Graduate Texts in Mathematics 105

Springer
New York
Berlin
Heidelberg
Barcelona
Budapest
Hong Kong
London
Milan
Paris
Santa Clara
Singapore
Tokyo

Springer Books on Elementary Mathematics by Serge Lang

MATH! Encounters with High School Students
1985, ISBN 96129-1

The Beauty of Doing Mathematics
1985, ISBN 96149-6

Geometry. A High School Course (with G. Murrow)
1991, ISBN 96654-4

Basic Mathematics
1988, ISBN 96787-7

A First Course in Calculus, Fifth Edition
1998, ISBN 96201-8

Calculus of Several Variables
1987, ISBN 96405-3

Introduction to Linear Algebra
1988, ISBN 96205-0

Linear Algebra
1989, ISBN 96412-6

Undergraduate Algebra, Second Edition
1990, ISBN 97279-X

Undergraduate Analysis, Second Edition
1997, ISBN 94841-4

Complex Analysis
1993, ISBN 97886-0

Real and Functional Analysis
1993, ISBN 94001-4

Serge Lang

$SL_2(\mathbf{R})$

With 33 Figures

Springer

Serge Lang
Department of Mathematics
Yale University
New Haven, Connecticut 06520
U.S.A.

AMS Subject Classification: 22E46

Library of Congress Cataloging in Publication Data
Lang, Serge
$SL_2(\mathbf{R})$.
(Graduate texts in mathematics; 105)
Originally published: Reading, Mass.: Addison-Wesley, 1975.
Bibliography: p.
Includes index.
1. Lie groups. 2. Representations of groups. I. Title. II. Series.
QA387.L35 1985 512′.55 85-14802

Softcover reprint of the hardcover 1st edition 1985

9 8 7 6 5 4 3 2 (Corrected second printing, 1998)

ISBN-13: 978-1-4612-9581-5 e-ISBN-13: 978-1-4612-5142-2
DOI: 10.1007/978-1-4612-5142-2

Foreword

Starting with Bargmann's paper on the infinite dimensional representations of $SL_2(\mathbf{R})$, the theory of representations of semisimple Lie groups has evolved to a rather extensive production. Some of the main contributors have been: Gelfand–Naimark and Harish-Chandra, who considered the Lorentz group in the late forties; Gelfand–Naimark, who dealt with the classical complex groups, while Harish-Chandra worked out the general real case, especially through the derived representation of the Lie algebra, establishing the Plancherel formula (Gelfand–Graev also contributed to the real case); Cartan, Gelfand–Naimark, Godement, Harish-Chandra, who developed the theory of spherical functions (Godement gave several Bourbaki seminar reports giving proofs for a number of spectral results not accessible otherwise); Selberg, who took the group modulo a discrete subgroup and obtained the trace formula; Gelfand, Fomin, Pjateckii-Shapiro, and Harish-Chandra, who established connections with automorphic forms; Jacquet–Langlands, who pushed through the connection with L-series and Hecke theory. This history is so involved and so extensive that I am incompetent to give a really good account, and I refer the reader to bibliographies in the books by Warner, Gelfand–Graev–Pjateckii-Shapiro, and Helgason for further information. A few more historical comments will be made in the appropriate places in the book.

It is not easy to get into representation theory, especially for someone interested in number theory, for a number of reasons. First, the general theorems on higher dimensional groups require massive doses of Lie theory. Second, one needs a good background in standard and not so standard analysis on a fairly broad scale. Third, the experts have been writing for each other for so long that the literature is somewhat labyrinthine.

I got interested because of the obvious connections with number theory, principally through Langlands' conjecture relating representation theory to elliptic curves [La 2]. This is a global conjecture, in the adelic theory. I realized soon enough that it was best to acquire a good understanding of the real theory before getting everything on the adeles. I think most people who have worked in representations have looked at $SL_2(\mathbf{R})$ first, and I know this is the case for both Harish and Langlands.

Therefore, as I learned the theory myself it seemed a good idea to write up $SL_2(\mathbf{R})$. The topics are as follows:

1. We first show how a representation decomposes over the maximal compact subgroup K consisting of all matrices

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

and see that an irreducible representation decomposes in such a way that each character of K (indexed by an integer) occurs at most once.

2. We describe the Iwasawa decomposition $G = ANK$, from which most of the structure and theorems on G follow. In particular, we obtain representations of G induced by characters of A.

3. We discuss in detail the case when the trivial representation of K occurs. This is the theory of spherical functions. We need only Haar measure for this, thereby making it much more accessible than in other presentations using Lie theory, structure theory, and differential equations.

4. We describe a continuous series of representations, the induced ones, some of which are unitary.

5. We discuss the derived representation on the Lie algebra, getting into the infinitesimal theory, and proving the uniqueness of any possible unitarization. We also characterize the cases when a unitarization is possible, thereby obtaining the classification of Bargmann. Although not needed for the Plancherel formula, it is satisfying to know that any unitary irreducible representation is infinitesimally isomorphic to a subrepresentation of an induced one from a quasicharacter of the diagonal group. The derived representation of the Lie algebra on the algebraic space of K-finite vectors plays a crucial role, essentially algebraicizing the situation.

6. The various representations are related by the Plancherel inversion formula by Harish-Chandra's method of integrating over conjugacy classes.

7. We give a method of Harish-Chandra to unitarize the "discrete series," i.e. those representations admitting a highest and lowest weight vector in the space of K-finite vectors.

8. We discuss the structure of the algebra of differential operators, with special cases of Harish-Chandra's results on $SL_2(\mathbf{R})$ giving the center of the universal enveloping algebra and the commutator of K. At this point, we have enough information on differential equations to get the one fact about spherical functions which we could not prove before, namely that there are no other examples besides those exhibited in Chapter IV.

The above topics in a sense conclude a first part of the book. The second part deals with the case when we take the group modulo a discrete subgroup. The classical case is $SL_2(\mathbf{Z})$. This leads to inversion formulas and spectral decomposition theorems on $L^2(\Gamma\backslash G)$, which constitute the remaining chapters.

I had originally intended to include the Selberg trace formula over the reals, but in the case of non-compact quotient this addition would have been sizable, and the book was already getting big. I therefore decided to omit it, hoping to return to the matter at a later date.

A good portion of the first part of the book depends only on playing with Haar measure and the Iwasawa decomposition, without infinitesimal considerations. Even when we use these, we are able to carry out the Plancherel formula and the discussion of the various representations without caring whether we have "all" irreducible unitary representations, or "all" spherical functions (although we prove incidentally that we do). A separate chapter deals with those theorems directly involving partial differential equations via the Casimir operator, and analytical considerations using the regularity theorem for elliptic differential equations. The organization of the book is therefore designed for maximal flexibility and minimal *a priori* knowledge. The methods used and the notation are carefully chosen to suggest the approach which works in the higher dimensional case.

Since I address this book to those who, like me before I wrote it, don't know anything, I have made considerable efforts to keep it self-contained. I reproduce the proofs of a lot of facts from advanced calculus, and also several appendices on various parts of analysis (spectral theorem for bounded and unbounded hermitian operators, elliptic differential equations, etc.) for the convenience of the reader. These and my *Real Analysis* form a *sufficient* background.

The Faddeev paper on the spectral decomposition of the Laplace operator on the upper half-plane is an exceedingly good introduction to analysis, placing the latter in a nice geometric framework. Any good senior undergraduate or first year graduate student should be able to read most of it, and I have reproduced it (with the addition of many details left out to more expert readers by Faddeev) as Chapter XIV. Faddeev's method comes from perturbation theory and scattering theory, and as such is interesting for its own sake, as well as to analysts who may know the analytic part and may want to see how it applies in the group theoretic context. Kubota's recent book on Eisenstein series (which appeared while the present book was in production) uses a different method (Selberg–Langlands), and assumes most of the details of functional analysis as known. Therefore, neither Kubota's book nor mine makes the other unnecessary.

It would have been incoherent to expand the present book to a global context with adeles. I hope nevertheless that the reader will be well prepared to move in that direction after having gotten acquainted with $SL_2(\mathbf{R})$. The book by Gelfand–Graev–Pjateckii-Shapiro is quite useful in that respect.

I have profited from discussions with many people during the last two years, some of them at the Williamstown conference on representation theory in 1972. Among them I wish to thank specifically Godement, Harish-Chandra, Helgason, Labesse, Lachaud, Langlands, C. Moore, Sally, Wilfried Schmid, Stein. Peter Lax and Ralph Phillips were of great help in teaching me some PDE. I also thank those who went through the class at Yale and made helpful contributions during the time this book was evolving. I am especially grateful to R. Bruggeman for his careful reading of the manuscript. I also want to thank Joe Repka for helping me with the proofreading.

New Haven, Connecticut
September 1974

Serge Lang

Contents

Notation

To denote the fact that a function is bounded, we write $f = O(1)$. If f, g are two functions on a space X and $g \geqslant 0$, we write $f = O(g)$ if there exists a constant C such that $|f(x)| \leqslant Cg(x)$ for all $x \in X$. If $X = \mathbf{R}$ is the real line, say, the above relation may hold for x sufficiently large, say $x \geqslant x_0$, and then we express this by writing $x \longrightarrow \infty$. Instead of $f = O(g)$, we also use the Vinogradov notation,

$$f \ll g.$$

On a topological space X, $C(X)$ is the space of continuous functions. If X is a C^∞ manifold (nothing worse than open subsets of euclidean space, or something like $SL_2(\mathbf{R})$, with obvious coordinates, will occur), we let $C^\infty(X)$ be the space of C^∞ functions. We put a lower index c to indicate compact support. Hence $C_c(X)$ and $C_c^\infty(X)$ are the spaces of continuous and C^∞ functions with compact support, respectively.

By the way, $SL_2(\mathbf{R})$ is the group of 2×2 real matrices with determinant 1.

An isomorphism is a morphism (in a category) having an inverse in this category. An automorphism is an isomorphism of an object with itself. For instance, a continuous linear automorphism of a normed vector space H is a continuous linear map $A: H \longrightarrow H$ for which there exists a continuous linear map $B: H \longrightarrow H$ such that $AB = BA = I$. A C^∞ isomorphism is a C^∞ mapping having a C^∞ inverse.

If H is a Banach space, we let $\mathrm{En}(H)$ denote the Banach space of continuous linear maps of H into itself. If H is a Hilbert space, we let $\mathrm{Aut}(H)$ be the group of **unitary** automorphisms of H. We let $GL(H)$ be the group of continuous linear automorphisms of H with itself.

If G' is a subgroup of a group G we let

$$G' \backslash G$$

be the space of right cosets of G'. If Γ operates on a set $\mathfrak{H}$, we let

$$\Gamma\backslash\mathfrak{H}$$

be the space of Γ-orbits. Certain right wingers put their discrete subgroup Γ on the right. Gelfand–Graev–Pjateckii-Shapiro and Langlands put it on the left. I agree with the latter, and hope to turn the right wingers into left wingers.

For the convenience of the reader we also include a summary of objects used frequently throughout the book, with a very brief indication of their respective definitions at the end of the book for quick reference.

I General Results

§1. THE REPRESENTATION ON $C_c(G)$

Let G be a locally compact group, always assumed Hausdorff. Let H be a Banach space (which in most of our applications will be a Hilbert space). A **representation** of G in H is a homomorphism

$$\pi: G \longrightarrow GL(H)$$

of G into the group of continuous linear automorphisms of H, such that for each vector $v \in H$ the map of G into H given by

$$x \longmapsto \pi(x)v$$

is continuous. One may say that the homomorphism is **strongly continuous**, the strong topology being the norm topology on the Banach space. [We recall here that the **weak topology** on H is that topology having the smallest family of open sets for which all functionals on H are continuous.]

A representation is called **bounded** if there exists a number $C > 0$ such that $|\pi(x)| \leqslant C$ for all $x \in G$. If H is a Hilbert space and $\pi(x)$ is unitary for all $x \in G$, i.e. preserves the norm, then the representation π is called **unitary**, and is obviously bounded by 1.

For a representation, it suffices to verify the continuity condition above on a dense subset of vectors; in other words:

Let $\pi: G \longrightarrow GL(H)$ be a homomorphism and assume that for a dense set of $v \in H$ the map $x \longmapsto \pi(x)v$ is continuous. Assume that the image of some neighborhood of the unit element e in G under π is bounded in $GL(H)$. Then π is a representation.

This is trivially proved by three epsilons. Indeed, it suffices to verify the continuity at the unit element. Let $v \in H$ and select v_1 close to v such that

$x \mapsto \pi(x)v_1$ is continuous. We then use the triangle inequality

$$|\pi(x)v - v| \leqslant |\pi(x)v - \pi(x)v_1| + |\pi(x)v_1 - v_1| + |v_1 - v|$$

to prove our assertion.

A representation $\pi\colon G \longrightarrow GL(H)$ is locally bounded, i.e. given a compact subset K of G, the set $\pi(K)$ is bounded in $GL(H)$.

Proof. Let K be a compact subset of G. For each $v \in H$ the set $\pi(K)v$ is compact, whence bounded. By the uniform boundedness theorem (*Real Analysis*, VIII, §3) it follows that $\pi(K)$ is bounded in $GL(H)$.

For the convenience of the reader, we recall briefly the uniform boundedness theorem.

Let $\{T_i\}_{i \in I}$ be a family of bounded operators in a Banach space E, and assume that for each $v \in E$ the set $\{T_i v\}_{i \in I}$ is bounded. Then the family $\{T_i\}_{i \in I}$ is bounded, as a subset of $\mathrm{End}(E)$.

Proof. Let C_n be the set of elements $v \in E$ such that

$$|T_i v| \leqslant n, \qquad \text{all } i \in I.$$

Then C_n is closed, and E is the union of the sets C_n. It follows by Baire's theorem that some C_n contains an open ball. Translating this open ball to the origin yields an open ball B such that the union of the sets $T_i(B)$, $i \in I$, is bounded, whence the family $\{T_i\}_{i \in I}$ is bounded, as desired.

We let $C_c(G)$ denote the space of continuous functions on G with compact support. It is an algebra under convolution, i.e. the product is defined by

$$\varphi * \psi(x) = \int_G \varphi(xy^{-1})\psi(y)\, dy,$$

where dy is a Haar measure on G. We shall assume throughout that G is **unimodular**, meaning that left Haar measure is equal to right Haar measure. For any function f on G we denote by f^- the function $f^-(x) = f(x^{-1})$. Then

$$\int f(x)\, dx = \int f(x^{-1})\, dx = \int f^-(x)\, dx.$$

Remark. When G is not unimodular, then by uniqueness of Haar measure, there is a modular function $\Delta\colon G \longrightarrow \mathbf{R}^+$ which is a continuous homomorphism into the positive reals, such that

$$\int_G f(xa)\,dx = \Delta(a)\int_G f(x)\,dx.$$

One then has

$$\int_G f(x^{-1})\Delta(x)\,dx = \int_G f(x)\,dx$$

by an obvious argument. It follows that $\Delta(x)\,dx$ is right Haar measure. The typical non-unimodular group which will concern us, but not until Chapter III, is the group of triangular matrices

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}$$

For this chapter, you can forget about the non-unimodular case.

The modular function occurs in a slightly more general context than above. Let $\tau\colon G \longrightarrow G$ be either an automorphism (group and topological) of G, or an anti-automorphism, meaning

$$(xy)^\tau = y^\tau x^\tau.$$

We write either x^τ or ${}^\tau x$ for the effect of τ on an element $x \in G$. By the invariance of Haar measure, there exists a positive number $\Delta(\tau)$ such that

$$\int_G f(x^\tau)\,dx = \Delta(\tau)\int_G f(x)\,dx,$$

because the expression on the left is a non-trivial invariant positive functional on $C_c(G)$. We have the obvious composition rule

$$\Delta(\tau\sigma) = \Delta(\tau)\,\Delta(\sigma).$$

In many applications, we have $\tau^2 = Id$, and therefore $\Delta(\tau) = 1$, i.e. τ is unimodular. This occurs in the context of matrices, when for instance τ is the transpose.

The basic example of a unimodular group is the group of matrices

$$G = GL_n(\mathbf{R}).$$

The change of variables formula shows that Haar measure on G is equal to

$$\frac{d^+x}{|\det x|^n}$$

where d^+x is Lebesgue measure on the additive space of $n \times n$ matrices. The above measure on $GL_n(\mathbf{R})$ is therefore both right and left Haar measure. Since

$$GL_2^+(\mathbf{R}) = SL_2(\mathbf{R}) \times \mathbf{R}^+,$$

where $GL_2^+(\mathbf{R})$ is the group of 2×2 matrices with positive determinant, and $\mathbf{R}^+$ is the group of positive reals, it follows that left Haar measure on $SL_2(\mathbf{R})$ is also right invariant, i.e. $SL_2(\mathbf{R})$ is unimodular. A better proof is to observe that left and right Haar measures differ by a continuous homomorphism of the group into the positive reals, and that $SL_2(\mathbf{R})$ has no such non-trivial homomorphism. (By looking at conjugacy classes of elements and using various decompositions of $SL_2(\mathbf{R})$ given later in the book, you should be able to work this out as an exercise.) Later we shall give explicit descriptions of the Haar measure on $SL_2(\mathbf{R})$ in terms of various choices of coordinates, and hence we do not stop here for a more thorough discussion.

We return to an arbitrary locally compact group G. Let π be a representation of G in H, and let $\varphi \in C_c(G)$. We define what will be an algebra homomorphism

$$\pi^1 \colon C_c(G) \to \operatorname{End}(H)$$

by letting

$$\pi^1(\varphi)v = \int_G \varphi(x)\pi(x)v\, dx.$$

The integral is defined because $x \mapsto \varphi(x)\pi(x)v$ is a continuous map with compact support from G into H. [If one develops ordinary integration theory in a natural way over the real or complex numbers, one sees that positivity is not needed, only linearity and completeness in the space of values of the functions to be integrated. Cf. my *Real Analysis*, for instance. Thus the integral is the ordinary integral, with values in H.]

Let $a \in G$ and define in this section $\tau_a\varphi(x) = \varphi(a^{-1}x)$. Then the left invariance of Haar measure immediately yields

$$\pi(a)\pi^1(\varphi) = \pi^1(\tau_a\varphi). \tag{1}$$

Furthermore one also sees that π^1 is a homomorphism for the convolution product, i.e.

$$\pi^1(\varphi * \psi) = \pi^1(\varphi)\pi^1(\psi). \tag{2}$$

Indeed,

$$\pi^1(\varphi * \psi) = \int_G (\varphi * \psi)(x)\pi(x)\,dx$$

$$= \int_G\int_G \varphi(xy^{-1})\psi(y)\pi(x)\,dy\,dx.$$

Reversing the order of integration and letting $x \mapsto xy$, this is

$$= \int_G\int_G \varphi(x)\psi(y)\pi(x)\pi(y)\,dx\,dy$$

$$= \pi^1(\varphi)\pi^1(\psi).$$

In the above proof, for simplicity, we omitted placing a vector v to the right of $\pi^1(\varphi * \psi)$, and to the right of every expression inside the integral signs. The integrals are meant in this sense.

Since φ has compact support and π is locally bounded, it follows that $\pi^1(\varphi)$ is a bounded operator, i.e. $\pi^1(\varphi) \in \mathrm{End}(H)$.

If π is a bounded representation, then instead of using functions $\varphi \in C_c(G)$, we could have taken functions $f \in \mathfrak{L}^1(G)$ and formulas (1), (2) remain valid. In other words, π^1 extends to $\mathfrak{L}^1(G)$, and furthermore we have the inequality

$$|\pi^1(f)| \leqq C\|f\|_1. \tag{3}$$

Thus π^1 is a continuous linear homomorphism (**representation**) of $\mathfrak{L}^1(G)$ into $\mathrm{End}(H)$, as Banach algebras.

If H is a Hilbert space, and π is unitary, then we also have the formula

$$\pi^1(\varphi^*) = \pi^1(\varphi)^*, \tag{4}$$

where φ^* is the function such that $\varphi^*(x) = \overline{\varphi(x^{-1})}$. This follows at once from the definition of the symbols involved.

One can recover the values $\pi(a)$ for $a \in G$ by knowing the values $\pi^1(\varphi)$ for $\varphi \in C_c(G)$, as follows. By a **Dirac sequence** on G we mean a sequence of functions $\{\varphi_n\}$, real valued, in $C_c(G)$, satisfying the following properties:

DIR 1. *We have $\varphi_n \geqq 0$ for all n.*

DIR 2. *For all n, we have $\int_G \varphi_n(x)\,dx = 1$.*

DIR 3. *Given a neighborhood V of e in G, the support of φ_n is contained in V for all n sufficiently large.*

The third condition shows that for large n, the area under φ_n is concentrated near the origin. A Dirac sequence looks like Fig. 1.

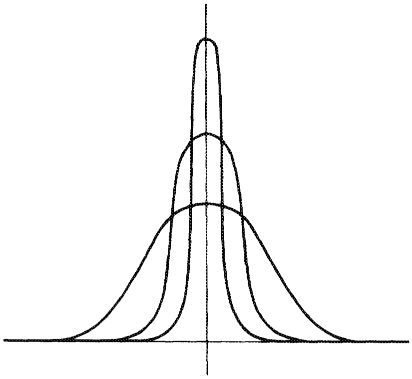

Figure 1

It is obvious that Dirac sequences exist. If G has a C^∞ structure, like $SL_2(\mathbf{R})$, one can even take the functions φ_n to be C^∞. It is frequently convenient to use a slightly weaker condition than **DIR 3**, namely

DIR 3′. *Given a neighborhood V of e in G, and its complement Z, and ϵ, we have*

$$\int_Z \varphi_n(x)\, dx < \epsilon$$

for all n sufficiently large.

In other words, instead of assuming that the supports of the functions φ_n shrink to e, we merely assume the corresponding L^1 condition. It is slightly more intuitive to work with the stronger condition which suffices for almost all applications. When the need arises for the condition **DIR 3′**, we shall assume that the reader can verify for himself the needed convergence statements valid with the same proof as for the other case.

As will be mentioned later when we discuss analytic vectors, the condition **DIR 3′** becomes essential if we want the function φ_n to be analytic functions (they cannot have compact support).

At the beginning of this book, and for several chapters, we are principally interested in the measure theoretic aspects, or the C^∞ aspects, of representations. Consequently we don't need any more about Dirac sequences than their definitions. It may nevertheless be helpful to realize explicitly that some convolutions arising in the classical literature are taken with Dirac sequences. We have the following examples on $\mathbf{R}$.

i) Let φ be a C^∞ function on $\mathbf{R}$ which is positive, has compact support, and is such that

$$\int_{-\infty}^{\infty} \varphi(t)\, dt = 1.$$

Then the sequence $\varphi_n(t) = n\varphi(nt)$ is a Dirac sequence.

ii) Let $\varphi(t) = \pi^{-1/2}e^{-t^2}$. Let φ_n be defined by the same formula as in (i). Then $\{\varphi_n\}$ is a Dirac sequence.

iii) Let

$$\varphi_\epsilon(t) = \frac{1}{\pi}\frac{\epsilon}{t^2 + \epsilon^2}.$$

Then $\{\varphi_\epsilon\}$ is a Dirac family for $\epsilon \longrightarrow 0$ (in an obvious sense, to get the Dirac sequence take $\epsilon = 1/n$).

In cases (ii) and (iii) the factor involving π is there to insure that the integral is equal to 1. The verification that the above are Dirac sequences is at the level of freshman calculus. Note that the examples (ii) and (iii) do not have compact support. Example (ii) is the one which is useful in the discussion of analytic vectors. For a use of Example (iii), see Appendix 2, §3. The Fejer and Poisson kernels in the theory of Fourier series also provide examples of Dirac sequences. The explicit formulas are irrelevant for the basic properties, and we now return to the general properties of Dirac sequences, even reproducing some basic approximation results from *Real Analysis*.

Let $\{\varphi_n\}$ be a Dirac sequence. Then for each $v \in H$, the sequence $\{\pi^1(\varphi_n)v\}$ converges to v.

Proof. We have

$$\int_G \varphi_n(x)\pi(x)v\, dx - v = \int_G [\varphi_n(x)\pi(x) - \varphi_n(x)]v\, dx$$

$$= \int_{S_n} \varphi_n(x)[\pi(x)v - v]\, dx,$$

where S_n is the support of φ_n. From the continuity condition on a representation, it is clear that this last integral tends to 0 as $n \longrightarrow \infty$.

Let $a \in G$. If $\{\varphi_n\}$ is a Dirac sequence, then $\{\tau_a\varphi_n\}$ is a Dirac sequence at a (in the obvious sense). It is clear from (1) that

$$\pi^1(\tau_a\varphi_n)v \longrightarrow \pi(a)v$$

as $n \to \infty$. The value $\pi(a)v$ is therefore obtained as a limit of values $\pi^1(\varphi)v$ for suitable functions $\varphi \in C_c(G)$.

Let W be a subspace of H. (By **subspace** we shall always mean closed subspace unless otherwise specified, in which case we say an **algebraic subspace.**) By a **dense subspace** we mean a dense algebraic subspace. We say that W is G-**invariant** if $\pi(x)W \subset W$ for all $x \in G$. We make a similar definition for $C_c(G)$-invariant.

Quite generally, let S be a family of operators on H. We say that W is S-**invariant** if $AW \subset W$ for every $A \in S$. Let W_0 be a dense algebraic subspace of W. If W_0 is S-invariant, then it is clear that W is also S-invariant.

From the limiting property obtained above, we conclude:

A subspace W of H is G-invariant if and only if W is $C_c(G)$-invariant.

Let $\mathfrak{A}$ be a dense subspace of $\mathfrak{L}^1(G)$ and assume that π is bounded. A subspace W of H is G-invariant if and only if W is also $\mathfrak{A}$-invariant.

For the convenience of the reader, we also recall convergence properties of Dirac convolutions in $\mathfrak{L}^1(G)$.

Let $f \in \mathfrak{L}^1(G)$ and let Z be a compact set on which f is continuous. Let $\{\varphi_n\}$ be a Dirac sequence. Then $\varphi_n * f$ converges to f uniformly on Z.

Proof. We have

$$\varphi_n * f(x) = \int \varphi_n(xy^{-1})f(y)\,dy = \int \varphi_n(y)f(y^{-1}x)\,dy$$
$$f(x) = \int \varphi_n(y)f(x)\,dy.$$

Hence

$$\varphi_n * f(x) - f(x) = \int [f(y^{-1}x) - f(x)]\varphi_n(y)\,dy.$$

There exists a neighborhood U of e in G such that if $y \in U$, then for all $x \in Z$, we have

$$|f(y^{-1}x) - f(x)| < \epsilon.$$

For n large, the support of φ_n is contained in U, whence our integral is concentrated in U, and is obviously estimated by ϵ. This proves our assertion.

The support of $\varphi_n * f$ is contained in (supp φ_n)(supp f), because in the integral for the convolution, we can limit the integral to $xy^{-1} \in \operatorname{supp} \varphi$ and $y \in \operatorname{supp} f$. Hence:

If f is continuous with compact support, then $\{\varphi_n * f\}$ converges uniformly to f on the compact set (supp φ_n)(supp f) and is 0 outside this set. Hence $\{\varphi_n * f\}$ is L^1-convergent to f.

Since $C_c(G)$ is L^1-dense in $\mathfrak{L}^1(G)$, we obtain also:

Let $f \in \mathfrak{L}^1(G)$. Then $\{\varphi_n * f\}$ is L^1-convergent to f.

Proof. First find $\varphi \in C_c(G)$ such that $\|\varphi - f\|_1 < \epsilon$. Then

$$\|\varphi_n * f - f\|_1 \leqslant \|\varphi_n * f - \varphi_n * \varphi\|_1 + \|\varphi_n * \varphi - \varphi\|_1 + \|\varphi - f\|_1.$$

Since $\|g * h\|_1 \leqslant \|g\|_1\|h\|_1$ for two functions $g, h \in \mathfrak{L}^1(G)$, and since $\|\varphi_n\|_1 = 1$ by **DIR 2**, our statement is proved by three epsilons.

The same argument applies to L^p instead of L^1, $1 \leqslant p < \infty$. For our purposes, the most we would want it for is L^2.

§2. A CRITERION FOR COMPLETE REDUCIBILITY

Let

$$\pi: G \longrightarrow GL(H) \qquad \text{and} \qquad \pi': G \longrightarrow GL(H')$$

be representations. A **morphism** of π into π' is a continuous linear map $A: H \longrightarrow H'$ such that for every $x \in G$ the following diagram is commutative.

$$\begin{array}{ccc} H & \xrightarrow{A} & H' \\ {\scriptstyle \pi(x)}\downarrow & & \downarrow{\scriptstyle \pi'(x)} \\ H & \xrightarrow[A]{} & H' \end{array}$$

(In the literature, a morphism is sometimes called an **intertwining operator**.) We say that A is an **embedding** if A is a topological linear isomorphism of H onto a subspace of H'. We say that A is an **isomorphism** if there is a morphism B of π' into π such that AB and BA are the identities of H' and H respectively. An isomorphism is also called an **equivalence**. When H, H' are Hilbert spaces, and π, π' are unitary, then we may deal exclusively with unitary maps, i.e. require that A be unitary. The context will always make it clear whether this additional restriction is intended. We say that π **occurs** in π' if there exists an embedding of π in π'.

A representation $\rho: G \longrightarrow GL(E)$ is called **irreducible** if E has no invariant subspace other than $\{0\}$ and E itself. Let S be a set of operators on E.

We say that E is S-**irreducible** if E has no S-invariant subspace other than $\{0\}$ and E itself.

Let H be a Hilbert space. If there exist irreducible subspaces $E_1, \ldots, E_m$ of H which are all G-isomorphic (under π) to (ρ, E), and such that H can be expressed as a direct sum

$$H = E_1 \oplus E_2 \oplus \cdots \oplus E_m \oplus F,$$

and F contains no subspace $\pi(G)$-isomorphic to the E_i, then we say that E **occurs with multiplicity** m in H. It is easy to see that if this is the case, then in any expression of H as a direct sum,

$$H = E_1' \oplus E_2' \oplus \cdots \oplus E_r' \oplus F',$$

where the E_j' are $\pi(G)$-isomorphic to the E_i, and (ρ, E) does not occur in F', then $r = m$. For the needed technique to reduce the proof to standard algebraic arguments of semi-simplicity, see *Real Analysis*, Chapter VII, Exercise 19. We call m the **multiplicity** of ρ in π (or of E in H).

Let H be a Hilbert space and π a representation of G in H. We say that H is **completely reducible** for π, or that π is completely reducible, if H is the orthogonal direct sum of irreducible subspaces. We write such a direct sum as

$$H = \widehat{\bigoplus_{i \in I}} H_i,$$

where $\{i\}$ ranges over a set of indices I, the H_i are subspaces invariant under G, mutually orthogonal, and H is the closure of the algebraic space generated by the H_i. This closure is indicated by the roof over the direct sum sign, which signifies algebraic direct sum. We also say that the family $\{H_i\}$ is an **orthogonal decomposition** of H.

Let $A: H \longrightarrow H$ be an operator (continuous linear map). We recall that A is called **compact** if A maps bounded sets into relatively compact sets (sets whose closure is compact). Alternatively, we could say that if $\{v_n\}$ is a bounded sequence, then $\{Av_n\}$ has a convergent subsequence. A vector $v \in H$ is called an **eigenvector** for A if $Av = \lambda v$ for some complex number λ. Given $\lambda \in \mathbf{C}$, the set of elements $v \in H$ such that $Av = \lambda v$, together with 0, is a subspace H_λ, called the λ-**eigenspace** of A.

Spectral theorem for compact operators. *Let A be a compact hermitian operator on the Hilbert space E. Then the family of eigenspaces $\{E_\lambda\}$, where λ ranges over all eigenvalues* (*including* 0), *is an orthogonal decomposition of E.*

Proof. Let F be the closure of the subspace generated by all E_λ. Let H be the orthogonal complement of F. Then H is A-invariant, and A induces a compact hermitian operator on H, which has no eigenvalue. We must show that $H = \{0\}$. This will follow from the next lemma.

Lemma. *Let A be a compact hermitian operator on the Hilbert space $H \neq \{0\}$. Let $c = |A|$. Then c or $-c$ is an eigenvalue for A.*

Proof. There exists a sequence $\{x_n\}$ in H such that $|x_n| = 1$ and

$$|\langle Ax_n, x_n\rangle| \longrightarrow |A|.$$

Selecting a subsequence if necessary, we may assume that

$$\langle Ax_n, x_n\rangle \longrightarrow \alpha$$

for some number α, and $\alpha = \pm|A|$. Then

$$\begin{aligned} 0 \leqslant |Ax_n - \alpha x_n|^2 &= \langle Ax_n - \alpha x_n, Ax_n - \alpha x_n\rangle \\ &= |Ax_n|^2 - 2\alpha\langle Ax_n, x_n\rangle + \alpha^2|x_n|^2 \\ &\leqslant \alpha^2 - 2\alpha\langle Ax_n, x_n\rangle + \alpha^2. \end{aligned}$$

The right-hand side approaches 0 as n tends to infinity. Since A is compact, after selecting a subsequence, we may assume that $\{Ax_n\}$ converges to some vector y, and then $\{\alpha x_n\}$ must converge to y also. If $\alpha = 0$, then $|A| = 0$ and $A = O$, so we are done. If $\alpha \neq 0$, then $\{x_n\}$ itself must converge to some vector x, and then $Ax = \alpha x$ so that α is the desired eigenvalue for A, thus proving our lemma, and the theorem.

We observe that each E_λ has a Hilbert basis consisting of eigenvectors, namely any Hilbert basis of E_λ because all non-zero elements of E_λ are eigenvectors. Hence E itself has a Hilbert basis consisting of eigenvectors. Thus we recover precisely the analog of the theorem in the finite dimensional case. Furthermore, we have some additional information, which follows trivially:

Each E_λ is **finite dimensional** if $\lambda \neq 0$, otherwise a denumerable subset from a Hilbert basis would provide a sequence contradicting the compactness of A. For a similar reason, *given $r > 0$, there is only a finite number of eigenvalues λ such that $|\lambda| \geqslant r$.* Thus 0 is a limit of the sequence of eigenvalues if E is infinite dimensional. If H is a Hilbert space and A a compact operator on H, we may therefore write

$$H = \hat{\bigoplus_{\lambda}} H_\lambda = \hat{\bigoplus_{i=1}^{\infty}} H_{\lambda_i},$$

where the eigenvalues λ_i are so ordered that $|\lambda_{i+1}| \leqslant |\lambda_i|$, and $\lim \lambda_i = 0$.

A subalgebra $\mathfrak{A}$ of operators on H is said to be **$*$-closed** if whenever $A \in \mathfrak{A}$, then $A^* \in \mathfrak{A}$.

Theorem 1. *Let $\mathfrak{A}$ be a $*$-closed subalgebra of compact operators on a Hilbert space H. Then H is completely reducible for $\mathfrak{A}$, and each irreducible subspace occurs with finite multiplicity.*

Proof. Let $\{E_i\}$ be a maximal orthogonal family of $\mathfrak{A}$-irreducible subspaces, and let F be the orthogonal complement of the subspace generated by the E_i. Since $\mathfrak{A}$ is $*$-closed, it follows that F is $\mathfrak{A}$-invariant, and therefore we are reduced to proving, under the hypotheses of the theorem, that there exists an $\mathfrak{A}$-irreducible subspace. We do this as follows.

If $A \in \mathfrak{A}$, then

$$A = \frac{A + A^*}{2} + i\frac{A - A^*}{2i},$$

so there exists an element $A = A^* \neq O$ in $\mathfrak{A}$. If M is an invariant subspace $\neq \{0\}$, then the restriction of A to M satisfies the hypotheses of the theorem. Let $\lambda \neq 0$ be an eigenvalue for A. Among all invariant subspaces $M \neq \{0\}$, select one such that the eigenspace

$$M_\lambda = \{v \in M, Av = \lambda v\}$$

has minimal dimension. Let $v \in M$, $v \neq 0$. Then $\overline{\mathfrak{A}v} \subset M$, and $\overline{\mathfrak{A}v}$ is invariant. We contend that $\overline{\mathfrak{A}v}$ is irreducible. Suppose that $E \neq \{0\}$ is an invariant subspace of $\overline{\mathfrak{A}v}$. We can write

$$v = v_E + v_E'$$

where v_E is the E-component of v, and $v_E' \in \overline{\mathfrak{A}v}$ is perpendicular to E. Note that

$$\lambda v = Av = Av_E + Av_E' = \lambda v_E + \lambda v_E'.$$

So v_E and v_E' are λ-eigenvectors for A. If v_E or $v_E' = 0$, say $v_E' = 0$, then $v \in E$, whence $\overline{\mathfrak{A}v} = E$. This must necessarily happen, for otherwise, $v_E \neq 0$ and $v_E' \neq 0$ imply that $E_\lambda \subset M_\lambda$ and $E_\lambda \neq M_\lambda$, so $\dim E_\lambda < \dim M_\lambda$, contradiction. Hence $\overline{\mathfrak{A}v}$ is irreducible, and our theorem is proved.

Remark. To find the irreducible subspace, we needed only *one* compact hermitian operator in the algebra.

§3. L^2 KERNELS AND OPERATORS

A certain type of kernels and operators will recur sufficiently often so that it is worthwhile to mention them independently here, rather than in an appendix, or when we use them for the first time. They give examples of compact operators.

Theorem 2. *Let $(X, \mathfrak{M}, dx)$ and $(Y, \mathfrak{N}, dy)$ be measured spaces, and assume that $L^2(X)$, $L^2(Y)$ have countable orthogonal bases. Let $q \in \mathfrak{L}^2(dx \otimes dy)$. Then the operator $f \mapsto Qf$ such that*

$$Qf(x) = \int_Y q(x, y) f(y)\, dy$$

is a bounded operator from $L^2(Y)$ into $L^2(X)$, and is compact. We have $|Q| \leqslant \|q\|_2$.

Proof. Let $f \in \mathfrak{L}^2(Y)$. For almost all x our assumption implies that the function q_x such that $q_x(y) = q(x, y)$ is also in $\mathfrak{L}^2(Y)$. Hence the product $fq_x \in \mathfrak{L}^1(Y)$. We get by Schwarz

$$|Qf(x)|^2 \leqslant \|f\|_2^2\, \|q_x\|_2^2,$$

and integrating,

$$\|Qf\|_2^2 = \int |Qf(x)|^2\, dx \leqslant \|f\|_2^2 \iint |q(x, y)|^2\, dy\, dx$$
$$\leqslant \|q\|_2^2\, \|f\|_2^2.$$

This proves that $|Q| \leqslant \|q\|_2$, so that Q is a bounded operator.

Let $\{\varphi_i\}$, $\{\psi_j\}$ be orthonormal bases for $L^2(X)$ and $L^2(Y)$ respectively. Let

$$\theta_{ij}(x, y) = \varphi_i(x)\psi_j(y).$$

Then $\{\theta_{ij}\}$ is an orthonormal basis for $L^2(X \times Y)$. To see this, it is first clear that the θ_{ij} are of norm 1, and mutually orthogonal. Let $g \in \mathfrak{L}^2(X \times Y)$ be perpendicular to all θ_{ij}. Then

$$\int_X \varphi_i(x)\, dx \int_Y \psi_j(y) g(x, y)\, dy = 0$$

for all i, j. Hence

$$x \mapsto \int \psi_j(y) g(x, y)\, dy$$

is 0 except for x in a null set S in X. If $x \notin S$, then for almost all y, we have $g(x, y) = 0$. Hence $g(x, y) = 0$ for almost all $(x, y) \in X \times Y$ by Fubini's theorem.

Let

$$q = \sum a_{ij}\theta_{ij}$$

be the expression of q as a series in $L^2(X \times Y)$, with constants a_{ij}. Let

$$q_n = \sum_{i,j<n} a_{ij}\theta_{ij}$$

be a finite truncation of the series. It is immediately verified that the corresponding operator Q_n has finite dimensional image. In fact, if θ is a function on $X \times Y$ such that $\theta(x, y) = \varphi(x)\psi(y)$, then the image of the corresponding operator has dimension 1.

We have already proved the inequality

$$|Q_n - Q| \leqslant \|q_n - q\|_2,$$

and the expression on the right tends to 0 as $n \longrightarrow \infty$. Hence the operators Q_n, which are compact, tend in operator norm to Q, which is therefore also compact. This proves Theorem 2.

We now make some comments of a formal nature on the trace of operators represented by kernels as above. Observe that $\varphi_i \otimes \bar{\varphi}_j$ is an orthonormal basis for $L^2(X \times X)$. Take $Y = X$, and write the Fourier expansion for q in terms of $\varphi_i \otimes \bar{\varphi}_j$,

$$q(x, y) = \sum c_{ij}\, \varphi_i \otimes \bar{\varphi}_j.$$

Formally, we then expect the trace of Q to be given by

$$\operatorname{tr}(Q) = \sum_n \langle Q\varphi_n, \varphi_n \rangle = \sum_n \sum_{i,j} \iint c_{ij}\varphi_i(x)\bar{\varphi}_j(y)\varphi_n(y)\, dy\, dx.$$

By the orthogonality among the functions φ_j, φ_n we see that this last expression reduces to

$$\operatorname{tr}(Q) = \sum_n c_{nn}.$$

On the other hand

$$\int q(x, x)\, dx = \sum_{i,j} c_{ij} \int \varphi_i(x)\bar{\varphi}_j(x)\, dx = \sum_n c_{nn}.$$

Hence we find, formally,

$$\boxed{\operatorname{tr}(Q) = \sum_n c_{nn} = \int_X q(x, x)\, dx.}$$

What we need to make sense of these computations are sufficient conditions to make all the series converge, and the sum

$$\sum \langle Q\varphi_n, \varphi_n \rangle$$

independent of the choice of the orthonormal basis. We shall return to this when we discuss the Plancherel formula. Until then, we take the integral

$$\int_X q(x, x)\, dx$$

as the definition of the **trace**, whenever an operator can be defined by a continuous kernel q.

§4. PLANCHEREL MEASURES

Let X, Y be measured spaces, with measures dx and $d\mu(y)$ respectively. Let $\varphi = \varphi(x, y)$ be a function on the product. Then φ gives rise to an operator Φ from functions on X to functions on Y by the formula

$$\Phi f(y) = \int_X f(x)\varphi(x, y)\, dx,$$

and a transpose operator,

$$^t\Phi g(x) = \int_Y \varphi(x, y) g(y)\, d\mu(y).$$

(On occasion, we use the reverse convention, interchanging Φ and $^t\Phi$.) We also write Φ^* for $^t\overline{\Phi}$, i.e.

$$\Phi^* g(x) = \int_Y \overline{\varphi(x, y)}\, g(y)\, d\mu(y).$$

Then Φ^* is the adjoint for the scalar products defined by

$$\langle f_1, f_2 \rangle = \int f_1(x)\overline{f_2(x)}\, dx \qquad \text{on } X,$$

$$\langle g_1, g_2 \rangle_\mu = \int g_1(x)\overline{g_2(x)}\, d\mu(y) \qquad \text{on } Y;$$

in other words, we have

$$\langle \Phi f, g \rangle_\mu = \langle f, \Phi^* g \rangle.$$

This is immediately seen by the formal computation

$$\begin{aligned}\langle \Phi f, g\rangle_\mu &= \int_Y \int_X f(x)\varphi(x,y)\overline{g(y)}\, dx\, d\mu(y)\\ &= \int f(x)\, \overline{\int \overline{\varphi(x,y)}\, g(y)\, d\mu(y)}\, dx\\ &= \langle f, \Phi^* g\rangle.\end{aligned}$$

The above formalism applies to various situations in which the integrals converge (absolutely); for instance, if $\varphi \in \mathfrak{L}^2(X \times Y)$ and $f \in \mathfrak{L}^2(X)$; or if $\varphi \in C_c(X \times Y)$ and $f \in C_c(X)$. In any given situation it is incumbent on us to make the domain of the operator clear.

We have

$$\Phi^* = \Phi^{-1} \textit{ if and only if } \Phi \textit{ is unitary.}$$

This follows at once. In some applications, we prove that Φ is unitary, and then we conclude that Φ^{-1} is given by the starred kernel merely by applying the above formalism to functions g which obviously make the integrals converge, e.g. continuous with compact support on locally compact spaces.

In the applications of the above formalism, we always specify the function spaces on which the integrals converge. In representation theory, we start with $X = G$ ($SL_2(\mathbf{R})$ for this book), and dx is Haar measure.

We shall be given measured spaces (X, dx) and (Y, dy) in a "natural" way. We then want to find a positive function P on Y such that

$$\Phi^* = \Phi^{-1}$$

for the measure $d\mu(y) = P(y)\, dy$. In other words, interpret the transpose to be with respect to $d\mu(y)$ by

$$^t\Phi g(x) = \int \varphi(x,y) g(y) P(y)\, dy.$$

Then we want $\Phi^*\overline{\Phi} = Id$, on a suitable space of functions on X. If this happens, we call $P(y)\, dy$ the **Plancherel measure for** φ, and the formula $\Phi^*\Phi = Id$ is called the **Plancherel inversion formula**.

Actually, in the Plancherel formula on a non-commutative group (as on $SL_2(\mathbf{R})$ later) the situation is slightly more complicated, even though formally quite similar, because the map φ is operator valued, and in the inversion, we have to insert a trace. Cf. the end of the chapter on the Plancherel Formula.

Let $\hat{G}$ be the unitary equivalence classes of irreducible unitary representations of G. It is usually possible to parametrize $\hat{G}$, or an appropriate subset of $\hat{G}$, by means of an analytic space Y (set of zeros of analytic equations), with a "natural" measure dy (for instance if Y is a piece of Euclidean space, dy is Lebesgue measure). For $SL_2(\mathbf{R})$ we shall see that this space Y consists of two vertical lines and isolated points, looking like Fig. 2. The space $\hat{G}$ is actually bigger. Lebesgue measure is then our dy on the line, while the points have discrete measure.

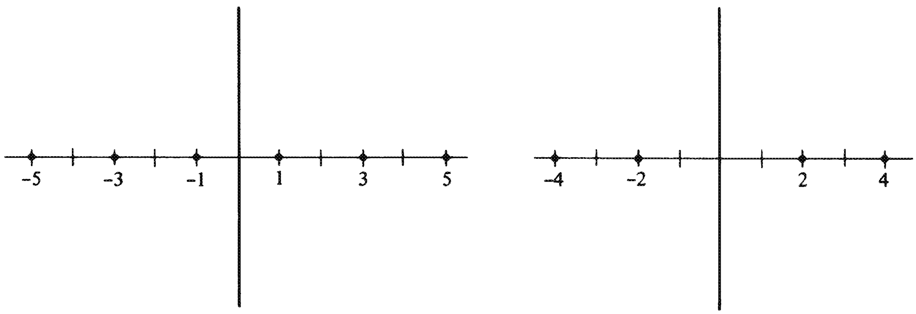

Figure 2

Before dealing with the formula in full generality, we shall deal with a somewhat simpler situation of a Plancherel measure only for a special class of functions, bi-invariant under an appropriate compact subgroup of G. This leads us to considering the representations of compact groups first, in the next chapter.

II Compact Groups

§1. DECOMPOSITION OVER K FOR $SL_2(\mathbf{R})$

In this section we essentially work out a special case of representation theory over compact groups, but in the context of $SL_2(\mathbf{R})$, providing a good introduction for what follows. We bring out immediately the important role of a maximal compact subgroup, the circle group K, i.e. the group of matrices

$$r(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

A **character** of K is by definition a continuous homomorphism of K into the unit circle, and the characters are indexed by the integers,

$$\chi_n(r(\theta)) = e^{in\theta}.$$

We let $G = GL_2^+(\mathbf{R})$ (matrices with positive determinant) or $SL_2(\mathbf{R})$. For each $y \in G$ and $f \in C_c(G)$ we let

$$f^y(\theta, \theta') = f(r(\theta)yr(\theta')).$$

Let $S_{n,m}$ be the subspace of $C_c(G)$ consisting of those functions f satisfying the condition

$$f(r(\theta)yr(\theta')) = e^{-in\theta}f(y)e^{-im\theta'}$$

for all $y \in G$ and all real θ, θ'.

Lemma 1. *The algebraic sum $\sum_{n,m} S_{n,m}$ is L^1-dense in $C_c(G)$. In fact, given ϵ and $f \in C_c(G)$, there exists a function $g \in \sum S_{n,m}$ such that the support of g is contained in $K(\operatorname{supp} f)K$, and such that $\|f - g\|_\infty < \epsilon$.*

Proof. Let

$$f_{n,m}(y) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f(r(\theta)yr(\theta'))e^{in\theta}e^{im\theta'}\,d\theta\,d\theta'$$

be the (n, m) Fourier coefficient of f^y. Then $f_{n,m}$ has support contained in $K(\operatorname{supp} f)K$. The Cesaro–Fejer kernels in one variable

$$\left\{\frac{1}{M}\sum_{N=0}^{M-1}\sum_{|n|<N} e^{in\theta}\right\}$$

form a Dirac sequence. The product of the kernel in θ and the kernel in θ' form a Dirac sequence in two variables, say $\{C_M(\theta, \theta')\}$, $M = 1, 2, \ldots$. From the definitions, we see that $(f^y * C_M)(0, 0)$ (convolution taken on the product of the circle with itself) is the sum of terms $c_{n,m}f_{n,m}(y)$ with appropriate constant coefficients $c_{n,m}$ arising from the sum in the Cesaro kernel. We check that the argument giving the convergence of the convolution toward

$$f^y(0, 0) = f(y)$$

is uniform in y. We have to estimate the difference

$$\iint [f^y(\theta, \theta')C_M(-\theta, -\theta') - f^y(0, 0)C_M(-\theta, -\theta')]\,d\theta\,d\theta'.$$

Given ϵ, there exists a neighborhood U of $(0, 0)$ such that

$$\iint_{\complement U} C_M < \epsilon \qquad \text{and} \qquad \int C_M = 1$$

where $\complement U$ is the complement of U. Therefore the integral of the difference above is estimated by integrals over U and $\complement U$, i.e. by

$$\sup_{(\theta,\theta')\in U} |f^y(\theta, \theta') - f^y(0, 0)| + 2\|f\|_\infty \iint_{\complement U} C_M.$$

Since f has compact support, we have

$$|f(r(\theta)yr(\theta')) - f(y)| < \epsilon$$

if U is sufficiently small, uniformly in all y, for $(\theta, \theta') \in U$. This proves what we wanted.

For one of the formulas of the next lemma, we recall that for any function φ on G, we defined

$$\varphi^*(x) = \overline{\varphi(x^{-1})}\,.$$

***Lemma* 2.** *We have*:

i) $S_{n,m} * S_{l,q} = 0$ *if* $m \neq l$.
ii) $S^*_{n,m} = S_{m,n}$.
iii) $S_{n,m} * S_{m,q} \subset S_{n,q}$.

Proof. Consider the convolution integral

$$f * g(x) = \int_G f(xy^{-1})g(y)\, dy.$$

Since G is unimodular, an integral with respect to y over G is invariant under the transformation $y \mapsto y^{-1}$. Now let $y \mapsto r(\theta)y$. From this invariance and the invariance under right and left translations, it follows that the above value $f * g(x)$ remains the same when multiplied by factors $e^{im\theta}$ and $e^{-il\theta}$. This is possible only when it is equal to 0, so (i) is proved. The other two assertions are proved in an analogous way, left to the reader.

The above lemma shows that $S_{n,n}$ is an algebra under convolution. The arguments are quite formal. We now come to a more specialized property.

***Lemma* 3.** *The algebra* $S_{n,n}$ *is commutative.*

As we are concerned here with the arbitrary $S_{n,n}$, and not just $S_{0,0}$, we give the proof in a general context. The reader will find it profitable to look at the simpler case of bi-invariant functions given at the beginning of Chapter IV, due to Gelfand. The generalization we give here is due to Silberger, *Proc. AMS* 1969, p. 437. (The result was designed to work p-adically.)

Let σ be an automorphism of a unimodular group G, or an anti-automorphism. By the uniqueness of Haar measure, there exists a positive number $\Delta(\sigma)$ such that for all $f \in C_c(G)$ we have

$$\int_G f(x^\sigma)\, dx = \Delta(\sigma) \int_G f(x)\, dx.$$

We must have $\Delta(\sigma^2) = \Delta(\sigma)\,\Delta(\sigma)$, and therefore if $\sigma^2 = 1$, it follows that $\Delta(\sigma) = 1$. Thus the Haar integral is invariant under the transformation $x \mapsto x^\sigma$. We also write ${}^\sigma x$ instead of x^σ.

***Theorem* 1.** *Let G be a a unimodular locally compact group. Let K be a compact subgroup. Assume*:

i) *That there exists an anti-automorphism τ of G, of order 2, such that $k^\tau = k^{-1}$ for all $k \in K$.*
ii) *If S is the set of elements $s \in G$ such that $s^\tau = s$, then $G = SK$.*
iii) *There exists an automorphism σ of order 2 such that $k^\sigma = k^{-1}$ for all $k \in K$, and if $s \in S$, then*

$$s^\sigma = k_1 s k_1^{-1}$$

for some $k_1 \in K$.

Let $\rho: K \mapsto \mathbf{C}^1$ be a character of K, and let $S_{\rho,\rho}$ be the set of functions $f \in C_c(G)$ such that

$$f(k_1 x k_2) = \rho(k_1) f(x) \rho(k_2)$$

for all $x \in G$ and $k_1, k_2 \in K$. Then $S_{\rho,\rho}$ is commutative.

Proof. Define $f^*(x) = f(x^\tau)$. Then

$$(f * g)^* = g^* * f^*.$$

On the other hand, define $f'(x) = f(x^\sigma)$. Then

$$(f * g)' = f' * g'.$$

We prove the former (the latter is easier). We have

$$(g * f)^*(x) = (g * f)(x^\tau) = \int g(x^\tau y^{-1}) f(y)\, dy$$

and

$$(f^* * g^*)(x) = \int f^*(xy^{-1}) g^*(y)\, dy = \int f(y^{-\tau} x^\tau) g(y^\tau)\, dy.$$

Letting successively $y \mapsto y^\tau$, $y \mapsto x^\tau y$, and $y \mapsto y^{-1}$ proves the formula. Also, for $f \in S_{\rho,\rho}$ we have $f^* = f'$. Indeed, it suffices to prove that $f(x^\sigma) = f(x^\tau)$. But write $x = sk$. Then

$$f(x^\tau) = f(k^{-1} s^\tau) = \rho(k)^{-1} f(s)$$

$$f(x^\sigma) = f(k^{-1} s^\sigma) = \rho(k)^{-1} f(k_1 s k_1^{-1}) = \rho(k^{-1}) f(s),$$

thus proving our assertion. It now follows that

$$f * g = g * f$$

as desired.

Example. For $G = GL_2(\mathbf{R})$ or $SL_2(\mathbf{R})$, we let K be the circle group as before. We let

$$x^\tau = {}^t x \qquad \text{(transpose of } x\text{)}$$

$$x^\sigma = \gamma x \gamma \qquad \text{where} \quad \gamma = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

The conditions of Theorem 1 are verified, in view of the standard polar decomposition of a matrix, which we recall. If $x \in GL_n(\mathbf{R})$, we let $y = x\,{}^t x$, so

y is symmetric positive definite. There is a basis for $\mathbf{R}^n$ consisting of eigenvectors such that

$$yv_i = \lambda_i v_i, \qquad \lambda_i > 0.$$

Let $s^2 = y$, so that s has eigenvalues $\pm\lambda_i$ on v_i, and choose s so that

$$\text{sign det } s = \text{sign det } x.$$

Let $k = s^{-1}x$. Then $x = sk$, and $\det k = 1$. Also

$${}^t kk = {}^t x {}^t s^{-1} s^{-1} x = {}^t x y^{-1} x = {}^t x\, {}^t x^{-1} x^{-1} x = 1.$$

Hence k is real unitary and we are done.

Let us return to $G = SL_2(\mathbf{R})$ or $GL_2^+(\mathbf{R})$. Let

$$\pi \colon G \longrightarrow GL(H)$$

be a representation of G into a Banach space H. For each integer n let H_n be the set of elements $v \in H$ such that

$$\pi(r(\theta))v = e^{in\theta}v.$$

Then H_n is a subspace (obviously closed).

Lemma 4. *Assume that H is a Hilbert space and π is unitary on K. If $m \neq n$, then H_n is perpendicular to H_m.*

Proof. For $v \in H_n$ and $w \in H_m$ we have $\pi(r(\theta))^* = \pi(r(-\theta))$, so

$$\begin{aligned}\langle \pi(r(\theta))v, w\rangle &= e^{in\theta}\langle v, w\rangle \\ &= \langle v, \pi(r(-\theta))w\rangle = e^{im\theta}\langle v, w\rangle.\end{aligned}$$

The assertion follows.

Lemma 5. *We have:*

i) $\pi^1(S_{n,m})H \subset H_n$,
ii) $\pi^1(S_{n,m})H_q = \{0\}$ *if* $m \neq q$.

Proof. If $m \neq q$, then we use the invariance of

$$\int_G f(y)\pi(y)v\, dy, \qquad v \in H_q,$$

under translations $y \mapsto yr(\theta)$. If $f \in S_{n,m}$, we find that the above value of the integral is equal to itself multiplied by $e^{-im\theta}e^{iq\theta}$, whence must be equal to 0. Statement (i) is equally clear, namely let $q = m$ and let $v \in H$, $f \in S_{n,m}$. Then for $k = r(\theta)$,

$$\begin{aligned}\pi(k)\pi^1(f)v &= \pi(k)\int_G f(y)\pi(y)v\,dy \\ &= \int_G f(y)\pi(ky)v\,dy \\ &= \int_G f(k^{-1}y)\pi(y)v\,dy \\ &= e^{in\theta}\pi^1(f)v.\end{aligned}$$

This proves our lemma.

If w lies in a finite direct sum of spaces H_q, we let w_q denote its component in H_q. Lemma 5 shows that $\pi^1(f)$ for $f \in \sum S_{n,m}$ maps H into such a direct sum.

Lemma 6. *Assume that π is irreducible. Then the space H_q is irreducible for $S_{q,q}$, and if $H_q \neq \{0\}$, then $\pi^1(S_{q,q})H_q \neq \{0\}$.*

Proof. Let W be a proper subspace of H_q, invariant for $\pi^1(S_{q,q})$. If $w \in W$ and f is a finite sum of functions $f_{n,m} \in S_{n,m}$, then by Lemma 5,

$$(\pi^1(f)w)_q = \pi^1(f_{q,q})w \in W.$$

The algebra $\mathfrak{C} = \sum S_{n,m}$ is L^1-dense in $C_c(G)$ by Lemma 1, and the algebraic space of elements $\pi^1(f)w$ with $f \in \mathfrak{C}$ has its q-component contained in W. This is impossible because of the possibility of Dirac sequence approximations (cf. I, §1).

Theorem 2. *Let π be an irreducible representation of G on a Banach space H. Let H_n be the subspace of vectors v such that*

$$\pi(r(\theta))v = e^{in\theta}v.$$

If $\dim H_n$ *is finite, then* $\dim H_n = 0$ *or* 1. *This is always the case if π is unitary irreducible.*

Proof. We know that H_n is irreducible for $\pi^1(S_{n,n})$ and finite dimensional linear algebra shows that $\dim H_n = 0$ or 1, since $S_{n,n}$ is commutative. On the other hand, if π is unitary, and $f \in S_{n,n}$, then $\pi^1(f)^* = \pi^1(f^*)$, where $f^*(x) = f(x^{-1})$. It is immediately verified that $f^* \in S_{n,n}$ (cf. Lemma 2, ii). Hence, $\pi^1(S_{n,n})$ is $*$-closed, and Schur's lemma implies that $\dim H_n = 0$ or 1, cf. Appendix 1.

Theorem 3. *Let π be an irreducible representation of G on a Banach space*

H. Then the sum $\sum H_n$ is dense in H. If H is a Hilbert space and π is unitary on K, this sum is an orthogonal decomposition of H.

Proof. Let E be the (closed) subspace generated by the H_n. By Lemma 5 and the fact that the sum $\sum S_{m,n}$ is dense in $C_c(G)$, we conclude that E is $C_c(G)$-invariant, whence is G-invariant. Since π is irreducible, it follows that $E = H$. If π is unitary on K, we know from Lemma 4 that the H_n are mutually orthogonal. This proves our theorem.

Theorems 2 and 3 give us an indication of what will happen to the representations of $SL_2(\mathbf{R})$. Up to a point, they will be classified by the presence or absence of appropriate H_n. In the theory of spherical functions, we study the case when H_0 occurs. This is equivalent to the existence of a **fixed vector** under K, i.e. a vector $v \in H$, $v \neq 0$ such that $\pi(K)v = v$. In the alternative case, we are led to the discrete series.

In this section we dealt with the K-decomposition of the representation by means of the abstract nonsense of Haar measure and convolution. In Chapter VI we return to this decomposition from the point of view of the derived representation on the Lie algebra, and get much more precise information on the way the group operates, via the exponential map. This later chapter is mostly logically independent of the material on spherical functions, and the reader can easily read most of it immediately following the present discussion, to see how differentiability can be used.

Let π be a representation of G in a Banach space H, and suppose that H is a direct sum

$$H = \hat{\bigoplus} H_n,$$

where H_n is the n-th eigenspace of K as defined above. Then the *algebraic sum*

$$\sum H_n$$

is an algebraic subspace of H, dense in H. It has an algebraic characterization. Let us say that an element $v \in H$ is **K-finite** if $\pi(K)v$ generates a finite dimensional vector space.

The algebraic space $\sum H_n$ is the space of K-finite vectors.

Proof. It is clear that every element of $\sum H_n$ is K-finite. Conversely, suppose that an element $v \in H$ is K-finite. A finite dimensional representation of K in a space W decomposes into a direct sum of spaces W_n, and $W_n \subset H_n$. It is therefore clear that v is contained in $\sum H_n$.

The algebraic sum $\sum H_n$ will be denoted by $H(K)$. Theorem 2 shows the importance of knowing that the dimensions of the components H_n are finite.

Because of this, we define a representation π to be **admissible** if $\dim H_n$ is finite for all n. Theorem 2 with this terminology then implies that every irreducible unitary representation is admissible. We say that the representation is **strictly admissible** if the dimensions $\dim H_n$ are bounded.

§2. COMPACT GROUPS IN GENERAL

In the case of $SL_2(\mathbf{R})$, the circle group discussed in §1 is commutative, and consequently one does not need the general theory of compact groups (which, however, follows closely the pattern given in the commutative case). However, the non-commutative aspects illustrate other principles which will arise in a much more complicated fashion for the non-compact $SL_2(\mathbf{R})$, e.g. the formalism of the trace. Hence it is worthwhile to go through the theory of compact groups as an introduction to the other.

Let K be a compact group with Haar measure equal to 1, and let

$$\pi: K \longrightarrow GL(H)$$

be a representation in a Hilbert space H. By a remark at the beginning of Chapter I, §1, we know that π is bounded.

We shall now see that we can find an equivalent norm on H such that π is unitary with respect to this norm. For $v \in H$ define

$$|v|_\pi^2 = \int_K |\pi(k)v|^2 \, dk.$$

Then $|v|_\pi^2 \leqslant C^2|v|^2$ if C is a bound for π. Hence $|v|_\pi \leqslant C|v|$. On the other hand, for $k \in K$,

$$|v| = |\pi(k)^{-1}\pi(k)v| \leqslant C|\pi(k)v|,$$

whence

$$|\pi(k)v| \geqslant C^{-1}|v| \qquad \text{and} \qquad |v|_\pi \geqslant C^{-1}|v|.$$

This proves that $|\ |_\pi$ is equivalent to $|\ |$, and it is clear that π is unitary with respect to the norm $|\ |_\pi$. This proves what we wanted.

On $L^2(K)$ (with respect to Haar measure), we have an operation of right translation T, defined by

$$T(y)f(x) = f(xy).$$

Then T is unitary because

$$\int_K |f(xy)|^2 \, dx = \int_K |f(x)|^2 \, dx,$$

since a compact group is unimodular (a homomorphism of a compact group into the positive reals must be trivial). We also call T the **regular representation** (on the right).

Let $\varphi \in C_c(G)$. Then

$$\begin{aligned} T^1(\varphi)f(x) &= \int_K f(xy)\varphi(y)\,dy \\ &= \int_K f(y)\varphi(x^{-1}y)\,dy \\ &= f * \varphi^-\,(x) \end{aligned}$$

where $\varphi^-\,(x) = \varphi(x^{-1})$. We see that $T^1(\varphi)$ arises from a kernel

$$(x, y) \mapsto \varphi(x^{-1}y),$$

which is continuous on $K \times K$. By the Weierstrass–Stone theorem, any continuous function on $K \times K$ can be uniformly approximated by finite sums

$$\sum \varphi_i(x)\psi_i(y)$$

and the operator arising from the kernel $\varphi_i \otimes \psi_i$, i.e. the function

$$(x, y) \mapsto \varphi_i(x)\psi_i(y)$$

for each i, has a one-dimensional image. Consequently, $T^1(\varphi)$ can be approximated in norm by operators with finite dimensional image, whence $T^1(\varphi)$ is compact. By I, §2, Th. 1 we get:

Theorem 1. *Under the regular representation, $L^2(K)$ is the orthogonal direct sum of irreducible subspaces, i.e. the regular representation is completely reducible.*

Theorem 2. *Let π: $K \mapsto \mathrm{Aut}(H)$ be a unitary irreducible representation of a compact group K. Then H is finite dimensional.*

Proof. Let u be a unit vector in H and let P be the orthogonal projection on the one-dimensional space (u). Let Q: $H \longrightarrow H$ be the continuous linear map defined by

$$Qv = \int_K \pi(x)^{-1} P\pi(x)v\,dx.$$

Then Q commutes with all operators $\pi(y)$, $y \in K$ (immediate by the right and left invariance of Haar measure); and $Q = Q^*$, since

$$\langle Qv, w\rangle = \int_K \langle \pi(x)^{-1} P\pi(x)v, w\rangle \, dx$$

$$= \int_K \langle v, \pi(x)^{-1} P\pi(x)w\rangle \, dx = \langle v, Qw\rangle.$$

By App. 1, Th. 4 we conclude that $Q = \lambda I$ for some scalar λ, and $\lambda \neq 0$ because the integrand defining $\langle Qu, u\rangle$ is $\geqslant 0$, and > 0 for x near e. Let $\{u_i\}$ be an orthonormal basis for H. Then

$$\sum_{i=1}^{n} \int_K \langle \pi(x)^{-1} P\pi(x)u_i, u_i\rangle = n\lambda.$$

For each x, $\{\pi(x)u_i\}$ is an orthonormal basis. Hence

$$\sum_{i=1}^{n} \langle P\pi(x)u_i, \pi(x)u_i\rangle \leqslant \sum_{i=1}^{\infty} \langle P\pi(x)u_i, \pi(x)u_i\rangle$$

$$\leqslant \sum_{i=1}^{\infty} \langle Pu_i', u_i'\rangle$$

where $u_i' = \pi(x)u_i$. But $Pv = \langle v, u\rangle u$. So

$$\langle Pu_i', u_i'\rangle = |\langle u_i', u\rangle|^2.$$

It follows that

$$\sum_{i=1}^{n} \langle P\pi(x)u_i, \pi(x)u_i\rangle \leqslant \sum_{i=1}^{\infty} |\langle u_i', u\rangle|^2 = 1.$$

Integrating over K proves our theorem.

Remark Let $\pi(x) = (\pi_{ij}(x))$ be a matrix representation of a group in a finite dimensional space. Let $\{e_1, \ldots, e_n\}$ be a basis and let λ_i be the projection on the i-th coordinate. Then the coefficient function $\pi_{ij}(x)$ is $\lambda_i(\pi(x)e_j)$. The corresponding multiplication of matrices looks like

$$\begin{pmatrix} \pi_{11} & \cdots & \pi_{1n} \\ \vdots & & \vdots \\ \pi_{n1} & \cdots & \pi_{nn} \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} \pi_{1j} \\ \vdots \\ \pi_{jj} \\ \vdots \\ \pi_{nj} \end{pmatrix}.$$

In general, if π is a representation of a group G in a Banach space H, if $v \in H$ and λ is a functional on H, then we call

$$x \mapsto \lambda(\pi(x)v) = \pi_{v,\lambda}(x)$$

a **coefficient function**. If H is a Hilbert space, we can of course represent λ by an element w of H, so that the coefficient functions are given as

$$x \mapsto \langle \pi(x)v, w \rangle = \pi_{v,w}(x).$$

If $\dim H$ is finite, $\{e_i\}$ is a basis of H, and $\{\lambda_i\}$ is the dual basis, then the trace of the representation is given by

$$\chi_\pi(x) = \operatorname{tr} \pi(x) = \sum_i \lambda_i(\pi(x)e_i).$$

We used the trace in the proof of Theorem 2. In the case of infinite dimensional representations, the convergence of the series defining the trace becomes a problem, which will be discussed later in connection with specific representations.

Let π, σ be representations of the compact group K on Banach spaces H_π and H_σ. Let $a \in K$, and let λ be a functional on H_σ. Then for $w \in H_\pi$, the map

$$L\colon v \mapsto \int_K \lambda(\sigma(ax)v)\pi(x^{-1})w\,dx \tag{1}$$

of H_σ into H_π is a K-homomorphism.

Proof. By definition,

$$L\sigma(y)v = \int_K \lambda(\sigma(a)\sigma(x)\sigma(y)v)\pi(x^{-1})w\,dx.$$

Let $x \mapsto xy^{-1}$. The expression on the right transforms into $\pi(y)Lv$, as desired.

Schur's lemma (Appendix 1) then yields:

Theorem 3. *If π, σ are inequivalent irreducible representations of K, then for all $v \in H_\sigma$, $w \in H_\pi$, $a \in K$ we have*

$$\int_K \lambda(\sigma(ax)v)\pi(x^{-1})w\,dx = 0, \tag{2}$$

i.e. $\pi^1(\sigma_{\bar{\lambda}, v}) = 0$; *the coefficients of one representation operate trivially on the other. If μ is a functional on H_π, then*

$$\int_K \lambda(\sigma(ax)v)\mu(\pi(x^{-1})w)\,dx = 0. \tag{3}$$

Note. We obtain (3) from (2) by applying the functional μ. Integration commutes with continuous linear maps on the space of values.

It is convenient to deal with the symmetric scalar product arising from the integrals, so we let

$$[f,g] = \int_K f(x)g(x^{-1})\,dx.$$

Theorem 3 shows that the coefficient functions of two inequivalent representations are orthogonal with respect to this scalar product. This is why Theorem 3 is called an orthogonality relation.

Corollary. *Under the hypotheses of Theorem* 3, *we have*

$$\pi^1(\chi_\sigma^-) = 0$$

where we recall that $\chi_\sigma^-(x) = \chi_\sigma(x^{-1})$.

Proof.

$$\pi^1(\chi_\sigma^-)w = \int_K \sum_i \lambda_i(\sigma(x^{-1})e_i)\pi(x)w\,dx = 0.$$

Let χ be the character of a finite dimensional representation σ, and let d_χ or d_σ, or $d(\sigma)$ be the dimension of σ. For any π let P_χ^π or P_σ^π be defined by

$$P_\chi^\pi = d_\chi \pi^1(\chi^-) = d_\chi \int_K \chi(x^{-1})\pi(x)\,dx = d_\chi \int_K \chi(x)\pi(x^{-1})\,dx.$$

If π, σ are *unitary*, then P_χ^π is self-adjoint, because in this case,

$$\chi^- = \bar{\chi}$$

Note that P_χ^π commutes with all $\pi(y)$, $y \in G$, i.e.

$$P_\sigma^\pi \pi(y) = \pi(y)P_\sigma^\pi.$$

The proof is immediate:

$$\begin{aligned} P_\sigma^\pi \pi(y) &= \int_K \chi(x)\pi(x^{-1}y)\,dx \\ &= \int_K \chi(yx)\pi(x^{-1})\,dx && (\text{by } x \mapsto yx) \\ &= \int_K \chi(yxy^{-1})\pi(y)\pi(x^{-1})\,dx && (\text{by } x \mapsto xy^{-1}) \\ &= \pi(y)P_\sigma^\pi. \end{aligned}$$

From Schur's lemma, we conclude that *if π is irreducible, then*

$$P_\sigma^\pi = c_\sigma I$$

for some complex number c_σ, which we shall compute.

If σ is not equivalent to π, then $c_\sigma = 0$ and $P_\sigma^\pi = 0$ by Theorem 3.

Lemma. *Let $\lambda \neq 0$ be a functional on the finite dimensional space H. Let*

$$\varphi_{\lambda, v}(w) = \lambda(w)v.$$

Then

$$\operatorname{tr} \varphi_{\lambda, v} = \lambda(v).$$

Proof. If $v = 0$, the assertion is clear. Let $v \neq 0$, $v = v_1$, and extend v to a basis $\{v_1, v_2, \ldots, v_n\}$ of H. Then

$$\varphi_{\lambda, v}(v_1) = \lambda(v_1)v_1$$

$$\varphi_{\lambda, v}(v_j) = \lambda(v_j)v_1 \qquad \text{for } j > 1.$$

The matrix of $\varphi_{\lambda, v}$ is non-zero only in the first row, and the expression for the trace is clearly the desired one.

The trace being a continuous linear functional (on operators), we find that

$$\operatorname{tr} \int_K \pi(x^{-1})\varphi_{\lambda, v}\pi(x)\, dx = \operatorname{tr} \varphi_{\lambda, v} = \lambda(v). \tag{4}$$

Theorem 4. *Let π be an irreducible representation of K on H. Let $v, w \in H$ and let λ be a functional on H. Then*

$$\int_K \lambda(\pi(x^{-1})w)\pi(x)v\, dx = \frac{1}{d(\pi)}\lambda(v)w.$$

Proof. For v fixed, consider the map $L: H \to H$ such that $L(w)$ is the expression on the left-hand side of the formula to be proved. Then

$$L(w) = \int_K \pi(x)\varphi_{\lambda, v}[\pi(x^{-1})w]\, dx.$$

The trace of L is the trace of $\varphi_{\lambda, v}$, namely $\lambda(v)$. Furthermore, L is a K-homomorphism, so $L = tI$ for some number t by Schur's lemma. Hence

$$\lambda(v) = t \dim \pi = t\, d(\pi),$$

so

$$L(w) = \frac{\lambda(v)}{d(\pi)} w.$$

This proves our theorem.

Corollary 1. *For any $a \in K$ and any functional μ on H, we have*

$$\int_K \lambda(\pi(a)\pi(x^{-1})w)\pi(x)v\, dx = \frac{1}{d(\pi)}\lambda(\pi(a)v)w$$

and

$$\int_K \lambda(\pi(a)\pi(x^{-1})w)\mu(\pi(x)v)\, dx = \frac{1}{d(\pi)}\lambda(\pi(a)v)\mu(w).$$

Proof. Replace v by $\pi(a)v$ in the theorem, let $x \mapsto xa^{-1}$, and apply the functional μ to the relation of the theorem.

Assume now that π is unitary and that $\{e_i\}$ is an orthonormal basis of H. Let

$$\lambda_i(v) = \langle v, e_i \rangle, \qquad \pi_{ij}(x) = \langle \pi(x)e_i, e_j \rangle.$$

Then Theorem 4 and its corollary show that

$$\begin{aligned}[\pi_{ij}, \pi_{kl}] &= \frac{1}{d(\pi)}\langle e_l, e_i \rangle \langle e_j, e_k \rangle \\ &= 0 \qquad \text{unless} \qquad i = l \quad \text{and} \quad j = k.\end{aligned}$$

So the scalar product is 0 unless $\pi_{kl} = \pi_{ji}$. But *in the unitary case, we have*

$$\boxed{\pi_{ji}(x) = \overline{\pi_{ij}(x^{-1})}\,,}$$

and

$$[\pi_{ij}, \pi_{kl}] = \int_K \pi_{ij}(x^{-1})\pi_{kl}(x)\, dx = \int_K \overline{\pi_{ji}(x)}\,\pi_{kl}(x)\, dx.$$

Hence we get orthogonality for coefficient functions of the same representation:

Corollary 2. *Assume that π is unitary and let π_{ij} be the coefficient functions relative to an orthonormal basis of H. Then for the hermitian product*

$$\langle f, g \rangle = \int_K f(x)\overline{g(x)}\, dx,$$

π_{ij} is orthogonal to π_{kl} unless $i = k$ and $j = l$.

Theorem 5. *Let π be an irreducible representation of K. Then*

$$\pi^1(\chi_\pi^-) = \frac{1}{d(\pi)} I.$$

Proof. We have:

$$\begin{aligned}
\pi^1(\chi^-)v &= \int_K \chi(x^{-1})\pi(x)v\,dx \\
&= \int_K \sum_i \lambda_i(\pi(x^{-1})e_i)\pi(x)v\,dx \\
&= \sum_i \int_K \lambda_i(\pi(x^{-1})e_i)\pi(x)v\,dx \\
&= \sum_i \frac{1}{d(\pi)}\lambda_i(v)e_i = \frac{1}{d(\pi)}v,
\end{aligned}$$

as was to be shown.

We may therefore summarize one orthogonality relation in the precise form

$$\boxed{\pi^1(d_\sigma\chi_\sigma^-) = \begin{cases} 0 & \text{if} \quad \pi \not\sim \sigma \\ I & \text{if} \quad \pi \sim \sigma \end{cases}}$$

whenever π, σ are irreducible.

Theorem 6. *Every irreducible representation of K occurs in the regular representation on $L^2(K)$.*

Proof. By complete reducibility, we know that

$$L^2(K) = \hat{\bigoplus_\pi} m_\pi H_\pi.$$

Let σ be an irreducible representation and ψ its character. If σ does not occur, then for all π occurring in $L^2(K)$ we get

$$\pi^1(\psi^-) = 0.$$

Hence if T is the regular representation (by right translation), we see that $T^1(\psi^-)$ annihilates every H_π, and hence that

$$T^1(\psi^-) = 0$$

on $L^2(K)$. But then for all $f \in L^2(\mathrm{K})$,

$$\begin{aligned}
0 = T^1(\psi^-)f(x) &= \int \psi^-(y)T(y)f(x)\,dy \\
&= \int \psi(y^{-1})f(xy)\,dy \\
&= (f * \psi)(x).
\end{aligned}$$

This is impossible, for instance because of Dirac sequence approximations, and proves our theorem.

Theorem 7. *Let π, σ be irreducible representations of K. Then*:

$$\chi_\sigma * \chi_\pi = \begin{cases} 0 & \text{if} \quad \sigma \not\sim \pi, \\ d_\pi^{-1}\chi_\pi & \text{if} \quad \sigma \sim \pi. \end{cases}$$

Proof. To avoid subscripts, let χ and ψ be the characters of inequivalent irreducible representations of K, say on spaces H and H' respectively. Let $\{e_i\}$ be a basis of H and $\{\lambda_i\}$ the dual basis, and similarly $\{e_j'\}$ and $\{\lambda_j'\}$ for H'. Then

$$\begin{aligned}
\chi * \psi(a) &= \int_K \chi(ax)\psi(x^{-1})\,dx \\
&= \sum_{i,j} \int_K \lambda_i(\pi(a)\pi(x)e_i)\lambda_j'(\pi'(x^{-1})e_j')\,dx \\
&= 0.
\end{aligned}$$

On the other hand,

$$\begin{aligned}
\chi * \chi(a) &= \sum_{i,j} \int_K \lambda_i(\pi(a)\pi(x^{-1})e_i)\lambda_j(\pi(x)e_j)\,dx \\
&= \frac{1}{d(\pi)} \sum_{i,j} \lambda_i(\pi(a)e_j)\lambda_j(e_i) \\
&= \frac{1}{d(\pi)} \sum_{i} \lambda_i(\pi(a)e_i) \\
&= \frac{1}{d(\pi)} \chi(a).
\end{aligned}$$

This proves our theorem.

Corollary. *The function $d_\pi\chi_\pi$ is an idempotent in $C_c(K)$ (equal to $C(K)$), for any irreducible representation π of K.*

Let T be right translation as before, giving rise to the regular unitary representation

$$T\colon K \longrightarrow \operatorname{Aut} L^2(K)$$

by

$$T(y)f(x) = f(xy).$$

If χ is the character of a finite dimensional representation, let

$$P_\chi = T^1(d_\chi \chi^-).$$

Then

$$\begin{aligned} T^1(d_\chi \chi^-)f(y) &= d_\chi \int \chi(x^{-1})f(xy)\,dx \\ &= d_\chi \int \chi(yx^{-1})f(x)\,dx. \end{aligned}$$

Thus we see that

$$\boxed{(P_\chi f)(y) = d_\chi(\chi * f)(y).}$$

In this way we see that the projection operator P_χ amounts to a convolution by χ in the L^2-algebra on G.

III Induced Representations

§1. INTEGRATION ON COSET SPACES

We shall study $SL_2(\mathbf{R})$ by decomposing it as a product of certain closed subgroups (not normal). Here we recall the general foundations for integration on coset spaces.

Let K be a closed subgroup of the locally compact group G, both assumed unimodular. Then G operates as a group of topological automorphisms of the coset space G/K, by

$$(x, yK) \mapsto xyK.$$

A measure μ on G/K is said to be **G-invariant** if $\mu(A) = \mu(xA)$ for every Borel set A in G/K and every $x \in G$. For the correspondence between measures and integrals, we refer to *Real Analysis*, XIII, §4.

If $f \in C_c(G)$, we denote by f^K the function

$$f^K(x) = \int_K f(xk)\, dk.$$

Then $f^K \in C_c(G/K)$ (right invariance of Haar measure). We refer to *Real Analysis*, XIII, §4, Theorem 3, for the proof that

$$f \mapsto f^K$$

maps $C_c(G)$ *onto* $C_c(G/K)$.

Theorem 1. *Let K be a closed subgroup of G, both assumed unimodular. There exists a unique invariant measure $\mu_{G/K}$ on G/K such that for any $f \in C_c(G)$ we have*

$$\int_{G/K} f^K\, d\mu_{G/K} = \int_G f\, d\mu_G,$$

where μ_G is Haar measure on G.

Proof. The uniqueness is obvious. Given $\varphi \in C_c(G/K)$, let $f \in C_c(G)$ such that $f^K = \varphi$. The invariant integral on G/K can be defined by means of the formula in the theorem, provided we show that if $f^K = 0$, then

$$\int_G f(x)\,dx = 0.$$

Let pr: $G \longrightarrow G/K$ be the canonical map. Let $\psi \in C_c(G/K)$ be such that $\psi = 1$ on pr(supp f). Let $g \in C_c(G)$ be such that $g^K = \psi$. Then assuming that $f^K = 0$, we get

$$\begin{aligned} 0 &= \int_G \int_K g(x)f(xk)\,dk\,dx = \int_K \int_G g(x)f(xk)\,dx\,dk \\ &= \int_G f(x)g^K(x)\,dx \\ &= \int_G f(x)\,dx. \end{aligned}$$

This proves our theorem.

Note. Although not needed, we point out that a similar proof shows that an invariant measure exists when G, K are not necessarily unimodular, provided that $\Delta_G | K = \Delta_K$, where Δ is the "modular" function relating left and right Haar measures.

Let P, K be closed subgroups of G such that $G = PK$, and such that the map

$$(p, k) \longmapsto pk$$

gives a topological isomorphism (not group isomorphism) from $P \times K$ onto G. Assume that G, K are unimodular.

Then a Haar integral on G is given by

$$f \longmapsto \int_K \int_P f(pk)\,dp\,dk.$$

Indeed, there exists a left invariant measure on G/K, and G/K is P-isomorphic to P itself as a transformation space, under left translation. Hence this measure is a Haar measure on P. Symbolically,

$$\boxed{dx = dp\,dk.} \tag{1}$$

The order pk when taking the value $f(pk)$ is essential, but by Fubini's theorem, we can also write

$$\int_G f(x)\, dx = \int_P \int_K f(pk)\, dk\, dp,$$

i.e. reverse the order of integration.

In the applications, P can be expressed as a further product,

$$P = AN,$$

where A, N are closed unimodular subgroups, and A normalizes N (see the example below), i.e. $ana^{-1} \in N$ for $a \in A$ and $n \in N$. In other words, the map of $A \times N \longrightarrow P$ given by

$$(a, n) \longmapsto an$$

is a topological isomorphism. A group G admitting such a decomposition $G = ANK$ is called an **Iwasawa group**, with an **Iwasawa decomposition**. Let da, dn, dk denote the Haar measures on A, N, K respectively. Then symbolically we have

$$\boxed{dp = da\, dn} \tag{2}$$

i.e. for Haar measure on P suitably normalized by a constant factor, we have

$$\int_P f(p)\, dp = \int_A \int_N f(an)\, dn\, da = \int_N \int_A f(an)\, da\, dn.$$

Proof. The measure $da\, dn$ is clearly left invariant under A. Let $n_1 \in N$. We get

$$\int_A \int_N f(n_1 an)\, dn\, da = \int_A \int_N f(aa^{-1}n_1 an)\, dn\, da.$$

But $a^{-1}n_1 a \in N$, so we can cancel it in the inner integral by left invariance of Haar measure on N. This proves our assertion.

Combining (1) *and* (2), *we see that given a decomposition*

$$G = ANK, \qquad x = ank = a_x n_x k_x,$$

into unimodular, closed subgroups such that A normalizes N, we have

$$\boxed{dx = da\, dn\, dk.} \tag{3}$$

Under the assumption that A normalizes N, given $a \in A$, the map

$$n \mapsto ana^{-1}$$

is an automorphism of N, and hence there exists a continuous homomorphism

$$\alpha \colon A \to \mathbf{R}^{+}$$

such that

$$\int_N f(ana^{-1})\, dn = \alpha(a)^{-1} \int_N f(n)\, dn.$$

Replacing f by its right a-translate, this formula is equivalent to

$$\int_N f(an)\, dn = \alpha(a)^{-1} \int_N f(na)\, dn. \tag{4}$$

Combining (4) with (3) yields

$$\boxed{dx = \alpha(a)^{-1}\, dn\, da\, dk.} \tag{5}$$

We can also compute the modular function on $P = AN$, and we contend that

$$\boxed{\Delta(p) = \Delta(an) = \alpha(a).} \tag{6}$$

Indeed, the integral

$$f \mapsto \int_N \int_A f(an)\, da\, dn$$

is obviously invariant on the right by N. Let $a_1 \in A$. Then

$$\begin{aligned} \iint f(ana_1)\, da\, dn &= \iint f(aa_1a_1^{-1}na_1)\, da\, dn \\ &= \alpha(a_1) \iint f(an)\, da\, dn. \end{aligned}$$

This proves our contention.

Example. Let $G = GL_2^{+}(\mathbf{R})$ (2×2 matrices with positive determinant).

A is the diagonal group of matrices $\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$ with $a_1, a_2 > 0$.

N is the unipotent group of matrices $\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}$.

P is the triangular group of matrices $\begin{pmatrix} a_1 & b \\ 0 & a_2 \end{pmatrix}$. In the case of $SL_2(\mathbf{R})$, $a_2 = a_1^{-1}$. We have the commutation rule

$$\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & a_1a_2^{-1}b \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}.$$

Making the change of variables $t = a_1b/a_2$ and $dt = (a_1/a_2)\,db$, we get the value for $\alpha(a)$, namely

$$\alpha\left(\begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}\right) = \frac{a_1}{a_2},$$

so that on $SL_2(\mathbf{R})$,

$$\alpha\left(\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}\right) = a^2. \tag{7}$$

One usually does not resist the temptation to indulge the incorrect notation

$$\boxed{\alpha(a) = a^2.}$$

If we want to be correct, we should let, for instance, $h_a = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$ and then write

$$\boxed{\alpha(h_a) = a^2.}$$

The upper half-plane representation. Let $\mathfrak{H}$ be the upper half plane, i.e. the set of complex numbers

$$z = x + iy, \qquad y > 0.$$

Then $G = GL_2^+(\mathbf{R})$ operates on $\mathfrak{H}$. Namely, let

$$\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2^+(\mathbf{R}).$$

We define

$$\sigma z = \frac{az + b}{cz + d}.$$

A brute force computation shows that $\sigma(\sigma'(z)) = (\sigma\sigma')(z)$. Observe that the matrices

$$\begin{pmatrix} a & 0 \\ 0 & a \end{pmatrix}$$

operate trivially. Furthermore, a trivial computation shows that

$$\operatorname{Im} \sigma(z) = \frac{\operatorname{Im}(z)(ad - bc)}{|cz + d|^2}.$$

Thus the condition $ad - bc > 0$ guarantees that if $z \in \mathfrak{H}$, then $\sigma z \in \mathfrak{H}$ also. The operation of $GL_2^+(\mathbf{R})$ factors through $SL_2(\mathbf{R})$ in view of the trivial action of scalar multiples of the identity.

Let K be the isotropy group of i, in other words the group of matrices such that

$$\frac{ai + b}{ci + d} = i,$$

in $SL_2(\mathbf{R})$. This amounts to the conditions

$$a^2 + b^2 = 1, \qquad c^2 + d^2 = 1, \qquad ad - bc = 1.$$

In other words, K is the group of matrices

$$r(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

The mapping $\sigma \mapsto \sigma i$ from $SL_2(\mathbf{R})$ into $\mathfrak{H}$ therefore induces a bijection

$$NA \to \mathfrak{H}.$$

In fact we see that

$$\begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \mapsto x + iy$$

with $y = a^2$. Observe here that the convenient order in the decomposition of P is

$$P = NA,$$

because $a \to \infty$ as $y \to \infty$ in the upper half plane. Let d^+a be additive Lebesgue measure on the line. Then

$$dy = 2ad^+a$$

and hence

$$\frac{dx\,dy}{y^2} = \frac{2dx\;a\;d^+a}{a^4} = 2\alpha(a)^{-1}dx\;d^*a,$$

where

$$d^*a = \frac{d^+a}{a}$$

is the Haar measure on the positive multiplicative group.

If we let a represent the variable in A in $SL_2(\mathbf{R})$, then with our formulas (5), (6), (7), we find in terms of the (x, y)-coordinates that

$$(8) \qquad 2\alpha(a)^{-1}\,dn\,da \text{ on } G/K \text{ is } \frac{dx\,dy}{y^2} \text{ on } \mathfrak{H}.$$

Moral of the story: The decomposition ank with measure $da\,dn\,dk$ is most useful for formal Haar measure computations on G. The decomposition nak with measure $\alpha(a)^{-1}dn\,da\,dk$ is most useful when we want to deal also with the homogeneous space G/K and its representation as the upper half plane.

§2. INDUCED REPRESENTATIONS

Let K be a closed subgroup of G, both assumed unimodular. Assume also that $G = PK$ with a closed subgroup P, and that the map $(p, k) \mapsto pk$ of $P \times K \to G$ is a topological isomorphism.

Then we may form the homogeneous space $P\backslash G$ on the *left*, and

$$P\backslash G \approx K,$$

as spaces on which K operates on the *right*.

Let σ be a representation of P on a Hilbert space V. Let $H(\sigma)$ be the space of mappings

$$f\colon G \to V$$

whose restriction to K is in $L^2(K)$, and satisfying the condition

$$f(py) = \Delta(p)^{1/2}\sigma(p)f(y),$$

where $\Delta = \Delta_P$ is the modular function on P. Define

$$\|f\|_K^2 = \int_K |f(k)|^2\,dk$$

to be the L^2-norm on K. The representation π of G on $H(\sigma)$ given by right translation, i.e.

$$\pi(y)f(x) = f(xy),$$

is called the **induced representation** of σ to G. The point of the extra factor $\Delta(p)^{1/2}$ in the definition of the induced representation is to make the next statement true.

Theorem 2. *If σ is bounded, then the induced representation π is bounded. If σ is unitary, then the induced representation π is unitary.*

Proof. Fix y. Write $ky = p'_k k'$, so that

$$f(ky) = f(p'_k k') = \Delta(p'_k)^{1/2}\sigma(p'_k)f(k').$$

Then

$$\int_K |f(ky)|^2\, dk = \int_K \Delta(p'_k)|\sigma(p'_k)f(k')|^2\, dk.$$

If σ is bounded, then the right-hand side is bounded by a constant times

$$\int_K \Delta(p'_k)|f(k')|^2\, dk,$$

and equality holds if σ is unitary. We shall now verify that for $\psi \in C_c(K)$,

$$\int_K \psi(k')\Delta(p'_k)\, dk = \int_K \psi(k)\, dk. \tag{1}$$

Let $f \in C_c(G)$. Since G is unimodular, we get

$$\begin{aligned}\int_P\int_K f(pk)\, dk\, dp &= \int_G f(x)\, dx = \int_G f(xy)\, dx\\ &= \int_P\int_K f(pky)\, dk\, dp\\ &= \int_K\int_P f(pp'_k k')\, dp\, dk\\ &= \int_K\int_P f(pk')\,\Delta(p'_k)\, dp\, dk\\ &= \int_P\int_K f(pk')\,\Delta(p'_k)\, dk\, dp.\end{aligned}$$

We reverse the order of integration and get what we want, by taking

$$f = \varphi \otimes \psi, \quad \text{with} \quad \varphi \in C_c(P), \psi \in C_c(K)$$

and

$$\int_P \varphi(p)\, dp = 1.$$

We also observe that we proved the boundedness of π from that of σ, if we only assume that σ is bounded.

Let us assume for simplicity that σ is one-dimensional, i.e. is a homomorphism into the multiplicative group of complex numbers. Then so is σ^{-1} or $\bar{\sigma}^{-1}$, where $\bar{\sigma}$ is the complex conjugate. We write

$$\sigma^* = \bar{\sigma}^{-1}.$$

To avoid complex conjugations, define the symmetric product on $L^2(K)$ by

$$[f, g] = \int_K f(k)g(k)\, dk.$$

Theorem 3. *The spaces $H(\sigma)$ and $H(\sigma^{-1})$ are dual to each other under this symmetric product. For $y \in G$, $f \in H(\sigma)$, $g \in H(\sigma^{-1})$, we have*

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

Proof. This is an easy computation:

$$\begin{aligned}\int_K \pi(y)f(k)g(k)\, dk &= \int_K f(ky)g(k)\, dk \\ &= \int_K \Delta(p_k')^{1/2}\sigma(p_k')f(k')g(k)\, dk \\ &= \int_K \Delta(p_k')\sigma(p_k')f(k')g(p_k'k'y^{-1})\, dk \\ &= \int_K \Delta(p_k')\sigma(p_k')f(k')\sigma(p_k')^{-1}g(k'y^{-1})\, dk \\ &= \int_K f(k)g(ky^{-1})\, dk\end{aligned}$$

as was to be shown.

By taking the hermitian product with the complex conjugate, one has to replace σ^{-1} by σ^*. Then $H(\sigma^*)$ is antidual to $H(\sigma)$.

§3. ASSOCIATED SPHERICAL FUNCTIONS

Let $G = PK$ as before, and $P = AN$ where A normalizes N. We assume that K is compact and has measure 1. As before

$$\Delta(p) = \Delta(an) = \alpha(a).$$

We let

$$\rho(a) = \alpha(a)^{1/2}.$$

Let s be a complex number, and define

$$\rho_s(x) = \rho_s(ank) = \rho(a)^{s+1}.$$

Then

$$\rho_s(k) = 1 = \rho_s(n).$$

The function

$$\mu_s : P \to \mathbf{C}^*$$

given by

$$\mu_s(an) = \rho(a)^s$$

is obviously a **character** (continuous homomorphism into $\mathbf{C}^*$). We do not require that its absolute value be 1. If it is, then we say that μ_s is a **unitary character**.

Example. In the case of $SL_2(\mathbf{R})$,

$$\mu_s\left(\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}\right) = a^s.$$

We denote by $\boldsymbol{H(s)}$ **the space of the representation** $\boldsymbol{\pi_s}$ **induced by** $\boldsymbol{\mu_s}$. It is the space of functions on G such that

i) $f(any) = \rho(a)^{s+1}f(y)$;

ii) the restriction of f to K is in $L^2(K)$.

Then $H(s)$ is a Hilbert space under the $L^2(K)$-norm.

We contend that ρ_s is a unit vector in the induced represention π_s.

Proof. First we show that ρ_s transforms properly under P. We have

$$\begin{aligned} \rho_s(nay) &= \rho_s(nan_y a_y) \\ &= \rho_s(nan_y a^{-1} a a_y) \\ &= \rho_s(a)\rho_s(y) \\ &= \rho(a)\rho(a)^s \rho_s(y) \\ &= \Delta(na)^{1/2}\mu_s(a)\rho_s(y), \end{aligned}$$

so ρ_s transforms as it should. Furthermore

$$\|\rho_s\|_K^2 = \int_K |\rho_s(k)|^2\, dk = 1,$$

and therefore ρ_s is a unit vector in the induced representation, which we denote by π. We also have

$$\pi(k)\rho_s = \rho_s,$$

so ρ_s is called a **fixed vector** by $\pi(K)$, or by K for short.

Even though we are now in an infinite dimensional representation, we may form the **coefficient function**

$$\begin{aligned}\langle \pi(x)\rho_s, \rho_s\rangle &= \int_K \pi(x)\rho_s(k)\,\overline{\rho_s(k)}\, dk \\ &= \int_K \rho_s(kx)\, dk.\end{aligned}$$

This turns out to be an important function associated with the character μ_s, and will be studied in detail in the chapter on spherical functions. We shall use the notation

$$\boxed{\varphi_s(x) = \int_K \rho(kx)^{s+1}\, dk.}$$

The character μ_s is unitary if and only if s is pure imaginary. In that case, we shall see later how the translates of φ_s generate an irreducible unitary representation. We call φ_s the **spherical function associated with** μ_s.

The family $\{\pi_s\}$ is often called the **principal series** of representations of $SL_2(\mathbf{R})$. Conventions differ, and sometimes the use of the term "principal series" is restricted to the unitary case when s is pure imaginary. We shall always specify which range of s we intend when using this terminology. Besides, as we shall see, the induced representations decompose into irreducible components according to parity, and here again, one may use the term "principal series" only for those irreducible components. Cf. Chapter VII, §3.

§4. THE KERNEL DEFINING THE INDUCED REPRESENTATION

Let $s \in \mathbf{C}$. Let $H(s)$ be as before, and π_s be the representation of G on $H(s)$ by right translation, so that for $f \in H(s)$ we have

$$\pi_s(x)f(y) = f(yx).$$

Let $\psi \in C_c(G)$. We derive an expression for $\pi_s^1(\psi)$ on f as follows:

$$\begin{aligned}
\pi_s^1(\psi)f(y) &= \int_G \psi(x)\pi_s(x)f(y)\,dx \\
&= \int_G \psi(x)f(yx)\,dx \\
&= \int_A \int_N \int_K \psi(k_y^{-1}n_y^{-1}a_y^{-1}\,ank)f(ank)\,da\,dn\,dk \\
&= \int_A \int_N \int_K \psi(k_y^{-1}n_y^{-1}a_y^{-1}ank)\rho(a)^{s+1}f(k)\,da\,dn\,dk \\
&= \int_K q_\psi(k,y)f(k)\,dk
\end{aligned}$$

where

(1) $$q_\psi(k,y) = \int_A \int_N \psi(k_y^{-1}n_y^{-1}a_y^{-1}ank)\rho(a)^{s+1}\,da\,dn$$

and

(2) $$q_\psi(k,k') = \int_A \int_N \psi(k'^{-1}ank)\rho(a)^{s+1}\,da\,dn.$$

Of course, $\pi_s^1(\psi)f$ is determined by its values on K, and we have

(3) $$\boxed{\pi_s^1(\psi)f(k') = \int_K q_\psi(k,k')f(k)\,dk.}$$

Thus we see that $\pi_s^1(\psi)$ is represented by the kernel $q_\psi(k,y)$. At the moment, we do not want to go into questions of convergence of a trace defined in terms of coefficient functions, and we prefer to rush as neatly and as fast as possible into the theory of spherical functions. Therefore, on an ad hoc basis, we define the **trace** of the operator $\pi_s^1(\psi)$ to be

(4) $$\operatorname{tr} \pi_s^1(\psi) = \int_K q_\psi(k,k)\,dk$$

where

(5) $$q_\psi(k,k) = \int_A \int_N \psi(k^{-1}ank)\rho(a)^{s+1}\,da\,dn.$$

[We shall discuss the relation of this and the usual trace when we deal with Plancherel's formula later.]

Assume that ψ is invariant under inner automorphism by K. This means that

$$\psi(kxk^{-1}) = \psi(x)$$

for all $k \in K$ and $x \in G$. The space of such functions is denoted by $C_c(G, K)$. Then the formula for q_ψ simplifies even more, namely

If $\psi \in C_c(G, K)$ and K has measure 1, then

$$q_\psi(k, k) = \int_A \int_N \psi(an)\rho(a)^{s+1}\, da\, dn \tag{6}$$

is constant, independent of k.

Let us continue to assume that K has measure 1. In Chapter V we shall study in detail the **Harish transform** on $C_c(G, K)$ defined by the integral

$$\mathbf{H}\psi(a) = \rho(a)\int_N \psi(an)\, dn, \qquad \text{where } \rho\left(\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}\right) = a.$$

With this definition, we get

Theorem 4. *For $\psi \in C_c(G, K)$ and $s \in \mathbf{C}$ we have*

$$\operatorname{tr} \pi_s^1(\psi) = \int_A \mathbf{H}\psi(a)\rho(a)^s\, da.$$

In the sequel, the integral as above with s as a parameter will be called a **Mellin transform M**, and thus we could abbreviate still further the formula for the trace by

$$\operatorname{tr} \pi_s^1(\psi) = \mathbf{MH}\psi(s).$$

If ψ is bi-invariant under K, i.e. $\psi(kxk') = \psi(x)$ for $k, k' \in K$ and $x \in G$, then this last integral is equal to

$$\int_G \psi(x)\rho(x)^{s+1}\, dx.$$

The whole situation of bi-invariant functions will be discussed systematically in the next chapter.

IV Spherical Functions

In this chapter and the next we study the algebra of functions on G which are invariant on the left and on the right by K, and relate the characters of this algebra to representation theory. This amounts to studying those representations which contain a K-fixed vector. We cover §3, §4 of the last chapter of Helgason's book [He 2]. We work with the abstract nonsense of Haar measure and convolution, without differential operators. This point of view was emphasized by Godement [Go 6]; see also Tamagawa [Tam], which we follow in part. To prove that all spherical functions are those which we exhibit explicitly, we need the differential equations, and the proof is postponed to Chapter X, §3.

For the p-adic theory, see McDonald [McD].

Throughout this chapter we let G be a unimodular group and K a compact subgroup with measure 1, i.e.

$$\int_K dk = 1.$$

§1. BI-INVARIANCE

A function f on G is said to be **K-bi-invariant (bi-invariant** for short) if

$$f(k_1 x k_2) = f(x)$$

for all $k_1, k_2 \in K$ and $x \in G$. Bi-invariance is denoted by a double bar, so $C_c(G/\!/K)$ denotes the bi-invariant continuous functions with compact support. For any function f on G we let

$$f^K(x) = \int_K f(xk)\, dk,$$

and we define ${}^K f$ similarly, averaging over K on the left. Then f^K is right K-invariant and ${}^K f$ is left K-invariant by the invariance of Haar measure. If f happens to be already right K-invariant, then $f^K = f$, so that $f \mapsto f^K$ is a projection operator on right K-invariant functions. A similar statement holds for left invariance. Thus ${}^K f^K$ is bi-invariant.

