Aspects of Harmonic Analysis and Representation Theory

Jean Gallier and Jocelyn Quaintance
Department of Computer and Information Science
University of Pennsylvania
Philadelphia, PA 19104, USA
e-mail: jean@cis.upenn.edu

© Jean Gallier

November 27, 2018
Preface

The question that motivated writing these notes is:

What is the Fourier transform?

We were quite surprised by how involved the answer is, and how much mathematics is needed to answer it, from measure theory, integration theory, some functional analysis, to some representation theory.

First we should be a little more precise about our question. We should ask two questions:

(1) What is the input domain of the Fourier transform?

(2) What is the output domain of the Fourier transform?

The answer to (1) is that the domain of the Fourier transform, denoted by $F$, is a set of functions on a group $G$. Now, in order for the Fourier transform to be useful, it should behave well with respect to convolution (denoted $f \ast g$) on the set of functions on $G$, which implies that these functions should be integrable.

This leads to the first subtopic, which is what is integration on a group. The technical answer involves the Haar measure on a locally compact group. Thus, any serious effort to understand what the Fourier transform is entails learning a certain amount of measure theory and integration theory, passing through versions of the Radon–Riesz theorem relating Radon functionals and Borel measures, and culminating with the construction of the Haar measure. The two candidates for the domain of the Fourier transform are the spaces $L^1(G)$ and $L^2(G)$. Unfortunately, convolution and the Fourier transform are not necessarily defined for functions in $L^2(G)$, so the domain of the Fourier transform is $L^1(G)$. Then the equation $F(f \ast g) = F(f)F(g)$ holds, as desired. If $G$ is a compact group, $L^2(G)$ is a suitable (and better) domain.

The answer to Question (2) is more complicated, and depends heavily on whether the group $G$ is commutative or not. The answer is much simpler if $G$ is commutative. In both cases, the output domain of the Fourier transform should be a set of functions from a space $Y$ to a space $Z$. 
If $G$ is commutative, then we can pick $Z = \mathbb{C}$. However, the space $Y$ is rarely equal to $G$ (except when $G = \mathbb{R}$). It turns out that a good theory (which means that it covers all cases already known) is obtained by picking $Y$ to be the group $\hat{G}$, the Pontrjagin dual of $G$, which consists of the characters of the group $G$. A character of $G$ is a continuous homomorphism $\chi : G \to U(1)$ from $G$ to the group of complex numbers of absolute value 1. For any function $f \in L^1(G)$, the Fourier transform $\mathcal{F}(f)$ of $f$ is then a function

$$\mathcal{F}(f) : \hat{G} \to \mathbb{C}.$$  

In general, $\hat{G}$ is completely different from $G$, and this creates problems. For the familiar cases, $G = \mathbb{T} \cong U(1)$, $G = \mathbb{Z}$, $G = \mathbb{R}$, and $G = \mathbb{Z}/n\mathbb{Z}$, the characters are well known. The case $G = \mathbb{Z}/n\mathbb{Z}$ corresponds to the discrete Fourier transform.

For the groups listed above, we know that under some suitable restriction, we have Fourier inversion, which means that there is some transform $\overline{\mathcal{F}}$ (called Fourier cotransform) such that

$$f = \overline{\mathcal{F}}(\mathcal{F}(f)). \quad (\ast)$$

We have to be a bit careful because the domain of $\overline{\mathcal{F}}$ is $L^1(\hat{G})$, and not $L^1(G)$, are they are usually very different because in general $G$ and $\hat{G}$ are not isomorphic. Then (assuming that it makes sense), $\overline{\mathcal{F}}(\mathcal{F}(f))$ is a function with domain $\hat{\hat{G}}$, so there seems no hope, except in very special cases such as $G = \mathbb{R}$, that $(\ast)$ could hold. Fortunately, Pontrjagin duality asserts that $G$ and $\hat{\hat{G}}$ are isomorphic, so $(\ast)$ holds (under suitable conditions) in the form

$$f = \overline{\mathcal{F}}(\mathcal{F}(f)) \circ \eta,$$

where $\eta : G \to \hat{\hat{G}}$ is a canonical isomorphism.

If $G$ is a commutative abelian group, there is a beautiful and well understood theory of the Fourier transform based on results of Gelfand, Pontrjagin, and André Weil. In particular, even though the Fourier transform is not defined on $L^2(G)$ in general, for any function $f \in L^1(G) \cap L^2(G)$, we have $\mathcal{F}(f) \in L^2(\hat{G})$, and by Plancherel’s theorem, the Fourier transform extends in a unique way to an isometric isomorphism between $L^2(G)$ and $L^2(\hat{G})$. Furthermore, if we identify $G$ and $\hat{\hat{G}}$ by Pontrjagin duality, then $\mathcal{F}$ and $\overline{\mathcal{F}}$ are mutual inverses.

If $G$ is not commutative, things are a lot tougher. Characters no longer provide a good input domain, and instead one has to turn to unitary representations. A unitary representation is a homomorphism $U : G \to U(H)$ satisfying a certain continuity property, where $U(H)$ is the group of unitary operators on the Hilbert space $H$. Then $\hat{G}$ is the set of equivalence classes of irreducible unitary representations of $G$, but it is no longer a group.

If $G$ is compact, an important theorem due to Peter and Weyl gives a nice decomposition of $L^2(G)$ as a Hilbert sum of finite-dimensional matrix algebras corresponding to the irreducible unitary representations of $G$. As a consequence, there is good notion of Fourier
transform, such that the Fourier transform $\mathcal{F}(f)$ is a function with domain $\hat{G}$, but its output domain is no longer $\mathbb{C}$. Instead, it is a finite-dimensional hermitian space depending on the irreducible representation given as input. In general, it is very difficult to find the irreducible representations of a compact group, so this Fourier transform does seem to be very useful in practice.

If the compact group is a Lie group, then the whole machinery of Lie algebras and Lie groups developed by Élie Cartan and Hermann Weyl involving weights and roots becomes available. In particular, if $G$ is a connected semisimple Lie group, the finite-dimensional irreducible representations are determined by highest weights. There is a beautiful and extensive theory of representations of semisimple Lie groups, and many books have been written on the subject.

A way to deal with noncommutativity, due to Gelfand, is to work with pairs $(G, K)$, where $K$ is a compact subgroup of $G$. Then, instead of working with functions on $G$, which is “too big,” we work with functions on the homogeneous space $G/K$, the space of left cosets. Then, under certain assumptions on $G$ and $K$, which makes $(G, K)$ a Gelfand pair, it is possible to consider a commutative algebra of functions on the set of double cosets $KsK$ ($s \in G$), so that some results from the commutative theory can be used. The domain of the Fourier transform is a set of functions called spherical functions, and this set happens to be homeomorphic to the set of characters on the commutative algebra mentioned above. There is a very nice theory of the Fourier transform and its inverse, but how useful it is in practice remains to be seen.

**Acknowledgement:** Many thanks to the participants of the “underground” tuesday meetings, Christine Allen-Blanchette, Carlos Esteves, Stephen Phillips, and João Sedoc, for catching mistakes and for many helpful comments. We also thank Kostas Daniilidis for being a source of inspiration.
## Contents

1 Function Spaces Often Encountered .............................. 11  
1.1 Spaces of Bounded Functions ........................................ 11  
1.2 Convergence: Pointwise, Uniform, Compact ....................... 13  
1.3 Equicontinuous Sets of Continuous Functions ................. 21  
1.4 Regulated Functions ................................................. 25  
1.5 Neighborhood Bases and Filters ................................. 31  
1.6 Topologies Defined by Semi-Norms; Fréchet Spaces .......... 35  

2 The Riemann Integral .................................................. 39  
2.1 Riemann Integral of a Continuous Function .................... 40  
2.2 The Riemann Integral of Regulated Functions ................. 45  

3 Measure Theory; Basic Notions ..................................... 49  
3.1 $\sigma$-Algebras, Measures .......................................... 50  
3.2 Null Subsets and Properties Holding Almost Everywhere .... 59  
3.3 Construction of a Measure from an Outer Measure .............. 61  
3.4 The Lebesgue Measure on $\mathbb{R}$ ............................... 64  

4 Integration ................................................................. 69  
4.1 Measurable Maps ...................................................... 71  
4.2 Step Maps on a Measurable Space .................................. 75  
4.3 $\mu$-Measurable Maps and $\mu$-Step Maps ................. 77  
4.4 The Integral of $\mu$-Step Maps ..................................... 84  
4.5 Integrable Functions; the Spaces $L_\mu(X, \mathcal{A}, F)$ and $L_\mu(X, \mathcal{A}, F)$ .... 90  
4.6 Fundamental Convergence Theorems .............................. 99  
4.7 The Spaces $L_\mu^p(X, \mathcal{A}, F)$ and $L_\mu^p(X, \mathcal{A}, F)$; $p = 1, 2, \infty$ ................. 106  
4.8 Products of Measure Spaces and Fubini’s Theorem ............ 113  
4.9 The Lebesgue Measure in $\mathbb{R}^n$ ............................... 118  

5 Radon Measures on Locally Compact Spaces .................... 121  
5.1 Positive Radon Functionals Induced by Borel Measures ....... 122  
5.2 The Radon–Riesz Theorem and Positive Radon Functionals ... 128
5.3 Regular Borel Measures ..................................................... 132
5.4 Complex and Real Measures ........................................... 136
5.5 Real Measures and the Hahn–Jordan Decomposition ............ 139
5.6 Total Variation of a Radon Functional ............................... 142
5.7 The Radon–Riesz Theorem and Bounded Radon Functionals . 145

6 The Haar Measure and Convolution .............................. 151
6.1 Topological Groups .......................................................... 153
6.2 Existence of the Haar Measure; Preliminaries ..................... 163
6.3 Existence of the Haar Measure .......................................... 169
6.4 Uniqueness of the Haar Measure ....................................... 173
6.5 Examples of Haar Measures ............................................. 176
6.6 The Modular Function ...................................................... 178
6.7 More Examples of Haar Measures ..................................... 184
6.8 The Modulus of an Automorphism .................................... 185
6.9 Some Properties and Applications of the Haar Measure ....... 190
6.10 G-Invariant Measures on Homogeneous Spaces ................. 193
6.11 Convolution ................................................................. 199
6.12 Regularization ............................................................... 209

7 The Fourier Transform and Cotransform on $T^n, Z^n, R^n$ .... 219
7.1 Fourier Analysis on $T$ ...................................................... 222
7.2 The Fourier Transform and Cotransform on $T^n$ and $Z^n$ .... 236
7.3 The Fourier Transform and the Fourier Cotransform on $R$ .... 241
7.4 The Sampling Theorem .................................................... 247
7.5 The Fourier Transform and the Fourier Cotransform on $R^n$ .. 249
7.6 The Schwartz Space ......................................................... 251
7.7 The Poisson Summation Formula ...................................... 256
7.8 Pointwise Convergence of Fourier Series on $T$ ................. 257
7.9 The Heisenberg Uncertainly Principle ............................... 262
7.10 Fourier’s Life; a Brief Summary ...................................... 264

8 Normed Algebras and Spectral Theory .......................... 267
8.1 Normed Algebras, Banach Algebras .................................. 271
8.2 Two Algebra Constructions ............................................. 276
8.3 Spectrum, Characters, Gelfand Transform, I ...................... 280
8.4 Spectrum, Characters, II; For a Banach Algebra ............... 286
8.5 Gelfand Transform, II; For a Banach Algebra ................. 290
8.6 Banach Algebras with Involution; $C^*$-Algebras .............. 293
8.7 Characters and Gelfand Transform in a $C^*$-Algebra .......... 299
8.8 The Enveloping $C^*$-Algebra of an Involutive Banach Algebra 302

9 Analysis on Locally Compact Abelian Groups .................. 305
9.1 Characters and The Dual Group ........................................ 310
9.2 The Fourier Transform and the Fourier Cotransform .............. 323
9.3 The Fourier Transform on a Finite Abelian Group ................. 331
9.4 Dirichlet Characters .................................................... 335
9.5 Fourier transform and Cotransform in Terms of Matrices .......... 338
9.6 The Discrete Fourier Transform (on $\mathbb{Z}/n\mathbb{Z}$) ............ 345
9.7 Some Properties of the Fourier Transform .......................... 350
9.8 Plancherel’s Theorem and Fourier Inversion ....................... 352
9.9 Pontrjagin Duality and Fourier Inversion .......................... 355

10 Hilbert Algebras ....................................................... 361
10.1 Representations of Algebras with Involution ....................... 364
10.2 Positive Linear Forms and Positive Hilbert Forms ............... 371
10.3 Traces, Bitraces, Hilbert Algebras .................................. 373
10.4 Complete Separable Hilbert Algebras ............................... 378
10.5 The Plancherel–Godement Theorem ⊛ ............................ 390

11 Representations of Locally Compact Groups .......................... 401
11.1 Finite-Dimensional Group Representations ......................... 402
11.2 Unitary Group Representations ..................................... 406
11.3 Unitary Representations of $G$ and $L^1(G)$ ........................ 411
11.4 Functions of Positive Type and Unitary Representations ......... 419
11.5 The Gelfand–Raikov Theorem ....................................... 425
11.6 Measures of Positive Type and Unitary Representations ........ 429

12 Analysis on Compact Groups and Representations .................. 437
12.1 The Peter–Weyl Theorem, I .......................................... 441
12.2 Characters of Compact Groups ...................................... 450
12.3 The Peter–Weyl Theorem, II ......................................... 456
12.4 The Fourier Transform ............................................... 462

13 Induced Representations ............................................... 473
13.1 Cocycles and Induced Representations ............................... 476
13.2 Cocycles on a Homogeneous Space $X = G/H$ ..................... 480
13.3 Converting Induced Representations of $G$ From $E^X$ to $E^G$ .... 485
13.4 Construction of the Hilbert Space $L^2_p(X; E)$ ..................... 487
13.5 Induced Representations, I; $G/H$ has a $G$-Invariant Measure .... 490
13.6 Quasi-Invariant Measures on $G/H$ ............................... 493
13.7 Induced Representations, II; Quasi-Invariant Measures .......... 496
13.8 Partial Traces, Induced Representations of Compact Groups .... 504

14 Harmonic Analysis on Gelfand Pairs .................................. 513
14.1 Gelfand Pairs ......................................................... 516
## CONTENTS

14.2 Real Forms of a Complex Semi-Simple Lie Group ................................................. 521
14.3 Spherical Functions ................................................................................................. 538
14.4 Examples of Gelfand Pairs ...................................................................................... 547
14.5 The Fourier Transform ............................................................................................ 560
14.6 The Plancherel Transform ...................................................................................... 563
14.7 Extension of the Plancherel Transform; $P(G)$ and $P'(Z) \circledast$ ......................... 570
14.8 Spherical Functions of Positive Type and Representations ........................................ 575

15 Topology .................................................................................................................. 579
15.1 Metric Spaces and Normed Vector Spaces ............................................................... 579
15.2 Topological Spaces .................................................................................................. 586
15.3 Continuous Functions, Limits .................................................................................. 596
15.4 Connected Sets ....................................................................................................... 603
15.5 Compact Sets and Locally Compact Spaces .............................................................. 613
15.6 Second-Countable and Separable Spaces ............................................................... 625
15.7 Sequential Compactness ........................................................................................... 629
15.8 Complete Metric Spaces and Compactness ............................................................. 635
15.9 Completion of a Metric Space .................................................................................... 638
15.10 The Contraction Mapping Theorem ......................................................................... 645
15.11 Continuous Linear and Multilinear Maps ............................................................... 649
15.12 Completion of a Normed Vector Space ................................................................... 656
15.13 Further Readings ..................................................................................................... 658

16 Vector Norms and Matrix Norms ............................................................................. 659
16.1 Normed Vector Spaces ............................................................................................ 659
16.2 Matrix Norms ......................................................................................................... 665

17 Basics of Groups and Group Actions ........................................................................ 679
17.1 Groups, Subgroups, Cosets ..................................................................................... 679
17.2 Group Actions: Part I, Definition and Examples ..................................................... 692
17.3 Group Actions: Part II, Stabilizers and Homogeneous Spaces ................................. 705
17.4 The Grassmann and Stiefel Manifolds ................................................................... 713

18 Hilbert Spaces .......................................................................................................... 719
18.1 The Projection Lemma, Duality ................................................................................ 719
18.2 Total Orthogonal Families, Fourier Coefficients .................................................... 732
18.3 The Hilbert Space $l^2(K)$ and the Riesz-Fischer Theorem ....................................... 740

Bibliography .................................................................................................................. 751
Chapter 1

Function Spaces Often Encountered

Various spaces of functions \( f : E \to F \) from a topological space \( E \) to a metric space or a normed vector space \( F \) come up all the time. The most frequently encountered are the spaces \( (F^E)_b \) of bounded functions, the spaces \( \mathcal{K}(E; F) \) of continuous functions with compact support, the spaces \( \mathcal{C}_0(E; F) \) of continuous functions which tend to zero at infinity, and the spaces \( \mathcal{C}_b(E; F) \) of bounded continuous functions. When \( F \) is a normed vector space, all these spaces are normed vector spaces with the sup norm. An important issue about function spaces is the convergence of sequences of functions. We review the main three notions, pointwise convergence (also known as simple convergence), uniform convergence, and compact convergence. A sequence of continuous functions may converge pointwise to a function which is not continuous. Uniform convergence has a better behavior. If \( F \) is a complete normed vector space, then both spaces \( \mathcal{C}_b(E; F) \) and \( (F^E)_b \) are also complete. An interesting family of functions in \( (F^{[a,b]})_b \) is the space \( \text{Reg}([a, b]; F) \) of regulated functions. These functions only have at most countably many simple kinds of discontinuities called discontinuities of the first kind. If \( F \) is a complete normed vector space, then the space \( \text{Step}([a, b]; F) \) is complete. It contains a subspace \( \text{Step}([a, b]; F) \) consisting of very simple functions called step functions, which take finitely many different values on consecutive open intervals. The space \( \text{Step}([a, b]; F) \) is dense in \( \text{Reg}([a, b]; F) \). If \( E \) is a locally compact space, then the space \( \mathcal{C}_0(E; \mathbb{C}) \) is the closure of \( \mathcal{K}_C(E) \) in \( \mathcal{C}_b(E; \mathbb{C}) \).

1.1 Spaces of Bounded Functions

In this section, we are dealing with functions \( f : E \to F \), where \( F \) is either a metric space or a normed vector space. Recall that the set of all functions from \( E \) to \( F \) is denoted by \( F^E \).

First assume that \( F \) is a metric space with metric \( d \). We would like to make \( F^E \) into a metric space. It is natural to define a metric on \( F^E \) by setting

\[
d_\infty(f, g) = \sup_{x \in E} d(f(x), g(x))
\]

for any two functions \( f, g : E \to F \), but if \( d(f(x), g(x)) \) is unbounded as \( x \) ranges over \( E \), the
expression \( \sup_{x \in E} d(f(x), g(x)) \) is undefined. Therefore, we consider the space of bounded functions defined as follows.

**Definition 1.1.** If \((F, d)\) is a metric space, a function \(f: E \to F\) is *bounded* if its image \(f(E)\) is bounded in \(F\), which means that \(f(E) \subseteq B(a, \alpha)\), for some closed ball \(B(a, \alpha)\) of center \(a\) and radius \(\alpha > 0\). See Figure 1.1. The space of bounded functions \(f: E \to F\) is denoted by \((F^E)_b\).

![Figure 1.1](image)

Figure 1.1: Let \(E = F = \mathbb{R}\) with the Euclidean metric. In Figure (i), \(f\) is unbounded since \(f(E) = \mathbb{R}\). In Figure (ii), \(f \in (F^E)_b\) since \(f(E) = (0, 1]\) and \((0, 1] \subseteq B(0, 1) = [-1, 1]\).

If \(f: E \to F\) and \(g: E \to F\) are bounded functions then it is easy to see that if \(f(E) \subseteq B(a, \alpha)\) and if \(g(E) \subseteq B(b, \beta)\), then

\[
d(f(x), g(x)) \leq \alpha + \beta + d(a, b) \quad \text{for all } x \in E.
\]

Therefore, \(\sup_{x \in E} d(f(x), g(x))\) is well defined. It is easy to check that if we define

\[
d_\infty(f, g) = \sup_{x \in E} d(f(x), g(x))
\]

for any two bounded functions \(f, g\), then \(d\) is indeed a metric on \((F^E)_b\).

**Definition 1.2.** If \((F, \|\|)\) is a normed metric space, then \((F^E)_b\) is a vector space, and it is easy to check that \((F^E)_b\) is also a vector space. For any bounded function \(f: E \to F\) (which means that \(f(E) \subseteq B(0, \alpha)\), for some closed ball \(B(0, \alpha)\)), then

\[
\|f\|_\infty = \sup_{x \in E} \|f(x)\|
\]

is a norm on the vector space \((F^E)_b\).
1.2. CONVERGENCE: POINTWISE, UNIFORM, COMPACT

Figure 1.2: Let $E = F = \mathbb{R}$ with the Euclidean metric. Both $f, g \in (F^E)_b$ since $f(E) = (0, 1)$, while $g(E) = [-1, 0)$. The concatenation of the vertical dashed red lines is $d_\infty(f, g) = \sup_{x \in E} d(f(x), g(x)) = 1 - (-1) = 2$.

Definition 1.3. If $(F, \| \|)$ is a normed vector space, then for any bounded function $f \in (F^E)_b$, the quantity

$$\|f\|_\infty = \sup_{x \in E} \|f(x)\|$$

is a norm on $(F^E)_b$, often called the sup norm.

The following important theorem can be shown; see Schwartz [69] (Chapter XV, Section 1, Theorem 1).

Theorem 1.1. (1) If $(F, d)$ is a complete metric space, then $((F^E)_b, d_\infty)$ is also a complete metric space.

(2) If $(F, \| \|)$ is a complete normed vector space, then $((F^E)_b, \| \|_\infty)$ is also a complete normed vector space.

1.2 Convergence: Pointwise, Uniform, Compact

When dealing with spaces of functions, a crucial issue is to identify notions of limit that preserve certain desirable properties, such as continuity. There are primarily three such notions of convergence: pointwise, uniform, and compact, that we now review.

Definition 1.4. Let $(F, d)$ be a metric space. A sequence $(f_n)_{n \geq 1}$ of functions $f_n : E \to F$ converges pointwise (or converges simply) to a function $f : E \to F$ if for every $x \in E$, for every $\epsilon > 0$, there is some $N > 0$ such that

$$d(f_n(x), f(x)) < \epsilon \quad \text{for all } n \geq N.$$
Figure 1.3: A schematic illustration of $f_n(x)$ converging pointwise $f(x)$, where $E = F = \mathbb{R}$. As $n$ increases, the graph of $f_n(x)$ near $x$ must be in the band determined by the graphs of $f(x) - \epsilon$ and $f(x) + \epsilon$.

Definition 1.4 says that for every $x \in E$, the sequence $(f_n(x))_{n \geq 1}$ converges to $f(x)$. Observe that the above $\epsilon$ depends on $x$.

Let $F$ be a topological space. The product topology on $F^E$ is defined as follows: a subset of functions in $F^E$ is open if it is the union of subsets $U_A$ of functions $f : E \to F$ for which there is some finite subset $A$ of $E$ such that $f(x) \in U_x$ for all $x \in A$, where $U_x$ is an open subset of $F$, and $f(x) \in F$ is arbitrary for all $x \in E - A$. Equivalently, for any $x \in E$ and any open subset $U$ of $F$, let $S(x, U)$ be the set

$$S(x, U) = \{ f \in F^E, f(x) \in U \}.$$  

Then observe that

$$U_A = \bigcap_{x \in A} S(x, U_x), \quad A \text{ finite},$$

that is, the sets $S(x, U)$ form a subbasis of the product topology on $F^E$.

For every $x \in E$, if $\pi_x : F^E \to F$ is the projection map given by

$$\pi_x(f) = f(x), \quad f \in F^E,$$

(evaluation at $x$), then the product topology on $F^E$ is the weakest topology that makes all the $\pi_x$ continuous. Indeed, the weakest topology on $F^E$ making all the $\pi_x$ continuous consists of all unions of finite intersections of subsets of $F^E$ of the form $\pi_x^{-1}(U_x)$, for any open subset $U_x$ of $F$, but

$$\pi_x^{-1}(U_x) = S(x, U_x),$$

is one of the sets in the subbasis defined above. For this reason, the product topology on $F^E$ is also called the weak topology induced by the family of functions $(\pi_x)_{x \in E}$; see Rudin [67] (Chapter 3, Section 3.8). Then it is not hard to see that a sequence $(f_n)_{n \geq 1}$ of elements
of \( F^E \) converges pointwise to \( f \in F^E \) iff the sequence \((f_n)_{n\geq 1}\) converges to \( f \) in the product topology; see Munkres [62] (Chapter 7, Section 46, Theorem 46.1), or Folland [32] (Chapter 4, Proposition 4.2). This is sometimes referred to as weak convergence.

Definition 1.5. If \( F \) is any topological space and \( E \) is any set, the topology on \( F^E \) having the sets
\[
S(x, U) = \{ f \mid f \in F^E, f(x) \in U \}, \quad x \in E, U \text{ open in } F,
\]
as a subbasis is the topology of pointwise convergence. An open subset of \( F^E \) in this topology is any union (possibly infinite) of finite intersections of subsets of the form \( S(x, U) \) as above.

If \( F \) is Hausdorff, so is the topology of pointwise convergence. Indeed, if \( f, g \in F^E \) and \( f \neq g \), then there is some \( x \in E \) such that \( f(x) \neq g(x) \), and since \( F \) is Hausdorff, there exist two disjoint open subsets \( U_{f(x)} \) and \( U_{g(x)} \) with \( f(x) \in U_{f(x)} \) and \( g(x) \in U_{g(x)} \). Then \( \pi_x^{-1}(U_{f(x)}) \) and \( \pi_x^{-1}(U_{g(x)}) \) are disjoint open subsets with \( f \in \pi_x^{-1}(U_{f(x)}) \) and \( g \in \pi_x^{-1}(U_{g(x)}) \).

Definition 1.6. Let \((F,d)\) be a metric space. A sequence \((f_n)_{n \geq 1}\) of functions \( f_n : E \to F \) converges uniformly to a function \( f : E \to F \) if for every \( \epsilon > 0 \), there is some \( N > 0 \) such that
\[
d(f_n(x), f(x)) < \epsilon \quad \text{for all } n \geq N \text{ and all } x \in E.
\]
See Figure 1.4.

---

**Figure 1.4:** A schematic illustration of \( f_n \) converging uniformly to \( f \), where \( E = F = \mathbb{R} \). As \( n \) increases, the graph of \( f_n \) must lie entirely in the band determined by the graphs of \( f - \epsilon \) and \( f + \epsilon \).
The difference between simple and uniform convergence is that in uniform convergence, $\epsilon$ is independent of $x$. For example the functions $f_n : [0, 2\pi] \to \mathbb{R}$ defined by $f_n(x) = n \sin \left( \frac{x}{n} \right)$ converges uniformly to $f(x) = x$, as evidenced by Figure 1.5. Consequently, uniform convergence implies simple convergence, but the converse is false, as the following examples illustrate.

![Figure 1.5: The colored functions $f_n(x) = n \sin \left( \frac{x}{n} \right)$, over the domain $[0, 2\pi]$, converge uniformly to the black line $f(x) = x$.]

**Example 1.1.** Let $g : \mathbb{R} \to \mathbb{R}$ be the function given by

$$g(x) = \frac{1}{1 + x^2},$$

and for every $n \geq 1$, let $f_n : \mathbb{R} \to \mathbb{R}$ be the function given by

$$f_n(x) = \frac{1}{1 + (x - n)^2}.$$

The function $f_n$ is obtained by translating $g$ to the right using the translation $x \mapsto x + n$; see Figure 1.6

Since

$$\lim_{n \to \infty} \frac{1}{1 + (x - n)^2} = 0,$$

the sequence $(f_n)_{n \geq 1}$ converges pointwise to the zero function $f$ given by $f(x) = 0$ for all $x \in \mathbb{R}$. However, since the the maximum of each $f_n$ is 1, we have

$$d_\infty(f_n, f) = 1 \quad \text{for all } n \geq 1,$$

so the sequence $(f_n)_{n \geq 1}$ does not converge uniformly to the zero function.
Figure 1.6: The bell curve graphs of Example 1.1; \( g(x) \) in brown; \( f_1(x) \) in red; \( f_2(x) \) in purple; \( f_3(x) \) in blue.

**Example 1.2.** Pick any positive real \( \alpha > 0 \). For each \( n \geq 1 \), let \( f_n : \mathbb{R} \to \mathbb{R} \) be the piecewise affine function defined as follows:

\[
 f_n(x) = \begin{cases} 
 0 & \text{if } x \leq 0 \text{ or } x \geq 1/n \\
 (2n)^\alpha x & \text{if } 0 \leq x \leq 1/(2n) \\
 2n^\alpha (1-nx) & \text{if } 1/(2n) \leq x \leq 1/n.
\end{cases}
\]

See Figure 1.7.

Figure 1.7: The piecewise affine functions of Example 1.2 with \( \alpha = 3 \); \( f_1(x) \) in magenta; \( f_2(x) \) in red; \( f_3(x) \) in purple; \( f_3(x) \) in blue. Each \( f_n(x) \) has a symmetrical triangular peak. As \( n \) increases, the peak becomes taller and thinner.

For every \( x > 0 \), there is some \( n \) such that \( 1/n < x \), so \( \lim_{n \to \infty} f_n(x) = 0 \) for \( x > 0 \), and since \( f_n(x) = 0 \) for \( x \leq 0 \), we see that the sequence \( (f_n)_{n \geq 1} \) converges pointwise to the zero
function \( f \). However, the maximum of \( f_n \) is \( n^\alpha \) (for \( x = 1/(2n) \)) so
\[
d_\infty(f_n, f) = n^\alpha,
\]
and \( \lim_{n \to \infty} d_\infty(f_n, f) = \infty \), so the sequence \((f_n)_{n \geq 1}\) does not converge uniformly to the zero function.

Observe that convergence in the metric space of bounded functions \(((F^E)_b, d_\infty)\) is the uniform convergence of sequences of functions. Similarly, convergence in the normed vector space of bounded functions \(((F^E)_b, \| \|_\infty)\) is the uniform convergence of sequences of functions. For this reason, the topology on \((F^E)_b\) induced by the metric \(d_\infty\) (or the norm \(\| \|_\infty\)) is sometimes called the topology of uniform convergence.

If \( E \) is a topological space, it is useful to define the following notion of convergence.

**Definition 1.7.** Let \( E \) be a topological space and let \((F, d)\) be a metric space. A sequence \((f_n)_{n \geq 1}\) of functions \(f_n: E \to F\) converges locally uniformly to a function \(f: E \to F\) if for every \( x \in E \), there is some open subset \( U \) of \( E \) containing \( x \) such that for every \( \epsilon > 0 \), there is some \( N > 0 \) such that
\[
d(f_n(x), f(x)) < \epsilon \quad \text{for all } n \geq N \text{ and for all } x \in U.
\]

If \( E \) is locally compact, it is easy to see that a sequence \((f_n)_{n \geq 1}\) converges locally uniformly iff it converges uniformly on every compact subset of \( E \).

If a sequence \((f_n)_{n \geq 1}\) of continuous functions converges pointwise to a function \( f \), the limit \( f \) is not necessarily continuous. For example, the functions \( f_n: [0, 1] \to \mathbb{R} \) given by \( f_n(x) = x^n \) are continuous, and the sequence \((f_n)_{n \geq 1}\) converges pointwise to the discontinuous function \( f: [0, 1] \to \mathbb{R} \) given by
\[
f(x) = \begin{cases} 
0 & \text{if } 0 \leq x < 1 \\
1 & \text{if } x = 1,
\end{cases}
\]
as evidenced by Figure 1.8.

The following theorem gives sufficient conditions for the limit of a sequence of continuous functions to be continuous.

**Theorem 1.2.** Let \( E \) be a topological space, and \((F, d)\) be a metric space, and let \((f_n)_{n \geq 1}\) be a sequence of functions \(f_n: E \to F\) converging locally uniformly to a function \(f: E \to F\). Then the following properties hold:

1. If the functions \( f_n \) are continuous at some point \( a \in E \), then the limit \( f \) is also continuous at \( a \).

2. If the functions \( f_n \) are continuous (on the whole of \( E \)), then the limit \( f \) is also continuous (on the whole of \( E \)).
1.2. CONVERGENCE: POINTWISE, UNIFORM, COMPACT

Figure 1.8: The sequence of functions $f_n(x) = x^n$ over $[0, 1]$ converges pointwise to the discontinuous green graph.

(3) If $E$ is metric space, the sequence $(f_n)_{n \geq 1}$ converges uniformly to $f$, and the $f_n$ are uniformly continuous on $E$, then the limit $f$ is also uniformly continuous on $E$.

The proof of Theorem 1.2 can be found in Schwartz [69] (Chapter XV, Section 4, Theorem 1).

Here are a few applications of Theorem 1.2.

Definition 1.8. Let $E$ be a topological space, and let $(F, d)$ be a metric space. The metric subspace of $((F^E)_b, d_\infty)$ consisting of all continuous bounded functions $f: E \to F$ is denoted $C_b(E; F)$. If $(E, \|\|)$ is a normed vector space, the normed subspace of $((F^E)_b, \|\|_\infty)$ consisting of all continuous bounded functions $f: E \to F$ is also denoted $C_b(E; F)$.

Proposition 1.3. Let $E$ be a topological space, and let $(F, d)$ be a metric space. The metric subspace $C_b(E; F)$ of $((F^E)_b, d_\infty)$ is closed. If $(F, d)$ is a complete metric space, then $(C_b(E; F), d_\infty)$ is also complete.

Proposition 1.4. Let $E$ be a topological space, and let $(F, \|\|)$ be a normed vector space. The normed subspace $C_b(E; F)$ of $((F^E)_b, \|\|_\infty)$ is closed. If $(F, \|\|)$ is a complete normed vector space, then $(C_b(E; F), \|\|_\infty)$ is also complete.

An important special case of Proposition 1.4 is the case where $F = \mathbb{R}$ or $F = \mathbb{C}$, namely, our functions are real-valued continuous and bounded functions $f: E \to \mathbb{R}$, or complex-valued continuous and bounded functions $f: E \to \mathbb{C}$. The spaces of functions $(C_b(E; \mathbb{R}), d_\infty)$ and $(C_b(E; \mathbb{C}, \|\|_\infty)$ are complete.
If $E$ is compact and if $(F, \|\|)$ is a complete normed vector space, then every continuous function $f: E \to F$ is bounded. As a consequence, the space $C(E; F)$ of continuous functions $f: E \to F$ is complete.

Another topology on a function space that will be used in the definition of the dual of an abelian locally compact group is the topology of compact convergence.

**Definition 1.9.** Let $E$ be any topological space and let $(F, d)$ be a metric space. For any $\epsilon > 0$, any $f \in FE$, and any compact subset $K$ of $E$, define the set $B_K(f, \epsilon)$ by

$$B_K(f, \epsilon) = \left\{ g \in FE \mid \sup_{x \in K} d(f(x), g(x)) < \epsilon \right\}.$$

The family of sets $B_K(f, \epsilon)$ is a subbasis of the topology of compact convergence; that is, an open set of $FE$ in this topology is any union (possibly infinite) of finite intersections of subsets of the form $B_K(f, \epsilon)$.

The difference between this topology and the topology of pointwise convergence is that a general basis subset containing a function $f$ contains functions that are close to $f$ not just at finitely many points, but at all points of some compact subset. Thus the topology of pointwise convergence is weaker than the topology of compact convergence, which itself is weaker than the topology of uniform convergence. It is easy to see the sets $B_K(f, \epsilon)$ actually form a basis of the topology of compact convergence (they are closed under finite intersections).

It is easy to show that a sequence $(f_n)$ of functions in $FE$ converges to a function $f$ in the topology of compact convergence iff for every compact subset $K$ of $E$, the sequence $(f_n)$ converges uniformly to $f$ on $K$.

If the space $E$ is compactly generated, then the topology of compact convergence is even better behaved.

**Definition 1.10.** A topological space $E$ is *compactly generated* if any subset $U$ of $E$ is open if and only if $U \cap K$ is open in $K$ for every compact subset $K$.

The following result is shown in Munkres [62] (Chapter 7, Section 46, Lemma 46.3).

**Proposition 1.5.** If a topological space $E$ is locally compact or first countable, then it is compactly generated.

A nice consequence of $E$ being compactly generated is that, as in the case of uniform convergence, the limit of a sequence of continuous functions that converges to a function $f$ in the topology of compact convergence is continuous.

**Proposition 1.6.** Let $E$ be a compactly generated topological space and let $(F, d)$ be a metric space. Then the space $C(E; F)$ of continuous functions from $E$ to $F$ is closed in $FE$ in the topology of compact convergence.
Proposition 1.6 is proven in Munkres [62] (Chapter 7, Section 46, Theorem 46.5).

In many applications we are interested in the space $C(E; F)$ of continuous functions from $E$ to $F$, and in this case there is another way to define the topology of compact convergence.

**Definition 1.11.** Let $E$ and $F$ be two topological spaces. For any compact subset $K$ of $E$ and any open subset $U$ of $F$, let $S(K, U)$ be the set of continuous functions

$$S(K, U) = \{ f \mid f \in C(E; F), \ f(K) \subseteq U \}.$$ 

The sets $S(K, U)$ form a subbasis for a topology on $C(E; F)$ called the **compact-open topology**. An open set in the topology is any union (possibly infinite) of finite intersections of subsets of the form $S(K, U)$.

It is immediately verified that if $F$ is Hausforff, then the compact-open topology on $C(E; F)$ is Hausdorff.

**Remark:** Observe that the open subsets $S(x, U)$ of the topology of pointwise convergence can be viewed as the result of restricting $K$ to be a single point but relaxing $f$ to belong to $F^{E}$.

The compact-open topology is interesting in its own right and coincides with the topology of compact convergence when $F$ is a metric space. The following result is proven in Munkres [62] (Chapter 7, Section 46, Theorem 46.8).

**Proposition 1.7.** If $E$ is a topological space and if $(F, d)$ is a metric space, then on the space $C(E; F)$ of continuous functions from $E$ to $F$ the compact-open topology and the topology of compact convergence coincide.

A subspace of $(F^{\mathbb{R}})_b$, where $F$ is a Banach space (a complete normed vector space), that plays an important role in the theory of the Riemann integral, is the space of regulated functions.

Another notion that is often useful to show that a sequence of continuous functions converges pointwise to a continuous function is the notion of an equicontinuous set of functions.

### 1.3 Equicontinuous Sets of Continuous Functions

Intuitively speaking equicontinuity is of sort of uniform continuity for sets of functions.

**Definition 1.12.** Let $E$ be a topological space and let $(F, d_F)$ be a metric space. A subset $S \subseteq C(E; F)$ of the set of continuous functions from $E$ to $F$ is **equicontinuous at a point** $x_0 \in E$ if for every $\epsilon > 0$, there is some open subset $U \subseteq E$ containing $x_0$ such that $d_F(f(x), f(x_0)) \leq \epsilon$ for all $x \in U$ and for all $f \in S$. If $E$ is also a metric space with metric $d_E$, then, the above condition says that for every $\epsilon > 0$, there is some $\eta > 0$, such that $d_F(f(x), f(x_0)) \leq \epsilon$ for all $x \in E$ such that $d_E(x, x_0) \leq \eta$, and for all $f \in S$. The set of functions $S$ is **equicontinuous** if it is equicontinuous at every point $x \in E$. 


For example, if $E$ is a metric space and if there exists two constant $c, \alpha > 0$ such that we have the Lipschitz condition
\[ d_F(f(x) - f(y)) \leq c(d_E(x, y))^\alpha, \quad \text{for all } f \in S \text{ and all } x, y \in E, \]
then $S$ is equicontinuous.

**Proposition 1.8.** Let $(f_n)$ be a sequence of functions $f_n \in \mathcal{C}(E; F)$, and let $(x_n)$ be a sequence of points $x_n \in E$. If the set \{f_n\} is equicontinuous, the sequence $(x_n)$ converges to $x \in E$, and the sequence $(f_n)$ converges pointwise to some function $f : E \to F$, then the sequence $(f_n(x_n))$ converges to $f(x) \in F$.

**Proof.** We have the inequality
\[ d_F(f_n(x_n), f(x)) \leq d_F(f_n(x_n), f_n(x)) + d_F(f_n(x), f(x)). \]
For every $\varepsilon > 0$, since the sequence $(f_n)$ converges pointwise to $f$, there is some $N_2 > 0$ such that $d_F(f_n(x), f(x)) \leq \varepsilon/2$ for all $n \geq N_2$. Since \{f_n\} is equicontinuous, there is some open subset $U \subseteq E$ containing $x$ such that
\[ d_F(f_n(y), f_n(x)) \leq \varepsilon/2 \quad \text{for all } n \geq 1 \text{ and all } y \in U. \]
Since $(x_n)$ converges to $x$, there is some $N_1 > 0$ such that $x_n \in U$ for all $n \geq N_1$, so
\[ d_F(f_n(x_n), f_n(x)) \leq \varepsilon/2 \quad \text{for all } n \geq N_1, \]
and for all $n \geq \max\{N_1, N_2\}$, we have $d_F(f_n(x_n), f(x)) \leq \varepsilon$, which proves that $(f_n(x_n))$ converges to $f(x)$. \(\square\)

There are various results about equicontinuous sets of functions usually known as variants of Ascoli's theorem. Schwartz [69] (Chapter XX) gives one of the most complete expositions we are aware of. We only consider three variants of Ascoli’s theorem that we will need.

**Theorem 1.9.** (Ascoli I) Let $E$ be a topological space, let $(F, d_F)$ be a metric space, and let $S \subseteq \mathcal{C}(E; F)$ be set of equicontinuous functions at some $x_0 \in E$. Then the closure $\overline{S}$ of $S$ in $F^E$ with the topology of pointwise convergence is also equicontinuous at $x_0$. As a corollary, if $S \subseteq \mathcal{C}(E; F)$ is a set of equicontinuous functions, then every function $f \in \overline{S}$ is continuous, and for every sequence $(f_n)$ of functions $f_n \in S$, if $(f_n)$ converges pointwise to a function $f \in F^E$, then $f$ is continuous.

**Proof.** Since $S$ is equicontinuous at $x_0$, for every $\varepsilon > 0$, there is some open subset $U$ containing $x_0$ such that
\[ d_F(f(x_0), f(x)) \leq \varepsilon, \quad \text{for all } f \in S \text{ and all } x \in U. \]
But for $x \in U$ fixed, the map $f \mapsto (f(x_0), f(x))$ from $F^E$ to $F^2 = F \times F$ is continuous (this is a projection onto a product), and $d_F$ is continuous on $F^2$. As a consequence, the set
\[ \{ f \in F^E \mid d_F(f(x_0), f(x)) \leq \varepsilon \} \]
1.3. EQUICONTINUOUS SETS OF CONTINUOUS FUNCTIONS

is closed in $F^E$, and since it contains $S$, it also contains $\overline{S}$. Thus, for every $\epsilon > 0$, we found an open subset $U$ containing $x_0$ such that $d_F(f(x_0), f(x)) \leq \epsilon$ for all $x \in U$ and all $f \in \overline{S}$, which means that $\overline{S}$ is equicontinuous.

Since every function in an equicontinuous set of functions is continuous, every function $f \in \overline{S}$ is continuous. By definition of the pointwise topology, if a sequence $(f_n)$ of functions $f_n \in S$ converges pointwise to a function $f \in F^E$, then $f \in \overline{S}$, so $f$ is continuous.

Dieudonné proves a weaker version of Theorem 1.9, namely that for every subset $S$ of the space of bounded continuous functions $C_b(E; F)$, if $S$ is equicontinuous, then its closure $\overline{S}$ is also equicontinuous. This is Proposition 7.5.4 in [23] (Chapter 7, Section 5).

The second version of Ascoli’s theorem involves a dense subset $E_0$ of $E$. We need the following variant of Definition 1.5. The topology of pointwise convergence in $E_0$ is the topology on $F^E$ having the sets

$$S(x, U) = \{ f \mid f \in F^E, f(x) \in U \}, \quad x \in E_0, U \text{ open in } F,$$

as a subbasis.

**Theorem 1.10.** (Ascoli II) Let $E$ be a topological space, let $(F, d_F)$ be a metric space, $E_0$ be a dense subset of $E$, and $S \subseteq C(E; F)$ be set of equicontinuous functions. Then the topology of pointwise convergence in $E_0$, the topology of pointwise convergence, and the topology of compact convergence (all three topologies being defined in $F^E$), induce identical topologies on $S$.

Theorem 1.10 is proven in Schwartz [69] (Chapter XX, Theorem XX.3.1). The following corollaries of Theorem 1.10 are particularly useful. The first of these two propositions is an immediate consequence of Theorem 1.10.

**Proposition 1.11.** Let $E$ be a topological space and let $(F, d_F)$ be a metric space. If a sequence $(f_n)$ of continuous functions $f_n \in C(E; F)$ converges pointwise for every $x \in E_0$; (2) The set $\{f_n\}$ is equicontinuous; (3) The set $\{f_n(x) \mid n \geq 1\}$ is contained in a complete subset of $F$ for every $x \in E$; then the sequence $(f_n)$ converges pointwise (for all $x \in E$) to a continuous function $f$, and $(f_n)$ converges uniformly to $f$ on every compact subset. If $F$ complete, then Condition (3) is automatically satisfied and can be omitted.
Proof. Since by Theorem 1.10, the topology of pointwise convergence on $E_0$ is identical to the topology of pointwise convergence on $E$, as the sequence $(f_n)$ converges pointwise for every $x \in E_0$, it also converges pointwise for every $x \in E$. This implies that for every $x$, the sequence $(f_n(x))$ is a Cauchy sequence in $F$, but since by (3) the set $\{f_n(x) \mid n \geq 1\}$ is contained in a complete subset of $F$, the sequence $(f_n(x))$ converges. Thus $(f_n)$ converges pointwise to a function $f \in F^E$, and since $\{f_n\}$ is equicontinuous, by Proposition 1.11, the function $f$ is continuous, and $(f_n)$ converges uniformly to $f$ on every compact subset. □

Dieudonné proves a special case of Proposition 1.12 where $E$ is a metric space, $F$ is a complete normed vector space (a Banach space), the functions $f_n$ are continuous and bounded, and $\{f_n\}$ is equicontinuous; see Proposition 7.5.5 and Proposition 7.5.6 in [23] (Chapter 7, Section 5).

In most applications, $E$ is a metric space and $F$ is a (complete) normed vector space. The following result about sets of continuous linear maps will be needed.

**Proposition 1.13.** Let $E$ be a metrizable vector space and $F$ be a normed vector space. A subset of continuous linear maps $S \subseteq L(E; F)$ is equicontinuous if and only if there is some open subset $V \subseteq E$ containing 0 and some real $c > 0$ such that $\|f(x)\| \leq c$ for all $x \in V$ and all $f \in S$.

Proof. If $S$ is equicontinuous then obviously the property of the proposition holds. Conversely, for any $\epsilon > 0$, the condition $\|f(x)\| \leq c$ for all $x \in V$ and all $f \in S$ implies that $\|f(x)\| \leq \epsilon$ for all $x \in (\epsilon/c)V$ and all $f \in S$, so $S$ is equicontinuous at 0. For any $x_0 \in E$, and for all $x \in x_0 + (\epsilon/c)V$, we have

$$\|f(x) - f(x_0)\| = \|f(x - x_0)\| \leq \epsilon$$

for all $f \in S$, that is, $S$ is equicontinuous at $x_0$. □

A third version of Ascoli’s theorem involving relative compactness will be needed in Section 14.1. Recall that a subset $A$ of a Hausforff space $X$ is relatively compact if its closure $\overline{A}$ is compact in $X$.

**Theorem 1.14. (Ascoli III)** Let $E$ be a topological space, let $(F,d_F)$ be a metric space, and let $S \subseteq C(E; F)$ be set of continuous functions. Assume the following two conditions hold:

1. The set $S$ is equicontinuous.

2. For every $x \in E$, the set $S(x) = \{f(x) \mid f \in S\}$ is relatively compact in $F$.

Then the set $S$ is relatively compact in the space $(F^E)_c$ of continuous functions from $E$ to $F$ with the topology of compact convergence. Conversely, if $E$ is locally compact and if the set $S$ is relatively compact in the space $(F^E)_c$, then Conditions (1) and (2) hold.
1.4. REGULATED FUNCTIONS

Proof. A complete proof is given in Schwartz [69] (Chapter XX, Theorem XX.4.1). We only prove the first part of the theorem. The proof uses Tychonoff’s powerful theorem. By hypothesis, for every $x \in E$, the closure $\overline{S(x)}$ of $S(x)$ is compact in $F$, so by Tychonoff’s theorem, the product $\prod_{x \in E} \overline{S(x)}$ is compact in $F$. By definition of the above product, this means that the set $\hat{S}$ of functions $f \in F^E$ such that $f(x) \in \overline{S(x)}$ for all $x \in E$ is compact in $F^E$ with the topology of pointwise convergence. Since $S$ is contained in the compact set $\hat{S}$, we deduce that its closure $\overline{S}$ is compact in $F^E$ with the topology of pointwise convergence. Since $S$ is equicontinuous, the set $S$ is also equicontinuous. By Ascoli I (Theorem 1.9), since $S$ is equicontinuous, the set $S$ is also equicontinuous. By Ascoli II (Theorem 1.10), since the restriction to $S$ of the topology of pointwise convergence on $F^E$ coincides with the restriction to $\overline{S}$ of the topology of compact convergence on $F^E$, the set $S$ is compact in $(F^E)_c$, and thus $S$ is relatively compact in $(F^E)_c$.

The special case of Theorem 1.14 in which $E$ is compact and $F$ is a Banach space is proven in Dieudonné [23] (Chapter 7, Section 5, Theorem 5.7.5). Because $F$ is complete the proof is simpler and does not use Tychonoff’s theorem.

1.4 Regulated Functions

Recall that there are four kinds of intervals of $\mathbb{R}$: $(a,b)$, $[a,b)$, $(a,b]$, and $[a,b]$, with $a < b$.

By convention, $(a,b) = [a,b]$ if $a = \infty$, and $(a,b) = (a,b]$ if $b = \infty$.

Definition 1.13. Let $I$ be an interval of $\mathbb{R}$, and let $F$ be a metric space (or a normed vector space). Given a function $f : I \to F$, for any $x \in I$ with $x \neq b$, we say that $f$ has a limit to the right in $x$ if $\lim_{y \in I, y \to x} f(y)$ exists as $y \in I$ tends to $x$ from above. This limit is denoted by $f(x^+)$. For any $x \in I$ with $x \neq a$, we say that $f$ has a limit to the left in $x$ if $\lim_{y \in I, y \to x} f(y)$ exists as $y \in I$ tends to $x$ from below. This limit is denoted by $f(x^-)$. Given any interval $I$, a function $f : I \to F$ is a regulated function (or ruled function) if it has a left limit and a right limit for every $x \in I$. If $F$ is a normed vector space, a function $f : \mathbb{R} \to F$ is a regulated function (or ruled function) if there is some interval $I$ such that $f$ vanishes outside $I$, and the restriction $f : I \to F$ of $f$ to $I$ is regulated. See Figure 1.9.

The notion of a regulated function can also be defined in terms of certain kinds of discontinuities.

Definition 1.14. Let $I$ be an interval of $\mathbb{R}$, and let $F$ be a metric space (or a normed vector space). Given a function $f : I \to F$, we say that a point $x \in I$ is a discontinuity of the first kind if the left limit $f(x^-)$ and the right limit $f(x^+)$ both exist, but $f(x-) \neq f(x)$ or $f(x+) \neq f(x)$.

It is clear that a function $f : I \to F$ is regulated iff for every $x \in I$, either $f$ is continuous or $x$ is a discontinuity of the first kind. Thus every continuous function is a regulated function. It is also easy to see that a monotonic function $f : I \to \mathbb{R}$ is a regulated function.
CHAPTER 1. FUNCTION SPACES OFTEN ENCOUNTERED

Figure 1.9: An illustration of a regulated function $f : \mathbb{R} \to \mathbb{R}$. This function has three discontinuities $x_1$, $x_2$, and $x_3$, each of the first kind. Note that $f(x_1-) = y_4$, $f(x_1+) = f(x_1) = y_2$, $f(x_2-) = f(x_2) = y_3$, $f(x_2+) = y_6$, $f(x_3-) = y_3$, $f(x_3+) = y_5$, yet $f(x_3) = y_1$.

The function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(x) = \begin{cases} \sin \left( \frac{1}{x} \right) & \text{if } x \neq 0 \\ 0 & \text{if } x = 0 \end{cases}$$

is discontinuous at $x = 0$, but this is not a discontinuity of the first kind. See Figure 1.10.

Figure 1.10: The graph of $f(x) = \sin \left( \frac{1}{x} \right)$, $x \neq 0$.

The following result is shown in Schwartz [71] (Chapter III, Section 2, Theorem 3.2.3).

**Proposition 1.15.** If $f : I \to F$ is a regulated function (where $F$ is a metric space), then $f$ has at most countably many discontinuities of the first kind.
Regulated functions on a closed and bounded interval $[a, b]$ must be bounded. As a consequence, they arise as limits of uniformly convergent sequences of step functions.

**Definition 1.15.** A function $f: \mathbb{R} \to F$ (where $F$ is any set) is a step function if there is a finite sequence $(a_0, a_1, \ldots, a_n)$ of reals such that $a_k < a_{k+1}$ for $k = 0, \ldots, n$, and $f$ is constant on each of the open intervals $(-\infty, a_0)$, $(a_k, a_{k+1})$ for $k = 0, \ldots, n$, and $(a_n, +\infty)$. The sequence $(a_0, a_1, \ldots, a_n)$ is called an admissible subdivision for $f$. See Figure 1.11.

![Figure 1.11: An illustration of a step $f: \mathbb{R} \to \mathbb{R}$ with admissible subdivision $(a_0, a_1, a_2, a_3, a_4)$.](image)

Observe that Definition 1.15 does not make any restriction on the values $f(a_k)$, but a step function is regulated. Also, a given step function admit infinitely many admissible subdivisions, by refining a given subdivision. By a step function $f: [a, b] \to F$, we mean a step function such that $f(x) = 0$ for all $x \leq a$ and for all $x \geq b$.

The following result is easy to prove.

**Proposition 1.16.** If $F$ is a vector space, then the set of step functions $f: \mathbb{R} \to F$ is a vector space denoted by Step($\mathbb{R}; F$). The set of step functions $f: [a, b] \to F$ is also vector space denoted by Step([a, b]; F).

The following proposition is much more interesting.

**Proposition 1.17.** Let $F$ be a metric space and let $[a, b]$ be a closed and bounded interval. Then every regulated function $f: [a, b] \to F$ is the limit of a uniformly convergent sequence $(f_n)_{n \geq 1}$ of step functions $f_n: [a, b] \to F$. Furthermore, if $F$ is a complete metric space, then the limit of any uniformly convergent sequence $(f_n)_{n \geq 1}$ of step functions is a regulated function.
The proof of Proposition 1.17 is given in Schwartz [71] (Chapter III, Section 2, Theorem 3.2.9).

As a corollary of Proposition 1.17, if $F$ is a complete metric space, then the space of regulated functions on $[a, b]$ is closed in $(F[a, b])_b$, and the space of step functions on $[a, b]$ is dense in the space of regulated functions on $[a, b]$. Thus if $F$ is complete, since $(F[a, b])_b$ is complete, the space of regulated function on $[a, b]$ is also complete.

Another corollary of Proposition 1.17 is that every continuous function $f : [a, b] \to F$ to a metric space $F$ is the limit of a uniformly convergent sequence $(f_n)_{n \geq 1}$ of step functions $f_n : [a, b] \to F$.

If $F$ is a vector space, the set of regulated functions defined on the closed and bounded interval $[a, b]$ is a vector space denoted by $\text{Reg}([a, b]; F)$. Then, Proposition 1.17 implies the following result.

**Proposition 1.18.** Let $F$ be a complete normed vector space. The space $\text{Reg}([a, b]; F)$ of regulated functions on $[a, b]$ is complete, and the space $\text{Step}([a, b]; F)$ is dense in $\text{Reg}([a, b]; F)$.

Step functions can be used to define the Riemann integral. To do so it is convenient to consider functions of finite support.

**Definition 1.16.** Given any function $f : E \to F$, where $E$ is a topological space and $F$ is a vector space, the support $\text{supp}(f)$ of $f$ is the closure of the subset of $E$ where $f$ is nonzero, that is, $\text{supp}(f) = \{x \in E \mid f(x) \neq 0\}$. The function $f$ has compact support if its support $\text{supp}(f)$ is compact. If $E$ is Hausdorff, this is equivalent to saying that $f$ vanishes outside some compact subset $K$ of $E$. See Figure 1.12.

If a step function $f$ has compact support, then we assume that $f$ vanishes on $(-\infty, a_0)$ and on $(a_n, +\infty)$ for any admissible subdivision $(a_0, a_1, \ldots, a_n)$ for $f$.

It is easy to see that the set of continuous functions $f : E \to F$ with compact support is a vector space.

**Definition 1.17.** The vector space of continuous functions $f : E \to F$ with compact support is denoted by $C_c(E; F)$, or $\mathcal{K}(E; F)$. For every compact subset $K$ of $E$, we denote by $\mathcal{K}(K; F)$ the space of continuous functions whose support is contained in $K$. Then

$$\mathcal{K}(E; F) = \bigcup_{K \subseteq E, \text{compact}} \mathcal{K}(K; F).$$

Observe that every function in $\mathcal{K}(E; F)$ is bounded, that is, $\mathcal{K}(E; F) \subseteq C_b(E; F)$.

If $F = \mathbb{R}$ or $F = \mathbb{C}$, then we write $\mathcal{K}_\mathbb{R}(E)$ or $\mathcal{K}_\mathbb{C}(E)$ for $\mathcal{K}(E; F)$. Radon functionals are certain kinds of linear forms on $\mathcal{K}_\mathbb{C}(\mathbb{C})$.

If $(F, \| \|)$ is a Banach space and $K$ is a fixed compact subset of $E$, then so is $\mathcal{K}(K; F)$ (for the sup norm $\| \|_\infty$), because it is closed in $C_b(E; F)$. However, the normed vector space $(\mathcal{K}(E; F), \| \|_\infty)$ is not complete!
Example 1.3. For every \( n \geq 1 \), consider the function \( u_n : \mathbb{R} \to \mathbb{R} \) defined as follows:

\[
    u_n(x) = \begin{cases} 
        1 & \text{if } -n \leq x \leq n \\
        x + n + 1 & \text{if } -(n+1) \leq x \leq -n \\
        -x + n + 1 & \text{if } n \leq x \leq n+1 \\
        0 & \text{if } |x| \geq n + 1.
    \end{cases}
\]

Now consider the sequence of functions \( (f_n) \) given by

\[
    f_n(x) = u_ne^{-|x|}.
\]

Each function \( f_n \) is continuous and has compact support \([- (n + 1), n + 1]\), and it is easy to show that the sequence \( (f_n) \) converges uniformly to the function \( f \) given by \( f(x) = e^{-|x|} \), but \( f \) does not have compact support. The problem is that the domains of the functions \( f_n \), although compact, keep growing as \( n \) goes to infinity. See Figure 1.13.
Figure 1.13: The functions of Example 1.3. Figure (i) illustrates the \( u_1(x) \) in magenta; \( u_2(x) \) in red, \( u_3(x) \) in orange, \( u_4(x) \) in purple, and \( u_5(x) \) in blue. Figure (ii) uses the same color scheme to illustrate the corresponding \( f_n(x) \). Note these \( f_n(x) \) converge uniformly to green \( f(x) = e^{-|x|} \).

support, there is some compact subset \( K \) of \( E \) such that \( \|f(x)\| \leq \epsilon \) for all \( x \in E - K \). This suggests the following definition.

**Definition 1.18.** The subspace of \( C_b(E; F) \), denoted \( C_0(E; F) \), consisting of the continuous functions \( f \) such that for every \( \epsilon > 0 \), there is some compact subset \( K \) of \( E \) such that \( \|f(x)\| \leq \epsilon \) for all \( x \in E - K \), is called the space of **continuous functions which tend to 0 at infinity**.

Observe that if \( X = \mathbb{R} \), then a function \( f \in C_0(\mathbb{R}; F) \) does indeed have the property that \( \lim_{x \to -\infty} f(x) = \lim_{x \to +\infty} f(x) = 0 \).

We showed that \( \mathcal{K}(E; F) \subseteq C_0(E; F) \). If \( E \) is locally compact, then we have the following result from Dieudonné [22] (Chapter XIII, Section 20) and Rudin [66] (Chapter 3, Theorem 3.17).

**Proposition 1.19.** If \( E \) is locally compact, then \( C_0(E; \mathbb{C}) \) is the closure of \( \mathcal{K}(E; \mathbb{C}) \) in \( C_b(E; \mathbb{C}) \). Consequently, \( C_0(E; \mathbb{C}) \) is complete.

**Proof.** We already showed just before definition 1.18 that if a function \( f \) belongs to the closure of \( \mathcal{K}(E; \mathbb{C}) \), then it tends to zero at infinity. Conversely, pick any \( f \) in \( C_0(E; \mathbb{C}) \).
For every \( \varepsilon > 0 \), there is a compact subset \( K \) of \( E \) such that \( |f(x)| < \varepsilon \) outside of \( K \). By Proposition 15.39, there is continuous function \( g: E \to [0,1] \) with compact support such that \( g(x) = 1 \) for all \( x \in K \). Clearly \( fg \in \mathcal{K}_c(E) \), and \( \|fg - f\|_\infty < \varepsilon \). This shows that \( \mathcal{K}(E; \mathbb{C}) \) is dense in \( \mathcal{C}_0(E; \mathbb{C}) \).

In summary, if \( E \) is locally compact, then we have the strict inclusions
\[
\mathcal{K}(E; \mathbb{C}) \subset \mathcal{C}_0(E; \mathbb{C}) \subset \mathcal{C}_b(E; \mathbb{C}),
\]
with \( \mathcal{C}_0(E; \mathbb{C}) \) and \( \mathcal{C}_b(E; \mathbb{C}) \) complete, and \( \mathcal{K}(E; \mathbb{C}) \) dense in \( \mathcal{C}_0(E; \mathbb{C}) \). It turns out that the space of continuous linear forms on \( \mathcal{C}_0(E; \mathbb{C}) \) is isomorphic to the space of bounded Radon functionals.

A modified version of step functions involving a measure will be used to define the integral on a measure space.

### 1.5 Neighborhood Bases and Filters

When dealing with convolution we will need a notion of convergence more general than the notion of convergence of a sequence. There are two equivalent definitions of such a general notion of convergence. One in terms of nets, and the other in terms of filters. For our purposes, the definition in terms of filters is more convenient.

First let us review the notion of neighborhood and neighborhood base.

**Definition 1.19.** Let \( X \) be a topological space whose topology is specified by a set \( \mathcal{O} \) of open sets. For any subset \( A \subseteq X \), a **neighborhood** of \( A \) is any subset \( N \) containing some open subset \( U \) containing \( A \); in short, there is some \( U \in \mathcal{O} \) such that \( A \subseteq U \subseteq N \). If \( A = \{x\} \), a neighborhood of \( \{x\} \) is called simply a **neighborhood of \( x \)**.

A **neighborhood base** of a point \( x \) (resp. of a subset \( A \)) is a family \( \mathcal{N} \) of neighborhoods of \( x \) (resp. of neighborhoods of \( A \)), such for every neighborhood \( V \) of \( x \) (resp. neighborhood of \( A \)), there is some \( N \in \mathcal{N} \) such that \( N \subseteq V \).

In many cases, a neighborhood base consists of open sets. Let us now define the notion of filter and filter base. This notion is defined for any set, not just for a topological space.

**Definition 1.20.** Let \( X \) be any set. A **filter** \( \mathcal{F} \) on \( X \) is a family of subsets of \( X \) satisfying the following properties.

1. For any two subset \( A, B \) of \( X \), if \( A \in \mathcal{F} \) and if \( A \subseteq B \), then \( B \in \mathcal{F} \) (**\( \mathcal{F} \) is upward-closed**).
2. For any two subsets \( A, B \) of \( X \), if \( A \in \mathcal{F} \) and \( B \in \mathcal{F} \), then \( A \cap B \in \mathcal{F} \) (**closure under intersection**).
3. We have \( X \in \mathcal{F} \).
(4) The empty set does not belong to \( F \).

The axioms of a filter show that filters only exist on nonempty sets. In particular, Axiom (4) prevents \( F = 2^X \) to be a filter.

**Example 1.4.**

1. If \( X \neq \emptyset \), for any nonempty subset \( A \) of \( X \), the family of all subsets of \( X \) containing \( A \) is a filter.

2. If \((X, \mathcal{O})\) is a topological space, then for any \( x \in X \) (resp. any nonempty subset \( A \) of \( X \)), the family of neighborhoods of \( x \) (resp. \( A \)) is a filter.

3. If \( X \) is an infinite set, the family of complements of finite subsets of \( X \) is a filter. If \( X = \mathbb{N} \), then the filter of complements of finite subsets of \( \mathbb{N} \) is called the Fréchet filter.

4. Let \( F \) be a filter on \( X \). For any \( A \in F \), let \( S(A) \) be the family

\[
S(A) = \{ B \in F \mid B \subseteq A \},
\]

called a section. It is easy to check that the family of sections \( S(A) \) (for all \( A \in F \)) is a filter on the set \( F \), called the filter of sections of \( F \).

Filters are compared as follows.

**Definition 1.21.** Let \( X \) be any nonempty set. Given two filters \( F \) and \( F' \) on \( X \), we say that \( F' \) is finer than \( F \) if \( F \subseteq F' \).

A convenient way to generate a filter is to use a filter base.

**Definition 1.22.** Let \( X \) be any nonempty set. A filter base \( B \) on \( X \) is a family of subsets of \( X \) satisfying the following properties.

1. For any two subsets \( A, B \) of \( X \), if \( A \in B \) and \( B \in B \), then there is some \( C \in B \) such that \( C \subseteq A \cap B \).

2. The family \( B \) is nonempty.

3. The empty set does not belong to \( B \).

It is immediately verified that if \( B \) is a filter base on \( X \), then the family of subsets of \( X \) containing some subset in \( B \) is a filter called the filter generated by \( B \).

If \((X, \mathcal{O})\) is a topological space, for any \( x \in X \), the filter bases of neighborhoods of \( x \) are exactly the neighborhood bases of \( x \).

The main reason for introducing filters is to define the following general notion of convergence.
Definition 1.23. Let $X$ be a topological space whose topology is specified by a set $\mathcal{O}$ of open sets. For any $x \in X$, a filter $\mathcal{F}$ converges to $x$, or $x$ is a limit of the filter $\mathcal{F}$, if every neighborhood $N$ of $x$ belongs to $\mathcal{F}$; equivalently, the filter $\mathcal{B}(x)$ of neighborhoods of $x$ is a subset of the filter $\mathcal{F}$; that, is the filter $\mathcal{F}$ is finer than the filter $\mathcal{B}(x)$. A filter base $\mathcal{B}$ converges to $x$ if the filter generated by $\mathcal{B}$ converges to $x$.

