE & WALDE [9] and PONTRJAGIN [8] are textbooks on beginning graduate level; CHEVALLEY [2], VARADARAJAN [10] and HELGASON [4, Ch.2] are advanced works; finally KIRILLOV [6, §6] gives a concise introduction with most of the theorems being presented in the form of exercises. All these books except MILLER [7] contain a chapter on manifolds and tangent spaces. See also BRÖCKER & JÄNICH [1, §1 and 2] for this topic. HUMPHREYS [5] is a good introduction to Lie algebras.

[1] BRÖCKER, Th. & K. JANICH, *Einführung in die Differentialtopologie*, Springer-Verlag, Berlin, 1973.

[2] CHEVALLEY, C., *Theory of Lie groups*, Vol. I, Princeton University Press, Princeton, 1946.

[3] CODDINGTON, E.A. & N. LEVINSON, *Theory of ordinary differential equations*, McGraw-Hill, New York, 1955.

[4] HELGASON, S., *Differential geometry and symmetric spaces*, Academic Press, New York, 1962.

[5] HUMPHREYS, J.E., *Introduction to Lie algebras and representation theory*, Springer-Verlag, Berlin, 1972.

[6] KIRILLOV, A.A., *Elements of the theory of representations*, Springer-Verlag, Berlin, 1976.

[7] MILLER, W., Jr., *Symmetry groups and their applications*, Academic Press, New York, 1972.

[8] PONTRJAGIN, L.S., *Topological groups*, Princeton University Press, Princeton, 1958.

[9] SAGLE, A.A. & R.E. WALDE, *Introduction to Lie groups and Lie algebras*, Academic Press, New York, 1973.

[10] VARADARAJAN, V.S., *Lie groups, Lie algebras and their representations*, Prentice-Hall, Englewood Cliffs, N.J., 1974.

[11] WARNER, F.W., *Foundations of differentiable manifolds and Lie groups*, Scott, Foresman and Company, Glenview, Illinois, 1971.

V

INTEGRATION ON LOCALLY COMPACT GROUPS

J. DE VRIES

Mathematisch Centrum

CONTENTS

There are several approaches to integration theory, each with its own particular advantage, and for each there exist excellent textbooks. However, since there does not exist a universally accepted terminology, this diversity in methods makes it difficult to avoid mistakes if one wants to switch over from one textbook to another (for example, in order to apply a theorem from one book in the framework of another). In the past, there has been a deep controversy between the two main schools in integration theory (viz. the σ-algebra approach of, say HALMOS, and the approach by means of Radon measures à la BOURBAKI; cf.2.8 below). Although in recent textbooks the stress is more on the unity of integration theory than on the differences of both approaches, up to now (1978) both approaches are widely in use.

For integration on locally compact groups a BOURBAKI-like approach via Radon-measures (integrals as functionals on function spaces) would seem most natural, but we have chosen first to review the σ-algebra approach. So in Section 1, we present all definitions and most facts which are needed in this and the subsequent chapters. Since this section is only inserted for purposes of reference, we have not included here proofs, examples or motivational remarks. However, vector-valued integration is treated in some detail, partly because it shows how the preceding techniques can be applied, and partly because elementary textbooks do not treat this material. Then, in Section 2, we discuss integration on locally compact spaces: a version of the RIESZ Representation theorem is mentioned which characterizes integrals as certain functionals on spaces of continuous functions with compact support. Finally, in Section 3 we discuss the Haar integral on locally compact topological groups and Lie groups. (Note, that for this latter class of groups, we have neglected the approach via differential forms.) In order to simplify the presentation and to make our discussion compatible with most of the cited references, we restrict our attention in Sections 2 and 3 to σ-compact spaces (i.e. locally compact spaces which are unions of countably many compact subsets). Since all Lie groups are σ-compact and locally compact (cf. IV. 1.5) this does not restrict the applicability of this chapter.

The following notation will be used throughout this chapter: $\mathbb{R}_* :=$ *) $\mathbb{R} \cup \{-\infty,\infty\}$ with its usual ordering, topology and algebraic operations. Thus, for all $x \in \mathbb{R}$, $-\infty < x < \infty$, and in addition:

---

*) P := Q or Q =: P means that P is *defined* as Q.

$$x + (\pm\infty) = (\pm\infty) + x = \pm\infty; \ (\pm\infty) + (\pm\infty) = \pm\infty;$$

$$x\cdot(\pm\infty) = (\pm\infty)\cdot x = \pm\infty \text{ if } x > 0;$$

$$x\cdot(\pm\infty) = (\pm\infty)\cdot x = \mp\infty \text{ if } x < 0.$$

Expressions like $\infty + (-\infty)$ or $0\cdot(\pm\infty)$ are not defined. If $A \subseteq \mathbb{R}_*$, then $A^+ := \{a \in A | a \geq 0\}$; in particular, $\mathbb{R}_*^+ = \mathbb{R}^+ \cup \{\infty\} = \{a \in \mathbb{R}_* \mid a \geq 0\}$. All other notation is more or less standard. Intervals in $\mathbb{R}_*$ are defined in the usual way, for example, if $a,b \in \mathbb{R}_*$, then $(a,b] := \{t \in \mathbb{R}_* | a < t \leq b\}$, etc. Our use of arrows in the notation of functions is as follows: a function f, defined on a set X and with values in a set Y is denoted $f: X \to Y$ or $x \mapsto f(x): X \to Y$. It should be clear now what we mean by an expression like: consider the function $f: x \mapsto x^2: \mathbb{R} \to \mathbb{R}^+$. Words like "function", "mapping" or "transformation" will be used as synonyms.

1. MEASURES AND INTEGRALS

In this section we give all relevant definitions from the theory of measure and ingeration. Thus, in 1.1 through 1.3 we define concepts like (non-negative) measures, measurability and integrability of extended-real-valued functions. In 1.4 through 1.12 the most useful properties of integrals are collected, among others, Lebesgue's dominated convergence theorem (§1.5), definition and properties of L^p-spaces for $1 \leq p \leq \infty$ (§1.6), the Radon-Nikodym theorem for σ-finite measures (§1.9), the behaviour of measures and integrals under measurable transformations (§1.11), and Fubini's theorem for products of σ-finite measures (§1.12).

Until §1.12 all definitions and properties are stated only for real-valued functions and positive measures. In 1.13 we make a few remarks about complex-valued functions. In view of later applications, we define also the concept of a complex measure and we consider a few of its properties.

The final part of this section is devoted to measurability and integrability of functions with values in a Banach space or in a space of operators.

1.1. Measurable spaces and functions

Let X be a non-empty set. A collection $\mathcal{S}$ of subsets of X which is closed under countable unions and set theoretic differences, and which contains X as an element, will be called a *σ-algebra*, and the pair $(X,\mathcal{S})$ is then called a *measurable**) space (i.e. a space, able to carry a measure; not to be confused with a measure space, which concept we shall define in 1.2 below). The elements of S are called *S-measurable* sets or just *measurable* sets. Notice that $\emptyset$ and X are S-measurable for any σ-algebra S in X.

If E is a collection of subsets of X, then the intersection of the set of all σ-algebras in X in which E is included is a σ-algebra: it is the least σ-algebra in which E is included, and it is called the *σ-algebra generated by E*. In particular, the subsets of $\mathbb{R}$, respectively $\mathbb{R}_*$, which are elements of the σ-algebra in $\mathbb{R}$, respectively $\mathbb{R}_*$, generated by the collection of all open subsets of $\mathbb{R}$, respectively $\mathbb{R}_*$, are called *Borel measurable sets*, or just *Borel sets*, in $\mathbb{R}$, respectively $\mathbb{R}_*$ (see also 2.2 below). It is rather easy to show that the σ-algebra of Borel sets in $\mathbb{R}$, respectively $\mathbb{R}_*$, coincides with the σ-algebra generated by all open intervals in $\mathbb{R}$, respectively $\mathbb{R}_*$. (Notice that, for example, $(a,\infty]$ and $[-\infty,a)$ are open in $\mathbb{R}_*$ for every $a \in \mathbb{R}$.)

If (X,S) and (Y,T) are measurable spaces, then a function $f\colon X \to Y$ is called *measurable* whenever $f^{-1}[T] \subseteq S$, that is, $f^{-1}[B] \in S$ for every $B \in T$. Of particular interest is the case that $Y = \mathbb{R}_*$ and T is the σ-algebra of Borel sets in $\mathbb{R}_*$. When we speak of a S-measurable (or just measurable) function $f\colon X \to \mathbb{R}_*$ without specifying the σ-algebra in $\mathbb{R}_*$, we shall always mean the σ-algebra of Borel sets in $\mathbb{R}_*$. It is not difficult to show that a function $f\colon X \to \mathbb{R}_*$ is measurable if and only if $f^{-1}[U] \in S$ for every open interval U in $\mathbb{R}_*$.

A *simple function* is a measurable function $f\colon X \to \mathbb{R}$ which assumes only a finite number of (finite!) values. Simple functions are just the functions that can be written in the form $f = \Sigma_{i=1}^{n} a_i\chi_{A_i}$ with $A_i \in S$ and $a_i \in \mathbb{R}$. Here χ_B is the *characteristic function* of a subset B of X, that is, $\chi_B(x) = 1$ if $x \in B$ and $\chi_B(x) = 0$ if $x \notin B$. For the easy proof of the following statement, which is of fundamental importance, we refer to HEWITT & STROMBERG [4;11.35]:

---

*) In the literature also the words *Borel space* and *Borel sets* are used in the meaning of measurable space and measurable set, respectively. Instead of measurable function (to be defined below) the term Borel function is then used. We shall use these terms in another sense: for our definitions of the notions of Borel set and Borel function, see the following paragraph and also 2.2 below.

If $f: X \to \mathbb{R}_*^+$ *is a measurable function then there exists a non-decreasing sequence* $\{f_n\}$ *of non-negative simple functions which converges pointwise to* f; *that is, each of* f_n *is a simple function,* $0 \le f_n(x) \le f_{n+1}(x)$ *and* $\lim_{n\to\infty} f_n(x) = f(x)$ *for every* $x \in X$. Notation: $0 \le f_n \uparrow f$.
If f *is bounded, then the convergence may assumed to be uniform on* X.

1.2. Measures

If $(X,\mathcal{S})$ is a measurable space, then a *measure on* $\mathcal{S}$ is a function $\mu: \mathcal{S} \to \mathbb{R}_*^+$ with the following properties:

(i) $\mu(\emptyset) = 0$;

(ii) If $A_1, A_2, \ldots$ is a sequence of mutually disjoint elements of $\mathcal{S}$, then $\mu(\bigcup_{n=1}^{\infty} A_n) = \Sigma_{n=1}^{\infty} \mu(A_n)$.

Since all terms of the series in (ii) are non-negative, the order of summation is irrelevant; in case of divergence, the sum is understood to be $+\infty$. If $(X,\mathcal{S})$ is a measurable space and μ is a measure on $\mathcal{S}$, then the triple $(X,\mathcal{S},\mu)$ will be called a *measure space*.

A property P, applicable to points of a measure space $(X,\mathcal{S},\mu)$, is said to hold μ-*almost everywhere* whenever the set $\{x \in X | P(x) \text{ is false}\}$ is included in a null-set, that is, a set $A \in \mathcal{S}$ with $\mu(A) = 0$. Stated otherwise, P holds μ-almost everywhere if and only if there is a null-set $A \in \mathcal{S}$ such that P(x) is true for all $x \in X\backslash A$. As an abbreviation of "P holds μ-almost everywhere" we shall use "P a.e.[μ]". If P a.e.[μ], then the set $\{x \in X | P(x) \text{ is false}\}$ need not be a null-set itself, because it may be not an element of $\mathcal{S}$. However, if the measure space $(X,\mathcal{S},\mu)$ is such that all subsets of null-sets are measurable, then all subsets of null-sets are again null-sets. Such a measure is called *complete*. Thus, in a complete measure space, if P a.e.[μ], then the set $\{x \in X | P(x) \text{ is false}\}$ is a null-set.

Let $(X,\mathcal{S},\mu)$ be a measure space. The *completion* $(X,\bar{\mathcal{S}},\bar{\mu})$ of the given measure space is defined as follows. Let $\mathcal{N}$ denote the family of all subsets of (μ-) null-sets in X. Then $\bar{\mathcal{S}}$ is the σ-algebra, generated by $\mathcal{S} \cup \mathcal{N}$. It is not difficult to show that for a subset E of X, $E \in \bar{\mathcal{S}}$ if and only if there is $A \in \mathcal{S}$ such that $A \subseteq E$ and $E\backslash A \in \mathcal{N}$, if and only if there are $A,B \in \mathcal{S}$ with $A \subseteq E \subseteq B$ and $\mu(B\backslash A) = 0$. In that case, set $\bar{\mu}(E) := \mu(A) = \mu(B)$. Then $\bar{\mu}$ is well-defined, $\bar{\mu}$ is the unique measure on $\bar{\mathcal{S}}$ such that $\bar{\mu}|_{\mathcal{S}} = \mu$ and $\bar{\mu}(E) = 0$ for all $E \in \mathcal{N}$. Finally, $(X,\bar{\mathcal{S}},\bar{\mu})$ is a complete measure space. Sometimes we shall refer to the completion of a measure space $(X,\mathcal{S},\mu)$ without explicitly introducing a notation for it (like $\bar{\mathcal{S}}$ and $\bar{\mu}$). Thus, the elements of $\bar{\mathcal{S}}$ are called μ-*measurable sets,* and a function $f\colon X \to \mathbb{R}_*$ is called a μ-*measurable function* whenever it is an $\bar{S}$-measurable function. Clearly, every measurable*) function is μ-measurable, but not always conversely. However, it is an easy exercise to show that *a function* $f\colon X \to \mathbb{R}_*$ *is* μ-*measurable if and only if there is a measurable function* $g\colon X \to \mathbb{R}_*$ *such that* $f = g$ *a.e.* $[\mu]$.

1.3. Integrals

Let (X,S,μ) be a measure space (not necessarily complete), and let ϕ be a non-negative simple function on X. Then ϕ has a unique representation as $\phi = \Sigma_{i=1}^{n} a_i \chi_{A_i}$ such that $a_i > 0$ and $A_i = \{x \in X \mid \phi(x) = a_i\}$ for $i = 1,\dots,n$; notice that the sets A_i are mutually disjoint elements of S. The *integral of* ϕ is defined as the (possibly infinite) number

$$\int_X \phi d\mu := \sum_{i=1}^{n} a_i \mu(A_i).$$

If $f\colon X \to \mathbb{R}_*^+$ is a measurable function then there exists a sequence $\{\phi_n\}$ of simple functions such that $0 \le \phi_n \uparrow f$ (see the last paragraph in 1.1). Then *the integral of* f (which may be $+\infty$) is defined as the number

$$\int_X f d\mu := \sup_{n\in\mathbb{N}} \int_X \phi_n d\mu.$$

Its value turns out to be independent of the particular choice of the sequence of simple functions $\{\phi_n\}$(provided, of course, $0 \le \phi_n \uparrow f$). If $\int_X f d\mu < \infty$, we say that f is *integrable* (or sometimes: μ-*integrable*).

If $f\colon X \to \mathbb{R}_*$ is an arbitrary measurable function, then we can write $f = f^+ - f^-$, where

$$f^+(x) := \max\{f(x),0\};\quad f^-(x) := \max\{-f(x),0\}$$

for every $x \in X$. Then both f^+ and f^- can be shown to be measurable, so that their integrals exist by the above definition. If f^+ and f^- are both integrable (i.e. have finite integrals) then we say that f is *integrable*, and the (finite!) number

---

*) "measurable" without prefix will always mean "S-measurable".

$$\int_X f d\mu := \int_X f^+ d\mu - \int_X f^- d\mu$$

is called *the integral of* f. Instead of $\int_X f d\mu$ we shall mostly write $\int f d\mu$. Sometimes it is advisable to write $\int_X f(x)d\mu(x)$ or $\int f(x)d\mu(x)$. It should be noticed, that for every non-negative measurable function the integral is defined, but only in the case that the integral has a finite value the function is called integrable.

If $f\colon X \to \mathbb{R}_*$ is integrable, then the set $\{x \in X \mid f(x) \in \{-\infty,\infty\}\}$ is measurable and has measure zero, i.e. , it is a null-set (cf. HEWITT & STROMBERG [4; 12.15]). It follows easily that there exists an integrable function $f'\colon X \to \mathbb{R}$ (finite-valued!) such that $|f'| \le |f|$ and $f' = f$ a.e. $[\mu]$ (so that, by 1.4(ii) below, $\int_X f d\mu = \int_X f' d\mu$).

1.4. <u>The space $L^1(X,\mu)$ of all integrable functions</u>

The set of all integrable functions defined on a measure space (X,S,μ) with values in $\mathbb{R}_*$ will be denoted by $L^1(X,\mu)$, and sometimes by $L^1(\mu)$ or $L^1(X)$, and when X and μ are understood, we shall simply write L^1. The following statements about L^1 are fundamental:

(i) Under pointwise addition and scalar multiplication, L^1 is a (usually infinite-dimensional) vector space over $\mathbb{R}$, and the mapping $f \mapsto \int f d\mu\colon L^1 \to \mathbb{R}$ is linear.

(ii) The linear mapping $f \mapsto \int f d\mu$ is positive on L^1, that is, if $f \in L^1$ and $f \ge 0$ a.e. $[\mu]$, then $\int f d\mu \ge 0$. Consequently, if $f,g \in L^1$ and $f \le g$ a.e. $[\mu]$, then $\int f d\mu \le \int g d\mu$. In particular, if $f,g \in L^1$ and $f = g$ a.e. $[\mu]$, then $\int f d\mu = \int g d\mu$.

(iii) If $f \in L^1$, $f \ge 0$ a.e. $[\mu]$ and $\int f d\mu = 0$, then $f = 0$ a.e. $[\mu]$. Consequently, if $f\colon X \to \mathbb{R}_*$ is any measurable function such that $\int \chi_A f d\mu = 0$ for every $A \in S$, then $f = 0$ a.e. $[\mu]$.

(iv) For every measurable function $f\colon X \to \mathbb{R}_*$ we have $f \in L^1$ if and only if $|f| \in L^1$, and in that case we have

$$\left|\int f d\mu\right| \le \int |f| d\mu.$$

(v) If $f\colon X \to \mathbb{R}_*$ is measurable and if there exists $g \in L^1$, $g \ge 0$, such that $|f| \le g$ a.e. $[\mu]$, then $f \in L^1$.

REMARK. Since $(+\infty) + (-\infty)$ is not defined, addition of integrable functions might cause difficulties. Therefore, the pointwise addition in (i) above should be understood as follows. If $f_1, f_2 \in L^1$, then by the final remark in 1.3, there are integrable functions f_1' and f_2' of X into $\mathbb{R}$ such that $f_i' = f_i$ a.e. $[\mu]$ for $i = 1,2$. The function $f_1' + f_2'$ is defined everywhere on X, and (i) above states that it is an element of L^1. Notice that $f_1(x) + f_2(x)$ is defined and equals $f_1'(x) + f_2'(x)$ for almost every $x \in X$. Therefore, we define $f_1 + f_2 := f_1' + f_2'$ on all of X. Notice that $f_1 + f_2$ is *almost everywhere* uniquely defined, and that its definition depends on the more or less arbitrary choice of f_1' and f_2'. This ambiguity will be removed in 1.6 below.

1.5. Convergence theorems

Let (X,S,μ) be a measure space and let $\{f_n\}$ be a sequence of measurable functions on X with values in $\mathbb{R}_*$. Then:

(i) FATOU's LEMMA: *If each* $f_n \geq 0$, *then*

$$\int (\liminf f_n)\,d\mu \leq \liminf \int f_n\,d\mu \;.$$

[Here $\liminf f_n$ can easily be shown to be a non-negative measurable function so that both sides of the inequality are defined (but possibly $+\infty$).]

(ii) MONOTONE CONVERGENCE THEOREM: *If* $0 \leq f_n \uparrow f$, *then* f *is a measurable function, and*

$$\int f\,d\mu = \lim_{n\to\infty} \int f_n\,d\mu \;.$$

[Here $f_n \uparrow f$ means that for every $x \in X$, $f_n(x) \leq f_{n+1}(x)$ for all $n \in \mathbb{N}$, and $f(x) = \lim_{n\to\infty} f_n(x)$.]

(iii) LEBESGUE's DOMINATED CONVERGENCE THEOREM: *If the sequence* $\{f_n\}$ *converges pointwise to a function* $f\colon X \to \mathbb{R}_*$, *and if there exists* $g \in L^1$ *such that* $|f_n| \leq g$ *a.e.* $[\mu]$, *then* $f \in L^1$, *and* $\lim_{n\to\infty} \int |f - f_n|\,d\mu = 0$. *In particular,*

$$\int f\,d\mu = \lim_{n\to\infty} \int f_n\,d\mu \;.$$

For proofs of these statements, see HEWITT & STROMBERG [4; § 12].