If f is right invariant, and φ is any function, then

$$\int_G f(x)\varphi(x)\,dx = \int_G f(x)\varphi^K(x)\,dx$$

because

$$\int_G f(x)\varphi(x)\,dx = \int_K \int_G f(x)\varphi(x)\,dx\,dk$$

(by $x \mapsto xk$)

$$= \int_K \int_G f(x)\varphi(xk)\,dx\,dk$$

(by Fubini)

$$= \int_G f(x)\varphi^K(x)\,dx.$$

A similar statement holds on the left. In particular, if f is bi-invariant, then

$$\int_G f(x)\varphi(x)\,dx = \int_G f(x)^K\varphi^K(x)\,dx.$$

These relationships hold whenever the integrals are absolutely convergent, e.g. if f is continuous and φ has compact support, or if f is in $\mathfrak{L}^1$ and φ is bounded. In practice, such convergence will always be clearly satisfied. The theory to be developed is not delicate from this point of view.

Let $\pi: G \to GL(H)$ be a representation of G on a Hilbert space H. We denote by H^K the subspace of elements $v \in H$ which are fixed under K, i.e. such that

$$\pi(k)v = v, \qquad \text{all } k \in K.$$

Let $P_K: H \to H$ be the map

$$P_K v = \int_K \pi(k)v\,dk.$$

Then P_K is the orthogonal projection on H^K. Let $v \in H^K$ and $\varphi \in C_c(G)$. Then

$$\pi^1(\varphi)v = \pi^1(\varphi^K)v.$$

Conversely, if $\varphi \in C_c(G//K)$ and $v \in H$, then

$$\pi^1(\varphi)v \in H^K.$$

If $v, w \in H^K$ and $\pi(K)^* = \pi(K)$ (e.g. if π is unitary on K), then

$$\langle \pi^1(\varphi)v, w\rangle = \langle \pi^1({}^K\varphi^K)v, w\rangle.$$

These statements are trivially verified from the definitions, and will be used freely without reference.

Next we come to the basic commutativity first discovered by Gelfand.

Theorem 1. *Let G be locally compact unimodular, and let K be a compact subgroup. Let τ be an anti-automorphism of G of order 2 such that given $x \in G$ there exist $k_1, k_2 \in K$ satisfying*

$$x^\tau = k_1 x k_2.$$

Then the algebra $C_c(G//K)$ is commutative.

Proof. Haar measure is invariant under $x \mapsto x^\tau$ because

$$1 = \Delta(\tau^2) = \Delta(\tau)\,\Delta(\tau),$$

so $\Delta(\tau) = 1$. Also $f(x) = f(x^\tau)$ for any $f \in C_c(G//K)$. Then for

$$f, g \in C_c(G//K)$$

we obtain

$$\begin{aligned} f * g(x) &= \int f(xy^{-1})g(y)\,dy \\ &= \int f({}^\tau y^{-1}x^\tau)g(y^\tau)\,dy && \text{(take } \tau) \\ &= \int f({}^\tau y^{-1})g(x^\tau y^\tau)\,dy && (y \mapsto yx) \\ &= g * f(x). \end{aligned}$$

For this last step, let $y \mapsto y^{-1}$ and replace x^τ by k_1xk_2, using the invariance of Haar measure. This proves Theorem 1.

Example. The hypotheses of Theorem 1 are of course satisfied for $G = SL_2(\mathbf{R})$ and K the circle group. We take τ to be the transpose. The decomposition of a matrix $x = sk$ into a product of a symmetric matrix and an element $k \in K$ immediately shows that ${}^t x = {}^t k s = k_1 x k_2$ because ${}^t k = k^{-1}$.

§2. IRREDUCIBILITY

One of the applications of Theorem 1 will be to irreducible representations, especially unitary ones. Instead of assuming that a representation $\pi: G \to GL(H)$ is unitary, it sometimes suffices to assume that it is **star**

closed, i.e. $\pi(G) = \pi(G)^*$; or **star closed on** K, i.e. $\pi(K) = \pi(K)^*$. Furthermore, it is clear that the closure of $\pi(G)H^K$ is G-invariant, and hence if a representation is irreducible, then H is equal to this closure.

Theorem 2. *Let $\pi\colon G \longrightarrow GL(H)$ be a representation which is star closed on G and K. Assume $H^K \neq 0$ and H is equal to the closure of $\pi(G)H^K$. Then H^K is $C_c(G//K)$-irreducible if and only if H is $C_c(G)$ (and so G) irreducible.*

Proof. $\Rightarrow$: Let W be a closed G-stable subspace $\neq 0$ of H, so that $W^\perp$ is also G-stable (because of star closure). Let $P = P_K$ be the orthogonal projection on H^K. We consider two cases. First, suppose that $PW = 0$. From

$$H = W \oplus W^\perp \qquad \text{and} \qquad PH = PW \oplus PW^\perp,$$

we get $PW^\perp = H^K$, so $H^K \subset W^\perp$, whence $W^\perp = H$, $W = 0$, and we are done. On the other hand, if $PW = W^K \neq H^K$ and $\neq 0$, then W^K is

$$C_c(G//K)\text{-invariant},$$

because if $\varphi \in C_c(G//K)$ and $w \in W$, then

$$\pi^1(\varphi)w \in W,$$

and also $\pi^1(\varphi)w$ is fixed under K. So we are also done in this case, thereby proving the implication in one direction.

$\Leftarrow$: Conversely, assume that H is irreducible. Let $W \neq 0$ be a subspace of H^K, invariant under $C_c(G//K)$, and $W \neq H^K$. Then there exists $v \in H^K$ such that $v \neq 0$, $v \perp W$. We show that

$$\pi^1(C_c(G))v \perp W,$$

whence it follows that the closure of $\pi^1(C_c(G))v$ is a proper subspace. Let $w \in W$ and $\varphi \in C_c(G)$. Then

$$\langle \pi^1(\varphi)v, w\rangle = \int_K\int_K\int_G \langle \varphi(x)\pi(x)v, w\rangle \, dx\, dk_1\, dk_2.$$

Let $x \longmapsto xk_1$. The term $\pi(xk_1)$ splits into $\pi(x)\pi(k_1)$, and $\pi(k_1)$ disappears because v is in H^K. Let $x \longmapsto k_2x$. Then $\pi(k_2x) = \pi(k_2)\pi(x)$, and we move $\pi(k_2)$ over to w, with a star, where it disappears because $w \in H^K$. This shows that the above expression is

$$= \int_G \langle {}^K\varphi^K(x)\pi(x)v, w\rangle \, dx$$

$$= \langle \pi^1({}^K\varphi^K)v, w\rangle = 0,$$

because the orthogonal complement of W in H^K is invariant under $\pi^1(f)$, for $f \in C_c(G//K)$. This contradicts the irreducibility of H.

Theorem 3. *Let $\pi: G \longrightarrow GL(H)$ be a unitary irreducible representation. If $C_c(G//K)$ is commutative, then dim $H^K \leqslant 1$.*

Proof. Suppose dim $H^K \neq 0$. By Theorem 2 we know that H^K is irreducible for $C_c(G//K)$ which is a commutative star closed algebra of operators. Schur's lemma shows that dim $H^K = 1$, as desired.

Naturally, Theorems 2 and 3 apply to $G = SL_2(\mathbf{R})$ and K equal to the circle group. Specializing to this case would not have simplified any proof we have given.

§3. THE SPHERICAL PROPERTY

We continue to assume G unimodular and K compact. We say that a function f on G is **K-spherical**, or **spherical** for short, if it satisfies the following properties.

SPH 1. *f is bi-invariant and continuous.*

SPH 2. *f is an eigenfunction of $C_c(G//K)$ on the right, i.e.*

$$f * \psi = \lambda(f, \psi)f$$

for $\psi \in C_c(G//K)$ and some complex number $\lambda(f, \psi)$.

SPH 3. $f(e) = 1$, *where e is the unit element of G.*

The third condition is a normalization. A function satisfying the first two properties, and such that $f(e) \neq 0$, can be divided by $f(e)$ to yield a function satisfying all three properties.

Note that the eigenvalue $\lambda(f, \psi)$ is

$$\lambda(f, \psi) = (f * \psi)(e),$$

which we see from conditions **SPH 2** and **SPH 3**, evaluating at e.

The next theorem gives a fundamental example of spherical functions.

Theorem 4. *Assume that $G = PK$, where P is a closed subgroup, and $P \times K \longrightarrow PK = G$ is a topological isomorphism. Let*

$$\rho: P \longrightarrow \mathbf{C}^*$$

be a character (continuous homomorphism), which we extend to a function *on G by setting* $\rho(pk) = \rho(p)$. *Then* ρ *is a right eigenvector of* $C_c(G//K)$, *i.e.*

$$\rho * \psi(x) = \lambda(\rho, \psi)\rho(x),$$

and the function f such that

$$f(x) = \int_K \rho(kx)\, dk$$

is K-spherical with eigenvalues $\lambda(\rho, \psi) = \rho * \psi(e) = \lambda(f, \psi)$.

Proof. Write $x = p_1 k_1$. Then for $\psi \in C_c(G//K)$ we get

$$\begin{aligned}\rho * \psi(x) &= \int_G \rho(xy^{-1})\psi(y)\, dy \\ &= \int_G \rho(p_1 y)\psi(y^{-1})\, dy.\end{aligned}$$

Writing $y = pk$ we have $\rho(p_1 y) = \rho(p_1 p) = \rho(p_1)\rho(p) = \rho(p_1)\rho(y)$, so our last expression is

$$\begin{aligned}&= \rho(p_1)\int_G \rho(y)\psi(y^{-1})\, dy \\ &= \lambda(\rho, \psi)\rho(x).\end{aligned}$$

This proves that ρ is an eigenvector, and also gives us the explicit expression for the eigenvalue $\lambda(\rho, \psi)$. For f, we now have

$$\begin{aligned}f * \psi(x) &= \int_G f(xy^{-1})\psi(y)\, dy \\ &= \int_K \int_G \rho(kxy^{-1})\psi(y)\, dy\, dk \\ &= \int_K \lambda(\rho, \psi)\rho(kx)\, dk \\ &= \lambda(\rho, \psi)f(x),\end{aligned}$$

so that f is also an eigenvector, with the same eigenvalue as ρ. Clearly $f(e) = 1$, and f is bi-invariant since ρ is invariant on the right, while the integral takes care of left invariance. This proves Theorem 4.

Example. The abstract nonsense of Theorem 4 has a concrete form in the case we keep in mind for this book, namely $G = SL_2(\mathbf{R})$. The group P is the group of triangular matrices

$$p = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}$$

with $a > 0$ and

$$\rho(h_a) = \rho\left(\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}\right) = a.$$

Let s be a complex number. Then the function

$$\rho^s(p) = a^s$$

is a continuous homomorphism of P into $\mathbf{C}^*$, and in this way we obtain the spherical functions

$$\boxed{x \mapsto \int_K \rho(kx)^s \, dk.}$$

The notation for these will be used systematically in the next chapter. For the moment, we continue with more theorems which are valid under the general setup with G, K.

Theorem 5. *Let f be a continuous function on G, not identically* 0. *Then f is spherical if and only if for all $x, y \in G$ we have*

$$\int_K f(xky) \, dk = f(x)f(y).$$

Proof. Assume that f is spherical. For each x let

$$F_x(y) = \int_K f(xky) \, dk.$$

Let $\varphi \in C_c(G//K)$. Then

$$\begin{aligned} F_x * \varphi(y) &= \int_G F_x(yz^{-1})\varphi(z) \, dz \\ &= \int_G \int_K f(xkyz^{-1})\varphi(z) \, dk \, dz. \end{aligned}$$

Interchange the integrals, let $z \mapsto zy$, then let $z \mapsto zk$. We see that the last expression is

$$\begin{aligned} &= \int_K \int_G f(xz^{-1})\varphi(zky) \, dz \, dk \\ &= (f * \varphi_y')(x) \end{aligned}$$

where

$$\varphi_y'(z) = \int_K \varphi(zky) \, dk.$$

Since φ_y' is bi-invariant, we finally obtain

$$F_x * \varphi(y) = \lambda(f, \varphi_y')f(x),$$

where $\lambda(f, \varphi_y')$ is the eigenvalue. Let $x = e$. Then

$$F_e(y) = f(y),$$

so $F_e = f$. We get

$$(f * \varphi)(y) = \lambda(f, \varphi_y')f(e) = \lambda(f, \varphi_y'),$$

so that

$$F_x * \varphi(y) = (f * \varphi)(y)f(x).$$

On the other hand, let $\{\varphi_n\}$ be a Dirac sequence, and apply what we just obtained to $\varphi = \varphi_n$. We know that

$$F_x * \varphi_n \longrightarrow F_x \quad \text{and} \quad f * \varphi_n \longrightarrow f.$$

Since F_x, f are both bi-invariant, we can replace φ_n by ${}^K\varphi_n^K$. Hence

$$\int_K f(xky)\,dk = F_x(y) = f(y)f(x).$$

This proves half of our theorem.

Conversely, assume that f satisfies the stated functional equation. Let x_0 be such that $f(x_0) \neq 0$. Then

$$f(x_0)f(y) = \int_K f(x_0kk_1y)\,dk = f(x_0)f(k_1y),$$

so $f(y) = f(k_1y)$ for all $k_1 \in K$, and f is left invariant. A similar argument shows that f is right invariant, so f is bi-invariant. Then

$$f(x_0) = \int_K f(x_0k)\,dk = f(x_0)f(e),$$

so that $f(e) = 1$. Finally, let $\varphi \in C_c(G/\!/K)$. By definition,

$$f * \varphi(x) = \int_G f(xy^{-1})\varphi(y)\,dy.$$

Integrate over K on the outside, let $y \longmapsto yk^{-1}$, change the order of integration, to get this expression

$$= \int_G \int_K f(xky^{-1})\varphi(y)\,dk\,dy$$

$$= \int_G f(x)f(y^{-1})\varphi(y)\,dy$$

$$= (f * \varphi(e))f(x).$$

Therefore f is an eigenvector of $C_c(G//K)$, thereby proving the second half of our theorem.

Theorem 6. *Let $f \in C(G//K)$. Then f is spherical if and only if the map*

$$L\colon \varphi \longmapsto \int_G \varphi(x)f(x)\,dx$$

is an algebra homomorphism of $C_c(G//K)$ into $\mathbf{C}$.

Proof. By definition,

$$L(\varphi * \psi) = \int_G \int_G \varphi(xy^{-1})\psi(y)f(x)\,dy\,dx.$$

Interchange $dy\,dx$ to $dx\,dy$, let $x \longmapsto xy$, get the right-hand side

$$= \int_G \int_G \varphi(x)\psi(y)f(xy)\,dx\,dy.$$

Integrate with respect to K on the outside, let $x \longmapsto xk$, move the integral with respect to K inside, getting

$$(1) \qquad L(\varphi * \psi) = \int_G \int_G \varphi(x)\psi(y)\int_K f(xky)\,dk\,dx\,dy.$$

On the other hand,

$$(2) \qquad L(\varphi)L(\psi) = \int_G \int_G \varphi(x)\psi(y)f(x)f(y)\,dx\,dy,$$

so the implication $\Rightarrow$ in Theorem 6 is clear.

Conversely, assume that L is an algebra homomorphism, i.e.

$$L(\varphi * \psi) = L(\varphi)L(\psi)$$

for all $\varphi, \psi \in C_c(G//K)$. Then the functional equation for f follows at once from the equality between (1) and (2), as was to be shown.

Note. If we assume f bounded in Theorem 6, then

$$L\colon \varphi \longmapsto \int_G \varphi(x)f(x)\,dx$$

extends to an algebra homomorphism of $L^1(G/\!/K)$ into $\mathbf{C}$.

Theorem 7. *Any continuous algebra homomorphism of $L^1(G/\!/K)$ into $\mathbf{C}$ is of the form*

$$\varphi \longmapsto (f * \varphi)(e)$$

for some bounded spherical function f.

Proof. By measure theory, given a character $L \neq 0$ of the algebra $L^1(G/\!/K)$, there exists a bounded measurable function f such that

$$L(\varphi) = \int_G \varphi(x)f(x)\,dx, \qquad \text{all } \varphi \in L^1(G/\!/K).$$

Replace $\varphi(x)$ by $\varphi(k_1xk_2)$, integrate with respect to K, let $x \longmapsto k_1^{-1}xk_2^{-1}$. This shows that we can replace f by

$$\int_K \int_K f(k_1xk_2)\,dk_1\,dk_2,$$

i.e. we may assume that f is bi-invariant. From (1) and (2) we get

$$\int_K f(xky)\,dk = f(x)f(y)$$

for almost all $(x, y) \in G \times G$. To show that f can be replaced by a continuous function, let $\psi \in C_c(G)$ be such that

$$\int_G \psi(y)f(y)\,dy \neq 0,$$

and assume without loss of generality that this last integral is equal to 1 (after multiplying ψ by a constant if necessary). Then

$$\int_G \varphi(x)f(x)\,dx \int_G \psi(y)f(y)\,dy = \int_G \int_G \varphi(x)\psi(y) \int_K f(xky)\,dk\,dy\,dx$$

which, after using Fubini and letting $y \longmapsto k^{-1}y$ and $y \longmapsto x^{-1}y$ is

$$= \int_G \int_K \int_G \varphi(x)\psi(k^{-1}x^{-1}y)f(y)\,dy\,dk\,dx$$

$$= \int_G \int_G \int_K \varphi(x)\psi(kx^{-1}y)f(y)\,dk\,dy\,dx.$$

We can replace f by

$$g(x) = \int_G \int_K \psi(kx^{-1}y)f(y)\, dk\, dy,$$

which is continuous. This proves our theorem.

Remark. If G is a Lie group, with a C^∞ structure, then the above function is C^∞ in x if one takes $\psi \in C_c^\infty(G)$.

§4. CONNECTION WITH UNITARY REPRESENTATIONS

Let $\pi\colon G \longrightarrow \mathrm{Aut}(H)$ be a unitary representation of G in a Hilbert space H. Let $u \in H^K$ be a unit vector. We shall consider the coordinate function

$$f(x) = \pi_{u,u}(x) = \langle \pi(x)u, u\rangle.$$

Clearly, f is bi-invariant and $f(e) = 1$.

A vector $v \in H$ is said to **generate H topologically (under π)** if H is the closure of the algebraic subspace generated by the translates $\pi(x)v$ for all $x \in G$.

Theorem 8. *Let $\pi\colon G \longrightarrow \mathrm{Aut}(H)$ be a unitary representation and assume that there exists a unit vector $u \in H^K$ which generates H topologically under π. Then*

$$\dim H^K = 1 \Leftrightarrow \textit{the function } f(x) = \langle \pi(x)u, u\rangle \textit{ is spherical.}$$

Proof. We have seen that $f(e) = 1$ and f is bi-invariant. Assume that $H^K = \mathbf{C}u$ has dimension 1. For any $\varphi \in C_c(G/\!/K)$, $\pi^1(\varphi)u$ is fixed by K, so $\pi^1(\varphi)u = \lambda(\varphi)u$ for some $\lambda(\varphi) \in \mathbf{C}$, and

$$\varphi \longmapsto \lambda(\varphi)$$

is a homomorphism. But

$$\begin{aligned} f * \varphi^-(e) &= \int_G f(x)\varphi(x)\, dx \\ &= \int_G \langle \varphi(x)\pi(x)u, u\rangle\, dx \\ &= \langle \pi^1(\varphi)u, u\rangle. \end{aligned}$$

So $\varphi \longmapsto f * \varphi^-(e)$ is a homomorphism of $C_c(G/\!/K)$ into $\mathbf{C}$, whence f is spherical, by Theorem 6. Note that for this side of the implication, we did not need that u is a topological generator of H.

Conversely, assume that f is spherical. Let

$$Pv = \int_K \pi(k)v\,dk$$

be the projection on H^K. For any $b \in G$, we have

$$\begin{aligned}\langle P\pi(x)u, \pi(b)u\rangle &= \langle \pi(b^{-1})P\pi(x)u, u\rangle \\ &= \int_K f(b^{-1}kx)\,dk \\ &= f(b^{-1})f(x) \qquad \text{(Theorem 5)} \\ &= \langle f(x)u, \pi(b)u\rangle.\end{aligned}$$

Since the vectors $\pi(b)u$, $b \in G$, generate a dense subspace of H, it follows that

$$P\pi(x)u = f(x)u.$$

Hence $PH = \mathbf{C}u$ and $\dim H^K = 1$. This proves our theorem.

§5. POSITIVE DEFINITE FUNCTIONS

A function on G is called **positive definite** if and only if it is continuous, $\neq 0$, and for all $x_1, \dots, x_n \in G$ and $\alpha_1, \dots, \alpha_n \in \mathbf{C}$ we have

$$\sum_{i,j} f(x_i x_j^{-1})\alpha_i \overline{\alpha_j} \geqslant 0.$$

This last condition can also be written

$$\sum_{i,j} f(x_i^{-1} x_j)\alpha_i \overline{\alpha_j} \geqslant 0,$$

replacing x_i by x_i^{-1}.

Example. Let $\pi: G \longrightarrow \mathrm{Aut}(H)$ be a unitary representation on a Hilbert space H, and let u be a unit vector. Then

$$f(x) = \langle \pi(x)u, u\rangle$$

is positive definite, because

$$\sum_{i,j} \langle \pi(x_i x_j^{-1})\alpha_i \overline{\alpha_j}\, u, u \rangle = \left\langle \sum_i \pi(x_i)\alpha_i u, \sum_j \pi(x_j)\alpha_j u \right\rangle$$

$$\geqslant 0.$$

We shall see that the example is essentially the only one.

First we enumerate three simple properties.

$$f(e) \text{ is real } \geqslant 0 \tag{1}$$

$$f(x^{-1}) = \overline{f(x)} \tag{2}$$

$$|f(x)| \leqslant f(e) \tag{3}$$

To prove the first, take $n = 1$, $x = e$, and $\alpha_1 = 1$. For the second, take $x_1 = e$, $x_2 = x$, $\alpha_1 = \alpha_2 = 1$. Write

$$f(x) = r(x) + it(x).$$

We know that

$$f(e)\alpha_1\overline{\alpha_1} + f(x^{-1})\alpha_1\overline{\alpha_2} + f(x)\alpha_2\overline{\alpha_1} + f(e)\alpha_2\overline{\alpha_2} \geqslant 0.$$

We see that $f(x^{-1}) + f(x)$ is real, so $t(x^{-1}) = -t(x)$. Take $\alpha_1 = i$ and $\alpha_2 = 1$. Then

$$r(x^{-1}) + r(x)\overline{\alpha_1} = r(x^{-1})i - r(x)i$$

is supposed to be real, so $r(x^{-1}) = r(x)$. This proves the second property. For the third, let

$$\alpha_2 = -|f(x)| \qquad \text{and} \qquad \alpha_1 = f(x).$$

We get

$$f(e)|f(x)|^2 - 2|f(x)|^3 + f(e)|f(x)|^2 \geqslant 0,$$

whence

$$|f(x)|^3 \leqslant f(e)|f(x)|^2.$$

If $f(x) = 0$, we are done because $f(e) \geqslant 0$; and if $f(x) \neq 0$, we cancel $|f(x)|^2$ to get our result.

Given a positive definite function φ, we shall now construct a unitary representation as in the example. Let V_φ be the vector space generated by right translates of φ, i.e. for $a \in G$, by functions $\pi(a)\varphi$ such that

$$\pi(a)\varphi(x) = \varphi(xa).$$

Then $\pi(ab) = \pi(a)\pi(b)$. If $f, g \in V_\varphi$, then

$$f(x) = \sum \alpha_i \varphi(xa_i) \qquad \text{and} \qquad g(x) = \sum \beta_j \varphi(xb_j)$$

with $\alpha_i, \beta_j \in \mathbf{C}$. Define

$$\begin{aligned}\langle f, g\rangle &= \sum \alpha_i \overline{\beta_j} \varphi(b_j^{-1}a_i) \\ &= \sum \alpha_i \overline{\beta_j}\, \overline{\varphi(a_i^{-1}b_j)} \\ &= \sum \alpha_i \overline{g(a_i^{-1})} \\ &= \sum \overline{\beta_j}\, \overline{f(b_j^{-1})}\,.\end{aligned}$$

Given an expression for g as linear combination of translates of φ, these equalities show that $\langle f, g\rangle$ is independent of the expression of f as a linear combination of translates of φ. Similarly on the other side (f and g interchanged), so our symbol $\langle f, g\rangle$ is well defined, and clearly gives a positive hermitian product on V_φ, not necessarily definite.

Let V_φ^0 be the null space of the hermitian product, and let H_φ be the completion of V_φ / V_φ^0, so that H_φ is a Hilbert space. Then the translation operators of G in V_φ, which are unitary, give unitary operators on V_φ / V_φ^0, and therefore extend to unitary operators of H_φ. We therefore obtain an algebraic homomorphism

$$\pi_\varphi = \pi \colon G \longrightarrow \operatorname{Aut}(H).$$

We shall now verify that the continuity condition for a representation is satisfied. We had already pointed out at the beginning of the book that it suffices to verify this condition on a dense subset, and thus it suffices to verify it on elements $f \in V_\varphi$.

Write $f(x) = \sum \alpha_i \varphi(xa_i)$, so that for $y \in G$,

$$\pi(y)f(x) = \sum \alpha_i \varphi(xya_i).$$

Then

$$\begin{aligned}\|\pi(y)f - f\|^2 &= \langle \pi(y)f - f, \pi(y)f - f\rangle \\ &= 2\langle f, f\rangle - \langle \pi(y)f, f\rangle - \langle f, \pi(y)f\rangle \\ &= 2\sum \alpha_i \overline{\alpha_j} \varphi(a_j^{-1}a_i) - \sum \alpha_i \overline{\alpha_j} \varphi(a_i^{-1}y^{-1}a_j) - \sum \alpha_i \overline{\alpha_j} \varphi(a_i^{-1}ya_j).\end{aligned}$$

This last expression tends to 0 as $y \to e$. Hence

$$y \mapsto \pi(y)f$$

is continuous, as a map from G into H, in a neighborhood of e. Hence the map is continuous everywhere, and π is a unitary representation.

Taking for our unit vector u the function φ itself, we see that

$$\varphi(x) = \langle \pi(x)\varphi, \varphi \rangle.$$

In this way we see that every positive definite function arises as in the example. More formally, consider triples (π, H, u) consisting of a unitary representation $\pi: G \to \mathrm{Aut}(H)$, and a unit vector u. An **isomorphism** from (π, H, u) to (π', H', u') is a unitary isomorphism from H to H', commuting with the action of G, and sending u to u'.

Theorem 9. *The association*

$$\varphi \mapsto (\pi_\varphi, H_\varphi, \varphi)$$

is a bijection from the set of positive definite functions on G to the isomorphism classes of triples (π, H, u) consisting of a unitary representation

$$\pi: G \to \mathrm{Aut}(H)$$

and a unit vector u which generates H topologically under π.

Corollary. *Let K be a compact subgroup of G such that $C_c(G//K)$ is commutative. In the bijection of Theorem 9, the positive definite spherical functions correspond to the irreducible unitary representations having a K-fixed vector u.*

Proof. This is an immediate consequence of Theorem 8, §4 and the irreducibility theorem of §2.

The preceding result of course applies to $SL_2(\mathbf{R})$. It would not have been simpler to give the proof in the concrete case rather than in abstract nonsense.

V The Spherical Transform

In this chapter we study the spectral decomposition of the algebra $C_c^\infty(G//K)$, consisting of those C^∞ functions with compact support, bi-invariant under K. We shall also determine the bounded spherical functions. We let $G = SL_2(\mathbf{R})$ throughout.

§1. INTEGRAL FORMULAS

Harish-Chandra put in evidence the important role played by the two subgroups A and K in G, and the manner in which $A\backslash G$ and $K\backslash G$ give rise to the Plancherel inversion. In this chapter, we deal only with $A\backslash G$, and postpone the relations between G, $A\backslash G$, and $K\backslash G$ to the later chapter on the Plancherel formula.

Functions on $A\backslash G/K$ amount to functions on N, and so we begin by integral formulas relating integration on $A\backslash G$ and N. The first formula requires the computation of a "Jacobian", and cannot be handled by abstract nonsense. Its proof will therefore be given by using the explicit matrix representation of $G = SL_2(\mathbf{R})$.

For any $f \in C_c(G)$ and $a \in A$ such that $\alpha(a) \neq 1$, we have

I 1. $$\int_N f(ana^{-1}n^{-1})\,dn = \frac{1}{|\alpha(a) - 1|}\int_N f(n)\,dn.$$

Proof. Matrix multiplication shows that

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a^{-1} & 0 \\ 0 & a \end{pmatrix}\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & (a^2-1)n \\ 0 & 1 \end{pmatrix}$$

Let $t = (a^2 - 1)n$, change variables. The formula drops out.

(Cf. Helgason [He 2], Chapter X, Proposition 1.13, for the proof in the general case of Lie groups. The matrix computation requires a more elaborate analogue in the Lie algebra.)

The next formula luckily depends only on **I 1** and a general Iwasawa decomposition $G = ANK$, $dx = da\,dn\,dk$. Even though we should use the notation $\dot{x}$ for a "coset" variable in, say, $A\backslash G$, we shall sometimes use x for simplicity of notation. We do write $d\dot{x}$, however, for the Haar measure in $A\backslash G$, such that $dx = da\,d\dot{x}$.

Let $f \in C_c(G)$ and $a \in A$ be such that $\alpha(a) \neq 1$. Then the function $x \mapsto f(x^{-1}ax)$ has compact support on $A\backslash G$, and we have

I 2.
$$\int_{A\backslash G} f(x^{-1}ax)\,d\dot{x} = \frac{1}{|\alpha(a^{-1}) - 1|}\int_K \int_N f(kank^{-1})\,dn\,dk$$
$$= \frac{\alpha(a)^{1/2}}{|\alpha(a)^{1/2} - \alpha(a)^{-1/2}|}\int_K \int_N f(kank^{-1})\,dn\,dk$$

Proof. Let $\varphi(x) = f(x^{-1}ax)$. It is clear that φ has compact support on $A\backslash G$. By general theorems on homogeneous spaces, if $g \in C_c(G)$ is such that ${}^A g = \varphi$, we get, using $dx = da\,dn\,dk$,

$$\int_{A\backslash G} \varphi(x)\,d\dot{x} = \int_G g(x)\,dx = \int_N \int_K \varphi(nk)\,dn\,dk$$
$$= \int_N \int_K f(k^{-1}n^{-1}ank)\,dn\,dk$$
$$= \int_N \int_K f(k^{-1}aa^{-1}n^{-1}ank)\,dn\,dk$$
$$= \frac{1}{|\alpha(a^{-1}) - 1|}\int_K \int_N f(k^{-1}ank)\,dn\,dk \quad (\text{by } \mathbf{I\ 1}).$$

This proves the formula.

We use the notation

$$D(a) = \alpha(a)^{1/2} - \alpha(a)^{-1/2} = \rho(a) - \rho(a)^{-1}.$$

Then

$$|D(a)| = |D(a^{-1})|$$

and

$$\frac{\alpha(a)^{1/2}}{|\alpha(a)^{1/2} - \alpha(a)^{-1/2}|} = \frac{\alpha(a)^{1/2}}{|D(a)|}.$$

Depending on situations, it is often convenient to change the coefficient of the integral in **I 2** according to these identities.

§2. THE HARISH TRANSFORM

Let $C_c(G, K)$ denote the space of functions with compact support which are invariant under conjugation by K, i.e. satisfy

$$f(k^{-1}xk) = f(x)$$

for all $k \in K$ and $x \in G$.

For $f \in C_c(G, K)$ we define the **Harish transform**

$$\mathbf{H}f(a) = \rho(a)\int_N f(an)\,dn = |D(a)|\int_{A\backslash G} f(x^{-1}ax)\,d\dot{x}.$$

The first integral expression is valid for all $a \in A$, the second only for a such that $\rho(a) \neq 1$. Later we shall define a Harish transform for the other Cartan subgroup K, and hence the above transform will be written $\mathbf{H}^A f$ to denote its dependence on A. For this chapter, we shall deal almost exclusively with the integral expression over N rather than that over $A\backslash G$ for the Harish transform. It shows that if $f \in C_c^\infty(G, K)$, then $\mathbf{H}f \in C_c^\infty(A)$.

Let

$$w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

be called the **Weyl element**, and let the group of order 2 which it generates (mod ± 1) be called the **Weyl group**. Then w operates on A by conjugation, and we have

$$a^w = a^{-1},$$

or in matrix form

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a^{-1} & 0 \\ 0 & a \end{pmatrix}.$$

Note that $w^2 = -1$. A set of representatives for A modulo the Weyl group is the set A^+, consisting of $a \in A$ with $\rho(a) \geqslant 1$, i.e. of matrices

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, \quad \text{with} \quad a \geqslant 1.$$

The Weyl group also operates on functions in the obvious way.

Theorem 1. *$\mathbf{H}f$ is invariant under the Weyl group, i.e.*

$$\mathbf{H}f(a) = \mathbf{H}f(a^{-1}).$$

Proof. By continuity it suffices to prove the assertion when $D(a) \neq 0$, and so $|D(a)| = |D(a^{-1})|$. Note that

$$x \longmapsto wxw^{-1}$$

is an inner automorphism of G, of order 2, sending $a \longmapsto a^{-1}$. This map preserves the measure on $A \backslash G$, and hence, using the second integral expression for the Harish transform, the invariance of $\mathbf{H}f$ under the Weyl group is clear.

The Harish transform is therefore a linear map

$$\mathbf{H}\colon C_c(G, K) \longrightarrow C_c(A)^w.$$

where the upper index w means the space of functions invariant under w.

Theorem 2. *If $f, g \in C_c(G//K)$, then*

$$\mathbf{H}(f * g) = \mathbf{H}f * \mathbf{H}g,$$

i.e. on $C_c(G//K)$, the Harish transform is an algebra homomorphism.

Proof. We have

$$\begin{aligned} \mathbf{H}(f * g)(a) &= \rho(a) \int_N (f * g)(an)\, dn \\ &= \rho(a) \int_N \int_G f(any) g(y^{-1})\, dy\, dn \qquad (\text{by } y \longmapsto y^{-1}) \\ &= \rho(a) \int_N \int_G f(ay) g(y^{-1}n)\, dy\, dn. \end{aligned}$$

Letting $y = bmk$, we obtain

$$\begin{aligned}
\mathbf{H}(f*g)(a) &= \rho(a)\int_N \int_A \int_N f(abm)g(m^{-1}b^{-1}n)\, db\, dm\, dn \\
&= \rho(a)\int_N \int_A \int_N f(ab^{-1}m)g(m^{-1}bn)\, db\, dn\, dm \\
&= \rho(a)\rho(b)^{-2}\int\int\int f(ab^{-1}m)g(m^{-1}nb)\, db\, dn\, dm \\
&= \rho(a)\rho(b)^{-2}\int\int\int f(ab^{-1}m)g(nb)\, db\, dn\, dm \\
&= \rho(a)\int\int\int f(ab^{-1}m)g(bn)\, db\, dn\, dm.
\end{aligned}$$

But

$$\begin{aligned}
\mathbf{H}f * \mathbf{H}g(a) &= \int_A \mathbf{H}f(ab^{-1})\mathbf{H}g(b)\, db \\
&= \int\int\int \rho(ab^{-1})f(ab^{-1}m)\rho(b)g(bn)\, dm\, dn\, db,
\end{aligned}$$

which is the same as the last expression obtained above, thus proving our theorem.

The main theorem about the Harish transform is:

Theorem 3.

$$\mathbf{H}\colon C_c^\infty(G//K) \to C_c^\infty(A)^w$$

is an isomorphism.

The proof will give an explicit inversion for the Harish transform, due to Godement [Go 6]; see also Harish [H-C 6].

We use P to denote the set of positive definite symmetric 2×2 matrices. Any element $p \in P$ can be diagonalized, and there exists an orthonormal basis of $\mathbf{R}^2$ consisting of eigenvectors for p. Consequently we have

$$p = k^{-1}ak$$

for some $k \in K$ and $a \in A$.

On the other hand, any $x \in G$ can be written as a product

$$x = pk_1$$

with some $p \in P$ and $k_1 \in K$. Hence

$${}^t xx = k_1^{-1}p^2k_1 = k_1^{-1}k^{-1}a^2kk_1.$$

If f is a bi-invariant function, then

$$f(x) = f(p) = f(a) = f(a^{-1})$$

because $w \in K$ and $waw^{-1} = a^{-1}$. Thus f depends only on $\alpha(a)$, and we note that the eigenvalues of ${}^t xx$ are $\alpha(a)$ and $\alpha(a^{-1})$. We have

$$\frac{\operatorname{tr}({}^t xx)}{2} = \frac{\alpha(a) + \alpha(a^{-1})}{2}.$$

Now we switch notation in order to deal with coordinates. We write.

$$x = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

If x is diagonal, then x is of the form

$$h_a = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$$

and $\alpha(h_a) = a^2$. The mapping

$$a \mapsto \frac{a^2 + a^{-2}}{2} = \frac{\alpha(a) + \alpha(a^{-1})}{2}$$

is a bijection

$$[1, \infty) \to [1, \infty)\,.$$

If f_G is a bi-invariant function on G, then f_G depends only on the value $\operatorname{tr}({}^t xx)$, and therefore we use the new variable

$$v = \frac{\operatorname{tr}({}^t xx)}{2} = \frac{a^2 + b^2 + c^2 + d^2}{2}.$$

If x is in diagonal form as above, then

$$v = \frac{a^2 + a^{-2}}{2}.$$

The function f_G corresponds to a function of a real variable $\geqslant 1$,

$$f_G(x) = f_G\begin{pmatrix} a & b \\ c & d \end{pmatrix} = f\left(\frac{a^2 + b^2 + c^2 + d^2}{2}\right) = f(v).$$

In terms of these coordinates, we can write the Harish transform as

$$\begin{aligned} \mathbf{H}f_G(h) &= \alpha(h)^{1/2}\int_N f_G(hn)\,dn \\ &= \alpha(h)^{1/2}\int_{-\infty}^{\infty} f_G\begin{pmatrix} a & au \\ 0 & a^{-1} \end{pmatrix} du. \end{aligned}$$

Therefore we have the ordinary real integral expression

$$\mathbf{H}f_G(h_a) = a\int_{-\infty}^{\infty} f\left(\frac{a^2 + a^{-2}}{2} + \tfrac{1}{2}a^2u^2\right) du$$

which, after a change of variables, yields

$$(1) \qquad \boxed{\mathbf{H}f(a) = \mathbf{H}f_G(h_a) = \int_{-\infty}^{\infty} f\left(\frac{a^2 + a^{-2}}{2} + \tfrac{1}{2}u^2\right) du.}$$

Lemma. *If*

$$F(v) = \int_{-\infty}^{\infty} f(v + \tfrac{1}{2}u^2)\, du, \qquad v \geqslant 1,$$

then

$$f(v) = -\frac{1}{2\pi}\int_{-\infty}^{\infty} F'(v + \tfrac{1}{2}w^2)\, dw,$$

and conversely.

Proof. (Freshman calculus). Differentiating under the integral sign yields

$$F'(v) = \int_{-\infty}^{\infty} f'(v + \tfrac{1}{2}u^2)\, du,$$

so

$$\begin{aligned}\int_{-\infty}^{\infty} F'(v + \tfrac{1}{2}w^2)\, dw &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f'\left(v + \frac{u^2 + w^2}{2}\right) du\, dw \\ &= \int_0^{2\pi}\int_0^{\infty} f'\left(v + \frac{r^2}{2}\right) r\, dr\, d\theta \\ &= 2\pi\int_0^{\infty} f'(v + x)\, dx \\ &= -2\pi f(v),\end{aligned}$$

as was to be shown. (The converse is equally clear.)

Observe that the lemma also proves Theorem 3 by giving the inversion formula for the Harish transform explicitly in terms of the matrix coordinates.

In the literature one sometimes finds another change of variables which corresponds to parametrizing an element of A by the exponential map from the Lie algebra. In our case, put

$$a = e^{t/2} \qquad \text{so} \qquad v = \frac{e^t + e^{-t}}{2} = \cosh t.$$

Recall that

$$\sinh t = \frac{e^t - e^{-t}}{2} \qquad \text{and} \qquad \cosh^2 t - \sinh^2 t = 1.$$

We have an expression for the Harish transform in terms of t, namely

$$\text{(2)} \qquad \mathbf{H}f_G(h_a) = F(v) = \Phi(t) = \int_{-\infty}^{\infty} f\left(\frac{e^t + e^{-t}}{2} + \tfrac{1}{2}u^2\right) du$$

and

$$\Phi'(t) = F'(v)\sinh t.$$

Therefore

$$\text{(3)} \qquad -2\pi f_G(1_G) = \int_{-\infty}^{\infty} F'(1 + \tfrac{1}{2}w^2)\, dw.$$

Let $w = e^{t/2} - e^{-t/2}$, change variables. We find the

Inversion formula. *If $f \in C_c^\infty(G//K)$ and $\Phi(t)$ is the expression for the Harish transform as in* (2), *then*

$$-2\pi f_G(1_G) = \int_{-\infty}^{\infty} \Phi'(t) \frac{dt}{e^{t/2} - e^{-t/2}}.$$

This a special case of the Plancherel inversion formula to be proved later for functions which are not necessarily bi-invariant.

§3. THE MELLIN TRANSFORM

Let A be as before, the group of diagonal matrices with determinant 1 and positive diagonal elements. On A we have the **Mellin transform**, defined for $g \in C_c(A)$ by

$$\mathbf{M}g(s) = \int_A g(a)\rho(a)^s\, da, \qquad s \in \mathbf{C}.$$

This is obviously an entire function of s. If $g \in C_c(A)^w$, i.e. $g(a) = g(a^{-1})$, then letting $a \mapsto a^{-1}$ in the integral shows that

$$\mathbf{M}g(s) = \mathbf{M}g(-s).$$

Functions on the complex numbers which are even will be said to be **invariant under the Weyl group**, i.e. we denote by

$$\mathrm{Hol}(\mathbf{C})^w$$

the space of entire functions $h(s)$ satisfying $h(s) = h(-s)$. Then the Mellin transform gives us a linear map

$$\mathbf{M}\colon C_c(A)^w \to \mathrm{Hol}(\mathbf{C})^w.$$

The group A amounts to the multiplicative group, and we can write the Mellin transform of functions on $\mathbf{R}^+ = (0, \infty)$ by

$$Mf(s) = \int_0^\infty f(a)a^s \frac{da}{a}, \qquad f \in C_c(\mathbf{R}^+)$$

where a is a multiplicative variable. We wish to characterize its image. We define the **Paley–Wiener space** $PW(\mathbf{C})$ to consist of those entire functions f for which there exists a positive number C such that given an integer $N > 0$ we have

$$|f(\sigma + it)| \ll \frac{C^{|\sigma|}}{(1 + |t|)^N},$$

where the implied constant in $\ll$ depends on f and N (could be taken of the form C^N). In words, we may say that f has at most exponential growth with respect to σ, and is rapidly decreasing, uniformly in every strip of finite width.

If f is C^∞ in addition to having compact support, then its Mellin transform lies in the Paley–Wiener space. To see this, we integrate by parts:

$$\int_0^\infty f(a)a^{s-1}\,da = f(a)\frac{a^s}{s}\Big|_0^\infty - \frac{1}{s}\int_0^\infty f'(a)a^s\,da.$$

Since f has compact support on the open interval $(0, \infty)$, the first term on the right is 0. Continuing to integrate by parts, we pick up successive derivatives of f, which all have compact support, and we get successive products

$$\frac{1}{s(s+1)(s+2)\cdots(s+n)}$$

in the denominator, which show that the Mellin transform goes to 0 rapidly in a strip of fixed width. The exponential growth in σ is clear.

If f is C^∞, differentiating under the integral sign shows that the n-th derivative is given by

$$(Mf)^{(n)}(s) = \int_0^\infty f(a)a^s(\log a)^n \frac{da}{a},$$

and therefore that all the derivatives of the Mellin transform also lie in the Paley–Wiener space.

We want to invert the Mellin transform, and we prove:

Theorem 4. *The Mellin transform*

$$\mathbf{M}\colon C_c^\infty(A)^w \to PW(\mathbf{C})^w$$

is an isomorphism.

Proof. Actually, we prove that it is an isomorphism omitting the action of w. Fix any real number σ, and for $F \in PW(\mathbf{C})$ define

$${}^t M_\sigma F(a) = \int_{\mathrm{Re}\, s = \sigma} F(s)a^s \frac{ds}{i}.$$

Then ${}^t M_\sigma F(a)$ is independent of σ. Indeed, let $\sigma_1 < \sigma_2$ and integrate around a rectangle as shown in the diagram of Fig. 1.

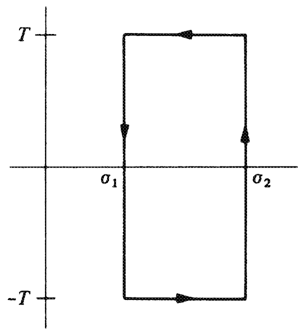

Figure 1

The integral over the rectangle R of the function below is 0:

$$\int_R F(s)a^s \frac{ds}{i} = 0.$$

If $s = \sigma_1 + it$, then $ds = i\,dt$ (explaining our dividing by i). On the other hand, we have the estimate

$$\left|\int_{\text{top}} |F(\sigma + iT)|a^\sigma \,d\sigma\right| \ll \frac{1}{T^n} \frac{e^{\sigma \log a}}{\log a}\Bigg|_{\sigma_1}^{\sigma_2}$$

and similarly for the bottom, showing that the integrals over the top and bottom tend to 0 as $T \to \infty$. This proves that

$$'M_\sigma F(a)$$

is independent of σ.

The function $'M_\sigma F$ has compact support on $\mathbf{R}^+$. To see this, let $a > 0$. For σ large, we have

$$\left|\int_{\operatorname{Re} s=\sigma} F(\sigma + it)a^\sigma \frac{ds}{i}\right| \ll \int_0^\infty \frac{C^\sigma a^\sigma}{(1 + |t|^N)}\,dt$$

$$\ll (aC)^\sigma.$$

Take $a < 1/C$ and let $\sigma \to \infty$. We see that $'M_\sigma F(a) = 0$. If a is large, we integrate over $\operatorname{Re} s = -\sigma$ and use a similar estimate to see that $'M_\sigma F(a) = 0$ also. This proves our assertion that $'M_\sigma F$ has compact support.

There remains to prove the inversion formula. Define

$$'M^- F(a) = \int_{\operatorname{Re} s=\sigma} F(s)a^{-s} \frac{ds}{i}.$$

We shall prove that

$$\boxed{'M^- Mf = 2\pi f}$$

for $f \in C_c(\mathbf{R}^+)$, and similarly on the other side. In fact, up to a change of variables, this is merely the Fourier inversion formula. Indeed, write $a = e^x$, so that

$$Mf(s) = \int_{-\infty}^{\infty} f(e^x)e^{\sigma x}e^{itx}\,dx = \hat{F}_\sigma(-t),$$

where

$$F_\sigma(x) = f(e^x)e^{\sigma x}.$$

Then

$$\begin{aligned} {}^t M^- F(a) &= \int_{\mathrm{Re}\, s=\sigma} F(s)a^{-s}\,\frac{ds}{i} \\ &= \int_{-\infty}^{\infty} F(\sigma + it)a^{-\sigma - it}\,dt \\ &= a^{-\sigma}\int_{-\infty}^{\infty} F(\sigma + it)a^{-it}\,dt. \end{aligned}$$

We see that this is again the inverse Fourier transform, and our inversion follows by the Fourier inversion.

§4. THE SPHERICAL TRANSFORM

Let, as before,

$$\boxed{\varphi_s(x) = \int_K \rho(kx)^{s+1}\,dk = \varphi(x, s),}$$

where K is the circle group and x the variable in $G = SL_2(\mathbf{R})$. For

$$f \in C_c(G/\!/K),$$

define the **spherical transform** by means of the kernel $\varphi(x, s)$, namely

$$\mathbf{S}f(s) = \int_G f(x)\varphi_s(x)\,dx.$$

Theorem 5. *On $C_c^\infty(G/\!/K)$ we have a a commutative diagram*

$$\begin{tikzcd} C_c^\infty(G/\!/K) \arrow[rr, "\mathbf{H}"] \arrow[dr, "\mathbf{S}"'] & & C_c^\infty(A)^w \arrow[dl, "\mathbf{M}"] \\ & PW(\mathbf{C})^w & \end{tikzcd}$$

i.e. $\mathbf{S} = \mathbf{MH}$, *and all the arrows are isomorphisms.*

Proof.

$$\begin{aligned}
\mathbf{S}f(s) &= \int_G f(x)\varphi_s(x)\,dx \\
&= \int_K \int_G f(x)\rho(kx)^{s+1}\,dx\,dk \\
&= \int_A \int_N \int_K f(an)\rho(ka)^{s+1}\,da\,dn\,dk \\
&= \mathbf{MH}f(s).
\end{aligned}$$

This proves our theorem, because we have already shown that **M** and **H** are isomorphisms.

Corollary 1. $\varphi_s = \varphi_{-s}$.

Proof. For every $f \in C_c^\infty(G//K)$ we get $\mathbf{S}f(s) = \mathbf{S}f(-s)$, so the integrals of φ_s and φ_{-s} against f give the same value. Hence $\varphi_s = \varphi_{-s}$.

Corollary 2. *φ_s is bounded by* 1 *for* $-1 \leqslant \mathrm{Re}\, s \leqslant 1$.

Proof. We shall give two proofs, each one illustrating a useful technique. The first is that of Helgason–Johnson [He, Jo], who proved the result in general. Let $f \in \mathfrak{L}^1(G)$ be bi-invariant. Then

$$\int |f(x)|\,dx = \int |f(kan)|\rho(a)^2\,dk\,da\,dn = \int_A \rho(a)(\mathbf{H}|f|)(a)\,da$$

is finite. If $-1 \leqslant \sigma = \mathrm{Re}\, s \leqslant 1$, then

$$\begin{aligned}
\int |\varphi_s(x)f(x)|\,dx \leqslant \int \varphi_\sigma(x)|f(x)|\,dx &= (\mathbf{MH}|f|)(\sigma) \\
&= \int_{A^+} (\mathbf{H}|f|)(a)(\rho(a)^\sigma + \rho(a)^{-\sigma})\,da \\
&\leqslant 2\int_{A^+} \rho(a)(\mathbf{H}|f|)(a)\,da < \infty.
\end{aligned}$$

Thus the integral of φ_s against any function in $\mathfrak{L}^1(G)$ is finite, and therefore φ_s is bounded.

The second proof is due to Eli Stein. Fix a, and view $\varphi_s(a)$ as a function of s. For $\sigma = -1$ we get the bound of 1 trivially. By Corollary 1, we conclude that the bound of 1 also applies at $\sigma = 1$. The growth is obviously at most exponential in the strip. By the Phragmen–Lindelöf theorem, we conclude that φ_s is bounded in the strip, as was to be shown.

It will be proved in the next section that the values of s in the above strip are the only ones for which the spherical function φ_s is bounded.

For the convenience of the reader, we reproduce the proof of the Phragem–Lindelöf theorem.

Phragmen–Lindelöf Theorem. *Let f be holomorphic in a strip $\sigma_1 \leqslant \sigma \leqslant \sigma_2$, and bounded by 1 in absolute value on the sides of the strip. Assume that there is a number $\alpha \geqslant 1$ such that $f(s) = O(e^{|s|^\alpha})$ in the strip. Then f is bounded by 1 in the whole strip.*

Proof. For all sufficiently large $|t|$ we have

$$|f(\sigma + it)| \leqslant e^{|t|^\lambda}$$

if we take $\lambda > \alpha$. Select an integer $m \equiv 2 \pmod 4$ such that $m > \lambda$. If $s = re^{i\theta}$, then

$$s^m = r^m(\cos m\theta + i \cdot \sin m\theta),$$

and $m\theta$ is close to π. Consider the function

$$g_\epsilon(s) = g(s) = f(s)e^{\epsilon s^m},$$

with $\epsilon > 0$.

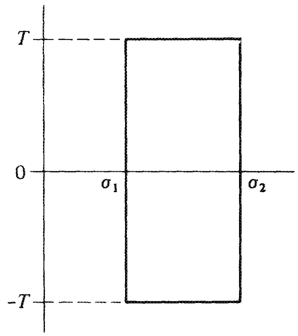

Figure 2

Then for s in the strip we get

$$|g(s)| \leqslant e^{|t|^\lambda} e^{\epsilon r^m \cos m\theta}.$$

Consequently for large T the function $g(s)$ is bounded by 1 on the horizontal segment $t = T$, $\sigma_1 \leqslant \sigma \leqslant \sigma_2$. It is also clear that $|g(s)|$ is bounded by 1 on the sides of the rectangle, as shown in Fig. 2. Hence

$$|f(s)| \leqslant e^{-\epsilon r^m \cos m\theta}$$

inside the rectangle. This is true for every $\epsilon > 0$, and hence

$$|f(s)| \leqslant 1$$

inside the rectangle, thus proving the theorem.

We end this section by making explicit the inversion formula for the spherical transform, which we know exists by Theorem 5. We want to see that the kernel $\varphi(x, s)$ gives us the inversion for s on the line $\sigma = 0$, so viewing $\varphi(x, s)$ as $\varphi(x, it)$.

We write our spherical transform as

$$\mathbf{S}f(s) = \int_G f(x)\varphi(x, s)\, dx.$$

Theorem 6. *Let $f \in C_c^\infty(G//K)$. Then*

$$f(1) = \int_{-\infty}^{\infty} \mathbf{S}f(i\tau)\tau \tanh(\pi\tau)\, \frac{d\tau}{2\pi}.$$

Proof. We keep our old notation, letting

$$h_a = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, \qquad v = \frac{a^2 + a^{-2}}{2}, \qquad a = e^{t/2}$$

and

$$\mathbf{H}f(h_a) = F(v).$$

By Mellin inversion we find

$$F\left(\frac{a^2 + a^{-2}}{2}\right) = \int_{-\infty}^{\infty} \mathbf{S}f(i\tau)a^{-2i\tau}\, \frac{d\tau}{2\pi} = \int_{-\infty}^{\infty} \mathbf{S}f(i\tau)\frac{a^{2i\tau} + a^{-2i\tau}}{2}\, \frac{d\tau}{2\pi}.$$

Hence

$$F(\cosh t) = \int_{-\infty}^{\infty} \mathbf{S}f(i\tau) \cos(t\tau)\, \frac{d\tau}{2\pi}$$

and

$$F'(\cosh t) = \int_{-\infty}^{\infty} \mathbf{S}f(i\tau)\frac{\sin(t\tau)}{\sinh t}(-\tau)\,\frac{d\tau}{2\pi}.$$

From the definition of the Harish transform, we get

$$\begin{aligned} 2\pi f(1) &= \int_{-\infty}^{\infty} F'(1+\tfrac{1}{2}u^2)\,du \\ &= \int_{-\infty}^{\infty} F'(\cosh t)\cosh\frac{t}{2}\,dt \\ &= \int_{-\infty}^{\infty} \mathbf{S}f(i\tau)\tau\,\frac{d\tau}{2\pi}\int_{-\infty}^{\infty}\frac{\sin t\tau}{\sinh t}\cosh\frac{t}{2}\,dt. \end{aligned}$$

The integral on the right is equal to

$$\int_{-\infty}^{\infty}\frac{\sin t\tau}{e^{t/2}-e^{-t/2}}\,dt = \pi\tanh\pi\tau.$$

This proves the inversion formula for $f(1)$, i.e. Theorem 6.

Theorem 7. *Let*

$$P(s)\,ds = \frac{s}{i}\tanh\left(\frac{\pi s}{i}\right)\frac{ds}{2\pi i}.$$

If g is in the Paley–Wiener space and is even, then

$$\mathbf{S}^{-1}g(x) = \int_{\mathrm{Re}\,s=0} g(s)\varphi(x,s)P(s)\,ds.$$

In other words, $\Phi^* = \Phi^{-1}$ for the Plancherel measure $d\mu(s) = P(s)\,ds$.

Proof. From the inversion formula for $f(1)$, we shall get the general inversion for $f(x)$ by a reduction, using the formalism of spherical functions. For any $f \in C_c^\infty(G//K)$, let f_x be defined by

$$f_x(y) = \int_K f(xky)\,dk.$$

Then f_x is in $C_c^\infty(G//K)$ and $f_x(1) = f(x)$. Applying the special case of inversion in Theorem 6 to f_x, we see that the general inversion formula follows from the next lemma.

Lemma. $\mathbf{S}f_x(s) = \mathbf{S}f(s)\varphi(x,s).$

Proof. We have

$$\begin{aligned} \mathbf{S}f_x(s) &= \int_G f_x(y)\varphi(y, s)\, dy = \int_K \int_G f(xky)\varphi(y, s)\, dy \\ &= \int_G f(xy)\varphi(y, s)\, dy \\ &= \int_G f(y)\varphi(x^{-1}y, s)\, dy = \int_G f(y)\varphi(x^{-1}ky, s)\, dy \end{aligned}$$

for any $k \in K$. Averaging over K, and using the functional equation, we obtain

$$\mathbf{S}f_x(s) = \int_G f(y)\varphi(x^{-1}, s)\varphi(y, s)\, dy = \mathbf{S}f(s)\varphi(x, s),$$

as was to be shown.

§5. EXPLICIT FORMULAS AND ASYMPTOTIC EXPANSIONS

Write an element $x \in G = SL_2(\mathbf{R})$ as

$$x = \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} e & 0 \\ 0 & e^{-1} \end{pmatrix} \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Then $\alpha(x) = e^2$. Multiplying out the matrices yields

$$c = -\frac{\sin\theta}{e} \qquad \text{and} \qquad d = \frac{\cos\theta}{e}.$$

Hence

$$\alpha\left(\begin{pmatrix} * & * \\ c & d \end{pmatrix}\right) = \frac{1}{c^2 + d^2}.$$

Now if

$$h_a = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \qquad \text{and} \qquad r(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix},$$

we get the explicit formula

$$(1) \qquad \boxed{\alpha(r(\theta)h_a) = \frac{1}{a^2 \sin^2\theta + a^{-2}\cos^2\theta}.}$$

Therefore, if we put $\alpha = \alpha(a) = a^2$, we find the advanced calculus integral expression for the spherical function,

$$\boxed{\varphi_s(a) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{1}{(\alpha \sin^2\theta + \alpha^{-1}\cos^2\theta)^{(s+1)/2}}\, d\theta.} \tag{2}$$

By symmetry, we can evaluate the integral from $-\pi/2$ to $\pi/2$, and multiply by 2. Change variables; let

$$u = \tan\theta, \qquad \theta = \arctan u, \qquad d\theta = \frac{du}{1+u^2}.$$

Then for $\alpha = \alpha(a)$ we have

$$\boxed{\varphi_{2s-1}(a) = \frac{1}{\pi}\int_{-\infty}^{\infty} \frac{\alpha^s(1+u^2)^{s-1}}{(1+\alpha^2u^2)^s}\, du.} \tag{3}$$

Remark. In general, as in Harish's papers, the variable u comes from

$$\bar{n}(u) = \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix}$$

and (3) is an integral expression for φ_s taken as an integral over $\bar{N}$.

Changing variables again, with $v = \alpha u$, we find

$$\boxed{\varphi_{2s-1}(a) = \frac{1}{\pi}\alpha^{s-1}\int_{-\infty}^{\infty} \frac{(1+v^2/\alpha^2)^{s-1}}{(1+v^2)^s}\, dv.} \tag{4}$$

In absolute value, the integral is bounded by

$$\int_{-\infty}^{\infty} \frac{(1+v^2/\alpha^2)^{\sigma-1}}{(1+v^2)^{\sigma}}\, dv.$$

Let us now suppose that $\sigma > 1$. This last real integral is decreasing as α increases, $\alpha \geqslant 1$. For $\alpha = 1$, the integral has the value

$$\int_{-\infty}^{\infty} \frac{1}{1+v^2}\, dv.$$

Hence we can apply the dominated convergence theorem, to get

$$\lim_{\alpha\to\infty}\int_{-\infty}^{\infty}\frac{(1+v^2/\alpha^2)^{s-1}}{(1+v^2)^s}\,dv=\int_{-\infty}^{\infty}\frac{1}{(1+v^2)^s}\,dv.$$

We recall the advanced calculus identity

$$\int_{-\infty}^{\infty}\frac{1}{(1+v^2)^s}\,dv=\sqrt{\pi}\,\frac{\Gamma(s-\frac{1}{2})}{\Gamma(s)}\,. \tag{5}$$

This identity is proved by considering the product

$$\Gamma(s)\int_{-\infty}^{\infty}\frac{1}{(1+v^2)^s}\,dv=\int_{-\infty}^{\infty}\int_0^{\infty}e^{-t}\frac{1}{(1+v^2)^s}t^s\,\frac{dt}{t}\,dv.$$

Let $t\mapsto(1+v^2)t$ and use the invariance of the integral with respect to dt/t relative to multiplicative translations. The desired expression drops out.

Our expression for the limit is $\neq 0$. Consequently, we obtain the asymptotic expansion for $\sigma>1$ and $\alpha(a)$ or $\rho(a)\to\infty$:

$$\boxed{\varphi_s(a)\sim\frac{1}{\sqrt{\pi}}\rho(a)^{s-1}\frac{\Gamma\left(\frac{s}{2}\right)}{\Gamma\left(\frac{s+1}{2}\right)}}\,. \tag{6}$$

From it, we see:

Theorem 8. *If* $Re(s)$ *does not lie in the interval* $[-1,1]$, *then the spherical function* φ_s *is not bounded.*

The proof we have given is that in Helgason–Johnson [He, Jo], for arbitrary semisimple Lie groups, but collapsing to advanced calculus in the case of $SL_2(\mathbf{R})$.

In the theory of second order linear differential equations, when finding an eigenfunction expansion for the solutions, certain asymptotic estimates play an important role, realized by Bargmann explicitly in his original paper [Ba]. Although we shall not discuss this aspect of the question here, for the convenience of the reader, we give in Theorem 9 the first two terms for the asymptotic expansion of the spherical functions (4), carried out in general by Harish-Chandra [H-C 7], Lemma 37, actually a central result. I am indebted to Eli Stein for the elementary exposition in the rest of this section.

Let $c(s)$ be the analytic continuation of the function

$$\int_0^\infty \frac{1}{(1+v^2)^{(s+1)/2}}\,dv, \qquad \operatorname{Re} s > 0.$$

Formula (5) expresses this integral in terms of the gamma function.

Theorem 9. *We have the asymptotic behavior for $\epsilon \longrightarrow 0$:*

$$\int_0^\infty \frac{1}{(1+v^2)^{\frac{1}{2}(1+it)}(1+\epsilon^2 v^2)^{\frac{1}{2}(1-it)}}\,dv = c(it) + c(-it)\epsilon^{it} + O(\epsilon^{1/2}).$$

Proof. We need some lemmas.

Lemma 1. *There exists a function c_t of t alone such that for $N \longrightarrow \infty$ we have*

$$\int_0^N \frac{dv}{(1+v^2)^{\frac{1}{2}(1+it)}} = -\frac{N^{-it}}{it} + c_t + O(N^{-1}).$$

Proof. Write

$$\int_0^N = \int_0^a + \int_a^N$$

where $a > 0$ is a fixed constant. We have first

$$\int_a^N \frac{1}{(1+v^2)^{\frac{1}{2}(1+it)}}\,dv = \int_a^N \left[\frac{1}{(1+v^2)^{\frac{1}{2}(1+it)}} - \frac{1}{v^{1+it}}\right] dv + \int_a^N \frac{1}{v^{1+it}}\,dv.$$

For the first integral on the right we write

$$\int_a^N = \int_a^\infty - \int_N^\infty,$$

and by the mean value theorem,

$$\int_N^\infty \left[\frac{1}{(1+v^2)^{\frac{1}{2}(1+it)}} - \frac{1}{v^{1+it}}\right] dv = O(N^{-1}).$$

On the other hand,

$$\int_a^N \frac{1}{v^{1+it}}\, dv = -\frac{N^{-it}}{it} + \frac{a^{-it}}{it}.$$

This proves the lemma, with

$$c_t = \int_0^a \frac{1}{(1+v^2)^{\frac{1}{2}(1+it)}}\, dv + \int_a^\infty \left[\frac{1}{(1+v^2)^{\frac{1}{2}(1+it)}} - \frac{1}{v^{1+it}}\right] dv + \frac{a^{-it}}{it}.$$

Lemma 2. *The value c_t is equal to $c(it)$.*

Proof. We consider

$$\int_0^a \frac{1}{(1+v^2)^{\frac{1}{2}(1+s)}}\, dv + \int_a^\infty \left[\frac{1}{(1+v^2)^{\frac{1}{2}(1+s)}} - \frac{1}{v^{1+s}}\right] dv + \frac{a^{-s}}{s},$$

which is clearly analytic in a neighborhood of $s = it$, in fact for $\operatorname{Re} s > -1$, $s \neq 0$. When $\operatorname{Re} s > 0$ we can continue the integrals and obtain the appropriate value for c_t.

We now prove Theorem 9. Split the integral of Theorem 9 into two integrals,

$$\int_0^\infty = \int_0^{\epsilon^{-1/2}} + \int_{\epsilon^{-1/2}}^\infty = \mathrm{I} + \mathrm{II}.$$

In I, $0 \leqslant v \leqslant \epsilon^{-1/2}$, so that $\epsilon^2 v^2 \leqslant \epsilon \ll 1$. Thus

$$(1 + \epsilon^2 v^2)^{\frac{1}{2}(1+it)} = 1 + O(\epsilon^2 v^2).$$

Hence

$$\begin{aligned} \mathrm{I} &= \int_0^{\epsilon^{-1/2}} \frac{1}{(1+v^2)^{\frac{1}{2}(1+it)}(1+\epsilon^2 v^2)^{\frac{1}{2}(1-it)}}\, dv \\ &= \int_0^{\epsilon^{-1/2}} \frac{1}{(1+v^2)^{\frac{1}{2}(1+it)}}\, dv + O\left(\epsilon^2 \int_0^{\epsilon^{-1/2}} \frac{v^2}{(1+v^2)^{1/2}}\, dv\right) \\ &= -\frac{\epsilon^{it/2}}{it} + c_t + O(\epsilon^{1/2}) \end{aligned}$$

for $\epsilon \longrightarrow 0$, by Lemma 1.

Next,

$$\mathrm{II} = \int_{\epsilon^{-1/2}}^{\infty} \frac{1}{(1+v^2)^{\frac{1}{2}(1+it)}(1+\epsilon^2 v^2)^{\frac{1}{2}(1-it)}}\,dv$$

$$= \int_0^{\epsilon^{1/2}} \frac{1}{(1+v^2)^{\frac{1}{2}(1+it)}(\epsilon^2+v^2)^{\frac{1}{2}(1-it)}}\,dv,$$

making the change of variables $v \mapsto 1/v$. Next let $v \mapsto v/\epsilon$. This yields

$$\mathrm{II} = \epsilon^{it} \int_0^{\epsilon^{-1/2}} \frac{1}{(1+\epsilon^2 u^2)^{\frac{1}{2}(1+it)}(1+u^2)^{\frac{1}{2}(1-it)}}\,du$$

$$= \epsilon^{it}\left[-\frac{\epsilon^{-it/2}}{-it} + c_{-t} + O(\epsilon^{1/2}) \right]$$

by the previous arguments. Hence

$$\mathrm{I} + \mathrm{II} = c_t + \epsilon^{it} c_{-t} + O(\epsilon^{1/2}),$$

as was to be shown.

VI The Derived Representation on the Lie Algebra

In this chapter for the first time we begin to deal with the C^∞ or (real) analytic structure of G, rather than with just measure theory. We shall see how a representation of G gives rise to an *algebraic* representation of the Lie algebra on a dense subspace, for an arbitrary Lie group G. In the case of $SL_2(\mathbf{R})$, this representation has an especially simple form, as shown in §2.

§1. THE DERIVED REPRESENTATION

Let G be a Lie group. For our purposes, you can assume that $G = SL_2(\mathbf{R})$ or $GL_2^+(\mathbf{R})$, or $GL_n^+(\mathbf{R})$. The important thing is that G can be coordinatized in a C^∞ manner. Recall that for $SL_2(\mathbf{R})$ we have our coordinates (x, y, θ) arising from the upper half plane representation. For $GL_2^+(\mathbf{R})$, we would have the four coordinates (u, x, y, θ) where u is a diagonal scalar factor. Of course, for $GL_2(\mathbf{R})$ we can also use the four coordinates of the 2×2 matrix.

If G is a closed subgroup of $GL_n(\mathbf{R})$, one can define its **Lie algebra** as the set of matrices X such that

$$\exp(tX) = \sum_{n=0}^{\infty} \frac{t^n X^n}{n!} \qquad \text{all } t \in \mathbf{R}$$

lies in G. The Lie algebra is denoted by $\mathfrak{g}$. For $G = SL_2(\mathbf{R})$, it is an exercise to verify that $\mathfrak{g}$ consists of all real 2×2 matrices whose trace is 0. (Use the Jordan normal form, for instance.) Thus for $SL_2(\mathbf{R})$, a basis of the Lie algebra over $\mathbf{R}$ is given by the three matrices

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

If X lies in the Lie algebra, then the map

$$t \mapsto \exp(tX)$$

is called a **one-parameter** subgroup of G. It can be proved in general that

$$\exp\colon \mathfrak{g} \longrightarrow G$$

is a local real analytic isomorphism in a neighborhood of 0 (not a group isomorphism, of course). It is trivially verified for $SL_2(\mathbf{R})$. If X, $Y \in \mathfrak{g}$ and X, Y commute, then we do have

$$\exp(X + Y) = \exp(X)\exp(Y).$$

In particular, the parametrization of the one-parameter subgroup above is a group homomorphism.

A function is C^∞ on G if and only if it is C^∞ in terms of coordinates. The notion is clear for $SL_2(\mathbf{R})$ and we won't bother the reader with general definitions of manifolds and Lie groups at this point. Let him use his imagination or look it up elsewhere. For Banach valued maps, and differentiability, cf. *Real Analysis*. There is essentially no difference with real functions.

Let G be a Lie group and let H be a Banach space. Let

$$f\colon G \longrightarrow H$$

be a C^∞ mapping. For each X in the Lie algebra $\mathfrak{g}$, and $y \in G$, we define the **Lie derivative**

$$\mathfrak{L}_X f(y) = \frac{d}{dt} f(y \exp(tX))\Big|_{t=0}.$$

Then $\mathfrak{L}_X f$ is also C^∞, and so we get a linear map

$$\mathfrak{L}_X\colon C^\infty(G, H) \longrightarrow C^\infty(G, H).$$

It is clear that $\mathfrak{L}_X$ is left G-invariant, i.e. commutes with left translations. If $H = \mathbf{C}$ is the scalar field, then $\mathfrak{L}_X$ is easily verified to be a derivation, that is

$$\mathfrak{L}_X(fg) = f\mathfrak{L}_X g + (\mathfrak{L}_X f)g.$$

For fixed $y \in G$ let

$$F_y(X) = f(y \exp(X)).$$

Then F_y is C^∞ on $\mathfrak{g}$, and by definition,

$$\mathfrak{L}_X f(y) = \lim_{t \to 0} \frac{F_y(tX) - F_y(0)}{t}.$$

For any C^∞ map F on a neighborhood of 0 in $\mathfrak{g}$, we have, by Taylor's formula,

$$F(X) = F(0) + F'(0)X + O(|X|^2),$$

where

$$F'(0)\colon \mathfrak{g} \longrightarrow H$$

is the derivative of F at 0 (linear map). Therefore

$$\mathfrak{L}_X f(y) = F_y'(0)X,$$

and $\mathfrak{L}_X$ is linear in X, that is, for $X, Y \in \mathfrak{g}$ and $c \in \mathbf{R}$, we have

$$\mathfrak{L}_{X+Y} = \mathfrak{L}_X + \mathfrak{L}_Y \qquad \text{and} \qquad \mathfrak{L}_{cX} = c\mathfrak{L}_X.$$

We extend this formula by linearity to complex coefficients. We let $\mathfrak{g}_{\mathbf{C}}$ consist of linear combinations of elements of $\mathfrak{g}$ with complex numbers, so that elements of $\mathfrak{g}_{\mathbf{C}}$ are matrices of the form

$$X + iY, \qquad X, Y \in \mathfrak{g}.$$

If $\alpha = a + ib$ with a, b real, then we define

$$\mathfrak{L}_{\alpha X} = \mathfrak{L}_{aX} + i\mathfrak{L}_{bX},$$

and thus obtain an extension of the Lie derivatives to elements of $\mathfrak{g}_{\mathbf{C}}$, which is a **complex Lie algebra**.

When dealing with C^∞ functions and integral representations, we must frequently take a limit or differentiate under an integral sign, and for the convenience of the reader we reproduce the lemmas allowing us to do this.

Lemma 1. *Let X be a measured space with positive measure μ. Let U be an open subset of $\mathbf{R}^n$. Let $f\colon X \times U \longrightarrow E$ be a mapping into a Banach space. Assume*:

i) *For each $y \in U$ the map $x \longmapsto f(x, y)$ is in $\mathfrak{L}^1(\mu, E)$.*
ii) *For each $x \in X$ and $y_0 \in U$, we have*

$$\lim_{y \to y_0} f(x, y) = f(x, y_0).$$

iii) *There exists a function $f_1 \in \mathfrak{L}^1(\mu)$ such that for all $y \in U$,*

$$|f(x, y)| \leqslant |f_1(x)|.$$

Then the function

$$y \longmapsto \int_X f(x, y)\, d\mu(x)$$

is continuous.

Proof. It suffices to prove that for any sequence $\{y_k\}$ converging to y,

$$\int_X f(x, y_k)\, d\mu(x) \qquad \text{converges to} \qquad \int_X f(x, y)\, d\mu(x).$$

Let $f_k(x) = f(x, y_k)$. Then $\{f_k\}$ converges pointwise to the function

$$x \longmapsto f(x, y),$$

and by (iii), we can apply the dominated convergence theorem to conclude the proof.

Lemma 2. *Let X be a measured space with positive measure μ. Let U be an open subset of $\mathbf{R}^n$. Let $f\colon X \times U \longrightarrow E$ be a mapping into a Banach space. Assume:*

i) *For each $y \in U$ the map $x \longmapsto f(x, y)$ is in $\mathfrak{L}^1(\mu, E)$.*
ii) *For each $y \in U$, each partial $D_j f(x, y)$ (taken with respect to the j-th y-variable) is in $\mathfrak{L}^1(\mu, E)$.*
iii) *There exists a function $f_1 \in \mathfrak{L}^1(\mu)$ such that for all $y \in U$,*

$$|D_j f(x, y)| \leqslant |f_1(x)|.$$

Let

$$\Phi(y) = \int_X f(x, y)\, d\mu(x).$$

Then $D_j\Phi$ exists and we have

$$D_j\Phi(y) = \int_X D_j f(x, y)\, d\mu(x).$$

Proof. We have

$$\frac{\Phi(y + he_j) - \Phi(y)}{h} = \int_X \frac{1}{h}\left[f(x, y + he_j) - f(x, y)\right] d\mu(x).$$

Using the mean value theorem and (iii), together with the dominated convergence theorem, we conclude that the right-hand side has a limit, equal to

$$\int_X D_j f(x, y)\, d\mu(x).$$

[As in the previous proof, we have to use the device of taking a sequence $\{h_k\}$ to apply the dominated convergence theorem in its standard form.]

Aside from hypothesis insuring that all the symbols make sense, the essential hypothesis in Lemma 2, allowing us to differentiate under the integral sign, is that the partial derivative with respect to y is uniformly dominated by a function in $\mathfrak{L}^1$, independently of x.

Example. Let $f \in \mathcal{L}^1(\mathbf{R}^n)$ and let $\varphi \in C_c^\infty(\mathbf{R}^n)$. We see that $f * \varphi$ is C^∞, and for any monomial $D^p = D_1^{p_1} \cdots D_n^{p_n}$ of partial differential operators, we have

$$D^p(f * \varphi) = f * D^p\varphi.$$

Indeed, by definition,

$$f * \varphi(y) = \int_{\mathbf{R}^n} f(-x)\varphi(x+y)\,dx.$$

It is clear that Lemma 2 applies, and we can differentiate φ under the integral sign repeatedly to get the above formula.

Lemma 3. Let G be a Lie group and $f \in \mathcal{L}^1(G, H)$ where H is a Banach space. Let $\varphi \in C_c^\infty(G, \mathbf{C})$ and let $X \in \mathfrak{g}_{\mathbf{C}}$. Then $f * \varphi$ is C^∞, and

$$\mathcal{L}_X(f * \varphi) = f * \mathcal{L}_X \varphi.$$

Proof. Exactly the same as above for $\mathbf{R}^n$. By definition,

$$f * \varphi(y) = \int_G f(x^{-1})\varphi(xy)\,dx.$$

In the neighborhood of a point y we can choose local coordinates identifying this neighborhood with an open set in Euclidean space. The uniform domination of Lemma 2 is valid.

Let G be a Lie group and $\pi: G \longrightarrow GL(H)$ a representation in a Banach space. We define the algebraic subspace H_π^∞ to consist of all those vectors $v \in H$ such that the map

$$x \longmapsto \pi(x)v$$

is C^∞. We call H_π^∞ the space of **C^∞ vectors**. It is stable under the action of G, i.e. if $a \in G$ and $v \in H_\pi^\infty$, then

$$\pi(a)v \in H_\pi^\infty,$$

and it is also stable under the action of smooth functions with compact support. More is even true:

If $\varphi \in C_c^\infty(G)$ and $v \in H$, then $\pi^1(\varphi)v \in H_\pi^\infty$.

Proof. By definition,

$$\begin{aligned}
\pi^1(\varphi)v &= \int \varphi(y)\pi(y)v\,dy \\
\pi(x)\pi^1(\varphi)v &= \int \varphi(y)\pi(xy)v\,dy \\
&= \int \varphi(x^{-1}y)\pi(y)v\,dy.
\end{aligned}$$

We can differentiate under the integral sign by Lemma 2, with respect to x. This is essentially a special case of Lemma 3.

Recall that if $\{\theta_n\}$ is a Dirac sequence, then for $v \in H$, $\pi^1(\theta_n)v$ converges to v. Consequently, since we can take a Dirac sequence to consist of C^∞ functions with compact support, we conclude:

$$H_\pi^\infty \textit{ is dense in } H.$$

In other words, there is an ample supply of C^∞ vectors, even of the form $\pi^1(\varphi)v$, where $\varphi \in C_c^\infty(G)$.

We now come to the derived representation. If $v \in H_\pi^\infty$, we define

$$d\pi(X)v = \frac{d}{dt}\pi(\exp(tX)v)\bigg|_{t=0}.$$

We shall prove that the right-hand side lies also in H_π^∞, and furthermore we get the formula

DER 1. *If $f(x) = \pi(x)v$, then for $a \in G$, $v \in H_\pi^\infty$ we have*

$$d\pi(X)v = (\mathfrak{L}_X f)(e) \qquad \pi(a)d\pi(X)v = (\mathfrak{L}_X f)(a).$$

Since f is C^∞ by assumption, we see that the formula implies that $d\pi(X)v$ is also C^∞. As for the truth of the formula, we have by definition

$$(\mathfrak{L}_X f)(a) = \frac{d}{dt}f(a \exp(tX))\bigg|_{t=0}.$$

Applying the continuous operator $\pi(a)$ to the limit

$$\lim_{t \to 0} \frac{\pi(\exp(tX))v - v}{t},$$

we see that we get our formula.

We see that $d\pi(X)\colon H_\pi^\infty \to H_\pi^\infty$ is a linear map, depending linearly on X. We call it the **derived representation** of π on the Lie algebra.

DER 2. *For $X, Y \in \mathfrak{g}$ and $f(x) = \pi(x)v$, $v \in H_\pi^\infty$, we have*

$$d\pi(X)d\pi(Y)v = \mathfrak{L}_X \mathfrak{L}_Y f(e).$$

Proof. We have

$$\pi(x)d\pi(Y)v = \mathfrak{L}_Y f(x)$$

by **DER 1**; also, putting $a = e$, we know that

$$d\pi(X)w = (\mathfrak{L}_X f_w)(e)$$

if $w \in H_\pi^\infty$ and $f_w(x) = \pi(x)w$. We apply this to $w = d\pi(Y)v$ to get **DER 2.**

Let $X, Y \in \mathfrak{g}$. There exists a unique element $[X,Y] \in \mathfrak{g}$ such that for all $f \in C^\infty(G,H)$ we have

$$\boxed{\mathfrak{L}_{[X,Y]}f = \mathfrak{L}_X \mathfrak{L}_Y f - \mathfrak{L}_Y \mathfrak{L}_X f.}$$

The bracket is defined by using multiplication of matrices, namely

$$[X, Y] = XY - YX.$$

We shall prove this below. From **DER 2** we then get for $d\pi$,

DER 3. $$d\pi([X, Y]) = [d\pi(X), d\pi(Y)],$$

where the bracket on the right is $d\pi(X)d\pi(Y) - d\pi(Y)d\pi(X)$.

The proof for the commutation formula of the Lie derivative will use two lemmas. Note that if g is an invertible matrix, then

$$g^{-1}(\exp X)g = \exp(g^{-1}Xg).$$

This is trivial, because $g^{-1}X^n g = (g^{-1}Xg)^n$.

Lemma 4. *Let s be a real number, and put $g_s = \exp(sY)$. Let*

$$\varphi(X) = f(\exp X).$$

Then

$$\frac{d}{ds}\varphi(g_s^{-1}Xg_s)\Big|_{s=0} = \varphi'(X)(XY - YX).$$

Proof. We have

$$g_s^{-1}Xg_s = (I - sY + O(s^2))X(I + sY + O(s^2))$$

$$= X + s(XY - YX) + O(s^2).$$

Hence

$$\varphi(g_s^{-1}Xg_s) = \varphi(X) + \varphi'(X)s(XY - YX) + O(s^2).$$

The assertion of the lemma follows at once.

***Lemma* 5.** *Let $y = \exp(uY)$, where u is a fixed real number. Let*

$$J(u,s,t) = f\big(y \exp(tg_s^{-1}Xg_s)\big).$$

Then

$$D_3D_2J(u, 0, 0) = \varphi'(O)(XY - YX).$$

Proof. The expression on the left is equal to

$$\frac{d}{dt}D_2J(u, 0, t)\bigg|_{t=0} = \frac{d}{dt}[\varphi'(tX)t(XY - YX)]\bigg|_{t=0}.$$

The lemma follows from the rule for the derivative of a product.

Note that in Lemma 4, we could consider $\varphi_y(X) = f(y \exp X)$. We wrote the y explicitly in Lemma 5 in order to have the numbering of the variables fit the application we have in mind.

Next, let $g_s = \exp(sY)$, as before, and let

$$F(u, s, t) = f\big(\exp(uY) \exp(tg_s^{-1}Xg_s)\big).$$

Then

$$f(\exp(tX) \exp(uY)) = f\big(\exp(uY) \exp(tg_u^{-1}Xg_u)\big) = F(u, u, t),$$

so that

$$\mathfrak{L}_X \mathfrak{L}_Y f(e) = D_3D_1F(0, 0, 0) + D_3D_2F(0, 0, 0).$$

Furthermore,

$$f(\exp(uY) \exp(tX)) = F(u, 0, t),$$

so that

$$\mathfrak{L}_Y \mathfrak{L}_X f(e) = D_3D_1F(0, 0, 0).$$

Therefore

$$\mathfrak{L}_X \mathfrak{L}_Y f(e) - \mathfrak{L}_Y \mathfrak{L}_X f(e) = D_3D_2F(0, 0, 0).$$

By Lemma 5, we find

$$D_3D_2F(u, 0, 0) = \varphi_y'(O)(XY - YX).$$

We obtain

$$\begin{aligned} D_2D_2F(0, 0, 0) &= \varphi'(O)(XY - YX) \\ &= \mathfrak{L}_{XY-YX}f(e). \end{aligned}$$

This proves the desired relation of commutation between the Lie derivatives.

We also want to know how the derived representation behaves when composed with $\pi^1(\varphi)$ for $\varphi \in C_c^\infty(G)$. We introduce the **right Lie derivative**

$$\mathfrak{R}_X f(x) = \frac{d}{dt} f(\exp(tX)x)\Big|_{t=0},$$

which commutes with right translations.

In what follows, we let $\varphi \in C_c^\infty(G)$. For the next property **DER 4**, we need not assume $v \in H_\pi^\infty$, merely that $v \in H$.

DER 4. $\quad d\pi(X)\pi^1(\varphi) = \pi^1(\mathfrak{R}_{-X}\varphi)$ *on* H.

Proof. We have:

$$\begin{aligned}\pi(\exp(tX))\pi^1(\varphi)v &= \pi(\exp(tX))\int \varphi(x)\pi(x)v\,dx \\ &= \int \varphi(x)\pi(\exp(tX)x)v\,dx \\ &= \int \varphi(\exp(-tX)x)\pi(x)v\,dx\end{aligned}$$

by letting $x \mapsto \exp(-tX)x$. We consider this expression for small values of t near 0, and we differentiate under the integral sign, valid because φ has compact support. Let

$$\varphi(\exp(-tX)x) = F(t, x).$$

Then $D_1F(t, x)$ has compact support. Applying d/dt to our expression yields

$$\int \frac{d}{dt}\varphi(\exp(-tX)x)\pi(x)v\,dx.$$

At $t = 0$ this yields precisely $\pi^1(\mathfrak{R}_{-X}\varphi)v$, thus proving the formula **DER 4.**

DER 5. $\quad \pi^1(\varphi)d\pi(X) = \pi^1(\mathfrak{L}_{-X}\varphi)$

Proof. Entirely similar to the preceding proof, moving d/dt inside and out of an integral and using the definitions. We don't bother to write it out.

Partly because of the minus signs occurring in **DER 4** and **DER 5**, and partly for other suggestive reasons, it is sometimes useful to introduce the notation

$$(X * \varphi)(y) = \frac{d}{dt}\varphi(\exp(-tX)y)\Big|_{t=0} = -\mathfrak{R}_X\varphi(y),$$

$$(\varphi * X)(y) = \frac{d}{dt}\varphi(y\exp(-tX))\Big|_{t=0} = -\mathfrak{L}_X\varphi(y).$$

It is easily verified that

$$(\varphi * X) * \psi = \varphi * (X * \psi)$$

for $\varphi, \psi \in C_c^\infty(G)$. Then we have

$$\pi^1(\varphi * X) = \pi^1(\varphi)\, d\pi(X) \qquad \text{and} \qquad \pi^1(X * \varphi) = d\pi(X)\pi^1(\varphi).$$

By an abuse of notation, one sometimes writes $\pi(X)$ and $\pi(\varphi)$ instead of $d\pi(X)$ and $\pi^1(\varphi)$ (a notation which confused me a lot when I learned the subject, but which does become convenient), in which case one sees π as a multiplicative homomorphism for the convolution operation on the algebra generated by elements of the Lie algebra and functions in $C_c^\infty(G)$, acting as operators. In other words, we have in this notation

$$\pi(\varphi * X) = \pi(\varphi)\pi(X) \qquad \text{and} \qquad \pi(X * \varphi) = \pi(X)\pi(\varphi).$$

We are now finished with the formalism of C^∞ vectors, and say a few additional words about analyticity. Let H be a Banach space. If $f\colon G \longrightarrow H$ is a mapping, we say that f is **analytic** (i.e. real analytic) if it has a power series expansion in the neighborhood of every point, in terms of local coordinates.

Taylor's formula. *Let $f\colon G \longrightarrow H$ be analytic. Let $y \in G$. For all X sufficiently small in $\mathfrak{g}$ (with respect to any norm on the finite dimensional $\mathbf{R}$-space $\mathfrak{g}$), and all t with $0 \leqslant t \leqslant 1$, we have the Taylor series expansion*

$$f(y \exp(tX)) = \sum_{n=0}^{\infty} \frac{1}{n!}(\mathfrak{L}_X^n f)(y)t^n.$$

Proof. By definition,

$$\mathfrak{L}_X f(y \exp(uX)) = \frac{d}{dt} f(y \exp(u + t)X)\Big|_{t=0} = \frac{d}{du} f(y \exp(uX)).$$

By induction, let $F = \mathfrak{L}_X^n f$ and assume

$$F(y \exp(uX)) = \left(\frac{d}{du}\right)^n f(y \exp(uX)).$$

Then

$$\mathfrak{L}_X F(y \exp(uX)) = \frac{d}{du} F(y \exp uX) = \left(\frac{d}{du}\right)^{n+1} f(y \exp uX).$$

In other words, we have proved the formula

$$(*) \qquad \mathfrak{L}_X^n f(y \exp uX) = \frac{d^n}{du^n} f(y \exp uX).$$

Let $(x_1, \ldots, x_r)$ be the coordinates of X with respect to a fixed basis of $\mathfrak{g}$ over $\mathbf{R}$. Since f is analytic, we have for small $|X|$,

$$f(y \exp X) = P(x_1, \ldots, x_r)$$

where P is a convergent power series in a neighborhood of 0. Hence

$$f(y \exp uX) = \sum_{n=0}^{\infty} a_n \frac{1}{n!} u^n, \qquad a_n \in \mathbf{R},$$

for $0 < u < 1$. Hence, evaluating $(*)$ at $u = 0$, we get

$$\mathfrak{L}_X^n f(y) = a_n.$$

Putting $u = t$ proves our Taylor formula.

Let $\pi\colon G \to GL(H)$ be a representation in a Banach space. An element $v \in H$ is called **analytic** if the map $x \mapsto \pi(x)v$ is analytic.

Corollary. *If v is analytic, then*

$$\pi(\exp(X))v = \sum_{n=0}^{\infty} \frac{1}{n!} d\pi(X)^n v.$$

Proof. Immediate from Taylor's formula.

It is clear that the analytic vectors in H form an algebraic subspace, denoted by H_π^{an}.

Theorem 1. *Let G be a connected Lie group and let*

$$\pi\colon G \to GL(H)$$

be a representation on a Banach space. Let V be an algebraic subspace of H consisting of analytic vectors, and invariant under $d\pi(\mathfrak{g})$. Then the closure of V is invariant under G.

Proof. Let $v \in V$. For $X \in \mathfrak{g}$ and $|X|$ small, we apply the corollary to the analytic map $f(x) = \pi(x)v$, and get

$$\pi(\exp(X))v = \sum \frac{1}{n!} d\pi(X)^n v \in \bar{V}.$$

But exp of a small neighborhood of 0 in $\mathfrak{g}$ is a neighborhood of e in G. Hence the closure of V is invariant under a neighborhood of e in G. The products of such a neighborhood with itself generate G, so our theorem is proved.

Theorem 1 provides an important criterion for irreducibility. As we shall see later, the algebraic space of K-finite vectors in a certain representation, or appropriate algebraic subspaces, will provide examples of the space V in Theorem 1.

In concretely given situations, as we shall see below, one has direct means to prove that certain vectors are analytic. This arises from the fact that certain representations are in function spaces on G, and that these function spaces contain plenty of analytic functions on G. One has then to prove that such functions are analytic vectors, when viewed as elements of the function spaces.

The existence of a dense subspace of analytic vectors for completely arbitrary Lie groups was proved by Nelson [Ne].

On the other hand, a general proof for $SL_2(\mathbf{R})$, following the original ideas of Harish–Chandra for general semisimple Lie groups, falls easily within our range of ideas. One can construct a Dirac sequence consisting of analytic functions, using the weak definition of Dirac sequences, where **DIR 3** is replaced by the analogous L^1 condition. Indeed, analytically, G is merely a product $A \times N \times K$, where A is isomorphic to $\mathbf{R}^+$, N is isomorphic to $\mathbf{R}$, and K is the circle group. Fourier series provide analytic Dirac sequences on K. Starting with the function

$$\varphi(t) = \frac{1}{\sqrt{\pi}} e^{-t^2}$$

one forms the sequence of functions

$$\varphi_n(t) = n\varphi(nt)$$

whose integral is still 1, which are positive, and which are easily proved to satisfy the third Dirac sequence condition. Taking $\pi^1(\varphi_n)v$ for an arbitrary vector v is easily shown to yield an analytic vector. There is therefore no great difficulty in handling analytic vectors in the cases of interest to us, just like C^∞ vectors. An even better way will be to see later that in all cases of interest to us, **K-finite vectors are analytic**. See X, §2, Th. 7.

§2. THE DERIVED REPRESENTATION DECOMPOSED OVER K

We return to the considerations of Chapter II, §1, where we obtained a decomposition of an arbitrary representation of $G = SL_2(\mathbf{R})$ over its circle group K. We shall relate it to the derived representation. As usual, we let

$$r(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

For the convenience of the reader, we recall that $f \in S_{n,m}$ if and only if $f \in C_c(G)$ and satisfies

$$f(r(\theta)yr(\theta')) = e^{-in\theta}f(y)e^{-im\theta'}$$

for all $y \in G$ and all real θ, θ'. We note that the space of C^∞ functions in $S_{n,m}$, denoted by $S^\infty_{n,m}$, is dense in $S_{n,m}$. This is essentially obvious: Any continuous function with compact support can be uniformly approximated by a C^∞ function whose support is very close, and then the arguments given to prove Lemma 1, Chapter II, §1, amounting to Fourier series convergence, yield C^∞ functions. From now on, when we quote Lemma 1 from Chapter II, we allow ourselves to do it with $S^\infty_{n,m}$ replacing $S_{n,m}$.

We use the same notation as in Chapter II, §1, for the K-decomposition. Let $\pi: G \longrightarrow GL(H)$ be a representation in a Banach space. Then H_n is the subspace of H consisting of those vectors v such that

$$\pi(r(\theta))v = e^{in\theta}v,$$

i.e. it is the n-th eigenspace of K in H.

The denseness of $S^\infty_{q,q}$ in $S_{q,q}$ shows that if $\pi^1(S_{q,q})H_q \neq 0$, then

$$\pi^1(S^\infty_{q,q})H_q \neq 0.$$

As before, we let

$$W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

so that the one-parameter subgroup of G generated by W is precisely K.

Lemma 1. *Let $\pi: G \longrightarrow GL(H)$ be a representation. Let $X \in \mathfrak{g}$. Let $v \in H^\infty_\pi$ be an eigenvector for $d\pi(X)$ with eigenvalue λ, i.e. assume that*

$$d\pi(X)v = \lambda v.$$

Then

$$\pi(\exp(tX))d = e^{\lambda t}v$$

for all real t.

Proof. Let $f(t) = \pi(\exp(tX))d$. Then f is differentiable by assumption, and

$$\frac{f(t+h) - f(t)}{h} = \frac{\pi(\exp(t+h)X)v - \pi(\exp(tX))v}{h}$$

$$= \pi(\exp(tX))\frac{\pi(\exp hX)v - v}{h}.$$

Taking the limit as $h \longrightarrow 0$ yields $\pi(\exp tX)\, d\pi(X)v = \pi(\exp tX)\lambda v$. Hence $f'(t) = \lambda f(t)$. Considering $f(t)/e^{\lambda t}$ yields $f(t) = e^{\lambda t}v$, as desired.

The lemma will be applied when $X = W$, $t = \theta$, $\exp \theta W = r(\theta)$.

Theorem 2. *Let $\pi\colon G \longrightarrow GL(H)$ be an irreducible admissible representation of $G = SL_2(\mathbf{R})$ in a Banach space. Let n be an integer such that $H_n \neq \{0\}$. Then H_n has dimension 1, any element of H_n is a C^∞ vector, and H_n is the eigenspace of $d\pi(W)$ with eigenvalue in.*

Proof. We know from II, §1, Th. 2 that H_n has dimension 1. Since $\pi^1(S^\infty_{n,n})H_n \neq 0$, if $\{v_n\}$ is a basis of H_n over $\mathbf{C}$, then

$$\pi^1(S^\infty_{n,n})v_n = \mathbf{C}v_n.$$

Thus v_n is a C^∞ vector. From the definition of H_n it is clear that its character on K is the n-th character, and that it is an eigenspace for $d\pi(W)$ with eigenvalue in. The lemma shows that any such eigenvector must lie in H_n, and our theorem is proved.

Remark. As mentioned before, the analyticity of elements in H_n will be proved in X, §2, Th. 6.

We shall now investigate the action of other basis elements of the Lie algebra on the spaces H_n. It is convenient to deal with complex elements. We let

$$E^- = \begin{pmatrix} 1 & -i \\ -i & -1 \end{pmatrix}, \qquad E^+ = \overline{E^-} = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$$

Then by matrix multiplication, we get

$$[E^+, E^-] = -4iW, \qquad [W, E^+] = 2iE^+, \qquad [W, E^-] = -2iE^-.$$

Bracketing with some X in the Lie algebra yields a linear endomorphism called $\mathrm{ad}(X)$ of the Lie algebra, and we see that E^+, E^- are eigenvectors of $\mathrm{ad}(W)$. [Note that $\mathrm{ad}(X)$ is the **regular representation** of X in its own algebra. For some unknown horrible reason, it is called the **adjoint representation**, which is very confusing because there is no adjoint operator in the sense of scalar products anywhere in sight. It would have been better to call it $\mathrm{reg}(X)$, but it's too late to change.]

Occasionally we shall also write E_+ and E_- instead of E^+ and E^-, especially when these symbols occur in the context of powers.

Theorem 3. *Let $\pi\colon G \longrightarrow GL(H)$ be an irreducible representation of G in a Banach space. Let m be an integer. Then the sum*

$$\sum_{n \equiv m \ (\mathrm{mod}\ 2)} H_n$$

is stable under $d\pi(\mathfrak{g})$. Furthermore, we have

$$d\pi(E^+)\colon H_n \longrightarrow H_{n+2},$$

$$d\pi(E^-)\colon H_n \longrightarrow H_{n-2}.$$

Proof. Let $v \in H_n$. Then **DER 3** shows

$$\begin{aligned} d\pi(W)\, d\pi(E^+)v &= d\pi[W, E^+]v + d\pi(E^+)\, d\pi(W)v \\ &= 2i\, d\pi(E^+)v + in\, d\pi(E^+)v \\ &= i(n+2)\, d\pi(E^+)v. \end{aligned}$$

Hence $d\pi(E^+)$ is an eigenvector for $d\pi(W)$ with eigenvalue $i(n+2)$. Theorem 2, or the lemma, shows that $d\pi(E^+)v$ lies in H_{n+2}. The argument for E^- is the same, replacing $+$ by $-$, thereby proving our theorem.

The above arguments, due to Bargmann [Ba], determine the action of the Lie algebra on irreducible representations of $SL_2(\mathbf{R})$.

Suppose that $H_m = 0$ for some integer m. We form the sums

$$H_m^+ = \sum_{q>m} H_q \qquad \text{and} \qquad H_m^- = \sum_{q<m} H_q,$$

over all integers $q \equiv m \pmod 2$. Then H_m^+ and H_m^- are stable algebraic subspaces for the operation of the Lie algebra, $d\pi(\mathfrak{g}_{\mathbf{C}})$. If we know that the vectors in H_n are analytic vectors, then we can apply Theorem 1 to conclude that if H_m^+, say, is $\neq 0$, then its closure is invariant under $\pi(G)$, whence the irreducibility of π implies that this closure is exactly H.

Bargmann [Ba] following up the above arguments, also showed the uniqueness of the irreducible unitary representations behaving like the above with respect to the Lie algebra, extending to infinity in each direction, so that there are no others besides the ones which we shall exhibit. We shall prove these uniqueness statements below, and in the next section.

From the action of the Lie algebra, we see that there are four possible classes of irreducible infinite dimensional representations of G, according to the behavior with respect to $d\pi(\mathfrak{g})$ extending to infinity in at least one direction:

Case 1. There exists an integer m such that H is the closure of the space

$$\sum_{\substack{q>m \\ q\equiv m \pmod 2}} H_q$$

and $H_m \neq 0$.

Case 2. There exists an integer m such that H is the closure of the space

$$\sum_{\substack{q < m \\ q \equiv m \ (\mathrm{mod}\ 2)}} H_q,$$

and $H_m \neq 0$.

Case 3. The space H_0 is not 0 and H is the closure of the sum

$$\sum_{q \text{ even}} H_q.$$

Case 4. H is the closure of the sum

$$\sum_{q \text{ odd}} H_q.$$

In Case 1, we call m the **lowest weight**, and call a non-zero element of H_m a **lowest weight vector**. In Case 2, we call m the **highest weight**, and call a non-zero element of H_m a **highest weight vector**. In Case 3, we note that H_0 consists of the elements of H which are invariant under K. It is therefore no surprise that the theory of Case 3 will be a continuation of the theory of spherical functions. In the other three cases, bi-invariant functions operate trivially on H, i.e. the trivial representation of K does not occur in these three cases.

Visually, the direct sum of the non-zero spaces H_n in the four cases looks like this:

$$H_m \oplus H_{m+2} \oplus H_{m+4} \oplus \cdots,$$

$$\cdots \oplus H_{m-4} \oplus H_{m-2} \oplus H_m,$$

$$\cdots \oplus H_{-4} \oplus H_{-2} \oplus H_0 \oplus H_2 \oplus H_4 \oplus \cdots,$$

$$\cdots \oplus H_{-3} \oplus H_{-1} \oplus H_1 \oplus H_3 \oplus H_5 \oplus \cdots.$$

There are also finite dimensional representations (occurring in a natural way between H_{-m} and H_m!). We shall deal with these later in connection with the Plancherel formula.

The fact that the K-components of an irreducible subspace of a representation must have a given parity makes it useful to decompose functions in the group algebra in a similar fashion. For any function f on G we let

$$f^+(x) = \frac{f(x) + f(-x)}{2} \qquad \text{and} \qquad f^-(x) = \frac{f(x) - f(-x)}{2}.$$

Thus f^+ is **even** and f^- is **odd**. (The context will always make it clear whether f^- is the odd function defined above, or $f^-(x) = f(x^{-1})$.)

Let π be a representation of G on a Banach space H. We call the space

$$H^+ = \widehat{\bigoplus_{n \text{ even}}} H_n$$

the space of **even** elements in H, and we call the space

$$H^- = \widehat{\bigoplus_{n \text{ odd}}} H_n$$

the space of **odd** elements in H.

Lemma 2. *If $\psi \in C_c^\infty(G)$ is odd (resp. even), then $\pi^1(\psi)$ maps H into H^- (resp. H^+) and annihilates H^+ (resp. H^-).*

Proof. This is immediate from the definitions of $\pi^1(\psi)$.

Writing a function $\psi = \psi^+ + \psi^-$, we can then study the effect of $\pi(\psi)$ on each one of the spaces H^+ and H^- separately.

We end this section by considering the uniqueness of the derived representation in certain cases.

Let π be a representation of G on a Banach space H. Let

$$H(K) = \sum H_n$$

be the algebraic space of K-finite vectors, where H_n is the n-th eigenspace of K. We called the representation π **admissible** if the dimensions of the spaces H_n are finite. If this is the case, then an argument similar to that used to prove Theorem 2, §2, shows that the elements of H_n are C^∞ vectors. We leave this as an exercise to the reader. [It is also true that they are analytic vectors.] The proof will be given for the case dim $H_n = 1$ in Chapter X, §2.

Let $y \in G$ and $X \in \mathfrak{g}$. We define

$$\mathrm{Ad}(y)X = yXy^{-1},$$

so that $\mathrm{Ad}(y)$ is a linear endomorphism of $\mathfrak{g}$. We call $y \longmapsto \mathrm{Ad}(y)$ the **adjoint representation** of G. Recall that

$$\exp(yXy^{-1}) = y\exp(X)y^{-1}.$$

If $y \in G$, then we get trivially $\pi(y)d\pi(X) = d\pi(\mathrm{Ad}(y)X)\pi(y)$.