The following proposition is an immediate consequence of the definition.

Proposition 1.20. Let $X$ be a topological space whose topology is specified by a set $\mathcal{O}$ of open sets. For any $x \in X$, a filter base $\mathcal{B}$ converges to $x$ iff every neighborhood base of $x$ contains some set in $\mathcal{B}$.

Intuitively, $x$ is a limit of a filter base $\mathcal{B}$ if there are sets in $\mathcal{B}$ as close to $x$ as desired.

The limit of a sequence $(x_n)_{n \geq 0}$ of points $x_n \in X$ is a special case of Definition 1.23. Indeed, if we define for every $n \geq 0$ the set $S_n$ given by

$$S_n = \{x_p \mid p \geq n\},$$

then the family of sets $S_n$ forms a filter base, and an element $y \in X$ is a limit of the sequence $(x_n)$ iff the filter base $\{S_n\}$ converges to $y$. Indeed, by Proposition 1.20, the filter base $\{S_n\}$ converges to $y$ iff for every neighborhood $V$ of $y$, there is some $S_n \subseteq V$, in other words, there is some $n \geq 1$ such that $x_p \in V$ for all $p \geq n$, which is the standard definition of convergence of a sequence.

We can also define the notion of limit of a function. Let $f : X \to Y$ be a function where $X$ is any nonempty set and $Y$ is a topological space. Then if $\mathcal{F}$ is any filter on $X$, it is immediately verified that the family of sets $f(U)$, with $U \in \mathcal{F}$, is a filter base on $Y$.

Definition 1.24. Let $f : X \to Y$ be a function where $X$ is any nonempty set and $Y$ is a topological space. For any filter $\mathcal{F}$ on $X$, and for any $y \in Y$, we say that $y$ is a limit of $f$ according to $\mathcal{F}$ (or simply that $y$ is a limit of $\mathcal{F}$) if the filter basis $f(\mathcal{F})$ (consisting of the subsets $f(U)$ of $Y$ with $U \in \mathcal{F}$) converges to $y$. We write

$$\lim_{x,\mathcal{F}} f(x) = y.$$

If we view a sequence $(x_n)_{n \geq 0}$ of points in a topological space $X$ as a function $x : \mathbb{N} \to X$, then $(x_n)$ converges to $y$ in the traditional sense iff $x$ converges to $y$ according to the Fréchet filter on $\mathbb{N}$ (the family of subsets of $\mathbb{N}$ having a finite complement).

The following useful characterization of a limit of a filter is immediate from the definitions.

Proposition 1.21. Let $f : X \to Y$ be a function where $X$ is any nonempty set and $Y$ is a topological space. A point $y \in Y$ is a limit of a filter $\mathcal{F}$ on $X$ if and only if for every neighborhood $V$ of $y$, there is some $W \in \mathcal{F}$ such that $f(W) \subseteq V$, or equivalently, $f^{-1}(V) \in \mathcal{F}$ for every neighborhood $V$ of $y$. 
Filters also provide a useful characterization of the notion of compactness. First we define ultrafilters.

**Definition 1.25.** Let $X$ be any nonempty set. A filter $\mathcal{F}$ on $X$ is an *ultrafilter* if it is a maximal filter; that is, there is no filter different from $\mathcal{F}$ and finer than $\mathcal{F}$.

For example, for any $x \in X$, the filter of subsets containing $x$ is an ultrafilter. The following important result shows that there many ultrafilters, but they are very nonconstructive in nature. The proof uses Zorn’s lemma.

**Theorem 1.22.** Let $X$ be any nonempty set. Every filter $\mathcal{F}$ on $X$ is contained in a finer ultrafilter.

Observe that an ultrafilter $\mathcal{F}$ has the following completeness property: for any subset $A$ of $X$, either $A \in \mathcal{F}$ or its complement $X - A \in \mathcal{F}$, but not both.

Indeed, if $A \notin \mathcal{F}$ and $X - A \notin \mathcal{F}$, then it is easy to see that the family $\mathcal{G}$ of subsets of $X$ given by

$$\mathcal{G} = \{B \subseteq X \mid A \cup B \in \mathcal{F}\}$$

is a filter finer than $\mathcal{F}$ and containing $X - A$, thus strictly finer than $\mathcal{F}$, contradicting the maximality of $\mathcal{F}$. This property of ultrafilters is used in logic to prove completeness results.

We also need to define a notion weaker than the notion of limit of a filter.

**Definition 1.26.** Let $X$ be a topological space whose topology is specified by a set $\mathcal{O}$ of open sets. A point $x \in X$ is a *cluster point* (or *cluster*) of a filter base $\mathcal{B}$ if every neighborhood of $x$ has a nonempty intersection with every set in $\mathcal{B}$ (equivalently, if $x \in \bigcap_{V \in \mathcal{B}} V$).

A limit $x$ of a filter is a cluster point, but the converse is false in general.

We see immediately that $x$ is a cluster point of a filter $\mathcal{F}$ iff there is a filter $\mathcal{G}$ finer than $\mathcal{F}$ and the filter $\mathcal{G}$ converges to $x$. An ultrafilter $\mathcal{F}$ converges to a limit $x$ iff $x$ is a cluster point of $\mathcal{F}$.

Finally, we have the following characterizations of compactness.

**Theorem 1.23.** Let $X$ be a topological space whose topology is specified by a set $\mathcal{O}$ of open sets. The following properties are equivalent.

1. Every filter $\mathcal{F}$ on $X$ has some cluster point.
2. Every ultrafilter $\mathcal{F}$ on $X$ converges to some limit.
3. Every open cover $(U_\alpha)_{\alpha \in I}$ of $X$ contains some finite subcover; that is, if $\bigcup_{\alpha \in I} U_\alpha = X$, then there is a finite subset $J$ of $I$ such that $\bigcup_{\alpha \in J} U_\alpha = X$.
4. For every family $(F_\alpha)_{\alpha \in I}$ of closed subsets of $X$, if $\bigcap_{\alpha \in I} F_\alpha = \emptyset$, then there is a finite subset $J$ of $I$ such that $\bigcap_{\alpha \in J} F_\alpha = \emptyset$. 

Let us also mention that a topological space $X$ is Hausdorff if and only if every filter has at most one limit.

The theory of filters and their use in topology is discussed quite extensively in Bourbaki [11] (Chapter 1).

### 1.6 Topologies Defined by Semi-Norms; Fréchet Spaces

Certain function spaces, such as the space $\mathcal{C}(X; \mathbb{C})$ of continuous functions on a topological space $X$, do not come with “natural” topologies defined by a norm or a metric for which they are complete. However, the weaker notion of semi-norm can be used to define a topology, and under certain conditions, although such topologies are not defined by any norm, they are metrizable, and complete. In this section, we briefly discuss the use of semi-norms to define topologies. It turns out that the corresponding spaces are locally convex.

Recall from Definition 16.1 that a semi-norm satisfies Properties (N2) and (N3) of a norm, but in general does not satisfy Condition (N1), so $\|x\| = 0$ does not necessarily imply that $x = 0$. Here is a method for defining a topology on a vector space using a family of semi-norms.

**Definition 1.27.** Let $X$ be a vector space and let $(p_\alpha)_{\alpha \in I}$ be a family of semi-norms on $X$. For every $x \in X$, every $\epsilon > 0$, and every $\alpha \in I$, let

$$U_{x,\alpha,\epsilon} = \{ y \in X \mid p_\alpha(y - x) < \epsilon \}.$$  

The **topology induced by the family of semi-norms** $(p_\alpha)_{\alpha \in I}$ is the weakest (coarsest) topology whose open sets are arbitrary unions of finite intersections of subsets of the form $U_{x,\alpha,\epsilon}$.

We can think of the subset $U_{x,\alpha,\epsilon}$ as an open ball of center $x$ and radius $\epsilon$ in $X$, determined by the semi-norm $p_\alpha$. We have made our vector space $X$ into a topological space but it is not clear that the operations (addition and scalar multiplication) are continuous. Also, in general, this topology is not Hausdorff. The following proposition addresses these issues.

**Proposition 1.24.** Let $X$ be a vector space and let $(p_\alpha)_{\alpha \in I}$ be a family of semi-norms on $X$.

1. With the topology induced by the family of semi-norms $(p_\alpha)_{\alpha \in I}$, addition and scalar multiplication are continuous, so $X$ is a topological vector space.

2. For every $x \in X$, the finite intersections of subsets of the form $U_{x,\alpha,\epsilon}$ is a neighborhood base of $x$.

3. Every open set $U_{x,\alpha,\epsilon}$ is convex.

4. Every $p_\alpha$ is continuous.
(5) The topology induced by the family of semi-norms is Hausdorff if and only if, for every \( x \neq 0 \), there is some \( \alpha \in I \) such that \( p_\alpha(x) \neq 0 \).

Proposition 1.24 is proven in Folland [32] (Chapter 5, Section 5.4, Theorem 5.14), or Rudin [67] (Chapter 1, Theorem 1.37). In view of Property (3), the topological space \( X \) is said to be **locally convex**.

Two good examples of topologies induced by families of seminorms are the topology of pointwise convergence and the topology of compact convergence on a normed vector space \( F \). If we define the family of semi-norms \((p_x)_{x \in E}\) by

\[
p_x(f) = \|f(x)\|, \quad f \in F^E, \; x \in E,
\]

then it is easy to see that the topology defined by the family \((p_x)_{x \in E}\) is the topology of pointwise convergence on \( F^E \), which has the subsets

\[
S(x, U) = \{f \mid f \in F^E, f(x) \in U\}, \quad x \in E, \; U \text{ open in } F,
\]

as a subbasis.

If we define the family of semi-norms \( \{p_K \mid K \text{ compact in } E\} \), by

\[
p_K(f) = \sup_{x \in K} \|f(x)\|, \quad f \in F^E, \; K \text{ compact in } E,
\]

then it is easy to see that the topology defined by the family \((p_K)\) is the topology of compact convergence on \( F^E \), which has the subsets

\[
B_K(f, \epsilon) = \left\{ g \in F^E \mid \sup_{x \in K} d(f(x), g(x)) < \epsilon \right\}
\]

as a subbasis.

In a vector space \( X \) whose topology defined by a family of semi-norms \((p_\alpha)_{\alpha \in I}\) is Hausdorff, it is easy to see that the convergence of a sequence \((x_n)\) to a limit \( x \) is expressed conveniently as follows: for every \( \alpha \in I \), for every \( \epsilon > 0 \), there is some \( N_\alpha > 0 \) such that \( p_\alpha(x - x_n) \leq \epsilon \) for all \( n \geq N_\alpha \); equivalently, \( \lim_{n \to \infty} p_\alpha(x - x_n) = 0 \), for every \( \alpha \in I \).

When the index family \( I \) is countable and the topology induced by a family of semi-norm is Hausdorff, then \( X \) is actually metrizable.

**Proposition 1.25.** Let \( X \) be a vector space and let \((p_\alpha)_{\alpha \in I}\) be a family of semi-norms on \( X \). If the topology induced by \((p_\alpha)_{\alpha \in I}\) is Hausdorff and if \( I \) is countable, then \( X \) is metrizable with a translation-invariant metric \( d \) (this means that \( d(a, b) = d(a+u, b+u) \) for all \( a, b, u \in X \)). In fact, we can use the metric \( d \) given by

\[
d(x, y) = \sum_{m=0}^{\infty} \frac{1}{2^m} \frac{p_m(y-x)}{1 + p_m(y-x)}.
\]
Proposition 1.24 is proven in Dieudonné [22] (Chapter 12, Section 4, Theorem 12.4.6), and in Rudin [67] (Chapter 1, Page 29, with an equivalent metric).

**Definition 1.28.** A vector space $X$ whose topology is defined by a countable family of semi-norms, and which is Hausdorff and complete for some translation-invariant metric defining the topology of $X$ is called a Fréchet space.

If a topological vector space $X$ is Hausdorff and complete for some translation-invariant metric $d$ defining the topology of $X$, then it is also complete for every translation-invariant metric $d'$ defining the topology of $X$, as shown by the following proposition.

**Proposition 1.26.** Let $X$ be a metrizable Hausdorff topological vector space. For any translation-invariant metric $d$ defining the topology of $X$, a sequence $(x_n)$ is a Cauchy sequence if and only if for every neighborhood $V$ of 0, there is some $N > 0$ such that $x_m - x_n \in V$ for all $m, n$ such that $m \geq N$ and $n \geq N$.

**Proof.** A slightly more general result is proven for topological groups in Dieudonné [22] (Chapter 12, Section 9, Theorem 12.9.2) and Rudin [67] (Chapter 1, page 21). If a metric $d$ defining the topology of $X$ is translation-invariant, then $d(x_n, x_m) = d(0, x_m - x_n)$, and the sequence $(x_n)$ is a Cauchy sequence if and only if for every $\epsilon > 0$, there is some $N > 0$ such that $d(0, x_m - x_n) < \epsilon$ for all $m \geq N$ and $n \geq N$, which is equivalent to saying that $x_m - x_n \in V$, where $V$ is the open ball of center 0 and radius $\epsilon$, which is an open subset of $X$, by definition of the metric topology. Conversely, since the topology of $X$ is defined by the metric $d$, every open ball of center 0 is an open set, so the condition of the proposition implies that $(x_n)$ is a Cauchy sequence, for every translation-invariant metric defining the topology of $X$. 

A prime example of a Fréchet space is the space $C(X; \mathbb{C})$ of continuous functions on a separable, locally compact, metrizable space $X$.

Recall from Proposition 15.45 that since $X$ is metrizable, there is a sequence $(U_n)_{n \geq 0}$ of open subsets such that for all $n \in \mathbb{N}$, $U_n \subseteq U_{n+1}$, $\overline{U_n}$ is compact, $\overline{U_n} \subseteq U_{n+1}$, and $X = \bigcup_{n \geq 0} U_n = \bigcup_{n \geq 0} \overline{U_n}$. For every $n \in \mathbb{N}$, define the function $p_n : C(X; \mathbb{C}) \to \mathbb{R}$ by

$$p_n(f) = \sup_{x \in U_n} |f(x)|, \quad f \in C(X; \mathbb{C}).$$

It is immediately verified that the $p_n$ are semi-norms (but they are not norms if $X$ is not compact). For each $f \in C(X; \mathbb{C})$, if $f \neq 0$, then there is some $n$ such that $x \in U_n$, hence $p_n(f) \neq 0$. Thus, the space $C(X; \mathbb{C})$ with the topology induced by the family of semi-norms $(p_n)$ is Hausdorff. By Proposition 1.25, this topology is metrizable. Note that the restriction of $p_{n+1}$ to the compact subset $\overline{U_n}$ is actually a norm, and by definition of the metric $d$ given by Proposition 1.25, the restriction of $d$ to $\overline{U_n}$ is equivalent to $p_{n+1}$.
Proposition 1.27. Let $X$ be a separable, locally compact, metrizable space. The space $\mathcal{C}(X; \mathbb{C})$ with the topology induced by the family of semi-norms $(p_n)$ is complete. Therefore, it is a Fréchet space.

Proof. Since the restriction of the metric $d$ to $\overline{U_n}$ is equivalent to $p_{n+1}$, by Proposition 1.26, a sequence $(f_k)$ of functions in $\mathcal{C}(X; \mathbb{C})$ is a Cauchy sequence if for every $n$, the sequence of restrictions $f_k|_{\overline{U_n}}$ is a Cauchy sequence in the Banach space $\mathcal{C}(\overline{U_n}; \mathbb{C})$, hence converges uniformly in $\overline{U_n}$ to a continuous function $g_n \in \mathcal{C}(\overline{U_n}; \mathbb{C})$. Since $g_{n+1}|_{\overline{U_n}} = g_n$, there exists a continuous function $f \in \mathcal{C}(X; \mathbb{C})$ whose restriction to each $U_n$ agrees with the restriction of $g_n$ to $U_n$. It is clear that $\lim_{m \to \infty} p_n(f - f_m) = 0$ for all $n \geq 0$, hence $f$ is the limit of the Cauchy sequence $(f_k)$, and $\mathcal{C}(X; \mathbb{C})$ is complete. \hfill $\square$

The following result is shown in Dieudonné [22] (Chapter 12, Section 14, Theorem 12.14.6.2).

Proposition 1.28. Let $X$ be a separable, locally compact, metrizable space. The Fréchet space $\mathcal{C}(X; \mathbb{C})$ is separable. In fact, there is a countable dense set consisting of continuous functions with compact support.

Another good example of a Fréchet space is the Schwartz space; see Section 7.6.
Chapter 2

The Riemann Integral

Let \( f : [a, b] \rightarrow \mathbb{R} \) be a continuous function. Intuitively, the Riemann integral \( \int_a^b f(t) \, dt \) is the area of the surface “under the curve” \( t \mapsto f(t) \) from \( x = a \) to \( x = b \). It can be approximated by the sum \( s_T(f) \) (called Cauchy-Riemann sum) of the areas \((t_{k+1} - t_k)f(t_k)\) of \( n \geq 1 \) narrow rectangles, where \( T = (t_0, t_1, \ldots, t_n) \) is any sequence of reals such that \( t_0 = a \), \( t_n = b \) and \( t_k < t_{k+1} \), for \( k = 0, \ldots, n - 1 \); see Figure 2.1. The fact that the function \( f \) is continuous on the compact interval \([a, b] \) implies that the sums \( s_T(f) \) have a limit when the diameter of the subdivision tends to zero (see Definition 2.1), which means the maximum of the distances \( t_{k+1} - t_k \) tends to zero (as \( n \) goes to infinity), and this limit is independent of the subdivision. Thus we can define the Riemann integral \( \int_a^b f(t) \, dt \) as this common limit. The mapping \( f \mapsto \int_a^b f(t) \, dt \) is a positive linear form on the space of continuous functions on \([a, b]\). This procedure applies unchanged to continuous functions \( f : [a, b] \rightarrow F \), where \( F \) is a complete normed vector space.

The method for constructing the integral of a continuous function can be adapted to define the integral of regulated functions (see Definition 1.13). We proceed in two steps:

1. The method of Cauchy-Riemann sums is easily adapted to define the notion of integral for a step function (see Definition 1.15). This yields a mapping \( \tilde{f} : \text{Step}([a, b]; F) \rightarrow F \) which is easily seen to be linear and continuous.

2. By Proposition 1.18, the vector space \( \text{Step}([a, b]; F) \) of steps functions over \([a, b] \) is dense in \( \text{Reg}([a, b]; F) \), the space of regulated functions over \([a, b] \), and \( \text{Reg}([a, b]; F) \) is complete. By Theorem 15.69, the continuous linear map \( \tilde{f} : \text{Step}([a, b]; F) \rightarrow F \) has a unique extension \( \tilde{f} : \text{Reg}([a, b]; F) \rightarrow F \) to \( \text{Reg}([a, b]; F) \), which is also continuous and linear. This is how the integral of a regulated function is defined.

In summary, we define an “obvious” notion of integral on the simple set \( \text{Step}([a, b]; F) \). It is a linear and continuous mapping, so we extend it by continuity to the bigger space \( \text{Reg}([a, b]; F) \) in which \( \text{Step}([a, b]; F) \) is dense.
2.1 Riemann Integral of a Continuous Function

Definition 2.1. Let $a < b$ be any two reals. A set $T = \{t_0, t_1, \ldots, t_n\}$ of reals such that $t_0 = a$, $t_n = b$ and $t_k < t_{k+1}$, for $k = 0, \ldots, n - 1$, is called a subdivision of $[a,b]$. The diameter $\delta(T)$ of $T$ is defined by

$$\delta(T) = \max_{0 \leq k \leq n-1} (t_{k+1} - t_k).$$

Given a continuous function $f : [a,b] \to \mathbb{R}$, define the Cauchy–Riemann sum $s_T(f)$ by

$$s_T(f) = \sum_{k=0}^{n-1} (t_{k+1} - t_k) f(t_k).$$

See Figure 2.1

![Figure 2.1](image)

Figure 2.1: The Cauchy–Riemann sum $s_T(f) = \sum_{k=0}^{n-1} (t_{k+1} - t_k) f(t_k)$ is the area represented by the pastel shaded boxes.

Observe that

$$\sum_{k=0}^{n-1} (t_{k+1} - t_k) = b - a.$$

We immediately check that $s_T$ is a linear form on the set of continuous functions on $[a,b]$. Furthermore, if $f \geq 0$, which means that $f(t) \geq 0$ for all $t \in [a,b]$, then $s_T(f) \geq 0$.

The question is, as the subdivision $T$ becomes finer and finer, in the sense that $\delta(T)$ becomes smaller and smaller (which means that $n$ gets bigger and bigger), do the sums $s_T(f)$ have a limit?

The answer is yes.
The reason is that a continuous function on a compact interval \([a, b]\) is uniformly continuous, and this implies that for any sequence \((T_m)\) of subdivisions such that \(\delta(T_m) \to 0\) as \(m\) goes to infinity, the sums \(s_{T_m}(f)\) form a Cauchy sequence, as we now explain.

Proposition 2.1. Let \(f : [a, b] \to \mathbb{R}\) be a continuous function defined on a closed and bounded (compact) interval \([a, b]\). For every \(\epsilon > 0\), there is some \(\eta > 0\) such that for any two subdivisions \(T\) and \(T'\) of \([a, b]\) such that \(\delta(T) < \eta\) and \(\delta(T') < \eta\), we have

\[
|s_T(f) - s_{T'}(f)| < \epsilon.
\]

Proof. Since a continuous function on \([a, b]\) is actually uniformly continuous, for any \(\epsilon > 0\), we can find some \(\eta > 0\) such that

\[
|f(x) - f(x')| < \epsilon/2(b - a) \quad \text{for all } x, x' \in [a, b] \text{ such that } |x - x'| < \eta.
\]

If \(T = \{t_0, t_1, \ldots, t_n\}\) and \(T' = \{t'_0, t'_1, \ldots, t'_{n'}\}\), let \(T'' = T \cup T'\) and let \(T''_k\) be the subdivision \(T''_k = T' \cap [t_k, t_{k+1}]\), more precisely, \(T''_k = \{s_0, s_1, \ldots, s_r\}\), with \(s_0 = t_k\), \(s_r = t_{k+1}\), and

\[
\{s_1, \ldots, s_{r-1}\} = \{t'_j | t_k < t'_j < t_{k+1}\},
\]

with \(r = 0\) if the above set on the right-hand side is empty, for \(k = 0, \ldots, n - 1\).

Then we immediately check that

\[
T'' = \bigcup_{k=0}^{n-1} T''_k, \quad \text{and} \quad s_{T''}(f) = \sum_{k=0}^{n-1} s_{T''_k}(f).
\]

See Figure 2.2.

Figure 2.2: An illustration of the refinement \(s_{T''_k}(f)\) utilized in the proof of Proposition 2.1. Note that \(T\) is given by the black dots while \(T'\) is given by the brown dots.
CHAPTER 2. THE RIEMANN INTEGRAL

Since $s_{T_k}^\nu(f)$ is of the form

$$s_{T_k}^\nu(f) = \sum_{i=0}^{r-1} (s_{i+1} - s_i) f(s_i),$$

where $t_k \leq s_i \leq t_{k+1}$ for $i = 0, \ldots, r$, and since $\sum_{i=0}^{r-1} (s_{i+1} - s_i) = s_r - s_0 = t_{k+1} - t_k$, we have

$$|s_{T_k}^\nu(f) - (t_{k+1} - t_k)f(t_k)| = \left| \sum_{i=0}^{r-1} (s_{i+1} - s_i) f(s_i) - (t_{k+1} - t_k)f(t_k) \right|$$

$$= \left| \sum_{i=0}^{r-1} (s_{i+1} - s_i) (f(s_i) - f(t_k)) \right|$$

$$\leq \sum_{i=0}^{r-1} |(s_{i+1} - s_i)||f(s_i) - f(t_k)|$$

$$< \sum_{i=0}^{r-1} |(s_{i+1} - s_i)| \frac{\epsilon}{2(b-a)}$$

$$= (t_{k+1} - t_k) \frac{\epsilon}{2(b-a)}.$$

As a consequence, we obtain

$$|s_T(f) - s_{T_k}^\nu(f)| = \left| \sum_{k=0}^{n-1} (t_{k+1} - t_k)f(t_k) - \sum_{k=0}^{n-1} s_{T_k}^\nu(f) \right|$$

$$\leq \sum_{k=0}^{n-1} |s_{T_k}^\nu(f) - (t_{k+1} - t_k)f(t_k)|$$

$$< \sum_{k=0}^{n-1} (t_{k+1} - t_k) \frac{\epsilon}{2(b-a)}$$

$$\leq \frac{\epsilon}{2},$$

that is,

$$|s_T(f) - s_{T_k}^\nu(f)| < \frac{\epsilon}{2}.$$

By a similar argument applied to $T'$, we obtain

$$|s_{T'}(f) - s_{T_k}^\nu(f)| < \frac{\epsilon}{2}.$$

But then we obtain

$$|s_T(f) - s_{T'}(f)| = |s_T(f) - s_{T_k}^\nu(f) + s_{T_k}^\nu(f) - s_{T'}(f)|$$

$$\leq |s_T(f) - s_{T_k}^\nu(f)| + |s_{T_k}^\nu(f) - s_{T'}(f)|$$

$$< \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$
Remark: It is easy to check that the proof of Proposition 2.1 is still valid if we use more general Cauchy-Riemann sums. Namely, given a subdivision $T = \{t_0, t_1, \ldots, t_n\}$ of $[a, b]$, and any choice of reals $\theta_1, \ldots, \theta_n$ such that $t_k \leq \theta_{k+1} \leq t_{k+1}$ for $k = 0, \ldots, n-1$, define $s_T(f)$ as

$$s_T(f) = \sum_{k=0}^{n-1} (t_{k+1} - t_k) f(\theta_k).$$

Proposition 2.1 implies the following result, which establishes the existence of the Riemann integral of a continuous function defined on a closed and bounded (compact) interval $[a, b]$. 

Theorem 2.2. Let $f : [a, b] \to \mathbb{R}$ be a continuous function defined on a closed and bounded (compact) interval $[a, b]$. For every sequence $T = (T_m)_{m \geq 1}$ of subdivisions of $[a, b]$ such that $\lim_{m \to \infty} \delta(T_m) = 0$, the sequence $(s_{T_m}(f))_{m \geq 1}$ is a Cauchy sequence, and thus has a limit $S_T(f)$. For any two sequences $T = (T_m)_{m \geq 1}$ and $T' = (T'_m)_{m \geq 1}$ of subdivisions of $[a, b]$, if $\lim_{m \to \infty} \delta(T_m) = 0$ and $\lim_{m \to \infty} \delta(T'_m) = 0$, then $S_T(f) = S_{T'}(f)$, that is, the limit of the sequence $(s_{T_m}(f))$ is independent of the sequence $T = (T_m)_{m \geq 1}$ such that $\lim_{m \to \infty} \delta(T_m) = 0$.

Proof. Pick any $\epsilon > 0$, an let $\eta > 0$ be some number given by Proposition 2.1, such that for any two subdivisions $T$ and $T'$ of $[a, b]$ such that $\delta(T) < \eta$ and $\delta(T') < \eta$, we have

$$|s_T(f) - s_{T'}(f)| < \epsilon.$$ 

Since $\lim_{m \to \infty} \delta(T_m) = 0$, there is some $N > 0$ such that for all $m, n \geq N$, we have $\delta(T_m) < \eta$ and $\delta(T_n) < \eta$, which by the definition of $\eta$, implies that

$$|s_{T_m}(f) - s_{T_n}(f)| < \epsilon \quad \text{for all } m, n \geq N.$$ 

Therefore, $(s_{T_m}(f)))$ is a Cauchy sequence. Since $\mathbb{R}$ is a complete metric space, this sequence has a limit $S_T(f)$.

Since by hypothesis $\lim_{m \to \infty} \delta(T_m) = 0$ and $\lim_{m \to \infty} \delta(T'_m) = 0$, there is some $N > 0$ such that for all $m \geq N$, we have $\delta(T_m) < \eta$ and $\delta(T'_m) < \eta$, so by Proposition 2.1,

$$|s_{T_m}(f) - s_{T'_m}(f)| < \epsilon \quad \text{for all } m \geq N,$$

which shows that the Cauchy sequences $(s_{T_m}(f))$ and $(s_{T'_m}(f))$ have the same limit. 

Theorem 2.2 also holds for the more general Cauchy-Riemann sums defined in the Remark after Proposition 2.1.

Theorem 2.2 justifies the following definition.
Definition 2.2. Let \( f : [a, b] \to \mathbb{R} \) be a continuous function defined on a closed and bounded (compact) interval \([a, b]\). The common limit \( S_T(f) \) of the Cauchy sequences \((s_{T_m}(f))_{m \geq 1}\), for all sequences \( T = (T_m)_{m \geq 1} \) of subdivisions of \([a, b]\) such that \( \lim_{m \to \infty} \delta(T_m) = 0 \), is called the Riemann integral of \( f \), and is denoted by \( \int_a^b f(t) \, dt \).

The following are basic properties of the Riemann integral, which are easy to prove (using suitable subdivisions of \([a, b]\)):

1. The mapping \( f \mapsto \int_a^b f(t) \, dt \) is a linear form on the space of continuous functions on \([a, b]\). This means that for any two continuous functions \( f, g : [a, b] \to \mathbb{R} \) and any scalar \( \lambda \in \mathbb{R} \),

\[
\int_a^b (f + g)(t) \, dt = \int_a^b f(t) \, dt + \int_a^b g(t) \, dt,
\]

\[
\int_a^b (\lambda f)(t) \, dt = \lambda \int_a^b f(t) \, dt,
\]

where, as usual, \( f + g \) is the function given by \((f + g)(t) = f(t) + g(t)\), and \( \lambda f \) is the function given by \((\lambda f)(t) = \lambda f(t)\), for all \( t \in [a, b] \). Furthermore, it is a positive linear form, which means that if \( f \geq 0 \), then \( \int_a^b f(t) \, dt \geq 0 \). These seemingly innocuous properties turn out to be very important. Indeed, we will see later how the notion of integral on a locally compact space can be defined in terms of such linear forms (Radon functionals).

2. \[
\left| \int_a^b f(t) \, dt \right| \leq \int_a^b |f(t)| \, dt \leq (b - a) \max_{t \in [a, b]} |f(t)|.
\]

3. If \( f \geq 0 \) and \( f(t) > 0 \) for some \( t \in [a, b] \), then \( \int_a^b f(t) \, dt > 0 \).

4. If \( a < b < c \), then

\[
\int_a^b f(t) \, dt + \int_b^c f(t) \, dt = \int_a^c f(t) \, dt.
\]

5. If \( H : [a, b] \to \mathbb{R} \) is the function given by

\[
H(x) = \int_a^x f(t) \, dt,
\]

then \( H \) is differentiable on \([a, b]\) and \( H'(x) = f(x) \) (the so-called first fundamental theorem of calculus).

The process that we just described only requires that the codomain be complete and that a continuous function \( f : [a, b] \to F \) be uniformly continuous. We also need the linear
combinations $\sum_{k=0}^{n-1}(t_{k+1} - t_k)f(t_k)$ to make sense, so $F$ should be a vector space. If we assume that $F$ is a complete normed vector space (a Banach space), then the Riemann integral of a vector-valued function $f : [a, b] \to F$ can be defined by using the method that we just presented.

In the next section, we show how to define the integral of function with discontinuities, provided that these discontinuities are “reasonable.” For this, a new crucial idea is needed: to define the integral on a class of simple functions with a finite number of reasonable discontinuities, and then to extend the integral to a bigger class of functions by taking limits of simple functions. For this process to work, the bigger space of functions should be complete.

### 2.2 The Riemann Integral of Regulated Functions

In this section we show how to define the integral of regulated functions $f : [a, b] \to F$, where $F$ is any complete normed vector space, in particular $\mathbb{R}$ or any finite-dimensional vector space (real or complex).

The first key ingredient is that the method of Cauchy-Riemann sums can be immediately adapted to define the notion of integral for a step function. The mapping $\int : \text{Step}([a, b]; F) \to F$ is easily seen to be linear and continuous.

The second key ingredient is that, by Proposition 1.18, the vector space $\text{Step}([a, b]; F)$ of steps functions over $[a, b]$ is dense in $\text{Reg}([a, b]; F)$, the space of regulated functions over $[a, b]$, and $\text{Reg}([a, b]; F)$ is complete, where $[a, b]$ is a closed and bounded interval.

Then, because $\text{Step}([a, b]; F)$ is dense in $\text{Reg}([a, b]; F)$, and $\text{Reg}([a, b]; F)$ is complete, by Theorem 15.69, the continuous linear map $\int : \text{Step}([a, b]; F) \to F$ has a unique extension $\int : \text{Reg}([a, b]; F) \to \text{Reg}([a, b]; F)$, which is also continuous and linear. This is how the integral of a regulated function is defined.

Thus it remains to define the integral of a step function.

#### Definition 2.3

Let $f : [a, b] \to F$ be a step function. For any admissible subdivision $T = (a_0, a_1, \ldots, a_n)$ for $f$, for any sequence $\xi = (\xi_1, \ldots, \xi_n)$ or reals such that $\xi_{k+1} \in (a_k, a_{k+1})$ for $k = 0, \ldots, n - 1$, define $s_T\xi(f)$ by

$$s_T\xi(f) = \sum_{k=0}^{n-1}(a_{k+1} - a_k)f(\xi_{k+1}).$$

See Figure 2.3.

The above is a linear combination of vectors in $F$, and since $F$ is a vector space, it is well defined. Note that because $\xi_{k+1} \in (a_k, a_{k+1})$, $s_T\xi(f)$ does not depend on the value of $f$ at the $a_k$. For simplicity of language, we refer to a pair $(T, \xi)$ as in Definition 2.3 as an admissible pair for $f$. 
CHAPTER 2. THE RIEMANN INTEGRAL

Figure 2.3: An illustration of \( s_{T,\xi}(f) = \sum_{k=0}^{3}(a_{k+1} - a_k)f(\xi_{k+1}) \) for step function \( f: [a, b] \to \mathbb{R} \).

The problem with the above definition of \( s_{T,\xi}(f) \) is that it depends on the admissible subdivision \( T \), and on \( \xi \), but because \( f \) is a step function, it is constant on each interval \((a_k, a_{k+1})\), so in fact \( s_{T,\xi}(f) \) is independent of the admissible pair \((T, \xi)\).

**Proposition 2.3.** Given a step function \( f: [a, b] \to F \), for any two admissible pairs \((T, \xi)\) and \((T', \xi')\) for \( f \), we have \( s_{T,\xi}(f) = s_{T',\xi'}(f) \).

Proposition 2.3 is proved by using an admissible pair which is finer than both \((T, \xi)\) and \((T', \xi')\). The details are left to the reader, or see Schwartz [72] (Chapter V, Section §1).

**Definition 2.4.** Let \( f: [a, b] \to F \) be a step function. The integral of \( f \), denoted \( \int_{[a,b]} f \), is the common value of the sum \( s_{T,\xi}(f) \), for any any admissible pair \((T, \xi)\) for \( f \).

The following proposition follows almost immediately from the definitions.

**Proposition 2.4.** The map \( \int: \text{Step}([a,b]; F) \to F \), where \( \int f = \int_{[a,b]} f \) is the integral defined in Definition 2.4, is linear. Furthermore, we have
\[
\| \int_{[a,b]} f \| \leq \int_{[a,b]} \| f \| \quad \text{and} \quad \| \int_{[a,b]} f \| \leq (b-a) \| f \|_{\infty},
\]
where \( \| f \| \) means the real-valued function \( x \mapsto \| f(x) \| \). If \( f = \mathbb{R} \) and if \( f \geq 0 \), then \( \int_{[a,b]} f \geq 0 \).

Proposition 2.4 shows that the map \( \int: \text{Step}([a,b]; F) \to F \) is linear and continuous. As we explained earlier, by Theorem 15.69, the map \( \int: \text{Step}([a,b]; F) \to F \) has a unique extension \( \int: \text{Reg}([a,b]; F) \to F \) to \( \text{Reg}([a,b]; F) \), which is also linear and continuous.
Definition 2.5. The integral \( \int_{[a,b]} f \) of any regulated function \( f \in \text{Reg}([a,b]; F) \) is equal to \( f \), where \( f : \text{Reg}([a,b]; F) \to F \) is the unique linear and continuous extension of the linear and continuous map \( f : \text{Step}([a,b]; F) \to F \). This integral is called the Riemann integral of the regulated function \( f \).

Definition 2.5 is not very constructive. It turns out that the Riemann integral of a regulated function can be defined more directly in terms of generalized Riemann sums. This approach is presented in Schwartz [72] (Chapter V, Section §1).

Note that we actually haven’t defined the notion of Riemann-integrable function. What we did is to exhibit a family of functions, the regulated functions, which are Riemann-integrable. The notion of Riemann-integrable function is defined in various books, including Schwartz [72]. This can be done using the notion of upper integral \( \int^* f \), which is defined for a positive function \( f \in \mathcal{K}(\mathbb{R}, F) \) as the infimum of the integrals of the step functions that bound \( f \) from above.

The space of Riemann-integrable functions contains other functions besides the regulated functions. For example, functions with compact support which are continuous except at finitely many points, are Riemann-integrable. The function \( x \mapsto \sin(1/x) \) is such a function (with value 0 at \( x = 0 \)). It is Riemann-integrable on \([0, 1]\), even though 0 is not a discontinuity of the first kind.

The method of this section, which consists in defining the notion of integral for a “big” set of functions, such as \( \text{Reg}([a,b]; F) \), by first defining a notion of integral on a very simple set of functions for which the definition is obvious, such as \( \text{Step}([a,b]; F) \), and then to extend the integral on \( \text{Step}([a,b]; F) \) to a notion of integral on \( \text{Reg}([a,b]; F) \) using a completion process, is a key idea. In this situation, we are lucky that \( \text{Reg}([a,b]; F) \) is complete.

In order to define a notion of integral for functions defined on a domain \( X \) which is more general than a compact interval \([a,b]\) of \( \mathbb{R} \), we can proceed as above, but some additional structure on \( X \) is needed to define step functions and the notion of integral of step functions. This new ingredient is the notion of measure. The other technical difficulty is that the completion of the space of generalized step functions is not a space identifiable with a space of familiar functions. By Theorem 15.68, the completion always exists, but its elements are equivalence classes of functions, so it will take some work to exhibit this space as a set of functions.
Chapter 3

Measure Theory; Basic Notions

Let $X$ be a nonempty set. Intuitively, a measure on $X$ is a function $\mu$ that assigns a nonnegative real number $\mu(A)$ to every subset $A$ in some specified nonempty collection $\mathcal{A}$ of subsets of $X$, where $\mu(A)$ is a generalization of the notion of length, area, or volume. For example, any natural measure $\mu$ on $\mathbb{R}$ should have the property that $\mu([a, b]) = b - a$ for all $a \leq b$. It is natural to require that if a subset $A$ is sliced into countably many pairwise disjoint small pieces $A_i$, then $\mu(A) = \mu(\bigcup_{n=1}^{\infty} A_i) = \sum_{n=1}^{\infty} \mu(A_i)$. This property is called $\sigma$-additivity. Then the family $\mathcal{A}$ of subsets on which $\mu$ is defined should be closed under countable unions. It is also natural to require $\mathcal{A}$ to be closed under complementation. This leads to the important notion of a $\sigma$-algebra, which is closed under complementation and countable unions. The weaker notion in which only closure under complementation and closure under finite union is that of an algebra. In general, it is not easy to construct nontrivial $\sigma$-algebras, so it is useful to have tools to do so. A pair $(X, \mathcal{A})$ consisting of a nonempty set and a $\sigma$-algebra $\mathcal{A}$, is called a measurable space.

Given any nonempty family $S$ of subsets of $X$, there is a smallest $\sigma$-algebra $\mathcal{A}(S)$ containing $S$. If $X$ is a topological space, then the $\sigma$-algebra $\mathcal{B}(X)$ containing the open subsets of $X$ is an important $\sigma$-algebra called the Borel $\sigma$-algebra.

The notion of monotone class is also useful to construct $\sigma$-algebras. Given any nonempty family $S$ of subsets of $X$, there is a smallest monotone class $\mathcal{M}(S)$ containing $S$. Given an algebra $\mathcal{B}$, the smallest $\sigma$-algebra $\mathcal{A}(\mathcal{B})$ containing $\mathcal{B}$ and the smallest monotone class $\mathcal{M}(\mathcal{B})$ containing $\mathcal{B}$ are identical: $\mathcal{A}(\mathcal{B}) = \mathcal{M}(\mathcal{B})$.

Next we define (positive) measures on a $\sigma$-algebra. A triple $(X, \mathcal{A}, \mu)$ consisting of a nonempty set, a $\sigma$-algebra $\mathcal{A}$, and a measure $\mu$ on $\mathcal{A}$, is called a measure space. We investigate a few properties of measures. In particular, we show that every measure can be extended to a complete measure, which means that all $A \in \mathcal{A}$, if $\mu(A) = 0$, then $B \in \mathcal{A}$ for all $B \subseteq A$.

As we said earlier, it is not easy to construct nontrivial measures. A very useful concept to achieve this is the notion of outer measure, introduced in Section 3.3. Outer measures are defined for all subsets of $X$, which makes them much easier to construct. In particular, we construct the Lebesgue outer measure $\mu^*_L$. 

49
A fundamental theorem due to Carathéodory shows that every outer measure induces a measure space; see Theorem 3.11.

By applying Theorem 3.11 to the outer measure \( \mu^*_L \) we obtain the \( \sigma \)-algebra \( L(\mathbb{R}) \) of Lebesgue measurable sets and the Lebesgue measure \( \mu_L \); see Section 3.4. The Borel \( \sigma \)-algebra \( B(\mathbb{R}) \) is properly contained in the \( \sigma \)-algebra \( L(\mathbb{R}) \) of Lebesgue measurable sets, and there are subsets of \( \mathbb{R} \) that are not Lebesgue measurable sets (assuming the axiom of choice). We also discuss various regularity properties of the Lebesgue measure.

### 3.1 \( \sigma \)-Algebras, Measures

Let \( X \) be a nonempty set. We would like to define the notion of “measure” for the subsets of \( X \), in such a way that familiar properties of the notion of length, area, or volume, of polyhedral objects in \( \mathbb{R} \), \( \mathbb{R}^2 \) or \( \mathbb{R}^3 \), hold. The measure \( m(A) \) of a subset of \( X \) should be nonnegative, but we have to allow “big” objects to have infinite measure so it is desirable to extend the nonnegative real numbers by adding a new element corresponding to infinity.

Technically, we define \( \mathbb{R}_+ \) as the union

\[
\mathbb{R}_+ = \{ \alpha \in \mathbb{R} \mid \alpha \geq 0 \} \cup \{ +\infty \} = \mathbb{R}_+ \cup \{ +\infty \},
\]

where \(+\infty\) is not in \( \mathbb{R}_+ \), and we assume that the following properties hold:

(a) \( \alpha < +\infty \), for all \( \alpha \in \mathbb{R}_+ \),

(b) \( \alpha + (+\infty) = (+\infty) + \alpha = +\infty \), for all \( \alpha \in \mathbb{R}_+ \),

(c) \( \alpha \cdot (+\infty) = (+\infty) \cdot \alpha = +\infty \), for all \( \alpha \in \mathbb{R}_+ - \{0\} \),

(d) \( 0 \cdot (+\infty) = (+\infty) \cdot 0 = 0 \),

(e) If \( (\alpha_i)_{i \geq 1} \) is a sequence with \( \alpha_i \in \mathbb{R}_+ \), and if \( \alpha_i = +\infty \) for some \( i \), then \( \sum_{i=1}^{\infty} \alpha_i = +\infty \).

The set \( \mathbb{R}_+ \) is also denoted by \([0, +\infty] \).

In this chapter, we closely follow Halmos [39]. Other nice (but concise) presentations can be found in Rudin [66] and Lang [52]. A very detailed presentation is given in Schwartz [72].

An “ideal measure” should be a function \( m \) satisfying the following properties:

1. \( m : 2^X \rightarrow [0, +\infty] \), that is, \( m \) is defined for all subsets of \( X \).

2. \( m(\emptyset) = 0 \).

3. For any countable sequence \( (A_i)_{i \geq 1} \) of subsets \( A_i \) of \( X \) such that \( A_i \cap A_j = \emptyset \) for all \( i \neq j \),

\[
m \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} m(A_i).\]

This property is called \( \sigma \)-additivity.
The intuition behind \( \sigma \)-additivity is that if we slice an object \( A \) into countably many pairwise disjoint small pieces \( A_i \), then the measure \( m(A) \) of \( A \) should be the sum of the measures \( m(A_i) \) of the pieces \( A_i \).

Observe that by choosing a sequence \( (A_i)_{i \geq 1} \) such that \( A_j = \emptyset \) for all \( j > n \), and \( A_i \cap A_j = \emptyset \) if \( i \neq j \), we obtain the property

\[
m\left( \bigcup_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} m(A_i),
\]

known as \textit{additivity}.

For any two subsets \( A \) and \( B \), if \( A \subseteq B \), we can write \( B = A \cup (B - A) \), with \( A \cap (B - A) = \emptyset \), so by additivity,

\[
m(B) = m(A) + m(B - A),
\]

and since \( m(B - A) \geq 0 \), we obtain

\[
m(A) \leq m(B).
\]

We claim that the following property holds.

\textbf{Proposition 3.1.} If a function \( m \) satisfies Properties (1–3) above, then for any countable sequence \( (A_i)_{i \geq 1} \) of (not necessarily pairwise disjoint) subsets \( A_i \) of \( X \),

\[
m\left( \bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} m(A_i).
\]

\textit{Proof.} Define the sequence \( (B_i) \) of subsets of \( X \) as follows: \( B_1 = A_1, B_2 = A_2 - A_1, \ldots, B_i = A_i - \left( \bigcup_{j=1}^{i-1} A_j \right) \), for all \( i \geq 2 \). See Figure 3.1.

\begin{figure}[h]
\centering
\includegraphics[width=0.7\textwidth]{set_construction.png}
\caption{A schematic illustration of the set construction \((B_i)\).}
\end{figure}
It is easy to check that \( \bigcup_{i=1}^{\infty} B_i = \bigcup_{i=1}^{\infty} A_i, B_i \cap B_j = \emptyset \) for all \( i \neq j \), and \( m(B_i) \leq m(A_i) \) for all \( i \geq 1 \), so by \( \sigma \)-additivity,

\[
m\left( \bigcup_{i=1}^{\infty} A_i \right) = m\left( \bigcup_{i=1}^{\infty} B_i \right) = \sum_{i=1}^{\infty} m(B_i) \leq \sum_{i=1}^{\infty} m(A_i),
\]

as claimed. \( \square \)

In general, for an arbitrary set \( X \), there may be no function \( m \) satisfying Properties (1–3) for all subsets of \( X \), as well as certain desirable properties. For example, there is no such translation invariant function on \( 2^\mathbb{R} \) such that \( m([0,1)) \neq 0 \) or \( m([0,1)) \neq +\infty \), and no such function on \( 2^\mathbb{R} \) such that \( m([a,b]) = b - a \) for every interval \([a,b] \); see Section 3.4.

Thus we are led to relax some of these conditions. There are two options:

1. The first option is to relax (3) by replacing it by the result of Proposition 3.1, namely

   (3’) For any countable sequence \((A_i)_{i \geq 1}\) of subsets \( A_i \) of \( X \),

   \[
m\left( \bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} m(A_i).
\]

   This approach leads to outer measures, and is discussed in Section 3.3.

2. Condition (3) is highly desirable, so the second option is to restrict the domain of \( m \) to be a proper family of subsets of \( X \); the right notion is that of a \( \sigma \)-algebra.

The notion of a \( \sigma \)-algebra is more important that the notion of outer measure, which is needed for technical reasons. Thus we now consider Option 2, and define \( \sigma \)-algebras. Once the notion of \( \sigma \)-algebra is defined, we will be able to define the abstract notion of a measure (see Definition 3.9), which is the crucial ingredient in defining a general notion of integral.

**Definition 3.1.** Let \( X \) be any nonempty set. A family \( \mathcal{A} \) of subsets of \( X \) is a \( \sigma \)-algebra if it satisfies the following conditions:

(A1) \( X \in \mathcal{A} \).

(A2) For every subset \( A \) of \( X \), if \( A \in \mathcal{A} \), then \( X - A \in \mathcal{A} \) (closure under complementation).

(\( \sigma \)-A3) For every countable family \((A_i)_{i \geq 1}\) of subsets of \( X \), if \( A_i \in \mathcal{A} \) for all \( i \geq 1 \), then \( \bigcup_{i=1}^{\infty} A_i \in \mathcal{A} \) (closure under countable unions).

From (A1) and (A2), we see that \( \emptyset \in \mathcal{A} \). From (A2) and (\( \sigma \)-A3) and the fact that \( A = X - (X - A) \) and \( \bigcap_{i=1}^{\infty} A_i = X - (\bigcup_{i=1}^{\infty} (X - A_i)) \), if \( A_i \in \mathcal{A} \) for all \( i \geq 1 \), then \( \bigcap_{i=1}^{\infty} A_i \in \mathcal{A} \) (closure under countable intersections). In particular, if we let \( A_i = \emptyset \) for all \( i \geq 3 \), we see that if \( A_1 \in \mathcal{A} \) and \( A_2 \in \mathcal{A} \), then \( A_1 \cup A_2 \in \mathcal{A} \) and \( A_1 \cap A_2 \in \mathcal{A} \). Since \( A_1 - A_2 = A_1 \cap (X - A_2) \), we also have \( A_1 - A_2 \in \mathcal{A} \).

Axiom (\( \sigma \)-A3) is a strong condition, and this is the reason why it is not easy to construct nontrivial \( \sigma \)-algebras. There are two extreme \( \sigma \)-algebras:
1. \( \mathcal{A} = \{ \emptyset, X \} \).

2. \( \mathcal{A} = 2^X \), the family of all subsets of \( X \).

Interesting \( \sigma \)-algebra lie in-between.

**Remarks:**

1. Some authors use the term \( \sigma \)-field instead of \( \sigma \)-algebra. This is a rather unfortunate terminology, because in algebra, a field is a set with two operations that have identity elements. Here the operations are union and intersection, but there is no identity element for intersection.

2. If we weaken Condition \( \sigma \)-A3 to finite unions, then we obtain a structure called an algebra (or boolean algebra of sets).

**Definition 3.2.** Let \( X \) be any nonempty set. A family \( \mathcal{A} \) of subsets of \( X \) is an algebra (or boolean algebra of sets) if it satisfies the following conditions:

(A1) \( X \in \mathcal{A} \).

(A2) For every subset \( A \) of \( X \), if \( A \in \mathcal{A} \), then \( X - A \in \mathcal{A} \) (closure under complementation).

(A3) For every finite family \( (A_i)_{i=1}^n \) of subsets of \( X \), if \( A_i \in \mathcal{A} \) for all \( i = 1, \ldots, n \), then \( \bigcup_{i=1}^n A_i \in \mathcal{A} \) (closure under finite unions).

As in the case of \( \sigma \)-algebras, an algebra contains \( \emptyset \) and is closed under (finite) union and intersection. In the construction of a product of measurable spaces, another notion of algebra will come up. These are the semi-algebras.

**Definition 3.3.** Let \( X \) be any nonempty set. A family \( \mathcal{S} \) of subsets of \( X \) is a semi-algebra if it satisfies the following conditions:

(S1) \( X, \emptyset \in \mathcal{S} \).

(S2) For all \( A, B \in \mathcal{S} \), we have \( A \cap B \in \mathcal{S} \).

(S3) For all \( A \in \mathcal{S} \), we have \( X - A = X_1 \cup \cdots \cup X_n \), for finitely many pairwise disjoint subsets \( X_i \in \mathcal{S} \).

**Example 3.1.** First consider the family of intervals of \( \mathbb{R} \) of the form \( [a, b) \), with \( a \leq b \), where \( a = -\infty \) or \( b = \infty \) is allowed. By convention, let \( [a, b) = \emptyset \) if \( a > b \). This is a semi-algebra, because

\[
[a_1, b_1) \cap [a_2, b_2) = [\max(a_1, a_2), \min(b_1, b_2)),
\]

and

\[
X - [a, b) = (-\infty, a) \cup [b, \infty).
\]
CHAPTER 3. MEASURE THEORY; BASIC NOTIONS

Next, let $X$ and $Y$ be two nonempty sets, and let $\mathcal{A}$ be an algebra on $X$ and let $\mathcal{B}$ be an algebra on $Y$. Define the set $\mathcal{R}$ of rectangles in $X \times Y$ as follows:

$$\mathcal{R} = \{A \times B \in X \times Y \mid A \in \mathcal{A}, B \in \mathcal{B}\}.$$  

It is easy to check that $\mathcal{R}$ is a semi-algebra. For example,

$$(A_1 \times B_1) \cap (A_2 \times B_2) = (A_1 \cap A_2) \times (B_1 \cap B_2),$$

and

$$(X \times Y) - (A \times B) = ((X - A) \times (Y - B)) \cup ((X - A) \times B) \cup (A \times (Y - B)).$$

Then it can be shown that the set $\mathcal{B}(\mathcal{R})$ of finite unions of pairwise disjoint sets in $\mathcal{R}$ is the smallest algebra containing the semi-algebra $\mathcal{R}$. This algebra will be used to construct the product of measurable spaces.

The following result can be shown.

**Proposition 3.2.** Given a semi-algebra $\mathcal{S}$, the smallest algebra $\mathcal{B}(\mathcal{S})$ containing $\mathcal{S}$ is the family of finite unions of pairwise disjoint subsets in $\mathcal{S}$.

**Definition 3.4.** Let $X$ be any nonempty set. A pair $(X, \mathcal{A})$ where $\mathcal{A}$ is a $\sigma$-algebra of subsets of $X$ is called a measurable space. The subsets of $X$ that belong to $\mathcal{A}$ are called the measurable subsets of $X$.

**Proposition 3.3.** Let $X$ be any nonempty set, and let $\mathcal{S}$ be any nonempty family of subsets of $X$. Then there is a $\sigma$-algebra $\mathcal{A}(\mathcal{S})$ with the following properties:

(a) $\mathcal{S} \subseteq \mathcal{A}(\mathcal{S})$.

(b) If $\mathcal{A}'$ is any $\sigma$-algebra such that $\mathcal{S} \subseteq \mathcal{A}'$, then $\mathcal{A}(\mathcal{S}) \subseteq \mathcal{A}'$.

This means that $\mathcal{A}(\mathcal{S})$ is the smallest $\sigma$-algebra containing $\mathcal{S}$.

**Definition 3.5.** Let $X$ be any nonempty set, and let $\mathcal{S}$ be any nonempty family of subsets of $X$. The smallest $\sigma$-algebra $\mathcal{A}(\mathcal{S})$ containing $\mathcal{S}$ is called the $\sigma$-algebra generated by $\mathcal{S}$.

The $\sigma$-algebra $\mathcal{A}(\mathcal{S})$ is the intersection of the family of all $\sigma$-algebras containing $\mathcal{S}$. This family is nonempty since $2^X$ belongs to it. This way of defining $\mathcal{A}(\mathcal{S})$ is highly nonconstructive. A bottom-up construction of $\mathcal{A}(\mathcal{S})$ can be performed, but to guarantee closure under infinite unions, transfinite induction is required; see Schwartz [72] (Chapter V, Section §2).

An important example arises when $X$ is a topological space $(X, \mathcal{O})$.

**Definition 3.6.** Let $(X, \mathcal{O})$ be a topological space. The $\sigma$-algebra $\mathcal{B}(X)$ generated by the family $\mathcal{O}$ of open sets is called the Borel $\sigma$-algebra of $X$. The subsets in $\mathcal{B}(X)$ are called Borel sets.
All open subsets and all closed sets are Borel sets. Countably infinite unions of closed sets and countable infinite intersections of open sets are Borel sets. But there are many more Borel sets.

Another way to construct σ-algebras is to use algebras and monotone classes.

**Definition 3.7.** Let \( X \) be any nonempty set. A nonempty family \( \mathcal{M} \) of subsets of \( X \) is a **monotone class** if for every countable family \( (A_i)_{i \geq 1} \) of subsets of \( X \), if \( A_i \in \mathcal{M} \) for all \( i \geq 1 \) then:

1. If \( A_i \subseteq A_{i+1} \) for all \( i \geq 1 \), then \( \bigcup_{i=1}^{\infty} A_i \in \mathcal{M} \). See Figure 3.2, Diagram (i).
2. If \( A_{i+1} \subseteq A_i \) for all \( i \geq 1 \), then \( \bigcap_{i=1}^{\infty} A_i \in \mathcal{M} \). See Figure 3.2, Diagram (ii).

![Figure 3.2](image_url): The rose colored sets of Figure (i) satisfy the increasing nesting condition of \( A_i \subseteq A_{i+1} \), while the periwinkle sets of Figure (ii) satisfy the decreasing nesting condition \( A_{i+1} \subseteq A_i \).

**Proposition 3.4.** Let \( X \) be any nonempty set. For any algebra \( \mathcal{B} \), if \( \mathcal{B} \) is a monotone class, then \( \mathcal{B} \) is a σ-algebra.

**Proof.** Let \( (A_i)_{i \geq 1} \) be a countable family of subsets of \( X \), such that \( A_i \in \mathcal{B} \) for all \( i \geq 1 \). Since \( \mathcal{B} \) is an algebra, it is closed under finite unions, so \( B_n = \bigcup_{i=1}^{n} A_i \in \mathcal{B} \), and obviously \( B_n \subseteq B_{n+1} \) for all \( n \geq 1 \), and \( \bigcup_{i=1}^{\infty} A_i = \bigcup_{n=1}^{\infty} B_n \). Since \( \mathcal{B} \) is a monotone class, \( \bigcup_{i=1}^{\infty} A_i = \bigcup_{n=1}^{\infty} B_n \in \mathcal{B} \).

Here is a version of Proposition 3.3 for monotone classes.

**Proposition 3.5.** Let \( X \) be any nonempty set, and let \( \mathcal{S} \) be any nonempty family of subsets of \( X \). Then there is a monotone class \( \mathcal{M}(\mathcal{S}) \) with the following properties:

(a) \( \mathcal{S} \subseteq \mathcal{M}(\mathcal{S}) \).
(b) If $\mathcal{M}$ is any monotone class such that $\mathcal{S} \subseteq \mathcal{M}$, then $\mathcal{M}(\mathcal{S}) \subseteq \mathcal{M}$.

This means that $\mathcal{M}(\mathcal{S})$ is the smallest monotone class containing $\mathcal{S}$.

**Definition 3.8.** Let $X$ be any nonempty set, and let $\mathcal{S}$ be any nonempty family of subsets of $X$. The smallest monotone class $\mathcal{M}(\mathcal{S})$ containing $\mathcal{S}$ is called the *monotone class generated by $\mathcal{S}$*.

The following theorem yields another way of generating a $\sigma$-algebra from an algebra.

**Theorem 3.6.** Let $X$ be any nonempty set. For any algebra $\mathcal{B}$, the $\sigma$-algebra $\mathcal{A}(\mathcal{B})$ generated by $\mathcal{B}$ and the monotone class $\mathcal{M}(\mathcal{B})$ generated by $\mathcal{B}$ are identical; that is,

$$\mathcal{A}(\mathcal{B}) = \mathcal{M}(\mathcal{B}).$$

We now come to the very important notion of measure.

**Definition 3.9.** Let $X$ be any nonempty set. A *measure* on $X$ is a map $\mu$ satisfying the following properties:

1. $(\mu_1)$ $\mu: \mathcal{A} \to [0, +\infty]$, where $\mathcal{A}$ is a $\sigma$-algebra of subsets of $X$.
2. $(\mu_2)$ $\mu(\emptyset) = 0$.
3. $(\mu_3)$ For any countable sequence $(A_i)_{i \geq 1}$ of subsets $A_i$ of $\mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$,

$$\mu \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu(A_i).$$

This property is called *$\sigma$-additivity*.

A *measure space* is a triple $(X, \mathcal{A}, \mu)$, where $(X, \mathcal{A})$ is a measurable space and $\mu$ is a measure on $X$.

**Remarks:**

1. The degenerate situation where $\mu(A) = +\infty$ for all nonempty subsets in $\mathcal{A}$ is allowed. If $\mu$ is nontrivial, which means that $\mathcal{A}$ possesses some nonempty subset $A$ such that $\mu(A)$ is finite, then by letting $A_1 = A$ and $A_i = \emptyset$ for all $i \geq 2$, by $(\mu_3)$ we get $\mu(A) = \mu(A) + \sum_{i=2}^{\infty} \mu(\emptyset)$, which implies $\mu(\emptyset) = 0$. In this situation Axiom $(\mu_2)$ is unnecessary. Rudin makes the assumption that a measure is nontrivial; see [66].

2. A measure $\mu$ is also called a *positive measure*, to stress that its range is nonnegative. There are more general measures taking their values in $\mathbb{R}$ or $\mathbb{C}$, even a in Banach space. For such measures, Condition $(\mu_3)$ needs to be slightly strengthened.
3. Some authors used the term measured space instead of measure space!

**Definition 3.10.** Let \((X, \mathcal{A}, \mu)\) be a measure space. The measure \(\mu\) is finite if \(\mu(X)\) is finite. If \(\mu: \mathcal{A} \to [0, 1]\) and if \(\mu(X) = 1\), then \((X, \mathcal{A}, \mu)\) is called a probability space. The measure \(\mu\) is a \(\sigma\)-finite if there exist a countable family \((A_i)_{i \geq 1}\) of subsets \(A_i \in \mathcal{A}\) such that \(X = \bigcup_{i=1}^{\infty} A_i\), and \(\mu(A_i)\) is finite for all \(i \geq 1\). The measure \(\mu\) is complete if for all \(A \in \mathcal{A}\), if \(\mu(A) = 0\), then \(B \in \mathcal{A}\) for all \(B \subseteq A\). A subset \(A \in \mathcal{A}\) such that \(\mu(A) = 0\) is called a set of measure zero.

**Example 3.2.** Let \(X\) be any nonempty set, and consider the \(\sigma\)-algebra \(\mathcal{A} = 2^X\). The map \(\mu: 2^X \to [0, +\infty]\) given by

\[
\mu(A) = \begin{cases} |A| & \text{if } A \text{ is finite} \\ +\infty & \text{if } A \text{ is infinite} \end{cases}
\]

is a measure called the counting measure on \(X\).

Here is a summary of useful properties of measures.

**Proposition 3.7.** Let \((X, \mathcal{A}, \mu)\) be a measure space. The following properties hold:

1. For any finite sequence \((A_1, \ldots, A_n)\) of subsets \(A_i \in \mathcal{A}\) such that \(A_i \cap A_j = \emptyset\) whenever \(i \neq j\), we have

\[
\mu \left( \bigcup_{i=1}^{n} A_i \right) = \sum_{i=1}^{n} \mu(A_i).
\]

2. For any two subsets \(A, B\) of \(X\), if \(A, B \in \mathcal{A}\) and if \(A \subseteq B\), then \(\mu(A) \leq \mu(B)\).

3. For any countable sequence \((A_i)_{i \geq 1}\) of subsets \(A_i \in \mathcal{A}\),

\[
\mu \left( \bigcup_{i=1}^{\infty} A_i \right) \leq \sum_{i=1}^{\infty} \mu(A_i).
\]

4. For any countable sequence \((A_i)_{i \geq 1}\) of subsets \(A_i \in \mathcal{A}\), if \(A_{i+1} \subseteq A_i\) for all \(i \geq 1\) and if \(\mu(A_1)\) is finite, then

\[
\mu \left( \bigcap_{i=1}^{n} A_i \right) = \lim_{n \to \infty} \mu(A_i).
\]

5. For any countable sequence \((A_i)_{i \geq 1}\) of subsets \(A_i \in \mathcal{A}\), if \(A_i \subseteq A_{i+1}\) for all \(i \geq 1\), then

\[
\mu \left( \bigcup_{i=1}^{n} A_i \right) = \lim_{n \to \infty} \mu(A_i).
\]
CHAPTER 3. MEASURE THEORY; BASIC NOTIONS

Proof. The proof of (1) and (2) is identical to the proof given just before Proposition 3.1, and (3) is Proposition 3.1. We prove (4), leaving the proof of (5) as an exercise.

We can write
\[ A_n = \left( \bigcap_{i=1}^{\infty} A_i \right) \cup \left( \bigcup_{i=n}^{\infty} (A_i - A_{i+1}) \right), \]
a union of pairwise disjoint subsets since \( A_{i+1} \subseteq A_i \) for all \( i \geq 1 \). By (\( \mu_3 \)), we have
\[ \mu \left( \bigcap_{i=1}^{\infty} A_i \right) + \sum_{i=n}^{\infty} \mu(A_i - A_{i+1}) = \mu(A_n) \leq \mu(A_1) < +\infty, \]
since \( A_{i+1} \subseteq A_i \) for all \( i \geq 1 \) and since \( \mu(A_1) \) is assumed to be finite. See Figure 3.3.

![Figure 3.3: Decomposing the decreasing nested sequences of periwinkle sets into disjoint rings. Note \( A_1 = (A_1 - A_2) \cup (A_2 - A_3) \cup \cdots \cup (A_n - A_{n+1}) \cup A_{n+1} \).](image)

Consequently, for \( n = 1 \) we deduce that the series \( \sum_{i=1}^{\infty} \mu(A_i - A_{i+1}) \) converges, which implies that
\[ \lim_{n \to \infty} \sum_{i=n}^{\infty} \mu(A_i - A_{i+1}) = 0. \]
Since
\[ \mu \left( \bigcap_{i=1}^{\infty} A_i \right) + \sum_{i=n}^{\infty} \mu(A_i - A_{i+1}) = \mu(A_n), \]
we conclude that \( \mu \left( \bigcap_{i=1}^{n} A_i \right) = \lim_{n \to \infty} \mu(A_n) \).

The following result shows that every measure can be completed; this is technically useful.

**Proposition 3.8.** Let \((X, \mathcal{A}, \mu)\) be a measure space. A measure space \((X, \overline{\mathcal{A}}, \overline{\mu})\) with the following properties can be constructed:
(a) $\mathcal{A} \subseteq \overline{\mathcal{A}}$.

(b) $\overline{\mu}$ extends $\mu$; that is, $\overline{\mu}(A) = \mu(A)$ for all $A \in \mathcal{A}$.

(c) The measure $\overline{\mu}$ is complete.

We force the completeness property by defining $\overline{\mathcal{A}}$ as follows:

$$\overline{\mathcal{A}} = \{ \overline{A} \subseteq X \mid (\exists A, A' \in \mathcal{A})(\exists B \subseteq A')(\overline{A} = A \cup B, \mu(A') = 0) \}.$$ 

The measure $\overline{\mu}$ is defined such that

$$\overline{\mu}(\overline{A}) = \overline{\mu}(A \cup B) = \mu(A).$$

See Figure 3.4.