1.6. The space $\mathcal{L}^p$

Let $(X,\mathcal{S},\mu)$ be a measure space and let $p \in \mathbb{R}$, $1 \le p < \infty$. Define $\mathcal{L}^p = \mathcal{L}^p(X,\mu)$ as the space of all measurable functions $f\colon X \to \mathbb{R}_*$ for which the expression

$$(1.1) \qquad \|f\|_p := \left(\int |f|^p d\mu\right)^{1/p}$$

is finite. For $p = 1$, this definition is in accordance with 1.4 (see also 1.4(iv)), and clearly a measurable function f is in $\mathcal{L}^p$ if and only if $|f|^p \in \mathcal{L}^1$. Define an equivalence relation ~ in $\mathcal{L}^p$ by

$$f \sim g \quad \text{if and only if} \quad f = g \text{ a.e.}[\mu].$$

Observe, that $f \sim g$ in $\mathcal{L}^p$ if and only if $|f|^p \sim |g|^p$ in $\mathcal{L}^1$. Denote the equivalence class of $f \in \mathcal{L}^p$ by $[f]$. By the final remark in 1.3, every equivalence class $[f]$ contains a finite-valued member f', that is, $f' \in \mathcal{L}^p$, $f'(x)$ is finite for every $x \in X$, and $f' \in [f]$, hence $[f'] = [f]$. Thus every equivalence class can be labeled by a finite-valued representant. It can be shown that for finite-valued elements f and g of $\mathcal{L}^p$, $\alpha f + \beta g \in \mathcal{L}^p$ for all $\alpha,\beta \in \mathbb{R}$, and that the class $[\alpha f + \beta g]$ does not depend on the choice of the finite-valued representants of the classes $[f]$ and $[g]$. Hence by $\alpha[f] + \beta[g] := [\alpha f+\beta g]$, addition and scalar multiplication of equivalence classes in $\mathcal{L}^p$ can be defined in an unambiguous way, and *the set of all equivalence classes, equipped with these linear operations, is a vector space*. This vector space will be denoted $L^p(X,\mathcal{S},\mu)$, or just L^p when no confusion is likely to arise (sometimes, we write $L^p(\mu)$ or $L^p(X)$). It follows from 1.4(ii) that $\|f\|_p = \|g\|_p$ whenever $f \in \mathcal{L}^p$ and $g \in [f]$. Therefore we can unambiguously define $\|[f]\|_p := \|f\|_p$ for every $[f] \in L^p$. In order to facilitate notation, we shall follow from now on the generally accepted convention to denote the elements of L^p (which are equivalence classes of functions) so as if they where functions themselves (on X, of course). Thus, two functions f and g in $\mathcal{L}^p$ are distinct as elements of L^p if and only if they differ from each other on a set of positive measure.

The mapping $f \mapsto \|f\|_p$ which is, as we have noticed above, well-defined on L^p, turns out to be a *norm* on the vector space L^p. In particular, for $f \in L^p$ we have $f = 0$ (as an element of L^p) if and only if $\|f\|_p = 0$; this is an easy consequence of 1.4(ii) and 1.4(iii). Moreover, the triangle inequality $\|f+g\|_p \le \|f\|_p+\|g\|_p$ (which is known in this particular case as the

MINKOWSKI inequality) is valid for all $f,g \in L^p$: see HEWITT & STROMBERG [4;13.7]. *With this norm,* L_p *turns out to be a Banach space* (*i.e. a complete normed space*).

At this place it seems appropriate to stress the fact that convergence of a sequence in a space L^p is convergence with respect to the aforementioned norm. Thus, if $\{f_n\}$ is a sequence in L^p, then $f = \lim_{n\to\infty} f_n$ means that $f \in L^p$ and $\lim_{n\to\infty} \|f-f_n\|_p = 0$, that is,

$$\lim_{n\to\infty} \int |f(x)-f_n(x)|^p d\mu(x) = 0.$$

Similarly, a sequence $\{f_n\}$ in L^p is a *Cauchy sequence* whenever for every $\varepsilon > 0$ there exists $N_\varepsilon > 0$ such that

$$\int |f_n(x)-f_m(x)|^p d\mu(x) < \varepsilon \quad \text{for all } n,m \ge N_\varepsilon.$$

[Recall, that *completeness* of L^p means, that every Cauchy sequence in L^p has a limit in L^p according to the definition given above.] Also, a series $\Sigma_{k=1}^\infty f_k$ is said to *converge* in L^p with sum f whenever $f = \lim_{n\to\infty} \Sigma_{k=1}^n f_k$, that is,

$$\lim_{n\to\infty} \int |f(x) - \sum_{k=1}^{n} f_k(x)|^p d\mu(x) = 0.$$

<u>CAUTION</u>: if $f = \lim_{n\to\infty} f_n$ in L^p then it is, in general, not true that $f(x) = \lim_{n\to\infty} f_n(x)$ in $\mathbb{R}_*$ for any given $x \in X$. However, it can be shown (cf. HEWITT & STROMBERG [4;pp.192,193]) that *if a sequence* $\{f_n\}$ *converges in* L^p (*with respect to the norm* $\|\cdot\|_p$) *to an element* $f \in L^p$, *then there is a subsequence* $\{f_{n_k}\}$ *such that* $f(x) = \lim_{k\to\infty} f_{n_k}(x)$ *a.e.*[μ].

1.7. <u>The dual space of L^p</u>

Let $(X,\mathcal{S},\mu)$ be a measure space, and let $1 < p < \infty$, $1 < q < \infty$, and $p^{-1} + q^{-1} = 1$. Then for every $f \in L^p$ and $g \in L^q$ we have $fg \in L^1$ and

$$(1.2) \qquad \left|\int fg d\mu\right| \le \|f\|_p \|g\|_q$$

(HÖLDER's inequality). [The definition of fg, which has difficulties which are quite similar to those in the definition of a sum in L^p, is left to the reader.] It follows from 1.2 that the mapping

$$\phi_g\colon f \mapsto \int fgd\mu\colon L^p \to \mathbb{R}$$

is a bounded linear functional on the Banach space L^p, and that $\|\phi_g\| := \sup\{|\phi_g(f)| \mid f \in L^p \ \&\ \|f\|_p \le 1\} \le \|g\|_q$. In fact, it can be shown that $\|\phi_g\| = \|g\|_q$ for all $g \in L^q$, and that, in addition, $g \mapsto \phi_g$ is a linear, norm preserving mapping of the space L^q *onto* the dual space $(L^p)^*$ of L^p (thus, every bounded linear functional on L^p can be represented as ϕ_g for a unique element $g \in L^q$). See HEWITT & STROMBERG [4;§15].

It follows from HÖLDER's *inequality for the case* $p = q = 2$ *that* L^2 *is a Hilbert space, with inner product defined by*

$$(1.3)\qquad (f,g) := \int fgd\mu$$

for $f,g \in L^2$ (the completeness of L^2 is known as the RIESZ-FISCHER theorem). It is instructive, to re-read 1.6 and 1.7 with p replaced by 2; we leave this to the reader. Also, 1.6 should be re-read with p replaced by 1; in order to formulate 1.7 with $p = 1$, we have to define the space L^∞.

Let $\mathcal{L}^\infty$ be the space of all measurable functions f which are *essentially bounded*, that is, the expression

$$\|f\|_\infty := \inf\{c > 0 \mid |f(x)| \le c \text{ a.e.}[\mu]\}$$

is finite. Similar to 1.6, the space L^∞ is defined as the collection of all equivalence classes of functions in $\mathcal{L}^\infty$ which are equal a.e.[μ], and, again similar to the case $1 \le p < \infty$, L^∞ can be given the structure of a normed vector space with norm $\|\cdot\|_\infty$. The space L^∞ turns out to be complete, that is, it is a Banach space. It is not difficult to prove that, for all $f \in L^1$ and $g \in L^\infty$, we have $fg \in L^1$, and

$$\left|\int fgd\mu\right| \le \|g\|_\infty \int |f|d\mu = \|g\|_\infty \|f\|_1.$$

It follows, that the mapping

$$\phi_g\colon f \mapsto \int fgd\mu\colon L^1 \to \mathbb{R}$$

is a bounded, linear functional on the Banach space L^1, and $\|\phi_g\| = \|g\|_\infty$. It is easily seen that $g \mapsto \phi_g$ is a norm preserving linear mapping of L^∞ into the dual space $(L^1)^*$ of L^1. Using the RADON-NIKODYM theorem (cf. 1.9 below) it can be shown that this mapping is *onto* (i.e. every bounded linear functional on L^1 can be represented as a ϕ_g with unique $g \in L^\infty$), provided the measure μ is σ-finite; cf. HEWITT & STROMBERG [4;20.20]. Here a measure μ (also: the measure space $(X,\mathcal{S},\mu)$) is called *σ-finite* whenever $X = \bigcup_{n=1}^{\infty} A_n$ with each $A_n \in \mathcal{S}$ and $\mu(A_n) < \infty$. (In the reference, given above, the proof is given for so-called decomposable measure spaces, but every σ-finite measure space is decomposable.)

1.8. Dominated and equivalent measures

Let $(X,\mathcal{S},\mu)$ be a measure space, and consider a second measure ν, defined on a σ-algebra $\mathcal{T}$ in X. Then ν is said to be *dominated by* μ whenever $\mathcal{T} \supseteq \mathcal{S}$ and, moreover, for every $E \in \mathcal{S}$ with $\mu(E) = 0$ also $\nu(E) = 0$. Notation: $\nu \ll \mu$ (some authors call this "absolute continuity"). If $\mu \ll \nu$ and $\nu \ll \mu$, then μ and ν are called *equivalent*; notation: $\mu \equiv \nu$.
An example of domination is obtained as follows: let $f: X \to \mathbb{R}_*^+$ be an $\mathcal{S}$-measurable function. Define a mapping $\nu: \mathcal{S} \to \mathbb{R}_*^+$ by setting

$$(1.4) \qquad \nu(E) := \int \chi_E f d\mu \qquad \text{for } E \in \mathcal{S}.$$

Then ν is a measure on $\mathcal{S}$, and $\nu \ll \mu$. The following theorem states, that in the case of a σ-finite measure μ all dominated measures are obtained in this way.

1.9. The Radon-Nikodym theorem

THEOREM. *Let $(X,\mathcal{S},\mu)$ be a σ-finite measure space and let ν be any σ-finite measure on $\mathcal{S}$ such that $\nu \ll \mu$. Then there exists a finite-valued measurable function $f: X \to \mathbb{R}^+$ with the following properties:*

(i) $\forall E \in \mathcal{S} : \nu(E) = \int \chi_E f d\mu$.

(ii) *For every $\mathcal{S}$-measurable function* $g: X \to \mathbb{R}_*^+$,

$$\int g d\nu = \int g f d\mu.$$

(iii) *For every $\mathcal{S}$-measurable function* $g: X \to \mathbb{R}_*$, *one has* $g \in \mathcal{L}^1(\nu)$ *if and only if* $gf \in \mathcal{L}^1(\mu)$, *and, for every* $g \in \mathcal{L}^1(\nu)$,

$$\int g d\nu = \int g f d\mu.$$

The function f is unique in the following sense: if f_1 is also a measurable function such that $\nu(E) = \int \chi_E f_1 d\mu$ for every $E \in S$, then $f_1 = f$ a.e.[μ]. In addition, $f \in L^1(\mu)$ if and only if ν is a finite measure *).

For a proof of this theorem, we refer to HEWITT & STROMBERG [4; Corollary 19.28]. The function f is sometimes called the *Radon-Nikodym derivative of* ν *with respect to* μ. Notation: $f = d\nu/d\mu$. If we write $\int f(x)d\mu(x)$ instead of $\int f d\mu$, then the formulas in (i) and (iii) of the Radon-Nikodym theorem are to be written as

$$\nu(E) = \int \chi_E(x)f(x)d\mu(x) \text{ and } \int g(x)d\nu(x) = \int g(x)f(x)d\mu(x),$$

respectively. This can also be expressed by writing symbolically

$$d\nu(x) = f(x)d\mu(x).$$

The following "chain rule" is valid in this context, which is an easy consequence of the uniqueness of the Radon-Nikodym derivative: let μ, ν and λ be σ-finite measures (on S, say), and let $\lambda \ll \nu$ and $\nu \ll \mu$. Then $\lambda \ll \mu$ and

$$\frac{d\lambda}{d\mu} = \frac{d\lambda}{d\nu} \cdot \frac{d\nu}{d\mu} \text{ a.e.}[\mu] .$$

Or in the other notation: if $d\nu(x) = f(x)d\mu(x)$ and $d\lambda(x) = h(x)d\nu(x)$, then

$$d\lambda(x) = h(x)f(x)d\mu(x) .$$

1.10. A criterion for equivalence of measures

Let (X,S,μ) and ν be as in 1.9. In particular, $\nu \ll \mu$; set $f := d\nu/d\mu$. If $f(x) > 0$ for every $x \in X$, then by 1.4(iii) and 1.9(i), we have $\mu \ll \nu$, so that $\mu \equiv \nu$. Conversely, if $\mu \equiv \nu$, say $\nu(x) = f(x)d\mu(x)$ and $\mu(x) = g(x)d\nu(x)$, then $f(x)g(x) = 1$ for μ-almost every $x \in X$ (chain rule!). Hence $f(x) > 0$ a.e.[μ], and using 1.4(ii) we may assume that $f(x) > 0$ for for every $x \in X$. Thus, we have shown that $\mu \equiv \nu$ if and only if $d\nu(x) = f(x)d\mu(x)$ with f a measurable function such that $f(x) > 0$ for every $x \in X$.

*) This means, of course, that $\nu(X) < \infty$.

1.11. The behaviour of integrals under a measurable transformation

Let $(X,\mathcal{S},\mu)$ be a measure space and let $T: (X,\mathcal{S}) \to (X,\mathcal{S})$ be a measurable transformation (not necessarily bijective). Thus for every $A \in \mathcal{S}$ we have $T^{-1}[A] \in \mathcal{S}$. So we can define a mapping $\mu_T : \mathcal{S} \to \mathbb{R}^+_*$ by setting $\mu_T[A] := \mu(T^{-1}[A])$ for $A \in \mathcal{S}$. *Then* μ_T *is a measure on* $\mathcal{S}$, *and for every* $\mathcal{S}$*-measurable function* $f: X \to \mathbb{R}^+_*$ *it turns out that*

$$(1.5) \qquad \int f d\mu_T = \int f \circ T d\mu .$$

Moreover, a measurable function $f: X \to \mathbb{R}_*$ *is* μ_T*-integrable if and only if* $f \circ T$ *is* μ*-integrable and, again, the above formula holds.* (See HALMOS [1;§39,Thm.D] for a proof.)

Of particular interest is the case that the transformation T is such that $\mu(A) = 0$ implies $\mu(T^{-1}[A]) = 0$, that is, such that $\mu_T \ll \mu$. In that case, if we assume μ and μ_T to be σ-finite*), there exists by the Radon-Nikodym theorem a non-negative, finite-valued measurable function $R_T: X \to \mathbb{R}^+$ such that $d\mu_T(x) = R_T(x)d\mu(x)$, that is,

$$(1.6) \qquad \int f(x)R_T(x)d\mu(x) = \int f(T(x))d\mu(x)$$

for every $f \in L^1(\mu)$. The measures μ_T and μ have exactly the same sets of measure zero, that is $\mu_T \equiv \mu$, if and only if $R_T(x) > 0$ for every $x \in X$ (see 1.10).

1.12. Product measures and the Fubini-Tonelli theorem

Let $(X,\mathcal{S},\mu)$ and $(Y,\mathcal{T},\nu)$ be two σ-finite measure spaces, and let $\mathcal{S} \otimes \mathcal{T}$ denote the σ-algebra of subsets of $X \times Y$, generated by the family of all sets $A \times B$ with $A \in \mathcal{S}$, $B \in \mathcal{T}$. Then there exists a unique measure on $\mathcal{S} \otimes \mathcal{T}$, which we shall denote by $\mu \otimes \nu$, such that

$$(\mu\otimes\nu)(A\times B) = \mu(A)\nu(B)$$

for $A \in \mathcal{S}$, $B \in \mathcal{T}$. The measure $\mu \otimes \nu$ is called the *product measure* of μ and

---

*) If μ is σ-finite and T is *bijective*, then μ_T is certainly σ-finite: if $X = \bigcup_{n=1}^{\infty} A_n$ with each $\mu(A_n) < \infty$, then setting $B_n := T[A_n]$, we have $X = \bigcup_{n=1}^{\infty} B_n$ and $\mu_T(B_n) = \mu(A_n) < \infty$.

ν. The following notational conventions will be convenient.

If $f\colon X \times Y \to R_*$ is a mapping, then we say that the *iterated integral* $\int_X(\int_Y f(x,y)d\nu(y))d\mu(x)$ exists whenever there exist a null-set N in X and a mapping $h\colon X \to R_*$ such that:

(i) For every $x \in X \setminus N$ the integral $\int_Y f(x,y)d\nu(y)$ exists (either because the function $y \mapsto f(x,y)$ is ν-integrable, or because it is non-negative and $\mathcal{T}$-measurable) and $h(x) = \int_Y f(x,y)d\nu(y)$.

(ii) The integral $\int_X h(x)d\mu(x)$ exists (either because h is μ-integrable or because h is $\mathcal{S}$-measurable and non-negative).

In that case, the value of the iterated integral is, by definition, $\int_X h(x)d\mu(x)$.

In a similar way the iterated integral $\int_Y(\int_X f(x,y)d\mu(x))d\nu(y)$ is defined. For a proof of the following theorem see HEWITT & STROMBERG [4;pp. 379-387]. With the above notation and terminology we have (for σ-finite measures μ and ν):

THEOREM (FUBINI-TONELLI). *Let* $f\colon X \times Y \to R_*$ *be an* $\mathcal{S} \times \mathcal{T}$*-measurable function. Then all functions* $y \mapsto f(x_0,y)$, $x_0 \in X$ *are* $\mathcal{T}$*-measurable, and all functions* $x \mapsto f(x,y_0)$, $y_0 \in Y$, *are* $\mathcal{S}$*-measurable. Moreover,*

(i) *If f is non-negative then both iterated integrals exist, and*

$$\int_{X\times Y} f d\mu \otimes \nu = \int_X \left(\int_Y f(x,y)d\nu(y)\right)d\mu(x) = \int_Y \left(\int_X f(x,y)d\mu(x)\right)d\nu(y).$$

In particular, if one of the iterated integrals is finite, then f is $\mu \otimes \nu$*-integrable (TONELLI's theorem).*

(ii) *If f is* $\mu \otimes \nu$*-integrable, then both iterated integrals exist and are equal to* $\int f d\mu\otimes\nu$ *(FUBINI's theorem).*

1.13. Complex-valued functions; signed and complex measures

A few remarks about *complex-valued functions* on a measure space $(X,\mathcal{S},\mu)$. If $f\colon X \to \mathbb{C}$ is a function, then two real-valued functions $\mathcal{Re}f$ and $\mathcal{Im}f\colon X \to \mathbb{R}$ are defined by setting $f(x) = \mathcal{Re}f(x)+i\,\mathcal{Im}f(x)$ for $x \in X$. Then f is said to be *$\mathcal{S}$-measurable, μ-measurable,* or *μ-integrable* whenever both $\mathcal{Re}f$ and $\mathcal{Im}f$ are so. If $f\colon X \to \mathbb{C}$ is μ-integrable, then we define

$$\int f(x)d\mu(x) := \int \mathcal{Re}f(x)d\mu(x) + i \int \mathcal{Im}f(x)d\mu(x),$$

where both integrals in the right-hand side do exist. With these concepts of measurability and integrability all of the preceding theory remains valid for complex-valued functions. Of course, some obvious modifications must be made, e.g. L^p for $1 \leq p \leq \infty$ is now a *complex* Banach space, and L^2 is a complex Hilbert space if the inner product is defined by

$$(f,g) := \int f(x)\overline{g(x)}\,d\mu(x) \quad \text{for } f, g \in L^2,$$

where the bar denotes complex conjugation.

Next, we come to the generalisation of the preceding theory to arbitrary real-valued and complex-valued measures.