The space $H(K)$ is stable under $d\pi(\mathfrak{g})$.

Proof. Let $k \in K$, $v \in H(K)$ and $X \in \mathfrak{g}$. Then

$$\pi(k)\, d\pi(X)v = d\pi(\mathrm{Ad}(k)X)\pi(k)v.$$

Hence, $\pi(K)\, d\pi(X)v$ is contained in the image under $d\pi(\mathfrak{g})$ of the finite dimensional space generated by $\pi(K)v$, and is therefore finite dimensional, as contended.

Let π_1, π_2 be admissible representations on Banach spaces H_1, H_2. We shall say that they are **infinitesimally isomorphic** if there exists a linear isomorphism

$$L\colon H_1(K) \longrightarrow H_2(K)$$

which commutes with the derived representation, i.e. such that

$$L \circ d\pi_1(X) = d\pi_2(X) \circ L \qquad \text{on} \quad H_1(K).$$

We also say in this case that L is a $\mathfrak{g}$-**isomorphism** on $H_1(K)$, or is an **infinitesimal isomorphism.**

Let $S^\infty = \sum S^\infty_{n,m}$ as before. We say that π_1 is S^∞**-isomorphic to** π_2 if there exists a linear isomorphism L such that for $\varphi \in S^\infty$,

$$L \circ \pi_1(\varphi) = \pi_2(\varphi) \circ L \qquad \text{on} \quad H_1(K).$$

If L is an S^∞-isomorphism, then it is a $\mathfrak{g}$-isomorphism.

Proof. Let $v \in H_1(K)$. There exists $\varphi \in S^\infty$ and $w \in H_1(K)$ such that $v = \pi_1(\varphi)w$. For $X \in \mathfrak{g}$ we get:

$$\begin{aligned} L\pi_1(X)v = L\pi_1(X)\pi_1(\varphi)w &= L\pi_1(X * \varphi)w \\ &= \pi_2(X * \varphi)Lw \\ &= \pi_2(X)\pi_2(\varphi)Lw \\ &= \pi_2(X)L\pi_1(\varphi)w \\ &= \pi_2(X)Lv, \end{aligned}$$

as was to be shown.

Observe that since a $\mathfrak{g}$-isomorphism on $H(K)$ commutes with $d\pi(W)$, it necessarily commutes with the action of K, i.e. preserves the eigenspace direct sum decomposition of the K-finite vectors.

Theorem 4. *Let π, π' be irreducible admissible representations of G on Banach spaces H, H'. Assume that there exists a postive integer m such that π, π' have a lowest weight vector of weight m, say u_m, u'_m respectively. If H, H' are infinite dimensional or if they have the same finite dimension, then* *there exists an infinitesimal isomorphism*

$$L: H(K) \to H'(K)$$

such that $Lu_m = u'_m$.

Proof. For simplicity write Xv instead of $d\pi(X)v$, and similarly for Xv', $v' \in H'$. Let

$$E_+^r u_m = u_{m+2r}, \qquad E_+^r u'_m = u'_{m+2r}.$$

By irreducibility, u_{m+2r} and u'_{m+2r} are $\neq 0$ and are basis elements of H_{m+2r}, H'_{m+2r} respectively for the same values of r. We define L such that

$$L(u_{m+2r}) = u'_{m+2r}.$$

Then L commutes with W and E_+. There remains to prove that L commutes with E_-. This is obvious inductively, using the commutation rule

$$E_+ E_- = E_- E_+ - 4iW,$$

and starting the induction with the assumption that $E_- u_m = 0$, $E_- u'_m = 0$. Thus our theorem is proved.

Theorem 4 has an analogue for irreducible representations having highest weight vectors of weight $-m$, where m is still a positive integer, and the proof is the same, mutatis mutandis. If $m \geqslant 2$, such infinite dimensional representations are called **discrete series** representations, because they are infinitesimally isomorphic with irreducible subspaces of $L^2(G)$ on which G acts by translation. We shall prove this later, by exhibiting an appropriate subspace of $L^2(G)$ having a lowest weight vector. As for the finite dimensional case, we get:

Corollary. *For each positive integer d there exists one, and up to isomorphism only one, irreducible representation of $SL_2(\mathbf{R})$ of dimension d.*

Proof. Existence will be proved at the end of §5. Infinitesimal uniqueness is a special case of the theorem, so let

$$L: H \to H'$$

be an infinitesimal isomorphism. In the finite dimensional case, we automatically have Taylor's formula

$$\pi(\exp X)v = \sum \frac{1}{n!} d\pi(X)^n v,$$

whence L is necessarily also an isomorphism with respect to the group action, as desired.

§3. UNITARIZATION OF A REPRESENTATION

Let π be a representation of G on a Banach space E. We say that π is **unitarizable** if π is infinitesimally isomorphic to a unitary representation on a Hilbert space H. An infinitesimal isomorphism

$$L\colon E(K) \longrightarrow H(K)$$

induces a Hilbert space scalar product on the algebraic space $E(K)$. If $E(K)_n$ denotes the n-th eigenspace of $E(K)$ (equal to the n-th eigenspace of E), then any such Hilbert space product must make distinct eigenspaces orthogonal to each other. In particular, if $\dim E(K)_n = \dim E_n = 1$, then the scalar product on E_n is determined by the value $\langle u, u \rangle$ for any basis vector u in E_n. If u is a unit vector for one scalar product, say

$$\langle u, u \rangle_1 = 1,$$

and if $\langle v, w \rangle_2$ is another scalar product, such that $\langle u, u \rangle_2 = c^2$ with $c > 0$, then $c^{-1}u$ is a unit vector for the second scalar product. If $\{u_n\}$ is a family of unit vectors for the spaces E_n such that $\dim E_n = 1$, in the first scalar product, then $\{c_n^{-1}u_n\}$ is a family of unit vectors for these spaces E_n in the second scalar product. We therefore see that a unitarization of π determines such a family of positive numbers $\{c_n\}$.

Infinitesimally, there is a simple necessary condition for a representation to be unitary.

Lemma 1. *If π is a unitary representation of G, and $X \in \mathfrak{g}$, then $d\pi(X)$ is skew symmetric on H_π^∞.*

Proof. Obvious from the definition of the derived representation.

The next lemmas lead to a theorem giving us the uniqueness of a unitary representation infinitesimally isomorphic to a given representation.

Lemma 2. *If π is irreducible admissible on the Banach space H, and if $\mathfrak{A}$ is the algebra of operators on $H(K)$ generated over $\mathbf{C}$ by the elements of $d\pi(\mathfrak{g})$, then given any non-zero element $u \in H_m$ for some m, we have $H(K) = \mathfrak{A}u$. In other words, the algebra $\mathfrak{A}$ operates transitively on $H(K)$.*

Proof. This is obvious from Theorem 3, §2, and Theorem 1, §1.

Lemma 3. *Let $V = \sum H_n$ be a vector space over $\mathbf{C}$, expressed as a direct sum of subspaces H_n. Let $\mathfrak{A}$ be an algebra of linear endomorphisms of V generated by elements $X_1, \ldots, X_d$. Let m be an integer such that H_m has dimension 1, generated by an element u, and assume $\mathfrak{A}u = V$ (i.e. $\mathfrak{A}$ acts* *transitively). Any two positive definite scalar products on V which preserve the orthogonality of the spaces H_n and for which $X_1, \ldots, X_d$ are skew hermitian differ by a positive scalar multiple of each other.*

Proof. After multiplying one scalar product by a constant, we may assume that the two scalar products are equal on H_m. What we have to show is that a scalar product satisfying the conditions of the theorem is determined by its values on H_m. It suffices to consider the scalar product of elements v, w for which there exist $Z, Z' \in \mathfrak{A}$ such that $Zu = v$ and $Z'u = w$, and

$$Z = Z_1 \cdots Z_r \quad \text{and} \quad Z' = Z_1' \cdots Z_s'$$

with elements Z_i, Z_j' which are equal to some of $X_1, \ldots, X_d$. Then

$$\langle v, w\rangle = (-1)^r \langle u, \tilde{Z}Z'u\rangle, \quad \text{where} \quad \tilde{Z} = Z_r \cdots Z_1.$$

By orthogonality, the scalar product in this last expression depends only on the projection of $\tilde{Z}Z'u$ on H_m. Therefore, if we know the value of the scalar product on H_m, it is determined on all of V, as was to be shown.

Lemma 4. *Let π be an irreducible admissible representation of G on a Banach space H. Any two positive definite hermitian products on $H(K)$ for which the elements of $d\pi(\mathfrak{g})$ are skew symmetric are positive scalar multiples of each other.*

Proof. We take the algebra $\mathfrak{A}$ to be the one generated by $d\pi(X)$, $X \in \mathfrak{g}$, and apply Lemmas 2, 3.

Theorem 5. *Let π_1, π_2 be irreducible unitary representations of G, and let $L\colon H_1(K) \to H_2(K)$ be an infinitesimal isomorphism. Then there exists $c > 0$ such that cL is unitary, and its unitary extension $H_1 \to H_2$ is a unitary isomorphism between π_1 and π_2.*

Proof. The first statement is immediate from Lemma 4. After multiplying L by a positive number, let us assume that L is unitary. For X sufficiently small, we have Taylor's formula,

$$\langle \pi(\exp X)v, w\rangle = \sum \frac{1}{n!}\langle d\pi(X)^n v, w\rangle.$$

Putting $\pi = \pi_1$, the formula shows that for all x close to e,

$$\langle L\pi_1(x)v, Lw\rangle_2 = \langle \pi_2(x)Lv, Lw\rangle_2.$$

Hence L commutes with the group action near e, and therefore with the group action itself, since G is connected. This proves Theorem 4.

Remark. The essential ingredient of the above arguments is that in the K-decomposition of $H(K)$, one of the irreducible representations of K occurs with multiplicity 1. A K-invariant positive definite hermitian product on this component is the generalization of our present situation, and similar arguments in the higher dimensional case yield an analogous theorem.

Let π be a representation of G in a Hilbert space H. We do *not* assume π unitary. If $v \in H$ is a non-zero vector, then we consider the coefficient function

$$f_v(x) = \langle \pi(x)u, u \rangle,$$

where u is a unit vector in the direction of v. If H_n has dimension 1, then we can define the n-th coefficient function to be f_v for any vector $v \neq 0$ in H_n.

Theorem 6. *If two representations of G which are unitary on K are infinitesimally isomorphic, and the n-th eigenspace for the representations has dimension* 1, *then the n-th coefficient functions of the two representations are equal.*

Proof. Let $H(K)$ be the space of K-finite vectors for the first representation, let u be a unit vector in H_n, and let

$$f(x) = \langle \pi(x)u, u \rangle.$$

Then the function f is analytic on G (see below), and for all small $X \in \mathfrak{g}$ we have

$$\pi(\exp X)u = \sum \frac{1}{q!} d\pi(X)^q u$$

whence, putting $x = \exp X$,

$$f(x) = \sum \frac{1}{q!} \langle d\pi(X)^q u, u \rangle.$$

Each term $\langle d\pi(X)^q u, u \rangle$ is determined by knowing the derived representation algebraically. Hence f is uniquely determined locally in a neighborhood of the origin. By analyticity, f is determined on all of G.

Remark 1. We used the analyticity result already mentioned, and to be proved in Chapter X, §2.

Remark 2. We shall use Theorem 6 later to see that the trace, properly defined, is independent of the infinitesimal isomorphism class of representations. Indeed, as a distribution, the trace will be taken as a sum of coefficient functions

$$\sum \langle \pi(x)u_n, u_n \rangle$$

and Theorem 6 will apply.

Coefficient functions can sometimes be used to embed a representation infinitesimally into $L^2(G)$, as follows. Let π be a representation of G in a Banach space H and let λ be a non-zero functional on H. Let $v \in H$. We let

$$f_{\lambda, v}(x) = f_v(x) = \lambda(\pi(x)v).$$

We shall keep λ fixed, and hence index f only by v.

In preparation for the next result, we tabulate some more formalism. Let $\varphi \in C_c(G)$, and as usual, let $\varphi^-(x) = \varphi(x^{-1})$. Then

$$f_{\pi(\varphi)v} = f_v * \varphi^- \tag{1}$$

Let $X \in \mathfrak{g}$. Let $f \in L^2(G)$ and $\varphi \in C_c^\infty(G)$. Then

$$\mathfrak{L}_X(f * \varphi) = f * \mathfrak{L}_X \varphi \tag{2}$$

$$(X * \varphi)^- = \mathfrak{L}_X(\varphi^-) \tag{3}$$

$$f_{\pi(X)\pi(\varphi)v} = \mathfrak{L}_X(f_{\pi(\varphi)v}). \tag{4}$$

All these formulas are immediate from the definitions.

Lemma 5. *Let π be the representation by right translation on $L^2(G)$, i.e. $\pi(x)f(y) = f(yx)$. Let $f \in L^2(G)$ and $\varphi \in C_c^\infty(G)$. Then $f * \varphi$ is a C^∞ vector in $L^2(G)$, and*

$$d\pi(X)(f * \varphi) = \mathfrak{L}_X(f * \varphi).$$

Proof. It suffices to prove that the map $x \mapsto \pi(x)f$ is C^∞ near the origin in G. Let

$$\pi_{tX}f(y) = f(y \exp(tX)).$$

We first prove the formula, and we have to show that

$$\frac{\pi_{tX}(f * \varphi) - (f * \varphi)}{t}$$

approaches a limit in $L^2(G)$, given by $\mathfrak{L}_X(f * \varphi)$. Thus we have to investigate the limit of

$$\int_G \frac{f(y \exp(tX)x^{-1})\varphi(x) - f(yx^{-1})\varphi(x)}{t}\, dx$$

as a function of y, in $L^2(G)$. Change variables, letting

$$x \mapsto xy \exp(tX).$$

The above expression is equal to

$$\int_G \frac{f(x^{-1})\varphi(xy \exp(tX)) - f(x^{-1})\varphi(xy)}{t}\, dx.$$

Let $\psi(t) = \varphi(xy \exp(tX))$. By the mean value theorem,

$$\psi(t) - \psi(0) = \psi'(t_0)t = \mathfrak{L}_X \varphi(yx \exp(t_0 X))t$$

by the mean value theorem, for some t_0 between t and 0. Hence our expression is equal to

$$\int_G f(x^{-1})\mathfrak{L}_X \varphi(xy \exp(t_0 X))\, dx,$$

which is a function of y. We have to show that this function of y approaches

$$f * \mathfrak{L}_X \varphi(y) = \int_G f(yx^{-1})\mathfrak{L}_X \varphi(x)\, dx = \int_G f(x^{-1})\mathfrak{L}_X \varphi(xy)\, dx$$

in $L^2(G)$, and this amounts to showing that the integral

$$\int_G \left| \int_G f(x^{-1})[\mathfrak{L}_X \varphi(xy \exp(t_0 X)) - \mathfrak{L}_X \varphi(xy)]\, dx \right|^2 dy$$

approaches 0 as t_0 approaches 0. We bring the absolute value under the integral sign and use the Schwarz inequality. We can then take the limit under the integral sign to prove the formula. Induction and (2) show that f is a C^∞ vector, as was to be proved.

***Theorem* 7.** *Let π be an irreducible admissible representation of G in a Banach space E. Let $\lambda \in E'$, $\lambda \neq 0$, and let*

$$f_v(x) = \lambda(\pi(x)v), \qquad v \in E(K),$$

be the corresponding coordinate function. Let G operate by right translation on $L^2(G)$. If f_u is in $L^2(G)$ for some $u \neq 0$ in $E(K)$, then f_v is in $L^2(G)$ for all $v \in E(K)$, and the map

$$v \longmapsto f_v$$

is a $\mathfrak{g}$-embedding of $E(K)$ into $L^2(G)$. In particular, π is unitarized by this embedding.

Proof. By irreducibility, $\sum \pi(S^\infty_{m,n})$ acts transitively on $E(K)$. By formula (1), we conclude that $f_v \in L^2(G)$ for all $v \in E(K)$. Lemma 5 and formulas (2), (3), (4) immediately show that the map $v \mapsto f_v$ commutes with the derived action of the Lie algebra, and by irreducibility, $v \mapsto f_v$ is injective, i.e. is an embedding, thereby proving our theorem.

§4. THE LIE DERIVATIVES ON G

In this section, we compute the Lie derivatives explicitly for functions on G. As we know, $SL_2(\mathbf{R})$ operates on the upper half plane. It is more convenient here to deal with the operation of $GL_2^+(\mathbf{R})$, in which of course the scalar diagonal elements operate trivially, because we can use the two coordinates of the upper half plane (x, y) for $GL_2^+(\mathbf{R})$.

Thus we deal with $G = GL_2^+(\mathbf{R})$ and write an element of G as

$$g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

We let $z = x + iy \in \mathfrak{H}, y > 0$, so that $gi = z = x + iy$. Under this operation, we have

$$\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix} \mapsto x + iy,$$

and an element g has the unique decomposition

$$g = \begin{pmatrix} u & 0 \\ 0 & u \end{pmatrix}\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

with $u, y > 0$. With this notation we have, from matrix multiplication,

(1) $$ue^{i\theta} = d - ic \quad \text{and} \quad e^{i\theta} = \frac{d - ic}{|d - ic|};$$

(2) $$x + iy = \frac{1}{u}e^{i\theta}(a + ib).$$

Any function Φ on G can be written in terms of our coordinates (u, x, y, θ), namely

$$\Phi(g) = \Phi\left(u\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}r(\theta)\right) = F(u, x, y, \theta).$$

Since we want to look at $SL_2(\mathbf{R})$, we assume for simplicity that Φ is independent of u, whence we omit the u and write simply $F(x, y, \theta)$.

Let $X \in \mathfrak{g}$. Put

$$g_*(t) = g \exp(tX).$$

Then

$$\Phi(g_*(t)) = F(x_*(t), y_*(t), \theta_*(t))$$

and

$$\frac{d}{dt}\Phi(g_*(t))\Big|_{t=0} = \frac{\partial F}{\partial x}\frac{dx_*}{dt}\Big|_{t=0} + \frac{\partial F}{\partial y}\frac{dy_*}{dt}\Big|_{t=0} + \frac{\partial F}{\partial \theta}\frac{d\theta_*}{dt}\Big|_{t=0}.$$

We shall compute the explicit formulas for the Lie derivatives associated with the elements of the Lie algebra given by

$$X = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

$$V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Note that

(3) $$V = 2X - W, \qquad E^- = H - iV, \qquad E^+ = H + iV.$$

(4) $$\boxed{\mathfrak{L}_X = y \cos 2\theta \frac{\partial}{\partial x} + y \sin 2\theta \frac{\partial}{\partial y} + \sin^2\theta \frac{\partial}{\partial \theta}.}$$

Proof. We have

$$\exp(tX) = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix},$$

so

$$g_*(t) = \begin{pmatrix} a & at + b \\ c & ct + d \end{pmatrix},$$

$$g_*(t)i = \frac{ai + at + b}{ci + ct + d} = z_*(t) = x_*(t) + iy_*(t),$$

$$z_*'(0) = \frac{ad - bc}{(ci + d)^2} = ye^{2i\theta},$$

$$i\theta_*(t) = \log \frac{ct + d - ic}{\left((ct + d)^2 + c^2\right)^{1/2}}.$$

Hence

$$x_*'(0) = y \cos 2\theta,$$

$$y_*'(0) = y \sin 2\theta,$$

$$\theta_*'(0) = \sin^2 \theta = \frac{1 - \cos 2\theta}{2}.$$

This proves our formula.

Trivially, we have

$$\boxed{\mathfrak{L}_W = \frac{\partial}{\partial \theta}.} \tag{5}$$

Since $V = 2X - W$, we find

$$\boxed{\mathfrak{L}_V = 2y \cos 2\theta \frac{\partial}{\partial x} + 2y \sin 2\theta \frac{\partial}{\partial y} - \cos 2\theta \frac{\partial}{\partial \theta}.} \tag{6}$$

Since

$$\exp(tH) = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix},$$

we find

$$g_*(t) = \begin{pmatrix} ae^t & be^t \\ ce^t & de^t \end{pmatrix},$$

from which we get $g_*(t)i = z_*(t)$,

$$g_*(t)i = \frac{ae^ti + be^{-t}}{ce^ti + de^{-t}} = x_*(t) + iy_*(t).$$

We also have

$$i\theta_*(t) = \log \frac{de^{-t} - ice^t}{(d^2e^{-2t} + c^2e^{2t})^{1/2}}.$$

From this we obtain the Lie derivative

$$\boxed{\mathfrak{L}_H = -2y \sin 2\theta \frac{\partial}{\partial x} + 2y \cos 2\theta \frac{\partial}{\partial y} + \sin 2\theta \frac{\partial}{\partial \theta}.} \tag{7}$$

Since $E^- = H - iV$, we obtain

$$\boxed{\begin{aligned} \mathfrak{L}_{E^-} &= -2iye^{-2i\theta}\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) + ie^{-2i\theta}\frac{\partial}{\partial \theta} \\ &= -2iye^{-2i\theta}\frac{\partial}{\partial \bar{z}} + ie^{-2i\theta}\frac{\partial}{\partial \theta}. \end{aligned}} \tag{8}$$

Observe the $\partial/\partial\bar{z}$ which pops out, and is the Cauchy–Riemann operator. A function $f(z)$ is analytic if and only if $\partial f/\partial \bar{z} = 0$. This concludes our tabulation of the Lie derivatives on G.

§5. IRREDUCIBLE COMPONENTS OF THE INDUCED REPRESENTATIONS

Let s be a complex number. Let $H(s)$ be the space of functions on $G = SL_2(\mathbf{R})$ whose restriction to K lies in $L^2(K)$, and satisfying the condition

$$f(ank) = \rho(a)^{s+1}f(k)$$

in the Iwasawa decomposition $G = ANK$. Let π be the representation of G on $H(s)$ by right translation.

From the decomposition of a matrix g in the form

$$g = u\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix} r(\theta), \qquad u, y > 0,$$

it follows that

$$f(g) = y^{(s+1)/2}f(r(\theta)).$$

Let φ_n be the function in $H(s)$ such that

$$\varphi_n(r(\theta)) = e^{in\theta}.$$

Then

$$\varphi_n\left(u\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix} r(\theta)\right) = y^{(s+1)/2}e^{in\theta},$$

and by the formulas for the Lie derivatives which we found in the preceding section, we obtain the corresponding Lie derivatives of these functions φ_n, as follows:

$$
(1)\qquad \boxed{\begin{aligned} \mathcal{L}_W \varphi_n &= in\varphi_n, \\ \mathcal{L}_{E^-} \varphi_n &= (s+1-n)\varphi_{n-2}, \\ \mathcal{L}_{E^+} \varphi_n &= (s+1+n)\varphi_{n+2}. \end{aligned}}
$$

We now prove the two necessary lemmas identifying the effect of a Lie derivative $\mathcal{L}_X$ with $d\pi(X)$, and an analytic function with an analytic vector.

Lemma 1. *Let π be the representation by right translation of G on $H(s)$. Let $X \in \mathfrak{g}$, let $f \in H(s)$, and assume that f is C^∞ on G. Then f is a C^∞ vector as element of $L^2(K)$, and*

$$d\pi(X)f = \mathcal{L}_X f.$$

Proof. Let

$$\pi_{tX} f(g) = f(g \exp(tX)).$$

We have to verify that

$$\frac{\pi_{tX} f - f}{t} \to \mathcal{L}_X f$$

in $L^2(K)$, i.e. that

$$\int_K \left| \frac{f(k \exp(tX)) - f(k)}{t} - \mathcal{L}_X f(k) \right|^2 dk$$

approaches 0 as $t \to 0$. Let $\psi(t) = f(k \exp(tX))$. By the mean value theorem, there exists $0 < t_0 < t$ (say$t > 0$) such that

$$\frac{\psi(t) - \psi(0)}{t} = \psi'(t_0) = \mathcal{L}_X f(k \exp(t_0 X)).$$

The values $\mathcal{L}_X f(k \exp(t_0 X))$ are bounded ıor $k \in K$ and small t_0. By the dominated convergence theorem, we can take the limit under the integral sign to prove the desired formula.

Lemma 2. *Let f be an analytic function on G. Let $\pi(x)f$ be the right* *translation,*

$$\pi(x)f(y) = f(yx).$$

Then the map

$$x \longmapsto \pi(x)f|K$$

is an analytic map of G into $L^2(K)$.

Proof. Let $x_0 \in G$. If we have proved that $x \longmapsto \pi(x)f|K$ is analytic in a neighborhood of the origin e on G, then

$$x \longmapsto \pi(x_0^{-1}x)f = \pi(x_0^{-1})\pi(x)f|K$$

is seen to be analytic because $\pi(x_0^{-1})$ is a continuous linear map. Therefore, it suffices to prove our lemma in a neighborhood of the origin on G. Let $(x_1, \ldots, x_m)$ be local coordinates for x near the origin such that

$$x_1(e) = \cdots = x_m(e) = 0.$$

By the compactness of K, we can find intervals

$$0 = \theta_0 < \theta_1 < \cdots < \theta_r = 2\pi$$

such that on each interval $[\theta_j, \theta_{j+1})$ we have a power series expansion

$$f(r(\theta)x) = \sum_{(\nu)} \sum_{\mu} c_{(\nu),\mu}(\theta - \theta_j)^{\mu} x_1^{\nu_1} \cdots x_m^{\nu_m}.$$

Let

$$\psi_{(\nu)}^{(j)}(\theta) = \sum_{\mu} c_{(\nu),\mu}(\theta - \theta_j)^{\mu},$$

and let $\psi_{(\nu)}$ be the function of θ equal to $\psi_{(\nu)}^{(j)}$ on the j-th interval $[\theta_j, \theta_{j+1})$. Then

$$f(r(\theta)x) = \sum_{(\nu)} \psi_{(\nu)}(\theta) x_1^{\nu_1} \cdots x_m^{\nu_m},$$

in other words,

$$\pi(x)f = \sum_{(\nu)} x_1^{\nu_1} \cdots x_m^{\nu_m} \psi_{(\nu)}.$$

The absolute convergence of the series implies the $L^2(K)$-convergence, so we get the desired analytic expansion for $\pi(x)f$ as a power series with coefficients in $L^2(K)$.

Irreducible subspaces

We now see that our functions φ_n are analytic vectors, as elements of $H(s)$, and the values for the Lie derivatives with respect to W, E^-, E^+ on φ_n give us the values for $d\pi_s$ at these elements, namely:

$$(2) \qquad \boxed{\begin{aligned} d\pi_s(W)\varphi_n &= in\varphi_n, \\ d\pi_s(E^-)\varphi_n &= (s+1-n)\varphi_{n-2}, \\ d\pi_s(E^+)\varphi_n &= (s+1+n)\varphi_{n+2}. \end{aligned}}$$

Using Theorem 1, §1, we can therefore easily determine the irreducible subspaces of $H(s)$, identified by restriction with $L^2(K)$.

Case 1. s is not an integer.

Let H' be an irreducible subspace of $H(s)$. Then H' decomposes as an orthogonal direct sum of irreducible subspaces over K, which are one dimensional. In particular, H' contains φ_n for some n, and therefore, in view of the values

$$s+1-n \qquad \text{or} \qquad s+1+n$$

which cannot be 0, we conclude that H' contains either all φ_n with n even, or all φ_n with n odd, and that

$$\widehat{\bigoplus_{n \text{ even}}} (\varphi_n), \qquad \widehat{\bigoplus_{n \text{ odd}}} (\varphi_n)$$

are irreducible subspaces.

Case 2. $s = 0$.

In this case, we have three irreducible subspaces, as follows.

$$\widehat{\bigoplus_{n \text{ even}}} (\varphi_n), \qquad \widehat{\bigoplus_{\substack{n \text{ odd} \\ n \geqslant 1}}} (\varphi_n), \qquad \widehat{\bigoplus_{\substack{n \text{ odd} \\ n \leqslant -1}}} (\varphi_n).$$

Case 3. $s = m - 1$ where m is an integer $\geqslant 2$.

In this case we observe that the function φ_m is annihilated by $d\pi_s(E^-)$, is an eigenvector for $d\pi_s(W)$, and is sent on a scalar multiple of φ_{m+2} by

$d\pi_s(E^+)$. Consequently the space

$$H^{(m)} = \widehat{\bigoplus_{\substack{n \geqslant m \\ n \equiv m}}} (\varphi_n)$$

is irreducible. [We write $n \equiv m$ for $n \equiv m \pmod 2$.] On the other hand, for an analogous reason we see that the space

$$H^{(-m)} = \widehat{\bigoplus_{\substack{n \leqslant -m \\ n \equiv -m}}} (\varphi_n)$$

is irreducible, because $d\pi_s(E^+)\varphi_{-m} = 0$.

There is also an irreducible piece arising from parity, namely

$$\widehat{\bigoplus_{n \not\equiv m}} (\varphi_n).$$

The factor space

$$V(m-1) = H(m-1) \Big/ \left[H^{(m)} + H^{(-m)} + \widehat{\bigoplus_{n \not\equiv m}} (\varphi_n) \right]$$

is finite dimensional, of dimension $m - 1$. It has a basis represented by the elements

$$\{\varphi_{-m+2}, \varphi_{-m+4}, \ldots, \varphi_{m-2}\}$$

in $H(m-1)$. The action of the Lie algebra shows that this factor space is irreducible. It has both a highest and lowest weight vector.

The above infinite dimensional irreducible pieces of the representations in case 3 have a lowest weight vector of weight m, or a highest weight vector of weight $-m$. We shall prove later that for $m \geqslant 2$, there exists a unitary representation whose derived representation is algebraically isomorphic to these.

Recall that in case 2 with $s = 0$ we found two irreducible unitary representations with highest weight vector of weight -1 and lowest weight vector of weight 1, respectively.

A unitary representation is usually said to be in the **discrete series** if it occurs in the regular representation of $L^2(G)$, with the operation of left translation. We shall see that the unitary representations corresponding to those of Case 3 with $m \geqslant 2$ are in the discrete series. Those of Case 2 are not, but in many respects behave like discrete series representations. They could be called "**mock discrete**" representations.

Case 4. $s = -m + 1$ **where** m **is an integer** $\geqslant 2$.

This case is analogous to Case 3, and in fact is dual to it. Indeed, we know from the general theory of induced representations that for any complex number s, the space $H(s)$ is dual to $H(-s)$, letting $H(s)$ be the space of the representation induced from the character μ_s as in Chapter III, §2, Theorem 3. It is clear from the action of the Lie algebra that we find an irreducible finite dimensional subspace $V(-m+1)$ rather than a finite dimensional factor space, and this subspace is spanned by the eigenvectors for K with the same indices as for $V(m-1)$. Furthermore, the factor space

$$H(-m+1)/V(-m+1)$$

contains two irreducible subspaces, one with a highest weight vector of weight $-m$, and another with lowest weight vector of weight m. If we let $\{\varphi_n\}$ be the orthonormal basis of $H(m-1)$ as before, and if we let $\{\varphi'_{-n}\}$ be the corresponding orthonormal basis of $H(-m+1)$, then these bases are dual bases to each other, in the duality of III, §2, Th. 3. In particular, the finite dimensional spaces

$$V(m-1) \qquad \text{and} \qquad V(-m+1)$$

are dual to each other. For further use of this situation, cf. Chapter VII, §4, especially Lemma 2, which describes $V(-m+1)$ more explicitly.

Let $H^{(m)}$ and $H'^{(m)}$ be the spaces with lowest weight vector of weight m in $H(m-1)$ and $H(-m+1)/V(-m+1)$ respectively. We can compute explicitly an infinitesimal isomorphism

$$L: H^{(m)}(K) \to H'^{(m)}(K).$$

Let φ_n, φ'_n be the standard basis elements of H_n and H'_n, where

$$n \geqslant m \qquad \text{and} \qquad n \equiv m \pmod 2.$$

We want an isomorphism L such that $L\varphi_n = b_n\varphi'_n$, with some constant b_n. The known effect of $\mathbf{E}^+$ on φ_n, φ'_n shows that we have the recursion relation

$$b_{n+2} = b_n \frac{-m+2+n}{m+n}.$$

Thus the choice of b_m determines b_n uniquely for all n as above.

§6. CLASSIFICATION OF ALL UNITARY IRREDUCIBLE REPRESENTATIONS

Although it is irrelevant for the particular topics considered in the present book, it becomes necessary sometimes to know what all the unitary irreducible representations of G are. For instance, other such representations than those occurring in the Plancherel formula on G will occur in the representation on $L^2(\Gamma\backslash G)$ for various discrete subgroups Γ, and none of them can be dismissed a priori. We shall therefore give Bargmann's classification, which is now immediate since we have all the needed tools at hand.

Let π be an irreducible admissible representation of G on a space H, with derived representation $d\pi$ on $H(K)$. We have

$$H(K) = \sum H_n$$

where n ranges over a subset of integers of a given parity (i.e. in a congruence class mod 2).

Let us assume first that there is no highest or lowest weight vector.

Suppose first, to fix ideas, that H_0 occurs. Let v_0 be a basis element of H_0. We then select a basis for H_n (n even) such that

$$v_{n+2} = E_+ v_n,$$

where we abbreviate $d\pi(E_+)v_n$ by E_+v_n. This is immediately done by induction to the right and left of v_0. Let c_n be the number such that

$$E_- v_n = c_n v_{n-2}.$$

We ask for necessary conditions that there exist a positive definite scalar product on $H(K)$ compatible with this representation of $\mathfrak{g}_{\mathbf{C}}$, in particular such that $d\pi(X)$, $X \in \mathfrak{g}$, is skew hermitian. We can obviously choose the length of v_0 arbitrarily, say $a_0 > 0$, i.e.

(1) $$\langle v_0, v_0 \rangle = a_0^2.$$

From

$$(E_+ E_- - E_- E_+)v_n = -4iWv_n = 4nv_n,$$

we get the condition

(2) $$c_n - c_{n+2} = 4n.$$

From the skew hermitian condition

$$\langle E_+ v_n, v_{n+2} \rangle = \langle v_n, -E_- v_{n+2} \rangle$$

we get the condition

(3) $$a_{n+2}^2 = -\bar{c}_{n+2} a_n^2.$$

This shows that c_n is necessarily real and *negative*. Furthermore, from (2) we see that a choice of c_0 completely determines c_n for all n. From (3) we then see that a_n is completely determined for all n by such a choice of c_0. The possible unitarization therefore depends **uniquely** on the choice of a negative number c_0.

A similar discussion can be carried out for the case when the parity is odd, with a basis vector v_1 for H_1. Conditions (2) and (3) remain unchanged, and the possible unitarizations correspond uniquely to a choice of a negative number c_1.

Consider the induced representation π_s on $H(s)$, with a complex number s. Since

$$d\pi(E_+)\varphi_n = (s + 1 + n)\varphi_{n+2} \qquad \text{and} \qquad d\pi(E_-)\varphi_n = (s + 1 - n)\varphi_{n-2},$$

we see that the condition for c_0 to be negative amounts to

$$(s + 1)(s - 1) < 0.$$

This amounts to s being pure imaginary, or s real and $-1 < s < 1$. We already know that when s is pure imaginary, we obtain a unitary irreducible representation. The case when $-1 < s < 1$ can be unitarized by completing $H(K)$ according to the scalar product obtained from the necessary conditions above. The irreducible representations arising from this interval are called those of the **complementary series**.

In the case of odd parity, we note that the condition for c_1 to be negative amounts to $s^2 = c_1 < 0$. Therefore in this case, we cannot have a real value for s, only the pure imaginary values.

Finally, consider the possibility of a unitarization in case there is, say, a highest weight vector of weight m. We contend that m is necessarily < 0. Indeed, if $E_+ v_m = 0$, then the relation analogous to (2) in this case is

$$c_m = 4m,$$

and c_m must be negative. Similarly, we see that in the case of lowest weight vector of weight m, we must have $m > 0$.

We can therefore summarize Bargmann's classification as follows.

Theorem 8. *The irreducible unitary representations of $SL_2(\mathbf{R})$ must be infinitesimally isomorphic to the following components of π_s in $H(s)$, for the following values of s:*

i) *The discrete series, with $s = m - 1$ or $s = -m + 1$, and m is an integer $\geqslant 2$.*
ii) *The mock discrete series, with $s = 0$, consisting of the component with highest weight vector -1 and lowest weight vector $+1$.*
iii) *The principal series with $s = i\tau$, $\tau \neq 0$, both parities; and for $s = 0$, the component of even parity.*
iv) *The complementary series with $-1 < s < 1$, $s \neq 0$, components which are not in the mock discrete series at $s = 0$, and which have even parity.*

Theorem 8 justifies our cross mentioned in Chapter I, §4. The full space $\hat{G}$ is parametrized by the picture in Fig. 1, where the vertical line is a double line (one copy for each parity), while the dots at the integers, together with the horizontal segment, occur merely once. These dots index lowest or highest weight vectors.

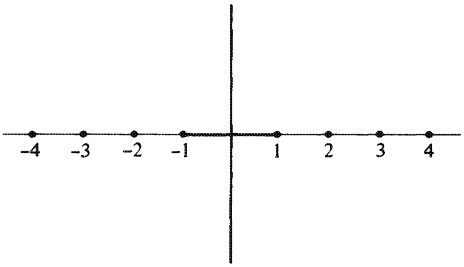

Figure 1

We also see that the irreducible unitary representations occur in induced representations. Theorem 5 of §3 gives us the uniqueness.

§7. SEPARATION BY THE TRACE

For compact groups, we know that the character of the representation, i.e. its trace, determines the representation. We are interested in a similar statement for Hilbert space unitary irreducible representations. Roughly speaking, we cannot hope to separate one representation in a continuous family from all the others. We can only hope to separate a representation occurring discretely. We now show how to do this for the representations described in §5.

We denote by $H^{(m)}$ for every integer $m \neq 0$, a representation space for a representation π_m having highest weight vector of weight m if m is negative, and lowest weight vector of weight m if m is positive. Then $H_m^{(m)}$ is the highest (resp. lowest) space in the orthogonal decomposition over K. As in Chapter V, we let $\mathbf{H}$ be the Harish transform.

Theorem 9. *Given m as above, there exists $\psi \in C_c^\infty(G)$ such that:*

i) $\pi^1(\psi)|H_m^{(m)} =$ *identity on* $H_m^{(m)}$.
ii) $\pi^1(\psi)$ *annihilates* $H_n^{(m)}$ *if* $n \neq m$ *and annihilates* $H^{(q)}$ *if* $q \neq m$.
iii) $\mathbf{H}\psi = 0$.

Proof. Say $m \geqslant 1$. For any $f \in S^{\infty}_{m,m}$ we have

$$\begin{aligned} \pi^1(f)H_m^{(m)} &= H_m^{(m)} \quad \text{or} \quad 0, \\ \pi^1(f)H_n^{(q)} &= 0 \quad \text{if} \quad n \neq m. \end{aligned}$$

We visualize the weighted series as follows, dealing first with those lying between 1 and m:

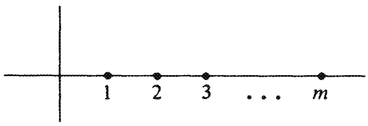

Figure 2

and separate these as follows. The image of $\pi^1(f)$ for $f \in S_{n,n}$ is at most one-dimensional and hence it makes sense to speak of the trace of this operator. Consider the functionals $\theta_1, \ldots, \theta_m$ on

$$\sum_{n=1}^{m} S^{\infty}_{n,n}$$

given by

$$\theta_j(f) = \operatorname{tr} \pi_j^1(f)$$

for $j = 1, \ldots, m$. They are obviously linearly independent, so the map

$$f \mapsto (\theta_1(f), \ldots, \theta_m(f))$$

sends

$$\sum_{n=1}^{m} S^{\infty}_{n,n} \to \mathbf{C}^m.$$

Consequently there exists a function $f \in \sum_{n=1}^{m} S^{\infty}_{n,n}$ such that

$$\operatorname{tr} \pi_m^1(f) = 1,$$

$$\operatorname{tr} \pi_n^1(f) = 0 \quad \text{if} \quad 1 \leqslant n < m.$$

Also, automatically because of the transformation properties of elements of $S_{n,n}$ and the orthogonality relations, we have

$$\pi_n^1(f) = 0 \quad \text{if} \quad n \leqslant -1 \quad \text{or} \quad n > m.$$

We have therefore succeeded in separating the representation π_m from all other π_n ($n \neq 0$, $n \neq m$) having highest and lowest weights.

There remains to prove that we can adjust our constructed function f to make the Harish transform equal to 0. Our function f lies in $C_c^\infty(G, K)$, i.e. is invariant under conjugation by elements of K. There exists a function $f_1 \in C_c^\infty(G//K)$ such that

$$\mathbf{H}f = \mathbf{H}f_1,$$

by the fundamental theorem on spherical transforms, i.e. the surjectivity of the Harish transform (V, §2, Th. 3). We have

$$\pi_n^1(f_1) = 0$$

because f_1 is bi-invariant, and hence $\pi_n^1(f_1)$ annihilates any eigenspace of K with non-trivial eigenvalues, i.e. $\pi_n^1(f_1)$ kills all $H_n^{(q)}$ if $n \neq 0$, for all q. We let $\psi = f - f_1$. Then $\mathbf{H}\psi = 0$ and

$$\pi_n^1(\psi) = \pi_n^1(f).$$

This proves our theorem.

Theorem 9 is due to Duflo–Labesse [Du, La]. I owe the above proof to Harish-Chandra.

Theorem 4 of Chapter III, §4 combined with Theorem 9 above give us a separation of various representations by the trace, but Theorem 9 is even stronger, since it even gives a separation at the operator level.

VII Traces

In this chapter we deal systematically with the trace in infinite dimensional representations, especially those we have explicitly constructed. We prove that in the induced representations, the trace expressed as the integral of a kernel over the diagonal can be identified with the usual sum of diagonal matrix coefficients. We then compute the trace in various representations.

The integral formulas provide a technical interlude, designed to give a tabulation of various measures involving several decompositions of G, some of which are adapted to the study of conjugacy classes. Except for a set of measure 0 (the conjugacy classes of elements of N), we have the conjugacy classes of elements of A and K, and the traces can be expressed as integrals over such classes. This systematic approach, due to Harish-Chandra, is pursued afterwards to get the Plancherel formula, and gave rise to the name I have chosen for the Harish transform.

The Plancherel formula will be seen to involve the traces of the discrete series in pairs. Harish-Chandra's theorem giving the trace in each component will be omitted; cf. the comments of Theorem 5, §4. This is one aspect of Harish's proof and of Harish's work in more general cases which should be emphasized: The determination of the trace goes beyond the Plancherel formula, and there are higher dimensional instances when the representations have not yet been determined, even though the Plancherel formula is available.

§1. OPERATORS OF TRACE CLASS

Let A be an operator on a Hilbert space H (all operators here are assumed bounded). We call A **Hilbert–Schmidt** if for some orthonormal basis $\{u_i\}$ of H we have

$$\sum |Au_i|^2 < \infty.$$

The same then holds for any other orthonormal basis $\{v_j\}$. Indeed, note that for any vector w we have

$$|w|^2 = \sum |\langle w, v_j\rangle|^2,$$

so that

$$\sum |Au_i|^2 = \sum_{i,j} |\langle Au_i, v_j\rangle|^2 = \sum_{i,j} |\langle u_i, A^* v_j\rangle|^2 = \sum_j |A^* v_j|^2.$$

An operator A is said to be of **trace class** if it is the product of two Hilbert–Schmidt operators, say $A = B^* C$ where B, C are Hilbert–Schmidt. This being the case, we define the **trace** of A to be

$$\operatorname{tr}(A) = \sum \langle Au_i, u_i\rangle = \sum \langle Cu_i, Bu_i\rangle.$$

The first sum shows that it is independent of the choice of C, B. The second sum gives us the absolute convergence of the series, in view of the Schwartz inequality applied twice.

Let us now take K to be the circle group, and let $\{\varphi_n\}$ be the usual orthonormal basis, $\varphi_n(\theta) = e^{in\theta}$. Let q be a C^∞ function on $K \times K$. Write the Fourier series expansion

$$q = \sum c_{mn} \varphi_m \otimes \bar{\varphi}_n.$$

Then the Fourier coefficients c_{mn} tend to 0 very rapidly, as one sees at once upon integrating by parts. In fact, for every positive integer d we have an estimate of type

$$|c_{mn}| \ll \frac{1}{(1 + |m| + |n|)^d},$$

where the constant implied in the estimate symbol $\ll$ depends on d. Thus the Fourier series of q converges rapidly to the values of q, and the computations of I, §3, are valid. In addition, we also have

Theorem 1. *Let q be a C^∞ function on $K \times K$. Then the integral operator Q defined by q is of trace class.*

Proof. We have to express Q as a product of two Hilbert–Schmidt operators. Let d be a large positive integer. Let $P_{m,n}$ be the integral operator defined by the kernel $\varphi_m \otimes \bar{\varphi}_n$, so that

$$P_{m,n}\varphi_j = 0 \quad \text{if} \quad j \neq n,$$

$$P_{m,n}\varphi_n = \varphi_m.$$

Let P_j be the projection on the one-dimensional space (φ_j). Let

$$B = \sum_{m,n} c_{m,n}(1 + n^{2d})P_{m,n} \qquad \text{and} \qquad C = \sum_j \frac{1}{1 + j^{2d}} P_j.$$

The series for B and C converge rapidly, defining Hilbert–Schmidt operators. It is clear that $BC = Q$, as desired.

Let $G = SL_2(\mathbf{R})$ and let π be a representation of G in a Hilbert space H. For each integer n we let H_n be the subspace of H on which K acts by the n-th character, so that

$$H = \hat{\bigoplus} H_n.$$

We say that π is **strictly admissible** if the dimensions of the spaces H_n are bounded independently of n. By II, §1, Th. 2, we know that an irreducible unitary representation of G is strictly admissible (the dimension of H_n is 0 or 1 for all n).

***Theorem* 2.** *Let π be a strictly admissible representation of G on a Hilbert space H. If $f \in C_c^\infty(G)$, then $\pi^1(f)$ is of trace class.*

Proof. The idea is similar to the idea used in the proof of Theorem 1. Write $G = ANK$ and let $B = AN$. Then by definition,

$$\pi^1(f) = \int_G f(x)\pi(x)\,dx = \int_B \int_K f(bk)\pi(b)\pi(k)\,dk\,db$$

$$= \int_B \pi(b) \int_0^{2\pi} f(bk_\theta)\pi(k_\theta)\,d\theta\,db.$$

[We don't care here that K has measure 2π.] But $\pi(k_\theta)$ has an expansion

$$\pi(k_\theta) = \sum_n e^{in\theta}P_n$$

where P_n is the projection on the space H_n, and consequently

$$\pi^1(f) = \int_B \pi(b) \int_0^{2\pi} \sum_n e^{2\pi in\theta} f(bk_\theta)\,d\theta\,P_n\,db$$

$$= \int_B \pi(b) \sum_n f_n(b) P_n\,db,$$

where

$$f_n(b) = \int_0^{2\pi} f(bk_\theta)e^{2\pi in\theta}\,d\theta.$$

Integrating by parts shows that given an integer $d > 0$ we have

$$|f_n(b)| \ll \frac{1}{1 + n^{2d}},$$

and f_n has compact support. Let

$$Q_n = \int_B \pi(b) f_n(b)\, db, \qquad \text{so that} \qquad |Q_n| \ll \frac{1}{1 + n^{2d}}.$$

(Remember that π is locally bounded, I, §1.) Then

$$\pi^1(f) = \sum_n Q_n P_n.$$

Let

$$A_1 = \sum_n (1 + n^{2d}) Q_n P_n,$$

$$A_2 = \sum_n \frac{1}{1 + n^{2d}} P_n.$$

Then the series defining A_1, A_2 converge rapidly, thus defining Hilbert–Schmidt operators. Furthermore, $\pi^1(f) = A_1 A_2$, thereby proving our theorem.

Let π be a strictly admissible representation of G on a Hilbert space H. Then the association

$$\psi \longmapsto \operatorname{tr} \pi^1(\psi)$$

is a functional on $C_c^\infty(G)$, and it is easy to verify that it is a distribution (for the definition, cf. the end of Appendix 4). For the convenience of the reader, we prove this fact when π is unitary, although we won't use it. If $\psi \in C_c^\infty(G)$ we write for simplicity $W\psi$ instead of $\mathfrak{L}_W \psi$, where W is the usual generator of the Lie algebra of K. It suffices to prove that for all ψ with support in a compact set Ω, we have

$$|\operatorname{tr} \pi^1(\psi)| \leqslant C_\Omega \|(1 - W^2)\psi\|_\Omega.$$

Let u_n be a unit vector in H_n. We let

$$f_n(x) = \langle \pi(x) u_n, u_n \rangle$$

be the corresponding coefficient function, and we shall give an estimate for each term

$$\int_G \psi(x) f_n(x)\, dx$$

which will imply what we want. We have by a trivial integration by parts:

$$\int_G (1 - W^2)\psi(x)f_n(x)\,dx = \int_G \psi(x)(1 - W^2)f_n(x)\,dx$$
$$= (1 + n^2)\int_G \psi(x)f_n(x)\,dx.$$

Consequently

$$\left|\int_G \psi(x)f_n(x)\,dx\right| \leqq \frac{1}{(1 + n^2)}\int_\Omega |(1 - W^2)\psi(x)f_n(x)|\,dx.$$

Since $|f_n(x)| \leqq 1$ (having assumed π unitary), the desired estimate follows at once.

In this chapter, it turns out that the trace functional can be "represented" by a function $T_\pi(x)$, namely

$$\operatorname{tr} \pi^1(\psi) = \int_G \psi(x)T_\pi(x)\,dx.$$

Since the same Haar measure occurs on the left-hand side and right-hand side of the above equation, we see that the function is independent of the choice of Haar measure. If it exists, it is a priori only defined almost everywhere. However, we shall see that this function is continuous on an open set, and can be chosen to be 0 outside this open set, so even this ambiguity will be removed. No other type of "distribution" will occur in this chapter.

Suppose that H is finite dimensional. The sum expressing the trace

$$\operatorname{tr} \pi^1(\psi) = \sum_i \int_G \psi(x)\langle \pi(x)u_i, u_i\rangle\,dx$$

is finite, and can be taken under the integral sign. Therefore $T_\pi(x)$ is the genuine ordinary trace of elementary algebra.

For simplicity, we shall often write $\pi(\psi)$ instead of $\pi^1(\psi)$.

For the rest of this section, we assume that the representations are strictly admissible.

Invariance under conjugation

Let π be a representation of G on a Hilbert space H. Let $\psi \in C_c^\infty(G)$. Let $y \in G$ and let ψ^y be the function such that

$$\psi^y(x) = \psi(y^{-1}xy).$$

Then

$$\operatorname{tr} \pi(\psi) = \operatorname{tr} \pi(\psi^y).$$

Proof. We have trivially $\pi(\psi^y) = \pi(y)\pi(\psi)\pi(y^{-1})$, and

$$\operatorname{tr}(XAX^{-1}) = \operatorname{tr}(A)$$

(cf. the appendix to this chapter).

Assume that K has measure 1. *Define*

$$\psi_K(x) = \int_K \psi(k^{-1}xk)\, dk.$$

Then tr $\pi(\psi)$ = tr $\pi(\psi_K)$.

Proof. Consider the integral

$$\operatorname{tr} \pi(\psi_K) = \sum_i \int_G \int_K \psi(k^{-1}xk)\langle \pi(x)u_i, u_i\rangle\, dk\, dx.$$

Interchange the order of integration, and move the integral with respect to K on the outside. The transformation

$$x \longmapsto kxk^{-1}$$

has modular function equal to 1, and our trace is equal to

$$\int_K \sum_i \int_G \psi(x)\langle \pi(k)\pi(x)\pi(k^{-1})u_i, u_i\rangle\, dx\, dk.$$

Our assertion follows from the fact that $\operatorname{tr} \pi(\psi) = \operatorname{tr} \pi(\psi^k)$.

In view of the above, when computing the trace of operators $\pi^1(\psi)$, we shall usually assume that $\psi \in C_c^\infty(G, K)$, i.e. that ψ is invariant under conjugation by elements of K, in addition to being C^∞ with compact support.

Harish-Chandra proved for the general case of irreducible unitary representations of semisimple Lie groups that the trace is a distribution. This amounts to proving a condition of strict admissibility, i.e. giving bounds for the dimensions of K-irreducible subspaces in the K-finite vectors. (Precisely, $\dim H_\rho/(\dim \rho)^2$ is bounded for $\rho \in \hat{K}$.) He also proved that the trace can be represented by a locally L^1-function. On $SL_2(\mathbf{R})$, we shall follow Harish by splitting the integral over various conjugacy classes of elements of G. We call an element of G **regular** if it has distinct eigenvalues, and we denote by G' the set of regular elements. If S is a subset of G, we denote by S' the set of regular elements in S. We denote by S^G the set of elements conjugate to elements of S, i.e. the set of elements $g^{-1}xg$ with $g \in G$, $x \in S$.

Lemma 1. *Let $G = ANK$. Then*

$$G' = \pm A'^G \cup K'^G.$$

The complement of G' in G has measure 0.

Proof. If an element $g \in SL_2(\mathbf{R})$ has a single eigenvalue of multiplicity 2, then by the Jordan normal form it is conjugate to an element in $\pm N$. Since AN normalizes N, the image of N under conjugation by G has dimension 2, and therefore has measure 0. If an element g has distinct eigenvalues in $\mathbf{R}$, then the Jordan normal form shows that it is conjugate to an element of $\pm A'$. If the eigenvalues are complex conjugate and not real, then standard two-dimensional linear algebra shows that the element is conjugate to an element in K', as asserted.

The above determination of the regular elements implies that

$$\boxed{\int_G f(x)\,dx = \int_{\pm A'^G} f(x)\,dx + \int_{K'^G} f(x)\,dx.}$$

This is the decomposition which we shall use to compute traces, and formulas in §2 will describe various forms for these integrals, depending on representatives for the conjugacy classes.

Infinitesimal invariance

Recall that two norms are **equivalent** if each is less than or equal to a positive scalar multiple of the other. Two positive definite scalar products are called **equivalent** if their norms are equivalent.

Lemma 2. *Let π_1 be a representation of G on a Hilbert space H_1, and let H_2 be the same space as H_1 but with an equivalent scalar product. Let π_2 be the same map as π_1, but viewed as a representation on H_2. Then*

$$\operatorname{tr} \pi_1(\psi) = \operatorname{tr} \pi_2(\psi).$$

Proof. Let $T\colon H_1 \to H_2$ be the identity map, which is bicontinuous. Then

$$\pi_2(x) = T\pi_1(x)T^{-1},$$

and the equality between the traces follows by the general theory of traces (cf. the appendix to this chapter, last theorem).

In particular, suppose that $\pi = \pi_1$ is not unitary on K. By averaging the scalar product over K we can get a scalar product and a Hilbert space H_2 such that π_2 is unitary on K. The traces will be the same in both representations.

Lemma 2 is a special case of the next theorem.

Theorem 3. *Let π_1, π_2 be representations of G on Hilbert spaces. Assume that the K-eigenspaces in each have dimension* 0 *or* 1, *and that π_1, π_2 are infinitesimally isomorphic. Then for every $\psi \in C_c^\infty(G)$ we have*

$$\operatorname{tr} \pi_1(\psi) = \operatorname{tr} \pi_2(\psi).$$

Proof. By the preceding remarks, we may assume that π_1, π_2 are unitary on K. We can then apply VI, §3, Th. 5, where we proved that the coefficient functions are equal. A fortiori, their sums are equal, to

$$\sum_i \int_G \psi(x)\langle \pi(x)u_i, u_i\rangle \, dx,$$

where π is either π_1 or π_2, and the scalar product is that corresponding to π_1 or π_2 respectively.

Theorem 3 implies that we can compute the trace of a unitary representation in any convenient model, not necessarily unitary, which is infinitesimally isomorphic to the given one. In particular, this is useful when dealing with unitary representations having a model coming from an induced representation, when the trace is given as a simple integral of a kernel on the diagonal.

§2. INTEGRAL FORMULAS

Preliminaries

In this section we tabulate various change of variable formulas, and for the convenience of the reader, we recall some elementary facts about manifolds and integration. We assume that the reader is acquainted with the basic information of *Real Analysis*, Chapters XVI, XVII, and the parts of Chapter XVIII giving the general theorems concerning integration on manifolds and the measures arising from differential forms.

In the present section, our manifolds will be either $SL_2(\mathbf{R})$ or products of one-dimensional groups, like N, A, K (so the reals, or the multiplicative group of reals > 0, or the circle group). The tangent space at the origin will be taken as the standard domain of charts by means of the exponential map. If G is a Lie group like the above, and $\mathfrak{g}$ its Lie algebra, then for each $g \in G$ we have a chart in a neighborhood of g given by

$$\exp\colon\ U \longrightarrow G, \qquad X \longmapsto g\exp(X)$$

for X in a small neighborhood of 0 in $\mathfrak{g}$. If G' is another Lie group with Lie algebra $\mathfrak{g}'$, and if $F\colon G' \longrightarrow G$ is a C^∞ mapping (not group homomorphism necessarily), then we have at each point $g' \in G'$ a corresponding differential

$$dF(g')\colon \mathfrak{g}' \longrightarrow \mathfrak{g}$$

which is the tangent linear map of F at g'. Let $\{X_1, \dots, X_d\}$ be a basis of $\mathfrak{g}$ and $\{X_1', \dots, X_d'\}$ a basis of $\mathfrak{g}'$. In the applications, $d = 3$, and G, G' have the same dimension, so we assume this here. The wedge products

$$X_1 \wedge \cdots \wedge X_d \qquad \text{and} \qquad X_1' \wedge \cdots \wedge X_d'$$

form a basis of the one-dimensional spaces $\bigwedge^d \mathfrak{g}$ and $\bigwedge^d \mathfrak{g}'$ respectively. If $Y_1, \dots, Y_d \in \mathfrak{g}$ then

$$Y_1 \wedge \cdots \wedge Y_d = \omega(Y_1, \dots, Y_d) X_1 \wedge \cdots \wedge X_d,$$

where $\omega(Y_1, \dots, Y_d)$ is a real number, and

$$(Y_1, \dots, Y_d) \longmapsto \omega(Y_1, \dots, Y_d)$$

is a differential form. We have a similar situation with Y_j', X_j' replacing Y_j, X_j respectively, giving rise to a form ω'. These forms of course depend on our choice of bases $\{X_j\}$, $\{X_j'\}$ respectively.

Assume, as will be the case, that F is a local C^∞ isomorphism at each point (i.e. is locally differentiably invertible at each point) of an open set of G'. Then

$$\textstyle\bigwedge^d dF(g')\colon \bigwedge^d \mathfrak{g}' \to \bigwedge^d \mathfrak{g}$$

is an isomorphism, and is therefore given as multiplication by a real number, depending on g', say $c(g')$. Thus

$$dF(g')X_1' \wedge \cdots \wedge dF(g')X_d' = c(g')X_1 \wedge \cdots \wedge X_d.$$

By definition, the inverse image of ω by F is given by

$$(F^*\omega)(g') = c(g')\omega'(g').$$

The number $c(g')$ is the "Jacobian" of the transformation F, and if μ, μ' are the positive measures associated with the differential forms ω, ω' respectively, then locally at each point,

$$(F^*\omega)(g') = c(g')\omega'(g').$$

In terms of integrals, this means that in an open set V' where F is a C^∞ isomorphism,

$$\int_{F(V')} f(g)\, d\mu(g) = \int_{V'} f(F(g'))|c(g')|\, d\mu'(g').$$

If F is a covering of degree m from an open set V' in G' onto an open set V in G, then the right-hand side of the above relation has to be divided by m in order to make the relation valid in this case.

The rest of this section is devoted to computing the stretching factor (Jacobian) in four special cases. The effect of $dF(g')$ can be computed by taking Lie derivatives. Let $X' \in \mathfrak{g}'$. We want to determine $dF(g')X'$. It is that vector X having the following property. Let $\varphi \in C_c^\infty(G)$. Then

$$(\mathfrak{L}_X \varphi)(F(g')) = \mathfrak{L}_{X'}(\varphi \circ F)(g').$$

The computations of Jacobians

Throughout the rest of this section, we select as a basis of the Lie algebra $\mathfrak{g}$ of $SL_2(\mathbf{R}) = G$ the elements

$$H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad X^+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X^- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}.$$

Then $H \wedge X^+ \wedge X^-$ is a basis of $\bigwedge^3 \mathfrak{g}$, and there is a unique differential form ω on G, invariant under translations, such that

$$\omega(H, X^+, X^-) = 1.$$

This differential form gives rise to a positive measure μ_ω, and we shall compare other measures with this standard one, thus getting a comparison between the other measures among themselves.

The one-parameter subgroup having H as tangent vector is A, and if we write its elements as

$$a = h_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix},$$

then the measure da on A is that corresponding to Lebesgue measure dt.

Similarly, the one-parameter subgroup having X^+ as tangent vector is N, and if we write its elements as

$$n = n_u = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix},$$

then the measure dn on N is that corresponding to Lebesgue measure du. The situation is similar for the group $\bar{N}$ whose elements are

$$\bar{n}_u = \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix},$$

with measure $d\bar{n}$ corresponding to du.

If $g \in G$ and $Z \in \mathfrak{g}$ we write

$$Z^g = g^{-1}Zg = \mathrm{Ad}(g^{-1})Z.$$

The map $g \mapsto \mathrm{Ad}(g)$ is a representation of G on $\mathfrak{g}$, called, unfortunately, the **adjoint representation**. Observe that

$$g^{-1}\exp(Z)g = \exp(g^{-1}Zg).$$

Hence,

$$\exp(Z)g = g\exp(Z^g) = g\exp(\mathrm{Ad}(g^{-1})Z).$$

This shows how to move a group element across an exponential.

As usual, we let

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

We view W as a distinguished basis for the Lie algebra of K, and the circle is parametrized by $\exp(\theta W)$. The corresponding measure on the circle is $d\theta$.

In the following products $A \times N \times K$ or $K \times A \times K$, the measures on A, N, K are $da = dt$, $dn = du$, $dk = d\theta$ respectively.

INT 1. *Let $F: A \times N \times K \to G$ be $F(a,n,k) = ank$. Then for $\varphi \in C_c^\infty(G)$,*

$$\int_G \varphi(g)\, d\mu_\omega(g) = \int_A \int_N \int_K \varphi(ank)\, da\, dn\, dk.$$

Proof. Let φ be a C^∞ function on G. We view

$$\{H, X^+, W\}$$

as a basis for the tangent space at the origin of the group $A \times N \times K$, and we compute $dF(a,n,k)$ applied to H, X^+, W respectively by using the Lie derivative. We first look at

$$\frac{d}{dt}\bigg|_{t=0} \varphi(a \exp(tH)nk).$$

Then

$$\varphi(a \exp(tH)nk) = \varphi(ank \exp(tH^{nk})).$$

This means that

$$dF(a,n,k)H = \mathrm{Ad}(k^{-1})\,\mathrm{Ad}(n^{-1})H.$$

Similarly,

$$dF(a,n,k)X^+ = \mathrm{Ad}(k^{-1})X^+,$$

$$dF(a,n,k)W = W = X^+ - X^-.$$

But $\mathrm{Ad}(n^{-1})H = H + 2uX^+$ by a direct matrix computation if

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

Hence,

$$\mathrm{Ad}(k)\, dF(a,n,k)\colon \begin{cases} H \mapsto H + 2uX^+ \\ X^+ \mapsto X^+ \\ W \mapsto X^+ - X^-. \end{cases}$$

The map $k \mapsto \bigwedge^3 \mathrm{Ad}(k)$ yields a representation of K into the multiplicative group, and is therefore trivial. Therefore

$$\textstyle\bigwedge^3 dF(a,n,k)\colon H \wedge X^+ \wedge W \mapsto -\, H \wedge X^+ \wedge X^-.$$

This means that locally the pull back of the measure μ_ω is precisely $da\,dn\,dk$. Since F is bijective, our first formula is proved .

Next we consider the measure relative to what is called a Cartan decomposition. Let A^+ consist of all those matrices

$$a = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}$$

with $a > 1$. We consider the map

$$K \times A \times K \to G$$

given by $(k_1, a, k_2) \mapsto k_1 a k_2$. This map is surjective, because the polar decomposition of a matrix x allows us to write uniquely

$$x = sk$$

where s is symmetric positive and $k \in K$. Furthermore, if the eigenvalues of s are distinct, then s is conjugate under K to an element of A. Conjugating with w if necessary transforms such an element of A into an element of A^+. Finally, if $a \neq \pm 1$, then an expression $s = k_1^{-1} a k_1$ for a positive symmetric matrix s, with $a \in A^+$ and $k_1 \in K$, determines k_1 up to ± 1. Hence a decomposition

$$x = k_1 a k_2$$

with $a \in A^+$, $k_1, k_2 \in K$ is uniquely determined up to a factor of ± 1 on k_1, and a corresponding factor ± 1 on k_2.

We call the mapping

$$F \colon K \times A^+ \times K \to G, \qquad (k_1, a, k_2) \mapsto k_1 a k_2,$$

the **Cartan decomposition**. The above remarks show that **F is of degree 2** over its image, which in fact is precisely the complement of K in G. Indeed, if $a \in A^+$, then KaK cannot intersect K (because $A \cap K = \pm 1$), and on the other hand we have seen that every $x \in G$ can be written as $x = k_1 a k_2$ for some $k_1, k_2 \in K$, $a \in A$. In particular, the image of $K \times A^+ \times K$ in G is open and its complement has measure 0.

INT 2. *Let $F \colon K \times A^+ \times K \to G$ be the Cartan decomposition. For any function $\varphi \in C_c(G)$ we have*

$$\int_G \varphi(x)\, d\mu_\omega(x) = \int_K \int_{A^+} \int_K \varphi(k_1 a k_2) \frac{\alpha(a) - \alpha(a^{-1})}{2}\, dk_1\, da\, dk_2.$$

Letting $a = h_t$ and $k_1 = r(\theta_1)$, $k_2 = r(\theta_2)$ we can also write the above relation as

$$\int_G \varphi(x)\, d\mu_\omega(x) = \int_0^{2\pi} \int_0^{\infty} \int_0^{2\pi} \varphi(r(\theta_1) h_t r(\theta_2)) \sinh 2t\, d\theta_1\, dt\, d\theta_2.$$

The integral over t is taken from 0 to ∞ because this is the interval for t parametrizing A^+. Recall that

$$\alpha(a) = e^{2t}.$$

To compute the differential of F we use $W = W_1$, H, $W = W_2$ as basis elements for the Lie algebras of $K = K_1$, A, $K = K_2$ respectively, thus determining the desired measure on $K \times A \times K$. We have to see what the image of $W_1 \wedge H \wedge W_2$ is under $dF(k_1, a, k_2)$. We take as before a function $\varphi \in C_c^\infty(G)$ and compute

$$\varphi(k_1 \exp(tW_1)ak_2), \qquad \varphi(k_1 a \exp(tH)k_2), \qquad \varphi(k_1 a k_2 \exp(tW_2))$$

by moving the exponential term to the right, using the commutation rule with the group element next to it. We then find

$$\mathrm{Ad}(k_2)\, dF(k_1, a, k_2)\colon \begin{cases} W_1 \mapsto \mathrm{Ad}(a^{-1})W = a^{-2}X^+ - a^2X^- \\ H \mapsto H \\ W_2 \mapsto W = X^+ - X^-. \end{cases}$$

Taking the wedge product of the three elements on the right yields

$$(a^{-2} - a^2)H \wedge X^+ \wedge X^-.$$

This means that locally the pull back $F^*(d\mu_\omega)$ is given by

$$(\alpha(a) - \alpha(a^{-1}))dk_1\, da\, dk_2.$$

Since F is a covering of degree 2, we must divide by 2 in order to get the integral over the whole product $K \times A^+ \times K$.

The next two formulas deal with integrals over conjugacy classes. We let G' denote the set of elements in G with distinct eigenvalues, and similarly for A' and K'. We call such elements the **regular elements**.

We first deal with A'. The set A'^G consisting of all elements $g^{-1}ag$ with $a \in A'$ and $g \in G$ is open in G. In fact, suppose $g \in G$ is such that $g^{-1}ag$ also lies in A. Then $g^{-1}ag$ and a have the same eigenvalues, whence

$$g^{-1}ag = a \qquad \text{or} \qquad g^{-1}ag = a^{-1}.$$

In the first case, g centralizes a, and we leave it to the reader to prove that $g \in MA$, where $M = \{\pm 1\}$. In the second case, $gw \in MA$ because conjugation by w sends a to a^{-1}. We interpret this in terms of the mapping

$$F\colon A' \times A\backslash G \to G'$$

given by

$$(a, \dot{g}) \mapsto g^{-1}ag.$$

The above remarks show that **this mapping is of degree 4:** the four elements

$$(a, \pm 1), \qquad (a^{-1}, \pm w)$$

have the same image a in G' under F. The decomposition

$$G = ANK$$

gives us representatives NK for $A \backslash G$, and so we shall consider F as the mapping given in terms of these representatives for the next integral formula.

INT 3. *Let* $F\colon A' \times N \times K \longrightarrow G'$ *be the mapping*

$$(a, n, k) \longmapsto k^{-1}n^{-1}ank.$$

Let $D(a) = \alpha(a)^{1/2} - \alpha(a)^{-1/2}$. *Then for any function* φ *with support in* A'^G *we have*

$$\int_G \varphi(g)\, d\mu_\omega(g) = \frac{1}{4}\int_A \int_N \int_K \varphi(k^{-1}n^{-1}ank)|D(a)|^2\, da\, dn\, dk.$$

Proof. The factor $1/4$ is due to the fact that F has degree 4. There remains to prove that locally, the pull back of the measure $d\mu_\omega$ picks up the factor $|D(a)|^2$. We consider first the expression

$$\varphi(k^{-1}n^{-1}a \exp(tH)nk).$$

Pulling $\exp(tH)$ across nk picks up $\mathrm{Ad}(k^{-1})\,\mathrm{Ad}(n^{-1})$ on H. As before, we multiply all the way through by $\mathrm{Ad}(k^{-1})$. We compute $\mathrm{Ad}(n^{-1})H$ explicitly with 2×2 matrices to find

$$\mathrm{Ad}(k)\, dF(a, n, k)H = \mathrm{Ad}(n^{-1})H = H + 2uX^+$$

if

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

Next we consider

$\varphi(k^{-1} \exp(-tX^+)n^{-1}an \exp(tX^+)k)$

$$= \varphi\big(k^{-1}n^{-1}an \exp(-t\, \mathrm{Ad}(n^{-1})\,\mathrm{Ad}(a^{-1})\,\mathrm{Ad}(n)X^+) \exp(tX^+)k\big).$$

To differentiate this with respect to t and evaluate at $t = 0$ we replace the second occurrence of t with a new variable, say v. Then we use the chain rule, summing the two partial derivatives with respect to t and v respectively, and substituting $t = v = 0$. For the derivative with respect to t, we note that

$$\mathrm{Ad}(n)X^+ = X^+$$

(elements of N commute with each other!), and

$$\mathrm{Ad}(a^{-1})X^+ = a^{-2}X^+.$$

As before, the final occurrence of k on the right can be disregarded provided we multiply our differential by $\mathrm{Ad}(k)$. We then obtain

$$\mathrm{Ad}(k)\, dF(a,n,k)X^+ = (1 - a^{-2})X^+.$$

Finally, we consider

$$\varphi(\exp(-tW)k^{-1}n^{-1}ank \exp(tW))$$
$$= \varphi\big(k^{-1}n^{-1}an \exp(-t\, \mathrm{Ad}(n^{-1})\mathrm{Ad}(a^{-1})\, \mathrm{Ad}(n)W) \exp(tW)k\big),$$

and again use the chain rule, replacing the second occurrence of t by the new variable v. We find

$$\mathrm{Ad}(k)\, dF(a,n,k)W = -\mathrm{Ad}(n^{-1}a^{-1}n)W + W.$$

Ultimately, we shall wedge the three image vectors of H, X^+, W. We observe that the image of X^+ is a scalar multiple of X^+. Therefore we can read the image of $H \bmod X^+$, and we can read the image of $W \bmod H$, X^+. We compute explicitly the matrix $n^{-1}a^{-1}n = g$, and then compute the X^- component of gWg^{-1}. We find:

$$\mathrm{Ad}(k)\, dF(a,n,k)W = (a^2 - 1)X^- \pmod{H, X^+}.$$

Wedging the three image vectors of H, X^+, W with each other yields

$$(1 - a^{-2})(a^2 - 1)H \wedge X^+ \wedge X^-.$$

But

$$(1 - a^{-2})(a^2 - 1) = (a - a^{-1})^2 = |D(a)|^2.$$

This proves our formula.

Let A_- be the coset of A consisting of all elements $-a$ with $a \in A$, and let $m = -1$. Then **INT 3** has a counterpart for the conjugacy set of elements in A_-.

INT 3. *Let* $F\colon A'_- \times N \times K \longrightarrow G'$ *be the mapping*

$$(a, n, k) \longmapsto k^{-1}n^{-1}ank.$$

For any function φ *with support in* A'^G *we have*

$$\int_{A'^G} \varphi(g)\, d\mu_\omega(g) = \frac{1}{4}\int_A \int_N \int_K \varphi(k^{-1}n^{-1}mank)|D(a)|^2\, da\, dn\, dk.$$

This is obvious from **INT 3**.

The final formula deals with $K\backslash G$. In the higher dimensional theory, the analogue of K in the present context is a Cartan subgroup, denoted by B. Here it happens that $B = K$, but we sometimes write B instead of K.

The Cartan decomposition KA^+K gives us unique representatives for the set $(K\backslash G)' = K\backslash KA^+K$, namely

$$(K\backslash G)' \approx A^+K/M$$

where, as before, $M = \{\pm 1\}$. We now consider the mapping

$$F\colon K' \times A^+ \times K \to G'$$

given by

$$(k', a, k) \mapsto k^{-1}a^{-1}k'ak,$$

where K' consists of those elements of K which are $\neq \pm 1$, i.e. the set of regular elements in K. The essentially unique Cartan decomposition shows that if $k' \in K'$ and $g \in G$ is such that

$$g^{-1}k'g \in K,$$

then $g \in K$. It follows at once that **the above map F is of degree 2** (corresponding to the factor ± 1 on the K-component on the right).

If $k = k_\theta = r(\theta)$, we put

$$\boxed{D(k_\theta) = D(\theta) = e^{i\theta} - e^{-i\theta} = 2i \sin \theta.}$$

This is the difference of the eigenvalues of k_θ.

In computing the change of variables formula in the present case, it will be useful to use a complex basis for the Lie algebra which consists of eigenvectors for K. We recall the matrices

$$V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad E^+ = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}, \quad E^- = \begin{pmatrix} 1 & -i \\ -i & -1 \end{pmatrix},$$

so that $\{H, V\}$ form a basis for the symmetric matrices, and

$$E^+ = H + iV \quad \text{and} \quad E^- = H - iV.$$

Furthermore, $\{W\}$ is a basis for the skew-symmetric matrices,

$$[W, E^+] = 2iE^+ \quad \text{and} \quad [W, E^-] = -2iE^-.$$

The following lemma will show that E^+ and E^- are also eigenvectors for $\mathrm{Ad}(k)$, $k \in K$. The lemma is valid on Lie groups; we just look at matrices.

Let Z, $Y \in \mathrm{Mat}_n(\mathbf{C})$ be $n \times n$ complex matrices, and let $g \in GL_n(\mathbf{C})$. By definition,

$$\mathrm{ad}(Z)Y = [Z, Y] \quad \text{and} \quad \mathrm{Ad}(g)Y = gYg^{-1}.$$

Thus $\mathrm{ad}(Z)$ is a linear endomorphism of the vector space of matrices. We can form $\exp \mathrm{ad}(Z)$ by the usual power series,

$$\exp \mathrm{ad}(Z) = I + \mathrm{ad}(Z) + \frac{\mathrm{ad}(Z)^2}{2!} + \cdots .$$

Furthermore, $\exp Z \in GL_n(\mathbf{C})$, and thus $\mathrm{Ad}(\exp Z)$ is defined.

Lemma. *We have* $\mathrm{Ad}(\exp Z) = \exp(\mathrm{ad}\, Z)$.

Proof. Let

$$f(t) = \mathrm{Ad}(\exp(tZ)) \quad \text{and} \quad g(t) = \exp(\mathrm{ad}(tZ)).$$

Then f, g are homomorphisms of $\mathbf{R}$ into $GL(\mathrm{Mat}_n(\mathbf{C}))$ having the same value at 0, namely

$$f(0) = g(0) = I.$$

We shall see that their derivative at 0 is $\mathrm{ad}(Z)$, whence it will follow that $f(t) = g(t)$ for all t as desired. We have

$$\begin{aligned}\mathrm{Ad}(\exp(tZ))Y &= (I + tZ + O(t^2))Y(I - tZ + O(t^2)) \\ &= Y + t[Z, Y] + O(t^2).\end{aligned}$$

This proves that $f'(0) = \mathrm{ad}\, Z$. Furthermore,

$$(\exp(\mathrm{ad}\, tZ))Y = I + t(\mathrm{ad}\, Z)Y + O(t^2).$$

from which we see at once that $g'(0) = \mathrm{ad}\, Z$ also, and the lemma follows.

As an application of the lemma, suppose that Y is an eigenvector of $\mathrm{ad}(Z)$, say

$$\mathrm{ad}(Z)Y = \lambda Y.$$

Then we see that

$$\mathrm{Ad}(\exp Z)Y = e^{\lambda}Y,$$

and therefore Y is also an eigenvector for $\mathrm{Ad}(\exp Z)$.

This will be used when $Y = E^+$ or E^- and $Z = \theta W$. We have

$$k_\theta = \exp(\theta W).$$

Consequently we obtain the formulas

$$\boxed{\mathrm{Ad}(k_\theta)E^+ = e^{2i\theta}E^+ \qquad \text{and} \qquad \mathrm{Ad}(k_\theta)^- = e^{-2i\theta}E^-.}$$

INT 4. *Let $F\colon K' \times A^+ \times K \to G'$ be the map*

$$(k', a, k) \mapsto k^{-1}a^{-1}k'ak.$$

Then for any function φ with support in K'^G we have

$$\int_{K'^G} \varphi(g)\, d\mu_\omega(g)$$

$$= \int_K \int_{A^+} \int_K \varphi(k^{-1}a^{-1}k'ak)|D(k')|^2 \frac{\alpha(a) - \alpha(a^{-1})}{2}\, dk'\, da\, dk$$

$$= \int_0^{2\pi} \int_0^\infty \int_0^{2\pi} \varphi(k_\theta^{-1}h_t^{-1}k_{\theta'}h_tk_\theta)|D(\theta')|^2 \frac{e^{2t} - e^{-2t}}{2}\, d\theta'\, dt\, d\theta.$$

Proof. The proof will follow the same pattern as before. We look at the effect of the differential of F on $W' = W$, H, and W.

For W' we consider $\varphi(k^{-1}a^{-1}k' \exp(tW')ak)$ and have to move $\exp(tW')$ through a and k. This implies that

$$\mathrm{Ad}(a)\, \mathrm{Ad}(k)\, dF(k', a, k)W' = W.$$

We can prove that

$$\mathrm{Ad}(a)\, \mathrm{Ad}(k)\, dF(k', a, k)H = (1 - \mathrm{Ad}(k'^{-1}))H,$$

$$\mathrm{Ad}(a)\, \mathrm{Ad}(k)\, dF(k', a, k)W = (1 - \mathrm{Ad}(k'^{-1}))\, \mathrm{Ad}(a)W,$$

by using the same technique as previously, moving an expression $\exp(-tH)$ or $\exp(-tW)$ through group elements. As in **INT 3**, there are two occurrences of such exponential terms, and we have to use the chain rule in two variables, replacing one occurrence of t by a new variable v. The above values for the differential then follow immediately.

We must compute their wedge product. Since W occurs as the first vector on the right, we use the symmetric elements H, V or E^+, E^- for a complementary basis of the Lie algebra $\mathfrak{g}$. Thus by matrix multiplication,

$$\mathrm{Ad}(a)W = \begin{pmatrix} 0 & a^2 \\ -a^{-2} & 0 \end{pmatrix},$$

and we write this as

$$\mathrm{Ad}(a)W = a^2X^+ - a^{-2}X^- = \frac{a^2 - a^{-2}}{2}V + \frac{a^2 - a^{-2}}{2}W.$$

For purposes of taking our final wedge product, the W-component in this last expression can be omitted, and the wedge product of the images of W', H, W under $\mathrm{Ad}(a)\ \mathrm{Ad}(k)\ dF(k',a,k)$ is

$$W\wedge(1 - \mathrm{Ad}(k'^{-1}))H\wedge(1 - \mathrm{Ad}(k'^{-1}))\frac{a^2 - a^{-2}}{2}V.$$

We use $H = E^+ - iV = E^- + iV$, or in other words

$$H = \frac{E^+ + E^-}{2} \qquad \text{and} \qquad V = \frac{E^+ - E^-}{2i}.$$

Using the fact that E^+ and E^- are eigenvectors for $\mathrm{Ad}(k')$, we see that our wedge product is equal to (writing $k' = \exp(\theta W)$)

$$\frac{1}{2}\frac{a^2 - a^{-2}}{2}(1 - e^{2i\theta})(1 - e^{-2i\theta})W\wedge E^- \wedge E^+.$$

Writing $W = X^+ - X^-$ and using

$$E^+ = H + iV, \qquad E^- = H - iV, \qquad V = X^+ + X^-,$$

we find

$$W\wedge E^-\wedge E^+ = -4iH\wedge X^+\wedge X^-.$$

Furthermore,

$$(1 - e^{2i\theta})(1 - e^{-2i\theta}) = |D(\theta)|^2.$$

Up to a factor of absolute value 1, this gives us precisely the factor

$$2|D(\theta)|^2\frac{\alpha(a) - \alpha(a^{-1})}{2}$$

as the stretching factor in our Jacobian computation, which has been computed for

$$\mathrm{Ad}(a)\ \mathrm{Ad}(k)\ dF(k',a,k).$$

On $\bigwedge^3\mathfrak{g}$ the representation $g \mapsto \bigwedge^3 \mathrm{Ad}(g)$ is a one-dimensional representation of G, and we also know that $\bigwedge^3 \mathrm{Ad}(k)$ operates trivially. Since $G = SL_2(\mathbf{R})$ has no non-trivial continuous homomorphism into the multiplicative group of real numbers, it follows that the triple wedge product of

$\mathrm{Ad}(a)\,\mathrm{Ad}(k)$ is equal to the identity on $\bigwedge^3 \mathfrak{g}$. Hence, the stretching factor which we have found is also the one associated with

$$dF(k', a, k).$$

This proves our last formula **INT 4.**

Remark 1. We normalized dk in the product decompositions so that it is the measure $d\theta$ on the circle parametrized by $e^{i\theta}$. This means that in the formulas above, K has measure 2π. One can divide all the integral expressions above by 2π a posteriori to get relations between integrals. However, in those cases when K occurs twice, for instance in the Cartan decomposition of **INT 2**, and the conjugacy class integration of **INT 4**, this factor of $1/2\pi$ on the right will apply only to one of the integrals over K, and consequently the other integral is left without a factor of 2π. In view of this lack of symmetry, I preferred to state the above formulas with $d\theta$ throughout, and divide by 2π only in specific cases later when such a normalization is warranted in a natural fashion.

Remark 2. One can prove other integral formulas in a systematic way following the above patterns. For instance, we may use the unique decomposition $G = KAN$ and use AN or NA to represent the cosets of $K\backslash G$ for the conjugation map considered in **INT 4**,

$$K' \times K\backslash G \to G'$$

such that

$$(k', \dot{g}) \mapsto g^{-1}k'g.$$

Since it should now be obvious to the reader how to derive this formula trivially when he needs it, and since we won't need it, we shall omit it.

§3. THE TRACE IN THE INDUCED REPRESENTATION

We let s be a complex number, and recall that $H(s)$ is the Hilbert space of functions f on G whose restriction to K is in $L^2(K)$, and such that

$$f(ank) = \rho(a)^{s+1} f(k).$$

Then G operates by right translation, and we saw in Chapter III, §4 that for $\psi \in C_c^\infty(G, K)$ we have, for the corresponding representation π,

$$\pi_s(\psi)f(k') = \int_K q_\psi(k, k') f(k)\, dk$$

where

$$q_\psi(k, k') = \int_A \int_N \psi(k'^{-1}ank)\rho(a)^{s+1}\, da\, dn.$$

We assume that the measure is $dx = da\, dn\, dk$, and that K has measure 1, so that

$$dk_\theta = \frac{d\theta}{2\pi}.$$

Then as we had seen,

$$\operatorname{tr} \pi_s(\psi) = \int_K q_\psi(k, k)\, dk.$$

We define the **Harish transform** as before,

$$\mathbf{H}^A\psi(a) = |D(a)| \int_{A\backslash G} \psi(x^{-1}ax)\, d\dot{x} = \rho(a) \int_N \psi(an)\, dn.$$

Let $\mu = \mu_s$ be the character

$$\mu(a) = \mu_s(a) = \rho(a)^s.$$

Let π_μ denote the representation of G in $H(s)$. Then we obtain

$$(1) \qquad \operatorname{tr} \pi_\mu(\psi) = \int_A \mathbf{H}^A\psi(a)\mu(a)\, da.$$

By V, §2, Th. 1, the Harish transform is invariant under $a \mapsto a^{-1}$. Consequently we also have

$$(2) \qquad \operatorname{tr} \pi_\mu(\psi) = \int_A \mathbf{H}^A\psi(a) \frac{\mu(a) + \mu(a^{-1})}{2}\, da.$$

We recall that G' denotes the set of regular elements (distinct eigenvalues).

Theorem 4. *The trace* $\operatorname{tr} \pi_\mu$ *is a distribution, which can be represented by the function* T_μ*, defined by*:

$$T_\mu(a) = 2\frac{\mu(a) + \mu(a^{-1})}{|D(a)|} \quad \text{if} \quad a \in A',$$

$$T_\mu(x) = 0 \quad \text{if} \quad x \notin A'^G,$$

and $T_\mu(x^{-1}ax) = T_\mu(a)$ *for all* $x \in G$, $a \in A'$. *In other words,*

$$\operatorname{tr} \pi_\mu(\psi) = \int_G \psi(x) T_\mu(x)\, dx.$$

Proof. According to the integral formula **INT 3**, we have

$$\int_{A'^G} \psi(x)\,dx = \frac{1}{4}\int_A \int_{A\backslash G} \psi(x^{-1}ax)|D(a)|^2\,da\,d\dot{x}.$$

$$= \frac{1}{4}\int_A \mathbf{H}^A\psi(a)|D(a)|\,da.$$

There is a function f equal to 0 outside of A'^G, invariant under conjugation by elements of G, and such that

$$f(a) = \frac{\mu(a) + \mu(a^{-1})}{|D(a)|}.$$

Such a function exists because

$$x^{-1}ax \in A \text{ if and only if } x = \pm 1 \text{ or } \pm w \pmod{A},$$

and $waw^{-1} = a^{-1}$. The integral expression for $\mathbf{H}^A\psi$ over the conjugacy classes shows that

$$\mathbf{H}^A(f\psi) = f\mathbf{H}^A(\psi).$$

Replacing ψ by ψf shows that

$$\int_{A'^G} \psi(x)f(x)\,dx = \frac{1}{4}\int_A \mathbf{H}^A\psi(a)[\mu(a) + \mu(a^{-1})]\,da.$$

Comparing with (2) yields our theorem.

The reason for the factor 2 in the trace of the induced representation lies in the fact that we have not separated H^+ and H^-. When we do this, the factor disappears in the following manner.

Let μ be a continuous homomorphism of A into $\mathbf{C}^*$. Let ϵ be a character of the group $M = \{\pm 1\}$. Then the pair (μ, ϵ) defines a character of MA by letting

$$za \mapsto \epsilon(z)\mu(a), \qquad z \in M, a \in A.$$

We can form an induced representation from the subgroup MAN, and the functions in the induced representation space $H(\mu, \epsilon)$ or $H(s, \epsilon)$ are precisely those functions which are even or odd according as ϵ is trivial or non-trivial. The two possible representations are denoted by $\pi_{\mu,+}$ and $\pi_{\mu,-}$. Similarly, the traces are denoted by $T_{\mu,+}$ or T_μ^+ and $T_{\mu,-}$ or T_μ^-.

Corollary. *Let $\pi_{\mu,\epsilon}$ be the representation on $H(\mu, \epsilon)$. Then the trace* tr $\pi_{\mu,\epsilon}$ *is* *a distribution, which can be represented by the function $T_{\mu,\epsilon}$, defined by*:

$$T_{\mu,\epsilon}(za) = \epsilon(z)\frac{\mu(a) + \mu(a^{-1})}{|D(a)|} \quad \text{if} \quad z \in M, a \in A',$$

$$T_{\mu,\epsilon}(x) = 0 \quad \text{if} \quad x \notin \pm A'^G,$$

and $T_{\mu,\epsilon}$ is invariant under conjugation by elements of G.

Proof. For definiteness, consider the case when ϵ is the non-trivial character. Let $f(x)$ be as in Theorem 4. Then

$$\begin{aligned}\operatorname{tr} \pi_\mu(\psi^-) &= \int_{A'^G} \frac{\psi(x) - \psi(-x)}{2} 2f(x)\,dx \\ &= \int_{A'^G} \psi(x)f(x)\,dx - \int_{A'^G} \psi(-x)f(x)\,dx \\ &= \int_{(MA')^G} \psi(x)\epsilon(x)f(x)\,dx.\end{aligned}$$

This proves our assertion in the present case. The other case is proved the same way.

§4. THE TRACE IN THE DISCRETE SERIES

Let m be an integer $\geqslant 2$ and let $s = m - 1$. We have seen in Chapter VI, §5, that $H(m - 1)$ decomposes into certain irreducible subspaces, and we are interested in the spaces

$$H^{(m)} = \widehat{\bigoplus_{\substack{n \geqslant m \\ n \equiv m}}} H_n, \qquad H^{(-m)} = \widehat{\bigoplus_{\substack{n \leqslant -m \\ n \equiv m}}} H_n$$

$$V(m - 1) = H(m - 1, \epsilon) \bmod [H^{(m)} + H^{(-m)}], \text{ where } \quad \epsilon(-1) = (-1)^m.$$

It was clear from the derived representation that $H^{(m)}$, $H^{(-m)}$, $V(m - 1)$ are irreducible, and $V(m - 1)$ is finite dimensional. They look like Fig. 1.

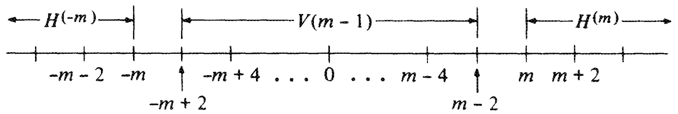

Figure 1

We shall compute the trace in the direct sum

$$H^{(m)} + H^{(-m)}$$

by using the trace in $H(m-1)$ found in the preceding section, and subtracting the trace in the finite dimensional space $V(m-1)$.

We let

$$h_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \qquad z = \pm 1.$$

Lemma 1. *Let n be an integer $\geqq 1$, let $m = n + 1$, and let*

$$\epsilon(-1) = (-1)^m.$$

The trace of the representation in $H(n, \epsilon)$ is a distribution represented by the function invariant under conjugation such that

$$T(zh_t) = \epsilon(z)\frac{e^{nt} + e^{-nt}}{|e^t - e^{-t}|},$$

$$T(x) = 0 \qquad \text{if} \quad x \text{ is not conjugate to some } \pm h_t.$$

Proof. This is a special case of the trace found in the last section for an arbitrary induced representation.

Lemma 2. *For each pair of integers $p, q \geqq 0$ such that $p + q = n - 1$, let $f_{p,q}$ be the function of*

$$x = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

such that $f_{p,q}(x) = c^p d^q$. Then $f_{p,q} \in H(-n, \epsilon)$. The functions $f_{p,q}$ form a basis of a finite dimensional irreducible space of dimension $n = m - 1$,

$$V'(n) = V(-n) = V(-m+1).$$

Let ρ' be the representation of G in $V'(n)$. Then

$$\operatorname{tr} \rho'(zh_t) = \epsilon(z)\frac{e^{nt} - e^{-nt}}{e^t - e^{-t}}.$$

Proof. The functions $f_{p,q}$ obviously lie in $H(-n)$, they are linearly independent, and matrix multiplication shows that the space they generate is stable under right translation by G. Each function is an eigenvector of $\rho(h_t)$, and in fact

$$\rho(h_t)f_{p,q} = e^{(p-q)t}f_{p,q}.$$

Hence the trace is trivially computed to be that stated in the lemma.

We know from VI, §2, Th. 4, that two finite dimensional irreducible representations of the same dimension are infinitesimally isomorphic, or even isomorphic by the corollary of that theorem, although for traces to be equal, an infinitesimal isomorphism suffices. The representation of Lemma 2 or $V(m-1)$ can therefore be used as a model in a given dimension.

***Lemma* 3.** *Let V be the finite dimensional irreducible representation of dimension $n = m - 1$, and let ρ be the representation in V. Then*

$$\operatorname{tr} \rho(k_\theta) = \frac{e^{in\theta} - e^{-in\theta}}{e^{i\theta} - e^{-i\theta}}.$$

Proof. We know that V is generated by the K-eigenvectors having eigenvalues

$$e^{i(-m+2)\theta}, e^{i(-m+4)\theta}, \ldots, e^{i(m-2)\theta}.$$

We find the trace on elements of K by summing these eigenvalues, i.e. summing part of a geometric series, which obviously gives the stated result.

***Lemma* 4.** *Let n be an integer $\geqslant 1$ and let $m = n + 1$. Let*

$$\epsilon(-1) = (-1)^m.$$

Then the trace of the representation on $H^{(m)} + H^{(-m)}$ is a distribution represented by the function:

$$T(zh_t) = \epsilon(z)\frac{2e^{-|nt|}}{|e^t - e^{-t}|},$$

$$T(k_\theta) = \frac{-(e^{in\theta} - e^{-in\theta})}{e^{i\theta} - e^{-i\theta}},$$

$$T(x) = 0 \qquad \text{if} \quad x \text{ is not conjugate to some } \pm h_t \text{ or } k_\theta.$$

Proof. This is obvious from Lemmas 1, 2, 3, subtracting the values found in Lemmas 2, 3 from the values found in Lemma 1.

We have obtained the trace in the sum $H^{(m)} + H^{(-m)}$. We must now separate these two irreducible pieces. The final theorem is:

Theorem 5. *Let n be an integer $\neq 0$ and let $m = |n| + 1$. Let $z = \pm 1$, and $\epsilon(-1) = (-1)^m$. Let σ_n be the representation in $H^{(n+1)}$ if $n > 0$ and $H^{(-n-1)}$ if $n < 0$. Then the trace* tr σ_n *is a distribution represented by the function S_n invariant under conjugation such that*

$$S_n(k_\theta) = \frac{-(\operatorname{sign} n)e^{in\theta}}{e^{i\theta} - e^{-i\theta}},$$

$$S_n(zh_t) = \epsilon(z)\frac{e^{-|nt|}}{|e^t - e^{-t}|},$$

and zero otherwise.

Proof. The proof is a little elaborate and is due to Harish–Chandra. The main difficulty is to eliminate the possibility that there is a contribution from the singular set (the complement of G' in G) which may cancel in the sum of the discrete series with positive and negative weight. I shall omit it. (Perhaps I shall include it in a second volume dealing with other matters involving differential equations and distributions on $SL_2(\mathbf{R})$). It is an easy exercise to see that on the regular set, i.e. on functions with compact support in G', the trace is given by the expected function. This is a simple consequence of the integral formulas, and the expressions for the Harish transforms on A and K found in the next chapter, §2. For the applications to Plancherel's formula this is not important because only the sum of the two discrete series occurs. However, for further investigations on $SL_2(\mathbf{R})$, it is an important fact.

Also note that the integral formulas show that the function representing the trace is locally L^1.

§5. RELATION BETWEEN THE HARISH TRANSFORMS ON A AND K

For $\psi \in C_c^\infty(G, K)$ we had defined the Harish transform on A by

$$\text{(1)} \qquad \mathbf{H}^A\psi(a) = |D(a)| \int_{A\backslash G} \psi(x^{-1}ax)\, d\dot{x},$$

where $d\dot{x}$ is the measure on $A\backslash G$ such that $dx = da\, d\dot{x}$, and dx is normalized as

$$dx = da\, dn\, dk, \qquad dk_\theta = \frac{d\theta}{2\pi}.$$

Thus by definition, for any function f,

$$\int_{A\backslash G} \int_A f(a\dot{x})\, da\, d\dot{x} = \int_G f(x)\, dx.$$

In view of **INT 3**, §2, we have

(2) $$\int_{A'G} \psi(x)\,dx = \frac{1}{4}\int_A \mathbf{H}^A\psi(a)|D(a)|\,da.$$

Analogously, we define the **Harish transform** on $K' = K - \{\pm 1\}$ by

(3) $$\mathbf{H}^K\psi(k) = D(k)\int_{K\backslash G} \psi(x^{-1}kx)\,d\dot{x},$$

where $d\dot{x}$ is now the measure on $K\backslash G$ such that

$$\int_{K\backslash G}\int_K f(k\dot{x})\,dk\,d\dot{x} = \int_G f(x)\,dx,$$

and

$$D(k_\theta) = e^{i\theta} - e^{-i\theta}.$$

Then by **INT 2**, we find that $K\backslash G$ is represented uniquely by $A^+K/\pm 1$, and therefore that

(4) $$\mathbf{H}^K\psi(k) = D(k)\int_{A^+} \psi(a^{-1}ka)\frac{\alpha(a) - \alpha(a^{-1})}{2}\,da.$$

By **INT 4** we then obtain

(5) $$\int_{K'G} \psi(x)\,dx = \int_0^{2\pi} \mathbf{H}^K\psi(k_\theta)\,\overline{D(k_\theta)}\,d\theta.$$

[In **INT 4**, dividing both sides by 2π still leaves one integration over K where K has measure 2π.]

For each integer $n \neq 0$ let S_n be the function invariant under conjugation by G, given by

$$S_n(k_\theta) = \frac{-(\operatorname{sign} n)e^{in\theta}}{e^{i\theta} - e^{-i\theta}},$$

$$S_n(zh_t) = \epsilon(z)\frac{e^{-|nt|}}{|e^t - e^{-t}|},$$

and zero otherwise. As before, $z = \pm 1$ and $\epsilon(z) = z^{n+1}$. We view S_n again as a functional, so that for any function f,

$$S_n(f) = \int_G f(x)S_n(x)\,dx.$$

Theorem 6. *We have the relation*

$$S_n(\psi) = -(\operatorname{sign} n)\int_0^{2\pi} \mathbf{H}^K\psi(k_\theta)e^{in\theta}\,d\theta$$
$$+\frac{1}{2}\int_0^{\infty}\left[\mathbf{H}^A\psi(h_t) + (-1)^{n+1}\mathbf{H}^A\psi(-h_t)\right]e^{-|nt|}\,dt.$$

Proof. We compute $S_n(\psi)$ by integrating over K'^G and $\pm A'^G$ respectively. Let us start with K'^G. We have

$$\int_{K'^G} \psi(x)(-\operatorname{sign} n)\frac{e^{in\theta(x)}}{e^{i\theta(x)} - e^{-i\theta(x)}}\,dx$$

if x is conjugate to $k_{\theta(x)}$. We now use **INT 4** applied to the function under the integral sign, and obtain from (4),

$$-\operatorname{sign} n\int_0^{2\pi} \mathbf{H}^K\psi(k_\theta)e^{in\theta}\,d\theta,$$

which is the first term on the right of our relation.

For the second term, if x is conjugate to $\pm h_{t(x)}$, we have

$$\int_{\pm A'^G}\psi(x)\epsilon(x)\frac{e^{-|nt(x)|}}{|e^{t(x)} - e^{-t(x)}|}\,dx$$
$$= \int_{A'^G}\left[\psi(x) + (-1)^{n+1}\psi(-x)\right]\frac{e^{-|nt(x)|}}{|e^{t(x)} - e^{-t(x)}|}\,dx$$
$$= \frac{1}{4}\int_{-\infty}^{\infty}\left[\mathbf{H}^A\psi(h_t) + (-1)^{n+1}\mathbf{H}^A\psi(-h_t)\right]e^{-|nt|}\,dt$$

whence the theorem follows at once from the invariance of the Harish transform on A under $a \mapsto a^{-1}$.

APPENDIX. GENERAL FACTS ABOUT TRACES

Polar decomposition

Let H be a Hilbert space with countable Hilbert basis and let A be an operator on H. [All operators in this appendix are assumed bounded.] Then A^*A is symmetric positive and has a unique symmetric positive square root, denoted by $P_A = (A^*A)^{1/2}$. We can define a linear map $U = U_A$ on Im P_A by the formula

$$U(A^*A)^{1/2}v = Av, \qquad \text{for } v \in H.$$

To show this is well defined, it suffices to prove that if $(A^*A)^{1/2}v = 0$, then $Av = 0$. But under the stated assumption, we have

$$0 = |(A^*A)^{1/2}v|^2 = \langle A^*Av, v\rangle = \langle Av, Av\rangle,$$

so what we want is true. It is then immediate that

$$U\colon \operatorname{Im} P_A \to \operatorname{Im} A$$

is a unitary map, which can therefore be extended by continuity to the closure of Im P_A. We define U to be 0 on the orthogonal complement of Im P_A (such an operator is often called a **partial isometry**). Then we have the obvious formulas

$$U^*U = I \quad \text{on} \quad \operatorname{Im} P_A \qquad \text{and} \qquad UU^* = I \quad \text{on} \quad \operatorname{Im} A.$$

The decomposition

$$A = UP$$

into a partial isometry U (relative to Im P) and a positive operator P is unique. Indeed, if $A = WQ$, then $A^* = QW^*$ and

$$P^2 = A^*A = QW^*WQ = Q^2$$

whence $P = Q$ because the symmetric positive square root of a positive operator is uniquely determined (spectral theorem!). The above decomposition is called the **polar decomposition** of A. The polar decomposition of A^* is easily obtained in terms of that for A, namely

$$A^* = U_{A^*}P_{A^*}, \qquad U_{A^*} = U^*, \qquad P_{A^*} = UP_AU^*.$$

To see this, note that UPU^* is positive, and

$$(UPU^*)^2 = UPU^*UPU^* = UP^2U^* = AA^*.$$

This gives us P_{A^*}, and the expression for U_{A^*} follows at once.

Hilbert–Schmidt operators

An operator A is called **Hilbert–Schmidt** if for some orthonormal basis $\{u_i\}$ we have

$$\sum |Au_i|^2 < \infty.$$

The same then holds for any other orthonormal basis $\{v_j\}$. Indeed, note that for any vector w we have $|w|^2 = \sum |\langle w, v_j\rangle|^2$, so that

$$\sum_i |Au_i|^2 = \sum_{i,j} |\langle Au_i, v_j\rangle|^2 = \sum_{i,j} |\langle u_i, A^*v_j\rangle|^2 = \sum_j |A^*v_j|^2.$$

For Hilbert–Schmidt operators A and B, we define their scalar product

$$\langle A, B\rangle = \sum \langle Au_i, Bu_i\rangle$$

with some orthonormal basis $\{u_i\}$. This sum is convergent (i.e. absolutely convergent), as we see by the Schwarz inequality (applied twice!),

$$\sum |\langle Au_i, Bu_i\rangle| \leqslant \sum |Au_i||Bu_i|.$$

Therefore B^*A is Hilbert–Schmidt, and the scalar product is independent of the choice of $\{u_i\}$. The corresponding norm is denoted by

$$N_2(A) = \|A\|_2,$$

and we have

$$\|A\|_2^2 = \sum |Au_i|^2.$$

The Hilbert–Schmidt operators form a normed vector space under this L^2-norm, denoted sometimes by $L^2(H)$. For the following further properties, A, B are assumed to be Hilbert–Schmidt, and X denotes an arbitrary operator.