Figure 3.4: A schematic illustration of a set in $\overline{\mathcal{A}}$. The magenta set $A$ has positive measure, while the grayish set $A'$, and all of its subsets, including $B$, have zero measure. Then $\overline{A} = A \cup B$.

The verification that $\overline{\mathcal{A}}$ is a $\sigma$-algebra with the required properties and that $\overline{\mu}$ is a measure with the required properties is tedious (among other things, one need to check that $\overline{\mu}(\overline{A})$ does not depend on the representation of $\overline{A}$).

The measure $\overline{\mu}$ given by Proposition 3.8 is called the completed measure of $\mu$.

### 3.2 Null Subsets and Properties Holding Almost Everywhere

One of the secrets of measure theory is that subsets of measure zero should be ignored. Since a measure is not necessarily complete the correct technical definition is as follows.
**Definition 3.11.** Let \((X, \mathcal{A}, \mu)\) be a measure space. A subset \(E \subseteq X\) is null\(^1\) (or negligible) if there is some \(A \in \mathcal{A}\) such that \(E \subseteq A\) and \(\mu(A) = 0\). A property \(P\) of the elements of \(X\) holds almost everywhere, abbreviated holds a.e., if the subset were it fails is null; that is, the set \(\{x \in X \mid P(x) = \text{false}\}\) is null.

To be very precise, we should say \(\mu\)-null and holds \(\mu\)-a.e., since these notions depend on the measure \(\mu\). In most cases, there is no risk of confusion and we drop \(\mu\).

Observe that if the measure \(\mu\) is complete, then a subset \(E \subseteq X\) is null iff \(\mu(E) = 0\), and a property \(P\) holds a.e. iff \(\mu(\{x \in X \mid P(x) = \text{false}\}) = 0\). In general, a null set may either be measurable or nonmeasurable, and a nonmeasurable has no reason to be null, but may be null.

Here are a few properties of null sets.

**Proposition 3.9.** Let \((X, \mathcal{A}, \mu)\) be a measure space. Every subset of a null set is null. Every countable union of null sets is a null set.

**Proof.** The first property follows immediately from the definition. Let \((A_i)_{i \geq 1}\) be a countable family of null sets. There are subsets \(B_i \in \mathcal{A}\) such that \(A_i \subseteq B_i\) and \(\mu(B_i) = 0\) for all \(i \geq 1\). We have
\[
\bigcup_{i=1}^{\infty} A_i \subseteq \bigcup_{i=1}^{\infty} B_i \in \mathcal{A}
\]
because \(\mathcal{A}\) is a \(\sigma\)-algebra, so it remains to show that \(\bigcup_{i=1}^{\infty} B_i\) has measure zero. For this, observe that
\[
0 \leq \mu \left( \bigcup_{i=1}^{\infty} B_i \right) \leq \sum_{i=1}^{\infty} \mu(B_i) = 0,
\]
so \(\mu(\bigcup_{i=1}^{\infty} B_i) = 0\), as desired. \(\square\)

Let \(P\) and \(P'\) be two properties of \(X\). If \(P\) implies \(P'\) and if \(P\) holds a.e., then \(P'\) holds a.e.

**Definition 3.12.** Consider the set of functions \(f : X \to \mathbb{R}\), where \((X, \mathcal{A}, \mu)\) is a measure space. We say that \(f\) and \(g\) are equal a.e. if the set \(\{x \in X \mid f(x) \neq g(x)\}\) is null. Write \(f = g\) (a.e.).

It is an easy exercise to show that equality a.e. is an equivalence relation.

It should be observed that the notion of equality a.e. is more subtle than one might think.

---

\(^{1}\)Beware that in measure theory, the notion of null set has more than one meaning. Some authors mean something different from what we define here.
Example 3.3. For example, consider the function $\chi_Q : \mathbb{R} \to \mathbb{R}$, given by

$$
\chi_Q(x) = \begin{cases} 
1 & \text{if } x \in \mathbb{Q} \\
0 & \text{if } x \notin \mathbb{Q}.
\end{cases}
$$

In other words, $\chi_Q$ is the characteristic function of the rationals. It is easy to see that $\chi_Q$ is discontinuous at every point $x \in \mathbb{R}$ (if $x$ is irrational, then every small interval containing $x$ contains some rational number; similarly, if $x$ is rational, then every small interval containing $x$ contains some irrational number, say of the form $x + \frac{\sqrt{2}}{2^n}$ for $n$ large enough). Now, the Lebesgue measure $\mu_L$ discussed in Section 3.4 has the property that every countable set has measure zero, so in particular $\mathbb{Q}$ has Lebesgue measure zero. It follows that $\chi_Q$ is equal to the zero function on a.e., and the zero function is a “very nice” function; it is infinitely differentiable.

This is the beauty of equality a.e. Given a “very bad” function, we can ignore its bad behavior on a set of measure zero, as least from the point of view of integration.

An interesting variation of $\chi_Q$ is the following function $D_Q : \mathbb{R} \to \mathbb{R}$, given by

$$
D_Q(x) = \begin{cases} 
\frac{1}{p} & \text{if } x = p/q \in \mathbb{Q}, q \neq 0, \gcd(p, q) = 1 \\
0 & \text{if } x \notin \mathbb{Q}.
\end{cases}
$$

It is easy to show that $D_Q$ is discontinuous at every rational point $x$, but is continuous at every irrational point $x$. In fact, $D_Q$ is a regulated function. Again $D_Q$ is equal to the zero function a.e. (with respect to the Lebesgue measure).

A property that will play an important role is pointwise convergence a.e.

Definition 3.13. Let $(X, \mathcal{A}, \mu)$ be a measure space, and let $F$ be any topological space (in most cases a normed vector space). A sequence $(f_n)_{n \geq 1}$ of functions $f_n : X \to F$ converges pointwise a.e. to a function $f : X \to F$ if there is a null set $Z \subseteq X$ such that the sequence $(f_n(x))_{n \geq 1}$ converges to $f(x)$ for all $x \in X - Z$. See Figure 3.5.

3.3 Construction of a Measure from an Outer Measure

It turns out that defining explicitly a function $m$ satisfying Conditions (2) and (3) from the beginning of Section 3.1 on a $\sigma$-algebra is not easy, but defining a function $\mu^*$ on $2^X$ satisfying (1), (2), and (3'), is quite easy. Furthermore, given such a function $\mu^*$, called an outer measure, there is a way of generating a $\sigma$-algebra and a measure on it.

If $X$ is a locally compact topological space, then there is a way to construct a $\sigma$-algebra and a function $m$ satisfying (2) and (3) on this $\sigma$-algebra using Radon functionals. This method will be explored in Chapter 5.

We now consider Option 1 from Section 3.1 and define outer measures.
Figure 3.5: A schematic illustration of pointwise convergence a.e. Let $X$ be the solid black line, $F = \mathbb{R}$, and $Z = Z_1 \cup Z_2 \cup Z_3$, where each $Z_i$ has measure zero. The sequence $(f_n)$ converges pointwise to the graph $f$ (in red) for all $x \in X - Z$. As shown in the bottom right corner, the magenta graph $f_{n+1}$ is "close" to $f$ outside of $Z$.

**Definition 3.14.** Given a nonempty set $X$, an outer measure $\mu^*$ on $X$ is a function satisfying the following properties:

$(\mu^*1)$ $\mu^*: 2^X \to [0, +\infty]$, that is, $\mu^*$ is defined for all subsets of $X$.

$(\mu^*2)$ $\mu^*(\emptyset) = 0$.

$(\mu^*3)$ For any countable sequence $(A_i)_{i \geq 1}$ of subsets $A_i$ of $X$,

$$\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu^*(A_i).$$

This property is called $\sigma$-subadditivity.

$(\mu^*4)$ If $A \subseteq B$, then $\mu^*(A) \leq \mu^*(B)$. 
3.3. CONSTRUCTION OF A MEASURE FROM AN OUTER MEASURE

Example 3.4. (Outer measure of Dirac) Let $X$ be any nonempty set, and let $a$ be any point chosen in $X$. The map $\mu^*_a: 2^X \to [0, +\infty]$ given by

$$
\mu^*_a(A) = \begin{cases} 
1 & \text{if } a \in A \\
0 & \text{if } a \notin A
\end{cases}
$$

is an outer measure called the outer measure of Dirac.

Here is a simple way to construct outer measures.

Proposition 3.10. Let $X$ be a nonempty set, and $\mathcal{I} \subseteq 2^X$ be a family of subsets with the following properties:

(a) $\emptyset \in \mathcal{I}$.

(b) For every subset $A$ of $X$, there is a countably infinite sequence $(I_n)_{n \geq 1}$ of subsets $I_n \in \mathcal{I}$ such that $A \subseteq \bigcup_{n=1}^{\infty} I_n$.

Moreover, let $\lambda: \mathcal{I} \to [0, +\infty]$ be any function such that

(c) $\lambda(\emptyset) = 0$.

Then the map $\mu^*$ given by

$$
\mu^*(A) = \inf \left\{ \sum_{n=1}^{\infty} \lambda(I_n) \mid A \subseteq \bigcup_{n=1}^{\infty} I_n, I_n \in \mathcal{I} \right\}
$$

is an outer measure on $X$.

As an application of Proposition 3.10, we obtain the outer Lebesgue measure.

Example 3.5. Let $\mathcal{I}$ consist of the set of all open intervals $(a, b)$, where $a = -\infty$ or $b = +\infty$ is allowed. It is easy to see that Properties (a) and (b) of Proposition 3.10 are satisfied. Let $\lambda: \mathcal{I} \to [0, +\infty]$ be given by $\lambda((a, b)) = b - a$. Obviously, Property (c) holds. The outer measure given by Proposition 3.10 is called the outer Lebesgue measure $\mu^*_L$ on $\mathbb{R}$.

A similar construction could be performed on $\mathbb{R}^n$ by using products of open intervals $(a_1, b_1) \times \cdots \times (a_n, b_n)$ and $\lambda((a_1, b_1) \times \cdots \times (a_n, b_n)) = \prod_{i=1}^{n}(b_i - a_i)$.

We now state a fundamental theorem due to C. Carathéodory which gives a method for constructing a measure space from an outer measure.

Theorem 3.11. (Carathéodory) Let $\mu^*: 2^X \to [0, +\infty]$ be an outer measure. Define the family $\mathcal{A}$ of subsets of $X$ as follows:

$$
\mathcal{A} = \{ A \in 2^X \mid \mu^*(E) = \mu^*(E \cap A) + \mu^*(E \cap (X - A)), \text{ for all } E \subseteq X \}.
$$

See Figure 3.6. Then the following properties hold:
(a) $\mathcal{A}$ is a $\sigma$-algebra which contains all subsets $A \subseteq X$ such that $\mu^*(A) = 0$.

(b) The restriction $\mu$ of $\mu^*$ to $\mathcal{A}$ is a measure. Furthermore, $\mu$ is a complete measure.

The proof of Theorem 3.11 is not really difficult but quite long and involves some tedious verifications.

**Example 3.6.** If we apply Theorem 3.11 to the Dirac outer measure $\mu^*_\delta$ of Example 3.4, we find easily that $\mathcal{A} = 2^X$ and that $\mu = \mu^*_\delta$. The Dirac measure $\mu^*_\delta$ is usually denoted by $\delta_\delta$.

If we apply Theorem 3.11 to the Lebesgue outer measure of Example 3.5, we obtain the Lebesgue measure on $\mathbb{R}$. It can be shown that the $\sigma$-algebra of Lebesgue-measurable sets obtained from the construction contains the $\sigma$-algebra of Borel sets of $\mathbb{R}$. This example is considered in slightly more details in the next section.

### 3.4 The Lebesgue Measure on $\mathbb{R}$

Recall that in Example 3.5 we defined the outer Lebesgue measure $\mu^*_L$ on $\mathbb{R}$. For this, we considered the set $\mathcal{I}$ consisting of all open intervals $(a, b)$, where $a = -\infty$ or $b = +\infty$ is allowed. By Proposition 3.10 applied to the function $\lambda: \mathcal{I} \to [0, +\infty]$ given by $\lambda((a, b)) =$...
b - a, we obtained the outer Lebesgue measure \( \mu^*_L \) given by

\[
\mu^*_L(A) = \inf \left\{ \sum_{i=1}^{\infty} \lambda(I_n) \mid A \subseteq \bigcup_{n=1}^{\infty} I_n, \ I_n \in \mathcal{J} \right\}.
\]

By applying Theorem 3.11 to the outer measure \( \mu^*_L \), we obtain the \( \sigma \)-algebra \( \mathcal{L}(\mathbb{R}) \) of Lebesgue-measurable sets, and the Lebesgue measure \( \mu_L \).

The construction used by Theorem 3.11 yields very little explicit information regarding what the Lebesgue-measurable sets look like, but it is possible to describe some of them. In particular, if \( \mathcal{B}(\mathbb{R}) \) denotes the Borel \( \sigma \)-algebra generated by the open sets of \( \mathbb{R} \), it turns out that \( \mathcal{B}(\mathbb{R}) \subseteq \mathcal{L}(\mathbb{R}) \), a proper inclusion. Actually, every open subset of \( \mathbb{R} \) can be expressed as a countable disjoint union of open intervals, so the Borel \( \sigma \)-algebra is generated by the open intervals \((a, b)\). The following proposition gives convenient characterizations of the Borel sets.

**Proposition 3.12.** The \( \sigma \)-algebra \( \mathcal{B}(\mathbb{R}) \) of Borel sets of \( \mathbb{R} \) is generated by the following intervals:

1. \([a, b] \), with \( a \leq b \) finite.
2. \([a, b) \), with \( a \leq b \) finite.
3. \((a, \infty) \), with \( a \) finite.

**Proof.** (1) We know that the open intervals \((a, b)\) generate \( \mathcal{B}(\mathbb{R}) \). We have

\[
(a, \infty) = \bigcup_{n=1}^{\infty} (a, a + n), \quad (-\infty, a) = \bigcup_{n=1}^{\infty} (a - n, a).
\]

Observe that

\[
[a, b] = (-\infty, a) \cap (b, \infty) = [a, \infty) \cap (-\infty, b],
\]

so \([a, b] \in \mathcal{B}(\mathbb{R})\). We also have

\[
(a, b) = \bigcup_{n=1}^{\infty} \left[ a + \frac{1}{n}, b - \frac{1}{n} \right],
\]

so the closed intervals \([a, b] \) generate \( \mathcal{B}(\mathbb{R}) \).

(2) We have

\[
[a, b) = \bigcup_{n=1}^{\infty} \left[ a, b - \frac{1}{n} \right],
\]

so \([a, b) \in \mathcal{B}(\mathbb{R})\). Then

\[
(a, b) = \bigcup_{n=1}^{\infty} \left[ a + \frac{1}{n}, b \right),
\]
so the intervals \([a, b)\) generate \(\mathcal{B}(\mathbb{R})\).

(3) We already know from (1) that \((a, \infty) \in \mathcal{B}(\mathbb{R})\). This implies that

\[ (-\infty, a] = (a, \infty) \in \mathcal{B}(\mathbb{R}), \]

so

\[ (-\infty, a) = \bigcup_{n=1}^{\infty} \left(-\infty, a - \frac{1}{n}\right) \in \mathcal{B}(\mathbb{R}), \]

and thus

\[ (a, b) = (-\infty, b) \cap (a, \infty), \]

so the intervals \((a, \infty)\) generate \(\mathcal{B}(\mathbb{R})\).

Let use the notation \(\wr(a, b)\) to denote any of the four types of intervals \((a, b), [a, b), (a, b], [a, b]\) (with \(a = -\infty\) or \(b = +\infty\) allowed, and \(a = b\) allowed). The following result can be shown.

Proposition 3.13. Let \(\mathcal{B}(\mathbb{R})\) be the Borel \(\sigma\)-algebra of open sets, \(\mathcal{L}(\mathbb{R})\) the \(\sigma\)-algebra of Lebesgue-measurable sets, and \(\mu_L\) be the Lebesgue measure, for \(\mathbb{R}\).

1. \(\mathcal{L}(\mathbb{R}) \neq 2^\mathbb{R}\); that is, there exist non-measurable sets. The proof requires the axiom of choice.

2. \(\mathcal{B}(\mathbb{R}) \subset \mathcal{L}(\mathbb{R})\), the inclusion being strict. This is because \(|\mathcal{B}(\mathbb{R})| = \aleph_1 = 2^{\aleph_0}\), but \(|\mathcal{L}(\mathbb{R})| = \aleph_2 = 2^{\aleph_1}\).

3. The Borel \(\sigma\)-algebra contains all four types of intervals, and

\[ \mu_L(\wr(a, b)) = \begin{cases} b - a & \text{if } a \neq -\infty \text{ and } b \neq +\infty \\ +\infty & \text{if } a = -\infty \text{ or } b = +\infty. \end{cases} \]

4. The restriction of the Lebesgue measure \(\mu_L\) to the Borel \(\sigma\)-algebra is a measure \(\mu_B\). The completion of the measure space \((\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_B)\) given by Proposition 3.8 gives back the measure space \((\mathbb{R}, \mathcal{L}(\mathbb{R}), \mu_L)\) of Lebesgue-measurable sets.

It is surprising how much work it takes to prove Part (3) of Proposition 3.13. As a corollary, every one-point set \(\{a\}\) has Lebesgue measure 0, and thus every countable subset has Lebesgue measure 0. There are also uncountable subsets of Lebesgue measure 0.

The Lebesgue measure also has the following regularity properties which show that every Lebesgue-measurable set can be approximated either by an open set or by a closed set.

Proposition 3.14. For any subset \(A\) of \(\mathbb{R}\), we have

\[ \mu_L^*(A) = \inf \{\mu_L(O) \mid A \subseteq O, O \text{ is open}\}. \]

For every Lebesgue-measurable set \(A \in \mathcal{L}(\mathbb{R})\), the following facts hold:
(a) For every $\epsilon > 0$, there is some open subset $O$ such that $A \subseteq O$ and $\mu_L(O - A) < \epsilon$.

(a) For every $\epsilon > 0$, there is some closed subset $F$ such that $F \subseteq A$ and $\mu_L(A - F) < \epsilon$.

As a corollary of Proposition 3.14 we have the following facts: For every Lebesgue-measurable set $A \in \mathcal{L}(\mathbb{R})$:

(a’) $\mu_L(A) = \inf\{\mu_L(O) \mid A \subseteq O, \text{ } O \text{ is open}\}$.

(b’) $\mu_L(A) = \sup\{\mu_L(F) \mid F \subseteq A, \text{ } F \text{ is closed}\}$.

See Figure 3.7

Figure 3.7: A Lebesgue-measurable set $A$ of $\mathbb{R}$ is approximated from the “outside” by an open set $O$; it is also approximated from the “inside” by a closed set $F$.

It should be noted that Properties (a’) and (b’) are weaker than Properties (a) and (b), because they imply Properties (a) and (b) only when $\mu(A)$ is finite.

It can also be shown that for every Lebesgue-measurable set $A \in \mathcal{L}(\mathbb{R})$, we have

$$\mu_L(A) = \sup\{\mu_L(K) \mid K \subseteq A, \text{ } K \text{ is compact}\}.$$ 

Proposition 3.14 also holds for the Lebesgue-measurable subsets of $\mathbb{R}^n$.

Another important property of the Lebesgue measurable is that it is translation-invariant.

**Proposition 3.15.** For any Lebesgue measurable set $A \in \mathcal{L}(\mathbb{R})$, we have $\mu_L(x + A) = \mu_L(A)$ for all $x \in \mathbb{R}$, where $x + A = \{x + a \mid a \in \mathbb{R}\}$. This property is called translation-invariance.
For a proof, see Section 6.5, Example 6.1.

It can be shown that there is no translation-invariant measure $\mu$ defined on all subsets of $\mathbb{R}$ such that $\mu([0,1]) \neq 0$ or $\mu((0,1)) \neq +\infty$. As a consequence, there is no measure defined on all subsets of $\mathbb{R}$ such that $\mu([a,b]) = b - a$.

To prove this, we can consider the quotient set $\mathbb{R}/\mathbb{Q}$ of the reals modulo the equivalence relation $x \sim y$ iff $x - y \in \mathbb{Q}$. Using the axiom of choice, we can form a subset $E \subseteq [0,1)$ which contains exactly one number from each equivalence class of $\mathbb{R}/\mathbb{Q}$. Then one can express $[0,1)$ as a countable disjoint union of translates of $E$ modulo 1 by the rationals in $[0,1)$, and deduce that $E$ is not Lebesgue measurable, as well as the other facts that we stated.

We conclude by mentioning that if $X$ is a topological space, given a function $\mu$ defined on the open subsets and the compact subsets of $X$, we can define the following maps for every subset $A$ of $X$:

$$\mu^*(A) = \inf\{\mu(O) \mid A \subseteq O, \ A \text{ is open}\}$$

$$\mu_*(A) = \sup\{\mu(K) \mid K \subseteq A, \ K \text{ is compact}\}.$$

Then the measurable subsets are those subsets $A$ of $X$ such that

$$\mu^*(A) = \mu_*(A).$$

It can be shown that these subsets form a $\sigma$-algebra $\mathcal{A}$, and that the map $\mu$ with domain $\mathcal{A}$ given by $\mu(A) = \mu^*(A) = \mu_*(A)$ is a measure. This is the approach using Radon measures.
Chapter 4
Integration

Given a measure space \((X, \mathcal{A}, \mu)\), we would like to define the integral of a real-valued function \(f: X \to \mathbb{R}\), or more generally of a complex-valued function \(f: X \to \mathbb{C}\), or even of a function \(f: X \to F\), where \(F\) is a normed vector space. The key idea is that the integral of a very simple function \(f\), such as a function taking only a finite number of nonzero values \(y_1, \ldots, y_n\), should be “obvious.” Namely, if \(A_i = f^{-1}(y_i)\) is the subset of \(X\) over which \(f\) has the value \(y_i\), then each \(A_i\) should be measurable (that is, \(A_i \in \mathcal{A}\)), and \(A_i\) should have finite measure, so that the expression

\[
\sum_{i=1}^{n} y_i \mu(A_i) \in F
\]

makes sense. Then we define the integral \(\int f \, d\mu\) of our simple function \(f\) as

\[
\int f \, d\mu = \sum_{i=1}^{n} \mu(A_i) y_i. \tag{*}
\]

Observe that the function \(f\) can be written as

\[
f = \sum_{i=1}^{n} y_i \chi_{A_i},
\]

where \(\chi_{A_i}\) is the characteristic function of the subset \(A_i\). Such a function is called a \(\mu\)-step function.

Observe that \((*)\) is a generalization of the notion of area under the curve. If the subsets \(A_i\) are closed adjacent intervals, then we are back to the notion of Riemann integral. However, in our new setting, the subsets \(A_i\) can be very complicated, but as long as they are measurable and have finite measure, the integral \((*)\) makes sense.

If we define \(\|f\|\) as

\[
\|f\| = \sum_{i=1}^{n} \|y_i\| \chi_{A_i},
\]

69
If we define $N_1(f) = \int \|f\| \, d\mu$, then $N_1$ satisfies all the properties of a norm, except that $N_1(f) = 0$ does not necessarily imply that $f = 0$. However, $N_1(f) = 0$ iff $f = 0$ almost everywhere. The set $\text{Step}_\mu(X, \mathcal{A}, F)$ of $\mu$-step functions is a vector space, and $N_1$ is almost a norm on it; it is a semi-norm. The integral given by $(*)$ is a linear continuous map on $\text{Step}_\mu(X, \mathcal{A}, F)$. However, the space $\text{Step}_\mu(X, \mathcal{A}, F)$ is not Cauchy-complete under the semi-norm $N_1$ (there are Cauchy sequences with respect to $N_1$ that do not have a limit). The problem then is to complete the space $\text{Step}_\mu(X, \mathcal{A}, \mu)$ and to extend the integral $(*)$ to this bigger set of functions.

There are several ways to proceed.

1. If we let $SN$ be subspace of $\text{Step}_\mu(X, \mathcal{A}, F)$ consisting of the $\mu$-step functions equal to $0$ a.e., then the quotient space $\text{Step}_\mu(X, \mathcal{A}, F) = \text{Step}_\mu(X, \mathcal{A}, \mu)/SN$ is a vector space and $N_1$ induces a (true) norm on it. Therefore we can apply the general completion theorem (Theorem 15.68) to obtain a complete normed vector space $(L_\mu(X, \mathcal{A}, F), \|\|_1)$. Since integration is a linear continuous map on $\text{Step}_\mu(X, \mathcal{A}, F)$, it extends uniquely to a linear continuous map on $L_\mu(X, \mathcal{A}, F)$.

In theory, we have achieved our goal of defining a complete normed vector space of functions containing the $\mu$-step functions for which every function is integrable. However, the completion $(L_\mu(X, \mathcal{A}, F), \|\|_1)$ is a very complicated object. It consists of equivalence classes of Cauchy sequences of functions in the quotient space $\text{Step}_\mu(X, \mathcal{A}, F)$. It would be much more convenient if the objects in $L_\mu(X, \mathcal{A}, F)$ could be described as functions, and this is indeed possible.

2. The second approach is to first define a set $L_\mu(X, \mathcal{A}, F)$ of functions using a limit process. Every function $f$ in $L_\mu(X, \mathcal{A}, F)$ is the limit pointwise a.e. of a $N_1$-Cauchy sequence $(f_n)_{n \geq 1}$ (called an approximation sequence) of functions $f_n$ in $\text{Step}_\mu(X, \mathcal{A}, F)$. We also define the space $M_\mu(X, \mathcal{A}, F)$ of $\mu$-measurable functions, and $L_\mu(X, \mathcal{A}, F)$ is the subspace of $M_\mu(X, \mathcal{A}, F)$ consisting of the functions for which the integral is well defined.

It turns out that $L_\mu(X, \mathcal{A}, F)$ is complete with respect to an extension $\|\|_1$ of the semi-norm $N_1$, and the integral $\int f \, d\mu$ of any function $f \in L_\mu(X, \mathcal{A}, F)$ can be defined by a limit process. There are technical complications when $F$ is infinite-dimensional, and it also takes some work to show that the integral of a function $f \in L_\mu(X, \mathcal{A}, F)$ does not depend on the approximation sequence used to define $f$, but all difficulties can be overcome. Finally, the subspace $\mathcal{N}$ of functions $f$ such that $\|f\|_1 = 0$ is the set of functions equal to $0$ a.e., and we obtain the complete space $(L_\mu(X, \mathcal{A}, F), \|\|_1)$ of
the first approach as the quotient space $L_\mu(X, \mathcal{A}, F)/\mathcal{N}$. However, the construction of $L_\mu(X, \mathcal{A}, F)$ is much more informative.

We also investigate convergence properties of $L_\mu(X, \mathcal{A}, F)$, as well as other related spaces (the spaces $L_p^\mu(X, \mathcal{A}, F)$, $p = 1, 2, \infty$). We conclude with the construction of the integral on a product space.

The vector valued-integral defined in this chapter (where the space $F$ of values is a Banach space) was first discovered by Bochner in 1933. The version discussed here is due to Dunford (1935), and is presented in detail in Dunford and Schwartz [28]. More recent expositions of this method are given in Lang [52] and Marle [56].

4.1 Measurable Maps

Measurable functions are functions between measurable spaces that are the analog of continuous functions between topological spaces, but as we will see, they are a lot more flexible, especially in terms of convergence properties. In this chapter, our presentation follows Marle [56] and Lang [52] very closely.

**Definition 4.1.** Given any two measurable spaces $(X, \mathcal{A})$ and $(Y, \mathcal{B})$, a function $f: X \rightarrow Y$ is measurable if $f^{-1}(B) \in \mathcal{A}$ for every $B \in \mathcal{B}$. A measurable function is also called a measurable map.

If $(X, \mathcal{A})$ is a measurable space, then obviously the identity id: $X \rightarrow X$ is measurable.

The composition of two measurable maps is also measurable.

**Proposition 4.1.** Given three measurable spaces $(X, \mathcal{A})$, $(Y, \mathcal{B})$, and $(Z, \mathcal{C})$, if $f: X \rightarrow Y$ and $g: Y \rightarrow Z$ are measurable maps, then $g \circ f: X \rightarrow Z$ is a measurable map.

**Proof.** Recall that one of the properties of inverse images is that $(g \circ f)^{-1}(C) = f^{-1}(g^{-1}(C))$, for any subset $C$ of $Z$. But if $C \in \mathcal{C}$, since $g$ is measurable, $g^{-1}(C) \in \mathcal{B}$, and since $f$ is measurable, $f^{-1}(g^{-1}(C)) \in \mathcal{A}$, which shows that $g \circ f$ is measurable. \[\square\]

**Remark:** The above properties show that measurable spaces are the objects of a category whose morphisms are the measurable maps.

**Proposition 4.2.** Let $X$ and $Y$ be any two nonempty sets, and let $f: X \rightarrow Y$ be a function between them.

1. If $\mathcal{A}$ is a $\sigma$-algebra on $X$, then we can define $\mathcal{A}_f$ as the family of subsets of $Y$ given by

   $$\mathcal{A}_f = \{B \in 2^Y \mid f^{-1}(B) \in \mathcal{A}\}.$$ 

   Then $\mathcal{A}_f$ is the largest $\sigma$-algebra on $Y$ which makes $f$ measurable.
(2) If $\mathcal{B}$ is a $\sigma$-algebra on $Y$, then let $f^{-1}(\mathcal{B})$ be the family of subsets of $X$ given by

$$f^{-1}(\mathcal{B}) = \{ f^{-1}(B) \in 2^X \mid B \in \mathcal{B} \}.$$ 

Then $f^{-1}(\mathcal{B})$ is the smallest $\sigma$-algebra on $X$ which makes $f$ measurable.

The proof of Proposition 4.2 is left as an exercise.

Using Proposition 4.2 we obtain the following proposition which gives simple criteria to check that a map is measurable.

**Proposition 4.3.** Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be two measurable spaces.

(1) If $S$ generates the $\sigma$-algebra $\mathcal{B}$ (which means that the smallest $\sigma$-algebra containing $S$ is $\mathcal{B}$), then a function $f : X \to Y$ is measurable iff $f^{-1}(S) \in \mathcal{A}$ for all $S \in S$.

(2) If $Y$ is a topological space and if $\mathcal{B}$ is its Borel $\sigma$-algebra of open subsets, then a function $f : X \to Y$ is measurable iff $f^{-1}(U) \in \mathcal{A}$ for every open subset $U$ of $Y$ (or $f^{-1}(U) \in \mathcal{A}$ for every closed subset $U$ of $Y$).

(3) If $X$ and $Y$ are both topological spaces and if $\mathcal{A}$ and $\mathcal{B}$ are their respective Borel $\sigma$-algebras, then every continuous map $f : X \to Y$ is measurable.

Given any subset $A$ of $X$, recall that the characteristic function $\chi_A$ of $A$ is defined by

$$\chi_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A. \end{cases}$$

Then, as illustrated in Figure 4.1, it is easy to show that for any subset $A$ of $X$, the function $\chi_A : X \to \mathbb{R}$ (where $\mathbb{R}$ is equipped with its $\sigma$-algebra of Borel sets) is measurable iff $A \in \mathcal{A}$, that is, $A$ is measurable.

In the theory of integration, all maps of interest will be measurable mapsootnote{Actually, not quite in the most general case, but they will be equal to a measurable map a.e.} $f : X \to F$ where $(X, \mathcal{A})$ is a measurable space, and $(F, \mathcal{B})$ is a measurable space such that either $F = \mathbb{R}$, or $F = \mathbb{C}$, or more generally $F$ is a Banach space (a complete normed vector space over $\mathbb{R}$ or $\mathbb{C}$), and $\mathcal{B}$ is the Borel $\sigma$-algebra of open subsets of $F$. In this case, various operations can be performed on functions $f : X \to F$.

Assume that $F$ is a normed vector space over the field $K$, where $K = \mathbb{R}$ or $K = \mathbb{C}$, and that $f : X \to F$ is any function, not necessarily measurable.

1. Given any function $f : X \to F$, for any $\lambda \in K$, let $\lambda f : X \to F$ be the function given by

$$(\lambda f)(x) = \lambda f(x), \quad x \in X.$$
4.1. MEASURABLE MAPS

Figure 4.1: The upper figure illustrates $\chi_A: X \to \mathbb{R}$. If $S_1 \subset \mathbb{R}$ contains 1 but not 0, $\chi_A^{-1}(S_1) = A$. If $S_2 \subset \mathbb{R}$ contains 0 but not 1, $\chi_A^{-1}(S_1) = X - A$. Finally, if $S_3 \subset \mathbb{R}$ contains both 0 and 1, $\chi_A^{-1}(S_1) = A \cup (X - A) = X$.

2. Given any function $f: X \to F$, let $\|f\| : X \to \mathbb{R}_+$ be the function given by

$$\|f\|(x) = \|f(x)\|, \quad x \in X.$$ Beware that $\|f\|$ is not the norm of the function $f$, where $\|\|$ is the norm on some function space consisting of functions from $X$ to $F$. Instead, $\|f\|$ is the function defined pointwise as $\|f(x)\|$ for every $x \in X$, where $\|f(x)\|$ is the norm of $f(x)$ in $F$. This notation is somewhat confusing but appears to be standard. Later on, we will equip our space of functions from $X$ to $F$ with a norm, but it will be denoted $\|\|_1$, or more generally $\|\|_p$, so there will be no risk of confusion.

3. For any two functions $f: X \to \mathbb{R}$ and $g: X \to \mathbb{R}$, let $\sup(f,g)$ and $\inf(f,g)$ be the functions given by

$$\sup(f,g)(x) = \max(f(x), g(x)), \quad x \in X,$$
$$\inf(f,g)(x) = \min(f(x), g(x)), \quad x \in X.$$
4. For any two functions $f : X \to \mathbb{R}$, let $f^+$ and $f^-$ be the functions given by

$$f^+(x) = \begin{cases} 0 & \text{if } f(x) \leq 0 \\ f(x) & \text{if } f(x) > 0, \end{cases}$$

$$f^-(x) = \begin{cases} 0 & \text{if } f(x) \geq 0 \\ -f(x) & \text{if } f(x) < 0. \end{cases}$$

We also define $|f|$ as $|f| = f^+ + f^- = \sup(f, -f)$. Observe that $f = f^+ - f^-$. 

5. For any two functions $f : X \to F$ and $g : X \to F$, let $f + g : X \to F$ be the function given by

$$(f + g)(x) = f(x) + g(x), \quad x \in X.$$ 

6. For any two functions $f : X \to K$ and $g : X \to K$, where $K = \mathbb{R}$ or $K = \mathbb{C}$, let $fg : X \to K$ be the function given by

$$(fg)(x) = f(x)g(x), \quad x \in X.$$ 

**Definition 4.2.** Let $(X, \mathcal{A})$ be a measurable space, and let $(F, \mathcal{B})$ be a measurable space such that either $F = \mathbb{R}$, or $F = \mathbb{C}$, or more generally $F$ is a Banach space (a complete normed vector space over $\mathbb{R}$ or $\mathbb{C}$), and $\mathcal{B}$ is the Borel $\sigma$-algebra of open subsets of $F$. The set of measurable maps $f : X \to F$ is denoted by $\mathcal{M}(X, \mathcal{A}, F)$.

The following technical result is needed.

**Proposition 4.4.** Let $(X, \mathcal{A})$ be any measurable space, and let $(F_1, \mathcal{B}_1)$, $(F_2, \mathcal{B}_2)$, and $(G, \mathcal{G})$ be three measurable spaces, where $F_1, F_2, G$ are topological spaces, and $\mathcal{B}_1, \mathcal{B}_2, \mathcal{G}$ are their respective Borel $\sigma$-algebras. Let $h : F_1 \times F_2 \to G$ be a continuous map, and let $f_1 : X \to F_1$ and $f_2 : X \to F_2$ be two measurable maps. If the subspace topologies on $f_1(X) \subseteq F_1$ and $f_2(X) \subseteq F_2$ are second-countable (which means that they have a countable basis of open subsets), then $h \circ (f_1, f_2) : X \to G$ is measurable.

Recall that a topological space $E$ is separable if it contains a countable subset which is dense in $E$ (see Definition 15.34). If $E$ is a metric space, then by Proposition 15.42, the space $E$ is separable if and only if it is second-countable. Using Proposition 4.4 we obtain the following important result stating various closure properties of $\mathcal{M}(X, \mathcal{A}, F)$.

**Proposition 4.5.** Let $(X, \mathcal{A})$ be any measurable space, and assume that $F$ is a normed vector space over the field $K$, where $K = \mathbb{R}$ or $K = \mathbb{C}$. The following properties hold:

1. For any $f \in \mathcal{M}(X, \mathcal{A}, F)$ and any $\lambda \in K$, we have $\lambda f \in \mathcal{M}(X, \mathcal{A}, F)$.

2. For any $f \in \mathcal{M}(X, \mathcal{A}, F)$, we have $\|f\| \in \mathcal{M}(X, \mathcal{A}, \mathbb{R})$. 

4.2. **STEP MAPS ON A MEASURABLE SPACE**

3. For any $f \in \mathcal{M}(X, \mathcal{A}, \mathbb{R})$ and any $g \in \mathcal{M}(X, \mathcal{A}, \mathbb{R})$, we have $\sup(f, g), \inf(f, g), f^+, f^-, |f| \in \mathcal{M}(X, \mathcal{A}, \mathbb{R})$.

4. For any $f \in \mathcal{M}(X, \mathcal{A}, F)$ and any $g \in \mathcal{M}(X, \mathcal{A}, F)$, if $f(X)$ and $g(X)$ are separable subsets of $F$, then $f + g \in \mathcal{M}(X, \mathcal{A}, F)$. In particular, if $F$ is separable, then $\mathcal{M}(X, \mathcal{A}, F)$ is a vector space over $K$.

5. For any $f \in \mathcal{M}(X, \mathcal{A}, K)$ and any $g \in \mathcal{M}(X, \mathcal{A}, K)$, we have $fg \in \mathcal{M}(X, \mathcal{A}, K)$. This implies that $\mathcal{M}(X, \mathcal{A}, K)$ is actually a $K$-algebra.

One will observe that in (4), if $F$ is infinite-dimensional, the sum of two measurable maps may *not* be measurable. This is the first technical difficulty of the general theory of integration (with values in an infinite-dimensional vector space). As we will see, a second technical difficulty has to do with the approximation of a measurable map by step functions. Fortunately these technical difficulties can be overcome in a simple way.

The following important result shows that measurable maps behave better than continuous maps in terms of simple convergence.

**Theorem 4.6.** Let $(X, \mathcal{A})$ and $(F, \mathcal{B})$ be two measurable spaces, where $F$ is a metric space and $\mathcal{B}$ is the Borel $\sigma$-algebra on $F$. If $(f_n)_{n \geq 1}$ is a sequence of measurable maps $f_n \in \mathcal{M}(X, \mathcal{A}, F)$ which converges pointwise to a function $f : X \to F$, then $f \in \mathcal{M}(X, \mathcal{A}, F)$; that is, $f$ is measurable.

A proof of Theorem 4.6 can be found in Lang [52] (Chapter VI, Section 1, Property M7).

Our next goal is to generalize the notion of step function given in Definition 1.15 to the framework of measure spaces.

### 4.2 Step Maps on a Measurable Space

Let $(X, \mathcal{A})$ be a measurable space. The generalization of the notion of step map is obtained by replacing the intervals $(a_i, a_{i+1})$ by *arbitrary measurable sets*.

**Definition 4.3.** Let $(X, \mathcal{A})$ be a measurable space, and let $F$ be any set. A function $f : X \to F$ is a *step map* (with respect to $\mathcal{A}$) if there is a finite partition $(A_1, \ldots, A_n)$ of $X$ by pairwise disjoint nonempty subsets $A_i \in \mathcal{A}$ such that $X = \bigcup_{i=1}^{n} A_i$, and such that the restriction of $f$ to each $A_i$ is a constant function with some value $y_i \in F$. The partition $(A_1, \ldots, A_n)$ is said to be *adapted to $f$*. See Figure 4.2. The set of all step maps is denoted by $\text{Step}(X, \mathcal{A}, F)$.

Observe that every constant function is a step map, and that $f(X)$ is a finite subset of $F$. At this stage, no measure $\mu$ is involved, but for the theory of integration, we will have a measure space $(X, \mathcal{A}, \mu)$ and we will need to require each $A_i$ for which $y_i \neq 0$ to have finite measure (this makes sense since in this case $F$ is a vector space).

We gather some useful properties of step maps in the following proposition.
Figure 4.2: Let \((X, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))\) and \(F = \mathbb{R}\). A step map is shown in blue with values \(\{y_i\}_{i=1}^n\). The partition \((A_1, \ldots, A_n)\) adapted to \(f\) is shown underneath the peach box.

**Proposition 4.7.** Let \((X, \mathcal{A})\) be a measurable space, and let \(F\) be any set.

1. For any \(\sigma\)-algebra \(\mathcal{B}\) on \(F\), every step map \(\text{Step}(X, \mathcal{A}, F)\) is measurable.

2. Let \(F_1, F_2, G\) be three sets, and let \(h: F_1 \times F_2 \to G\) be any function. For any \(f_1 \in \text{Step}(X, \mathcal{A}, F_1)\) and any \(f_2 \in \text{Step}(X, \mathcal{A}, F_2)\), we have \(h \circ (f_1, f_2) \in \text{Step}(X, \mathcal{A}, G)\).

3. If \(K = \mathbb{R}\) or \(K = \mathbb{C}\), then \(\text{Step}(X, \mathcal{A}, K)\) is a vector space and a ring under pointwise multiplication of functions. Thus, \(\text{Step}(X, \mathcal{A}, K)\) is an algebra over \(K\).

4. If \(F\) is a vector space over \(K\) (with \(K = \mathbb{R}\) or \(K = \mathbb{C}\)), then \(\text{Step}(X, \mathcal{A}, F)\) is a vector space over \(K\), and a module over \(\text{Step}(X, \mathcal{A}, K)\), which means that if \(f \in \text{Step}(X, \mathcal{A}, F)\) and \(g \in \text{Step}(X, \mathcal{A}, K)\), then \(gf \in \text{Step}(X, \mathcal{A}, F)\).

5. If \(F\) is a normed vector space, and if \(f \in \text{Step}(X, \mathcal{A}, F)\), then \(\|f\| \in \text{Step}(X, \mathcal{A}, \mathbb{R})\).

6. If \(f \in \text{Step}(X, \mathcal{A}, \mathbb{R})\) and \(g \in \text{Step}(X, \mathcal{A}, \mathbb{R})\), then we have \(\sup(f, g), \inf(f, g), f^+, f^-, |f| \in \text{Step}(X, \mathcal{A}, \mathbb{R})\).
By Theorem 4.6, if $F$ is a metric space equipped with its $\sigma$-algebra of Borel sets, if a function $f: X \to F$ is the limit of a sequence $(f_n)_{n \geq 1}$ of step functions $f_n \in \text{Step}(X, \mathcal{A}, F)$ that converges pointwise, then the function $f: X \to F$ must be measurable, since the $f_n$ are measurable by Proposition 4.7.

Unfortunately, in general, a measurable map $f: X \to F$ may not be the pointwise limit of a sequence of step maps if $F$ has infinite dimension. For one thing, such a limit of steps maps has its image contained in the closure of a countable subset of $F$. This is the second technical difficulty of the general theory.

To overcome this second difficulty, one needs to define a more refined notion of measurable map and of step map. We will do so shortly, but first we observe that if we only need to consider values in a finite-dimensional vector space, then there is no problem.

**Proposition 4.8.** Let $(X, \mathcal{A})$ and $(F, \mathcal{B})$ be two measurable spaces, where $F$ is a topological space and $\mathcal{B}$ is its Borel $\sigma$-algebra, and let $f: X \to F$ be a measurable map.

1. If $F$ is either a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, or $F = \mathbb{R}_+$, then there is a sequence $(f_n)$ of step maps $f_n \in \text{Step}(X, \mathcal{A}, F)$ that converges pointwise to $f$. If $F = \mathbb{R}_+$, we may assume that the $f_n$ take finite values.

2. If $F = \mathbb{R}$ or $F = \mathbb{R}_+$, and if $f \geq 0$, then we may assume that $f_n \geq 0$ and $f_n \leq f_{n+1}$ for all $n \geq 1$.

A proof of Proposition 4.8 can be found in Lang [52] (Chapter VI, Section 1, Properties M8 and M9).

### 4.3 $\mu$-Measurable Maps and $\mu$-Step Maps

We explained in the previous section that in general, the space $\mathcal{M}(X, \mathcal{A}, F)$ of measurable maps from $X$ to $F$ is not a vector space, and that a measurable map $f: X \to F$ may not be the pointwise limit of a sequence of step maps. This suggests modifying the notion of measurable map and the notion of step map to recover these properties. The second property is crucial in extending the notion of integral to more general functions.

So far, the space $X$ was only a measurable space, but no measure was involved. The new ingredient is to define a suitable notions of step map and measurable map relative to a *measure space* $(X, \mathcal{A}, \mu)$, where the measure $\mu$ plays a role.

The main trick is to relax the notion of pointwise convergence to pointwise convergence almost everywhere, and more generally, to consider that two functions are equivalent if they are equal almost everywhere (they differ on a null set). The plan is the following:

1. Define the space $\text{Step}_\mu(X, \mathcal{A}, F)$ of $\mu$-step maps.
2. Define the space $\mathcal{M}_\mu(X, A, F)$ of $\mu$-measurable maps, where a $\mu$-measurable map is the limit of a sequence $(f_n)$ of $\mu$-step maps $f_n \in \text{Step}_\mu(X, A, F)$ converging pointwise almost everywhere.

3. Prove that if $F$ is a vector space, then $\mathcal{M}_\mu(X, A, F)$ is a vector space.

Our presentation of the method that we just sketched follows Marle [56] and Lang [52] very closely. It is a generalization (with some simplifications) to functions with values in a Banach space of the approach followed by Halmos [39]. The results that we state without proof are proved either in Marle [56] or in Lang [52].

**Definition 4.4.** Let $(X, A, \mu)$ be a measure space, and let $F$ be any vector space (over $\mathbb{R}$ or $\mathbb{C}$). A function $f : X \to F$ is a $\mu$-step map if it is a step map, and if $\{x \in X \mid f(x) \neq 0\} \in A$ and has finite measure. See Figure 4.3. The set of $\mu$-step maps is denoted by $\text{Step}_\mu(X, A, F)$.

![Figure 4.3: Let $(X, A) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and $F = \mathbb{R}$. A $\mu$-step map is shown in red where $A_1 \cup A_2 \cup A_3 = \{x \in X \mid f(x) \neq 0\} \in A$.](image)

For technical reasons, it is useful to have the following equivalent characterization of a $\mu$-step map.

**Proposition 4.9.** Let $(X, A, \mu)$ be a measure space, and let $F$ be any vector space (over $\mathbb{R}$ or $\mathbb{C}$). A function $f : X \to F$ is a $\mu$-step map iff there is a nonempty subset $A \in A$ of finite measure such that $f$ vanishes outside $A$, that is, $f(x) = 0$ for all $x \in X - A$, and if there is a finite partition $(A_1, \ldots, A_n)$ of $A$ of subsets $A_i \in A$ (nonempty pairwise disjoint subsets) such that the restriction of $f$ to each $A_i$ has a constant value $y_i$.

**Proof.** Let $f$ be a step map with respect to a partition $(A_1, \ldots, A_n)$ of $X$ such that $\{x \in X \mid f(x) \neq 0\} \in A$ and has finite measure. Then any $A_i$ on which $f$ has value $y_i \neq 0$ must
have finite measure. If \( f = 0 \) on \( X \), then pick \( A \) to be any \( A_i \) and the partition to be \((A_i)\). Otherwise, let \( J = \{ j \in \{1, \ldots, n \} \mid f \neq 0 \text{ on } A_j \} \), and let \( A = \bigcup_{j \in J} A_j \). Then, \((A_j)_{j \in J}\) is a partition of \( A \) with \( A_j \in \mathcal{A} \), where \( A \) is a nonempty set of finite measure, and \( f \) vanishes on \( X - A \). See Figure 4.4.

![Figure 4.4](image)

Figure 4.4: Let \((X, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))\) and \( F = \mathbb{R} \). A step map is shown in red with adapted partition \( \mathbb{R} = \bigcup_{i=1}^{7} A_i \). To interpret this step map as a \( \mu \)-step map, let \( A = A_2 \cup A_4 \cup A_6 \), where \( \mu(A) < \infty \), \( A = \{ x \in X \mid f(x) \neq 0 \} \in \mathcal{A} \). Then \( f(x) = 0 \) on \( X - A \) where \( X - A = A_1 \cup A_3 \cup A_5 \cup A_7 \).

Conversely, since \( A \) has finite measure and since the \( A_i \) belongs to \( \mathcal{A} \), each \( A_i \) has finite measure, so \( \{ x \in X \mid f(x) \neq 0 \} \in \mathcal{A} \) is a set of finite measure. If \( A = X \), then we already have a step map (as defined in Definition 4.3). Otherwise, \( X - A \in \mathcal{A} \) and \( f \) vanishes on \( X - A \), so \((A_1, \ldots, A_n, X - A)\) is partition of \( X \), and \( f \) is a step map with respect to this partition. See Figure 4.5.

The condition that a \( \mu \)-step map must vanish outside of a measurable set of finite measure is the measure-theoretic analog of the topological notion of compact support.

Proposition 4.9 suggests the following equivalent definition of a \( \mu \)-step map.

**Definition 4.5.** Let \((X, \mathcal{A}, \mu)\) be a measure space, and let \( F \) be any vector space (over \( \mathbb{R} \) or \( \mathbb{C} \)). A function \( f: X \to F \) is a \( \mu \)-step map if there is a nonempty subset \( A \in \mathcal{A} \) of finite measure such that \( f \) vanishes outside \( A \), that is, \( f(x) = 0 \) for all \( x \in X - A \), and if there is a finite partition \((A_1, \ldots, A_n)\) of \( A \) consisting of nonempty pairwise disjoint subsets in \( \mathcal{A} \), such that the restriction of \( f \) to each \( A_i \) has a constant value \( y_i \) (possibly zero). The partition \((A_1, \ldots, A_n)\) of \( A \) is said to be adapted to \( f \).
CHAPTER 4. INTEGRATION

Figure 4.5: Let \((X, \mathcal{A}) = (\mathbb{R}, \mathcal{B}(\mathbb{R}))\) and \(F = \mathbb{R}\). A \(\mu\)-step map is shown in red with \(A = A_1 \cup A_2 \cup A_3 \cup A_4\). The turquoise set is \(X - A\) and \((A_1, A_2, A_3, A_4, X - A)\) forms an adapted partition for the corresponding step map.

Technically, Definition 4.5 appears to be more convenient. Observe that a \(\mu\)-step map can be expressed as a (necessarily finite) linear combination

\[
\tilde{f} = \sum_{i=1}^{n} y_i \chi_{A_i},
\]

for some \(y_i \in F\) and for some nonempty pairwise disjoint measurable sets \(A_i \in \mathcal{A}\) of finite measure, a concise and convenient representation.

**Remark:** The proof of Proposition 4.9 shows that if a \(\mu\)-step function \(f\) is not identically zero, then we can find a subset \(A\) in \(\mathcal{A}\) of finite measure, and a partition \((A_1, \ldots, A_n)\) of \(A\) of subsets in \(\mathcal{A}\) such that the value of \(f\) on each \(A_i\) is nonzero, and \(f\) is zero outside of \(A\). However, it turns out to be more convenient for certain proofs to allow \(f\) to be zero on some of the \(A_i\), and this is why we allow this possibility in Definition 4.5.

**Example 4.1.** Consider the function \(f: \mathbb{R} \to \mathbb{R}\), given by

\[
f(x) = \begin{cases} 
0 & \text{if } x < 0 \text{ or } x > 1 \\
1 & \text{if } x \in [0, 1/2] - \mathbb{Q} \\
0 & \text{if } x \in [0, 1/2] \cap \mathbb{Q} \\
2 & \text{if } x \in [1/2, 1] - \mathbb{Q} \\
0 & \text{if } x \in [1/2, 1] \cap \mathbb{Q}.
\end{cases}
\]

See Figure 4.6 If we let \(A_1 = [0, 1/2] - \mathbb{Q}\), \(A_2 = [0, 1/2] \cap \mathbb{Q}\), \(A_3 = [1/2, 1] - \mathbb{Q}\), \(A_4 = [1/2, 1] \cap \mathbb{Q}\), and \(A = [0, 1]\), with the Lebesgue measure \(\mu_L\) on \(\mathbb{R}\), then \(A_1, A_2, A_3, A_4\) are
4.3. \(\mu\)-MEASURABLE MAPS AND \(\mu\)-STEP MAPS

Lebesgue measurable, \(\mu(A_1) = 1/2, \mu(A_2) = 0, \mu(A_3) = 1/2, \mu(A_3) = 0\), \((A_1, A_2, A_3, A_4)\) is a partition of \(A\), a set of measure 1. Thus \(f\) is a \(\mu\)-L-step function.

![Figure 4.6: The \(\mu\)-step map of Example 4.1](image)

This example shows that a \(\mu\)-step function can be very complicated, unlike the step functions of Definition 1.15.

**Proposition 4.10.** Let \((X, \mathcal{A}, \mu)\) be a measure space, and let \(F\) be any vector space

1. Let \(F_1, F_2, G\) be three Banach spaces over \(\mathbb{R}\) or \(\mathbb{C}\), and let \(h: F_1 \times F_2 \to G\) be any function. If \(h\) satisfies \(h(0,0) = 0\), then for any \(f_1 \in \text{Step}_\mu(X, \mathcal{A}, F_1)\) and any \(f_2 \in \text{Step}_\mu(X, \mathcal{A}, F_2)\), we have \(h \circ (f_1, f_2) \in \text{Step}_\mu(X, \mathcal{A}, G)\).

2. If \(K = \mathbb{R}\) or \(K = \mathbb{C}\), then \(\text{Step}_\mu(X, \mathcal{A}, K)\) is a subspace of \(\text{Step}(X, \mathcal{A}, K)\), and for any \(g \in \text{Step}(X, \mathcal{A}, K)\) and any \(f \in \text{Step}_\mu(X, \mathcal{A}, K)\) we have \(gf \in \text{Step}_\mu(X, \mathcal{A}, K)\). Thus \(\text{Step}_\mu(X, \mathcal{A}, K)\) is an ideal in \(\text{Step}(X, \mathcal{A}, K)\).

3. If \(F\) is a vector space over \(K\) (with \(K = \mathbb{R}\) or \(K = \mathbb{C}\)), then \(\text{Step}_\mu(X, \mathcal{A}, F)\) is a subspace of \(\text{Step}(X, \mathcal{A}, F)\) and a module over \(\text{Step}(X, \mathcal{A}, K)\), which means that if \(f \in \text{Step}_\mu(X, \mathcal{A}, F)\) and \(g \in \text{Step}(X, \mathcal{A}, K)\), then \(gf \in \text{Step}_\mu(X, \mathcal{A}, F)\).

4. If \(F\) is a normed vector space, and if \(f \in \text{Step}_\mu(X, \mathcal{A}, F)\), then \(\|f\| \in \text{Step}_\mu(X, \mathcal{A}, \mathbb{R})\). In fact, if \(f = \sum_{i=1}^n y_i \chi_{A_i}\), then \(\|f\| = \sum_{i=1}^n \|y_i\| \chi_{A_i}\).

5. If \(f \in \text{Step}_\mu(X, \mathcal{A}, \mathbb{R})\) and \(g \in \text{Step}_\mu(X, \mathcal{A}, \mathbb{R})\), then \(\sup(f,g), \inf(f,g), f^+, f^-, |f| \in \text{Step}_\mu(X, \mathcal{A}, \mathbb{R})\).

We now come to the crucial notion of \(\mu\)-measurable map.
Definition 4.6. Let \((X, \mathcal{A}, \mu)\) be a measure space, and let \(F\) be any vector space (over \(\mathbb{R}\) or \(\mathbb{C}\)). A function \(f: X \to F\) is a \(\mu\)-measurable if there is a sequence \((f_n)_{n \geq 1}\) of \(\mu\)-step maps \(f_n \in \text{Step}_\mu(X, \mathcal{A}, F)\) which converges pointwise to \(f\) almost everywhere. See Figures 4.7 and 4.8. Recall that this means that there is a null set \(Z \subseteq X\) such that for every \(x \in X - Z\), the sequence \((f_n(x))\) converges to \(f(x)\). The set of \(\mu\)-measurable maps is denoted by \(\mathcal{M}_\mu(X, \mathcal{A}, F)\).

![Figure 4.7: Let \(X = F = \mathbb{R}\). Assume \(\mu\) is the Lebesgue measure on \(\mathbb{R}\). The graph of the \(\mu\)-measurable \(f\) is shown in the upper left corner. The middle figure illustrates the sequence \((f_n)_{n \geq 1}\) of \(\mu\)-step maps \(f_n \in \text{Step}_\mu(X, \mathcal{A}, F)\) which converges pointwise to \(f\) almost everywhere.](image1)

Observe that a \(\mu\)-measurable map is not necessarily measurable, so \(\mathcal{M}_\mu(X, \mathcal{A}, F)\) is not a subspace of \(\mathcal{M}(X, \mathcal{A}, F)\). However, we will see shortly that a \(\mu\)-measurable map is equal to a measurable map almost everywhere, and this is good enough to construct the Lebesgue integral. The following proposition can be proved using Proposition 4.10 by passing to the limit (carefully).

Proposition 4.11. Let \((X, \mathcal{A}, \mu)\) be a measure space, and let \(F\) be any vector space

1. Let \(F_1, F_2, G\) be three Banach spaces over \(\mathbb{R}\) or \(\mathbb{C}\), and let \(h: F_1 \times F_2 \to G\) be any function. If \(h\) satisfies \(h(0, 0) = 0\), then for any \(f_1 \in \mathcal{M}_\mu(X, \mathcal{A}, F_1)\) and any \(f_2 \in \mathcal{M}_\mu(X, \mathcal{A}, F_2)\), we have \(h \circ (f_1, f_2) \in \mathcal{M}_\mu(X, \mathcal{A}, G)\).

2. If \(K = \mathbb{R}\) or \(K = \mathbb{C}\), then \(\mathcal{M}_\mu(X, \mathcal{A}, K)\) is a vector space, and for all \(f, g \in \mathcal{M}_\mu(X, \mathcal{A}, K)\) we have \(fg \in \mathcal{M}_\mu(X, \mathcal{A}, K)\). Thus \(\mathcal{M}_\mu(X, \mathcal{A}, K)\) is an algebra over \(K\). For any \(g \in \mathcal{M}(X, \mathcal{A}, K)\) and any \(f \in \mathcal{M}_\mu(X, \mathcal{A}, K)\) we have \(gf \in \mathcal{M}_\mu(X, \mathcal{A}, K)\).
4.3. \(\mu\)-MEASURABLE MAPS AND \(\mu\)-STEP MAPS

Figure 4.8: A more detailed look at the pointwise convergence of converges the \((f_n)_{n \geq 1}\). The pointwise limit only differs from \(f\) on a set of measure zero, namely \(\{3/2\}\).

3. If \(F\) is a vector space over \(K\) (with \(K = \mathbb{R}\) or \(K = \mathbb{C}\)), then \(\mathcal{M}_\mu(X, \mathcal{A}, F)\) is a vector space over \(K\) and a module over \(\mathcal{M}(X, \mathcal{A}, K)\), which means that if \(f \in \mathcal{M}_\mu(X, \mathcal{A}, F)\) and \(g \in \mathcal{M}(X, \mathcal{A}, K)\), then \(gf \in \mathcal{M}_\mu(X, \mathcal{A}, F)\). The space \(\mathcal{M}_\mu(X, \mathcal{A}, F)\) is also a module over \(\mathcal{M}_\mu(X, \mathcal{A}, K)\).

4. If \(F\) is a normed vector space, and if \(f \in \mathcal{M}_\mu(X, \mathcal{A}, F)\), then \(\|f\| \in \mathcal{M}_\mu(X, \mathcal{A}, \mathbb{R})\).

5. If \(f \in \mathcal{M}_\mu(X, \mathcal{A}, \mathbb{R})\) and \(g \in \mathcal{M}_\mu(X, \mathcal{A}, \mathbb{R})\), then we have \(\sup(f, g), \inf(f, g), f^+, f^-, |f| \in \mathcal{M}_\mu(X, \mathcal{A}, \mathbb{R})\).

The following result gives a characterization of a \(\mu\)-measurable map which shows that a \(\mu\)-measurable map is equal to a measurable map almost everywhere, and that there are strong countability restrictions on its domain and its range.

**Proposition 4.12.** Let \((X, \mathcal{A}, \mu)\) be a measure space, and let \(F\) be any Banach space. A function \(f: X \to F\) is \(\mu\)-mesurable iff there is a null set \(Z\) such that the following three conditions hold:

1. There is a measurable map \(g \in \mathcal{M}(X, \mathcal{A}, F)\) such that \(f\) and \(g\) are equal on \(X - Z\).

2. The function \(f\) vanishes outside of a measurable \(\sigma\)-finite subset of \(X\).

3. The image \(f(X - Z)\) is separable in \(F\), which means that \(f(X - Z)\) contains a countable dense subset.
In particular, if \( \mu \) is \( \sigma \)-finite and if \( F \) is separable, then \( f : X \to F \) is \( \mu \)-measurable iff \( f \) is measurable almost everywhere (there is a null set \( Z \) such that \( f \) agrees with a measurable map on \( X - Z \)).

A proof of Proposition 4.12 can be found in Lang [52] (Chapter VI, Section 1, Property M11). Again, Condition (2) is a measure-theoretic analog of the notion of compact support.

The version of Theorem 4.6 for \( \mu \)-measurable maps is stated below.

**Theorem 4.13.** Let \((X, \mathcal{A}, \mu)\) be a measure space and let \((F, \mathcal{B})\) be a measurable space, where \(F\) is a metric space and \(\mathcal{B}\) is the Borel \(\sigma\)-algebra on \(F\). If \((f_n)_{n \geq 1}\) is a sequence of \(\mu\)-measurable maps \(f_n \in \mathcal{M}_\mu(X, \mathcal{A}, F)\) which converges pointwise to a function \(f : X \to F\), then \(f \in \mathcal{M}_\mu(X, \mathcal{A}, F)\); that is, \(f\) is \(\mu\)-measurable.

A proof of Theorem 4.6 can be found in Lang [52] (Chapter VI, Section 1, Property M12).

We are now ready to construct a very general version of the integral. The original construction was first proposed by Lebesgue, but the more general version presented here applying to functions with values in a Banach space is due to Bochner and Dunford.

### 4.4 The Integral of \(\mu\)-Step Maps

Let \((X, \mathcal{A}, \mu)\) be a measure space and let \((F, \mathcal{B})\) be a measurable space consisting of a Banach space \(F\) and its Borel \(\sigma\)-algebra \(\mathcal{B}\). There is an “obvious” definition of the integral of a \(\mu\)-step map \(f = \sum_{i=1}^{n} y_i \chi_{A_i}\) (where \(y_i \in F\)), namely

\[
I(f) = \int f d\mu = \sum_{i=1}^{n} \mu(A_i)y_i.
\]

Since by definition the \(A_i\) belong to \(\mathcal{A}\) and have finite measure, the linear combination \(\sum_{i=1}^{n} \mu(A_i)y_i\) is a well-defined vector in \(F\). The only problem is that \(I(f)\) seems to depend on the subset \(A\) (and its partition) chosen to express \(f\), but it is easy to show that \(I(f)\) is independent of the representation of \(f\). Then it is easy to show that \(I : \text{Step}_\mu(X, \mathcal{A}, F) \to F\) is a linear map. Furthermore, by Proposition 4.11, we have \(\|f\| \in \text{Step}_\mu(X, \mathcal{A}, \mathbb{R})\), so we can define

\[
N_1(f) = \int \|f\| d\mu,
\]

and we have

\[
\left\| \int f d\mu \right\| \leq \int \|f\| d\mu = N_1(f).
\]

It turns out that \(N_1\) satisfies all the axioms of a norm, except that \(N_1(f) = 0\) does not necessarily imply that \(f = 0\). We say that \(N_1\) is a *semi-norm*, see Definition 15.3. Fortunately,
for any $f \in \text{Step}_\mu(X, \mathcal{A}, F)$, we have $N_1(f) = 0$ iff $f = 0$, except on a subset of measure zero.

We can define the notion of $N_1$-*Cauchy sequence* of a sequence $(f_n)$ of functions $f_n \in \text{Step}_\mu(X, \mathcal{A}, F)$ as follows: for all $\epsilon > 0$, there is some $N > 0$, such that for all $m, n \geq N$, we have $N_1(f_m - f_n) < \epsilon$. We can also define the notion of $N_1$-*convergence* of a sequence $(f_n)$ of functions $f_n \in \text{Step}_\mu(X, \mathcal{A}, F)$ to a limit $f \in \text{Step}_\mu(X, \mathcal{A}, F)$ as follows: for all $\epsilon > 0$, there is some $N > 0$, such that for all $n \geq N$, we have $N_1(f - f_n) < \epsilon$. A convergent $N_1$-sequence does not necessarily have a unique limit, but we will see that any two limits are equal a.e.

The problem is that an $N_1$-Cauchy sequence may not have a limit in $\text{Step}_\mu(X, \mathcal{A}, F)$. Thus we are led to completing $\text{Step}_\mu(X, \mathcal{A}, F)$ with respect to the semi-norm $N_1$. This can be done and we obtain a vector space $\mathcal{L}_\mu(X, \mathcal{A}, F)$ which is a subspace of $\mathcal{M}_\mu(X, \mathcal{A}, F)$. The integral map $I$ and the semi-norm $N_1$ can be extended to $\mathcal{L}_\mu(X, \mathcal{A}, F)$ as a semi-norm denoted $\| \|_1$, the space $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is Cauchy-complete with respect to the semi-norm $\| \|_1$, and $\text{Step}_\mu(X, \mathcal{A}, F)$ is dense in $\mathcal{L}_\mu(X, \mathcal{A}, F)$ with respect to the semi-norm $\| \|_1$.

It also turns out that the subspace $SN$ of $\text{Step}_\mu(X, \mathcal{A}, F)$ consisting of all functions $f$ such that $N_1(f) = 0$ is the set of functions in $\text{Step}_\mu(X, \mathcal{A}, F)$ that are equal to 0 a.e. Similarly, the subspace $N$ of $\mathcal{L}_\mu(X, \mathcal{A}, F)$ consisting of all functions $f$ such that $\| f \|_1 = 0$ is the set of functions in $\mathcal{L}_\mu(X, \mathcal{A}, F)$ that are equal to 0 a.e. Thus, we can form the quotients spaces $\text{Step}_\mu(X, \mathcal{A}, F) = \text{Step}_\mu(X, \mathcal{A}, F)/SN$ and $\mathcal{L}_\mu(X, \mathcal{A}, F) = \mathcal{L}_\mu(X, \mathcal{A}, F)/N$. In $\text{Step}_\mu(X, \mathcal{A}, F)$ and in $\mathcal{L}_\mu(X, \mathcal{A}, F)$ the semi-norm $\| \|_1$ is really a norm, and $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is the completion of $\text{Step}_\mu(X, \mathcal{A}, F)$.