Let (X,S) be a measurable space. A mapping $\mu : S \to \mathbb{R}$ is called a *signed measure* whenever $\mu(\emptyset) = 0$ and μ is σ-additive, that is, if $A_1, A_2, \ldots$ is a sequence of mutually disjoint elements of S, then the series $\Sigma_{n=1}^{\infty} \mu(A_n)$ converges absolutely in $\mathbb{R}$, and

$$\sum_{n=1}^{\infty} \mu(A_n) = \mu(\bigcup_{n=1}^{\infty} A_n).$$

Clearly, if μ_1 and μ_2 are finite (non-negative) measures on S and a, b are arbitrary real numbers, then $\mu(A) := a\mu_1(A) + b\mu_2(A)$, $A \in S$, defines a signed measure on S. *Conversely, it can be shown that every signed measure* $\mu : S \to \mathbb{R}$ *is the difference of two non-negative measures* μ^+ *and* μ^- *on* S:

$$\mu(A) = \mu^+(A) - \mu^-(A),$$

where

$$\mu^+(A) := \sup\{\mu(B) \mid B \in S \;\&\; B \subseteq A\},$$

$$\mu^-(A) := -\inf\{\mu(B) \mid B \in S \;\&\; B \subseteq A\}$$

(JORDAN decomposition). The measures μ^+ and μ^- are called the *positive* and *negative variations of* μ, respectively, and the measure $|\mu| := \mu^+ + \mu^-$ is called the *total variation of* μ. It can be shown that μ^+ and μ^- are the unique non-negative measures on S such that

(i) $\mu = \mu^+ - \mu^-$,

(ii) $\mu^+ \perp \mu^-$, that is, there exists $E \in S$ such that $\mu^+(E) = \mu^-(X \backslash E) = 0$.

(for details, cf.[9;Ch.11]). Using this, it is rather easy to see that the total variation $|\mu|$ satisfies the relation

$$|\mu|(A) = \sup\{\sum_{n=1}^{\infty} |\mu(A_n)| \mid A_n \in S,\ A_n \cap A_m = \emptyset \text{ for } n \neq m,\ \bigcup_{n=1}^{\infty} A_n = A\}$$

for every $A \in S$.

If μ is a signed measure on S, then for arbitrary real- or complex-valued functions f on X we define

$$\int f d\mu := \int f d\mu^+ - \int f d\mu^-$$

provided $f \in L^1(X,\mu^+) \cap L^1(X,\mu^-)$. Note, that measurability and boundedness of f is sufficient for the existence of $\int f d\mu$. Obviously, this integral with respect to the signed measure μ has the usual linear properties. In addition, one has

$$\left|\int f d\mu\right| \leq \left|\int f d\mu^+\right| + \left|\int f d\mu^-\right| \leq \int |f| d(\mu^+ + \mu^-) = \int |f| d|\mu|.$$

Finally, we come to the defintion of a complex measure. A *complex measure* on S is a mapping $\mu : S \to \mathbb{C}$ which has the property that $\mu(\emptyset) = 0$ and which is σ-additive, that is, if $A_1, A_2, \ldots$ is a disjoint sequence of members of S, then the series $\Sigma_{n=1}^{\infty} \mu(A_n)$ converges absolutely, and

$$\mu(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n).$$

If μ is a complex measure on S, then two signed measures μ_1 and μ_2 are defined by

$$\mu_1(A) := \mathcal{Re}\,\mu(A), \quad \mu_2(A) := \mathcal{Im}\,\mu(A), \quad A \in S,$$

and clearly $\mu = \mu_1 + i\mu_2$. Conversely, if μ_1 and μ_2 are signed measures on S, then $\mu_1 + i\mu_2$ is a complex measure on S. The *total variation* of a complex measure is defined by the formula

$$|\mu|(A) := \sup\{\sum_{n=1}^{\infty} |\mu(A_n)| \mid A_n \in S,\ A_n \cap A_m = \emptyset \text{ for } n \neq m,\ \bigcup_{n=1}^{\infty} A_n = A\}$$

for $A \in S$. It turns out that $|\mu|$ is a finite measure on S.

If $\mu = \mu_1 + i\mu_2$ is a complex measure, then for every measurable function $f : X \to \mathbb{C}$ for which $\int f d\mu_1$ and $\int f d\mu_2$ both exist, the integral of f with respect to μ is defined by

$$\int f d\mu : = \int f d\mu_1 + i \int f d\mu_2 =$$
$$= \int f d\mu_1^+ - \int f d\mu_1^- + i \int f d\mu_2^+ - i \int f d\mu_2^- .$$

Again, the integral with respect to a complex measure has the usual linear properties and, in addition, it can be shown that

$$\left| \int f d\mu \right| \le \int |f| d|\mu| .$$

1.14. Vector-valued integrals

We shall discuss now some elementary first principles of integration of vector-valued functions with respect to non-negative measures.

Let $(X, \mathcal{S}, \mu)$ be a measure space, let E be a complex Banach space[1)] and let E^* be its dual (that is, E^* is the space of all continuous linear functionals). A function $f: X \to E$ is called *weakly measurable* whenever $\phi \circ f: X \to \mathbb{C}$ is measurable for every $\phi \in E^*$. If $f: X \to E$ has the property that $\phi \circ f \in L^1(\mu)$ for every $\phi \in E^*$ and, in addition, there exists $\xi \in E$ such that $\langle \xi, \phi \rangle = \int \langle f(x), \phi \rangle d\mu(x)$ for every $\phi \in E^*$, then f is said to be *weakly integrable* [2)]. Since E^* separates the points of E by the Hahn-Banach Theorem, the vector ξ (if it exists) is determined uniquely. It will be denoted $\int f d\mu$. Thus, if f is weakly integrable, then by definition:

$$(1.7) \qquad \left\langle \int f d\mu, \phi \right\rangle = \int \langle f(x), \phi \rangle d\mu(x), \qquad \phi \in E^* .$$

Clearly, if $f: X \to E$ is weakly integrable, then it is weakly measurable. Moreover, if also $g: X \to E$ is weakly integrable then so is $af + bg$ $(a, b \in \mathbb{C})$, and $\int (af+bg) d\mu = a\int d\mu + b\int g d\mu$.

---

1) For real Banach spaces, the definitions are similar. Much of what we do here can also be done for locally convex spaces.

2) We shall use the following notation: $\langle \xi, \phi \rangle : = \phi(\xi)$ for $\xi \in E$, $\phi \in E^*$.

Let F be a second complex Banach space and $T: E \to F$ a continuous linear mapping. If $T^*: F^* \to E^*$ is the dual mapping *), then for any weakly integrable function $f: X \to E$ we have in virtue of the relation $\langle Tf(x),\psi\rangle = \langle f(x),T^*\psi\rangle$ ($\psi \in F^*$, hence $T^*\psi \in E^*$) that $x \mapsto \langle Tf(x),\psi\rangle: X \to \mathbb{C}$ is integrable, and

$$\langle T\left(\int f d\mu\right),\psi\rangle = \langle\int f d\mu, T^*\psi\rangle = \int\langle f(x),T^*\psi\rangle d\mu(x) = \int\langle Tf(x),\psi\rangle d\mu(x).$$

This shows that $T\circ f: X \to F$ is weakly integrable and that

$$T\left(\int f d\mu\right) = \int (T\circ f) d\mu. \tag{1.8}$$

We shall now briefly discuss a class of functions which are weakly integrable, viz. the Bochner integrable functions. If $f: X \to E$ has finite range, say $f[X] = \{\xi_1,\ldots,\xi_n\}$ and if for $i = 1,\ldots,n$ the set $A_i := \{x \in X \mid f(x) = \xi_i\}$ is an element of S, whereas $\mu(A_i) < \infty$ if $\xi_i \neq 0$, then f is weakly integrable and $\int f d\mu = \Sigma_{i=1}^n \mu(A_i)\xi_i$. We leave the straightforward proof to the reader. Such a function f will be called a *simple integrable function*. If f is a simple integrable function, then the mapping $x \mapsto \|f(x)\|: X \to \mathbb{R}^+$ is an integrable simple function, and

$$\|\int f d\mu\| \leq \int \|f(x)\| d\mu(x). \tag{1.9}$$

Now suppose that $f: X \to E$ is a function such that there exists a sequence $\{h_n\}$ of simple integrable functions such that for every $n \in \mathbb{N}$ the function $x \mapsto \|f(x)-h_n(x)\|: X \to \mathbb{R}^+$ is integrable, and

$$\lim_{n\to\infty} \int \|f(x)-h_n(x)\| d\mu(x) = 0.$$

Then f is called *Bochner integrable*. In that case, the sequence $\{\int h_n d\mu\}$ in E turns out to be a Cauchy sequence in E, and its limit ξ is called the *Bochner integral* of f. It satisfies the relation $\langle\xi,\phi\rangle = \int\langle f(x),\phi\rangle d\mu(x)$ for every $\phi \in E^*$. (All integrals $\int\langle f(x),\phi\rangle d\mu(x), \phi \in E^*$, turn out to exist.) *Thus, if f is Bochner integrable, then it is weakly integrable, and its Bochner integral equals its weak integral. It can also be shown that* $x \mapsto \|f(x)\|: X \to \mathbb{R}^+$ *is integrable and that* (1.9) *holds for all Bochner integrable functions.*

*) By definition, $T^*(\psi) := \psi\circ T$ for $\psi \in F^*$. Stated otherwise, $\langle\xi,T^*(\psi)\rangle = \langle T\xi,\psi\rangle$ for $\xi \in E$, $\psi \in F^*$.

The converse may be not true: if $f : X \to E$ is weakly integrable, then f may be not Bochner integrable, not even if the mapping $x \mapsto \|f(x)\| : X \to \mathbb{R}^+$ is integrable; the problem is that these hypotheses may not imply the existence of a sequence of integrable simple functions $\{h_n\}$ such that $h_n(x) \to f(x)$ for almost every $x \in X$. If such a sequence exists and if the mapping $x \mapsto \|f(x)\|$ is in $L^1(\mu)$, then f is Bochner integrable [9;pp.222].

The collection $B_1 = B_1(X,S,\mu;E)$ of all Bochner integrable functions on X with values in E is easily seen to be a normed vector space with pointwise linear operations and norm $\|f\|_1 := \int\|f(x)\|d\mu(x)$. It can be shown that B_1 *with this norm is complete*. In particular, it follows that if $\{f_n\}$ is a sequence of Bochner integrable functions (not necessarily simple integrable functions) and if $f : X \to E$ is an arbitrary function such that $\lim_{n\to\infty} \int\|f(x) - f_n(x)\|d\mu(x) = 0$, then $f \in B_1$. [Indeed, $\{f_n\}$ is easily seen to be a Cauchy sequence in B_1 which, by completeness, converges in B_1 with limit $f' \in B_1$, say. Since then

$$\int\|f(x) - f'(x)\|d\mu(x) \le \int\|f(x) - f_n(x)\|d\mu(x) + \int\|f_n(x) - f'(x)\|d\mu(x)$$

where the right-hand member tends to zero for $n \to \infty$, it follows that $f = f'$ a.e.[μ], which implies that $f \in B_1$.]

For details about Bochner integrable functions, see ZAANEN [9;§31].

1.15. Operator-valued integrals

Let E and F be Banach spaces and (X,S,μ) a measurable space. Then the space $L(E,F)$ of all continuous linear operators from E into F is a Banach space, so the definitions given in 1.14 could be applied. Usually, this is done only for Bochner integrability of functions from X into $L(E,F)$. As to weak integrability, then one considers not the norm topology on $L(E,F)$, but the *weak operator topology* (i.e. the weakest topology making all mappings $T \mapsto \langle T\xi,\psi\rangle : L(E,F) \to \mathbb{C}$ continuous for $\xi \in E$, $\psi \in F^*$). Since the dual of $L(E,F)$ with this topology is $E \otimes F^*$, or less sophisticated, since all continuous linear functionals on $L(E,F)$ with the weak topology are finite sums of functionals of the form $T \mapsto \langle T\xi,\psi\rangle$ with $\xi \in E$ and $\psi \in F^*$, the following definitions are quite natural.

A function $A: X \to L(E,F)$ is said to be *weakly measurable* if the function $x \mapsto \langle A(x)\xi,\psi\rangle: X \to \mathbb{C}$ is measurable for every $\xi \in E$, $\psi \in F^*$. The function $A: X \to L(E,F)$ is called *weakly integrable* if all functions

$x \mapsto \langle A(x)\xi,\psi\rangle: X \to \mathbb{C}$ are integrable and, in addition, there exists $T \in L(E,F)$ such that $\langle T\xi,\psi\rangle = \int\langle A(x)\xi,\psi\rangle d\mu(x)$ for all $\xi \in E$, $\psi \in F^*$. In that case, such an operator T is uniquely determined, and will be denoted $\int A d\mu$. Thus, if $A: X \to L(E,F)$ is weakly integrable, then, by definition, *)

$$(1.10) \qquad \langle\left(\int A d\mu\right)\xi,\psi\rangle = \int\langle A(x)\xi,\psi\rangle d\mu(x), \qquad \xi \in E,\ \psi \in F^*.$$

The usual properties of this type of integral are easily established. It is also easily verified that

$$(1.11) \qquad B\circ\int A d\mu = \int(B\circ A)d\mu \quad \text{and} \quad \int A d\mu\circ C = \int(A\circ C)d\mu$$

for all continuous linear operators B and C, defined on F, resp. with values in E (here $(B\circ A)(\xi) := B(A(\xi))$ and $(A\circ C)(\xi) := A(\xi)\circ C$ for $\xi \in E$). Finally, it follows easily from the definitions that *if* $A: X \to L(E,F)$ *is weakly integrable, then for every* $\xi \in E$ *the function* $x \mapsto A(x)\xi: X \to E$ *is weakly integrable, and*

$$(1.12) \qquad \int A(x)(\xi)d\mu(x) = \left(\int A d\mu\right)(\xi).$$

Conversely, suppose that $A: X \to L(E,F)$ *is such that for every* $\xi \in E$, *the function* $x \mapsto A(x)\xi: X \to E$ *is weakly integrable and that, in addition, the mapping* $x \mapsto \|A(x)\| : X \to \mathbb{R}^+$ *is integrable. Then* $A : X \to L(E,F)$ *is weakly integrable and*

$$(1.13) \qquad \left\|\int A d\mu\right\| \leq \int\|A(x)\|d\mu(x) .$$

<u>PROOF</u>. Clearly $T: \xi \mapsto \int A(x)\xi d\mu(x): E \to F$ is a well-defined linear mapping. In virtue of the definition of the weak integral $\int A(x)\xi d\mu(x)$ we have for every $\psi \in F^*$:

$$\langle T(\xi),\psi\rangle = \langle\int A(x)\xi d\mu(x),\psi\rangle = \int\langle A(x)\xi,\psi\rangle d\mu(x).$$

---

*) In finite-dimensional spaces this means exactly that, with respect to given bases in E and F, the matrix of $\int A d\mu$ is obtained by entry-wise integration of the matrix function associated to the operator valued function A.

So we need only to show that T is continuous, i.e. $T \in L(E,F)$. Now $|\langle T(\xi),\psi\rangle| \le \int |\langle A(x)\xi,\psi\rangle| d\mu(x) \le \|\xi\| \ \|\psi\| \int \|A(x)\| d\mu(x)$, so that $\|T(\xi)\| = \sup_{\|\psi\|\le 1} |\langle T(\xi),\psi\rangle| \le \|\xi\| \int \|A(x)\| d\mu(x)$. It follows that T is bounded (i.e. continuous) and, in fact, $\|T\| \le \int \|A(x)\| d\mu(x)$. This proves our assertion.

1.16. Vector- and operator-valued integrals in the case of a Hilbert space

Let (X,S,μ) be a measure space and let H be a Hilbert space. Then the dual of H can be identified with H, and the pairing $\langle .,. \rangle$ between H and its dual can be replaced by the inner product $(.,.)$ of H. Now the definitions, given in 1.14 and 1.15, can easily be rewritten in the present context. For example, a function $f: X \to H$ or a function $A: X \to L(H) := L(H,H)$ is weakly measurable if all functions $x \mapsto (f(x),\eta) : X \to \mathbb{C}$, respectively $x \mapsto (A(x)\xi,\eta): X \to \mathbb{C}$ are measurable $(\xi,\eta \in H)$, etc.

We shall prove now some (weak) integrability criterions. First, let $f : X \to H$ be weakly measurable and let $x \mapsto \|f(x)\| : X \to \mathbb{R}^+$ be integrable. In view of the inequality $|f(x),\eta)| \le \|\eta\| \ \|f(x)\|$ it follows from 1.4(v) that the function $x \mapsto (f(x),\eta) : X \to \mathbb{C}$ is integrable, so that

$$\phi_f: \eta \mapsto \int (f(x),\eta) d\mu(x): H \to \mathbb{C}$$

is a well-defined conjugate linear functional on H such that $|\phi_f(\eta)| \le \int |(f(x),\eta)| d\mu(x) \le \int \|f(x)\| d\mu(x) \ \|\eta\|$. So ϕ_f is a bounded conjugate linear functional with norm $\|\phi_f\| \le \int \|f(x)\| d\mu(x)$. Consequently, by the RIESZ Representation Theorem there exists a unique vector $\xi_f \in H$ such that

$$(\xi_f,\eta) = \phi_f(\eta) = \int (f(x),\eta) d\mu(x)$$

for all $\eta \in H$. This means exactly that f is weakly integrable, with $\int f d\mu = \xi_f$. Moreover, since $\|\xi_f\| = \|\phi_f\| \le \int \|f(x)\| d\mu(x)$ we have in this special case

$$\|\int f d\mu\| \le \int \|f(x)\| d\mu(x) \ . \tag{1.14}$$

Resuming, we have proved that *if* $f: X \to H$ *is weakly measurable and if the mapping* $x \mapsto \|f(x)\| : X \to \mathbb{R}^+$ *is integrable, then* f *is weakly integrable, and* (1.14) *holds.*

Next, consider a mapping $A: X \to L(H)$ which is weakly measurable and assume, in addition, that $x \to \|A(x)\| : X \to \mathbb{R}^+$ is integrable. In virtue of 1.4(v) the inequality $|(A(x)\xi,\eta)| \leq \|A(x)\|.\|\xi\|.\|\eta\|$ implies that $\int (A(x)\xi,\eta)d\mu(x)$ exists for every $\xi,\eta \in H$. So

$$\psi_A: (\xi,\eta) \mapsto \int (A(x)\xi,\eta)d\mu(x): H \times H \to \mathbb{C}$$

is a well-defined bounded sesquilinear form on H with norm $\|\psi_A\| \leq \int \|A(x)\|d\mu(x)$. Hence there exists a unique bounded linear operator $T_A \in L(H)$ such that

$$(T_A\xi,\eta) = \psi_A(\xi,\eta) = \int (A(x)\xi,\eta)d\mu(x)$$

for all $\xi,\eta \in H$. Hence $A: X \to L(H)$ is weakly integrable, and $\int A d\mu = T_A$. Moreover, $\|T_A\| = \|\psi_A\|$, so that

$$\|\int A d\mu\| \leq \int \|A(x)\|d\mu(x). \tag{1.15}$$

Thus, we have proved: *if* $A : X \to L(H)$ *is weakly measurable and if the mapping* $x \mapsto \|A(x)\| : X \to \mathbb{R}^+$ *is integrable, then* A *is weakly integrable, and* (1.15) *holds*. [Observe, that we could also prove this by using the last statement in 1.15 and the criterion stated above for weak integrability of functions from X to H. Formula (1.15) then follows from formula (1.13).]

1.17. The Hilbert space $L^2(X,\mu;H)$

Let (X,S,μ) be a measure space, and let H be a complex Hilbert space with norm $\|.\|$ and inner product $(.,.)$. Motivated by 1.6 one may consider, for every $p \geq 1$, the collection $L^p(X,H)$ of all functions $f: X \to H$ which are weakly measurable and for which the mapping $x \mapsto \|f(x)\|^p : X \to \mathbb{R}^+$ is integrable. In particular, $L^1(X,H)$ is the space of all weakly measurable functions $f: X \to H$ for which the mapping $x \mapsto \|f(x)\| : X \to \mathbb{R}^+$ is integrable. By the first part of 1.16, each $f \in L^1(X,H)$ is weakly integrable.

We shall discuss now $L^2(X,H)$ for the case that H is a *separable* Hilbert space. Let $\{e_n | n \in \mathbb{N}\}$ be an orthonormal basis in H. First, observe that if $f, g : X \to H$ are two weakly measurable functions, then we have for all $x \in X$

$$(f(x),g(x)) = \sum_{n\in\mathbb{N}} (f(x),e_n)\overline{(g(x),e_n)}, \tag{1.16}$$

where all terms of this series are products of measurable, complex-valued functions of the variable x. *It follows that the function* $x \mapsto (f(x),g(x))$: $X \to \mathbb{C}$ *is measurable. In particular, the function* $x \mapsto \|f(x)\|^2 = (f(x),f(x))$: $X \to \mathbb{R}^+$ *is measurable.*

Let $L^2(X,\mu;H)$ be the set of all weakly measurable functions $f\colon X \to H$ for which

$$\|f\|_2 := \Big(\int_X \|f(x)\|^2 d\mu(x) \Big)^{1/2} < \infty.$$

In the usual way, identifying functions which are equal a.e.[μ], one obtaines the space $L^2(X,\mu;H)$ (compare 1.6), and the elements of $L^2(X,\mu;H)$ will be treated as if it were functions instead of equivalence classes of functions.