HS 1. $$\|A^*\|_2 = \|A\|_2.$$

HS 2. *XA and AX are Hilbert–Schmidt, and*

$$\|XA\|_2 \leqslant |X|\|A\|_2, \qquad \|AX\|_2 \leqslant |X|\|A\|_2.$$

HS 3. *A Hilbert–Schmidt operator is compact.*

Before going any further, we prove the above. Note that **HS 1** follows from the identity already proved, where we can interchange A and A^*.

Property **HS 2** is due to the obvious inequality $|XAu_i| \leqslant |X||Au_i|$, and the fact that $(AX)^* = X^*A^*$. To prove the compactness of **HS 3**, there exists N such that

$$\sum_{i>N} |Au_i|^2 < \epsilon.$$

Let P_N be the projection on the space spanned by $u_1, \ldots, u_N$. Then the above inequality can be written

$$\sum_{i=N+1}^{\infty} |(A - AP_N)u_i|^2 = \sum_{N+1}^{\infty} |Au_i|^2 < \epsilon.$$

For any operator T we have $|T| \leqslant \sum |Tu_i|^2$. This proves that A can be approximated uniformly by operators with finite dimensional image, and hence that A is compact.

The polar form of the function $A \longmapsto \|A\|_2^2$ is given by

HS 4. $$\|A + B\|_2^2 - \|A\|_2^2 - \|B\|_2^2 = 2\,\mathrm{Re}\langle A, B\rangle.$$

This is immediate from

$$\langle (A + B)u_i, (A + B)u_i\rangle = \langle Au_i, Au_i\rangle + 2\,\mathrm{Re}\langle Au_i, Bu_i\rangle + \langle Bu_i, Bu_i\rangle.$$

Since $\|A\|_2 = \|A^*\|_2$, we get

HS 5. $$\mathrm{Re}\sum \langle Au_i, Bu_i\rangle = \mathrm{Re}\sum \langle A^*u_i, B^*u_i\rangle$$

HS 6. $$\langle A^*, B^*\rangle = \langle A, B\rangle.$$

HS 7. $\langle XA, B\rangle = \langle A, X^*B\rangle$ *and* $\langle AX, B\rangle = \langle A, BX^*\rangle$.

HS 5 and **HS 6** are clear from **HS 4**. For **HS 7**, we have

$$\langle AX, B\rangle = \overline{\langle X^*A^*, B^*\rangle} = \overline{\langle A^*, XB^*\rangle} = \langle A, BX^*\rangle.$$

The other part of **HS 7** is obtained by starring each term in this last identity.

Trace class operators

An operator A will be said to be of **trace class** if it is the product of two Hilbert–Schmidt operators. We have to choose where to write stars, and in expressing A as such a product, we write $A = B^*C$ where B, C are Hilbert–Schmidt, in order to avoid stars elsewhere. This being the case, we can define the **trace** of A,

$$\mathrm{tr}(A) = \sum \langle Au_i, u_i\rangle = \sum \langle Cu_i, Bu_i\rangle = \langle C, B\rangle.$$

The first sum over i shows that the trace is independent of the choice of B, C. We have trivially

TR 1. $$|\mathrm{tr}(A)| \leqslant \|B\|_2\|C\|_2.$$

TR 2. *If A is of trace class, so are AX and XA, and we have*

$$\mathrm{tr}(AX) = \mathrm{tr}(XA).$$

Indeed, AX and XA are of trace class by **HS 2**. Furthermore,

$$\begin{aligned}\mathrm{tr}(AX) &= \mathrm{tr}(B^*CX) = \langle CX, B\rangle = \langle C, BX^*\rangle = \mathrm{tr}(BX^*)^*C \\ &= \mathrm{tr}(XB^*C) = \mathrm{tr}(XA).\end{aligned}$$

We return to the polar decomposition

$$A = U_A P_A = UP,$$

where U is a partial isometry and P is symmetric positive. We call P the **absolute value** of A, sometimes written $P = \mathrm{Abs}(A)$. Since $P = U^*A$, we see that

TR 3. *A is of trace class if and only if P_A is of trace class, and*

$$\mathrm{tr}(P_A) = \mathrm{tr}(P_{A^*}).$$

This last identity is due to

$$\mathrm{tr}(P_A) = \mathrm{tr}(U^*A) = \mathrm{tr}(AU^*) = \mathrm{tr}(UA^*) = \mathrm{tr}(P_{A^*}).$$

It is not true in general, according to our definition, that A is of trace class if and only if the sum

$$\sum |\langle Au_i, u_i\rangle|$$

converges. On the other hand, we do have:

TR 4. *Let P be a symmetric positive operator. Then P is of trace class if and only if $\sum \langle Pu_i, u_i\rangle$ converges.*

The proof is clear, using $P^{1/2}$. In particular, A is of trace class if and only if its absolute value is of trace class.

If A is an operator of trace class, we define

$$N_1(A) = \|A\|_1 = \mathrm{tr}\, P_A = \|P_A^{1/2}\|_2^2.$$

TR 5. *The operators of trace class form a vector space, the function $A \longmapsto \|A\|_1$ is a norm, satisfying $\|A\|_1 = \|A^*\|_1$.*

Proof. Write $P_{A+B} = U^*(A + B)$ where U is a partial isometry. Assume that both A, B are of trace class. The sum $A + B$ is of trace class if and only if

$$\sum \langle P_{A+B}u_i, u_i\rangle < \infty.$$

However, this sum is equal to

$$\sum \langle U^*Au_i, u_i\rangle + \sum \langle U^*Bu_i, u_i\rangle = \mathrm{tr}(U^*A) + \mathrm{tr}(U^*B)$$
$$\leqslant \|A\|_1 + \|B\|_1.$$

This proves both that operators of trace class form a vector space, and that

$\| \ \|_1$ is a norm (the scalar homogeneity property is obvious). Furthermore, $\|A\|_1 = \|A^*\|_1$ is merely **TR 3**.

TR 6. *If A is of trace class, then so are XA, AX, and we have*

$$\|XA\|_1 \leqslant |X| \|A\|_1 \quad \textit{and} \quad \|AX\|_1 \leqslant |X| \|A\|_1.$$

Proof. It suffices to consider XA, since $AX = (X^*A^*)^*$. There is a partial isometry V such that $P_{XA} = V^*XA$, and letting $A = UP_A$, we have

$$P_{XA} = V^*XA = V^*XUP_A = YP_A, \quad \text{with} \quad Y = V^*XU.$$

Then $|Y| \leqslant |X|$. Furthermore,

$$\begin{aligned} \|XA\|_1 &= \sum \langle YP_A u_i, u_i \rangle = \langle P_A^{1/2}, P_A^{1/2} Y^* \rangle \\ &\leqslant \|P_A^{1/2}\|_2 \, \|P_A^{1/2} Y^*\|_2 \\ &\leqslant \|P_A^{1/2}\|_2^2 \, |Y^*| \\ &\leqslant \|A\|_1 \, |X|, \end{aligned}$$

as desired.

TR 7. *If A is of trace class, then* $|\text{tr } A| \leqslant \|A\|_1$.

Proof. Write $A = UP_A$. Then

$$\begin{aligned} |\text{tr } A| &= \left| \sum \langle P_A^{1/2} u_i, P_A^{1/2} U^* u_i \rangle \right| \\ &= |\langle P_A^{1/2}, P_A^{1/2} U^* \rangle| \\ &\leqslant \|P_A^{1/2}\|_2 \, \|P_A^{1/2} U^*\|_2 \\ &\leqslant \|P_A^{1/2}\|_2^2 = \text{tr } P_A. \end{aligned}$$

Theorem 7. *Let $\{T_n\}$ be a sequence of operators on H, converging weakly to an operator T. In other words, for each $v, w \in H$, $\langle T_n v, w \rangle \longrightarrow \langle Tv, w \rangle$. Let A be of trace class. Then*

$$\text{tr}(TA) = \lim_{n \to \infty} \text{tr}(T_n A),$$

and similarly on the other side.

Proof. Assume first that $A = P$ is positive symmetric. Since A is compact (because A is Hilbert–Schmidt and **HS 3**), there is an orthonormal basis of H

consisting of eigenvectors, say $\{u_i\}$, with $Au_i = c_i u_i$. For fixed $v, w \in H$ the set $\{\langle T_n v, w\rangle\}$ is bounded. Viewing v as fixed and w variable, we see from the uniform boundedness theorem that the set $\{T_n u\}$ is bounded. Again the uniform boundedness theorem shows that the norms $|T_n|$ are bounded, say by a number N. Then by the absolute convergence

$$\sum_i |c_i| < \infty$$

we get

$$\begin{aligned}\lim_{n\to\infty} \operatorname{tr}(T_n A) &= \lim_{n\to\infty} \sum_i \langle T_n A u_i, u_i\rangle = \lim_{n\to\infty} \sum_i c_i \langle T_n u_i, u_i\rangle \\ &= \sum_i \lim_{n\to\infty} c_i \langle T_n u_i, u_i\rangle \\ &= \operatorname{tr}(TA),\end{aligned}$$

which proves the assertion in the present case. In general, write the polar decomposition $A = UP$ where U is a partial isometry and P is symmetric positive. Then $TA = (TU)P$, and $T_n U$ converges weakly to TU, so the first part of the proof applies to give the general case also.

Corollary. *Let A be an operator of trace class in a Hilbert space H_1. Let $T\colon H_1 \to H_2$ be a topological linear isomorphism between H_1 and another Hilbert space H_2. Then*

$$\operatorname{tr} TAT^{-1} = \operatorname{tr} A.$$

Proof. The assertion is true for finite dimensional spaces. Let P_n be the projection on the space generated by $u_1, \ldots, u_n$ (assuming that $\{u_1, u_2, \ldots\}$ is an orthonormal basis). Then

$$\operatorname{tr} A = \lim \operatorname{tr} P_n A P_n \qquad \text{and} \qquad \operatorname{tr} TAT^{-1} = \lim \operatorname{tr} TP_n A P_n T^{-1}.$$

The corollary follows from the finite dimensional case.

The corollary shows that the trace is independent of the choice of positive definite scalar product in an equivalence class, i.e. defining equivalent norms.

Remark. Define $L^0(H)$ to be the Banach space of compact operators. It can be shown fairly easily that the pairing

$$(A, K) \mapsto \operatorname{tr}(AK) \qquad \text{of} \qquad L^1(H) \times L^0(H) \to \mathbf{C}$$

induces a norm preserving isomorphism of $L^1(H)$ onto the dual space of $L^0(H)$. We won't need this in the present book, and leave it as an exercise.

VIII The Plancherel Formula

We shall put together the facts we have learned about traces in order to prove the Plancherel formula, giving an expansion of a function in terms of its characters. The proof is due to Harish-Chandra [H-C 6]. It consists in expanding out the Fourier series of the Harish transform $\mathbf{H}^K\psi(\theta)$, using the relation between the trace of the discrete series and the Harish transform on A given in Theorem 6 of the preceding chapter, and then performing a Fourier transform on some of the terms to get the final formula. It turns out that $\mathbf{H}^K\psi(k_\theta)$ is not continuous, the discontinuity occurring at those elements of k which also lie in A, i.e. at ± 1. The first calculus lemma serves to determine the jumps, and also shows that the derivative $(\mathbf{H}^K\psi)'(k_\theta)$ is continuous at those points, and that the Fourier series converges for the derivative at those points. This gives us the value $\psi(1)$, in terms of a series involving traces in the discrete and principal series.

Pukanszky [Pu] gives the Plancherel formula for the universal covering group of $SL_2(\mathbf{R})$. For the p-adic case, cf. Gelfand *et al*, [Ge, Gr] and Sally–Shalika [Sa, Sh 2]. For connections with special classical functions, see Vilenkin [Vi].

Bargmann [Ba] gave various completeness relations, without ever stating exactly the Plancherel formula. He realized the connection with certain asymptotic expansions (cf. Chapter V, §5). It seems that the function $c(s)$ which appears in such asymptotic expansions also determines the Plancherel measure, and that this phenomenon, relating an asymptotic expansion with eigenfunction expansion, has long been known, more or less explicitly, in connection with second order linear differential equations. An exposition of the Plancherel formula for $SL_2(\mathbf{R})$ following the general pattern of eigenfunction expansions for a second order operator, along Bargmann's lines, is given in Vilenkin [Vi], pp. 336–337, but not carried out in detail. Vilenkin merely refers to the 2-volumes books of Titchmarsh and Levitan, without being more specific than saying: "Employing the usual techniques of expansion in eigenfunctions of selfadjoint operators, we get the following result." (The Plancherel Formula). Each inexperienced reader has to work out the details for himself.

The original Harish proof for the Plancherel formula takes an entirely different approach, but more recent papers of his have returned to the asymptotic expansion point of view. The estimates required on arbitrary groups are of considerable difficulty. Cf. his papers [H-C 7], pp. 576 and 582; [H-C 3], p. 71. See also Knapp–Stein [Kn, St] for other connections.

The last chapter of the present book treats the spectral decomposition of $\Gamma\backslash G$ by a method starting from the asymptotic expansion, and then making a perturbation to get the precise spectral decomposition relation.

§1. A CALCULUS LEMMA

Lemma. *Let* $F \in C_c^\infty(\mathbf{R}^2)$. *Let*

$$g(\theta) = \int_0^\infty F(\theta e^t, \theta e^{-t})(e^t - e^{-t})\, dt.$$

Then:

i) $\displaystyle\lim_{\theta\to 0+} \theta g(\theta) = \int_0^\infty F(u, 0)\, du;$

ii) $\displaystyle\lim_{\theta\to 0-} \theta g(\theta) = -\int_{-\infty}^0 F(u, 0)\, du;$

iii) $\displaystyle\lim_{\theta\to 0} \frac{d}{d\theta}(\theta g(\theta)) = -2F(0, 0).$

iv) *There exist* $a, b > 0$ *such that for* $0 < \theta \leqslant 1$,

$$\left|\frac{d}{d\theta}(\theta g(\theta)) + 2F(0, 0)\right| \leqslant a|\theta| + b|\theta \log \theta|.$$

Proof. First we note that trivially

$$\theta \int_0^\infty F(\theta e^t, \theta e^{-t})e^{-t}\, dt \to 0 \quad \text{as} \quad \theta \to 0. \tag{1}$$

To get (i), we consider the other integral

$$\int_0^\infty \theta F(\theta e^t, \theta e^{-t})e^t\, dt$$

with $\theta > 0$. Change variables, let $u = \theta e^t$, $du = \theta e^t\,dt$. Then

$$e^{-t} = \theta/u,$$

and the integral transforms to

$$\int_\theta^\infty F(u, \theta^2/u)\,du.$$

As $\theta \longrightarrow 0$, the integral approaches the desired limit of (i). The limit of (ii) is done in a similar way. To prove (iii), we have

$$\begin{aligned}\frac{d}{d\theta}(\theta g(\theta)) &= g(\theta) + \theta g'(\theta)\\ &= \int_0^\infty F(\theta e^t, \theta e^{-t})e^t\,dt - \int_0^\infty F(\theta e^t, \theta e^{-t})e^{-t}\,dt\\ &\quad + \theta\int_0^\infty [D_1F(\theta e^t, \theta e^{-t})e^t + D_2F(\theta e^t, \theta e^{-t})e^{-t}](e^t - e^{-t})\,dt.\end{aligned}$$

The second integral goes to $-F(0, 0)$ as θ goes to 0. Any integral having a factor of θ in front and no factor e^t inside tends to 0 as θ tends to 0. So we have to consider the integrals of terms which are not of these two types. We have

$$\begin{aligned}\frac{d}{dt}[e^tF(\theta e^t, \theta e^{-t})] = e^tF(\theta e^t, \theta e^{-t}) &+ \theta e^{2t}D_1F(\theta e^t, \theta e^{-t})\\ &- \theta D_2F(\theta e^t, \theta e^{-t}).\end{aligned}$$

Plugging this exact expression into the integral immediately shows that the remaining integrals give the appropriate contribution $-F(0, 0)$ after making the proper cancellations. To estimate the derivative as stated in (iv), let T be such that $D_2F(x, y) = 0$ if $|x| \geqslant T$ or $|y| \geqslant T$. If $|\theta e^t| \leqslant T$, then $e^t \leqslant T/|\theta|$, i.e.

$$t \leqslant \log T/|\theta|.$$

Hence we get an estimate of type

$$\left|\int_0^\infty D_2F(\theta e^t, \theta e^{-t})\,dt\right| \leqslant \|D_2F\|\,|\log T/|\theta||,$$

and similarly for the integral of D_1F. From this, (iv) is obvious.

The estimate of (iv) is intended to be used to prove the convergence of a certain Fourier series.

Remark. In taking the various limits of the lemma, we can replace θ by $\sin\theta$ since these two functions are equal up to order 3, which does not affect our limits.

§2. THE HARISH TRANSFORMS DISCONTINUITIES

Write an element x of $G = SL_2(\mathbf{R})$ as

$$x = ank = h_t n_u k_\theta,$$

where we use the parametrizations

$$a = h_t = \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}, \qquad n = n_u = \begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}, \qquad k_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

We let the measure on G be

$$\boxed{dx = da\,dn\,dk = dt\,du\,\frac{d\theta}{2\pi},}$$

so that ***K* has measure 1**. We let

$$D_K(\theta) = e^{i\theta} - e^{-i\theta} = 2i\sin\theta,$$

and define as before the **Harish transform** on K, for $\theta \neq 0, \pi$ and $f \in C_c^\infty(G,K)$ by

(1) $$\mathbf{H}^K f(k_\theta) = \mathbf{H}^K f(\theta) = D_K(\theta)\int_0^\infty f(h_t^{-1}k_\theta h_t)\,\frac{(e^{2t} - e^{-2t})}{2}\,dt.$$

Up to a constant factor, this is precisely

$$D_K(k)\int_{K\backslash G} f(g^{-1}kg)\,d\dot{g}.$$

We want to put the expression for the Harish transform on K into a form which fits our calculus lemma. For this we write our element $a^{-1}ka$ as an exponential,

$$h_t^{-1}k_\theta h_t = h_t^{-1}\exp(\theta W)h_t = \exp(h_t^{-1}\theta W h_t),$$

and note that

$$\begin{pmatrix} e^{-t} & 0 \\ 0 & e^t \end{pmatrix}\begin{pmatrix} 0 & \theta \\ -\theta & 0 \end{pmatrix}\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix} = \begin{pmatrix} 0 & \theta e^{-2t} \\ -\theta e^{2t} & 0 \end{pmatrix}.$$

Change variables, replacing $2t$ by t. The limits of integration do not change, and we get

$$\text{(2)} \qquad \mathbf{H}^K f(\theta) = \tfrac{1}{4}\, D_K(\theta) \int_0^\infty F(\theta e^t, \theta e^{-t})(e^t - e^{-t})\, dt,$$

where

$$F(u, v) = f\left(\exp\begin{pmatrix} 0 & v \\ -u & 0 \end{pmatrix}\right).$$

We obviously have

$$F(0, 0) = f(1), \qquad 1 = 1_G.$$

On the other hand, we had the Harish transform on A, namely

$$\mathbf{H}^A f(a) = |D(a)| \int_{A\backslash G} f(x^{-1}ax)\, d\dot{x} = \rho(a) \int_N f(an)\, dn.$$

We needed values only on A. However, we may view this transform to take values also on $-A$. Thus for $a \in A$,

$$\mathbf{H}^A f(-a) = \rho(a) \int_N f(-an)\, dn.$$

Theorem 1. *Let $f \in C_c^\infty(G, K)$. Then*

$$\mathbf{H}^K f \Big]_{0-}^{0+} = \mathbf{H}^K f(0+) - \mathbf{H}^K f(0-) = \frac{i}{2}\, \mathbf{H}^A f(1),$$

$$\mathbf{H}^K f \Big]_{\pi-}^{\pi+} = \mathbf{H}^K f(\pi+) - \mathbf{H}^K f(\pi-) = \frac{i}{2}\, \mathbf{H}^A f(-1).$$

Proof. By the calculus lemma and $wn_u w^{-1} = \bar{n}_{-u}$ we find

$$\mathbf{H}^K f(0+) = \frac{i}{2} \int_0^\infty f\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} du$$

and

$$\mathbf{H}^K f(0-) = -\frac{i}{2} \int_{-\infty}^0 f\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} du.$$

Hence

$$\mathbf{H}^K f(0+) - \mathbf{H}^K f(0-) = \frac{i}{2}\int_{-\infty}^{\infty} f\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} du$$

$$= \frac{i}{2}\int_N f(n)\,dn$$

$$= \frac{i}{2}\mathbf{H}^A f(1).$$

The second assertion is proved by replacing f with the function $f(-x)$, noting that

$$f(-x) = f(xk_\pi).$$

The points $\theta = 0$ and $\theta = \pi$ are the points of discontinuity of the Harish transform on K, and the preceding theorem gives the jumps in terms of the Harish transform on A.

We know from the calculus lemma that the derivative of the Harish transform $(\mathbf{H}^K f)'(\theta)$ is continuous at 0.

Theorem 2. *For $f \in C_c^\infty(G)$, we have* $(\mathbf{H}^K f)'(0) = -if(1)$.

Proof. Immediate from (iii) of the calculus lemma.

Theorem 3. *The Fourier series of $(\mathbf{H}^K f)'$ converges to the function at* 0.

Proof. From (iv) of the calculus lemma, we know that

$$(\mathbf{H}^K f)'(\theta) = \text{constant} + O(|\theta \log|\theta||).$$

The Dirichlet kernel from which the Fourier series is obtained by convolution is equal to (up to a constant factor)

$$\frac{1}{2\pi n}\,\frac{\sin(n+\frac{1}{2})\theta}{\sin(\theta/2)}\,.$$

The convergence of the Fourier series is a local phenomenon near the point, and depends on the integral

$$\frac{1}{n}\int_{-\epsilon}^{\epsilon} \theta \log|\theta|\,\frac{\sin(n+\frac{1}{2})\theta}{\sin(\theta/2)}\,d\theta$$

tending to 0 as $\epsilon \longrightarrow 0$. But this integral is estimated by

$$\left|\int_{-\epsilon}^{\epsilon} \log|\theta|\, d\theta\right| \ll \epsilon\, |\log \epsilon|,$$

so our theorem follows.

§3. SOME LEMMAS

In this section, we gather some easy lemmas so as not to interrupt the course of the argument of the Plancherel formula. The first one amounts to summing a geometric series.

Lemma 1. *We have the two identities:*

$$\sum_{n \text{ odd}} \sin(|n|\theta)e^{-|n|t} = \frac{2\sin\theta\cosh t}{\cosh 2t - \cos 2\theta},$$

$$\sum_{n \text{ even}} \sin(|n|\theta)e^{-|n|t} = \frac{\sin 2\theta}{\cosh 2t - \cos 2\theta}.$$

The sums are taken over odd and even integers respectively; both positive and negative integers are included.

Proof. Consider n positive, odd, $n = 2d + 1$. We have

$$\sum_{d=0}^{\infty} e^{i(2d+1)\theta}e^{-(2d+1)t} = e^{i\theta - t}\sum_{d=0}^{\infty} e^{2d(i\theta - t)}$$

$$= \frac{1}{e^{-(i\theta - t)} - e^{(i\theta - t)}}.$$

We multiply numerator and denominator by the complex conjugate of the denominator to get

$$\frac{e^{t+i\theta} - e^{-(t+i\theta)}}{2[\cosh 2t - \cos 2\theta]} = \frac{e^{t}(\cos\theta + i\sin\theta) - e^{-t}(\cos\theta - i\sin\theta)}{2[\cosh 2t - \cos 2\theta]}.$$

Taking the imaginary part of the numerator yields $e^{t}\sin\theta + e^{-t}\sin\theta$, which is precisely $2\cosh t\sin\theta$, and proves our first identity.

For the second, summing over positive even integers yields

$$1 + 2\sum_{d=1}^{\infty} e^{2id\theta}e^{-2dt} = 1 + \frac{2e^{2(i\theta - t)}}{1 - e^{2(i\theta - t)}}$$

$$= \frac{e^{-(i\theta - t)} + e^{(i\theta - t)}}{e^{-(i\theta - t)} - e^{(i\theta - t)}}.$$

Again we multiply by the complex conjugate of the denominator, to get

$$\frac{e^{2t} + e^{2i\theta} - e^{-2i\theta} - e^{-2t}}{2[\cosh 2t - \cos 2\theta]} = \frac{\sinh 2t + i \sin 2\theta}{\cosh 2t - \cos 2\theta}.$$

This proves the second identity, by considering the imaginary part.

Next we have to compute a couple of Fourier transforms.

Lemma 2. *For* $0 < \theta < \pi$ *and* $\lambda > 0$ *we have*

$$\lim_{c\to\infty} \int_{-c}^{c} \frac{e^{i\lambda t}}{\cosh 2t - \cos 2\theta}\, dt = \frac{\pi}{\sin 2\theta} \frac{\sinh\left(\frac{\pi}{2} - \theta\right)\lambda}{\sinh \frac{\pi\lambda}{2}}.$$

Proof. We integrate around the rectangle as shown.

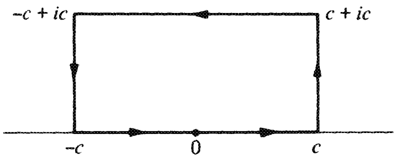

Figure 1

The integral

$$\oint \frac{e^{i\lambda z}}{\cosh 2z - \cos 2\theta}\, dz$$

is equal to $2\pi i$ times the sum of the residues. We have

$$\cosh 2z = \cos 2\theta \Leftrightarrow z = (n\pi \pm \theta)i, \qquad n \in \mathbf{Z}.$$

Also,

$$\frac{d}{dz}\left(\frac{e^{2z}+e^{-2z}}{2}-\cos 2\theta\right)=2\sinh 2z.$$

We evaluate this at the points $z_n=(n\pi\pm\theta)i$ to find $\pm 2i\sin 2\theta$. Hence the residue of the function under the integral sign at z_n is

$$\frac{e^{-\lambda(n\pi\pm\theta)}}{2i\sin(\pm 2\theta)}.$$

The limit of the integral is therefore equal to

$$\frac{2\pi i}{2i\sin 2\theta}\left\{\sum_{n\geqq 0}e^{-\lambda(n\pi+\theta)}-\sum_{n\geqq 1}e^{-\lambda(n\pi-\theta)}\right\}$$

$$=\frac{\pi}{\sin 2\theta}\left[\frac{e^{-\lambda\theta}}{1-e^{-\lambda\pi}}-\frac{e^{\lambda\theta}e^{-\lambda\pi}}{1-e^{-\lambda\pi}}\right].$$

Multiplying numerator and denominator by $e^{\lambda\pi/2}$ proves the lemma.

Of course, in taking the limit, we pick c not equal to a value which would cause a pole of the function to fall on the boundary of the rectangle.

Lemma 3. *For $0<\theta<\pi$ and $\lambda>0$ we have*

$$\lim_{c\to\infty}\int_{-c}^{c}\frac{e^{i\lambda t}\cosh t}{\cosh 2t-\cos 2\theta}\,dt=\frac{\pi}{2\sin\theta}\,\frac{\cosh\left(\frac{\pi}{2}-\theta\right)\lambda}{\cosh\frac{\pi\lambda}{2}}.$$

Proof. We integrate

$$\oint\frac{e^{i\lambda z}\cosh z}{\cosh 2z-\cos 2\theta}\,dz$$

over the same contour as before. The residue at $z_n=(n\pi\pm\theta)i$ is

$$\frac{e^{-\lambda(n\pi\pm\theta)}\cosh(n\pi\pm\theta)i}{2i\sin(\pm 2\theta)}=(-1)^n\frac{\cos\theta e^{-\lambda(n\pi\pm\theta)}}{2i\sin(\pm 2\theta)}.$$

Hence our Fourier transform is equal to $2\pi i$ times the sum of the residues, namely

$$\frac{2\pi i}{4i \sin \theta}\left\{\sum_{n=0}^{\infty}(-1)^n e^{-\lambda(n\pi+\theta)} - \sum_{n=1}^{\infty}(-1)^n e^{-\lambda(n\pi-\theta)}\right\}$$

$$= \frac{\pi}{2 \sin \theta}\left[\frac{e^{-\lambda\theta}}{1+e^{-\lambda\pi}} + \frac{e^{\lambda\theta}e^{-\lambda\pi}}{1+e^{-\lambda\pi}}\right]$$

$$= \frac{\pi}{2 \sin \theta}\,\frac{\cosh\left(\frac{\pi}{2}-\theta\right)\lambda}{\cosh \frac{\pi\lambda}{2}},$$

as was to be proved.

The above lemmas will be applied to a Plancherel inversion situation under the following conditions. Let $\varphi \in C_c^\infty(\mathbf{R})$, and let g be one of the two functions whose Fourier transform is computed in Lemma 2 or 3, so that in particular, g is real. Let g_c be the function equal to 0 outside the interval $[-c, c]$, and equal to g on this interval. Then the Fourier transform of g_c is actually computed in the above lemmas, and is equal to a partial sum of the residues, plus a term which tends to 0 as c tends to ∞. Since φ has compact support, we obtain

$$\int_{-\infty}^{\infty} \varphi g = \int_{-\infty}^{\infty} \varphi \overline{g_c} = \int_{-\infty}^{\infty} \hat{\varphi}\overline{\hat{g}_c}.$$

As $c \longrightarrow \infty$, we see from Lemmas 2 and 3 that $\overline{\hat{g}_c}$ tends to a bounded continuous function of λ. If g denotes the function on the right-hand side of the formulas in Lemmas 2 and 3, we obtain the Plancherel formula

$$\int_{-\infty}^{\infty} \varphi g = \int_{-\infty}^{\infty} \hat{\varphi}\hat{g}.$$

§4. THE PLANCHEREL FORMULA

We put everything together. We recall that by definition

$$f^+(x) = \frac{f(x)+f(-x)}{2}, \qquad f^-(x) = \frac{f(x)-f(-x)}{2}.$$

Put $s = i\lambda$ with λ real. The representation π_s on $H(i\lambda)$ is the representation of the principal series, and we have seen that for $\lambda \neq 0$ it splits into two irreducible components according to parity. By formula (1) of VII, §3, the trace in these components is given by the two expressions (Fourier transforms!)

$$T_\lambda^+(f) = \int_{-\infty}^{\infty} \mathbf{H}^A f^+(h_t) e^{i\lambda t}\, dt, \tag{1}$$

$$T_\lambda^-(f) = \int_{-\infty}^{\infty} \mathbf{H}^A f^-(h_t) e^{i\lambda t}\, dt. \tag{2}$$

For $0 < \theta < \pi$ we have the Fourier expansion for $f \in C_c^\infty(G)$:

$$\mathbf{H}^K f(\theta) = \sum_n e^{-in\theta} \int_0^{2\pi} \mathbf{H}^K f(\varphi) e^{in\varphi} \frac{d\varphi}{2\pi},$$

$$\mathbf{H}^K f(-\theta) = \sum_n e^{in\theta} \int_0^{2\pi} \mathbf{H}^K f(\varphi) e^{in\varphi} \frac{d\varphi}{2\pi}.$$

Subtract to find the average

$$J^K f(\theta) = \tfrac{1}{2}[\mathbf{H}^K f(\theta) - \mathbf{H}^K f(-\theta)].$$

Theorem 4. *Let S_n be the function defined in* VII, §5, *Th.* 6. *Let $f \in C_c^\infty(G, K)$. For $0 < \theta < \pi$ we have*

$$\frac{2\pi}{i} J^K f(\theta) = -\sum_{n \neq 0} S_n(f) \sin|n|\theta$$

$$+ \frac{1}{2} \int_0^\infty T_\lambda^+(f) \frac{\cosh\left(\frac{\pi}{2} - \theta\right)\lambda}{\cosh \frac{\pi\lambda}{2}}\, d\lambda + \frac{1}{2} \int_0^\infty T_\lambda^-(f) \frac{\sinh\left(\frac{\pi}{2} - \theta\right)\lambda}{\sinh \frac{\pi\lambda}{2}}\, d\lambda.$$

Proof. When subtracting the Fourier series above, the constant terms cancel, and we find

$$\frac{2\pi}{i} J^K f(\theta) = -\sum_{n \neq 0} \sin n\theta \int_0^{2\pi} \mathbf{H}^K f(\varphi) e^{in\varphi}\, d\varphi. \tag{3}$$

Substituting the value found in the last theorem of the preceding chapter, VII,

§5, Th. 6, we find that

$$(4) \qquad \frac{2\pi}{i} J^K f(\theta) = -\sum_{n \neq 0} S_n(f) \sin|n|\theta$$

$$+ \sum_{n \neq 0} \int_0^\infty \frac{1}{2}\left[\mathbf{H}^A f(h_t) + (-1)^{n+1}\mathbf{H}^A f(-h_t)\right] \cdot \sin|n|\theta \, e^{-|nt|}\, dt.$$

We split the second sum into terms with n odd and n even respectively. This yields an expression for the second sum, namely

$$\int_0^\infty \mathbf{H}^A f^+(h_t) \frac{2\sin\theta\cosh t}{\cosh 2t - \cos 2\theta}\, dt + \int_0^\infty \mathbf{H}^A f^-(h_t) \frac{\sin 2\theta}{\cosh 2t - \cos 2\theta}\, dt$$

in view of the computation of Lemma 1, §3. By Plancherel's formula as discussed in the preceding section, and the fact that $\mathbf{H}^A f$ has compact support, this becomes

$$\frac{1}{2}\int_0^\infty T_\lambda^+(f) \frac{\cosh\left(\frac{\pi}{2} - \theta\right)\lambda}{\cosh \frac{\pi\lambda}{2}}\, d\lambda + \frac{1}{2}\int_0^\infty T_\lambda^-(f) \frac{\sinh\left(\frac{\pi}{2} - \theta\right)\lambda}{\sinh \frac{\pi\lambda}{2}}\, d\lambda.$$

Note that the two functions for which we take the Fourier transform are even functions of t, and hence that the Fourier transforms taken as integrals from $-\infty$ to ∞ can be replaced by twice integrals from 0 to ∞. We have therefore obtained the expansion stated in the theorem.

Plancherel Formula. *Let $f \in C_c^\infty(G)$. Then*

$$2\pi f(1) = \sum_{n \neq 0} |n| S_n(f)$$

$$+ \frac{1}{2}\int_0^\infty T_\lambda^+(f)\lambda \tanh\frac{\pi\lambda}{2}\, d\lambda + \frac{1}{2}\int_0^\infty T_\lambda^-(f)\lambda\coth\frac{\pi\lambda}{2}\, d\lambda.$$

Proof. We differentiate the expression of Theorem 4 and put $\theta = 0$. The Plancherel formula drops out.

Remark. We know that the traces T_λ come from unitary representations. It also turns out that the traces S_n come from unitary representations, as we shall prove in the next chapter. Hence in Plancherel formula, only unitary representations appear. This was irrelevant in the preceding considerations, but we shall use this below in the formalism which can be derived from the above stated formula.

Let $\hat{G}$ be the set of isomorphism classes of unitary representations. We let μ be the measure on $\hat{G}$ defined as follows:

For a point of the discrete series of weight m, we let the measure be discrete, and we give this point measure $|n|/2\pi$, where $n = m - 1$ or $m + 1$ according as $m > 0$ or $m < 0$.

On the subset of $\hat{G}$ in the principal series parametrized by $s = i\lambda$, $\lambda > 0$, with even parity, we take the measure

$$d\mu(\lambda) = \frac{1}{2\pi}\frac{\lambda}{2}\tanh\frac{\pi\lambda}{2}\,d\lambda.$$

On the subset of $\hat{G}$ in the principal series parametrized by $s = i\lambda$, $\lambda > 0$, with odd parity, we take the measure

$$d\mu(\lambda) = \frac{1}{2\pi}\frac{\lambda}{2}\coth\frac{\pi\lambda}{2}\,d\lambda.$$

We let the complement of the above sets have measure 0.

We call μ the **Plancherel measure**. We shall write $d\mu(\pi)$ with $\pi \in \hat{G}$, viewing $\hat{G}$ itself as the measured space, instead of using parameters n and λ.

On the product $\hat{G} \times G$ we have an operator-valued mapping

$$\varphi(\pi, x) = \pi(x).$$

The corresponding map on functions $f \in C_c^\infty(G)$ is

$$\hat{f}(\pi) = \int f(x)\pi(x)\,dx = \pi^1(f).$$

On the other hand, let $\{A(\pi)\}$ be a family of operators, with $A(\pi)$ operating on H_π. Define (formally) the adjoint Φ^* by

$$\Phi^* A(x) = \int_{\hat{G}} \operatorname{tr} A(\pi)\pi(x)^*\,d\mu(\pi).$$

We shall see in a moment that the inversion formula

$$\Phi^*\Phi = id \quad \text{on} \quad C_c^\infty(G)$$

follows from the above Plancherel formula, which can be written in the form

PL 1. $$f(1) = \int_{\hat{G}} \operatorname{tr} \pi^1(f)\,d\mu(\pi),$$

and is valid for $f \in C_c^\infty(G)$ or $f \in C_c^\infty(G, K)$. Indeed, if $f \in C_c^\infty(G)$, let

$$f_K(x) = \int_K f(k^{-1}xk)\,dk.$$

Then both sides of **PL 1** remain unchanged if f is replaced by f_K. Let $r(x)f$ be the right translate of f by x, so that

$$r(x)f(y) = f(yx).$$

Replacing f by $r(x)f$ in **PL 1**, and noting that

$$\pi^1(r(x)f) = \pi^1(f)\pi(x^{-1}),$$

we obtain the inversion formula at an arbitrary point, namely

PL 2.
$$f(x) = \int_{\hat{G}} \operatorname{tr} \pi^1(f)\pi(x)^*\,d\mu(\pi).$$

Furthermore, let $\varphi, \psi \in C_c^\infty(G)$ and $\varphi^*(x) = \overline{\varphi(x^{-1})}$. Apply **PL 1** to

$$f = \varphi * \psi^*.$$

Then:

$$\pi^1(f) = \pi^1(\varphi)\pi^1(\psi)^*,$$

$$(\varphi * \psi^*)(1) = \langle \varphi, \psi \rangle_G = \int_G \varphi(x)\,\overline{\psi(x)}\,dx,$$

$$\operatorname{tr} \pi^1(\varphi * \psi^*) = \operatorname{tr} \pi^1(\varphi)\pi^1(\psi)^* = \langle \pi^1(\varphi), \pi^1(\psi) \rangle = \langle \hat{\varphi}, \hat{\psi} \rangle.$$

Therefore we find the L^2-version of Plancherel inversion, namely:

PL 3. *For $\varphi, \psi \in C_c^\infty(G)$, we have*

$$\langle \varphi, \psi \rangle_G = \int_{\hat{G}} \operatorname{tr} \pi^1(\varphi)\pi^1(\psi)^*\,d\mu(\pi) = \langle \hat{\varphi}, \hat{\psi} \rangle.$$

Thus the Plancherel formula **PL 1** implies the other versions by means of simple formal arguments.

Observe that in the above formulas, we frequently take the trace of a product of operators. Let $\{u_i\}$ be an orthonormal basis of H. If A, B are operators, let $A_{ij} = \langle Au_i, u_j \rangle$. Then

$$\operatorname{tr} AB = \sum_{i,j} A_{ij}B_{ji},$$

as in the finite dimensional case. Indeed, in H we have the convergent sum

$$Bu_i = \sum_j B_{ij} u_j,$$

so

$$\langle ABu_i, u_i \rangle = \sum_j \langle AB_{ij} u_j, u_i \rangle$$

and

$$\operatorname{tr} AB = \sum_{i,j} B_{ij} A_{ji}.$$

In practice, one sometimes needs such expansions, when $\{u_i\}$ is a basis of the K-finite vectors, in order to estimate certain traces.

IX Discrete Series

In this chapter we give various unitary realizations of the discrete series, i.e. those irreducible representations which admit a lowest weight vector of weight $\geqslant 2$ and highest weight vector of weight $\leqslant -2$. It turns out that in each case, one is the complex conjugate of the other, so essentially we need only look at those with a lowest weight vector. We shall see that they admit an infinitesimal embedding in $L^2(G)$, with the action of left translation, and also that they can be represented as operations on certain function spaces in the upper half plane.

The uniqueness theorem of Chapter VI, §3, leaves no ambiguity concerning such unitarizations, yielding a natural unitary isomorphism between them.

§1. DISCRETE SERIES IN $L^2(G)$

The unitarization of a discrete series in $L^2(G)$ in this section is due to Harish-Chandra.

Let

$$T = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix}, \qquad T^{-1} = \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix}.$$

If

$$x = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G = SL_2(\mathbf{R}),$$

we change the coordinates by considering the matrix coefficients of TxT^{-1}, namely we put

$$(1) \qquad \frac{1}{\sqrt{2}}\begin{pmatrix} 1 & -i \\ -i & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & i \\ i & 1 \end{pmatrix} = \begin{pmatrix} \alpha & \bar{\beta} \\ \beta & \bar{\alpha} \end{pmatrix}.$$

(We put bars this way to minimize the number of bars in this section.) Then

(2) $\alpha = \frac{1}{2}(a + d - ic + ib)$ and $\beta = \frac{1}{2}(c + b - ia + id)$.

Note that α, β are function on the group, namely $\alpha(x)$, $\beta(x)$. They are easily computed for elements of A and K, namely:

$$\alpha\left(\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}\right) = \frac{e^t + e^{-t}}{2} = \cosh t, \tag{3}$$

$$\beta\left(\begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix}\right) = -i\frac{e^t - e^{-t}}{2} = -i \sinh t. \tag{4}$$

Furthermore, these transform in a simple way by the operation of K on the right and left. We note that conjugation by T as above on elements of K yields

$$Tk_\theta T^{-1} = T\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}T^{-1} = \begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}.$$

Thus we see that

$$\alpha(k_\theta x) = e^{i\theta}\alpha(x), \qquad \alpha(xk_\theta) = \alpha(x)e^{i\theta}, \tag{5}$$

$$\beta(k_\theta x) = e^{-i\theta}\beta(x), \qquad \beta(xk_\theta) = \beta(x)e^{i\theta}. \tag{6}$$

From the KA^+K decomposition of G (Cartan decomposition), we now see that the values of α and β are determined by the above tabulation, and especially the absolute values are $\cosh t$ and $\sinh t$ respectively if x is conjugate to h_t. *Note that* $|\alpha(x)| \geqslant 1$, all $x \in G$.

Lemma 1. *Let m be an integer $\geqslant 2$. Then the function α^{-m} is in $L^2(G)$.*

Proof. By the integral formula **INT 2**, we know that

$$\int_G |\alpha|^{-2m}\, dx = \int_0^\infty (\cosh t)^{-2m} \sinh 2t\, dt,$$

up to a constant factor. Since $\sinh 2t = (e^t + e^{-t})(e^t - e^{-t})/2$, the integral is of the form

$$\int u^{-2m+1}\, du = \frac{u^{-2m+2}}{-2m+2},$$

which converges for $m \geqslant 2$, as desired.

Let π be the representation by left translation on $L^2(G)$, so that

$$\pi(y)f(x) = f(y^{-1}x).$$

Abbreviate $\pi(y)f$ by f_y. Then

$$\alpha_y = \alpha(y^{-1})\alpha + \bar{\beta}(y^{-1})\beta, \tag{7}$$

$$\beta_y = \beta(y^{-1})\alpha + \bar{\alpha}(y^{-1})\beta. \tag{8}$$

Theorem 1. *Let $\varphi_{m+2r} = \alpha^{-m-r}\beta^r$ for $r = 0, 1, 2, \ldots$. Then the functions φ_{m+2r} are eigenvectors of K with eigenvalue $e^{i(m+2r)\theta}$. The closed subspace of $L^2(G)$ which they generate is invariant under left translation by G, it is irreducible, and has lowest weight vector equal to α^{-m}.*

Proof. The eigenvalue property is immediate from (5) and (6), i.e. essentially directly from the definition of α and β. Observe that β/α is a function of absolute value < 1, and that for any given y, $|\beta(y)/\alpha(y)| < 1$. The function $\pi(y^{-1})\varphi_{m+2r}$ lies in the vector space generated by the functions

$$\frac{1}{\alpha^m} \frac{1}{\left(1 + \dfrac{\bar{\beta}(y)}{\alpha(y)} \dfrac{\beta}{\alpha}\right)^{m+r}} \left(\frac{\beta}{\alpha}\right)^{\nu}.$$

Let $\lambda = \bar{\beta}(y)\beta/\alpha(y)\alpha$, so that $|\lambda| \leqslant |\beta(y)/\alpha(y)| < 1$. We can expand $\alpha^{-m}(1 + \lambda)^{-m-r}$ into a power series which converges in L^2 by Lemma 1. This proves that the closed subspace generated by the prescribed functions is invariant under translation, and proves our theorem.

§2. REPRESENTATION IN THE UPPER HALF PLANE

We shall here describe another model for the unitary representation of a discrete series, by a function space on the upper half plane.

Let m be an integer $\geqslant 2$. We recall that on the upper half plane $\mathfrak{H}$ we have a measure

$$d\mu(x, y) = \frac{dx\, dy}{y^2}$$

invariant under the action of G. We let

$$d\mu_m = y^m \frac{dx\,dy}{y^2}.$$

Let

$$\boxed{H = L^2_{\mathrm{hol}}(\mathfrak{H}, \mu_m)}$$

be the space of holomorphic functions on $\mathfrak{H}$ which are in $\mathfrak{L}^2$ with respect to the measure μ_m. The scalar product is the usual hermitian product given by the integral over $\mathfrak{H}$. We need a lemma to insure that H is complete.

Lemma 1. *If a sequence of holomorphic functions $\{f_n\}$ is L^2-convergent in an open set in the complex plane, then it is uniformly convergent to a holomorphic function on any compact set. In fact, locally, we have domination of norms*:

$$\|\ \| \ll \|\ \|_1 \ll \|\ \|_2.$$

Proof. We work in the neighborhood of a point, which we may assume to be the origin without loss of generality. Our estimates will depend on some disc of fixed radius δ around any point, and again it suffices to bound $|f(0)|$ in terms of the L^1 and L^2 norm of f in a δ-disc around 0. Cauchy's formula gives

$$f(0) = \frac{1}{2\pi i}\int \frac{f(\zeta)}{\zeta - z}\,d\zeta = \frac{1}{2\pi}\int_0^{2\pi} \frac{f(re^{i\theta})}{re^{i\theta}}\,d\theta,$$

whence

$$|f(0)| \leqslant \frac{1}{2\pi}\int_0^{2\pi} \frac{|f(re^{i\theta})|}{r}\,d\theta.$$

Since

$$|f(0)|\frac{\delta^3}{3} = \int_0^{\delta} |f(0)|r^2\,dr,$$

we obtain

$$\begin{aligned}
|f(0)|\frac{\delta^3}{3} &\leqslant \frac{1}{2\pi}\int_0^{\delta}\int_0^{2\pi} \frac{|f(re^{i\theta})|}{r}\,r^2\,dr\,d\theta \\
&= \frac{1}{2\pi}\|f\|_{1,\mathrm{loc}} \\
&\leqslant \frac{1}{2\pi}\|f\|_{2,\mathrm{loc}}\|1\|_{2,\mathrm{loc}}
\end{aligned}$$

by Schwarz. The desired estimate drops out.

Let

$$\sigma^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G = SL_2(\mathbf{R}).$$

On H define

$$(\pi_m(\sigma)f)(z) = f(\sigma^{-1}z)(cz+d)^{-m}.$$

It is verified by brute force that π_m is a group homomorphism of G into the linear automorphisms of H.

Theorem 2. *π_m is a unitary representation.*

Proof. We first verify the unitary property. Let $w = \sigma^{-1}z$. Recall that

$$\operatorname{Im} \sigma^{-1}z = \frac{y}{|cz+d|^2}.$$

We have

$$\begin{aligned} \|\pi_m(\sigma)f\|_2^2 &= \int_0^\infty \int_{-\infty}^\infty |f(\sigma^{-1}z)|^2 |cz+d|^{-2m} y^m \frac{dx\,dy}{y^2} \\ &= \int_0^\infty \int_{-\infty}^\infty |f(w)|^2 (\operatorname{Im} w)^m \, d\mu(w) \\ &= \|f\|_2^2, \end{aligned}$$

as desired.

To prove the continuity condition for a representation, it suffices to do it for a dense subset of elements of the Hilbert space. Our representation occurs in a subspace of $L^2(\mathfrak{H}, \mu_m)$, and it suffices to prove continuity for the operation in this bigger space. We may therefore take a vector f which lies in $C_c(\mathfrak{H})$, and we have to verify that as $\sigma \to 1$, we have

$$\|\pi(\sigma)f - f\|_2 \to 0.$$

This is immediate from the dominated convergence theorem.

Lemma 2. *Let n be an integer $\geqq 0$ and let*

$$\psi_n(z) = \left(\frac{z-i}{z+i}\right)^n (z+i)^{-m}.$$

Then $\psi_n \in H$.

Proof. As $|z| \to \infty$, $\psi_n(z) \to 0$, and we have

$$\left|\frac{z-i}{z+i}\right| \leqslant 1$$

for all z in the upper half plane. Let $\delta > 0$. The proof consists in proving that the L^2-integrals above δ and below δ both converge. As to the first, we have the estimate:

$$\begin{aligned}
\int_\delta^\infty \int_{-\infty}^\infty |\psi_n(z)|^2\, d\mu_m &\ll \int_\delta^\infty \int_{-\infty}^\infty \frac{1}{|z+i|^{2m}} y^m \frac{dx\,dy}{y^2} \\
&\ll \int_\delta^\infty \int_{-\infty}^\infty \frac{1}{(x^2+y^2)^m} y^{m-2}\, dx\,dy \\
&\ll \int_\delta^\infty \int_{-\infty}^\infty \frac{1}{\left((x/y)^2+1\right)^m} \frac{y^m}{y^{2m+2}}\, dx\,dy \\
&\ll \int_\delta^\infty \int_{-\infty}^\infty \frac{1}{(u^2+1)^{1+\epsilon}} \frac{1}{y^{m+1}}\, du\,dy,
\end{aligned}$$

which converges. As to the second,

$$\begin{aligned}
\int_0^\delta \int_{-\infty}^\infty \frac{1}{|z+i|^{2m}} y^m \frac{dx\,dy}{y^2} &\ll \int_0^\delta \int_{-\infty}^\infty \frac{1}{(x^2+1)^{1+\epsilon}} y^{m-2}\, dx\,dy \\
&\ll \int_0^\delta y^{m-2}\, dy,
\end{aligned}$$

which is bounded if $m > 1$. This proves our lemma.

Theorem 3. *The representation π_m on $H = L^2_{\text{hol}}(\mathfrak{H}, \mu_m)$ is irreducible. Let H_{m+2n} be the one-dimensional subspace generated by the function ψ_n. Then H_{m+2n} is an eigenspace of K, with character $m + 2n$, and*

$$H = \widehat{\bigoplus_{n \geqslant 0}} H_{m+2n}$$

is an orthogonal decomposition, with lowest weight vector ψ_0, of weight m.

The proof of Theorem 3 is best carried out by changing the model for the representation under the analytic isomorphism between the upper half plane and the unit disc. We shall do this in the next section.

Theorem 3 gives us another example of a representation with arbitrary lowest weight vector $\geqslant 2$. To get similar representations on the other side, i.e. with highest weight vectors of weight $\leqslant -2$, we merely let $SL_2(\mathbf{R})$ operate on the antiholomorphic functions in the lower half plane.

§3. REPRESENTATION ON THE DISC

It is often easier to work on the disc where the functions ψ_n have an easier expression. Let

$$w = \frac{z - i}{z + i}.$$

The map $z \mapsto w$ is an analytic isomorphism between the upper half plane and the disc D of radius 1, centered at the origin. The inverse mapping is

$$z = -i\frac{w + 1}{w - 1}.$$

If f is a function on $\mathfrak{H}$, we let

$$T_m f(w) = f\left(-i\frac{w+1}{w-1}\right)\left(\frac{-2i}{w-1}\right)^m.$$

Then

$$T_m\colon \mathcal{F}(H) \to \mathcal{F}(D)$$

is a linear map from the functions on $\mathfrak{H}$ to the functions on D. On the disc, let $w = u + iv$, and let

$$\boxed{\begin{aligned} dv_m &= \frac{4}{4^m}(1 - |w|^2)^m \frac{du\,dv}{(1 - |w|^2)^2} \\ &= \frac{4}{4^m}(1 - r^2)^{m-2} r\,dr\,d\theta. \end{aligned}}$$

Lemma 1. *The map*

$$T_m\colon L^2_{\text{hol}}(\mathfrak{H}, \mu_m) \to L^2_{\text{hol}}(D, \nu_m)$$

is an isometry.

Proof. We have $dw\wedge d\bar{w} = -2\,du\wedge dv$ and $dz\wedge d\bar{z} = -2i\,dx\wedge dy$. We get

$$\frac{4\,du\,dv}{(1 - |w|^2)^2} = \frac{dx\,dy}{y^2}.$$

The isometry amounts to

$$\iint \left| f\left(-i\frac{w+1}{w-1}\right)\left(\frac{1}{w-1}\right)^m \right|^2 (1-|w|^2)^m \frac{4\,du\,dv}{(1-|w|^2)^2}$$
$$= \iint |f(z)|^2 y^m \frac{dx\,dy}{y^2},$$

which is clear.

We now see the advantage of having gone over to the unit disc, namely

$$T_m\psi_n(w) = w^n.$$

Thus our functions go over to the powers of the variable in the unit disc, and they are trivially verified to be orthogonal.

Theorem 4. *The functions* $\{1, w, w^2, \ldots\}$ *form a complete orthogonal basis for* $L^2_{\text{hol}}(D, \nu_m)$.

Proof. Let $f \in L^2_{\text{hol}}(D, \nu_m)$. Then f has a power series expansion

$$f(w) = \sum_{n=0}^{\infty} a_n w^n.$$

It suffices to prove that the series converges in $L^2(\nu_m)$. Let $0 < r' < 1$, and let $D_{r'}$ be the disc of radius r'. Then

$$\int_{D_{r'}} |f|^2\,d\nu_m = \sum_{q,\,n} a_n \overline{a_q} w^n \overline{w^q}\,d\nu_m(w)$$
$$= \frac{4}{4^m} \sum_{n=1}^{\infty} |a_n|^2\, 2\pi \int_0^{r'} r^{2n}(1-r^2)^{m-2} r\,dr$$

because

$$\int_0^{2\pi} e^{in\theta} e^{-iq\theta}\,d\theta = 0$$

if $n \neq q$. Take the limit as $r' \to 1$. We get

$$\|f\|_2 = \int_D |f|^2\,d\nu_m = \frac{2\pi\cdot 4}{4^m} \sum |a_n|^2 \int_0^1 r^{2n}(1-r^2)^{m-2} r\,dr$$
$$= c\sum |a_n|^2 \|w^n\|^2_{2,\,\nu_m}$$

for some constant c. Hence $\sum a_n w^n$ is $L^2(\nu_m)$-convergent for $m > 1$. We are done.

The group action on the disc is transformed from the upper half plane as follows. $SL_2(\mathbf{R})$ goes to the group of matrices

$$\sigma^{-1} = \begin{pmatrix} \alpha & \beta \\ \bar{\beta} & \bar{\alpha} \end{pmatrix} \qquad \text{with} \quad \alpha\bar{\alpha} - \beta\bar{\beta} = 1.$$

Denote by $\tilde{\pi}_m$ the transformed representation. Then

$$\tilde{\pi}_m(\sigma)f(w) = f\left(\frac{\alpha w + \beta}{\bar{\beta}w + \bar{\alpha}}\right)(\bar{\beta}w + \bar{\alpha})^{-m}.$$

Let $\tilde{H} = L^2_{\text{hol}}(D, \nu_m)$. Then the one-dimensional subspace generated by w^n is our old $\tilde{H}_{m+2n}$. Indeed,

$$\tilde{\pi}_m\left(\begin{pmatrix} e^{i\theta} & 0 \\ 0 & e^{-i\theta} \end{pmatrix}\right)(w^n) = e^{i\theta(m+2n)}w^n$$

according to the formula describing $\tilde{\pi}_m$. Therefore $\tilde{H}_{m+2n}$ is an eigenspace of K, with character $m + 2n$. The constant function 1 has character m.

In order to finish the proof of Theorem 2, we have to check:

Lemma 2. *The elements* $1, w, w^2, \ldots$ *in* $\tilde{H}$ *are analytic vectors.*

Proof. The proof is entirely similar to the proof of the analogous statement for the continuous series, VI, §5, Lemma 2, and will be left to the reader.

We can then apply the general irreducibility discussion of Chapter VI, §2 to finish the proof of Theorem 2.

§4. THE LIFTING OF WEIGHT m

This section shows how to establish an isomorphism between the representation in §1 and that in §2 and §3 by relating functions on $\mathfrak{H}$ with functions on G through a formal lifting process.

Let $\mathfrak{H}$ be the upper half plane and let m be a fixed integer. To each function φ on $\mathfrak{H}$ we associate a function Φ on $G = GL_2^+(\mathbf{R})$ by letting

$$\Phi\left(u\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}r(\theta)\right) = \varphi(x + iy)y^{m/2}e^{im\theta},$$

where

$$g = \begin{pmatrix} u & 0 \\ 0 & u \end{pmatrix}\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

is our usual unique decomposition of an element of $GL_2^+(\mathbf{R})$, with $u, y > 0$. Our function Φ is independent of u and so gives rise to a function on $SL_2(\mathbf{R})$, but we like to work with the coordinates (x,y) which have the convenient interpretation in the upper half plane.

The associated function Φ satisfies the conditions:

$\mathfrak{F}_m$**1**. *Φ is independent of u.*

$\mathfrak{F}_m$**2**. $\Phi(gr(\theta)) = \Phi(g)e^{im\theta}$.

Let $\mathfrak{F}_m(G)$ be the space of functions on G satisfying these two conditions, and let $\mathfrak{F}(\mathfrak{H})$ be the space of all functions on $\mathfrak{H}$. We have the lifting map

$$J_m\colon \mathfrak{F}(\mathfrak{H}) \to \mathfrak{F}_m(G)$$

which to each φ associates Φ. This map is bijective, because it has an inverse, which to each function Φ satisfying the two conditions $\mathfrak{F}_m$**1** and $\mathfrak{F}_m$**2** associates the function φ on H such that

$$\varphi(x + iy) = \Phi\left(\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}\right) y^{-m/2}.$$

To each element

$$\sigma^{-1} = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in SL_2(\mathbf{R})$$

we associate an operator $\pi_m(\sigma)$ on the function space $\mathfrak{F}(\mathfrak{H})$ by letting $\pi_m(\sigma)\varphi$ be the function such that

$$\pi_m(\sigma)\varphi(z) = \varphi(\sigma^{-1}z)(\gamma z + \delta)^{-m}.$$

Then $\sigma \mapsto \pi_m(\sigma)$ is an algebraic representation. Let L_σ be left translation by σ, so that $(L_\sigma\Phi)(g) = \Phi(\sigma^{-1}g)$, for any function Φ on G. *Then the following diagram commutes.*

$$\begin{array}{ccc} \mathfrak{F}(\mathfrak{H}) & \xrightarrow{J_m} & \mathfrak{F}_m(G) \\ {\scriptstyle \pi_m(\sigma)}\downarrow & & \downarrow {\scriptstyle L_\sigma} \\ \mathfrak{F}(\mathfrak{H}) & \xrightarrow[J_m]{} & \mathfrak{F}_m(G) \end{array}$$

Proof. Matrix multiplication shows that

(1) $$ue^{i\theta} = d - ic \quad \text{and} \quad e^{i\theta} = \frac{d - ic}{|d - ic|},$$

(2) $$x + iy = \frac{1}{u} e^{i\theta}(ai + b).$$

Let

$$\sigma^{-1}g = u'\begin{pmatrix} y' & x' \\ 0 & 1 \end{pmatrix} r(\theta') = \begin{pmatrix} * & * \\ \gamma a + \delta c & \gamma b + \delta d \end{pmatrix},$$

$$z' = \frac{\alpha z + \beta}{\gamma z + \delta} = x' + iy'.$$

Then (1) and (2) applied to $\sigma^{-1}g$ show that

$$\theta' = \theta - \arg(\gamma z + \delta).$$

We now get

$$\begin{aligned}
\Phi(\sigma g) &= \varphi(\sigma^{-1}gi)y'^{m/2}e^{im\theta'} \\
&= \varphi(z')y^{m/2}|\gamma z + \delta|^{-m} e^{im\theta} e^{-im \arg(\gamma z + \delta)} \\
&= \varphi(\sigma^{-1}z)(\gamma z + \delta)^{-m} y^{m/2} e^{im\theta} \\
&= \pi_m(\sigma)\varphi(z)y^{m/2}e^{im\theta} \\
&= J_m^{-1}\pi_m(\sigma)J_m\Phi(g).
\end{aligned}$$

This proves the commutativity.

§5. THE HOLOMORPHIC PROPERTY

In the correspondence between functions on $\mathfrak{H}$ and functions on G, it is clear that C^∞ functions correspond to each other under the lifting of weight m. We now want to see what is the condition on Φ corresponding to analyticity of φ. As before, we let $\mathfrak{g}$ be the Lie algebra of $SL_2(\mathbf{R})$, and we recall our Lie derivative associated with an element $X \in \mathfrak{g}$,

$$\mathfrak{L}_X \Phi(g) = \frac{d}{dt}\Phi(g \exp(tX))\Big|_{t=0}.$$

We had computed the Lie derivative in terms of the coordinates (x,y,θ) in Chapter VI, §4, and especially we had found the formula for $\mathfrak{L}_{E^-}$, which we now apply to the function

$$\Phi(g) = F(u, x, y, \theta) = \varphi(x + iy)y^{m/2}e^{im\theta},$$

independent of u. We find

$$\mathfrak{L}_{E^-}F(x, y, \theta) = -2iy^{(m/2)+1}e^{2i((m/2)-1)\theta}\frac{\partial\varphi}{\partial\bar{z}}. \tag{3}$$

Consequently, we get

Theorem 5. *The function φ is holomorphic if and only if*

$$\mathfrak{L}_{E^-}\Phi = 0.$$

X Partial Differential Operators

So far we have avoided to a large extent the more refined behavior of functions with respect to Lie derivatives. For the theory of spherical functions, we dealt with eigenvectors of convolution operators. The time has come to relate some invariants we have found in the representation theory with some of the invariant differential operators on G. Bargmann [Ba] saw how coefficient functions are eigenfunctions of such operators. Harish-Chandra got a complete insight into the situation by determining the center of the algebra of invariant differential operators, the centralizer of K in this algebra. Gelfand characterized spherical functions as eigenfunctions of this centralizer. In this chapter, we give Harish-Chandra's result that there are no other spherical functions, besides those described in Chapter IV, on $SL_2(\mathbf{R})$ where the proofs are short and easy.

§1. THE UNIVERSAL ENVELOPING ALGEBRA

Let $\mathfrak{g}$ be the Lie algebra of $SL_2(\mathbf{R})$. It has a basis,

$$X_+ = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X_- = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

and its complexification $\mathfrak{g}_{\mathbf{C}}$ has another basis

$$E_+ = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}, \quad E_- = \begin{pmatrix} 1 & -i \\ -i & -1 \end{pmatrix}, \quad W = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

We recall the eigenrelations

$$[H,X_+] = 2X_+, \qquad [H,X_-] = -2X_-, \qquad [X_+,X_-] = H,$$
$$[W,E_+] = 2iE_+, \qquad [W,E_-] = -2iE_-, \qquad [E_+,E_-] = -4iW.$$

To each $X \in \mathfrak{g}$ associate a "variable" L_X, and form the universal associative algebra over $\mathbf{C}$ generated by these variables. We introduce the ideal generated by all elements

$$L_X L_Y - L_Y L_X - L_{XY-YX},$$

where $XY - YX = [X, Y]$ is given by matrix multiplication. The factor algebra by this ideal is called the **universal enveloping algebra** of $\mathfrak{g}$, and is denoted by $\mathfrak{U}(\mathfrak{g})$.

It is clear that the use of the extra symbol L is clumsy notation. On the other hand, we have a natural homomorphism of $\mathfrak{U}(\mathfrak{g})$ into any associative algebra generated by elements indexed by $\mathfrak{g}$, and satisfying the above relation. In particular, if $\mathfrak{L}_X$ denotes the left invariant differential operator previously defined on $C^\infty(G)$, then we have a unique homomorphism of $\mathfrak{U}(\mathfrak{g})$ into the algebra generated by these operators such that

$$L_X \mapsto \mathfrak{L}_X.$$

Similarly, let π be a representation into a Banach space H, and let, as before, H_π^∞ be the space of C^∞ vectors. Let End H_π^∞ be the algebra of endomorphisms. We have a unique homomorphism

$$\mathfrak{U}(\mathfrak{g}) \to \operatorname{End} H_\pi^\infty$$

such that $L_X \mapsto d\pi(X)$.

It turns out that the first homomorphism $L_X \mapsto \mathfrak{L}_X$ is in fact an isomorphism. We shall prove this, and also use a notation which is somewhat less clumsy. For $X \in \mathfrak{g}_{\mathbf{C}}$ we denote by $\boldsymbol{X}$ the operator $\mathfrak{L}_X$. Then $\boldsymbol{X}f = \mathfrak{L}_X f$.

Theorem 1. *The map from $\mathfrak{U}(\mathfrak{g})$ into the algebra of differential operators on $C^\infty(G)$ is injective. If we identify $\mathfrak{U}(\mathfrak{g})$ with its image, then $\mathfrak{U}(\mathfrak{g})$ has a basis consisting of the elements*

$$\boldsymbol{X}_1^p \boldsymbol{X}_2^q \boldsymbol{X}_3^r, \qquad p, q, r \text{ integers} \geqq 0,$$

if $\{X_1, X_2, X_3\}$ is a basis of $\mathfrak{g}$ over $\mathbf{R}$ or $\mathfrak{g}_{\mathbf{C}}$ over $\mathbf{C}$.

Proof. Without loss of generality it suffices to prove the assertion for one particular basis, say E_+, E_-, W. In any associative algebra A, given $x \in A$, the mapping

$$y \mapsto [x,y] = xy - yx$$

is a derivation, i.e. satisfies $D(yz) = yDz + Dy\cdot z$. Furthermore,

$$D(x_1x_2 \cdots x_r) = \sum_{j=1}^{r} x_1 \cdots Dx_j \cdots x_r.$$

The defining relations in $\mathfrak{U}(\mathfrak{g})$ show that the monomials as indicated generate $\mathfrak{U}(\mathfrak{g})$ linearly. It suffices to prove that any differential operator written in the form

$$\sum c_{pqr} E_+^p W^q E_-^r$$

which is the operator 0 must have all coefficients c_{pqr} equal to 0. Write such an operator as a polynomial

$$f_0(E_+, W) + f_1(E_+, W)E_- + \cdots + f_r(E_+, W)E_-^r.$$

In Chapter VI, §5, take $s = n - 1$, and apply the operator to φ_n, with $n \geqslant 1$. Then E_- kills φ_n, and the rules (1) show that $f_0(E_+, W)$ is identically 0 as a polynomial in E_+, W. Proceeding inductively using $\varphi_{n+2}, \varphi_{n+4}, \ldots$ shows that the other polynomials $f_j(E_+, W)$ are identically 0, as desired.

In view of Theorem 1, we identify $\mathfrak{U}(\mathfrak{g})$ with the algebra of differential operators whenever necessary.

***Theorem* 2.** *The centralizer of W in $\mathfrak{U}(\mathfrak{g})$ is the set of all linear combinations of monomials*

$$E_+^p W^q E_-^p, \qquad p, q \geqslant 0$$

Proof. The rule for the derivative of a product shows that

$$(1) \qquad [W, E_+^p W^q E_-^r] = 2i(p - r)E_+^p W^q E_-^r.$$

This shows that all the above monomials commute with W. Conversely, suppose that a linear combination

$$\sum c_{pqr} E_+^p W^q E_-^r$$

commutes with W. Take the bracket with W. We obtain

$$\sum c_{pqr}(p - r)2iE_+^p W^q E_-^r = 0.$$

Using the linear independence of the monomials occurring in these linear relations, we conclude that $c_{pqr} = 0$ for $p \neq r$, thereby proving our theorem.

The centralizer of W in $\mathfrak{U} = \mathfrak{U}(\mathfrak{g})$ is denoted by $\mathfrak{Z}(W)$. The center of $\mathfrak{U}$ is denoted by $\mathfrak{Z}(\mathfrak{U})$ or $\mathfrak{Z}(\mathfrak{g})$. It is clear that $\mathfrak{Z}(W) \supset \mathfrak{Z}(\mathfrak{g})$. Since W

generates the Lie algebra of K, we also denote $\mathfrak{Z}(W)$ by $\mathfrak{Z}(\mathfrak{k})$.

We now proceed to determine the center of $\mathfrak{U}(\mathfrak{g}) = \mathfrak{U}$. There exists a unique linear map

$$h\colon \mathfrak{Z}(W) \longrightarrow \mathbf{C}[W]$$

such that

$$\begin{aligned} h(E_+^p W^q E_-^p) &= W^q \qquad \text{if} \qquad p = 0, \\ &= 0 \qquad \text{if} \qquad p > 0. \end{aligned}$$

The image $h(Y)$ of an element Y under h is characterized by the congruence

$$Y \equiv h(Y) \bmod \mathfrak{U}E_-.$$

Lemma 1. *The map h is a multiplicative homomorphism on $\mathfrak{Z}(W)$.*

Proof. Let $Y_1 \equiv h(Y_1) \bmod \mathfrak{U}E_-$ and let $Y_2 \equiv h(Y_2) \bmod \mathfrak{U}E_-$. Then

$$Y_2 Y_1 \equiv h(Y_2)h(Y_1) \bmod \mathfrak{U}E_-\, W + \mathfrak{U}E_-.$$

Using the commutation rule between E_- and W, and the fact that E_- is an eigenvector for $\mathrm{ad}(W)$, our assertion is clear.

Lemma 2. *The map h is injective on $\mathfrak{Z}(\mathfrak{U})$.*

Proof. Suppose

$$Y = \sum_q c_q E_+^r W^q E_-^r + \sum_{p>r} c_{pq} E_+^p W^q E_-^p,$$

some coefficient $c_{rq} \neq 0$ and $r \geqslant 1$. We show that Y cannot commute with E_-. Consider the irreducible representation with lowest weight vector $m \geqslant 1$, as in Chapter VI, §5, so that we can use the formulas for the derived representation $d\pi(X)$, selecting $s = m - 1$. We look at the effect of monomials on φ_{m+2r}. For $p > r$ the monomial $E_+^p W^q E_-^p$ annihilates φ_{m+2r} and $E_-\,\varphi_{m+2r}$. Therefore the effect of the terms involving only the r-th power of E_+ and E_- is the same as the effect of Y, and is

$$Y\varphi_{m+2r} = \lambda(r) \sum c_q (im)^q \varphi_{m+2r}, \qquad \lambda(r) \in \mathbf{C},\ \lambda(r) \neq 0$$

Applying E_- yields a non-zero vector for an appropriate value of m. On the other hand, $YE_-\varphi_{m+2r} = 0$. Hence Y and E_- do not commute. Consequently, if $Y \in \mathfrak{Z}(\mathfrak{U})$, and $h(Y) = 0$, then $Y = 0$, as desired.

We exhibit explicitly an element of the center of $\mathfrak{U}$, namely the **Casimir operator**

$$\omega = H^2 + 2(X_+ X_- + X_- X_+).$$

It is verified by a trivial direct computation that ω commutes with $\boldsymbol{H}$, $\boldsymbol{X}_+$, $\boldsymbol{X}_-$ and therefore lies in the center. One can normalize ω by a constant multiple, often to be taken equal to $-1/4$ in order to make the operator "positive", but here we are not interested in this, so we minimize denominators. Let

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

Then Casimir also has the expressions (obtained by direct computation)

$$\boxed{\begin{aligned} \omega &= \boldsymbol{H}^2 + \boldsymbol{V}^2 - \boldsymbol{W}^2 = 2i\boldsymbol{W} - \boldsymbol{W}^2 + \boldsymbol{E}_+\boldsymbol{E}_-, \\ \omega &= -1 - (\boldsymbol{W} - i)^2 + \boldsymbol{E}_+\boldsymbol{E}_-. \end{aligned}} \tag{2}$$

In particular,

$$-h(\omega + 1) = (\boldsymbol{W} - i)^2. \tag{3}$$

Theorem 3. *The center of* $\mathfrak{U}(\mathfrak{g})$ *is the polynomial algebra in one variable* $\mathbf{C}[\omega]$. *Its image under* h *is* $\mathbf{C}[(\boldsymbol{W} - i)^2]$.

Proof. Since h is injective on $\mathfrak{Z}(\mathfrak{U})$ and since we have already seen that $(\boldsymbol{W} - i)^2$ already occurs in the image of the center under h, it will suffice to prove that only polynomials in even powers of $(\boldsymbol{W} - i)$ can occur. This amounts to proving that if Y is in the center of $\mathfrak{U}$ and

$$Y = \sum c_q(\boldsymbol{W} - i)^q + \sum_{p \geqq 1} c_{pq}\boldsymbol{E}_+^p\boldsymbol{W}^q\boldsymbol{E}_-^p,$$

then only even powers of q occur in the first sum with $p = 0$. We look at the action of the universal enveloping algebra arising from the induced representations $\pi = \pi_s$ where $s = -m + 1$, m is an integer $\geqq 2$. Then the finite dimensional space $V(-m + 1)$ generated by the K-eigenvectors

$$\varphi_{-m+2}, \varphi_{-m+4}, \ldots, \varphi_{m-2}$$

is stable under the action of the Lie algebra. Since $d\pi(Y)$ commutes with $d\pi(\mathfrak{g})$, and since $V(-m + 1)$ is irreducible for $d\pi(\mathfrak{g})$, it follows that $d\pi(Y)$ is a scalar multiple of the identity. Looking at the effect of Y on φ_{-m+2}, which is annihilated by $\mathbf{E}_-$, we see that

$$Y\varphi_{-m+2} = P(-m + 1)\varphi_{-m+2}, \qquad \text{where} \quad P(T) = \sum c_q i^q T^q.$$

On the other hand, using the hypothesis that Y commutes with $\boldsymbol{E}_-$, and noting that $E_-\varphi_m$ is a scalar multiple of φ_{m-2}, and is $\neq 0$, we find

$$E_-Y\varphi_m = YE_-\varphi_m = P(-m-1)E_-\varphi_m = E_-P(-m+1)\varphi_m.$$

Since E_- is injective on $H(-m+1)_m$ [the m-th eigenspace of $H(-m+1)$], it follows that $Y\varphi_m = P(-m+1)\varphi_m$. In other words, $P(-m+1)$ is also an eigenvalue of $d\pi(Y)$ on $H(-m+1)_m$.

Now instead of taking $s = -m+1$, take $s = m-1$ and let π be π_{m-1}. Then $d\pi(E_-)$ annihilates $H(m-1)_m$, the space of lowest weight in $H(m-1)$. Note that $W\varphi_m = im\varphi_m$. Consequently

$$Y\varphi_m = \sum c_q i^q (m-1)^q \varphi_m = P(m-1)\varphi_m.$$

It follows that $P(-m+1) = P(m-1)$ for infinitely many m, whence P is an even polynomial, thereby proving our theorem.

Theorem 4. *The centralizer of* $\mathfrak{k}$ *(i.e. of* W*) in* $\mathfrak{U}$ *is the commutative polynomial algebra* $\mathbf{C}[\omega, W]$.

Proof. The expression of ω in terms of W, E_+, E_- shows that the above commutative algebra contains E_+E_-, whence $(E_+E_-)^p$ for every integer $p \geqslant 0$. We then prove by induction that it contains all monomials $E_+^p W^q E_-^p$, using the commutation rules

$$[W, E_+] = 2iE_+, \qquad [W, E_-] = -2iE.$$

By Theorem 2, we see that $\mathbf{C}[\omega, W] = \mathfrak{Z}(\mathfrak{k})$, as desired.

The adjoint representation of G on $\mathfrak{g}$ extends in a natural way to an action on $\mathfrak{U}(\mathfrak{g})$. If

$$X = X_1 \cdots X_n$$

and $X_i = \mathfrak{L}_{X_i}$ with $X_i \in \mathfrak{g}$, then

$$X^g = X_1^g \cdots X_n^g, \qquad \text{where} \quad X_i^g = g^{-1}X_i g.$$

We also write

$$X^g = \mathrm{Ad}_{\mathfrak{U}}(g)X.$$

Lemma 3. *For* $X \in \mathfrak{g}$ *we have*

$$\mathrm{Ad}_{\mathfrak{U}}(\exp X) = \exp(\mathrm{ad}\, X),$$

where $(\mathrm{ad}\, X)(Y) = XY - YX$.

Proof. Entirely similar to the proof of the lemma in Chapter VII, §2. We leave it to the reader.

Theorem 5. *Let $Y \in \mathfrak{U}(\mathfrak{g})$ and assume that Y commutes with some X with $X \in \mathfrak{g}$. Then Y commutes with translation on the right and left by* $\exp X$ *(as an operator on $C^\infty(G, H)$, where H is a Banach space). If Y is in the center of $\mathfrak{U}(\mathfrak{g})$, then Y commutes with all right and left translations.*

Proof. By definition, Y is a left invariant differential operator, and we have only to prove that Y is also right invariant. Let $g = \exp X$. Then

$$f(x \exp(tY)g) = f(xg \exp(tg^{-1}Yg)),$$

and therefore, if R_g is right translation by g, we get

$$YR_g f = R_g Y^g f.$$

By the lemma, we know that $Y^g = \mathrm{Ad}_{\mathfrak{U}}(g)Y = Y$. This proves our theorem.

Remark. The theorem for Banach valued mappings is also a consequence of the theorem for complex valued mappings, by composing the relation to be proved with enough functionals, using the Hahn–Banach theorem.

Let π be a representation of G in a Banach space H. Let v be a C^∞ vector and let f_v be the map

$$f_v(x) = \pi(x)v,$$

so that by definition, f_v is a C^∞ map of G into H.

If Y is in $\mathfrak{U}(\mathfrak{g})$, then we can denote correctly by $d\pi(Y)$ its natural image in the space of endomorphisms of H_π^∞ (the space of C^∞ vectors). However, we shall from now on often abbreviate our notation, and write

$$Yv = d\pi(Y)v.$$

If $X \in \mathfrak{g}$, then from the definitions we obtain at once the formula

$$\mathfrak{L}_X f_v = f_{d\pi(X)v}. \tag{4}$$

By induction and linerarity, this formula extends to arbitrary elements of $\mathfrak{U}(\mathfrak{g})$, and in abbreviated notation, it reads

$$Yf_v = f_{Yv}, \qquad \text{all } Y \in \mathfrak{U}(\mathfrak{g}). \tag{5}$$

In particular, if v is an eigenvector of Y, then f_v is an eigenvector of Y also.

Theorem 6. *Let π be a representation of G on a Banach space H. If*

$Y \in \mathfrak{U}(\mathfrak{g})$ and $X \in \mathfrak{g}$, and if Y commutes with X, then $d\pi(Y)$ commutes with $\pi(\exp X)$ on H_π^∞.

Proof. The theorem is proved just like Theorem 5. We leave it to the reader.

§2. ANALYTIC VECTORS

Let us use the coordinates (x, y, θ) on G as in Chapter VI, §4, where we expressed a group element as

$$g = \begin{pmatrix} u & 0 \\ 0 & u \end{pmatrix}\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Then from the computations of that chapter, we find an expression for the Casimir operator in terms of these coordinates, namely

$$\boxed{\omega = 4y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) - 4y\,\frac{\partial^2}{\partial x\,\partial\theta}\,.} \tag{1}$$

In particular, on functions which are independent of θ on the right, i.e. functions on G/K (the upper half plane), the Casimir operator, up to a constant factor, is the operator

$$y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right).$$

We saw in Theorem 5 that it commutes with right translations on G, and hence that it is an invariant differential operator on the upper half plane. It will be studied in detail in the last chapter of this book.

Furthermore, suppose that f is a C^∞ function on G satisfying

$$f(gk_\theta) = e^{in\theta}f(g).$$

Then $\partial f/\partial\theta = inf$. From formula (1), giving the coordinate expression for Casimir, we see that ω has the same effect on f as the differential operator

$$4y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) - 4yin\,\frac{\partial}{\partial x}\,,$$

which is an elliptic operator (cf. Appendix 4).

***Theorem* 7.** *Let π be a representation of G in a Banach space H. Assume that for some integer n, the space H_n has dimension 1, and let $\{v\}$ be a basis of H_n. Then v is an analytic vector.*

Proof. Let $f_v(g) = \pi(g)v$. Then

$$f_v(gk_\theta) = e^{in\theta}f_v(g).$$

It follows that f_v is an eigenvector of W with eigenvalue in. Since ω commutes with $\pi(k)$ for all $k \in K$, it follows that $\omega v \in H_n$, and since H_n is assumed one dimensional, v is an eigenvector of ω, namely

$$\omega v = cv$$

for some complex c. Thus $(\omega - c)v = 0$, and therefore by (5) of §1,

$$(\omega - c)f_v = 0.$$

Therefore f_v is the solution of an elliptic differential equation. For any functional λ on H, the function $\lambda \circ f_v$ is a solution of the same elliptic equation, and is therefore analytic, by the regularity theorem (Appendix 4 and the analytic reference). Since weak analyticity implies strong analyticity (Appendix 5), we conclude that f_v is analytic, as was to be shown.

§3. EIGENFUNCTIONS OF $\mathfrak{Z}(\mathfrak{k})$

We first consider spherical functions as in Chapter IV. We recall that a spherical function is defined to be a continuous function on G, which is bi-invariant under K; is an eigenfunction of convolution with elements of $C_c^\infty(G//K)$ on the right; and is normalized to take the value 1 at the unit element of G. By the convolution property, a spherical function is obviously C^∞.

***Theorem* 8.** *Let f be a C^∞ function on G, bi-invariant under K and taking the value 1 at e. Then f is a spherical function if and only if there exists $\lambda \in \mathbf{C}$ such that $\omega f = \lambda f$.*

Proof. Assume first that f is a spherical function. Let $\varphi \in C_c^\infty(G//K)$. By definition,

$$f * \varphi(x) = \int_G f(xy^{-1})\varphi(y)\, dy = \int_G f(y^{-1})\varphi(yx)\, dy.$$

Since ω is composed of left invariant differential operators, it is clear that

$$\omega(f * \varphi) = f * (\omega\varphi).$$

On the other hand, since ω is also right invariant, we get

$$\omega(f * \varphi) = (\omega f) * \varphi.$$

(The differentiations under the integral sign are obviously legitimate since φ has compact support.) Let $\{\psi_n\}$ be a Dirac sequence in $C_c^\infty(G)$ and let $\varphi_n = {}^K\psi_n^K$ be its average on the left and right by K. Then

$$\omega f * \varphi_n = f * \omega\varphi_n = \lambda_n f,$$

where

$$\lambda_n = (f * \omega\varphi_n)(e) = (\omega f * \varphi_n)(e).$$

Since $\omega f * \varphi_n$ converges pointwise to ωf, it follows that $\{\lambda_n\}$ converges to $\omega f(e)$, and we see that $\omega f = \lambda f$, where

$$\lambda = \omega f(e),$$

thus proving half of our theorem.

The converse will follow from a characterization of eigenfunctions of the Casimir operator, which we now discuss. The treatment of the bi-invariant case will follow from results on functions which are conjugate invariant under K, so we look at conjugation by K.

We recall that an eigenfunction of the Casimir operator which is right invariant under K (i.e. a function on G/K) is necessarily analytic, by the regularity theorem for elliptic equations (cf. §2)).

We had defined the adjoint representation of G on $\mathfrak{U}(\mathfrak{g})$. If

$$\boldsymbol{X} = \boldsymbol{X}_1 \cdots \boldsymbol{X}_n$$

and $\boldsymbol{X}_i = \mathfrak{L}_{X_i}$ with $X_i \in \mathfrak{g}$, then

$$\boldsymbol{X}^g = \boldsymbol{X}_1^g \cdots \boldsymbol{X}_n^g, \qquad \text{where} \quad X_i^g = g^{-1}X_i g.$$

The association $\boldsymbol{Y} \mapsto \boldsymbol{Y}^g$ is an algebra automorphism of $\mathfrak{U}(\mathfrak{g})$ for every $g \in G$.

In particular, taking $g = k \in K$, we can average an element $\boldsymbol{Y} \in \mathfrak{U}(\mathfrak{g})$ over K; that is, let

$$\boldsymbol{Y}^K = \int_K \boldsymbol{Y}^k \, dk.$$

Then it is trivially verified that $\boldsymbol{Y}^K \in \mathfrak{Z}(\mathfrak{k})$, i.e. $\boldsymbol{Y}^K$ commutes with $\boldsymbol{W}$.

For any function f we let f^k be the function such that

$$f^k(x) = f(kxk^{-1}).$$

We also let $x^g = g^{-1}xg$. Let $X \in \mathfrak{g}$. We have

$$f^k(x \exp(tX)) = f(kx \exp(tX)k^{-1})$$
$$= f(kxk^{-1} \exp(tkXk^{-1})).$$

From this and induction we obtain the formula

$$\boxed{Y^k f^k(x^k) = Yf(x)} \tag{1}$$

for all $Y \in \mathfrak{U}(\mathfrak{g})$. In particular, if $f \in C^\infty(G, K)$, then

$$Y^K f(e) = Yf(e). \tag{2}$$

Theorem 9. *Let $\lambda: \mathfrak{Z}(\mathfrak{k}) \to \mathbf{C}$ be a character, i.e. an algebra homomorphism. Let f be an analytic function on G such that f is invariant under conjugation by K, i.e.*

$$f(kxk^{-1}) = f(x), \qquad x \in G,\ k \in K,$$

and such that

$$Yf = \lambda(Y)f$$

for all $Y \in \mathfrak{Z}(\mathfrak{k})$. Then for all small $X \in \mathfrak{g}$, we have

$$f(\exp X) = \sum_{m \geqq 0} \frac{\lambda\big((X^m)^K\big)}{m!} f(e).$$

Proof. This is an immediate consequence of Taylor's formula

$$f(\exp X) = \sum \frac{X^m f(e)}{m!},$$

and formula (2) above, together with the assumption that f is an eigenvector for $\mathfrak{Z}(\mathfrak{k})$ with eigencharacter λ.

Theorem 9 gives us the value of f near the origin in G, and hence determines f on all of G by analyticity. In particular:

Theorem 10. (*i*) *If two functions on G are conjugate invariant under K and are eigenfunctions of $\mathfrak{Z}(\mathfrak{k})$ with the same character, normalized to have the value 1 at the origin, then they are equal.*

(*ii*) *If two functions on G are bi-invariant under K, are eigenfunctions of the*

Casimir operator with the same eigenvalue, and are normalized to have the value 1 *at the origin, then they are equal.*

Proof. The first part has already been proved. The second follows from the next remarks.

Let $f \in C^{\infty}(G//K)$ be a smooth bi-invariant function. Then

$$Wf = 0.$$

If at the same time f is an eigenfunction of ω, then by Theorem 4, for every $Y \in \mathfrak{Z}(\mathfrak{k}) = \mathbf{C}[\omega, W]$ there exists a complex number $\lambda(Y)$ such that

$$Yf = \lambda(Y)f,$$

and the map

$$Y \longmapsto \lambda(Y)$$

is an algebra homomorphism of $\mathfrak{Z}(\mathfrak{k})$ into $\mathbf{C}$, i.e. a character of $\mathfrak{Z}(\mathfrak{k})$. Furthermore, since $\lambda(W) = 0$, such a character is determined by its value on the Casimir operator, i.e. by $\lambda(\omega)$.

The analyticity of the functions as eigenfunctions of the Casimir operator can either be assumed, or be observed to follow from the regularity theorem for elliptic equations.

We also know enough to conclude the proof of Theorem 8. Indeed, our explicit construction of spherical functions

$$f_s(x) = \int_K \rho(kx)^s \, dk$$

provides us with bi-invariant functions which are eigenfunctions of Casimir with arbitrary eigenvalues $s(s - 1) = \lambda$, taking the value 1 at the origin. In addition to finishing the proof of Theorem 8, we also have finished Harish-Chandra's proof of his classification of spherical functions:

Theorem 11. *The only spherical functions are the ones which we have already exhibited, i.e. the functions f_s above, $s \in \mathbf{C}$.*

Remark. One can take another approach to the classification of spherical functions as follows. We start with the second order linear differential equation for a bi-invariant function f,

$$\omega f = s(s - 1)f.$$

Since f is bi-invariant, we can view f as a function of the A-variable only in the Cartan decomposition $G = KA^{+}K$. The above differential equation is a second order linear differential equation, and it has two linearly independent solutions. Given an eigenvalue λ and a spherical function f such that $\omega f = \lambda f$, we can find s such that $\lambda = s(s - 1)$. Thus the special spherical function f_s is a solution of the same differential equation as f. [Solutions of this equation are classical functions.] In Chapter XIV, §2, Theorem 1, we shall exhibit another solution in terms of another variable, but which is seen to have a logarithmic singularity in terms of the variable y in the upper half-plane representation (actually in terms of the variable u as described there). Since f must be a linear combination of f_s and the function exhibited in Chapter XIV, §2, and since f does not have a singularity, it follows that f is a constant multiple of f_s, as desired.

XI The Weil Representation

There is a whole aspect of $SL_2(\mathbf{R})$ into which we shall not go, namely the various models which can be found in an infinitesimal equivalence class of representations, and the possibility of finding canonical models, e.g. the Whittaker model in such a class. We refer the reader to Jacquet–Langlands [Ja, La], Knapp–Stein [Kn, St], and Stein [St 2] for more information in this direction, and a discussion of intertwining operators among various models. Helgason [He 3] gives a particularly interesting model of representations in eigenspaces of the Laplacian. I include here just the special model of the Weil representation because of its particular interest in number theoretic applications, and the possibility of constructing automorphic forms with it, as in Shalika–Tanaka [Sh, Ta]. Besides, since Weil's *Acta* paper [We] is written in an extremely general context, it may be useful to have a naive treatment of the special case as an introduction. Finally, the way the Weil representation is constructed provides an excuse for giving generators and relations for SL_2, and for mentioning the Bruhat decomposition. I did not want to get very much involved in the matters discussed here, and so the chapter is somewhat arbitrary.

§1. SOME CONVOLUTIONS

For $b \in \mathbf{R}$, $b \neq 0$, consider the function on $\mathbf{R}^n$ defined by

$$h_b(x) = e^{-\pi i b x^2},$$

where $x^2 = x \cdot x$ is the dot product. This function has absolute value 1, and hence its Fourier transform does not exist in a "naive" sense. However, by a suitable limiting procedure, we shall be able to operate with it just as with functions in the Schwartz space in the ordinary theory of convolutions and Fourier transforms. (Cf. *Real Analysis*, Chapter XIV. We assume that the reader is acquainted with this elementary reference and will use the concepts and theorems there without further reference.) Observe that h_b is an even function,

$$h_b(-x) = h_b(x).$$

We shall show that defining

h 1.
$$\hat{h}_b = \frac{1}{(ib)^{n/2}} h_{-1/b}$$

is "natural". Furthermore, we shall prove that if f is in the Schwartz space, then the function $h_b * f$ defined by the usual formula

$$h_b * f(x) = \int_{\mathbf{R}^n} h_b(x-y)f(y)\,dy$$

is in the Schwartz space, so is $h_b f$, and the following formalism is satisfied, the Fourier transform of f being defined in the usual manner,

$$\hat{f}(x) = \int_{\mathbf{R}^n} f(y)e^{-2\pi i x\cdot y}\,dy.$$

h 2.
$$(h_b * f)\hat{} = \hat{h}_b \hat{f}.$$

h 3.
$$(h_b f)\hat{} = \hat{h}_b * \hat{f}.$$

h 4.
$$h_b * f = h_b((h_b f)\hat{} \circ M_{-b})$$

where, for $b \in \mathbf{R}$, $b \neq 0$ we defined $M_b(x) = bx$ (multiplication by b), and $f \circ M_b(x) = f(bx)$. In particular,

h 5.
$$h_{-b}(h_b * f) = (h_b f)\hat{} \circ M_{-b}.$$

In other words, our definition **h 1** is such that all the usual properties of the Fourier transform and convolution are satisfied by the function h_b.

We now proceed to prove this. To avoid a subscript, we fix b and let

$$\varphi(x) = e^{-\pi i b x^2}.$$

For $a \in \mathbf{R}$, $a > 0$ we define

$$\varphi_a(x) = e^{-\pi i b x^2} e^{-\pi a x^2}.$$

Thus we multiply φ by one of the standard functions in the Schwartz space. It is then clear that φ_a is in the Schwartz space, and we can take its Fourier transform,

$$\hat{\varphi}_a(y) = \int e^{-\pi(a+ib)x^2} e^{-2\pi ixy}\, dx.$$

Let $g(x) = e^{-\pi x^2}$. Then g is self dual, $\hat{g} = g$. For α complex let

$$g_\alpha(x) = e^{-\pi\alpha x^2}.$$

If α is real > 0, then

$$\hat{g}_\alpha(x) = \frac{1}{\alpha^{n/2}} e^{-\pi x^2/\alpha}.$$

By analytic continuation, this same relation holds for α complex with positive real part, i.e. $\alpha = a + ib$, $a > 0$. Therefore

$$\hat{\varphi}_a(x) = \frac{1}{(a + ib)^{n/2}} e^{-\pi x^2/(a+ib)}.$$

It is then natural to look at the limit as $a \to 0$. The limit cannot be taken under the integral sign, but it can be taken in this last expression, and yields our definition for $\hat{\varphi} = \hat{h}_b$.

Let $f \in S$. Then $f * \varphi \in S$.

Proof. We have

$$(f * \varphi_a)^\wedge = \hat{f}\hat{\varphi}_a.$$

We shall prove below that the limit as $a \to 0$ can be taken under the integral sign, and we obtain

$$(f * \varphi)^\wedge = \hat{f}\hat{\varphi}.$$

It is clear that the ordinary product $\hat{f}\hat{\varphi}$ is in S, and taking Fourier transforms shows that $f * \varphi$ is the Fourier transform of a function in the Schwartz space. Consequently it lies in the Schwartz space, as desired.

We also have the analogous relation

$$(f\varphi)^\wedge = \hat{f} * \hat{\varphi},$$

replacing f by $\hat{f}$ and φ by $\hat{\varphi}$.

We must now justify the limit under the integral sign

$$\lim_{a\to 0} \int (f * \varphi_a)(x) e^{-2\pi ixy}\, dx = \int (f * \varphi)(x) e^{-2\pi ixy}\, dx.$$

For this we need to find a function in $\mathfrak{L}^1$ which dominates $f * \varphi_a$ in absolute value independently of a, and it suffices to prove that the function

$$x \longmapsto x^p(f * \varphi_a)(x)$$

is bounded for all p. Let M^p be multiplication by the monomial, that is

$$M^p f(x) = x_1^{p_1} \cdots x_n^{p_n} f(x).$$

By the theory of Fourier transforms, loc. cit., we have

$$\begin{aligned} M^p(f * \varphi_a) &= M^p\left(\left(\hat{f}^- \hat{\varphi}_a^-\right)^\wedge\right) \\ &= \left(D^p\left(\hat{f}^- \hat{\varphi}_a^-\right)\right)^\wedge. \end{aligned}$$

We may replace f by f^-. Also we have $(\hat{\varphi}_a)^- = (\varphi_a^-)^\wedge$ and $\varphi_a^- = \varphi_a$. Hence it suffices to prove that

$$\left(D^p\left(\hat{f}\hat{\varphi}_a\right)\right)^\wedge$$

is bounded. But $D^p(\hat{f}\hat{\varphi}_a)$ is a sum of terms,

$$\sum c_{qr} D^q \hat{f} D^r \hat{\varphi}_a$$

and $D^q \hat{f} = f_1$ is in the Schwartz space. Also,

$$D^r \hat{\varphi}_a(x) = \frac{1}{(a+ib)^{n/2}} \left(\frac{-2\pi}{a+ib}\right)^r x^r e^{-\pi x^2/(a+ib)}.$$

We have the estimate

$$\begin{aligned} \left|D^q f(x) x^r e^{-\pi x^2 a/(a^2+b^2)}\right| &\leqslant |D^q f(x)||x|^r \\ &\leqslant |f_2(x)|, \end{aligned}$$

where f_2 is in S. Hence each function

$$f_{q,r} = \left(D^q \hat{f}\right)\left(D^r \hat{\varphi}_a\right)$$

is bounded by a function g in S independent of a, and

$$\begin{aligned} |\hat{f}_{q,r}(x)| &\leqslant \int |f_{q,r}(y) e^{-2\pi i x y}|\, dy \\ &\leqslant \|g\|_1. \end{aligned}$$

This justifies taking the desired limit under the integral sign.

The formal properties stated at the beginning of the section which have not already been proved are then immediate.

§2. GENERATORS AND RELATIONS FOR SL_2

Let F be a field. We shall give generators and relations for $SL_2(F)$. If we put for $b \in F$ and $a \in F$, $a \neq 0$:

$$u(b) = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \quad s(a) = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, \quad w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

then

$$(*) \qquad s(a) = wu(a^{-1})wu(a)wu(a^{-1}).$$

SL 1. *u is an additive homomorphism.*

SL 2. *s is a multiplicative homomorphism.*

SL 3. $w^2 = s(-1)$.

SL 4. $s(a)u(b)s(a^{-1}) = u(ba^2)$.

Suppose that G is an arbitrary group with generators $u(b)$ ($b \in F$) and w, such that if we define $s(a)$ for $a \neq 0$ by (*), then conditions **SL 1** through **SL 4** are satisfied. Then **SL 3** and **SL 4** show that $s(-1)$ is in the center, and $w^4 = 1$. Furthermore,

SL 5. $ws(a) = s(a^{-1})w$.

Indeed,

$$w^{-1}s(a) = u(a^{-1})wu(a)wu(a^{-1}),$$

whence

$$\begin{aligned} ws(a)w^{-1}s(a) &= ws(a)u(a^{-1})s(a^{-1})u(-a)w^{-1} \\ &= 1 \qquad [\text{using (4) with } b = a^{-1}], \end{aligned}$$

as desired.

Let G be the free group with generators $u(b)$, w and relations **SL 1** through **SL 4**, defining $s(a)$ as in (*). Then we have a natural homomorphism from G onto $SL_2(F)$.

Theorem 1. *The natural homomorphism of the above free group onto $SL_2(F)$ is an isomorphism.*

Proof. Any group with generators $u(b)$ ($b \in F$), w, and $s(a)$ defined by (*), satisfying our stated relations, whether it is the free group or not, consists of all elements of the form

$$u(b)s(a) \qquad \text{or} \qquad u(b)s(a)wu(c).$$

Indeed, let N be the subgroup of all elements $u(b)$ and let A be the subgroup of all elements $s(a)$. Then **SL 4** shows that $NA = AN$. Let G' be the subset $G' = NA \cup NAwN$. Multiplying NA on the right by w or N maps NA into G'. Multiplying $NAwN$ by N on the right does the same thing. From the identity

$$wu(c)w = u(-c^{-1})w^{-1}s(c)u(-c^{-1})$$

(immediate from the definition of $s(c)$), and **SL 5**, we conclude that

$$NAwNw \subset NAwN.$$

Hence $G' = G$. To show that $SL_2(F)$ is the free group, it now suffices to prove the next lemma.

Lemma. *Every element of $SL_2(F)$ has a unique decomposition of the form*

$$u(a)s(b) \qquad \text{or} \qquad u(a)s(b)wu(c),$$

i.e. $SL_2(F)$ can be decomposed uniquely into the products

$$NA \cup NAwN.$$

Proof. Let F^2 be the vector space of row vectors, and let $SL_2(F)$ operate as matrices on the right of F^2. Let e_2 be the unit vector

$$e_2 = (0, 1).$$

Then we see at once that N is the isotropy group of e_2, i.e. is the set of matrices $g \in SL_2(F)$ such that $e_2 g = e_2$. Hence $N \backslash SL_2(F)$ is in bijection with its orbit of e_2. The image of an element $s(a)$ is

$$e_2 s(a) = (0, a^{-1}),$$

and consists of those vectors whose first component is 0. The element a is uniquely determined. If we multiply this by w on the right, we get $(-a^{-1}, 0)$. Multiplying further by $u(c)$ on the right yields

$$(-a^{-1}, -ca^{-1}).$$

This yields a vector with non-zero first coordinate, determining a uniquely, and then c is uniquely determined by the second coordinate. This proves that the decomposition is unique, and proves the lemma, as well as Theorem 1.

The decomposition of the lemma is called the **Bruhat decomposition**.

Remark. Over an arbitrary field, there is no question of having positive or negative elements in the field, and so A was taken to be the full image of the multiplicative group of F under the homomorphism s. Over the reals, of course, we have the slightly finer distinction arising from the parity.

§3. THE WEIL REPRESENTATION

Theorem 2. *Let $S = S(\mathbf{R}^2)$ be the Schwartz space of $\mathbf{R}^2$. There exists a unique algebraic representation r of $SL_2(\mathbf{R})$ on S (so we impose no continuity condition) satisfying the following properties:*

$$r(w)f = -i\hat{f}; \tag{1}$$

$$r(u(b))f = h_b f \quad \text{where} \quad h_b(x) = e^{-\pi i b x^2} \qquad b \in \mathbf{R} \tag{2}$$

This representation also satisfies

$$r(s(a))f(x) = af(ax), \qquad a \neq 0. \tag{3}$$

It is unitary for the ordinary hermitian product on S.

Proof. We have to check that the above operations defined in terms of a, b, w satisfy the relations of §2. Note that

$$r(w^2)f = -f^-.$$

The easiest way to proceed is to verify first that if we define $r(s(a))$ by formula (3), then the relation amounting to (*),

$$r(w)r(u(b))r(w)r(u(b^{-1})) = r(u(-b^{-1}))r(w^{-1})r(s(b))$$

holds. This consists in using the properties **h 1** through **h 4** of §1, and is a routine matter, albeit a little tedious. The rest of the theorem is then clear, since with the above definition of $r(s(a))$, we see immediately, for instance, that $a \mapsto r(s(a))$ is a homomorphism.