Theoretically, we could define $\mathcal{L}_\mu(X, \mathcal{A}, F)$ directly as the Cauchy completion (see Theorem 15.58 and Theorem 15.68) of $\text{Step}_\mu(X, \mathcal{A}, F)$, but we obtain equivalence classes of Cauchy sequences of equivalence classes of functions in $\text{Step}_\mu(X, \mathcal{A}, F)$, which are not easily interpretable as functions. The same space $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is obtained, see the diagram below.

\[
\text{Step}_\mu(X, \mathcal{A}, F) \xrightarrow{\text{completion}} \mathcal{L}_\mu(X, \mathcal{A}, F) \\
\text{Step}_\mu(X, \mathcal{A}, F) / SN \xrightarrow{\text{completion}} \mathcal{L}_\mu(X, \mathcal{A}, F) / N.
\]

The construction that we alluded to, although involving some extra work, yields a very clear description of these equivalence classes in terms of functions (in $\mathcal{L}_\mu(X, \mathcal{A}, F)$). The completeness of $\mathcal{L}_\mu(X, \mathcal{A}, F)$ (under the $\| \|_1$-norm) is also immediately obtained.

As in the previous section the results that we state without proof are proved either in Marle [56] or in Lang [52].

We now return to the definition of the integral of a $\mu$-step maps.
Proposition 4.14. Let \((X, \mathcal{A}, \mu)\) be a measure space and let \((F, \mathcal{B})\) be a measurable space, with \(F\) a Banach space and \(\mathcal{B}\) its Borel \(\sigma\)-algebra. For any \(\mu\)-step map \(f \in \text{Step}_\mu(X, \mathcal{A}, F)\), for any two partitions \((A_1, \ldots, A_m)\) and \((B_1, \ldots, B_n)\) adapted to \(f\), so that \(f = \sum_{i=1}^{m} y_i \chi_{A_i} = \sum_{j=1}^{n} z_j \chi_{B_j}\), we have
\[
\sum_{i=1}^{m} \mu(A_i) y_i = \sum_{j=1}^{n} \mu(B_j) y_j.
\]

Proposition 4.14 justifies the following definition.

Definition 4.7. Let \((X, \mathcal{A}, \mu)\) be a measure space and let \((F, \mathcal{B})\) be a measurable space, with \(F\) a Banach space and \(\mathcal{B}\) its Borel \(\sigma\)-algebra. For any \(\mu\)-step map \(f \in \text{Step}_\mu(X, \mathcal{A}, F)\), the common value
\[
\int f \, d\mu
\]
of the expression
\[
I(f) = \sum_{i=1}^{n} \mu(A_i) y_i
\]
for any partition \((A_1, \ldots, A_n)\) adapted to \(f\) is called the integral of \(f\) (relative to the measure \(\mu\)).

Recall that if the \(\mu\)-step map \(f\) is expressed as \(f = \sum_{i=1}^{n} y_i \chi_{A_i}\), then the \(\mu\)-set map \(\|f\|\) is expressed as \(\|f\| = \sum_{i=1}^{n} \|y_i\| \chi_{A_i}\). We also define the semi-norm \(N_1(f)\) of \(f\) as
\[
N_1(f) = \int \|f\| \, d\mu = \sum_{i=1}^{n} \mu(A_i) \|y_i\|.
\]

For any measurable subset \(E \in \mathcal{A}\), since \(\chi_E f \in \text{Step}_\mu(X, \mathcal{A}, F)\), we let
\[
\int_E f \, d\mu = \int \chi_E f \, d\mu.
\]

For simplicity of notation, we often write \(\int_E f\) instead of \(\int_E f \, d\mu\), and if \(E = X\), we write \(\int f\) instead of \(\int f \, d\mu\).

We stress that the integral \(\int f \, d\mu\) or \(\int_E f \, d\mu\) is always finite; that is, an element of \(F\), but not \(\infty\). This is in contrast with the approach where the integral of a step function may have the value \(+\infty\), as in Rudin [66] (Chapter 1). At some later stage, in defining the space \(L^1(X, \mathcal{A}, F)\), it is necessary to require the integral to be finite anyway. We find the approach where the integral is finite in the first place less confusing. It also yields a more explicit definition of \(L^1(X, \mathcal{A}, F)\).

\[\text{2This integral is usually called the Lebesgue integral or Bochner integral. A more appropriate name might be the Bochner–Dunford integral.}\]
Example 4.2. The special case in which $X$ is a countable set, $A = 2^X$, $\mu$ is the counting measure defined in Example 3.2, and $F = \mathbb{C}$, is of particular interest. Say $X = \mathbb{N}$. A $\mu$-step function is of the form

$$f = \sum_{i=1}^{n} y_i \chi_{A_i},$$

where $A_i$ must be a finite subset of $\mathbb{N}$, and $y_i \in \mathbb{C}$. By definition of $\mu$, we have $\mu(A_i) = |A_i|$, so

$$\int f \, d\mu = \sum_{i=1}^{n} y_i |A_i|.$$ 

But $f$ is the function with finite support $A = \bigcup_{i=1}^{n} A_i$, such that $f(j) = y_i$ for all $j \in A_i$, and $f(j) = 0$ for all $j \notin A$, so

$$\int f \, d\mu = \sum_{j \in A} f(j) = \sum_{j \in \mathbb{N}} f(j),$$

the sum of the (finite) sequence $(f(j))_{j \in \mathbb{N}}$. Similarly, if $X = \mathbb{Z}$, then for any sequence $(f_j)_{j \in \mathbb{Z}}$ with only finitely many nonzero entries,

$$\int f \, d\mu = \sum_{j \in \mathbb{Z}} f(j).$$

Here are some of the main properties of the integral.

**Proposition 4.15.** Let $(X, A, \mu)$ be a measure space and let $(F, B)$ be a measurable space, with $F$ a Banach space and $B$ its Borel $\sigma$-algebra. The following properties hold:

1. The integral map $\int : \text{Step}_\mu(X, A, F) \to F$ is a linear map.

2. If $A$ and $B$ are any two disjoint measurable subsets, then

$$\int_{A \cup B} f \, d\mu = \int_A f \, d\mu + \int_B f \, d\mu$$

3. For any map $f \in \text{Step}_\mu(X, A, F)$, we have $\|f\| \in \text{Step}_\mu(X, A, \mathbb{R})$, and

$$\left\| \int f \, d\mu \right\| \leq \int \|f\| \, d\mu = N_1(f).$$

We also have

$$\int \|f\| \, d\mu \leq \mu(\{x \in X \mid f(x) \neq 0\}) \|f\|_{\infty}.$$ 

4. For any two maps $f, g \in \text{Step}_\mu(X, A, F)$, if $f = g$ a.e., then $\int f \, d\mu = \int g \, d\mu$. 

5. For any two maps $f, g \in \text{Step}_\mu(X, \mathcal{A}, \mathbb{R})$, if $f \leq g$ a.e., then $\int f \, d\mu \leq \int g \, d\mu$. In particular, if $f \geq 0$ a.e., then $\int f \, d\mu \geq 0$.

6. $N_1$ is a semi-norm on $\text{Step}_\mu(X, \mathcal{A}, F)$. Furthermore, for any $f \in \text{Step}_\mu(X, \mathcal{A}, F)$, we have $N_1(f) = 0$ iff $f = 0$, except on a subset of measure zero.

7. If $F_1$ and $F_2$ are two Banach spaces over $\mathbb{R}$ or $\mathbb{C}$, and if $f : F_1 \to F_2$ is a continuous linear map, then for any $f \in \text{Step}_\mu(X, \mathcal{A}, F_1)$, we have $h \circ f \in \text{Step}_\mu(X, \mathcal{A}, F_2)$, and

$$\int (h \circ f) \, d\mu = h \left( \int f \, d\mu \right)$$

If $F = \mathbb{C}$, the above property holds for any semi-linear map.

Proof. We prove (3) and (6), leaving the other properties as exercises.

(3) If $f = \sum_{i=1}^n y_i \chi_{A_i}$ then $\int f \, d\mu = \sum_{i=1}^n \mu(A_i) y_i$. We also have $\|f\| = \sum_{i=1}^n \|y_i\| \chi_{A_i}$ and $\int \|f\| \, d\mu = \sum_{i=1}^n \mu(A_i) \|y_i\|$. It follows that

$$\left\| \int f \, d\mu \right\| = \left\| \sum_{i=1}^n \mu(A_i) y_i \right\| \leq \sum_{i=1}^n \|\mu(A_i) y_i\| = \sum_{i=1}^n \mu(A_i) \|y_i\| = \int \|f\| \, d\mu.$$

We also have

$$\int \|f\| \, d\mu = \sum_{i=1}^n \mu(A_i) \|y_i\| \leq \mu(\{x \in X \mid f(x) \neq 0\}) \max_{1 \leq i \leq n} \|y_i\|$$

$$= \mu(\{x \in X \mid f(x) \neq 0\}) \|f\|_\infty.$$

(6) Since by (1) the integral is linear, we have

$$N_1(\lambda f) = \int \|\lambda f\| \, d\mu = \int |\lambda| \|f\| \, d\mu = |\lambda| \int \|f\| \, d\mu = |\lambda| N_1(f).$$

Since $\|(f + g)(x)\| \leq \|f(x)\| + \|g(x)\|$ for all $x \in X$, by (5) we have

$$N_1(f + g) = \int \|f + g\| \, d\mu \leq \int \|f\| \, d\mu + \int \|g\| \, d\mu = N_1(f) + N_1(g).$$

Assume that $N_1(f) = 0$, which means that $\int \|f\| \, d\mu = 0$. Since $f$ is a $\mu$-step function, we can write

$$f = \sum_{i=1}^n y_i \chi_{A_i},$$

for a finite sequence $(A_1, \ldots, A_n)$ of nonempty pairwise disjoint subsets $A_i \in \mathcal{A}$ of finite measure. Since

$$\|f\| = \sum_{i=1}^n \|y_i\| \chi_{A_i},$$
4.4. THE INTEGRAL OF $\mu$-STEP MAPS

so

$$N_1(f) = \int \|f\| \, d\mu = \sum_{i=1}^{n} \|y_i\| \mu(A_i) = 0.$$ 

Since $\|y_i\| \geq 0$ and $\mu(A_i) \geq 0$, the following must hold:

- if $\mu(A_i) \neq 0$, then $\|y_i\| = 0$, that is, $y_i = 0$.
- if $y_i \neq 0$, that is, $\|y_i\| \neq 0$, then $\mu(A_i) = 0$.

Consequently

$$\{ x \in X \mid f(x) \neq 0 \} = \bigcup_{i \in I} A_i, \quad \text{with} \quad I = \{ i \mid 1 \leq i \leq n \mid y_i \neq 0 \},$$

where $\bigcup_{i \in I} A_i \in \mathcal{A}$ is a set of measure 0, since $i \in I$ implies that $\mu(A_i) = 0$.

By Proposition 4.15(6), the set

$$SN = \{ f \in \text{Step}_\mu(X, \mathcal{A}, F) \mid N_1(f) = 0 \} = \{ f \in \text{Step}_\mu(X, \mathcal{A}, F) \mid f = 0 \text{ a.e.} \}$$

is a subspace of $\text{Step}_\mu(X, \mathcal{A}, F)$.

**Definition 4.8.** Let $\text{Step}_\mu(X, \mathcal{A}, F)$ be the quotient space $\text{Step}_\mu(X, \mathcal{A}, F)/SN$.

For every equivalence class $f \in \text{Step}_\mu(X, \mathcal{A}, F)$, we can define

$$\int f d\mu = \int f d\mu$$

for any function $f \in \text{Step}_\mu(X, \mathcal{A}, F)$ in the equivalence class of $f$, because if $f = g$ a.e., then $\int f d\mu = \int g d\mu$, so $\int f d\mu$ does not depend on the representative chosen in the equivalence class $f$. Similarly, we define $N_1(f)$ by

$$N_1(f) = N_1(f) = \int \|f\| \, d\mu,$$

for any function $f \in \text{Step}_\mu(X, \mathcal{A}, F)$ in the equivalence class of $f$. Again if $f = g$ a.e., then $\|f\| = \|g\|$ a.e., so $N_1(f) = N_1(g)$, which means that $N_1(f)$ is well defined. It is immediately verified that $N_1$ is a semi-norm, and in fact a norm, since $N_1(f) = 0$ iff $N_1(f) = 0$ for any representative $f \in \text{Step}_\mu(X, \mathcal{A}, F)$ in the equivalence class $f$ iff $f = 0$ a.e., which means that $f = 0$. Therefore, $(\text{Step}_\mu(X, \mathcal{A}, F), N_1)$ is a normed vector space. It is easy to see that the inequality

$$\left\| \int f d\mu \right\| \leq \int \|f\| \, d\mu = N_1(f)$$

holds, which shows that the map $\int : \text{Step}_\mu(X, \mathcal{A}, F) \to F$ is continuous (in fact, uniformly continuous). The space $(\text{Step}_\mu(X, \mathcal{A}, F), N_1)$ is not complete, so we can apply Theorem
15.68) to form its completion \(L_\mu(X, A, F)\) and extend the map \(\int\) to it. Theoretically we have achieved our goal of defining a notion of integral on a normed vector space \(L_\mu(X, A, F)\) which is complete and in which \(\text{Step}_\mu(X, A, F)\) is dense, but the elements in this abstract completion are equivalence classes of Cauchy sequences, and are not easily identifiable with functions.

We will follow a different path, still very much inspired by the completion method involving Cauchy sequences, the twist being that we consider Cauchy sequences whose limit is known ahead of time, but where we use pointwise convergence \emph{almost everywhere}, instead of pointwise convergence.

### 4.5 Integrable Functions; the Spaces \(L_\mu(X, A, F)\) and \(\text{L}_\mu(X, A, F)\)

In this section we construct the completion \(L_\mu(X, A, F)\) of the vector space \(\text{Step}_\mu(X, A, F)\) equipped with the semi-norm \(N_1\), and construct the integral of a function in \(L_\mu(X, A, F)\).

The semi-norm \(N_1\) is extended to \(L_\mu(X, A, F)\) as a semi-norm \(\|\cdot\|_1\) called the \(L^1\)-semi-norm, and we find that the space of functions such that \(\|f\|_1 = 0\) is the set \(\mathcal{N}\) of functions in \(L_\mu(X, A, F)\) that are zero a.e. Then we define the quotient space \(L_\mu(X, A, F) = L_\mu(X, A, F)/\mathcal{N}\). The space \(L_\mu(X, A, F)\) is the completion of \(\text{Step}_\mu(X, A, F)\); this is one of the most important results of this section (the Fischer–Riesz theorem).

As in the previous section the results that we state without proof are proved either in Marle [56] or in Lang [52].

Recall the following definitions.

**Definition 4.9.** A sequence \((f_n)\) of functions \(f_n \in \text{Step}_\mu(X, A, F)\) is a \(N_1\)-Cauchy sequence if for every \(\epsilon > 0\), there is some \(N > 0\), such that for all \(m, n \geq N\), we have \(N_1(f_m - f_n) < \epsilon\), where \(N_1(f_m - f_n) = \int \|f_m - f_n\| d\mu\). A sequence \((f_n)\) of maps \(f_n \in \text{Step}_\mu(X, A, F)\) converges pointwise \emph{almost everywhere} to a limit \(f : X \to F\) if there is a null set \(Z\) such that for every \(x \in X - Z\), for every \(\epsilon > 0\), there is some \(N > 0\), such that \(\|f(x) - f_n(x)\| < \epsilon\) for all \(n \geq N\).

We define the space \(L_\mu(X, A, F)\) as follows.

**Definition 4.10.** Let \((X, A, \mu)\) be a measure space and let \((F, \mathcal{B})\) be a measurable space, with \(F\) a Banach space and \(\mathcal{B}\) its Borel \(\sigma\)-algebra. The set \(L_\mu(X, A, F)\) of \(\mu\)-integrable functions consists of all functions \(f : X \to F\) such that there is some \(N_1\)-Cauchy sequence \((f_n)_{n \geq 1}\) of \(\mu\)-step maps \(f_n \in \text{Step}_\mu(X, A, F)\) which converges pointwise almost everywhere to \(f\). A sequence \((f_n)_{n \geq 1}\) of \(\mu\)-step maps as above is called an approximation sequence for \(f\).

Observe that not only do we require that the sequence \((f_n)_{n \geq 1}\) converges pointwise to \(f\) a.e., which makes \(f\) a \(\mu\)-measurable map, but also that this sequence is \(N_1\)-Cauchy. This is
4.5. INTEGRABLE FUNCTIONS; THE SPACES $\mathcal{L}_\mu(X, \mathcal{A}, F)$ AND $L_\mu(X, \mathcal{A}, F)$

the key to defining the notion of integral of the function $f$, as shown technically in Proposition 4.16.

We will see that $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is a vector space containing $\text{Step}_\mu(X, \mathcal{A}, F)$, and a subspace of $\mathcal{M}_\mu(X, \mathcal{A}, F)$. Also, and this is the point of the construction, $\mathcal{L}_\mu(X, \mathcal{A}, F)$ is complete with respect to the extension $\| \cdot \|_1$ of the semi-norm $N_1$ to $\mathcal{L}_\mu(X, \mathcal{A}, F)$, a fact that is not obvious at all from the definition.

The crucial point is that Definition 4.10 is designed so that the following fact holds.

**Proposition 4.16.** For any $N_1$-Cauchy sequence $(f_n)_{n \geq 1}$ of $\mu$-step maps, the sequence of integrals $(\int f_n \, d\mu)_{n \geq 1}$ is a Cauchy sequence in $F$.

**Proof.** Indeed, by Proposition 4.15(3), we have

$$\left\| \int f_n \, d\mu - \int f_m \, d\mu \right\| = \left\| (f_n - f_m) \, d\mu \right\| \leq \int \| f_n - f_m \| \, d\mu = N_1(f_n - f_m),$$

and since by hypothesis $(f_n)$ is an $N_1$-Cauchy sequence, the sequence $(\int f_n \, d\mu)_{n \geq 1}$ is a Cauchy sequence in $F$. Indeed, for every $\epsilon > 0$, since the sequence $(f_n)$ is $N_1$-Cauchy, there is some $N > 0$ such that $N_1(f_n - f_m) < \epsilon$ for all $m, n \geq N$, which implies that $\left\| \int f_n \, d\mu - \int f_m \, d\mu \right\| < \epsilon$ for all $m, n \geq N$.

Then, since $F$ is complete, the sequence $(\int f_n \, d\mu)_{n \geq 1}$ converges to an element of $f$, and if $(f_n)_{n \geq 1}$ is an approximation sequence for $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, it is natural to define the integral of $f$ as

$$\int f \, d\mu = \lim_{n \to \infty} \int f_n \, d\mu.$$

The problem is that the definition of $\int f \, d\mu$ depends on the approximation sequence $(f_n)_{n \geq 1}$ chosen for $f$.

Actually, the definition of $\int f \, d\mu$ does not depend on the approximation approximation sequence $(f_n)_{n \geq 1}$ chosen for $f$, but proving this is nontrivial. The proof relies on a remarkable fact called the fundamental lemma of integration by Serge Lang; see [52], Chapter VI, §3.

**Proposition 4.17.** Let $(f_n)_{n \geq 1}$ be any $N_1$-Cauchy sequence of maps $f_n \in \text{Step}_\mu(X, \mathcal{A}, F)$. There exists a subsequence $(f_{n_i})$ which converges pointwise almost everywhere to a limit $f : X \to F$. Furthermore, for any $\epsilon > 0$, there is a measurable subset $Z \in \mathcal{A}$ such that $\mu(Z) \leq \epsilon$, and the subsequence $(f_{n_i})$ converges uniformly to $f$ on $X - Z$ (recall Definition 1.6).

The proof of Proposition 4.17 can be found in Lang [52] (Chapter VI, §3, Lemma 3.1). It should be mentioned that in general, the original sequence $(f_n)$ may not converge pointwise, even a.e. An example of such a sequence $(f_n)$ which is $N_1$-Cauchy, yet $(f_n(x))$ diverges for every $x \in X$, is given in Schwartz [72] (Chapter 5, §6).

Using Proposition 4.17, the following result can be shown. This result implies that the integral $\int f \, d\mu$ is well defined.
CHAPTER 4. INTEGRATION

Proposition 4.18. Let \((f_n)_{n \geq 1}\) and \((g_n)_{n \geq 1}\) be two \(N_1\)-Cauchy sequences of \(\mu\)-step maps \(f_n, g_n \in \text{Step}_\mu(X, \mathcal{A}, F)\) which approximate the same function \(f\). The sequences \((\int f_n d\mu)_{n \geq 1}\) and \((\int g_n d\mu)_{n \geq 1}\) converge to the same limit, and
\[
\lim_{n \to \infty} \int \|f_n - g_n\| d\mu = 0,
\]
that is, \(\lim_{n \to \infty} N_1(f_n - g_n) = 0\).

The proof of Proposition 4.18 can be found in Lang [52] (Chapter VI, §3, Lemma 3.2).

Proposition 4.18 justifies the following definition.

Definition 4.11. Let \((X, \mathcal{A}, \mu)\) be a measure space and let \((F, \mathcal{B})\) be a measurable space, with \(F\) a Banach space and \(\mathcal{B}\) its Borel \(\sigma\)-algebra. For any function \(f \in L_\mu(X, \mathcal{A}, F)\), we define the integral\(^3\) of \(f\) (with respect to \(\mu\)) by
\[
\int f d\mu = \lim_{n \to \infty} \int f_n d\mu,
\]
where \((f_n)_{n \geq 1}\) is any approximation sequence of \(f\) by \(\mu\)-step maps.

Proposition 4.19. For any function \(f \in L_\mu(X, \mathcal{A}, F)\) and any approximation sequence \((f_n)\) of \(f\) with \(f_n \in \text{Step}_\mu(X, \mathcal{A}, F)\), we have \(\|f\| \in L_\mu(X, \mathcal{A}, \mathbb{R})\), and the sequence \((\|f_n\|)\) is an approximation sequence of \(\|f\|\) with \(\|f_n\| \in \text{Step}_\mu(X, \mathcal{A}, \mathbb{R})\). Furthermore,
\[
\int \|f\| d\mu = \lim_{n \to \infty} \int \|f_n\| d\mu = \lim_{n \to \infty} N_1(f_n).
\]

Proof. Since the sequence \((f_n)\) converges pointwise to \(f\) a.e., we verify immediately that the sequence \((\|f_n\|)\) converges pointwise to \(\|f\|\) a.e. Since
\[
\|\|f_n\| - \|f_m\|\| \leq \|f_n - f_m\|
\]
(see just after Definition 15.3), by Proposition 4.15(5) we have
\[
N_1(\|f_n\| - \|f_m\|) = \int \|f_n\| - \|f_m\| d\mu \leq \int \|f_n - f_m\| d\mu = N_1(f_n - f_m),
\]
and since \((f_n)\) is an \(N_1\)-Cauchy sequence, the sequence \((\|f_n\|)\) is an \(N_1\)-Cauchy sequence. Therefore \(\|f\| \in L_\mu(X, \mathcal{A}, \mathbb{R})\), and \((\|f_n\|)\) is an approximation sequence of \(\|f\|\). By definition of the integral,
\[
\int \|f\| d\mu = \lim_{n \to \infty} \int \|f_n\| d\mu = \lim_{n \to \infty} N_1(f_n),
\]
as claimed.

\(^3\)This integral is usually called the Lebesgue integral or Bochner integral.
Definition 4.12. For any function \( f \in \mathcal{L}_\mu(X, \mathcal{A}, F) \), we define the \( L^1 \)-semi-norm \( \| f \|_1 \) of \( f \) as
\[
\| f \|_1 = \int \| f \| d\mu.
\]

Observe that if \( f \in \text{Step}_\mu(X, \mathcal{A}, F) \), then \( \| f \|_1 = N_1(f) \). The following proposition is easily shown by passing to the limit.

Proposition 4.20. The set \( \mathcal{L}_\mu(X, \mathcal{A}, F) \) is a vector space, and \( \| \|_1 \) is a semi-norm on \( \mathcal{L}_\mu(X, \mathcal{A}, F) \). The space \( \text{Step}_\mu(X, \mathcal{A}, F) \) is a subspace of \( \mathcal{L}_\mu(X, \mathcal{A}, F) \), which is a subspace of \( \mathcal{M}_\mu(X, \mathcal{A}, F) \).

We are almost ready to prove that \( \mathcal{L}_\mu(X, \mathcal{A}, F) \) is complete with respect to the \( L^1 \)-semi-norm, but first we need the following result.

Proposition 4.21. The subspace \( \text{Step}_\mu(X, \mathcal{A}, F) \) is dense in \( \mathcal{L}_\mu(X, \mathcal{A}, F) \) with respect to the \( L^1 \)-semi-norm \( \| \|_1 \). Furthermore, any approximation sequence \( (f_n)_{n \geq 1} \) of \( f \) by \( \mu \)-step maps converges to \( f \) according to the semi-norm \( \| \|_1 \).

Proof. Pick any \( f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R}) \) and let \( (f_n) \) be any approximation sequence for \( f \). This means that the sequence \( (f_n) \) is a \( N_1 \)-Cauchy sequence of \( \mu \)-step maps which converges pointwise to \( f \) a.e. We will prove that
\[
\lim_{n \to \infty} \| f - f_n \|_1 = 0,
\]
which shows that the sequence \( (f_n) \) converges to \( f \) in the \( L^1 \)-semi-norm.

First, we claim that for any fixed \( n \geq 1 \), the sequence \( (\| f_p - f_n \|)_{p \geq 1} \) is an \( N_1 \)-Cauchy sequence which converges to \( \| f - f_n \| \) a.e. Indeed, we have
\[
\int \| f_p - f_n \| - \| f_q - f_n \| \, d\mu \leq \int \| f_p - f_n - (f_q - f_n) \| \, d\mu = \int \| f_p - f_q \| \, d\mu = N_1(f_p - f_q),
\]
and since \( (f_n) \) is a \( N_1 \)-Cauchy sequence, for every \( \epsilon > 0 \), there is some \( N > 0 \) such that \( N_1(f_p - f_q) < \epsilon \) for all \( p, q \geq N \), which shows that \( (\| f_p - f_n \|)_{p \geq 1} \) is a \( N_1 \)-Cauchy sequence (in \( \mathbb{R} \)). The fact that \( (f_p)_{p \geq 1} \) converges pointwise a.e. to \( f \) immediately implies that \( \| f_p - f_n \| \) converges to \( \| f - f_n \| \) a.e. By definition of \( \| \|_1 \) and of the integral
\[
\| f - f_n \|_1 = \int \| f - f_n \| \, d\mu = \lim_{p \to \infty} \int \| f_p - f_n \| \, d\mu = \lim_{p \to \infty} N_1(f_p - f_n).
\]
Thus for every \( \epsilon > 0 \), there is some \( M_1 > 0 \) such that
\[
\| f - f_n \|_1 - N_1(f_p - f_n) < \frac{\epsilon}{2} \text{ for all } p \geq M_1,
\]
and since \((f_n)\) is an \(N_1\)-Cauchy sequence, there is some \(M_2 > 0\) such that
\[
N_1(f_p - f_n) < \frac{\epsilon}{2} \quad \text{for all } n, p \geq M_2,
\]
so for all \(n, p \geq \max(M_1, M_2)\) we have
\[
\|f - f_n\|_1 \leq \|f - f_n\|_1 - N_1(f_p - f_n) + N_1(f_p - f_n) < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,
\]
which proves that \(\lim_{n \to \infty} \|f - f_n\|_1 = 0\); that is, the sequence \((f_n)\) converges to \(f\) in the \(L^1\)-semi-norm.

\[\square\]

**Remark:** It appears that Lang [52] skipped this step, which is used in the proof his Theorem 3.4, and the proof of the next theorem.

Now we can prove one of our main theorems.

**Theorem 4.22.** (Fischer–Riesz) The space \(L_\mu(X, A, F)\) is complete with respect to the \(L^1\)-semi-norm. This means that for every sequence \((f_n)_{n \geq 1}\) of functions \(f_n \in L_\mu(X, A, F)\), if \((f_n)\) is \(\|\cdot\|_1\)-Cauchy, then there is some function \(f \in L_\mu(X, A, F)\) such that for every \(\epsilon > 0\), there is some \(N > 0\) such that \(\|f - f_n\|_1 < \epsilon\) for all \(n \geq N\).

**Proof.** Let \((f_n)_{n \geq 1}\) be an \(\|\cdot\|_1\)-Cauchy sequence of functions \(f_n \in L_\mu(X, A, F)\). By Proposition 4.21, for every \(n\) there is an approximation sequence \((g_{n,m})_{m \geq 1}\) of \(\mu\)-step map that converges to \(f_n\) pointwise a.e., and in the \(\|\cdot\|_1\)-semi-norm. Thus, for every \(n \geq 1\), there is some \(m(n)\) such that
\[
\|f_n - g_{n,m(n)}\|_1 \leq \frac{1}{n}.
\]
Each sequence \((g_{n,m(n)})_{n \geq 1}\) is \(N_1\)-Cauchy, because
\[
N_1(g_{p,m(p)} - g_{q,m(q)}) = \|g_{p,m(p)} - g_{q,m(q)}\|_1 \\
\leq \|g_{p,m(p)} - f_p\|_1 + \|f_p - f_q\|_1 + \|f_q - g_{q,m(q)}\|_1 \\
\leq \frac{1}{p} + \frac{1}{q} + \|f_p - f_q\|_1,
\]
and the right-hand side tends to 0 when \(p\) and \(q\) tend to \(+\infty\), since the sequence \((f_n)\) is \(\|\cdot\|_1\)-Cauchy. By Proposition 4.17, for each sequence \((g_{n,m(n)})_{m \geq 1}\), we can extract a subsequence \((g_{n_k,m(n_k)})_{k \geq 1}\) that converges pointwise a.e. to some function \(f \in L_\mu(X, A, F)\), and is also \(N_1\)-Cauchy. By the second part of Proposition 4.21, the subsequence \((g_{n_k,m(n_k)})_{k \geq 1}\) converges to \(f\) for the semi-norm \(\|\cdot\|_1\). Since \((g_{n,m(n)})_{n \geq 1}\) is \(N_1\)-Cauchy and has a subsequence \((g_{n_k,m(n_k)})_{k \geq 1}\) \(\|\cdot\|_1\)-convergent to \(f\), it also \(\|\cdot\|_1\)-converges to the function \(f\). Using the inequality
\[
\|f - f_n\|_1 \leq \|f - g_{n,m(n)}\|_1 + \|g_{n,m(n)} - f_n\|_1 \\
\leq \|f - g_{n,m(n)}\|_1 + \frac{1}{n},
\]
we can extract a subsequence \((g_{n_k,m(n_k)})_{k \geq 1}\) that converges pointwise a.e. to \(f\) in the \(L^1\)-semi-norm.
and since the sequence \((g_{n,m(n)})_{n \geq 1}\) converges to the function \(f\), we deduce that the sequence \((f_n)_{n \geq 1}\) converges to \(f\) for the semi-norm \(\|\cdot\|_1\).

In the diagram below, the original sequence \((f_n)_{n \geq 1}\) is shown as the top horizontal row. Below each \(f_n\), we have the approximation sequence \((g_{n,m})_{m \geq 1}\) shown as an ascending column. The sequence of \(g_{n,m(n)}\) chosen for each \(n\) is shown in boldface, and its subsequence in red.

\[
\begin{array}{ccccccccc}
  f_1 & f_2 & f_3 & f_4 & f_5 & f_6 & \cdots & f_n & \cdots \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\
  g_{1,m(n)} & g_{2,m(n)} & g_{3,m(n)} & g_{4,m(n)} & g_{5,m(n)} & g_{6,m(n)} & \cdots & g_{n,m(n)} & \cdots \\
  \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \ddots \\
  g_{1,6} & g_{2,6} & g_{3,6} & g_{4,6} & g_{5,6} & g_{6,6} & \cdots & g_{n,6} & \cdots \\
  g_{1,5} & g_{2,5} & g_{3,5} & g_{4,5} & g_{5,5} & g_{6,5} & \cdots & g_{n,5} & \cdots \\
  g_{1,4} & g_{2,4} & g_{3,4} & g_{4,4} & g_{5,4} & g_{6,4} & \cdots & g_{n,4} & \cdots \\
  g_{1,3} & g_{2,3} & g_{3,3} & g_{4,3} & g_{5,3} & g_{6,3} & \cdots & g_{n,3} & \cdots \\
  g_{1,2} & g_{2,2} & g_{3,2} & g_{4,2} & g_{5,2} & g_{6,2} & \cdots & g_{n,2} & \cdots \\
  g_{1,1} & g_{2,1} & g_{3,1} & g_{4,1} & g_{5,1} & g_{6,1} & \cdots & g_{n,1} & \cdots \\
\end{array}
\]

This concludes the proof.

The following properties of the integral are easily obtained by passing to the limit.

**Proposition 4.23.** Let \((X,A,\mu)\) be a measure space and let \((F,B)\) be a measurable space, with \(F\) a Banach space and \(B\) its Borel \(\sigma\)-algebra. The following properties hold:

1. For any \(f \in \mathcal{L}_\mu(X,A,F)\), if \(f = 0\) a.e., then \(\int f \, d\mu = 0\). More generally, if \(f,g \in \mathcal{L}_\mu(X,A,F)\) and if \(f = g\) a.e., then \(\int f \, d\mu = \int g \, d\mu\).

2. For any \(f \in \mathcal{L}_\mu(X,A,F)\), and for any measurable subset \(A \in A\), the integral \(\int_A f \, d\mu = \int f \chi_A \, d\mu\) exists, and

\[
\left\| \int_A f \, d\mu \right\| \leq \int_A \|f\| \, d\mu \leq \|f\|_\infty \mu(A).
\]

Furthermore, if \(A,B \in A\) are disjoint, then

\[
\int_{A \cup B} f \, d\mu = \int_A f \, d\mu + \int_B f \, d\mu.
\]

3. The integral \(\int : \mathcal{L}_\mu(X,A,F) \to F\) is linear.

4. For any \(f \in \mathcal{L}_\mu(X,A,F)\), we have \(\|f\| \in \mathcal{L}_\mu(X,A,\mathbb{R})\), and

\[
\left\| \int f \, d\mu \right\| \leq \int \|f\| \, d\mu = \|f\|_1.
\]
5. If \( f, g \in \mathcal{L}_\mu(X,\mathcal{A},\mathbb{R}) \), then \( \sup(f,g), \inf(f,g), f^+, f^-, |f| \in \mathcal{L}_\mu(X,\mathcal{A},\mathbb{R}) \). Since \( f^+ = (|f| + f)/2 \) and \( f^- = (|f| - f)/2 \), we have \( f \in \mathcal{L}_\mu(X,\mathcal{A},\mathbb{R}) \) iff \( f^+ \in \mathcal{L}_\mu(X,\mathcal{A},\mathbb{R}) \) and \( f^- \in \mathcal{L}_\mu(X,\mathcal{A},\mathbb{R}) \).

6. If \( f, g \in \mathcal{L}_\mu(X,\mathcal{A},\mathbb{R}) \) and \( f \leq g \) a.e., then \( \int f \, d\mu \leq \int g \, d\mu \). In particular, if \( f \geq 0 \) a.e., then \( \int f \, d\mu \geq 0 \).

7. Let \( F_1 \) and \( F_2 \) be two Banach spaces, and let \( h: F_1 \to F_2 \) be a linear map (or semi-linear map when the field is \( \mathbb{C} \)). If \( f \in \mathcal{L}_\mu(X,\mathcal{A},F_1) \), then \( h \circ f \in \mathcal{L}_\mu(X,\mathcal{A},F_2) \), and

\[
\int (h \circ f) \, d\mu = h \left( \int f \, d\mu \right).
\]

8. Let \( F_1 \) and \( F_2 \) be two Banach spaces, and let \( F_1 \times F_2 \) be the product space (under any of the product norms defined just before Definition 15.13). Then there is an isomorphism between \( \mathcal{L}_\mu(X,\mathcal{A},F_1 \times F_2) \) and \( \mathcal{L}_\mu(X,\mathcal{A},F_1) \times \mathcal{L}_\mu(X,\mathcal{A},F_2) \), and if \( f = (f_1, f_2) \), then

\[
\int f \, d\mu = \left( \int f_1 \, d\mu, \int f_2 \, d\mu \right).
\]

In particular, since \( \mathbb{C} \) is isomorphic to \( \mathbb{R} \times \mathbb{R} \), a function \( f \in \mathcal{L}_\mu(X,\mathcal{A},\mathbb{C}) \) corresponds uniquely to a function \( f = u + iv \) with \( u, v \in \mathcal{L}_\mu(X,\mathcal{A},\mathbb{R}) \), and we have

\[
\int f \, d\mu = \int u \, d\mu + i \int v \, d\mu.
\]

Remark: Observe that in our approach, if \( f \) is a real-valued function or a complex-valued function, the integral \( \int f \, d\mu \) is defined directly. There is another approach in which the integral is first defined for real-valued positive functions. Then the integral of a real-valued function \( f \) is defined in terms of the integrals of \( f^+ \) and \( f^- \), and the integral of a complex valued function \( f = u + iv \) is defined in terms of the integrals of \( u^+, u^-, v^+, v^- \). See Rudin [66], Definition 1.31.

The next step is to identify the functions \( f \in \mathcal{L}_\mu(X,\mathcal{A},\mathbb{R}) \) such that \( \|f\|_1 = 0 \). For this, we need two propositions.

Proposition 4.24. For any function \( f \in \mathcal{L}_\mu(X,\mathcal{A},F) \), and for any real \( a > 0 \), the subset \( E_a = \{ x \in X \mid \|f(x)\| \geq a \} \) can be written as \( E_a = (B - Z) \cup N \), with \( B \) a measurable subset of finite measure, and \( Z \) and \( N \) two null subsets. The function \( f \) vanishes outside of a \( \sigma \)-finite measurable set.

Proof. By Proposition 4.17, there is an \( N_1 \)-Cauchy sequence \((f_n)\) of \( \mu \)-step maps which converges pointwise to \( f \) a.e., and for every \( \epsilon > 0 \), there is a measurable subset \( Z_1 \) of measure \( \mu(Z_1) < \epsilon \) such that \( f_n \) converges uniformly to \( f \) on \( X - Z_1 \). Pick \( \epsilon' \) such that
0 < \epsilon' < a. The uniform convergence implies that there is some $M > 0$ such that for all $n \geq M$ and all $x \in X - Z_1$,

$$\|f(x) - f_n(x)\| \leq \epsilon',$$

thus

$$\|f(x)\| \leq \|f_n(x)\| + \epsilon',$$

which implies that

$$E_a \subseteq Z_1 \cup \{x \in X \mid \|f_n(x)\| \geq a - \epsilon'\},$$

where both sets on the right-hand side have finite measure (the second one because $f_n$ is a $\mu$-step function, and so is $\|f_n\|$, and a $\mu$-step function vanishes outside of a set of finite measure). Since the function $f$ is $\mu$-measurable, by Proposition 4.12(1), it is equal a.e. to a measurable function $g$, so there is a null set such $Z$ that $\|f\|(x) = g(x)$ for all $x \in X - Z$. Then we have

$$E_a = \{x \in X \mid \|f(x)\| \geq a\} = \{x \in X - Z \mid \|f(x)\| \geq a\} \cup \{x \in Z \mid \|f(x)\| \geq a\} = \{x \in X \mid g(x) \geq a\} \cup \{x \in Z \mid \|f(x)\| \geq a\} = (\{x \in X \mid g(x) \geq a\} - Z) \cup \{x \in Z \mid \|f(x)\| \geq a\} = (B - Z) \cup N,$$

with $B = \{x \in X \mid g(x) \geq a\}$ and $N = \{x \in Z \mid \|f(x)\| \geq a\}$. Since $g$ is measurable and $[a, \infty)$ is closed, $B$ is measurable, and $N$ as a subset of a null set is a null set.

What we showed above with $\|f\|$ replaced by $g$ measurable implies that $B$ has finite measure. The second statement of the proposition follows from Proposition 4.12(2). \qed

Proposition 4.17 can be promoted to $L_\mu(X, \mathcal{A}, F)$ as follows.

**Proposition 4.25.** Let $(f_n)_{n \geq 1}$ be any $\|\cdot\|_1$-Cauchy sequence of maps $f_n \in L_\mu(X, \mathcal{A}, F)$ that converges to some function $f \in L_\mu(X, \mathcal{A}, F)$ in the semi-norm $\|\cdot\|_1$. There exists a subsequence $(f_{n_i})$ which converges pointwise almost everywhere to $f$. Furthermore, for any $\epsilon > 0$, there is a measurable subset $Z \in \mathcal{A}$ such that $\mu(Z) \leq \epsilon$, and the subsequence $(f_{n_i})$ converges uniformly to $f$ on $X - Z$ (recall Definition 1.6).

Here are some corollaries of Proposition 4.25.

**Proposition 4.26.** For any function $f \in L_\mu(X, \mathcal{A}, F)$, we have $\|f\|_1 = 0$ iff $f = 0$ a.e.

**Proof.** If $f = 0$ a.e., then $\|f\| = 0$ a.e., and by Proposition 4.23(1), we have $\|f\|_1 = 0$. Conversely, the sequence $(f_n)$ where $f_n$ is the zero function is $\|\cdot\|_1$-Cauchy and converges to $f$ in the $\|\cdot\|_1$-norm. By Proposition 4.25 there is a subsequence that converges pointwise a.e. to $f$. But since $f_n$ is the zero function for all $n$, this subsequence also converges pointwise a.e. to the zero function, so $f = 0$ a.e. \qed
Proposition 4.26 is the second main important result of this section because it provides a very natural characterization of the functions $f$ such that $\|f\|_1 = 0$.

**Proposition 4.27.** Let $(f_n)$ be a sequence of functions $f_n \in L_\mu(X, A, F)$. If $(f_n)$ is an $\|\|_1$-Cauchy sequence which converges pointwise a.e. to a function $f : X \to F$, then $f \in L_\mu(X, A, F)$, and $(f_n)$ converges to $f$ in the semi-norm $\|\|_1$.

**Proof.** Since the sequence $(f_n)$ is an $\|\|_1$-Cauchy sequence, by the Fischer–Riesz theorem (Theorem 4.22) it converges to some function $g \in L_\mu(X, A, F)$ in the $\|\|_1$-semi-norm. By Proposition 4.25, some subsequence $(f_{n_k})_{k \geq 1}$ of $(f_n)$ converges pointwise a.e. to $g$. Since $(f_n)$ converges pointwise a.e. to $f$, the subsequence $(f_{n_k})_{k \geq 1}$ also converges pointwise a.e. to $f$, so $f = g$ a.e., and since $g \in L_\mu(X, A, F)$ and $(f_n)$ converges to $g$ in the semi-norm $\|\|_1$, we also have $f \in L_\mu(X, A, F)$, and $(f_n)$ converges to $f$ in the semi-norm $\|\|_1$. \qed

The main disadvantage of the space $L_\mu(X, A, F)$ is that it is not a normed vector space under the semi-norm $\|\|_1$. Thus it is natural to consider the quotient of $L_\mu(X, A, F)$ by the subspace $\mathcal{N}$ consisting of the functions such that $\|f\|_1 = 0$.

**Definition 4.13.** Let $\mathcal{N}$ be the subspace of $L_\mu(X, A, F)$ given by

$$\mathcal{N} = \{ f \in L_\mu(X, A, F) \mid \|f\|_1 = 0 \},$$

which is just the subspace of function equal to 0 a.e. Then we define $L_\mu(X, A, F)$ as the quotient space

$$L_\mu(X, A, F) = L_\mu(X, A, F)/\mathcal{N}.$$ 

For any equivalence class $f \in L_\mu(X, A, F)$, since for any two representatives $f, g \in L_\mu(X, A, F)$ in the equivalence class $f$, we have $f = g$ a.e., by Proposition 4.23(1),

$$\int f d\mu = \int g d\mu,$$

so we can define $\int f d\mu$ as

$$\int f d\mu = \int f d\mu.$$

Similarly, $\|f\|_1$ is defined as

$$\|f\|_1 = \|f\|_1,$$

for any $f \in L_\mu(X, A, F)$ in the equivalence class $f$.

The following theorem is immediately obtained from Theorem 4.22 by passing to the quotient.

**Theorem 4.28.** (Fischer–Riesz) The semi-norm $\|\|_1$ on $L_\mu(X, A, F)$ induced by the semi-norm $\|\|_1$ on $L_\mu(X, A, F)$ by passing to the quotient is a norm on $L_\mu(X, A, F)$ called the $L^1$-norm. With this norm, the space $L_\mu(X, A, F)$ is complete (it is a Banach space). The subspace $\text{Step}_\mu(X, A, F)$ is dense in $L_\mu(X, A, F)$. 

Finally, the following proposition confirms one of our earlier claims.

**Proposition 4.29.** The space $L_\mu(X, A, F) = \mathcal{L}_\mu(X, A, F)/\mathcal{N}$ is isomorphic to the Cauchy completion of the space $\text{Step}_\mu(X, A, F)/\mathcal{SN}$; see the diagram

\[
\begin{align*}
\text{Step}_\mu(X, A, F) & \xrightarrow{\text{completion}} \mathcal{L}_\mu(X, A, F) \\
\text{Step}_\mu(X, A, F)/\mathcal{SN} & \xrightarrow{\text{completion}} L_\mu(X, A, F) = \mathcal{L}_\mu(X, A, F)/\mathcal{N}.
\end{align*}
\]

In the next section, we consider some fundamental convergence theorems. A very useful corollary of these theorems is that a function $f$ belongs to $L_\mu(X, A, F)$ iff it belongs to $\mathcal{M}_\mu(X, A, F)$ (it is $\mu$-measurable), and if $\int \|f\| \, d\mu$ exists. By Proposition 4.20, the space $L_\mu(X, A, F)$ is a subspace of $\mathcal{M}_\mu(X, A, F)$, and we already know from Proposition 4.23(4) that if $f \in L_\mu(X, A, F)$ then $\|f\| \in L_\mu(X, A, \mathbb{R})$. The converse is not trivial, but it will be shown as a corollary of the dominated convergence theorem discussed in Section 4.6.

### 4.6 Fundamental Convergence Theorems

Besides the fact that the Lebesgue–Bochner integral is defined for a much bigger class of functions than the regulated functions (or the Riemann-integrable functions), one of its main advantages is that it leads to simple and flexible criteria to tell whether the limit of a sequence of integrable functions is integrable. Most of these results allow interchanging a limit and an integral. We begin with criteria applying to real-valued functions. These results actually apply to extended functions with values in $\mathbb{R} \cup \{+\infty\}$, but for simplicity we stick to functions $f : X \to \mathbb{R}$. As in the previous section the results that we state without proof are proven either in Marle [56] or in Lang [52].

**Theorem 4.30.** (Monotone Convergence Theorem) Let $(f_n)_{n \geq 1}$ be a sequence of functions $f_n \in \mathcal{L}_\mu(X, A, \mathbb{R})$ such that $f_n \leq f_{n+1}$ for all $n \geq 1$, and assume that there is some $M > 0$ such that

$$\left| \int f_n d\mu \right| \leq M \quad \text{for all } n \geq 1.$$  

Then the sequence $(f_n)_{n \geq 1}$ converges pointwise a.e., and also in the $\| \|_1$-norm, to a function $f \in \mathcal{L}_\mu(X, A, \mathbb{R})$. We also have $\lim_{n \to \infty} \|f_n\|_1 = \|f\|_1$ and

$$\lim_{n \to \infty} \int f_n d\mu = \int f d\mu.$$  

The same result applies to a nonincreasing sequence $(f_n)$ (with $f_n \geq f_{n+1}$ for all $n \geq 1$).
Proof. We follow Lang [52] (Chapter VI, §5, Theorem 5.5). Let
\[ \alpha = \sup_k \int f_k d\mu, \]
which is well defined since
\[ \left| \int f_n d\mu \right| \leq M \quad \text{for all } n \geq 1. \]
For \( n \geq m \) we have
\[ \| f_n - f_m \|_1 = \int (f_n - f_m) d\mu = \int f_n d\mu - \int f_m d\mu \leq \alpha - \int f_m d\mu, \]
which implies that \((f_n)\) is a \( \| \cdot \|_1\)-Cauchy sequence. By the Fischer–Riesz theorem (Theorem 4.22), the sequence converges to some limit \( f \in L_\mu(X,A,\mathbb{R}) \) in the \( \| \cdot \|_1 \)-norm. By Proposition 4.25, there is a subsequence \((f_{n_k})_{k \geq 1}\) of \((f_n)\) that converges a.e. to \( f \in L_\mu(X,A,\mathbb{R}) \), and since the sequence \((f_n)\) is increasing, it also converges a.e. to \( f \).

The following theorem has a different flavor. It asserts the existence of the sup of a sequence of functions.

**Theorem 4.31.** (Beppo–Levi) Let \((f_n)\) be a sequence of functions \( f_n \in L_\mu(X,A,\mathbb{R}) \). If there is a function \( g \in L_\mu(X,A,\mathbb{R}) \) such that \( g \geq 0 \) and \( |f_n| \leq g \) for all \( n \geq 1 \), then \( \sup_{n \geq 1} f_n \) and \( \inf_{n \geq 1} f_n \) belong to \( L_\mu(X,A,\mathbb{R}) \), and we have
\[ \sup_{n \geq 1} \int f_n d\mu \leq \int \left( \sup_{n \geq 1} f_n \right) d\mu \quad \text{and} \quad \int \left( \inf_{n \geq 1} f_n \right) d\mu \leq \inf_{n \geq 1} \int f_n d\mu. \]

Proof. We follow Lang [52] (Chapter VI, §5, Corollary 5.6). The functions
\[ g_n = \sup \{ f_1, \ldots, f_n \} \]
belong to \( L_\mu(X,A,\mathbb{R}) \) and form an increasing sequence bounded by \( g \). Since
\[ \int g_n d\mu \leq \int g_{n+1} d\mu \quad \text{and} \quad \int g_n d\mu \leq \int g d\mu, \]
there is some \( M > 0 \) such that \( \| g_n \|_1 \leq M \) for all \( n \geq 1 \). Therefore, by the monotone convergence theorem (Theorem 4.30), the sequence \((g_n)\) converges pointwise a.e. to some function in \( L_\mu(X,A,\mathbb{R}) \), but \((g_n)\) converges pointwise to \( \sup_{n \geq 1} f_n \), so the sequence \((g_n)\) converges pointwise a.e. to \( \sup_{n \geq 1} f_n \). Since \( f_n \leq \sup_{n \geq 1} f_n \), we have
\[ \int f_n d\mu \leq \int (\sup_{n \geq 1} f_n) d\mu, \]
4.6. FUNDAMENTAL CONVERGENCE THEOREMS

which implies
\[ \sup_{n \geq 1} \int f_n d\mu \leq \int (\sup_{n \geq 1} f_n) d\mu, \]
as claimed.

Given a sequence \((f_n)_{n \geq 1}\) of functions \(f_n : X \to \mathbb{R}\) such that \(f_n \geq 0\), recall that
\[ \lim inf f_n = \lim_{k \to \infty} \inf_{n \geq k} f_n. \]

**Theorem 4.32.** (Fatou’s Lemma) Let \((f_n)_{n \geq 1}\) be a sequence of functions \(f_n \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})\) such that \(f_n \geq 0\). If \(\lim inf \|f_n\|_1 = \lim inf \int f_n d\mu\) exists, then there is a function \(f \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})\) such that \(\lim inf f_n\) converges pointwise to \(f\) a.e., and
\[ \int f d\mu \leq \lim inf \int f_n d\mu. \]

A proof of Theorem 4.32 is given in Lang [52] (Chapter VI, §5).

The next theorem applies to functions with values in any Banach space \(F\), and is the most important convergence theorem.

**Theorem 4.33.** (Lebesgue Dominated Convergence Theorem) Let \((f_n)_{n \geq 1}\) be a sequence of functions \(f_n \in \mathcal{L}_\mu(X, \mathcal{A}, F)\). If \((f_n)\) converges pointwise a.e. to a function \(f : X \to F\), and if there is some function \(g \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})\) such that \(g \geq 0\) and \(\|f_n\| \leq g\) for all \(n \geq 1\), then \(f \in \mathcal{L}_\mu(X, \mathcal{A}, F)\) and \((f_n)_{n \geq 1}\) converges to \(f\) in the \(\|\cdot\|_1\)-norm. Consequently
\[ \lim_{n \to \infty} \int f_n d\mu = \int f d\mu. \]

**Proof.** We follow Marle [56] (Chapter 2, Section 4, Theorem 2.4.7). For each \(n, p \geq 1\), let
\[ g_{n,p} = \sup_{n \leq m \leq n + p} \sup_{n \leq r \leq n + p} \|f_m - f_r\| \]
\[ g_n = \sup_{m,r \geq n} \|f_m - f_r\| = \sup_{p \geq 1} g_{n,p}. \]

For all \(n, p \geq 1\), we have \(g_{n,p} \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})\), \(g_{n,p} \leq g_{n,p+1}\), and
\[ 0 \leq g_{n,p} \leq 2g. \]

We get
\[ \left| \int g_{n,p} d\mu \right| \leq 2 \int g d\mu, \]
so by the monotone convergence theorem (Theorem 4.30), the sequence \((g_{n,p})_{p \geq 1}\) converges pointwise a.e. to a limit in \(L_\mu(X, A, \mathbb{R})\). However, by construction this limit is \(g_n\). Thus \(g_n \in L_\mu(X, A, \mathbb{R})\), and we also have

\[
\int g_n \, d\mu \leq 2 \int g \, d\mu.
\]

The sequence \((g_n)_{n \geq 1}\) is nonincreasing and since by hypothesis \((f_n)\) converges pointwise a.e. to \(f\), the sequence \((g_n)\) converges pointwise a.e. to 0. By the monotone convergence theorem (Theorem 4.30),

\[
\lim_{n \to \infty} \int g_n \, d\mu = 0.
\]

Hence by definition of \(g_n\) the sequence \((f_n)\) is actually a Cauchy sequence, and by Proposition 4.27, we have \(f \in L_\mu(X, A, F)\) and \((f_n)\) converges to \(f\) in the \(\|\cdot\|_1\)-norm. We have

\[
\left\| \int f_n \, d\mu - \int f \, d\mu \right\| = \left\| \int (f_n - f) \, d\mu \right\|
\leq \int \|f_n - f\| \, d\mu
= \|f_n - f\|_1,
\]

and since \((f_n)\) converges to \(f\) in the \(\|\cdot\|_1\)-norm, this implies that

\[
\lim_{n \to \infty} \int f_n \, d\mu = \int f \, d\mu,
\]

as claimed. \(\square\)

The first important application of Theorem 4.33 is to provide a characterization of the integrability of a function \(f \in M_\mu(X, A, F)\) in terms of \(\int \|f\| \, d\mu\).

**Theorem 4.34.** A function \(f : X \to F\) is integrable, that is, \(f \in L_\mu(X, A, F)\), iff \(f \in M_\mu(X, A, F)\) and \(\|f\| \in L_\mu(X, A, \mathbb{R})\). More generally, if \(f \in M_\mu(X, A, F)\) and if there is a function \(g \in L_\mu(X, A, \mathbb{R})\) such that \(g \geq 0\) and \(\|f\| \leq g\), then \(f \in L_\mu(X, A, F)\).

**Proof.** By Proposition 4.20, the space \(L_\mu(X, A, F)\) is a subspace of \(M_\mu(X, A, F)\), and we already know from Proposition 4.23(4) that if \(f \in L_\mu(X, A, F)\) then \(\|f\| \in L_\mu(X, A, \mathbb{R})\).

For the converse, we may assume that \(f\) and \(g\) are measurable, since a \(\mu\)-measurable function is equal a.e. to a measurable function. There is a sequence \((h_n)_{n \geq 1}\) of \(\mu\)-step maps that converges pointwise a.e. to \(f\). For every \(x \in X\) and every \(n \geq 1\), let

\[
h_n'(x) = \begin{cases} h_n(x) & \text{if } \|h_n(x)\| \leq 2g(x) \\ 0 & \text{if } \|h_n(x)\| > 2g(x). \end{cases}
\]
For every \( n \geq 1 \), the function \( h'_n \) is a \( \mu \)-step function and \( \|h'_n\| \leq 2g \in L_\mu(X, \mathcal{A}, \mathbb{R}) \). We claim that for every \( x \in X \) such that \( (h_n(x)) \) converges to \( f(x) \), the sequence \( (h'_n(x)) \) also converges to \( f(x) \). If \( g(x) = 0 \), then \( \|f(x)\| = 0 \), so \( f(x) = 0 \), and then \( h'_n(x) = 0 \) for all \( n \geq 1 \). If \( g(x) \neq 0 \), since the sequence \( (h_n(x)) \) converges to \( f(x) \) and since \( \|f(x)\| < 2g(x) \), there is some \( M > 0 \) such that \( \|h_n(x)\| \leq 2g(x) \) for all \( n \geq M \), which implies that \( h'_n(x) = h_n(x) \).

It follows that the sequence \( (h'_n)_{n \geq 1} \) converges pointwise a.e. to \( f \). By Theorem 4.33, since \( \|h'_n\| \leq 2g \), we conclude that \( f \in L_\mu(X, \mathcal{A}, F) \).

A useful corollary of Theorem 4.34 is the following result.

**Proposition 4.35.** The following facts hold:

1. If \( f \in L_\mu(X, \mathcal{A}, F) \), \( g \in M_\mu(X, \mathcal{A}, K) \) with \( K = \mathbb{R} \) or \( K = \mathbb{C} \), and \( \|g\| \) is bounded, then \( fg \in L_\mu(X, \mathcal{A}, F) \).

2. Let \( h : E \times F \rightarrow G \) be a continuous bilinear map, where \( E, F, G \) are Banach spaces. If \( f \in L_\mu(X, \mathcal{A}, E) \) and \( g \in M_\mu(X, \mathcal{A}, F) \) with \( \|g\| \) bounded, then \( h(f, g) \in L_\mu(X, \mathcal{A}, G) \).

3. Let \( f \in L_\mu(X, \mathcal{A}, \mathbb{R}) \), with \( f \geq 0 \), and let \( g \in M_\mu(X, \mathcal{A}, \mathbb{R}) \), with values in an interval \([m, M]\). Then \( fg \in L_\mu(X, \mathcal{A}, \mathbb{R}) \) and we have

\[
m \int f \, d\mu \leq \int gf \, d\mu \leq M \int f \, d\mu.
\]

Another corollary involves series of functions in \( L_\mu(X, \mathcal{A}, F) \).

**Proposition 4.36.** Let \( (f_n)_{n \geq} \) be a sequence of functions \( f_n \in L_\mu(X, \mathcal{A}, F) \). If the series

\[
\sum_{n=1}^{\infty} \int \|f_n\| \, d\mu
\]

converges, then the series

\[
f(x) = \sum_{n=1}^{\infty} f_n(x)
\]

converges a.e., \( f \in L_\mu(X, \mathcal{A}, F) \), and

\[
\int f \, d\mu = \sum_{n=1}^{\infty} \int f_n \, d\mu
\]

The following proposition is needed for the proof of several results stated in Chapter 5.

**Proposition 4.37.** (Averaging Theorem) Let \( f \in L_\mu(X, \mathcal{A}, F) \) be any function and let \( S \) be any closed subset of \( X \). If for any measurable subset \( A \) of finite measure \( \mu(A) > 0 \) we have

\[
\frac{1}{\mu(A)} \int_A f \, d\mu \in S,
\]

and if \( 0 \in S \) or if \( X \) is \( \sigma \)-finite, then \( f(x) \in S \) for all almost all \( x \in X \).
Proposition 4.37 is proven in Lang [52] (Chapter VI, Theorem 5.15). By applying Proposition 4.37 to the set $S = \{0\}$, we obtain the following useful corollary.

**Proposition 4.38.** For any function $f \in \mathcal{L}_\mu(X, \mathcal{A}, F)$, if

$$\int_A f \, d\mu = 0$$

for every measurable subset $A$ of finite measure, then $f = 0$ almost everywhere.

We conclude this section with two results about the continuity and the differentiability of a function defined by an integral.

**Proposition 4.39.** Let $(X, \mathcal{A}, \mu)$ be a measure space, let $U$ be metric space, let $F$ be a Banach space (over $\mathbb{R}$ or $\mathbb{C}$), and let $f: U \times X \to F$ be a function.

1. (Continuity of the integral) Assume that $f$ has the following properties:
   
   (a) For every $u \in U$, the map $f_{u,-}: X \to F$ given by
   
   $$f_{u,-}(x) = f(u, x) \quad x \in X,$$
   
   belongs to $\mathcal{L}_\mu(X, \mathcal{A}, F)$,
   
   (b) For every $x \in X$, the map $f_{-x}: U \to F$ given by
   
   $$f_{-x}(u) = f(u, x) \quad u \in U,$$
   
   is continuous.
   
   (c) There is some $g \in \mathcal{L}_\mu(X, \mathcal{A}, \mathbb{R})$, $g \geq 0$, such that
   
   $$\|f(u, x)\| \leq g(x) \quad \text{for all } u \in U, \text{ and all } x \in X.$$

   Then the map $h: U \to F$ given by
   
   $$h(u) = \int f_{u,-} \, d\mu$$

   is continuous.

2. (Taking a derivative under the integral sign) Suppose $U$ is an open subset of a Banach space $G$, and let $\mathcal{L}(G; F)$ be the space of linear continuous maps from $G$ to $F$, with the operator norm (see Definition 15.42). Assume that $f$ has the following properties:

   (d) For every $u \in U$, the map $f_{u,-}: X \to F$ given by
   
   $$f_{u,-}(x) = f(u, x) \quad x \in X,$$
   
   belongs to $\mathcal{L}_\mu(X, \mathcal{A}, F)$,
(e) For every \( x \in X \), the map \( f_{-x} : U \to F \) is differentiable, and let \( Df_{-x} \) be this derivative (a map from \( U \) to \( \mathcal{L}(G; F) \)).

(f) For every \( u \in U \), the map from \( X \) to \( \mathcal{L}(G; F) \) given by

\[
x \mapsto Df_{-x}(u)
\]

belongs to \( \mathcal{L}_{\mu}(X, \mathcal{A}, \mathcal{L}(G; F)) \), and there is some \( g \in \mathcal{L}_{\mu}(X, \mathcal{A}, \mathbb{R}) \), \( g \geq 0 \), such that

\[
\|Df_{-x}(u)\| \leq g(x) \quad \text{for all } u \in U, \text{ and all } x \in X.
\]

Then the map \( h : U \to F \) given by

\[
h(u) = \int f_{u,-} \, d\mu
\]

is differentiable in \( U \), and its derivative at \( u \in U \) is given by

\[
Dh_u = \int Df_{-x}(u) \, d\mu.
\]

More could be said about the applications of the convergence theorems, but we have everything we need.

**Remark:** There is another approach to the definition of the integral that applies only to real and complex-valued functions, presented in various texts such as Rudin [66]. In this approach, positive functions play a central role. This approach relies on the fact that for any measurable function \( f : X \to [0, +\infty] \) there is a monotonic sequence \((f_n)\) of positive step functions that converges pointwise to \( f \); see Rudin [66] (Chapter 1, Theorem 1.17). The integral of a step function is defined in the usual way. Then, given any measurable function \( f : X \to [0, +\infty] \), the integral of \( f \) is defined as

\[
\int f \, d\mu = \sup_{0 \leq s \leq f} \int s \, d\mu,
\]

where \( s \) is a step function.

A main difference with the approach we followed is that this definition of the integral allows it to take the value \(+\infty\). Of course, later on, in order to define what it means for a measurable complex-valued function \( f : X \to \mathbb{C} \) to be integrable, the condition

\[
\int |f| \, d\mu < +\infty
\]

is required. Thus in this approach, the space \( \mathcal{L}_{\mu}(X, \mathcal{A}, \mathbb{C}) \) is defined as the space of measurable functions such that the positive function \(|f|\) has a finite integral.
In the approach that we followed, due to Bochner and Dunford, the space $L_\mu(X, \mathcal{A}, \mathbb{C})$ is defined in terms of various Cauchy sequences, and the fact that if a function $f : X \to \mathbb{C}$ is measurable and if $|f|$ has a finite integral, then $f \in L_\mu(X, \mathcal{A}, \mathbb{C})$, is a theorem (Theorem 4.34). Ultimately, it is proved that $L_\mu(X, \mathcal{A}, \mathbb{C})$ is complete (see Rudin [66] Chapter 3, Theorem 3.11), and it is observed that as a corollary, from a $\|\cdot\|_1$-Cauchy sequence, one can extract a subsequence that converges pointwise a.e. (Rudin [66] Chapter 3, Theorem 3.12).

It is also shown that the $\mu$-step functions are dense in $L_\mu(X, \mathcal{A}, \mathbb{C})$ (Rudin [66] Chapter 3, Theorem 3.13).

The circle is closed. What we took as a definition of $L_\mu(X, \mathcal{A}, \mathbb{C})$ is obtained as a corollaries in the other approach, and the two approaches yield the same notion of integrability (the same space $L_\mu(X, \mathcal{A}, \mathbb{C})$).

One might argue that the approach relying on the integral of positive functions is simpler, or at least takes less efforts. For one thing, it does not need the refined notion of $\mu$-step maps and $\mu$-measurable maps. However, our feeling is that the approach we followed provides a better understanding of the structure of $L_\mu(X, \mathcal{A}, \mathbb{C})$. Also, it can’t be avoided in one wants to integrate functions with values in an infinite-dimensional vector space.

4.7 The Spaces $L^p_\mu(X, \mathcal{A}, F)$ and $L^p_\mu(X, \mathcal{A}, F)$; $p = 1, 2, \infty$

Theorem 4.34 suggests the definition of other families of integrable functions.

**Definition 4.14.** Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(F, \mathcal{B})$ be a measurable space, with $F$ a Banach space and $\mathcal{B}$ its Borel $\sigma$-algebra. For any $p \geq 1$, the set of functions $L^p_\mu(X, \mathcal{A}, F)$ is the set of functions $f \in M_\mu(X, \mathcal{A}, F)$ such that $\|f\|^p \in L_\mu(X, \mathcal{A}, \mathbb{R})$, or equivalently

$$\int \|f\|^p \, d\mu < +\infty.$$ 

By Theorem 4.34, we have $L^1_\mu(X, \mathcal{A}, F) = L_\mu(X, \mathcal{A}, F)$, and we know that $L^1_\mu(X, \mathcal{A}, F)$ is a vector space. Although it is possible to develop a theory of $L^p$ spaces for any $p \geq 1$, for our applications to harmonic analysis we only need the cases $p = 1, 2$. The case where $p = \infty$ arises when we consider duality, but we postpone the definition of $L^\infty_\mu(X, \mathcal{A}, F)$.