If $f,g \in L^2(X,\mu;H)$ then we have seen above that $x \mapsto (f(x),g(x))$: $X \to \mathbb{C}$ is measurable. Moreover, in virtue of the inequality

$$(1.17) \qquad |(f(x),g(x))| \le \|f(x)\|\|g(x)\| \le \tfrac{1}{2}(\|f(x)\|^2+\|g(x)\|^2),$$

where the right-hand side represents an integrable real-valued function, it follows from 1.4(v) that the function $x \mapsto (f(x),g(x))\colon X \to \mathbb{C}$ is integrable. Thus, for $f,g \in L^2(X,\mu;H)$ we can define the complex number

$$(1.18) \qquad (f,g) := \int_X (f(x),g(x))d\mu(x),$$

and clearly also (g,f) is defined, and $(g,f) = \overline{(f,g)}$. In addition, if $f,g \in L^2(X,\mu;H)$, then $x \mapsto f(x)+g(x)$ is weakly measurable. Since then $x \mapsto \|f(x)+g(x)\|^2\colon X \to \mathbb{R}^+$ is measurable and

$$\|f(x)+g(x)\|^2 \le \|f(x)\|^2+(f(x),g(x))+(g(x),f(x))+\|g(x)\|^2$$

$$\le 2(\|f(x)\|^2+\|g(x)\|^2)$$

by (1.17), it follows from 1.4(v) that $x \mapsto \|f(x)+g(x)\|^2\colon X \to \mathbb{R}^+$ is integrable. Consequently, $f + g \in L^2(X,\mu;H)$. It is obvious that complex multiples of elements of $L^2(X,\mu;H)$ are again in $L^2(X,\mu;H)$, so we may conclude that with pointwise addition and scalar multiplication $L^2(X,\mu;H)$ is a vector space. Now it is easily seen that (1.18) defines an inner product on $L^2(X,\mu;H)$, such that the norm, derived from it is just $\|\cdot\|_2$. *Thus,*

$L^2(X,\mu;H)$ *has the structure of a complex pre-Hilbert space.*

In order to show that $L^2(X,\mu;H)$ is a Hilbert space (i.e. that it is complete in its norm) we proceed as follows. For $f \in L^2(X,\mu;H)$, set

$$f_n(x) := (f(x),e_n), \qquad x \in X.$$

Observe, that $f_n: X \to \mathbb{C}$ is measurable, and that

$$(1.19) \qquad \|f(x)\|^2 = \sum_{n\in\mathbb{N}} |(f(x),e_n)|^2 = \sum_{n\in\mathbb{N}} |f_n(x)|^2$$

so that $|f_n(x)|^2 \leq \|f(x)\|^2$. It follows from 1.4(v) that $f_n \in L^2(\mu)$, the "ordinary" L^2-space, defined in 1.6. In virtue of (1.19) and the Monotone Convergence Theorem 1.5(ii) the following is true:

$$(1.20) \qquad \int \|f(x)\|^2 d\mu(x) = \int \lim_{k\to\infty} \sum_{n=1}^{k} |f_n(x)|^2 \, d\mu(x)$$

$$= \lim_{k\to\infty} \sum_{n=1}^{k} \int |f_n(x)|^2 \, d\mu(x) = \sum_{n=1}^{\infty} \|f_n\|_2^2.$$

Since the left-hand side of this equality is finite, the right-hand side is finite, that is, $(f_n)_{n\in\mathbb{N}}$ is an element of the direct sum $\Sigma^{\oplus}_{n\in\mathbb{N}} L^2(\mu)$ of countably many copies of $L^2(\mu)$. If we set $\Phi(f) := (f_n)_{n\in\mathbb{N}}$, then we may derive also from (1.20) that the norm of $\Phi(f)$ in this direct sum satisfies*)

$$\|\Phi(f)\| = \left(\sum_{n=1}^{\infty} \|f_n\|_2^2\right)^{1/2} = \|f\|_2.$$

Hence $\Phi: f \mapsto (f_n)_{n\in\mathbb{N}}: L^2(X,\mu;H) \to \Sigma^{\oplus}_{n\in\mathbb{N}} L^2(\mu)$ is a norm-preserving mapping, which is easily seen to be linear. We show that Φ is a *surjection,* as follows: if $(g_n)_{n\in\mathbb{N}} \in \Sigma^{\oplus}_{n\in\mathbb{N}} L^2(\mu)$ then by the Monotone Convergence Theorem (which holds for all non-negative measurable functions) we have, similar to (1.20),

$$\int \sum_{n=1}^{\infty} |g_n(x)|^2 \, d\mu(x) = \sum_{n=1}^{\infty} \|g_n\|_2^2 < \infty.$$

---

*) The reader should distinguish the norms of $\Sigma^{\oplus}_{n\in\mathbb{N}} L^2(\mu)$ and of H, both denoted $\|\cdot\|$, from each other, and also the norms of $L^2(X,\mu;H)$ and $L^2(\mu)$, both denoted $\|\cdot\|_2$.

It follows that the function $x \mapsto \Sigma_{n=1}^{\infty} |g_n(x)|^2 : X \to \mathbb{R}^+$ is μ-integrable, so that it is finite for almost every $x \in X$ (cf. the last paragraph of 1.3). Changing the values of the functions g_n on a suitable null-set, we may assume that $\Sigma_{n=1}^{\infty} |g_n(x)|^2 < \infty$ for every $x \in X$. (Notice that such a change does not alter the equivalence class of g_n in $L^2(\mu)$, i.e., it does not affect the element g_n of $L^2(\mu)$.) So by completeness of H, we may define for every $x \in X$ an element $g(x) \in H$ by

$$g(x) := \sum_{n=1}^{\infty} g_n(x)e_n.$$

Then $g: X \to H$ is easily seen to be weakly measurable, and a computation, similar to (1.20), shows that $g \in L^2(X,\mu;H)$. Now it is obvious that $\Phi(g) = (g_n)_{n\in\mathbb{N}}$ so, indeed, Φ is surjective. *We have shown that* Φ *is an isomorphism of pre-Hilbert spaces from* $L^2(X,\mu;H)$ *onto* $\Sigma_{n\in\mathbb{N}}^{\oplus} L^2(\mu)$. However, each $L^2(\mu)$ is complete (cf.1.6 or 1.7), so the direct sum $\Sigma_{n\in\mathbb{N}}^{\oplus} L^2(\mu)$ is complete. *Therefore,* $L^2(X,\mu;H)$ *is complete, i.e., it is a Hilbert space.*

2. INTEGRATION ON LOCALLY COMPACT SPACES

In this section, the concepts considered in the previous section are applied in the context of a special σ-algebra (Borel sets) on a locally compact space. We shall point out that in this context the class of continuous functions with compact support is of basic importance.

2.1. The space $K(X)$ of continuous functions with compact support

Recall that a topological space X is *locally compact* whenever X is a Hausdorff space (i.e., different points have disjoint neighbourhoods) such that for every $x \in X$ and for every neighbourhood U of x there exists an open neighbourhood V of x such that $\bar{V} \subseteq U$ and $\bar{V}$ is compact. Here $\bar{V}$ denotes the *closure* of V, that is, the smallest closed set in which V is included. (If X is locally a metric space, in particular, if X is an analytic manifold, then $\bar{V}$ consists of all points of V together with all possible limits in X of sequences in V.)

Let X be a locally compact space. If f is a real-valued or complex-valued continuous function on X, then the *support of* f is defined as $\mathrm{supp}(f) := \overline{\{x \in X \mid f(x) \neq 0\}}$. The set of all continuous real-valued*)

*) We shall treat only integration of real-valued functions. For complex-valued functions, use the recipe of 1.13: consider both the real and imaginary part of a complex-valued function.

functions with compact support will be denoted $K(X)$. The follwoing property is fundamental:

If $K \subseteq U \subseteq X$ *with* K *compact and* U *open, then there exists* $f \in K(X)$ *such that* $0 \leq f(x) \leq 1$ *for all* $x \in X$, $f(x) = 1$ *for all* $x \in K$ *and* $f(x) = 0$ *for all* $x \in X\backslash U$. *If* X *is an analytical manifold, then* f *can be chosen, in addition, in* $C^{\infty}(X)$.

2.2. Borel sets and Borel functions

The *Borel sets* in a topological space X are the elements of the σ-algebra $\mathcal{B}(X)$ (or shortly: $\mathcal{B}$), generated by the collection of all open subsets of X. As σ-algebras are invariant under complementation, the σ-algebra $\mathcal{B}$ in X is also generated by the collection of all *closed* subsets of X. If X is σ-compact, i.e., X is a countable union of compact sets, then $\mathcal{B}$ is also genarated by the collection of all *compact* subsets of X. (So in that case, our definition, which is the one given in HEWITT & ROSS [3] or HEWITT & STROMBERG [4], agrees with HALMOS [1]*).) This applies to all Lie groups!

A *Borel measurable function* (or shortly, a *Borel function*) on X is a real- (or complex-) valued function on X that is measurable with respect to the σ-algebra $\mathcal{B}$ of Borel sets in X.

Every continuous real- (or complex-) valued function on X *is a Borel function*. [Indeed, if $f: X \to \mathbb{R}$ is continuous, then $f^{-1}[U]$ is an open subset of X for every open set $U \subseteq \mathbb{R}$. In particular, $f^{-1}[U] \in \mathcal{B}$ for every open interval U in $\mathbb{R}$. Now the last statement of the third paragraph in 1.1 gives the desired result. If f is complex-valued, then apply the preceding to the functions $\mathcal{R}e f$ and $\mathcal{I}m f$.] Consequently, if E is a real or complex Banach space and $f: X \to E$ is weakly continuous (i.e., f is continuous with respect to the weak topology on E or, equivalently, $\phi \circ f: X \to \mathbb{R}$ (resp. $\mathbb{C}$) is continuous for every $\phi \in E^*$) then f is weakly measurable. Similarly, if an operator-valued function $A: X \to L(E)$ is weakly continuous (i.e., A is continuous with respect to the weak operator topology on $L(E)$ or, equivalently, $x \mapsto \langle A(x)\xi,\phi\rangle: X \to \mathbb{R}$ (resp. $\mathbb{C}$) is continuous for every $\xi \in E$, $\phi \in E^*$) then A is weakly measurable.

---

*) For a detailed discussion of the several σ-algebras which are important in the theory of integration on topological spaces, we refer to [5;Ch.7].

2.3. Borel measures

Let X be a locally compact space. A *Borel measure* on X is a measure μ defined on the σ-algebra $\mathcal{B}$ of Borel sets such that $\mu(K) < \infty$ for every compact subset K of X. A Borel measure μ is called *regular* whenever for every open subset U of X,

$$\mu(U) = \sup\{\mu(K) \mid K \subseteq U \text{ \& } K \text{ compact}\} \tag{2.1}$$

and, in addition, for every $A \in \mathcal{B}$,

$$\mu(A) = \inf\{\mu(V) \mid V \supseteq A \text{ \& } V \text{ open}\}. \tag{2.2}$$

Clearly, two regular Borel measures are identical if and only if they are equal on the compact sets.

2.4. Borel measures on a σ-compact space

From now on we shall assume that X is not only *locally compact, but also σ-compact*, that is, X is a union of countably many compact subsets. Examples of such spaces are all locally compact spaces satisfying the second axiom of countability (including all Lie groups; cf. IV.1.5). We mention two important consequences of this assumption:

(i) Every Borel measure on X is σ-finite. (Indeed, X is union of countably many sets of finite measure.)

(ii) The two conditions (2.1) and (2.2) in the definition of regularity are equivalent.

A third consequence has already been mentioned at the end of the first paragraph of 2.2, namely, that we can refer freely both to HALMOS [1] and to HEWITT & STROMBERG [4], since in the case of σ-compact spaces their definitions of Borel set, etc., coincide with ours. The following remark is related to this. (In fact, this makes the theory of integration on locally compact spaces slightly simpler as compared with the presentation in HEWITT & ROSS [3] or HEWITT & STROMBERG [4], because the distinction between null-sets (= sets of measure zero) and sets which are "locally null" disappears.)

(iii) A Borel set A in X has $\mu(A) = 0$ for a given Borel measure μ if and only if $\mu(A \cap K) = 0$ for every compact subset K of X.

["Only if": obvious. "If": $X = \bigcup_{n=1}^{\infty} K_n$ with every K_n compact, so $A = \bigcup_{n=1}^{\infty} A \cap K_n$, whence $\mu(A) \leq \sum_{n=1}^{\infty} \mu(A \cap K_n) = 0$.]

2.5. The role of metrizability

If X is, in addition to being locally compact and σ-compact, also *metrizable*, then every compact set is a countable intersection of open sets (that is, every compact set is a G_δ-set). This means that every Borel measure on X is a so-called *Baire-measure* (see HALMOS [1] for the definition). Since every Baire-measure is regular (HALMOS [1;p.229]) it follows that, *on a locally compact, σ-compact metrizable space* X, *every Borel measure is regular. In particular, this applies to the case that* X = G/H *where* G *is a locally compact group satisfying the second axiom of countabilty and* H *is a closed subgroup of* G. (Indeed, then G is metrizable by HEWITT & ROSS [3;5.3]; by [3;5.22 & 8.14(b)], G/H is locally compact and metrizable; that G/H is second countable is an obvious consequence of the fact that the quotient map $x \mapsto xH$: $G \to G/H$ sends open sets in G onto open sets in G/H; cf.[3;5.17].) In particular, every Borel measure on a Lie group is regular.

2.6. Integration with respect to a Borel measure

Let μ be a Borel measure on X. Since every $f \in K(X)$ is a Borel function, and $\int|f|d\mu \leq \|f\|_\infty . \mu(\mathrm{supp}(f)) < \infty$, it follows that *every* $f \in K(X)$ *is μ-integrable*.

Thus, we may define a mapping I_μ: $K(X) \to \mathbb{R}$ by

$$I_\mu(f) := \int f d\mu \quad \text{for } f \in K(X).$$

Now $K(X)$ is a vector space (pointwise addition and scalar multiplication) and 1.4(i),(ii) imply that the mapping I_μ is *linear* and *positive*, that is:

(i) $I_\mu(c_1f_1+c_2f_2) = c_1I(f_1) + c_2I(f_2)$ for $f_i \in K(X)$ and $c_i \in \mathbb{R}$ (i=1,2).

(ii) $I_\mu(f) \geq 0$ for every $f \in K(X)$ such that $f \geq 0$ (that is, such that $f(x) \geq 0$ for all $x \in X$).

Moreover, if μ is a finite measure, then $|I_\mu(f)| \leq \mu(X) . \|f\|_\infty$ for all $f \in K(X)$; since $\mu(X)$ is a finite non-negative number, independent of $f \in K(X)$, this shows that I_μ is a *bounded* linear functional on $K(X)$, with $\|I_\mu\| \leq \mu(X)$; in fact, it is not difficult to show that $\|I_\mu\| = \mu(X)$.

The following shows that a regular Borel measure is completely determined by its integral.

2.7. The Riesz representation theorem: uniqueness

If μ *and* ν *are two regular Borel measures on* X *such that* $\int f d\mu = \int f d\nu$ *for every* $f \in K(X)$, *then* $\mu = \nu$.

PROOF. Let A be a Borel set of X. Let us first assume that $\mu(A) < \infty$ and $\nu(A) < \infty$. Let $\varepsilon > 0$. Then by regularity there exist a compact subset K of A and an open set $U \supseteq A$ such that

$$\mu(A) - \varepsilon < \mu(K) \leq \mu(U) < \mu(A) + \varepsilon,$$

$$\nu(A) - \varepsilon < \nu(K) \leq \nu(U) < \nu(A) + \varepsilon.$$

Stated otherwise, $\chi_K \leq \chi_A \leq \chi_U$ and

$$\int (\chi_U - \chi_K) d\mu < 2\varepsilon; \int (\chi_U - \chi_K) d\nu < 2\varepsilon.$$

By 2.1 there exists $f \in K(X)$ such that $\chi_K \leq f \leq \chi_U$. It follows that

$$\left|\int (\chi_A - f) d\mu\right| \leq \int |\chi_A - f| d\mu \leq \int (\chi_U - \chi_K) d\mu < 2\varepsilon$$

and, similarly, $\left|\int (\chi_A - f) d\nu\right| < 2\varepsilon$. Hence

$$\left|\int \chi_A d\mu - \int \chi_A d\nu\right| \leq \left|\int (\chi_A - f) d\mu\right| + \left|\int (\chi_A - f) d\nu\right| + \left|\int f d\mu - \int f d\nu\right|$$

$$< 2\varepsilon + 2\varepsilon + 0.$$

Since this holds for every $\varepsilon > 0$, it follows that $\int \chi_A d\mu = \int \chi_A d\nu$, that is, $\mu(A) = \nu(A)$. In particular, this holds if A is a compact set.

Next, suppose that $\mu(A) = \infty$. Then by regularity, A contains compact subsets K of arbitrary large μ-measure; since $\nu(A) \geq \nu(K) = \mu(K)$, it follows that $\nu(A) = \infty$. Similarly, if $\nu(A) = \infty$, then $\mu(A) = \infty$. So in all cases, $\mu(A) = \nu(A)$. □

2.8. The Riesz representation theorem: existence

The following theorem is of fundamental interest. It states that every positive linear functional on $K(X)$ comes from a measure. (The theorem is also valid if X is not σ-compact, but local compactness is essential.)

RIESZ REPRESENTATION THEOREM. *If* I: $K(X) \to \mathbb{R}$ *is a positive linear functional, then there exists a unique regular Borel measure* μ *on* X, *such that, for every* $f \in K(X)$,

$$(2.3) \qquad I(f) = \int f d\mu.$$

If I *is bounded then* μ *is finite, and* $\mu(X) = \|I\|$.

For a proof, we refer to HEWITT & STROMBERG [4;§9 and Theorem 12.36].

REMARK. 1. Uniqueness in the above theorem is a consequence of regularity: see 2.7 above. If X is also metrizable, regularity of μ follows from 2.5.
2. Motivated by this representation theorem, a positive linear functional on $K(X)$ is often called an *integral* or even a *measure* (sometimes it is called a *Radon measure*). In the approach of integration theory of BOURBAKI (see also REITER [7]), Radon measures are the starting point, and measures according to our definition in §1 are derived from it.
3. The discussion of the relation between complex linear functionals and measures will be postponed to 2.13 below.

2.9. Lebesgue measure

By means of the construction presented in 1.2, the regular Borel measure μ that corresponds to a given positive linear functional on $K(X)$ can be extended to a complete measure $\bar{\mu}$, defined on a σ-algebra $\bar{B} \supseteq B$. It should be observed, that $\bar{B}$ depends on μ.

EXAMPLE. Let $X = \mathbb{R}^n$ and define I: $K(\mathbb{R}^n) \to \mathbb{R}$ by $I(f) := \int_{\mathbb{R}^n} f(x_1,\dots,x_n) dx_1\dots dx_n$ (ordinary Riemann integral). Then I is linear and positive, so there exists a unique measure λ on a σ-algebra L such that

(i) all Borel sets of $\mathbb{R}^n$ belong to L, that is, $B \subseteq L$;
(ii) $\lambda|_B$ is a regular Borel measure;
(iii) $(\mathbb{R}^n, L, \lambda)$ is a complete measure space and L is the completion of B;
(iv) for every $f \in K(\mathbb{R}^n)$ one has

$$\int f d\lambda = \int_{\mathbb{R}^n} f(x_1,\dots,x_n) dx_1\dots dx_n.$$

In this particular case, the measure λ is called the n-*dimensional Lebesgue measure*, and the elements of L are called the Lebesgue measurable sets.

It is known that $\mathcal{B} \neq \mathcal{L}$, that is, not every Lebesgue measurable set is a Borel set. It is also known that not every subset of $\mathbb{R}^n$ is Lebesgue-measurable.

2.10. $\underline{K(X)}$ is dense in $L^p(X,\mu)$

Let I and μ be as in 2.8 and let for $1 \le p < \infty$ the Banach spaces $L^p(\mu)$ be defined according to 1.6. Recall that $L^2(\mu)$ is a Hilbert space. A consequence of the following statement is that for many relations, in order to hold on the whole of $L^p(\mu)$, it is sufficient to be valid on $K(X)$:

The space $K(X)$ *is a dense linear subspace of the Banach space* $L^p(\mu)$ *for every* p *with* $1 \le p < \infty$, *that is, for every* $f \in L^p(\mu)$ *and every* $\varepsilon > 0$ *there exists* $g \in K(X)$ *such that*

$$\int |f(x)-g(x)|^p \, d\mu(x) < \varepsilon.$$

[Sketch of proof for the case p = 1. It is a straightforward consequence of the definition of the integral in 1.3 that the set of integrable simple functions is dense in $L^1(\mu)$. By essentially the same method as used in the proof of 2.7 one shows that every integrable simple function can be approximated by elements of $K(X)$. So $K(X)$ is dense in $L^1(\mu)$.]