We now identify $\mathbf{R}^2$ with the complex numbers $\mathbf{C}$. Let $\mathbf{C}^1$ be the group of complex numbers of absolute value 1. Let χ be a character of $\mathbf{C}^1$, and let

$$S(\mathbf{C}, \chi)$$

be the subspace of $S(\mathbf{C})$ consisting of those functions f satisfying

$$f(\alpha z) = \chi(\alpha)^{-1} f(z), \qquad \alpha \in \mathbf{C}^1, z \in \mathbf{C}.$$

Lemma 1. *$S(\mathbf{C}, \chi)$ is stable under the representation r.*

Proof. Let α' denote the complex conjugate of α. Then

$$h_b(\alpha z) = h_b(z)$$

if $\alpha\alpha' = 1$, and so $r(u(b))f$ lies in $S(\mathbf{C}, \chi)$ if f does. Also,

$$\hat{f}(\alpha x) = \int f(y) e^{-\pi i \operatorname{Tr}(\alpha x y')} dy.$$

Let $y \mapsto \alpha y$. The measure on $\mathbf{R}^2 = \mathbf{C}$ is unchanged since $|\alpha| = 1$. Hence

$$\hat{f}(\alpha x) = \int f(\alpha y) e^{-\pi i \operatorname{Tr}(xy')} dy$$

$$= \chi(\alpha)^{-1} \hat{f}(x),$$

thus proving the stability of $S(\mathbf{C}, \chi)$ under the representation.

We want to extend the representation r to $GL_2^+(\mathbf{R})$, and first make some remarks on abstract algebra.

Let G be a group and let A, B be subgroups such that $G = AB$. Let $\varphi: A \longrightarrow G_1$ and $\psi: B \longrightarrow G_1$ be homomorphisms into another group. Assume that whenever an element $g \in G$ can be written

$$g = ab = b'a'$$

with $a, a' \in A$ and $b, b' \in B$, then

$$\varphi(a)\psi(b) = \psi(b')\varphi(a'),$$

and also the restrictions of φ, ψ to $A \cap B$ are equal. Then we can extend φ, ψ uniquely to a homomorphism h of G into G_1, by defining

$$h(ab) = \varphi(a)\psi(b).$$

The verification is trivial and left to the reader.

Suppose in addition that B is normal in G and $A \cap B = \{1\}$. Then the relation $ab = b'a'$ implies

$$aba^{-1} = b'a'a^{-1},$$

whence $a' = a$ and $ab = b'a$. Thus for the existence of the extension h above, it suffices to verify that if $b' = aba^{-1}$, then

$$\varphi(a)\psi(b)\varphi(a)^{-1} = \psi(b').$$

We apply this to our representation r. Let ω be a character of $\mathbf{C}^*$ (not necessarily of absolute value 1) restricting to χ on $\mathbf{C}^1$. For $a > 0$ let

$$v(a) = \begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}.$$

Define

$$r_\omega(v(a))f(z) = |\alpha|\omega(\alpha)f(\alpha z) \tag{4}$$

for any complex number α such that $\alpha\alpha' = a$. This is well defined, i.e. independent of the choice of α. We contend that r_ω is a homomorphism of $GL_2^+(\mathbf{R})$ into the group of algebraic automorphisms of $S(\mathbf{C}, \omega)$. In view of the remarks on algebra which we just made, it suffices to prove that for $x \in SL_2(\mathbf{R})$,

$$r_\omega\left(\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}\right) r(x) r_\omega\left(\begin{pmatrix} a^{-1} & 0 \\ 0 & 1 \end{pmatrix}\right) = r\left(\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix} x \begin{pmatrix} a^{-1} & 0 \\ 0 & 1 \end{pmatrix}\right).$$

If this relation is proved for $x, x' \in SL_2(\mathbf{R})$, it follows for xx'. Hence it suffices to prove the relation for

$$x = \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad x = w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

It is then immediate from the definitions, using the expressions

$$\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} a^{-1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

$$\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a^{-1} & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & ab \\ 0 & 1 \end{pmatrix}.$$

We also observe that r_ω is unitary, i.e. each operator $r_\omega(x)$ for $x \in GL_2^+(\mathbf{R})$ is unitary, with respect to the ordinary hermitian product on $S(\mathbf{C})$.

Theorem 3. *Let $L^2(\mathbf{C}, \omega)$ be the completion of $S(\mathbf{C}, \omega)$ with respect to the L^2-norm. Then $L^2(\mathbf{C}, \omega)$ is irreducible for r_ω.*

Proof. The algebraic subspace $S(\mathbf{C}, \omega)$ is dense in $L^2(\mathbf{C}, \omega)$. Map the space $S(\mathbf{C}, \omega)$ into a function space on $\mathbf{C}$ by the mapping T such that

$$Tf(z) = |z|\omega(z)f(z). \tag{5}$$

Note that

$$Tf(\epsilon z) = Tf(z)$$

for any complex number ϵ of absolute value 1. Thus Tf has the advantage of being a function of the distance from the origin, and can therefore be viewed as a function on $\mathbf{R}^+$.

Furthermore, let

$$r_\omega^T = T \circ r_\omega \circ T^{-1},$$

i.e.

$$r_\omega^T(x)f = T(r_\omega(x)(T^{-1}f)).$$

Then on a function f such that $f(\epsilon z) = f(z)$ for all ϵ with $|\epsilon| = 1$, we have the operation of r_ω^T described by

$$r_\omega^T\left(\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}\right)f(z) = e^{-\pi ibzz'}f(z), \tag{6}$$

$$r_\omega^T\left(\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}\right)f(z) = f(\alpha z) \tag{7}$$

for any complex α such that $\alpha\alpha' = a > 0$.

We view Tf as a function Tf^* on $\mathbf{R}^+$ by defining

$$(Tf)^*(t) = Tf(t^{1/2}) = Tf(z) \tag{8}$$

for any complex z such that $zz' = t$. Then one verifies at once

***Lemma* 2.** *The association*

$$f \mapsto (Tf)^*$$

is a unitary isomorphism with respect to the hermitian product taken with Lebesgue measure on $\mathbf{C}$ *and* $\pi dt/t$ *on* $\mathbf{R}^+$.

Proof. Immediate, and left to the reader.

We have now shifted the study of r_ω to the study of the representation r_ω^T on the space $TS(\mathbf{C}, \omega)$, which is dense in $L^2(\mathbf{R}^+)$, with respect to the measure dt/t, i.e. multiplicative Haar measure.

Let $B^+(\mathbf{R})$ be the group of matrices

$$\begin{pmatrix} a & b \\ 0 & 1 \end{pmatrix}, \quad a > 0, \quad b \in \mathbf{R}.$$

We denote the restriction of r_ω^T to $B^+(\mathbf{R})$ by π. It is an algebraic representation of $B^+(\mathbf{R})$, which extends to a unitary representation on $L^2(\mathbf{R}^+)$, given by

$$\pi\left(\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}\right) f(t) = e^{-\pi i b t} f(t), \tag{9}$$

$$\pi\left(\begin{pmatrix} a & 0 \\ 0 & 1 \end{pmatrix}\right) f(t) = f(at). \tag{10}$$

In order to prove Theorem 3, it suffices therefore to prove

Theorem 4. *The representation π on $B^+(\mathbf{R})$ defined by* (9) *and* (10) *on $L^2(\mathbf{R}^+)$ is irreducible.*

The proof comes from a couple of abstract nonsense lemmas in L^2 theory.

Lemma 3. *Let (X, μ) be a σ-finite measure space. Let*

$$A\colon L^2(X, \mu) \longrightarrow L^2(X, \mu)$$

be a continuous linear map which commutes with multiplication by all functions in $\mathfrak{L}^\infty(X, \mu)$. Then $A = M_g$ for some $g \in \mathfrak{L}^\infty(X, \mu)$, where M_g is multiplication by g, that is, $M_g f = gf$.

Proof. Let $\varphi \in \mathfrak{L}^2(X, \mu)$ be an essentially positive function. For instance, decompose

$$X = \bigcup X_n$$

as a disjoint union of sets of finite measure, and let φ on X_n be the constant function

$$\frac{1}{n^2 \mu(X_n)^{1/2}}.$$

For any $f \in \mathfrak{L}^\infty \cap \mathfrak{L}^2$ we have $A(\varphi f) = \varphi A(f) = f A(\varphi)$. Hence

$$Af = \frac{A\varphi}{\varphi} f.$$

Let $g = A\varphi/\varphi$. It suffices to prove that g is bounded. If g is not bounded, given N there exists a set E with measure

$$0 < \mu(E) < \infty$$

such that $|g(x)| > N$ for all $x \in E$. Let χ_E be the characteristic function of E.

Let

$$f = \frac{\bar{g}}{|g|}\chi_E$$

whenever $g \neq 0$, and 0 whenever $g = 0$. Then

$$A\left(\frac{\bar{g}}{|g|}\chi_E\right) = \frac{g\bar{g}}{|g|}\chi_E = |g|\chi_E \geqslant N\chi_E.$$

The L^2-norm of this last function is $N\mu(E)^{1/2}$. But

$$\left\|\frac{\bar{g}}{|g|}\chi_E\right\|_2^2 = \int_E \left(\frac{|\bar{g}|}{|g|}\right)^2 d\mu \leqslant \mu(E).$$

This contradicts the boundedness of A, and proves our lemma.

Lemma 4. *Let* $\mathbf{R}^+$ *be taken with its Haar measure* dt/t. *Let*

$$A: L^2(\mathbf{R}^+) \longrightarrow L^2(\mathbf{R}^+)$$

be a bounded linear map, which commutes with multiplication by all functions e^{ibt} *for all* $b \in \mathbf{R}$. *Then* A *commutes with* M_φ *for all* $\varphi \in \mathcal{L}^\infty(\mathbf{R}^+)$. *Hence* $A = M_g$ *for some* $g \in \mathcal{L}^\infty(\mathbf{R}^+)$ *by Lemma* 3.

Proof. We first prove the assertion of the theorem when $\varphi \in C_c^\infty(\mathbf{R}^+)$. Let N be a large integer, and let ψ_N be the extension of φ to $\mathbf{R}^+$ by periodicity over the interval $(0, N]$, so that the graph of ψ_N looks like this.

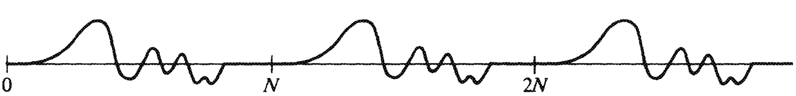

Figure 1

Let E be an interval not containing 0 and contained in $(0, N]$. Let χ_E be its characteristic function. Then

$$A(\varphi\chi_E) = A(\psi_N\chi_E) = \psi_N A(\chi_E).$$

Keeping E fixed and letting $N \longrightarrow \infty$ we get

$$A(\varphi\chi_E) = \varphi A(\chi_E).$$

This is true for all choices of intervals E in $\mathbf{R}^+$ not containing 0. Linear combinations of characteristic functions of such intervals are dense in L^2. By continuity, it follows that

$$A(\varphi f) = \varphi A(f)$$

for all $f \in \mathfrak{L}^2$.

We now extend the validity of this relation to all bounded measurable functions. Let $\varphi \in \mathfrak{L}^\infty$. There exists a sequence $\{\varphi_n\}$ in $C_c^\infty(\mathbf{R}^+)$ such that $\varphi_n \longrightarrow \varphi$ almost everywhere, and such that the functions φ_n are uniformly bounded. (For instance, approximate φ first over a finite interval $[1/N, N]$, and let $N \longrightarrow \infty$.) If $f \in \mathfrak{L}^2$, then $\varphi_n f \longrightarrow \varphi f$ in L^2 by the dominated convergence theorem. Hence

$$\varphi_n Af = A(\varphi_n f) \longrightarrow A(\varphi f)$$

and also

$$\varphi_n Af \longrightarrow \varphi Af.$$

This proves our lemma.

To prove the desired irreducibility of Theorem 4, we consider the projection A on an invariant subspace. Write $f \circ a$ for the function such that $(f \circ a)(x) = f(ax)$. Then by Lemma 3, for given $a > 0$,

$$A(f \circ a) = g(f \circ a) = (Af) \circ a$$

or in other words,

$$A(f \circ a)(t) = g(t)f(at) = g(at)f(at).$$

This implies that $g(t) = g(at)$ for almost all t. By Fubini's theorem (in both directions), the function $g(at) - g(t)$ is 0 for almost all (a, t). Hence g is essentially constant, and this implies that A is the identity, as was to be shown.

XII Representation on ${}^0L^2(\Gamma \backslash G)$

The algebraic and arithmetic properties of SL_2 begin to be felt when we consider the representation on $\Gamma \backslash G$ for some discrete subgroup Γ. In this chapter, after a general discussion of the nature of the factor space $\Gamma \backslash G$ or $\Gamma \backslash G/K = \Gamma \backslash \mathfrak{H}$, which is essentially classical, we prove that on a certain subspace ${}^0L^2(\Gamma \backslash G)$ of $L^2(\Gamma \backslash G)$ the representation is completely reducible when $\Gamma = SL_2(\mathbf{Z})$. The method works just as well for any "arithmetic" subgroup, i.e. a subgroup of finite index in $SL_2(\mathbf{Z})$. It uses the Poisson summation formula, in addition to some estimates. It has the advantage of being very rapid and of using a minimum of analysis.

On the other hand, the method fails for more general discrete subgroups, and a discussion of that situation is given in Chapter XIV, where the general case is treated by a method of Faddeev.

The main result of this chapter, the complete reducibility of ${}^0L^2(\Gamma \backslash G)$, is due to Gelfand-Graev-Pjateckii-Shapiro [Ge, Gr], where an adelic version is given. I shall follow essentially without change Godements's article [Go 5]. Godement treats a more general case, but in line with our general policy, we feel that it is easier to read a proof first for $SL_2(\mathbf{Z}) = \Gamma$, and then observe that the proof holds in greater generality.

§1. CUSPS ON THE GROUP

Let Γ be a discrete subgroup of $SL_2(\mathbf{R})$. A one-parameter subgroup N of $G = SL_2(\mathbf{R})$ is called **unipotent** if N is the subgroup consisting of all elements

$$\exp tX$$

for some nilpotent 2×2 real matrix X, and all $t \in \mathbf{R}$. We say that N is **cuspidal** for Γ if $N/N \cap \Gamma$ is compact, or equivalently, N contains a non-identity element of Γ.

The typical example of a unipotent subgroup is the usual N_0 consisting of all matrices

$$\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}, \qquad b \in \mathbf{R}.$$

Straightforward matrix multiplication shows that the normalizer of N_0 in $GL_2(\mathbf{R})$ consists of all triangular matrices

$$\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}, \qquad ad \neq 0.$$

We let $\mathrm{Norm}(N)$ denote the normalizer of N in $SL_2(\mathbf{R})$.

Theorem 1. *All the one-parameter unipotent subgroups of $SL_2(\mathbf{R})$ are conjugate. Let N, N' be two unipotent subgroups. We have $N = N'$ if and only if* $\mathrm{Norm}(N) = \mathrm{Norm}(N')$.

Proof. Given X nilpotent, there exists a matrix $M \in GL_2(\mathbf{R})$ such that

$$MXM^{-1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

by the Jordan normal form theorem. We can change M by

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

if necessary, so that without loss of generality we may assume that M lies in $GL_2^+(\mathbf{R})$. Adjustment by a positive scalar even allows us to assume that M lies in $SL_2(\mathbf{R})$, so that

$$MXM^{-1} = \begin{pmatrix} 0 & \pm 1 \\ 0 & 0 \end{pmatrix}.$$

Then the group of elements $\exp(tMXM^{-1})$ with $t \in \mathbf{R}$ is conjugate to $\{\exp(tX)\}$ in $SL_2(\mathbf{R})$. This shows that all unipotent one-parameter subgroups of $SL_2(\mathbf{R})$ are conjugate in $SL_2(\mathbf{R})$.

Let B be the Borel subgroups of triangular matrices in $SL_2(\mathbf{R})$, namely B consists of all matrices

$$\begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix}.$$

The set of one-parameter unipotent subgroups is the orbit of N_0 under conjugation by $SL_2(\mathbf{R})$, and is therefore in bijection with the coset space

$$SL_2(\mathbf{R})/B.$$

For each $x \in \mathbf{R}$ let

$$N_x = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} N_0 \begin{pmatrix} 1 & 0 \\ -x & 1 \end{pmatrix}$$

be the conjugation of N_0 by

$$\begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix}.$$

The group N_x consists of all matrices

$$(1) \qquad \begin{pmatrix} 1 - ux & u \\ x^2u & 1 + ux \end{pmatrix}, \qquad u \in \mathbf{R}.$$

Since any 2×2 matrix can be written in the form

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ x & 1 \end{pmatrix} \begin{pmatrix} \alpha & \beta \\ 0 & \delta \end{pmatrix}$$

with $x = c/a$, provided $a \neq 0$, we see that the groups N_x $(x \in \mathbf{R})$ form a set of representatives for those cosets of $SL_2(\mathbf{R})/B$ containing a matrix as above with $a \neq 0$. It is then immediately seen that there is only one other coset represented by the element

$$w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

giving rise to the unipotent subgroup

$$N_\infty = \left\{ \begin{pmatrix} 1 & 0 \\ u & 1 \end{pmatrix} \right\} = \overline{N_0}\,.$$

Thus $\{x, \infty\}$ parametrizes the one-parameter unipotent subgroups of $SL_2(\mathbf{R})$, bijectively.

Theorem 2. *A subgroup N_x is cuspidal for $SL_2(\mathbf{Z})$ if and only if x is rational, or ∞.*

Proof. This is immediately seen from the explicit description of matrices in N_x given in (1) above. If the intersection of N_x and $SL_2(\mathbf{Z})$ contains an element other than 1, then $u = n$ is an integer and $1 - ux$ is also an integer, so x is rational. The converse is also clear.

Let Γ be a subgroup of $SL_2(\mathbf{Z})$, of finite index. Then N is cuspidal for Γ if and only if it is cuspidal for $SL_2(\mathbf{Z})$.

Proof. If the intersection of N with $SL_2(\mathbf{Z})$ contains an element other than 1, then a power of this element lies in Γ. Furthermore N does not contain any elements of finite order, so N is cuspidal for Γ. The converse is clear.

Two unipotent subgroups N, N' are called Γ-conjugate if there exists an element $\gamma \in \Gamma$ such that $\gamma N \gamma^{-1} = N'$. Consider the set of all unipotent subgroups which are cuspidal for Γ. Then Γ operates on this set by conjugation, and a Γ-conjugacy class (orbit under Γ) is called a **cusp** of Γ.

Example. For $\Gamma \subset SL_2(\mathbf{Z})$, the cusps are in bijection with the double cosets

$$\Gamma\backslash SL_2(\mathbf{Q})/B(\mathbf{Q}),$$

where $B(\mathbf{Q})$ is the group of rational triangular matrices in $SL_2(\mathbf{Q})$. This is immediate from the parametrization by means of x, ∞ with x rational. When $\Gamma = SL_2(\mathbf{Z})$, then there is only one cusp. Otherwise, the number of cusps is bounded by the index

$$(SL_2(\mathbf{Z}) : \Gamma).$$

Theorem 3. *Let Γ be a discrete subgroup of $SL_2(\mathbf{R})$ such that the quotient $\Gamma\backslash SL_2(\mathbf{R})$ is compact. Then there is no cuspidal group for Γ.*

Proof. We have to show that Γ contains no unipotent matrix other than 1. We first prove that the conjugacy class of any element γ under $SL_2(\mathbf{R})$ is closed in $SL_2(\mathbf{R})$. Indeed, let W be a compact set such that

$$SL_2(\mathbf{R}) = W\Gamma.$$

Consider the map of $SL_2(\mathbf{R})$ into itself given by

$$g \longmapsto g\gamma g^{-1}.$$

Its image consists of the conjugation by elements of W of the set

$$S = \{\sigma\gamma\sigma^{-1}, \sigma \in \Gamma\}.$$

This set S is contained in Γ, and is closed, discrete. If there is a sequence of elements $g_n \in W$ and $\sigma_n \in \Gamma$ such that

$$g_n\sigma_n\gamma\sigma_n^{-1}g_n \quad \text{converges to} \quad h$$

for some element $h \in SL_2(\mathbf{R})$, then after selecting a subsequence if necessary, we may assume that g_n converges to an element $g \in W$. Then

$$\sigma_n\gamma\sigma_n^{-1} \longrightarrow g^{-1}hg,$$

and hence $\sigma_n \gamma \sigma_n^{-1}$ is constant for all n sufficiently large (but we are not saying that σ_n itself is constant!), because S is discrete. Therefore

$$g_n \sigma \gamma \sigma^{-1} g_n^{-1} \longrightarrow g \sigma \gamma \sigma^{-1} g^{-1}.$$

This proves that the conjugacy class is closed. The proof is concluded by observing that any unipotent element $\neq 1$ in $SL_2(\mathbf{R})$ has a non-closed conjugacy class. Indeed, the element is conjugate to

$$\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}, \qquad u \neq 0.$$

Conjugating this by a diagonal matrix, we see that

$$\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a^{-1} & 0 \\ 0 & a \end{pmatrix} \rightarrow \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

as $a \longrightarrow 0$, and the conjugacy class of our unipotent element is not closed. This proves Theorem 3.

The word cusp comes from the geometric situation arising on the upper half plane. Let $G = SL_2(\mathbf{R})$ and $\mathfrak{H} = G/K$ as usual. Let Γ be a discrete subgroup of G. Then $\Gamma\backslash\mathfrak{H}$ is an interesting object. Let us call a **fundamental domain** F for Γ in H a subset of $\mathfrak{H}$ which contains a representative for each orbit of Γ in $\mathfrak{H}$ (operating by multiplication of the left), and such that if two points z, z' lie in the same orbit, then they lie on the boundary of F. If $\Gamma = SL_2(\mathbf{Z})$, then it is an easy matter to prove that a fundamental domain is given by the illustration in Fig. 1. Cf., for instance, Serre's *Cours d'Arithmétique*, or practically any book on elliptic functions, e.g. mine.

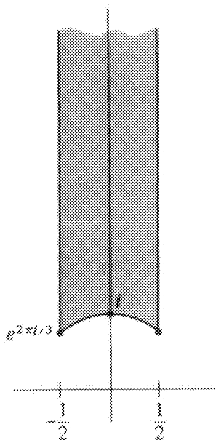

Figure 1

The single cusp with respect to Γ is determined by the standard unipotent group N_0. We could consider the space consisting of $\Gamma\backslash\mathfrak{H}$ and the cusp, topologizing this space by defining a system of neighborhoods of the cusp to consist of all elements of $\Gamma\backslash\mathfrak{H}$ having a representative in the set F_a consisting of all $x + iy$ such that

$$-\tfrac{1}{2} \leqslant x \leqslant \tfrac{1}{2} \quad \text{and} \quad y \geqslant a.$$

(Cf. Shimura [Sh], Chapter 1.) The element

$$w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

transforms the fundamental domain into the figure shown in Fig. 2(a), i.e. it reflects the fundamental domain across the arc of circle. This shape now looks like a cusp (the sides coming down to the real axis have a common vertical tangent at 0). This is the origin of the name *cusp*.

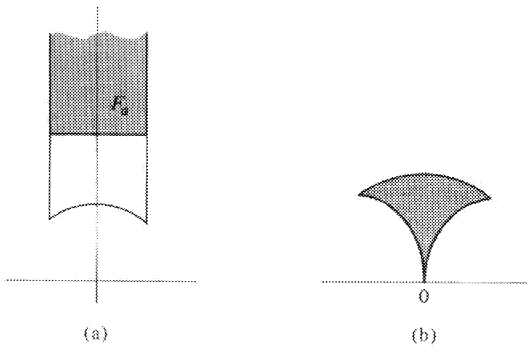

Figure 2

In general, Siegel [Si] has shown that if Γ is a discrete subgroup of $SL_2(\mathbf{R})$ such that $\Gamma\backslash\mathfrak{H}$ has finite volume, then we can always find a fundamental domain F of the following form. It consists of a finite number of pieces

$$F_0 \cup \bigcup_{\alpha=1}^{n} F_\alpha$$

where:

i) F_0 is compact with piecewise analytic boundary.

ii) There exists $x_0 > 0$ such that F is contained in the strip

$$-x_0 \leqslant x \leqslant x_0.$$

iii) There is a number a (large) and elements g_α $(\alpha = 1, \ldots, n)$ in $SL_2(\mathbf{R})$ such that $F_\alpha = g_\alpha F_a$.

Cf. Gelfand-Graev-Pjateckii-Shapiro [Ge, Gr], Chapter 1. In this book, we wish to emphasize representation theoretic topics and analysis rather than Riemann surfaces, and so we do not include the proof of this result. For other facts concerning fundamental domains and Riemann surfaces, cf. Petersson [Pe 1]. We shall return to these ideas when we discuss the Faddeev paper in Chapter XIV.

We shall now discuss growth conditions and how they correspond under the lifting map $\varphi \longmapsto \Phi$.

We consider a special case. Let Γ_0 be the group of matrices

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}$$

with $n \in \mathbf{Z}$. It is the intersection of $SL_2(\mathbf{Z})$ with our standard unipotent group N_0. Suppose that a function f on the upper half plane is invariant under Γ_0. This means that

$$f(z+1) = f(z), \qquad z \in \mathfrak{H}.$$

Suppose also that f is holomorphic on $\mathfrak{H}$. The map

$$z \longmapsto e^{2\pi i z} = q_z$$

is a complex analytic map which factors through Γ_0. For a fixed value of y, it maps the segment

$$-\tfrac{1}{2} \leqslant x \leqslant \tfrac{1}{2}, \qquad y = y_0$$

on the circle

$$e^{-2\pi y_0} e^{2\pi i x}.$$

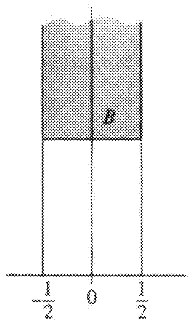

Figure 3

As $y \to \infty$ the radius of the circle shrinks. Thus $z \mapsto q_z$ for $\mathrm{Im}(z) \geqslant B$ is a complex analytic isomorphism of $\Gamma_0\backslash\mathfrak{H}$ onto the punctured disc

$$0 < |q| \leqslant e^{-2\pi B}.$$

Let f be a holomorphic function on $\mathfrak{H}$, invariant under Γ_0. Then under the analytic isomorphism $z \mapsto q_z$ the function becomes a function $f^*(q)$, defined in a neighborhood of 0 in the q-plane, excluding 0. If f^* has a pole at 0, we say that f is **meromorphic at infinity**. If f^* is holomorphic at 0, we say that f is holomorphic at infinity. In the case of a pole, we have a power series expansion

$$f(z) = f^*(q) = \sum_{-r}^{\infty} a_n q^n.$$

When f^* is holomorphic at 0, this power series starts with a constant term or higher. If there is a pole, then the first term

$$a_r q^{-r} = a_r e^{2\pi r y} e^{2\pi i x}$$

with $a_r \neq 0$ grows exponentially like $e^{2\pi r y}$ as $y \to \infty$. The condition that there be no pole is therefore that f is bounded at infinity, i.e. as $y \to \infty$.

Consider now one of our lifted functions Φ defined by

$$\Phi\left(u\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix} r(\theta)\right) = \varphi(x + iy)y^m e^{2im\theta}.$$

The above discussion shows that if φ is holomorphic on $\mathfrak{H}$ and invariant under Γ_0, then

φ is holomorphic at infinity if and only if

$$|\Phi(g)| \ll y_g^m \qquad \textit{for} \quad y_g \to \infty.$$

In other words, $\Phi(g)$ has at most **polynomial growth at infinity**.

A similar discussion applies to a subgroup of Γ_0 of finite index, necessarily consisting of matrices of the form

$$\begin{pmatrix} 1 & dn \\ 0 & 1 \end{pmatrix}, \qquad n \in \mathbf{Z},$$

where d is some fixed positive number. Instead of a strip of width 1 we have to use a strip of width d. Otherwise, the discussion is entirely similar.

A similar discussion also applies in the neighborhood of an arbitrary cusp, by taking an inner automorphism with an element of $SL_2(\mathbf{R})$.

§2. CUSP FORMS

Let Γ be a discrete subgroup of $G = SL_2(\mathbf{R})$ and assume for simplicity that Γ has only a finite number of cusps, i.e. that there is only a finite number of Γ-conjugacy classes of unipotent subgroups N of G whose intersection with Γ is not trivial. We let m be a positive integer.

We denote by $\mathcal{Q}(\Gamma \backslash G, m)$ the space of C^∞ functions on G satisfying the conditions for all $g \in G$ and $\gamma \in \Gamma$:

AUT 1. $\quad f(gr(\theta)) = e^{im\theta}f(g).$

AUT 2. $\quad f(\gamma g) = f(g).$

AUT 3. $\quad \mathfrak{L}_{E^-}f = 0.$

AUT 4. *At every cusp, the function has at most polynomial growth.*

The third condition is the condition corresponding to analyticity on the upper half plane, and shows that f is at any rate real analytic on G, cf. Chapter IX, §4. The fourth condition says that our function is analytic at infinity. Conditions **AUT 1** and **AUT 2** can be interpreted in terms of the lifting procedure of Chapter IX, §4 as saying that if $f = \Phi$ is lifted from the function φ, then

$$\varphi(\gamma z) = \varphi(z)(cz + d)^m$$

for all $\gamma \in \Gamma$, with

$$\gamma^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Often in the classical literature, the action of γ is written on the **right**, to avoid having to take an inverse. Thus if

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then we write

$$\varphi \mid [\gamma]_m(z) = \varphi(\gamma z)(cz + d)^{-m}.$$

We then have the right-hand rule

$$[\gamma\gamma']_m = [\gamma]_m[\gamma']_m.$$

The subspace of $\mathscr{Q}(\Gamma\backslash G, m)$ consisting of those f satisfying the additional condition

$$\int_{\Gamma_N\backslash N} f(ng)\, dn = 0$$

for every cuspidal subgroup N of G is called the space of **analytic cusp forms**, and is denoted by $\mathscr{Q}^0(\Gamma\backslash G, m)$. We have abbreviated $\Gamma \cap N$ by Γ_N.

We also consider $L^2(\Gamma\backslash G)$, and let ${}^0L^2(\Gamma\backslash G)$ be the closure in $L^2(\Gamma\backslash G)$ of the space spanned by all bounded continuous functions f such that for all $g \in G$, we have

CUSP 1. $$\int_{\Gamma_N\backslash N} f(ng)\, dn = 0$$

for all cuspidal N. Let

$$\pi : G \longrightarrow \text{Aut } L^2(\Gamma\backslash G)$$

be the unitary representation by right translation, i.e.

$$\pi(s)f(g) = f(gs).$$

We observe that ${}^0L^2(\Gamma\backslash G)$ is stable under π, because

$$\int_{\Gamma_N\backslash N} (\pi(s)f)(ng)\, dn = \int_{\Gamma_N\backslash N} f(ngs)\, dn = 0.$$

Also note that if the cuspidal condition (1) is satisfied for one cuspidal subgroup N, then it is satisfied for $\gamma N \gamma^{-1}$ if $\gamma \in \Gamma$. Indeed,

$$\gamma(\Gamma \cap N)\gamma^{-1} = \Gamma \cap (\gamma N \gamma^{-1}),$$

so if we let

$$M = \gamma N \gamma^{-1},$$

we obtain

$$\begin{aligned}\int_{\Gamma_M \backslash M} f(mg)\, dm &= \int_{\Gamma_N \backslash N} f(\gamma n \gamma^{-1} g)\, dn \\ &= \int_{\Gamma_N \backslash N} f(n\gamma^{-1} g)\, dn \\ &= 0\end{aligned}$$

by assumption.

We now assume for the rest of the section that

$$\Gamma = SL_2(\mathbf{Z}),$$

in order to deal with the q-expansion without a more elaborate language (throwing back the situation at an arbitrary cusp to the standard one by an inner automorphism). Thus Γ_0 is the group of matrices

$$\begin{pmatrix} 1 & n \\ 0 & 1 \end{pmatrix}, \qquad n \in \mathbf{Z}.$$

We let $\Phi \in \mathfrak{A}(\Gamma \backslash G, m)$ correspond to the function φ on $\mathfrak{H}$. In view of **AUT 3**, we know that φ is holomorphic on $\mathfrak{H}$, and the growth condition **AUT 4** tells us that φ is holomorphic at infinity, i.e. has a q-expansion

$$\varphi^*(q) = \sum_{n=0}^{\infty} a_n q^n$$

where $q = e^{2\pi i z} = e^{-2\pi y} e^{2\pi i x}$. This is essentially a Fourier series. We have

$$e^{-2\pi n y} a_n = \int_0^1 \varphi(x + iy) e^{-2\pi i n x}\, dx$$

and therefore

$$a_0 = \int_0^1 \varphi(x + iy)\, dx.$$

Using the definition of the m-th lifting

$$\Phi\left(u\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix} r(\theta)\right) = \varphi(x + iy),$$

we find

$$a_0 = \int_0^1 e^{-im\theta}y^{-m/2}\Phi\left(u\begin{pmatrix} y & x \\ 0 & 1\end{pmatrix}r(\theta)\right)dx$$

$$= e^{im\theta}y^{-m/2}\int_0^1 \Phi\left(\begin{pmatrix} 1 & x \\ 0 & 1\end{pmatrix}\begin{pmatrix} y & 0 \\ 0 & 1\end{pmatrix}ur(\theta)\right)dx$$

$$= C(g)\int_0^1 \Phi\left(\begin{pmatrix} 1 & x \\ 0 & 1\end{pmatrix}g\right)dx$$

where $C(g)$ is a number depending on g. Therefore we find a condition equivalent to the cusp condition (1) in terms of a_0, namely:

CUSP 2. *Let $\Phi \in \mathcal{Q}(\Gamma\backslash G, m)$, and let a_0 be the 0-th coefficient in the q-expansion of the associated function φ on $\mathfrak{H}$. We have $a_0 = 0$ if and only if*

$$\int_0^1 \Phi\left(\begin{pmatrix} 1 & x \\ 0 & 1\end{pmatrix}g\right)dx = 0$$

for all $g \in G$.

Of course, the integral in **CUSP 2** can be rewritten as

$$\int_{\Gamma_N\backslash N}\Phi(ng)\,dn = 0.$$

Theorem 4. *We have $\mathcal{Q}^0(\Gamma\backslash G, m) \subset {}^0L^2(\Gamma\backslash G)$.*

Proof. Let φ be the function on $\mathfrak{H}$ corresponding to a lifted function Φ in $\mathcal{Q}^0(\Gamma\backslash G, m)$. Since $a_0 = 0$, we see that the power series in q starts with a term

$$a_1 e^{-2\pi y}e^{2\pi i x}$$

and hence that

$$|\varphi(x+iy)| \ll e^{-2\pi y},$$

that is, φ decreases exponentially for $y \to \infty$. The order of growth of Φ is at most like

$$y^{m/2}|\varphi(x + iy)|,$$

and consequently Φ is bounded. Since $\Gamma \backslash G$ has finite measure (modulo K the measure is that of a fundamental domain with respect to

$$\frac{dx\,dy}{y^2}$$

on the upper half plane), it follows that Φ is in $\mathfrak{L}^2$, as was to be shown.

We end by giving a classical estimate of Fourier coefficients of cusp forms due to Hecke.

Theorem 5. *Let f be a holomorphic function on $\mathfrak{H}$, satisfying*

$$f|[\gamma]_m = f, \qquad \textit{all } \gamma \in \Gamma.$$

Assume that its 0*-th coefficient in the q-expansion is* 0, *i.e.*

$$f^*(q) = \sum_{n=1}^{\infty} a_n q^n.$$

Then: (i) $|f(z)| \ll y^{-m/2}$ *for* $y \to \infty$.
(ii) $|a_n| = O(n^{m/2})$.

Proof. The function $h(z) = |f(z)|y^{m/2}$ is invariant under Γ, as is seen at once from the definitions of the operation $[\gamma]_m$, $\gamma \in \Gamma$. Also, $h(z) \to 0$ as $y \to \infty$. Hence h is bounded on the fundamental domain. This proves the first assertion. Next, we have

$$a_n = \frac{1}{2\pi i} \int f^*(q) \frac{dq}{q^{n+1}}$$

where the integral is taken over a small circle around 0 in the q-plane. Note that $dq/q = 2\pi i\, dx$. Hence

$$\begin{aligned} |a_n| &\leqslant \int_0^1 |f(x + iy)| e^{2\pi n y}\, dx \\ &\ll e^{2\pi n y} y^{-m/2} \end{aligned}$$

for all values of $y > 0$, using the first part of the proof. Let $y = 1/n$. The second inequality concerning a_n drops out.

§3. A CRITERION FOR COMPACT OPERATORS

This section contains some abstract nonsense lemmas designed for application in the next section. We give a criterion for an operator to be compact.

Theorem 6. *Let X be a locally compact space with a finite positive measure μ. Let H be a closed subspace of $L^2(X, \mu) = L^2(X)$, and let T be a linear map of H into the vector space of bounded continuous functions on X. Assume that there exists $C > 0$ such that*

$$\|Tf\| \leqslant C\|f\|_2, \qquad \text{all } f \in H,$$

where $\| \ \|$ is the sup norm. Then

$$T\colon H \longrightarrow L^2(X)$$

is a compact operator, which can be represented by a kernel in $L^2(X \times X)$.

Proof. We know that an operator which can be represented by a kernel in L^2 is compact, cf. Chapter I, §3. To get hold of the kernel, we use the hypothesis that for each $x \in X$, the map

$$f \longmapsto Tf(x), \qquad f \in H,$$

is continuous linear on H. Hence there exists a function $q_x \in H$ such that for all $f \in H$, we have

$$(Tf)(x) = \langle f, q_x \rangle = \int f(y)\, \overline{q_x(y)}\, dy.$$

Since Tf is a continuous function on X, this also shows that

$$x \longmapsto q_x$$

as a map from X into $H \subset L^2(X)$ is weakly continuous, and therefore weakly measurable. Also, its image is bounded in $L^2(X)$. Our theorem will therefore result from the next lemma.

Lemma 1. *Let X, Y be finite measured spaces such that the σ-algebra of measurable sets in Y is generated by a countable subalgebra. Given a weakly measurable map*

$$q\colon X \longrightarrow L^2(Y)$$

whose image is L^2-bounded, there exists $Q \in \mathfrak{L}^2(X \times Y)$ such that for almost all $x \in X$ we have

$$q_x(y) = Q(x, y)$$

for all $y \notin S_x$, where S_x is a set of measure 0 in Y, depending on x.

Proof. In the next lemma, we shall prove that the map

$$x \mapsto \langle q_x, q_x \rangle = \int |q_x(y)|^2 \, dy$$

is measurable. It is bounded, whence in $\mathfrak{L}^1$. For any step function g on $X \times Y$ with respect to rectangles (i.e. products of measurable sets in X and Y respectively) we have

$$\left| \int \int g(x, y) \overline{q_x(y)} \, dy \, dx \right|^2 \leq \|g\|_2^2 \int \int |q_x(y)|^2 dy \, dx.$$

The proof is the same as the usual proof for the Schwarz inequality, expanding $0 \leq (uA + vB)^2$ with $u = A \cdot B$ and $v = -|A|^2$. Hence the map

$$g \mapsto \int \int g(x, y) \overline{q_x(y)} \, dy \, dx$$

is L^2-continuous. Hence there exists $Q \in \mathfrak{L}^2(dx \otimes dy)$ such that for all characteristic functions φ, ψ of measurable sets in X, Y respectively, we have

$$\int \varphi(x) \int \psi(y) \overline{q_x(y)} \, dy \, dx = \int \varphi(x) \int \psi(y) Q(x, y) \, dy \, dx.$$

For each ψ there exists a null set Z_ψ in X such that if $x \notin Z_\psi$, then

$$\int \psi(y) \overline{q_x(y)} \, dy = \int \psi(y) Q(x, y) \, dy.$$

Using countably many ψ's, we get this relation for all x outside a set Z of measure 0 in X. Hence for $x \notin Z$, we get

$$\overline{q_x(y)} = Q(x, y),$$

for all $y \notin S_x$, where S_x is a set of measure 0 in Y. This proves our lemma.

There remains the little point about measurability.

Lemma 2. *Let H be a Hilbert space with countable base, and let X be a measured space. If $f, g \colon X \to H$ are weakly measurable maps, then the map $x \mapsto \langle f(x), g(x) \rangle$, i.e. the map $\langle f, g \rangle$, is measurable.*

Proof. Let $\{u_i\}$ be a Hilbert basis of H, and let

$$f(x) = \sum f_i(x)u_i,$$
$$g(x) = \sum g_i(x)u_i$$

be the Fourier expansions of f and g. Then by hypothesis, each f_i, g_i is measurable, and

$$\langle f(x), g(x)\rangle = \sum f_i(x)\overline{g_i(x)}$$

is a limit of measurable functions, whence measurable, as desired.

§4. COMPLETE REDUCIBILITY OF ${}^0L^2(\Gamma\backslash G)$

We consider $\Gamma = SL_2(\mathbf{Z})$ and let π be the right translation representation of $G = SL_2(\mathbf{R})$ on ${}^0L^2(\Gamma\backslash G)$, so that

$$\pi(y)f(x) = f(xy).$$

If $\varphi \in C_c^\infty(G)$, then

$$\pi^1(\varphi)f(x) = \int_G \pi(y)f(x)\varphi(y)\,dy$$
$$= \int_G f(xy)\varphi(y)\,dy.$$

For simplicity we write $\pi(\varphi)$ instead of $\pi^1(\varphi)$. The main result of this section is

Theorem 7. *If $\varphi \in C_c^\infty(G)$, then there exists a number C_φ such that for all $f \in {}^0L^2(\Gamma\backslash G)$ we have*

$$\|\pi(\varphi)f\| \leqslant C_\varphi\|f\|_2,$$

where $\|\ \|$ is the sup norm.

In view of Theorem 6 in the last section, we obtain

Corollary. *The operator $\pi(\varphi)$ is compact.*

Using I, §2, Th. 1 we get:

Theorem 8. *The representation π by right translation on ${}^0L^2(\Gamma\backslash G)$ is completely reducible, and each irreducible component occurs only a finite number of times in it.*

As usual, $G = NAK$ and $\Gamma_N = \Gamma \cap N$. We proceed with the proof of Theorem 7. We have

$$\begin{aligned}\pi(\varphi)f(x) &= \int_G f(y)\varphi(x^{-1}y)\,dy \\ &= \int_{\Gamma_N\backslash G} \sum_{\eta\in\Gamma_N} \varphi(x^{-1}\eta\dot{y})f(\eta\dot{y})\,d\dot{y} \\ &= \int_{\Gamma_N\backslash G} J_\varphi(x,y)f(y)\,dy\end{aligned}$$

where

$$J_\varphi(x,y) = \sum_{m\in\mathbf{Z}} \varphi\left(x^{-1}\begin{pmatrix}1 & m\\ 0 & 1\end{pmatrix}y\right).$$

Letting $\varphi_{x,y}(m)$ abbreviate each term in this sum, we find

$$J_\varphi(x,y) = \sum_{m\in\mathbf{Z}} \varphi_{x,y}(m) = \sum_{m\in\mathbf{Z}} \hat{\varphi}_{x,y}(m)$$

by the Poisson summation formula. Thus the effect of $\pi(\varphi)$ is given by the kernel J_φ, which we shall estimate.

We write the NAK decomposition in the form

$$x = \begin{pmatrix}t & 0\\ 0 & t\end{pmatrix}\begin{pmatrix}v_x & u_x\\ 0 & 1\end{pmatrix}r(\theta)$$

with t, $v_x > 0$, so that in the upper half plane G/K, we have

$$x\cdot i = u_x + iv_x.$$

We also write $x = n_x a_x k_x$.

We fix a number $c > 0$ and let the **Siegel set** F_c consist of all $x \in G$ such that u_x lies in a compact set Ω_N in N, and $v_x \geqslant c$. Choosing c sufficiently small, we have

$$G = \Gamma F_c.$$

The picture of F_c in G/K is shown in Fig. 4(a). To estimate $\pi(\varphi)f(x)$, we may assume without loss of generality that $x \in F_c$, because f is invariant on the left by Γ. Note that the integral

$$\int_{\Gamma_N\backslash G}$$

can be viewed as an integral over the "fundamental domain" for Γ_N which is the inverse image in G of the region drawn in Fig. 4(b). We have to estimate $J_\varphi(x,y)$ for $x \in F_c$.

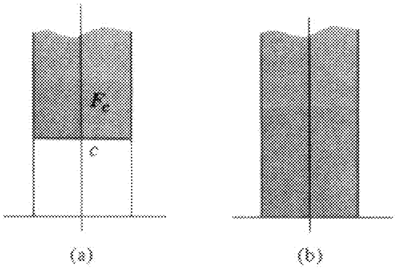

Figure 4

As a matter of convenient notation, we denote by Ω a compact set, and index Ω to denote in what space, e.g. Ω_G is a compact set in G, etc., fixed in advance. All constants will depend on such compact sets.

We first observe that

$$\text{(1)} \qquad \text{if } x \in F_c, \qquad \text{then} \qquad x \in \Omega_N a_x K \subset a_x \Omega_G.$$

Indeed, matrix multiplication gives the commutation

$$\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}\begin{pmatrix} v_x & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} v_x & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1 & u' \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} v_x & v_x u' \\ 0 & 1 \end{pmatrix},$$

so that $v_x u' = u$ and $u' = v_x^{-1}u$. Since $v_x \geqslant c$, it follows that $\Omega_N a_x \subset a_x \Omega'_N$ and our assertion is obvious.

Let $\omega_x = x^{-1}a_x$. Then we also obtain

$$\text{(2)} \qquad \omega_x = x^{-1}a_x \in \Omega_G.$$

Furthermore, letting $\omega_{x,y} = a_x^{-1}y$, we contend that if $\varphi(x^{-1}\eta y) \neq 0$ for some $\eta \in \Gamma_N$, then we may assume that

$$\text{(3)} \qquad \omega_{x,y} = a_x^{-1}y \in \Omega_G.$$

To see this, write

$$x^{-1}\eta y = x^{-1}a_x \cdot a_x^{-1}\eta n_y a_x \cdot a_x^{-1}a_y k_y.$$

The left-hand side lies in a compact set, and we have seen that $x^{-1}a_x$ lies in a compact set. The two elements

$$a_x^{-1}\eta n_y a_x \qquad \text{and} \qquad a_x^{-1}a_y k_y$$

lie in N and AK respectively, and have to lie in compact sets. Hence $a_y \in a_x\Omega_A$, and $y \in Na_x\Omega_A K$. Since y can be changed on the left by an element of Γ_N, our assertion follows at once.

Lemma 1. *For each positive integer d there exists a constant $C(\varphi, d, F_c)$ such that*

$$|\hat{\varphi}_{x,y}(\lambda)| \leqslant C(\varphi, d, F_c)\, v_x^{1-d} |\lambda|^{-d}.$$

Proof. By definition,

$$\begin{aligned}
\hat{\varphi}_{x,y}(\lambda) &= \int_{\mathbf{R}} \varphi\left(x^{-1}\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} y\right) e^{-2\pi i\lambda u}\, du \\
&= \int_{\mathbf{R}} \varphi\left(x^{-1}a_x \cdot a_x^{-1}\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} a_x \cdot a_x^{-1} y\right) e^{-2\pi i\lambda u}\, du \\
&= \int_{\mathbf{R}} \varphi\left(\omega_x \cdot a_x^{-1}\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix} a_x \cdot \omega_{x,y}\right) e^{-2\pi i\lambda u}\, du.
\end{aligned}$$

Carrying out the matrix multiplication, we change variables, let

$$w = v_x^{-1}u, \qquad dw = v_x^{-1}\, du.$$

We obtain

$$\begin{aligned}
\hat{\varphi}_{x,y}(\lambda) &= v_x \int_{\mathbf{R}} \varphi\left(\omega_x\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}\omega_{x,y}\right) e^{-2\pi i\lambda v_x u}\, du \\
&= v_x \hat{\varphi}_{\omega,x,y}(v_x\lambda),
\end{aligned}$$

where

$$\varphi_{\omega,x,y}(u) = \varphi\left(\omega_x\begin{pmatrix} 1 & u \\ 0 & 1 \end{pmatrix}\omega_{x,y}\right), \qquad \omega_x, \omega_{x,y} \in \Omega_G.$$

This last function is a C^∞ function of u, depending on parameters in a compact set. Integrating by parts d times yields the desired estimate

$$|\hat{\varphi}_{\omega,x,y}(\lambda)| \leqslant C(\varphi, d, F_c)|\lambda|^{-d},$$

whence the lemma follows by replacing λ with $v_x\lambda$.

We apply this to estimate the sum for $J_\varphi(x, y)$, using the cusp condition

$$\int_{\Gamma_N\backslash N} f(ny)\, dn = 0 \quad \text{for almost all} \quad y \in G.$$

We had seen that

$$\pi(\varphi)f(x) = \int_{\Gamma_N\backslash G} \sum_{m\in\mathbf{Z}} \hat{\varphi}_{x,y}(m)f(y)\,dy.$$

The term with $m = 0$ is equal to

$$\int_{\Gamma_N\backslash G}\int_N \varphi(x^{-1}ny)f(y)\,dn\,dy = \int_{N\backslash G}\int_{\Gamma_N\backslash N}\int_N \varphi(x^{-1}nn'y)f(n'y)\,dn\,dn'\,dy.$$

Invariance of integration under translation shows that the term $\varphi(x^{-1}nn'y)$ can be replaced by $\varphi(x^{-1}ny)$, and our last expression is equal to

$$\int_{N\backslash G}\left[\int_N \varphi(x^{-1}ny)\,dn\cdot\int_{\Gamma_N\backslash N} f(n'y)\,dn'\right]dy = 0$$

by the cusp condition.

For $m \neq 0$ we use the estimate of Lemma 1, to find

$$\begin{aligned}|\pi(\varphi)f(x)| &\ll \sum_{m\neq 0} \int_{\substack{\Gamma_N\backslash G\\ y\in a_x\Omega_G}} v_x^{1-d}\frac{1}{|m|^d}|f(y)|\,dy\\ &\ll v_x^{1-d}\int_{\substack{\Gamma_N\backslash G\\ y\in a_x\Omega_G}} |f(y)|\,dy\\ &\ll v_x^{1-d}\mu(a_x\Omega_G)^{1/2}\|f\|_2\end{aligned}$$

by Schwarz, thereby proving our theorem.

XIII The Continuous Part of $L^2(\Gamma \backslash G)$

We now look at the orthogonal complement of ${}^0L^2(\Gamma \backslash G)$ and prove a spectral decomposition theorem following Godement's paper [Go 2], using the Poisson summation formula. The method works for arithmetic subgroups Γ, and has the advantage of being rapid and easy. It fails for more general discrete subgroups, and the question is reconsidered by other methods in the next and last chapter. The spectral decomposition is achieved by the Eisenstein transform, which maps the orthogonal complement of ${}^0L^2(\Gamma \backslash G)$ and the constant functions on the L^2 space of a positive real line—with our normalization, the upper half of the imaginary line $\operatorname{Re} s = \frac{1}{2}$.

For simplicity, throughout this chapter, we let

$$\Gamma = SL_2(\mathbf{Z}).$$

In the Iwasawa decomposition, the group $\Gamma_N = \Gamma \cap N$ is then the group of matrices

$$\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix},$$

with $m \in \mathbf{Z}$. The factor space $\Gamma \backslash G$ has finite measure. It is useful to keep in mind that except for the K-component on the right, $\Gamma \backslash G$ is like $\Gamma \backslash \mathfrak{H}$, where $\mathfrak{H}$ is the upper half plane, $\mathfrak{H} = G/K$.

§1. AN ORTHOGONALITY RELATION

The theta transform

Let $G = NAK$ be the Iwasawa decomposition of $G = SL_2(\mathbf{R})$. In Chapter XI, §2, we have already studied the coset space $N \backslash G$. We let G operate on $\mathbf{R}^2$

on the right. The isotropy group of the unit vector $e_2 = (0, 1)$ is precisely N, and thus $N\backslash G$ is in bijection with $\mathbf{R}^2 - \{0\}$, i.e. $\mathbf{R}^2$ from which the origin has been deleted. We shall denote by $S(N\backslash G)$ (Schwartz space of $N\backslash G$) the space of functions on $N\backslash G$ which are restrictions of functions in the Schwartz space of $\mathbf{R}^2$. We can therefore identify $S(N\backslash G)$ with $S(\mathbf{R}^2)$. (For the Schwartz space, cf. *Real Analysis*.)

We let the **theta transform** (theta series)

$$T\colon \text{functions on } N\backslash G \longrightarrow \text{functions on } \Gamma\backslash G$$

be defined by

$$T\varphi(x) = \sum_{\Gamma_N\backslash\Gamma} \varphi(\gamma x).$$

The transform is defined only on the space of functions for which this series converges, absolutely. We look into this convergence. If

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix},$$

then

$$e_2\gamma = (0, 1)\gamma = (c, d).$$

If $\gamma \in \Gamma$, then $ad - bc = 1$ and therefore (c, d) is a primitive vector of integers, that is c, d are relatively prime. Conversely, any relatively prime pair of integers (c, d) can be completed to a 2×2 matrix in $SL_2(\mathbf{Z})$. We see that the coset space $\Gamma_N\backslash\Gamma$ is in bijection with primitive pairs

$$(c, d) = (m_1, m_2) = m \in \mathbf{Z}^2.$$

If φ is a function on $N\backslash G$, we denote by $\varphi[\ ]$ the corresponding function on $\mathbf{R}^2$. Then

$$T\varphi(y) = \sum_{\Gamma_N\backslash\Gamma} \varphi(\gamma y) = \sum_{m \text{ prim}} \varphi[my]$$

$$= \sum_{(c, d)=1} \varphi[(c, d)y],$$

For φ in the Schwartz space, and therefore rapidly decreasing at infinity, we see that the series converges.

Let φ have compact support on $N\backslash G$. Then for each y, the sum

$$T\varphi(y) = \sum_{\Gamma_N\backslash\Gamma} \varphi(\gamma y)$$

has only a finite number of non-zero terms.

Proof. Let Ω be the compact support of φ. If $\varphi(\gamma y) \neq 0$, then $\gamma y \in \Omega$. There exists $\gamma' \in \Gamma_N$ such that $\gamma'\gamma y \in \Omega_N \Omega$, where Ω_N is compact in N. Hence $\gamma'\gamma \in \Omega_N \Omega y^{-1}$, which proves our assertion.

It is clear that if $\varphi \in C_c(N \backslash G)$, then the function $T\varphi$ is continuous, on $\Gamma \backslash G$.

If $\varphi \in C_c(N \backslash G)$, then $T\varphi$ has compact support.

Proof. Let $\Omega \subset G$ be compact such that $N \backslash N\Omega$ contains the support of φ. If $\varphi(\gamma y) \neq 0$, then $\gamma y \in N\Omega$. There exists $\gamma' \in \Gamma_N$ such that $\gamma'\gamma y \in \Omega_N \Omega$, where Ω_N is a compact set in N. Hence $y \in \Gamma\Omega_G$, where Ω_G is compact in G, as was to be shown.

From the above, we may view the theta transform as a mapping

$$T\colon C_c(N \backslash G) \to C_c(\Gamma \backslash G).$$

Since Γ contains -1, we note that the sum over primitive elements in $\mathbf{Z}^2$ is symmetric with respect to the origin. Expressing a function φ as a sum of an even and an odd function on $\mathbf{R}^2 - \{0\}$, we see that if φ is odd, then $T\varphi = 0$. Consequently, without loss of generality, in what follows when considering the theta transform, we may assume that φ is even, namely

$$\varphi(x) = \varphi(-x).$$

Adjoint of the theta transform

We shall also consider the mapping

$$T^0\colon \text{functions on } \Gamma \backslash G \to \text{functions on } N \backslash G$$

given by

$$T^0 f(y) = \int_{\Gamma_N \backslash N} f(ny)\, dn.$$

Thus T^0 is defined by the integral with which we defined cusp forms. In any given situation, we shall always specify the function space to which we apply T^0. The space $L^2(\Gamma \backslash G)$ will be a frequent candidate, and ${}^0L^2(\Gamma \backslash G)$ consists of those functions f such that $T^0 f = 0$.

The mappings T and T^0 are adjoint to each other. In other words, under conditions of absolute convergence,

$$\langle T\varphi, f\rangle_{\Gamma \backslash G} = \langle \varphi, T^0 f\rangle_{N \backslash G}, \tag{1}$$

where the scalar product is that given by the usual hermitian integral.

Remark. The above adjointness relation relates any pair of closed subgroups Γ, N of a group G. We do not need that Γ is discrete, for instance. It is essentially an abstract nonsense relation on a locally compact group with two closed subgroups. We carry out first the formal computation, winding and unwinding an integral. We use the Haar measure of a subgroup and factor group, with respect to the sequence of groups:

$$\begin{tikzcd}[column sep=small, row sep=small] & & N \arrow[dr] & \\ \Gamma\cap N = \Gamma_N \arrow[urr, phantom] & \Gamma_N \arrow[ur] \arrow[dr] & & G \\ & & \Gamma \arrow[ur] & \end{tikzcd}$$

We have:

$$\begin{aligned}
\langle T\varphi, f\rangle_{\Gamma\backslash G} &= \int_{\Gamma\backslash G} T\varphi(y)\overline{f(y)}\, dy \\
&= \int_{\Gamma\backslash G}\int_{\Gamma_N\backslash\Gamma} \varphi(\gamma y)\overline{f(y)}\, d\gamma\, dy \\
&= \int_{\Gamma_N\backslash G} \varphi(y)\overline{f(y)}\, dy \\
&= \int_{N\backslash G}\int_{\Gamma_N\backslash N} \varphi(ny)\overline{f(ny)}\, dn\, dy \\
&= \int_{N\backslash G} \varphi(y)\int_{\Gamma_N\backslash N} \overline{f(ny)}\, dn\, dy \\
&= \langle \varphi, T^0 f\rangle_{N\backslash G}.
\end{aligned}$$

Theorem 1.

i) *The above adjointness relation holds if $f \in L^2(\Gamma\backslash G)$ and $\varphi \in C_c(N\backslash G)$.*

ii) *The function $T\varphi$ is orthogonal to ${}^0L^2(\Gamma\backslash G)$.*

iii) *If $f \in L^2(\Gamma\backslash G)$ and f is orthogonal to $T\varphi$ for all $\varphi \in C_c(N\backslash G)$, then $f \in {}^0L^2(\Gamma\backslash G)$.*

iv) *Let $\hat{\varphi}$ be the Fourier transform of φ, viewed as a function on $\mathbf{R}^2$. We have $\hat{\varphi}[0] = 0$ if and only if $T\varphi$ is orthogonal to the constant 1 on $\Gamma\backslash G$.*

Proof. All of these statements are immediate from the adjointness relation. The Fourier transform of a function on $\mathbf{R}^2$ is normalized to be, in the present case,

$$\hat{\varphi}[z] = \int_{\mathbf{R}^2} \varphi[\xi] e^{-2\pi i\xi\cdot z}\, d\xi.$$

The invariant measure on $N\backslash G$ under right translation by G is the same as Lebesgue measure on $\mathbf{R}^2$, and therefore

$$\hat{\varphi}[0] = \int_{N\backslash G} \varphi(y)\,dy.$$

The condition $\hat{\varphi}[0] = 0$ is equivalent with

$$0 = \langle \varphi, 1\rangle_{N\backslash G} = \langle \varphi, T^0 1\rangle_{N\backslash G} = \langle T\varphi, 1\rangle_{\Gamma\backslash G}$$

i.e. $T\varphi$ is orthogonal to 1 on $\Gamma\backslash G$.

§2. THE EISENSTEIN SERIES

For any function φ on $N\backslash G$, and any s for which the following integral converges, we define the **zeta transform**

$$Z(\varphi, y, 2s) = \int_A \varphi(ay)\rho(a)^{-2s}\,da = \int_0^\infty \varphi(h_a y)a^{-2s}\,\frac{da}{a}.$$

We use previous notation, i.e.

$$\rho\left(\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}\right) = a \quad \text{for} \quad a \in \mathbf{R}^+$$

and

$$h_a = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}.$$

If $\varphi \in S(N\backslash G)$ and $\sigma = \operatorname{Re} s > 0$, then the integral defining the zeta transform $Z(\varphi, y, 2s)$ converges absolutely.

Proof. Replacing φ by its right translate by y, and viewing φ as a function on $\mathbf{R}^2$, our integral becomes

$$\int_0^\infty f[0, a^{-1}]a^{-2s}\,\frac{da}{a}.$$

Let $a \mapsto a^{-1}$ and use the invariance of Haar measure under this transformation. The convergence is immediate.

Before going on to other convergence properties, we point out that the above zeta transform satisfies the conditions of an induced function, as in the theory of induced representations, namely

$$Z(\varphi, ny, 2s) = Z(\varphi, y, 2s), \qquad n \in N,$$

$$Z(\varphi, ay, 2s) = Z(\varphi, y, 2s)\rho(a)^{2s}, \qquad a \in A.$$

Any function f on G which satisfies the conditions

$$f(ny) = f(y) \quad \text{and} \quad f(ay) = f(y)\rho(a)^{2s}$$

will be said to be of **type** $2s$. This decomposition into types will be important in the arguments of Theorem 3, §7.

We now turn to analyticity properties of zeta transforms under various conditions.

Lemma 1. *If $\varphi \in C_c(N\backslash G)$, then $Z(\varphi, y, 2s)$ is entire in s.*

This is essentially obvious since nothing horrible occurs either near 0 or near ∞. We had already encountered this situation in Chapter V, §3.

Lemma 2. *For $\varphi \in S(N\backslash G)$, the **Eisenstein** series*

$$E(\varphi, y, s) = \sum_{\Gamma_N\backslash\Gamma} Z(\varphi, \gamma y, 2s) = TZ(\varphi, y, 2s)$$

converges absolutely for Re $s > 1$.

Proof. Replacing φ by its y-translate on the right, it suffices to prove the convergence with $y = e$, i.e. deal with the sum

$$\sum_{\Gamma_N\backslash\Gamma} Z(\varphi, \gamma, 2s).$$

But if

$$\gamma = \begin{pmatrix} * & * \\ c & d \end{pmatrix},$$

then

$$Z(\varphi, \gamma, 2s) = \int_0^\infty \varphi[e_2 h_a \gamma] a^{-2s} \frac{da}{a}$$

$$= \int_0^\infty \varphi[c/a, d/a] a^{-2s} \frac{da}{a}$$

$$= \int_0^\infty \varphi[ac, ad] a^{2s} \frac{da}{a}.$$

Let $m=(c,d)$ range over primitive pairs (or even all pairs) of integers. We decompose the sum over such m by partitioning those which lie in an annulus of radius n and width 1, say with respect to the Euclidean norm. Pick C such that if $\xi\in\mathbf{R}^2$ and $|\xi| > C$, then $\varphi[\xi]\ll 1/\xi^{2\sigma+\epsilon}$. For $m\in\mathbf{Z}^2$ and m in the annulus of radius n, width 1, our integral is estimated by the sum

$$\int_0^{C/n}+\int_{C/n}^{\infty}.$$

The first integral has the bound

$$\int_0^{C/n}|\varphi[am]|\,a^{2\sigma-1}\,da \ll a^{2\sigma}\Big|_0^{C/n} \ll \frac{1}{n^{2\sigma}}\,.$$

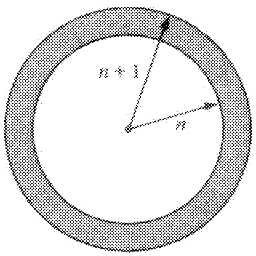

Figure 1

The second integral has the bound

$$\int_{C/n}^{\infty}|\varphi[am]|\,a^{2\sigma-1}\,da \ll \int_{C/n}^{\infty}\frac{a^{2\sigma-1}}{|am|^{2\sigma+\epsilon}}\,da \ll \frac{1}{n^{2\sigma}}\,.$$

The number of lattice points in our annulus is $\ll n$. Hence the sum over such lattice points is $\ll 1/n^{\sigma}$. Summing over n proves the desired convergence.

§3. ANALYTIC CONTINUATION AND FUNCTIONAL EQUATION

Let $\varphi\in S(N\backslash G) = S(\mathbf{R}^2)$. Viewing φ as a function on $\mathbf{R}^2$, we have its usual Fourier transform $\hat{\varphi}$, defined by

$$\hat{\varphi}(z) = \int_{\mathbf{R}^2}\varphi[\xi]e^{-2\pi i z\cdot\xi}\,d\xi,$$

where $d\xi$ is Lebesgue measure. Then $\hat{\hat{\varphi}} = \varphi^-$, where $\varphi^-[z] = \varphi[-z]$. (For the Fourier transform, and other elementary matters pertaining to it, cf. *Real Analysis.*)

For $y \in G$ we let

$$\hat{y} = {}^t y^{-1}.$$

The change of variables formula shows that the Fourier transform of the function

$$\xi \mapsto \varphi[\xi a y]$$

is the function

$$\xi \mapsto \hat{\varphi}[\xi a^{-1}\hat{y}]a^{-2}.$$

The Poisson formula then yields

$$(1) \qquad \sum_{m \in \mathbf{Z}^2} \varphi[may] = \sum_{m \in \mathbf{Z}^2} \hat{\varphi}[ma^{-1}\hat{y}]a^{-2},$$

which will be used in the proof of the next theorem.

Observe that if $\varphi \in C_c(N\backslash G)$, then $\varphi[0] = 0$. However, it does not follow that $\hat{\varphi}[0] = 0$ also. In some applications of the next theorem, we shall take $\varphi \in C_c(N\backslash G)$, but need the next theorem also to be applied to $\hat{\varphi}$.

Theorem 2. *Let $\varphi \in S(N\backslash G)$ be an even function, and assume that*

$$\varphi[0] = \hat{\varphi}[0] = 0.$$

Define

$$E^*(\varphi, y, s) = \zeta(2s)E(\varphi, y, s).$$

Then E^* is an entire function of s, and

$$E^*(\varphi, y, s) = E^*(\hat{\varphi}, \hat{y}, 1 - s).$$

Proof. By definition

$$Z(\varphi, y, 2s) = \int_0^\infty \varphi(h_a y)a^{-2s}\,\frac{da}{a}.$$

Let

$$\gamma = \begin{pmatrix} * & * \\ c & d \end{pmatrix}.$$

Then

$$Z(\varphi, \gamma y, 2s) = \int_0^\infty \varphi[(ac, ad)y]a^{2s} \frac{da}{a}$$

$$= \int_0^\infty \varphi[may]a^{2s} \frac{da}{a}$$

and

$$E(\varphi, y, s) = \sum_{m \text{ prim}} \int_0^\infty \varphi[may]a^{2s} \frac{da}{a}.$$

From

$$\zeta(2s) = \sum \frac{1}{n^{2s}},$$

multiplying $\zeta(2s)$ with $E(\varphi, y, s)$ yields

$$\zeta(2s)E(\varphi, y, s) = \sum_{n=1}^\infty \sum_{m \text{ prim}} \int_0^\infty \varphi[may](a/n)^{2s} \frac{da}{a}$$

$$= \int_0^\infty \sum_{m \in \mathbf{Z}^2} \varphi[may]a^{2s} \frac{da}{a}$$

$$= \int_1^\infty \sum_m \varphi[may]a^{2s} \frac{da}{a} + \int_0^1 \sum_m \varphi[may]a^{2s} \frac{da}{a}$$

$$= \int_1^\infty \sum_m \varphi[may]a^{2s} \frac{da}{a} + \int_0^1 \sum_m \hat{\varphi}[ma^{-1}\hat{y}]a^{2s-2} \frac{da}{a}$$

$$= \int_1^\infty \sum_m \varphi[may]a^{2s} \frac{da}{a} + \int_1^\infty \sum_m \hat{\varphi}[ma\hat{y}]a^{2-2s} \frac{da}{a},$$

where this last step is obtained by letting $a \mapsto a^{-1}$ in the second integral. Each one of these last integrals is entire in s, and the sum of the two integrals is invariant under

$$(\varphi, y, s) \mapsto (\hat{\varphi}, \hat{y}, 1 - s),$$

using the fact that $\hat{\hat{\varphi}} = \varphi^-$ and $\varphi^- = \varphi$ because φ is assumed to be an even function. This proves our theorem.

Define

$$\check{\varphi}(y) = \check{\varphi}[e_2 y] = \hat{\varphi}[e_2 yw]$$

where as usual

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

Note that $wyw^{-1} = \hat{y}$. Furthermore, $m \mapsto mw$ permutes $\mathbf{Z}^2$. Hence

$$\sum_{m \in \mathbf{Z}^2} \hat{\varphi}[ma\hat{y}] = \sum_{m \in \mathbf{Z}^2} \hat{\varphi}[mw\hat{y}w^{-1}aw] = \sum_{m \in \mathbf{Z}^2} \check{\varphi}[mya].$$

This yields

Corollary. $$\boxed{E^*(\varphi, y, s) = E^*(\check{\varphi}, y, 1 - s).}$$

The advantage of the formulation in the corollary is that y is unchanged in both sides.

§4. MELLIN AND ZETA TRANSFORMS

Let f be a function in the Schwartz space of $\mathbf{R}$. In Chapter V we had already considered the Mellin transform

$$Mf(2s) = \int_0^\infty f(a)a^{2s} \frac{da}{a}$$

for functions with compact support not containing 0, whence Mf is entire. In the present case, the above integral converges absolutely for $\operatorname{Re} s > 0$. Furthermore the integral

$$\int_1^\infty f(a)a^{2s} \frac{da}{a}$$

converges for all complex s, and defines an entire function of s. The possible poles are due to the behaviour of f near 0.

Lemma 1. *If $f \in S(\mathbf{R})$ and $f(0) = 0$ then $Mf(2s)$ is a meromorphic function of s with at most simple poles at*

$$s = -\tfrac{1}{2}, -1, -\tfrac{3}{2}, \dots .$$

Proof. Assuming $f(0) = 0$ implies that $f(a) = af_1(a)$ where f_1 is C^∞.

Integrating by parts the integral

$$\int_0^1 f(a)a^{2s}\,\frac{da}{a} = \int_0^1 f(a)a^{2s-1}\,da$$

gives

$$\left.\frac{f_1(a)a^{2s+1}}{2s+1}\right|_0^1 - \frac{1}{2s+1}\int_0^1 f_1'(a)a^{2s+1}\,da.$$

We continue to integrate by parts to get the lemma.

The above lemma applies to a function f formed with some

$$\varphi \in C_c^\infty(N\backslash G) = C_c^\infty(\mathbf{R}^2 - \{0\}),$$

and some $y \in G$, letting

$$f(a) = \varphi[(0, a)y].$$

In this case,

$$Mf(2s) = Z(\varphi, y, 2s)$$

is entire. However, when we pass to the Fourier transform, the function $\hat{\varphi}$ does not necessarily have support away from 0, while the lemma still applies to the function

$$Z(\check{\varphi}, y, 2s) = Mf(2s), \qquad \text{where} \quad f(a) = \hat{\varphi}[(0, a)yw]$$

under the assumption that $\hat{\varphi}[0] = 0$.

Consider a function g defined in a strip

$$\sigma_0 \leqslant \sigma = \operatorname{Re} s < \sigma_1.$$

We say that g is **rapidly decreasing on every vertical line, uniformly in the strip**, if for every positive integer d, there exists a constant C such that

$$|g(\sigma + i\tau)| \leqslant \frac{C}{1 + \tau^{2d}}$$

for all $\sigma + i\tau$ in the strip. (We write $2d$ to have an even exponent, killing the parity of τ.) If a finite number of poles exist in the strip. and the above estimate holds outside a neighborhood of the poles, then we extend the terminology accordingly.

Lemma 2. Let $f \in S(\mathbf{R})$ and assume $f(0) = 0$. Then $Mf(2s)$ is rapidly decreasing on every vertical line uniformly in each strip

$$\sigma_0 \leqslant \sigma \leqslant \sigma_1.$$

outside a neighborhood of the poles.

Proof. Suppose first $\operatorname{Re} s > 0$. For any integer $n > 0$ we have, after integrating by parts,

$$(2s + 1)(2s + 2) \cdots (2s + n)Mf(2s) = Mf^{(n)}(2s + n),$$

where $f^{(n)}$ is the n-th derivative of f. This expression is bounded. Dividing by the n-fold product shows that Mf is rapidly decreasing. Furthermore, the estimate is also valid for all s outside a neighborhood of the poles, by analytic continuation. The result holds for all s because for s in a strip, we can always select n sufficiently large so that $s + n$ is large positive.

We had also seen in Chapter V that a change of variables makes the Mellin transform into a Fourier transform. We have:

Lemma 3. Let $\sigma = \operatorname{Re} s > 1$. If the functions f, g on $\mathbf{R}^+$ (with respect to Haar measure da/a) are such that $f(a)a^{2s-2}$ and $g(a)a^{-2\bar{s}}$ are in $\mathfrak{L}^1 \cap \mathfrak{L}^2(\mathbf{R}^+)$, then

$$\int_0^\infty f(a)\overline{g(a)}\,a^{-2}\,\frac{da}{a} = \int_0^\infty f(a)a^{2s-2}\,\overline{g(a)a^{-2\bar{s}}}\,\frac{da}{a}$$

$$= \frac{1}{\pi i}\int_{\operatorname{Re} s = \sigma} Mf(2\bar{s} - 2)\,\overline{Mg(-2s)}\,ds.$$

(*On the vertical line, $ds = i\,d\tau$ and so $i^{-1}\,ds$ is real.*)

Proof. Put $a^2 = e^b$, so that $a = e^{b/2}$. Then

$$\frac{da}{a} = \frac{1}{2}\,db.$$

Let $s = \sigma + i\tau$. Let

$$f_1(b) = f(e^{b/2})e^{b(\sigma-1)} \qquad \text{and} \qquad g_1(b) = g(e^{b/2})e^{-b\sigma}.$$

Then $f_1, g_1 \in \mathfrak{L}^1 \cap \mathfrak{L}^2(\mathbf{R})$ and the ordinary Plancherel formula applies, namely

$$\langle f_1, g_1\rangle = \langle \hat{f}_1, \hat{g}_1\rangle,$$

with a normalization of the Fourier transform *for this proof* such that

$$\hat{f}_1(\tau) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} f_1(b)e^{-ib\tau}\, db,$$

and similarly for $\hat{g}_1$. This means in terms of the Mellin transform that

$$\hat{f}_1(\tau) = \frac{2}{\sqrt{2\pi}} Mf(2\bar{s} - 2) \qquad \text{and} \qquad \hat{g}_1(\tau) = \frac{2}{\sqrt{2\pi}} Mg(-2s).$$

Hence the integral expressing the scalar product is equal to

$$\frac{1}{2}\int_{-\infty}^{\infty} \hat{f}_1(\tau)\overline{\hat{g}_1(\tau)}\, d\tau = \frac{1}{\pi i}\int_{\mathrm{Re}\, s=\sigma} Mf(2\bar{s} - 2)\,\overline{Mg(-2s)}\, ds$$

as desired.

Warning. When applying the Mellin transform as a zeta transform, watch out for signs. If θ is a function on $N\backslash G$ and $f(a) = \theta(h_a y)$, then

$$Mf(2s) = \int_0^{\infty} \theta(h_a y)a^{2s}\, \frac{da}{a} = Z(\theta, y, -2s).$$

§5. SOME GROUP THEORETIC LEMMAS

Double cosets

As before, $\Gamma_N = \Gamma \cap N$. Write a double coset decomposition,

$$\Gamma = \bigcup_i \Gamma_N \gamma_i \Gamma_N.$$

If $\Gamma_N\gamma_i\Gamma_N \neq \pm\Gamma_N$, then the elements $\{\gamma_i\eta\}$, $\eta\in\Gamma_N$, form a system of single coset representatives for $\Gamma_N\gamma_i\Gamma_N$, that is we have the single coset decomposition

$$\Gamma_N\gamma_i\Gamma_N = \bigcup_{\eta\in\Gamma_N} \Gamma_N\gamma_i\eta.$$

Proof. If $\eta \neq \eta' \in \Gamma_N$, then

$$\Gamma_N\gamma_i\eta \neq \Gamma_N\gamma_i\eta',$$

for otherwise, there exists $\eta_1 \in \Gamma_N$ such that $\gamma_1\gamma_i\eta = \gamma_i\eta'$, so that

$$\gamma_i^{-1}\eta_1\gamma_i = \eta'\eta^{-1} \in \Gamma_N,$$

whence $\gamma_i \in \pm\Gamma_N$.

It is also immediately verified that the double cosets

$$\Gamma_N\gamma\Gamma_N \qquad \text{and} \qquad -\Gamma_N\gamma\Gamma_N$$

are distinct for any $\gamma \in \Gamma$, so the double cosets occur in pairs.

Bruhat decomposition

We recall briefly the Bruhat decomposition, which was discussed in detail in Chapter XI, §2.

The group G has a unique decomposition

$$G = \pm NA \cup \pm NAwN,$$

where

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

This is obtained by looking at the coset space $N\backslash G \subset \mathbf{R}^2$, noting that N is the isotropy group of $e_2 = (0, 1)$. We write an element $y \in NAwN$ uniquely as

$$y = n'_y h_y w n''_y,$$

so that h_y is the Bruhat representative of y in A.

Observe that if

$$y = \begin{pmatrix} 1 & u' \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1/c & 0 \\ 0 & c \end{pmatrix}\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & u'' \\ 0 & 1 \end{pmatrix},$$

then

$$y = \begin{pmatrix} * & * \\ c & * \end{pmatrix}, \qquad \text{with} \quad c = c(y).$$

Consequently

$$-h_y = \begin{pmatrix} 1/c(y) & 0 \\ 0 & c(y) \end{pmatrix}.$$

§6. AN EXPRESSION FOR $T^0T\varphi$

Let $\varphi \in C_c(N \backslash G)$ and assume that φ is an even function, that is

$$\varphi(x) = \varphi(-x).$$

We denote by (γ) non-trivial pairs of double cosets

$$(\gamma) = \pm\Gamma_N\gamma\Gamma_N \neq \pm\Gamma_N.$$

We let h_γ be the Bruhat representative of γ.

We consider the expression

$$T^0T\varphi(y) = \int_{\Gamma_N\backslash N} \sum_{\Gamma_N\backslash\Gamma} \varphi(\gamma n y)\, dn.$$

In the previous section we had determined representatives of $\Gamma_N\backslash\Gamma$ in terms of double coset representatives. In the above sum, the term corresponding to $\pm$ the trivial coset yields

$$2\int_{\Gamma_N\backslash N}\varphi(ny)\, dn = 2\varphi(y).$$

Therefore we find

$$\begin{aligned}
\frac{1}{2}T^0T\varphi(y) &= \varphi(y) + \sum_{(\gamma)} \int_N \varphi(\gamma n y)\, dn \\
&= \varphi(y) + \sum_{(\gamma)} \int_N \varphi(h_\gamma w n_\gamma'' n y)\, dn \\
&= \varphi(y) + \sum_{(\gamma)} \int_N \varphi(h_\gamma w n y)\, dn \\
&= \varphi(y) + \sum_{(\gamma)} \int_N \varphi(w h_\gamma^{-1} n y)\, dn \\
&= \varphi(y) + \sum_{(\gamma)} \int_{\mathbf{R}} \varphi\left(w\begin{pmatrix} 1 & uc_\gamma^2 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} c_\gamma & 0 \\ 0 & 1/c_\gamma \end{pmatrix} y\right) du.
\end{aligned}$$

This yields, after a change of variables,

$$\boxed{\frac{1}{2}T^0T\varphi(y) = \varphi(y) + \sum_{(\gamma)} c(\gamma)^{-2}\int_N \varphi(wnh_\gamma^{-1}y)\,dn.} \tag{1}$$

The sum $\sum c(\gamma)^{-2}$ is related to the zeta function as follows. Define

$$F(s) = \sum_{(\gamma)} \frac{1}{c(\gamma)^{2s}}. \tag{2}$$

Letting ϕ be the Euler function, we have the identity

$$F(s) = \sum_{m=1}^{\infty} \frac{\phi(m)}{m^{2s}} = \frac{\zeta(2s-1)}{\zeta(2s)}. \tag{3}$$

To see this, note that each pair (c, d) of relatively prime integers,

$$c > 0,\ 0 \leqslant d < c,$$

represents exactly one of a pair of double cosets, as is clear from the formulas

$$\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} * & * \\ c & d \end{pmatrix}$$

and

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 1 & m \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} * & * \\ c & d+mc \end{pmatrix}.$$

This shows that $F(s) = \sum \phi(m)/m^{2s}$.

On the other hand, let μ be the Moebius function and let f be the function of the positive integers such that $f(m) = m$. Then under "convolution" product,

$$\mu * f(m) = \sum_{d|m} \mu(d)f(m/d) = \phi(m),$$

that is, $\mu * f = \phi$. Since

$$\zeta(2s-1) = \sum \frac{m}{m^{2s}} = \sum \frac{f(m)}{m^{2s}},$$

our second expression for $F(s)$ in terms of the zeta function follows, because the product of Dirichlet series arises from the convolution of their coefficients.

Remark. The Moebius function, the Euler function, and the function f are multiplicative. Convolution product relations among them are therefore verified by checking them on prime powers, in which case the verification is easy. Furthermore, μ and the constant function 1 are inverse to each other, so that

$$\frac{1}{\zeta(2s)} = \sum \frac{\mu(m)}{m^{2s}}.$$

From these remarks, the reader can prove for himself the relation needed above, namely $f = \phi * 1$.

§7. ANALYTIC CONTINUATION OF THE ZETA TRANSFORM OF $T^0T\varphi$

We are interested in the function $Z(T^0T\varphi, y, 2s)$. For this, we begin by some remarks concerning $Z(T^0\theta, y, 2s)$ for a more general function θ.

We view T^0 as a map

$$T^0\colon C_c(\Gamma\backslash G) \longrightarrow BC(N\backslash G),$$

where BC means bounded continuous.

Lemma 1. *If $\theta \in C_c(\Gamma\backslash G)$, then $T^0\theta(h_a y) = 0$ for large a.*

Proof. By definition,

$$T^0\theta(h_a y) = \int_{\Gamma_N\backslash N} \theta(nh_a y)\, dn.$$

As $a \longrightarrow \infty$, nh_a tends to infinity (i.e. lies outside a given compact set). Hence the expression under the integral sign is 0 for large a, as desired.

Lemma 2. *The integral*

$$Z(T^0\theta, y, 2 - 2s) = \int_0^\infty T^0\theta(h_a y) a^{2s-2} \frac{da}{a}$$

converges absolutely for $\mathrm{Re}\, s > 1$.

Proof. Using Lemma 1 to cut off at infinity, our integral is estimated by

$$\int_0^B a^{2\sigma-2}\,\frac{da}{a} = \left.\frac{a^{2\sigma-2}}{2\sigma-2}\right|_0^B$$

which exists, as desired.

Theorem 3. *Let $\varphi \in C_c^\infty(N\backslash G)$ be even, and assume that $\varphi[0] = \hat{\varphi}[0] = 0$. Then for* Re $s > 1$,

$$\frac{1}{2}\,Z(T^0T\varphi, y, 2-2s) = Z(\varphi, y, 2-2s) + \frac{\zeta(2-2s)}{\zeta(2s)}\,Z(\check{\varphi}, y, 2-2s).$$

The right-hand side is a meromorphic function in the whole plane, giving the analytic continuation of the left-hand side. It is holomorphic for Re $s \geqslant \frac{1}{2}$.

The proof of the above theorem will result from a number of computations, and transformations, giving rise to relations (1), (2), and (3) below. We shall use the following notation.

Let Re $s > 1$, *let h_γ be the Bruhat representative of γ in A. If a is a positive real number, let h_a be the corresponding element*

$$h_a = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix}, \qquad a = a(h)$$

in A. No ambiguity can arise, since a and γ lie in different kinds of sets.

We apply the Z-operator to the relation found in the preceding section, and we find by Lemma 2, for $\varphi \in C_c(N\backslash G)$

$$\begin{aligned}
&\tfrac{1}{2}Z(T^0T\varphi, y, 2-2s)\\
&\quad = \frac{1}{2}\int_A T^0T\varphi(hy)a(h)^{2s-2}\,dh\\
&\quad = Z(\varphi, y, 2-2s) + \sum_{(\gamma)} c(\gamma)^{-2}\int_N\int_A \varphi(wnh_\gamma^{-1}hy)a(h)^{2s-2}\,dn\,dh.
\end{aligned}$$

Let $h \mapsto h_\gamma h$ in the second term, so that $a(h_\gamma) = c(\gamma)^{-1}$. Then we see that the second term in this sum is equal to

$$\sum_{(\gamma)} c(\gamma)^{-2s} \int_N \int_A \varphi(wnhy) a(h)^{2s-2}\, dn\, dh$$

$$= F(s) \int_N \int_A \varphi(whny) a(h)^{2s}\, dn\, dh$$

$$= F(s) \int_N \int_A \varphi(h^{-1}wny) a(h)^{2s}\, dh\, dn$$

$$= F(s) \int_N \int_A \varphi(hwny) a(h)^{-2s}\, dh\, dn.$$

Therefore we obtain for $\mathrm{Re}(s) > 1$, and $\varphi \in C_c(N \backslash G)$:

$$(1) \qquad \boxed{\frac{1}{2} Z(T^0T\varphi, y, 2-2s) = Z(\varphi, y, 2-2s) + F(s) \int_N Z(\varphi, wny, 2s)\, dn.}$$

Furthermore, $Z(T^0T\varphi, y, 2 - 2s)$ and $Z(\varphi, y, 2 - 2s)$ are of type $2 - 2s$, so that the last term on the right of (1) is also of that type, so far for $\mathrm{Re}(s) > 1$. We shall obtain the analytic continuation of this last term by approaching it from another track, and are aiming for formula (2) below.

We had for $\mathrm{Re}\, s > 1$

$$E(\varphi, y, s) = TZ(\varphi, y, 2s) = \sum_{\Gamma_N \backslash \Gamma} Z(\varphi, \gamma y, 2s).$$

Then

$$T^0TZ(\varphi, y, 2s) = \int_{\Gamma_N \backslash N} \sum_{\Gamma_N \backslash \Gamma} Z(\varphi, \gamma n y, 2s)\, dn,$$

and, as above, we decompose the sum over double cosets separately for the trivial double coset and the non-trivial ones. We find:

$$\frac{1}{2} T^0TZ(\varphi, y, 2s) = Z(\varphi, y, 2s) + \sum_{(\gamma)} \int_N Z(\varphi, \gamma n y, 2s)\, dn$$

$$= Z(\varphi, y, 2s) + \sum_{(\gamma)} \int_N Z(\varphi, h_\gamma w n_\gamma'' n y, 2s) dn$$

$$= Z(\varphi, y, 2s) + \sum_{(\gamma)} \int_N Z(\varphi, h_\gamma w n y, 2s)\, dn.$$

Pulling out $c(\gamma)^{-2s}$ yields

$$\text{(2)} \qquad \boxed{\frac{1}{2}\, T^0TZ(\varphi, y, 2s) = Z(\varphi, y, 2s) + F(s)\int_N Z(\varphi, wny, 2s)\, dn.}$$

We shall now prove the relation completing the proof of Theorem 3, namely, for $\operatorname{Re} s > 1$,

$$\text{(3)} \qquad \boxed{F(s)\int_N Z(\varphi, wny, 2s)\, dn = \frac{\zeta(2-2s)}{\zeta(2s)}\, Z(\check{\varphi}, y, 2-2s).}$$

Proof. We prove this by a study of the various types, combined with whatever we know about analytic continuation.

Let us multiply relation (2) by $\zeta(2s)$, and let

$$L(\varphi, y, 2s) = \tfrac{1}{2}\zeta(2s)T^0TZ(\varphi, y, 2s).$$

By §3 giving the analytic continuation of the Eisenstein series, we see that $L(\varphi, y, 2s)$ is entire, and invariant under the transformation

$$(\varphi, s) \mapsto (\check{\varphi}, 1-s)$$

by the Corollary of Theorem 2.

The expression

$$L(\varphi, y, 2s) - \zeta(2s)Z(\varphi, y, 2s)$$

is equal to the last term in (2), multiplied by $\zeta(2s)$. It is meromorphic, and by (1) it is of type $2-2s$ for all complex s by analytic continuation. Therefore, making the transformation $(\varphi, s) \mapsto (\check{\varphi}, 1-s)$, we see that

$$L(\check{\varphi}, y, 2-2s) - \zeta(2-2s)Z(\check{\varphi}, y, 2-2s)$$

has type $2s$. Directly from the definition, we know that $Z(\check{\varphi}, y, 2-2s)$ has type $2-2s$, and is meromorphic by the theory of Mellin transforms, §4. We have

$$\begin{aligned} L(\check{\varphi}, y, 2-2s) &- \zeta(2-2s)Z(\check{\varphi}, y, 2-2s) \\ &= L(\varphi, y, 2s) - \zeta(2-2s)Z(\check{\varphi}, y, 2-2s) \end{aligned}$$

or in other words, the left-hand side of (2) multiplied by $\zeta(2s)$ is

$$L(\varphi, y, 2s) = \zeta(2-2s)Z(\check{\varphi}, y, 2-2s) + \text{an expression of type } 2s.$$

Since such a decomposition into a sum of terms of type $2s$ and $2 - 2s$ is uniquely determined, we conclude that the terms of type $2 - 2s$ are equal, and this amounts precisely to relation (3), thus concluding the proof of Theorem 3 also.

Theorem 4. *Let $\varphi \in C_c^\infty(N \backslash G)$ and assume $\hat{\varphi}[0] = 0$. Then*

$$Z(T^0 T\varphi, y, 2 - 2s)$$

is meromorphic, and holomorphic for $\operatorname{Re} s \geqq \frac{1}{2}$. *This function and*

$$Z(\check{\varphi}, y, 2 - 2s)$$

are rapidly decreasing on every vertical line for $\operatorname{Re} s \geqq \frac{1}{2}$, *uniformly in each finite strip.*

Proof. The Riemann zeta function $\zeta(2s)$ has no zero for $\operatorname{Re} 2s \geqq 1$, or in other words, for $\operatorname{Re} s \geqq \frac{1}{2}$. The denominator $\zeta(2s)$ in Theorem 3 therefore introduces no singularity in our stated region. The rapid decrease on vertical lines is a consequence of Lemma 2, §4.

In the applications of Theorem 4 and Theorem 3, we consider an integral taken over a vertical line with $\sigma > 1$ at first. Using the analytic continuation, holomorphicity up to the line $\sigma = \frac{1}{2}$, and the rapid decrease of the expressions under the integral sign, we shall be able to shift the line of integration to the line $\sigma = \frac{1}{2}$. For $\operatorname{Re} s = \sigma = \frac{1}{2}$ we have $1 - s = \bar{s}$, and the formula of Theorem 3 reads

$$\boxed{\frac{1}{2} Z(T^0 T\varphi, y, 2s) = Z(\varphi, y, 2s) + \frac{\zeta(2s)}{\zeta(2\bar{s})} Z(\check{\varphi}, y, 2s).} \tag{4}$$

§8. THE SPECTRAL DECOMPOSITION

Let f be a function of type $2s$ on G. Then f is determined by its values on K in the decomposition $G = NAK$. Functions like

$$Z(\varphi, y, 2s)$$

can therefore be viewed as functions on K, and as such will be denoted by omitting the variable, i.e. by writing $Z(\varphi, 2s)$. Thus the value of $Z(\varphi, 2s)$ at $k \in K$ is $Z(\varphi, k, 2s)$. The L^2-norm on K is denoted by

$$\|f\|_K^2 = \int_K |f(k)|^2 \, dk.$$

We use a similar notation for the L^2-norm on any other group or coset space, indicated as an index, and similarly for the scalar product.

We had seen in §1, Theorem 1, that the theta transform sends $C_c^\infty(N\backslash G)$ into a space orthogonal to the cusp forms, and the additional condition $\hat{\varphi}[0] = 0$ yielded the additional property that $T\varphi$ is orthogonal to the constant functions. We shall now obtain a completeness relation showing that nothing else is lost.

Theorem 5. *Let $\varphi, \psi \in C_c^\infty(N\backslash G)$ be even functions such that*

$$\hat{\varphi}[0] = \hat{\psi}[0] = 0.$$

Then

$$\langle T\varphi, T\psi\rangle_{\Gamma\backslash G} = \frac{1}{2\pi i}\int_{\mathrm{Re}\, s=\frac{1}{2},\ \mathrm{Im}\, s>0} \langle Z(T^0T\varphi, 2s), Z(T^0T\psi, 2s)\rangle_K \, ds$$

Proof. The formal computation preceding Theorem 1, §1, converges in the present case to yield

$$\langle T\varphi, T\psi\rangle_{\Gamma\backslash G} = \langle T^0T\varphi, \psi\rangle_{N\backslash G} = \int_{N\backslash G} T^0T\varphi(y)\overline{\psi(y)}\, dy$$

$$= \int_K\int_A T^0T\varphi(ak)\overline{\psi(ak)}\,\rho(a)^{-2}\, da\, dk.$$

By the Plancherel formula, expressed as Lemma 3, §4, this yields

$$(1)\qquad \langle T\varphi, T\psi\rangle_{\Gamma\backslash G} = \frac{1}{\pi i}\int_K\int_{\sigma>1} Z(T^0T\varphi, k, 2-2\bar{s})\,\overline{Z(\psi, k, 2s)}\, ds\, dk$$

where the integral over $\sigma > 1$ means the integral on the vertical line $\sigma + i\tau$, with some $\sigma > 1$. By Theorem 4 of the last section, we can now move the line of integration from $\sigma > 1$ to $\sigma = \frac{1}{2}$ because the function under the integral sign is antiholomorphic and rapidly decreasing in the strip between $\sigma = \frac{1}{2}$ and $\sigma > 1$.

On the other hand, for $\mathrm{Re}\, s > 1$,

$$E(\psi, \bar{s}) = TZ(\psi, 2\bar{s}).$$

Using the formal adjointness relation in a situation where it clearly converges, we find

$$\begin{aligned}\langle T\varphi, E(\psi, \bar{s})\rangle_{\Gamma\backslash G} &= \langle T\varphi, TZ(\psi, 2\bar{s})\rangle_{\Gamma\backslash G}\\ &= \langle T^0T\varphi, Z(\psi, 2\bar{s})\rangle_{N\backslash G}\\ &= \int_K\int_A T^0T\varphi(ak)\,\overline{Z(\psi, ak, 2\bar{s})}\,\rho(a)^{-2}\,da\,dk.\end{aligned}$$

Since $Z(\psi, y, 2\bar{s})$ is of type $2\bar{s}$, this last expression is

$$= \int_K Z(T^0T\varphi, k, 2-2s)\,\overline{Z(\psi, k, 2\bar{s})}\;dk. \tag{2}$$

This is true so far for $\operatorname{Re} s > 1$, but is also true for $\operatorname{Re} s = \frac{1}{2}$ by analytic continuation. We multiply this relation by $\zeta(2\bar{s})$. For $\operatorname{Re} s = \frac{1}{2}$ we have $1 - s = \bar{s}$. Therefore, recalling that we defined

$$E^*(\psi, s) = \zeta(2s)E(\psi, s),$$

we obtain for $\operatorname{Re} s = \frac{1}{2}$,

$$\langle T\varphi, E^*(\psi, \bar{s})\rangle_{\Gamma\backslash G} = \int_K Z(T^0T\varphi, k, 2\bar{s})\zeta(2\bar{s})\,\overline{Z(\psi, k, 2\bar{s})}\;dk. \tag{3}$$

The invariance of E^* under the map $(\psi, s) \mapsto (\check{\psi}, \bar{s})$ (on the line $\operatorname{Re} s = \frac{1}{2}$) implies that this last expression is invariant under this transformation also. Hence it is equal to

$$\int_K Z(T^0T\varphi, k, 2s)\zeta(2s)\,\overline{Z(\check{\psi}, k, 2s)}\;dk, \tag{4}$$

and consequently, dividing (3) and (4) by $\zeta(2\bar{s})$, we find

$$\int_K Z(T^0T\varphi, k, 2\bar{s})\,\overline{Z(\psi, k, 2\bar{s})}\;dk = \int_K Z(T^0T\varphi, k, 2s)\,\frac{\zeta(2s)}{\zeta(2\bar{s})}\,\overline{Z(\check{\psi}, k, 2s)}\,dk.$$

Finally, in the integral (1) for the scalar product $\langle T\varphi, T\psi\rangle_{\Gamma\backslash G}$ we can interchange the integration $dk\,ds$ because of the fast decreasing integrand. Hence

$$\langle T\varphi, T\psi\rangle_{\Gamma\backslash G} = \frac{1}{\pi i}\int_{\sigma=\frac{1}{2}}\int_K Z(T^0T\varphi, k, 2s)\,\overline{Z(\psi, k, 2s)}\;dk\,ds.$$

But

$$\int_{\frac{1}{2}}^{-} f(s)\,ds = \int_{\frac{1}{2}}^{+} f(\bar{s})\,ds,$$

where the symbols

$$\int_{\frac{1}{2}}^{-} \quad \text{and} \quad \int_{\frac{1}{2}}^{+}$$

indicate the integrals over the half vertical lines $\sigma = \frac{1}{2} + i\tau$, with

$$\tau < 0, \qquad \text{and} \qquad \tau > 0$$

respectively. Consequently

$$\langle T\varphi, T\psi\rangle_{\Gamma\backslash G} = \frac{1}{\pi i}\int_{\frac{1}{2}}^{+}\int_{K} Z(T^0T\varphi, k, 2s)\,\overline{Z(\psi, k, 2s)}\;dk\,ds$$

$$+ \frac{1}{\pi i}\int_{\frac{1}{2}}^{+}\int_{K} Z(T^0T\varphi, k, 2\bar{s})\,\overline{Z(\psi, k, 2\bar{s})}\;dk\,ds,$$

and by Theorem 3 (cf. also the end of the last section) this is

$$= \frac{1}{\pi i}\int_{\frac{1}{2}}^{+}\int_{K} Z(T^0T\varphi, k, 2s)\left[\overline{Z(\psi, k, 2s) + \frac{\zeta(2s)}{\zeta(2\bar{s})}Z(\check{\psi}, k, 2s)}\right] dk\,ds$$

$$= \frac{1}{2\pi i}\int_{\frac{1}{2}}^{+}\int_{K} Z(T^0T\varphi, k, 2s)\,\overline{Z(T^0T\psi, k, 2s)}\;dk\,ds$$

$$= \frac{1}{2\pi i}\int_{\frac{1}{2}}^{+}\langle Z(T^0T\varphi, 2s), Z(T^0T\psi, 2s)\rangle_K\,ds.$$

This proves our theorem.

Remark. We could have given the arguments so as to integrate on the lower part of the line $\sigma = \frac{1}{2}$ just as well.

The theorem shows that T^0 is unitary on the image of T, namely

Corollary. *We have* $\langle T\varphi, T\psi\rangle_{\Gamma\backslash G} = \langle T^0T\varphi, T^0T\psi\rangle_{N\backslash G}$.

Proof. The argument is formal and left to the reader.

XIV Spectral Decompositon of the Laplace Operator on $\Gamma\backslash\mathfrak{H}$

This chapter reproduces, with a number of details added, a paper of Faddeev [Fa 1].

Eigenfunctions of the Laplace operator on the upper half plane, automorphic with respect to a discrete subgroup of $SL_2(\mathbf{R})$, were introduced by Maass [Ma 1], [Ma 2] as an analogue of the classical automorphic forms. They were then discussed in two papers of Roelcke [Roe], and, very recently, Elstrodt [El]. They found important applications in Selberg's paper [Se 2] devoted to the trace formula. Connections with the theory of infinite dimensional representations were developed by Gelfand and Pjateckii-Shapiro [Ge, P-S], [Ge, Gr] and Gelfand–Fomin [Ge, Fo].

Let Γ be a discrete subgroup of $SL_2(\mathbf{R})$ such that $\Gamma\backslash\mathfrak{H}$ has finite volume. A fundamental domain F consists of a compact part F_0, and "cuspidal" parts, obtained by transforming the upper part of a strip by a finite number of elements of $SL_2(\mathbf{R})$. The Laplace operator L can be extended to a self-adjoint operator A in the Hilbert space $L^2(\Gamma\backslash\mathfrak{H})$. We want to describe the spectral decomposition of A, i.e. describe the eigenspaces, and find a kernel $\eta(z, s)$ with $z \in F$ and s in an appropriate space, called an Eisenstein function, such that the corresponding operator, the Eisenstein transform, sends A on a simple "multiplication" operator, as it turns out, multiplication by the function $\frac{1}{4} + t^2$, where t is a real variable. The Eisenstein functions satisfy a certain functional equation, which is intimately tied up with this spectral theory. The spectral decomposition is the subject of Kubota's book [Ku].

The most classical subgroups are $SL_2(\mathbf{Z})$ and its congruence subgroups. Roelcke investigated the general case, and observed that the spectral decomposition theorem could be proved if it were known that the corresponding Eisenstein functions had the same type of analytic continuation as those associated with $SL_2(\mathbf{Z})$. Selberg stated the appropriate theorem [Se 1], without proof.

Godement gave essentially classical proofs, relying on the Poisson summation formula for the arithmetic case, still not covering the general case. Langlands [La 1] gave very general proofs in the context of semisimple Lie groups, in an unpublished but fairly widely distributed manuscript. The method of Selberg and Langlands does not use the Poisson summation formula, but is more in the framework of operator theory. A summarized account occurs in [La 1]. The operator $M(\lambda)$ of Langlands is essentially the operator $\mathfrak{c}(s)$ of Harish-Chandra [H-C 1].

Faddeev reconsiders the question from a quite different point of view, that of perturbation theory originated by Friedrichs [Fr] and developed by Povzner [Po]. We sketch Faddeev's method. The source of the functional equation and analytic continuation lies in the resolvant equation

$$R(s) - R(s') = (s(1-s) - s'(1-s'))R(s')R(s)$$

for the resolvant of the Laplace operator. The parameter s is the same as that in the usual theory of Dirichlet series, and the corresponding eigenvalues are

$$\lambda_s = s(1-s).$$

With the above normalization, the critical line is on $\operatorname{Re} s = \frac{1}{2}$.

We select a large number $\kappa > 0$, and analyze the resolvant equation for $s' = \kappa$, as a function of s. Putting $R = R(\kappa)$, we have the equation

$$R(s) - R = \omega(s)RR(s).$$

We then select an appropriate Green's function $q_s(y, y')$ for the differential equation

$$\psi''(y) = -\frac{s(1-s)}{y^2}\psi(y),$$

satisfying a certain boundary condition for $y \geqslant a$. Let $Q(s)$ be the corresponding cuspidal operator. Instead of studying $R(s)$ directly, we make the transformation

$$R(s) = Q(s) + (I + \omega Q(s))B(s)(I + \omega Q(s)),$$

which can be shown to be solvable for an operator $B(s)$, which turns out to be analytic in s for $0 < \operatorname{Re} s < 2$, except for a discrete set of poles. The resolvant equation for $R(s)$ has a corresponding equation for $B(s)$. One can then construct an operator

$$W(s) = \omega(s)(I + \omega(s)Q(s))B(s)$$

which one uses to perturb the identity.

The Green's function q_s is the form

$$q_s(z, z') = \frac{1}{2s-1}\begin{cases} \theta(y, s)y'^{1-s} & \text{if } y < y', \\ y^{1-s}\theta(y', s) & \text{if } y > y', \end{cases}$$

where

$$\theta(y, s) = y^s + c(s)y^{1-s},$$

$$c(s) = a^{2s-1}\frac{s-\kappa}{s+\kappa-1}.$$

These basic functions satisfy the Eisenstein formalism

$$\theta(y,s) = \theta(y,1-s)c(s),$$

$$c(s)c(1-s) = 1.$$

If we define

$$\eta(z,s) = [I + W(s)]\theta(z,s),$$

then the Eisenstein functions η_s satisfy an analogous formalism, where the functional equation and analytic continuation come from the resolvant equation for $B(s)$ and the analyticity properties of $W(s)$.