The space $L^2_\mu(X, \mathcal{A}, F)$ is particularly interesting because if $F$ is a Hilbert space, then it can be given a Hilbert space structure which uses the inner product on $F$ (not quite, because the Hermitian form $\langle - , - \rangle_\mu$ that we obtain is not positive definite, which means that $\langle f, f \rangle_\mu = 0$ does necessarily imply that $f = 0$).

Let us start with the simple case where $F = \mathbb{C}$. If $f : X \to \mathbb{C}$ is a complex-valued function, then by $|f|^2$ we mean the function defined such that

$$|f|^2(x) = f(x)\overline{f(x)} \quad \text{for all } x \in X.$$
4.7. THE SPACES $\mathcal{L}_p^\mu(X, A, F)$ AND $L_p^\mu(X, A, F)$; $p = 1, 2, \infty$

For any two functions $f, g : X \to \mathbb{C}$, by $\langle f, g \rangle$ we mean the function defined such that

$$\langle f, g \rangle(x) = f(x)\overline{g(x)} \quad \text{for all } x \in X.$$  

For the more general case where $F$ is a Hilbert space with inner product $\langle -, - \rangle_F$, for any two functions $f, g : X \to F$, then by $\langle f, g \rangle$ we mean the function defined such that

$$\langle f, g \rangle(x) = \langle f(x), g(x) \rangle_F \quad \text{for all } x \in X.$$  

In particular,

$$\|f\|^2 = \langle f, f \rangle.$$  

From now on, when dealing with $\mathcal{L}_2^\mu(X, A, F)$, we assume that $F$ is a Hilbert space (over $\mathbb{C}$). If the reader feels more comfortable, he/she may assume that $F = \mathbb{C}$, but significant simplifications do not arise.

**Proposition 4.40.** The set $\mathcal{L}_2^\mu(X, A, F)$ is a vector space. For any two maps $f, g \in \mathcal{L}_2^\mu(X, A, F)$, we have $\langle f, g \rangle \in \mathcal{L}_1^\mu(X, A, \mathbb{C})$, and the map

$$(f, g) \mapsto \int \langle f, g \rangle d\mu$$

is a Hermitian positive map (not necessarily positive definite).

**Proof.** It is easy to see that $\langle f, g \rangle$ is a limit of step maps a.e., so $\langle f, g \rangle \in \mathcal{M}_\mu(X, A, \mathbb{C})$, and we have the standard inequality

$$2|\langle f, g \rangle| \leq \|f\|^2 + \|g\|^2,$$

with $\|f\|^2 + \|g\|^2 \in \mathcal{L}^1_\mu(X, A, \mathbb{R})$ since $f, g \in \mathcal{L}^2_\mu(X, A, F)$. By Theorem 4.34, we have $\langle f, g \rangle \in \mathcal{L}^1_\mu(X, A, \mathbb{C})$, so $\int \langle f, g \rangle d\mu$ is well defined. If $f \in \mathcal{L}^2_\mu(X, A, F)$ and $g \in \mathcal{L}^2_\mu(X, A, F)$, then since

$$\|f + g\|^2 \leq \|f\|^2 + 2|\langle f, g \rangle| + \|g\|^2,$$

as all the functions on the right-hand side are in $\mathcal{L}^1_\mu(X, A, \mathbb{R})$, we have $f + g \in \mathcal{L}^2_\mu(X, A, F)$. For any $\lambda \in \mathbb{C}$, we have $|\lambda|^2 \|f\| \in \mathcal{L}^1_\mu(X, A, \mathbb{R})$, so $\lambda f \in \mathcal{L}^2_\mu(X, A, F)$. Thus, $\mathcal{L}^2_\mu(X, A, F)$ is a vector space. Using the linearity of the integral, it is easy to check that the map

$$(f, g) \mapsto \int \langle f, g \rangle d\mu$$

is a Hermitian positive map. $\square$

**Definition 4.15.** For any two functions $f, g \in \mathcal{L}^2_\mu(X, A, F)$, the Hermitian map $\langle f, g \rangle_\mu$ is defined by

$$\langle f, g \rangle_\mu = \int \langle f, g \rangle d\mu$$

The $L^2$-semi-norm $\|f\|_2$ is given by

$$\|f\|_2 = \sqrt{\langle f, f \rangle_\mu} = \left( \int \langle f, f \rangle d\mu \right)^{1/2} = \left( \int \|f\|^2 d\mu \right)^{1/2}.$$
It is a standard result of linear algebra that the Cauchy–Schwarz inequality holds:
\[ |\langle f, g \rangle_\mu| \leq \|f\|_2 \|g\|_2. \]

As a consequence \(\|\|\|_2\) is a semi-norm.

**Proposition 4.41.** For any \(f \in \mathcal{L}_\mu^2(X, \mathcal{A}, F)\), we have \(\|f\|_2 = 0\) iff \(f = 0\) a.e.

**Proof.** If \(f = 0\) a.e., then \(\langle f, f \rangle = 0\) a.e., so \(\|f\|_2^2 = \int \langle f, f \rangle d\mu = 0\). Conversely, if \(\|f\|_2 = 0\), then this means that \(\int \langle f, f \rangle d\mu = 0\), but \(\langle f, f \rangle \in \mathcal{L}_\mu^1(X, \mathcal{A}, \mathbb{R})\) is a positive function, so we know from Proposition 4.26 that \(\langle f, f \rangle = 0\) a.e., that is, \(f = 0\) a.e. \(\square\)

If \(X\) has finite measure, then \(\mathcal{L}_\mu^2(X, \mathcal{A}, F)\) is contained in \(\mathcal{L}_\mu^1(X, \mathcal{A}, F)\).

**Proposition 4.42.** If \(X\) has finite measure, then for any \(f \in \mathcal{L}_\mu^2(X, \mathcal{A}, F)\), we have \(\|f\|_1 \leq \|f\|_2 \|1_X\|_2\), and \(\mathcal{L}_\mu^2(X, \mathcal{A}, F)\) is contained in \(\mathcal{L}_\mu^1(X, \mathcal{A}, F)\).

**Proof.** We apply the Cauchy–Schwarz inequality to \(\|f\|\) and to the constant function \(1_X\) equal to 1 on \(X\). To be more specific, since
\[
\langle \|f\|, 1_X \rangle = \|f\| \cdot \overline{1_X} = \|f\|,
\]
we have
\[
\langle \|f\|, 1_X \rangle_\mu = \int \langle \|f\|, 1_X \rangle d\mu = \int \|f\| = \|f\|_1,
\]
and
\[
\|\|f\|\|_2^2 = \langle \|f\|, \|f\| \rangle_\mu = \int \langle \|f\|, \|f\| \rangle d\mu = \int \|f\|^2 d\mu = \|f\|_2^2,
\]
and so the Cauchy–Schwarz inequality implies that \(\|f\|_1 \leq \|f\|_2 \|1_X\|_2\). Obviously, this inequality shows that \(\mathcal{L}_\mu^2(X, \mathcal{A}, F)\) is contained in \(\mathcal{L}_\mu^1(X, \mathcal{A}, F)\). \(\square\)

It should be noted that if \(X\) has finite measure then the inclusion can be strict, and if \(X\) has infinite measure, then in general there are no inclusion properties.

**Example 4.3.**

1. If \(X = (0, 1)\), with the Lebesgue measure, then \(\frac{1}{\sqrt{x}} \in \mathcal{L}^1((0, 1), \mu_L)\) but \(\frac{1}{\sqrt{x}} \notin \mathcal{L}^2((0, 1), \mu_L)\).
2. If \(X = (1, \infty)\) with the Lebesgue measure, then \(\frac{1}{x} \in \mathcal{L}^2((1, \infty), \mu_L)\) but \(\frac{1}{x} \notin \mathcal{L}^1((1, \infty), \mu_L)\).
3. If \(X = (0, \infty)\) with the Lebesgue measure, then \(\frac{1}{(x+1)\sqrt{x}} \in \mathcal{L}^1((0, \infty), \mu_L)\) but \(\frac{1}{(x+1)\sqrt{x}} \notin \mathcal{L}^2((0, \infty), \mu_L)\).
One of the main properties of $L_2^\mu(X, \mathcal{A}, F)$ is that it is complete for the semi-norm $\| \|_2$. By taking the quotient of $L_2^\mu(X, \mathcal{A}, F)$ by the space of function equal to 0 a.e., we obtain a Hilbert space.

**Theorem 4.43.** Let $(f_n)_{n \geq 1}$ be an $\| \|_2$-Cauchy sequence of functions $f_n \in L_2^\mu(X, \mathcal{A}, F)$. Then there is a function $f \in L_2^\mu(X, \mathcal{A}, F)$ with the following properties:

1. The sequence $(f_n)_{n \geq 1}$ converges to $f$ in the $\| \|_2$-semi-norm. Thus $L_2^\mu(X, \mathcal{A}, F)$ is complete.

2. There is a subsequence $(f_{n_k})_{k \geq 1}$ of $(f_n)_{n \geq 1}$ with the following properties:

3. For every $\epsilon > 0$, there is a subset $Z$ such that $\mu(Z) < \epsilon$ and the subsequence $(f_{n_k})_{k \geq 1}$ converges uniformly to $f$ on $X - Z$.

A proof of Theorem 4.43 is given in Lang [52] (Chapter VII, §1).

In view of Proposition 4.41, we make the following definition.

**Definition 4.16.** Let $L_2^\mu(X, \mathcal{A}, F)$ be the quotient of the vector space $L_2^\mu(X, \mathcal{A}, F)$ by the subspace of functions equal to 0 a.e. (which is the set of functions $f$ such that $\| f \|_2 = 0$). The norm induced by the semi-norm $\| \|_2$ on $L_2^\mu(X, \mathcal{A}, F)$ is called the $L^2$-norm.

Obviously the positive Hermitian form $\langle f, g \rangle_\mu$ induces a positive definite Hermitian form on $L_2^\mu(X, \mathcal{A}, F)$. Theorem 4.43 immediately implies the following result.

**Theorem 4.44.** (Fischer–Riesz) The space $L_2^\mu(X, \mathcal{A}, F)$ is a Hilbert space under the positive definite Hermitian form induced by $\langle - , - \rangle_\mu$.

**Definition 4.17.** The norm $\| \|_2$ associated with the inner product $\langle - , - \rangle_\mu$ on $L_2^\mu(X, \mathcal{A}, F)$ is called the $L^2$-norm.

In the special case where $X = \mathbb{N}$ (or $X = \mathbb{Z}$), $\mathcal{A} = 2^X$, $\mu$ is the counting measure, and $F = \mathbb{C}$, as in Example 4.2, we see that for $p = 1, 2$, we have

$$L_p^\mu(X, \mathcal{A}, \mathbb{C}) = \{(x_n)_{n \in \mathbb{N}} \mid x_n \in \mathbb{C}, \sum_{n \in \mathbb{N}} |x_n|^p < \infty\}.$$ 

It is customary to denote this space by $l^p(\mathbb{N})$. We define $l^p(\mathbb{Z})$ similarly by replacing $\mathbb{N}$ by $\mathbb{Z}$.

We will show shortly that the space of $\mu$-step functions is dense in $L_2^\mu(X, \mathcal{A}, F)$ (for the $L^2$-norm). First here is a corollary of Theorem 4.44.
Proposition 4.45. If \((f_n)_{n \geq 1}\) is a \(\|\cdot\|_2\)-Cauchy sequence of functions \(f_n \in L^2_μ(X, A, F)\), and if \((f_n)_{n \geq 1}\) converges pointwise a.e. to a function \(f: X \to F\), then \(f \in L^2_μ(X, A, F)\), and \((f_n)_{n \geq 1}\) converges to \(f\) in the \(\|\cdot\|_2\)-semi-norm.

The Lebesgue dominated convergence theorem also holds for \(L^2_μ(X, A, F)\).

Theorem 4.46. (Lebesgue Dominated Convergence Theorem for \(L^2_μ\)) Let \((f_n)_{n \geq 1}\) be a sequence of functions \(f_n \in L^2_μ(X, A, F)\). If \((f_n)\) converges pointwise a.e. to a function \(f: X \to F\), and if there is some function \(g \in L^2_μ(X, A, R)\) such that \(g \geq 0\) and \(\|f_n\| \leq g\) for all \(n \geq 1\), then \(f \in L^2_μ(X, A, F)\) and \((f_n)_{n \geq 1}\) converges to \(f\) in the \(\|\cdot\|_2\)-norm. Consequently

\[
\lim_{n \to \infty} \int f_n dμ = \int f dμ.
\]

A proof of Theorem 4.46 is given in Lang [52] (Chapter VII, §1).

Theorem 4.34 also holds for \(L^2_μ(X, A, F)\).

As a corollary of Theorem 4.46 we can show that the \(μ\)-step functions are dense in \(L^2_μ(X, A, F)\).

Proposition 4.47. The subspace \(\text{Step}_μ(X, A, F)\) is dense in \(L^2_μ(X, A, F)\) with respect to the \(L^2\)-semi-norm.

Proof. Let \(f \in L^2_μ(X, A, F)\). Since \(f\) is \(μ\)-measurable, there is a sequence \((f_n)_{n \geq 1}\) of \(μ\)-step functions \(f_n\) that converges pointwise a.e. to \(f\). For every \(n \geq 1\) and every \(x \in X\), define \(g_n\) by

\[
g_n(x) = \begin{cases} 
  f_n(x) & \text{if } \|f_n(x)\| \leq 2\|f(x)\| \\
  0 & \text{if } \|f_n(x)\| > 2\|f(x)\|.
\end{cases}
\]

We may assume that \(f\) is measurable since it differs from a measurable function on a set of measure zero. Then the functions \(g_n\) are \(μ\)-step functions, they satisfy the inequality \(\|g_n\| \leq 2\|f\|\) with \(2\|f\| \in L^2_μ(X, A, R)\), and the sequence \((g_n)\) converges a.e. to \(f\). By Theorem 4.46, the sequence \((g_n)\) converges to \(f\) in the \(\|\cdot\|_2\)-norm, which proves that \(\text{Step}_μ(X, A, F)\) is dense in \(L^2_μ(X, A, F)\) with respect to the \(L^2\)-semi-norm. ∎

We now would like to understand the duals of \(L^1_μ(X, A, F)\) and \(L^2_μ(X, A, F)\), that is, the spaces of continuous linear forms on \(L^1_μ(X, A, F)\) and \(L^2_μ(X, A, F)\) (with values in \(C\)). In the case of \(L^2_μ(X, A, F)\), it is a classical theorem (the Riesz representation theorem) that the dual of a Hilbert space is isomorphic to itself, so the dual of \(L^2_μ(X, A, F)\) is isomorphic to \(L^2_μ(X, A, F)\) itself. In the case of \(L^1_μ(X, A, F)\), it turns out that its dual is isomorphic to a space denoted \(L^∞_μ(X, A, F)\). Here we assumed that \(F\) is a Hilbert space.

To define \(L^∞_μ(X, A, F)\), we only need the fact that \(F\) is a Banach space. The space \(L^∞_μ(X, A, F)\) consists of all functions \(f: X \to F\) that are equal to a bounded \(μ\)-measurable function a.e. We can define a semi-norm on \(L^∞_μ(X, A, F)\) as follows.
Definition 4.18. For any function \( f \in \mathcal{M}_\mu(X, \mathcal{A}, F) \), define the essential sup or semi-norm \( \|f\|_\infty \) of \( f \) by

\[
\|f\|_\infty = \inf \{ \alpha \in \mathbb{R}^+ \mid \mu(\{ x \in X \mid \|f(x)\| \geq \alpha \}) = 0 \}.
\]

The space \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \) is the set of functions \( f \in \mathcal{M}_\mu(X, \mathcal{A}, F) \) such that \( \|f\|_\infty < +\infty \).

The definition of \( \|f\|_\infty \) makes it clear that \( \|f\|_\infty = 0 \) iff \( f = 0 \) a.e. Observe that \( \|f\|_\infty \) is the greatest lower bound of the numbers \( \alpha \geq 0 \) such that a \( \mu \)-measurable function \( f \) has the property that \( \|f(x)\| \geq \alpha \) on a set of measure zero, in other words, such a \( \mu \)-measurable function is bounded a.e.

The space \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \) is a vector space. We also have the following result showing that \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \) is complete in the semi-norm \( \|\|_\infty \), but unless \( X \) has finite measure, the \( \mu \)-step maps are not dense in \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \).

Theorem 4.48. The following properties hold.

1. The space \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \) is complete in the semi-norm \( \|\|_\infty \). Furthermore, if \( (f_n)_{n \geq 1} \) is an \( \|\|_\infty \)-Cauchy sequence, then there is a set \( Z \) of measure zero such that \( (f_n)_{n \geq 1} \) converges uniformly to \( f \) on \( X - Z \).

2. If \( F \) is finite-dimensional, then the step maps (not the \( \mu \)-step maps) are dense in \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \).

3. If \( X \) has finite measure, then for every \( \epsilon > 0 \) and every \( f \in \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \), there is a \( \mu \)-step map \( s \) and a subset \( Z \) with \( \mu(Z) < \epsilon \) such that

\[
\|f - s\| < \epsilon \quad \text{on} \quad X - Z.
\]

Note that the constant with value 1 belongs to \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \), so if \( X \) has infinite measure, there is no way that it is a uniform limit of \( \mu \)-step maps, since a \( \mu \)-step map vanishes outside of a set of finite measure.

Remark: If \( X \) has finite measure, then we have the inclusion \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \subseteq \mathcal{L}_\mu^p(X, \mathcal{A}, F) \). In fact, \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \subseteq \mathcal{L}_\mu^p(X, \mathcal{A}, F) \subseteq \mathcal{L}_\mu^q(X, \mathcal{A}, F) \) for all \( p, q \geq 1 \) with \( p > q \); see Marle [56] (Chapter 4, Proposition 4.5.7).

Definition 4.19. Let \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \) be the quotient of the vector space \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \) by the subspace of functions equal to 0 a.e. (which is the set of functions \( f \) such that \( \|f\|_\infty = 0 \)). The norm induced by the semi-norm \( \|\|_\infty \) on \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \) is called the \( L^\infty \)-norm.

It should be noted that both the monotone convergence theorem and the dominated convergence theorem fail for \( \mathcal{L}_\mu^\infty(X, \mathcal{A}, F) \). Convergence in the \( \|\|_\infty \)-semi-norm fails; see Marle [56] (Chapter 4, Section 2).
We now consider the duality between the spaces $L^1_\mu(X,\mathcal{A},F)$ and $L^\infty_\mu(X,\mathcal{A},F)$. The field $\mathbb{C}$ is a Hilbert space, but for the general case we need to assume that $F$ is a Hilbert space. The key point is that by Proposition 4.35(2), for any $f \in L^1_\mu(X,\mathcal{A},F)$ and any $g \in L^\infty_\mu(X,\mathcal{A},F)$, then $\langle f,g \rangle \in L^1_\mu(X,\mathcal{A},F)$.

**Definition 4.20.** For any functions $f \in L^1_\mu(X,\mathcal{A},F)$ and $g \in L^\infty_\mu(X,\mathcal{A},F)$, define $[f,g]_\mu$ by

$$[f,g]_\mu = \int \langle f,g \rangle d\mu.$$  

We obtain a map $[-,-]_\mu: L^1_\mu(X,\mathcal{A},F) \times L^\infty_\mu(X,\mathcal{A},F) \to \mathbb{C}$ which is a sesquilinear pairing.

For simplicity, let us consider the special case where $F = \mathbb{C}$. In this case, we can define a bilinear (as opposed to sesquilinear) pairing $[-,-]_\mu: L^1_\mu(X,\mathcal{A},\mathbb{C}) \times L^\infty_\mu(X,\mathcal{A},\mathbb{C}) \to \mathbb{C}$ given by

$$[f,g]_\mu = \int fg d\mu.$$  

Observe that we intentionally used $fg$ instead of $f\overline{g}$, because we simply want a bilinear pairing.

Whenever we have a bilinear pairing $\varphi: E \times F \to \mathbb{C}$, recall that we define the linear maps $l_\varphi: E \to F^*$ and $r_\varphi: F \to E^*$ such that, for every $u \in E$,

$$l_\varphi(u)(y) = \varphi(u,y) \quad \text{for all } y \in F,$$

and for every $v \in F$,

$$r_\varphi(v)(x) = \varphi(x,v) \quad \text{for all } x \in E.$$

**Definition 4.21.** A bilinear pairing $\varphi$ is nondegenerate if for every $u \in E$, if $\varphi(u,v) = 0$ for all $v \in F$, then $u = 0$, and for every $v \in F$, if $\varphi(u,v) = 0$ for all $u \in E$, then $v = 0$.

Then if $\varphi$ is nondegenerate, then the maps $l_\varphi$ and $r_\varphi$ are injective. They are not surjective in general.

If $E$ is a normed vector space, then its dual $E'$ is the space of all continuous linear maps from $E$ to $\mathbb{C}$. We have $E' \subseteq E^*$, and the inclusion is strict if $E$ is infinite-dimensional.

The following result holds. For simplicity of notation, we drop $\varphi$ when writing $l_\varphi$ and $r_\varphi$.

**Theorem 4.49.** Assume $(X,\mathcal{A},\mu)$ is a measure space and that $\mu$ is $\sigma$-finite. Then the bilinear pairing $[-,-]_\mu: L^1_\mu(X,\mathcal{A},\mathbb{C}) \times L^\infty_\mu(X,\mathcal{A},\mathbb{C}) \to \mathbb{C}$
is nondegenerate. It satisfies the inequality
\[ |[f, g]_\mu| \leq \|f\|_1 \leq \|f\|_1 \|g\|_\infty. \]

The map \( l \) is a norm-preserving injective linear map between \( L^1_\mu(X, \mathcal{A}, \mathbb{C}) \) and the dual \( L^\infty_\mu(X, \mathcal{A}, \mathbb{C})' \) of \( L^\infty_\mu(X, \mathcal{A}, \mathbb{C}) \), and the map \( r \) is a norm-preserving injective linear between \( L^\infty_\mu(X, \mathcal{A}, \mathbb{C}) \) and the dual \( L^1_\mu(X, \mathcal{A}, \mathbb{C})' \) of \( L^1_\mu(X, \mathcal{A}, \mathbb{C}) \). Furthermore, the map \( r: L^\infty_\mu(X, \mathcal{A}, \mathbb{C}) \to L^1_\mu(X, \mathcal{A}, \mathbb{C})' \) is an isomorphism.

A proof of Theorem 4.49 is given in Lang [52] (Chapter VII, §2). Theorem 4.49 can be generalized to a Hilbert space \( F \), one just has to exercise caution in defining \( l \) and \( r \) to deal with sequilinearity.

The map \( l: L^1_\mu(X, \mathcal{A}, \mathbb{C}) \to L^\infty_\mu(X, \mathcal{A}, \mathbb{C})' \) is not surjective, and understanding which linear forms in \( L^\infty_\mu(X, \mathcal{A}, \mathbb{C})' \) can be represented by functions in \( L^1_\mu(X, \mathcal{A}, \mathbb{C}) \) is a natural question. A partial answer to this question is the Radon–Nikodym theorem, but will this would lead us too far. The interested reader is referred to Lang [52] or Rudin [66].

### 4.8 Products of Measure Spaces and Fubini’s Theorem

The purpose of this section is to define the notion of product measure space and product measure, given two measure spaces \((X, \mathcal{A}, \mu)\) and \((Y, \mathcal{B}, \nu)\). Then we will state Fubini’s theorem (also known as the theorem of Lebesgue–Fubini), which allows us to compute the integral on a product space as two successive integrals. The technical details are surprisingly involved.

We begin by recalling what we did in Example 3.1. We defined the set \( \mathcal{R} \) of rectangles in \( X \times Y \) as follows:
\[
\mathcal{R} = \{ A \times B \in X \times Y \mid A \in \mathcal{A}, \ B \in \mathcal{B} \}.
\]
The set \( \mathcal{R} \) is a semi-algebra, and it can be shown that the set \( \mathcal{B}(\mathcal{R}) \) of finite unions of pairwise disjoint sets in \( \mathcal{R} \) is the smallest algebra containing the semi-algebra \( \mathcal{R} \).

**Definition 4.22.** Let \( \mathcal{A} \otimes \mathcal{B} \) be the smallest \( \sigma \)-algebra generated by \( \mathcal{R} \) (and thus by \( \mathcal{B}(\mathcal{R}) \)); see Proposition 3.3.\(^4\)

The hard part is now to define a product measure \( \lambda \) on \( \mathcal{A} \otimes \mathcal{B} \) which satisfies the natural identity
\[
\lambda(A \times B) = \mu(A)\nu(B)
\]
for all rectangles \( A \times B \). Here as in Section 3.1 we use extended multiplication on \( \mathbb{R}_+ \), where
\[
a \cdot (+\infty) = (+\infty) \cdot a = +\infty
\]

\(^4\)The meaning of the tensor sign \( \otimes \) in the notation \( \mathcal{A} \otimes \mathcal{B} \) is a completely different from its meaning in a tensor product of vector spaces. Hopefully, the two notions will never appear together!
if $0 < a \leq +\infty$, and
\[0 \cdot (+\infty) = (+\infty) \cdot 0 = 0.\]

We need a few definitions.

**Definition 4.23.** Given any subset $E \subseteq X \times Y$, for any $x \in X$, we define the section of $E$ (determined by $x$) as the subset $E_x$ given by
\[E_x = \{y \in Y \mid (x, y) \in E\} \subseteq Y.\]

Similarly, for any $y \in Y$, we define the section of $E$ (determined by $y$) as the subset $E_y$ given by
\[E_y = \{x \in X \mid (x, y) \in E\} \subseteq X.\]

See Figure 4.9.

![Diagram of sections](image)

Figure 4.9: Let $X = Y = \mathbb{R}$. The top figure illustrates an $x$-section of the peach set $E$, while the bottom figure illustrates a $y$-section.

**Proposition 4.50.** The sections of any subset $E \in \mathcal{A} \otimes \mathcal{B}$ are measurable.
**Proof idea.** Let $\mathcal{E}$ be the family of subsets of $X \times Y$ defined as follows:

$$\mathcal{E} = \{ F \subseteq X \times Y \mid F_x \in \mathcal{B} \text{ for all } x \in X, \text{ and } F_y \in \mathcal{A} \text{ for all } y \in Y \}.$$ 

These are the subsets of $X \times Y$ whose sections are measurable. Then prove that $\mathcal{E}$ is a $\sigma$-algebra containing $\mathcal{R}$, which implies that $\mathcal{E} = \mathcal{A} \otimes \mathcal{B}$.

**Definition 4.24.** Given any function $f: X \times Y \to \mathbb{R}$, for any $x \in X$, we define the section of $f$ (determined by $x$) as the function $f_x: Y \to \mathbb{R}$ given by

$$f_x(y) = f(x, y) \text{ for all } y \in Y.$$ 

Similarly, for any $y \in Y$, we define the section of $f$ (determined by $y$) as the function $f_y: X \to \mathbb{R}$ given by

$$f_y(x) = f(x, y) \text{ for all } x \in X.$$ 

**Proposition 4.51.** If $f: X \times Y \to \mathbb{R}$ is a measurable function (on $(X \times Y, \mathcal{A} \otimes \mathcal{B})$, then every section of $f$ is measurable.

**Proof.** By Proposition 3.12, it suffices to show that the inverse image of every open subset of the form $(\alpha, \infty)$ is measurable.

For any $x \in X$, for any $\alpha \in \mathbb{R}$, we have

$$\{ y \in Y \mid f_x(y) < \alpha \} = \{ y \in Y \mid f(x, y) < \alpha \} = \{ (x, y) \in X \times Y \mid f(x, y) < \alpha \}_x,$$

and this last subset is measurable by Proposition 4.50. The proof for $f_y$ is similar.

**Definition 4.25.** Given an algebra $\mathfrak{A}$ of sets, a measure on $\mathfrak{A}$ satisfies the same axioms as a measure on a $\sigma$-algebra; see Definition 3.9.

The next two results take a lot more work.

**Proposition 4.52.** Let $(X, \mathcal{A}, \mu)$ and $(Y, \mathcal{B}, \nu)$ be two measure spaces, and assume that $\mu$ and $\nu$ are $\sigma$-finite. Then the map $\lambda: \mathcal{R} \to [0, +\infty]$ given by

$$\lambda(A \times B) = \mu(A)\nu(B)$$

has a unique extension to a $\sigma$-finite measure on the algebra $\mathcal{B}(\mathcal{R})$.

A proof of Proposition 4.52 can be found in course notes given by Philippe G. Ciarlet in 1970-1971 at ENPC (Paris, France). Interestingly, the proof uses the Monotone Convergence Theorem. A related treatment is given in Halmos [39] (Chapter VII); see also Lang [52] (Chapter VI, §8) and Marle [56] (Chapter 5, Section 2).
Theorem 4.53. Let \((X, \mathcal{A}, \mu)\) and \((Y, \mathcal{B}, \nu)\) be two measure spaces, and assume that \(\mu\) and \(\nu\) are \(\sigma\)-finite. Then the map \(\lambda: \mathcal{R} \to [0, +\infty]\) given by
\[
\lambda(A \times B) = \mu(A)\nu(B)
\]
has a unique extension to a measure \(\lambda = \mu \otimes \nu\) on the \(\sigma\)-algebra \(\mathcal{A} \otimes \mathcal{B}\). The measure \(\mu \otimes \nu\) is \(\sigma\)-finite.

The following properties hold for any measurable subset \(E \in \mathcal{A} \times \mathcal{B}\):

1. We have
\[
(\mu \otimes \nu)(E) = \inf \left\{ \sum_{i=1}^{\infty} \mu(A_i)\nu(B_j) \mid E \subseteq \bigcup_{i=1}^{\infty} (A_i \times B_j), A_i \in \mathcal{A}, B_j \in \mathcal{B} \right\}.
\]

2. The map \(\nu_E\) from \(X\) to \(\mathbb{R}_+\) given by \(x \mapsto \nu(E_x)\) is measurable (w.r.t. \(\mathcal{A}\)), and the map \(\mu_E\) from \(Y\) to \(\mathbb{R}_+\) given by \(y \mapsto \mu(E_y)\) is measurable (w.r.t. \(\mathcal{B}\)). One of these maps is integrable iff the other is integrable.

3. We have
\[
(\mu \otimes \nu)(E) = \begin{cases} 
\int \nu_E \, d\mu = \int \mu_E \, d\nu & \text{if both } \nu_E \text{ and } \mu_E \text{ are integrable} \\
+\infty & \text{otherwise.}
\end{cases}
\]

A proof of Theorem 4.53 can be found in course notes given by Philippe G. Ciarlet in 1970-1971 at ENPC (Paris, France). Again, the proof uses the Monotone Convergence Theorem. A related treatment is given in Halmos [39] (Chapter VII); see also Lang [52] (Chapter VI, §8) and Marle [56] (Chapter 5, Section 2, Proposition 5.2.3).

If \((X, \mathcal{A}, \mu)\) and \((Y, \mathcal{B}, \nu)\) are two measure spaces with \(\mu\) and \(\nu\) both \(\sigma\)-finite. Then for any Banach space \(F\), we have the space of integrable functions \(L_{\mu \otimes \nu}(X \times Y, \mathcal{A} \otimes \mathcal{B}, F)\). The problem is to find a way to compute an integral \(\iint f \, d(\mu \otimes \nu)\), also written \(\iint f \, d\mu \otimes d\nu\), as two successive integrals. The answer is given by a theorem known as Fubini’s theorem. The first version of this theorem was proved by Lebesgue, and then was generalized by Fubini. For this reason some authors refer to this theorem as the Lebesgue–Fubini theorem, but it seems more common to call it simply Fubini’s theorem.

Theorem 4.54. (Fubini’s Theorem, Part 1) Let \((X, \mathcal{A}, \mu)\) and \((Y, \mathcal{B}, \nu)\) be two measure spaces with \(\mu\) and \(\nu\) both \(\sigma\)-finite. Consider a function \(f: X \times Y \to F\), where \(F\) is a Banach space. If \(f \in L_{\mu \otimes \nu}(X \times Y, \mathcal{A} \otimes \mathcal{B}, F)\) then:

1. The section \(f_x: Y \to F\) is \(\nu\)-integrable for almost all \(x \in X\), the section \(f_y: X \to F\) is \(\mu\)-integrable for almost all \(y \in Y\).
2. The map from $X$ to $F$ defined a.e. by

$$x \mapsto \int f_x \, d\nu$$

is $\mu$-integrable, and the map from $Y$ to $F$ defined a.e. by

$$y \mapsto \int f_y \, d\mu$$

is $\nu$-integrable.

Then

$$\int \int f \, d\mu \otimes d\nu = \int_X \left( \int_Y f_x \, d\nu \right) \, d\mu = \int_Y \left( \int_X f_y \, d\mu \right) \, d\nu.$$
Theorem 4.55 is proved in Lang [52] (Chapter VI, §8, Theorem 8.7); see also Marle [56] (Chapter 5, Section 2).

In practice, it is customary to use a less formal notation to express Fubini’s theorem, namely

$$\int \int f(x,y) d\mu(x) \otimes d\nu(y) = \int_X \left( \int_Y f(x,y) d\nu(y) \right) d\mu(x) = \int_Y \left( \int_X f(x,y) d\mu(x) \right) d\nu(y),$$

and the measure $d\mu(x) \otimes d\nu(y)$ is often denoted simply by $d\mu(x)d\nu(y)$.

As an application of the product measure, we define the Lebesgue measure in $\mathbb{R}^n$.

### 4.9 The Lebesgue Measure in $\mathbb{R}^n$

As an application of Theorem 4.53, since the Lebesgue measure $\mu_L$ on $\mathbb{R}$ is $\sigma$-finite, we see that the product measure $\mu_{L,n}$ of $n$ copies of $\mu_L$ is a measure on $\mathbb{R}^n$. The completed $\sigma$-algebra (see Proposition 3.8) obtained from the product algebra $\mathcal{L}(\mathbb{R}) \otimes \cdots \otimes \mathcal{L}(\mathbb{R})$ is called the $\sigma$-algebra of Lebesgue measurable subsets of $\mathbb{R}^n$; it is denoted $\mathcal{L}(\mathbb{R}^n)$. To simplify notation, we may write $\mu_n$ instead of $\mu_{L,n}$, and $\mathcal{L}^1(\mu_n)$ instead of $\mathcal{L}^1_{\mu_n}(\mathbb{R}^n; \mathcal{L}(\mathbb{R}^n), \mathbb{C})$.

A crucial property of the Lebesgue integral is that the space $\mathcal{K}_C^\infty(\mathbb{R}^n)$ of smooth functions with compact support is dense in $\mathcal{L}^1(\mu_n)$. To prove this, one needs to show the existence of smooth “bump functions” in order to approximate the characteristic function $\chi_A$ of a rectangle $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$. The following results are shown in Lang [52] (Chapter VI, Section 9).

**Proposition 4.56.** For any function $f \in \mathcal{L}^1(\mu_n)$, if

$$\int f \varphi d\mu_n = 0 \quad \text{for all } \varphi \in \mathcal{K}_C^\infty(\mathbb{R}^n),$$

then $f = 0$ a.e.

**Proposition 4.57.** For every rectangle $A = [a_1, b_1] \times \cdots \times [a_n, b_n]$, for every $\epsilon > 0$, there exists some functions $\varphi, \psi \in \mathcal{K}_C^\infty(\mathbb{R}^n)$ such that

1. $0 \leq \varphi \leq \chi_A \leq \psi \leq 1$.
2. $\int (\psi - \varphi)d\mu_n < \epsilon$.

Furthermore, $\psi$ vanishes outside the rectangle $[a_1 - \epsilon, b_1 + \epsilon] \times \cdots \times [a_n - \epsilon, b_n + \epsilon]$, and $\varphi \equiv 1$ on the rectangle $[a_1 + \epsilon, b_1 - \epsilon] \times \cdots \times [a_n + \epsilon, b_n - \epsilon]$.

Using the above results, we obtain the following theorem.
4.9. THE LEBESGUE MEASURE IN $\mathbb{R}^n$

**Theorem 4.58.** The space $K^\infty_\infty(\mathbb{R}^n)$ is dense in $L^1(\mu_n)$ (for the $L^1$-semi-norm).

Theorem 4.58 is proven in Lang [52] (Chapter VI, Section 9).

The Lebesgue measure $\mu_n$ on $\mathbb{R}^n$ has the same regularity properties as the Lebesgue measure on $\mathbb{R}$, and we have the following version of Proposition 3.14.

**Proposition 4.59.** For every Lebesgue-measurable set $A \in \mathcal{L}(\mathbb{R}^n)$, the following facts hold:

(a) \[
\mu_n(A) = \inf \{ \mu_n(O) \mid A \subseteq O, \text{ O is open} \}
\]

(b) For every $\epsilon > 0$, if $\mu_n(A)$ has finite measure then there is some open subset $O$ such that $A \subseteq O$ and $\mu_n(O - A) < \epsilon$, and there is some compact subset $F$ such that $F \subseteq A$ and $\mu_n(A - F) < \epsilon$.

Proposition 4.59 is proved in Lang [52] (Chapter VI, Section 9).

The Lebesgue measure on $\mathbb{R}^n$ is translation-invariant, which means that $\mu_n(x + A) = \mu_n(A)$ for all $x \in \mathbb{R}^n$ and all $A \in \mathcal{L}(\mathbb{R}^n)$, where $x + A = \{ x + a \mid a \in A \}$. This will be proved in Section 6.9.

We conclude this section with the change of variables formula. Without this formula, it would be basically impossible to compute the integrals of familiar functions. The proof is not really difficult but quite long and tedious. The interested reader is referred to Lang [52] (Chapter XXI, Section 2, Theorem 2.6).

Given an injective $C^1$ function $f: U \to \mathbb{R}^n$ where $U$ is some open subset of $\mathbb{R}^n$, which means that the derivative $df: U \to \mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ is defined and continuous on $U$, we denote the Jacobian matrix of $df_x$ at $x \in U$ (in the canonical basis of $\mathbb{R}^n$) by $J_f(x)$ (where $\mathcal{L}(\mathbb{R}^n, \mathbb{R}^n)$ denotes the vector space of linear maps from $\mathbb{R}^n$ to itself).

**Theorem 4.60.** (Change of variables formula, I) Let $U$ be an open subset of $\mathbb{R}^n$, and let $f: U \to \mathbb{R}^n$ be an injective $C^1$ function. For every function $g \in L^1(f(U), \mu_n)$, we have $(g \circ f)| \det(J_f| \in L^1(U, \mu_n)$, and

$$
\int_{f(U)} g(x) d\mu_n(x) = \int_{U} (g \circ f)(x)| \det(J_f(x))| d\mu_n(x).
$$

In some cases, for example using polar coordinates, we deal with a $C^1$ function $f: U \to \mathbb{R}^n$ which is only injective on the interior of a measurable subset $A$ of $U$ whose boundary has measure zero. In this case, the following theorem can be used. For a proof, see Lang [52] (Chapter XXI, Section 2, Corollary 2.67).
**Theorem 4.61.** (Change of variables formula, II) Let $U$ be an open subset of $\mathbb{R}^n$, and let $f: U \to \mathbb{R}^n$ be an injective $C^1$ function. Let $A$ be a measurable subset of $U$ whose boundary has measure zero, and such that $f$ is injective on the interior of $A$. For every function $g \in L^1(f(A), \mu_n)$, we have $(g \circ f)\lvert \det(J_f) \in L^1(A, \mu_n)$, and

$$\int_{f(A)} g(x) d\mu_n(x) = \int_A (g \circ f)(x)\lvert \det(J_f(x))\lvert d\mu_n(x).$$
Chapter 5

Radon Functionals and Radon Measures on Locally Compact Spaces

After having considered a very general theory of integration of functions defined on an arbitrary measure space and taking their values in any Banach space, we turn to the special case of complex-valued or real-valued functions defined on a locally compact space \( X \). This corresponds to measure spaces \((X, \mathcal{B}, \mu)\), where \( X \) is a locally compact space, \( \mathcal{B} \) is the \( \sigma \)-algebra of Borel set (which is the smallest \( \sigma \)-algebra containing the open subsets of \( X \)), and \( \mu \) is any (positive) measure on \( \mathcal{B} \), which we call a Borel measure.

The theme of this chapter is that a Borel measure \( \mu \) can be used to define linear forms on various function spaces. For example, pick the space \( K_C(X) \) of continuous functions on \( X \) with compact support. For every function \( f \in K_C(X) \) we can compute the integral

\[
\varphi_\mu(f) = \int f \, d\mu.
\]

We have to check that functions in \( f \in K_C(X) \) are integrable, which is indeed true if \( \mu(K) \) is finite for every compact subset. We obtain a map \( \varphi_\mu : K_C(X) \to \mathbb{C} \), and since the integral is a linear operator, the map \( \varphi_\mu \) is linear. In general it is not continuous, but it satisfies some weaker continuity properties. It is also a positive map, which means that \( \varphi_\mu(f) \geq 0 \) for every positive function \( f \geq 0 \).

What F. Riesz and J. Radon and discovered is that, in some sense to be made precise, a special class of Borel measures is in one-to-one correspondence with the positive linear forms on the space \( K_C(X) \). This means that for every positive linear form \( \Phi \) on \( K_C(X) \), there is a (unique) Borel measure \( m_\Phi \) with some special properties such that \( \Phi \) is represented by \( m_\Phi \), in the sense that

\[
\Phi(f) = \int f \, dm_\Phi \quad \text{for all } f \in K_C(X).
\]

There are two versions of this correspondence theorem known as the Radon–Riesz theorem, depending on the conditions imposed on the Borel measures.
These results are similar in flavor to the fact known from linear algebra that, in a finite-dimensional vector space \( E \) with an inner product \( \langle -, - \rangle \), every linear form \( \varphi \in E^* \) is represented by a unique vector \( u \in E \), in the sense that
\[
\varphi(v) = \langle v, u \rangle \quad \text{for all } v \in E.
\]
If \( (E, \langle -,- \rangle) \) is an infinite-dimensional vector space which is a Hilbert space (it is complete for the norm \( u \mapsto \sqrt{\langle u, u \rangle} \)), then by the Riesz representation theorem, every continuous linear form \( \varphi \in E' \) is represented by a unique vector \( u \in E \), in the sense that
\[
\varphi(v) = \langle v, u \rangle \quad \text{for all } v \in E.
\]

The Radon–Riesz theorems show that certain kinds of (possibly discontinuous) linear forms on \( K_\mathbb{C}(X) \) can be represented using integration instead of an inner product.

The main limitation of this approach is that the linear forms \( \Phi \) induced by a positive measure are positive, which means that \( \Phi(f) \geq 0 \) if \( f \geq 0 \). In particular, it is impossible to represent an arbitrary continuous linear form on \( K_\mathbb{C}(X) \) using integration. The solution to overcome this limitation is to generalize the notion of measure so that a measure can take negative, or even complex, values! We will show how to do this. We will also see that, in the end, complex measures can be expressed in terms of four positive measures, but these positive measures only take finite values in \( \mathbb{R}_+ \). Then we will obtain a third Radon–Riesz correspondence between the continuous linear forms on \( K_\mathbb{C}(X) \) and certain kinds of complex Borel measures. This correspondence plays a crucial role in defining the notion of convolution on a locally compact group.

In this chapter every topological space \( X \) is assumed to be locally compact (and Hausdorff).

## 5.1 Positive Radon Functionals Induced by Borel Measures

One direction of the correspondence (Borel measures \(\mapsto\) linear forms) is easy to describe. It is the observation that the linear forms induced by Borel measures are positive.

**Definition 5.1.** For any function \( f: X \to \mathbb{C} \), we write \( f \geq 0 \) if \( f(X) \subseteq [0, \infty) \). If \( f, g: X \to \mathbb{R} \), we write \( f \leq g \) if \( g - f \geq 0 \). A linear form \( \Phi: K_\mathbb{C}(X) \to \mathbb{C} \) is positive if for every \( f \in K_\mathbb{C}(X) \), if \( f \geq 0 \), then \( \Phi(f) \in \mathbb{R} \) and \( \Phi(f) \geq 0 \).

A positive linear form has the following properties.

**Proposition 5.1.** If \( \Phi: K_\mathbb{C}(X) \to \mathbb{C} \) is a positive linear form, then the following properties hold:

1. For any real-valued function \( f \in K_\mathbb{R}(X) \) we must have \( \Phi(f) \in \mathbb{R} \).
5.1. POSITIVE RADON FUNCTIONALS INDUCED BY BOREL MEASURES

(2) For any two real-valued functions \( f, g \in \mathcal{K}_R(X) \), if \( f \leq g \), then \( \Phi(f) \leq \Phi(g) \).

Proof. Indeed, a real-valued function \( f \) can be written uniquely as \( f = f^+ - f^- \), with \( f^+, f^- \in \mathcal{K}_R(X) \), \( f^+ \geq 0 \) and \( f^- \geq 0 \). Since \( \Phi \) is linear,

\[
\Phi(f) = \Phi(f^+) - \Phi(f^-) \in \mathbb{R},
\]

since \( \Phi(f^+) \geq 0 \) and \( \Phi(f^-) \geq 0 \) as \( \Phi \) is positive.

We have \( f \leq g \) iff \( g - f \geq 0 \), and since \( \Phi \) is positive, \( \Phi(g - f) \geq 0 \), but since \( \Phi \) is linear and positive, \( \Phi(g) - \Phi(f) \geq 0 \) with \( \Phi(f), \Phi(g) \in \mathbb{R} \), that is, \( \Phi(f) \leq \Phi(g) \).

The following proposition yields the mapping from Borel measures to positive linear forms.

**Proposition 5.2.** Assume that the Borel measure \( \mu \) has the property that \( \mu(K) \) is finite for every compact subset of \( X \) (since \( X \) is Hausdorff, a compact set is closed, and thus a Borel set). Every function \( f \in \mathcal{K}_C(X) \) is integrable. Furthermore, the map \( \varphi_\mu : \mathcal{K}_C(X) \to \mathbb{C} \) given by

\[
\varphi_\mu(f) = \int f \, d\mu
\]

is a positive linear form.

Proof. Since \( f \) has compact support, say \( K \), and since it is continuous, it is bounded, say \( |f| \leq M_{\chi_K} \). Since \( f \) is continuous, it is measurable, and the function \( M_{\chi_K} \) is a step function which is integrable since \( \mu(K) \) is finite. By Theorem 4.34, the function \( f \) is integrable. By Proposition 4.23, the map \( \varphi_\mu \) is linear and positive.

Remark: As a point of terminology, the map \( \varphi_\mu : \mathcal{K}_C(X) \to \mathbb{C} \) just is just a linear form, but since its domain is a function space (\( \mathcal{K}_C(X) \)), it is customary to call it a linear functional.

The remarkable fact is that any positive linear functional \( \Phi : \mathcal{K}_C(X) \to \mathbb{C} \) determines a Borel measure \( m_\Phi \) (with some special properties) such that

\[
\Phi(f) = \int f \, dm_\Phi \quad \text{for all } f \in \mathcal{K}_C(X).
\]

Knowing how to integrate functions in \( \mathcal{K}_C(X) \) is sufficient to determine the measure \( m_\Phi \) completely. In some sense, continuous functions with compact support play the role of \( \mu \)-step functions.

Recall that for any compact subset \( K \) of \( X \), we denote by \( \mathcal{K}(K; \mathbb{C}) \) the set of complex-valued continuous functions whose support is contained in \( K \) (and similarly \( \mathcal{K}(K; \mathbb{R}) \) for real-valued functions). Interestingly, every positive linear functional on \( \mathcal{K}_C(X) \) is continuous on \( \mathcal{K}(K; \mathbb{C}) \) for every compact subset \( K \) of \( X \).
Proposition 5.3. If $\Phi: K_C(X) \to \mathbb{C}$ is a positive linear functional on $K_C(X)$ then for every compact subset $K$ of $X$, there is some real number $c_K \geq 0$ such that $|\Phi(f)| \leq c_K \|f\|_{\infty}$ for all $f \in K(K; \mathbb{C})$.

Proof. Every function $f$ in $K(K; \mathbb{C})$ can be written uniquely as $f = f_1 + if_2$ with $f_1, f_2 \in K_{\mathbb{R}}(X)$. Since $\Phi$ is positive linear, we have $\Phi(f_1) \in \mathbb{R}$, $\Phi(f_2) \in \mathbb{R}$ and $\Phi(f) = \Phi(f_1 + if_2) = \Phi(f_1) + i\Phi(f_2)$, so

$$|\Phi(f)| = \sqrt{\Phi(f_1)^2 + \Phi(f_2)^2}.$$

If we can show that $|\Phi(f_1)| \leq c_1 \|f_1\|_{\infty}$ and $|\Phi(f_2)| \leq c_2 \|f_2\|_{\infty}$, then we get

$$|\Phi(f)| \leq \sqrt{c_1^2 \|f_1\|_{\infty}^2 + c_2^2 \|f_2\|_{\infty}^2} \leq \sqrt{c_1^2 + c_2^2} \|f\|_{\infty},$$

since

$$\|f\|_{\infty} = \sup_{x \in K} |f(x)| = \sup_{x \in K} \sqrt{(f_1(x))^2 + (f_2(x))^2} = \sqrt{\sup_{x \in K} (f_1(x))^2 + \sup_{x \in K} (f_2(x))^2},$$

which implies that

$$\|f_1\|_{\infty} = \sup_{x \in K} |f_1(x)| \leq \|f\|_{\infty},$$

and

$$\|f_2\|_{\infty} = \sup_{x \in K} |f_2(x)| \leq \|f\|_{\infty}.$$

Therefore we may assume that $f \in K(K; \mathbb{R})$. By Proposition 15.39, there is a continuous function with compact support $g \in K_C(X)$ (a bump function) such that $g(x) = 1$ for all $x \in K$. For any $f \in K_{\mathbb{R}}(X)$, we have

$$-g \|f\|_{\infty} \leq f \leq g \|f\|_{\infty}$$

and since $\Phi$ is a positive linear functional, by Proposition 5.1(2), we get

$$-\Phi(g) \|f\|_{\infty} \leq \Phi(f) \leq \Phi(g) \|f\|_{\infty}$$

that is

$$|\Phi(f)| \leq \Phi(g) \|f\|_{\infty},$$

as desired \qed

Proposition 5.3 suggests that the linear functionals $\Phi: K_C(X) \to \mathbb{C}$ satisfying the conclusion of the proposition are of particular interest, and they are. In fact the measure theory and the integration theory for complex-valued functions on a locally compact space can be developed entirely in terms of these functionals. This approach is presented in Dieudonné [22], Bourbaki [5, 7, 9], and Schwartz [72]. Dieudonné and Bourbaki even go so far as calling such functionals *measures*, which we feel is unfortunate because this term already has a well
established meaning. Unlike these two previous sources, Schwartz actually develops in parallel both the theory of integration using measure theory, and the theory of integration using certain linear functionals that he calls Radon measures. Again, we find this terminology unfortunate because these are functionals and not measures in the traditional sense. We propose to use the term Radon functional.

**Definition 5.2.** A linear functional $\Phi: K_C(X) \to \mathbb{C}$ is a Radon functional if for every compact subset $K$ of $X$, there is some real number $c_K \geq 0$ such that $|\Phi(f)| \leq c_K \|f\|_{\infty}$ for all $f \in K_C(K; \mathbb{C})$. The set of Radon functionals is denoted $M_C(X)$, or simply, $M(X)$. The set of positive Radon functionals is denoted $M^+(X)$, and the set of continuous (or bounded) Radon functionals is denoted $M^1(X)$. See Figure 5.1.

Equivalently, a linear functional is a Radon functional if it is continuous when restricted to $K_C(K; \mathbb{C})$, for every compact subset $K$ of $X$.

In general, a Radon functional is *not* continuous on $K_C(X)$ for the sup norm $\|\cdot\|_{\infty}$. For a continuous Radon functional, there is a *uniform constant* $c \geq 0$ such that

$$|\Phi(f)| \leq c \|f\|_{\infty} \quad \text{for all } f \in K_C(X).$$

Continuous Radon functionals are often called bounded Radon functionals.

**Proposition 5.4.** Any positive linear functional $\Phi: K_C(X) \to \mathbb{C}$ is a positive Radon functional.
Observe that a Radon functional $\Phi : \mathcal{K}_C(X) \to \mathbb{C}$ is completely determined by its restriction $\Phi_R : \mathcal{K}_R(X) \to \mathbb{C}$ to the space of real-valued functions in $\mathcal{K}_R(X)$. Indeed, every function $f \in \mathcal{K}_C(X)$ can be written uniquely as $f = f_1 + if_2$ with $f_1, f_2 \in \mathcal{K}_R(X)$, and by $\mathbb{C}$-linearity,

$$\Phi(f) = \Phi(f_1 + if_2) = \Phi_R(f_1) + i\Phi_R(f_2).$$

Furthermore, if $\Phi$ is a positive Radon functional, then by Proposition 5.1 we have $\Phi(f) \in \mathbb{R}$ for all $f \in \mathcal{K}_R(X)$, so $\Phi_R : \mathcal{K}_R(X) \to \mathbb{R}$. Therefore, there is a bijection between the space $\mathcal{M}^+(X)$ of positive linear functionals $\Phi : \mathcal{K}_C(X) \to \mathbb{C}$ and the space $\mathcal{M}_R^+(X)$ of positive linear functionals $\Psi : \mathcal{K}_R(X) \to \mathbb{R}$ as illustrated by Figure 5.2.

![Figure 5.2: The correspondence between $\mathcal{M}^+(X)$ and $\mathcal{M}_R^+(X)$.](image)

Also observe that $\mathcal{M}(X)$ and $\mathcal{M}^1(X)$ are vector spaces. The operator norm $\|\|_\Phi$ is well defined on the vector space $\mathcal{M}^1(X)$. For any bounded linear functional $\Phi$, by definition

$$\|\Phi\| = \sup\{|\Phi(f)| \mid f \in \mathcal{K}_C(X), \|f\|_\infty = 1\}.$$  

Using Proposition 1.19 it is easy to show that $\mathcal{M}^1(X)$ is isomorphic to the dual $\mathcal{C}_0(X; \mathbb{C})'$ of the space $\mathcal{C}_0(X; \mathbb{C})$, that is, the space of all continuous linear forms on $\mathcal{C}_0(X; \mathbb{C})$. Recall that $\mathcal{C}_0(X; \mathbb{C})$ is the space of continuous functions which tend to 0 at infinity; see Definition 1.18.
Proposition 5.5. Let \( X \) be a locally compact space. The space \( M^1(X) \) of bounded Radon functionals is isomorphic to the dual \( C_0(X; \mathbb{C})' \) of \( C_0(X; \mathbb{C}) \), that is, the space of all continuous linear forms on \( C_0(X; \mathbb{C}) \). Consequently \( M^1(X) \) is a Banach space (w.r.t. the sup norm).

Proof. By Proposition 1.19, the space \( C_0(X; \mathbb{C}) \) is the closure of \( \mathcal{K}_C(X) \). By definition, \( M^1(X) \) is the space of continuous linear forms on \( \mathcal{K}_C(X) \). By Theorem 15.69, every continuous linear form has a unique continuous extension to \( C_0(X; \mathbb{C}) \). Therefore \( M^1(X) \) is isomorphic to the dual of \( C_0(X; \mathbb{C}) \). Since \( \mathbb{C} \) is complete, it is known that the set of continuous linear maps from any vector space into \( \mathbb{C} \) is complete.

Here are some example of Radon functionals.

Example 5.1.

1. Pick any \( a \in X \). The map \( \delta_a \) given by
   \[
   \delta_a(f) = f(a)
   \]
   for all \( f \in \mathcal{K}_C(X) \) is a Radon functional called (with an abuse of terminology) the Dirac measure. Since \( |f(a)| \leq \|f\|_\infty \), it is a bounded Radon functional.

2. Consider the space \( \mathcal{K}_C(\mathbb{R}) \) of continuous functions \( f : \mathbb{R} \to \mathbb{C} \) with compact support. For each function \( f \in \mathcal{K}_C(\mathbb{R}) \), there is a compact interval \([a, b]\) such that \( f \) vanishes outside of \([a, b]\), and from Section 2.1, the Riemann integral
   \[
   I(f) = \int_a^b f(t)dt
   \]
   is defined. We obtain a map \( I : \mathcal{K}_C(\mathbb{R}) \to \mathbb{C} \) which is obviously linear. Since
   \[
   \left| \int_a^b f(t)dt \right| \leq (b - a) \|f\|_\infty,
   \]
   this map is a Radon functional. Actually, this functional is positive. We will see later that this Radon functional corresponds to the Lebesgue measure.

3. Let \( \Phi \) be any Radon functional and pick any continuous function \( g \in C(X; \mathbb{C}) \). It is clear that if \( f \in \mathcal{K}_C(X) \), then \( gf \in \mathcal{K}_C(X) \), and we have a map \( \Psi \) given by
   \[
   \Psi(f) = \Phi(gf) \quad \text{for all } f \in \mathcal{K}_C(X).
   \]
   Clearly, this is a linear functional. For any compact subset \( K \) of \( X \), if \( f \in \mathcal{K}_C(X) \), then we have
   \[
   \|gf\|_\infty \leq \|f\|_\infty \sup_{x \in K} |g(x)|.
   \]
Since $\Phi$ is a Radon functional, there is some real $c_K \geq 0$ such that

$$|\Phi(gf)| \leq c_K \|gf\|_{\infty},$$

so we obtain

$$|\Phi(gf)| \leq c_K \sup_{x \in K} |g(x)| \|f\|_{\infty},$$

which shows that $\Psi$ is a Radon functional. The Radon functional $\Psi$ is called the Radon functional with density $g$ relative to $\Phi$, and it is denoted $g \cdot \Phi$. Such Radon functionals play an important role in the definition of the notion of convolution in the theory of integration based on Radon functionals developed in Dieudonné [22] and Bourbaki [5, 7, 9, 6].

In the next section, we state the most important theorem of the theory of Radon functionals, which is that every positive Radon functional arises from a unique Borel measure with some regularity properties.

### 5.2 The Radon–Riesz Theorem and Positive Radon Functionals

In this section we deal with the direction of the correspondence positive Radon functionals $\Rightarrow$ Borel measures. Our first goal is to show that for every positive Radon functional $\Phi$, there is a $\sigma$-algebra $\mathcal{M}$ and a unique positive measure $m_\Phi$ on $\mathcal{M}$ (with certain properties) representing $\Phi$ as an integral, which means that

$$\Phi(f) = \int f \, dm_\Phi \quad \text{for all } f \in \mathcal{K}_C(X).$$

For instance, the positive Radon functional of Example 5.1(2) yields the Lebesgue measure. In a second stage, by imposing some reasonable conditions on the measure we obtain a bijective correspondence.

Complete proofs of these results are quite long and intricate. Such proofs can be found in Rudin [66] (Chapter 2), Lang [52] (Chapter IX), Folland [32] (Chapter 7, Theorem 7.2), and Schwartz [72] (Chapters 5 and 7). Going back and forth between Rudin, Folland, and Lang is a possible strategy to understanding the proof.

Theorem 5.6 is often referred to as the Riesz representation theorem. A version of this theorem for $X = [0, 1]$ was first proven by Frigyes Riesz\(^1\) in 1909. In 1913, Radon extended Riesz’ result to a compact subset of $\mathbb{R}^n$ in terms of regular measures rather than a Stieltjes integral. Following Malliavin [55], it seems appropriate to call it the Radon–Riesz theorem, but it should be noted that other versions of this theorem were obtained by Banach, Saks, Markov, and Kakutani, which gives the most general version stated in Theorem 5.29; see Dunford and Schwartz [28].

\(^1\)Not to be confused with his younger brother Marcel Riesz.
5.2. THE RADON–RIESZ THEOREM AND POSITIVE RADON FUNCTIONALS

**Theorem 5.6.** (Radon–Riesz) Let $X$ be a locally compact (Hausdorff) space. For every positive linear functional $\Phi: \mathcal{K}_c(X) \to \mathbb{C}$, there is a $\sigma$-algebra $\mathcal{M}$ containing the Borel $\sigma$-algebra, and there is a unique positive measure $m_\Phi$ on $\mathcal{M}$ with the following properties:

1. The linear functional $\Phi$ is represented by $m_\Phi$, that is,
   \[ \Phi(f) = \int f \, dm_\Phi \quad \text{for all } f \in \mathcal{K}_c(X). \]

2. The measure $m_\Phi(K)$ is finite for every compact subset $K$ of $X$.

3. We have
   \[ m_\Phi(E) = \inf \{ m_\Phi(V) \mid E \subseteq V, \text{ V open} \} \]
   for every $E \in \mathcal{M}$.

4. We have
   \[ m_\Phi(E) = \sup \{ m_\Phi(K) \mid K \subseteq E, \text{ K compact} \} \]
   for every open subset $E$, and for every $E \in \mathcal{M}$ with $m_\Phi(E) < +\infty$.

5. For any $E \in \mathcal{M}$ and any $A \subseteq E$, if $m_\Phi(E) = 0$, then $m_\Phi(A) = 0$, in other words, $m_\Phi$ is a complete measure.

Let us make a few comments about the proof. The uniqueness of $m_\Phi$ is not so bad. Observe that by (3) and (4), the measure $m_\Phi$ is determined by its values on compact subsets. Hence it suffices to prove that if two measures $\mu_1$ and $\mu_2$ satisfy the theorem, then they agree on all compact subsets.

Pick any compact $K$ and any $\epsilon > 0$. By (3) and (4), there is some open subset $V$ such that $K \subseteq V$ and $\mu_2(V) < \mu_2(K) + \epsilon$. By Proposition 15.39, there is a continuous function $f: X \to [0, 1]$ such that $f(x) = 1$ for all $x \in K$, and such that $\text{supp}(f)$ is compact and $\text{supp}(f) \subseteq V$; this implies that

\[
\mu_1(K) = \int \chi_K \, d\mu_1 \\
\leq \int f \, d\mu_1 = \Phi(f) = \int f \, d\mu_2 \\
\leq \int \chi_V \, d\mu_2 = \mu_2(V) \\
< \mu_2(K) + \epsilon.
\]

Therefore, $\mu_1(K) \leq \mu_2(K)$. By swapping $\mu_1$ and $\mu_2$, we obtain $\mu_2(K) \leq \mu_1(K)$, and thus $\mu_1(K) = \mu_2(K)$. Observe that the above derivation also shows that $m_\Phi$ is finite for every compact subset $K$.

To construct $m_\Phi$, we proceed as follows; for simplicity of notation, write $\mu$ instead of $m_\Phi$. 

(a) For every open set $V$ in $X$, for every continuous function $g: X \to \mathbb{R}$, write $g \prec V$ if $g: X \to [0,1]$, supp($g$) is compact, and supp($g$) $\subseteq V$. Let

$$
\mu(V) = \sup \{ \Phi(g) \mid g \prec V \}.
$$

This will force Condition (4).

(b) Next, to force Condition (3), we extend $\mu$ to arbitrary subsets. For every $E \subseteq X$, let

$$
\mu(E) = \inf \{ \mu(V) \mid E \subseteq V, \text{ V open} \}.
$$

It can be checked that $\mu$ is an outer measure.

(c) In order to obtain a $\sigma$-algebra and a measure, we need to cut down the family of subsets, still forcing Conditions (3) and (4). Let $\mathcal{A}$ be the family of all subsets $A$ of $X$ such that $\mu(A) < +\infty$ and

$$
\mu(A) = \sup \{ \mu(K) \mid K \subseteq A, \ K \text{ compact} \}.
$$

Then $\mathcal{A}$ is an algebra containing all compact sets and all open sets of finite measure. The map $\mu$ is a measure on $\mathcal{A}$, and if $\mu(A) < +\infty$ then $A \in \mathcal{A}$.

(d) Let $\mathcal{M}$ be the family of all subsets $Y$ of $X$ such that $Y \cap K$ lies in $\mathcal{A}$ for all compact subsets $K$. Then $\mathcal{M}$ is the desired $\sigma$-algebra containing the Borel sets, and $\mu$ is a positive measure on $\mathcal{M}$. The algebra $\mathcal{A}$ consists of the sets of finite measure in $\mathcal{M}$.

Having done all this, one still needs to check that Conditions (1), (3), and (4) hold. Proposition 15.40 (existence of finite partitions of unity) is used for some of these checks.

Theorem 5.6 shows that the measure that arises from a positive linear functional has special regularity properties that we already encountered when we met the Lebesgue measure in Section 3.4.

**Definition 5.3.** A Borel measure $\mu$ on the Borel $\sigma$-algebra $\mathcal{B}$ of a locally compact space $X$ is $\sigma$-regular if the following two conditions hold:

For every $E \in \mathcal{B}$,

$$
\mu(E) = \inf \{ \mu(V) \mid E \subseteq V, \text{ V open} \}. \quad (*)
$$

For every open subset $E$, and for every $E \in \mathcal{B}$ with $\mu(E) < +\infty$,

$$
\mu(E) = \sup \{ \mu(K) \mid K \subseteq E, \ K \text{ compact} \}. \quad (**_{\sigma})
$$

Condition (*) is called outer regularity, and Condition (**$_{\sigma}$) is called $\sigma$-inner regularity.

We say that $\mu$ is locally finite if $\mu(K)$ is finite for every compact subset $K$.

The following proposition justifies the terminology $\sigma$-inner regularity.
5.2. THE RADON–RIESZ THEOREM AND POSITIVE RADON FUNCTIONALS

**Proposition 5.7.** Let $X$ be a locally compact (Hausdorff) space. If a Borel measure $\mu$ is $\sigma$-inner regular, then

$$\mu(E) = \sup \{\mu(K) \mid K \subseteq E, K \text{ compact}\}$$

(\ast \ast_{\sigma})

holds for every $\sigma$-finite subset $E \in \mathcal{B}$.

**Proof.** Say $E = \bigcup_{i=1}^{\infty} E_i$ with $E_i \in \mathcal{B}$ and $\mu(E_i) < +\infty$. We may assume that $\mu(E) = +\infty$, since if $\mu(E) < +\infty$ then we already have $\sigma$-inner regularity by definition. For every $M > 0$, there is some $n \geq 1$ such that $\mu\left(\bigcup_{i=1}^{n} E_i\right) > M$. Since $\bigcup_{i=1}^{n} E_i$ has finite measure, $\sigma$-inner regularity applies, so there is some compact subset $K$ such that $\mu(K) > M$. This shows that

$$\sup \{\mu(K) \mid K \subseteq E, K \text{ compact}\} = +\infty = \mu(E),$$

which shows $\sigma$-inner regularity for $E$. 

**Definition 5.4.** Let $X$ be a locally compact (Hausdorff) space. A Borel measure $\mu$ is called a (positive) $\sigma$-Radon measure if it is $\sigma$-regular and locally finite. The space of $\sigma$-Radon measures is denoted by $\mathcal{M}_{\sigma}^{+}(X)$. See Figure 5.3.

Theorem 5.6 immediately implies the following correspondence which we illustrate in Figure 5.4.