2.11. The behaviour of a Borel measure under a homeomorphism

Suppose $T: X \to X$ is a homeomorphism. Then it is easy to see that $A \subseteq X$ is a Borel set if and only if $T^{-1}(A)$ is a Borel set. So T is a measurable transformation in the sense of 1.11. If μ is a regular Borel measure on X, then μ_T can be defined according to 1.11, and μ_T is easily seen to be a regular Borel measure as well. Assume that μ is T-invariant, that is, $\mu = \mu_T$ or

$$\mu(A) = \mu(T^{-1}(A)) \tag{2.4}$$

for every Borel set A. Then 1.11 implies immediately that for any Borel function $f: X \to \mathbb{R}_*$ the following is true: $f \in L^1(\mu)$ if and only if $f \circ T \in L^1(\mu)$ and in that case

$$\int f(T(x))d\mu(x) = \int f(x)d\mu(x). \tag{2.5}$$

In particular, (2.5) is valid for every $f \in K(X)$.

Conversely, assume μ is a regular Borel measure and T is a homeomorphism of X onto itself such that (2.5) holds for every $f \in \mathcal{K}(X)$. Using 1.11, we can restate (2.5) as follows:

$$\int f d\mu_T = \int f d\mu$$

for every $f \in \mathcal{K}(X)$. Then 2.7 (regularity of μ!) implies that $\mu = \mu_T$, that is, (2.4) holds.

We have proved:

For every regular Borel measure μ *and every homeomorphism* $T: X \to X$ *the following are equivalent*:

(i) $\mu(A) = \mu(T^{-1}[A])$ *for every Borel set* A;

(ii) $\int f(T(x))d\mu(x) = \int f(x)d\mu(x)$ *for every* $f \in \mathcal{K}(X)$.

If these conditions are fulfilled, then for any Borel function f *we have* $f \in L^1(\mu)$ *if and only if* $f \circ T \in L^1(\mu)$. *Furthermore,* (ii) *holds for every* $f \in L^1(\mu)$.

In essentially the same way one can show that the following are equivalent:

(i) $\mu_T \equiv \mu$, that is, for any Borel set A, $\mu(A) = 0$ if and only if $\mu(T^{-1}[A]) = 0$.

(ii) There is a finite-valued Borel function $R_T > 0$ such that $d\mu_T(x) = R_T(x)d\mu(x)$, i.e. for every $f \in \mathcal{K}(X)$, $f \geq 0$, we have

$$\int f(T(x))d\mu(x) = \int f(x)R_T(x)d\mu(x). \tag{2.6}$$

If these conditions are fulfilled then for any Borel function f we have $f \circ T \in L^1(\mu)$ if and only if $f.R_T \in L^1(\mu)$, and in that case formula (2.6) holds as well.

2.12. Products of Borel measures

Let X and Y be locally compact, σ-compact topological Hausdorff spaces, let μ be a Borel measure on X and let ν be a Borel measure on Y. According to 1.12, $\mathcal{B}(X) \otimes \mathcal{B}(Y)$ is the σ-algebra in $X \times Y$, generated by the collection of all sets $A \times B$ with $A \in \mathcal{B}(X)$, $B \in \mathcal{B}(Y)$. On the other hand, we have the σ-algebra $\mathcal{B}(X \times Y)$ of all Borel sets in the product space $X \times Y$. It is rather obvious that $\mathcal{B}(X) \otimes \mathcal{B}(Y) \subseteq \mathcal{B}(X \times Y)$, and there are examples showing that the inclusion may be strict. On the other hand, in applying FUBINI's theorem, it would be very convenient if we had equality: then the product measure $\mu \otimes \nu$ on $\mathcal{B}(X) \otimes \mathcal{B}(Y)$ would be a Borel measure on $X \times Y$. Therefore, the following observation is useful: *if* X *and* Y *are locally compact,* σ*-compact metrizable spaces, then* $\mathcal{B}(X) \otimes \mathcal{B}(Y) = \mathcal{B}(X\times Y)$ *and* $\mu \otimes \nu$ *is a Borel measure on* $X \times Y$. [For a direct proof, cf.[5;Prop.7.10]. This result is related to our remark in 2.5, that in this case all Borel sets are Baire sets: indeed, Baire sets behave well with respect to the formation of products; cf.[1] or [5;Cor.7.3].] Using this, we obtain the following reformulation of 1.12:

Let X *and* Y *be locally compact,* σ*-compact metrizable spaces and let* μ *and* ν *be Borel measures on* X *and* Y*, respectively. Let* $f\colon X \times Y \to \mathbb{R}_*$ *be a Borel function on* $X \times Y$*. Then* $y \mapsto f(x_0,y)$ *and* $x \mapsto f(x,y_0)$ *are Borel functions on* Y *and* X *respectively* $(x_0\in X, y_0\in Y)$*, and in addition:*

(i) *If* f *is non-negative, then* $x \mapsto \int_Y f(x,y)d\nu(y)$ *and* $y \mapsto \int_X f(x,y)d\mu(y)$ *are Borel functions on* X*, resp.* Y*, and*

$$\int_{X\times Y} fd\mu \otimes \nu = \int_X\left(\int_Y f(x,y)\,d\nu(y)\right)d\mu(x) = \int_Y\left(\int f(x,y)\,d\mu(x)\right)d\nu(y).$$

In particular, if one of the iterated integrals is finite, then f *is* $\mu \otimes \nu$*-integrable.*

(ii) *If* f *is* $\mu \otimes \nu$*-integrable, then both iterated integrals of* f *exist and the equation in* (i) *is valid.*

2.13. The Riesz representation theorem for complex measures

As before, let $K(X)$ denote the space of all real-valued continuous functions with compact support, and let $K(X)^c$ be the space of all complex-valued continuous functions with compact support. Then $K(X) \subset K(X)^c$ and, in fact, $K(X)^c = K(X) \oplus iK(X)$: if $f \in K(X)^c$, then $f = \mathcal{R}e f + i\mathcal{I}m f$ with $\mathcal{R}e f$, $\mathcal{I}m f \in K(X)$. If μ is a complex Borel measure, that is, μ is a complex measure on the Borel sets, then the equation (cf.1.13 for the definition of integral in this case)

$$J_\mu(f) := \int fd\mu, \quad f \in K(X)^c,$$

defines a linear functional $J_\mu\colon K(X)^c \to \mathbb{C}$ with the property that

$$|J_\mu(f)| = \left|\int fd\mu\right| \le \int |f|\,d|\mu| \le \|f\|_\infty . |\mu|(X),$$

where $|\mu|(X)$ is a finite non-negative number. This shows, that J_μ is a bounded linear functional on the normed space $K(X)^c$ ($K(X)^c$ endowed with its supremum norm $\|.\|_\infty$), and that $\|J_\mu\| \le |\mu|(X)$; in fact, it can be shown that $\|J_\mu\| = |\mu|(X)$.

Conversely, it can be shown (cf. HEWITT & ROSS [3;B.37]) that every bounded linear functional $J: K(X)^c \to C$ can be decomposed as $J = J_1^+ - J_1^- + i(J_2^+ - J_2^-)$, where J_k^+ and J_k^- (k=1,2) are positive linear functionals (i.e., they are real-valued on $K(X)$ and non-negative on $K(X)^+$). Application of the RIESZ representation theorem (cf. 2.8) to each of them then gives the existence part of the following statement.

If $J : K(X)^c \to \mathbb{C}$ *is a bounded linear functional then there exists a unique regular complex Borel measure* μ *such that* $J(f) = \int f d\mu$ *for all* $f \in K(X)^c$. *That is, there exist four non-negative, finite regular Borel measures* μ_1, μ_2, μ_3 *and* μ_4 *such that*

$$J(f) = \int f d\mu_1 - \int f d\mu_2 + i\left(\int f d\mu_3 - \int f d\mu_4\right)$$

for all $f \in K(X)^c$, *and if* ν_1, ν_2, ν_3 *and* ν_4 *are four non-negative, finite regular Borel measures with the same property, then*

$$\nu_1(E) - \nu_2(E) + i(\nu_3(E) - \nu_4(E)) = \mu_1(E) - \mu_2(E) + i(\mu_3(E) - \mu_4(E))$$

for every Borel set E.

PROOF. Existence: see above.
Unicity: it is straightforward to show that $K(X)^c$ is dense in the space $L^1(X,\mu)$ of all Borel functions $f: X \to \mathbb{C}$ for which $\int f d\mu$ exists (cf. also 2.10). Similarly, $K(X)^c$ is dense in $L^1(X,\nu)$. Therefore, it is dense in $L^1(X,\nu) \cap L^1(X,\mu)$. Since this intersection contains all functions χ_A for all Borel sets A, it follows from the equality $\int f d\mu = \int f d\nu$ for all $f \in K(K)^c$ that $\int \chi_E d\mu = \int \chi_E d\nu$, that is, $\mu(E) = \nu(E)$, for all Borel sets E. □

2.14. Integrability of continuous vector-valued functions with compact support

Let E *be a Banach space. Then every continuous function* $f: X \to E$ *with compact support is Bochner integrable.* The proof is quite straightforward: we shall construct a sequence $\{f_n\}$ of simple integrable functions from X into E such that $\lim_{n\to\infty} \int \|f - f_n\| d\mu = 0$ (cf. 1.14). For given $n \in \mathbb{N}$, the function f_n is defined as follows. For every $x \in X$, the set

$U(n,x) := \{y \in X \mid \|f(y) - f(x)\| < 1/n\}$ is open in X and contains the point x. Hence the (compact!) set supp(f) can be covered by finitely many of the sets $U(n,x)$, say $\text{supp}(f) \subseteq \cup\{U(n,x_{ni}) \mid i = 1,\dots,k_n\}$, $x_{ni} \in X$ for $i = 1,\dots,k_n$. If we put inductively

$$A_{n1} := U(n,x_{n1}) \cap \text{supp}(f) ,$$

$$A_{nj} := (U(n,x_{nj}) \setminus \bigcup_{i=1}^{j-1} A_{ni}) \cap \text{supp}(f), \quad j = 2,\dots,k_n,$$

then $\{A_{n1},\dots,A_{nk_n}\}$ is a disjoint collection of Borel sets whose union is supp(f). In particular, $\mu(A_{nj}) \leq \mu(\text{supp}(f)) < \infty$, so $f_n := \Sigma_{j=1}^{k_n} f(x_{nj})\chi_{A_{nj}}$ is a simple integrable function. In addition, it follows from the construction that

$$0 \leq \|f(x) - f_n(x)\| = \begin{cases} 0 & \text{if } x \notin \text{supp}(f) ; \\ \|f(x) - f(x_{nj})\| < \frac{1}{n} & \text{if } x \in A_{nj}. \end{cases}$$

Since supp(f) is covered by the sets A_{nj}, it follows that

$$\int \|f - f_n\| d\mu \leq \frac{1}{n}\, \mu(\text{supp}(f))$$

so that, indeed, $\lim_{n\to\infty} \int \|f - f_n\| d\mu = 0$. This proves that f is Bochner integrable.

Consequently, every continuous function $f: X \to E$ *with compact support is weakly integrable* (cf.1.14). This could also be shown, using the last paragraph of 2.2 and 1.16. If we consider mappings $f: X \to H$, where H is a separable Hilbert space then more can be shown: *in that case every weakly continuous function* $f: X \to H$ *with compact support is weakly integrable.* Indeed, by 1.16 it is sufficient to show that $\|f\|: X \to \mathbb{R}^+$ is μ-integrable. In order to do so, use the method of the first part of 1.17 to show that $\|f\|^2$ is a Borel function. (Indeed, $\|f(x)\|^2 = \Sigma_{n\in\mathbb{N}} (f(x),e_n)\overline{(f(x),e_n)}$, so $\|f\|^2$ is limit of a sequence of continuous functions.) It follows that $\|f\|$ is a Borel function. Moreover, the range $f(X)$ of f is a compact subset of H in the weak topology. It is well-known that this implies that $f(X)$ is norm-bounded. Hence $\|f\| \leq c\chi_K$ for some constant $c > 0$, where $K := \text{supp}(f)$. Now 1.4(v) implies that $\|f\|$ is μ-integrable, and the proof that f is weakly integrable is completed. Notice that with the notation used above, we have also $\|f\|^2 \leq c^2\chi_K$, so that $\|f\|^2: X \to \mathbb{R}^+$ is μ-integrable. This implies: *every weakly continuous function* $f: X \to H$, *H is separable Hilbert space, is in* $L^2(X,\mu;H)$ (cf.1.17).

3. INTEGRATION ON LOCALLY COMPACT GROUPS

In this section we shall consider a special measure on a locally compact group G. We shall always assume that G satisfies the second axiom of countability (for reasons to make this assumption, cf.2.4 and 2.5; see also 2.12). Such a group G is always σ-compact, separable and metrizable.

3.1. Left Haar measures

If $y \in G$, then $L_y: x \mapsto yx: G \to G$ is homeomorphism of G onto itself with inverse $(L_y)^{-1} = L_{y^{-1}}$; cf.IV.2.3. Restating the relevant part of 2.11 for this particular homeomorphism, we see that for a regular Borel measure μ on G the following conditions are equivalent:

(i) $\mu(A) = \mu(y^{-1}A)$ for every Borel set A in G;

(ii) $\int f(yx)d\mu(x) = \int f(x)d\mu(x)$ for every $f \in \mathcal{K}(G)$.

If μ satisfies these conditions for every $y \in G$, then μ is called *left invariant*. For a left invariant regular Borel measure μ, the following are equivalent:

(iii) $\mu \neq 0$;
(iv) There is an open set $U_0 \subseteq G$ such that $\mu(U_0) > 0$;
(v) For every open set $\emptyset \neq U \subseteq G$, $\mu(U) > 0$.

[PROOF: (v) ⇒ (iv) ⇒ (iii) is trivial. That (iii) ⇒ (iv) follows from condition (2.2)(regularity of μ), and the proof of (iv) ⇒ (v) is as follows: suppose $\mu(U) = 0$ for some open set U in G, and let K be any compact subset of G. Then K can be covered by finitely many left translates of U, each of which is a null-set by left invariance of μ. Hence $\mu(K) = 0$. So every compact set is a null-set. Now condition (2.1) of regularity implies that every open set is a null-set; in particular $\mu(U_0) = 0$: a contradiction with (iv).]

A regular Borel measure μ on G that is left invariant and non-zero is called a *left Haar measure* on G. Similarly, we define a *left Haar integral* as a linear mapping $I: \mathcal{K}(G) \to \mathbb{R}$ with the following properties:

(vi) I is strictly positive, that is, $\forall f \in \mathcal{K}(G): f \geq 0 \;\&\; f \neq 0 \Rightarrow I(f) > 0$.
(vii) I is left invariant: $\forall f \in \mathcal{K}(G), \forall y \in G: I(f \circ L_y) = I(f)$.

If μ is a left Haar measure on G, then the mapping

$$I_\mu: f \mapsto \int f(x)d\mu(x): \mathcal{K}(G) \to \mathbb{R}$$

is a left Haar integral. Conversely, if I *is a left Haar integral on* $\mathcal{K}(G)$ *then the regular Borel measure* μ *which is associated with* I *according to the RIESZ Representation Theorem (see* 2.8) *is a left Haar measure.* [Note that (ii) $\Longleftrightarrow$ (vii); (vi) $\Rightarrow$ (iii) obviously. In order to prove (v) $\Rightarrow$ (vi), let $f \in \mathcal{K}(G)$, $f \geq 0$ and $f \neq 0$, say $f(x_0) > 0$. Then the set

$$U := \{x \in G \mid f(x) > \tfrac{1}{2}f(x_0)\}$$

is non-void and open, so that $\mu(U) > 0$ by (v). Since $f \geq \frac{1}{2}f(x_0)\chi_U$, it follows that $I(f) = \int f d\mu \geq \frac{1}{2}f(x_0)\int \chi_U d\mu > 0$.]
A similar argument can be used to prove the following property of a Haar integral:

(viii) If $f: G \to \mathbb{R}$ is continuous, $f \geq 0$ and $I(f) = 0$, then $f = 0$.

3.2. Existence and unicity of left Haar measures

We have observed above that there is a 1-1 correspondence between left Haar measures and left Haar integrals. As to existence and unicity, we have the following:

THEOREM. *There exists a left Haar measure* μ *on* G. *Moreover,* μ *is unique in the following sense: if* ν *is another left Haar measure on* G *then there exist a constant* $c > 0$ *such that* $\nu = c\mu$, *or equivalently:*

$$\int f(x)d\nu(x) = c\int f(x)d\mu(x), \qquad f \in \mathcal{K}(G).$$

For a proof of this theorem, we refer to HEWITT & ROSS [3;pp.186-194] or to NACHBIN [6]. It should be noticed, that this theorem holds for every locally compact group (not necessarily satisfying the second axiom of countability). For historical remarks about this theorem see [3; the *Notes* at the end of §15].

In the sequel, we shall use the following notation: if $f:G \to S$ is any function on G (S an arbitrary set), then

$$f_y := f\circ L_y : x \mapsto f(yx) : G \to S$$

for any $y \in G$.

3.3. Existence and unicity of right Haar measures

Similar to the left Haar measure and integral one can also introduce the concept of a *right Haar measure* and a *right Haar integral*. We leave the definitions to the reader. The following argument shows how existence and unicity for right Haar integral (or measure) follow from the corresponding properties of left Haar integral (measure).

Let us denote in this subsection integration with respect to left Haar measure with $d_\ell x$. If $f \in K(G)$ then the function $\tilde{f}: G \to \mathbb{R}$, defined by

$$\tilde{f}(x) := f(x^{-1}), \quad x \in G ,$$

is continuous and has compact support. So we can define a mapping $I_r: K(G) \to \mathbb{R}$ by

$$I_r(f) := \int \tilde{f}(x) d_\ell x, \quad f \in K(G) .$$

Then I_r is a non-zero, positive linear functional, and

$$I_r(f \circ R_a) = \int \widetilde{f \circ R_a}(x) d_\ell x = \int f(x^{-1}a) d_\ell x = \int \tilde{f}(a^{-1}x) d_\ell x =$$

$$= \int \tilde{f}(x) d_\ell x = I_r(f)$$

for every $f \in K(G)$ and $a \in G$. (Here $R_y(x) := xy$, $x, y \in G$.) Hence I_r is a right Haar integral.

Next, suppose that J' is any right Haar integral. Then similar to the above argument one shows that $f \mapsto J'(\tilde{f}): K(G) \to \mathbb{R}$ is a left Haar integral. Hence there exists a constant $c > 0$ such that

$$J'(\tilde{f}) = c \int f(x) d_\ell (x)$$

for all $f \in K(G)$. In particular,

$$c\, I_r(f) = c \int \tilde{f}(x) d_\ell x = J'(\tilde{\tilde{f}}) = J'(f)$$

for all $f \in K(G)$, that is, $J' = c\, I_r$. This proves unicity of the right Haar integral.

3.4. Examples of Haar integrals

We shall give now some examples of left and right Haar integrals. Observe that for certain groups the left and the right Haar integrals are identical (cf. the examples (a), (b), (d) and (f)) while on other groups they are different.

(a). Let $G := \mathbb{R}$, with its ordinary additive group structure and its usual topology. Then a left and right ($\mathbb{R}$ is abelian!) Haar integral is given by

$$I(f) := \int_{\mathbb{R}} f(x)\,dx, \qquad f \in K(\mathbb{R}),$$

where dx denotes ordinary Riemann (or, if you want, Lebesgue) integration.

(b). Let $G := \mathbb{T} = \{e^{i\alpha} \mid 0 \le \alpha < 2\pi\}$. With ordinary multiplication of complex numbers, $\mathbb{T}$ is a compact topological group. Then a left and right ($\mathbb{T}$ is abelian!) Haar integral is given by

$$I(f) := \frac{1}{2\pi}\int_0^{2\pi} f(e^{i\alpha})\,d\alpha, \qquad f \in K(G),$$

where $d\alpha$ denotes ordinary Riemann integration.

(c). Let G be any topological group with the following three properties:

(i) As a topological space G is an open subset of $\mathbb{R}^n$ for some $n \ge 1$.