The general scheme of the above arguments is quite similar to that used by Faddeev himself in a previous paper [Fa 2], in other connections.

The analytic continuation of the families of operators is done simultaneously with the continuation of their kernels. It is important to distinguish those regions where the kernel of the resolvant has an analytic continuation, where it does not represent the resolvant. Essentially nothing is known about these in general. For the special case of $SL_2(\mathbf{Z})$, one sees that the possible poles of the analytic continuation of the kernel of the resolvant on the left on the line $\operatorname{Re} s = \frac{1}{2}$ coincide with the zeros of Riemann's zeta function $\zeta(2s)$. Thus the line $\operatorname{Re} s = \frac{1}{4}$ becomes another critical line (much more critical), as had already been observed by Selberg [Se 1].

If I were to teach someone analysis I would tell him to read the Faddeev paper in complete detail. Many techniques of analysis are brought to bear in a coherent and fascinating general context, albeit in a concrete no nonsense situation. As I already said in the foreword, all auxiliary results from general analysis are reproduced in appendices, to simplify the reader's task.

The arguments are carried through first in the case of $SL_2(\mathbf{Z})$, when there is only one cusp to deal with. A final section points out those places where a linguistic change has to be made to deal with the general case. Essentially it does not go beyond inserting n indices here and there. The function $\mathbf{c}(s)$ then becomes a matrix (operator) whose size is $n \times n$, where n is the number of cusps for the discrete subgroup Γ. It is closely related to the scattering operator in physics, as is shown in a forthcoming paper by Faddeev and Povzner.

Faddeev's method has been used by Lachaud to carry out the spectral decomposition theorem for so-called groups of rank 1. Hopefully the method also extends to groups of higher rank.

In the exposition, I have added a lot of details left out by Faddeev. Perhaps the necessary expansion of pages makes it a little more difficult to follow through the main trend of thought. The reader is therefore encouraged to look at the original paper.

§1. GEOMETRY AND DIFFERENTIAL OPERATORS ON $\mathfrak{H}$

Geometry

The group $G = SL_2(\mathbf{R})$ acts on the upper half plane $\mathfrak{H}$ in the usual way,

$$z \mapsto \frac{az + b}{cz + d} = \gamma z \qquad \text{if} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The function

$$\boxed{u(z, z') = \frac{|z - z'|^2}{4yy'}}$$

of pairs of points $z = x + iy$ and $z' = x' + iy'$ is obviously invariant under G, i.e. for $\gamma \in G$,

$$u(\gamma z, \gamma z') = u(z, z').$$

In what follows we deal exclusively with the function $u(z, z')$ (except for one incidental use of the invariant area), and the reader could skip the following discussion of the Poincaré metric. However, it is well to get acquainted with the Poincaré geometry for intuitive purposes.

In the upper half plane, the Poincaré metric is defined by

$$ds^2 = \frac{dx^2 + dy^2}{y^2} = \frac{dz\, d\bar{z}}{y^2}.$$

Since $d(\gamma z) = (cz + d)^{-2}\, dz$ and $\mathrm{Im}(\gamma z) = \mathrm{Im}(z)/|cz + d|^2$, it follows that the metric is invariant under G, and gives rise to a distance function $\rho(z, z')$, equal to the length of the shortest curve joining z and z'. If

$$z(t) = x(t) + iy(t), \quad a \leqslant t \leqslant b,$$

is a curve in $\mathfrak{H}$, then by definition, its length is

$$\int_a^b \frac{\sqrt{x'(t)^2 + y'(t)^2}}{y(t)}\, dt.$$

Given $z, z' \in \mathfrak{H}$, there exists $\gamma \in G$ such that $\gamma z = i$ and $\gamma z' = iy_0$ for some real number $y_0 \geqslant 1$. Then

$$\rho(z, z') = \rho(i, iy_0).$$

We give a brief argument that the vertical line segment between i and iy_0 is actually the shortest curve. Let $x(t) + iy(t)$ be a curve joining i and iy_0. Then its length is

$$\geqslant \int_a^b \frac{|y'(t)|}{y(t)}\, dt \geqslant \left| \int_a^b \frac{y'(t)}{y(t)}\, dt \right|.$$

If the curve has minimal length, then equality must hold everywhere, for otherwise the curve $iy(t)$, which joints i and iy_0, would have shorter length. Hence the x-component of the curve must be 0, and the argument is concluded.

The distance between a point it ($t > 1$) and i is now trivially computed. We have

$$\rho(i, it) = \int_0^t \frac{dy}{y} = \log t.$$

So the distance is given by

$$s = \log t \qquad \text{and} \qquad t = e^s.$$

Furthermore,

$$\boxed{u = \sinh^2 \frac{s}{2}}$$

because

$$\begin{aligned}\sinh^2 \frac{s}{2} &= \frac{\cosh s - 1}{2}\\ &= \frac{1}{2}\left(\frac{t + 1/t}{2} - 1\right)\\ &= \frac{(t-1)^2}{4t} = u.\end{aligned}$$

In the representation on the unit disc, the metric is given by

$$ds^2 = \frac{4(dx^2 + dy^2)}{(1 - r^2)^2}$$

where $r^2 = x^2 + y^2$. Hence in this case,

$$s(r) = 2\int_0^r \frac{d\rho}{1 - \rho^2} = \log \frac{1 + r}{1 - r}.$$

Therefore $r = (e^s - 1)/(e^s + 1)$, so that

$$\boxed{r = \tanh \frac{s}{2}.}$$

We shall also prove that the area $A(r)$ of a disc of radius r for the Poincaré distance is given by

$$\boxed{A(r) = 4\pi u.}$$

Indeed,

$$\begin{aligned}A(r) &= \int_0^r \int_0^{2\pi} \frac{4\rho\, d\rho\, d\theta}{(1 - \rho^2)^2} = 4\pi \frac{r^2}{1 - r^2}\\ &= 4\pi \frac{\tanh^2(s/2)}{1 - \tanh^2(s/2)} = 4\pi \sinh^2(s/2) = 4\pi u.\end{aligned}$$

Going back to the representation on the upper half plane, let $D_R(z)$ be the set of points z' such that $u(z, z') \leqslant R$, in other words, the disc of radius R

and center z for the "distance" $u(z, z')$. Then D_R is defined by the equation for an ordinary Euclidean disc,

$$x^2 + (y - b)^2 \leqslant a^2,$$

where $b = 2R + 1$ and $a^2 = 2R(2R + 2)$. As R becomes large, this disc comes closer and closer to the real line, and can be drawn as in Fig. 1.

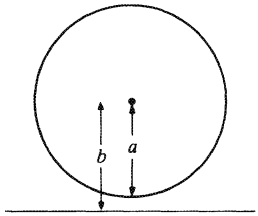

Figure 1

Differential operators

Let $\mathfrak{g}$ be the Lie algebra of $G = SL_2(\mathbf{R})$, i.e. the set of real matrices with trace 0. For $X \in \mathfrak{g}$ we have differential operator L_X on $C^\infty(\mathfrak{H})$ given by

$$L_X f(z) = \frac{d}{dt} f((\exp tX)z)|_{t=0}.$$

The operators corresponding to

$$X_1 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$$

are denoted by L_1, L_2, L_3. In terms of the coordinates $z = (x, y)$ they are represented by the following combinations of partial derivatives:

$$L_1 = \frac{\partial}{\partial x},$$

$$L_2 = (y^2 - x^2)\frac{\partial}{\partial x} - 2xy\frac{\partial}{\partial y},$$

$$L_3 = 2x\frac{\partial}{\partial x} + 2y\frac{\partial}{\partial y}.$$

These relations are easily proved. We indicate how the proof goes, say for X_2 and L_2. Let

$$f_*(t) = f\left(\begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix} z\right) = f(g(t))$$

where

$$g(t) = \frac{x + iy}{t(x + iy) + 1}.$$

Then

$$g'(0) = (y^2 - x^2) - 2ixy.$$

If $g(t) = u(t) + iv(t)$, then

$$f_*'(0) = \frac{\partial f}{\partial u}\frac{\partial u}{\partial t}\bigg|_{t=0} + \frac{\partial f}{\partial v}\frac{\partial v}{\partial t}\bigg|_{t=0},$$

and the desired formula drops out.

We shall be concerned with the **Laplacian**,

$$L = -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) = -L_3^2 - \frac{1}{2}(L_2L_3 + L_3L_2).$$

The minus sign is used in order to make L "positive definite" rather than "negative definite" in the following context. Let φ, ψ be in $C_c^\infty(\mathfrak{H})$, and let them be real functions. In the scalar product integral

$$\langle L\varphi, \psi\rangle = \iint_{\mathfrak{H}} L\varphi \cdot \psi \frac{dx\, dy}{y^2}$$

one can integrate by parts, and transfer one partial from φ to ψ, with a change of sign which cancels the minus sign in front. Thus

$$\langle L\varphi, \varphi\rangle = \iint_{\mathfrak{H}} \left[\left(\frac{\partial\varphi}{\partial x}\right)^2 + \left(\frac{\partial\varphi}{\partial y}\right)^2\right] dx\, dy > 0 \quad \text{if} \quad \varphi \neq 0.$$

The measure $dx\, dy/y^2$ is invariant under $SL_2(\mathbf{R})$, and will be denoted by dz all the way through this chapter.

The operator L is G-invariant. This means that if

$$\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbf{R})$$

and τ_σ is the translation

$$\tau_\sigma f(z) = f(\sigma z),$$

then

$$\boxed{\tau_\sigma \circ L = L \circ \tau_\sigma.}$$

This can be checked by a direct computation using brute force, and the repeated application of the chain rule from freshman calculus, having to take the partials of

$$z' = \frac{a(x + iy) + b}{c(x + iy) + d}$$

with respect to x and y. On the other hand, we already know it from X, §1, Th. 4, that is, from general considerations concerning the structure of the algebra of differential operators on $SL_2(\mathbf{R})$.

The Laplacian will be applied to functions of the u-distance, and hence it has an expression as an ordinary differential operator in terms of u only, namely:

Let φ be a C^∞ function on the positive reals, and let z_0 be a point of $\mathfrak{H}$. Let $f(z) = \varphi(u(z, z_0))$. Then

$$Lf = l\varphi,$$

where

$$l\varphi(u) = -(u^2 + u)\varphi''(u) - (1 + 2u)\varphi'(u).$$

Proof. We have:

$$-y^2\left[\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}\right] = -y^2\{\varphi''[u_x^2 + u_y^2] + \varphi'[u_{xx} + u_{yy}]\},$$

and we have to show that

$$y^2(u_x^2 + u_y^2) = u^2 + u,$$

$$y^2(u_{xx} + u_{yy}) = 1 + 2u.$$

But

$$u(z, z') = \frac{(x - x')^2 + (y - y')^2}{4yy'},$$

and we compute

$$u_x = \frac{x - x'}{2yy'}, \qquad u_y = \frac{y^2 - y'^2 - (x - x')^2}{4y^2y'},$$

$$u_{xx} = \frac{1}{2yy'}, \qquad u_{yy} = \frac{y'^2 + (x - x')^2}{2y^3y'}.$$

What we want drops out.

§2. A SOLUTION OF $l\varphi = s(1 - s)\varphi$

At the end of the last section, we computed the Laplace operator on functions depending only on the distance. We now exhibit a solution of the homogeneous differential equation for this operator. It is given by the classical integral

$$\varphi_s(u) = \varphi(u, s) = \frac{1}{4\pi}\int_0^1 [t(1 - t)]^{s-1}(t + u)^{-s}\, dt$$

absolutely convergent for complex $s = \sigma + i\tau$, $\sigma > 0$, and $u > 0$.

Theorem 1. *In the above region, φ is analytic in s, C^∞ in u, and:*

i) $l\varphi = s(1 - s)\varphi$

ii) $\varphi(u, s) = -\dfrac{1}{4\pi}\log u + O(1)$ for fixed s, $u \to 0$;

iii) $\varphi_s'(u) = -\dfrac{1}{4\pi u} + O(1)$ for fixed s, $u \to 0$;

iv) $\varphi_s(u) = O(u^{-\sigma})$ for fixed σ, $u \to \infty$.

Proof. Property (iv) is clear. To prove that φ satisfies the desired differential equation, we may differentiate under the integral sign, and it turns out that applying l_u to this integrand turns it into an exact differential, with zero boundary value. More precisely, a trivial direct computation shows that if

$$M_u = (u^2 + u)\left(\frac{d}{du}\right)^2 + (1 + 2u)\frac{d}{du} + s(1 - s),$$

then

$$M_u[t(1-t)]^{s-1}(t+u)^{-s} = s\,\frac{d}{dt}\left\{[t(1-t)]^s(t+u)^{-s-1}\right\}.$$

Since the boundary values of the function on the right at 0 and 1 are 0, it follows that

$$(l - s(1-s))\varphi = 0.$$

Remark. In general, let $\mathfrak{L}_u$ be a differential operator with respect to a variable u in a manifold, and let $\omega(t, u)$ be a function of u and a differential form with respect to a variable t in a manifold with boundary T. We call $\omega(t, u)$ a **resolving form** for $\mathfrak{L}$ if there exists a form $\eta(t, u)$ such that $\eta(t, u) = 0$ for t on the boundary, and such that

$$\mathfrak{L}_u\omega(t, u) = d_t\eta(t, u),$$

where d_t is ordinary exterior differentiation. Stokes theorem shows that

$$\int_T \omega(t, u)$$

is a solution of $\mathfrak{L}\varphi = 0$. This technique of resolving forms will again be used in connection with the Whittaker equation later in this chapter. It would be interesting to study it quantitatively for appropriate differential operators on higher dimensional manifolds.

Next, we prove that

$$\int_0^1 [t(1-t)]^{s-1}(t+u)^{-s}\,dt = -\log u + O(1)$$

for $u \to 0$. It is clear that the integral from $\frac{1}{2}$ to 1 is bounded as a function of u for u near 0. Hence let

$$I = \int_0^{1/2} (1-t)^{s-1}\frac{t^{s-1}}{(t+u)^s}\,dt$$

and

$$A(r, u) = \int_0^r \frac{t^{s-1}}{(t+u)^s}\,dt.$$

Changing variables, $t = u\tau$ yields

$$A(r, u) = \int_0^{r/u} \left(1 + \frac{1}{\tau}\right)^{-s} \frac{d\tau}{\tau} = A\left(\frac{r}{u}\right)$$

where

$$A(x) = \int_0^{x} \left(1 + \frac{1}{\tau}\right)^{-s} \frac{d\tau}{\tau} = \begin{cases} \log x + O(1) & \text{if } \ 1 \leqslant x, \\ O(1) & \text{if } \ x \leqslant 1. \end{cases}$$

We have

$$\begin{aligned} I &= \int_0^{1/2} (1 - t)^{s-1} \frac{dA(t, u)}{dt}\, dt \\ &= \left(1 - \frac{1}{2}\right)^{s-1} A\left(\frac{1}{2}, u\right) - \int_0^{1/2} \frac{d}{dt}(1 - t)^{s-1} A(t, u)\, dt. \end{aligned}$$

We note that

$$A\left(\frac{1}{2}, u\right) = A\left(\frac{1}{2u}\right) = -\log u + O(1).$$

We write

$$\int_0^{1/2} = \int_0^{u} + \int_u^{1/2}.$$

If $0 \leqslant t \leqslant u$, then $A(t, u) = A(t/u)$ is bounded, and hence the integral from 0 to u is $O(1)$. For the other integral, we use the other estimate for $A(t, u) = A(t/u)$ to get

$$\begin{aligned} &-\int_u^{1/2} \frac{d}{dt}(1 - t)^{s-1}[\log t - \log u + O(1)]\, dt \\ &= \log u \int_u^{1/2} \frac{d}{dt}(1 - t)^{s-1}\, dt + O(1) \\ &= (\log u)\left[(1 - \tfrac{1}{2})^{s-1} - (1 - u)^{s-1}\right] + O(1). \end{aligned}$$

The terms $(\log u)(1 - \frac{1}{2})^{s-1}$ cancel, to leave

$$\begin{aligned} I &= -(\log u)(1 - u)^{s-1} + O(1) \\ &= -\log u + O(1), \end{aligned}$$

as desired. (I owe this proof to Lax.)

The asymptotic estimate for $\varphi_s'(u)$ as $u \longrightarrow 0$ is proved in an analogous manner, after differentiating under the integral sign.

§3. THE RESOLVANT OF THE LAPLACE OPERATOR ON $\mathfrak{H}$ FOR $\sigma > 1$

Let $BC^\infty(\mathfrak{H})$ be the space of bounded C^∞ functions on $\mathfrak{H}$.

Theorem 2. *Let $f \in BC^\infty(\mathfrak{H})$. For* $\mathrm{Re}(s) > 1$, *let*

$$R_{\mathfrak{H}}(s)f(z) = \int_{\mathfrak{H}} \varphi_s(u(z, z'))f(z')\,dz'$$

where dz is the G-invariant measure $dx\,dy/y^2$. Then $R_{\mathfrak{H}}(s)f$ is also bounded, C^∞, and we have

$$(L - s(1 - s)I)R_{\mathfrak{H}}(s)f = f.$$

In other words, we have found a right inverse for $L - s(1 - s)I$, where I is the identity. It is given by a kernel

$$r_{\mathfrak{H}}(z, z'; s) = \varphi(u(z, z'); s),$$

which we also write $r_{\mathfrak{H}}(s) = \varphi_s \circ u$ if we want to omit (z, z').

The proof will be carried out in three steps.

First, we show that $R_{\mathfrak{H}}(s)f$ is bounded.

Second, we prove the special case of the theorem when f has compact support, by potential theory.

Third, we prove the general case by applying the fundamental theorem on elliptic operators, whose proof is recalled in an appendix for the convenience of the reader.

Lemma 1. *If $f \in BC^\infty(\mathfrak{H})$, then $R_{\mathfrak{H}}(s)f$ is bounded, if $\sigma > 1$.*

Proof. Let

$$h(z) = \int_{\mathfrak{H}} \varphi_s(u(z, z'))f(z')\,dz'.$$

Then

$$|h(z)| \leqslant \|f\| \int_{\mathfrak{H}} |\varphi_s(u(z, z'))|\,dz',$$

and the integral on the right is independent of z. Indeed, if $\gamma \in G$,

$$\begin{aligned}\int_{\mathfrak{H}} |\varphi_s(u(\gamma z, z'))|\, dz' &= \int_{\mathfrak{H}} |\varphi_s(u(z, \gamma^{-1}z'))|\, dz' \\ &= \int_{\mathfrak{H}} |\varphi_s(u(z, z'))|\, dz'\end{aligned}$$

because dz' is G-invariant. We may therefore assume that $z = i$. Let D be a small disc around i, and write

$$\int_{\mathfrak{H}} = \int_{\mathfrak{H}-D} + \int_{D}.$$

For small u, $\varphi_s(u) \sim \log u$ which is locally integrable, so the integral over D exists. On the other hand, on $\mathfrak{H} - D$, we have:

$$\int_{\mathfrak{H}-D} |\varphi_s(u(z, z'))|\, dz' \ll \int_{\mathfrak{H}-D} \frac{1}{(1+u)^{\sigma}}\, dz'.$$

So it suffices to prove the next estimate.

Lemma 2. *The integral*

$$\int_{\mathfrak{H}} \frac{1}{(1 + u(z, z'))^{\sigma}}\, dz'$$

converges for $\sigma > 1$.

Proof. We may assume $z = i$. The integral is then equal to

$$\int_{-\infty}^{\infty}\int_{0}^{\infty} \frac{y^{\sigma-2}}{\left[x^2 + (y-1)^2\right]^{\sigma}}\, dy\, dx.$$

We write

$$\int_0^{\infty} = \int_0^{\delta} + \int_{\delta}^{\infty}.$$

The matter is routine, and we actually have already carried out the estimate in full in Chapter IX, §2.

Potential theory

For the convenience of the reader we recall some elementary potential theory in the plane. Let f, g be two functions on $\mathbf{R}^2$. Let U be a region with piecewise smooth boundary. For any vector field F on U, we have the

Stokes–Green theorem in dimension 2,

$$\iint_U \operatorname{div} F\, dx\, dy = \int_{\partial U} F \cdot n\, ds.$$

Let $F = g(\operatorname{grad} f) = (gf_x, gf_y)$ or $f(\operatorname{grad} g) = (fg_x, fg_y)$. Let

$$\Delta = \operatorname{div} \operatorname{grad} = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

We get the formula

$$\iint_U (g\,\Delta f - f\,\Delta g)\, dx\, dy = \int_{\partial U} \left(g \frac{\partial f}{\partial n} - f \frac{\partial g}{\partial n} \right) ds.$$

Let z_0 be a point in the plane, and take for $U = U(\epsilon)$ the outside of the circle $S(\epsilon)$ centered at z_0, with radius ϵ. *Assume that f has compact support.* Then we have **Green's formula**

$$\iint_{U(\epsilon)} (g\,\Delta f - f\,\Delta g)\, dx\, dy = -\int_{S(\epsilon)} \left(g \frac{\partial f}{\partial n} - f \frac{\partial g}{\partial n} \right) ds.$$

Example. Let z_0 be fixed, and let

$$g(z) = \frac{1}{2\pi} \log|z - z_0|.$$

In polar coordinates,

$$\frac{\partial f}{\partial n} = \frac{\partial f}{\partial r}$$

if $f = f(r, \theta)$, where $r = |z - z_0|$. We have $ds = d\theta$ if we parametrize the circle by $(\epsilon \cos \theta/\epsilon,\ \epsilon \sin \theta/\epsilon)$, $0 \leqslant \theta \leqslant 2\pi\epsilon$. Also, $\Delta g = 0$. The right-hand side of Green's formula gives

$$\int_{S(\epsilon)} \left(f \frac{\partial g}{\partial n} - g \frac{\partial f}{\partial n} \right) ds = \int_0^{2\pi\epsilon} f(\epsilon, \theta) \frac{1}{2\pi\epsilon}\, d\theta - \int_0^{2\pi\epsilon} \frac{1}{2\pi} (\log \epsilon) \frac{\partial f}{\partial r}\, d\theta.$$

As $\epsilon \to 0$, the first term goes to $f(0, 0)$ and the second term goes to 0. Hence

$$\iint_U g\, \Delta f\, dx\, dy = f(z_0).$$

A similar argument will be given in the upper half plane.

Lemma 3. *Let $\sigma > 1$ and $f \in C_c^\infty(\mathfrak{H})$. Let*

$$h(z) = \int_{\mathfrak{H}} \varphi_s(u(z, z'))f(z')\, dz'.$$

then h is C^∞. Let $M_s = L - s(1 - s)I$. Then $M_s h = f$, and

$$M_s h = \int_{\mathfrak{H}} \varphi_s(u(z, z'))(M_s f)(z')\, dz' = \int_{\mathfrak{H}} r_{\mathfrak{H}}(z, z'; s)(M_s f)(z')\, dz'.$$

Proof. We omit the subscript s which is fixed. For z varying in a small open set, let

$$z \longmapsto \gamma(z) \in G$$

be a C^∞ map such that $\gamma(z)i = z$ (so $z \longmapsto \gamma(z)$ is a C^∞ section). Abbreviate $\varphi(u(z, z'), s) = k(z, z')$. Then

$$\begin{aligned} h(z) &= \int k(\gamma(z)i, z')f(z')\, dz' \\ &= \int k\big(i, \gamma(z)^{-1}z'\big)f(z')\, dz' \\ &= \int k(i, z')f(\gamma(z)z')\, dz'. \end{aligned}$$

Since f has compact support, we can differentiate under the integral sign, and we see that h is C^∞.

Lemma. $L_j h(z) = \int k(z, z')L_j f(z')\, dz'.$

Proof. We have

$$h\big(\exp(tX_j)z\big) = \int k\big(\exp(tX_j)z, z'\big)f(z')\, dz'.$$

We can move $\exp(tX_j)$ over to z' after taking its inverse because

$$k(\gamma z, \gamma z') = k(z, z'), \qquad \text{all } \gamma \in G,$$

Using the invariance of dz' under G, we find

$$h\big(\exp(tX_j)z\big) = \int k(z, z')f\big(\exp(tX_j)z'\big)\, dz'.$$

We can differentiate under the integral sign since f has compact support, and our lemma follows at once.

In particular, we find the intermediate formula

$$\boxed{M_s h(z) = \int k(z, z')(M_s f)(z')\, dz'.}$$

We apply it to the relation of potential theory:

$$\int_U\int (fLg - gLf)\, dz = \int_U\int (g\,\Delta f - f\,\Delta g)\, dx\, dy$$

$$= \int_{\partial U}\left(g\frac{\partial f}{\partial n} - f\frac{\partial g}{\partial n}\right) ds.$$

We let $U = U(\epsilon)$ be the outside of a small Euclidean disc of radius ϵ centered at z, so that $\partial U(\epsilon) = S(\epsilon)$ is a circle. Omit the subscript s from M for simplicity. We get

$$\int_{U(\epsilon)}\int k(z, z')Mf(z')dz' - \int_{U(\epsilon)}\int M_{z'}k(z, z')f(z')\, dz' = -\int_{S(\epsilon)}\left(k\frac{\partial f}{\partial n} - f\frac{\partial k}{\partial n}\right)ds.$$

But $M_{z'}k(z, z') = 0$ away from the diagonal, so the term involving $M_{z'}k(z, z')$ vanishes. Since $k(z, z')$ behaves like $\log u(z, z')$ for z' near z, it follows that the integral of

$$k\frac{\partial f}{\partial n}$$

over the circle tends to 0 as $\epsilon \longrightarrow 0$. Finally, we know from §2, Th. 1, that $\varphi'(u) = -1/4\pi u + O(1)$, and

$$\frac{\partial k(z, z')}{\partial r'} = \varphi'(u)\frac{\partial u}{\partial r'}\,.$$

It follows at once that the integral of $\partial k/\partial n$ tends to 1 as $\epsilon \longrightarrow 0$. This proves Theorem 2 when f has compact support.

Finally, we deal with the general case when f does not necessarily have compact support, but $f \in BC^{\infty}(\mathfrak{H})$. We know from Lemma 1 that $R_{\mathfrak{H}}(s)f = h$ is bounded. Furthermore, $M = M_s$ is an elliptic operator on $\mathfrak{H}$. Let $\psi \in C_c^{\infty}(\mathfrak{H})$ and assume ψ real. Then

$$\langle h, M\psi\rangle = \int\int k(z, z')M\psi(z)\, dz\, dz',$$

and we can apply Fubini. By the special case proved for functions with compact support, we obtain

$$\int k(z, z')M\psi(z)\,dz = \psi(z').$$

Hence

$$\langle h, M\psi\rangle = \langle f, \psi\rangle.$$

By the regularity theorem for elliptic operators, this implies that h is C^∞ and

$$Mh = f.$$

This proves Theorem 2.

§4. SYMMETRY OF THE LAPLACE OPERATOR ON $\Gamma\backslash\mathfrak{H}$

The symmetry of the Laplace operator on $\mathfrak{H}$ was already mentioned briefly in §1. We now want to see that the Laplace operator is also symmetric on $\Gamma\backslash\mathfrak{H}$. We shall need a cutoff function for technical purposes, i.e. a function $\zeta_Y(Y)$, for large positive numbers Y whose graph is that in Fig. 2.

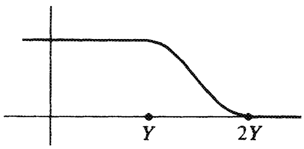

Figure 2

In other words:

1. $\zeta_Y(y) = 1$ if $y \leqslant Y$, $\zeta_Y(y) = 0$ if $y \geqslant 2Y$.
2. $\zeta_Y''(y) \ll 1/Y^2$.

Such a function is easily constructed. Let $\psi(t)$ have the graph shown in Fig. 3.

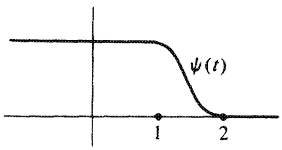

Figure 3

Then ψ' and ψ'' are bounded. It is clear that the function

$$\zeta_Y(t) = \psi(t/Y)$$

satisfies our requirements.

We let $\Gamma = SL_2(\mathbf{Z})$. A fundamental domain F for Γ is given by the illustrated region in Fig. 4.

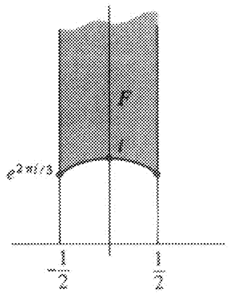

Figure 4

We omit the proof that this is a fundamental domain, easily accessible in various references. We can identify $\Gamma\backslash\mathfrak{H}$ to F.

As before, we let

$$L = -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) \quad \text{and} \quad \Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}.$$

We recall that L is invariant under $G = SL_2(\mathbf{R})$. Hence if $f \in C^\infty(\Gamma\backslash\mathfrak{H})$, i.e. f is a C^∞ function on $\mathfrak{H}$ which is invariant under Γ, then Lf is also C^∞ on $\mathfrak{H}$, and is invariant under Γ.

Lemma 1. *If f, Lf are in $BC^\infty(\Gamma\backslash\mathfrak{H})$ and real, then the integral*

$$[f, f] = \iint_F \left(\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2\right) dx\, dy$$

over a fundamental domain F is finite, and we have

$$\langle Lf, f\rangle = [f, f].$$

Proof. Let Y be a large positive number, and $\zeta = \zeta_Y$ a cutoff function as above, between Y and $2Y$. Our first goal is to prove formula (*) below. Let

$$\omega(x,y) = \frac{\partial f}{\partial x} f\zeta\, dy - \frac{\partial f}{\partial y} f\zeta\, dx = \frac{\partial f}{\partial n} f\zeta\, ds.$$

By Stokes–Green we have

$$-\iint_{\Gamma\backslash\mathfrak{H}} (Lf)f\zeta \frac{dx\,dy}{y^2} = \iint_{\Gamma\backslash\mathfrak{H}} (\Delta f)f\zeta\, dx\,dy$$

$$= \iint_{\Gamma\backslash\mathfrak{H}} d\omega - \iint_{\Gamma\backslash\mathfrak{H}} \left[\frac{\partial f}{\partial x}\frac{\partial(f\zeta)}{\partial x} + \frac{\partial f}{\partial y}\frac{\partial(f\zeta)}{\partial y} \right] dx\,dy$$

$$= -\iint_F \left[\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 \right] \zeta\, dx\,dy - \iint_F f \frac{\partial f}{\partial y}\frac{\partial \zeta}{\partial y}\, dx\,dy,$$

because $\Gamma\backslash\mathfrak{H}$ has no boundary. [If you want, work entirely on F, and use the periodicity of $\partial f/\partial n$, together with the fact that pieces of the boundary correspond to each other with reversed orientation under Γ.] We also used $\partial\zeta/\partial x = 0$. But

$$\frac{1}{2} d\left(f^2 \frac{\partial \zeta}{\partial y}\, dx\right) = -\left[f \frac{\partial f}{\partial y}\frac{\partial \zeta}{\partial y} + \frac{1}{2} f^2 \frac{\partial^2 \zeta}{\partial y^2} \right] dx\wedge dy.$$

However, $\partial\zeta/\partial y = 0$ whenever $dx \neq 0$ since Y is large. Hence

$$(*) \qquad \iint_F (Lf)f\zeta \frac{dx\,dy}{y^2} = \iint_F \left[\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 \right] \zeta(y)\, dx\wedge dy$$

$$- \frac{1}{2}\int_0^1 \int_Y^{2Y} f^2(z)\zeta_Y''(y)\, dx\,dy.$$

As $Y \to \infty$, the second term on the right is $O(1/Y)$ and therefore

$$\iint_F \left[\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2 \right] \zeta_Y(y)\, dx\,dy = \iint_F (Lf)f\zeta_Y \frac{dx\,dy}{y^2} + O(1/Y).$$

As $Y \to \infty$ the left-hand side increases and remains bounded because the right-hand side tends to

$$\iint_F (Lf)f \frac{dx\,dy}{y^2}.$$

This proves the lemma.

Theorem 3. *Let $f, g \in BC^\infty(\Gamma\backslash\mathfrak{H})$ be real functions such that Lf and Lg are also in $BC^\infty(\Gamma\backslash\mathfrak{H})$. Then L is symmetric, i.e.*

$$\langle Lf, g\rangle = \langle Lg, f\rangle.$$

Proof. Let a be a large positive number, and F_a a cutoff fundamental domain as shown in Fig. 5. We may assume that f, g are real. Let

$$I = \langle Lf, g\rangle - \langle Lg, f\rangle.$$

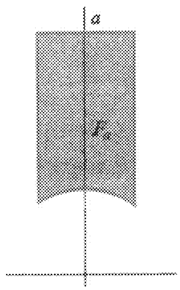

Figure 5

Then by Stokes, we have a truncated integral

$$I_a = \iint_{F_a} [(\Delta f)g - (\Delta g)f]\,dx\,dy = \int_{\partial F_a} \left(g\frac{\partial f}{\partial n} - f\frac{\partial g}{\partial n}\right) ds.$$

But on the pieces which are paired under Γ, the integrals cancel. Hence only the top piece of the boundary integral does not vanish, namely

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\frac{\partial f}{\partial y} g - \frac{\partial g}{\partial y} f\right]_{y=a} dx.$$

Using the boundedness of f, g, whence that of f^2, g^2, and the Schwarz inequality, we get

$$|I_a|^2 \ll \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2\right]_{y=a} dx.$$

Integrate with respect to a between large numbers A and B. You find

$$\int_A^B |I_a|^2\, da \ll \int_A^B \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[\left(\frac{\partial f}{\partial y}\right)^2 + \left(\frac{\partial g}{\partial y}\right)^2\right] dx\, dy.$$

We know by Lemma 1 that this integral is bounded independently of A and B. On the other hand,

$$I_a = \iint_{F_a} (g\,\Delta f - f\,\Delta g)\, dx\, dy$$

approaches a limit as $a \to \infty$ because both Lf, Lg are bounded and F has finite measure under $dx\,dy/y^2$. This limit must be 0, for otherwise there is some $c > 0$ such that for all large a we have

$$|I_a|^2 \geqslant c > 0,$$

whence

$$\int_A^B |I_a|^2\, da \geqslant c(B - A)$$

is not bounded, contradicting our previous estimate. This proves Theorem 3.

§5. THE LAPLACE OPERATOR ON $\Gamma\backslash\mathfrak{H}$

Let D_L be the space of functions $f \in BC^\infty(\Gamma\backslash\mathfrak{H})$ such that Lf is also bounded C^∞. Denote the Hilbert space $L^2(\Gamma\backslash\mathfrak{H})$ by H. This convention remains in force throughout the chapter. Then D_L is a dense subspace of H. We view L as an operator on this dense subspace, and we know that L is symmetric by Theorem 3. We shall see that L can be extended to a self adjoint operator (cf. Appendix 2) by constructing its resolvant, averaging over Γ the resolvant of L on $\mathfrak{H}$ itself. For this we need an estimate.

Lemma 1. *If* $\sigma > 1$, *then the series*

$$\sum_{\gamma\in\Gamma} \frac{1}{[1 + u(z, \gamma z')]^{\sigma}}$$

is convergent uniformly for z, z' *in compact domains.*

Proof. Let M_R be the number of elements $\gamma\in\Gamma$ such that $u(z, \gamma z') \leqslant R$. We contend that $M_R \ll R$. Indeed, let D be a disc of fixed small radius around z', such that $D \cap \gamma D$ is empty if $\gamma z' \neq z'$. Let D_R be the set of points z'' such that $u(z, z'') \leqslant R$. We know that D_R has area $\ll R$ (cf. §1). If $\gamma z' \in D_R$, then $\gamma D \subset D_{2R}$ for R large. Hence the number of translates $\gamma D \subset D_{2R}$ with $\gamma z' \neq z'$ is essentially bounded by

$$\frac{\text{Area } D_{2R}}{\text{Area } D} \ll R.$$

The number of elements $\gamma\in\Gamma$ such that $\gamma z' = z'$ is uniformly bounded (in the case of $SL_2(\mathbf{Z})$, bounded by 6, and is actually 2 for all points except those equivalent to i and $e^{2\pi i/3}$ under $SL_2(\mathbf{Z})$, as is easy to prove). So our contention is proved.

Now we split our sum into partial sums over those γ such that $\gamma z'$ lies in the annulus

$$\frac{R}{2^{n+1}} \leqslant u(z, \gamma z') \leqslant \frac{R}{2^n}.$$

Let $m = [\log_2 R]$. Then our sum is dominated by

$$\frac{R}{R^{1+\epsilon}} + \frac{R/2}{(R/2)^{1+\epsilon}} + \cdots + \frac{R/2^m}{(R/2^m)^{1+\epsilon}}$$

$$\ll \frac{1 + 2^{\epsilon} + 2^{2\epsilon} + \cdots + 2^{m\epsilon}}{R^{\epsilon}} \ll \frac{2^{(m+1)\epsilon}}{R^{\epsilon}} = O(1).$$

The estimates are obviously uniform for z, z' in compact sets. This proves the lemma.

For $\sigma > 1$ let

$$\boxed{r(z, z'; s) = \frac{1}{2} \sum_{\gamma\in\Gamma} \varphi(u(z, \gamma z'); s).}$$

We occasionally write $r_s(z, z')$ instead of $r(z, z'; s)$. We have multiplied the sum by $\frac{1}{2}$ to take the trivial action of ± 1 into account. If $z \notin \Gamma z'$, then we conclude from Lemma 1 that the series converges absolutely for $\sigma > 1$. For $f \in BC^\infty(\Gamma\backslash\mathfrak{H})$ we have

$$\int_F r(z, z'; s) f(z')\, dz' = \frac{1}{2} \int_F \sum_{\gamma \in \Gamma} \varphi(u(z, \gamma z'); s) f(z')\, dz'.$$

By Lemma 1 we can interchange the integral and the sum, to get the above exp

$$= \int_{\mathfrak{H}} \varphi(u(z, z'); s) f(z')\, dz'$$

$$= R_{\mathfrak{H}}(s) f(z).$$

Thus we obtain the old resolvant on $\mathfrak{H}$ itself, and using Theorem 2, we have proved

Theorem 4. *Let $\sigma > 1$. The kernel $r(z, z'; s)$ defines an operator*

$$R(s) \quad BC^\infty(\Gamma\backslash\mathfrak{H}) \to BC^\infty(\Gamma\backslash\mathfrak{H})$$

satisfying

$$(L - s(1 - s)I) R(s) f = f.$$

From Theorem 4, we see that

$$D_L \supset R(s) BC^\infty(\Gamma\backslash\mathfrak{H}),$$

i.e. the domain of L contains the image of $R(s)$, and also that

$$(L - s(1 - s)) D_L$$

is dense in $\mathfrak{H}$. By abstract nonsense concerning unbounded operators (Appendix 2) we get

Theorem 5. *The operator L with domain D_L has a closure, denoted by A, with domain D_A. The operator (A, D_A) is self adjoint.*

Using appropriate estimates, we shall prove in §7:

Theorem 6. *For $\sigma > 3/2$, $R(s)$ is a bounded operator on $H = L^2(\Gamma\backslash\mathfrak{H})$.*

This will involve decomposing the kernel $r(z, z'; s)$ into various components. Thus $R(s)$ is what is usually called the **resolvant** of A.

It follows from Theorem 5 that

$$R(s) H \subset D_A,$$

i.e. the image of $R(s)$ is contained in the domain of A. Indeed, let $f_n \to f$ in H, and $f_n \in BC^\infty(\Gamma\backslash\mathfrak{H})$. Let $\lambda = s(1-s)$. From

$$LR(s)f_n = \lambda R(s)f_n + f_n,$$

it follows that $LR(s)f_n \to \lambda R(s)f + f$. Hence, $R(s)f$ is in the domain of A, and the relation

$$\boxed{(A - s(1-s)I)R(s)f = f}$$

holds for all $f \in H$, $\mathrm{Re}(s) > 3/2$.

§6. GREEN'S FUNCTIONS AND THE WHITTAKER EQUATION

This section recalls some advanced calculus and is an interlude preparing the ground for the decomposition of the resolvant $R(s)$. We deal with special cases of second order linear differential equations, sufficient for the applications we have in mind.

Let (a, b) be an open interval, which may be $(0, \infty)$. Let

$$M_y = -\left(\frac{d}{dy}\right)^2 + p(y)$$

where p is a C^∞ function on (a, b). By a **Green's function** for the differential operator M, we mean a function $g(y, y')$ on $(a, b) \times (a, b)$ such that

$$M_y \int_a^b g(y, y')f(y')\,dy' = f(y)$$

for all $f \in C_c^\infty(a, b)$. In other words, the Green's operator inverts M on the right. For this section, we assume in addition that the Green's function satisfies

GF 0. *The function g is continuous. It is C^∞ in each variable except on the diagonal.*

We shall see in a moment that it is essential that the partial derivatives not be continuous on the diagonal. In fact, suppose that g also satisfies the following additional condition.

GF 1. *If $y \neq y'$, then $M_y g(y, y') = 0$.*

In other words, g satisfies the homogeneous differential equation away from the diagonal.

Let g be a function satisfying **GF 0** *and* **GF 1**. *Then g is a Green's function for the operator M if and only if g also satisfies the jump condition*

GF 2. $$D_1 g(y, y+) - D_1 g(y, y-) = 1.$$

As usual,

$$D_1 g(y, y+) = \lim_{\substack{y' \to y \\ y' > y}} D_1 g(y, y'),$$

and similarly for $y-$ instead of $y+$, we take the limit with $y' < y$.

To prove the above assertion, write

$$\int_a^b = \int_a^y + \int_y^b .$$

Take d/dy and use the continuity of g. We obtain

$$g(y, y)f(y) + \int_a^y D_1 g(y, y')f(y')\,dy'$$

$$- g(y, y)f(y) + \int_y^b D_1 g(y, y')f(y')\,dy'.$$

So the term $g(y, y)f(y)$ cancels. Taking $(d/dy)^2$ yields

$$D_1 g(y, y-)f(y) + \int_a^y D_1^2 g(y, y')f(y')\,dy'$$

$$- D_1 g(y, y+)f(y) + \int_y^b D_1^2 g(y, y')f(y')\,dy'.$$

Taking $-(d/dy)^2 + p(y)$ yields

$$M_y \int_a^b g(y, y')f(y')\,dy'$$

$$= [D_1 g(y, y+) - D_1 g(y, y-)]f(y) + \int_a^b M_y g(y, y')f(y')\,dy'.$$

By **GF 1** the second term on the right vanishes. Hence **GF 2** is equivalent to g being a Green's function, as was to be shown.

Let J, K be two linearly independent solutions of the homogeneous equation, i.e.

$$J'' + pJ = 0$$

$$K'' + pK = 0.$$

There exists a unique Green's function of the form

$$g(y, y') = \begin{cases} A(y')J(y) & \text{if } \; y' < y \\ B(y')K(y) & \text{if } \; y' > y, \end{cases}$$

and the functions $A(y')$, $B(y')$ necessarily have the values computed below.

To prove this, we note that for $g(y, y')$ as given, condition **GF 1** is satisfied by definition. The continuity and **GF 2** amount to the linear equations

$$A(y)J(y) - B(y)K(y) = 0,$$

$$-A(y)J'(y) + B(y)K'(y) = 1.$$

Let $W = JK' - J'K$ (called the **Wronskian**). Note that $W' = 0$ (immediate from the differential equation), and therefore W is constant, $\neq 0$ because J, K are linearly independent. Hence

$$A = K/W \qquad \text{and} \qquad B = J/W.$$

Therefore we get the Green's function necessarily to be

$$\boxed{g(y, y') = \begin{cases} \dfrac{K(y')J(y)}{W} & \text{if } \; y' < y, \\[2ex] \dfrac{J(y')K(y)}{W} & \text{if } \; y' > y. \end{cases}}$$

The next two examples are those used in the applications.

Example 1. On $(0, \infty)$ let

$$M_y = -\left(\frac{d}{dy}\right)^2 - \frac{s(1-s)}{y^2}.$$

The associated homogeneous equation is

$$\psi''(y) = -\frac{s(1-s)}{y^2}\psi(y).$$

For $s \neq \frac{1}{2}$ we take the two linearly independent solutions

$$y^{1-s} \quad \text{and} \quad y^s.$$

Their Wronskian is $2s - 1$, and therefore a Green's function is given by

$$t(y, y'; s) = \frac{1}{2s-1}\begin{cases} y'^s y^{1-s} & \text{if } y' < y, \\ y'^{1-s} y^s & \text{if } y' > y. \end{cases}$$

This notation will remain in force throughout the rest of the chapter.

Example 2. (Whittaker's equation) Let $s \in \mathbf{C}$, $\mathrm{Re}(s) > 0$. Let c be a real number $\geqslant 1$. Let

$$M_y = -\left(\frac{d}{dy}\right)^2 + \left(c^2 - \frac{s(1-s)}{y^2}\right).$$

The homogeneous equation for M_y is

$$\psi''(y) = \left(c^2 - \frac{s(1-s)}{y^2}\right)\psi(y),$$

and for large y can be viewed as a perturbation of the simpler equation $\psi'' = c^2\psi$, having as independent solutions e^{cy} and e^{-cy}. We shall get the existence of corresponding perturbations as solutions of the Whittaker equation, having analogous asymptotic properties.

We shall prove that there exist two solutions

$$J = J_{s,c} \quad \text{and} \quad K = K_{s,c}$$

having the following asymptotic behavior, uniformly under the stated conditions.

For $y \to \infty$, $c \geqslant 1$, s in a compact set, $\mathrm{Re}(s) > 0$, except for $J'(y)$ where $\mathrm{Re}(s) > 1$.

$$J(y) \sim e^{-cy}, \qquad J'(y) \sim -ce^{-cy},$$
$$K(y) \sim e^{cy}, \qquad K'(y) \sim ce^{cy}.$$

For $y \to 0$, given c, uniformly for s in a compact set.

$$J_{s,c}(y) \sim \frac{\Gamma(2s-1)}{\Gamma(s)}(2cy)^{1-s}, \qquad \text{for} \quad \mathrm{Re}(s) > \tfrac{1}{2},$$

$$K_{s,c}(y) \sim \frac{\beta(s)}{\Gamma(s)}(2cy)^{s}, \qquad \text{for} \quad \mathrm{Re}(s) > 0.$$

where

$$\beta(s) = \int_0^1 [t(1-t)]^{s-1}\, dt.$$

From the asymptotic behavior for $y \to \infty$, we conclude that the Wronskian of J, K is asymptotic to $2c$, and hence equal to $2c$, since it is constant. Therefore there exists a Green's function given by the formulas

$$\boxed{g_{s,c}(y, y') = \begin{cases} \dfrac{K(y')J(y)}{2c} & \text{if} \quad y' < y, \\[2ex] \dfrac{J(y')K(y)}{2c} & \text{if} \quad y' > y. \end{cases}}$$

This is the unique Green's function satisfying **GF 0** and **GF 1**, and having the form $A(y')J(y)$ if $y' < y$ and $B(y')K(y)$ if $y' > y$. By the uniform estimates stated above, this function satisfies

$$\boxed{|g_{s,c}(y, y')| \leqslant C_1 \frac{e^{-c|y-y'|}}{c}}$$

where C_1 is a constant, uniform for s in a compact set, $\mathrm{Re}(s) > 0$, and $0 < a \leqslant y, y' < \infty$.

Proofs. There remains to give the proof of the existence of $J_{s,c}$ and $K_{s,c}$ having the desired properties. This is done by recalling some classical results.

Let

$$W_s(y) = \frac{1}{\Gamma(s)} y^s e^{-y/2} \int_0^\infty e^{-ty}[t(1+t)]^{s-1}\, dt,$$

which, after a change of variables, is also expressible as

$$W_s(y) = \frac{1}{\Gamma(s)} e^{-y/2} \int_0^\infty e^{-t} t^{s-1} \left(1 + \frac{t}{y}\right)^{s-1} dt.$$

Then W_s satisfies the Whittaker equation

$$\psi''(y) = \left(\frac{1}{4} - \frac{s(1-s)}{y^2}\right)\psi(y).$$

Indeed, if we apply the Whittaker differential operator

$$\left(\frac{d}{dy}\right)^2 - \left(\frac{1}{4} - \frac{s(1-s)}{y^2}\right)$$

to

$$y^s e^{-y/2} e^{-ty} [t(1+t)]^{s-1},$$

we turn it into the exact expression (with respect to t)

$$-\frac{d}{dt}\left\{y^{s-1} e^{-y/2} e^{-ty} [t(1+t)]^s\right\},$$

which has boundary values equal to 0 at 0 and ∞. Differentiating under the integral sign shows that W_s satisfies the Whittaker equation.

Let

$$\begin{aligned} J_{s,c}(y) &= W_s(2cy) \\ &= \frac{1}{\Gamma(s)} e^{-cy} \int_0^\infty e^{-t} t^s \left(1 + \frac{t}{2cy}\right)^{s-1} \frac{dt}{t}. \end{aligned}$$

It is clear by the dominated convergence theorem that the integral on the right tends to the Gamma integral

$$\int_0^\infty e^{-t} t^s \frac{dt}{t}$$

uniformly for $c \geqslant 1$ and s in a compact set, $\operatorname{Re}(s) > 0$. Hence

$$J_{s,c}(y) \sim e^{-cy}$$

uniformly under these conditions. Differentiating under the integral sign gives

$$J'_{s,c}(y) = -cJ_{s,c}(y) + \frac{1}{\Gamma(s)} e^{-cy} \int_0^\infty e^{-t} t^s (s-1)\left(1 + \frac{t}{2cy}\right)^{s-2} \left(\frac{-t}{2cy^2}\right) \frac{dt}{t}.$$

Again we get the desired uniform asymptotic behavior of $J'_{s,c}(y)$ for $y \to \infty$.

On the other hand, we also have

$$J_{s,c}(y) = \frac{1}{\Gamma(s)} e^{-cy}(2cy)^{1-s} \int_0^\infty e^{-t} t^s (2cy + t)^{s-1} \frac{dt}{t}.$$

Fix c. As $y \to 0$, we see that the integral on the right approaches $\Gamma(2s-1)$ uniformly for s in a compact set. Hence for fixed c, we get

$$J_{s,c}(y) \sim \frac{\Gamma(2s-1)}{\Gamma(s)} (2cy)^{1-s}.$$

This concludes our analysis of the first solution.

The second solution is handled by similar means as follows. Let

$$V_s(y) = \frac{1}{\Gamma(s)} y^s e^{-y/2} \int_0^1 e^{yt}[t(1-t)]^{s-1}\, dt.$$

The same technique as before with the resolving form shows that V_s is a solution of the Whittaker equation, namely applying the Whittaker operator to

$$y^s e^{-y/2} \int_0^1 e^{yt}[t(1-t)]^{s-1}\, dt$$

turns it into the exact form

$$-\frac{d}{dt}\left\{y^{s-1} e^{-y/2} e^{yt}[t(1-t)]^s\right\},$$

which has boundary values equal to 0.

For $y \to 0$ we have immediately

$$V_s(y) \sim \frac{\beta(s)}{\Gamma(s)} y^s \qquad \text{where} \quad \beta(s) = \int_0^1 [t(1-t)]^{s-1}\, dt.$$

Let

$$K_{s,c}(y) = V_s(2cy) = \frac{1}{\Gamma(s)} (2cy)^s e^{-cy} \int_0^1 e^{2cyt}[t(1-t)]^{s-1}\, dt.$$

Then $K_{s,c}$ has the desired behavior for $y \to 0$. Make the change of variables $t \mapsto 1 - t$ and then $u = 2cyt$, $du = 2cy\,dt$. We find

$$K_{s,c}(y) = \frac{1}{\Gamma(s)} e^{cy} \int_0^{2cy} e^{-u}\left[u\left(1 - \frac{u}{2cy}\right)\right]^{s-1} du.$$

This last integral converges to $\Gamma(s)$ uniformly in the desired region for $y \to \infty$. Hence

$$K_{s,c}(y) \sim e^{cy}$$

for $y \to \infty$, uniformly as stated. On the other hand, differentiating $V_s(2cy)$ with respect to y yields

$$K'_{s,c}(y) = \frac{s}{y} K_{s,c}(y) - cK_{s,c}(y)$$

$$+ \frac{1}{\Gamma(s)} (2cy)^s e^{-cy} \int_0^1 e^{2cyt}[t(1 - t)]^{s-1} 2ct\,dt.$$

Again let $t \mapsto 1 - t$, and then $u = 2cyt$. The third term on the right becomes

$$\frac{2c}{\Gamma(s)} e^{cy} \int_0^{2cy} e^{-u}\left(1 - \frac{u}{2cy}\right)^s u^{s-1}\,du,$$

which is uniformly asymptotic to $2ce^{cy}$. Since $-cK_{s,c}(y) \sim -ce^{cy}$, we obtain

$$K'_{s,c}(y) \sim ce^{cy}$$

uniformly, as desired.

§7. DECOMPOSITION OF THE RESOLVANT ON $\Gamma\backslash\mathfrak{H}$ for $\sigma > 3/2$

In this section, we study the resolvant for $\sigma > 3/2$ or even $\sigma > 3$. The sum over all $\gamma \in \Gamma$ will be split into two sums. Let Γ_0 be the group of matrices

$$\begin{pmatrix} \pm 1 & n \\ 0 & \pm 1 \end{pmatrix}$$

with $n \in \mathbf{Z}$. We write

$$r(z, z'; s) = \frac{1}{2} \sum_{\gamma \in \Gamma_0} \varphi(u(z, \gamma z'; s)) + \frac{1}{2} \sum_{\gamma \notin \Gamma_0} \varphi(u(z, \gamma z'); s)$$

[The $\frac{1}{2}$ is to cancel the trivial effect of ± 1.] The first sum will be called the **cuspidal part of the resolvant**, and the second sum will be called the **non-cuspidal part**. We study first the cuspidal. We shall see that we can also decompose it into a sum of several kernels exhibiting various boundedness conditions. In particular, we shall introduce other function spaces besides $L^2(\Gamma\backslash\mathfrak{H})$ on which the effect of these kernels will be more transparent than on the Hilbert space itself.

We break up the fundamental domain into two pieces. Let a be a large positive number. Let F_0 be the part of the fundamental domain with $y \leqslant a$ and let F_1 be the part with $y \geqslant a$, as illustrated in Fig. 6.

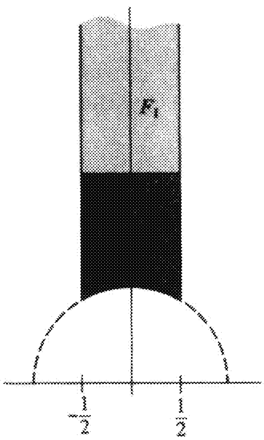

Figure 6

This means that a function f on F has two components,

$$f = (f_0, f_1)$$

which are also written vertically,

$$f = \begin{pmatrix} f_0 \\ f_1 \end{pmatrix},$$

to allow for the operation of a matrix on the left. If, for instance, f_0, f_1 are viewed as elements of certain function spaces, then any operator Q will be written as a 2×2 matrix of operators

$$Qf = \begin{pmatrix} Q_{00} & Q_{01} \\ Q_{10} & Q_{11} \end{pmatrix} \begin{pmatrix} f_0 \\ f_1 \end{pmatrix}.$$

The spaces we consider are the following. First, naturally,

$$H = L^2(\Gamma\backslash\mathfrak{H}) = L^2(F_0) \oplus L^2(F_1).$$

(The Hilbert scalar product is taken with the measure $dz = dx\,dy/y^2$.)

Second, for any real number μ, we have the space

$$\mathfrak{B}_\mu(F) = \mathfrak{B}_\mu(F_0) \oplus \mathfrak{B}_\mu(F_1)$$

where $\mathfrak{B}_\mu(F_0)$ is simply the space of continuous functions on F_0, with the sup norm, and $\mathfrak{B}_\mu(F_1)$ is the space of continuous functions f on F_1 having the property that

$$|f(x+iy)| \ll y^\mu,$$

with the μ-norm,

$$\|f\|_\mu = \sup_{z \in F_1} \frac{|f(z)|}{y^\mu}.$$

Thus H is a Hilbert space, while $\mathfrak{B}_\mu(F)$ is a Banach space. Observe that

$$\mathfrak{B}_0 \subset H,$$

i.e. any bounded function is in $\mathfrak{L}^2(\Gamma\backslash\mathfrak{H})$ because $\Gamma\backslash\mathfrak{H}$ has finite measure. This is especially true of $\mathfrak{B}_{-1} \subset \mathfrak{B}_0$.

Let $k(z, z')$ be a function defined on the product $F_i \times F_j$. We usually use the notation $k_{ij}(z, z')$ to denote this property, when the variables z, z' range over $z \in F_i$ and $z' \in F_j$. A kernel will be said to be of **type** $\mathfrak{B}_\mu$ if it is continuous, and if

$$|k(z, z')| \ll (yy')^\mu.$$

If, for instance, $k = k_{01}$, then the variable z (hence y) ranges over a compact set, and consequently the inequality has bearing only for the second variable, so that in this case it is equivalent to

$$|k_{01}(z, z')| \ll y'^\mu.$$

We say that an operator is of **type** $\mathfrak{B}_\mu$ if it has a kernel of this type. Such operators have obvious continuity properties applied to spaces $\mathfrak{B}_\nu$. For instance

Lemma 0. *Let $k(z, z')$ be a kernel of type $\mathfrak{B}_{-\mu}$. Then the associated operator K is defined on $\mathfrak{B}_{1+\mu-\epsilon}$ and maps it continuously into $\mathfrak{B}_{-\mu}$.*

The proof is trivial, by freshman integration applied to the integral, say on F_1:

$$\int_a^\infty (yy')^{-\mu} y'^{1+\mu-\epsilon} \frac{dy'}{y'^2} \ll y^{-\mu}.$$

We shall also need to know that certain operators between $\mathfrak{B}_\mu$ spaces are compact. For this, we need some additional remarks along the lines of Ascoli's theorem, whose statement we recall for the convenience of the reader.

Ascoli's theorem. *Let X be a compact space and Φ a family of continuous functions on X. Then Φ is relatively compact (compact closure) in the space of continuous functions on X, with sup norm, if and only if Φ is equicontinuous and bounded.*

(For the proof, cf. *Real Analysis*.) By **equicontinuous**, one means that given $x_0 \in X$ and ϵ there exists δ such that if $|x - x_0| < \delta$ then $|f(x) - f(x_0)| < \epsilon$ for all $f \in \Phi$.

We shall combine Ascoli's theorem with the following statement to get compactness.

Let $\nu < \mu$. A set E of equicontinuous functions on F_1, bounded in $\mathfrak{B}_\nu$, is relatively compact in $\mathfrak{B}_\mu$.

Proof. It suffices to prove that E is totally bounded in $\mathfrak{B}_\mu$, i.e. can be covered by a finite number of balls of given radius $r > 0$. By assumption, there exists $C > 0$ such that for any $f \in E$,

$$|f(y)| \leqslant Cy^\nu \qquad \text{that is} \qquad |f(y)|y^{-\nu} \leqslant C.$$

Pick Y so large that if $y \geqslant Y$, then $|f(y)|y^{-\mu} < \epsilon$ for all $f \in E$. Write

$$f = (f^+, f^-)$$

where f^- is the restriction of f to the domain $y \leqslant Y$, and f^+ is its restriction to the domain $y \geqslant Y$. We have an inclusion

$$\mathfrak{B}_\nu(F_1) \subset \mathfrak{B}_\nu(F_Y^-) \times \mathfrak{B}_\nu(F_Y^+).$$

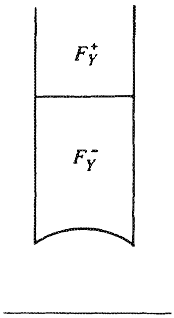

Figure 7

The function f^+ lies in a ball of radius ϵ in $\mathfrak{B}_\mu(F_Y^+)$. On the other hand, the family of functions $\{f^-\}$, for $f \in E$, is equicontinuous and bounded on F_Y^-. Hence it is relatively compact in $\mathfrak{B}_\mu(F_Y^-)$ by the ordinary Ascoli theorem, since F_Y^- is compact. This proves our assertion.

In the applications, we obtain a set of equicontinuous functions by means of a kernel $k(y, y')$, say on $F_1 \times F_1$ applied to the unit ball in $\mathfrak{B}_\nu$. Let K be the associated operator. Let f be a function in the unit ball in $\mathfrak{B}_\nu$. Then at a point z_0 we have

$$|Kf(z) - Kf(z_0)| \ll \int_a^\infty |k(y, y') - k(y_0, y')| y'^\nu \frac{dy'}{y'^2},$$

and we see that the uniform $\mathfrak{B}_\nu$-bound rids us of the function f in the estimate, transferring any needed estimates to the kernel. Thus we shall obtain equicontinuity with any kernel for which we can take the limit under the integral sign, as $y \longrightarrow y_0$. In practice, the needed estimates will be obvious if k is a kernel of type $\mathfrak{B}_\mu$ for some μ, but we shall need to expand somewhat more work for other types of kernel which arise in the cuspidal part.

The cuspidal part

We look at the sum

$$\begin{aligned} r^0(z, z'; s) &= \frac{1}{2} \sum_{\gamma\in\Gamma_0} \varphi(u(z, \gamma z'); s) \\ &= \sum_{-\infty}^{\infty} \varphi(u(z, z' + n); s) \\ &= \sum_{-\infty}^{\infty} \varphi\left(\frac{|z - z' + n|^2}{4yy'} ; s\right). \end{aligned}$$

Then r^0 is periodic in x, and is an even function of $x - x'$. Hence it has a Fourier expansion of the form

$$(1) \qquad r^0(z, z'; s) = m_0(y, y'; s) + 2 \sum_{k=1}^{\infty} m_k(y, y'; s)\cos 2\pi k(x - x')$$

where the Fourier coefficients for $k \geqslant 0$ are given by

$$(2) \qquad m_k(y, y'; s) = \int_{-\frac{1}{2}}^{\frac{1}{2}} r^0(z, z'; s)\cos 2\pi k(x - x')\, dx.$$

It is clear that $m_k(y, y'; s)$ is symmetric in (y, y'), and the complex conjugate is given by

$$\overline{m_k(y, y'; s)} = m_k(y, y'; \bar{s}).$$

Also, the integral expression shows that m_k is continuous, even on the diagonal.

We shall determine explicitly m_0, and show that it is equal to the kernel t of Example 1 in our discussion of Green's functions. We shall also see that m_k for $k \geqslant 1$ is the Green's function of Example 2. Finally we shall give estimates which describe continuity properties for these kernels on various spaces.

The kernel $r^0(z, z'; s)$ gives rise to an operator $R^0(s)$ defined on $BC^\infty(\Gamma_0\backslash\mathfrak{H})$, and the same argument that showed

$$(L - s(1 - s)I)R(s)f = f$$

for $f \in BC^\infty(\Gamma\backslash\mathfrak{H})$ also proves the formula

$$\boxed{(L - s(1-s)I)R^0(s)f = f}$$

for $f \in BC^\infty(\Gamma_0\backslash\mathfrak{H})$. The fundamental domain for $\Gamma_0\backslash\mathfrak{H}$ is the strip S as shown on Fig. 8.

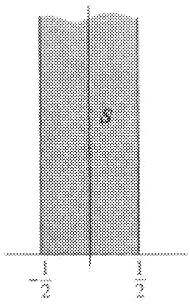

Figure 8

Lemma 1. *For $k \geqslant 0$ the function $m_k(y, y'; s)$ is a Green's function for the differential operator*

$$l_{k,y} = -\left(\frac{d}{dy}\right)^2 + (2\pi k)^2 - \frac{s(1-s)}{y^2}$$

on $(0, \infty)$. Furthermore, m_k is continuous, and satisfies the homogeneous differential equation away from the diagonal. Hence m_k satisfies **GF 0, GF 1, GF 2.**

Proof. Abbreviate

$$M_s = L - s(1-s)I.$$

Let $f, g \in C_c^\infty(\Gamma_0\backslash\mathfrak{H})$ be test functions. The inversion formula for $R^0(s)$ and the symmetry of M_s show that

$$\int_S \int_S r^0(z, z'' | s) f(z'') M_s g(z)\, dz''\, dz = \int_S f(z'') g(z'')\, dz''.$$

Inverting the order of integration shows that

$$\int_S r^0(z, z''; s) M_s g(z)\, dz = g(z'').$$

Fix z' and take for g a function of the form

$$g(z) = \cos 2\pi k(x - x')h(y),$$

where $h \in C_c^\infty(\mathbf{R}^+)$. Then

$$M_s g(z) = -y^2\left[-(2\pi k)^2 \cos 2\pi k(x - x')h(y) + \cos 2\pi k(x - x')h''(y)\right]$$
$$-s(1 - s)\cos 2\pi k(x - x')h(y).$$

Since $dz = dx\, dy/y^2$ and

$$\int_S dz = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_0^\infty \frac{dx\, dy}{y^2},$$

integrating first with respect to x, and noting that $g(z') = h(y')$, we obtain

$$(*) \qquad \int_0^\infty m_k(y, y')\ell_k h(y)\, dy = h(y').$$

First, if we fix y' and pick test functions h such that the support of h does not contain y', then (*) reads

$$\int_0^\infty m_k(y, y')\ell_k h(y)\, dy = 0.$$

By the regularity theorem for elliptic operators, we conclude that $m_k(y, y')$ is C^∞ away from the diagonal, and satisfies the homogeneous differential equation $\ell_{k,y} m_k(y, y') = 0$.

Second, integrating (*) against any test function $\psi \in C_c^\infty(\mathbf{R}^+)$ yields

$$\int_0^\infty \int_0^\infty m_k(y, y')\ell_k h(y)\psi(y')\, dy\, dy' = \int_0^\infty h(y)\psi(y)\, dy.$$

Changing the order of integration, the left-hand side is equal to

$$\int_0^\infty \left[\ell_{k,y} \int_0^\infty m_k(y, y')\psi(y')\, dy'\right] h(y)\, dy.$$

It follows that

$$\ell_{k,y} \int_0^\infty m_k(y, y')\psi(y')\, dy' = \psi(y).$$

Hence m_k is a Green's function as desired.

Lemma 2. *For* $\mathrm{Re}(s) > 1$ *we have*

$$m_0(y, y'; s) = t(y, y'; s) = \frac{1}{2s-1} \begin{cases} y'^s y^{1-s} & \text{if } y' < y, \\ y'^{1-s} y^s & \text{if } y' > y. \end{cases}$$

Furthermore for $k > 0$, $m_k(y, y'; s) = g_{s,2\pi k}(y, y')$, *where* $g_{s,c}$ *is the Green's function of* §6, *Example* 2.

Proof. We have

$$m_0(y, y'; s) = \int_{-\frac{1}{2}}^{\frac{1}{2}} r^0(z, z'; s)\, dx.$$

Fix y'. Let $y > y'$. Since m_0 satisfies the homogeneous differential equation away from the diagonal, we must have

$$m_0(y, y'1\, s) = a(y')y^{1-s} + b(y')y^s.$$

As $y \longrightarrow \infty$, $r^0(z, z'; s) \longrightarrow 0$ because $\varphi(u) \sim u^{-\sigma}$ for $u \longrightarrow \infty$. Hence $b(y') = 0$. Now let $y < y'$. Then

$$m_0(y, y'; s) = c(y')y^{1-s} + d(y')y^s.$$

As $y \longrightarrow 0$, $u(z, z') \longrightarrow \infty$, so again $\varphi(u(z, z'); s) \longrightarrow 0$. Hence $c(y') = 0$. It follows that

$$m_0(y, y'1\, s) = \begin{cases} a(y')y^{1-s} & \text{if } y' < y, \\ d(y')y^s & \text{if } y' > y. \end{cases}$$

By the general theory of Green's functions, we conclude that $m_0 = t$. Exactly the same arguments show that $m_k = g_{s,2\pi k}$. Note that we needed to know the asymptotic behavior of the solutions of the homogeneous equation both for $y \longrightarrow \infty$ and $y \longrightarrow 0$, in order to have the unique determination of m_k as a specific Green's function.

From the general estimates of the Green's function $g_{s,c}$ we obtain uniform estimates for $m_k(y, y'; s)$ $(k \geqslant 1)$, stated below.

We let for $\sigma > 1$,

$$\boxed{m(z, z'; s) = 2\sum_{k=1}^{\infty} m_k(y, y'; s)\cos 2\pi k(x - x').}$$

Lemma 3. *For $\sigma > 1$, we have an estimate*

$$|m_k(y, y'; s)| \leqslant C\frac{e^{-2\pi k|y-y'|}}{k}$$

where C is a constant, uniform for s in a compact set, and

$$0 < a \leqslant y, y' < \infty.$$

Therefore $m(z, z'; s)$ lies in $\mathfrak{L}^2(F_1 \times F_1)$, and the corresponding operator

$$\mathfrak{L}^2(F_1) \rightarrow \mathfrak{L}^2(F_1)$$

is compact.

We denote by $M(s)$ the operator whose only non-zero component is $M_{11}(s)$, given by the kernel $m(z, z'; s)$ above. Thus

$$m_{11}(z, z'; s) = m(z, z'; s)$$

$$m_{01} = m_{10} = m_{00} = 0.$$

We have obtained the decomposition

$$\boxed{r^0(z, z'; s) = t(y, y'; s) + m(z, z'; s).} \tag{3}$$

The next lemmas are concerned with further continuity properties of the operators associated with these kernels on various spaces.

In considering operators arising from the cuspidal part, we agree that they have only a 11-component. This holds in particular for $T(s)$ and $M(s)$. Thus the kernel matrix $t_{ij}(s)$ for $T(s)$ has components $t_{ij}(z, z'1\ s) = 0$ if $i \neq j$ and if $i = j = 0$, while

$$t_{11}(z, z'; s) = t(y, y'; s).$$

Similarly for $M(s)$.

Lemma 4. *Assume $\sigma > \frac{1}{2}$ and let $1 - \sigma < \mu < \sigma$. Then for $y \geqslant a$,*

$$\int_a^\infty |t(y, y';\ s)| y'^{\mu} \frac{dy'}{y'^2} \ll y^{\mu}.$$

Thus the associated operator

$$T(s)\colon \mathfrak{B}_\mu(F_1) \longrightarrow \mathfrak{B}_\mu(F_1)$$

is a bounded operator.

Proof. Write the integral on the left as

$$\int_a^\infty = \int_a^y + \int_y^\infty .$$

It is estimated by

$$\int_a^y y^{1-\sigma} y'^{\sigma} y'^{\mu-2}\, dy' + \int_y^\infty y^{\sigma} y'^{1-\sigma} y'^{\mu-2}\, dy'$$

which is $\ll y^{\mu}$ by freshman integration, as was to be proved.

Lemma 5. *The operator $T(s)$ having kernel $t(y, y';\ s)$ is bounded on $L^2(F_1)$ for $\sigma > \frac{1}{2}$.*

Proof. Select μ as in Lemma 4. Without loss of generality, we may take a function f in $\mathfrak{L}^2([a, \infty))$ and see what $T(s)$ does to f. Apply the Schwarz inequality to the functions

$$|t(y, y') y'^{\mu}|^{1/2} \qquad \text{and} \qquad |t(y, y') y'^{-\mu}|^{1/2} |f(y')|.$$

We get

$$|Tf(y)|^2 \leqslant \int_a^\infty |t(y, y')| y'^{\mu} \frac{dy'}{y'^2} \cdot \int_a^\infty |t(y, y')|\, y'^{-\mu} |f(y')|^2 \frac{dy'}{y'^2} .$$

We use Lemma 4, multiply the first integral on the right by $y^{-\mu}$ and the second integral by y^{μ}, and obtain the estimate

$$\ll \int_a^\infty y^{\mu} |t(y, y')|\, y'^{-\mu}\, |f(y')|^2 \frac{dy'}{y'^2} .$$

Now integrate to get

$$\|Tf\|_2^2 \ll \int_a^\infty \int_a^\infty y^{\mu} |t(y, y')| y'^{-\mu} |f(y')|^2 \frac{dy'}{y'^2} \frac{dy}{y^2}$$
$$\ll \|f\|_2^2 \quad \text{(again by Lemma 4).}$$

This concludes the proof.

Lemma 6. *Let $c \geqslant 1$. We have for any real number μ:*

$$\int_a^\infty e^{-c|y-y'|}y'^{\mu}\,dy' \leqslant C_1\frac{1}{c}y^{\mu},$$

where C_1 is a constant independent of c and y.

Proof. Integrate by parts, and split the integral from a to y and y to ∞. For instance, the integral from a to y yields

$$\frac{y'^{\mu}e^{-c(y-y')}}{-c}\Bigg|_a^y + \frac{1}{c}\int_a^y e^{-c(y-y')}\mu y'^{\mu-1}\,dy'.$$

The first term is $\ll \dfrac{1}{c}y^{\mu}$. For the second term, split it again, over the intervals $a \leqslant y' \leqslant 2p$ and $2p \leqslant y' \leqslant y$. The first of these behaves like e^{-cy}/c, and the second is smaller than the original integral divided by 2. Transferring this second integral to the left-hand side concludes the proof.

Lemma 7. *For $\sigma > 1$, the operator $M(s)$ whose kernel is $m(z, z'; s)$ maps $L^2(F_1)$ continuously into $\mathfrak{B}_{-1}(F_1)$.*

Proof. The assertion reduces to the following estimates, for a function $f \in \mathfrak{L}^2([a, \infty))$.

$$\sum_{k=1}^{\infty}\int_a^\infty \frac{e^{-2\pi k|y-y'|}}{k}|f(y')|\frac{dy'}{y'^2}$$

$$\ll \sum_{k=1}^{\infty}\frac{\|f\|_2}{k}\left[\int_a^\infty e^{-4\pi k|y-y'|}\frac{dy'}{y'^2}\right]^{1/2} \qquad \text{(by Schwarz)}$$

$$\ll \sum_{k=1}^{\infty}\frac{\|f\|_2}{k}\frac{1}{k^{1/2}y} \qquad \text{(by Lemma 6)}$$

$$\ll \|f\|_2\frac{1}{y}.$$

This proves the lemma.

Lemma 8. *For $\sigma > 1$ and any real number μ, the operator $M(s)$ whose* *kernel is $m(z, z'; s)$ gives a bounded linear map*

$$M(s): \mathfrak{B}_\mu(F_1) \longrightarrow \mathfrak{B}_{\mu-2}(F_1),$$

and the induced linear map of $\mathfrak{B}_{-1}$ into $\mathfrak{B}_{-1}$ is compact.

Proof. Again we may consider a function $f \in \mathfrak{B}_\mu([a, \infty))$, and for the first assertion, we estimate the sum

$$\sum_{k=1}^{\infty} \int_a^{\infty} m_k(y, y'; s) y'^{\mu-2}\, dy',$$

using

$$|m_k(y, y'; s)| \ll \frac{e^{-2\pi k|y-y'|}}{2\pi k}.$$

The fact that $M(s)$ maps $\mathfrak{B}_\mu$ into $\mathfrak{B}_{\mu-2}$ continuously then results from Lemma 6. For the compactness statement, we follow the pattern already mentioned at the beginning of the section, and prove:

The image under $M(s)$ of the unit ball in $\mathfrak{B}_\nu$ $(\nu < 1)$ is equicontinuous on F_1.

Proof. A function in the unit ball in $\mathfrak{B}_\nu$ is bounded. Fix $z_0 \in F_1$. We get the estimates:

$$|M(s)f(z) - M(s)f(z_0)| \ll \int_a^{\infty} \sum_{k=1}^{\infty} |m_k(y, y'; s) - m_k(y_0, y'; s)| y'^{-2}\, dy'.$$

We can take the limit under the integral sign to get the equicontinuity. Observe that the effect of the kernel is to make the function f disappear from under the integral sign.

The non-cuspidal part

We had our original kernel $r(z, z'; s)$ for $\sigma > 1$ as the average of the kernel $r_{\mathfrak{H}}(z, z'; s)$ over Γ. We let $n(z, z'; s)$ be the kernel whose components are given as follows:

$$n_{00} = r_{00}, \qquad n_{01} = r_{01}, \qquad n_{10} = r_{10},$$

$$n_{11}(z, z'; s) = \frac{1}{2} \sum_{\gamma \notin \Gamma_0} \varphi(u(z, \gamma z'); s),$$

so that we may write

$$N(s) = \begin{pmatrix} R_{00}(s) & R_{01}(s) \\ R_{10}(s) & N_{11}(s) \end{pmatrix}.$$

The corresponding operators are denoted by capital letters.

Lemma 9. *For $\sigma > 3/2$ each kernel $n_{ij}(z, z'; s)$ lies in $\mathfrak{L}^2(F_i \times F_j)$, and $N(s)$ is therefore a compact operator on $L^2(F)$. In fact, for $(i, j) \neq (0, 0)$, for $\sigma > 1$, and sufficiently small ϵ, $r_{ij}(z, z'; s)$ is of type $\mathfrak{B}_{2+\epsilon-\sigma}$, that is:*

$$|r_{01}(z, z'; s)| \ll y'^{2+\epsilon-\sigma},$$

$$|r_{10}(z, z'; s)| \ll y^{2+\epsilon-\sigma},$$

$$|n_{11}(z, z'; s)| \ll (yy')^{2+\epsilon-\sigma}.$$

Finally

$$r_{00}(z, z'; s) = -\frac{1}{4\pi} \log|z - z'| + \text{continuous function of } (z, z').$$

Proof. It is clear that the estimates imply that the kernels are in $\mathfrak{L}^2$. For the estimates, we need another lemma.

Lemma 10. *Let $y_0 > 0$. For $\sigma > 1$, uniformly in $-\frac{1}{2} \leqslant x \leqslant \frac{1}{2}$ and $y, y' \geqslant y_0$, we have*

$$\sum_{\gamma \not\in \Gamma_0} \frac{1}{[1 + u(z, \gamma z')]^\sigma} \ll (yy')^{2+\epsilon-\sigma}.$$

Proof. Write $\gamma \not\in \Gamma_0$ as

$$\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \qquad c \neq 0.$$

Then

$$\begin{aligned} 4yy'u(z, z') &= czz' + dz - az' - b^2 \\ &= (cxx' + dx - ax' - b)^2 + y^2(cx' + d)^2 \\ &\quad + y'^2(cx - a)^2 + c^2y^2y'^2 - 2yy'. \end{aligned}$$

Let $z_0 = x + iy_0$ and $z_0' = x' + iy_0$, as illustrated in Fig. 9. Then $4y_0y_0'$ $u(z_0, \gamma z_0')$ is equal to an expression similar to the above, with y_0, y_0 replacing y, y' respectively. Note that the first term in the expression involves only x, x' and is common to both. We also have obvious inequalities like

$$y^2(cx' + d)^2 \geqslant y_0^2(cx' + d)^2.$$

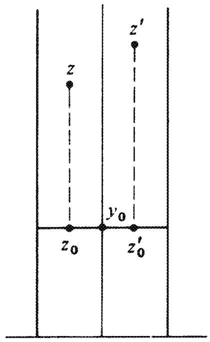

Figure 9

Therefore we obtain the inequality

$$4yy'u(z, \gamma z') - 4y_0y_0u(z_0, \gamma z_0') \geqslant c^2y^2y'^2 - c^2y_0^2y_0^2 - 2yy' + 2y_0y_0$$

whence

$$4yy'[1 + u(z, \gamma z')] \geqslant 4y_0y_0[1 + u(z_0, \gamma z_0')] + \tfrac{15}{16}c^2(yy')^2.$$

Divide by $(yy')^2$ to get (with the obvious abbreviation)

$$4\,\frac{1 + u}{yy'} \geqslant 4\,\frac{y_0y_0}{(yy')^2}\,[1 + u_0].$$

Raise the left-hand side to the σ power, and the right-hand side to the $1 + \epsilon$ power. Up to a constant factor depending on y_0, the inequality is preserved, and we end up with

$$\frac{1}{[1 + u(z, \gamma z')]^\sigma} \ll \frac{1}{[1 + u(z, \gamma z_0)]^{1+\epsilon}}\;\frac{1}{(yy')^{\sigma - 2 - 2\epsilon}}\,.$$

Our proof is concluded by Lemma 1, §5.

The estimates of Lemma 9 are obtained by combining Lemma 10 with estimates like

$$\sum_{\gamma \in \Gamma_0} |\varphi(u(z, \gamma z'); s)| \ll y'^{1-\sigma}$$

for z in a compact domain, arising from the sum

$$\sum_{n=1}^{\infty} \frac{y'^{\sigma}}{(y'^2 + n^2)^{\sigma}} \ll y'^{1-\sigma}.$$

This type of sum, or the analogous integral, had already been considered when we studied the discrete series.

Lemma 11. *For $\sigma > 3$, $N(s)$ maps $\mathfrak{B}_1$ into $\mathfrak{B}_{-1-\delta}$ for some $\delta > 0$, and*

$$N(s)\colon \mathfrak{B}_1 \to \mathfrak{B}_{-1}$$

is compact. Also $N(s)$ maps H into $\mathfrak{B}_{-1}$ continuously.

Proof. This comes from freshman integration applied to operators of type $\mathfrak{B}_{2+\epsilon-\sigma}$. For instance if f is in $\mathfrak{B}_1(F_1)$, then we evaluate the integral

$$\int_a^{\infty} (yy')^{2+\epsilon-\sigma} y' \frac{dy'}{y'^2} \ll y^{2+\epsilon-\sigma}.$$

The compactness of $N(s)$ when viewed as a map from $\mathfrak{B}_1$ into $\mathfrak{B}_{-1}$ is due to the same phenomenon as that already encountered. Applying the operator to the unit ball yields an equicontinuous set of functions. One sees that $N(s)$ maps H into $\mathfrak{B}_{-1}$ by a straightforward use of the Schwarz inequality.

§8. THE EQUATION $-\psi''(y) = \dfrac{s(1-s)}{y^2}\psi(y)$ ON $[a, \infty)$

We need still another operator in order to analyze the above differential equation on the half line, with boundary at a, which we select to be a fixed positive number, say $a \geqslant 1$. We let $\kappa > 3$. Let

$$c(s) = a^{2s-1} \frac{s-\kappa}{s+\kappa-1}$$

and let

$$\theta(y, s) = y^s + c(s)y^{1-s}.$$

Then $\theta(y, s)$ is a solution of the above differential equation, satisfying the boundary condition $\theta(a) = a\kappa^{-1}\theta'(a)$. We also verify directly and trivially from the definitions, that $\theta(y, s)$ satisfies the basic formalism

$$\boxed{\begin{aligned} c(s)c(1-s) &= 1, \\ \theta(y, s) &= \theta(y, 1-s)c(s), \end{aligned}}$$

which we call the **Eisenstein formalism.**

Using the two linearly independent solutions y^{1-s} and $\theta(y, s)$ in connection with the general discussion of Green's functions in §6, and noting that the Wronskian of these functions is $2s - 1$, we get the Green's function

$$q(y, y'; s) = \frac{1}{2s-1}\begin{cases} \theta(y', s)y^{1-s} & \text{if } y' < y, \\ y'^{1-s}\theta(y, s) & \text{if } y < y'. \end{cases}$$

It is clear that $q(y, y'; s)$ is symmetric in y, y' and satisfies the symmetry

$$\overline{q(y, y'; s)} = q(y, y'; \bar{s}).$$

We let $Q(s)$ be the operator whose only component is the "diagonal" component, with kernel

$$q_{11}(z, z'; s) = q(y, y'; s),$$

so that for suitable functions f on F we have

$$Q(s)f(z) = \int_{F_1} q(y, y'; s)f(z')\, dz'.$$

Note that $Q(s)f(z)$ is independent of x.

We shall specify below on what spaces $Q(s)$ acts, and with what continuity properties.

Lemma 1.

i) *For $\sigma > \frac{1}{2}$, $Q(s)$ is a bounded operator on H.*
ii) *For $0 < \sigma < 2$, $Q(s)$ maps $\mathfrak{B}_{-1}(F_1)$ into $\mathfrak{B}_{1-\sigma}(F_1)$ continuously, and*

$$Q(s)\colon \mathfrak{B}_{-1}(F_1) \to \mathfrak{B}_1(F_1)$$

is a compact linear map.

Proof. Except for the compactness assertion of (ii), the continuity assertions of (i) and (ii) are verified by the same elementary integrals as for the corresponding assertions concerning $T(s)$ in Lemmas 4 and 5, §7. It is also clear from these integrals that $Q(s)$ maps $\mathfrak{B}_{-1}$ into $\mathfrak{B}_{1-\sigma}$. The compactness of the operator from $\mathfrak{B}_{-1}$ into $\mathfrak{B}_1$ then follows from the next assertion.

The image by $Q(s)$ of the unit ball in $\mathfrak{B}_{-1}$ is equicontinuous.

Proof. Let $f \in \mathfrak{B}_{-1}$. We have

$$|Q(s)f(y) - Q(s)f(y_0)| \leqslant \int_{F_1} |q(y, y'; s) - q(y_0, y'; s)||f(z')|\, dz'.$$

This is estimated by replacing $|f(z')|$ by $1/y'$, and the expression under the integral sign is then

$$\ll y'^{\sigma} y'^{-3} \leqslant \frac{1}{y'^{1+\epsilon}},$$

integrable on $[a, \infty)$. Our assertion follows at once.

Relations for $q(s)$

Let $k_1(y, y')$ and $k_2(y, y')$ be two kernels. We define their convolution

$$k_1 * k_2(y, y') = \int_a^\infty k_1(y, y'')k_2(y'', y')\, \frac{dy''}{y''^2},$$

and in any specific applications we have to check the absolute convergence of the integral, over a specified range. The convolution corresponds to the composition of the associated operators, K_1K_2 when applied on the left (otherwise, it gets reversed on the right), valid on any spaces where the integrals converge absolutely.

We shall list certain properties of the convolutions of the kernels $q(s)$ and $t(s)$. We let $\kappa > 3$ as always, and

$$\omega(s) = s(1 - s) - \kappa(1 - \kappa).$$

We also abbreviate $T(\kappa)$ by T.

q 1. $q(y, y'; s) - t(y, y'; \kappa) = \omega(s)\displaystyle\int_a^\infty t(y, y''; \kappa)q(y'', y'; s)\, \frac{dy''}{y''^2}$

is valid for $\sigma > 1 - \kappa$.

We could also write this

$$q_s - t_\kappa = \omega(s) t_\kappa * q_s,$$

or as an operator relation

Q 1. $$Q(s) - T = \omega(s) T Q(s).$$

The kernel relation **q 1** is verified by direct integration, and the definitions of $q(s)$, $t(\kappa)$. For instance, say $y' < y$. We write the integral as

$$\int_a^\infty = \int_a^{y'} + \int_{y'}^{y} + \int_y^\infty .$$

We substitute the definitions of $t(y, y''; \kappa)$ and $q(y'', y'; s)$ and evaluate the integrals. The relation falls out.

Let $T = T(\kappa)$. Relation **Q 1** is equivalent with

Q 2. $$(I - \omega(s)T)(I + \omega(s)Q(s)) = I, \qquad \sigma > 1 - \kappa,$$

as one sees by distributivity. Furthermore, since the left-hand side kernel $q(s) - t(\kappa)$ is symmetric, and $q(s)$, $t(\kappa)$ are also symmetric, it follows that we have commutativity in the convolution of kernels

q 3. $$t_\kappa * q_s = q_s * t_\kappa, \qquad \sigma > 1 - \kappa,$$

so that the invertibility relation of **Q 2** also holds, as a kernel relation, on the other side, on whatever spaces the composition is defined. We then have commutativity of operators,

Q 3. $$T(\kappa)Q(s) = Q(s)T(\kappa), \qquad \sigma > 1 - \kappa.$$

***Lemma* 2.** *The operator relations* **Q 1, Q 2, Q 3** *hold*

i) *For* $\sigma > \frac{1}{2}$ *on* H;
ii) *For* $0 < \sigma < 2$ *on* $\mathfrak{B}_{-1}$.

Proof. The integrals involving the operator $Q(s)$ and T converge absolutely in the appropriate domains by Lemmas 0, 4, 5 of §7.

For the next section, we shall also use the relation

q 4. $$q(y, y'; s) - q(y, y'; 1 - s) = \frac{1}{2s - 1}\theta(y, s)\theta(y', 1 - s)$$

valid without restriction, and trivial from the definitions.

The next relations will not be used until §11, *and are inserted here only to have a complete tabulation of the relations involving $q(s)$ together.*

q 5. $q(y, y'; s) - q(y, y'; s')$
$$= [\omega(s) - \omega(s')]\int_a^\infty q(y, y''; s)q(y'', y'; s')\,\frac{dy''}{y''^2}$$

for $\sigma, \sigma' > \frac{1}{2}$.

Observe that

$$\omega(s) - \omega(s') = s(1-s) - s'(1-s').$$

The relation is again verified by direct computations. It is here essential to restrict the domain to $\sigma, \sigma' > \frac{1}{2}$, since otherwise for $s' = 1 - s$, say, we get $\omega(s') = \omega(s)$ and the whole right-hand side would vanish. On the other hand, observe that relation **Q 4** gave us the difference between $Q(s)$ and $Q(1-s)$ in terms of the functions $\theta(y, s)$. Note that **q 5** has the operator formulation

Q 5. $$Q(s) - Q(s') = [\omega(s) - \omega(s')]Q(s)Q(s').$$

Finally we have

q 6. $$\int_a^\infty q(y, y'; s)\theta(y', s')\,\frac{dy'}{y'^2} = \frac{1}{\omega(s') - \omega(s)}\,\theta(y, s')$$

for $\sigma' < \sigma$.

This is also verified by direct computations from the definitions, and has the operator formulation

Q 6. $$Q(s)\theta_{s'} = \frac{1}{\omega(s') - \omega(s)}\,\theta(y, s').$$

In other words, $\theta_{s'}$ is an eigenvector for $Q(s)$, with the stated restriction $\sigma' < \sigma$.

***Lemma* 3.** *Let $M = M(\kappa)$ for $\kappa > 3$. The composite operators*

$$MQ(s) \quad \text{and} \quad Q(s)M$$

are defined on $\mathfrak{B}_{-1}$ and $\mathfrak{B}_1$ respectively, and

$$MQ(s) = O, \qquad Q(s)M = O.$$

Proof. The operator $M(\kappa)$ has a kernel expressed as a Fourier series involving $\cos 2\pi k(x - x')$, which is therefore orthogonal to the kernel of the operator $Q(s)$, which is independent of x.

This concludes our list of relations for $Q(s)$. On an adjoining table, we have summarized some of the continuity properties for our various kernels, and some relations.

$T(s)$, $\sigma > \frac{1}{2}$, only a 11-component

i) Bounded on H.
ii) Bounded on $\mathfrak{B}_\mu$ for $1-\sigma<\mu<\sigma$.

$M(s)$, $\sigma > 1$, only a 11-component

i) Kernel $m(z,z';s)$ in $L^2(F_1 \times F_1)$ so compact on H.
ii) Maps $\mathfrak{B}_\mu \to \mathfrak{B}_{\mu-2}$ continuously, compact for $\mathfrak{B}_{-1} \to \mathfrak{B}_{-1}$.
iii) Maps H into $\mathfrak{B}_{-1}$ continuously.

$N(s)$ of type $2+\epsilon-\sigma$

i) Kernel $n(z,z';s)$ in $\mathfrak{L}^2(F\times F)$, for $\sigma>3/2$, compact on H.
ii) For $\sigma>3$, maps $\mathfrak{B}_1 \longrightarrow \mathfrak{B}_{-1-\delta}$ for some δ, and $N(s): \mathfrak{B}_1 \longrightarrow \mathfrak{B}_{-1}$ is compact.
iii) Maps H into $\mathfrak{B}_{-1}$ continuously for $\sigma>3$.

$V = M(\kappa) + N(\kappa), \kappa > 3$, only a 11-component

i) Compact on H, maps H into $\mathfrak{B}_{-1}$ continuously.
ii) Maps $\mathfrak{B}_1 \longrightarrow \mathfrak{B}_{-1}$, compact for $\mathfrak{B}_{-1} \longrightarrow \mathfrak{B}_{-1}$.

$Q(s)$, only a 11-component

i) For $\sigma>\frac{1}{2}$, bounded on H.
ii) For $0<\sigma<2$, maps $\mathfrak{B}_{-1}$ into $\mathfrak{B}_{1-\sigma}$ continuously, compact for $\mathfrak{B}_{-1} \longrightarrow \mathfrak{B}_1$.

$$T = T(\kappa), \qquad M = M(\kappa), \qquad N = N(\kappa)$$

$$R = R(\kappa) = T + V$$

---

§9. EIGENFUNCTIONS OF THE LAPLACIAN IN $L^2(\Gamma\backslash\mathfrak{H}) = H$

Recall that A is the closure of the Laplacian, with domain D_A in the Hilbert space $H = L^2(\Gamma\backslash\mathfrak{H})$. We are interested in the eigenvectors of A, i.e. the elements $\psi \in D_A$, $\psi \neq 0$, such that $A\psi = \lambda\psi$. Let $\kappa > 3$ as before. Observe that if $\psi \in H$, $\psi \neq 0$ and $R(\kappa)\psi = \alpha\psi$ for some complex number $\alpha \neq 0$, then $\psi \in D_A$ and

$$A\alpha\psi = (1 - \kappa(1 - \kappa)\alpha)\psi,$$

so ψ is a corresponding eigenvector of A itself. We shall analyze the eigenvectors of $R(\kappa)$ in H, and see that they correspond to eigenvectors of a certain compact operator in $\mathfrak{B}_{-1}$.

Let $0 < \sigma < 2$. Let as before $\omega(s) = s(1 - s) - \kappa(1 - \kappa)$ be abbreviated by ω, and let

$$\boxed{K(s) = V + \omega(s)VQ(s) = V(I + \omega Q(s)).}$$

From the sequence of operators

$$\mathfrak{B}_{-1} \xrightarrow{Q(s)} \mathfrak{B}_{1-\sigma} \xrightarrow{V} \mathfrak{B}_{-1}$$

and the knowledge that V is compact, we see that

$$K(s)\colon \mathfrak{B}_{-1} \to \mathfrak{B}_{-1}$$

is a compact operator. It will be used for the analytic continuation of the resolvant to the strip $0 < \sigma < 2$. However, for the moment we are interested in the discrete spectrum, and hence we look only at the region to the right of the line $\sigma = \frac{1}{2}$.

Let $\frac{1}{2} \leqslant \sigma < 2$. Let $\mathfrak{B}_{-1}(\omega(s), K(s))$ be the $\omega(s)^{-1}$-eigenspace of $K(s)$ in $\mathfrak{B}_{-1}$, i.e. the space of functions $f \in \mathfrak{B}_{-1}$ such that

$$\omega(s)K(s)f = f.$$

Similarly, let $H(\omega(s), R) = H(\omega(s), R(\kappa))$ be the $\omega(s)^{-1}$-eigenspace of R in H, i.e. the space of functions $\psi \in H$ such that

$$\omega(s)R(\kappa)\psi = \psi.$$

***Theorem* 7.** *For $\frac{1}{2} \leqslant \sigma < 2$ and $s \neq \frac{1}{2}$, the maps*

$$I + \omega(s)Q(s) \qquad \text{and} \qquad I - \omega(s)T(\kappa)$$

give inverse isomorphisms

$$\mathfrak{B}_{-1}(\omega(s), K(s)) \leftrightarrow H(\omega(s), R(\kappa)).$$

The proof of Theorem 7 involves formal steps, with relations among our various operators, and also involves estimates which we prove as separate lemmas. We begin with the formal steps. We abbreviate $Q(s)$ by Q.

Assume first that $f \in \mathfrak{B}_{-1}$ and $\omega K(s)f = f$, i.e. that

$$\omega V(I + \omega Q)f = f.$$

We have $R = T + V$. Then

$$\begin{aligned}\omega R(I + \omega Q)f &= \omega(T + V)(I + \omega Q)f \\ &= (\omega T + \omega^2 TQ + I)f \\ &= (I + \omega Q)f,\end{aligned}$$

thereby proving half of the theorem.

Conversely, assume that $\omega R\psi = \psi$. Then $V\psi \in \mathfrak{B}_{-1}$, and

$$\omega T\psi + \omega V\psi = \psi,$$

so that

$$(I - \omega T)\psi = \omega V\psi \in \mathfrak{B}_{-1}.$$

Then

$$\begin{aligned}\omega K(s)(I - \omega T)\psi &= \omega V(I + \omega Q(s))(I - \omega T)\psi \\ &= \omega V\psi \\ &= (I - \omega T)\psi.\end{aligned}$$

This proves the converse, and concludes the formal proof of Theorem 7.

The subsequent lemmas make it valid. Observe that for the first part we need to know that if $f \in \mathfrak{B}_{-1}$, then $Q(s)f$ lies in H. This will be proved in Lemmas 1 and 2, when $\operatorname{Re} s = \frac{1}{2}$. It is obvious if $\sigma > \frac{1}{2}$. In the second part of the proof, we used

$$(I + \omega Q)(I - \omega T) = I$$

when applied to ψ. Theorem 8 below justifies this.

Lemma 1. *Let $s = \frac{1}{2} + it$ but $s \neq \frac{1}{2}$. Let $f \in \mathfrak{B}_{-1}$ be an eigenvector for $K(s)$,*

$$\omega(s)K(s)f = f.$$

Then f is orthogonal on F_1 to $\theta(y, s)$, i.e.

$$\int_{F_1} \theta(y, s)f(z)\, dz = 0.$$

Proof. Let $\psi = (I + \omega Q(s))f$. Then by assumption and the definition of $K(s)$, $f = \omega V\psi$. The function $\bar{\psi}(z)f(z)$ is in $\mathfrak{L}^1(F)$, because $f \in \mathfrak{B}_{-1}$ and $\psi \in \mathfrak{B}_1$. Multiply $f(z)$ by $\overline{\psi(z)}$ and integrate. We get

$$\int_F \bar{\psi}(z)f(z)\,dz = \omega(s)\int_F \bar{\psi}(z)V\psi(z)\,dz$$

$$= \omega(s)\int_F \int_F \bar{\psi}(z)v(z, z';\ \kappa)\psi(z')\,dz'\,dz,$$

where the kernel $v(z, z';\ \kappa)$ of V is real and symmetric. It follows at once that the imaginary part of the right-hand side vanishes. Hence

$$\int_F (\bar{\psi}f - \psi\bar{f})\,dz = 0.$$

But $\bar{\psi}f = [\bar{f} + \omega(\bar{s})Q(\bar{s})\bar{f}]f$ and $\psi\bar{f} = [f + \omega(s)Q(s)f]\bar{f}$. Since $\sigma = \frac{1}{2}$, we have $\bar{s} = 1 - s$, and $\omega(s) = \omega(\bar{s})$ is real. Hence

$$0 = \int_{F_1}\int_{F_1} [q(y, y';\ s) - q(y, y';\ 1 - s)]f(z)\bar{f}(z')\,dz\,dz'$$

$$= \int_{F_1}\int_{F_1} \theta(y, s)\overline{\theta(y', s)}f(z)\overline{f(z')}\,dz\,dz' \qquad \text{(by } \mathbf{q\,4}\text{)}$$

$$= \left|\int_{F_1} \theta(y, s)f(z)\,dz\right|^2$$

as was to be shown.

Lemma 2. *Let $0 < \sigma < 2$ and $s \neq \frac{1}{2}$. If $f \in \mathfrak{B}_{-1}$ then*

$$Q(s)f(z) = \frac{1}{2s - 1}y^{1-s}\int_{F_1} \theta(y', s)f(z')\,dz' + O(1).$$

If $s \neq \frac{1}{2}$ and $\sigma \geqslant \frac{1}{2}$, and

$$\omega(s)K(s)f = f,$$

then $Q(s)f$ is bounded, and in particular lies in H.

Proof. We use the definition of the kernel $q(y, y'; s)$ and

$$(2s - 1)Q(s)f(z) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \int_{a}^{\infty} q(y, y'; s)f(x + iy') \frac{dy'}{y'^2}\, dx.$$

We write

$$\int_a^\infty = \int_a^y + \int_y^\infty .$$

The integrals with respect to y' then amount to

$$y^{1-s} \int_a^y \theta(y', s)f(z') \frac{dy'}{y'^2} + \int_y^\infty y'^{1-s}(y^s + c(s)y^{1-s})f(z') \frac{dy'}{y'^2} .$$

The tail end corresponding to the first integral is estimated by

$$y^{1-\sigma} \int_y^\infty (y'^{\sigma} + c(s)y'^{1-\sigma}) \frac{1}{y'} \frac{dy'}{y'^2} ,$$

which is bounded. Hence the first integral gives the main contribution in the lemma. The second integral is trivially seen to be bounded also. This proves the first assertion of the lemma. As to the second, if $\sigma > \frac{1}{2}$, then it is clear that $Q(s)f$ is bounded from the integrals. If $\sigma = \frac{1}{2}$ but $s \neq \frac{1}{2}$, then we use Lemma 1 to conclude that the integral expression vanishes under the eigenvalue assumption

$$\omega(s)K(s)f = f.$$

This concludes the proof.

Theorem 8. (Maass) *Let* $\operatorname{Re} s = \frac{1}{2}$ *and* $s \neq \frac{1}{2}$. *If* $\psi \in H = L^2(\Gamma\backslash\mathfrak{H})$ *and*

$$A\psi = s(1 - s)\psi,$$

then ψ *is analytic and satisfies the estimate*

$$|\psi(x + iy)| \ll e^{-2\pi y}.$$

In particular, $f = (I - \omega(s)T(\kappa))\psi \in \mathfrak{B}_{-1}$. *If* $\psi \in \mathfrak{B}_\mu$ *for some* $\mu > 0$ *and*

$$L\psi = s(1 - s)\psi$$

then there are constants b_0, c_0 *such that*

$$\psi(x + iy) = b_0 y^s + c_0 y^{1-s} + O(e^{-2\pi y}).$$

Proof. For any test function $g \in C_c^\infty(\Gamma\backslash\mathfrak{H})$ we have

$$\langle (A - s(1-s))\psi, g\rangle = 0 = \langle \psi, (A - s(1-s))g\rangle.$$

Hence by the regularity theorem for elliptic operators, we conclude that ψ is analytic and

$$(A - s(1-s))\psi = L\psi - s(1-s)\psi.$$

Then ψ has a convergent Fourier series expansion

$$\psi(x+iy) = a_0(y) + \sum_{n \neq 0} a_n(y)e^{2\pi inx}$$

where

$$a_n(y) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \psi(x+iy)e^{-2\pi inx}\,dx.$$

We then see, as for the Fourier series of the resolvant, that

$$a_n''(y) = \left(4\pi^2n^2 - \frac{s(1-s)}{y^2}\right)a_n(y),$$

i.e. that $a_n(y)$ is a solution of the Whittaker equation. Hence there exist constants b_n, c_n such that

$$a_n(y) = b_n W_s(4\pi|n|y) + c_n V_s(4\pi|n|y),$$

where W_s is the exponentially decreasing solution of the Whittaker equation, and V_s is the exponentially increasing solution. If $\psi \in L^2$ or $\mathfrak{B}_\mu$, it follows that $a_n(y)$ is in $L^1([a, \infty), dy/y^2)$, or is at most polynomially increasing, and hence that the exponentially increasing term must vanish, i.e. $c_n = 0$. Fix y_1 such that for $y \geqslant y_1$ we have

$$|W_s(y)| \geqslant \tfrac{1}{2}e^{-y/2}.$$

Since the Fourier series converges at every point, its coefficients are bounded. In particular, there exists $C_1 > 0$ such that for all $n \neq 0$ we have

$$|a_n(y_1)| = |b_n W_s(4\pi|n|y_1)| \leqslant C_1,$$

whence

$$|b_n| \ll e^{2\pi|n|y_1}.$$

If $y \geqslant 2y_1$ we get

$$\begin{aligned}|\psi(x + iy) - a_0(y)| &\ll \sum e^{2\pi|n|y_1}e^{-2\pi|n|y}\\ &\ll e^{-2\pi y},\end{aligned}$$

so we have shown that

$$\psi(x + iy) = a_0(y) + O(e^{-2\pi y}).$$

Now $a_0(y)$ is a solution of the equation

$$a_0''(y) = -\frac{s(1-s)}{y^2}a_0(y).$$

Therefore there exist constants b_0, c_0 such that

$$a_0(y) = b_0y^s + c_0y^{1-s}.$$

For $\operatorname{Re}(s) = \frac{1}{2}$ the only way this is possible if $\psi \in L^2$ is with $b_0 = c_0 = 0$, and Theorem 8 is proved.