**Theorem 5.8.** (Radon–Riesz Correspondence, I) Let $X$ be a locally compact (Hausdorff) space. The maps $m: \mathcal{M}^{+}(X) \to \mathcal{M}_{\sigma}^{+}(X)$ and $\varphi: \mathcal{M}_{\sigma}^{+}(X) \to \mathcal{M}^{+}(X)$ given by

$$m(\Phi) = m_{\Phi} \quad \text{for all } \Phi \in \mathcal{M}^{+}(X)$$

$$\varphi(\mu) = \varphi_{\mu} \quad \text{for all } \mu \in \mathcal{M}_{\sigma}^{+}(X)$$

are mutual inverses that define a bijection between the space $\mathcal{M}^{+}(X)$ of positive Radon functionals and the space $\mathcal{M}_{\sigma}^{+}(X)$ of (positive) $\sigma$-Radon measures.

![Figure 5.3: A Venn diagram classification of Borel measures.](image-url)
In the next section, we show that by requiring the locally compact space \( X \) to be also \( \sigma \)-compact, then we obtain Borel measures that are not only \( \sigma \)-regular, but regular as well, which means that inner regularity holds for all \( E \in \mathcal{B} \).

### 5.3 Regular Borel Measures

In Theorem 5.6, outer regularity holds, but \( \sigma \)-inner regularity holds only for open subsets and measurable sets of finite measure. It is often desirable for inner regularity to hold for arbitrary subsets \( E \in \mathcal{B} \), possibly not \( \sigma \)-finite. It turns out that making some mild restrictions on \( X \), we obtain a bijection between positive linear functionals and these regular measures. On this subject, Rudin’s exposition seems clearer than Lang’s exposition.

**Definition 5.5.** A Borel measure \( \mu \) on the Borel \( \sigma \)-algebra \( \mathcal{B} \) of a locally compact space \( X \) is **regular** if the following two conditions hold for every \( E \in \mathcal{B} \):

\[
\mu(E) = \inf\{\mu(V) \mid E \subseteq V, \ V \text{ open}\}
\]

\[(*)
\]

and

\[
\mu(E) = \sup\{\mu(K) \mid K \subseteq E, \ K \text{ compact}\}.
\]

\[(**)
\]
Condition (*) is called *outer regularity*, and Condition (***) is called *inner regularity*. See Figure 5.5.

![Figure 5.5: Another Venn diagram classification of Borel measures.](image)

Observe that if a Borel measure $\mu$ is $\sigma$-finite (on $X$) and if it is $\sigma$-regular, then it is actually regular. Another sufficient condition is given in Proposition 5.11.

A way to obtain the Radon–Riesz correspondence between positive Radon functionals and regular locally finite Borel measures is to require $X$ to be $\sigma$-compact, which means that $X$ is the countable union of compact subsets (see Definition 15.35).

**Theorem 5.9.** Let $X$ be a locally compact (Hausdorff), $\sigma$-compact space. For every positive linear functional $\Phi: K_\mathbb{C}(X) \to \mathbb{C}$, if $\mathcal{M}$ and $m_\Phi$ are the $\sigma$-algebra and the measure obtained in Theorem 5.6, then the following properties holds:

1. For any $E \in \mathcal{M}$ and any $\epsilon > 0$, there is a closed set $F$ and an open set $O$ such that $F \subseteq E \subseteq O$ and $\mu(O - F) < \epsilon$.
2. The measure $m_\Phi$ is a regular, locally finite Borel measure on the Borel $\sigma$-algebra $\mathcal{B}$.

Theorem 5.9 is proven in Rudin [66] (Chapter 2, Theorem 2.17). The following theorem allows us to get a bijective correspondence between positive linear functional and regular locally finite Borel measures, and to state this theorem it is convenient to introduce the following definition.

**Definition 5.6.** Let $X$ be a locally compact (Hausdorff) space. A Borel measure $\mu$ is called a *positive Radon measure* if it is regular and locally finite. The space of Radon measures is denoted by $\mathcal{M}^+_{\text{rad}}(X)$, or simply $\mathcal{M}^+(X)$. See Figure 5.5.
**Theorem 5.10.** (Radon–Riesz Correspondence, II) Let $X$ be a locally compact (Hausdorff), $\sigma$-compact space. The maps $m: \mathcal{M}^+(X) \to \mathcal{M}^+(X)$ and $\varphi: \mathcal{M}^+(X) \to \mathcal{M}^+(X)$ given by

$$m(\Phi) = m_\Phi \quad \text{for all } \Phi \in \mathcal{M}^+(X)$$

$$\varphi(\mu) = \varphi_\mu \quad \text{for all } \mu \in \mathcal{M}^+(X)$$

are mutual inverses that define a bijection between the space $\mathcal{M}^+(X)$ of positive Radon functionals and the space $\mathcal{M}^+(X)$ of (positive) Radon measures. See Figure 5.6.

The following proposition gives us a sufficient condition for a locally finite Borel measure to be regular.

**Proposition 5.11.** Let $X$ be a locally compact (Hausdorff) space in which every open subset is $\sigma$-compact. If $\mu$ is a locally finite Borel measure, then $\mu$ is a regular measure.

Proposition 5.11 is proven in Rudin [66] (Chapter 2, Theorem 2.18).

Observe that $X = \mathbb{R}^n$ satisfies the condition of Proposition 5.11. Thus a locally finite Borel measure on $\mathbb{R}^n$ is a regular measure.
An interesting application of Theorem 5.10 is obtained by choosing $X = \mathbb{R}$ and $\Phi$ to be the Radon functional $I$ induced by the Riemann integral defined in Example 5.1(2). The Radon measure $m_I$ given by Theorem 5.10 turns out to be the Lebesgue measure $\mu_L$. For details, see Rudin [66] (Chapter 2).

Measurable functions on a locally compact space with a $\sigma$-regular, locally finite, Borel measure are very close to being continuous as stated in the following theorem of Lusin.

**Theorem 5.12.** (Lusin’s Theorem) Let $X$ be a locally compact space equipped with a $\sigma$-regular, locally finite, Borel measure $\mu$, and let $f$ be any measurable function on $X$. If $f$ vanishes outside of a set $A$ of finite measure, for any $\epsilon > 0$, there is some function $g \in \mathcal{K}_C(X)$ and a measurable set $Z$ with $\mu(Z) < \epsilon$, such that $f(x) = g(x)$ for all $x \in X - Z$, and $\|g\|_\infty \leq \|f\|_\infty$.

Theorem 5.12 is proven in Rudin [66] (Chapter 2, Theorem 2.24) and Lang [52] (Chapter IX, Theorem 3.3).

The Vitali–Carathéodory theorem states that every function in $L^1_\mu(X, \mathcal{B}, \mathbb{C})$ can be approximated from below and from above by certain kinds of functions called upper semicontinuous and lower semicontinuous, see Rudin [66] (Chapter 2, Theorem 2.25).

We have the following density result which uses Lusin’s theorem (Theorem 5.12).

**Theorem 5.13.** Let $X$ be a locally compact space equipped with a $\sigma$-regular, locally finite, Borel measure $\mu$. The space $\mathcal{K}_C(X)$ is dense in $L^p_\mu(X, \mathcal{B}, \mathbb{C})$ for $p = 1, 2$.

Theorem 5.13 is proved in Rudin [66] (Chapter 3, Theorem 3.14) and Lang [52] (Chapter IX, Theorem 3.1). The following proposition is needed for proving the uniqueness of the Haar measure up to a constant.

**Proposition 5.14.** Let $X$ be a locally compact space equipped with a $\sigma$-regular, locally finite, Borel measure $\mu$. For any function $f \in L^1_\mu(X, \mathcal{B}, \mathbb{C})$, if

$$\int_A f g \, d\mu = 0 \quad \text{for all } g \in \mathcal{K}_C(X),$$

then $f = 0$ almost everywhere.

*Proof.* We use Proposition 4.38, recalling the fact that $\int_A f \, d\mu = \int f \chi_A \, d\mu$. Let $A$ be any subset of finite measure. Then $\chi_A$ is the $L^1$-limit of a sequence $(g_n)$ of functions $g_n \in \mathcal{K}_C(X)$ with $g_n(X) \subseteq [0, 1]$. By Proposition 4.25, there is a subsequence $(g_{n_k})_{k \geq 1}$ that converges pointwise to $\chi_A$ a.e., and thus $(fg_{n_k})$ converges pointwise to $f \chi_A$ a.e. By the dominated convergence theorem, we conclude that $\int_A f \, d\mu = \int f \chi_A \, d\mu = 0$ for all subsets $A$ of finite measure, and by Proposition 4.38, we have $f = 0$ a.e.

---

$^2$Even for all $p$ with $1 \leq p < +\infty$. 
5.4 Complex and Real Measures

By Proposition 5.2, the functionals induced by Borel measures are positive, but there are Radon functionals that are not positive, so it is natural to ask if such functionals arise from some generalized measures allowed to take negative values, or even complex values. The answer is yes. It is even possible to define measures with values in any Banach space. Such measures are discussed in Lang [52], Schwartz [72] and Marle [56], but for simplicity we will only consider real and complex measures. In this section we take a small detour to define complex measures. Then we will show how they relate to functionals on $K_C(X)$ that are not necessarily positive, but continuous.

Going back to Definition 3.9, a (positive) measure on a measurable set $(X, \mathcal{A})$ is a map $\mu$ satisfying the following properties:

\begin{enumerate}
\item[(\mu 1)] $\mu : \mathcal{A} \to [0, +\infty]$ where $\mathcal{A}$ is a $\sigma$-algebra of subsets of $X$.
\item[(\mu 2)] $\mu(\emptyset) = 0$.
\item[(\mu 3)] For any countable sequence $(A_i)_{i \geq 1}$ of subsets $A_i$ of $\mathcal{A}$ such that $A_i \cap A_j = \emptyset$ for all $i \neq j$,
\[ \mu \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu(A_i). \]
\end{enumerate}

Such a function may have the value $+\infty$, but in (\mu 3), if $A = \bigcup_{i=1}^{\infty} A_i$ and if $\mu(A)$ is finite, then the series $\sum_{i=1}^{\infty} \mu(A_i)$ converges, and since it consists of nonnegative numbers, it converges absolutely, and thus commutatively, which means that for any permutation $\sigma$ of $\mathbb{N}_+$, we have
\[ \mu(A) = \mu \left( \bigcup_{i=1}^{\infty} A_{\sigma(i)} \right) = \sum_{i=1}^{\infty} \mu(A_{\sigma(i)}). \]

If we replace $[0, +\infty]$ by $\mathbb{R}$ or $\mathbb{C}$, then a new problem arises, namely that the convergence of the sum $\sum_{i=1}^{\infty} \mu(A_i)$ generally depends on the order of the $A_i$. The solution is to require commutative convergence of the series arising in (\mu 3). It is known from analysis that for $\mathbb{R}$ or $\mathbb{C}$, a series is commutatively convergent iff it is absolutely convergent, so we require the latter. We also require $\mu(A)$ be an element of $\mathbb{R}$ or $\mathbb{C}$, that is, $\mu(A)$ must be “finite.” There is a way to define measures with values in $\mathbb{R} \cup \{+\infty\}$, and even in $\mathbb{R} \cup \{-\infty, +\infty\}$, but we have no need for such generality (see Schwartz [72], Chapter V, §9).

**Definition 5.7.** Let $(X, \mathcal{A})$ be a measurable space. A complex measure on $(X, \mathcal{A})$ is a map $\mu$ satisfying the following properties:

\begin{enumerate}
\item[(\mu 1)] $\mu : \mathcal{A} \to \mathbb{C}$.
\item[(\mu 2)] $\mu(\emptyset) = 0$.
\end{enumerate}
(μ3) For any countable family \((A_i)_{i \geq 1}\) of subsets \(A_i\) of \(\mathcal{A}\) such that \(A_i \cap A_j = \emptyset\) for all \(i \neq j\),

\[
\mu \left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mu(A_i),
\]

where the series on the right-hand side is absolutely convergent.

A real measure (or signed measure) is a complex measure such that \(\mu(A) \subseteq \mathbb{R}\).

Observe that a real measure which is also positive is a positive measure according to Definition 3.9, but since a positive measure may take the value \(+\infty\), there are positive measures that are not real measures in the sense of Definition 5.7. When we use the term positive real measure, we mean that this measure only takes finite values. By positive measure, we mean a measure that may take the value \(+\infty\).

One might wonder if interesting real or complex measures exist. Indeed, for any arbitrary measure space \((X, \mathcal{A}, \mu)\), every function \(f \in L_\mu(X, \mathcal{A}, \mathbb{C})\) gives rise to such a measure.

**Proposition 5.15.** Let \((X, \mathcal{A}, \mu)\) be a measure space (here, \(\mu\) is a positive measure). For every integrable map \(f \in L_\mu(X, \mathcal{A}, \mathbb{C})\), the function \(\mu_f : \mathcal{A} \to \mathbb{C}\) given by

\[
\mu_f(A) = \int_A f \, d\mu = \int f \chi_A \, d\mu \quad \text{for all } A \in \mathcal{A}
\]

is a complex measure.

What is not obvious is that (μ3) holds. This follows from Proposition 4.36 (a consequence of the Lebesgue dominated convergence theorem). A detailed proof is given in Marle [56] (Chapter 1, Proposition 2.5.2).

The new twist here is that given a measure \(\mu\), rather than defining a functional by *varying the function being integrated*, we fix a function but we integrate by *varying the subset over which we integrate*.

It is trivial to check that the complex measures (and the real measures) form a vector space.

Remarkably, every complex measure \(\mu\) arises as a measure of the form \(|\mu|h\) for some suitable positive measure \(|\mu|\) and some well chosen function \(h \in L_{|\mu|}(X, \mathcal{A}, \mathbb{C})\); see Theorem 5.20. The measure \(|\mu|\) is defined as follows.

**Definition 5.8.** Let \((X, \mathcal{A})\) be a measurable space, and let \(\mu\) be a complex measure on \((X, \mathcal{A})\). Define the map \(|\mu| : \mathcal{A} \to [0, +\infty]\) by

\[
|\mu|(A) = \sup \sum_{i=1}^{\infty} |\mu(A_i)|,
\]

for all \(A \in \mathcal{A}\) and for all countable partitions \((A_i)_{i \geq 1}\) of \(A\) with \(A_i \in \mathcal{A}\). The map \(|\mu|\) is called the total variation measure (for short total variation) of \(\mu\).
Obviously, if $\mu$ is a real positive measure, then $|\mu| = \mu$. It is easy to see that by definition, 

$$|\mu(A)| \leq |\mu|(A) \quad \text{for all } A \in \mathcal{A}.$$ 

In fact, it is minimal with this property. We have the following remarkable theorems.

**Theorem 5.16.** Let $(X, \mathcal{A})$ be a measurable space, and let $\mu$ be a complex measure on $(X, \mathcal{A})$. The map $|\mu| : \mathcal{A} \to [0, +\infty]$ is a positive measure. The positive measure $|\mu|$ is the minimal measure such that 

$$|\mu(A)| \leq |\mu|(A) \quad \text{for all } A \in \mathcal{A},$$ 

in the sense that if $\lambda$ is any positive measure such that 

$$|\mu(A)| \leq \lambda(A) \quad \text{for all } A \in \mathcal{A},$$ 

then $|\mu| \leq \lambda$ (which means that $|\mu|(A) \leq \lambda(A)$ for all $A \in \mathcal{A}$).

A proof of Theorem 5.16 is given in Rudin [66] (Chapter 6, Theorem 6.2) and Lang [52] (Chapter VII, Theorem 3.1).

The next theorem is even more surprising.

**Theorem 5.17.** Let $(X, \mathcal{A})$ be a measurable space, and let $\mu$ be a complex measure on $(X, \mathcal{A})$. The map $|\mu| : \mathcal{A} \to [0, +\infty]$ is a finite positive measure; that is, $|\mu|(X) < +\infty$.

A proof of Theorem 5.17 is given in Rudin [66] (Chapter 6, Theorem 6.4) and Lang [52] (Chapter VII, Theorem 3.2). Theorem 5.17 implies that $\mu(X)$ is bounded: it is contained in a closed disk of finite radius. This fact shows that the convergence requirement of Condition ($\mu3$) is quite strong.

Theorem 5.17 allows us to make the space of complex measures into a normed vector space.

**Definition 5.9.** Let $(X, \mathcal{A})$ be measurable space. For any complex measure $\mu$, define $\|\mu\|$ as $\|\mu\| = |\mu|(X)$. The vector space of complex measures equipped with the norm defined above is denoted $\mathcal{CM}^1(X, \mathcal{A})$.

It is not hard to show that $\mathcal{CM}^1(X, \mathcal{A})$ is a Banach space.

**Proposition 5.18.** Let $(X, \mathcal{A})$ be a measurable space. The normed vector space $\mathcal{CM}^1(X, \mathcal{A})$ is a Banach space (it is complete).

Another interesting fact is that if $\mu$ is a positive measure (possibly taking the value $+\infty$) then $L^1_\mu(X, \mathcal{A}, \mathbb{C})$ can be embedded in $\mathcal{CM}^1(X, \mathcal{A})$.

**Proposition 5.19.** Let $(X, \mathcal{A}, \mu)$ be a measure space. The map $f \mapsto \mu_f$ is a linear embedding of $L^1_\mu(X, \mathcal{A}, \mathbb{C})$ into $\mathcal{CM}^1(X, \mathcal{A})$, and 

$$\|\mu_f\| = \|f\|_1 \quad \text{for all } f \in L^1_\mu(X, \mathcal{A}, \mathbb{C}).$$
The next theorem shows an important fact that we mentioned earlier, namely that every complex measure $\mu$ arises as a measure of the form $|\mu|h$ for some well chosen function $h \in L_{|\mu|}(X, A, \mathbb{C})$. This result is a special case of the Radon–Nikodym theorem, but for now we prefer not discuss this theorem.

**Theorem 5.20.** For every complex measure $\mu$ on a measurable space $(X, A)$, there is a function $h \in L_{|\mu|}(X, A, \mathbb{C})$ such that $|h| = 1$ and

$$\mu(A) = \int_A h d|\mu| \quad \text{for all } A \in A.$$

In other words, $\mu = |\mu|h$ (recall that $|\mu|$ is a positive measure). Furthermore, any two functions $h_1, h_2 \in L_{|\mu|}(X, A, \mathbb{C})$ satisfying the conditions of the theorem are equal $|\mu|$-a.e.

For a proof of Theorem 5.20, see Rudin [66] (Chapter 6, Theorem 6.12) and Lang [52] (Chapter VII, §2 and §4).

Let us now turn our attention to real measures. We will see that any real measure can be expressed in terms of two positive real measures. This implies that any complex measure can be expressed in terms of four positive real measures. This will allow us to explain how to integrate with respect to a complex measure.

### 5.5 Real Measures and the Hahn–Jordan Decomposition

If $\mu$ is a real measure, since $|\mu|$ is a finite measure, the following real measures $\mu^+$ and $\mu^-$ are well defined.

**Definition 5.10.** If $\mu$ is a real measure, the real measures $\mu^+$ and $\mu^-$ are defined by

$$\mu^+ = \frac{1}{2}(|\mu| + \mu), \quad \mu^- = \frac{1}{2}(|\mu| - \mu).$$

It is immediately checked that $\mu^+$ and $\mu^-$ are finite positive measures, and we have

$$\mu = \mu^+ - \mu^-, \quad |\mu| = \mu^+ + \mu^-.$$

**Definition 5.11.** Given a real measure $\mu$, the positive real measures $\mu^+$ and $\mu^-$ are called the **positive variation** and **negative variation** of $\mu$. The expression of $\mu$ as $\mu = \mu^+ - \mu^-$ is called the **Jordan decomposition** of $\mu$.

The Jordan decomposition has certain minimality properties that we are going to describe.
Definition 5.12. Let \((X, \mathcal{A})\) be a measurable space. A complex measure \(\mu\) is concentrated (or carried by) a measurable subset \(A\) if \(\mu(E) = 0\) for all \(E \in \mathcal{A}\) such that \(E \cap A = \emptyset\). Two complex measures \(\mu_1\) and \(\mu_2\) are mutually singular if there exist two disjoint measurable subsets \(A_1\) and \(A_2\) such that \(\mu_1\) is concentrated on \(A_1\) and \(\mu_2\) is concentrated on \(A_2\). We sometimes write \(\mu_1 \perp \mu_2\).

Every real measure has a Hahn–Jordan decomposition as described by the following theorem.

Theorem 5.21. (Hahn–Jordan Decomposition) Let \((X, \mathcal{A})\) be a measurable space. For any real measure \(\mu\), there is a partition \((X^+, X^-)\) of \(X\) into two disjoint subsets of \(X\) such that

\[
\mu = \mu^+ - \mu^-
\]

is the Jordan decomposition of \(\mu\), then \(\mu^+\) is concentrated on \(X^+\), and \(\mu^-\) is concentrated on \(X^-\). Furthermore, for any \(E \in \mathcal{A}\), we have

\[
\mu^+(E) = \sup\{\mu(A) \mid A \subseteq E, A \in \mathcal{A}\}, \quad \mu^-(E) = \sup\{-\mu(A) \mid A \subseteq E, A \in \mathcal{A}\}.
\]

For any other partition \((Y^+, Y^-)\) of \(X\) such that \(\mu^+\) is concentrated on \(Y^+\) and \(\mu^-\) is concentrated on \(Y^-\),

\[
\mu^+(E \cap X^+) = \mu^+(E \cap Y^+), \quad \mu^-(E \cap X^-) = \mu^-(E \cap Y^-),
\]

for all \(E \in \mathcal{A}\).

Let us now consider a complex measure \(\mu: \mathcal{A} \to \mathbb{C}\).

Definition 5.13. Given a complex measure \(\mu: \mathcal{A} \to \mathbb{C}\), the function \(\overline{\mu}: \mathcal{A} \to \mathbb{C}\) called the conjugate of \(\mu\) is defined by \(\overline{\mu}(A) = \overline{\mu(A)}\) for all \(A \in \mathcal{A}\). We also define \(\mu_1: \mathcal{A} \to \mathbb{R}\) and \(\mu_2: \mathcal{A} \to \mathbb{R}\) by

\[
\mu_1(A) = \frac{1}{2}(\mu(A) + \overline{\mu(A)}), \quad \mu_2(A) = \frac{1}{2i}(\mu(A) - \overline{\mu(A)})
\]

for all \(A \in \mathcal{A}\). We call \(\mu_1\) the real part of \(\mu\) and \(\mu_2\) the imaginary part of \(\mu\).

It is immediately checked that \(\overline{\mu}\) is a complex measure, and that \(\mu_1\) and \(\mu_2\) are real measures such that

\[
\mu = \mu_1 + i\mu_2, \quad \overline{\mu} = \mu_1 - i\mu_2.
\]

Using the Hahn–Jordan decomposition of \(\mu_1\) an \(\mu_2\), we see that we can write \(\mu\) uniquely in terms of four positive real measure \(\mu_1^+, \mu_1^-, \mu_2^+, \mu_2^-\), as

\[
\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-).
\]
Definition 5.14. For any complex measure \( \mu : A \to \mathbb{C} \), the expression

\[
\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-)
\]

is called the Jordan decomposition of \( \mu \).

Proposition 5.22. For any complex measure \( \mu : A \to \mathbb{C} \), we have \( |\mu_1| \leq |\mu|, |\mu_2| \leq |\mu| \), and that \( |\mu| \leq |\mu_1| + |\mu_2| \). A function \( f \) is \( |\mu| \)-integrable iff it is integrable for all four positive real measures \( \mu_1^+, \mu_1^-, \mu_2^+, \) and \( \mu_2^- \).

Proof. It is easy to check that \( |\mu_1| \leq |\mu|, |\mu_2| \leq |\mu| \), and that \( |\mu| \leq |\mu_1| + |\mu_2| \). It follows easily that \( f \) is \( |\mu| \)-integrable if \( f \) is \( |\mu_1| \)-integrable and \( |\mu_2| \)-integrable. Since \( |\mu_1| = \mu_1^+ + \mu_1^- \) and \( |\mu_2| = \mu_2^+ + \mu_2^- \), it is also easy to see that \( f \) is \( |\mu_1| \)-integrable iff \( f \) is \( \mu_1^+ \)-integrable and \( \mu_1^- \)-integrable, and similarly \( f \) is \( |\mu_2| \)-integrable iff \( f \) is \( \mu_2^+ \)-integrable and \( \mu_2^- \)-integrable. Therefore, \( f \) is \( |\mu| \)-integrable iff it is integrable for all four positive measures \( \mu_1^+, \mu_1^-, \mu_2^+, \) and \( \mu_2^- \).

The Jordan decomposition of the complex measure \( \mu \) suggests defining the integral \( \int f d\mu \) for any function \( f \in L^1_{|\mu|}(X, A, \mathbb{C}) \); see Dieudonné [22] (Chapter XIII, Section 16, no. 13.16.2), or Folland [32] (end of Section 3.1 and Section 3.3).

Definition 5.15. Given any complex measure \( \mu : A \to \mathbb{C} \), for any function \( f \in L^1_{|\mu|}(X, A, \mathbb{C}) \), we define the integral \( \int f d\mu \) as

\[
\int f d\mu = \int f d\mu_1 + i \int f d\mu_2 = \int f d\mu_1^+ - \int f d\mu_1^- + i \int f d\mu_2^+ - i \int f d\mu_2^-.
\]

By Proposition 5.22, the above expression is well defined since \( f \) is \( |\mu| \)-integrable iff it is integrable for all four positive real measures \( \mu_1^+, \mu_1^-, \mu_2^+, \) and \( \mu_2^- \).

Remark: Alternatively, if \( \mu \) is a complex measure, \( \int f d\mu \) can be defined using Theorem 5.20 as \( \int fh d|\mu| \), as in Rudin [66] (Chapter 6, Section 6.18).

The following fact will be needed later.

Proposition 5.23. Given a complex measure \( \mu \), if \( \overline{\mu} \) is the conjugate measure of \( \mu \), for any function \( f \in L^1_{|\mu|}(X, A, \mathbb{C}) \) we have

\[
\int f d\overline{\mu} = \int \overline{f} d\mu, \quad \text{or equivalently} \quad \int \overline{f} d\mu = \int f d\overline{\mu}.
\]

As a consequence, \( \overline{\mu} \) is the unique complex measure such that

\[
\int f d\overline{\mu} = \int \overline{f} d\mu, \quad \text{for all } f \in C_0(X; \mathbb{C}).
\]
Proof. Write $\mu = \mu_1 + i\mu_2$ as above, where $\mu_1$ and $\mu_2$ are real measures. We have $\overline{\mu} = \mu_1 - i\mu_2$, and the measures $\mu_1$ and $\mu_2$ be written as $\mu_1 = \mu_1^+ - \mu_1^-$ and $\mu_2 = \mu_2^+ - \mu_2^-$, where $\mu_1^+, \mu_1^-, \mu_2^+$, and $\mu_2^-$, are positive measures. Now, for any function $f$ integrable for all four positive measures above it is obvious that
\[
\int \overline{f}d\mu_1^+ = \int \overline{f}d\mu_1^-, \quad \int \overline{f}d\mu_2^+ = \int \overline{f}d\mu_2^-,
\]
so
\[
\int \overline{f}d\mu_1 = \int f d\mu_1, \quad \int \overline{f}d\mu_2 = \int f d\mu_2,
\]
thus
\[
\int \overline{f}d\mu = \int f d\mu_1 + i \int f d\mu_2
\]
\[
= \int f d\mu_1 + i \int f d\mu_2
\]
\[
= \int f d\mu_1 - i \int f d\mu_2
\]
\[
= \int f d\overline{\mu},
\]
as claimed. Since $C_0(X; \mathbb{C})$ is obviously contained in $L^1_{|\mu|}(X, \mathcal{A}, \mathbb{C})$, the last statement follows from Radon–Riesz II.

Since the measures $\mu_1^+, \mu_1^-, \mu_2^+$, and $\mu_2^-$ are positive real measures, they are finite. This immediately implies that the Radon functional $\varphi_\mu$ induced by a complex measure $\mu$ is bounded. Therefore, complex measures represent only bounded Radon functionals. Actually they represent all of them, which is the object of Section 5.7.

To show the above fact, we need to decompose a bounded Radon functional in terms of (four) positive bounded Radon functionals, and for this we introduce the notion of total variation of a Radon functional.

### 5.6 Total Variation of a Radon Functional

The notion of total variation of a Radon functional allows the decomposition of a bounded Radon functional into four positive bounded functionals, in a way that is similar to the Jordan decomposition of a complex measure. This fact is the key to the representation of a bounded Radon functional by a complex measure.

Recall that for any function $g: X \to \mathbb{C}$, we denote by $|g|$ the function $|g|: X \to \mathbb{R}$ given by $|g|(x) = |g(x)|$ for all $x \in X$.

The following result is shown in Dieudonné [22] (Chapter XIII, Section 3).
Theorem 5.24. For any Radon functional \( \Phi: \mathcal{K}_C(X) \to \mathbb{C} \) on a locally compact space \( X \), there is a smallest positive Radon functional \( |\Phi|: \mathcal{K}_C(X) \to \mathbb{C} \) such that

\[
|\Phi(f)| \leq |\Phi(|f|) \quad \text{for all } f \in \mathcal{K}_C(X).
\]

The functional \( |\Phi| \) is completely defined by its restriction to positive functions \( f \geq 0 \) in \( \mathcal{K}_\mathbb{R}(X) \) by

\[
|\Phi|(f) = \sup\{|\Phi(g)| \mid g \in \mathcal{K}_C(X), \ |g| \leq f\}.
\]

Proof sketch. We know from the remark just after Proposition 5.4 that a Radon functional \( \Phi: \mathcal{K}_C(X) \to \mathbb{C} \) is completely determined by its restriction \( \Phi_\mathbb{R}: \mathcal{K}_\mathbb{R}(X) \to \mathbb{C} \) to the space of real-valued functions in \( \mathcal{K}_\mathbb{R}(X) \). The first step in the proof of Theorem 5.24 is to show that the formula

\[
|\Phi|(f) = \sup\{|\Phi(g)| \mid g \in \mathcal{K}_C(X), \ |g| \leq f\}
\]

defined on positive functions \( f \geq 0 \) in \( \mathcal{K}_\mathbb{R}(X) \) is yields a finite number. Let \( K \) be the support of \( f \), which is compact. Since \( |g| \leq f \), the support of \( g \) is contained in \( K \), so

\[
|\Phi(g)| \leq c_K \|g\| \leq c_k \|f\|
\]

which shows that \( |\Phi|(f) \) is finite. Next we show that \( |\Phi| \) is additive, which is left as an exercise.

The second step is to extend \( |\Phi| \) to arbitrary functions \( f \in \mathcal{K}_\mathbb{R}(X) \) by writing \( f = f' - f'' \), where \( f', f'' \in \mathcal{K}_\mathbb{R}(X) \) and \( f', f'' \geq 0 \), by setting

\[
|\Phi|(f) = |\Phi|(f') - |\Phi|(f'').
\]

This expression does not depend on the decomposition of \( f \) because if \( f = f'_1 - f'_2 = f'_2 - f''_2 \), then \( f'_1 + f''_2 = f'_2 + f''_2 \), so \( |\Phi|(f'_1) + |\Phi|(f''_2) = |\Phi|(f'_2) + |\Phi|(f''_2) \), which implies \( |\Phi|(f'_1) - |\Phi|(f'_2) = |\Phi|(f''_2) - |\Phi|(f''_2) \).

The last step is to prove that \( |\Phi|(\lambda f) = \lambda |\Phi|(f) \), which is clear \( \lambda \geq 0 \). For \( \lambda < 0 \), we write \( f = f' - f'' \) with \( f', f'' \geq 0 \), and then

\[
|\Phi|(\lambda f) = |\Phi|(|\lambda f' - \lambda f''|)
\]

\[
= |\Phi|(|\lambda f' - \lambda f''|)
\]

\[
= |\Phi|(-\lambda f' - \lambda f'')
\]

\[
= \lambda(|\Phi|(f') - |\Phi|(f''))
\]

In summary, \( |\Phi| \) is a positive linear functional. By Proposition 5.4, the functional \( |\Phi| \) is a positive Radon functional. \( \square \)
Definition 5.16. Given any Radon functional $\Phi : \mathcal{K}_C(X) \to \mathbb{C}$, the positive Radon functional $|\Phi|$ is called total variation (or absolute value) of $\Phi$.

If $\Phi$ is a positive Radon functional, then

$$|\Phi| = \Phi.$$

Definition 5.17. Given a Radon functional $\Phi : \mathcal{K}_C(X) \to \mathbb{C}$, we define the conjugate $\overline{\Phi}$ of $\Phi$ by

$$\overline{\Phi}(f) = \Phi(\overline{f}), \quad f \in \mathcal{K}_C(X).$$

If we write $f = f_1 + if_2$ with $f_1, f_2 \in \mathcal{K}_R(X)$, then we have

$$\overline{\Phi}(f) = \overline{\Phi}(f_1 + if_2) = \Phi(f_1 - if_2) = (\Phi(f_1) - i\Phi(f_2)) = \Phi(f_1) + i\Phi(f_2).$$

Definition 5.18. We say that a Radon functional $\Phi : \mathcal{K}_C(X) \to \mathbb{C}$ is real if $\Phi = \Phi$.

Proposition 5.25. A Radon functional $\Phi : \mathcal{K}_C(X) \to \mathbb{C}$ is real iff its restriction $\Phi_R$ to $\mathcal{K}_R(X)$ is a real-valued function $\Phi_R : \mathcal{K}_R(X) \to \mathbb{R}$.

Proof. In view of the above computation, a Radon functional $\Phi$ is real iff

$$\Phi(f_1) + i\Phi(f_2) = \overline{\Phi(f_1) + i\Phi(f_2)}$$

for all $f_1, f_2 \in \mathcal{K}_R(X)$, which by setting $f_2 = 0$ or $f_1 = 0$ means that $\Phi(f_i) \in \mathbb{R}$ for all $f_i \in \mathcal{K}_R(X)$, for $i = 1, 2$. Equivalently, a Radon functional $\Phi : \mathcal{K}_C(X) \to \mathbb{C}$ is real iff its restriction $\Phi_R$ to $\mathcal{K}_R(X)$ is a real-valued function $\Phi_R : \mathcal{K}_R(X) \to \mathbb{R}$. \qed

Since a Radon functional is $\Phi$ is completely determined by its restriction $\Phi_R$ to $\mathcal{K}_R(X)$, we often think of a real Radon functional as a linear map $\Phi : \mathcal{K}_R(X) \to \mathbb{R}$.

Definition 5.19. Given a Radon functional $\Phi : \mathcal{K}_C(X) \to \mathbb{C}$, we define $\Phi_r$ and $\Phi_i$ by

$$\Phi_r = \frac{1}{2}(\Phi + \overline{\Phi}), \quad \Phi_i = \frac{1}{2i}(\Phi - \overline{\Phi}).$$

It is immediately verified that $\Phi_r$ and $\Phi_i$ are real Radon functionals such that

$$\Phi = \Phi_r + i\Phi_i, \quad \overline{\Phi} = \Phi_r - i\Phi_i.$$

We also have

$$|\Phi_r| \leq |\Phi|, \quad |\Phi_i| \leq |\Phi|, \quad |\Phi| \leq |\Phi_r| + |\Phi_i|.$$

Definition 5.20. If $\Phi : \mathcal{K}_C(X) \to \mathbb{R}$ is a real Radon functional, then as in the case of real measures we can define $\Phi^+$ and $\Phi^-$ by

$$\Phi^+ = \frac{1}{2}(|\Phi| + \Phi), \quad \Phi^- = \frac{1}{2}(|\Phi| - \Phi).$$
5.7. THE RADON–RIESZ THEOREM AND BOUNDED RADON FUNCTIONALS

It is immediately checked that $\Phi^+$ and $\Phi^-$ are positive Radon functionals, and we have

$$\Phi = \Phi^+ - \Phi^-,$$

$$|\Phi| = \Phi^+ + \Phi^-.$$

In the end, we have the following decomposition result analogous to the Jordan decomposition for complex measures.

**Proposition 5.26.** Every Radon functional $\Phi: K_C(X) \to \mathbb{C}$ can be expressed in terms of four positive Radon functionals:

$$\Phi = \Phi^+_r - \Phi^-_r + i(\Phi^+_i - \Phi^-_i).$$

By the Radon–Riesz I theorem (Theorem 5.8), there exist four positive $\sigma$-Radon measures $m_1, m_2, m_3, m_4$ such that

$$\Phi(f) = \int f dm_1 - \int f dm_2 + i \left( \int f dm_3 - \int f dm_4 \right)$$

for all $f \in K_C(X)$.

It is tempting to define the complex measure $m$ by

$$m = m_1 - m_2 + i(m_3 - m_4),$$

but there is a problem, which is that the positive measures $m_i$ may take the value $+\infty$, so expressions of the form $+\infty - (+\infty)$ may arise, but they do not make any sense!

We are not aware of a way around this problem in general. If $X$ is compact, then the Radon–Riesz II theorem yields positive Radon measures $m_i$ such that $m_i(X)$ is finite for $i = 1, \ldots, 4$, in which case the expression $m$ is indeed a measure. It is even possible to define a bijective correspondence by adding disjointness conditions on the subsets over which the $m_i$ are concentrated. Such results are given in Malliavin [55] (Chapter II, Section 5).

Another situation where $m$ is a complex measure is the case where the Radon functional $\Phi$ is bounded (continuous). This is the object of the next section.

5.7 The Radon–Riesz Theorem and Bounded Radon Functionals

Let $\Phi: K_C(X) \to \mathbb{C}$ be a bounded Radon functional. In this case the operator norm $\|\Phi\|$ is finite. Recall that

$$\|\Phi\| = \sup\{|\Phi(f)| \mid f \in K_C(X), \|f\|_\infty \leq 1\} = \sup\{|\Phi(f)| \mid f \in K_C(X), \|f\|_\infty = 1\}.$$

The following result is shown in Dieudonné [22] (Chapter VII, Section 20).
Proposition 5.27. Given a Radon functional $\Phi: K_\mathcal{C}(X) \to \mathbb{C}$, the norm $\|\Phi\|$ is finite, that is, $\Phi$ is bounded, iff $|\Phi|$ is bounded. In this case, $\|\Phi\| = \||\Phi||$. We deduce that $\Phi = \Phi_r + i\Phi_i$ is bounded iff $\Phi_r$ and $\Phi_i$ are bounded (see Definition 5.19).

But we also see that a real bounded Radon functional $\Psi = \Psi_r - \Psi_i$ is bounded iff the positive Radon functionals $\Psi^+$ and $\Psi^-$ are bounded (see Definition 5.20).

Proposition 5.28. A Radon functional $\Phi$ is bounded iff the positive Radon functionals $\Phi^+_r, \Phi^-_r, \Phi^+_i, \Phi^-_i$ are bounded.

If $m_1, m_2, m_3, m_4$ are the positive $\sigma$-Radon measures representing $\Phi^+_r, \Phi^-_r, \Phi^+_i, \Phi^-_i$ given by the Radon–Riesz I theorem (Theorem 5.8), it turns out that they are all finite measures, so $m = m_1 - m_2 + i(m_3 - m_4)$ is a complex measure, and it represents $\Phi$ on functions in $C_0(X)$. In order to state a suitable version of the Radon–Riesz correspondence, we need the following definition.

Definition 5.21. Let $X$ be a locally compact (Hausdorff) space. A complex measure $\mu$ on the $\sigma$-algebra $\mathcal{B}$ of Borel sets of $X$ is a regular complex Borel measure if the positive measure $|\mu|$ is a finite Radon measure, that is, a positive Borel measure that is regular and finite ($|\mu|(X)$ is finite). We denote the vector space of regular complex Borel measures by $M^1_{\text{reg}, \mathbb{C}}(X)$. See Figure 5.7.

Since $|\mu|(X)$ is finite, the measure $|\mu|(K)$ of every compact subset $K$ of $X$ is also finite (since $X$ is Hausdorff, every compact subset $K$ of $X$ is closed and thus measurable, and since $K \subseteq X$, we have $|\mu|(K) \leq |\mu|(X)$). Thus the positive Borel measure $|\mu|$ is locally finite.

We have the following beautiful theorem. Theorem 5.29 is also often referred to as the Riesz representation theorem, which is somewhat confusing.

Theorem 5.29. (Radon–Riesz Correspondence, III) Let $X$ be a locally compact (Hausdorff) space. There are bijections $m: M^1(X) \to M^1_{\text{reg}, \mathbb{C}}(X)$ and $\varphi: M^1_{\text{reg}, \mathbb{C}}(X) \to M^1(X)$ between the Banach space $M^1(X) = C_0(X, \mathbb{C})'$ of bounded Radon functionals, the dual of the space $C_0(X, \mathbb{C})$ of continuous functions that tend to zero at infinity, and the Banach space $M^1_{\text{reg}, \mathbb{C}}(X)$ of regular complex Borel measures. For every bounded Radon functional $\Phi \in C_0(X, \mathbb{C})'$, the regular complex Borel measure $m_\Phi$ represents $\Phi$ in the sense that

$$\Phi(f) = \int fdm_\Phi$$

$$= \int fd(m_\Phi)^+_r - \int fd(m_\Phi)^-_r + i \left( \int fd(m_\Phi)^+_i - \int fd(m_\Phi)^-_i \right)$$

for all $f \in C_0(X, \mathbb{C})$.

Furthermore, these bijections are norm preserving, that is, $\|\Phi\| = \|m_\Phi\| = |m_\Phi|(X)$. See Figure 5.8.
Theorem 5.29 is proven in Lang [52] (Chapter IX, §4, Theorem 4.2), Rudin [66] (Chapter 6 Theorem 6.19), Folland [32] (Chapter 7, Theorem 7.17), and Marle [56] (Chapter 9, Section 7, Proposition 9.7.3). The proof is quite involved. Among other things it uses Lusin’s theorem (Theorem 5.12) and the corollary of the Radon–Nikodym theorem (Theorem 5.20).

To prove surjectivity, by Proposition 5.26 we express the bounded Radon functional \( \Phi : \mathcal{K}_C(X) \to \mathbb{C} \) in terms of four positive Radon functionals:

\[
\Phi = \Phi^+_r - \Phi^-_r + i(\Phi^+_i - \Phi^-_i).
\]

By Proposition 5.28, these positive Radon functionals are bounded. By the Radon–Riesz theorem I (Theorem 5.8), there exist four positive \( \sigma \)-Radon measures \( m_1, m_2, m_3, m_4 \) such that

\[
\Phi(f) = \int f dm_1 - \int f dm_2 + i \left( \int f dm_3 - \int f dm_4 \right) \text{ for all } f \in \mathcal{K}_C(X).
\]

The reason why the \( \sigma \)-Radon measure \( m \) corresponding to a positive bounded Radon functional \( \Phi \) is finite is that this measure is inner regular, that is,

\[
m_\Phi(E) = \sup\{\mu(K) \mid K \subseteq E, K \text{ compact}\}
\]
for every $E \in \mathcal{B}$. We use this to compute $m_\Phi(X)$. For every compact subset $K$, by Proposition 15.39, there is a continuous function $f : X \to [0, 1]$ of compact support such that $f(x) = 0$ for all $x \in K$. Then since $\Phi$ is bounded we have

$$m_\Phi(K) \leq \int f \, dm_\Phi = \Phi(f) \leq \|\Phi\| \|f\|_\infty = \|\Phi\|$$

since $f$ has maximum value 1. Therefore,

$$m_\Phi(X) = \sup \{\mu(K) \mid K \subseteq X, K \text{ compact} \} \leq \|\Phi\|$$

is indeed finite. Since $m_\Phi(X)$ is finite, by Proposition 5.11, the $\sigma$-regular measure $m_\Phi$ is actually regular.

We also need to check that $\int f \, dm$ is finite for every function $f \in \mathcal{C}_0(X; \mathbb{C})$ and every positive finite Borel measure $m$. Since $\mathcal{C}_0(X; \mathbb{C})$ is the closure of $\mathcal{K}_C(X)$, there is a sequence $(f_n)$ of functions $f_n \in \mathcal{K}_C(X)$ that converges to $f$ according to the sup norm, and thus converges pointwise to $f$. Also $f$ is a bounded function, so there is some $M > 0$ such that $|f_n| \leq M$ for all $n \geq 1$. Since $m(X)$ is finite, the constant function $M$ is integrable, and the continuous functions $f_n$ are integrable. By the dominated convergence theorem (Theorem 4.33), $f$ is integrable.
Theorem 5.29 plays a crucial role in defining the notion of convolutions of two measures in \( \mathcal{M}_{\text{reg},\mathbb{C}}(X) \). We will need the following simple fact.

**Proposition 5.30.** Let \( X \) be any locally compact space, and let \( \mu \) be any positive Borel measure on \( B \). For any function \( f \in L_1^\mu(X,B,\mathbb{C}) \), the functional \( \Phi_{f,\mu}: C_0(X;\mathbb{C}) \to \mathbb{C} \) given by

\[
\Phi_{f,\mu}(g) = \int fg\,d\mu \quad \text{for all } g \in C_0(X;\mathbb{C})
\]

is a bounded Radon functional.

**Proof.** Since \( f \in L_1^\mu(X,B,\mathbb{C}) \) and \( g \) is continuous, \( g \) is measurable, and \( |g| \) is bounded by some \( M > 0 \), so by Proposition 4.35(1) \( fg \in L_1^\mu(X,B,\mathbb{C}) \). We have

\[
|\Phi_{f,\mu}(g)| = \left| \int fg\,d\mu \right| \leq \int |fg|\,d\mu = \int |f||g|\,d\mu \leq \|g\|_{\infty} \int |f|d\mu,
\]

which shows that \( \Phi_{f,\mu} \) is bounded. \( \square \)

By Theorem 5.29, the bounded Radon functional \( \Phi_{f,\mu} \) corresponds to a unique regular complex Borel measure \( m \) such that

\[
\int fg\,d\mu = \int g\,dm \quad \text{for all } g \in C_0(X;\mathbb{C}).
\]

The measure \( m \) is usually denoted by \( fd\mu \); for example, see Folland [31] (Chapter 2, Section 2.5). Proposition 5.30 gives us an embedding of \( L_1^\mu(X,B,\mathbb{C}) \) into \( \mathcal{M}_{\text{reg},\mathbb{C}}(X) \) as stated in the next proposition.

**Proposition 5.31.** Let \( X \) be a locally compact space. For every positive Borel measure \( \mu \) on \( B \), the map \( f \mapsto fd\mu \) is an embedding of \( L_1^\mu(X,B,\mathbb{C}) \) into the space \( \mathcal{M}_{\text{reg},\mathbb{C}}(X) \) of regular complex Borel measures on \( X \), with the property that

\[
\int fg\,d\mu = \int g\,fd\mu \quad \text{for all } g \in C_0(X;\mathbb{C}).
\]

One of the reasons for using this embedding is that if \( X \) is a locally compact group and if \( \mu \) is a Haar measure, convolution can be defined on both \( L_1^\mu(X,B,\mathbb{C}) \) and \( \mathcal{M}_{\text{reg},\mathbb{C}}(X) \), but there is no identity element for convolution on \( L_1^\mu(X,B,\mathbb{C}) \) while there is one for convolution on \( \mathcal{M}_{\text{reg},\mathbb{C}}(X) \).
Chapter 6

The Haar Measure and Convolution

Let $G$ be a locally compact group. Haar proved (1933) the remarkable fact that there is a positive $\sigma$-regular locally finite Borel measure $\mu$ on $G$ such that $\mu(U) > 0$ for every nonempty open subset $U$, and such that $\mu$ is left-invariant, which means that

$$\mu(A) = \mu(sA) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B},$$

where $\mathcal{B}$ is the $\sigma$-algebra of Borel sets on $G$. Furthermore, such a left-invariant measure is unique up to a positive scalar.

Actually, Haar proved the existence of a left-invariant measure in a special case. This result was established in full generality later by André Weil [84]. All proofs we are aware of (Weil [84], Halmos [39], Bourbaki [6], Dieudonné [22], Lang [52], Folland [31]) make use of Haar’s original clever idea (1933). Except for Halmos who constructs directly a measure (as Haar did), all the other proofs are essentially André Weil’s proof (which constructs a Haar functional) from his famous little book [84] first published in 1940.

In this chapter, we sketch the existence of the (left) Haar measure, providing most details, and we also prove its uniqueness up to a scalar; see Sections 6.2, 6.3, 6.4. Some Examples are given in Section 6.5.

For any $s \in G$ and any measure $\mu$ on $G$, let $\rho_s(\mu)$ be the measure given by

$$(\rho_s(\mu))(A) = \mu(As) \quad \text{for all } A \in \mathcal{B}.$$
also a right Haar measure. If $G$ is abelian, compact, or a connected semisimple Lie group, then $G$ is unimodular. More examples of Haar measures are given in Section 6.7.

Let $G$ be a locally compact group, and let $u : G \to G$ be an automorphism of $G$. For every left Haar measure $\mu$, define the measure $u^{-1}(\mu)$ by

$$(u^{-1}(\mu))(A) = \mu(u(A)), \quad \text{for all } A \in \mathcal{B}.$$ 

It can be shown that there is a unique positive number $\text{mod}(u)$ such that

$$u^{-1}(\mu) = \text{mod}(u)\mu$$

for all left Haar measures $\mu$. The number $\text{mod}(u)$ is called the modulus of the automorphism $u$. Properties of the modulus of an automorphism are discussed Section 6.8. As an application, we obtain formulae for the measure (volume) of a parallelotope and of a simplex.

Some applications of the Haar measure are discussed in Section 6.9. In particular, we prove Theorem 6.35, a basic tool in representation theory.

Let $G$ be a locally compact group, let $X$ be a locally compact space, and let $\cdot : G \times X \to X$ be a continuous left action of $G$ on $X$. A Borel measure $\mu$ on $X$ is $G$-invariant if

$$\mu(s^{-1} \cdot A) = \mu(A) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B}.$$ 

Our goal is to find sufficient conditions to ensure that $X$ has some $G$-invariant measure. We will consider the case where $X = G/H$, with the left action of $G$ on $G/H$ given by

$$a \cdot (bH) = abH, \quad a, b \in G.$$ 

In this case, by Proposition 6.6, the space $X$ is also locally compact (and Hausdorff).

A $G$-invariant measure on $G/H$ does not always exist. It turns out that there is a necessary and sufficient condition for a $G$-invariant $\sigma$-Radon measure to exist on $G/H$ in terms of $\Delta_G$ and $\Delta_H$: $\Delta_H$ must the equal to the restriction of $\Delta_G$ on $H$. This topic is discussed in Section 6.10.

One of the main applications of the Haar measure is the definition of the notion of convolution on a locally compact group. Recall that $\mathcal{M}_{\text{reg},\mathbb{C}}^1(G)$ denotes the Banach space of complex regular Borel measures on $G$ (see Definition 5.21), and that $L^1_\lambda(G, \mathcal{B}, \mathbb{C})$ denotes the space of integrable functions on the measure space $(G, \mathcal{B}, \lambda)$, where $\mathcal{B}$ is the $\sigma$-algebra of Borel sets of $G$. To simplify notation, we write $\mathcal{M}_1^1(G)$ for $\mathcal{M}_{\text{reg},\mathbb{C}}^1(G)$, and $L_1^1(G)$ for $L_1^1(G, \mathcal{B}, \mathbb{C})$. The vector space $\mathcal{M}_1^1(G)$ is a Banach space with the norm $\|\mu\| = |\mu|(G)$, and $L_1^1(G)$ is a Banach space with the $L^1$-norm. There are three flavors of convolutions but we will use mostly two of them:

1. Convolutions of two measures $\mu, \nu \in \mathcal{M}_1^1(G)$. This makes $\mathcal{M}_1^1(G)$ into a Banach algebra with identity, and with an involution.
2. Convolution of two functions \( f, g \in L^1(G) \), which makes \( L^1(G) \) into a Banach algebra with involution, but without a multiplicative unit element, unless \( G \) is discrete.

3. There is also a notion of convolution of a measure \( \mu \in M^1(G) \) and of a function \( f \in L^1(G) \) and a measure \( \mu \in M^1(G) \), but we relagate these definitions to the end of this section.

These notions of convolution are discussed in Section 6.11.

Convolution applied to functions and measures can be used as a regularization (or filtering) process; see Section 6.12.

6.1 Topological Groups

Since locally compact groups (and Lie groups) are topological groups, it is useful to gather a few basic facts about topological groups.

Definition 6.1. A set \( G \) is a topological group iff

(a) \( G \) is a Hausdorff topological space;

(b) \( G \) is a group (with identity 1);

(c) Multiplication \( \cdot : G \times G \to G \), and the inverse operation \( G \to G : g \mapsto g^{-1} \), are continuous, where \( G \times G \) has the product topology.

It is easy to see that the two requirements of Condition (c) are equivalent to

\((c')\) The map \( G \times G \to G : (g, h) \mapsto gh^{-1} \) is continuous.

Proposition 6.1. If \( G \) is a topological group and \( H \) is any subgroup of \( G \), then the closure \( \overline{H} \) of \( H \) is a subgroup of \( G \). If \( H \) is a normal subgroup of \( G \), then \( \overline{H} \) is also a normal subgroup of \( G \).

Proof. We use the fact that if \( f : X \to Y \) is a continuous map between two topological spaces \( X \) and \( Y \), then \( f(A) \subseteq f(A) \) for any subset \( A \) of \( X \). For any \( a \in \overline{A} \), we need to show that for any open subset \( W \subseteq Y \) containing \( f(a) \), we have \( W \cap f(A) \neq \emptyset \). Since \( f \) is continuous, \( V = f^{-1}(W) \) is an open subset containing \( a \), and since \( a \in \overline{A} \), we have \( f^{-1}(W) \cap A \neq \emptyset \), so there is some \( x \in f^{-1}(W) \cap A \), which implies that \( f(x) \in W \cap f(A) \), so \( W \cap f(A) \neq \emptyset \), as desired. The map \( f : G \times G \to G \) given by \( f(x, y) = xy^{-1} \) is continuous, and since \( H \) is a subgroup of \( G \), \( f(H \times H) \subseteq H \). By the above property, if \( a \in \overline{H} \) and if \( b \in \overline{H} \), that is, \((a, b) \in \overline{H} \times \overline{H} \), then \( f(a, b) = ab^{-1} \in \overline{H} \), which shows that \( \overline{H} \) is a subgroup of \( G \).

For every \( g \in G \), the map \( C_g : G \to G \) given by \( C_g(x) = gxg^{-1} \) for all \( x \in G \) is continuous, and if \( H \) is a normal subgroup of \( G \), then \( C_g(H) \subseteq H \). It follows that \( C_g(\overline{H}) \subseteq \overline{H} \) for all \( g \in G \), which means that \( \overline{H} \) is a normal subgroup of \( G \).

\( \square \)
Given a topological group $G$, for every $a \in G$ we define the left translation $L_a$ as the map $L_a : G \to G$ such that $L_a(b) = ab$, for all $b \in G$, and the right translation $R_a$ as the map $R_a : G \to G$ such that $R_a(b) = ba$, for all $b \in G$. Observe that $L_a^{-1}$ is the inverse of $L_a$ and similarly, $R_a^{-1}$ is the inverse of $R_a$. As multiplication is continuous, we see that $L_a$ and $R_a$ are continuous. Moreover, since they have a continuous inverse, they are homeomorphisms. As a consequence, if $U$ is an open subset of $G$, then so is $gU = L_g(U)$ (resp. $Ug = R_gU$), for all $g \in G$. Therefore, the topology of a topological group is determined by the knowledge of the open subsets containing the identity $1$.

Given any subset $S \subseteq G$, let $S^{-1} = \{s^{-1} | s \in S\}$; let $S^0 = \{1\}$, and $S^{n+1} = S^nS$, for all $n \geq 0$. Property (c) of Definition 6.1 has the following useful consequences, which shows there exists an open subset containing 1 which has a special symmetrical structure.

**Proposition 6.2.** If $G$ is a topological group and $U$ is any open subset containing 1, then there is some open subset $V \subseteq U$, with $1 \in V$, so that $V = V^{-1}$ and $V^2 \subseteq U$. Furthermore, $\overline{V} \subseteq U$.

**Proof.** Since multiplication $G \times G \to G$ is continuous and $G \times G$ is given the product topology, there are open subsets $U_1$ and $U_2$, with $1 \in U_1$ and $1 \in U_2$, so that $U_1U_2 \subseteq U$. Let $W = U_1 \cap U_2$ and $V = W \cap W^{-1}$. Then $V$ is an open set containing 1, and clearly $V = V^{-1}$ and $V^2 \subseteq U_1U_2 \subseteq U$. If $g \in \overline{V}$, then $gV$ is an open set containing $g$ (since $1 \in V$) and thus, $gV \cap V \neq \emptyset$. This means that there are some $h_1, h_2 \in V$ so that $gh_1 = h_2$, but then, $g = h_2h_1^{-1} \in VV^{-1} = VV \subseteq U$. \hfill \Box

**Definition 6.2.** A subset $U$ containing 1 and such that $U = U^{-1}$ is called symmetric.

Proposition 6.2 is used in the proofs of many the propositions and theorems on the structure of topological groups. For example, it is key in verifying the following proposition regarding discrete topological subgroups.

**Definition 6.3.** A subgroup $H$ of a topological group $G$ is discrete iff the induced topology on $H$ is discrete; that is, for every $h \in H$, there is some open subset $U$ of $G$ so that $U \cap H = \{h\}$.

**Proposition 6.3.** If $G$ is a topological group and $H$ is a discrete subgroup of $G$, then $H$ is closed.

**Proof.** As $H$ is discrete, there is an open subset $U$ of $G$ so that $U \cap H = \{1\}$, and by Proposition 6.2, we may assume that $U = U^{-1}$. Our goal is to show $H = \overline{H}$. Clearly $H \subseteq \overline{H}$. Thus it remains to show $\overline{H} \subseteq H$. If $g \in \overline{H}$, as $gU$ is an open set containing $g$, we have $gU \cap H \neq \emptyset$. Consequently, there is some $y \in gU \cap H = gU^{-1} \cap H$, so $g \in yU$ with $y \in H$. We claim that $yU \cap H = \{y\}$. Note that $x \in yU \cap H$ means $x = yu_1$ with $yu_1 \in H$ and $u_1 \in U$. Since $H$ is a subgroup of $G$ and $y \in H$, $y^{-1}yu_1 = u_1 \in H$. Thus $u_1 \in U \cap H$, which implies $u_1 = 1$ and $x = yu_1 = y$, and we have $g \in yU \cap \overline{H} \subseteq \overline{yU \cap H} = \overline{\{y\}} = \{y\}$.

since $G$ is Hausdorff. Therefore, $g = y \in H$. \hfill \Box
6.1. TOPOLOGICAL GROUPS

Using Proposition 6.2, we can give a very convenient characterization of the Hausdorff separation property in a topological group.

**Proposition 6.4.** If $G$ is a topological group, then the following properties are equivalent:

1. $G$ is Hausdorff;
2. The set $\{1\}$ is closed;
3. The set $\{g\}$ is closed, for every $g \in G$.

**Proof.** The implication (1) $\rightarrow$ (2) is true in any Hausdorff topological space. We just have to prove that $G - \{1\}$ is open, which goes as follows: For any $g \neq 1$, since $G$ is Hausdorff, there exists disjoint open subsets $U_g$ and $V_g$ with $g \in U_g$ and $1 \in V_g$. Thus, $U_g = G - \{1\}$, showing that $G - \{1\}$ is open. Since $L_g$ is a homeomorphism, (2) and (3) are equivalent.

Let us prove that (3) $\rightarrow$ (1). Let $g_1, g_2 \in G$ with $g_1 \neq g_2$. Then, $g_1^{-1}g_2 \neq 1$ and if $U$ and $V$ are disjoint open subsets such that $1 \in U$ and $g_1^{-1}g_2 \in V$, then $g_1 \in g_1U$ and $g_2 \in g_1V$, where $g_1U$ and $g_1V$ are still open and disjoint. Thus, it is enough to separate $1$ and $g \neq 1$. Pick any $g \neq 1$. If every open subset containing $1$ also contained $g$, then $1$ would be in the closure of $\{g\}$, which is absurd since $\{g\}$ is closed and $g \neq 1$. Therefore, there is some open subset $U$ such that $1 \in U$ and $g \not\in U$. By Proposition 6.2, we can find an open subset $V$ containing $1$, so that $V \subseteq U$ and $V = V^{-1}$. We claim that $V$ and $gV$ are disjoint open sets with $1 \in V$ and $g \in gV$.

Since $1 \in V$, it is clear that $g \in gV$. If we had $V \cap gV \neq \emptyset$, then by the last sentence in the proof of Proposition 6.2 we would have $g \in VV^{-1} = VV \subseteq U$, a contradiction. □

If $H$ is a subgroup of $G$ (not necessarily normal), we can form the set of left cosets $G/H$, and we have the projection $p: G \rightarrow G/H$, where $p(g) = gH = \overline{g}$. If $G$ is a topological group, then $G/H$ can be given the quotient topology, where a subset $U \subseteq G/H$ is open iff $p^{-1}(U)$ is open in $G$. With this topology, $p$ is continuous. The trouble is that $G/H$ is not necessarily Hausdorff. However, we can neatly characterize when this happens.

**Proposition 6.5.** If $G$ is a topological group and $H$ is a subgroup of $G$, then the following properties hold:

1. The map $p: G \rightarrow G/H$ is an open map, which means that $p(V)$ is open in $G/H$ whenever $V$ is open in $G$.
2. The space $G/H$ is Hausdorff iff $H$ is closed in $G$.
3. If $H$ is open, then $H$ is closed and $G/H$ has the discrete topology (every subset is open).
4. The subgroup $H$ is open iff $1 \in \hat{H}$ (i.e., there is some open subset $U$ so that $1 \in U \subseteq H$).
Proof. (1) Observe that if \( V \) is open in \( G \), then \( VH = \bigcup_{h \in H} Vh \) is open, since each \( Vh \) is open (as right translation is a homeomorphism). However, it is clear that

\[
p^{-1}(p(V)) = VH,
\]
i.e., \( p^{-1}(p(V)) \) is open which, by definition of the quotient topology, means that \( p(V) \) is open.

(2) If \( G/H \) is Hausdorff, then by Proposition 6.4, every point of \( G/H \) is closed, i.e., each coset \( gH \) is closed, so \( H \) is closed. Conversely, assume \( H \) is closed. Let \( \pi \) and \( \gamma \) be two distinct point in \( G/H \) and let \( x, y \in G \) be some elements with \( p(x) = \pi \) and \( p(y) = \gamma \). As \( \pi \neq \gamma \), the elements \( x \) and \( y \) are not in the same coset, so \( x \notin yH \). As \( H \) is closed, so is \( yH \), and since \( x \notin yH \), there is some open containing \( x \) which is disjoint from \( yH \), and we may assume (by translation) that it is of the form \( Ux \), where \( U \) is an open containing 1. By Proposition 6.2, there is some open \( V \) containing 1 so that \( V \cap V^{-1} \) is open and \( VV \subseteq U \) and \( V = V^{-1} \). Thus, we have

\[
V^2 x \cap yH = \emptyset
\]
and in fact,

\[
V^2 xH \cap yH = \emptyset,
\]
since \( H \) is a group; if \( z \in V^2 xH \cap yH \), then \( z = \nu_1 \nu_2 xh_1 = yh_2 \) for some \( \nu_1, \nu_2 \in V \), and some \( h_1, h_2 \in H \), but then \( \nu_1 \nu_2 x = yh_2 h_1^{-1} \) so that \( V^2 x \cap yH \neq \emptyset \), a contradiction. Since \( V = V^{-1} \), we get

\[
VxH \cap VyH = \emptyset,
\]
and then, since \( V \) is open, both \( VxH \) and \( VyH \) are disjoint, open, so \( p(VxH) \) and \( p(VyH) \) are open sets (by (1)) containing \( \pi \) and \( \gamma \) respectively and \( p(VxH) \) and \( p(VyH) \) are disjoint (because \( p^{-1}(p(VxH)) = VxHH = VxH \), \( p^{-1}(p(VyH)) = VyHH = VyH \), and \( VxH \cap VyH = \emptyset \)). See Figure 6.1.

(3) If \( H \) is open, then every coset \( gH \) is open, so every point of \( G/H \) is open and \( G/H \) is discrete. Also, \( \bigcup_{g \notin H} gH \) is open, i.e., \( H \) is closed.

(4) Say \( U \) is an open subset such that \( 1 \in U \subset H \). Then for every \( h \in H \), the set \( hU \) is an open subset of \( H \) with \( h \in hU \), which shows that \( H \) is open. The converse is trivial. \( \square \)

Recall that a topological space \( X \) is \textit{locally compact} iff for every point \( p \in X \), there is a compact neighborhood \( C \) of \( p \); that is, there is a compact \( C \) and an open \( U \), with \( p \in U \subset C \). For example, manifolds are locally compact.

The next two propositions will be needed.

**Proposition 6.6.** Let \( G \) be a topological group and let \( H \) be a closed subgroup of \( G \). The following properties hold.

1. If \( G \) is locally compact, then so is \( G/H \).
(2) If $H$ is a normal subgroup of $G$, then $G/H$ is a topological group.

Proof. (1) Since $H$ is closed, we already know from Proposition 6.5(2) that $G/H$ is Hausdorff. Let $K$ be a compact neighborhood of 1 in $G$, so that there is an open subset $U$ such that $1 \in U \subseteq K$ with $K$ compact. By Proposition 6.5(1) the quotient map $p: G \to G/H$ is an open map, and it is continuous, so for any $g \in G$, we have $g \in gU \subseteq gK$ with $gU$ open and $gK$ compact, so $p(g) \in p(gU) \subseteq p(gK)$ with $p(gU)$ open and $(gK)$ compact, which shows that $G/H$ is locally compact.

(2) If $H$ is a closed normal subgroup, then $G/H$ is a group, and we already know from Proposition 6.5(2) that $G/H$ is Hausdorff. We have to show that multiplication and inversion in $G/H$ are continuous. For any two cosets $g_1H$ and $g_2H$ in $G/H$, if $W$ is an open subset in $G/H$ containing $p(g_1g_2) = p(g_1)p(g_2) = (g_1H)(g_2H) = g_1g_2H$, then because the projection map $p$ is continuous, there are open subsets $U_1$ and $U_2$ of $G$ with $g_1 \in U_1$ and $g_2 \in U_2$, such that $p(U_1U_2) \subseteq W$. Since $p$ is an open map, $p(U_1)$ is an open subset containing $p(g_1) = g_1H$ and $p(U_2)$ is an open subset containing $p(g_2) = g_2H$, and we have $p(U_1)p(U_2) \subseteq W$, so multiplication in $G/H$ is continuous. A similar proof shows that inversion is continuous in $G/H$. \hfill \Box

**Proposition 6.7.** Let $G$ be a locally compact topological group and let $H$ be a closed subgroup of $G$. For any compact subset $K'$ in $G/H$, there is compact subset $K$ of $G$ such that $p(K) = K'$.