(ii) For $x = (x_1,\dots,x_n)$ and $y = (y_1,\dots,y_n)$ in G, the product xy is thus a function $F(x_1,\dots,x_n,y_1,\dots,y_n)$ which maps $G \times G \subseteq \mathbb{R}^{2n}$ into $G \subseteq \mathbb{R}^n$. Our hypothesis is, that all of the partial derivatives

$$\frac{\partial F_j}{\partial x_k}, \qquad \frac{\partial F_j}{\partial y_k},$$

exist and are continuous on $G \times G$. (Here F_j is the j-th coordinate function of F.)

Now for every $a \in G$ homeomorphisms L_a and R_a of G onto intself are defined by $L_a(x) := ax$ and $R_a(x) = xa$, respectively. Observe, that $L_e = R_e = id_G$, and

$$(3.1) \qquad L_{ab} = L_a \circ L_b, \qquad R_{ab} = R_b \circ R_a.$$

Since each L_a is continuously differentiable by (ii), we can define a continuous function $\lambda\colon G \times G \to \mathbb{R}^+$ by

$$\lambda(a,x) := |J(L_a)(x)| = \left|\det\left(\frac{\partial F_j(a_1,\ldots,a_n,y_1,\ldots,y_n)}{\partial y_k}\bigg|_x\right)\right| .$$

(Here $J(\tau)(x)$ denotes the Jacobian of a transformation $\tau\colon G \to G$, evaluated at $x \in G$.) Since L_e is the identity mapping on G, it follows that

(3.2) $\qquad \lambda(e,x) = 1$

for every $x \in G$. Using the familiar Jacobian identity $J(\tau'\circ\tau)(x) = J(\tau')(\tau x)\,.\,J(\tau)(x)$ for transformations $\tau,\tau'\colon G \to G$ and $x \in G$, we conclude from (3.1) that

(3.3) $\qquad \lambda(ab,x) = \lambda(a,bx)\lambda(b,x)$.

In particular, (3.2) and (3.3) imply $1 = \lambda(b^{-1},bx)\lambda(b,x)$, so that $\lambda(b,x) > 0$ for every $b, x \in G$. It follows that the function $a \mapsto \lambda(a,e)$ is a continuous function of G into $\mathbb{R}^+\backslash\{0\}$, and (by (3.3))

(3.4) $\qquad \lambda(ab,e) = \lambda(a,b)\lambda(b,e)$.

We shall form now the left Haar integral on $K(G)$ by using the function λ together with Riemann integration in $\mathbb{R}^n$. Given a continuous real-valued function ϕ on G vanishing outside a compact set, let $\int_G \phi(x)dx$ denote the ordinary n-dimensional Riemann integral of ϕ over the open subset G of $\mathbb{R}^n$. Now the idea is, that there has to be a recompensation for the change of a volume element under left translation, which is given by multiplication by the Jacobian of left translation. Thus, we divide beforehand the volume element by this Jacobian. In fact, we claim that

(3.5) $\qquad I_\ell(f) := \displaystyle\int_G f(x)\,\frac{1}{\lambda(x,e)}\,dx, \quad f \in K(G),$

is a left Haar integral on $K(G)$. To prove this assertion, we make use of the familiar formula for transformation of multiple integrals. Applied to G and L_a, this formula is

$$\int_{L_a(G)} \phi(x)dx = \int_G \phi(L_a(y))|J(L_a)(y)|dy,$$

where ϕ is, let us say, any function in $K(G)$. Applied to the function $\phi: x \mapsto \frac{f(a^{-1}x)}{\lambda(x,e)}$ with $f \in K(G)$ this yields

$$I_\ell(f_{a^{-1}}) = \int_G \frac{f(a^{-1}x)}{\lambda(x,e)}\, dx = \int_G \frac{f(y)}{\lambda(ay,e)}\, \lambda(a,y)dy$$

$$= \int_G f(y)\, \frac{1}{\lambda(y,e)}\, dy = I_\ell(f),$$

where we have used (3.4) and the fact that $L_a(G) = G$. Since I_ℓ is obviously a strictly positive linear function on $K(G)$, this proves our claim.

The construction of a right Haar integral proceeds in a completely similar way. Thus, we set

$$\rho(a,x) := |J(R_a)(x)| = \left|\det\left(\frac{\partial F_j(y_1,\ldots,y_n,a_1,\ldots,a_n)}{\partial y_k}\Big|_x\right)\right|$$

for a, $x \in G$. Then $\rho(e,x) = 1$ and, by (3.1),

(3.6) $\qquad \rho(ab,x) = \rho(b,xa)\rho(a,x)$

for all a, b, $x \in G$. Using this, one shows readily that

(3.7) $\qquad I_r(f) := \int_G f(x)\, \frac{1}{\rho(x,e)}\, dx, \quad f \in K(G)$,

defines a right Haar integral on $K(G)$.

(d). We apply (c) to the multiplicative group $\mathbb{R}\backslash\{0\}$. The transformation $L_a: x \mapsto ax$ has Jacobian a, so that the (left and right) Haar integral has the form

$$I(f) = \int_{-\infty}^{\infty} \frac{f(x)}{|x|}\, dx,$$

where for each $f \in K(G)$ the integral is actually extended over a union of two compact intervals, containing the support of f but not containing 0 (say, $[-\beta,-\alpha] \cup [\alpha,\beta]$ with $\beta > \alpha > 0$, depending on f).

(e). Next, we apply (c) to the group of all matrices $\begin{pmatrix} x & y \\ 0 & 1 \end{pmatrix}$ with $x, y \in \mathbb{R}$ and $x \neq 0$. For convenience we write the elements of G as points

$(x,y) \in \mathbb{R}^2$. Then G becomes an open subset of $\mathbb{R}^2$, and $(x,y)(u,v) = (xu, xv+y)$. So if we topologize G as a subset of $\mathbb{R}^2$, then G is a topological group, and the hypothesis of (c) is satisfied. For $(a,b) \in G$, we have $L_{(a,b)}(x,y) = (ax, ay+b)$, and the Jacobian of $L_{(a,b)}$ is $\det\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix} = a^2$. Thus, a left Haar integral on G is provided by

$$I_\ell(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f(x,y)}{x^2}\, dxdy, \quad f \in K(G),$$

where the integral is, for each fixed f, actually extended over a bounded open subset of $\mathbb{R}^2$ that does not intersect the line $\{(x,y) \in \mathbb{R}^2 \mid x = 0\}$ and that includes supp(f).

The transformation $R_{(a,b)}: G \to G$ has the form $R_{(a,b)}(x,y) = (ax, bx+y)$; the Jacobian is $\det\begin{pmatrix} a & 0 \\ b & 1 \end{pmatrix} = a$. So a right Haar integral is given by

$$I_r(f) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{f(x,y)}{|x|}\, dxdy.$$

(f). Let G be the group $GL(n,\mathbb{R})$. The element $X = (x_{ij})$ can be identified with a (column) vector in $\mathbb{R}^{n^2}$ by writing the columns of X below each other. Then G is identified with the open subset $\{X \in \mathbb{R}^{n^2} \mid \det(X) \neq 0\}$ of $\mathbb{R}^{n^2}$. (Since det is a continuous function on $\mathbb{R}^{n^2}$, this set is open in $\mathbb{R}^{n^2}$ indeed.) For $A \in GL(n,\mathbb{R})$ we have

$$L_A(X) = AX \approx \begin{pmatrix} A & 0 & \dots & 0 \\ 0 & A & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & A \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix},$$

where X_i denotes the i-th column of X. This shows that the Jacobian of L_A equals $\det(A)^n$. So a Haar integral on $GL(n,\mathbb{R})$ has the form

$$I(f) = \int \frac{f(X)}{|\det(X)|^n}\, dX, \quad f \in K(GL(n,\mathbb{R})),$$

the integral for every $f \in K(GL(n,\mathbb{R}))$ being extended over a compact subset of $\mathbb{R}^{n^2}$ which is disjoint from the set $\{X \mid \det(X) = 0\}$.

Since the Jacobian of R_A is also $\det(A)^n$, the above integral is not only a left, but also a right Haar integral.

(g). Let G be the group of all upper triangular n×n matrices

$$\begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ 0 & x_{22} & x_{23} & \cdots & x_{2n} \\ 0 & 0 & x_{33} & \cdots & x_{3n} \\ 0 & 0 & 0 & \cdots & x_{nn} \end{pmatrix}$$

with determinant $x_{11}x_{22} \ldots x_{nn} \neq 0$. Similar to what has been said under (f) for $GL(n,\mathbb{R})$, this group can be regarded as an open subset of $\mathbb{R}^{\frac{1}{2}n(n+1)}$ (omit all zero's). Then multiplication with respect to this parametrization takes the following form:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots\cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & x_{13} & \cdots & x_{1n} \\ 0 & x_{22} & x_{23} & \cdots & x_{2n} \\ 0 & 0 & x_{33} & \cdots & x_{3n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots\cdots & x_{nn} \end{pmatrix} \approx$$

$$\left(\begin{array}{c|c|c|c|c} a_{11} & 0 & 0 & \cdots & 0 \\ \hline 0 & \begin{matrix} a_{11} & a_{12} \\ 0 & a_{22} \end{matrix} & 0 & \cdots & 0 \\ \hline 0 & 0 & \begin{matrix} a_{11} & a_{12} & a_{13} \\ 0 & a_{22} & a_{23} \\ 0 & 0 & a_{33} \end{matrix} & & 0 \\ \hline & & & \ddots & \\ \hline 0 & 0 & 0 & & \begin{matrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \vdots \\ & & & a_{nn} \end{matrix} \end{array}\right) \cdot \left(\begin{array}{c} x_{11} \\ \hline x_{12} \\ x_{22} \\ \hline x_{13} \\ x_{23} \\ x_{33} \\ \hline \vdots \\ \hline x_{1n} \\ x_{2n} \\ \vdots \\ x_{nn} \end{array}\right).$$

Therefore for every $A \in G$ the Jacobian of L_A equals $a_{11}^{n} a_{22}^{n-1} \ldots . a_{nn}$. Consequently, the left Haar integral on G is given by

$$I_{\ell}(f) = \int \frac{f(X)}{|x_{11}^{n} x_{22}^{n-1} \cdot \ldots \cdot x_{nn}|}\, dX, \quad f \in \mathcal{K}(G),$$

where for each $f \in \mathcal{K}(G)$ the integral extends over some bounded open domain in $\mathbb{R}^{\frac{1}{2}n(n+1)}$ which is disjoint from the set $\{X \in G \mid x_{11}x_{22} \ldots x_{nn} = 0\}$.

A right Haar integral is

$$I_r(f) = \int \frac{f(X)}{|x_{11}x_{22}^2 \cdot \ldots \cdot x_{nn}^n|}\, dX, \quad f \in K(G).$$

3.5. The separability of $L^2(G)$

For the remainder of this section we make the following conventions. When we speak of Haar measure or Haar integral, we shall mean *left* Haar measure and *left* Haar integral respectively, unless stated otherwise. Integration with respect to any Haar measure is denoted by $\int f(x)dx$. The symbol μ will be reserved for any selected Haar measure.

If G is *compact*, then the function $\chi_G: x \mapsto 1: G \to \mathbb{R}^+$ is in $K(G)$, so $\mu(G) = \int \chi_G d\mu$ is finite and non-zero. Since the Haar measure is determined up to a multiplicative constant, we can choose it such that we have $\mu(G) = 1$. This (unique!) Haar measure is called *normed Haar measure*, and the corresponding integral *normed Haar integral*.

Observe that a Borel function on G which is integrable with respect to some Haar measure is integrable with respect to all other Haar measures. Thus, the space $L^1(G)$ (and similarly, $L^p(G)$ for $p \geq 1$) is independent of the choice of a particular Haar measure; only the norm in the space depends on this choice; however, this norm is obviously uniquely determined up to a multiplicative constant. Consequently, the properties of $L^p(G)$ as a Banach space (and, in particular, of $L^2(G)$ as a Hilbert space) do not depend on the choice of a particular Haar measure.

We shall show now that *the Hilbert space* $L^2(G)$ *is separable because* G *satisfies the second axiom of countability*. [Recall that a Hilbert space H is separable whenever there exists a countable dense subset in H. It is well-known that H is separable if and only if H has an orthonormal basis which is finite or countably infinite.]

Sketch of proof: Each $f \in L^2(G)$ can be written as the difference of two non-negative functions in $L^2(G)$, and for each $f \in L^2(G)$ such that $f \geq 0$ it follows easily from the statement at the end of 1.1 and Lebesgue's dominated convergence theorem that f can be approximated (in the L^2 -norm) with integrable simple functions. Therefore, it is sufficient to show that there is a countable set F of functions in $L^2(G)$ such that each χ_A with A a Borel set of finite measure can be approximated by members of F. Indeed, then the (countable!) set of all linear combinations of members of F with *rational* coefficients is dense in $L^2(G)$ by the above remarks.

Since G satisfies the second axiom of countability, there is a countable collection $\mathcal{U}$ of open sets in G such that each open set in G can be obtained as a union of members of $\mathcal{U}$. In view of local compactness of G we may and shall assume that $\overline{U}$ is compact for every $U \in \mathcal{U}$. Let $\mathcal{U}^*$ denote the collection of all pairs (V,U) such that V and U are unions of a finite number of elements of $\mathcal{U}$ and $\overline{V} \subset U$. Then $\mathcal{U}^*$ is countable. Note, that if $(V,U) \in \mathcal{U}^*$, then $\overline{V}$ is compact. Therefore, by 2.1 we can select for each $(V,U) \in \mathcal{U}^*$ an $f \in \mathcal{K}(G)$ such that $0 \leq f(x) \leq 1$ for all $x \in G$, $f(x) = 1$ for $x \in \overline{V}$ and $f(x) = 0$ for $x \notin U$. [N.B.: if G is a Lie group, then we can select such an f which is also a C^∞-function.] Let $\mathcal{F}$ denote the collection of functions in $\mathcal{K}(G)$, selected in this way (one for each pair (V,U) in $\mathcal{U}^*$); then $\mathcal{F}$ is countable.

Now let A be a Borel set in G with $\mu(A) < \infty$, and let $\varepsilon > 0$. By regularity of μ there are a compact subset K of G and an open subset W of G such that $K \subseteq A \subseteq W$ and $\mu(W\backslash K) < \varepsilon$. For each $x \in K$ there are $V_x \in \mathcal{U}$, $U_x \in \mathcal{U}$ with $x \in V_x \subseteq \overline{V}_x \subseteq U_x \subseteq W$ (an easy consequence of the properties of $\mathcal{U}$). Now K can be covered by finitely many of these V_x's, and in this way we obtain $(V,U) \in \mathcal{U}^*$ such that $K \subseteq V \subseteq U \subseteq W$. For the corresponding $f \in \mathcal{F}$ we have

$$\int |f-\chi_A|^2(x)dx \leq \int |\chi_W(x)-\chi_K(x)|^2 dx = \mu(W\backslash K) < \varepsilon.$$

Thus, $\mathcal{F}$ has the desired property. □

<u>REMARKS</u>. 1°. It follows from the proof that $L^2(G)$ has a countable dense subset, consisting of members of $\mathcal{K}(G)$, and if G is a Lie group, we have even members of $\mathcal{K}(G) \cap C^\infty(G)$. Now the Gram-Smith orthogonalization process implies:

The Hilbert space $L^2(G)$ *has a countable orthogonal basis consisting of continuous functions with compact support. If* G *is a Lie group the elements of the base may also supposed to be in* $C^\infty(G)$.

2°. The above proof applies to *any* Hilbert space $L^2(X,\mu)$ with X a locally compact space, satisfying the second axiom of countability, and μ a regular Borel measure on X. There is even a more general result: If $(X,\mathcal{S},\mu)$ is a σ-finite measure space such that the σ-algebra $\mathcal{S}$ is generated by a countable family of subsets of X, then $L^2(X,\mu)$ is separable. For a proof, see HALMOS [1; §40, Theorem B] together with either HALMOS [1; §42, Exercise 1] or ZAANEN [9; §20, Theorem 2].

3^o. In a similar way, one shows that $L^1(G)$ is a separable (Banach-) space.

3.6. The left invariance of the norm in $L^p(G)$

If we apply the result of 2.11 to the homeomorphisms $L_y: G \to G$, defined by $L_y(x) := yx\ (x,y \in G)$, we obtain the following theorem, which implies, among others, that $\|f_y\|_p = \|f\|_p$ for all $f \in L^p(G)$.

THEOREM. *Let $1 \le p < \infty$ and let f be a Borel function on G. Then $f \in L^p(G)$ if and only if for every $y \in G$ the function $f_y: x \mapsto f(yx)$ is in $L^p(G)$. In that case,*

$$\int |f(x)|^p dx = \int |f_y(x)|^p dx = \int |f(y\,x)|^p dx.$$

In particular, if $f \in L^1(G)$, then $\int f(y\,x)\,dx = \int f(x)dx$. Moreover (in the case $p = 2$), the mapping $U_y: f \mapsto f_y: L^2(G) \to L^2(G)$ is a linear, unitary operator for each $y \in G$.

PROOF. Immediate from 2.11, applied to the homeomorphisms $L_y: G \to G$, defined by $L_y(x) = yx\ (y \in G)$. □

3.7. The Haar modulus

For every $f \in K(G)$ and $y \in G$ the function $x \mapsto f(xy^{-1})$ is continuous and has compact support (namely Ky, where K is the support of f). Therefore, we can write down the expression

$$J_y(f) := \int f(xy^{-1})dx, \quad f \in K(G), \quad y \in G.$$

It is easily seen that $J_y: K(G) \to \mathbb{R}$ is a linear and strictly positive mapping. Finally, it is an easy exercise to show that J_y is left invariant, i.e., $J_y(f_z) = J_y(f)$ for all $z \in G$ and $f \in K(G)$. Stated otherwise, J_y is a left Haar integral. So the unicity of Haar integral implies that there is a constant $\Delta_G(y) > 0$ such that $J_y(f) = \Delta_G(y)J_e(f)$, where J_e equals, of course, our original Haar integral.

Thus, we have shown that there exists a function $\Delta_G: G \to \mathbb{R}^+$ such that $\Delta_G(y) > 0$ and

$$(3.8) \qquad \int f(xy^{-1})dx = \Delta_G(y) \int f(x)dx$$

for every $y \in G$ *and* $f \in K(G)$.

This function Δ_G is called the *modular function* on G (also: the *Haar modulus* on G). If no confusion will arise, we shall write Δ instead of Δ_G. It is easy to see that Δ does not depend on the particular choice of the left Haar measure μ. Obviously, $\Delta(y) = 1$ for every $y \in G$ if and only if the left Haar integral is also a right Haar integral. In that case the group G is called *unimodular*.

Every abelian group is unimodular (because right and left translation coalesce), and we shall show in 3.9 below that every compact group is unimodular.

3.8. Properties of the Haar modulus

The following properties of Δ *are important:*

(i) $\Delta(y) > 0$ *for every* $y \in G$.

(ii) $\Delta(yz) = \Delta(y)\Delta(z)$ *for every* $y, z \in G$, *and* $\Delta(e) = 1$.

(iii) $\Delta\colon G \to \mathbb{R}^+$ *is continuous.*

PROOF. The property mentioned in (i) follows from what has been said in 3.7. Property (ii) follows from a straightforward computation, which we leave to the reader. We present a proof of (iii):

It follows easily from (ii) that it is sufficient to show that Δ is continuous at e. Let U be any neighbourhood of e such that $\bar{U}$ is compact. Then, by continuity of the multiplication, there exists a neighbourhood V of e such that $V^2 := \{xy \mid x\in V \& y\in V\} \subseteq U$. By 2.1 there exists $f \in K(G)$ such that $f \geq 0$, $f(e) = 1$ and $f(x) = 0$ for $x \notin V$. Take $\varepsilon > 0$. Since f has compact support, f is *uniformly continuous* on G, that is, there exists a neighbourhood W of e (without restriction of generality $W \subseteq V$) such that

$$|f(u) - f(v)| < \varepsilon \quad \text{whenever} \quad u^{-1} v \in W.$$

If $y \in W$ then we have $(xy^{-1})^{-1} x \in W$, so $|f(xy^{-1}) - f(x)| < \varepsilon$ for all $x \in G$. Since $f(xy^{-1}) = f(x) = 0$ for all $x \in G\backslash\bar{U}$ and all $y \in W$ it follows that for all $y \in W$ we have

$$\left| \int f(xy^{-1})dx - \int f(x)dx \right| \le \int |f(xy^{-1}) - f(x)|dx =$$

$$= \int_{\bar{U}} |f(xy^{-1}) - f(x)|dx < \varepsilon\, \mu(\bar{U}).$$

Since $\Delta(y) = (\int f(xy^{-1})dx)(\int f(x)dx)^{-1}$, it follows that

$$|\Delta(y)-1| < \frac{\varepsilon\mu(\bar{U})}{\int f(x)dx} \quad \text{for all } y \in W.$$

As U and f are fixed in this discussion (independent of ε) this proves continuity of Δ at e. □

3.9. Unimodularity of compact groups

THEOREM. *If* H *is a compact subgroup of* G, *then* $\Delta(x) = 1$ *for every* $x \in H$. *In particular, it follows that every compact group is unimodular.*

PROOF. It follows from 3.8 that $\Delta(H)$ is a compact subgroup of the multiplicative group $R^+\backslash\{0\}$. Since $\{1\}$ is the only such subgroup, it follows that $\Delta(H) = \{1\}$, that is, $\Delta(x) = 1$ for every $x \in H$. If G is compact, apply this result to $H := G$. □

3.10. The transformation of an integration variable into its inverse

Let us denote in this subsection integration with respect to left Haar measure by $d_\ell x$. Recall from 3.3 that the equation

$$(3.9) \qquad I_r(f) := \int \tilde{f}(x)d_\ell(x) = \int f(x^{-1})d_\ell x, \quad f \in K(G),$$

defines a right Haar integral I_r, and that every right Haar integral on G is of the form $c\, I_r$ for some constant $c > 0$.