Note. Arguments like the ones above are typical of those used by Maass [Ma 1].

From the analytic definition of $Q(s)$, which maps $\mathfrak{B}_{-1}$ into $\mathfrak{B}_{1-\sigma}$, and the fact V is compact, we obtain also the following properties of the family of operators $K(s)$.

Theorem 9. *The map $s \mapsto K(s)$ is an analytic family of operators from the strip $0 < \sigma < 2$ into the space of compact operators on $\mathfrak{B}_{-1}$, except possibly at $s = \frac{1}{2}$ where a pole may occur due to the factor $2s - 1$ in the denominator of the definition of $Q(s)$. In any case $s \mapsto (2s - 1)K(s)$ is analytic in the whole strip. The set of points s in the strip where*

$$I - \omega(s)K(s)$$

is not invertible is discrete, and $s \mapsto [I - \omega(s)K(s)]^{-1}$ has, at most, poles at these points.

Proof. The analyticity of $s \mapsto K(s)$ is clear. A thorough discussion of analyticity of kernels and operators is given in Appendix 5 and in the end of the next section. The fact that

$$s \mapsto [I - \omega K(s)]^{-1}$$

is meromorphic follows from a general nonsense fact about families of compact operators, whose proof is recalled in Appendix 3.

In view of Theorem 8, a point s in the strip where $I-\omega K(s)$ is not invertible, or $s=\frac{1}{2}$, will be called a **singular point**. The set of singular points is discrete. Theorem 7 shows that the singular points on the line $\sigma=\frac{1}{2}$ (other than $s=\frac{1}{2}$) and those on the segment $\frac{1}{2}<s\leqslant 1$ correspond to eigenvalues of the Laplacian.

§10. THE RESOLVANT EQUATIONS FOR $0<\sigma<2$

We continue to use $\kappa>3$ and

$$R=R(\kappa),\qquad V=V(\kappa),\qquad T=T(\kappa).$$

We have the decomposition

$$R=T+V.$$

Also

$$\omega=\omega(s)=s(1-s)-\kappa(1-\kappa).$$

We wish to find an "analytic" expression for the kernel of the resolvant which will be valid throughout the strip $0<\sigma<2$ and will represent the resolvant for $\sigma>\frac{1}{2}$. The basic resolvant equation is

$$R(s)-R=\omega(s)RR(s). \tag{1}$$

We begin by a uniqueness lemma.

Lemma 1. *Let $\frac{1}{2}\leqslant\sigma<2$ but $s\neq\frac{1}{2}$. Assume also that s is non-singular. Then there exists at most one bounded operator X on $H=L^2(\Gamma\backslash\mathfrak{H})$ such that*

$$X-R=\omega(s)RX.$$

Proof. Let X, X' be two solutions for the above equation. Then

$$X-X'=\omega(s)R(X-X').$$

If $X-X'\neq O$, then any vector $\psi=(X-X')h\neq 0$, with $h\in H$, is a solution of the equation

$$\psi=\omega(s)R\psi,$$

which by definition means that s is singular, contradiction.

We observe that the kernel $r(z, z'; s)$ represents the resolvant for $\sigma > \frac{1}{2}$ and s not on the segment $\frac{1}{2} < \sigma \leqslant 1$. Even though we shall find an expression for the kernel valid in the strip, it will *not* represent the resolvant for other values of s.

We now transform the resolvant equation algebraically, and now perform some formal computations. Ultimately, we want to express $R(s)$ for s non-singular in the strip in the form

$$R(s) = Q(s) + (I + \omega Q(s))B(s)(I + \omega Q(s)) \tag{2}$$

for some operator $B(s)$. Let us see formally the necessary and sufficient conditions that $B(s)$ must satisfy for this to hold. Using formally

$$(I - \omega T) = (I + \omega Q(s))^{-1}$$

and $T + Q(s)T = Q(s)$, we have:

$$X - R = \omega RX \Leftrightarrow X - R = \omega TX + \omega VX \tag{3}$$

$$\Leftrightarrow (I - \omega T)X = R + \omega VX \tag{4}$$

$$\Leftrightarrow X = (I + \omega Q(s))R + \omega(I + \omega Q(s))VX \tag{5}$$

$$\Leftrightarrow X = Q(s) + (I + \omega Q(s))V + \omega(I + \omega Q(s))VX. \tag{6}$$

Assuming that $X = Q(s) + (I + \omega Q(s))B(s)(I + \omega Q(s))$, this is

$$\Leftrightarrow B(s) = V + \omega V(I + \omega Q(s))B(s) \tag{7}$$

$$\Leftrightarrow B(s) = V + \omega K(s)B(s) \tag{8}$$

$$\Leftrightarrow [I - \omega K(s)]B(s) = V. \tag{9}$$

Our intent is to reduce the study of the resolvent to the study of compact operators. As we now see, the assumption achieves it.

Lemma 2. *For s non-singular in the strip, the operator*

$$I - \omega K(s) \colon \mathfrak{B}_{-1} \longrightarrow \mathfrak{B}_{-1}$$

is invertible, and there exists a unique bounded operator

$$B(s) \colon \mathfrak{B}_{-1} \longrightarrow \mathfrak{B}_{-1}$$

such that

$$[I - \omega K(s)]B(s) = V.$$

Furthermore, $B(s)$ is compact.

Proof. We know that $K(s)$ is compact, and hence that $\omega K(s)$ is compact. Hence $I - \omega K(s)$ is Fredholm. The assumption that s is non-singular implies that the kernel of $I - \omega K(s)$ is 0. It follows that $I - \omega K(s)$ is invertible. [We are using here the fact that the index of a Fredholm operator is constant on connected components, and that for compact operator K, the segment $I - tK$, $0 \leqslant t \leqslant 1$, joins $I - K$ to the identity in the space of Fredholm operators. Cf. *Real Analysis*.] Since V is a compact operator on $\mathfrak{B}_{-1}$, it follows that $B(s)$ is compact.

The operator V is defined by a kernel v which is not continuous because of a logarithm appearing on the diagonal. For this reason it is convenient to make one more transformation on $B(s)$, namely to let

$$B_1(s) = B(s) - V. \tag{10}$$

The equation

$$[I - \omega K(s)]B(s) = V$$

is then equivalent with

$$[I - \omega K(s)]B_1(s) = B_0(s) \tag{11}$$

where

$$B_0(s) = \omega V(I + \omega Q(s))V. \tag{12}$$

We may now proceed backward, and define a kernel corresponding to the above equation. For each non-singular s in the strip, we define the operator $B_0(s)$ by the kernel

$$\boxed{b_{0,s} = \omega v * v + \omega^2 v * q_s * v} \tag{13}$$

where $v = v(\kappa)$, and q_s are the kernels for V and $Q(s)$ respectively. The definition applies "componentwise" for the ij-components of the kernel, corresponding to the product $F_i \times F_j$ of the basic parts of our fundamental domain.

Lemma 3. *Let $0 < \sigma < 2$ and assume s non-singular.*

i) *Each component of the kernels $v * v$ and $v * q_s * v$ is of type $\mathfrak{B}_{-1}$, and hence each component of the kernel $b_{0,s}$ is of this type.*
ii) *The family of functions $b_{0,z,s}$ such that*

$$b_{0,z,s}(z') = b_0(z, z'; s)$$

is equicontinuous for $z \in F_i$ (any i).

Proof. Let us look first at $v * v$, and more specifically at its component acting on $F_0 \times F_0$. This component has a kernel of the form

$$r_{00}(z, z'; \kappa) = \frac{1}{2} \sum_{\gamma \in \Gamma} \varphi(u(z, \gamma z'); \kappa).$$

Up to a constant factor, this kernel can be written in the form

$$\log |z - z'| + \text{continuous function of } (z, z'),$$

for $z, z' \in F_0$. In the composition $r_{00} * r_{00}$ the worst possibility comes from convoluting the logs, in which case we must see that the function of (z, z') given by

$$\int_{F_0} \log |z - z''| \log |z'' - z'| \, dz''$$

is continuous. Let $g(z, z') = \log |z - z'|$ and let

$$g_\epsilon(z, z') = \begin{cases} \log |z - z'| & \text{if } |z - z'| \geqslant \epsilon \\ \log \epsilon & \text{if } |z - z'| \leqslant \epsilon. \end{cases}$$

Then g_ϵ is continuous, and $g_\epsilon * g_\epsilon$ is clearly continuous. It suffices to prove that $g_\epsilon * g_\epsilon$ tends to $g * g$ uniformly. We have

$$\begin{aligned} &\int [g(z, z'')g(z'', z') - g_\epsilon(z, z'')g_\epsilon(z'', z')] \, dz'' \\ &= \int_{|z'' - z| < \epsilon} [g(z, z'')g(z'', z') - g_\epsilon(z, z'')g(z'', z')] \, dz'' \\ &\quad + \int_{|z'' - z'| < \epsilon} [g_\epsilon(z, z'')g(z'', z') - g_\epsilon(z, z'')g_\epsilon(z'', z')] \, dz'' \\ &\ll \epsilon^2 \int_{F_0} |\log |z'' - z| \log |z'' - z'|| \, dz'' \\ &\ll \epsilon^2 \int_{F_0} \log^2 |z - z''| \, dz'' \qquad \text{(by Schwarz)} \\ &\ll \epsilon^2, \end{aligned}$$

which proves the continuity of $r_{00} * r_{00}$.

The operator V is of the form $V = M + N$, where the components n_{01}, n_{10}, n_{11} of N are of type $\mathfrak{B}_{-1}$ and M_{11} is represented by a kernel

$$m(z, z'; \kappa) = \sum_{k \geqslant 1} m_k(y, y'; \kappa) \cos 2\pi k(x - x')$$

with

$$|m_k(y, y'; \kappa)| \ll \frac{1}{k} e^{-2\pi k|y - y'|}.$$

The other components M_{00}, M_{01}, M_{10} are equal to 0. The kernels giving NN, MN, and NM are easily verified to be of the desired type. We carry out the details for the last convolution.

Lemma 4. *If $m(z, z'; \kappa)$ is a kernel as above, then $m * m$ is of type $\mathfrak{B}_{-1}$.*

Proof. We have the orthogonality

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} \cos 2\pi k(x - x'') \cos 2\pi \ell(x - x'')\, dx'' = 0$$

unless $k = \ell$. Hence

$$|m * m(y, y')| \ll \sum_{k \geqslant 1} |m_k * m_k(y, y')|,$$

and

$$|m_k * m_k(y, y')| \ll \int_a^\infty \frac{e^{-2\pi k|y - y''|} e^{-2\pi k|y'' - y'|}}{k^2} \frac{dy''}{y''^2}.$$

We split the integral as before, for, say, $y' < y$:

$$\int_a^\infty = \int_a^{y'} + \int_{y'}^{y} + \int_y^\infty.$$

We find the estimate for the integral to be

$$\ll \frac{1}{k^2} e^{-2\pi k(y + y')} \left(\frac{e^{4\pi k y'}}{y'^2} + \frac{e^{4\pi k a}}{a^2} \right).$$

If $y \leqslant 2y'$, then the term involving the constant a goes to 0 exponentially in $y + y'$ and is therefore better than the desired type. The first term behaves like $1/y'^2$, i.e. like $1/yy'$ as we want. If $y \geqslant 2y'$ then $e^{-2\pi k y}$ wins over everything else.

We leave to the reader that $v * q_s * v$ is of type $\mathfrak{B}_{-1}$. Recall only that

$$MQ(s) = Q(s)M = 0$$

because of the orthogonality of cosines and the constant function, and the fact that by definition, $Q(s)$ only has a 11-component, depending only on y.

This takes care of the first part of Lemma 3.

As to the second part, I see no way but to check in each of the convolutions of the different components of the kernels that one can take a limit under the integral signs to show the equicontinuity. This is just more of the same boring routine techniques as above, and is left to the reader, just this once, as in Faddeev [Fa 1], p. 377, line 7.

Lemma 5. *Let $0 < \sigma < 2$ and assume s non-singular. The operator $B_1(s)$ defined by*

$$B_1(s) = [I - \omega K(s)]^{-1} B_0(s)$$

can be defined by a kernel $b_1(z, z'; s)$ which is of type $\mathfrak{B}_{-1}$.

Proof. We know from Lemma 4 that the kernel $b_0(z, z'; s)$ for $B_0(s)$ is of type $\mathfrak{B}_{-1}$. Write

$$b_0(z, z'; s) = b_{0, z', s}(z),$$

viewing $b_{0, z', s}$ as a function of z, in $\mathfrak{B}_{-1}$. Then

$$|b_{0, z', s}(z)| \ll \frac{1}{y'} \frac{1}{y}$$

and

$$\begin{aligned} \|[I - \omega K(s)]^{-1} b_{0, z', s}\|_{-1} &\ll \|b_{0, z', s}\|_{-1} \\ &\ll \frac{1}{y'}. \end{aligned}$$

If we define

$$b_1(z, z'; s) = \left((I - \omega K(s))^{-1} b_{0, z', s}\right)(z),$$

then the inequality we have just obtained means that

$$|b_1(z, z'; s)| \ll \frac{1}{yy'},$$

because for any function $f \in \mathfrak{B}_{-1}$ we have

$$\|f\|_{-1} = \sup |f(z)y|.$$

so that if $\|f\|_{-1} \ll 1/y'$, then $|f(z)| \ll 1/yy'$. This gives the desired asymptotic estimate. Continuity of b_1 is proved by decomposing

$$b_1(z_1, z_1') - b_1(z_2, z_2') = b_1(z_1, z_1') - b_1(z_2, z_1') + b_1(z_2, z_1') - b_1(z_2, z_2').$$

We fix (z_1, z_1') and let (z_2, z_2') approach (z_1, z_1'). The first difference on the right approaches 0 because $b_{1, z_1'}$ is continuous. The second difference approaches 0 because a similar continuity property holds for b_0 instead of b_1 according to the second part of Lemma 3, and this property is preserved by applying a bounded operator to the function space $\mathfrak{B}_{-1}$. This proves Lemma 5.

We are now finished with the continuity properties of the various kernels b_0 and b_1. We turn to the dependence on s. We have the analytic dependence of $B(s)$ on s.

Theorem 10. *For each non-singular s in the strip let $B(s)$ be defined by*

$$B(s) = [I - \omega(s)K(s)]^{-1}V.$$

i) *The map $s \longmapsto B(s)$ is a meromorphic map from the strip into the Banach space of bounded operators on $\mathfrak{B}_{-1}$. The poles occur only at the singular points.*
ii) *There is a decomposition*

$$B(s) = M + N^B(s)$$

where $N^B(s)$ is an operator defined by a kernel of type $\mathfrak{B}_{-1}$, and $M = M(\kappa)$ with $\kappa > 3$.
iii) *For $\sigma > \frac{1}{2}$, s not real, as operators on H, we have*

$$R(s) = Q(s) + (I + \omega Q(s))B(s)(I + \omega Q(s)).$$

Proof. The first assertion is immediate from Theorem 9 in the preceding section. The second follows from the lemmas, taking into account the equation

$$B(s) = B_1(s) + V = B_1(s) + M + N.$$

The third is clear from the sequence of equivalences (3) through (9), and the uniqueness of the resolvant satisfying the resolvant equation by Lemma 1. This proves our theorem.

We end this section by giving the form which the resolvant equation takes for $B(s)$.

Theorem 11.

i) *For* $0 < \sigma < 1$ *and* s, $1 - s$ *non-singular, we have, as operators on* $\mathfrak{B}_{-1}$,

$$B(s) - B(1 - s) = \omega(s)^2 B(s)[Q(s) - Q(1 - s)]B(1 - s).$$

ii) *For* $\sigma, \sigma' > \frac{1}{2}$ *we have, as operators on* H,

$$B - B' = (\omega - \omega')B(I + \omega Q)(I + \omega' Q')B',$$

where $B = B(s)$, $B' = B(s')$, $\omega = \omega(s)$, $\omega' = \omega(s')$, *etc.*

Proof. We start with $B = (I - \omega K)^{-1}V$. Let $0 < \sigma, \sigma' < 2$. On $\mathfrak{B}_{-1}$ we get

$$\begin{aligned} B - B' &= \left[(I - \omega K)^{-1} - (I - \omega' K')^{-1}\right]V \\ &= (I - \omega K)^{-1}(\omega K - \omega' K')(I - \omega' K')^{-1}V. \end{aligned}$$

Since $K = V(I + \omega Q)$ and $K' = V(I + \omega' Q')$ we obtain

$$(*) \qquad B - B' = B(\omega - \omega' + \omega^2 Q - \omega'^2 Q')B'.$$

Putting $s' = 1 - s$, in which case $\omega(s) = \omega(1 - s)$ we get our first assertion. On the other hand, the last assertion (*), valid on $\mathfrak{B}_{-1}$ for $0 < \sigma < 2$ and $0 < \sigma' < 2$ can now be read in the restricted domain $\sigma, \sigma' > \frac{1}{2}$, in which case the operators appearing in it are bounded operators on H, and (*) is valid on H. Using the formula

$$Q - Q' = (\omega - \omega')QQ',$$

valid on H, we see that (ii) is equivalent to (*) on H. This proves our theorem.

In the sequel we make use only of the first resolvant relation in the strip. This relation can be viewed as a "functional equation" for $B(s)$, from which the functional equation of the Eisenstein functions (to be defined in §12) will follow immediately.

§11. THE KERNEL OF THE RESOLVANT FOR $0 < \sigma < 2$

In this section, we shall see that even though the resolvant of the Laplace operator itself does not analytically cross the line $\sigma = \frac{1}{2}$, we can nevertheless make an analytic continuation of the *kernel* as a function of s into the entire strip $0 < \sigma < 2$, by giving an analytic continuation of the kernel $b(z, z'; s)$. I am indebted to R. Bruggeman for the exposition of this section.

Theorem 12. *Let U be an open set in $\mathbf{C}$. Let $\mu, \nu \in \mathbf{R}$. For each s in U, assume given a continuous operator*

$$R(s)\colon \mathfrak{B}_\mu \to \mathfrak{B}_\nu,$$

and assume that this operator is represented by a kernel $r(z, z'; s)$, which for each z is measurable on $F \times U$. For each compact set K in U assume that there is a bounded operator $R_K\colon \mathfrak{B}_\mu \to \mathfrak{B}_\nu$ represented by a kernel r_K satisfying

$$|r(z, z'; s)| \leqslant r_K(z, z')$$

for almost all (z, z') and all $s \in K$. Then $r(z, z'; s)$ is analytic in s for almost all (z, z') if and only if $s \mapsto R(s)$ is analytic.

Proof. We use Theorem 1 of Appendix 5 in connection with a statement similar to Theorem 2 of that appendix. Namely, we first prove

Lemma 1. *Let the hypotheses be as in Theorem 12. Then $r(z, z'; s)$ is analytic in s for almost all (z, z') if and only if for all $f \in \mathfrak{B}_\mu$ and all $z \in F$ the function*

$$s \mapsto R(s)f(z)$$

is analytic.

$\Rightarrow$: Let C be a circle enclosing a disc in U. Let $f \in \mathfrak{B}_\mu$ and $z \in F$. The function

$$z' \mapsto r_C(z, z')|f(z')|$$

is integrable and majorizes $z' \mapsto r(z, z'; s)f(z')$ for all $s \in C$. Hence $s \mapsto R(s)f(z)$ is continuous by the dominated convergence theorem. Now $(z', s) \mapsto r(z, z'; s)$ is measurable and bounded by the integrable function.

$$(z', s) \mapsto r_C(z, z')|f(z')|,$$

so that $(z', s) \mapsto r(z, z'; s)$ is integrable and

$$\begin{aligned}\int_C R(s)f(z)\,ds &= \int_C \int_F r(z, z'; s)f(z')\,dz'\,ds \\ &= \int_F \int_C r(z, z'; s)\,ds\,f(z')\,dz' \\ &= 0.\end{aligned}$$

This proves that $s \mapsto R(s)f(z)$ is analytic.

$\Leftarrow$: If $s \mapsto R(s)f(z)$ is analytic for each f, z, it is clear that we can again interchange an order of integration to find

$$\int_F \int_C r(z, z'; s)\, ds\, f(z')\, dz' = 0$$

for all $z \in F$ and $f \in \mathfrak{B}_\mu$. Then for almost all (z, z') we get

$$\int_C r(z, z'; s)\, ds = 0.$$

By the lemma of Appendix 5, §1, we conclude that $r(z, z'; s)$ is analytic in s.

We verify next that the hypotheses of Theorem 1, Appendix 5, are verified, with an appropriate set of functionals.

Since r is bounded by r_K for any compact K, we get

$$\sup_{s \in K} |R(s)| \leqslant |R_K|.$$

We let Λ be the set of functionals $\lambda_{f,z}$ on the Banach space

$$E = \operatorname{Hom}(\mathfrak{B}_\mu, \mathfrak{B}_\nu),$$

where f lies in the unit ball in $\mathfrak{B}_\mu$, $z \in F$, and

$$\lambda_{f,z}(A) = \frac{Af(z)}{y^\nu}$$

for any continuous linear map $A: \mathfrak{B}_\mu \longrightarrow \mathfrak{B}_\nu$. Then $\lambda_{f,z}$ is the composite of two maps,

$$A \mapsto Af \qquad \text{and} \qquad g \mapsto \frac{g(z)}{y^\nu}$$

which are continuous linear, of norms $\leqslant 1$, and norm determining. This shows that we can apply Theorem 1, Appendix 5, and concludes the proof of Theorem 12.

Corollary. *If we add to the assumptions of Theorem* 12 *that for each z and s the function*

$$z' \mapsto r(z, z'; s)$$

is continuous, and $z' \mapsto r_K(z, z')$ is locally bounded, then:

$$s \mapsto r(z, z'; s) \textit{ is analytic for all } (z, z') \Leftrightarrow s \mapsto R(s) \textit{ is analytic.}$$

Proof. Only the implication $\Leftarrow$ involves additional information, and amounts to proving that if

$$\int_C \int_F r(z, z'; s) f(z')\, dz'\, ds = 0$$

for all $z \in F$ and $f \in \mathfrak{B}_\mu$, then

$$\int_C r(z, z'; s)\, ds = 0$$

for all z, z'. Fix z and let $g(z')$ be the function of z' given by this last integral. For any real $f \in C_c(F)$, and therefore $f \in \mathfrak{B}_\mu$, we get

$$\langle g, f \rangle = \int_F \int_C r(z, z'; s) f(z')\, ds\, dz' = 0.$$

Hence g is equal to 0 almost everywhere. The assumptions of the theorem show that g is continuous, whence $g = 0$. This proves the corollary.

Let Ω be a compact subset of the strip $0 < \sigma < 2$. We can define a dominating kernel for $0 < \sigma_1 < \sigma < \sigma_2 < 2$ by

$$q_\Omega(z, z') = c_1 \begin{cases} (y^{\sigma_2} + c_2 y^{1-\sigma_1}) y'^{1-\sigma_1} & \text{if } \ y < y', \\ y^{1-\sigma_1}(y'^{\sigma_2} + c_2 y'^{1-\sigma_1}) & \text{if } \ y' < y, \end{cases}$$

with appropriate constants c_1, c_2, so that for all $s \in \Omega$, we have

$$|q(z, z'1\ s)| \leqslant q_\Omega(z, z').$$

We let

$$Q_\Omega \colon \mathfrak{B}_0 \longrightarrow \mathfrak{B}_1$$

be the bounded operator having only a 11-component represented by the above kernel. As with $Q(s)$, we see that Q_Ω maps $\mathfrak{B}_0$ continuously into $\mathfrak{B}_{1-\mu}$, if $0 < \mu < \sigma_1 < 2$, and we write

$$Q_\Omega \colon \mathfrak{B}_0 \longrightarrow \mathfrak{B}_{1-\mu}.$$

The next two lemmas give estimates needed to apply Theorem 12.

Lemma 2. *Let Ω be a compact neighborhood in the strip of a point s_0. Let m be the order of the pole of the operator*

$$[I - \omega(s) K(s)]^{-1}$$

on $\mathfrak{B}_{-1}$. Then there exists $C > 0$ such that for all $s \in \Omega$ we have

$$|s - s_0|^m \, |b_1(z, z'; s)| \leqslant C(yy')^{-1},$$

where, as before, $b_1(z, z'; s)$ is the kernel of $B_1(s)$.

Proof. The proof given for Lemma 5 in the preceding section works uniformly for s in the compact set Ω.

We give a symbol to the kernel which extends the kernel of the resolvant to the strip. We let

$$\boxed{\rho_s = q_s + (1 + \omega_s q_s) * b_s * (1 + \omega_s q_s)}$$

so that $\rho_s(z, z') = \rho(z, z'; s)$, and $\rho(z, z'; s)$ represents the resolvant for $\frac{1}{2} < \sigma < 2$, s not real, by Theorem 10, iii. Except possibly on the diagonal and on the boundaries of the regions F_i, used to decompose the fundamental domain, the two kernels $\rho(z, z'; s)$ and $r(z, z'; s)$ are continuous in (z, z') and coincide for $1 < \sigma < 2$.

Lemma 3. *Let Ω be a compact neighborhood in the strip of a point s_0, and assume Ω is contained in $0 < \sigma_1 < \sigma < \sigma_2$. Let m be as in Lemma 2. There exist constants c_i such that for any $s \in \Omega$ we have*

$$|s - s_0|^m \, |\rho(z, z'; s)|$$
$$\leqslant c_1 q_\Omega(z, z') + c_2(yy')^{-1} + c_3|m(z, z')| + c_4(yy')^{1-\sigma_1}.$$

Proof. This follows by estimating the integrals entering into the convolutions defining $\rho(z, z'1\, s)$.

We are now in a position to apply Theorem 12. Knowing that certain families of operators are meromorphic, and knowing an appropriate local boundedness condition on the corresponding kernels, we conclude that the kernels themselves are meromorphic, namely:

Theorem 13. *For $z \neq z'$ and z, z' not on the boundaries between the regions F_i of the fundamental domain F, the functions*

$$s \longmapsto \rho(z, z'; s) \quad \text{and} \quad s \longmapsto b(z, z'; s)$$

are meromorphic in the strip $0 < \sigma < 2$, with poles only at the singular points.

The orders of the poles and the principal parts will be determined by relating the situation to the self-adjointness of the operator which $\rho(z, z'; s)$ represents for $\sigma > \frac{1}{2}$, and we shall obtain:

Theorem 14. *Let s_0 be a singular point with $\sigma_0 \geqslant \frac{1}{2}$ and $s_0 \neq \frac{1}{2}$. Let $\psi_1, \ldots, \psi_n$ be a complete orthonormal system of eigenfunctions of the self-adjoint operator A in H, chosen to be real. Let $\lambda = s(1-s)$ and $\lambda_0 = s_0(1-s_0)$. Let*

$$f_i = (I - \omega(s_0)T)\psi_i.$$

Then $f_i \in \mathfrak{B}_{-1}$ also. Furthermore, identifying r and ρ,

$$r(z, z'; s) = \frac{1}{\lambda_0 - \lambda} \sum_{i=1}^{n} \psi_i(z)\psi_i(z') + r^+(z, z'; s),$$

$$b(z, z'; s) = \frac{1}{2s_0 - 1} \sum_{i=1}^{n} f_i(z)f_i(z') + b^+(z, z'; s)$$

where $r^+(z, z'; s)$ and $b^+(z, z'; s)$ are holomorphic in s near s_0.

Proof. The difficulty, such as it is, which prevents us from dealing exclusively with the operators is that on the line $\sigma = \frac{1}{2}$ we also have a continuous spectrum for A, so that the resolvant of A, as an operator, does not have a power series expansion in the neighborhood of a singular point on that line, even though we prove that the analytic continuation of the kernel does. This implies that we have to go through other spaces, e.g. $\mathfrak{B}_\mu$ spaces once again, and at the same time have to use properties of the resolvant to the right of $\sigma = \frac{1}{2}$, in particular the property

$$|R(s)_H| \ll \frac{1}{d(\lambda_s, \text{spectrum of } A)}$$

where "d" means the distance, and $\lambda_s = s(1-s)$. The spectrum of A lies on the real line, and corresponds to values of s such that

$$\tfrac{1}{2} < s \leqslant 1 \qquad \text{or} \qquad \operatorname{Re} s = \tfrac{1}{2}.$$

We put an index H on $R(s)$ to emphasize that we view it as operator in H. Note that

$$s(1-s) - s_0(1-s_0) = \lambda_s - \lambda_{s_0} = \omega(s) - \omega(s_0)$$

grows like the first power of $\lambda_s - \lambda_{s_0}$. The above property is merely a formulation in terms of the variable s of the immediate estimate in terms of λ given by

$$|(A - \lambda I)^{-1}| \leqslant \frac{1}{|\mathrm{Im}\,\lambda|}.$$

Cf. Appendix 2, §1, Theorem 2.

We now come to the proof proper. Let $\mu \geqslant 0$. For $\mu < \sigma < 2$ we have an operator

$$R(s)\colon \mathfrak{B}_0 \longrightarrow \mathfrak{B}_{1-\mu}$$

defined by the usual formula

$$R(s) = Q(s) + (I + \omega(s)Q(s))B(s)(I + \omega(s)Q(s)),$$

and the association $s \longmapsto R(s)$ is meromorphic. We pick a specific value of μ which allows us to work with $\sigma \geqslant \frac{1}{2}$, say $\mu = \frac{1}{4}$.

Lemma 4. *Let s_0 be a singular point $\neq \frac{1}{2}$ and $\sigma_0 \geqslant \frac{1}{2}$. Let*

$$R(s)\colon \mathfrak{B}_0 \longrightarrow \mathfrak{B}_{3/4}$$

be the operator defined by the kernel $r(z, z'; s)$, for s near s_0. Then $s \longmapsto R(s)$ has a pole of order at most 1 at s_0.

Proof. Let $\{s_n\}$ be a sequence of non-singular points with $\mathrm{Re}\, s_n > \frac{1}{2}$, converging to s_0, and such that $\mathrm{Re}(\lambda_{s_0} - \lambda_{s_n}) = 0$. Thus the imaginary parts of the λ_{s_n} tend to 0 with the same order as s_n tends to s_0. Let

$$R_{-m} = \lim_{s \to s_0} (s - s_0)^m R(s),$$

where m is the order of the pole. We have to show that $R_{-m} = 0$ if $m > 1$. Let $f \in \mathfrak{B}_0$. It suffices to prove $R_{-m}f = 0$. Suppose $R_{-m}f \neq 0$. Let

$$g_n = (\lambda_{s_n} - \lambda_{s_0})^m R(s_n)f.$$

Then $g_n \longrightarrow R_{-m}f$ in $\mathfrak{B}_{3/4}$. Therefore

$$|R_{-m}f(z)y^{-3/4}| \geqslant c_1 > 0$$

for some constant c_1 and all y in some open set. By the definition of the norm in $\mathfrak{B}_{3/4}$ we also obtain

$$\left|[g_n(z) - R_{-m}f(z)]y^{-3/4}\right| < \epsilon,$$

uniformly for y in some open set, for all n sufficiently large, whence for n large we get the inequality

$$|g_n(z)y^{-3/4}| \geqslant c_2 > 0,$$

and so for some constant c_3 and all y in some open set,

$$|g_n(z)| \geqslant c_3 > 0.$$

Since $\mathfrak{B}_0 \subset H$ and $\operatorname{Re} s_n > \frac{1}{2}$ it follows that $R(s_n)f \in H$ and so $g_n \in H$. The inequality we just obtained shows that there is a constant $c_4 > 0$ such that for all sufficiently large n we have

$$\| g_n \|_2^2 \geqslant c_4.$$

But

$$\|g_n\| \leqslant |\lambda_{s_n} - \lambda_0|^{m} |R(s_n)_H| \, \|f\|_2,$$

which contradicts the resolvant inequality in the Hilbert space H, and proves our lemma.

From Lemma 4 we shall now see that $r(z, z'; s)$ has a pole of order at most 1 at s_0. We know from Lemma 2 that it has a pole of order at most m. Say it has a pole of order $l \geqslant 2$. For any small contour C around s_0 corresponding to a circle around λ_0, and $f \in \mathfrak{B}_0$ we have by Lemma 4

$$\int_C \int_F (\lambda_0 - \lambda_s)^{l-1} r(z, z'; s) f(z') \, dz' \, ds = 0.$$

The estimate of Lemma 3 (the absolute value $|\lambda_0 - \lambda_s|$ is now fixed) shows that we can interchange the order of integration, and hence get

$$\int_C (\lambda_0 - \lambda_s)^{l-1} r(z, z'; s) \, ds = 0$$

for almost all (z, z'), whence for all (z, z') subject to the conditions of Theorem 13. This proves that $l = 1$.

We must now determine the principal part of the kernel, and the argument will follow a similar pattern, first dealing with the operator family, and then passing to the kernel.

We multiply the equation

$$R(s) - Q(s) = (I + \omega Q(s))B(s)(I + \omega Q(s))$$

on the right and on the left by $I - \omega T$, and get

$$B(s) = (I - \omega T)(R(s) - Q(s))(I - \omega T),$$

thus seeing that $s \mapsto B(s)$ has a pole of first order at s_0, viewing $B(s)$ as operator from $\mathfrak{B}_0$ to $\mathfrak{B}_{3/4}$. It follows from Theorem 12 that the kernel $b(z, z'; s)$ representing this operator has a pole of first order at s_0. But that same kernel actually defines an operator

$$\mathfrak{B}_{3/4} \longrightarrow \mathfrak{B}_{-1}$$

which therefore also has a pole of order 1, by that same theorem.

Lemma 5. *Let* $\mathfrak{B}_{-1}(s_0)$ *be the* $\omega(s_0)^{-1}$-eigenspace *of* $K(s_0)$ *in* $\mathfrak{B}_{-1}$. *Let*

$$B_{-1} = \lim_{s \to s_0} (\omega(s_0) - \omega(s)B(s)),$$

viewing $B(s)$ *as operator from* $\mathfrak{B}_{3/4}$ *to* $\mathfrak{B}_{-1}$. *Then*

$$B_{-1}\colon \mathfrak{B}_{3/4} \longrightarrow \mathfrak{B}_{-1}(s_0)$$

maps $\mathfrak{B}_{3/4}$ *into the above eigenspace.*

Proof. We have

$$[I - \omega(s)K(s)]B(s) = V.$$

Multiply both sides by $\omega(s_0) - \omega(s)$ and let $s \to s_0$. The limit makes sense and proves our lemma.

Lemma 6. *Define*

$$R_{-1} = \lim_{s \to s_0} (\omega(s_0) - \omega(s))R(s)\colon \mathfrak{B}_0 \longrightarrow \mathfrak{B}_{3/4}$$

as an operator relation among the spaces indicated. Then R_{-1} *maps* $\mathfrak{B}_0$ *into the eigenspace* $H(s_0)$ *with eigenvalue* λ_{s_0} *for* A, *and induces the identity mapping on this eigenspace.*

Proof. We multiply the relation

$$R(s) = Q(s) + (I + \omega(s)Q(s))B(s)(I + \omega(s)Q(s))$$

by $\omega(s_0) - \omega(s)$ on both sides and let $s \longrightarrow s_0$. On the left we obtain the limit R_{-1}. On the right we find

$$(I + \omega(s_0)Q(s_0))B_{-1}(I + \omega(s_0)Q(s_0)).$$

By the previous lemma and the fundamental theorem on eigenvectors, Theorem 7, §9, we see that R_{-1} maps $\mathfrak{B}_0$ into $H(s_0)$. Furthermore, if $\psi \in H(s_0)$, and therefore $\psi \in \mathfrak{B}_0$ by Theorem 8, §9, we know that for $\sigma > \frac{1}{2}$ we have

$$(\omega(s_0) - \omega(s))R(s)\psi = \psi.$$

This is also a relation in $\mathfrak{B}_{3/4}$. Let $s \longrightarrow s_0$, and take the limit in $\mathfrak{B}_{3/4}$. We get $R_{-1}\psi = \psi$, thereby proving the lemma.

***Lemma* 7.** *The operator R_{-1} maps $\mathfrak{B}_0 \cap H(s_0)^{\perp}$ into* 0.

Proof. Let $g \in \mathfrak{B}_0 \cap H(s_0)^{\perp}$. Let $\psi \in H(s_0)$. For $\sigma > \frac{1}{2}$ we have

$$\langle(\lambda_0 - \lambda_s)R(s)g, \psi\rangle = \langle g, (\lambda_0 - \lambda_{\bar{s}})R(\bar{s})\psi\rangle = 0.$$

The left-hand side is an integral

$$\int_F (\lambda_0 - \lambda_s)R(s)g(z)\overline{\psi(z)}\, dz,$$

and we view $R(s)g$ as an element of $\mathfrak{B}_{3/4}$. We take the limit as $s \longrightarrow s_0$ and $\sigma > \frac{1}{2}$. We then get

$$\langle R_{-1}g, \psi\rangle = 0.$$

Since by Lemma 6 we also know that $R_{-1}g \in H(s_0)$, it therefore follows that $R_{-1}g = 0$, as desired.

Let $\{\psi_1, \ldots, \psi_n\}$ be a complete orthonormal basis of $H(s_0)$, we may assume each ψ_i is a real function. The kernel

$$r_{-1}(z, z') = \sum_{i=1}^{n} \psi_i(z)\psi_i(z')$$

is the kernel of the projection of H on $H(s_0)$, and of $\mathfrak{B}_0$ on $H(s_0)$, because we have an orthogonal decomposition

$$\mathfrak{B}_0 = H(s_0) \oplus [\mathfrak{B}_0 \cap H(s_0)^{\perp}].$$

The kernel

$$r(z, z'; s) - \frac{1}{\lambda_0 - \lambda_s} r_{-1}(z, z')$$

is the kernel of the operator

$$R(s) - \frac{1}{\lambda_0 - \lambda_s} R_{-1}: \mathfrak{B}_0 \longrightarrow \mathfrak{B}_{3/4}.$$

The family of these operators is holomorphic in s at s_0. We then find that for a contour C around s_0,

$$\int_C \left[r(z, z'; s) - \frac{1}{\lambda_0 - \lambda_s} r_{-1}(z, z') \right] ds = 0,$$

using again the estimates of Lemma 3 and the corresponding fact for the operator family, taking a scalar product with an arbitrary function $f \in \mathfrak{B}_0$. This proves the part of Theorem 13 concerning $r(z, z'; s)$.

The part of Theorem 13 concerning $b(z, z'; s)$ is obtained by convolving the kernel $r(z, z'; s)$ with $1 - \omega(s)t$. This concludes the proof.

§12. THE EISENSTEIN OPERATOR AND EISENSTEIN FUNCTIONS

We let $\theta(z, s)$ be the function on F whose component on F_0 is 0, while its component on F_1 is

$$\theta_1(z, s) = \theta(y, s) = y^s + c(s)y^{1-s},$$

where

$$c(s) = a^{2s-1} \frac{s - \kappa}{s + \kappa - 1} .$$

This function lies in $\mathfrak{B}_\mu$, where $\mu = \max(\sigma, 1 - \sigma)$. We let $W(s)$ be the operator

$$W(s) = \omega[I + \omega Q(s)]B(s)$$

where as before, $\omega = \omega(s) = s(1 - s) - \kappa(1 - \kappa)$. This operator is defined for non-singular s in the strip $0 < \sigma < 2$ and maps

$$\mathfrak{B}_{2-\epsilon} \longrightarrow \mathfrak{B}_{1-\sigma}.$$

We let

$$I + W(s)$$

be the **Eisenstein operator**, which we can apply to the functions $\theta(z, s) = \theta_s(z)$ for $0 < \sigma < 2$, thereby getting the **Eisenstein functions**

$$\eta(z, s) = (I + W(s))\theta(z, s) = \theta(z, s) + W(s)\theta(z, s).$$

Theorem 14. *For fixed z the functions $\eta(z, s)$ are analytic in s in the strip $0 < \sigma < 2$, except for singular points for which either $\sigma < \frac{1}{2}$ or $\frac{1}{2} \leqslant s \leqslant 1$. In a neighborhood of the line $\sigma = \frac{1}{2}$, except possibly at $s = \frac{1}{2}$, these functions are analytic.*

Proof. If $\sigma \neq \frac{1}{2}$, our assertion is clear from the analyticity property of the kernels and functions involved. Let us look at the line $\sigma = \frac{1}{2}$. Let s_0 be such that $\sigma_0 = \frac{1}{2}$ but $s_0 \neq \frac{1}{2}$. From the analytic expression for the kernel $b(z, z'; s)$ (§11, Th. 13) and the definition of the eigenfunctions

$$\psi_i = (I + \omega Q(s))f_i,$$

we see that

$$\eta(z, s) = \frac{\omega(s_0)}{2s_0 - 1}\,\frac{1}{s - s_0}\sum_{i=1}^{m}\psi_i(z)\int_F f_i(z')\theta(z', s)\,dz' + \eta_1(z, s)$$

where $\eta_1(z, s)$ is analytic in a neighborhood of s_0. By Lemma 1 of §9 the integral on the right is equal to 0, for values of s of the form $\frac{1}{2} + it$, and t near t_0. Since the integral is analytic in s, it vanishes, thereby proving our theorem.

We shall continue to use the convolution notation for kernels, but now taken on F. In other words, suppose that $k_1(z, z')$ and $k_2(z, z')$ are kernels with variables $z, z' \in F$. Then by definition for what follows,

$$k_1 * k_2(z, z') = \int_F k_1(z, z'')k_2(z'', z')\,dz''.$$

A similar notation applies when convolving a kernel with a function, namely

$$k_1 * f(z) = \int_F k_1(z, z')f(z')\,dz'$$

$$f * k_1(z) = \int_F f(z')k_1(z', z)\,dz'.$$

Let b_s be the kernel for the operator $B(s)$, and define

$$\boxed{\mathbf{c}(s) = c(s) + \frac{\omega(s)^2}{2s-1}\theta_s * b_s * \theta_s}$$

where the convolution product written out reads

$$\theta_s * b_s * \theta_s = \int_F \int_F \theta(z, s)b(z, z'; s)\theta(z', s)\, dz\, dz'.$$

We then get an asymptotic description of the Eisenstein functions.

Theorem 15. *For fixed s non-singular, with $0 < \sigma < 2$, and $y \longrightarrow \infty$, we have*

$$\eta(z, s) = y^s + \mathbf{c}(s)y^{1-s} + O(1).$$

Proof. By definition,

$$\begin{aligned}\eta_s &= [I + \omega(I + \omega Q(s))B(s)]\theta_s \\ &= \theta_s + \omega b_s * \theta_s + \omega^2 q_s * b_s * \theta_s.\end{aligned}$$

We know that $B(s) = M + N^B(s)$, and we recall that M contains cosine terms in its series expansion, while θ_s is independent of x. Consequently

$$M\theta_s = 0 \quad \text{and} \quad B(s)\theta_s = N^B(s)\theta_s$$

by the orthogonality of cosine and the constant functions. Since $N(s)$ is of type $\mathfrak{B}_{-1}$, and θ_s is of type max $(\sigma, 1 - \sigma) < 2$, we conclude that $N^B(s)\theta_s$ is of type $\mathfrak{B}_{-1}$, and in particular is bounded. This accounts for the second term $B(s)\theta_s$ in our sum. For the third term, we use again that $B(s)\theta_s = N^B(s)\theta_s$ is in $\mathfrak{B}_{-1}$. By Lemma 2 of §9, we conclude that

$$Q(s)B(s)\theta_s = \frac{1}{2s-1}y^{1-s}\int_F \int_F \theta(z, s)b(z, z'; s)\theta(z', s)\, dz\, dz' + O(1).$$

This proves our theorem.

Theorem 16. *Let s be non-singular. If $0 < \sigma < 1$, there is one and only one solution to the equation*

$$\omega(s)R(\kappa)\eta = \eta$$

having the asymptotic behavior for fixed s and $y \to \infty$

$$\eta(z) = y^s + \mathbf{c}(s)y^{1-s} + O(1).$$

If $\sigma > 1$, *there is one and only one solution of the same equation having the asymptotic behavior*

$$\eta(z) = y^s + O(1).$$

Proof. The existence will be proved later, and we deal here only with the uniqueness. Take first the statement relating to the interval $0 < \sigma < 1$. If η_1, η_2 are two solutions of the given eigenvector equation for $R(\kappa)$, then their difference ψ is bounded, so in $L^2(\Gamma\backslash\mathfrak{H}) = H$, and is also a solution of the equation. By definition, this means that s is singular, unless $\psi = 0$, thus proving the uniqueness. The same argument works in the other case for $\sigma > 1$.

Corollary. *We have, for* $\sigma > 1$,

$$\eta(z, s) = \sum_{\gamma \in \Gamma_0\backslash\Gamma} \operatorname{Im}(\gamma z)^s.$$

Proof. The series on the right converges absolutely for $\sigma > 1$. The Laplace operator commutes with the action of $SL_2(\mathbf{R})$ on H, and hence the series on the right defines a function $\eta(z)$ such that

$$L\eta(z) = s(1 - s)\eta.$$

The term with $\gamma = 1$ yields y^s in the sum. All other terms, of the form

$$\frac{y^s}{|cz + d|^{2s}}, \quad \text{if } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad c \neq 0,$$

are bounded. The convergence of the series shows that the sum over the other γ yields a bounded function. We can therefore apply the uniqueness theorem, to conclude the proof.

Our function $\eta(z, s) = (I + W(s))\theta(z, s)$ which is defined in the strip $0 < \sigma < 2$ will now be shown to satisfy the eigenvector equation of the uniqueness thorem, thus showing that it coincides with the series expression for $\sigma > 1$. This equation will be the first of several formal properties of the Eisenstein functions, which follow directly from the corresponding formal properties of the Eisenstein operators, and are essentially resolvant relations.

Lemma. *For $\kappa > 3$ and $0 < \sigma < 2$ we have*

$$\omega(s)T(\kappa)\theta_s = \theta_s.$$

Proof. Freshman integration, and the integrals do converge.

From the lemma, we shall be able to replace ωT by the identity when applying it to the functions θ_s.

We have the operator relation

$$\omega R(I + W) = \omega T + W, \tag{1}$$

where as usual, $R = R(\kappa)$, $T = T(\kappa)$, and $W = W(s)$. This follows at once from the definition

$$W = \omega(I + \omega Q)B,$$

using $R = T + V$ and

$$B = V + \omega V(I + \omega Q)B.$$

Indeed, we get

$$\omega R(I + W) = \omega T + \omega(I + \omega Q)B = \omega T + W.$$

Therefore by the lemma,

$$\omega R(I + W)\theta_s = (I + W)\theta_s,$$

which gives us the first relation

ES 1. $$\omega(s)R(\kappa)\eta_s = \eta_s.$$

In words, the Eisenstein function is an eigenvalue of the resolvant of the Laplacian.

ES 2. $$\eta(\gamma z, s) = \eta(z, s) \quad \text{for} \quad \gamma \in \Gamma.$$

Proof. By **ES 1**,

$$\begin{aligned}\eta(\gamma z, s) &= \omega(s)R(\kappa)\eta(z, s) \\ &= \omega(s)\int_F r(\gamma z, z'; \kappa)\eta(z', s)\, dz' \\ &= \eta(z, s)\end{aligned}$$

because $r(\gamma z, z'; \kappa) = r(z, z'; \kappa)$.

ES 3. $$L\eta(z, s) = s(1 - s)\eta(z, s).$$

Proof. It suffices to prove this weakly because we can then apply the regularity theorem for elliptic operators. As before, let $M_s = L - s(1 - s)$. For $f \in C_c^\infty(\Gamma\backslash\mathfrak{H})$ we get

$$\omega(s)\int_F \eta(z,s)\, M_s f(z)\, dz = \int_F\int_F r(z, z'; \kappa)\,\eta(z',s)\, M_s f(z)\, dz\, dz'$$

$$= \int_F \eta(z', s)\, dz' \int_{\mathfrak{H}} r_{\mathfrak{H}}(z, z'; \kappa)\, M_s f(z)\, dz.$$

Use $M_s = M_\kappa - \omega(s)$, Lemma 1, §3, and Theorem 2, §3. Our last expression is

$$= \int_F \eta(z',s)\, M_\kappa \int_{\mathfrak{H}} r_{\mathfrak{H}}(z,z';\kappa) f(z)\; dz\, dz'$$

$$- \omega(s)\int_F\int_F \eta(z',s)\, r(z,z';\kappa) f(z)\, dz\, dz'$$

$$= \int_F \eta(z',s) f(z')\, dz' - \int_F \eta(z,s) f(z)\; dz\, dz'$$

$$= 0.$$

This proves the desired property.

ES 4. $$\overline{\eta(z, s)} = \eta(z, \bar{s})$$

This is clear from the fact that all functions and kernels we have considered satisfy the analogous property.

We shall now see how the various resolvant equations from the previous section can be interpreted as relations concerning the Eisenstein functions. Resolvant equations will be interpreted either as operator relations or as kernel relations.

First we consider the resolvant relations for $R(s)$. We know that in the domain $\sigma, \sigma' > \frac{1}{2}$ we have

$$R - R' = (\omega - \omega')RR',$$

where we abbreviate $R = R(s)$, $R' = R(s')$, $\omega' = \omega(s')$, and below, $B = B(s)$, $B' = B(s')$, etc. We recall two formulas for the q-kernel which will be used constantly in what follows

Q 5. $$Q - Q' = (\omega - \omega')QQ' \qquad \textit{for} \quad \sigma, \sigma' > \tfrac{1}{2}.$$

q 4. $$q_s - q_{1-s} = \frac{1}{2s - 1}\,\theta_s\theta'_{1-s}$$

where $\theta_s = \theta_s(z)$ and $\theta'_{1-s} = \theta_{1-s}(z')$. The two variables (z, z') correspond to the variables in $q_s(z, z')$ and $q_{1-s}(z, z')$.

ES 5. *For $0 < \sigma < 1$ and s, $1 - s$ non-singular, we have*

$$r(z, z'; s) - r(z, z'; 1 - s) = \frac{1}{2s - 1}\eta(z, s)\eta(z', 1 - s).$$

Proof. Using the relation

$$R = Q + (I + \omega Q)B(I + \omega Q),$$

We get

$$\begin{aligned} R - R' = QQ' &+ (I + \omega Q)B(I + \omega Q)Q' \\ &+ Q(I + \omega' Q')B'(I + \omega' Q') \\ &+ (I + \omega Q)B(I + \omega Q)(I + \omega' Q')B'(I + \omega' Q'). \end{aligned}$$

Observe that a word made up with B's and Q's in which the B's and Q's alternate makes sense for $0 < \sigma < 2$, as operators $\mathfrak{B}_{-1} \to \mathfrak{B}_1$, while if we take a product QQ', say, then it makes sense only for $\sigma, \sigma' > \frac{1}{2}$ as operator $\mathfrak{H} \to \mathfrak{H}$. Using **Q 5**, we can express the above relation with a term not containing $(\omega - \omega')$ as a factor, and a term Z having $(\omega - \omega')$ as a factor, and only alternating products as mentioned. Thus

$$\begin{aligned} R - R' = Q - Q' + (I + \omega Q)\omega B(Q - Q')\omega' B'(I + \omega' Q') \\ + (\omega - \omega')Z. \end{aligned}$$

In the expression on the right, we can then substitute $s' = 1 - s$, in which case

$$\omega(s) - \omega(1 - s) = 0,$$

and the term with factor $(\omega - \omega')$ vanishes. As for the other terms, we read them as a kernel relation, and use **Q 4**, valid for all s. By the symmetry of $q(z, z'; s)$ and $b(z, z'; s)$ our desired formula drops out. Note that convolution of kernels on the right corresponds to composition of operators in the opposite direction.

The further relations concerning the Eisenstein functions will be analogous to relations satisfied by $\theta(z, s)$. We recall that

$$\theta(z, s) = y^s + c(s)y^{1-s}$$

where

$$c(s) = a^{2s-1}\frac{s - \kappa}{s + \kappa - 1}.$$

We verify trivially that

$$\boxed{\begin{gathered} c(s)c(1-s) = 1, \\ \theta(z, s) = \theta(z, 1-s)c(s). \end{gathered}}$$

We shall prove analogous formulas for $\eta(z, s)$ and $\mathbf{c}(s)$, namely for

$$0 < \sigma < 1,$$

ES 6. $\qquad \mathbf{c}(s)\mathbf{c}(1-s) = 1.$

ES 7. $\qquad \eta(z, s) = \eta(z, 1-s)\mathbf{c}(s).$

Proofs. Let b_s as before be the kernel for the operator $B(s)$ and abbreviate

$$\rho(s) = \frac{\omega(s)^2}{2s-1}\,\theta_s * b_s * \theta_s$$

for s in the strip $0 < \sigma < 2$. We shall prove that

$$\rho(s)c(1-s) - \rho(1-s)c(s) = -\rho(s)\rho(1-s),$$

from which **ES 6** follows at once, because $\mathbf{c}(s) = c(s) + \rho(s)$.

The proof for the above relation is an immediate consequence of the resolvant relation for $B(s)$, expressed in terms of the kernel, Theorem 11 of §10, namely

$$\begin{aligned} b_s - b_{1-s} &= \omega(s)^2 b_s * (q_s - q_{1-s}) * b_{1-s} \\ &= \frac{\omega(s)^2}{2s-1}(b_s * \theta_s)(b_{1-s} * \theta_{1-s})' \end{aligned}$$

with the obvious notation that $(b_{1-s} * \theta_{1-s})'$ is a function of z', if (z, z') are the variables of the kernel $b_s(z, z')$. Convolve the above relation with θ_s on the left and θ_{1-s} on the right. Use

$$\theta_s = \theta_{1-s}c(s) \qquad \text{and} \qquad c(1-s) = c(s)^{-1}$$

to get

$$\theta_s * b_s * \theta_s c(1-s) - c(s)\theta_{1-s} * b_{1-s} * \theta_{1-s} = \rho(s)\rho(1-s)\,\frac{2s-1}{\omega(s)^2}\,.$$

The desired relation between $\rho(s)$ and $\rho(1-s)$ drops out from the definitions.

The proof for **ES 7** can be given following the same pattern using the resolvant equation for b_s. On the other hand, at this point it drops out of the uniqueness theorem. Indeed, we have for $0 < \sigma < 1$, by Theorem 15,

$$\eta_s(z) = y^s + \mathbf{c}(s)y^{1-s} + O(1)$$

$$\eta_{1-s}(z) = \mathbf{c}(1-s)y^s + y^{1-s} + O(1).$$

Multiply this second equation by $\mathbf{c}(s)$, use **ES 6** and the uniqueness theorem to get **ES 7**.

§13. THE CONTINUOUS PART OF THE SPECTRUM

Let $H_0 = L^2(0, \infty)$, with measure $\frac{1}{2\pi}\,dt$, where dt is Lebesgue measure. We shall prove that the orthogonal complement of the eigenspace of A in $H = L^2(\Gamma\backslash\mathfrak{H})$ is unitarily isomorphic to H_0, by specifically exhibiting a kernel achieving the isomorphism. This kernel is none other than the Eisenstein function $\eta(z, \frac{1}{2} + it)$ on the product space

$$\Gamma\backslash\mathfrak{H} \times [0, \infty).$$

We define the operator A_0 on H_0 by multiplication with $\frac{1}{4} + t^2$, i.e.

$$A_0\xi(t) = (\tfrac{1}{4} + t^2)\xi(t)$$

for those functions ξ such that the product is in $\mathfrak{L}^2$. Then A_0 is self-adjoint. We let the **Eisenstein transform** E be defined on the space of *bounded functions* on $\Gamma\backslash\mathfrak{H}$ by

$$Ef(t) = \int_F \eta(z, \tfrac{1}{2} + it)f(z)\,dz.$$

Theorem 17.

i) *Under the Eisenstein transform, the eigenspace of A in H goes to* 0.
ii) *The Eisenstein transform can be extended to a unitary isomorphism between the orthogonal complement of this eigenspace and H_0.*
iii) *Under the Eisenstein transform, A is carried to A_0, that is*

$$ER(\kappa) = R_0(\kappa)E, \qquad EA = A_0E,$$

where $R(\kappa)$ is the resolvant of A in H with $\kappa > 3$, and

$$R_0(s) = (A_0 - s(1-s)I_0)^{-1}.$$

Proof. Let $f \in C_c(\Gamma\backslash\mathfrak{H})$. Let $[a, b]$ be an interval on $(0, \infty)$ such that the interval $\frac{1}{2} + it$, $a \leqslant t \leqslant b$, does not contain any singular point. Then

$$\frac{1}{2\pi}\int_a^b |Ef(t)|^2\,dt$$

$$= \frac{1}{2\pi}\int_a^b \int_F \int_F \eta\left(z, \frac{1}{2} + it\right)\eta\left(z', \frac{1}{2} - it\right) f(z)\overline{f(z')}\,dz\,dz'\,dt$$

$$= \frac{1}{2\pi}\int_a^b 2it\,dt \int_F \int_F \left[r\left(z, z'; \frac{1}{2} + it\right) - r\left(z, z'; \frac{1}{2} - it\right)\right] f(z)\overline{f(z')}\,dz\,dz'$$

by **E 5**, and $2s - 1 = 2it$ for $s = \frac{1}{2} + it$. We have

$$\lambda = s(1 - s) = \tfrac{1}{4} + t^2 \qquad \text{and} \qquad d\lambda = 2t\,dt.$$

It is now convenient to express the above kernels of the resolvant in terms of the more usual λ because we are going to use the spectral theorem for unbounded operators. We write

$$r(z, z'; s) = \rho(z, z'; \lambda).$$

As t ranges over the interval $a \leqslant t \leqslant b$, the variable λ ranges over the interval

$$\tfrac{1}{4} + a^2 \leqslant \lambda \leqslant \tfrac{1}{4} + b^2.$$

Values of $\sigma + it$ with $\sigma > \frac{1}{2}$ correspond to values of λ with negative imaginary part, while values of $\sigma - it$ with $\sigma > \frac{1}{2}$ correspond to values of λ with positive imaginary part. Consequently our last expression is equal to

$$\frac{1}{2\pi}\int_{\lambda_a}^{\lambda_b} i\,d\lambda \int_F \int_F [\rho(z, z'; \lambda -) - \rho(z, z'; \lambda +)] f(z)\overline{f(z')}\,dz\,dz'$$

where

$$\rho(z, z'; \lambda -) = \lim_{\epsilon \to 0} \rho(z, z'; \lambda - i\epsilon) \qquad \text{and} \qquad \rho(z, z'; \lambda +) = \lim_{\epsilon \to 0} \rho(z, z'; \lambda + i\epsilon).$$

We may also write this expression as the limit for $\epsilon \to 0$ of

$$\frac{1}{2\pi i}\int_{\lambda_a}^{\lambda_b} d\lambda \int_F \int_F [\rho(z, z'; \lambda + i\epsilon) - \rho(z, z'; \lambda - i\epsilon)] f(z)\overline{f(z')}\,dz\,dz'.$$

For $\sigma \pm it$ with $\sigma > \frac{1}{2}$, the kernel $r(z,z';s)$ represents the resolvant. Consequently this last expression may be rewritten

$$\frac{1}{2\pi i} \lim_{\epsilon \to 0} \int_{\lambda_a}^{\lambda_b} \langle R(\lambda + i\epsilon) - R(\lambda - i\epsilon)f, f\rangle d\lambda,$$

which by functional analysis (Appendix 2) is equal to

$$\int_{\lambda_a}^{\lambda_b} d\mu_{f,f}(\lambda),$$

where $\mu_{f,f}$ is the spectral measure of A.

We divide the half line $s = \frac{1}{2} + it$, $t > 0$, into segments not containing singular points, sum the integrals over these segments, as the end points approach the singular points. We know by Appendix 2 that if P is the projection on the eigenspace of A, then the limit is equal to

$$\langle (I - P)f, f\rangle.$$

Thus we see that Ef is in H_0, that eigenvectors of A are transformed to 0 by the Eisenstein transform, and that if f is orthogonal to the eigenvectors of A, then

$$\|f\|_2^2 = \|Ef\|_2^2.$$

This proves (i) and (ii) of our theorem.

Formula **ES 1**, which says that the Eisenstein functions are eigenvectors of $R(\kappa)$ with eigenvalue $\omega(s)^{-1}$ is now merely a reformulation of (iii), or to be precise, it says that

$$EA \subset A_0 E$$

where the symbol $\subset$ has the usual meaning in the theory of unbounded operators, i.e. the domain of EA is contained in the domain of A_0E and A_0E restricts to EA on this domain.

It now follows that for any bounded measureable function h on the real line, we have

$$Eh(A) = h(A_0)E.$$

(We are assuming that the reader is familiar with the spectral theory of unbounded operators as in Appendix 2.)

There remains only to be proved that the image under E of the space of bounded functions on $\Gamma\backslash\mathfrak{H}$ is dense in H_0. It will follow that the extension of

E to H is an isometry between $(I - P)H$ and all of H_0. This denseness comes from the fact that we have considerable freedom in choosing both f and h.

Pick some value of $t_0 > 0$. We know that

$$\eta(z, s) = y^s + \mathbf{c}(s)y^{1-s} + O(1).$$

Let f be a function which is equal to the complex conjugate of $\eta(z, \frac{1}{2} + it_0)$ for z near some large value of iy, and then drops to 0 rapidly. Then

$$Ef(t) = \int_F \eta(z, s)f(z)\, dz$$

does not vanish at t_0, in fact has a positive value at t_0. Now select h so that $h(\frac{1}{4} + t^2)$ is equal to $1/Ef(t)$ for t near t_0 and is equal to 0 otherwise. Then

$$h(A_0)Ef = Ef(A)f$$

is the characteristic function of a small interval around t_0. This proves that the image of E is dense, and concludes the proof of our theorem.

§14. SEVERAL CUSPS

We have acted throughout this chapter as if $\Gamma = SL_2(\mathbf{Z})$. This was not essential, and the time has come to make the appropriate comments indicating which additional features show up in the general case of an arbitrary discrete subgroup Γ such that $\Gamma\backslash\mathfrak{H}$ has finite volume. We assume $-1 \in \Gamma$. We let $G = SL_2(\mathbf{R})$.

As already mentioned in Chapter XII, there is a fundamental domain F having the following properties

i) F lies in the strip $-x_0 \leqslant x \leqslant x_0$ for some x_0.

ii) There is a positive number a such that F is the union

$$F = F_0 \cup \bigcup_{\alpha=1}^{n} F_\alpha$$

of pieces, where F_0 is compact with piecewise smooth boundary, and F_α is the image of the upper strip $\mathbf{F}_1$ defined by the inequalities

$$-\frac{1}{2} \leqslant x \leqslant \frac{1}{2} \qquad \text{and} \qquad a \leqslant y,$$

under the mapping $z \mapsto g_\alpha z$ for some $g_\alpha \in SL_2(\mathbf{R}) = G$.

Each function f on F then has $n+1$ components instead of 2 components, $f_0, f_1, \ldots, f_n$ defined by

$$\begin{aligned} f_0(z) &= f(z) && \text{for} \quad z \in F_0 \\ f_\alpha(z) &= f(g_\alpha z) && \text{for} \quad z \in \mathbf{F}_1. \end{aligned}$$

Similarly, a kernel $k(z, z')$ on $F \times F$ has components $k_{\alpha\beta}(z, z')$ defined by

$$\begin{aligned} k_{00}(z, z') &= k(z, z') && \text{for} \quad z, z' \in F_0, \\ k_{0\alpha}(z, z') &= k(z, g_\alpha z') && \text{for} \quad z \in F_0 \quad \text{and} \quad z' \in \mathbf{F}_1, \end{aligned}$$

etc.

We let, for $\alpha = 1, \ldots, n$,

$$\Gamma_\alpha = g_\alpha^{-1}\Gamma g_\alpha, \qquad \Delta_\alpha = \Gamma g_\alpha, \qquad \Delta_{\alpha\beta} = g_\beta^{-1}\Gamma g_\alpha \qquad \text{for} \quad \alpha \neq \beta.$$

The subset of elements $\gamma \in \Gamma_\alpha$ for which

$$\gamma = \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}$$

is denoted by Γ_0. It *is* our old group Γ_0.

Essentially the first time that Λ entered in the computations was in proving the symmetry of the Laplace operator on $\Gamma\backslash\mathfrak{H}$, that is Lemma 1 of §4. It is typical of what follows that there is no change to be made in the general case other than replacing f^2 by f_α^2 in the right-hand side of (*), and summing over α. Thus each cusp contributes one term on the upper part of the strip.

The proof of the convergence of the series

$$\sum_{\gamma \in \Gamma} \frac{1}{[1 + u(z, \gamma z')]^\sigma}$$

in Lemma 1, §5, made no use of properties of Γ other than it operates discretely.

Slightly more serious is a fact needed for Lemma 9, §7, giving the estimate

$$\sum_{\gamma \notin \Gamma_0} \frac{1}{[1 + u(z, \gamma z')]^\sigma} \ll (yy')^{2+\epsilon-\sigma}.$$

One needs to know that in any subset of elements

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

in Γ_α, Δ_α, or $\Delta_{\alpha\beta}$ with $c \neq 0$ then $|c|$ is bounded from below. Faddeev refers to Petersson [Pe 2], and so do I.

All the dealings with cuspidal questions carried out for a single F_1 apply verbatim to the general case, since the elements g_α carry the standard upper strip to the corresponding part of the fundamental domain. We merely specify that certain kernels have only "diagonal" components, that is their components are 0 except for

$$t_{\alpha\alpha}(z, z'; s) = t(y, y'; s)$$

$$m_{\alpha\alpha}(z, z'; s) = m(z, z'; s)$$

$$q_{\alpha\alpha}(z, z'; s) = q(y, y'; s).$$

Until we come to §12, nothing more need to be said and no changes need be made except for replacing F_1 by $\mathbf{F}_1$.

In §12 we have to insert a few more indices. For each $\beta = 1, \ldots, n$ we let $\theta^\beta(z, s)$ be the piecewise smooth function on F whose components are given by

$$\theta^\beta_\alpha(z, s) = \delta_{\alpha\beta}\theta(y, s), \quad \theta_0(z, s) = 0.$$

Here $\delta_{\alpha\beta}$ is the usual Kronecker symbol. For each β we obtain an Eisenstein function

$$\eta^\beta(z, s) = \theta^\beta(z, s) + W(s)\theta^\beta(z, s),$$

where $W(s)$ is already defined in terms of its components. Theorem 14 applies to each Eisenstein function η^β. In the proof, one merely replaces η and θ by η^β and θ^β. Each η^β is a vertical vector, and we get an $(n + 1) \times n$ *matrix* η:

$$\begin{pmatrix} \eta^1_0 & \cdots & \eta^n_0 \\ \vdots & & \vdots \\ \eta^1_n & \cdots & \eta^n_n \end{pmatrix} = \eta,$$

formed by the component functions η^β_α, $\alpha = 0, \ldots, n$; $\beta = 1, \ldots, n$.

The function $\mathbf{c}(s)$ now becomes the **scattering matrix** whose components are

$$\mathbf{c}_{\alpha\beta}(s) = c(s)\delta_{\alpha\beta} + \frac{\omega^2(s)}{2s - 1}\ \theta^\alpha_s * n_s * \theta^\beta_s.$$

[We could write b_s instead of n_s because in the expression

$$B(s) = M + N^B(s),$$

the operator M annihilates the θ_s^β, i.e.

$$B(s)\theta_s^\beta = N^B(s)\theta_s^\beta.]$$

Theorem 15 giving the asymptotic behavior of $\eta(z, s)$ is unchanged, except that η is now the Eisenstein matrix as above. The term y^s is to be interpreted as $y^s I$, where I is the unit matrix.

In the corollary of Theorem 16, giving the series expression for the Eisenstein functions, one has to write

$$\eta^\beta(g_\beta z, s) = \sum_{\gamma\in\Gamma_0\backslash\Gamma_\beta} \mathrm{Im}(\gamma z)^s.$$

The formal properties **ES 1** through **ES 4** hold without change, for each component vertical vector η^β. The resolvant relation **ES 5** requires a summation,

ES 5. $$r(z,z';s)-r(z,z';1-s)=\frac{1}{2s-1}\sum_{\beta=1}^{n}\eta^\beta(z,s)\eta^\beta(z',1-s).$$

The Eisenstein properties **ES 6** and **ES 7** remain valid as stated, but with the matricial interpretation. In other words, $\mathbf{c}(s)$ is the inverse matrix of $\mathbf{c}(1-s)$. In **ES 7**, the matrix $\mathbf{c}(s)$ has been placed appropriately on the right of the matrix $\eta(z, 1-s)$.

In the spectral decomposition of the continuous spectrum, we have to introduce the components by letting H_0 be the Hilbert space of vector functions

$$\xi(t) = (\xi_1(t), \dots, \xi_n(t))$$

such that each ξ_α is in $\mathfrak{L}^2(0, \infty)$, and where the scalar product is given by

$$\langle \xi, \zeta \rangle = \frac{1}{2\pi}\sum_{\alpha=1}^{n}\int_0^\infty \xi_\alpha(t)\overline{\zeta_\alpha(t)}\, dt.$$

The operator A_0 is still multiplication by the function $\frac{1}{4}+t^2$ on each component. The Eisenstein transform Ef is defined componentwise,

$$E^\beta f(t) = \int_F \eta^\beta(z, \tfrac{1}{2}+it)f(z)\, dz$$

and Theorem 17 holds without further change.

Appendices

Appendix 1

Bounded Hermitian Operators and Schur's Lemma

§1. CONTINUOUS FUNCTIONS OF OPERATORS

We assume known only the most elementary facts about Hilbert space, and we reprove the next auxiliary result about a Hilbert space H.

Lemma. *Let A be an operator, and c a number such that*

$$|\langle Ax, x\rangle| \leqslant c|x|^2$$

for all $x \in H$. Then for all x, y we have

$$|\langle Ax, y\rangle| + |\langle x, Ay\rangle| \leqslant 2c|x||y|.$$

Proof. By the polarization identity,

$$2|\langle Ax, y\rangle + \langle Ay, x\rangle| \leqslant c|x+y|^2 + c|x-y|^2 = 2c(|x|^2 + |y|^2).$$

Hence

$$|\langle Ax, y\rangle + \langle Ay, x\rangle| \leqslant c(|x|^2 + |y|^2).$$

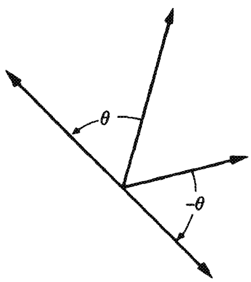

Figure 1

We multiply y by $e^{i\theta}$ and thus get on the left-hand side

$$|e^{-i\theta}\langle Ax, y\rangle + e^{i\theta}\langle Ay, x\rangle|.$$

The right-hand side remains unchanged, and for suitable θ, the left-hand side becomes

$$|\langle Ax, y\rangle| + |\langle Ay, x\rangle|.$$

(In other words, we are lining up two complex numbers by rotating one by θ and the other by $-\theta$.) Next we replace x by tx and y by y/t for t real and $t > 0$. Then the left-hand side remains unchanged, while the right-hand side becomes

$$g(t) = t^2|x|^2 + \frac{1}{t^2}|y|^2.$$

The point at which $g'(t) = 0$ is the unique minimum, and at this point t_0 we find that

$$g(t_0) = |x|\,|y|.$$

This proves our lemma.

Theorem 1. *Let A be a hermitian operator. Then $|A|$ is the greatest lower bound of all values c such that*

$$|\langle Ax, x\rangle| \leqslant c|x|^2$$

for all x, or equivalently, the sup of all values $|\langle Ax, x\rangle|$ taken for x on the unit sphere in H.

Proof. When A is hermitian we obtain

$$|\langle Ax, y\rangle| \leqslant c|x|\,|y|$$

for all $x, y \in H$, so that we get $|A| \leqslant c$ in the lemma. On the other hand, $c = |A|$ is certainly a possible value for c by the Schwarz inequality. This proves our theorem.

Theorem 1 allows us to define an ordering in the space of hermitian operators. If A is hermitian, we define $A \geqslant O$ and say that A is **positive** if $\langle Ax, x\rangle \geqslant 0$ for all $x \in H$. If A, B are hermitian we define $A \geqslant B$ if $A - B \geqslant O$. This is indeed an ordering, the usual rules hold: If $A_1 \geqslant B_1$ and $A_2 \geqslant B_2$, then

$$A_1 + A_2 \geqslant B_1 + B_2.$$

If c is a real number $\geqslant 0$ and $A \geqslant O$, then $cA \geqslant O$. So far, however, we say nothing about a product of positive hermitian operators AB, even if $AB = BA$. We shall deal with this question later.

Let c be a bound for A. Then $|\langle Ax, x\rangle| \leqslant c|x|^2$ and consequently

$$-cI \leqslant A \leqslant cI.$$

For simplicity, if α is real, we sometimes write $\alpha \leqslant A$ instead of $\alpha I \leqslant A$, and similarly we write $A \leqslant \beta$ instead of $A \leqslant \beta I$. If we let

$$\alpha = \inf_{|x|=1} \langle Ax, x\rangle \quad \text{and} \quad \beta = \sup_{|x|=1} \langle Ax, x\rangle,$$

then we have

$$\alpha \leqslant A \leqslant \beta,$$

and from Theorem 1,

$$|A| = \max(|\alpha|, |\beta|).$$

Let p be a polynomial with real coefficients, and let A be a hermitian operator. Write

$$p(t) = a_n t^n + \cdots + a_0.$$

We define

$$p(A) = a_n A^n + \cdots + a_0 I.$$

We let $\mathbf{R}[A]$ be the algebra generated over $\mathbf{R}$ by A, that is the algebra of all operators $p(A)$, where $p(t) \in \mathbf{R}[t]$. We wish to investigate the closure of $\mathbf{R}[A]$ in the (real) Banach space of all operators. We shall show how to represent this closure as a ring of continuous functions on some compact subset of the reals. First, we observe that the hermitian operators form a closed subspace of $\operatorname{End}(H)$, and that $\overline{\mathbf{R}[A]}$ is a closed subspace of the space of hermitian operators.

We shall prove that if p is a real polynomial which takes on positive values on the interval $[\alpha, \beta]$, then $p(A)$ is a positive operator. For this we need a purely algebraic lemma.

Lemma 1. *Let p be a real polynomial such that $p(t) \geqslant 0$ for all $t \in [\alpha, \beta]$. Then we can express p in the form*

$$p(t) = c\left[\sum Q_i(t)^2 + \sum (t-\alpha) Q_j(t)^2 + \sum (\beta - t) Q_k(t)^2\right]$$

where Q_i, Q_j, Q_k are real polynomials, and $c \geqslant 0$.

Proof. We first factor p into linear and irreducible quadratic factors over the real numbers. If p has a root γ such that $\alpha < \gamma < \beta$, then the multiplicity of γ is even (otherwise p changes sign near γ, which is impossible), and then $(t - \gamma)$ occurs in an even power. If a root γ is $\leqslant \alpha$ we have a linear factor $t - \gamma$ which we write

$$t - \gamma = (t - \alpha) + (\alpha - \gamma)$$

and note that $\alpha - \gamma$ is a real square. If γ is a root $\geqslant \beta$, then we write the linear factor as

$$\gamma - t = (\gamma - \beta) + (\beta - t)$$

and note that $\gamma - \beta$ is a real square. In a factorization of p we can take the factors to be of type $(t - \gamma)^{2m(\gamma)}$ if γ is root such that $\alpha < \gamma < \beta$, and otherwise to be of type $t - \gamma$ or $\gamma - t$ according as $\gamma < \alpha$ or $\gamma > \beta$. The quadratic factors are of type $(t - a)^2 + b^2$. The constant c (which can be taken as a constant factor) is then $\geqslant 0$ since p is positive on the interval. Multiplying out all these factors, and noting that a sum of squares times a sum of squares is a sum of squares, we conclude that p has an expression as stated in the lemma, except that there still appear terms of type

$$(t - \alpha)(\beta - t)Q(t)^2$$

where Q is a real polynomial. However, such terms can be reduced to terms of the other types, by using the identity

$$(t - \alpha)(\beta - t) = \frac{(t - \alpha)^2(\beta - t) + (t - \alpha)(\beta - t)^2}{\beta - \alpha}.$$

This proves our lemma.

Now to study $\overline{\mathbf{R}[A]}$, we observe that the map

$$p \longmapsto p(A)$$

is a ring-homomorphism of $\mathbf{R}[t]$ onto the ring $\mathbf{R}[A]$. Furthermore, if B, C are hermitian operators such the $BC = CB$ and $B \geqslant O$, then trivially, BC^2 is positive because

$$\langle BC^2x, x\rangle = \langle CBCx, x\rangle = \langle BCx, Cx\rangle \geqslant 0.$$

The sum of two positive hermitian operators is positive. Hence from the expression of p in the lemma, we obtain:

Lemma 2. *If p is positive on $[\alpha, \beta]$, then $p(A)$ is a positive operator. If p, q are polynomials such that $p \leqslant q$ on $[\alpha, \beta]$, then $p(A) \leqslant q(A)$. Finally,*

$$|p(A)| \leqslant \|p\|,$$

the sup norm being taken on $[\alpha, \beta]$.

Proof. The first assertion comes from the remarks preceding our lemma. The second follows at once by considering $q - p$. Finally, if we let

$$q(t) = \|p\| \pm p(t),$$

then $q \geqslant 0$ on $[\alpha, \beta]$ and hence $q(A) \geqslant O$, whence the last assertion follows from Theorem 1.

We conclude that the map

$$p \longmapsto p(A)$$

is a continuous linear map from the space of polynomial functions on $[\alpha, \beta]$ into $\mathbf{R}[A]$. By the linear extension theorem, we can extend this map to the Banach space of continuous functions on $[\alpha, \beta]$ by continuity, and thus we can define $f(A)$ for any continuous function f on $[\alpha, \beta]$, by the Stone-Weierstrass theorem. If $\{p_n\}$ is a sequence of polynomials converging uniformly to f on $[\alpha, \beta]$, then by definition,

$$f(A) = \lim p_n(A).$$

Furthermore, again by continuity, we have

$$|f(A)| \leqslant \|f\|,$$

the sup norm being taken on $[\alpha, \beta]$. If $p_n \longrightarrow f$ and $q_n \longrightarrow g$, then $p_n q_n \longrightarrow fg$. Hence we obtain $(fg)(A) = f(A)g(A)$ for any continuous functions f, g. In other words, our map is also a ring homomorphism.

Theorem 2. *If $A \geqslant O$, then there exists $B \in \overline{\mathbf{R}[A]}$ such that $B^2 = A$. The product of two commuting positive hermitian operators is again positive.*

Proof. The continuous function $t^{1/2}$ maps on a square root of A in $\overline{\mathbf{R}[A]}$, and it is clear that any element of $\overline{\mathbf{R}[A]}$ commutes with A. If A, C commute and we write $A = B^2$ with B in $\overline{\mathbf{R}[A]}$, then B and C also commute because C commutes with $p(A)$ for all real polynomials p, and hence C commutes with all elements of $\overline{\mathbf{R}[A]}$. But as we have seen, if $C \geqslant O$, then $B^2C \geqslant O$. This proves our theorem.

The kernel of our map $f \longmapsto f(A)$ is a closed ideal in the ring of continuous functions on $[\alpha, \beta]$. We forget for a moment the usual definition of the spectrum, and here define the **spectrum** $\sigma(A)$ to be the closed set of zeros of this ideal.

If f is any continuous function on $\sigma(A)$, we extend f to a continuous function on $[\alpha, \beta]$ having the same sup norm, say f_1, and define

$$f(A) = f_1(A).$$

If g is another extension of f to $[\alpha, \beta]$, then $g - f_1$ vanishes on $\sigma(A)$, and hence $g(A) = f_1(A)$. Hence $f(A)$ is well defined, independently of the particular extension of f to $[\alpha, \beta]$. We denote by $\| \; \|_A$ the sup norm with respect to $\sigma(A)$, thus

$$\|f\|_A = \sup_{t \in \sigma(A)} |f(t)|.$$

We then obtain a ring homomorphism from the ring of continuous functions on $\sigma(A)$ into $\overline{\mathbf{R}[A]}$, and we have

$$|f(A)| \leqslant \|f\|_A.$$

We now state the ***spectral theorem.***

Theorem 3. *The map $f \mapsto f(A)$ is a Banach-isomorphism from the algebra of continuous functions on $\sigma(A)$ onto the Banach algebra $\overline{\mathbf{R}[A]}$. A continuous function f is $\geqslant 0$ on $\sigma(A)$ if and only if $f(A) \geqslant O$.*

Proof. We had derived the norm inequality previously from the positivity statement. We do this again in the opposite direction. Thus we assume first that $f(A) \geqslant O$ and prove that f is $\geqslant 0$ on the spectrum of A. Assume that this is not the case. Then f is negative at some point c of the spectrum. Let g be a continuous function whose graph is that in Fig. 2.

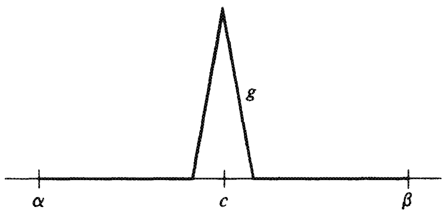

Figure 2

Thus g is $\geqslant 0$, and has a positive peak at c. Then fg is $\leqslant 0$ and fg is negative at the point c of the spectrum. Hence $-fg \geqslant 0$, and hence $-f(A)g(A) \geqslant O$. But $f(A) \geqslant O$ and $g(A) \geqslant O$, so that by Theorem 2 we also have

$$f(A)g(A) \geqslant O.$$

This implies that $f(A)g(A) = O$, which is impossible since fg does not vanish on the spectrum. We conclude that $f \geqslant 0$ on $\sigma(A)$, and in view of our previous result this proves the positivity statement of the theorem.

Now for the norm, let $b = |f(A)|$. Then $bI \pm f(A) \geqslant O$, whence

$$b \pm f(t) \geqslant 0$$

on the spectrum. This proves that

$$\|f\|_A \leqslant |f(A)|,$$

and hence a sequence $\{f_n(A)\}$ converges if and only if the sequence of continuous functions $\{f_n\}$ converges uniformly on the spectrum. This concludes the proof of the spectral theorem.

There remains to identify the spectrum as we have defined it in this section, and the general spectrum.

Corollary. *If A is hermitian, then the spectrum $\sigma(A)$ is equal to the set of complex numbers z such that $A - zI$ is not invertible.*

Proof. Let z be complex and such that $A - zI$ is not invertible. Then z is real, for otherwise, let

$$g(t) = (t - z)(t - \bar{z}).$$

Then $g(t) \neq 0$ on $\sigma(A)$, and hence $h(t) = 1/g(t)$ is its inverse. Then

$$h(A)(A - \bar{z}I)$$

would be an inverse for $A - zI$, a contradiction. This proves that z is real.

Let ξ be real and not in the spectrum $\sigma(A)$. Then $t - \xi$ is invertible on $\sigma(A)$, and hence so is $A - \xi I$.

Suppose that ξ is in the spectrum $\sigma(A)$. Let g be the continuous function whose graph is that in Fig. 3.

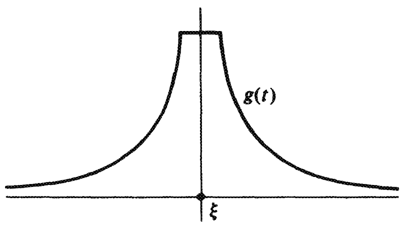

Figure 3

That is,

$$g(t)=\begin{cases} 1/|t-\xi| & \text{if } |t-\xi| \geqslant 1/N \\ N & \text{if } |t-\xi| \leqslant 1/N \end{cases}.$$

If $A-\xi I$ is invertible, let B be an inverse,

$$B(A-\xi I)=(A-\xi I)B=I.$$

Since $|(t-\xi)g(t)| \leqslant 1$ we get $|(A-\xi I)g(A)| \leqslant 1$, whence

$$|g(A)|=|B(A-\xi I)g(A)| \leqslant |B|.$$

But $g(t)$ has a large sup on the spectrum if we take N large, and hence $|g(A)|$ is equally large, a contradiction. Theorem 3 is proved.

The main idea to use the positivity to get the spectral theorem is due to F. Riesz. However, most treatments go from the positivity statement to an integral representation of A which we discuss in Appendix 2. Von Neumann always emphasized that it is much more efficient to prove at once the statement of Theorem 3, which suffices for many applications, and can be obtained quite simply from the positivity statement. In fact, the arguments used to derive Theorem 3 from the positivity statement are taken from a seminar of Von Neumann around 1950. The next theorem and its corollary are known as **Schur's lemma.**

Theorem 4. *Let S be a set of operators on the Hilbert space H, leaving no closed subspace invariant except $\{0\}$ and H itself. Let A be a hermitian operator such that $AB=BA$ for all $B \in S$. Then $A=cI$ for some real number c.*

Proof. It will suffice to prove that there is only one element in the spectrum of A. Suppose that there are two, $c_1 \neq c_2$. There exist continuous functions f, g on the spectrum such that neither is 0 on the spectrum, but fg is 0 on the spectrum. For instance, we can take for f, g the functions whose graphs are indicated in Fig. 4.

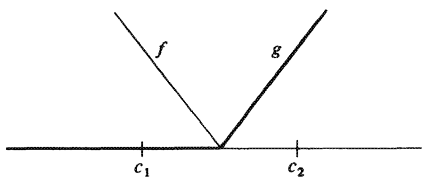

Figure 4

We have $f(A)B = Bf(A)$ for all $B \in S$ (because B commutes with real polynomials in A, hence with their limits). Hence $f(A)H$ is invariant under S because

$$Bf(A)H = f(A)BH \subset f(A)H.$$

Let F be the closure of $f(A)H$. Then $F \neq \{0\}$ because $f(A) \neq O$. Furthermore, $F \neq H$ because $g(A)f(A)H = \{0\}$ and hence $g(A)F = \{0\}$. Since F is invariant under S, we have a contradiction, thus proving our theorem.

Corollary. *Let S be a set of operators of the Hilbert space H, leaving no closed subspace invariant except $\{0\}$ and H itself. Let A be an operator such that $AA^* = A^*A$, $AT = TA$, and $A^*T = TA^*$ for all $T \in S$. Then $A = cI$ for some complex number c.*

Proof. Write $A = B + iC$ where B, C are hermitian and commute (e.g. $B = (A + A^*)/2$ and $C = (A - A^*)/2i$). Apply the theorem to each one of B and C to prove the corollary.

§2. PROJECTION FUNCTIONS OF OPERATORS

We need to extend the notion $f(A)$ to functions f which are not continuous, to include at least characteristic functions of intervals. We follow Riesz–Nagy more or less.

Lemma 1. *Let α be real, and let $\{A_n\}$ be a sequence of Hermitian operators such that $A_n \geqslant \alpha I$ for all n, and such that $A_n \geqslant A_{n+1}$. Given $v \in H$, the sequence $\{A_n v\}$ converges to an element of H. If we denote this element by Av, then $v \longmapsto Av$ is a bounded hermitian operator.*

Proof. From the inequality

$$\langle A_n v, v\rangle \geqslant \alpha\langle v, v\rangle$$

we conclude that $\langle A_n v, v\rangle$ converges, for each $v \in H$. Since

$$\langle A_n v, w\rangle = \tfrac{1}{2}\langle A_n(v+w), v+w\rangle - \tfrac{1}{2}\langle A_n(v-w), v-w\rangle,$$

it follows that $\langle A_n v, w\rangle$ converges for each pair of elements $v, w \in H$. Define

$$\lambda_v(w) = \lim_{n\to\infty} \langle A_n v, w\rangle.$$

Then λ_v is antilinear, and $|\langle A_n v, w\rangle| \leqslant C|v||w|$ for some C and all $v, w \in H$. Hence there exists an operator A such that

$$\langle Av, w\rangle = \lim_{n\to\infty} \langle A_n v, w\rangle.$$

Since $\langle A_n v, w\rangle = \langle v, A_n w\rangle$, it follows that A is hermitian.

Lemma 2. *Let f be a function on the spectrum of A, bounded from below, and which can be expressed as a pointwise convergent limit of a decreasing sequence of continuous functions, say $\{h_n\}$. Then*

$$\lim_{h\to\infty} h_n(A)$$

is independent of the sequence $\{h_n\}$.

Proof. Say $g_n(t)$ decreases also to $f(t)$. Given k, for large n we have

$$\max(g_n, h_k) \leqslant h_k + \epsilon,$$

so for all t we have $g_n(t) \leqslant h_k(t) + \epsilon$, and hence

$$g_n(A) \leqslant h_k(A) + \epsilon I.$$

This shows that

$$\lim g_n(A) \leqslant h_k(A) + \epsilon I,$$

and therefore that

$$\lim g_n(A) \leqslant \lim h_k(A) + \epsilon I.$$

This is true for all ϵ. Letting $\epsilon \to 0$ and using symmetry, we have proved our lemma.

From Lemma 2, we see that the association

$$f \mapsto f(A)$$

can be extended to the linear space generated by functions which can be obtained as limits from above of decreasing sequences of continuous functions, and are bounded from below. The map is additive, order preserving, and clearly multiplicative, i.e.

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

for f, g in this vector space.

The most important functions to which we apply this extension are characteristic functions like the function $\psi_c(t)$ whose graph is drawn in Fig. 5. It is a limit of the functions $h_n(t)$ drawn in Fig. 6.

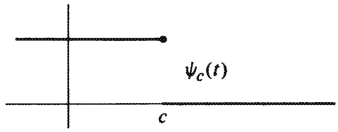

Figure 5

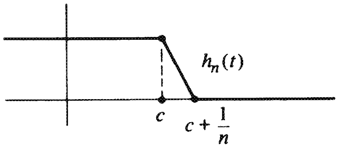

Figure 6

Lemma 3. *Let $\psi_c(A) = P_c$. If $\alpha I \leqslant A \leqslant \beta I$, then:*

i) $P_c = 0$ *if* $c \leqslant \alpha$, *and* $P_c = I$ *if* $c \geqslant \beta$.
ii) *If* $c \leqslant c'$, *then* $P_c \leqslant P_{c'}$

Proof. Clear from Lemma 2.

Observe that we also have $P_c^2 = P_c$, i.e. that P_c is a projection. We call $\{P_c\}$ the **spectral family** associated with A.

We keep the same notation, and we shall make use of the two functions f_c, g_c whose graph is drawn in Fig. 7. Thus $f_c(t) + g_c(t) = |t - c|$. We have

$$(t - c)(1 - \psi_c(t)) = f_c(t).$$

Hence

$$(A - cI)(I - P_c) = f_c(A) \tag{1}$$

$$A - cI = f_c(A) - g_c(A) \tag{2}$$

$$(A - cI)P_c = -g_c(A)P_c = -g_c(A). \tag{3}$$

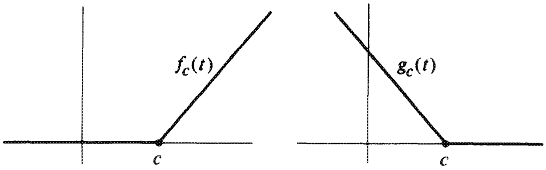

Figure 7

Theorem 5. *Let P_c be the spectral family associated with A. If $b \leqslant c$, then we have*

$$bI \leqslant A \leqslant cI, \text{ on } \operatorname{Im}(P_c - P_b).$$

Proof. From (1) above, we have $A - bI = f_b(A)$ on the orthogonal complement of P_b, whence the inequality $bI \leqslant A$ follows on this complement since $f_b \geqslant 0$. From (3) above, we have

$$A - cI = -g_c(A)$$

on the image of P_c, and since $-g_c$ is $\leqslant 0$, we get $A \leqslant cI$ on this image. This proves our theorem.

Theorem 6. *The family $\{P_t\}$ is strongly continuous from the right.*

Proof. Let $v \in H$. Our assertion means that $P_{c+\epsilon}v \longrightarrow P_c v$ as $\epsilon \longrightarrow 0$. It suffices to prove that

$$\langle P_{c+\epsilon}v, v\rangle \longrightarrow \langle P_c v, v\rangle$$

because

$$\langle (P_{c+\epsilon} - P_c)v, v\rangle = |(P_{c+\epsilon} - P_c)v|^2.$$

But $\langle P_c v, v\rangle$ is close to $\langle h(A)v, v\rangle$ where h is as shown in Fig. 8.

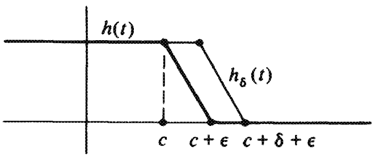

Figure 8

Let h_δ be the translation of h to the right by δ. Then

$$\psi_c(t) \leqslant h(t) \leqslant h_\delta(t)$$
$$\psi_c(t) \leqslant \psi_{c+\epsilon}(t) \leqslant h_\delta(t).$$

Since $h_\delta \longrightarrow h$ uniformly as $\delta \longrightarrow 0$, we conclude that $h_\delta(A)$ tends to $h(A)$. Then

$$P_c \leqslant h(A) \leqslant h(A) + \epsilon I,$$
$$P_c \leqslant P_{c+\epsilon} \leqslant h(A) + \epsilon I.$$

Applying this to (v, v), we get

$$\langle P_c v, v\rangle \leqslant \langle h(A)v, v\rangle \leqslant \langle h(A)v, v\rangle + \epsilon\langle v, v\rangle$$
$$\langle P_c v, v\rangle \leqslant \langle P_{c+\epsilon} v, v\rangle \leqslant \langle h(A)v, v\rangle + \epsilon\langle v, v\rangle,$$

and since $\langle h(A)v, v\rangle$ is ϵ-close to $\langle P_c v, v\rangle$, we get our theorem.

Theorem 7 (Lorch). *From the left,*

$$\lim_{\epsilon\to 0}(P_c - P_{c-\epsilon}) = Q_c$$

is the projection on the c-eigenspace of A.

Proof. Using Theorem 5, we have

$$(c - \epsilon)(P_c - P_{c-\epsilon}) \leqslant A(P_c - P_{c-\epsilon}) \leqslant c(P_c - P_{c-\epsilon})$$

whence

$$|(A - cI)(P_c - P_{c-\epsilon})| \leqslant \epsilon.$$

But for each v, $\lim_{\epsilon\to 0}(P_c - P_{c-\epsilon})v$ exists, say $= w$. It follows that $Aw = cw$, i.e. Q_c maps H into the c-eigenspace.

Conversely, if φ is a continuous function, then for any A-invariant closed subspace F, we have

$$\varphi(A|F) = \varphi(A)|F.$$

We want to show that Q_c is the identity on the c-eigenspace, and without loss of generality we may therefore assume that $H = H_c$ is this eigenspace. Then $P_c = 0$ because $f_c = 0$ on the spectrum of A. If $b < c$, then

$$f_b(A) = A - bI = (c - b)I$$

is invertible, and hence $P_b = O$. This proves our theorem.

Appendix 2
Unbounded Operators

§1. SELF-ADJOINT OPERATORS

Let H be a Hilbert space and A a linear map,

$$A\colon D_A \to H$$

defined on a dense subspace. Consider the set of vectors $v \in H$ such that there exists $w \in H$ such that

$$\langle u, w\rangle = \langle Au, v\rangle \qquad \text{all } u \in D_A,$$

or in other words, $\langle u, w\rangle - \langle Au, v\rangle = 0$. The set of such v is the projection on the first factor of the intersection of the kernels of

$$(v, w) \mapsto \langle u, w\rangle - \langle Au, v\rangle, \qquad u \in D_A.$$

It is a vector space. To each v in this vector space there is exactly one w, if it exists, having the above property, because

$$u \mapsto \langle Au, v\rangle$$

is a functional on a dense subspace. Hence we can define an operator A^* by the formula

$$A^*v = w,$$

on the space D_{A^*} of such vectors v. We call the pair (A^*, D_{A^*}) the **adjoint** of A.

Let $J\colon H \times H \to H \times H$ be the operator such that $J(x, y) = (-y, x)$. Then $J^2 = -I$. We note that the graph G_{A^*} of A^* is given by the formula

$$\boxed{G_{A^*} = (JG_A)^{\perp},}$$

where $\perp$ denotes orthogonal complement, and hence the graph of A^* is closed.

We say that A is **closed** if its graph G_A is closed.

If A is closed, then D_{A^*} is dense in H.

Proof. Let $h \in D_{A^*}^{\perp}$, so

$$(h, 0) \in (G_{A^*})^{\perp} = (JG_A)^{\perp\perp} = JG_A$$

because we assumed that A is closed. We conclude that $(0, h) \in G_A$, and hence $h = 0$, proving our assertion.

If A is closed, then $A^{**} = A$.

Proof. $G_{A^{**}} = (JG_{A^*})^{\perp} = (J(JG_A)^{\perp})^{\perp} = G_A$.

If D_A and D_{A^*} are dense, then $G_{A^{**}}$ = closure of G_A.

Proof. $G_{A^{**}} = (JG_{A^*})^{\perp} = (J(JG_A)^{\perp})^{\perp} = \overline{G_A}$.

If A is defined on D_A and B is defined on D_B, if $D_A \subset D_B$, and if the restriction of B to D_A is A, then one usually says that A is **contained in** B, and one writes $A \subset B$. The above assertion shows that $A \subset A^{**}$.

We say that A is **symmetric** if $\langle Au, v\rangle = \langle u, Av\rangle$ for all $u, v \in D_A$. We say that A is **self adjoint,** $A = A^*$, if in addition $D_A = D_{A^*}$.

If A is symmetric, then $A \subset A^*$.

This is clear. Recall that we assumed D_A dense in H.

If A, B are self-adjoint and $A \subset B$, then $A = B$.

This is also clear, because in general $B^* \subset A^*$, so in the self-adjoint case, $B \subset A$, whence $A = B$.

Let A be symmetric, defined on D_A dense as above. Let $\lambda \in \mathbf{C}$ not be real. Then $A - \lambda I$ is injective on D_A, because from

$$Au = \lambda u \qquad \text{and} \qquad \langle Au, u\rangle = \langle u, Au\rangle$$

we conclude

$$\lambda\langle u, u\rangle = \langle \lambda u, u\rangle = \langle u, \lambda u\rangle = \bar{\lambda}\langle u, u\rangle,$$

so $u = 0$. Hence we can define an operator

$$U = U_{A,\lambda} = (A + \bar{\lambda} I)(A + \lambda I)^{-1}$$

on the image $(A+\lambda I)D_A$. *We contend that U is unitary.* This amounts to verifying that for $u, v \in D_A$ we have

$$\langle Au + \bar{\lambda}u, Av + \bar{\lambda}v\rangle = \langle Au + \lambda u, Av + \lambda v\rangle,$$

which is obvious.

Lemma 1. *If A is symmetric, closed, and $\lambda \in \mathbf{C}$ is not real, then $(A + \lambda I)D_A$ is closed.*

Proof. Let $\{u_n\}$ be a sequence in D_A such that $\{(A + \lambda I)u_n\}$ is Cauchy. Since U is unitary, it follows that

$$\left\{(A + \bar{\lambda}I)u_n\right\}$$

is also Cauchy, hence $\{(\lambda - \bar{\lambda})u_n\}$ is Cauchy, and $\{u_n\}$ is Cauchy, say converging to u. But

$$\left\{2Au_n + (\lambda + \bar{\lambda})u_n\right\}$$

is Cauchy, whence also $\{Au_n\}$ is Cauchy. Since the graph of A is assumed closed, we conclude that $\{(u_n, Au_n)\}$ converges to an element (u, Au) in the graph, and the sequence

$$\{(A + \lambda I)u_n\}$$

converges to $(A + \lambda I)u$. This proves that $(A + \lambda I)D_A$ is closed.

Theorem 1. *Let A be symmetric, closed with dense domain. Let $\lambda \in \mathbf{C}$ be not real, and such that $(A + \lambda I)D_A$ and $(A + \bar{\lambda}I)D_A$ are dense (whence equal to H by the lemma). Then A is self-adjoint.*

Proof. Let $v \in D_{A^*}$. It suffices to show that $v \in D_A$. We have by definition

$$\langle Au, v\rangle = \langle u, A^*v\rangle, \qquad \text{all } u \in D_A.$$

Since $(A + \lambda I)D_A = H$, there exists $u_1 \in D_A$ such that

$$A^*v + \lambda v = Au_1 + \lambda u_1.$$

Then

$$\langle Au, v\rangle = \langle u, Au_1 + \lambda u_1 - \lambda v\rangle, \qquad \text{all } u \in D_A,$$

whence

$$\langle (A + \bar{\lambda}I)u, v\rangle = \langle (A + \bar{\lambda}I)u, u_1\rangle, \qquad \text{all } u \in D_A.$$

This proves that $v = u_1$, as was to be shown.

Remark. In the literature, you will find that the dimension of the cokernel of $(A + \lambda I)D_A$ is called a defect index. We are concerned here with a situation when the defect indices are 0.

Corollary. *Let A be symmetric with dense domain. Let $\lambda \in \mathbf{C}$ be not real, and such that $(A + \lambda I)D_A$ and $(A + \bar{\lambda} I)D_A$ are dense. Then the closure of G_A is the graph of an operator which is self-adjoint.*

Proof. Since A is symmetric, the domain of A^* is also dense, and we have shown above that $G_{A^{**}}$ is the closure of G_A, so A has a closure. It is immediate that this closure is also symmetric, and the theorem applies.

An operator A defined on D_A is called **essentially self-adjoint** if the closure of its graph is the graph of a self-adjoint operator. The corollary gives a sufficient condition for an operator to be essentially self-adjoint.

Theorem 2. *Let A be a self-adjoint operator. Let $z \in \mathbf{C}$ and z not real. Then $A - zI$ has kernel 0. There is a unique bounded operator*

$$R(z) = (A - zI)^{-1}: H \to D_A$$

which establishes a bijection between H and D_A, and is the inverse of $A - zI$. We have

$$R(z)^* = R(\bar{z}).$$

If $\operatorname{Im} z$, $\operatorname{Im} w \neq 0$, *then we have the resolvant equation*

$$(z - w)R(z)R(w) = R(z) - R(w) = (z - w)R(w)R(z),$$

so in particular, $R(z)$, $R(w)$ commute. We have $|R(z)| \leqslant 1/|Im\, z|$.

Proof. Let $z = x + iy$. If u is in the domain of A, then

$$|(A - zI)u|^2 = |(A - xI)u|^2 + y^2|u|^2 \geqslant y^2|u|^2$$

because A is symmetric, so the cross terms disappear. This proves that the kernel of $A - zI$ is 0, and that the inverse of $A - zI$ is continuous, when viewed as defined on the image of $A - zI$. If v is orthogonal to this image, i.e.

$$\langle Au - zu, v \rangle = 0$$

for all $u \in D_A$, then $\langle Au, v \rangle = \langle u, \bar{z}v \rangle$, and by the definition of being self-adjoint, it follows that v lies in the domain of A and that $Av = \bar{z}v$. Since the kernel of $A - \bar{z}I$ is 0, we conclude that $v = 0$. Hence the image of $A - zI$ is dense, so that by Theorem 2 this image is all of H and $R(z)$ is everywhere defined, equal to the inverse of $A - zI$. We then have

$$[(A - wI) - (A - zI)]R(w) = (z - w)R(w).$$

Multiplying this on the left by $R(z)$ yields the resolvant formula of the theorem, whose proof is concluded.