Proof. Since $G$ is locally compact, there is an open subset $U$ and a compact subset $V$ such that $1 \in U \subseteq V$. Since $p$ is an open map, the subsets of the form $p(gU)$ for $g \in G$ form an open cover of $K'$, and since $K'$ is compact, there is a finite subcover $\{p(g_1U), \ldots, p(g_nU)\}$ of $K'$. Since $p$ is continuous and $K'$ is compact and thus closed (since $G/H$ is Hausdorff), $p^{-1}(K')$ is closed and $g_1V \cup \cdots \cup g_nV$ is compact, $K = p^{-1}(K') \cap (g_1V \cup \cdots \cup g_nV)$ is compact in $G$, and we have $p(K) = K'$. \hfill \Box
We next provide a criterion relating the connectivity of $G$ with that of $G/H$.

**Proposition 6.8.** Let $G$ be a topological group and $H$ be any subgroup of $G$. If $H$ and $G/H$ are connected, then $G$ is connected.

**Proof.** It is a standard fact of topology that a space $G$ is connected iff every continuous function $f$ from $G$ to the discrete space $\{0,1\}$ is constant; see Proposition 15.17. Pick any continuous function $f$ from $G$ to $\{0,1\}$. As $H$ is connected and left translations are homeomorphisms, all cosets $gH$ are connected. Thus, $f$ is constant on every coset $gH$. It follows that the function $f: G \to \{0,1\}$ induces a continuous function $\overline{f}: G/H \to \{0,1\}$ such that $f = \overline{f} \circ p$ (where $p: G \to G/H$; the continuity of $\overline{f}$ follows immediately from the definition of the quotient topology on $G/H$). As $G/H$ is connected, $\overline{f}$ is constant, and so $f = \overline{f} \circ p$ is constant.

The next three propositions describe how to generate a topological group from its symmetric neighborhoods of 1.

**Proposition 6.9.** If $G$ is a connected topological group, then $G$ is generated by any symmetric neighborhood $V$ of 1. In fact, $G = \bigcup_{n \geq 1} V^n$.

**Proof.** Since $V = V^{-1}$, it is immediately checked that $H = \bigcup_{n \geq 1} V^n$ is the group generated by $V$. As $V$ is a neighborhood of 1, there is some open subset $U \subseteq V$, with $1 \in U$, and so $1 \in H$. From Proposition 6.5 (3), the subgroup $H$ is open and closed, and since $G$ is connected, $H = G$.

**Proposition 6.10.** Let $G$ be a topological group and let $V$ be any connected symmetric open subset containing 1. Then, if $G_0$ is the connected component of the identity, we have $G_0 = \bigcup_{n \geq 1} V^n$, and $G_0$ is a normal subgroup of $G$. Moreover, the group $G/G_0$ is discrete.

**Proof.** First, as $V$ is open, every $V^n$ is open, so the group $\bigcup_{n \geq 1} V^n$ is open, and thus closed, by Proposition 6.5 (3). For every $n \geq 1$, we have the continuous map $V \times \cdots \times V \longrightarrow V^n : (g_1, \ldots, g_n) \mapsto g_1 \cdots g_n$.

As $V$ is connected, $V \times \cdots \times V$ is connected, and so $V^n$ is connected; this follows from Proposition 15.18 because a finite product of connected spaces is connected. Since $1 \in V^n$ for all $n \geq 1$ and every $V^n$ is connected, we use Lemma 15.19 to conclude that $\bigcup_{n \geq 1} V^n$ is connected. Now, $\bigcup_{n \geq 1} V^n$ is connected, open and closed, so it is the connected component of 1. Finally, for every $g \in G$, the group $gGg^{-1}$ is connected and contains 1, so it is contained in $G_0$, which proves that $G_0$ is normal. Since $G_0$ is open, Proposition 6.5 (3) implies that the group $G/G_0$ is discrete.
Proposition 6.11. Let $G$ be a topological group and assume that $G$ is connected and locally compact. Then, $G$ is countable at infinity, which means that $G$ is the union of a countable family of compact subsets. In fact, if $V$ is any symmetric compact neighborhood of 1, then

$$G = \bigcup_{n \geq 1} V^n.$$ 

Proof. Since $G$ is locally compact, there is some compact neighborhood $K$ of 1. Then, $V = K \cap K^{-1}$ is also compact and a symmetric neighborhood of 1. By Proposition 6.9, we have

$$G = \bigcup_{n \geq 1} V^n.$$ 

An argument similar to the one used in the proof of Proposition 6.10 to show that $V^n$ is connected if $V$ is connected proves that each $V^n$ compact if $V$ is compact.

If $G$ is a locally compact group but $G$ is not connected, and if $G_0$ is the connected component of the identity, then $G$ is the disjoint union of the cosets $gG_0$, and each coset $gG_0$ is homeomorphic to $G_0$, connected, and countable at infinity ($\sigma$-compact). This observation plays a crucial role in the proof of the uniqueness of the Haar measure (Theorem 6.21), because it guarantees that the use of Fubini’s theorem is legitimate.

The notion of uniform continuity can be generalized to functions defined on a group.

Definition 6.4. Given a topological group $G$ and a subset $S$ of $G$, for any normed vector space $F$, a function $f: G \to F$ is left uniformly continuous on $S$ if for any $\epsilon > 0$, there is an open subset $U$ of $G$ containing 1 such that

$$\|f(y) - f(x)\| < \epsilon \quad \text{for all } x, y \in S \text{ such that } xy^{-1} \in U.$$ 

The function $f: G \to F$ is right uniformly continuous on $S$ if for any $\epsilon > 0$, there is an open subset $U$ of $G$ containing 1 such that

$$\|f(y) - f(x)\| < \epsilon \quad \text{for all } x, y \in S \text{ such that } x^{-1}y \in U.$$ 

Observe that if $xy^{-1} \in U$, then we can write $xy^{-1} = z$ for some $z \in U$, so $y = z^{-1}x$, and

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

It is customary to introduce a left action $\lambda$ of $G$ on functions $f: G \to F$ defined on $G$ by

$$(\lambda_s(f))(x) = f(s^{-1}x) \quad \text{for all } x, s \in G.$$ 

Observe that

$$\lambda_{st}(f)(x) = f((st)^{-1}x) = f(t^{-1}s^{-1}x) = \lambda_t(f)(s^{-1}x) = \lambda_s(\lambda_t(f))(x),$$

where $\lambda_t(f)(x) = f(t^{-1}x)$ for all $x \in G$. This action is continuous and respects the group structure of $G$. The left and right actions on functions are compatible with the group operations, and they are also compatible with the uniform continuity conditions defined above.
so
\[ \lambda_{st} = \lambda_s \circ \lambda_t, \]
which is the reason why we used \( s^{-1} \) instead of \( s \) in the definition of \( \lambda_s \).

Then \( \| f(z^{-1}x) - f(x) \| = \| \lambda_z(f)(x) - f(x) \| \), so the condition of the definition is equivalent to
\[ \| \lambda_z(f)(x) - f(x) \| < \epsilon \quad \text{for all } x \in S \text{ and all } z \in U. \]

Informally, the above condition can be written as
\[ \lim_{z \to 1} \sup_{x \in S} \| \lambda_z(f)(x) - f(x) \| = 0. \]

It is also customary to introduce a right action \( \rho \) of \( G \) on functions \( f: G \to F \) defined on \( G \) by
\[ (\rho_s(f))(x) = f(xs) \quad \text{for all } x, s \in G. \]

Observe that
\[ \rho_{st}(f)(x) = f(xst) = \rho_t(f)(x) = \rho_s(\rho_t(f))(x), \]
so
\[ \rho_{st} = \rho_s \circ \rho_t. \]

Observe that if \( x^{-1}y \in U \), then we can write \( x^{-1}y = z \) for some \( z \in U \), so \( y = xz \), and
\[ \| f(y) - f(x) \| = \| f(xz) - f(x) \| = \| \rho_z(f)(x) - f(x) \| < \epsilon. \]

Thus the condition of the definition is equivalent to
\[ \| \rho_z(f)(x) - f(x) \| < \epsilon \quad \text{for all } x \in S \text{ and all } z \in U. \]

Informally, the above condition can be written as
\[ \lim_{z \to 1} \sup_{x \in S} \| \rho_z(f)(x) - f(x) \| = 0. \]

**Proposition 6.12.** Let \( G \) be a topological group and let \( S \) be a subset of \( G \). For any function \( f: S \to F \), where \( F \) is any normed vector space, if \( f \) is continuous with compact support \( K \), then \( f \) is left (resp. right) uniformly continuous on \( K \).

*Proof.* We prove that \( f \) is left uniformly continuous, the proof that \( f \) is right uniformly continuous being similar and left as an exercise. Since \( f \) is continuous, for every \( y \in K \), there is some open subset \( U_y \) with \( 1 \in U_y \) such that
\[ \| f(y) - f(x) \| < \frac{\epsilon}{2} \quad \text{for all } x \in U_y y. \]

We can find an open subset \( V_y \) containing 1 such that \( V_y V_y \subseteq U_y \). The open subsets of the form \( V_y y \) for \( y \in K \) form an open cover of \( K \), and since \( K \) is compact, there is a finite subcover \( \{ V_{y_1} y_1, \ldots, V_{y_n} y_n \} \) of \( K \) with \( y_1, \ldots, y_n \in K \). Let
\[ V = V_{y_1} \cap \cdots \cap V_{y_n}. \]
Consider $x, y \in K$ such that $xy^{-1} \in V$, that is, $x \in Vy$. Then $y \in V_i y_i \subseteq U_i y_i$ for some $i$, and so 

$$x \in Vy \in VV_y y_i \subseteq V_y V_y y_i \subseteq U_y y_i,$$

which implies that 

$$\|f(y) - f(x)\| \leq \|f(y) - f(y_i)\| + \|f(y_i) - f(x)\| < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,$$

as desired. \hfill \Box

We end this section by combining the various properties of a topological group $G$ to characterize when $G/G_x$ is homeomorphic to $X$. In order to do so, we need two definitions.

**Definition 6.5.** Let $G$ be a topological group and let $X$ be a topological space. An action $\varphi: G \times X \rightarrow X$ is continuous (and $G$ acts continuously on $X$) if the map $\varphi$ is continuous.

If an action $\varphi: G \times X \rightarrow X$ is continuous, then each map $\varphi_g: X \rightarrow X$ is a homeomorphism of $X$ (recall that $\varphi_g(x) = g \cdot x$, for all $x \in X$). Indeed, the map $x \mapsto g \cdot x$ is a continuous bijection whose inverse $x \mapsto g^{-1} \cdot x$ is also continuous.

Under some mild assumptions on $G$ and $X$, the quotient space $G/G_x$ is homeomorphic to $X$. For example, this happens if $X$ is a Baire space.

**Definition 6.6.** A Baire space $X$ is a topological space with the property that if $\{F_i\}_{i \geq 1}$ is any countable family of closed sets $F_i$ such that each $F_i$ has empty interior, then $\bigcup_{i \geq 1} F_i$ also has empty interior. By complementation, this is equivalent to the fact that for every countable family of open sets $U_i$ such that each $U_i$ is dense in $X$ (i.e., $\overline{U_i} = X$), then $\bigcap_{i \geq 1} U_i$ is also dense in $X$.

**Remark:** A subset $A \subseteq X$ is rare if its closure $\overline{A}$ has empty interior. A subset $Y \subseteq X$ is meager if it is a countable union of rare sets. Then, it is immediately verified that a space $X$ is a Baire space iff every nonempty open subset of $X$ is not meager.

The following theorem shows that there are plenty of Baire spaces:

**Theorem 6.13.** (Baire) (1) Every locally compact topological space is a Baire space.

(2) Every complete metric space is a Baire space.

A proof of Theorem 6.13 can be found in Bourbaki [12], Chapter IX, Section 5, Theorem 1.

**Theorem 6.14.** Let $G$ be a topological group which is locally compact and countable at infinity, $X$ a Hausdorff topological space which is a Baire space, and assume that $G$ acts transitively and continuously on $X$. Then, for any $x \in X$, the map $\varphi: G/G_x \rightarrow X$ is a homeomorphism.
Proof. We follow the proof given in Bourbaki [12], Chapter IX, Section 5, Proposition 6 (Essentially the same proof can be found in Mneimné and Testard [60], Chapter 2). First, observe that if a topological group acts continuously and transitively on a Hausdorff topological space, then for every \( x \in X \), the stabilizer \( G_x \) is a closed subgroup of \( G \). This is because, as the action is continuous, the projection \( \pi_x : G \rightarrow X : g \mapsto g \cdot x \) is continuous, and \( G_x = \pi^{-1}(\{x\}) \), with \( \{x\} \) closed. Therefore, by Proposition 6.5, the quotient space \( G/G_x \) is Hausdorff. As the map \( \pi_x : G \rightarrow X \) is continuous, the induced map \( \varphi_x : G/G_x \rightarrow X \) is continuous, and by Proposition 17.14, it is a bijection. Therefore, to prove that \( \varphi_x \) is a homeomorphism, it is enough to prove that \( \varphi_x \) is an open map. For this, it suffices to show that \( \pi_x \) is an open map. Given any open \( U \) in \( G \), we will prove that for any \( g \in U \), the element \( \pi_x(g) = g \cdot x \) is contained in the interior of \( U \cdot x \). However, observe that this is equivalent to proving that \( x \) belongs to the interior of \( (g^{-1} \cdot U) \cdot x \). Therefore, we are reduced to the following case: if \( U \) is any open subset of \( G \) containing 1, then \( x \) belongs to the interior of \( U \cdot x \).

Since \( G \) is locally compact, using Proposition 6.2, we can find a compact neighborhood of the form \( W = V \), such that \( 1 \in W, W = W^{-1} \) and \( W^2 \subseteq U \), where \( V \) is open with \( 1 \in V \subseteq U \). As \( G \) is countable at infinity, \( G = \bigcup_{i \geq 1} K_i \), where each \( K_i \) is compact. Since \( V \) is open, all the cosets \( gV \) are open, and as each \( K_i \) is covered by the \( gV \)'s, by compactness of \( K_i \), finitely many cosets \( gV \) cover each \( K_i \), and so

\[
G = \bigcup_{i \geq 1} g_i V = \bigcup_{i \geq 1} g_i W,
\]

for countably many \( g_i \in G \), where each \( g_i W \) is compact. As our action is transitive, we deduce that

\[
X = \bigcup_{i \geq 1} g_i W \cdot x,
\]

where each \( g_i W \cdot x \) is compact, since our action is continuous and the \( g_i W \) are compact. As \( X \) is Hausdorff, each \( g_i W \cdot x \) is closed, and as \( X \) is a Baire space expressed as a union of closed sets, one of the \( g_i W \cdot x \) must have nonempty interior; that is, there is some \( w \in W \), with \( g_i w \cdot x \) in the interior of \( g_i W \cdot x \), for some \( i \). But then, as the map \( y \mapsto g \cdot y \) is a homeomorphism for any given \( g \in G \) (where \( y \in X \)), we see that \( x \) is in the interior of \( w^{-1}g_i^{-1} \cdot (g_i W \cdot x) = w^{-1}W \cdot x \subseteq W^{-1}W \cdot x = W^2 \cdot x \subseteq U \cdot x \), as desired. \(\square\)

By Theorem 6.13, we get the following important corollary:

**Theorem 6.15.** Let \( G \) be a topological group which is locally compact and countable at infinity, \( X \) a Hausdorff locally compact topological space, and assume that \( G \) acts transitively and continuously on \( X \). Then, for any \( x \in X \), the map \( \varphi_x : G/G_x \rightarrow X \) is a homeomorphism.

Readers who wish to learn more about topological groups may consult Sagle and Walde [68] and Chevalley [16] for an introductory account, and Bourbaki [11], Weil [84] and Pontryagin [64, 65], for a more comprehensive account (especially the last two references).
6.2 Existence of the Haar Measure; Preliminaries

Let $G$ be a locally compact group. We are going to show there is a positive $\sigma$-regular locally finite Borel measure $\mu$ on $G$ such that $\mu(U) > 0$ for every nonempty open subset $U$, and such that $\mu$ is left-invariant, which means that

$$\mu(A) = \mu(sA) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B},$$

where $\mathcal{B}$ is the $\sigma$-algebra of Borel sets on $G$.

Recall that for any $x \in G$, the maps $L_x : G \to G$ (left translation) and $R_x : G \to G$ (right translation) are defined by

$$L_x(z) = xz, \quad R_x(z) = zx, \quad \text{for all } x, z \in G.$$

It is obvious that

$$L_{xy} = L_x \circ L_y \quad \text{and} \quad R_{xy} = R_y \circ R_x,$$

and that $L_x$ and $R_y$ commute for all $x, y \in G$.

It is customary to introduce a left action $\lambda$ of $G$ and a right action $\rho$ of $G$ on functions $f : G \to F$. We did this in the previous section, but for the sake of completeness, we repeat these definitions.

**Definition 6.7.** Let $G$ be a group, and let $F$ be any set. The left action $\lambda$ of $G$ on a function $f : G \to F$ is the function $\lambda_s(f)$ is given by

$$(\lambda_s(f))(x) = f(s^{-1}x) \quad \text{for all } x, s \in G,$$

and the right action $\rho$ of $G$ on a function $f : G \to F$ is the function $\rho_s(f)$ given by

$$(\rho_s(f))(x) = f(xs) \quad \text{for all } x, s \in G.$$

Observe that

$$\lambda_{st}(f)(x) = f((st)^{-1}x) = f(t^{-1}s^{-1}x) = \lambda_t(f)(s^{-1}x) = \lambda_s(\lambda_t(f))(x),$$

so

$$\lambda_{st} = \lambda_s \circ \lambda_t,$$

which is the reason why we used $s^{-1}$ instead of $s$ in the definition of $\lambda_s$. Observe that

$$\rho_{st}(f)(x) = f(xst) = \rho_t(f)(xs) = \rho_s(\rho_t(f))(x),$$

so

$$\rho_{st} = \rho_s \circ \rho_t.$$

Given a subset $A$ of $G$, we usually write $sA$ for $L_s(A)$ and $As$ for $R_s(A)$.

We define a left action of $\lambda_s$ and a right action $\rho_s$ on measures and Radon functionals as follows.
**Definition 6.8.** Let $G$ be a locally compact topological group. The *left action* $\lambda$ of $G$ on a measure $\mu$ on $(G, \mathcal{B})$ is the measure $\lambda_s(\mu)$ given by

$$(\lambda_s(\mu))(A) = \mu(s^{-1}A) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B},$$

and the *right action* $\rho$ of $G$ on a measure $\mu$ on $(G, \mathcal{B})$ is the measure $\rho_s(\mu)$ given by

$$(\rho_s(\mu))(A) = \mu(As) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B}.$$  

The *left action* $\lambda$ of $G$ on a Radon functional $\Phi: \mathcal{K}_c(G) \to \mathbb{C}$ is the Radon functional $\lambda_s(\Phi)$ given by

$$(\lambda_s(\Phi))(f) = \Phi(\lambda_s^{-1}(f)) \quad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_c(G),$$

and the *right action* $\rho$ of $G$ on a Radon functional $\Phi: \mathcal{K}_c(G) \to \mathbb{C}$ is the Radon functional $\rho_s(\Phi)$ given by

$$(\rho_s(\Phi))(f) = \Phi(\rho_s^{-1}(f)) \quad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_c(G).$$

If $m_\Phi$ is the Borel measure corresponding to a positive Radon functional $\Phi$, $m_{\lambda_s(\Phi)}$ is the Borel measure corresponding to $\lambda_s(\Phi)$, and $m_{\rho_s(\Phi)}$ is the Borel measure corresponding to $\rho_s(\Phi)$, given by Theorem 5.8, then we have

$$\int f(x)dm_{\lambda_s(\Phi)}(x) = \int f(sx)dm_\Phi(x),$$

and

$$\int f(x)dm_{\rho_s(\Phi)}(x) = \int f(xs^{-1})dm_\Phi(x)$$

for all $f \in \mathcal{K}_c(G)$ and all $s \in G$.

Definition 6.8 has been designed so that for every measure $\mu$ on $G$ we have

$$\lambda_{st}(\mu) = \lambda_s(\lambda_t(\mu)), \quad \rho_{st}(\mu) = \rho_s(\rho_t(\mu)) \quad \text{for all } s, t \in G.$$  

For every Radon functional $\Phi$ and for every function $f \in \mathcal{K}_c(G)$, we have

$$(\lambda_{st}(\Phi))(f) = \Phi(\lambda_{st}^{-1}(f)) = \Phi(\lambda_t^{-1}s^{-1}(f))$$

$$= \Phi(\lambda_t^{-1}(\lambda_s^{-1}(f))) = (\lambda_t(\Phi))(\lambda_s^{-1}(f)) = (\lambda_s(\lambda_t(\Phi)))(f),$$

and a similar computation shows that

$$(\rho_{st}(\Phi))(f) = (\rho_s(\rho_t(\Phi)))(f).$$

Therefore, for every Radon functional $\Phi$, we have

$$\lambda_{st}(\Phi) = \lambda_s(\lambda_t(\Phi)), \quad \rho_{st}(\Phi) = \rho_s(\rho_t(\Phi)) \quad \text{for all } s, t \in G.$$
Definition 6.9. Let \( G \) be a locally compact group. A \textit{left Haar measure} \( \mu \) is a \( \sigma \)-regular, locally finite, Borel measure on the \( \sigma \)-algebra \( \mathcal{B} \) of Borel sets of \( G \), such that \( \mu(U) > 0 \) for all nonempty open subsets \( U \in \mathcal{B} \), and is \textit{left-invariant}, which means that
\[
\lambda_s(\mu) = \mu \quad \text{for all } s \in G.
\]
The above condition means that
\[
\mu(s^{-1}A) = \mu(A) \quad \text{for all } A \in \mathcal{B} \text{ and all } s \in G,
\]
or equivalently,
\[
\mu(sA) = \mu(A) \quad \text{for all } A \in \mathcal{B} \text{ and all } s \in G,
\]
A \textit{right Haar measure} \( \mu \) is a Borel measure satisfying the same conditions as a left Haar measure, except that it is \textit{right invariant}, which means that
\[
\rho_s(\mu) = \mu \quad \text{for all } A \in \mathcal{B} \text{ and all } s \in G.
\]
The above condition means that
\[
\mu(As) = \mu(A) \quad \text{for all } A \in \mathcal{B} \text{ and all } s \in G.
\]

Note that according to Definition 5.4, a left (resp.) Haar measure is a \( \sigma \)-Radon measure which is left-invariant (resp. right-invariant), and such that \( \mu(U) > 0 \) for all nonempty open subsets \( U \in \mathcal{B} \).

In order to prove that a left (resp. right) Haar measure exists, we will use Theorem 5.8, which motivates the following definition.

Definition 6.10. Let \( G \) be a locally compact group. A \textit{left Haar functional} \( \Phi \) is a positive non-zero Radon functional \( \Phi : \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C} \) which is \textit{left-invariant}, which means that
\[
\lambda_s(\Phi) = \Phi \quad \text{for all } s \in G.
\]
A \textit{right Haar functional} \( \Phi \) is a positive non-zero Radon functional \( \Phi : \mathcal{K}_{\mathbb{C}}(G) \to \mathbb{C} \) which is \textit{right-invariant}, which means that
\[
\rho_s(\Phi) = \Phi \quad \text{for all } s \in G.
\]

If \( m_{\Phi} \) is the Borel measure associated with \( \Phi \), then its left-invariance is equivalent to
\[
\int f(x)dm_{\Phi}(x) = \int f(sx)dm_{\Phi}(x),
\]
and its right-invariance is equivalent to
\[
\int f(x)dm_{\Phi}(x) = \int f(xs^{-1})dm_{\Phi}(x),
\]
for all \( f \in \mathcal{K}_{\mathbb{C}}(G) \) and all \( s \in G \).

The following operation will allow use to convert a left-invariant measure (resp. functional) to a right-invariant measure (resp. functional).
Definition 6.11. Let $G$ be any locally compact group, and $F$ be any set. For any function $f: G \to F$, define the function $\tilde{f}: G \to F$ by
\[
\tilde{f}(s) = f(s^{-1}) \quad \text{for all } s \in G.
\]
For any Borel measure $\mu$ on $(G, \mathcal{B})$, define the Borel measure $\tilde{\mu}$ by
\[
\tilde{\mu}(A) = \mu(A^{-1}) \quad \text{for all } A \in \mathcal{B}.
\]
For any Radon functional $\Phi: \mathcal{K}_c(G) \to \mathbb{C}$, define the Radon functional $\tilde{\Phi}: \mathcal{K}_c(G) \to \mathbb{C}$ by
\[
\tilde{\Phi}(f) = \Phi(\tilde{f}) \quad \text{for all } f \in \mathcal{K}_c(G).
\]
Observe that
\[
(\lambda_s(\mu))^{-1}(A) = (\lambda_s(\mu))(A^{-1}) = \mu(s^{-1}A^{-1}) = \mu((As)^{-1}) = \tilde{\mu}(As) = (\rho_s(\tilde{\mu}))(A),
\]
so
\[
\lambda_s(\mu)^\gamma = \rho_s(\tilde{\mu}).
\]
Similarly,
\[
(\rho_s(\mu))^{-1}(A) = (\rho_s(\mu))(A^{-1}) = \mu(A^{-1}s) = \mu((s^{-1}A)^{-1}) = \tilde{\mu}(s^{-1}A) = (\lambda_s(\tilde{\mu}))(A),
\]
so
\[
\rho_s(\mu)^\gamma = \lambda_s(\tilde{\mu}).
\]
For any function $f: G \to F$, we have
\[
(\lambda_s(f))^{-1}(x) = (\lambda_s(f))(x^{-1}) = f(s^{-1}x^{-1}) = f((xs)^{-1}) = \tilde{f}(xs) = (\rho_s(\tilde{f}))(x),
\]
and similarly
\[
(\rho_s(f))^{-1}(x) = (\rho_s(f))(x^{-1}) = f(x^{-1}s) = f((s^{-1}x)^{-1}) = \tilde{f}(s^{-1}x) = (\lambda_s(\tilde{f}))(x).
\]
Therefore,
\[
\lambda_s(f)^\gamma = \rho_s(\tilde{f}), \quad \rho_s(f)^\gamma = \lambda_s(\tilde{f}) \quad \text{for all } s \in G \text{ and all } f: G \to F.
\]
Using the above equations, for every Radon functional $\Phi: \mathcal{K}_c(G) \to \mathbb{C}$, we have
\[
(\lambda_s(\Phi))^{-1}(f) = (\lambda_s(\Phi))(\tilde{f}) = \Phi(\lambda_{s^{-1}}(\tilde{f})) = \Phi(\rho_{s^{-1}}(f)^\gamma) = \tilde{\Phi}(\rho_{s^{-1}}(f)) = (\rho_s(\tilde{\Phi}))(f).
\]
Similarly,
\[
(\rho_s(\Phi))^{-1}(f) = (\rho_s(\Phi))(\tilde{f}) = \Phi(\rho_{s^{-1}}(\tilde{f})) = \Phi(\lambda_{s^{-1}}(f)^\gamma) = \tilde{\Phi}(\lambda_{s^{-1}}(f)) = (\lambda_s(\tilde{\Phi}))(f).
\]
Therefore, we have
\[
\lambda_s(\Phi)^\gamma = \rho_s(\tilde{\Phi}), \quad \rho_s(\Phi)^\gamma = \lambda_s(\tilde{\Phi}) \quad \text{for all } s \in G.
\]
Proposition 6.16. Let $G$ be a locally compact group, and let $\mu$ be a $\sigma$-regular, locally finite, Borel measure on $G$. The following properties hold:

1. We have
   \[ \lambda_s(\mu) = \rho_s(\hat{\mu}), \quad \rho_s(\mu) = \lambda_s(\hat{\mu}) \]
   for all $s \in G$. Consequently, $\mu$ is a left-invariant measure iff $\hat{\mu}$ is a right-invariant measure. For any Radon functional $\Phi : K_C(G) \to \mathbb{C}$, we have
   \[ \lambda_s(\Phi) = \rho_s(\hat{\Phi}), \quad \rho_s(\Phi) = \lambda_s(\hat{\Phi}) \]
   for all $s \in G$. Consequently, $\Phi$ is left-invariant iff $\hat{\Phi}$ is right-invariant.

2. If the Haar measure $\mu$ is left-invariant then
   \[ \int \lambda_{s^{-1}}(f) \, d\mu = \int f \, d\mu \]
   for all $f \in L^1_\mu(G, \mathcal{B}, \mathbb{C})$ and all $s \in G$. If the Haar measure $\mu$ is right-invariant then
   \[ \int \rho_{s^{-1}}(f) \, d\mu = \int f \, d\mu \]
   for all $f \in L^1_\mu(G, \mathcal{B}, \mathbb{C})$ and all $s \in G.$

3. We have
   \[ \int \lambda_{s^{-1}}(f) \, d\mu = \int f \, d\lambda_s(\mu) \]
   for all $f \in L^1_\mu(G, \mathcal{B}, \mathbb{C})$ and all $s \in G.$ If
   \[ \int \lambda_{s^{-1}}(f) \, d\mu = \int f \, d\mu \]
   for all $f \in K_C(G)$ and all $s \in G$, then $\mu$ is left-invariant. We have
   \[ \int \rho_{s^{-1}}(f) \, d\mu = \int f \, d\rho_s(\mu) \]
   for all $f \in L^1_\mu(G, \mathcal{B}, \mathbb{C})$ and all $s \in G.$ If
   \[ \int \rho_{s^{-1}}(f) \, d\mu = \int f \, d\mu \]
   for all $f \in K_C(G)$ and all $s \in G$, then $\mu$ is right-invariant.

Proof. We already proved (1).

(2) Let $f$ be any $\mu$-step function

\[ f = \sum_{k=1}^{n} y_k \chi_{A_k}, \]
where the $A_k$ are measurable Borel sets of finite measure. For all $s, x \in G$, we see that $(\lambda_{s^{-1}} f)(x) = f(sx) = y_k$ iff $sx \in A_k$ iff $x \in s^{-1}A_k$, which means that

$$\lambda_{s^{-1}} f = \sum_{k=1}^{n} y_k \chi_{s^{-1}A_k}.$$ 

so

$$\int (\lambda_{s^{-1}} f) d\mu = \sum_{k=1}^{n} y_k \mu(s^{-1}A_k) = \sum_{k=1}^{n} y_k (\lambda_{s}(\mu))(A_k) = \int f d\lambda_{s}(\mu).$$

If $\mu$ is left-invariant, then $\lambda_{s}(\mu) = \mu$, so

$$\sum_{k=1}^{n} y_k (\lambda_{s}(\mu))(A_k) = \sum_{k=1}^{n} y_k \mu(A_k) = \int f d\mu,$$

and we deduce that

$$\int (\lambda_{s^{-1}} f) d\mu = \int f d\mu.$$ 

Every function $f \in L^{1}_{\mu}(G, \mathcal{B}, \mathbb{C})$ has some approximation sequence $(f_n)$ by $\mu$-step functions that converges to $f$ a.e. and in the $L^{1}$-norm. It follows that the sequence $(\lambda_{s^{-1}} f_n)$ converges a.e. to $\lambda_{s^{-1}}(f)$. We check immediately that it is a Cauchy sequence because

$$\int (\lambda_{s^{-1}} f_n) d\mu = \int f_n d\mu$$

for all $n$, and it follows that

$$\int \lambda_{s^{-1}}(f) d\mu = \int f d\mu.$$ 

If $f$ is any $\mu$-step function

$$f = \sum_{k=1}^{n} y_k \chi_{A_k},$$

we have $(\rho_{s^{-1}} f)(x) = f(xs^{-1}) = y_k$ iff $xs^{-1} \in A_k$ iff $x \in A_k s$, which means that

$$\rho_{s^{-1}} f = \sum_{k=1}^{m} y_k \chi_{A_k s},$$

so

$$\int (\rho_{s^{-1}} f) d\mu = \sum_{k=1}^{n} y_k \mu(A_k s) = \sum_{k=1}^{n} y_k (\rho_{s}(\mu))(A_k) = \int f d\rho_{s}(\mu).$$

We finish the argument as in the previous case.

(3) The proof in (2) actually shows that

$$\int \lambda_{s^{-1}}(f) d\mu = \int f d\lambda_{s}(\mu)$$


and
\[ \int \rho_{s^{-1}}(f) \, d\mu = \int f \, d\rho_s(\mu) \]
for all \( f \in L^1_\mu(G, \mathcal{B}, \mathbb{C}) \). If
\[ \int \lambda_{s^{-1}}(f) \, d\mu = \int f \, d\mu \]
then
\[ \int f \, d\lambda_s(\mu) = \int f \, d\mu \]
for all \( f \in K_C(G) \) and all \( s \in G \), and by the uniqueness of the Borel measure corresponding to the Radon functional \( f \mapsto \int f \, d\mu \) from Theorem 5.8, we see that \( \lambda_s(\mu) = \mu \) for all \( s \in G \), which means that \( \mu \) is left-invariant. The right-invariant case is similar.

The condition
\[ \int \lambda_{s^{-1}}(f) \, d\mu = \int f \, d\mu \quad \text{for all } s \in G \]
is also written as
\[ \int f(sx) \, d\mu(x) = \int f(x) \, d\mu(x) \quad \text{for all } s \in G. \]

The condition
\[ \int \rho_{s^{-1}}(f) \, d\mu = \int f \, d\mu \quad \text{for all } s \in G \]
is also written as
\[ \int f(xs^{-1}) \, d\mu(x) = \int f(x) \, d\mu(x) \quad \text{for all } s \in G. \]
Since \( G \) is a group and \( s \) is any arbitrary element of \( G \), the above condition is also equivalent to
\[ \int f(xs) \, d\mu(x) = \int f(x) \, d\mu(x) \quad \text{for all } s \in G. \]

### 6.3 Existence of the Haar Measure

We are now going to sketch the proof that a left-invariant Haar measure exists on any locally compact group. All proofs we are aware of (Weil [84], Halmos [39], Bourbaki [6], Dieudonné [22], Lang [52], Folland [31]) make use of Haar’s original clever idea (1933). Except for Halmos who constructs directly a measure (as Haar did), all the other proofs are essentially André Weil’s proof (which constructs a Haar functional) from his famous little book [84] first published in 1940.

As we noted just after Proposition 5.4, there is a bijection between the space \( M^+(X) \) of positive linear functionals \( \Phi: K_C(X) \to \mathbb{C} \) and the space of positive linear functionals \( \Psi: K_R(X) \to \mathbb{R} \), so is is enough to construct a left (or right) real Haar functional on \( K_R(G) \).
Theorem 6.17. (Haar) Every locally compact group \( G \) possesses a left-invariant Haar measure.

Proof sketch. Folland [31] (Chapter 2, Section 2.2) is kind enough to provide the intuition behind the construction. In this method, a measure is not constructed directly. Instead, a left Haar functional is constructed. Then Theorem 5.8 is used to obtain a left-invariant Borel measure which is a left Haar measure.

Suppose we have positive function \( \varphi \in K_R(G) \) bounded by 1, equal to 1 on a small open set \( U \), and whose support is a compact subset slightly larger than \( U \). If \( f \in K_R(G) \) is any other function slowly varying so that it is essentially constant on the left translates of \( U \), then \( f \) can be approximated by a linear combination \( f \approx \sum c_j \lambda_{s_j}(\varphi) \). If \( \mu \) were a left Haar measure, then we would have

\[
\int f \, d\mu \approx \left( \sum_j c_j \right) \int \varphi \, d\mu.
\]

This approximation gets better and better as the support of \( \varphi \) shrinks to a point, and if we introduce a normalization to cancel out the factor \( \int \varphi \, d\mu \), then we obtain \( \int f \, d\mu \) as the limit of the sums \( \sum_j c_j \). If \( G = \mathbb{R} \), and if \( \varphi \) is the characteristic function of a small interval, this is reminiscent of the approximation of \( \int f \, d\mu \) by Riemann sums. The issue is to make this idea precise and formal.

The first step is to pick some positive nonzero function \( \varphi \in K_R(G) \). Then we claim that for every function \( f \in K_R(G) \), there exists a finite set \( \{s_1, \ldots, s_n\} \) of elements of \( G \) and a finite sequence \((c_1, \ldots, c_n)\) of reals \( c_j \in \mathbb{R} \), such that

\[
f \leq \sum_{j=1}^n c_j \lambda_{s_j}(\varphi).
\]

This is because \( f \) has compact support, so its support can be covered by finitely many translates of the open subset \( U \) given by

\[
U = \left\{ s \in G \mid \varphi(s) > \frac{1}{2} \|\varphi\|_\infty \right\},
\]

and if we pick \( c_j = \|f\|_\infty / a \) where \( a = (1/2) \|\varphi\|_\infty \), then \( f \leq \sum_{j=1}^n c_j \lambda_{s_j}(\varphi) \).

Define \((f : \varphi)\) as the greatest lower bound of the sums \( \sum_{j=1}^n c_j \), over all sets \( \{s_1, \ldots, s_n\} \) of elements of \( G \) and all finite sequences \((c_1, \ldots, c_n)\) of reals \( c_j \in \mathbb{R} \) such that

\[
f \leq \sum_{j=1}^n c_j \lambda_{s_j}(\varphi).
\]

Then it is not hard to show that the quantity \((f : \varphi)\) has the following properties:
6.3. EXISTENCE OF THE HAAR MEASURE

\[ (f : \varphi) = (\lambda_s(f) : \varphi) \quad \text{for all } s \in G \]
\[ (f_1 + f_2 : \varphi) \leq (f_1 : \varphi) + (f_2 : \varphi) \quad \text{for all } f_1, f_2 \in \mathcal{K}_\mathbb{R}(G) \]
\[ (cf : \varphi) = c(f : \varphi) \quad \text{for all } c \geq 0 \]
\[ (f_1 : \varphi) \leq (f_2 : \varphi) \quad \text{whenever } f_1 \leq f_2 \]
\[ (f : \varphi) \geq \|f\|_\infty / \|\varphi\|_\infty \]
\[ (f : \varphi) \leq (f : \psi)(\psi : \varphi) \quad \text{for all positive } \psi \in \mathcal{K}_\mathbb{R}(G) \]
\[ 0 < \frac{1}{(f_0 : f)} \leq \frac{(f : \varphi)}{(f_0 : \varphi)} \leq (f : f_0) \quad \text{for all positive } f, f_0 \in \mathcal{K}_\mathbb{R}(G). \]

We now make a normalization by fixing some positive nonzero \( f_0 \in \mathcal{K}_\mathbb{R}(G) \), and defining
\[ I_\varphi(f) = \frac{(f : \varphi)}{(f_0 : \varphi)} \]
for every positive function \( f \in \mathcal{K}_\mathbb{R}(G) \). The above properties show that \( I_\varphi \) is a functional which is left-invariant, subadditive, homogeneous of degree 1, and monotone. It also satisfies the following property:
\[ \frac{1}{(f_0 : f)} \leq I_\varphi(f) \leq (f : f_0). \quad (*) \]

If \( I_\varphi \) were additive rather than subadditive, it would be the restriction to the positive functions in \( \mathcal{K}_\mathbb{R}(G) \) of a positive linear functional on \( \mathcal{K}_\mathbb{R}(G) \), and we would be done. To make a linear functional, we need to shrink to the domain of \( \varphi \), and this is the part of the argument which is the most subtle.

The following technical proposition whose proof is given in Folland [31] (Chapter 2, Lemma 2.18) is needed.

**Proposition 6.18.** If \( f_1 \) and \( f_2 \) are any two positive functions in \( \mathcal{K}_\mathbb{R}(G) \) and if \( \epsilon > 0 \), then there is an open subset \( V \) containing 1 such that \( I_\varphi(f_1) + I_\varphi(f_2) \leq I_\varphi(f_1 + f_2) + \epsilon \), whenever \( \text{supp}(\varphi) \subseteq V \).

To shrink the domain of \( \varphi \) we use a compactness argument. For every positive function \( f \in \mathcal{K}_\mathbb{R}(G) \), let \( X_f \) be the interval
\[ X_f = \left[ \frac{1}{(f_0 : f)}, (f : f_0) \right]. \]

Let \( X = \prod_f X_f \). Since each \( X_f \) is compact, by Tychonoff’s theorem \( X \) is also compact. We can think of \( X \) as the set of functions mapping a positive function \( f \) in \( \mathcal{K}_\mathbb{R}(G) \) to some number in the interval \( X_f \). By (\(*\)), we have \( I_\varphi(f) \in X \) for all positive nonzero \( \varphi \in \mathcal{K}_\mathbb{R}(G) \). For every compact neighborhood \( V \) containing 1, let \( K(V) \) be the closure in \( X \) of the subset \( \{ I_\varphi \mid \text{supp}(\varphi) \subseteq V \} \). The family of subsets \( K(V) \) has the finite intersection property, since
CHAPTER 6. THE HAAR MEASURE AND CONVOLUTION

Let \( K(\bigcap_{j=1}^n V_j) \subseteq \bigcap_{j=1}^n K(V_j) \). Since \( X \) is compact, there is some \( I \in X \) which lies in every \( K(V) \). This means that every neighborhood of \( I \) in \( X \) contains some \( I_\varphi \) with \( \text{supp}(\varphi) \) arbitrarily small. In other words, for any open subset \( V \) containing 1, any \( \epsilon > 0 \), and for any positive functions \( f_1, \ldots, f_n \in K_\mathbb{R}(G) \), there exist some positive nonzero \( \varphi \in K_\mathbb{R}(G) \) with \( \text{supp}(\varphi) \subseteq V \) such that \( |I(f_j) - I_\varphi(f_j)| < \epsilon \) for all \( j \). By the properties of \( I_\varphi \) listed above and by Proposition 6.18, we conclude that \( I \) commutes with left translation, addition, and multiplication by positive scalars.

We can extend \( I \) to arbitrary functions \( f \in K_\mathbb{R}(G) \) as follows. We can write \( f = f_1 - f_2 \) with \( f_1, f_2 \) positive functions in \( K_\mathbb{R}(G) \), and we let \( I(f) = I(f_1) - I(f_2) \). If we also have \( f = f'_1 - f'_2 \) with \( f'_1, f'_2 \) positive functions in \( K_\mathbb{R}(G) \), then

\[
f_1 + f'_2 = f_2 + f'_1,
\]

so by linearity of \( I \) on positive functions we get

\[
I(f_1) + (f'_2) = I(f_2) + I(f'_1),
\]

thus

\[
I(f_1) - I(f_2) = I(f'_1) - I(f'_2),
\]

which means that \( I(f) \) is well defined. The functional \( I \) is a left Haar functional, and we are done. By Proposition 6.16(3), since the Haar functional \( I \) is left-invariant, the corresponding \( \sigma \)-Radon measure is also left-invariant.

Remark: The proof in Bourbaki [6] uses an argument involving an ultrafilter instead of Tychonoff’s theorem, but otherwise it is identical. Dieudonné [22] assumes that the locally compact group \( G \) is separable and metrizable. This allows him to avoid using Tychonoff’s theorem, but does not make the proof simpler.

Let \( \mu \) be the left Haar measure associated with the left Haar functional \( I \) given by Theorem 6.17. Here is an immediate consequence of Theorem 6.17.

**Proposition 6.19.** If \( \mu \) is left Haar measure on \( G \), then for every nonempty open subset \( U \), we have \( \mu(U) > 0 \). For every positive nonzero function \( f \in K_\mathbb{R}(G) \), we have \( \int f d\mu > 0 \).

**Proof.** Assume \( U \) is a nonempty open set with \( \mu(U) = 0 \). Then since \( \mu \) is left-invariant \( \mu(gU) = 0 \) for all \( g \in G \), and since any compact subset \( K \) can be covered by finitely many translates of \( U \), we have \( \mu(K) = 0 \). Since \( \mu \) is a regular Borel measure, it is \( \sigma \)-inner regular, that is,

\[
\mu(G) = \sup\{\mu(K) \mid K \subseteq G, K \text{ compact}\}
\]

so \( \mu(G) = 0 \), contradicting the fact that \( \mu \) is not the zero measure.

For any positive nonzero function \( f \in K_\mathbb{R}(G) \), let \( U = \{g \in G \mid f(g) > \frac{1}{2} \|f\|_\infty\} \). Then \( \int f d\mu > \frac{1}{2} \|f\|_\infty \mu(U) > 0 \).
Remark: If $G$ is a Lie group, there are much simpler methods for obtaining a left Haar measure on $G$. Suppose $G$ has dimension $n$. Pick an $n$-differential form $\omega_0$ on $\mathfrak{g}$, and transport it on all tangent spaces by left translation, obtaining a left-invariant volume form $\omega$. Then $f \mapsto \int f \omega$ is a left-invariant Haar functional that induces a left Haar measure.

We now turn to the uniqueness of the Haar measure.

### 6.4 Uniqueness of the Haar Measure

Any two left Haar measures on a locally compact group are proportional up to a positive factor. All the proofs we are aware of use tricks involving a double integration and Fubini’s theorem. These proofs are attributed to von Neumann. In our opinion, the proof using the least devious trick is that of Dieudonné [22] (Chapter XIV, Section 1), also used in Bourbaki [6] in a slightly more concise form. Since Dieudonné uses a theory of integration based on Radon functionals rather than on measure theory, some minor adaptations need to be made; specifically, Proposition 5.14 is needed instead of Proposition 13.15.3 in Dieudonné [22]. The first step is the following crucial result.

**Proposition 6.20.** Given a left Haar functional $\Phi$ and a right Haar functional $\Psi$ on a locally compact group $G$, if $\nu$ is the corresponding right Haar measure, for any function $f \in K_{\mathbb{R}}(G)$, if $\Phi(f) \neq 0$, then the function $D_f$ given by

$$D_f(s) = \Phi(f)^{-1} \int f(t^{-1}s) \, d\nu(t)$$

for all $s \in G$ is continuous.

*Proof.* It suffices to show that the function

$$s \mapsto \int f(t^{-1}s) \, d\nu(t)$$

is continuous. Let $K$ be the compact subset of $G$ which is the support of $f$. Pick any $s_0 \in G$, and let $V_0$ by any compact neighborhood of $s_0$. For every $\epsilon > 0$, we have to find an open subset $V$ containing $s_0$ such that $V \subseteq V_0$ and

$$\left| \int (f(t^{-1}s) - f(t^{-1}s_0)) \, d\nu \right| < \epsilon$$

for all $s \in V$. Let $L = V_0 K^{-1}$, then

$$\int (f(t^{-1}s) - f(t^{-1}s_0)) \, d\nu = \int_L (f(t^{-1}s) - f(t^{-1}s_0)) \, d\nu.$$

By Proposition 6.12, the function $f$ is right uniformly continuous, so there is some open subset $W$ containing 1 such that

$$|f(t^{-1}s) - f(t^{-1}s_0)| < \frac{\epsilon}{\nu(L)}$$
for all \((t^{-1}s_0)^{-1}t^{-1}s \in W\), that is, \(s_0^{-1}s \in W\), namely \(s \in Ws_0\), and all \(t \in G\). If we take \(V = V_0 \cap W\), then

\[
\left| \int_L (f(t^{-1}s) - f(t^{-1}s_0)) d\nu \right| < \int_L |f(t^{-1}s) - f(t^{-1}s_0)| d\nu < \epsilon,
\]

as desired.

We are now ready to prove our uniqueness result.

**Theorem 6.21. (Haar)** If \(\mu\) and \(\nu\) are any two left-invariant Haar measures on a locally compact group \(G\), then there is some \(c > 0\) such that \(\mu = c \nu\).

**Proof.** Since a Haar functionl \(\Psi\) is right-invariant iff \(\bar{\Psi}\) is left-invariant, It suffices to prove that if \(\Phi\) is a left-invariant Haar functional and if \(\Psi\) is a right-invariant Haar functional, then there is some \(c > 0\) such that \(\Phi = c \bar{\Psi}\). Let \(\mu\) be the left Haar measure associated with \(\Phi\) and let \(\nu\) be right Haar measure associated with \(\Psi\).

Let \(f \in K_R(G)\) be any function such that \(\Phi(f) \neq 0\) and let \(g \in K_R(G)\) be any other function. The function from \(G \times G\) to \(\mathbb{R}\) given by \((s,t) \mapsto f(s)g(ts)\) is continuous and has compact support. Recall that \(D_f\) is given by

\[
D_f(s) = \Phi(f)^{-1} \int f(t^{-1}s) d\nu(t).
\]

For \(s = 1\), we have

\[
D_f(1) = \Phi(f)^{-1} \bar{\Psi}(f).
\]

Therefore, if we can show that \(D_f\) is independent of \(f\), we are done. We evaluate \(\Phi(f)\Psi(g)\) using Fubini’s theorem.

\[
\Phi(f)\Psi(g) = \left( \int f(s) d\mu(s) \right) \left( \int g(t) d\nu(t) \right)
= \int f(s) \left( \int g(t) d\nu(t) \right) d\mu(s)
= \int \left( \int f(s)g(t) d\nu(t) \right) d\mu(s)
= \int \left( \int f(s)g(ts) d\nu(t) \right) d\mu(s) \quad \text{by right-invariance of } \nu
= \int \left( \int f(s)g(ts) d\mu(s) \right) d\nu(t) \quad \text{by Fubini}
\]
6.4. UNIQUENESS OF THE HAAR MEASURE

\[
\begin{align*}
= & \int \left( \int f(t^{-1}s)g(s) \, d\mu(s) \right) \, d\nu(t) \quad \text{by left-invariance of } \mu \\
= & \int \left( \int f(t^{-1}s)g(s) \, d\nu(t) \right) \, d\mu(s) \quad \text{by Fubini} \\
= & \int g(s) \left( \int f(t^{-1}s) \, d\nu(t) \right) \, d\mu(s) \\
= & \int g(s) \Phi(f) D_f(s) \, d\mu(s) \quad \text{by definition of } D_f \\
= & \Phi(f) \Phi(D_f \cdot g),
\end{align*}
\]

where \( D_f \cdot g \) is the function given by \((D_f \cdot g)(s) = D_f(s)g(s)\) for all \( s \in G \). Since \( \Phi(f) \neq 0 \), we deduce that

\[ \Psi(g) = \Phi(D_f \cdot g). \]

The above equation shows that \( D_f \) is independent of \( f \) because if \( f' \) is another function \( f' \in \mathcal{K}_\mathbb{R}(G) \) such that \( \Phi(f') \neq 0 \), then

\[ \Phi(D_f \cdot g) = \int D_f(s)g(s) \, d\mu(s) = \int D_f'(s)g(s) \, d\mu(s) = \Phi(D_f' \cdot g) \quad \text{for all } g \in \mathcal{K}_\mathbb{R}(G). \]

By Proposition 5.14, we deduce that \( D_f \) and \( D_f' \) are equal a.e. (The version of Proposition 5.14 is stated for complex-valued functions, but it also holds for real-valued functions). However, \( D_f \) and \( D_f' \) are continuous and the subset \( N \) where they differ is open and a null set, thus empty by Proposition 6.19. Therefore \( D_f = D_f' = D \), and by definition of \( D \) we have

\[ \Phi(f) = D(1)^{-1} \tilde{\Psi}(f) \quad \text{for all } f \in \mathcal{K}_\mathbb{R}(G) \text{ with } \Phi(f) \neq 0. \]

Now \( \Phi \) and \( \tilde{\Psi} \) are two linear functionals that agree in the complement of the hyperplane \( H \) in \( \mathcal{K}_\mathbb{R}(G) \) of equation \( \Phi(f) = 0 \), so they agree everywhere. To see this, pick a basis of \( \mathcal{K}_\mathbb{R}(G) \) consisting of a basis \( (h_j)_{j \in J} \) of \( H \) and a function \( v \) not in \( H \). We claim that the family consisting of \( (h_j + v)_{j \in J} \) and \( v \) is a basis of \( \mathcal{K}_\mathbb{R}(G) \). This family obviously spans \( \mathcal{K}_\mathbb{R}(G) \) (since every \( h_j \) is obtained as \( h_j + v - v \)), and it is linearly independent because if we have a finite linear combination

\[ \sum_{i \in I} \lambda_i (h_i + v) + \mu v = 0, \]

for any finite subset \( I \) of \( J \), then

\[ \sum_{i \in I} \lambda_i h_i + \left( \mu + \sum_{i=1}^{|I|} \lambda_i \right) v = 0, \]

and by linear independence, \( \lambda_i = 0 \) for all \( i \in I \) and \( \mu + \sum_{i=1}^{|I|} \lambda_i = 0 \), which implies \( \mu = 0 \), and since this holds for any finite subset \( I \) of \( J \), the family consisting of \((h_j + v)_{j \in J} \) and \( v \)
is linearly independent. Since $\Phi$ and $\Psi$ are linear and they agree on a basis, they must be identical.

Since $\Psi \neq 0$, we must have $D(1) \neq 0$, thus, $\Phi = D(1)^{-1}\Psi$. Since $\Phi$ and $\Psi$ are positive functionals, we must have $D(1) > 0$.

As we observed earlier, since a locally compact group is the disjoint union of $\sigma$-compact cosets, it is legitimate to use Fubini’s theorem.

\section{Examples of Haar Measures}

Here are some examples of Haar measures on various locally compact groups. In most cases, a Haar measure $\mu$ on a locally compact group $G$ is defined indirectly by a Haar functional $f \mapsto \int f d\mu$, for all $f \in K_C(G)$. This Haar functional is denoted by $d\mu$.

\textbf{Example 6.1.} The additive group $\mathbb{R}$ is a locally compact group, and the Lebesgue measure $\mu_L$ is a left (and right) Haar measure on it. For a proof, see Lang [52] (Chapter VI, Theorem 9.7). An alternative is to use Proposition 6.16(3). By the simple change of variable $x = t + s$, for any function $f \in K_C(\mathbb{R})$, we have

$$\int_{-\infty}^{\infty} f(t)dt = \int_{-\infty}^{\infty} f(t + s)dt.$$ 

\textbf{Example 6.2.} The additive group $\mathbb{R}^n$ is a locally compact group, and the product Lebesgue measure $\mu_L$ on it (see Section 4.9) is a left (and right) Haar measure on it. This will be shown as an application of Proposition 6.36.

\textbf{Example 6.3.} The multiplicative group $\mathbb{R}_+^*$ is a locally compact group. We claim that $d\mu = dx/x$ is a left Haar measure, where $dx$ is the restriction of the Lebesgue measure to $\mathbb{R}_+^*$. Indeed, using the change of variable $t \mapsto st$, for any function $f \in K_C(\mathbb{R}_+^*)$, we have

$$\int_0^\infty \frac{f(t)dt}{t} = \int_0^\infty \frac{f(st)sd(t)}{st} = \int_0^\infty \frac{f(st)d(t)}{t},$$

establishing left-invariance. One might wonder what is the measure $\mu([a,b])$ of a closed interval, with $0 < a < b$. We have

$$\mu([a,b]) = \int \chi_{[a,b]} d\mu = \int_{[a,b]} dt = \int_a^b \frac{dt}{t} = \left[ \log t \right]_a^b = \log \frac{b}{a}.$$ 

For any $s > 0$, we have $s \cdot [a,b] = [sa, sb]$, and

$$\mu(s \cdot [a,b]) = \mu([sa, sb]) = \log \frac{sb}{sa} = \log \frac{b}{a}.$$ 

This measure is indeed left invariant.
Example 6.4. Let $\mathbb{T} = \mathbb{U}(1) = \{e^{it} \mid -\pi \leq t < \pi\}$, the circle group, that is, the group of complex numbers of unit length. Let $\sigma : \mathbb{T} \to \mathbb{R}$ be the injection given by

$$\sigma(e^{i\theta}) = \theta, \quad -\pi \leq \theta < \pi.$$  

Define the measure $\nu_1$ on $\mathbb{T}$ by

$$\nu_1(A) = \mu_L(\sigma(A)),$$

on the $\sigma$-algebra $\sigma^{-1}(\mathcal{B}(\mathbb{R}))$ defined in Proposition 4.2(2) (where $\mu_L$ is the Lebesgue measure on $\mathbb{R}$). For any $f \in L^1(\mathbb{T})$, we have

$$\int_{\mathbb{T}} f \, d\nu_1 = \int_{-\pi}^{\pi} f(e^{i\theta}) \, d\mu_L(\theta),$$

also written as $\int_{-\pi}^{\pi} f(e^{i\theta}) \, d\theta$. Observe that $\int_{-\pi}^{\pi} d\nu_1 = 2\pi$. It is easy to check that $\nu_1$ is left-invariant. Indeed, for $\theta_0 \in [-\pi, \pi)$, if we let $\varphi = \theta + \theta_0$, we have

$$\int_{-\pi}^{\pi} f(e^{i(\theta+\theta_0)}) \, d\theta = \int_{-\pi+\theta_0}^{\pi+\theta_0} f(e^{i\varphi}) \, d\varphi = \int_{-\pi+\theta_0}^{\pi} f(e^{i\varphi}) \, d\varphi + \int_{\pi}^{\pi+\theta_0} f(e^{i\varphi}) \, d\varphi.$$

Using the change of variable $\varphi = u + 2\pi$ in the second integral, we get

$$\int_{\pi}^{\pi+\theta_0} f(e^{i\varphi}) \, d\varphi = \int_{-\pi}^{-\pi+\theta_0} f(e^{iu}) \, du,$$

and so

$$\int_{-\pi}^{\pi} f(e^{i(\theta+\theta_0)}) \, d\theta = \int_{-\pi+\theta_0}^{\pi} f(e^{iu}) \, du + \int_{\pi}^{\pi+\theta_0} f(e^{iu}) \, du = \int_{-\pi}^{\pi} f(e^{i\theta}) \, d\theta.$$

Example 6.5. Let $G = \text{GL}(n, \mathbb{R})$, the group of invertible $n \times n$ real matrices. It can be shown that a left (and right) Haar measure on $\text{GL}(n, \mathbb{R})$ is given by

$$d\mu = \frac{dA}{|\det(A)|^{n}} = |\det(A)|^{-n} \bigotimes_{i,j} da_{ij},$$

with $A = (a_{ij})$, where $da_{ij}$ is the Lebesgue measure on $\mathbb{R}$, and $dA$ is the Lebesgue measure on $\mathbb{R}^{n^2}$.

In the next section, we explore the relationship between a left Haar measure $\mu$ and the left Haar measure $\rho_s(\mu)$. 

CHAPTER 6. THE HAAR MEASURE AND CONVOLUTION

6.6 The Modular Function

Let $\mu$ be a left Haar measure on the locally compact group $G$. For all $s, t \in G$, since $\lambda_t$ and $\rho_s$ commute (on functions, measures, and Radon functionals), since $\mu$ is a left Haar measure, $\lambda_t(\mu) = \mu$, so we have
$$\lambda_t(\rho_s(\mu)) = \rho_s(\lambda_t(\mu)) = \rho_s(\mu),$$
which means that $\rho_s(\mu)$ is also a left Haar measure. By the uniqueness result of Theorem 6.21, there is a constant $a > 0$ such that
$$\rho_s(\mu) = a\mu.$$
If $\nu$ is another left Haar measure, again by theorem Theorem 6.21, we have $\nu = c\mu$ for some $c > 0$, but $\rho_s(\nu)(A) = \nu(As) = c\mu(As) = c\rho_s(\mu)(A)$ for all $A \in \mathcal{B}$, that is, $\rho_s(\nu) = c\rho_s(\mu)$, so
$$\rho_s(\nu) = c\rho_s(\mu) = ca\mu = ac\mu = av.$$
Therefore, the number $a$ such that $\rho_s(\mu) = a\mu$ is independent of $\mu$. It is customary to denote this number by $\Delta(s)$.

**Definition 6.12.** Let $G$ be a locally compact group. For every $s \in G$, there is a unique positive number $\Delta(s)$ such that
$$\rho_s(\mu) = \Delta(s)\mu$$
for all left Haar measures $\mu$. The function $\Delta: G \to \mathbb{R}_+^*$ (given by $\Delta(s)$ for every $s \in G$) is called the **modular function** of $G$ (if necessary, we denote it by $\Delta_G$ to avoid ambiguities).

Observe that $(*)$ can be expressed as
$$\mu(As) = \Delta(s)\mu(A) \quad \text{for all } A \in \mathcal{B} \text{ and all } s \in G.$$

**Proposition 6.22.** Let $G$ be a locally compact group, and let $\mu$ be a left Haar measure on $G$. For any $f \in L^1(G, \mathcal{B}, \mathbb{C})$, we have
$$\int f \rho_s(f) \, d\mu = \Delta(s^{-1}) \int f \, d\mu.$$ 

*The function $\Delta: G \to \mathbb{R}_+^*$ is a continuous homomorphism.*

**Proof sketch.** Let $f = \sum_{i=1}^n y_i \chi_{A_i}$ be a $\mu$-step function. For all $x \in G$, we have $\rho_s(f)(x) = f(xs) = y_i$ iff $xs \in A_i$ iff $x \in A_is^{-1}$, which shows that
$$\rho_s(f) = \sum_{i=1}^n y_i \chi_{A_is^{-1}}.$$ Consequently,
$$\int \rho_s(f) \, d\mu = \sum_{i=1}^n y_i \mu(A_is^{-1}) = \Delta(s^{-1}) \sum_{i=1}^n y_i \mu(A_i) = \Delta(s^{-1}) \int f \, d\mu,$$
by (\ast). As in the proof of Proposition 6.16, every function $f \in \mathcal{L}_\mu^1(G, \mathcal{B}, \mathbb{C})$ has some approximation sequence $(f_n)$ by $\mu$-step functions that converges to $f$ a.e. and in the $L^1$-norm, which allows us to conclude that

$$\int \rho_s(f) \, d\mu = \Delta(s^{-1}) \int f \, d\mu$$

for every $f \in \mathcal{L}^1(G, \mathcal{B}, \mathbb{C})$.

By (\ast) we have

$$\Delta(st) \mu(A) = \mu(Ast) = \Delta(t) \mu(As) = \Delta(t) \Delta(s) \mu(A)$$

for all $A \in \mathcal{B}$, and since $\mu(A) > 0$ if $A$ is open and nonempty (and $\mathbb{R}_+$ is commutative under multiplication!), we deduce that

$$\Delta(st) = \Delta(s) \Delta(t).$$

Thus $\Delta$ is a homomorphism from $G$ to the multiplicative group $\mathbb{R}_+^*$. Proposition 6.12 implies that the map $s \mapsto \rho_s(f)$ is uniformly continuous, and so it can be shown that the map $s \mapsto \int \rho_s(f) \, d\mu$ is continuous, and since

$$\int \rho_s(f) \, d\mu = \Delta(s^{-1}) \int f \, d\mu,$$

we deduce that $\Delta$ is continuous. $\square$

The equation

$$\int \rho_s(f) \, d\mu = \Delta(s^{-1}) \int f \, d\mu$$

is also written as

$$\int f(xs) \, d\mu(x) = \Delta(s^{-1}) \int f(x) \, d\mu(x),$$

or equivalently as

$$\int f(xs^{-1}) \, d\mu(x) = \Delta(s) \int f(x) \, d\mu(x).$$

Since $\Delta: G \to \mathbb{R}_+^*$ is a group homomorphism, we have $\Delta(s^{-1}) = (\Delta(s))^{-1}$, and we write $\Delta^{-1}$ for the function given by $\Delta^{-1}(s) = \Delta(s^{-1})$.

Let $G$ and $G'$ be two locally compact groups. An isomorphism is map $\varphi: G \to G'$ which is a group isomorphism and a homeomorphism. Then it is easy to check that $\varphi$ maps Borel sets of $G$ to Borel sets of $G'$. Given a measure $\mu'$ on $G'$, we define the map $\varphi^{-1}(\mu')$ with domain $\mathcal{B}(G)$ by

$$(\varphi^{-1}(\mu'))(A) = \mu'(\varphi(A)) \text{ for all } A \in \mathcal{B}(G).$$
Proposition 6.23. Let $G$ and $G'$ be two locally compact groups, and $\varphi : G \to G'$ be an isomorphism. For any left Haar measure $\mu'$ on $G'$, the map $\varphi^{-1}(\mu')$ is a left Haar measure on $G$. We have
\[ \Delta_G = \Delta_{G'} \circ \varphi. \]

Proof. The fact that $\varphi^{-1}(\mu')$ is a measure follows from the fact that $\varphi$ maps Borel sets to Borel sets and is a bijection, so it preserves union and disjointness. The details are left as an exercise. Since $\mu'$ is a left Haar measure and $\varphi$ is a homomorphism,
\[ \varphi^{-1}(\mu')(sA) = \mu'(\varphi(sA)) = \mu'(\varphi(s)\varphi(A)) = \mu'(\varphi(A)) = \varphi^{-1}(\mu')(A), \]
so $\varphi^{-1}(\mu')$ is a left Haar measure.

We have
\[ (\varphi^{-1}(\mu')(A)) = \Delta_G(s)(\varphi^{-1}(\mu')(A)) = \Delta_G(s)\mu'(\varphi(A)), \]
and
\[ (\varphi^{-1}(\mu')(A)) = \mu'(\varphi(A)) = \mu'(\varphi(A)) = \Delta_{G'}(\varphi(s))\mu'(\varphi(A)), \]
which implies
\[ \Delta_{G'}(\varphi(s)) = \Delta_G(s) \quad \text{for all } s \in G \]
since we can pick a nonempty open subset $A$ of $G$, and $\varphi(A)$ is a nonempty open subset of $G'$.

Corollary 6.24. If $G' = G$, that is, $\varphi : G \to G$ is an automorphism, then $\Delta \circ \varphi = \Delta$.

Definition 6.13. Let $G$ be a locally compact group. We say that $G$ is unimodular if $\Delta(s) = 1$ for all $s \in G$, equivalently, if and only if a left Haar measure is also a right Haar measure.

Luckily, many familiar groups are unimodular (but unfortunately, not the affine groups of rigid motions). Obviously, abelian locally compact groups are unimodular.

Given a group $G$, recall that its commutator subgroup $[G, G]$ is the subgroup generated by all elements $[s, t] = sts^{-1}t^{-1}$. The group $[G, G]$ is a normal subgroup of $G$.

Proposition 6.25. Let $G$ be a locally compact group.

(1) If there is a compact neighborhood $V$ of 1 such that $sVs^{-1} = V$ for all $s \in G$, then $G$ is unimodular. Consequently, if $G$ is compact, discrete, or commutative, then $G$ is unimodular.
6.6. THE MODULAR FUNCTION

(2) If $K$ is any compact subgroup of $G$, then $\Delta | K \equiv 1$.

(3) If $G/[G,G]$ is compact, then $G$ is unimodular. As a consequence, every connected semisimple Lie group is unimodular. Recall that semisimple Lie group is a Lie group $G$ such that the Killing form on its Lie algebra $\mathfrak{g}$ is nondegenerate.

Proof. (1) Let $\mu$ be any left Haar measure. Since $\mu$ is left-invariant
\[
\mu(V) = \mu(sVs^{-1}) = \mu(Vs^{-1}) = \Delta(s)\mu(V),
\]
but $\mu(V) > 0$ because $V$ contains a nonempty open subset, so $\Delta(s) = 1$ for all $s \in G$. The corollaries are left as an easy exercise.

(2) Since $\Delta$ is continuous, $\Delta(K)$ is a compact subgroup of $\mathbb{R}^*$, which implies $\Delta(K) = \{1\}$.