For the particular right Haar integral I_r, defined by (3.9), we have

$$(3.10) \qquad \int f(x^{-1})d_\ell x = \int f(x)d_r x, \quad f \in K(G),$$

where $d_r x$ denotes integration with respect to the right Haar measure, corresponding to I_r. Now it follows from 3.7 that for every $f \in K(G)$ and $a \in G$ we have

$$I_r(f\circ L_a) = \int \widetilde{f\circ L_a}(x)d_\ell x = \int \tilde{f}(xa^{-1})d_\ell x = \Delta(a) \int \tilde{f}(x)d_\ell x = \Delta(a)I_r(f).$$

Stated otherwise,

$$(3.11) \qquad \int f(ax)d_r x = \Delta(a) \int f(x)d_r x, \quad f \in K(G), \quad a \in G.$$

Observe, that this is true for *any* right Haar integral, also for right Haar integrals that are not related to $d_\ell x$ as in formula (3.10). Indeed, this is an obvious consequence of unicity of the right Haar integral. Next, we shall show that the following is true: if $f \in K(G)$, then

$$(3.12) \qquad \int f(x^{-1})d_\ell x = \int f(x)\Delta(x^{-1})d_\ell x.$$

[This can also be written as

$$\int f(x)d_r x = \int f(x)\Delta(x^{-1})d_\ell x$$

provided the left and right Haar integrals are related as in (3.10). Notice that this means that $d_r x = \Delta(x)^{-1}d_\ell x$. Thus, if μ_r and μ_ℓ are the corresponding right and left Haar measure, then $\mu_r \equiv \mu_\ell$ (indeed, $\Delta > 0$; cf.1.10).]

The proof of (3.12) proceeds as follows. As we have seen above, the left-hand side of (3.12) defines a right Haar measure. Since, however, for $f \in K(G)$ and $a \in G$,

$$\int f(xa)\Delta(x^{-1})d_\ell x = \Delta(a^{-1})\int f(x)\Delta(ax^{-1})d_\ell x = \int f(x)\Delta(x^{-1})d_\ell x,$$

the right-hand side of (3.12) defines a right Haar integral as well. By unicity of right Haar integral it follows that there is $c > 0$ such that for all $f \in K(G)$,

$$(3.13) \qquad c\int f(x^{-1})d_\ell x = \int f(x)\Delta(x^{-1})d_\ell x.$$

We have to show now that $c = 1$. First, observe that for any $f \in K(G)$ such that $f(x) = f(x^{-1})$ for all $x \in G$ formula (3.13) implies

$$(3.14) \qquad (1-c)\int f(x)d_\ell x = \int f(x)(1-\Delta(x^{-1}))d_\ell x.$$

Now let $\varepsilon > 0$ and let U be a neighbourhood of e such that $|\Delta(x^{-1})-1| < \varepsilon$ for all $x \in U$ (continuity of Δ). By 2.1, there is $g \in K(G)$, $0 \le g \ne 0$, such that $\mathrm{supp}(g) \subseteq U \cap U^{-1}$. Then for $f := g+\tilde{g}$ we have, according to (3.14),

$$|c-1| \cdot \int f(x)d_\ell x \le \varepsilon \cdot \int f(x)d_\ell x.$$

Since $\int f(x)d_\ell x \neq 0$, it follows that $|c-1| < \varepsilon$. As this holds for every $\varepsilon > 0$, the desired result follows, that is, (3.12) is true.

Finally, observe that for any $f \in K(G)$, it follows from (3.10) and (3.12), applied to $\tilde{f}$ and $f.\Delta$:

$$\int f(x^{-1})d_r x = \int f(x)\Delta(x)d_r x. \tag{3.15}$$

Again by unicity of right Haar integral, this holds for every right Haar integral. If the left and right Haar integrals are related as in (3.10), this is equivalent to

$$\int f(x)d_\ell x = \int f(x)\Delta(x)d_r x, \quad f \in K(G). \tag{3.16}$$

Resuming, we have proved the following formulas for all $f \in K(G)$ and $y \in G$:

$$\int f(xy^{-1})d_\ell x = \Delta(y)\int f(x)d_\ell x; \tag{3.17}$$

$$\int f(yx)d_r x = \Delta(y)\int f(x)d_r x; \tag{3.18}$$

$$\int f(x^{-1})d_\ell x = \int f(x)\Delta(x^{-1})d_\ell x; \tag{3.19}$$

$$\int f(x^{-1})d_r x = \int f(x)\Delta(x)d_r x. \tag{3.20}$$

If the left and the right Haar integral are related by

$$\int f(x^{-1})d_\ell x = \int f(x)d_r x, \tag{3.21}$$

then we have also for all $f \in K(G)$:

$$\int f(x)d_r x = \int f(x)\Delta(x^{-1})d_\ell x; \tag{3.22}$$

$$\int f(x)d_\ell x = \int f(x)\Delta(x)d_r x. \tag{3.23}$$

3.11. Examples of Haar moduli

Let G satisfy the hypothesis of 3.4(c). Then it follows from (3.16) and the formulas (3.5), (3.7) that

$$\int f(x)\frac{1}{\lambda(x,e)}dx = \lambda \int f(x)\Delta(x)\frac{1}{\rho(x,e)}dx$$

for all $f \in K(G)$, where γ is some constant. (We do not know beforehand that these particular left and right Haar integral are related as in (3.10)!) By an argument very similar to the one used in 3.10 it follows that for every $x \in G$,

$$\left| \frac{1}{\lambda(x,e)} - \gamma \frac{\Delta(x)}{\rho(x,e)} \right| < \varepsilon$$

for every $\varepsilon > 0$. Hence $\gamma\Delta(x) = \rho(x,e)/\lambda(x,e)$. Since $\Delta(e) = \rho(e,e) = \lambda(e,e) = 1$, it follows that $\gamma = 1$, hence

$$(3.24) \qquad \Delta(x) = \frac{\rho(x,e)}{\lambda(x,e)}, \quad x \in G.$$

We apply this to a few groups, mentioned in 3.4.

(a) The groups $\mathbb{R}$, $\mathbb{T}$ and $\mathbb{R}\setminus\{0\}$ are unimodular (they are abelian). Although $GL(n,\mathbb{R})$ for $n \geq 2$ is not abelian, it is unimodular, because $\lambda(X,I) = \rho(X,I) = |\det(X)|^n$ (cf.3.4(f)).

(b) Let G be the group of all upper triangular $n\times n$ matrices X with $\det(X) \neq 0$ (cf.3.5(g)). Then

$$\Delta_G(X) := \frac{x_{11}x_{22}^2\cdot\ldots\cdot x_{nn}^n}{x_{11}^n x_{22}^{n-1}\cdot\ldots\cdot x_{nn}},$$

where $x_{11}, x_{22}, \ldots, x_{nn}$ are the entries of the main diagonal of $X \in G$. Clearly, G is not unimodular. Since G is a closed subgroup*) of $GL(n,\mathbb{R})$, this example shows that a closed subgroup of a unimodular group need not be unimodular.

(c) Let G be the group of all matrices $\begin{pmatrix} x & y \\ 0 & 1 \end{pmatrix}$ with $x \neq 0$ (see 3.4(e)). Then

---

*) As an abstract group, G is a subgroup of $GL(n,\mathbb{R})$, and it is not difficult to show that the topology of G as described in 3.4(g) coincides with the topology of G as a subgroup of $GL(n,\mathbb{R})$. See also the remarks after Definition IV.2.11.

$$\Delta_G\begin{pmatrix} x & y \\ 0 & 1 \end{pmatrix} = \frac{|x|}{x^2} = \frac{1}{|x|}$$

for all $(x,y) \in \mathbb{R}^2$, $x \neq 0$. So G is not unimodular.

3.12. The Haar measure on a direct product

A straightforward application of FUBINI's theorem as it is formulated in 2.12 yields the following.

THEOREM. *For* $i = 1,2$, *let* G_i *be a locally compact group, satisfying the second axiom of countability, with left Haar measure* μ_i *and modular function* Δ_i. *Then the product measure* $\mu_1 \otimes \mu_2$ *is a left Haar measure on the product group* $G_1 \times G_2$, *and for the modular functions the following relationship holds:*

$$\Delta_{G_1 \times G_2}(x,y) = \Delta_1(x)\Delta_2(y), (x,y) \in G_1 \times G_2.$$

COROLLARY. *A product of unimodular groups is unimodular.*

3.13. The Haar measure on a normal subgroup

For details of the following discussion, we refer to HEWITT & ROSS [3; 15.20 & 15.21] or REITER [7; III.3.3]. Let H be a closed normal subgroup of G. Then H is also a locally compact group, satisfying the second axiom of countability, so there is a (left) Haar measure ν on H. The corresponding Haar integral will be denoted $\int h(\xi)d\xi$, $h \in K(H)$. Notice, that ν is in general not the restriction of the Haar measure μ of G to H, that is, usually it is not true that $\nu(A) = \mu(A)$ for every Borel set A of G with $A \subseteq H$. Indeed, it may happen that $\mu(H) = 0$ (whereas we know that $\nu(H) > 0$, cf.3.1(v)). [EXAMPLE: $\mathbb{R}$ as a closed normal subgroup of $\mathbb{R}^2$.] We shall derive a relationship beteen Haar measure on G and H and on the quotient group G/H. Recall that the elements of G/H are the left cosets xH, $x \in G$, which we shall denote by [x]. If G/H is given the finest topology making the mapping $x \mapsto [x]: G \to G/H$ continuous, then G/H becomes a locally compact topological group satisfying the second axiom of countability. The left Haar integral on G/H will be denoted $\int_{G/H} h[x]d[x]$, $h \in K(G/H)$.

If $f \in K(G)$ and $x \in G$, then $f_x: \xi \mapsto f(x\xi): H \to \mathbb{R}$ is a continuous function on H with compact support, so

$$f'(x) := \int_H f_x(\xi)d\xi$$

is well-defined. Using left invariance of this integral (on H) it follows easily that $f'(x\eta) = f'(x)$ for every $x \in G$ and $\eta \in H$, that is, f' is constant on each coset xH. Hence there exists a function $f^*: G/H \to \mathbb{R}$ such that $f^*[x] = f'(x)$ for every $x \in G$. It can be shown that $f^* \in K(G/H)$ for every $f \in K(G)$. Hence we can define the mapping $f \mapsto \int_{G/h} f^*[x]d[x]$ $K(G) \to \mathbb{R}$, which is easily seen to be a left invariant, non-zero positive linear functional on $K(G)$. Hence it is proportional to the original Haar integral on G, and, after a suitable renorming of the Haar integral on G/H, we have equality, that is:

$$(3.25)\qquad \int_{G/H}\left(\int_H f(x\xi)d\xi\right)d[x] = \int_G f(x)dx, \quad f \in K(G).$$

(WEIL's formula).

3.14. The Haar modulus on a normal subgroup

THEOREM. *If H is a closed normal subgroup of G, then* $\Delta_H(\eta) = \Delta_G(\eta)$ *for all* $\eta \in H$. *In particular, if G is unimodular then so is H.*

PROOF. If $\eta \in H$ then for every $f \in K(G)$ we have, by (3.25):

$$\Delta_G(\eta) \int_G f(x)dx = \int_G f(x\eta^{-1})dx = \int_{G/H}\left(\int_H f(x\xi\eta^{-1})d\xi\right)d[x] =$$

$$= \int_{G/H}\left(\Delta_H(\eta) \int_H f(x\xi)d\xi\right)d[x] = \Delta_H(\eta) \int_G f(x)dx.$$

Taking $f \in K(G)$ such that $\int_G f(x)dx \neq 0$, the result follows. □

4. HAAR INTEGRALS ON LIE GROEPS

4.1. The Haar measure and modulus on a Lie group which can be covered by one chart

Essentially, the Haar integral on Lie groups has already been treated in the examples 3.4(c) and 3.11. In order to see this, we consider for a moment a Lie group G which can be covered by one chart (U,ϕ). Thus, $U = G$ and $\phi: G \to \mathbb{R}^n$ is a homeomorphism of G onto an open subset of $\mathbb{R}^n$. If we identify G with $\phi(G)$, then G becomes an open subset of $\mathbb{R}^n$, and the multiplication in G is such that the hypothesis of 3.4(c) applies to G. It should be clear now what the function $x \mapsto \lambda(x,e): G \to \mathbb{R}^+$ means:

(4.1) $\lambda(x,e) = |\det(dL_x)_e|, \quad x \in G,$

where $dL_x : T_e(G) \to T_x(G)$ is the differential of the left translation $L_x : y \mapsto xy : G \to G$ (compare the definition of $\lambda: G\times G \to \mathbb{R}^+$ in 3.4(c) with what has been said about differentials of mappings in and just above definition 2.3 in Chapter IV). Similarly, $\rho(x,e) = |\det(dR_x)_e|$, so that (see formula (3.24) in 3.11):

$$\Delta_G(x) = \frac{|\det(dR_x)_e|}{|\det(dL_x)_e|} = \frac{1}{|\det((dR_{x^{-1}})_x \circ (dL_x)_e)|} = \frac{1}{|\det d(R_{x^{-1}} \circ L_x)_e|}\ .$$

However, $(R_{x^{-1}} \circ L_x)(y) = xyx^{-1} = \alpha(x)(y)$ (see (IV.2.37)), hence $d(R_{x^{-1}} \circ L_x) = d(\alpha(x)) = \mathrm{Ad}(x)$. Consequently,

$$\Delta_G(x) = |\det \mathrm{Ad}(x)|^{-1}.$$

4.2. <u>The Haar measure on a general Lie group</u>

Now we consider the case that G can not be covered by one chart. Then the method of 3.4(c) and 3.11 to compute the Haar integral and the modular function, respectively, remains valid as long as we restrict ourselves to functions $f \in K(G)$ and elements $a \in G$ such that f and f_a have their supports in U, where (U,ϕ) is a chart such that $e \in U$.

Fix such a chart an let V be a symmetric neighbourhood of e with $V^2 \subseteq U$. We shall show now how a left invariant integral on $K(G)$ can be defined. The definition is in two steps.

<u>Step 1</u>. If $f \in K(G)$ has a support $\mathrm{supp}(f) \subseteq V$, then the integral (n-dimensional Lebesgue integral over open $U \subseteq \mathbb{R}^n$)

(4.2) $$I_\ell(f) := \int_U f(x)\frac{1}{\lambda(x,e)}\,dx$$

is well-defined. Moreover, if $a \in V$, then $f \circ L_a$ has support $a^{-1}\mathrm{supp}(f) \subseteq V^2 \subseteq U$, so $I_\ell(f \circ L_a)$ is well-defined, and the method of 3.4(c) shows that $I_\ell(f \circ L_a) = I_\ell(f)$.

<u>Step 2</u>. If $f \in K(G)$ is arbitrary, then by the use of a so-called *partition of unity* (see 4.3 below) it can be shown that f can be written as $f = f_1 + f_2 + \ldots + f_k$, with $f_i \in K(G)$ and such, that there exist points $x_1, x_2, \ldots, x_k \in G$ with $\mathrm{supp}(f_i) \subseteq x_iV$ for $i = 1,\ldots,k$. Then $\mathrm{supp}(f_i \circ L_{x_i}) \subseteq V$ so $I_\ell(f_i \circ L_{x_i})$ is defined according to (4.2). Now set

$$(4.3) \qquad I_\ell(f) := \sum_{i=1}^{k} I_\ell(f_i \circ L_{x_i}) = \sum_{i=1}^{k} \int_U f_i(x_i x) \frac{1}{\lambda(x,e)}\, dx .$$

It can be shown that *the value of the right-hand side of* (4.3) *is independent of the way in which* f *is written as a sum of members* f_i *of* $K(G)$ *with support contained in translates of* V. *In particular, it is independent of the choice of the points* $x_i \in G$, *provided* $\text{supp}(f_i) \subseteq x_i V$. For a proof, see REITER [7; III. 3.1(v)]; in the proof an essential use is made of the "local" left invariance of I_ℓ as described in Step 1. Hence the mapping I_ℓ: $K(G) \to \mathbb{R}$ is defined in an unambiguous way, and it is easily seen that I_ℓ is linear, strictly positive and left invariant. [We prove left invariance: if $f \in K(G)$, say $f = f_1+f_2+ \dots +f_k$ and $\text{supp}(f_i) \subseteq x_i V$ for $i = 1,\dots,k$, then for any $a \in G$ we have $f \circ L_a = f_1 \circ L_a + f_2 \circ L_a + \dots + f_k \circ L_a$ and $\text{supp}(f_i \circ L_a) \subseteq a^{-1}\text{supp}(f_i) \subseteq a^{-1} x_i V$ for $i = 1,\dots,k$. By the definition of I_ℓ (in particular, the aforementioned independence) we have

$$\begin{aligned} I_\ell(f \circ L_a) &= \sum_{i=1}^{k} I_\ell((f_i \circ L_a) \circ L_{a^{-1}x_i}) \\ &= \sum_{i=1}^{k} I_\ell(f_i \circ L_{x_i}) = I_\ell(f) . \end{aligned}$$

This proves left invariance of I_ℓ.]

Consequently, the functional I_ℓ: $K(G) \to \mathbb{R}$ constructed above is a left Haar integral. In particular, for functions $f \in K(G)$ with $\text{supp}(f) \subseteq V$ the Haar integral is given by (4.2).

4.3. Partition of unity of a Lie group

In the construction, outlined in 4.2, we have used the following: *Let* G *be a Lie group,* $f \in K(G)$ *and* V *a neighbourhood of* e *in* G. *Then there are* $f_i \in K(G)$ *and* $a_i \in G$ *such that* $f = f_1 + \dots + f_n$ *and* $\text{supp}(f_i) \subseteq a_i V$ *for* $i = 1,\dots,n$.

Sketch of proof: In view of 2.1 there is for every $x \in \text{supp}(f)$ an element $h_x \in K(G)$ such that $0 \le h_x \le 1$, $h_x(x) = 1$ and $\text{supp}(h_x) \subseteq xV$. Set $W_x := \{y \in G \mid h_x(y) > \tfrac{1}{2}\}$. Then $\{W_x \mid x \in \text{supp}(f)\}$ is an open covering of $\text{supp}(f)$, which has a finite subcovering, say $\{W_{a_1},\dots,W_{a_n}\}$. Set for $j = 1,\dots,n$ and $x \in G$:

$$f_j(x) := \frac{h_{a_j}(x) f(x)}{\sum_{i=1}^{n} h_{a_i}(x)} .$$

Then $f_1,\dots,f_n$ have the desired properties. (cf.[8], p.10, for more details).

4.4. The Haar measure on a Lie group which is almost covered by one chart

Let notation be as in 4.2. If U is such that $G\backslash U$ consists of a countable union of submanifolds of G of dimension $\le \dim(G)-1$, then $\mu(G\backslash U) = 0$ (μ is left Haar measure).

[Sketch of proof: it is sufficient to show that $\mu(M) = 0$ if M is a submanifold of G of dimension $\le \dim(G)-1$. Since such an M can be covered by countably many left translates of any sufficiently small neighbourhood V of e, and since μ is left invariant, it is sufficient to prove that $\mu(V\cap M) = 0$ if M is such a manifold and $e \in M$. In terms of the local coordinates in V, this means that we want to show that $\int_V \chi_M(x)\, \frac{1}{\lambda(x,e)}\, dx = 0$, where V is an open subset of $\mathbb{R}^n$ and M is an at most (n-1)-dimensional submanifold of $\mathbb{R}^n$. After a suitable transformation of coordinates, we obtain an integral of the form $\int \chi_N(y) g(y) dy$, where $N \subseteq \{x \in \mathbb{R}^n \mid x_n = 0\}$. See Def.IV. 2.10. This integral is clearly zero.]