We write

$$R(i) = (A - iI)^{-1} = C + iB$$

where B, C are bounded hermitian. From the resolvant equation between $R(i)$ and $R(-i)$ we conclude that B, C commute. We may call B the **imaginary part of** $(A - iI)^{-1}$, symbolically

$$B = \operatorname{Im}(A - iI)^{-1}.$$

***Lemma* 2.** *With the above notation, we have $C = AB$ and $BA \subset AB$. The kernel of B is* 0, *and $O \leqslant B \leqslant I$.*

Proof. We have from $R(i)^* = R(-i)$ that

$$(A - iI)^{-1} - (A + iI)^{-1} = 2iB.$$

We multiply this on the left with A, noting that

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

and

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

We then obtain $C = AB$. For BA we multiply the first relation on the right by A, so that we use

$$(A - iI)^{-1}(A - iI) = I_{D_A}$$

and similarly for $A + iI$. The relation $BA \subset AB$ follows. The kernel of B is 0, for any vector in the kernel is also in the kernel of $C = AB$, whence in the kernel of $(A - iI)^{-1}$, and therefore equal to 0. We leave the relation $B \geqslant O$ to the reader. That $B \leqslant I$ follows from $|R(i)| \leqslant 1$, a special case of the last inequality in Theorem 2.

We now give an example of a self-adjoint operator. It will be shown after that any self-adjoint operator is of this nature.

Theorem 3. *Let $\{H_n\}$ be a sequence of Hilbert spaces. Let A_n be a bounded self-adjoint operator on H_n. Let H be the orthogonal direct sum of the H_n, so that H consists of all series $\sum u_n$ with $\sum |u_n|^2 < \infty$. There exists a unique self-adjoint operator A on H such that each H_n is contained in the domain D_A and such that the restriction of A to H_n is A_n. Its domain is the vector space of series $u = \sum u_n$ such that*

$$\sum |A_n u_n|^2 < \infty,$$

and $Au = \sum A_n u_n$.

Proof. The uniqueness is clear from the property that if A, B are self-adjoint and $A \subset B$, then $A = B$. It suffices now to prove that if we let D_A be the domain described above, and define Au by $\sum A_n u_n$, then A is self-adjoint. It is clear that A is symmetric. Let $v \in D_{A^*}$. Then

$$\langle u, A^*v \rangle = \langle Au, v \rangle, \qquad \text{all } u \in D_A.$$

Say $u = \sum u_n$. Then

$$\sum \langle u_n, A^*v \rangle = \sum \langle Au_n, v \rangle.$$

If $u \in H_n$, then

$$\langle u_n, A^*v \rangle = \langle Au_n, v \rangle$$

$$\langle u_n, (A^*v)_n \rangle = \langle Au_n, v_n \rangle,$$

whence $(A^*v)_n = A_n v_n$. Then

$$\sum |A_n v_n|^2 = \sum |(A^*v)_n|^2 = |A^*v|^2,$$

whence $v \in D_A$, so $D_{A^*} \subset D_A$ and A is therefore self-adjoint. This proves the theorem.

In the situation of Theorem 3, we use the notation

$$A = \hat{\bigoplus} A_n.$$

We deal with the converse of Theorem 3. Let A be an arbitrary self-adjoint operator on the Hilbert space H, and let

$$(A - iI)^{-1} = C + iB$$

as above.

We are in a position to decompose our Hilbert space by means of B. Let θ_c be the function whose graph is given in Fig. 1., and which gives rise to a projection operator.

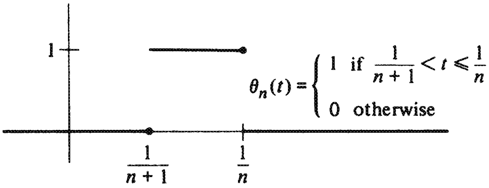

Figure 1

Let $\{P\}$ be the spectral family for B, and let

$$Q_n = \theta_n(B) = P_{1/n} - P_{1/(n+1)}.$$

Then Q_n is a projection operator, and we let

$$H_n = Q_n H = \operatorname{Im} Q_n.$$

Then

$$H = \hat{\bigoplus} H_n$$

is an orthogonal direct sum. In fact, let θ and η be the functions whose graphs are on Fig. 2 (a) and (b) respectively.

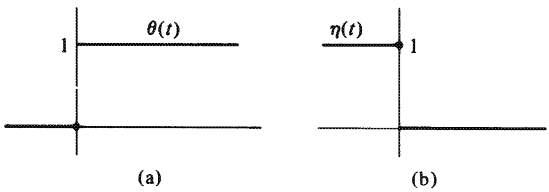

Figure 2

Then $1 - \theta = \eta$ and $\eta(B) = 0$ because the spectral family for B is continuous at 0, in view of Lemma 2 (kernel $B = 0$) and Lorch's Theorem 7.

Let $s_n(t)$ be the function whose graph is on Fig. 3. Then

$$Bs_n(B) = \theta_n(B) = Q_n.$$

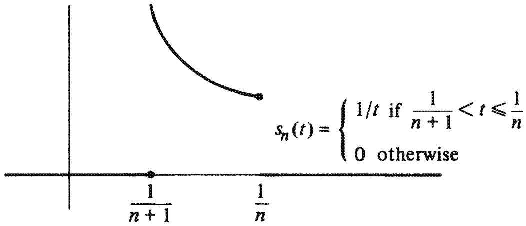

Figure 3

Theorem 4. *Let A be a self-adjoint operator and let $B = \operatorname{Im}(A - iI)^{-1}$. Let $Q_n = \theta_n(B)$ be the projection operator defined by the function θ_n above. Then A is defined on* Im Q_n, *and*

$$Q_n A \subset A Q_n = s_n(B)C.$$

Let $H_n = Q_n H$. Then H is the orthogonal direct sum of the spaces H_n, the restriction of A to H_n is a bounded operator A_n, and

$$A = \widehat{\bigoplus} A_n.$$

Proof. Since $ts_n(t) = \theta_n(t)$, we get $Bs_n(B) = \theta_n(B) = Q_n$. Then by Lemma 2,

$$AQ_n = ABs_n(B) = Cs_n(B) = s_n(B)C.$$

In particular, AQ_n is everywhere defined. On the other hand,

$$Q_n A = s_n(B)BA \subset s_n(B)AB \subset s_n(B)C.$$

This proves that $Q_n A \subset AQ_n$. It means that given $v \in D_A$, if

$$v = \sum v_n$$

is the decomposition of v according to the spaces H_n, and if

$$Av = \sum w_n,$$

then

$$Q_n A v = w_n = A Q_n v = A v_n.$$

So $Av = \sum Av_n$, and the theorem is proved.

§2. THE SPECTRAL MEASURE

We begin with a *bounded* self-adjoint operator A on H. For each $v \in H$ we obtain a functional on $C_c(\mathbf{R})$ by letting

$$\lambda_v(\varphi) = \langle \varphi(A)v, v\rangle,$$

and this functional is obviously positive. Hence there exists a unique positive measure μ_v on $\mathbf{R}$ such that

$$\langle \varphi(A)v, v\rangle = \int_{\mathbf{R}} \varphi \, d\mu_v.$$

By polarization, for $v, w \in H$ we get a complex measure $\mu_{v,w}$ such that

$$\langle \varphi(A)v, w\rangle = \int_{\mathbf{R}} \varphi \, d\mu_{v,w}.$$

It is clear that $\mu_{v,w}$ is $\mathbf{C}$-linear in v, antilinear in w. Furthermore by Theorem 3 of §1, we have

$$\langle \varphi(A)v, w\rangle \leqq \|\varphi\|_\infty |v| |w|,$$

so that in particular $\mu_v(\mathbf{R}) \leqq |v|^2$ and is finite.

Let $BM(\mathbf{R})$ be the Banach space of (Borel) bounded measurable functions on $\mathbf{R}$. For each $f \in BM(\mathbf{R})$ the association

$$(v, w) \mapsto \int f \, d\mu_{v,w}$$

is linear in v and antilinear in w. Furthermore

$$\left|\int f \, d\mu_{v,w}\right| \leqq \|f\|_\infty |v| |w|,$$

as one sees by using a sequence $\{\varphi_n\}$ in $C_c(\mathbf{R})$ approaching f pointwise almost everywhere with respect to the measure $|\mu_{v,w}|$, and such that $|\varphi_n| \leqq \|f\|_\infty$. Thus our association is continuous, and there exists a unique bounded operator, which we denote by $f(A)$, such that

$$\langle f(A)v, w\rangle = \int f \, d\mu_{v,w}.$$

The following properties are then satisfied for $f, g \in BM(\mathbf{R})$.

SPEC 1. $$(fg)(A) = f(A)g(A)$$

SPEC 2. $$f(A)^* = \bar{f}(A)$$

SPEC 3. *If f_1 is the function $f_1(t) = 1$, then $f_1(A) = I$.*

SPEC 4. *If the functions $f(t)$ and $g(t) = tf(t)$ are bounded measurable, then $g(A) = Af(A)$.*

SPEC 5. *We have $|f(A)| \leqslant \|f\|_\infty$. Furthermore, if $\{f_n\}$ is a bounded sequence in $BM(\mathbf{R})$ converging pointwise to f, then $\{f_n(A)\}$ converges strongly to $f(A)$.*

The above properties are either obvious from what precedes, or follow by applying the dominated convergence theorem, and taking limits. For instance, to prove **SPEC 1**, we use two sequences of functions in $C_c(\mathbf{R})$ approaching f and g respectively.

We shall use the structure theorem of the preceding section to extend the above results to an unbounded self-adjoint operator A, such that

$$A = \widehat{\bigoplus} A_n \qquad \text{and} \qquad H = \widehat{\bigoplus} H_n,$$

where A_n is bounded self-adjoint on H_n. Let $f \in BM(\mathbf{R})$ and $v \in H$. Since $|f(A_n)v_n| \leqslant \|f\|_\infty |v_n|$, there is a unique bounded operator $f(A)$ such that

$$f(A)v = \sum f(A_n)v_n.$$

To each H_n and $v_n \in H_n$ we can associate the measure $\mu_{v_n}^{(n)}$ as above. We let μ_v be the unique positive measure such that

$$\int f\, d\mu_v = \sum_n \int f\, d\mu_{v_n}^{(n)} = \sum_n \langle f(A)v_n, v_n \rangle.$$

Since

$$|\langle f(A)v_n, v_n\rangle| \leqslant \|f\|_\infty |v_n|^2,$$

we do get a positive measure, and the formalism of the five **SPEC** properties extends at once to the case of an unbounded operator A, in other words:

Theorem 5. *Let A be a self-adjoint operator. There exists a unique association $f \longmapsto f(A)$ from $BM(\mathbf{R})$ into the bounded operators on H satisfying* **SPEC 1** *through* **SPEC 5**.

Proof. The existence is essentially obvious from the above. Note that for

SPEC 4, in view of

$$|Af(A)v_n| \leqslant \|g\|_\infty |v_n|,$$

it follows that $\sum f(A)v_n$ lies in D_A and hence **SPEC 4** is valid. The uniqueness will follow from the considerations of the next section.

§3. THE RESOLVANT FORMULA

Suppose given an association $f \mapsto f(A)$ satisfying the five spectral properties of the preceding section. For each $v, w \in H$ there is a unique measure $\mu_{v,w}$ such that

$$\langle f(A)v, w\rangle = \int f\, d\mu_{v,w}.$$

Let z be complex and not real. The function $f(t)$ such that

$$f(t) = \frac{1}{t - z}$$

is bounded measurable, and $tf(t)$ is bounded. Also $(t - z)f(t) = 1$. Hence

$$(A - zI)f(A) = I.$$

This means that the resolvant has the integral expression

$$\boxed{\langle (A - zI)^{-1}v, w\rangle = \int_{\mathbf{R}} \frac{1}{t - z}\, d\mu_{v,w}(t).}$$

We write μ_v instead of $\mu_{v,v}$. Note that μ_v is a positive measure.

Theorem 6. *Let A be a self-adjoint operator on H and let $v \in H$. Let $R(z) = (A - zI)^{-1}$ for z not real. For any $\psi \in C_c(\mathbf{R})$ we have*

$$\lim_{\epsilon \to 0} \frac{1}{2\pi i} \int_{\mathbf{R}} \langle [R(\lambda + i\epsilon) - R(\lambda - i\epsilon)]v, v\rangle \psi(\lambda)\, d\lambda = \int_{\mathbf{R}} \psi(\lambda)\, d\mu_v(\lambda).$$

If $\lambda_1 < \lambda_2$ are real numbers which have μ_v-measure 0, then

$$\lim_{\epsilon \to 0} \frac{1}{2\pi i} \int_{\lambda_1}^{\lambda_2} \langle [R(\lambda + i\epsilon) - R(\lambda - i\epsilon)]v, v\rangle\, d\lambda = \int_{\lambda_1}^{\lambda_2} d\mu_v(\lambda).$$

The proof is based on the following lemma.

Lemma. *Let μ be a positive regular measure on* **R** *such that $\mu(\mathbf{R})$ is finite. Then for $\psi \in C_c(\mathbf{R})$ we have*

$$\lim_{\epsilon \to 0} \frac{1}{\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{\epsilon}{(t-\lambda)^2 + \epsilon^2} \psi(\lambda)\, d\mu(t)\, d\lambda = \int_{-\infty}^{\infty} \psi(\lambda)\, d\lambda.$$

Furthermore, if $\lambda_1 < \lambda_2$ are real and such that the set $\{\lambda_1, \lambda_2\}$ has μ-measure 0, then

$$\lim_{\epsilon \to 0} \frac{1}{\pi} \int_{\lambda_1}^{\lambda_2} \int_{-\infty}^{\infty} \frac{\epsilon}{(t-\lambda)^2 + \epsilon^2}\, d\mu(t)\, d\lambda = \int_{\lambda_1}^{\lambda_2} d\mu(\lambda).$$

Proof. First observe that the family of functions

$$\varphi_\epsilon(\lambda) = \frac{1}{\pi} \frac{\epsilon}{\lambda^2 + \epsilon^2}$$

is a Dirac family on **R** for $\epsilon \to 0$. (The functions don't have compact support, and so we are dealing with the case when condition **DIR 3′** is satisfied, rather than **DIR 3**.) The left-hand side integrals in our lemma can be written

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \varphi_\epsilon(t-\lambda) h(\lambda)\, d\mu(t)\, d\lambda$$

where h is either ψ or the characteristic function of the interval $[\lambda_1, \lambda_2]$. We apply Fubini's theorem to see that this expression is equal to

$$\int_{-\infty}^{\infty} \varphi_\epsilon * h(t)\, d\mu(t).$$

Note that $\varphi_\epsilon * h$ is bounded, and converges pointwise to h if $h = \psi$, and pointwise to h except at the end points λ_1, λ_2 in the other case. Since we picked our interval so that the end points have μ-measure 0, we can apply the dominated convergence theorem to conclude the proof.

The lemma obviously proves Theorem 6, because

$$\frac{1}{t-\lambda-i\epsilon} - \frac{1}{t-\lambda+i\epsilon} = \frac{2i}{(t-\lambda)^2 + \epsilon^2}.$$

Furthermore, Theorem 6 provides the desired uniqueness left hanging at the end of the last section, because it gives the value of the measure entirely in terms of the resolvant and Lebesgue measure, as on the left-hand side of the first formula on elements of $C_c(\mathbf{R})$.

It is possible to develop the spectral theory by starting with a direct proof of Theorem 6, showing that the limit on the left-hand side exists. One then defines the spectral measure as that associated with the corresponding functional, and one proves the other properties from there. Cf. Akniezer-Glazman, *Theory of Linear Operators in Hilbert Space*, Translated from the Russian, New York, Frederick Ungar, 1963, pp. 8 and 31, and also a forthcoming book of Hörmander, who gives a very elegant way of constructing the spectral measure directly from the resolvant formula.

Appendix 3

Meromorphic Families of Operators

§1. COMPACT OPERATORS

Let E be a Banach space, and let $K(E)$ be the Banach space of compact operators on E. We recall that T is **compact** means that T maps bounded sets on relatively compact sets. We have already recalled in Chapter II some properties of compact operators in Hilbert space. Let T be compact. If T has infinitely many eigenvalues $\neq 0$, then they form a sequence $\{\lambda_i\}$ such that

$$|\lambda_{i+1}| \leqslant |\lambda_i|$$

and $\lim \lambda_i = 0$. Let $\lambda_0 \neq 0$ be such that $T - \lambda_0 I$ is not invertible. Then λ_0 is an eigenvalue. Furthermore, for a sufficiently large integer n, there is a direct sum decomposition

$$E = \operatorname{Ker}(T - \lambda_0 I)^n \oplus W,$$

where W is closed, $\operatorname{Ker}(T - \lambda_0 I)^n$ is finite dimensional, $T - \lambda_0 I$ maps W into itself, and its restriction to W is invertible (we say that $T - \lambda_0 I$ is invertible on W). This is the basic fact about compact operators, and the reader will find a proof in most books on analysis, e.g. my *Real Analysis*. It then follows that $T - \lambda I$ is invertible on E (and also on W) for λ close to λ_0, $\lambda \neq \lambda_0$. We call $\operatorname{Ker}(T - \lambda_0 I)^n$ the **λ_0-Jordan space** of T in E. We shall now describe a natural way of constructing the projection of E on that space.

Theorem 1. *Let T be a compact operator on E. Let $\lambda_0 \neq 0$. Let W be the complement of the λ_0-Jordan space of T as above. Let*

$$P = \frac{1}{2\pi i}\int_C (\lambda I - T)^{-1}\, d\lambda$$

where C is a sufficiently small circle with center λ_0. Then P is the identity on *the λ_0-Jordan space of T, and $PW = 0$. Thus P is the projection of E on this Jordan space, killing the complement W.*

Proof. We first show that if $(T - \lambda_0 I)^n v = 0$, then $Pv = v$. We have

$$v = \frac{1}{2\pi i} \int_C \frac{v}{\lambda - \lambda_0}\, d\lambda.$$

Then

$$\begin{aligned} Pv - v &= \frac{1}{2\pi i} \int_C \left[(\lambda I - T)^{-1} - (\lambda - \lambda_0)^{-1}\right] v\, d\lambda \\ &= \frac{1}{2\pi i} \int_C \left[\frac{1}{(\lambda - \lambda_0 + \lambda_0 - T)} - \frac{1}{(\lambda - \lambda_0)} \right] v\, d\lambda \\ &= \frac{1}{2\pi i} \int_C \frac{1}{\lambda - \lambda_0} \left[\frac{1}{1 + \dfrac{\lambda_0 - T}{\lambda - \lambda_0}} - 1 \right] v\, d\lambda. \end{aligned}$$

We expand the expression in brackets by geometric series, which breaks off after $n - 1$ terms by assumption. The terms $1 - 1$ cancel. We are left with a series in $1/(\lambda - \lambda_0)$, with lowest term $1/(\lambda - \lambda_0)^2$, and so there is no residue. The integral is therefore equal to 0. This proves our first assertion.

Looking at the restriction of T to W, which is invertible, it suffices to prove:

If $A = T - \lambda_0 I$ is invertible, then $P = O$.

But we have

$$\begin{aligned} \int_C \frac{1}{\lambda - T}\, d\lambda &= \int_C \frac{1}{\lambda - \lambda_0 + A}\, d\lambda \\ &= \int_C \frac{1}{A} \frac{1}{(\lambda - \lambda_0)A^{-1} + I}\, d\lambda. \end{aligned}$$

We may now expand the expression inside the integral by the geometric series, and the integral of the series is 0 because the function is holomorphic in λ. This proves our theorem.

A family of operators $z \mapsto A(z)$ is said to be **meromorphic** at a point z_0 if there is a power series expansion

$$A(z) = \sum_{n \geqq -m} A_n (z - z_0)^n$$

with operator coefficients A_n. In general, if

$$f : U \to E$$

is a mapping of an open set in $\mathbf{C}$ into a Banach space E, we say that f is **meromorphic** at a point z_0 if in a neighborhood of z_0 f has a power series expansion

$$f(z) = \sum_{n \geqq -m} a_n (z - z_0)^n$$

with coefficients $a_n \in E$.

The next theorem is due to Stanley Steinberg, *Archs. ration. Mech. Analysis* **31** (1968), pp. 372-379.

Theorem 2. *Let E be a Banach space, U a connected open set in $\mathbf{C}$, and $T : U \to K(E)$ a holomorphic map. If there exists one point $z_0 \in U$ such that $I - T(z_0)$ is invertible, then*

$$z \mapsto (I - T(z))^{-1}$$

is meromorphic.

Proof. The set of z where $(I - T(z))^{-1}$ is meromorphic is open. We shall prove that the complement is also open, whence empty because of our assumption for $z = z_0$.

Let $z_1 \in U$. There exists a small circle C around 1 such that 1 is the only possible eigenvalue of $T(z_1)$ inside the circle, by the discreteness of the spectrum of compact operators. Then $\lambda I - T(z_1)$ is invertible for λ on the circle. The image

$$(\lambda, z_1) \mapsto \lambda I - T(z_1)$$

is compact in the open set of invertible operators, and hence $\lambda I - T(z)$ is invertible for all z close to z_1. Let

$$P(z) = \frac{1}{2\pi i} \int_C (\lambda I - T(z))^{-1} \, d\lambda,$$

Then $z \mapsto P(z)$ is holomorphic for z near z_1, and we know from Theorem 1 that $P(z)^2 = P(z)$ (it is a projection on a Jordan space).

We observe that $P(z_1)$ is the projection on the space E_1, the Jordan space with eigenvalue 1. Write

$$E = E_1 \oplus W.$$

Then $I - T(z_1)$ is invertible on W. We note that $P(z)$ commutes with $T(z)$. Let

$$Q(z) = I - P(z)$$

so that $Q(z_1)$ is the projection on W. Let

$$S(z) = P(z_1)P(z) + Q(z_1)Q(z).$$

Then $S(z_1) = I$, so $S(z)$ is invertible for z near z_1. Let

$$A(z) = S(z)(I - T(z))S(z)^{-1}$$

for z near z_1. Since obviously

$$S(z)P(z) = P(z_1)S(z),$$

we see that $P(z_1)$ *commutes with* $A(z)$, and so does $Q(z_1)$. Hence $A(z)$ "splits" into two operators, on E_1 and on W respectively, say

$$A(z)_1 = A(z)|E_1 \qquad \text{and} \qquad A(z)_W = A(z)|W,$$

so that

$$A(z) = A(z)_1 \oplus A(z)_W.$$

Since $I - T(z_1)$ is invertible on W, it follows that for z near z_1, $A(z)_W$ is invertible on W because $A(z)$ is close to

$$A(z_1) = I - T(z_1).$$

We let $A(z)_W^{-1}$ be the inverse of $A(z)_W$ on W, and be 0 on E_1. If $\det A(z)_1$ is identically zero for z near z_1, then $I - T(z)$ is not meromorphic. If $\det A(z)_1$ is not identically zero for z near z_1, then $A(z)_1$ is meromorphic. Thus we see from the connectedness of U and the invertibility of $I - T(z_0)$ that $I - T(z)$ is meromorphic.

We also get additional information, because we now know that $\det A(z)_1$ is not identically 0. Let $A(z)_1^{-1}$ be the inverse of $A(z)_1$ on E_1 and be 0 on W, for those points z near z_1 where the inverse exists. Since $A(z)_1$ operates on the finite dimensional space E_1, the map

$$z \mapsto A(z)_1^{-1}$$

is meromorphic in the trivial one-variable sense, and its poles are the zeros of the determinant $\det A(z)_1$. Then

$$[I - T(z)]^{-1} = S(z)A(z)_1^{-1}S(z)^{-1} + S(z)A(z)_W^{-1}S(z)^{-1}.$$

Locally, after applying a holomorphic family of inner automorphisms, we see that situation of Theorem 2 is reduced to that of a direct sum involving an operator in a fixed finite dimensional space, and a holomorphic family of invertible operators on a complementary subspace.

§2. BOUNDED OPERATORS

Theorem 3. *Let S be a closed set, contained in an open set U of $\mathbf{C}$. Let C be a simple closed curve in U whose interior contains S. Let*

$$\lambda \mapsto R(\lambda), \qquad \lambda \in U - S,$$

be a holomorphic family of operators in a Banach space E, satisfying the relation

$$R(\lambda) - R(\lambda') = (\lambda' - \lambda)R(\lambda)R(\lambda').$$

For any function f holomorphic on U, define

$$R(f) = \frac{1}{2\pi i}\int_C f(\lambda)R(\lambda)\, d\lambda.$$

Then $f \mapsto R(f)$ is a homomorphism, i.e. aside from the linearity, satisfies

$$R(fg) = R(f)R(g).$$

Proof. Let C' be a closed curve surrounding C as in Fig. 1. Then

$$\begin{aligned} R(f)R(g) &= \frac{1}{(2\pi i)^2}\int_C\int_{C'} f(\lambda)g(\lambda')R(\lambda)R(\lambda')\, d\lambda\, d\lambda' \\ &= \frac{1}{(2\pi i)^2}\int_C\int_{C'} f(\lambda)g(\lambda')\frac{R(\lambda) - R(\lambda')}{\lambda' - \lambda}\, d\lambda\, d\lambda'. \end{aligned}$$

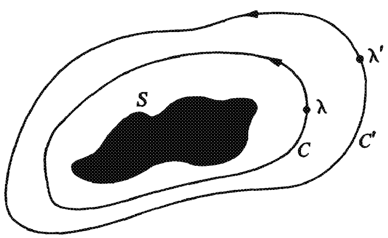

Figure 1

This decomposes $R(f)R(g)$ into a difference of two integrals. We have

$$\frac{1}{2\pi i}\int_{C'}\frac{g(\lambda')}{\lambda'-\lambda}\,d\lambda' = g(\lambda),$$

so the first integral is equal to $R(fg)$ by definition. On the other hand,

$$\int_{C}\frac{f(\lambda)}{\lambda'-\lambda}\,d\lambda = 0,$$

so the second integral is equal to 0, thus proving our assertion.

In particular, letting $f = g = 1$, we see that the integral

$$P = \frac{1}{2\pi i}\int_{C} R(\lambda)\,d\lambda$$

satisfies $P^2 = P$, and is therefore a projection on a subspace, corresponding to the "spectrum" S.

Appendix 4
Elliptic PDE

§1. SOBOLEV SPACES

We shall work on the torus, although some of the preliminary remarks would be valid on euclidean space as well. Thus we let $\mathbf{T}^d = \mathbf{R}^d/\mathbf{Z}^d$. We let $L^2 = L^2(\mathbf{T}^d)$. Any function $f \in L^2$ gives rise to a functional on $C^\infty(\mathbf{T})$ by

$$\varphi \longmapsto \int_{\mathbf{T}} f(x)\varphi(x)\,dx.$$

This functional is called the **distribution** associated with f, and is denoted by $\langle f \rangle$. Elements of $C^\infty(\mathbf{T})$ are called **test functions**. If $D_i = \partial/\partial x_i$, then we define $D_i\langle f \rangle$ to be the functional

$$\varphi \longmapsto -\int_{\mathbf{T}} f(x)D_i\varphi(x)\,dx.$$

The reason for the minus sign is to fit the formula for integration by parts in case f is C^1. The functional $D_i\langle f \rangle$ is then the same as $\langle D_i f \rangle$. We let as usual

$$D^p = D_1^{p_1} \cdots D_d^{p_d}$$

for any d-tuple $(p_1, \ldots, p_d)$ of integers $\geqslant 0$. We let

$$|p| = p_1 + \cdots + p_d,$$

and call $|p|$ the order of the differential operator D^p. Any finite linear combination

$$\sum \alpha_p D^p$$

with coefficients $\alpha_p \in C^\infty(\mathbf{T})$ is called a **differential operator**.

We use the scalar product

$$\langle f, g \rangle = \int_{\mathbf{T}} f(x)\overline{g(x)}\,dx.$$

Any differential operator then has an adjoint D^* with respect to this scalar product, in the obvious way. If $\varphi, \psi \in C^\infty(\mathbf{T})$, then

$$\langle D^p\varphi, \psi\rangle = (-1)^{|p|}\langle \varphi, D^p\psi\rangle.$$

One may define a Hilbert space norm on $C^\infty(\mathbf{T})$ by

$$|||\varphi|||_s^2 = \sum_{|p|<s} \langle D^p\varphi, D^p\varphi\rangle.$$

It is actually more convenient to work with an equivalent norm arising from the Fourier series as follows.

Let $\varphi \in C^\infty(\mathbf{T})$. Then φ has a Fourier series,

$$\varphi(x) = \sum c_n e^{in\cdot x}$$

where the sum is over $n \in \mathbf{Z}^d$. Integrating by parts when evaluating the Fourier coefficients, one sees that given any positive number k, the coefficients satisfy an inequality

$$|c_n| \ll \frac{1}{|n|^k}, \qquad n \to \infty,$$

and hence that the Fourier series converges absolutely to the function. If s is a positive integer, then we have trivially from the definitions

$$\sum_{n\in\mathbf{Z}^d} (1+n^2)^s |c_n|^2 \ll |||\varphi|||_s^2 \ll \sum_{n\in\mathbf{Z}^d} (1+n^2)^s |c_n|^2$$

for φ ranging over $C^\infty(\mathbf{T})$.

On $\mathbf{Z}^d$ we have a measure μ_s for $s \in \mathbf{Z}$ such that

$$\mu_s(n) = (1+n^2)^s.$$

Then

$$C^\infty(\mathbf{T}) \subset L^2(\mathbf{Z}^d, \mu_s)$$

for all $s \in \mathbf{Z}$, and $C^\infty(\mathbf{T})$ is dense in this Hilbert space, which we denote by H_s. Thus H_s by definition is the space of functions on $\mathbf{Z}^d$, written in the form of sequences $\{c_n\}$, or "formal Fourier series"

$$f(x) = \sum c_n e^{in\cdot x} = \sum f_n e^{in\cdot x}$$

such that the sum

$$\sum (1+n^2)^s |c_n|^2$$

converges. This sum is then by definition $\|f\|_s^2$. The scalar product in H_s is given by

$$\langle f, g\rangle_s = \sum f_n \overline{g_n}(1 + n^2)^s$$

where f_n and g_n are the n-th "Fourier coefficient" of f and g respectively in the above representation. The two norms $\| \ \|_s$ and $||| \ |||_s$ are equivalent on $C^\infty(\mathbf{T})$. By definition, H_s is complete. *Note that*

$$H_0 = L^2(\mathbf{T}).$$

The trigonometric polynomials, i.e. finite sums

$$\sum_{|n| < N} c_n e^{in \cdot x}$$

form a dense subspace of H_s. They are the sequences such that all but a finite number of terms are equal to 0.

Clearly, if $s < t$, then

$$\| \ \|_s < \| \ \|_t \qquad \text{and} \qquad H_s \supset H_t.$$

Intermediate inequality. *Let $r < s < t$. Given ϵ, there exists $C(\epsilon)$ such that for $f \in H_s$ we have*

$$\|f\|_s \leqslant \epsilon\|f\|_t + C(\epsilon)\|f\|_r.$$

Proof. For any positive number a, we have

$$a^s \leqslant \epsilon a^t + C(\epsilon)a^r$$

because

$$1 \leqslant \epsilon a^{t-s} + \frac{C(\epsilon)}{a^{s-r}}\ .$$

We apply this inequality to $a = (1 + n^2)$ in the definition of the norm, and we therefore find

$$\|f\|_s^2 \leqslant \epsilon\|f\|_t^2 + C(\epsilon)\|f\|_r^2.$$

Adding the cross term to the right-hand side and taking the square root yields the desired inequality.

We have a mapping of H_{-s} into the antidual space H_s^* under the scalar product

$$\langle f, g\rangle = \sum f_n \overline{g_n} = \sum f_n(1 + n^2)^{-s/2}\overline{g_n}(1 + n^2)^{s/2}.$$

(Convergence by Schwarz inequality.) Also by the Schwarz inequality we get

$$|\langle f, g\rangle| \leqslant \|f\|_{-s}\|g\|_{s}, \qquad f \in H_{-s},\ g \in H_{s}.$$

Hence we may define $\langle f\rangle$ as the antifunctional on H_s given by

$$g \mapsto \langle f, g\rangle.$$

Its norm is $\leqslant \|f\|_{-s}$, and we shall see in a moment that its norm is exactly equal to $\|f\|_{-s}$.

We define an operator D_i^2 which will correspond to

$$\frac{\partial^2 f}{\partial x_i^2} = \sum -n_i^2 f_n e^{in\cdot x}$$

on the C^∞ functions, and then $\Delta = \sum D_i^2$, so that for $f = \sum f_n e^{inx}$ we have by definition

$$(\Delta f)_n = -n^2 f_n.$$

For any real s we define the operator $(1 - \Delta)^s$ by

$$[(1 - \Delta)^s f]_n = (1 + n^2)^s f_n.$$

Then

$$\|(1 - \Delta)^{-s} f\|_s^2 = \sum |(1 + n^2)^{-s} f_n|^2 (1 + n^2)^s$$
$$= \|f\|_{-s}^2.$$

This shows that if $f \in H_{-s}$, then $(1 - \Delta)^{-s}f$ lies in H_s. Furthermore,

$$\langle f, (1 - \Delta)^{-s} f\rangle = \sum |f_n|^2 (1 + n^2)^{-s} = \|f\|_{-s}^2.$$

Taking $g = (1 - \Delta)^{-s}f$ shows that the norm of the functional $\langle f\rangle$ is exactly $\|f\|_{-s}$. Thus we have a norm-preserving embedding

$$H_{-s} \longrightarrow H_s^*$$

under the scalar product

$$(f, g) \mapsto \langle f, g\rangle.$$

If $\sum |f_n|$ converges, then the Fourier series for f converges uniformly to a continuous function which can be identified with f, and such that

$$\|f\| \leqq \sum |f_n|$$

where $\| \ \|$ is the sup norm.

Sobolev inequality. *Assume $s > d/2$. If $f \in H_s$, then $\|f\|$ is finite and there is a constant C such that for all $f \in H_s$,*

$$\|f\| \leqq C\|f\|_s.$$

Proof. We have

$$\|f\|^2 \leqq \left(\sum |f_n|\right)^2 \leqq \sum_n (1+n^2)^s |f_n|^2 \sum_m (1+m^2)^{-s}$$

by the Schwarz inequality, applied to the two functions of n,

$$|f_n|(1+n^2)^{s/2} \qquad \text{and} \qquad (1+n^2)^{-s/2}$$

over $\mathbf{Z}^d$, with measure 1 at each point. The assumption $s > d/2$ guarantees that the sum over m above converges. This proves the Sobolev inequality.

Under the assumption $s > d/2$ we see that the Fourier series of f converges uniformly, to a continuous function which may be identified with f. Thus f is continuous.

Let $D_i f$ be the formal Fourier series obtained by taking the partial derivative of the Fourier series for f formally. We have trivially

$$\|D_i f\|_{s-1} \leqq C\|f\|_s.$$

Consequently, if $s - 1 > d/2$ then $D_i f$ is continuous, and f is obtained by integrating the Fourier series for $D_i f$. Consequently f is of class C^1. Proceeding inductively, we find:

Sobolev regularity. *Let k be a positive integer. If $s > k + d/2$ and $f \in H_s$, then f is C^k and*

$$\|D^p f\| \leqq \|f\|_s, \qquad |p| \leqq k.$$

As a consequence we see that

$$\boxed{\bigcap_{s=1}^{\infty} H_s(\mathbf{T}) = C^{\infty}(\mathbf{T}).}$$

The union of all the spaces H_{-s} for $s \geqslant 0$ is called the space of **distributions** on the Torus. In the basic regularity theorem for elliptic operators, we prove that a certain distribution lies in all H_s for all s, and hence is C^∞.

We conclude this section by remarks which won't be used in the sequel, but which may be illuminating concerning the spaces H_s. Let $f \in L^2$. Then f gives rise to a functional on C^∞ functions by

$$\varphi \mapsto \int_{\mathbf{T}} f(x)\varphi(x)\, dx.$$

The functional

$$\varphi \mapsto (-1)^{|p|}\int_{\mathbf{T}} f(x)D^p\varphi(x)\, dx$$

is the distribution derivative $D^p\langle f\rangle$. We want to see directly without Fourier series that H_s for a positive integer s is the completion of the C^∞ functions under the s-norm whose square is

$$\sum_{|p|\leqslant s} \langle D^p\varphi, D^p\varphi\rangle.$$

For this purpose we now view H_s as the space of functions in L^2 whose distribution derivatives up to order s are represented by functions in L^2. This amounts to seeing that certain sequences of C^∞ functions converge in the s-norm, and is done in the next two remarks.

Remark 1. *Let $\{\varphi_n\}$ be a sequence of C^∞ functions such that $D^p\varphi_n \to g_p$ in L^2 for all $|p| \leqslant s$, and such that $\varphi_n \to f$ in L^2. Then*

$$D^p\langle f\rangle = \langle g_p\rangle.$$

Proof.

$$\begin{aligned}\langle g_p, \varphi\rangle &= \lim \langle D^p\varphi_n, \varphi\rangle\\ &= \lim \langle \varphi_n, D^{p*}\varphi\rangle\\ &= \langle f, D^{p*}\varphi\rangle.\end{aligned}$$

This proves our assertion.

Remark 2. *Given $f \in L^2$, assume that $D^p\langle f\rangle = \langle g\rangle$ for some $g \in L^2$. If $\{\alpha_n\}$ is a Dirac sequence of C^∞ functions, then*

$$D^p(\alpha_n * g) \to g$$

in L^2.

Proof. We shall prove that $D^p\langle \alpha_n * f\rangle = \langle \alpha_n * g\rangle$. From the elementary theory of Dirac sequences, we know that $\alpha_n * g$ converges to g in L^2. This proves our remark, provided that we now prove that for $\alpha \in C^\infty$ we have

$$D^p\langle \alpha * f\rangle = \langle \alpha * g\rangle.$$

Let $D = D^p$. We have:

$$\begin{aligned}
\langle \alpha * f, D^*\varphi\rangle &= \int\int \alpha(x-y)f(y)D^*\varphi(x)\,dy\,dx \\
(y \mapsto y+x) \qquad &= \int\int \alpha(-y)f(y+x)D^*\varphi(x)\,dy\,dx \\
(x \mapsto x-y) \qquad &= \int\int \alpha(-y)f(x)D^*\varphi(x-y)\,dx\,dy \\
&= \int\int \alpha(-y)g(x)\varphi(x-y)\,dx\,dy \\
&= \langle \alpha * g, \varphi\rangle,
\end{aligned}$$

as was to be shown.

The point of Remark 2 is that the differential operator can be moved inside the convolution sign when interpreted as a distribution derivative. Naturally, we have

$$D^p(\alpha * g) = (D^p\alpha) * g$$

(essentially by differentiating under the integral sign, cf. *Real Analysis*, XIV, §4, Th. 7), so that $\alpha * g$ is a smooth function because α is smooth, even though g is not.

§2. ORDINARY ESTIMATES

Let $\alpha \in C^\infty(\mathbf{T})$. We shall make estimates with a constant C_α depending only on α, and we use the notation O_α for these.

If $\varphi, \psi \in C^\infty(\mathbf{T})$, then

$$\langle D^p\varphi, \psi\rangle = (-1)^{|p|}\langle \varphi, D^p\psi\rangle,$$

so

$$\langle D^{2p}\varphi, \psi\rangle = \langle \varphi, D^{2p}\psi\rangle,$$

i.e. D^{2p} is self-adjoint, and

$$\langle \Delta\varphi, \psi\rangle = \langle \varphi, \Delta\psi\rangle.$$

Also we have

$$\langle (1-\Delta)^s\varphi, \psi\rangle = \langle \varphi, (1-\Delta)^s\psi\rangle,$$

whether s is negative or not.

We observe that the scalar product given formally by

$$\langle \varphi, \psi \rangle = \sum \varphi_n \overline{\psi_n}$$

coincides with the scalar product given by the usual integral, when φ, ψ are C^∞ functions. Hence the adjointness just stated for $(1 - \Delta)^s$ holds no matter how we interpret this scalar product. It is, of course, trivial in terms of the formal definition of the product. By the Schwarz inequality, and the integral form of the scalar product, we also find

$$\begin{aligned} |\langle D^q \alpha D^p \varphi, D^r \psi \rangle| &\leqslant \|D^q \alpha\| \, \|D^p \varphi\|_0 \, \|D^r \psi\|_0 \\ &\leqslant \|D^q \alpha\| \, \|\varphi\|_{|p|} \, \|\psi\|_{|r|}. \end{aligned}$$

We now come to various inequalities related to permuting taking derivatives and multiplying by functions, and transposing functions in a scalar product. They are easy but a little tedious.

Let D be a differential operator,

$$D = \sum \alpha_p D^p.$$

If the sum is over p with $|p| \leqslant k$, we say D has **order** $\leqslant k$. Then

$$\langle D(\alpha\varphi), \psi \rangle = \langle \alpha D\varphi, \psi \rangle + \langle E\varphi, \psi \rangle \tag{1}$$

where E is a differential operator of order $\leqslant \operatorname{ord} D - 1$, and the coefficients of E depend only on the derivatives of α.

For $s \geqslant 0$,

E 1. $\langle (1 - \Delta)^s(\alpha\varphi), \psi \rangle = \langle \alpha(1 - \Delta)^s\varphi, \psi \rangle + O_\alpha(\|\varphi\|_{s-1}\|\psi\|_s)$

Proof. In (1) we let $D = (1 - \Delta)^s$. Then D has order $2s$, and E has order $2s - 1$. Say $|p| = 2s - 1$ and β is C^∞ so that βD^p occurs in E. Write $p = q + p - q$ with $|q| = s - 1$. Then

$$\begin{aligned} \langle \beta D^p \varphi, \psi \rangle &= \langle D^p \varphi, \bar{\beta}\psi \rangle \\ &= \langle D^q \varphi, D^{p-q}(\bar{\beta}\psi) \rangle. \end{aligned}$$

Now $D^{p-q}(\bar{\beta}\psi) = \sum \gamma_r D^{p-q-r}\psi$, where the sum is taken for $|r| \leqslant p - q$, and the coefficients γ_r are derivatives of $\bar{\beta}$. Hence

$$\langle \beta D^q \varphi, \psi \rangle = O_\beta(\|\varphi\|_{s-1}\|\psi\|_s).$$

The case $|p| = 2s - 1$ is the worst that can occur, and **E 1** is proved.

E 2. $$\langle \alpha\varphi, \psi\rangle_s = \langle \varphi, \bar{\alpha}\psi\rangle_s + O_\alpha(\|\varphi\|_{s-1}\|\psi\|_s)$$

Proof. Consider first $s \geqslant 0$. We have by definition

$$\langle \alpha\varphi, \psi\rangle_s = \langle (1 - \Delta)^s(\alpha\varphi), \psi\rangle$$

and the desired inequality results from **E 1**. To deal with the negative case, let $s > 0$ and let

$$u = (1 - \Delta)^{-s}\varphi, \qquad v = (1 - \Delta)^{-s}\psi.$$

Then

$$\begin{aligned}\langle \alpha\varphi, \psi\rangle_{-s} &= \langle \alpha\varphi, (1 - \Delta)^{-s}\psi\rangle \\ &= \langle \alpha(1 - \Delta)^s u, v\rangle \\ &= \langle (1 - \Delta)^s(\alpha u), v\rangle + O_\alpha(\|u\|_{s-1}\|v\|_s)\end{aligned}$$

by **E 1**. Then

$$\begin{aligned}\langle (1 - \Delta)^s(\alpha u), v\rangle &= \langle \alpha u, (1 - \Delta)^s v\rangle \\ &= \langle (1 - \Delta)^{-s}\varphi, \bar{\alpha}\psi\rangle \\ &= \langle \varphi, \bar{\alpha}\psi\rangle_{-s}.\end{aligned}$$

As for the error term, we have

$$\begin{aligned}\|u\|_{s-1} &= \langle u, (1 - \Delta)^{s-1}u\rangle \\ &= \langle (1 - \Delta)^{-s}\varphi, (1 - \Delta)^{-1}\varphi\rangle \\ &= \langle (1 - \Delta)^{-s-1}\varphi, \varphi\rangle = \|\varphi\|_{-s-1}.\end{aligned}$$

Also $\|v\|_s = \|\psi\|_{-s}$, whence **E 2** follows.

E 3. $$\|\alpha\varphi\|_s \ll \|\alpha\|\|\varphi\|_s + O_\alpha(\|\varphi\|_{s-1})$$

Proof. First for $s \geqslant 0$. If $|p| \leqslant s$, then

$$D^p(\alpha\varphi) = \sum b_q\, D^q\alpha\, D^{p-q}\varphi$$

and by the triangle inequality,

$$\|D^p(\alpha\varphi)\|_0 \ll \sup_q \|D^q\alpha\, D^{p-q}\varphi\|_0.$$

If $|p| = s$ and $q = 0$, we get one term estimated by $\|\alpha\|\|\varphi\|_s$. If $|p| = s$ and $q \neq 0$, we get terms of order $\|\varphi\|_{s-1}$ times a constant depending on α. This takes care of the positive case.

In the negative case, let $\varphi, \psi \in C^\infty(\mathbf{T})$ and put

$$u = (1 - \Delta)^{-s}\varphi, \qquad v = (1 - \Delta)^{-s}\psi.$$

Then

$$\begin{aligned}\langle \alpha\varphi, \psi \rangle_{-s} &= \langle \alpha\varphi, (1 - \Delta)^{-s}\psi \rangle = \langle \alpha(1 - \Delta)^s u, v \rangle \\ &= \langle (1 - \Delta)^s(\alpha u), v \rangle + O_\alpha(\|u\|_{s-1}\|v\|_s) \\ &= \langle \alpha u, v \rangle_s + O_\alpha(\|u\|_{s-1}\|v\|_s).\end{aligned}$$

By the Schwarz inequality in the positive case,

$$|\langle \alpha u, v \rangle_s| \leqslant \|\alpha u\|_s\|v\|_s \ll [\|\alpha\|\|u\|_s + O_\alpha(\|u\|_{s-1})]\|v\|_s.$$

But

$$\|u\|_s = \|\varphi\|_{-s}, \qquad \|u\|_{s-1} = \|\varphi\|_{-s-1}, \qquad \|v\|_s = \|\psi\|_{-s}.$$

Hence

$$|\langle \alpha\varphi, \psi \rangle_{-s}| \ll [\|\alpha\|\|\varphi\|_{-s} + C_\alpha\|\varphi\|_{-s-1}]\|\psi\|_{-s},$$

and therefore

$$\|\alpha\varphi\|_{-s} \ll \|\alpha\|\|\varphi\|_{-s} + C_\alpha\|\varphi\|_{-s-1},$$

as desired.

E 4. *Let D be a differential operator of order $\leqslant k$. Then*

$$\|D\varphi\|_s \leqslant C_D\|\varphi\|_{s+k}$$

where C_D is a constant depending only on D. Furthermore,

$$\|D^p\varphi\|_s \ll \|\varphi\|_{s+k}.$$

Proof. Write $\varphi = \sum \varphi_n e^{in\cdot x}$. Then

$$D^p\varphi = \sum \varphi_n i^{|p|}n^p e^{in\cdot x},$$

so the estimate for $D^p\varphi$ comes out at once. The estimate for $\alpha D^p = D$ is then immediate, whence the result for an arbitrary D.

From **E 4** we see that D can be extended by continuity to a continuous linear map, again denoted by D,

$$D: H_s \longrightarrow H_{s-k},$$

for all integers s, positive or negative. Actually from **E 3** and **E 4** we have a more precise inequality, involving the leading coefficients of D, as distinguished from the others. If

$$D = \sum_{|p| \leqslant k} \alpha_p D^p,$$

we call those coefficients α_p with $|p| = k$ a leading coefficient. The next inequality **E 5** is immediate from **E 3** and **E 4**.

E 5. $$\|D\varphi\|_s \leqslant C_0(D)\|\varphi\|_{s+k} + C_1(D)\|\varphi\|_{s+k-1},$$

where $C_0(D)$ is a sup *bound for the leading coefficients of D, and $C_1(D)$ depends on all coefficients.*

E 6. *For any differential operator D of order $\leqslant k$,*

$$\langle D(\alpha\varphi), \psi\rangle_s = \langle \alpha D\varphi, \psi\rangle_s + O_{D,\alpha}(\|\varphi\|_{s+k-1}\|\psi\|_s).$$

Proof. If D is a function, the estimate is clear. It suffices to prove it for $D = D^p$, in which case we note that

$$D^p(\alpha\varphi) - \alpha D^p\varphi = \sum_{|q| \geqslant 1} c_q D^q\alpha\, D^{p-q}(\alpha\varphi),$$

and $|p - q| \leqslant s + k - 1$. By Schwarz,

$$\begin{aligned} |\langle D^q\alpha\, D^{p-q}(\alpha\varphi), \psi\rangle_s| &\leqslant \|D^q\alpha\, D^{p-q}(\alpha\varphi)\|_s\|\psi\|_s \\ &\leqslant C_{D,\alpha}\|\alpha\varphi\|_{s+k-1}\|\psi\|_s \qquad \text{by } \mathbf{E\ 4} \\ &\leqslant C_{D,\alpha}\|\varphi\|_{s+k-1}\|\psi\|_s \qquad \text{by } \mathbf{E\ 3}. \end{aligned}$$

This proves **E 6**.

E 7. *If D, D' are differential operators of order $\leqslant k$, then*

$$\begin{aligned} \langle D(\alpha\varphi), D'\psi\rangle_s &= \langle D\varphi, D'(\bar{\alpha}\psi)\rangle_s \\ &\quad + O_{D,D',\alpha}(\|\varphi\|_{s+k}\|\psi\|_{s+k-1} + \|\varphi\|_{s+k-1}\|\psi\|_k). \end{aligned}$$

Proof. This is clear by successive applications of **E 2** through **E 6**.

§3. ELLIPTIC ESTIMATES

Let

$$D = \sum_{|p| \leqq k} \alpha_p D^p$$

be a differential operator, with $\alpha_p \in C^\infty(U)$, where U is an open set in $\mathbf{R}^d$ or $\mathbf{T}^d$. We call

$$\sum_{|p|=k} \alpha_p D^p$$

its **leading term**, and k its **order**. For each point x we let

$$\sigma_D(x, \xi) = \sigma_{D,x}(\xi) = \sum_{|p|=k} \alpha_p(x)\xi^p$$

$$= \sum_{|p|=k} \alpha_p(x)\xi_1^{p_1} \cdots \xi_d^{p_d}$$

be the homogeneous polynomial in variables $\xi_1, \ldots, \xi_d$ formed by substituting ξ_j for D_j, and $\alpha_p(x)$ for α_p in the leading term of D. If $\sigma_{D,x}(\xi) \neq 0$ for all $\xi \in \mathbf{R}^d$, $\xi \neq (0, \ldots, 0)$, then we call D **elliptic**, and we call σ_D its **symbol**.

Example. The operator

$$\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$$

is elliptic. Its symbol is $\xi_1^2 + \xi_2^2$.

Example. The operator

$$\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}$$

is elliptic, and its symbol is $\xi_1 + i\xi_2$.

Example. The operator

$$y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$$

is elliptic of order 2 on the upper half plane.

Let L be an elliptic operator of order k. Then it is clear that for each x there exists a number $C_x > 0$ such that

$$|\sigma_D(x, \xi)| \geqq C_x|\xi|^k, \qquad \xi \neq 0.$$

Here, $|\xi|$ is the Euclidean norm on $\mathbf{R}^d$. In this section, we prove essentially a converse of the inequality **E 4** for elliptic operators.

Basic Estimate. *Let L be an elliptic operator of order k on $\mathbf{T}$. Let s_0 be any integer. Then for any $\varphi \in C^\infty(\mathbf{T})$ we have*

$$\|\varphi\|_{s+k} \underset{L}{\ll} \|L\varphi\|_s + \|\varphi\|_{s_0}.$$

The proof will proceed in three steps, and before giving it, we note that the extraneous integer s_0 is usually selected to be large negative. Furthermore, we shall see eventually that on a subspace of finite codimension, one can eliminate the term $\|\varphi\|_{s_0}$ completely from the estimate. In any case, this term makes the theorem non-trivial only if $s_0 \leqslant s + k - 1$, which we assume from now on. Also, we note that the constant implied in the inequality $\ll$ depends on L.

The crucial case of the basic estimate is when L has constant coefficients. We take care of this in the first lemma.

Lemma 1. *Let L have constant coefficients, and be homogeneous of order k. Then*

$$\|\varphi\|^2_{s+k} \underset{L}{\ll} \|L\varphi\|^2_s + \|\varphi\|^2_{s_0}.$$

Proof. We have

$$\|L\varphi\|^2_s = \sum |\varphi_n|^2 |\sigma_L(n)|^2 (1 + n^2)^s$$

$$\gg \sum |\varphi_n|^2 (n^2)^k (1 + n^2)^s$$

$$\|\varphi\|^2_{s_0} = \sum |\varphi_n|^2 (1 + n^2)^{s_0}$$

Hence

$$\|L\varphi\|^2_s + \|\varphi\|^2_{s_0} \gg \sum |\varphi_n|^2 (1 + n^2)^{s+k} = \|\varphi\|^2_{s+k}.$$

This proves Lemma 1.

Lemma 2. *Given an elliptic operator L of order k on T, there exists $\delta = \delta(L)$ such that if φ has support in a ball of radius $< \delta$, then*

$$\|\varphi\|^2_{s+k} \underset{L}{\ll} \|L\varphi\|^2_s + \|\varphi\|_{s+k-1}\|\varphi\|_{s+k}.$$

Proof. Let $L = \sum \alpha_p D^p$. By uniform continuity, we can pick δ so small that the oscillation of each α_p in a ball of radius δ is less than ϵ. Let x_0 be the center of the ball in which φ has its support. Let

$$L_0 = \sum_{|p|=k} \alpha_p(x_0)D^p$$

be the homogeneous elliptic operator with constant coefficients obtained by evaluating the coefficients of the leading term of L at x_0. Then

$$|\alpha_p(x) - \alpha_p(x_0)| < \epsilon,$$

for x in the ball $B_\delta(x_0)$. Now by the previous result,

$$\|\varphi\|_{s+k}^2 \ll \|L_0\varphi\|_s^2 + \|\varphi\|_{s_0}^2.$$

Note that $\|\varphi\|_{s_0}^2 \leqslant \|\varphi\|_{s+k-1}\|\varphi\|_{s+k}$. We have to compare $\|L\varphi\|_s^2$ and $\|L_0\varphi\|_s^2$. Their difference is

$$\langle (L_0 - L)\varphi, L_0\varphi\rangle_s + \langle L\varphi, (L_0 - L)\varphi\rangle_s.$$

We estimate the first term, say, by Schwarz. Let $a_p = \alpha_p(x_0)$. The absolute value of the p-th term is estimated by

$$\begin{aligned}
\langle(\alpha_p - a_p)D^p\varphi, L_0\varphi\rangle_s &\leqslant \|(\alpha_p - a_p)D^p\varphi\|_s\|L_0\varphi\|_s \\
\text{(by E 5)} \qquad &\leqslant [\epsilon\|\varphi\|_{s+k} + C_1\|\varphi\|_{s+k-1}]\,\|L_0\varphi\|_s \\
&\leqslant \epsilon C_2(L)\|\varphi\|_{s+k}^2 + C_3(L)\|\varphi\|_{s+k-1}\|\varphi\|_{s+k}.
\end{aligned}$$

By picking δ and then ϵ so that $\epsilon C_2(L) \leqslant 1/2$, we get

$$\frac{1}{2}\|\varphi\|_{s+k}^2 \leqslant C_4\|L\varphi\|_s^2 + C_5\|\varphi\|_{s+k-1}\|\varphi\|_{s+k}.$$

Our result follows by dividing throughout by $\|\varphi\|_{s+k}$, and using the intermediate inequality of §1, with the three numbers

$$s_0 < s + k - 1 < s + k.$$

We get a small factor times $\|\varphi\|_{s+k}$ on the right-hand side, which can be transposed to the left-hand side, leaving only the s_0 term, as desired.

Our final step is to use a partition of unity. Let $\{\alpha_i\}$ be a finite family of C^∞ functions on $\mathbf{T}$ such that

$$\sum \alpha_i^2 = 1,$$

and such that each α_i has support in a ball of radius $\delta = \delta(L)$ as in Lemma 2. (The usual construction of a partition of unity shows that the square roots can be taken, and yield C^∞ functions.) For any $\varphi \in C^\infty(\mathbf{T})$, the product $\alpha_i\varphi$ has such support also. We use $\ll$ to refer to constants depending on the partition of unity and L. (The partition of unity depends itself on L only.) Then

$$\begin{aligned}
\|\varphi\|_{s+k}^2 &= \langle \varphi, \varphi\rangle_{s+k} = \sum \langle \alpha_i^2\varphi, \varphi\rangle_{s+k} \\
\text{(by E 2)} \quad &= \sum \langle \alpha_i\varphi, \alpha_i\varphi\rangle_{s+k} + O_L(\|\varphi\|_{s+k}\|\varphi\|_{s+k-1}) \\
&\ll \sum \|L(\alpha_i\varphi)\|_s^2 \\
&\qquad + \sum \|\alpha_i\varphi\|_{s+k}\|\alpha_i\varphi\|_{s+k-1} + \|\varphi\|_{s+k}\|\varphi\|_{s+k-1} \\
\text{(by E 7)} \quad &\ll \sum \langle L(\alpha_i^2\varphi), L\varphi\rangle_s + \|\varphi\|_{s+k}\|\varphi\|_{s+k-1} \\
&\leqslant C_7(L)\|L\varphi\|_s^2 + C_8(L)\|\varphi\|_{s+k}\|\varphi\|_{s+k-1}.
\end{aligned}$$

Observe that $\|L\varphi\|_s \leqslant C_9(L)\|\varphi\|_{s+k}$. Divide both sides of the inequality by $\|\varphi\|_{s+k}$. Select ϵ such that

$$\epsilon C_8(L) < \tfrac{1}{2}.$$

Use the inequality

$$\|\varphi\|_{s+k-1} \leqslant \epsilon\|\varphi\|_{s+k} + C(\epsilon)\|\varphi\|_{s_0}.$$

Subtract $\frac{1}{2}\|\varphi\|_{s+k}$ from both sides. We get the basic inequality.

Remark. The basic inequality extends by continuity to elements of H_{s+k}. Indeed, if $h \in H_{s+k}$ and $\{\varphi_n\}$ is a sequence in $C^\infty(\mathbf{T})$ such that $\varphi_n \to h$ in H_{s+k}-norm, then

$$\|\psi_n\|_{s+k} \ll \|L\psi_n\|_s + \|\psi_n\|_{s_0}$$

implies that $\{\varphi_n\}$ is H_{s+k}-convergent also, and we get

$$\|h\|_{s+k} \underset{L}{\ll} \|Lh\|_s + \|h\|_{s_0},$$

or shifting indices,

$$\|f\|_s \underset{L}{\ll} \|Lf\|_{s-k} + \|f\|_{s_0}$$

for $f \in H_s$.

§4. COMPACTNESS AND REGULARITY ON THE TORUS

In this section, we eliminate the s_0 term from the basic inequality, at the cost of going to a subspace of finite codimension.

Lemma 1. *Let $r < s$. Then the unit ball in H_s is relatively compact in H_r, i.e. is totally bounded in H_r.*

Proof. Let

$$f = \sum f_n e^{in \cdot x}$$

be in the unit ball of H_s, so that

$$\sum |f_n|^2 (1 + n^2)^s \leqq 1.$$

Pick N such that $(1 + N^2)^{r-s} < \epsilon$. Write

$$f = \sum_{|n| \leqq N} f_n e^{in \cdot x} + \sum_{|n| > N} f_n e^{in \cdot x}.$$

Then the second sum is ϵ-close to 0 in H_{s-1}-norm, i.e.

$$\sum_{|n| > N} |f_n|^2 (1 + n^2)^r \leqq \sum_{|n| > N} |f_n|^2 (1 + n^2)^s \frac{1}{(1 + N^2)^{s-r}} < \epsilon.$$

The first sum belongs to a bounded set of the finite dimensional space of functions whose only non-zero coordinates are among those with $|n| \leqq N$. Such a set can be covered by a finite number of ϵ-balls for the H_{s-1}-metric. This proves total boundedness, i.e. it proves Lemma 1. It also proves the slightly stronger version stated in the next theorem.

Theorem 1. *Let $r < s$. Let $H_s(N)$ consist of all those elements $f \in H_s$ such that $f_n = 0$ if $|n| \leqq N$. Given ϵ, there exists N such that for all $f \in H_s(N)$ we have*

$$\|f\|_r \leqq \epsilon \|f\|_s.$$

The inclusion of H_s in H_r is a compact operator.

Note that $H_s(N)$ is closed and of finite codimension in H_s. Also, it is generated by smooth functions.

Remark. Theorem 1 was proved by the usual ad hoc manipulations with the formal Fourier series. It is illuminating to realize that it holds in somewhat greater generality in the following context.

Let H, E be Hilbert spaces and

$$A: H \longrightarrow E$$

a compact linear map. Let $\{e_i\}$ $(i = 1, 2, \dots)$ be an orthonormal basis in H. Let $H(N)$ be the closed subspace generated by the e_i with $i \geqslant N$. Given ϵ there exists N such that for all $h \in H(N)$ we have

$$\|Ah\|_E \leqslant \epsilon \|h\|_H.$$

Indeed, if this is not the case, we can find a unit vector $h_n \in H(n)$ with $n \longrightarrow \infty$, such that $\|Ah_n\|_E \geqslant \epsilon$. But $h_n \longrightarrow 0$ weakly, so $Ah_n \longrightarrow 0$ strongly, a contradiction.

We have used here the fact that an operator $A: H \longrightarrow E$ is compact if and only if A maps weakly convergent sequences into strongly convergent sequences. Furthermore, if A is compact and $v_n \longrightarrow 0$ weakly in H, then $Av_n \longrightarrow 0$ strongly in E.

We recall the easy proof. Assume that A is compact. Let $\{v_n\}$ be weakly convergent, say to v. Considering $\{v_n - v\}$ we may assume without loss of generality that $v_n \longrightarrow 0$ weakly. For each $w \in H$, $\langle v_n, w\rangle \longrightarrow 0$. If $\{Av_n\}$ is not convergent, then some subsequence $\{Av_{n_k}\}$ does not converge, contradicting compactness. Furthermore, $\{Av_n\}$ converges to 0, for if $Av_n \longrightarrow w \neq 0$, there exists a functional λ on E such that $\lambda(w) \neq 0$. Then $\lambda \circ A$ is a functional on H, and by assumption, $\lambda(Av_n) \longrightarrow 0$, contradiction. Conversely, assume that A maps weakly convergent sequences into strongly convergent sequences. Let $\{v_n\}$ be a sequence such that $|v_n| \leqslant 1$. Since the unit ball is weakly compact in Hilbert space, there is a weakly convergent subsequence $\{v_{n_k}\}$ and $\{Av_{n_k}\}$ converges. So A is a compact operator.

After this slightly general aside, we return to our special situation.

Theorem 2. *Let L be an elliptic operator on $\mathbf{T}$, of order k. Given s there exists N such that L induces a topological linear isomorphism from $H_s(N)$ to its image in H_{s-k}, i.e. we have*

$$\|h\|_s \underset{L}{\ll} \|Lh\|_{s-k}, \qquad \text{all } h \in H_s(N).$$

Furthermore, the space of $f \in H_{s-k}$ such that $\langle f, L\varphi\rangle = 0$ for all

$\varphi \in C^\infty(\mathbf{T})$ is finite dimensional, so the closure of the image $LC^\infty(\mathbf{T})$ in H_{s-k} has finite codimension.

Proof. By Theorem 1 and the basic inequality, given ϵ there is some N such that for $h \in H_s(N)$ we have

$$\begin{aligned}\|h\|_s &\leqslant C_1\|Lh\|_{s-k} + C_2\|h\|_{s_0}\\ &\leqslant C_1\|Lh\|_{s-k} + C_2\epsilon\|h\|_s.\end{aligned}$$

Pick ϵ such that $C_2\epsilon \leqslant \dfrac{1}{2}$. We get the first inequality of our theorem. Since trivially $\|Lh\|_{s-k} \ll \|h\|_s$, we get the first assertion also. In particular, the kernel of L in H_s is finite dimensional. To show that the image under L of the test functions has finite codimensional closure, let $h \in H_{-s+k}$ and suppose that

$$\langle h, L\varphi\rangle = 0, \qquad \text{all } \varphi \in C^\infty(\mathbf{T}).$$

We have $\langle L^*h, \varphi\rangle = 0$, so $L^*h = 0$, and again h has to be in a finite dimensional space, so that last assertion follows and the theorem is proved.

Theorem 3. *Let $h \in H_t$ for some t, and let L be an elliptic operator on $\mathbf{T}$, of order k. If $L\langle h\rangle = \langle g\rangle$ as functionals on $C^\infty(\mathbf{T})$, and $g \in H_s$, then $h \in H_{s+k}$. In particular, if $g \in C^\infty(\mathbf{T})$, then $h \in C^\infty(\mathbf{T})$.*

Proof. For sufficiently large N, we have

$$\|f\|_{-s} \ll \|L^*f\|_{-s-k}, \qquad \text{all } f \in H_{-s}(N).$$

By Theorem 2, for $\varphi \in H_{-s}(N)$ we have

$$\begin{aligned}\langle h, L^*\varphi\rangle &= \langle g, \varphi\rangle\\ &\leqslant \|g\|_s\,\|\varphi\|_{-s}\\ &\ll \|g\|_s\,\|L^*\varphi\|_{-s-k}.\end{aligned}$$

Hence $\langle h\rangle$ is continuous on $L^*C^\infty(\mathbf{T})$. The closure of $L^*C^\infty(\mathbf{T})$ in H_{-s-k} is finite codimensional by Theorem 2. The space generated by $L^*C^\infty(\mathbf{T})$ and a finite number of C^∞ functions is therefore dense in H_{-s-k}. Hence $\langle h\rangle$ is continuous on a dense subspace of H_{-s-k}, and is therefore induced by an element of the dual space, which is none other than H_{s+k}. Hence $h \in H_{s+k}$, as was to be shown.

Remark. The above proof of the regularity theorem is essentially due to

PETER LAX, "*On Cauchy's problem for hyperbolic equations and the differentiability of solutions of elliptic equations,*" *Comm. Pure App. Math.* **8** (1955), *pp.* 615-633.

However, Lax works with a slightly different definition of ellipticity (nowadays forgotten), and indicates only various parts of the proof. Nirenberg gave an exceedingly good self-contained exposition in

L. NIRENBERG, "*On elliptic partial differential equations,*" CIME Conference, Il principio di minimo e sue applicazioni alle equazioni funzionali, Rome, (1959), pp. 1-58.

However, instead of the short argument used in Theorem 3, Nirenberg uses "difference quotients" which I did not like. The finite dimensionality of the kernel and cokernel of the elliptic operator is also not brought out in the Lax and Nirenberg papers. Nirenberg's exposition has been copied several times since, e.g. by Bers and John in their *PDE* book, by Friedman in his *PDE* book, and Warner in his book on differential manifolds. On the whole, I thought it would still be worthwhile to give a complete proof here.

§5. REGULARITY IN EUCLIDEAN SPACE

Let U be an open set in Euclidean space $\mathbf{R}^d$. We now consider differential operators on U. If f is locally L^2 on U, it gives rise to the usual functional $\langle f\rangle$ on test functions, i.e. on $C_c^\infty(U)$. We shall see that the regularity theorem on $\mathbf{T}$ easily implies the regularity theorem on U.

Theorem 4. *Let f be locally L^2 on an open set U in $\mathbf{R}^d$. Let $g \in C^\infty(U)$, and let L be an elliptic operator of order k on U with C^∞ coefficients. Assume that $L\langle f\rangle = \langle g\rangle$. Then $f \in C^\infty(U)$.*

Proof. Let $x_0 \in U$. Let

$$L = \sum \alpha_p D^p \qquad \text{and} \qquad L_0 = \sum \alpha_p(x_0) D^p.$$

It suffices to prove that f is C^∞ in a neighborhood of x_0. Let $\alpha \in C_c^\infty(U)$ be equal to 1 in a ball of small radius r around x_0, and equal to 0 outside a ball of radius $2r$. Let

$$M = \alpha L + (1 - \alpha)L_0.$$

Then M is elliptic, has constant coefficients outside a small ball. Let α_1 be a C^∞ function which is equal to 1 near x_0, but decreases to 0 outside a ball of radius $r/2$. Then for $\varphi \in C_c^\infty(U)$ we have (the integrals are taken on all of $\mathbf{R}^d$ but the integrands have compact support):

$$\begin{aligned}\langle M\langle \alpha_1 f\rangle, \varphi\rangle &= \int \alpha_1 f M^*\varphi \\ &= \int \alpha_1 f L^*(\alpha\varphi) + \int \alpha_1 f L_0^*((1-\alpha)\varphi) \\ &= \int \alpha_1 f L^*(\alpha\varphi) \\ &= \int f L^*(\alpha_1\alpha\varphi) + \int f D^*\varphi,\end{aligned}$$

where D^* is a differential operator of order $< k$, whose coefficients have support contained in $\operatorname{supp}(\alpha)$. Hence we get the further equality

$$= \langle \alpha_1 g, \varphi\rangle + \langle f, D^*\varphi\rangle.$$

The relation

$$\int \alpha_1 f M^*\varphi = \langle \alpha_1 g, \varphi\rangle + \langle f, D^*\varphi\rangle$$

holds for periodic φ since each side depends only on the value of φ in $\operatorname{supp}(\alpha)$. Let $\beta \in C^\infty(\mathbf{T})$ be equal to 1 on $\operatorname{supp}(\alpha)$, and 0 outside a small ball. Then

$$\langle \beta f, D^*\varphi\rangle = \langle f, D^*\varphi\rangle$$

for all periodic φ. We can extend all of our "cut-off" objects by periodicity to functions on $\mathbf{T}$. In other words, we can find

$$g_\pi \in C^\infty(\mathbf{T}), \qquad (\beta f)_\pi \in L^2(\mathbf{T}),$$

an elliptic operator M_π on $\mathbf{T}$, $(\alpha_1 f)_\pi \in L^2(\mathbf{T})$, D_π of order $\leqslant k-1$ on $\mathbf{T}$ equal to g, βf, M, f, D respectively in a small neighborhood of x_0, such that

$$M_\pi\langle(\alpha_1 f)_\pi\rangle = \langle(\alpha_1 g)_\pi\rangle + D_\pi\langle(\beta f)_\pi\rangle = \langle h\rangle$$

for some element $h \in H_{-k+1}(\mathbf{T})$. Hence $(\alpha_1 f)_\pi \in H_1$ by the regularity theorem on $\mathbf{T}$.

We now repeat, and find that

$$M_\pi\langle(\alpha_1 f)_\pi\rangle = \langle h\rangle \quad \text{with some} \quad h \in H_{1-(k-1)} = H_{2-k}.$$

Hence $(\alpha_1 f)_\pi \in H_2$. Continuing in this way, we see that $(\alpha_1 f)_\pi \in H_\infty$, in other words, $(\alpha_1 f)_\pi$ gives rise to the same functional as a C^∞ function on $\mathbf{T}$, and is equal to such a function almost everywhere. Furthermore, $(\alpha_1 f)_\pi$ is equal to f locally at x_0. Hence we have proved what we wanted, that f is C^∞ locally at x_0.

If in the regularity theorem we make the assumption that g is real analytic, then it is true that f is also real analytic. The proof, by Morrey and Nirenberg, is reproduced elegantly and briefly in

Bers et al, *PDE*, Proceedings of the summer seminar, Boulder, Colorado, lecture by Bers and Schechter, *Interscience* (New York, 1964), pp. 207–210.

So the proof is only three pages long, and consists of a series of estimates to show that the C^∞ function obtained has a convergent power series. The proof actually uses parts of the C^∞ theory, and the basic estimate.

We stated the regularity in Theorem 4 for a locally L^2 function f. Actually the proof is equally applicable for any distribution. We recall that a **distribution** T on the open set U is a linear map

$$T\colon C_c^\infty(U) \to \mathbf{C}$$

having the following property. Given a compact set K in U, there exists a finite number of differential operators $D_1, \ldots, D_m$ such that if $\varphi \in C_c^\infty(U)$ and supp φ is contained in K, then

$$|T\varphi| \leqq \max_i \|D_i\varphi\|,$$

where $\| \;\|$ is the sup norm. The definition of a distribution on the torus $\mathbf{T}$ is the same, mutas mutandis, except that we don't have to mention the compact set K, and we take $\varphi \in C^\infty(\mathbf{T})$, i.e. we take φ to be periodic. It is then clear that the distributions on $\mathbf{T}$ are exactly the elements of the spaces H_s for all s. The regularity theorem on $\mathbf{R}^d$ then states:

Theorem 5. *Let T be a distribution on U. Let $g \in C^\infty(U)$, and let L be an elliptic operator of order k on U with C^∞ coefficients. Assume that $LT = g$. Then $T = T_h$ for some $h \in C^\infty(U)$.*

Proof. The same as that of Theorem 4. Instead of writing the integral of a function times f, i.e. instead say of writing

$$\int \alpha_1 f M^* \varphi,$$

we write

$$T(\alpha_1 M^* \varphi).$$

There is no other change.

Appendix 5
Weak and Strong Analyticity

§1. COMPLEX THEOREM

Theorem 1. *Let E be a Banach space over the complex numbers, and let Λ be a subset of the dual space E' which is norm determining, i.e. for $v \in E$ we have*

$$|v| = \sup \frac{|\lambda v|}{|\lambda|}, \qquad \lambda \in \Lambda, \lambda \neq 0$$

Let U be open in $\mathbf{C}$ and let $f : U \longrightarrow E$ be a mapping such that:

i) *For each $\lambda \in \Lambda$, the function $s \longmapsto \lambda \circ f(s)$ is analytic.*

ii) *For each compact set $K \subset U$, we have*

$$\sup_{s \in K} |f(s)| < \infty.$$

Then f is analytic.

Proof. (Cf. A. E. Taylor, *Functional Analysis*, p. 206.) We use the fact that we are in the complex case, and prove that the Newton quotients form a Cauchy family for $h \longrightarrow 0$. In other words, we prove that f is differentiable. It suffices to prove that

$$\left| \frac{f(s+h) - f(s)}{h} - \frac{f(s+k) - f(s)}{k} \right| < \frac{4M|h-k|}{r^2}$$

where M is a bound for f in a disc of radius r around s, and $|h|, |k| < r/2$. Since Λ is norm determining, it suffices to prove our inequality with $\lambda \circ f$ replacing f, and hence we may assume that f is complex valued, with values bounded by M, and we may also assume that f is analytic. In that case, Cauchy's formula or integral forms of the mean value theorem show that the desired inequality holds. The constant 4 is of course irrelevant. Any fixed number would do as well. The proof follows from the identity

$$\frac{f(s+h)-f(s)}{h}-\frac{f(s+k)-f(s)}{k}$$
$$=\frac{1}{2\pi i}\int_C \frac{f(z)(h-k)}{(z-s-h)(z-s-k)(z-s)}\,dz,$$

where C is a circle around s, of radius r.

In some applications, e.g. those in Chapter XIV, §11, one wants to prove analyticity of the kernel starting from the analyticity of the operator, in their dependence on s. For that one does not know a priori that the kernels are continuous in s, and one needs an additional lemma.

Lemma. *Let D be the open disc centered at 0 in $\mathbf{C}$, and say of radius 1. Let $f \in \mathfrak{L}^1_{\text{loc}}(D)$, and assume that for (Lebesgue) almost all pairs (a, r) with $a \in \mathbf{C}$, $r > 0$ and $|a| + r < 1$, we have*

$$\int_{C_r(a)} f\,ds = 0,$$

where $C_r(a)$ is the circle of center a and radius r. Then there exists a holomorphic function which is equal to f almost everywhere on D.

Proof. (I owe this proof to Dale Peterson.) For $n > 2$ define

$$f_n\colon \left(1-\frac{1}{n}\right)D \to \mathbf{C}$$

by

$$f_n(w) = \frac{n^2}{\pi}\iint_{\frac{1}{n}D} f(w+x+iy)\,dx\,dy.$$

Permuting integrals by Fubini shows that

$$\int_{C_r(a)} f_n\,ds = 0$$

for almost all (a, r) with $|a| + r < 1 - \dfrac{1}{n}$, $r > 0$. Since $f \in \mathfrak{L}^1_{\text{loc}}(D)$, one can take a limit under an integral sign to show that f_n is continuous on $(1 - \dfrac{1}{n})D$. By the usual criterion it follows that f_n is holomorphic on $(1 - \dfrac{1}{n})D$. On $\frac{1}{2}\,D$

we have the estimate

$$\begin{aligned}\iint\limits_{\frac{1}{2}D} |f - f_n|\,dx\,dy &= \iint\limits_{\frac{1}{2}D} \frac{n^2}{\pi}\left|\iint\limits_{\frac{1}{n}D} [f(w) - f(w + x + iy)]\,dx\,dy\right| dw \\ &\leqq \iint\limits_{\frac{1}{n}D} \frac{n^2}{\pi}\left[\iint\limits_{\frac{1}{2}D} |f(w) - f(w + x + iy)|\,dw\right] dx\,dy \\ &\leqq \sup \iint\limits_{\frac{1}{2}D} |f(w) - f(w + x + iy)|\,dw,\end{aligned}$$

where the sup is taken for all $x + iy \in \frac{1}{n}D$. By the dominated convergence theorem, as $n \to \infty$ the expression on the right tends to 0, thereby proving that $\{f_n\}$ is L^1-Cauchy on $\frac{1}{2}D$. By Lemma 1 of Chapter VII we know that $\{f_n\}$ converges uniformly on compact sets, whence converges uniformly to a holomorphic function, which is equal to f almost everywhere on $\frac{1}{2}D$. We have therefore proved our assertion locally in the neighborhood of a point, which suffices for our lemma, after making translations.

The lemma is used in contexts where one wants to deduce the analyticity of a kernel defining the operators, as in Theorem 12, Chapter XIV, §11.

Observe that the set Λ is *not* assumed to be a linear subspace of the dual space (actually a not too serious matter), but more importantly is *not closed.* In this respect, the complex theorem differs from the real theorem given in the next section, and these theorems find different applications in different contexts where neither seems to cover the other.

Even though the next theorem is not used in the book, it is illuminating to include it, and I owe it to J. Gamlen. The proof of Theorem 12 in Chapter XIV, §11 is modeled on it.

Let Z be a locally compact space with a positive measure μ. We write dz instead of $d\mu(z)$. Let U be an open set in $\mathbf{C}$, and let

$$s \mapsto R_s = R(s), \qquad s \in U,$$

be a family of bounded operators in $L^2(Z, \mu)$. We assume that each operator R_s can be defined by a kernel

$$r(z, z'; s)$$

where

$$r : Z \times Z \times U \to \mathbf{C}$$

is (Borel) measurable, i.e.

$$R_s f(z) = \int_Z r(z, z'; s) f(z')\,dz'$$

for $f \in \mathcal{L}^2(Z)$.

Theorem 2. *Assume that:*

i) *For* $(\mu \otimes \mu)$*-almost all* z, z' *the function*

$$s \longmapsto r(z, z'; s)$$

is analytic.

ii) *For any compact subset* $K \subset U$ *there exists a kernel*

$$M_K(z, z') \geqslant \sup_{s \in K} |r(z, z'; s)|$$

which when convolved with a function in $L^2(Z)$ *yields a function in* $L^2(Z)$.

Then $s \longmapsto R_s$ *is weakly analytic, in the sense that for each* $f, g \in \mathfrak{L}^2(Z)$, *the function*

$$s \longmapsto \langle R_s f, g \rangle$$

is analytic.

Proof. By the dominated convergence theorem applied to the function

$$r(z, z'; s) f(z') \overline{g(z)},$$

dominated by

$$M_K(z, z') |f(z')| |g(z)|$$

which lies in $\mathfrak{L}^1(Z \times Z)$, we conclude that $s \longmapsto \langle R_s f, g \rangle$ is continuous. It suffices now to show that for an arbitrary circle C surrounding a disc in U, we have

$$\int_C ds \int_Z \int_Z r(z, z'; s) f(z') \overline{g(z)} \, dz' \, dz = 0. \tag{1}$$

The integrand is majorized by

$$M_C(z, z') |f(z')| |g(z)|.$$

The assumption that M_C maps $L^2(Z)$ into itself implies that

$$\int_Z \int_Z M_C(z, z') |f(z')| |g(z)| \, dz' \, dz$$

exists. By Fubini's theorem, the expression under the integral sign is in $\mathfrak{L}^1(Z \times Z)$, and therefore the triple integral (1) can be taken in any order. Using hypothesis (i), we conclude that the integral is 0, as was to be shown.

Corollary. *Assumptions being as in Theorem 2, the operator valued map* $s \mapsto R_s$ *is analytic.*

Proof. This is a special case of Theorem 1.

In the complex case, complex differentiability implies the existence of a power series expansion. The proof for Banach valued mappings is the same as for complex valued functions. In the real case, analyticity must be defined in terms of the power series.

The fact that we dealt with a complex variable in Theorem 2 was essential. I owe to Gamlen also the following counterexample in the real case. We take $H = \ell^2(\mathbf{Z})$. Let

$$r(m, n; s) = \frac{e^{i(m+n)s}}{m^2 n^2} \qquad \text{for } m, n \neq 0$$

For $f, g \in \ell^2(\mathbf{Z})$ we have

$$\langle R(s)f, g\rangle = \sum_{mn \neq 0} \frac{e^{i(m+n)s}}{m^2 n^2} f(m)\overline{g(n)}\,.$$

Take

$$f = (\ldots, 0, 1, 0, \ldots)$$

$$g(n) = 1/n.$$

Then

$$\langle R(s)f, g\rangle = \sum_{n \neq 0} \frac{e^{ins}}{n^2}$$

is not analytic.

§2. REAL THEOREM

On the other hand, we have the following theorem of Browder, valid in the real analytic case ("Analyticity and partial differential equations," *Am. J. Math.* **84** (1962), pp. 666–710).

Theorem 3. *Let U be an open set in* $\mathbf{R}^d$ *and let*

$$f : U \to E$$

be a mapping into a real or complex Banach space E. Let H_1, H_2 be real or complex Banach spaces, and let

$$E \times H_1 \times H_2 \to \mathbf{R} \text{ or } \mathbf{C}, \quad (u, v, w) \mapsto \langle u, v, w\rangle$$

be a continuous trilinear map which induces an isometric embedding of E into the Banach space of bilinear forms $\mathrm{Bil}(H_1, H_2)$, i.e. *such that if* $u \in E$, *then* $|u| = \sup |\langle u, v, w\rangle|$ *for* $(v, w) \in H_1 \times H_2$, $|v|, |w| \leqslant 1$. *Assume that for each* $(v, w) \in H_1 \times H_2$ *the function*

$$x \longmapsto \langle f(x), v, w\rangle$$

is analytic. Then f is analytic.

Proof. The theorem is local, and so we may assume that U is a disc around the origin, with coordinates $x = (x_1, \ldots, x_d)$. We recall the trivial fact that a power series

$$\sum_p a_p x_1^{p_1} \cdots x_d^{p_d} = \sum_p a_p x^p$$

with coefficients a_p in a Banach space converges absolutely in a neighborhood of 0 if and only if there exists $C > 0$ and $r > 0$ such that for all p we have

$$|a_p| \leqslant Cr^{|p|}.$$

If this is the case, then the power series defines a C^∞ function, say $g(x)$, and

$$a_p = \frac{D^p g(0)}{p!}$$

where $p! = p_1! \cdots p_d!$.

Let B be the unit ball in $H_1 \times H_2$ (with the sup norm), and for $(v, w) \in B$ let $\lambda_{v,w}$ be the functional

$$\lambda_{v,w}(u) = \langle u, v, w\rangle.$$

By assumption, each function $\lambda \circ f$ is analytic for $\lambda \in B$, and therefore has a power series expansion, such that the derivatives satisfy

$$|D^p(\lambda \circ f)(0)| \leqslant C(\lambda) r(\lambda)^{|p|} p!.$$

The constant $C(\lambda)$ and the "radius" $r(\lambda)$ depend on λ, and we first show that they can be selected independently of λ. For each positive integers m, n let $B_{m,n}$ be the subset of $\lambda \in B$ such that for all p we have

$$|D^p(\lambda \circ f)(0)| \leqslant mn^{|p|} p!.$$

We contend that this subset is closed. It suffices to prove that for any x mapping

$$\lambda \longmapsto D^p(\lambda \circ f)(x)$$

is continuous (let $x = 0$). We do this by induction on $|p|$. It is clear for $|p| = 0$. Assume it for $|p|$, and let $|q| = |p| + 1$. Let e be a unit vector in the direction of the new partial derivative. Then

$$D^q(\lambda \circ f)(x) = \lim_{h \to 0} \frac{D^p(\lambda \circ f)(x + he) - D^p(\lambda \circ f)(x)}{h}.$$

We let h range over a sequence $\{h_n\}$ tending to 0. Each map

$$\lambda \mapsto \frac{D^p(\lambda \circ f)(x + he) - D^p(\lambda \circ f)(x)}{h}$$

is a continuous linear map by induction hypothesis. By the uniform boundedness theorem, the limit taken over the h_n for $n \to \infty$ is also continuous linear, thereby proving our contention.

We see that B is the union of the closed sets $B_{n,m}$, and consequently some $B_{n,m}$ contains an open ball in the product space by Baire's theorem. (Cf. *Real Analysis*.) If this open ball is centered at a point (v_1, v_2), the differences of vectors (w_1, w_2) near (v_1, v_2) will lie in B, and will contain a small ball around $(0, 0)$. This implies that we can pick our constants $C(\lambda)$, $r(\lambda)$ independently of $\lambda_{v,w}$ for (v, w) in a small ball in $H_1 \times H_2$.

Let $b_p : H_1 \times H_2 \to \mathbf{R}$ or $\mathbf{C}$ be the bilinear map

$$b_p : (v, w) \mapsto \frac{D^p(\lambda_{v,w} \circ f)(0)}{p!}.$$

The uniform estimate above implies that b_p is continuous, i.e.

$$b_p \in \mathrm{Bil}(H_1 \times H_2),$$

and there are constants C, r such that for all p,

$$|b_p| \leqq Cr^{|p|}.$$

Therefore the series

$$\sum_p b_p x^p$$

converges in a neighborhood of 0 in $\mathbf{R}^d$. Evaluating at enough

$$(v, w) \in H_1 \times H_2$$

shows that

$$f(x) = \sum_p b_p x^p$$

for all x close to 0 in $\mathbf{R}^d$. Hence f is analytic as a map into $\mathrm{Bil}(H_1 \times H_2)$. In particular, f is C^∞, and

$$b_p = \frac{D^p f(0)}{p!} .$$

Since E is a closed subspace of $\mathrm{Bil}(H_1 \times H_2)$, it follows that $b_p \in E$. This proves our theorem.

Remarks. The theorem is used in several contexts. First, when $E = \mathrm{End}(H)$ is the Banach space of endomorphisms of a Hilbert space H, and when the trilinear map is the natural one,

$$(A, v, w) \mapsto \langle Av, w \rangle$$

for $A \in \mathrm{End}(H)$, and $v, w \in H$.

Second, in the representation theory when one wants to prove that certain vectors are analytic, one usually does this in cases where they satisfy an elliptic differential equation, say weakly. In that case one can apply first the regularity theorem in the analytic case, to the functions obtained by composig a given map with functionals, and then one can apply Theorem 3 to conclude that the original map $x \mapsto \pi(x)v$ in the Hilbert space is analytic.

Bibliography

[AK] N. Akniezer and I. Glazman, *Theory of Linear Operators in Hilbert Space*, Translated from the Russian, New York: Ungar, 1961.

[AR] J. Arthur, *Harmonic Analysis of Tempered Distributions on Semisimple Lie Groups of Real Rank One*, New Haven: Yale, 1970.

[BA] V. Bargmann, "Irreducible unitary representations of the Lorentz group," *Ann. Math.* **48** (1947), pp. 568–640.

[BE, GE] Berezin-Gelfand-Graev-Naimark, "Group representations," *AMS transl.*, pp. 325–353.

[BR] F. Bruhat, "Représentations des groupes localement compacts," Université de Paris, 1969–1970, 1971, mimeographed.

[DU, LA] M. Duflo and J. P. Labesse, "Sur la formule des traces de Selberg," *Ann. Sci.* ENS (1971), pp. 193–284.

[EL] J. Elstrodt, "Die Resolvente zum Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene," Teil I, *Math. Ann.* **203** (1973), pp. 295–330; and Teil II, *Math. Zeitschr.* **132** (1973), pp. 99–134.

[FA 1] L. Faddeev, "Expansion in eigenfunctions of the Laplace operator on the fundamental domain of a discrete group on the Lobacevskii plane," *AMS Transl.* Trudy (1967), pp. 357–386.

[FA 2] ———, "On Friedrichs' model in the theory of perturbations of a continuous spectrum," *AMS Transl.*, Trudy (1964), pp. 177–203.

[FR] K. Friedrichs, "On the perturbation of continuous spectra," *Comm. Pure Appl. Math.* **1** (1948), pp. 361–406.

[GE, FO] I. Gelfand, S. Fomin, "Geodesic flows on manifolds of constant negative curvature," *Uspehi Mat. Nauk.* **7** (1952), pp. 49–65.

[GE, GR] I. Gelfand, M. Graev, I. Pjateckii-Shapiro, *Representation Theory and Automorphic Functions.* Moscow, 1966; translated, W. B. Saunders, 1969.

[GE, NA] I. Gelfand and Naimark, *Unitäre Darstellung klassischer Gruppen.* Akademie Verlag, 1957.

[GE, P-S] I. GELFAND and I. PJATECKII-SHAPIRO, "Theory of representations and the theory of automorphic functions," *Uspehi Mat. Nauk.* **14** (1959), *AMS Transl.* **26** (1963), pp. 173–200.

[Go 1] R. GODEMENT, *Analyse spectrale des fonctions modulaires,* Séminaire Bourbaki, 1964.

[Go 2] ———, "The decomposition of $L^2(G/\Gamma)$ for $\Gamma = SL_2(\mathbf{Z})$," *Proc. Symposia pure Math., AMS,* **9** (Providence, R. I., 1966), pp. 211–224.

[Go 3] ———, *La formule des traces de Selberg.* Séminaire Bourbaki, 1962.

[Go 4] ———, *Introduction aux travaux de A. Selberg.* Séminaire Bourbaki, 1957.

[Go 5] ———, "The spectral decomposition of cusp forms," *Proc. Symposia pure Math., AMS,* **9** (1966), pp. 225–234.

[Go 6] ———, "A theory of spherical functions," Trans. Am. Math. Soc. **73** (1952), pp. 496–556.

[H-C 1] HARISH-CHANDRA, "Automorphic forms on semisimple Lie groups," Springer Verlag Lecture Notes 62, 1968.

[H-C 2] ———, "Discrete series for semisimple Lie groups I," *Acta Math.* **113** (1965), pp. 241–318.

[H-C 3] ———, "Discrete series for semisimple Lie groups II," *Acta Math.* **116** (1966), pp. 1–111, especially p. 71.

[H-C 4] ———, "Harmonic analysis on semisimple Lie groups," *Bull. Am. Math. Soc.* **76** (1970), pp. 529–551.

[H-C 5] ———, "Invariant eigendistribution on a semisimple Lie algebra," *Publ. Math.* IHES No. 27 (1965), pp. 609–658 and 1–54.

[H-C 6] ———, "Plancherel formula for the 2×2 real unimodular group," *Proc. Nat. Acad. Sci. U.S.A.* **4** (1952), pp. 337–342.

[H-C 7] ———, "Spherical functions on a semisimple Lie group I," *Am. J. Math.* **80** (1958), pp. 241–310; and II, pp. 553–613, especially pp. 576 and 582.

[A fairly complete bibliography of Harish-Chandra's works occurs in Warner's book, *Harmonic Analysis on Semisimple Lie Groups*, Springer-Verlag, 1972.]

[HE 1] S. HELGASON, "Analysis on Lie groups and homogeneous spaces," Regional conference series, AMS, No. 14, 1972.

[HE 2] ———, *Differential Geometry and Symmetric Spaces*. New York: Academic Press, 1962.

[HE 3] ———, "A duality for symmetric spaces with applications to group representations," *Advances in Math.* (1970), pp. 1–154.

[HE, JO] S. HELGASON and K. JOHNSON, "The bounded spherical functions on symmetric spaces," *Advances in Math.* **3** (No. 4, 1969), pp. 586–593.

[JA, LA] H. JACQUET and R. LANGLANDS, "Automorphic forms on GL_2," Springer Verlag Lecture notes 114, 1970.

[KA] T. KATO, "Perturbation of continuous spectra by trace class operators," *Proc. Japan Acad.* **33** (1957), pp. 260–264.

[KN, ST] A. KNAPP and E. STEIN, "Intertwining operators for semisimple groups," *Ann. Math.* **93** (1971), pp. 489–578.

[KU] KUBOTA, *Introduction to Eisenstein series*, New York: Halsted Press, 1973.

[LA 1] R. LANGLANDS, "Eisenstein series," *Proc. Symposia pure Math., AMS*, **9** (1966), pp. 235–252. (There is also a dittoed unpublished manuscript giving complete proofs.)

[LA 2] ———, "Euler products," Yale lectures, 1967.

[LA 3] ———, "Problems in the theory of automorphic forms," Yale lectures, 1969.

[MA 1] H. MAASS, "Lectures on modular functions of one complex variable," Tate lecture notes, Bombay, 1964.

[MA 2] ———, "Uber eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen," *Math. Ann.* **121** (1949), pp. 141–183.

[McD] I. McDONALD, *Spherical functions on a group of p-adic type,* Ramanujan Institute Publications, Madras, 1971.

[NE] E. NELSON, "Analytic vectors," *Ann. Math.* **70** (1959), pp. 572–615.

[PE 1] H. PETERSON, "Über den Bereich absoluter Konvergenz der Poincareschen Reihen," *Acta. Math.* **80** (1948), pp. 23–63.

[PE 2] ———, "Zur analytischen Theorie der Grenzkreisgruppen," Teil I, *Math. Annln.* **115** (1937), pp. 23–67.

[PO] A. POVZNER, "On the expansion of arbitrary functions in characteristic functions of the operator $-\Delta u + cu$," *Math. Sb.* **32** (No. 74, 1953), pp. 109–156.

[PU] L. PUKANSZKY, "The Plancherel formula for the universal covering group of $SL(R, 2)$, *Math. Ann.* **156** (1964), pp. 96–143.

[ROE 1] W. ROELCKE, "Uber die Wellengleichung bei Grenzkreisgruppen erster Art," *Heidelberger Akad. Wiss. Math.* 1953–1955 (1956), pp. 159–267.

[ROE 2] W. ROELCKE, "Analytische Fortsetzung der Eisensteinreihen zu den parabolischen Spitzen von Grenzkreisgruppen erster Art," *Math. Annln.* **132** (1956), pp. 121–129.

[ROS] M. ROSENBLUM, "Perturbation of the continuous spectrum and unitary equivalence," *Pacif. J. Math.* **7** (1957), pp. 997–1010.

[SA, SH 1] P. SALLY and J. SHALIKA, "The Fourier transform of SL_2 over a non-archimedean local field," to be published.

[SA, SH 2] ———, "The Plancherel formula for $SL(2)$ over a local field," *Proc. Nat. Acad. Sci. U.S.A.* **63** (1969), pp. 661–667.

[SA] I SATAKE, "Theory of spherical functions on reductive algebraic groups over p-adic fields," *Publ. Math. IHES* **18** (1963), pp. 1–69.

[SE 1] A. SELBERG, "Discontinuous groups and harmonic analysis," *Proc. Internat. Congress,* (Stockholm, 1962), pp. 177–189.

[SE 2] ———, "Harmonic analysis and discontinuous groups in weakly symmetric Riemannian spaces with applications to Dirichlet series," *Colloquium on zeta functions*, (Tata Institute, 1956), pp. 47–87.

[SH] J. SHALIKA, "Representations of the 2×2 unimodular group over local fields to appear."

[SH, TA] J. SHALIKA and S. TANAKA, "On an explicit construction of a certain class of automorphic forms," *Am. J. Math.* **91** (No. 4 1969), pp. 1049–1076.

[SH] G. SHIMURA, *Introduction to the arithmetic theory of automorphic functions*, Iwanami Shoten and Princeton University Press, 1971.

[SI] C.L. SIEGEL, "Some remarks on discontinuous groups," *Ann. Math.* **46** (1945), pp. 708–718.

[ST 1] E. STEIN, "Analysis in matrix spaces and some new representations of $SL(N, \mathbb{C})$," *Ann. Math.* **86** (1967), pp. 461–490.

[ST 2] ———, "Analytic continuation of group representations," *Advances Math.* **4** (No. 2, 1970), pp. 172–207.

[TAK 1] R. TAKAHASHI, "Sur les fonctions sphériques et la formule de Plancherel dans le groupe hyperbolique," *Jap. J. Math.* **31** (1961), pp. 55–90.

[TAK 2] ———, "Sur les représentations unitaires des groupes de Lorentz généralisés," *Bull. Soc. Math. France* **91**(1963), pp. 289–433.

[TAM] T. TAMAGAWA, "On Selberg's trace formula," *J. Faculty Science, Tokyo,* Sec. I, **8** (Part 2, 1960), pp. 363–386.

[VI] N. VILENKIN, "Special functions and the theory of group representations," AMS translation of monographs 22, 1968.

[WA] G. WARNER, *Harmonic Analysis on Semisimple Lie Groups*, Springer Verlag, 1972.

[WE] A. WEIL, "Sur certains groupes d'operateurs unitaires," *Acta Math.* **111** (1964), pp. 143–211.

Note. The above bibliography makes no claim to completeness. Vilenkin has a good extensive bibliography, as does Warner.

Symbols Frequently Used

p. 2 $\varphi * \psi(x) = \int_G \varphi(xy^{-1})\psi(y)dy$

p. 2 $f^-(x) = f(x^{-1})$ or $f(-x)$ (specified in each context)

p. 4 $\pi^1(\varphi)v = \int_G \varphi(x)\pi(x)v\,dx$

p. 5 $\varphi^*(x) = \overline{\varphi(x^{-1})}$

p. 19 $S_{n,m}$: Functions f such that $f(r(\theta)yr(\theta')) = e^{-in\theta}f(y)e^{-in\theta'}$

p. 19 $$r(\theta) = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}$$

p. 23 H_n: Subspace of H consisting of those v such that $\pi(r(\theta))v = e^{in\theta}v$.

p. 41 $$\alpha\begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} = a^2$$

p. 46 $\rho(a) = \alpha(a)^{1/2}$, $\rho(x) = \rho(a)$ if $x = ank$.

p. 46 $\mu_s(an) = \rho(a)^s$

p. 46 $H(s)$ = space of functions whose restriction to K is in L^2 and satisfying $f(any) = \rho(a)^{s+1}f(y)$.
π_s is the representation on $H(s)$, $\pi_s(x)f(y) = f(yx)$.

p. 47 $\varphi_s(x) = \int_K \rho(kx)^{s+1}\,dx$ in the context of spherical functions. In the context of Fourier series, $\varphi_n(\theta) = e^{in\theta}$.

p. 49 $C_c(G, K)$ = continuous functions with compact support invariant under conjugation by elements of K.

p. 51 $C_c(G//K)$ = continuous functions with compact support and bi-invariant under K.

p. 68 $D(a) = \rho(a) - \rho(a)^{-1}$

p. 69
$$w = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

p. 69 **Harish transform**

$$\mathbf{H}f(a) = \rho(a)\int_N f(an)\,dn = D(a)\int_{A\backslash G} f(x^{-1}ax)\,d\dot{x}$$

p. 70 $A^+ = \{a \in A, \rho(a) \geqslant 1\}$

p. 74 **Mellin transform**

$$\mathbf{M}g(s) = \int_A g(a)\rho(a)^s\,da$$

p. 78 **Spherical transform**

$$\mathbf{S}f(s) = \mathbf{MH}f(s) = \int_G f(x)\varphi_s(x)\,dx$$

p. 89
$$\exp(X) = \sum_{n=0}^{\infty} \frac{X^n}{n!}$$

p. 90
$$\mathfrak{L}_X f(y) = \frac{d}{dt} f(y \exp(tX))\bigg|_{t=0}$$

p. 94
$$d\pi(X)v = \frac{d}{dt}\pi(\exp(tX)v)\bigg|_{t=0}$$

p. 102 E^+ or $E_+ = \begin{pmatrix} 1 & i \\ i & -1 \end{pmatrix}$; E^- or $E_- = \begin{pmatrix} 1 & -i \\ -i & -1 \end{pmatrix}$

$$[E_+, E_-] = -4iW, \qquad [W, E_+] = 2iE_+, \qquad [W, E_-] = -2iE_-$$

p. 102
$$\mathrm{ad}(X)Y = [X, Y]$$

p. 105
$$H^+ = \hat{\bigoplus_{n \text{ even}}} H_n \quad \text{and} \quad H^- = \hat{\bigoplus_{n \text{ odd}}} H_n$$

p. 105 $$\mathrm{Ad}(y)X = yXy^{-1}$$

p. 132 $$\psi_K(x) = \int_K \psi(k^{-1}xk)\,dk$$

p. 132 G' = set of regular elements, i.e. having distinct eigenvalues.

p. 149 $H(\mu, \epsilon)$ or $H(s, \epsilon)$, those functions in $H(s)$ which have character ϵ, where ϵ is a character of $\{\pm 1\}$.

p. 120, 150 $$H^{(m)} = \widehat{\bigoplus_{\substack{n \geq m \\ n \equiv m}}} H_n \quad \text{and} \quad H^{(-m)} = \widehat{\bigoplus_{\substack{n \leq -m \\ n \equiv m}}} H_n$$

p. 192 $\mathfrak{U}(\mathfrak{g})$ = algebra of differential operators generated by the Lie derivatives.

p. 193 $\mathfrak{Z}$ denotes the centralizer of whatever comes after it.

p. 228 ${}^0L^2(\Gamma \backslash G)$ consists of those functions f such that

$$\int_{\Gamma_N \backslash N} f(ng)\,dn = 0$$

for all g, and every cuspidal subgroup N with respect to Γ.

p. 243 $$Z(\varphi, y, 2s) = \int_A \varphi(ay)\rho(a)^{-2s}\,da$$

p. 244 $$E(\varphi, y, s) = \sum_{\Gamma_N \backslash \Gamma} Z(\varphi, \gamma y, 2s) = TZ(\varphi, y, 2s)$$

Index

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces.
4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician. 2nd ed.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.

continued after index

Graduate Texts in Mathematics

continued from page ii

119 ROTMAN. An Introduction to Algebraic Topology.
120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups. 2nd ed.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.
131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic *K*-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory: The Classical Bases.
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Program.
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.
177 NEWMAN. Analytic Number Theory.
178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.