In this case, $\int_G f d\mu = \int_U f d\mu$ for every $f \in K(G)$ (even for every $f \in L^1(G)$), so now the Haar integral on G is completely given by 4.1. An example of this situation is provided by the group SU(2): this group is homeomorphic to S^3, hence it has a chart (U,ϕ) such that $SU(2)\backslash U$ consists of one point. For integration with respect to the left Haar integral we are allowed to think of SU(2) as if it were covered by one chart.

4.5. The Haar modulus on a general Lie group

Let notation be as in 4.2. Using the fact that formula (4.2) presents the Haar integral for functions $f \in K(G)$ with $\mathrm{supp}(f) \subseteq V$, it follows easily from 4.1. that

$$\Delta_G(a) = |\det(\mathrm{Ad}\ a)|^{-1} \tag{4.4}$$

for all $a \in V$. Observe, that both $a \mapsto \Delta_G(a)$ and $a \mapsto |\det(\mathrm{Ad}\ a)|^{-1}$ are homeomorphisms from the group G to the multiplicative group $\mathbb{R}^+\backslash\{0\}$. It follows from Prop.IV.1.11 that for every $x \in G_0$, the connected component of e in G, there is a finite number of elements $a_1,\dots,a_p \in V$ such that $x = a_1 \dots a_p$. Consequently,

$$\Delta_G(x) = \prod_{i=1}^{p} \Delta_G(a_i)$$

$$= \prod_{i=1}^{p} |\det(\mathrm{Ad}\ a_i)|^{-1} = |\det(\mathrm{Ad}\ x)|^{-1}.$$

Stated otherwise, (4.4) holds for every $a \in G_0$. However, we shall show that (4.4) holds for every $a \in G$.

To this end, fix $a \in G$. Then aVa^{-1} is a neighbourhood of e, so there exists $f \in \mathcal{K}(G)$ such that

$$\text{(4.5)} \qquad \mathrm{supp}(f) \subseteq V \cap (aVa^{-1}),$$

and, in addition, $0 \le f(x) \le 1$ for all $x \in G$ and $f(e) > 0$.

Recall from formula (IV, 2.37) that $\alpha(a)(x) := axa^{-1}$ for all $x \in G$. It is easy to see that $\mathrm{supp}(f\circ\alpha(a)) = a^{-1}.\ \mathrm{supp}(f).a$, hence $\mathrm{supp}(f\circ\alpha(a)) \subseteq V$ by (4.5). So by (4.2),

$$I_{\ell}(f\circ\alpha(a)) = \int_U \frac{f(axa^{-1})}{\lambda(x,e)}\, dx$$

$$= \int_{aUa^{-1}} \frac{f(x)}{\lambda(a^{-1}xa,e)}\, |\det J(\alpha(a^{-1}))(x)|\, dx$$

Since the support of f is contained in U, the domain of integration can be taken to be U. Moreover, recall that $J(\alpha(a^{-1}))(e)$, the Jacobian of the mapping $\alpha(a^{-1})$ at e, is nothing but the differential $d\alpha(a^{-1})$ at e, that is, $J(\alpha(a^{-1}))(e) = \mathrm{Ad}(a^{-1}) = (\mathrm{Ad}\ a)^{-1}$; (see IV.2.43). Therefore, we can write the above result in the following form:

$$\text{(4.6)} \qquad I_{\ell}(f\circ\alpha(a)) = |\det(\mathrm{Ad}\ a)|^{-1} \int_U \frac{f(x)}{\lambda(x,e)}\ \frac{\lambda(x,e)\,|\det J(\alpha(a^{-1}))(x)|}{\lambda(a^{-1}xa,e)\,|\det J(\alpha(a^{-1}))(e)|}\, dx.$$

Now the second factor in the integrand is continuous in x, and has value 1 for x = e. So for every $\varepsilon > 0$ there exists a neighbourhood W_ε of e such that its value differs less than ε from 1, for every $x \in W_\varepsilon$. If we choose f such that, in addition to the earlier conditions, also $\mathrm{supp}(f) \subseteq W_\varepsilon$, then in (4.6) only the points x in W_ε contribute to the value of the integral. It follows then, that

(4.7) $$\left| I_\ell(f\circ\alpha(a)) - |\det(\mathrm{Ad}\ a)|^{-1} I_\ell(f) \right| \leq$$

$$\leq |\det(\mathrm{Ad}\ a)|^{-1} \int_U \frac{f(x)}{\lambda(x,e)} \cdot \left| \frac{\lambda(x,e)\,|\det J(\alpha(a^{-1}))(x)|}{\lambda(a^{-1}xa,e)\,|(\det J(\alpha(a^{-1}))(e)|} - 1 \right| dx$$

$$< |\det(\mathrm{Ad}\ a)|^{-1} .\ I_\ell(f).\varepsilon.$$

On the other hand, $\alpha(a) = R_{a^{-1}}\circ L_a$, and in view of the left invariance of the functional I_ℓ, it follows from the definition of $\Delta_G(a)$ that

$$I_\ell(f\circ\alpha(a)) = I_\ell(f\circ R_{a^{-1}}) = \Delta_G(a) I_\ell(f).$$

If we substitute this in (4.7), we see that

$$\left| \Delta_G(a) I_\ell(f) - |\det(\mathrm{Ad}\ a)|^{-1} I_\ell(f) \right| < \varepsilon \frac{I_\ell(f)}{|\det(\mathrm{Ad}\ a)|} .$$

Since $I_\ell(f) \neq 0$, we obtain

$$\left| \Delta_G(a) - |\det(\mathrm{Ad}\ a)|^{-1} \right| < \varepsilon\ |\det(\mathrm{Ad}\ a)|^{-1} ,$$

and, since ε is arbitrary, the desired result follows, namely, that (4.4) is valid for every $a \in G$.

4.6. The unimodularity of semisimple and nilpotent Lie groups

THEOREM. *The following types of Lie groups are unimodular:*

(i) *all semisimple Lie groups,*

(ii) *all nilpotent Lie groups.*

PROOF.

(i) Let G be a semisimple Lie group. It follows that the Killing form κ on the Lie algebra $\mathfrak{g}$ of G is non-degenerate (cf. Theorem IV.1.3). Therefore,

(4.8) $$\kappa(X,Y) = (TX,Y),\ X,\ Y \in \mathfrak{g},$$

where $(.,.)$ denotes an inproduct on $\mathfrak{g}$ and $T: \mathfrak{g} \to \mathfrak{g}$ is linear with $\det T \neq 0$. (Any non-degenerate bilinear form on a finite-dimensional linear space can be written in this form.) Since κ is invariant under all automorphisms of $\mathfrak{g}$, we have in particular

(4.9) $\kappa((\mathrm{Ad}x)X, (\mathrm{Ad}x)Y) = \kappa(X,Y)$

for all $X, Y \in \mathfrak{g}$ and $x \in G$. Using (4.8), it follows that

$$((T\circ \mathrm{Ad}x)X, (\mathrm{Ad}x)Y) = (TX,Y)$$

for all $X, Y \in \mathfrak{g}$, whence $(\mathrm{Ad}x)^*\circ T\circ(\mathrm{Ad}x) = T$ for every $x \in G$. Taking determinants we obtain $|\det(\mathrm{Ad}x)|^2 = 1$, because $\det T \neq 0$. By (4.4), this means that $\Delta(x)^2 = 1$ for every $x \in G$. Since $\Delta > 0$, the desired result follows.

(ii) Let W be a neighbourhood of e such that exp induces a bijection of a neighbourhood of 0 in $\mathfrak{g}$ onto W. Then for $x \in W$ there is a unique $X \in \mathfrak{g}$ such that $x = \exp X$. In view of (4.4) and Theorem IV.2.15(d) it follows that

$$\Delta(x) = |\det(\mathrm{Ad}\ x)| = |\det \mathrm{Ad}(\exp X)|$$
$$= |\det e^{\mathrm{ad}X}| = e^{\mathrm{tr}(\mathrm{ad}X)},$$

where $\mathrm{tr}(\mathrm{ad}X)$ denotes the trace of the linear operator adX on $\mathfrak{g}$. Since adX is a nilpotent operator, it follows that $\mathrm{tr}(\mathrm{ad}X) = 0$, Hence $\Delta(x) = 1$ for $x \in W$. However, nilpotent Lie groups are connected (by definition), so using Proposition IV. 1.11 and 3.8(ii), it follows that $\Delta(x) = 1$ for every $x \in G$. □

LITERATURE

[1] HALMOS, P.R., *Measure theory*, D. Van Nostrand Co., New York, 1950.

[2] HELGASON, S., *Differential geometry and symmetric spaces*, Academic Press, New York-London, 1962.

[3] HEWITT, E. & K.A. ROSS, *Abstract harmonic analysis I*, Springer-Verlag, Berlin, 1963.

[4] HEWITT, E. & K. STROMBERG, *Real and abstract analysis*, Springer-Verlag, Berlin, 1965.

[5] MUKHERJEA, A. & K. POTHOVEN, *Real and functional analysis*, Plenum Press, New York-London, 1978.

[6] NACHBIN, L., *The Haar integral,* Princeton University Press, Princeton, 1965.

[7] REITER, H., *Classical harmonic analysis and locally compact groups,* Oxford University Press, Oxford, 1968.

[8] WARNER, F.W., *Foundations of differentiable manifolds and Lie groups,* Scott, Foresman and Company, Glenview (Ill.), 1971.

UITGAVEN IN DE SERIE MC SYLLABUS

Onderstaande uitgaven zijn verkrijgbaar bij het Mathematisch Centrum, 2e Boerhaavestraat 49 te Amsterdam-1005, tel. 020-947272.

MCS 1.1 F. GÖBEL & J. VAN DE LUNE, *Leergang Besliskunde, deel 1: Wiskundige basiskennis*, 1965. ISBN 90 6196 014 2.

MCS 1.2 J. HEMELRIJK & J. KRIENS, *Leergang Besliskunde, deel 2: Kansberekening*, 1965. ISBN 90 6196 015 0.

MCS 1.3 J. HEMELRIJK & J. KRIENS, *Leergang Besliskunde, deel 3: Statistiek*, 1966. ISBN 90 6196 016 9.

MCS 1.4 G. DE LEVE & W. MOLENAAR, *Leergang Besliskunde, deel 4: Markovketens en wachttijden*, 1966. ISBN 90 6196 017 7.

MCS 1.5 J. KRIENS & G. DE LEVE, *Leergang Besliskunde, deel 5: Inleiding tot de mathematische besliskunde*, 1966. ISBN 90 6196 018 5.

MCS 1.6a B. DORHOUT & J. KRIENS, *Leergang Besliskunde, deel 6a: Wiskundige programmering 1*, 1968. ISBN 90 6196 032 0.

MCS 1.6b B. DORHOUT, J. KRIENS & J.TH. VAN LIESHOUT, *Leergang Besliskunde, deel 6b: Wiskundige programmering 2*, 1977. ISBN 90 6196 150 5.

MCS 1.7a G. DE LEVE, *Leergang Besliskunde, deel 7a: Dynamische programmering 1*, 1968. ISBN 90 6196 033 9.

MCS 1.7b G. DE LEVE & H.C. TIJMS, *Leergang Besliskunde, deel 7b: Dynamische programmering 2*, 1970. ISBN 90 6196 055 X.

MCS 1.7c G. DE LEVE & H.C. TIJMS, *Leergang Besliskunde, deel 7c: Dynamische programmering 3*, 1971. ISBN 90 6196 066 5.

MCS 1.8 J. KRIENS, F. GÖBEL & W. MOLENAAR, *Leergang Besliskunde, deel 8: Minimaxmethode, netwerkplanning, simulatie*, 1968. ISBN 90 6196 034 7.

MCS 2.1 G.J.R. FÖRCH, P.J. VAN DER HOUWEN & R.P. VAN DE RIET, *Colloquium Stabiliteit van differentieschema's, deel 1*, 1967. ISBN 90 6196 023 1.

MCS 2.2 L. DEKKER, T.J. DEKKER, P.J. VAN DER HOUWEN & M.N. SPIJKER, *Colloquium Stabiliteit van differentieschema's, deel 2*, 1968. ISBN 90 6196 035 5.

MCS 3.1 H.A. LAUWERIER, *Randwaardeproblemen, deel 1*, 1967. ISBN 90 6196 024 X.

MCS 3.2 H.A. LAUWERIER, *Randwaardeproblemen, deel 2*, 1968. ISBN 90 6196 036 3.

MCS 3.3 H.A. LAUWERIER, *Randwaardeproblemen, deel 3*, 1968. ISBN 90 6196 043 6.

MCS 4 H.A. LAUWERIER, *Representaties van groepen*, 1968. ISBN 90 6196 037 1.

MCS 5 J.H. VAN LINT, J.J. SEIDEL & P.C. BAAYEN, *Colloquium Discrete wiskunde*, 1968. ISBN 90 6196 044 4.

MCS 6 K.K. KOKSMA, *Cursus ALGOL 60*, 1969. ISBN 90 6196 045 2.

MCS 7.1 *Colloquium Moderne rekenmachines, deel 1*, 1969. ISBN 90 6196 046 0.

MCS 7.2 *Colloquium Moderne rekenmachines, deel 2*, 1969. ISBN 90 6196 047 9.

MCS 8 H. BAVINCK & J. GRASMAN, *Relaxatietrillingen*, 1969. ISBN 90 6196 056 8.

MCS 9.1 T.M.T. COOLEN, G.J.R. FÖRCH, E.M. DE JAGER & H.G.J. PIJLS, *Elliptische differentiaalvergelijkingen, deel 1*, 1970. ISBN 90 6196 048 7.

MCS 9.2 W.P. VAN DEN BRINK, T.M.T. COOLEN, B. DIJKHUIS, P.P.N. DE GROEN, P.J. VAN DER HOUWEN, E.M. DE JAGER, N.M. TEMME & R.J. DE VOGELAERE, *Colloquium Elliptische differentiaalvergelijkingen, deel 2*, 1970. ISBN 90 6196 049 5.

MCS 10 J. FABIUS & W.R. VAN ZWET, *Grondbegrippen van de waarschijnlijkheidsrekening*, 1970. ISBN 90 6196 057 6.

MCS 11 H. BART, M.A. KAASHOEK, H.G.J. PIJLS, W.J. DE SCHIPPER & J. DE VRIES, *Colloquium Halfalgebra's en positieve operatoren*, 1971. ISBN 90 6196 067 3.

MCS 12 T.J. DEKKER, *Numerieke algebra*, 1971. ISBN 90 6196 068 1.

MCS 13 F.E.J. KRUSEMAN ARETZ, *Programmeren voor rekenautomaten; De MC ALGOL 60 vertaler voor de EL X8*, 1971. ISBN 90 6196 069 X.

MCS 14 H. BAVINCK, W. GAUTSCHI & G.M. WILLEMS, *Colloquium Approximatietheorie*, 1971. ISBN 90 6196 070 3.

MCS 15.1 T.J. DEKKER, P.W. HEMKER & P.J. VAN DER HOUWEN, *Colloquium Stijve differentiaalvergelijkingen, deel 1*, 1972. ISBN 90 6196 078 9.

MCS 15.2 P.A. BEENTJES, K. DEKKER, H.C. HEMKER, S.P.N. VAN KAMPEN & G.M. WILLEMS, *Colloquium Stijve differentiaalvergelijkingen, deel 2*, 1973. ISBN 90 6196 079 7.

MCS 15.3 P.A. BEENTJES, K. DEKKER, P.W. HEMKER & M. VAN VELDHUIZEN, *Colloquium Stijve differentiaalvergelijkingen, deel 3*, 1975. ISBN 90 6196 118 1.

MCS 16.1 L. GEURTS, *Cursus Programmeren, deel 1: De elementen van het programmeren*, 1973. ISBN 90 6196 080 0.

MCS 16.2 L. GEURTS, *Cursus Programmeren, deel 2: De programmeertaal ALGOL 60*, 1973. ISBN 90 6196 087 8.

MCS 17.1 P.S. STOBBE, *Lineaire algebra, deel 1*, 1974. ISBN 90 6196 090 8.

MCS 17.2 P.S. STOBBE, *Lineaire algebra, deel 2*, 1974. ISBN 90 6196 091 6.

MCS 17.3 N.M. TEMME, *Lineaire algebra, deel 3*, 1976. ISBN 90 6196 123 8.

MCS 18 F. VAN DER BLIJ, H. FREUDENTHAL, J.J. DE IONGH, J.J. SEIDEL & A. VAN WIJNGAARDEN, *Een kwart eeuw wiskunde 1946-1971, Syllabus van de Vakantiecursus 1971*, 1974. ISBN 90 6196 092 4.

MCS 19 A. HORDIJK, R. POTHARST & J.Th. RUNNENBURG, *Optimaal stoppen van Markovketens*, 1974. ISBN 90 6196 093 2.

MCS 20 T.M.T. COOLEN, P.W. HEMKER, P.J. VAN DER HOUWEN & E. SLAGT, *ALGOL 60 procedures voor begin- en randwaardeproblemen*, 1976. ISBN 90 6196 094 0.

MCS 21 J.W. DE BAKKER (red.), *Colloquium Programmacorrectheid*, 1975. ISBN 90 6196 103 3.

MCS 22 R. HELMERS, F.H. RUYMGAART, M.C.A. VAN ZUYLEN & J. OOSTERHOFF, *Asymptotische methoden in de toetsingstheorie; Toepassingen van naburigheid*, 1976. ISBN 90 6196 104 1.

MCS 23.1 J.W. DE ROEVER (red.), *Colloquium Onderwerpen uit de biomathematica, deel 1*, 1976. ISBN 90 6196 105 X.

MCS 23.2 J.W. DE ROEVER (red.), *Colloquium Onderwerpen uit de biomathematica, deel 2*, 1976. ISBN 90 6196 115 7.

MCS 24.1 P.J. VAN DER HOUWEN, *Numerieke integratie van differentiaalvergelijkingen, deel 1: Eenstapsmethoden*, 1974. ISBN 90 6196 106 8.

MCS 25 *Colloquium Structuur van programmeertalen*, 1976. ISBN 90 6196 116 5.

MCS 26.1 N.M. TEMME (ed.), *Nonlinear analysis, volume 1*, 1976. ISBN 90 6196 117 3.

MCS 26.2 N.M. TEMME (ed.), *Nonlinear analysis, volume 2*, 1976. ISBN 90 6196 121 1.

MCS 27 M. BAKKER, P.W. HEMKER, P.J. VAN DER HOUWEN, S.J. POLAK & M. VAN VELDHUIZEN, *Colloquium Discretiseringsmethoden*, 1976. ISBN 90 6196 124 6.

MCS 28 O. DIEKMANN, N.M. TEMME (EDS), *Nonlinear Diffusion Problems*, 1976. ISBN 90 6196 126 2.

MCS 29.1 J.C.P. BUS (red.), *Colloquium Numerieke programmatuur, deel 1A, deel 1B*, 1976. ISBN 90 6196 128 9.

MCS 29.2 H.J.J. TE RIELE (red.), *Colloquium Numerieke programmatuur, deel 2*, 1976. ISBN 144 0.

* MCS 30 P. GROENEBOOM, R. HELMERS, J. OOSTERHOFF & R. POTHARST, *Efficiency begrippen in de statistiek*, ISBN 90 6196 149 1.

MCS 31 J.H. VAN LINT (red.), *Inleiding in de coderingstheorie*, 1976. ISBN 90 6196 136 X.

MCS 32 L. GEURTS (red.), *Colloquium Bedrijfssystemen*, 1976. ISBN 90 6196 137 8.

MCS 33 P.J. VAN DER HOUWEN, *Differentieschema's voor de berekening van waterstanden in zeeën en rivieren*, 1977. ISBN 90 6196 138 6.

MCS 34 J. HEMELRIJK, *Oriënterende cursus mathematische statistiek*, ISBN 90 6196 139 4.

MCS 35 P.J.W. TEN HAGEN (red.), *Colloquium Computer Graphics*, 1977. ISBN 90 6196 142 4.

MCS 36 J.M. AARTS, J. DE VRIES, *Colloquium Topologische Dynamische Systemen*, 1977. ISBN 90 6196 143 2.

MCS 37 J.C. van Vliet (red.), *Colloquium Capita Datastructuren*, 1978. ISBN 90 6196 159 9.

MCS 38.1 T.H. Koornwinder (ed.), *Representations of locally compact groups with applications*, 1979. ISBN 90 6196 161 0.

MCS 38.2 T.H. Koornwinder (ed.), *Representations of locally compact groups with applications*, 1979. ISBN 90 6196 181 5.

MCS 39 O.J. Vrieze & G.L. Waanrooij, *Colloquium Stochastic spelen*, 1978. ISBN 90 6196 167 X.

De met een * gemerkte uitgaven moeten nog verschijnen.