(3) Since $\mathbb{R}^*$ is abelian, we have
\[
\Delta([s,t]) = \Delta(st^{-1}t^{-1}) = \Delta(s)\Delta(t)\Delta(s)^{-1}\Delta(t)^{-1} = \Delta(s)\Delta(s)^{-1}\Delta(t)\Delta(t)^{-1} = 1,
\]
so $\Delta$ vanishes on $[G,G]$. It follows that $\Delta$ factors through $G/[G,G]$ as $\Delta = \pi \circ \theta$ where $\theta: G/[G,G] \to \mathbb{R}^*$ is a continuous homomorphism. Since $G/[G,G]$ is compact, we have $\theta(G/[G,G]) = \{1\}$, so $\Delta(G) = \{1\}$.

If $G$ is a connected semisimple Lie group, it is known that $G = [G,G]$, so $G$ is unimodular.

In order to discuss the behavior of the operator $\mu \mapsto \hat{\mu}$ we need the following proposition.

Proposition 6.26. Let $\mu$ and $\nu$ be two Radon measures on a locally compact topological space $X$. If there is a continuous function $g: X \to \mathbb{R}^*$ such that that
\[
\int f d\nu = \int fg \, d\mu \quad \text{for all } f \in \mathcal{C}(X),
\]
and if $\tilde{\nu}$ is the Radon measure given by
\[
\tilde{\nu}(E) = \int_E g \, d\mu \quad \text{for all } E \in \mathcal{B}(X),
\]
then $\nu = \tilde{\nu}$.

A proof of Proposition 6.26 is given in Folland [31] (Chapter 2, Proposition 2.23).

We propose to denote the Radon measure $\tilde{\nu}$ by $g \cdot \mu$, by analogy with the definition of the Radon functional $g \cdot \Phi$ in Example 5.1(3). The notation $gd\mu$ is also used.

The following proposition shows the behavior of the operator $\mu \mapsto \hat{\mu}$.
Proposition 6.27. Let $G$ be a locally compact group. For every left Haar measure $\mu$, the measure $\hat{\mu}$ is a right Haar measure, and we have

$$\int \hat{f} \, d\mu = \int f \, d\hat{\mu}, \quad \hat{\mu} = \Delta^{-1} \cdot \mu,$$

and

$$\int f(s^{-1})\Delta(s^{-1})d\mu(s) = \int f(s)d\mu(s)$$

for all $f \in \mathcal{L}^1(G,\mathcal{B},\mathbb{C})$.

Proof sketch. For every $\mu$-step function

$$\hat{f} = \sum_{i=1}^{n} y_i \chi_{A_i},$$

since $\hat{f}(s) = f(s^{-1})$, we immediately obtain

$$\hat{\hat{f}} = \sum_{i=1}^{n} y_i \chi_{A_i^{-1}},$$

and since $\hat{\mu}(A_i) = \mu(A_i^{-1})$, we get

$$\int \hat{\hat{f}} \, d\mu = \int f \, d\hat{\mu}.$$

Then by a familiar argument using approximations sequences, we deduce that

$$\int \hat{\hat{f}} \, d\mu = \int f \, d\hat{\mu}$$

for all $f \in \mathcal{L}^1(G,\mathcal{B},\mathbb{C})$.

Using the fact that $\Delta$ is a group homomorphism, for any $f \in \mathcal{K}_C(G)$, we have

$$\int (\rho_{t^{-1}}(f))(s)\Delta(s^{-1})d\mu(s) = \Delta(t^{-1})\int f(st^{-1})\Delta(t)\Delta(s^{-1})d\mu(s) \quad \text{by Definition of } \rho_{t^{-1}}(f)$$

$$= \Delta(t^{-1}) \int f(st^{-1})\Delta^{-1}(st^{-1})d\mu(s)$$

$$= \Delta(t^{-1})\Delta(t) \int f(s)\Delta^{-1}(s)d\mu(s) \quad \text{by Proposition 6.22}$$

$$= \int f(s)\Delta(s^{-1})d\mu(s),$$
which shows that the Radon functional \( f \mapsto \int f(s)\Delta(s^{-1})d\mu(s) \) is right-invariant. The corresponding Haar measure \( \nu \) is a right Haar measure, and since \( \check{\mu} \) is a right Haar measure, there is some \( a > 0 \) such that \( a\check{\mu} = \nu \). Then we have

\[
a \int f(s)d\check{\mu}(s) = \int f(s)d\nu = \int f(s)\Delta(s^{-1})d\mu(s),
\]

for all \( f \in \mathcal{K}_C(G) \), and since \( \Delta^{-1} \) is a positive continuous function, by Proposition 6.26, \( \nu = \Delta^{-1} \cdot \mu \), so

\[
a\check{\mu} = \Delta^{-1} \cdot \mu.
\]

It remains to show that \( a = 1 \).

Assume that \( a \neq 1 \). Since \( \Delta \) is continuous, there is a symmetric neighborhood \( U \) of 1 such that \( |\Delta(s^{-1}) - 1| \leq 1/(2a - 1) \) on \( U \). Since \( U \) is symmetric, \( \mu(U) = \check{\mu}(U) \), and we have

\[
|a - 1|\mu(U) = |a\check{\mu}(U) - \mu(U)| = \left| \int_U (\Delta(s^{-1}) - 1)d\mu(s) \right| \leq \frac{1}{2}|a - 1|\mu(U),
\]

a contradiction.

Therefore,

\[
\int f(s)d\check{\mu}(s) = \int f(s)\Delta(s^{-1})d\mu(s),
\]

for all \( f \in \mathcal{L}^1(G, \mathcal{B}, \mathbb{C}) \), so

\[
\int f(s)\Delta(s^{-1})d\mu(s) = \int \check{f}(s)d\mu(s),
\]

and by changing \( f \) to \( \check{f} \), we obtain the desired equation.

In summary, if \( G \) is unimodular, then we have

\[
\int f(sx)d\mu(x) = \int f(xs)d\mu(s) = \int f(x^{-1})d\mu(x) = \int f(x)d\mu(x)
\]

for all \( f \in \mathcal{L}^1(G, \mathcal{B}, \mathbb{C}) \), and

\[
\mu(A) = \mu(As) = \mu(sA) = \mu(A^{-1}),
\]

for all \( A \in \mathcal{B} \).

**Remark:** If \( G \) is a Lie group, then it can be shown that the modular function \( \Delta \) is given by

\[
\Delta(s) = |\det \text{Ad}(s^{-1})|;
\]

see Gallier and Quaintance [36] (Chapter 25, Proposition 25.23).
6.7 More Examples of Haar Measures

In the examples of Section 6.5, the groups under consideration were unimodular. The groups of the next examples are not unimodular.

**Example 6.6.** Let $G = \mathbf{GL}(n, \mathbb{R})$, the group of invertible upper triangular matrices. It can be shown that a left Haar measure on $\mathbf{GL}(n, \mathbb{R})$ is given by

$$d\mu_L = |\det(A)|^{-n-1} \prod_{i \leq j} da_{ij} \prod_i du_i,$$

with $A = (a_{ij})$, and $du_i$ is the Lebesgue measure on $\mathbb{R}$. A right Haar measure is given by

$$d\mu_R = |\det(A)|^{-n} \prod_{i \leq j} da_{ij} \prod_i du_i,$$

and the modular function is given by

$$\Delta(A) = |\det(A)|^{-1}.$$

A proof of these facts can be found in Bourbaki [6] (Chapter VII, Section 2, no. 10, Proposition 14, and Section 3, no. 3, Example 4).

**Example 6.7.** Let $G = \mathbf{T}(n, \mathbb{R})$, the group of invertible upper triangular matrices. It can be shown that a left Haar measure on $\mathbf{T}(n, \mathbb{R})$ is given by

$$d\mu_L = \prod_{i=1}^n |a_{ii}|^{-i} \prod_{i \leq j} da_{ij},$$

with $A = (a_{ij})$, and $da_{ij}$ is the Lebesgue measure on $\mathbb{R}$. A right Haar measure is given by

$$d\mu_R = \prod_{i=1}^n |a_{ii}|^{-i} \prod_{i \leq j} da_{ij},$$

and the modular function is given by

$$\Delta(A) = \prod_{i=1}^n |a_{ii}|^{2i-n-1}.$$

A proof of these facts can be found in Bourbaki [6] (Chapter VII, Section 2, no. 10, Proposition 14, and Section 3, no. 3, Example 4).

More examples can be found in Bourbaki [6] (Chapter VII, Section 3, no. 3). The group $\mathbf{SL}(n, \mathbb{R})$ is unimodular, but finding a Haar measure for it is nontrivial.
6.8 The Modulus of an Automorphism

We now consider the effect of an automorphism $u: G \to G$ on a Haar measure. Recall that $u$ is a group isomorphism and a homeomorphism.

**Definition 6.14.** Let $G$ be a locally compact group, and let $u: G \to G$ be an automorphism of $G$. For every function $f: G \to \mathbb{C}$, define the function $u(f)$ by

$$(u(f))(s) = f(u^{-1}(s)), \quad \text{for all } s \in G,$$

and for every left Haar measure $\mu$, define the measure $u^{-1}(\mu)$ by

$$(u^{-1}(\mu))(A) = \mu(u(A)), \quad \text{for all } A \in \mathcal{B}.$$ 

It is immediately verified that if $u$ and $v$ are two automorphisms of $G$, then

$$(u \circ v)(f) = (u(v(f))), \quad (u \circ v)(\mu) = (u(v(\mu))).$$

Also observe that since $\mu$ is left-invariant and $u$ is an automorphism,

$$(u^{-1}(\mu))(sA) = \mu(\mu(sA)) = \mu(\mu(s)u(A)) = \mu(u(A)) = u^{-1}(\mu)(A),$$

so $u^{-1}(\mu)$ is left-invariant. By the uniqueness of a left Haar measure up to a constant, there is a real $a > 0$ such that $u^{-1}(\mu) = a\mu$. For any other left Haar measure $\nu = c\mu$, we have

$$(u^{-1}(\nu))(A) = \nu(u(A)) = c\mu(u(A)) = cu^{-1}(\mu)(A) = ca\mu(A) = ac\mu(A) = a\nu(A).$$

Therefore, the constant $a$ is independent of the left Haar measure $\mu$.

**Definition 6.15.** Let $G$ be a locally compact group. For every automorphism $u: G \to G$, there is a unique positive number $\text{mod}(u)$ such that

$$u^{-1}(\mu) = \text{mod}(u)\mu$$

for all left Haar measures $\mu$. The number $\text{mod}(u)$ is called the **modulus of the automorphism** $u$.

Note that the condition of Definition 6.15 can also be expressed as

$$\mu(u(A)) = \text{mod}(u)\mu(A) \quad \text{for all } A \in \mathcal{B}. \quad (***)$$

**Proposition 6.28.** Let $G$ be a locally compact group, and let $\mu$ be any left Haar measure on $G$. For every automorphism $u: G \to G$, we have

$$\int u(f)d\mu = \int fdu^{-1}(\mu) = \text{mod}(u)\int fd\mu.$$ 

for all $f \in L^1(G, \mathcal{B}, \mathbb{C})$. 

Proof sketch. For every $\mu$-step function
\[ f = \sum_{i=1}^{n} y_i \chi_{A_i}, \]

since $(u(f))(s) = f(u^{-1}(s))$, we have $f(u^{-1}(s)) = y_i$ iff $u^{-1}(s) \in A_i$ iff $s \in u(A_i)$, which means that
\[ u(f) = \sum_{i=1}^{n} y_i \chi_{u(A_i)}; \]

Thus
\[ \int u(f) \, d\mu = \sum_{i=1}^{n} y_i \mu(u(A_i)) = \sum_{i=1}^{n} y_i u^{-1}(\mu)(A_i) = \int f \, du^{-1}(\mu). \]

Then by a familiar argument using approximations sequences, we deduce that
\[ \int u(f) d\mu = \int f \, du^{-1}(\mu) \]

for all $f \in L^1(G, \mathcal{B}, \mathbb{C})$. Since $u^{-1}(\mu) = \text{mod}(u)\mu$, we get the second equation. \qed

Proposition 6.15 can also be stated as
\[ \int f(u^{-1}(s)) \, d\mu(s) = \text{mod}(u) \int f(s) \, d\mu(s). \]

Suppose that $\mu$ is a right Haar measure. As in the case of a left Haar measure, we define the measure $u^{-1}(\mu)$ by
\[ (u^{-1}(\mu))(A) = \mu(u(A)) \quad \text{for all } A \in \mathcal{B}(G). \]

The measure $u^{-1}(\mu)$ is a right Haar measure because
\[ (u^{-1}(\mu))(As) = \mu(u(As)) = \mu(u(A)u(s)) = \mu(u(A)) = (u^{-1}(\mu))(A). \]

As in the left-invariant case, for every automorphism $u: G \to G$, there a constant $c > 0$ such that $u^{-1}(\mu) = c\mu$. Interestingly, $c = \text{mod}(u)$, so there is no difference between the left modulus and the right modulus of an automorphism.

Proposition 6.29. Let $G$ be a locally compact group, and let $u: G \to G$ be an automorphism. Then for all left Haar measure, and all right Haar measures, $\mu$ on $G$, we have
\[ u^{-1}(\mu) = \text{mod}(u)\mu, \]

where $\text{mod}(u)$ is the modulus of $u$ defined for left Haar measures (see Definition 6.15).
6.8. THE MODULUS OF AN AUTOMORPHISM

Proof. We use Corollary 6.24 which implies that \( \Delta \circ u^{-1} = \Delta \), since \( u^{-1} \) is also an automorphism when \( u \) is an automorphism. As a consequence,

\[
\Delta(s^{-1}) = \Delta(u^{-1}(s^{-1})) = \Delta((u^{-1}(s))^{-1}) = \Delta^{-1}(u^{-1}(s)),
\]

that is,

\[
\Delta(s^{-1}) = \Delta^{-1}(u^{-1}(s)). \quad (†)
\]

Recall that if \( \mu \) is a left Haar measure, then \( \tilde{\mu} \) is a right Haar measure, and by Proposition 6.27 we have \( \tilde{\mu} = \Delta^{-1} \cdot \mu \). Then for every \( f \in \mathcal{K}_C(G) \), we have

\[
\int f(s) \, du^{-1}(\Delta^{-1} \cdot \mu)(s) = \int (u(f))(s) \, d(\Delta^{-1} \cdot \mu)(s) \quad \text{by Proposition 6.28}
\]

\[
= \int f(u^{-1}(s)) \Delta(s^{-1}) \, d\mu(s) \quad \text{by Proposition 6.27}
\]

\[
= \int f(u^{-1}(s)) \Delta^{-1}(u^{-1}(s)) \, d\mu(s) \quad \text{by (†)}
\]

\[
= \text{mod}(u) \int f(s) \Delta^{-1}(s) \, d\mu(s) \quad \text{by Proposition 6.28}
\]

\[
= \text{mod}(u) \int f(s) \Delta(s^{-1}) \, d\mu(s)
\]

\[
= \text{mod}(u) \int f(s) \, d(\Delta^{-1} \cdot \mu)(s) \quad \text{by Proposition 6.27}
\]

By the uniqueness of the Radon measure associated with a Radon functional, this proves that

\[
u^{-1}(\Delta^{-1} \cdot \mu) = \text{mod}(u) \Delta^{-1} \cdot \mu,
\]

that is, \( u^{-1}(\tilde{\mu}) = \text{mod}(u)\tilde{\mu} \). Since every right Haar measure is of the form \( \tilde{\mu} \) for some left Haar measure \( \mu \), we proved our result.

For every \( s \in G \), if \( C_s \) is the automorphism conjugation by \( s \), namely \( C_s(t) = s^{-1}ts \), then we have the following result.

**Proposition 6.30.** Let \( G \) be a locally compact group. For every \( s \in G \), we have

\[
\text{mod}(C_s) = \Delta(s).
\]

**Proof.** We have

\[
C_s^{-1}(t) = s t s^{-1} = (R_{s^{-1}} \circ L_s)(t).
\]

Observe that for any left Haar measure \( \mu \),

\[
(L_s(\mu))(A) = \mu(L_s^{-1}(A)) = \mu(s^{-1}A) = \mu(A),
\]

and

\[
(R_{s^{-1}}(\mu))(A) = \mu(R_{s^{-1}}^{-1}(A)) = \mu(R_{s}(A)) = \mu(As) = \Delta(s)\mu(A).
\]
Then,
\[ C_{s}^{-1}(\mu) = R_{s}^{-1}(L_{s}(\mu)) = R_{s}^{-1}(\mu) = \Delta(s)\mu, \]
which shows that
\[ \mod(C_{s}) = \Delta(s), \]
as claimed.

**Proposition 6.31.** Let \( G \) be a locally compact group.

1. If \( G \) is compact or discrete, then \( \mod(u) = 1 \) for any automorphism \( u: G \to G \).

2. For any two automorphisms \( u: G \to G \) and \( v: G \to G \), we have
\[ \mod(u \circ v) = \mod(u) \mod(v). \]

**Proof.** (1) Since \( u \) is an automorphism \( u(G) = G \) and \( u(\{1\}) = \{1\} \). If \( G \) is compact, let \( A = G \) in the equation (**),
\[ \mu(u(A)) = \mod(u)\mu(A), \]
and if \( G \) is discrete, let \( A = \{1\} \), for any left Haar measure \( \mu \).

(2) Using equation (**), we have
\[ \mu((u \circ v)(A)) = \mu(u(v(A))) = \mod(u)\mu(v(A)) = \mod(u) \mod(v)\mu(A), \]
and we choose \( A \) to be any open nonempty subset.

If \( G = \mathbb{R}^{n} \) (which is a locally compact group under addition with the topology induced by any norm), and \( u \) a linear automorphism of \( \mathbb{R}^{n} \), that is, an invertible linear map of \( \mathbb{R}^{n} \), then we have the following interesting characterization of \( \mod(u) \).

**Proposition 6.32.** Let \( u: \mathbb{R}^{n} \to \mathbb{R}^{n} \) be an invertible linear map, with \( \mathbb{R}^{n} \) as an additive group with the Lebesgue measure. Then
\[ \mod(u) = |\det(u)|. \]

**Sketch of proof.** This result makes use of the following fact from linear algebra which is stated in Gallier (linalg.pdf), Chapter 4, Proposition 4.22, which can be restated as stating that every real \( n \times n \) invertible matrix can be expressed as the product of elementary matrices \( E_{i,j,\beta} = I_{n} + \beta E_{ij} \) and \( I_{n} + (\alpha - 1)E_{nn} \). Then one must check the formula of the proposition,
\[ \int u(f)(x_{1},\ldots,x_{n})d\mu(x_{1},\ldots,x_{n}) = \mod(u)\int f(x_{1},\ldots,x_{n})d\mu(x_{1},\ldots,x_{n}), \]
by integrating the functions of the form
\[ f(x_{1},\ldots,x_{n-1},\alpha x_{n}) \]
and
\[ f(x_{1},\ldots,x_{j} + \beta x_{i},\ldots,x_{n}), \]
with \( f \in K_{\mathbb{R}}(\mathbb{R}^{n}) \), using a change of variables. Details can be found in Dieudonné [22] (Chapter XIV, Proposition 14.3.9.1).
As an application of Proposition 6.32, we obtain formulae for the measure (volume) of a parallelotope, and of a simplex in $\mathbb{R}^n$.

Let $(v_1, \ldots, v_n)$ be $n$ linearly independent vectors in $\mathbb{R}^n$. Then the set
\[ P = \{\lambda_1 v_1 + \cdots + \lambda_n v_n \mid (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n, \ 0 \leq \lambda_i \leq 1\} \]
is called a parallelotope. The set
\[ S = \{\lambda_1 v_1 + \cdots + \lambda_n v_n \mid (\lambda_1, \ldots, \lambda_n) \in \mathbb{R}^n, \lambda_i \geq 0, \lambda_1 + \cdots + \lambda_n \leq 1\} \]
is called a simplex.

**Proposition 6.33.** Let $(v_1, \ldots, v_n)$ be $n$ linearly independent vectors in $\mathbb{R}^n$, and let $P$ be the parallelotope and $S$ be the simplex determined by $(v_1, \ldots, v_n)$. If $\mu$ is the Lebesgue measure on $\mathbb{R}^n$, then
\[ \mu(P) = |\det(v_1, \ldots, v_n)|, \quad \mu(S) = \frac{1}{n!} |\det(v_1, \ldots, v_n)|. \]

**Proof sketch.** Since $(v_1, \ldots, v_n)$ are linearly independent. There is a unique linear map $u$ such that $u(e_i) = v_i$, for $i = 1, \ldots, n$, where $e_i$ is the canonical basis vector of $\mathbb{R}^n$. Then $P = u(K)$, where $K$ is the $n$-cube determined by $(e_1, \ldots, e_n)$, and
\[ \mu(P) = \int \chi_P d\mu = \int \chi_K \circ u^{-1} d\mu = \int u(\chi_K) d\mu = \mod(u) \int \chi_K d\mu = |\det(u)| \mu(K). \]
But under the Lebesgue measure, $\mu(K) = 1$, and we get
\[ \mu(P) = |\det(u)| = |\det(v_1, \ldots, v_n)|, \]
as claimed.

For the simplex, write $S_m$ for simplex determined by the canonical basis vectors $e_1, \ldots, e_m$, and write $\mu_m$ for the Lebesgue measure in $\mathbb{R}^m$. Then $S = u(S_n)$, so
\[ \mu(S) = |\det(u)| \mu(S_n), \]
and we are reduced to computing $\mu(S_n)$. We view $\mathbb{R}^n$ as $\mathbb{R}^{n-1} \times \mathbb{R}$ and we consider the section $S_n(\lambda)$ of $S_n$ consisting of the set of points $(x_1, \ldots, x_{n-1}) \in \mathbb{R}^{n-1}$ such that
\[ x_1, \ldots, x_{n-1} \geq 0, \quad x_1 + \cdots + x_{n-1} \leq \lambda - 1, \]
where $0 \leq \lambda \leq 1$. This section is the image of $S_n(0)$ under the scaling by $\lambda - 1$, so we get
\[ \mu_{n-1}(S_n(\lambda)) = (1 - \lambda)^{n-1} \mu_{n-1}(S_{n-1}). \]
Then by Fubini we get
\[ \mu_n(S_n) = \int_0^1 (1 - \lambda)^{n-1} \mu_{n-1}(S_{n-1}) d\lambda = \frac{1}{n} \mu_{n-1}(S_{n-1}). \]
By induction
\[ \mu_{n-1}(S_{n-1}) = \frac{1}{(n-1)!}, \]
so we get
\[ \mu_n(S_n) = \frac{1}{n!} \]
as claimed. \(\square\)

As another application of Proposition 6.32, the computation of the measure (volume) of a closed ball in \(\mathbb{R}^n\) can be found in Dieudonné [22] (Chapter XIV, Proposition 14.3.11).

More on the modular function and the modulus of an automorphism can be found in Bourbaki [6], Chapter VII, Section 1.

6.9 Some Properties and Applications of the Haar Measure

The following result is proved in Dieudonné [22] (Chapter XIV, Proposition 14.2.3).

**Proposition 6.34.** Let \(G\) be a locally compact group, and let \(\mu\) be a left Haar measure on \(G\). Then \(G\) is discrete if and only if \(\mu(\{1\}) > 0\), and \(G\) is compact if and only if \(\mu(G) < +\infty\).

An interesting and important application of the Haar measure is the construction of a Hermitian inner product invariant under the representation of a compact group. The idea of such a construction originates with Hurwitz, and was generalized by H. Weyl.

Let \(K\) be a topological group, and let \(H\) be a (complex) finite-dimensional Hermitian space (with inner product \(\langle -, - \rangle\) and corresponding norm \(\|\|\)). A *representation* of \(K\) in \(H\) is a group homomorphism \(U: K \to \text{GL}(H)\), where \(\text{GL}(H)\) is the group of invertible linear maps on \(H\).

**Theorem 6.35.** Let \(K\) be a compact group, let \(H\) be a finite-dimensional Hermitian space, and let \(U: K \to \text{GL}(H)\) be a representation of \(K\) which is continuous when \(\text{GL}(H)\) is equipped with the operator norm. Then there is a Hermitian inner product \(\varphi\) on \(H\) such that
\[ \varphi(U_s(x), U_s(y)) = \varphi(x, y) \quad \text{for all } s \in K \text{ and all } x, y \in H. \]

In other words, the linear maps \(U_s\) are unitary transformations with respect to \(\varphi\). Furthermore, the norms \(\|\|\|\) and \(x \mapsto \sqrt{\varphi(x, x)}\) on \(H\) are equivalent.

**Proof.** Let \(\mu\) be a right Haar measure on \(K\) (which is also a left Haar measure since \(K\) is compact). By hypothesis the map
\[ s \mapsto \langle U_s(x), U_s(y) \rangle \]
is continuous for all $x, y \in H$. Define $\varphi$ by

$$\varphi(x, y) = \int \langle U_s(x), U_s(y) \rangle \, d\mu(s).$$

It is immediately verified that $\varphi$ is a sesquilinear form on $H$. Since $H$ is finite-dimensional, the sphere $S = \{x \in H \mid \|x\| = 1\}$ is compact. Since $U$ is continuous, the map $\theta: K \times S \to \mathbb{R}$ given by

$$\theta(s, x) = \|U_s(x)\|$$

is continuous, and since $K$ and $S$ are compact, $K \times S$ is compact so $\theta$ achieves a minimum $m > 0$ and a maximum $M > 0$. We deduce that for every $x \neq 0$,

$$m \|x\| \leq \|U_s(x)\| \leq M \|x\|,$$

that is,

$$m \|x\| \leq \|U_s(x)\| \leq M \|x\|.$$

As a consequence, we get

$$m^2 \mu(K) \|x\|^2 \leq \varphi(x, x) \leq M^2 \mu(K) \|x\|^2,$$

which shows that $\varphi$ is indeed positive definite, and that $\|\|$ is equivalent to the norm induced by $\varphi$. Finally, for every $t \in K$, since $\mu$ is right-invariant we have

$$\varphi(U_t(x), U_t(y)) = \int \langle U_s(U_t(x)), U_s(U_t(y)) \rangle \, d\mu(s)$$

$$= \int \langle U_{st}(x), U_{st}(y) \rangle \, d\mu(s)$$

$$= \int \langle U_s(x), U_s(y) \rangle \, d\mu(s)$$

$$= \varphi(x, y),$$

as claimed. \qed

Theorem 6.35 is a basic tool in representation theory. For example, if $G$ is a Lie group and if $V$ is a finite-dimensional vector space, for any representation $\rho: G \to \text{GL}(V)$, there is a $G$-invariant inner product on $V$ iff $\overline{\rho(G)}$ is compact; see Gallier and Quaintance [36] (Chapter 19, Theorem 19.5).

Theorem 6.35 will also be used in Section 11.2 to show that every linear representation of a compact group is the sum of irreducible representations.

Regarding the product of Haar measures, we have the following result.
**Proposition 6.36.** Let $G_1$ and $G_2$ be two locally compact groups, and let $\mu_1$ be a left Haar measure on $G_1$ and $\mu_2$ be a left Haar measure on $G_2$. Then the linear functional $\Phi_1 \otimes \Phi_2 : \mathcal{K}_C(G_1 \times G_2) \to \mathbb{C}$ given by

$$(\Phi_1 \otimes \Phi_2)(f) = \int f(x_1, x_2) d\mu_1(x_1) \otimes d\mu_2(x_2)$$

is a left-invariant positive Radon functional. If $G_1$ and $G_2$ are $\sigma$-compact, then the Radon measure $\mu_{\Phi_1 \otimes \Phi_2}$ on $G_1 \times G_2$ associated with $\Phi_1 \otimes \Phi_2$ given by Theorem 5.8 is a left Haar measure extending the product measure $\mu_1 \otimes \mu_2$. Furthermore, if $G_1$ and $G_2$ are also second-countable, then $\mu_{\Phi_1 \otimes \Phi_2} = \mu_1 \otimes \mu_2$.

**Proof sketch.** By Fubini’s Theorem (which applies since $f$ vanishes outside of a compact subset), we have

$$(\lambda_{(s_1, s_2)}(\Phi_1 \otimes \Phi_2))(f) = (\Phi_1 \otimes \Phi_2)(\lambda_{(s_1^{-1}, s_2^{-1})}(f))$$

$= \int (\lambda_{(s_1^{-1}, s_2^{-1})} f)(x_1, x_2) d\mu_1(x_1) \otimes d\mu_2(x_2)$$

$= \int f(s_1 x_1, s_2 x_2) d\mu_1(x_1) \otimes d\mu_2(x_2)$$

$= \int \left( \int f(s_1 x_1, s_2 x_2) d\mu_1(x_1) \right) d\mu_2(x_2)$$

$= \int \left( \int f(x_1, s_2 x_2) d\mu_2(x_2) \right) d\mu_1(x_1)$$

$= \int \left( \int f(x_1, x_2) d\mu_2(x_2) \right) d\mu_1(x_1)$$

$= \int \left( \int f(x_1, x_2) d\mu_1(x_1) \right) d\mu_2(x_2)$$

$= \int f(x_1, x_2) d\mu_1(x_1) \otimes d\mu_2(x_2) = (\Phi_1 \otimes \Phi_2)(f)$.

Therefore, $\Phi_1 \otimes \Phi_2$ is left-invariant. The other two statements are explained in Folland [31] (Chapter 2, Section 2.2).

**Remark:** Since $\Phi_1 \otimes \Phi_2$ is a positive linear functional, by Theorem 5.8, the corresponding Radon measure $\mu_{\Phi_1 \otimes \Phi_2}$ is a left Haar measure on $G_1 \times G_2$. But if $G_1$ or $G_2$ is not $\sigma$-compact then the product measure $\mu_1 \otimes \mu_2$ is not defined, and if $G_1$ or $G_2$ is not second-countable, then the $\sigma$-algebra associated with $\mu_{\Phi_1 \otimes \Phi_2}$ has more Borel subsets than the $\sigma$-algebra associated with $\mu_1 \otimes \mu_2$ (see Definition 4.22).
As an application of Proposition 6.36, since the Lebesgue measure $\mu_L$ on $\mathbb{R}$ is both left and right-invariant, we see that the product measure $\mu_{L,n}$ of $n$ copies of $\mu_L$ is a left and a right Haar measure on $\mathbb{R}^n$. To simplify notation, we may write $\mu_n$ instead of $\mu_{L,n}$, and $L^1(\mu_n)$ instead of $L^1_{\mu_n}(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), \mathbb{C})$.

As a Haar measure, $\mu_n$ is both inner and outer regular.

\section{G-Invariant Measures on Homogeneous Spaces}

Let $X$ be a locally compact space and let $G$ be a locally compact group. Suppose we have a continuous left action $\varphi: G \times X \to X$ of $G$ on $X$ (which means that the map $\varphi$ is continuous, see Definition 6.5). As usual, we write $g \cdot x$ instead of $\varphi(g, x)$. We would like to generalize the notion of left-invariance of a measure on $G$ to the notion of $G$-invariance of a measure $\mu$ on $X$.

\begin{definition}
Let $G$ be a locally compact group, let $X$ be a locally compact space, and let $\cdot: G \times X \to X$ be a continuous left action of $G$ on $X$. For every $s \in G$, define $L_s: X \to X$ by
\[
L_s(x) = s \cdot x \quad \text{for all } x \in C.
\]
For every subset $A$ of $X$ and every $s \in G$, let
\[
s \cdot A = \{ s \cdot a | a \in A \}.
\]
For every function $f: X \to \mathbb{C}$, the function $\lambda_s(f)$ is given by
\[
(\lambda_s(f))(x) = f(s^{-1} \cdot x) \quad \text{for all } x, s \in G,
\]
For every Borel measure $\mu$ on $(X, \mathcal{B}(X))$, the measure $\lambda_s(\mu)$ given by
\[
(\lambda_s(\mu))(A) = \mu(s^{-1} \cdot A) \quad \text{for all } s \in G \text{ and all } A \in \mathcal{B}(X).
\]
For every Radon functional $\Phi: \mathcal{K}_C(X) \to \mathbb{C}$, the Radon functional $\lambda_s(\Phi)$ given by
\[
(\lambda_s(\Phi))(f) = \Phi(\lambda_s^{-1}(f)) \quad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_C(X).
\]

It is immediately verified that
\[
L_{st} = L_s \circ L_t, \quad \lambda_{st}(f) = \lambda_s(\lambda_t(f)),
\]
and
\[
\lambda_{st}(\mu) = \lambda_s(\lambda_t(\mu)), \quad \lambda_{st}(\Phi) = \lambda_s(\lambda_t(\Phi)).
\]

The proof of Proposition 6.16 is immediately adapted to show that
\[
\int_X (\lambda_s^{-1}(f))(x) \, d\mu(x) = \int_X f(s \cdot x) \, d\mu(x) = \int_X f(x) \, d\lambda_s(\mu)(x),
\]
for every $s \in G$, every $f \in L^1_\mu(X, \mathcal{B}, \mathbb{C})$, and every Borel measure $\mu$ on $X$. 

**Definition 6.17.** Let $G$ be a locally compact group, let $X$ be a locally compact space, and let $*: G \times X \to X$ be a continuous left action of $G$ on $X$. A Borel measure $\mu$ on $X$ is $G$-invariant if

$$\lambda_s(\mu) = \mu \quad \text{for all } s \in G.$$  

A Radon functional $\Phi: \mathcal{C}(X) \to \mathbb{C}$ on $X$ is $G$-invariant if

$$\lambda_s(\Phi) = \Phi \quad \text{for all } s \in G.$$  

If $\mu$ is $G$-invariant, then

$$\int_X f(s \cdot x) \, d\mu(x) = \int_X f(x) \, d\mu(x),$$

for every $f \in L^1_\mu(X, \mathcal{B}, \mathbb{C})$ and all $s \in G$. The proof of Proposition 6.16 is immediately adapted to show that if

$$\int_X \lambda_s^{-1}(f)(x) \, d\mu(x) = \int_X f(s \cdot x) \, d\mu(x) = \int_X f(x) \, d\mu(x),$$

for all $f \in \mathcal{C}(X)$ and all $s \in G$, then $\mu$ is $G$-invariant.

Our goal is to find sufficient conditions to ensure that $X$ has some $G$-invariant measure. We will consider the case where $X = G/H$, with the left action of $G$ on $G/H$ given by

$$a \cdot (bH) = abH, \quad a, b \in G.$$  

In this case, by Proposition 6.6, the space $X$ is also locally compact (and Hausdorff).

A $G$-invariant measure on $G/H$ does not always exist. For example, if $G$ is the affine “$ax + b$” group (with $a \neq 0$) and $X = \mathbb{R}$, obviously $G$ acts transitively on $\mathbb{R}$ and the stabilizer of $0$ is $H = \mathbb{R}$. However, the only Borel measure on $\mathbb{R}$ invariant under translation is the Lebesgue measure, but it is not invariant under scaling transformations $x \mapsto ax$ with $a \neq 0, 1$.

It turns out that there is a necessary and sufficient condition for a $G$-invariant $\sigma$-Radon measure to exist on $G/H$ in terms of $\Delta_G$ and $\Delta_H$: $\Delta_H$ must be equal to the restriction of $\Delta_G$ on $H$. We proceed to explain this following Folland’s exposition [31] (Chapter 2, Section 2.6).

Suppose $\mu$ is a left Haar measure on $G$ and $\xi$ is a left Haar measure on $H$. The group $G$ is locally compact and $\sigma$-compact, and $H$ is a closed subgroup of $G$. Denote the quotient map by $\pi: G \to G/H$. The first step is to define a map $P$ from $\mathcal{C}(G)$ to $\mathcal{C}(G/H)$.

**Definition 6.18.** With $(G, \mu)$ and $(H, \xi)$ as above, let $P: \mathcal{C}(G) \to \mathcal{C}(G/H)$ be the function defined as follow: for every $f \in \mathcal{C}(G)$, for every $s \in G$, let

$$(P(f))(sH) = \int_H f(sh) \, d\xi(h).$$
We need to check that the map $P$ is well-defined, that is, if $sH = tH$, then $(P(f))(sH) = (P(f))(tH)$, but this follows from the left-invariance of $\xi$ since

$$(P(f))(sH) = \int_H f(sh) \, d\xi(h) = \int_H f(h) \, d\xi(h) = \int_H f(th) \, d\xi(h) = (P(f))(tH).$$

Roughly speaking, $P(f)(sH)$ is obtained by averaging over $H$. The following properties are immediately verified.

**Proposition 6.37.** The function $P: \mathcal{K}_C(G) \to \mathcal{K}_C(G/H)$ satisfies the following properties:

1. The function $P(f)$ is continuous.
2. We have $\text{supp}(P(f)) \subseteq \pi(\text{supp}(f))$.
3. For any $\varphi \in \mathcal{K}_C(G/H)$, we have
   $$P((\varphi \circ \pi)f) = \varphi P(f).$$

Our next goal is to show that $P$ is surjective. The following technical result is needed.

**Proposition 6.38.** For any compact subset $F$ of $G/H$, there is a positive function $f \in \mathcal{K}_C(G)$ such that $P(f) \equiv 1$ on $F$.

**Proof.** Let $E$ be a compact neighborhood of $F$ in $G/H$. By Proposition 6.7 there is a compact subset $K$ of $G$ such that $\pi(K) = E$. Since $G$ and $G/H$ are locally compact, by Proposition 15.39 we can find a positive function $g \in \mathcal{K}_C(G)$ such that $g$ is strictly positive on $K$ and a function $\varphi \in \mathcal{K}_C(G/H)$ such that $\varphi \equiv 1$ on $F$ and $\text{supp}(\varphi) \subseteq E$. Define $f$ by

$$f(s) = \begin{cases} \frac{\varphi(\pi(s))}{P(g)(\pi(s))} g(s) & \text{if } P(g(\pi(s)) \neq 0 \\ 0 & \text{if } P(g(\pi(s)) = 0. \end{cases}$$

Since $P(g) > 0$ on $\text{supp}(\varphi)$, the function $f$ is continuous, we have $\text{supp}(f) \subseteq \text{supp}(g)$, and by Proposition 6.37(3) applied to $\left(\frac{\varphi}{P(g)} \circ \pi\right) g$, we have $P(f) = (\varphi/P(g)) P(g) = \varphi$, as desired.

Using Proposition 6.38, we obtain the surjectivity of $P$.

**Proposition 6.39.** For any function $\varphi \in \mathcal{K}_C(G/H)$, there is some function $f \in \mathcal{K}_C(G)$ such that $P(f) = \varphi$. Furthermore, $\pi(\text{supp}(f)) = \text{supp}(\varphi)$, and if $\varphi \geq 0$ then $f \geq 0$.

**Proof.** If $\varphi \in \mathcal{K}_C(G/H)$, by Proposition 6.38 there is some function $g \geq 0$ in $\mathcal{K}_C(G)$ such that $P(g) \equiv 1$ on $\text{supp}(\varphi)$. Let $f = (\varphi \circ \pi)g$. Then by Proposition 6.37(3), we have $P(f) = \varphi P(g) = \varphi$ since $P(g) \equiv 1$ on $\text{supp}(\varphi)$. The other properties are immediately verified.
If $G$ is a locally compact group and if $H$ is a closed normal subgroup of $G$, then by Proposition 6.6, the group $G/H$ is also a locally compact group. As application of the surjectivity of the map $P: \mathcal{K}_c(G) \to \mathcal{K}_c(G/H)$, the following proposition shows how to integrate on $G$ by integrating on $H$ and $G/H$.

**Proposition 6.40.** Let $G$ be a locally compact group and let $H$ be a closed normal subgroup of $G$. If $\xi$ is a left Haar measure on $H$ and if $\gamma$ is a left Haar measure on $G/H$, then the functional

$$f \mapsto \int_{G/H} P(f)(sH) \, d\gamma(sH) = \int_{G/H} \int_H f(sh) \, d\xi(h) \, d\gamma(sH), \quad f \in \mathcal{K}_c(G)$$

is a left Haar functional on $G$. Consequently, for any left Haar measure $\mu$ on $G$, by rescaling $\xi$ or $\lambda$, we have

$$\int_G f(s) \, d\mu(s) = \int_{G/H} \int_H f(sh) \, d\xi(h) \, d\gamma(sH).$$

**Proof sketch.** The verification that the given functional is a positive and left-invariant functional is left as an easy exercise. The fact that the functional is not the zero functional follows immediately from the surjectivity of $P$. By Theorem 5.8, there is a left Haar measure corresponding to this positive Radon functional, and by uniqueness of the left Haar measure up to a scalar, we can rescale $\xi$ or $\gamma$ as desired. \hfill \Box

We now come to our main theorem.

**Theorem 6.41.** Let $G$ be a locally compact group, and let $H$ be closed subgroup of $G$. Suppose $\mu$ is a left Haar measure on $G$ and $\xi$ is a left Haar measure on $H$. There is a $G$-invariant $\sigma$-Radon measure $\gamma$ on $G/H$ if and only if $\Delta_H$ is equal to the restriction of $\Delta_G$ to $H$. In this case, $\gamma$ is unique up to a scalar, and with a suitable choice of this factor, we have

$$\int_G f(s) \, d\mu(s) = \int_{G/H} P(f)(sH) \, d\gamma(sH) = \int_{G/H} \int_H f(sh) \, d\xi(h) \, d\gamma(sH), \quad (\dagger)$$

for all $f \in \mathcal{K}_c(G)$.

**Proof.** First suppose that a $G$-invariant $\sigma$-Radon measure $\gamma$ on $G/H$ exists. The map $f \mapsto \int P(f) \, d\gamma$ is a nonzero left-invariant positive linear functional on $\mathcal{K}_c(G)$, so by the uniqueness of Haar measure on $G$ there is some $c > 0$ such that

$$\int P(f)(sH) \, d\gamma(sH) = c \int f(s) \, d\mu(s). \quad (*)$$

By Radon-Riesz I (Theorem 5.8), the Haar measure $\gamma$ is uniquely determined by the functional $\varphi \mapsto \int \varphi(sH) \, d\gamma(sH)$ (with $\varphi \in \mathcal{K}_c(G/H)$). By Proposition 6.39, since $P: \mathcal{K}_c(G) \to \mathcal{K}_c(G/H)$ intergal.
\( \mathcal{K}_C(G/H) \) is surjective, (*) determines this functional completely, and thus \( \gamma \) is also completely determined, so \( \gamma \) is unique up to the scalar determining a left Haar measure. By replacing \( \gamma \) by \( c^{-1}\gamma \), equation (†) holds. Then for any \( \eta \in H \) and and \( f \in \mathcal{K}_C(G) \), we have

\[
\Delta_G(\eta) \int_G f(s)d\mu(s) = \int_G \rho_{\eta^{-1}}(f)(s)d\mu(s) \quad \text{by Proposition 6.22}
\]

\[
= \int_{G/H} \int_H \rho_{\eta^{-1}}(f)(sh)d\xi(h)d\gamma(sH) \quad \text{by (†)}
\]

\[
= \int_{G/H} \int_H f(\eta^{-1})d\xi(h)d\gamma(sH) \quad \text{by definition of } \rho_{\eta^{-1}}(f)
\]

\[
= \Delta_H(\eta) \int_{G/H} \int_H f(\eta^{-1})d\xi(h)d\gamma(sH) \quad \text{by Proposition 6.22}
\]

\[
= \Delta_H(\eta) \int_G f(s)d\mu(s), \quad \text{by (†)}
\]

which implies that \( \Delta_G(\eta) = \Delta_H(\eta) \).

Conversely, assume that \( \Delta_H \) is equal to the restriction of \( \Delta_G \) to \( H \). We claim that if \( f \in \mathcal{K}_C(G) \) and if \( P(f) = 0 \), then \( \int f(s)d\mu(s) = 0 \).

By Proposition 6.38, there is a positive function \( \varphi \in \mathcal{K}_C(G) \) such that \( P(\varphi) \equiv 1 \) on \( \pi(\text{supp}(f)) \), we have

\[
0 = (P(f))(sH) = \int_H f(s)\varphi(s)f(sH)d\xi(h) \quad \text{by definition of } P(f)
\]

\[
= \int_H f(\eta^{-1})\Delta_H(h^{-1})d\xi(h) \quad \text{by Proposition 6.27}
\]

\[
= \int_H f(\eta^{-1})\Delta_G(h^{-1})d\xi(h), \quad \text{since } \Delta_H = \Delta_G | H
\]

which implies

\[
0 = \int_{G/H} \int_H \varphi(s)f(sH)\Delta_G(h^{-1})d\xi(h)d\mu(s)
\]

\[
= \int_{G/H} \int_H \varphi(s)f(sH)\Delta_G(h^{-1})d\mu(s)d\xi(h) \quad \text{by Fubini}
\]

\[
= \int_H \Delta_G(h^{-1}) \int_G \varphi(s)f(sH)d\mu(s)d\xi(h)
\]

\[
= \int_H \int_G \varphi(s)f(sH)d\mu(s)d\xi(h) \quad \text{by Proposition 6.22}
\]

\[
= \int_{G/H} \int_H \varphi(s)f(s)d\mu(s)d\xi(h) \quad \text{by Fubini}
\]

\[
= \int_G (P(\varphi))(sH)f(s)d\mu(s) = \int f(s)d\mu(s) \quad \text{since } P(\varphi) \equiv 1 \text{ on } \pi(\text{supp}(f)).
\]
CHAPTER 6. THE HAAR MEASURE AND CONVOLUTION

What we just showed implies that if \( P(f) = P(g) \) then \( \int f(s)d\mu(s) = \int g(s)d\mu(s) \). Since by Proposition 6.39 the map \( P : \mathcal{K}_C(G) \to \mathcal{K}_C(G/H) \) is surjective, we define a functional \( \Phi \) on \( \mathcal{K}_C(G/H) \) as follows: for every \( \varphi \in \mathcal{K}_C(G/H) \), let

\[
\Phi(\varphi) = \int_G f(s)d\mu(s) \quad \text{for any } f \in \mathcal{K}_C(G) \text{ such that } P(f) = \varphi.
\]

Since \( P(f) = P(g) \) implies that \( \int f(s)d\mu(s) = \int g(s)d\mu(s) \), the functional \( \Phi \) is well-defined, and it is immediately verified that \( \Phi \) is a \( G \)-invariant positive linear functional on \( \mathcal{K}_C(G/H) \).

By Radon–Riesz I (Theorem 5.8), this functional induces the desired \( G \)-invariant \( \sigma \)-Radon measure on \( G/H \).

If \( H \) is compact then by Proposition 6.25(2), we have \( \Delta_H = \Delta_G \mid H = 1 \), so we obtain the following useful corollary.

**Proposition 6.42.** If \( G \) is a locally compact group, for any compact subgroup \( H \) of \( G \), the space \( G/H \) admits a \( G \)-invariant \( \sigma \)-Radon measure (unique up to a scalar). In fact, if \( \pi : G \to G/H \) is the quotient map, then for any left Haar measure \( \lambda \) on \( G \), there is a unique \( G \)-invariant \( \sigma \)-Radon measure \( \mu \) on \( G/H \) such that

\[
\int_{G/H} f(x) d\mu(x) = \int_G (f \circ \pi)(s) d\lambda(s), \quad \text{for all } f \in \mathcal{K}_C(G/H).
\]

**Proof.** If \( f \in \mathcal{K}_C(G/H) \) has compact support \( K \), then \( f \circ \pi \) has support homeomorphic to \( K \times H \), and since \( K \) and \( H \) are compact, it is compact. Thus \( f \circ \pi \in \mathcal{K}_C(G) \). The functional \( \Phi: \pi(f) \mapsto \int_G (f \circ \pi)(s) d\lambda(s) \) is well-defined, clearly a positive linear functional on \( \mathcal{K}_C(G/H) \), and since \( \lambda \) is left-invariant, it is \( G \)-invariant. By Radon–Riesz I, there is a unique \( G \)-invariant \( \sigma \)-Radon functional corresponding to \( \Phi \).

Proposition 6.42 applies to the projective spaces and the Grassmannians with \( G = \text{SO}(n) \) and a suitable compact subgroup \( H \). It also applies to \( G = \text{GL}(n, \mathbb{R}) \) and \( X = \text{SPD}(n) \), where \( \text{SPD}(n) \) is the set of positive symmetric definite matrices discussed in Example 17.11.

**Example 6.8.** Recall that the group \( \text{GL}(n) = \text{GL}(n, \mathbb{R}) \) acts on \( \text{SPD}(n) \) as follows: for all \( A \in \text{GL}(n) \) and all \( S \in \text{SPD}(n) \),

\[
A \cdot S = ASA^\top.
\]

This action is transitive, and in Section 17.3, Example (d), we show that the stabilizer of \( I \) is \( \text{O}(n) \), so

\[
\text{GL}(n)/\text{O}(n) = \text{SPD}(n).
\]

It can be shown that the unique (up to a scalar) \( \text{GL}(n, \mathbb{R}) \)-invariant measure on \( \text{SPD}(n) \) is given by

\[
d\mu = (\det(H))^{-(n+1)/2}d\eta(H),
\]

where \( \eta \) is the Haar measure on the additive group \( \text{S}(n) \) of real symmetric matrices. For a proof, see Bourbaki [6] (Chapter VII, Section 3, no. 3, Example 8).
6.11. CONVOLUTION

When the condition $\Delta_H = \Delta_G \mid H$ fails, it is possible relax the notion of $G$-invariance and to obtain sufficient conditions for the existence of measures on $G/H$ satisfying weaker invariance conditions. Such notions are relative invariance, quasi-invariance, and strong quasi-invariance. The interested reader should consult Folland [31] (Chapter 2, Section 2.6) and Bourbaki [6] (Chapter VII, Section 2).

6.11 Convolution

Let $G$ be a locally compact group equipped with a left Haar measure $\lambda$. Recall that $\mathcal{M}^1_{\text{reg,}C}(G)$ denotes the Banach space of complex regular Borel measures on $G$ (see Definition 5.21), and that $L^1(G, B, C)$ denotes the space of integrable functions on the measure space $(G, B, \lambda)$, where $B$ is the $\sigma$-algebra of Borel sets of $G$. To simplify notation, we write $\mathcal{M}^1(G)$ for $\mathcal{M}^1_{\text{reg,}C}(G)$, and $L^1(G)$ for $L^1(G, B, C)$. The vector space $\mathcal{M}^1(G)$ is a Banach space with the norm $\|\mu\| = |\mu|(G)$, and $L^1(G)$ is a Banach space with the $L^1$-norm. There are three flavors of convolutions but we use mainly two of them:

1. Convolutions of two measures $\mu, \nu \in \mathcal{M}^1(G)$. This makes $\mathcal{M}^1(G)$ into a Banach algebra with identity, and with an involution.

2. Convolution of two functions $f, g \in L^1(G)$, which makes $L^1(G)$ into a Banach algebra with involution, but without a multiplicative unit element, unless $G$ is discrete.

3. There is also a notion of convolution of a measure $\mu \in \mathcal{M}^1(G)$ and of a function $f \in L^1(G)$, and of a function $f \in L^1(G)$ and a measure $\mu \in \mathcal{M}^1(G)$, but we relagate these definitions to the end of this section.

Convolution applied to functions can be used as a regularization (or filtering) process. We begin with the convolution of measures.

Let $\mu, \nu \in \mathcal{M}^1(G)$ be two complex measures, then for any function $f \in \mathcal{C}_0(G; \mathbb{C})$ (recall that $\mathcal{C}_0(G; \mathbb{C})$ is the space of continuous functions that tend to zero at infinity), we have a linear functional $\Phi: \mathcal{C}_0(G; \mathbb{C}) \to \mathbb{C}$ given by

$$\Phi(f) = \iint f(st) \, d\mu(s) \, d\nu(t), \quad f \in \mathcal{C}_0(G; \mathbb{C}).$$

Observe that

$$|\Phi(f)| \leq \|f\|_\infty \|\mu\| \|\nu\|,$$

so $\Phi$ is a bounded linear functional. By Radon–Riesz III (Theorem 5.29), there is a unique measure $\mu \ast \nu \in \mathcal{M}^1(G)$ such that

$$\Phi(f) = \iint f(st) \, d\mu(s) \, d\nu(t) = \int f \, d(\mu \ast \nu), \quad f \in \mathcal{C}_0(G; \mathbb{C}),$$
with \( \|\Phi\| = \|\mu * \nu\| \). Since
\[
\|\Phi\| = \sup \{|\Phi(f)| \mid f \in C_0(G; \mathbb{C}), \|f\|_\infty = 1\}
\]
and \( |\Phi(f)| \leq \|f\|_\infty \|\mu\| \|\nu\| \), we deduce that \( \|\mu * \nu\| \leq \|\mu\| \|\nu\| \).

**Definition 6.19.** Let \( G \) be locally compact group. If \( \mu, \nu \in M^1(G) \) are two measures, then the measure \( \mu * \nu \) called convolution of \( \mu \) and \( \nu \) is the unique measure such that
\[
\int f \, d(\mu * \nu) = \int \int f(st) \, d\mu(s) \, d\nu(t) \quad \text{for all } f \in C_0(G, \mathbb{C}).
\]
We have \( \|\mu * \nu\| \leq \|\mu\| \|\nu\| \).

Observe that by interchanging \( s \) and \( t \), \( \int \int f(st) \, d\mu(s) \, d\nu(t) = \int \int f(ts) \, d\mu(t) \, d\nu(s) \), and by Fubini’s theorem, \( \int \int f(ts) \, d\mu(t) \, d\nu(s) = \int \int f(ts) \, d\nu(s) \, d\mu(t) \), so we have
\[
\int \int f(st) \, d\mu(s) \, d\nu(t) = \int \int f(ts) \, d\nu(s) \, d\mu(t).
\]

Recall the Dirac measure \( \delta_s \) given by \( \delta_s(E) = 1 \) iff \( s \in E \), and \( \delta_s(E) = 0 \) otherwise (see Example 3.6).

**Proposition 6.43.** Let \( G \) be locally compact group. The following properties hold.

1. Convolution is associative; that is, if \( \mu, \nu, \sigma \in M^1(G) \), then
\[
(\mu * \nu) * \sigma = \mu * (\nu * \sigma).
\]

2. The measure \( \delta_1 \) (where 1 is the identity element of \( G \)) is an identity for convolution; that is,
\[
\mu * \delta_1 = \delta_1 * \mu = \mu \quad \text{for all } \mu \in M^1(G).
\]

We also have
\[
\delta_s * \mu = \lambda_s(\mu), \quad \mu * \delta_s = \rho_s^{-1}(\mu).
\]

3. For all \( s, t \in G \), we have \( \delta_s * \delta_t = \delta_{st} \), and convolution is commutative \( (\mu * \nu = \nu * \mu) \) if and only if \( G \) is abelian.

Most of Proposition 6.43 is proven in Folland [31] (Chapter 2, Section 2.5), and the other parts are proven using Proposition 6.16. See also Dieudonné [22] (Chapter XIV, Section 6).

We need to define \( \tilde{\mu} \) for complex measures. We simply use Definition 6.11.

**Definition 6.20.** Let \( G \) be locally compact group. For any complex measure \( \mu \in M^1(G) \), define \( \tilde{\mu} \) by
\[
\tilde{\mu}(A) = \mu(A^{-1}), \quad \text{for all } A \in \mathcal{B}(G).
\]

We also set \( \mu^* = \overline{\mu} \) and call it the adjoint of \( \mu \).
The complex measure $\tilde{\mu}$ can be characterized by a property similar to the property of Proposition 6.27. Recall that the complex measure $\mu$ has a unique Jordan decomposition

$$\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-),$$

where the measures $\mu_1^+, \mu_1^-, \mu_2^+, \text{ and } \mu_2^-$ are positive measures; see Theorem 5.21.

**Proposition 6.44.** Let $G$ be locally compact group. For any complex measure $\mu \in \mathcal{M}^1(G)$, for every function $f \in L_1(\mu)^1(X, A, \mathbb{C})$, we have

$$\int f d\tilde{\mu} = \int f(s^{-1})d\mu(s).$$

Consequently, the complex measure $\tilde{\mu}$ is the unique measure in $\mathcal{M}^1(G)$ such that

$$\int \varphi d\tilde{\mu} = \int \varphi(s^{-1})d\mu(s) \quad \text{for all } \varphi \in C_0(G; \mathbb{C}).$$

**Proof.** We know that $\mu$ can be expressed uniquely as

$$\mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-),$$

where the measures $\mu_1^+, \mu_1^-, \mu_2^+, \text{ and } \mu_2^-$ are positive measures, and we have

$$\tilde{\mu} = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-).$$

Then we have

$$\int f d\tilde{\mu} = \int f d\tilde{\mu}_1^+ - \int f d\tilde{\mu}_1^- + i \int f d\tilde{\mu}_2^+ - i \int f d\tilde{\mu}_2^-.$$

By Proposition 6.27,

$$\int f d\tilde{\mu}_1^+ = \int f(s^{-1})d\mu_1^+, \quad \int f d\tilde{\mu}_1^- = \int f(s^{-1})d\mu_1^-,$$

so we get

$$\int f d\tilde{\mu} = \int f d\tilde{\mu}_1^+ - \int f d\tilde{\mu}_1^- + i \int f d\tilde{\mu}_2^+ - i \int f d\tilde{\mu}_2^-$$

$$= \int f(s^{-1})d\mu_1^+ - \int f(s^{-1})d\mu_1^- + i \int f(s^{-1})d\mu_2^+ - i \int f(s^{-1})d\mu_2^-$$

$$= \int f(s^{-1})d\mu,$$

as claimed. The second fact is an immediate consequence of Radon–Riesz III theorem. □
Recall from Proposition 5.23 that \( \overline{\mu} \) is the unique measure in \( M^1(G) \) satisfying the equation

\[
\int \varphi \, d\overline{\mu} = \int \overline{\varphi(s)} \, d\mu(s), \quad \text{for all } \varphi \in C_0(G; \mathbb{C}).
\]

Since for any function \( \varphi \), the function \( \check{\varphi} \) is given by \( \check{\varphi}(s) = \varphi(s^{-1}) \) for all \( s \in G \), if we define \( \varphi^* \) by \( \varphi^*(s) = \varphi(s^{-1}) = \check{\varphi}(s) \), observe that the measure \( \mu^* = \check{\mu} \) is characterized by the equation

\[
\int \varphi \, d\mu^* = \int \check{\varphi}(s) \, d\mu(s), \quad \text{for all } \varphi \in C_0(G; \mathbb{C}).
\]

**Proposition 6.45.** Let \( G \) be locally compact group. For any measures \( \mu, \nu \in M^1(G) \), we have

\[
\begin{align*}
(\mu + \nu)^* &= \mu^* + \nu^* \\
(\alpha \mu)^* &= \overline{\alpha} \mu^* \quad (\alpha \in \mathbb{C}) \\
(\mu * \nu) &= \overline{\mu} * \nu \\
(\mu * \nu)^* &= \nu^* * \mu^* \\
(\mu^*)^* &= \mu \\
\|\mu^*\| &= \|\mu\|.
\end{align*}
\]

**Proof.** We prove the second, third, and fourth equations, leaving the others as easy exercises. For every \( \varphi \in C_0(G; \mathbb{C}) \) we have

\[
\int \varphi \, d(\alpha \mu)^* = \int \overline{\varphi(s^{-1})} \alpha \, d\mu(s) \\
= \overline{\alpha} \int \overline{\varphi(s^{-1})} \, d\mu(s) \\
= \overline{\alpha} \int \varphi \, d\mu^*,
\]
which, by Radon–Riesz III, shows that $\alpha \mu^* = \overline{\alpha \mu^*}$. We also have

$$
\int \varphi \, d(\mu * \nu) = \int \overline{\varphi} \, d(\mu * \nu)
$$

$$
= \int \left( \int \overline{\varphi(st)} \, d\mu(s) \right) \, d\nu(s)
$$

$$
= \int \left( \int \varphi(st) \, d\overline{\mu}(s) \right) \, d\nu(s)
$$

$$
= \int \left( \int \varphi(st) \, d\overline{\mu}(s) \right) \, d\overline{\nu}(s)
$$

$$
= \int \varphi \, d(\overline{\mu} * \overline{\nu}),
$$

which, by Radon–Riesz III, shows that $\overline{\mu * \nu} = \overline{\mu} * \overline{\nu}$. Finally, we have

$$
\int \varphi \, d(\mu * \nu)^* = \int \varphi(s^{-1}) \, d(\mu * \nu)
$$

$$
= \int \varphi(st^{-1}) \, d\mu(s) \, d\nu(t)
$$

$$
= \int \varphi(t^{-1}s^{-1}) \, d\nu(t) \, d\mu(s)
$$

$$
= \int \left( \int \varphi(ts^{-1}) \, d\nu(t) \right) \, d\mu(s)
$$

$$
= \int \left( \int \varphi(ts) \, d\nu(t) \right) \, d\overline{\mu}(s)
$$

$$
= \int \varphi \, d(\nu * \hat{\mu}),
$$

which, by Radon–Riesz III, shows that $(\mu * \nu)^* = \nu * \hat{\mu}$. \hfill \Box

The identities of Proposition 6.45 show that $\mathcal{M}^1(G)$ is a normed algebra with involution; see Example 8.3. Furthermore, $\mathcal{M}^1(G)$ has an identity element $\delta_1$ such that $\delta_1^* = \delta_1$, and it is complete. The algebra $\mathcal{M}^1(G)$ is called the measure algebra of $G$.

**Remark:** In general, the identity

$$
\| \mu^* * \mu \| = \| \mu \|^2
$$

fails.

Observe that until now, we had no need for a Haar measure on $G$. Now, we make use of the Haar measure.
We know from Proposition 5.31 that every \( f \in L^1(G) \) can be viewed as a complex measure \( fd\lambda \) in \( M^1(G) \), where \( fd\lambda \) is the unique complex regular Borel measure such that

\[
\int fg \, d\lambda = \int g (fd\lambda) \quad \text{for all } g \in C_0(G; \mathbb{C}).
\]

Thus we can see what happens when we convolve two complex measures of the form \( fd\lambda \) and \( gd\lambda \), where \( f, g \in L^1(G) \). We need to figure out what

\[
\Phi(h) = \iint h(ts)f(t)g(s) \, d\lambda(s) \, d\lambda(t)
\]

is, for any \( h \in C_0(G; \mathbb{C}) \), where we used the fact that

\[
\iint h(st) \, d\mu(s) \, d\nu(t) = \iint h(ts) \, d\nu(s) \, d\mu(t),
\]

with \( \mu = fd\lambda \) and \( \nu = gd\lambda \). Using the left-invariance of the Haar measure (changing \( s \) to \( t^{-1}s \)) and Fubini’s theorem, we have

\[
\Phi(h) = \iint h(ts)f(t)g(s) \, d\lambda(s) \, d\lambda(t) = \int \left( \iint h(ts)f(t)g(s) \, d\lambda(s) \right) \, d\lambda(t)
\]

\[
= \int \left( \iint h(s)f(t)g(t^{-1}s) \, d\lambda(s) \right) \, d\lambda(t)
\]

\[
= \int \left( \iint h(s)f(t)g(t^{-1}s) \, d\lambda(t) \right) \, d\lambda(s)
\]

\[
= \iint \left( \int f(t)g(t^{-1}s) \, d\lambda(t) \right) h(s) \, d\lambda(s).
\]

Fubini’s theorem implies that the integral \( \int f(t)g(t^{-1}s) \, d\lambda(t) \) is defined for almost all \( s \). We also have

\[
\left\| \int f(t)g(t^{-1}s) \, d\lambda(t) \right\|_1 \leq \| f \|_1 \| g \|_1.
\]

Indeed, by Fubini and by left-invariance of the Haar measure (changing \( s \) to \( ts \)), we have

\[
\left\| \int f(t)g(t^{-1}s) \, d\lambda(t) \right\|_1 = \int \left\| \int f(t)g(t^{-1}s) \, d\lambda(t) \right\|_1 \, d\lambda(s)
\]

\[
\leq \int \iint |f(t)g(t^{-1}s)| \, d\lambda(t) \, d\lambda(s)
\]

\[
= \int |f(t)| \left( \int |g(t^{-1}s)| \, d\lambda(s) \right) \, d\lambda(t)
\]

\[
= \int |f(t)| \left( \int |g(s)| \, d\lambda(s) \right) \, d\lambda(t)
\]

\[
= \| f \|_1 \| g \|_1.
\]
Then, by Proposition 5.30, if the function $f \ast g$ is the function in $L^1(G)$ given by

$$(f \ast g)(s) = \int f(t)g(t^{-1}s) \, d\lambda(t),$$

the equation

$$\Phi(h) = \int \int h(ts)f(t)g(s) \, d\lambda(s) \, d\lambda(t) = \int \left( \int f(t)g(t^{-1}s) \, d\lambda(t) \right) h(s) \, d\lambda(s)$$

shows that $\Phi$ is a bounded linear functional, and by Theorem 5.29, there is a unique complex measure $m = (f \ast g) \, d\lambda$ such that

$$\Phi(h) = \int \left( \int f(t)g(t^{-1}s) \, d\lambda(t) \right) h(s) \, d\lambda(s) = \int (f \ast g)(s)h(s) \, d\lambda(s) = \int h(s) \, dm(s),$$

which is the convolution of the measures $fd\lambda$ and $gd\lambda$. This suggests defining the convolution of functions as follows.

**Definition 6.21.** Let $G$ be a locally compact group equipped with a left Haar measure $\lambda$. For any two functions $f, g \in L^1(G)$, the function $f \ast g$ called the *convolution* of $f$ and $g$ is the function defined for all almost all $s$ by

$$(f \ast g)(s) = \int f(t)g(t^{-1}s) \, d\lambda(t).$$

It satisfies the inequality $\|f \ast g\|_1 \leq \|f\|_1 \|g\|_1$.

The following expression for the convolution $f \ast g$ of two functions $f$ and $g$ may shed some light on what convolution does. Note that

$$(f \ast g)(s) = \int_G f(t)g(t^{-1}s) \, d\lambda_G(t) = \int_G f(st)g(t^{-1}) \, d\lambda_G(t)$$

$$= \int_G (\lambda_{s^{-1}} f)(t)\tilde{g}(t) \, d\lambda_G(t).$$

Think of $\tilde{g}$ as the function $g$ flipped about the $y$-axis, which is what happens when $G = \mathbb{R}$, since in this case $\tilde{g}(t) = g(-t)$. Also, think of $\lambda_{s^{-1}} f$ as the function $f$ shifted along the $x$-axis by the amount $s$, which is what happens when $G = \mathbb{R}$, since $(\lambda_{s^{-1}} f)(t) = f(st) = f(t + s)$. Then $\int_G (\lambda_{s^{-1}} f)(t)\tilde{g}(t) \, d\lambda_G(t)$ is the “area” around $s$ corresponding to overlapping the functions $\lambda_{s^{-1}} f$ and $\tilde{g}$ for all $t$. We can think of $f$ as some kind of filter, and $f \ast g$ is the result of filtering, or smoothing $g$, using $f$. If $f$ is tall and narrow and decays quickly near the origin, $(f \ast g)(s)$ is almost $g(s)$. This is the idea behind the Dirac delta-function. If $f$ is wider, it tends to perform a better smoothing effect.

The following result will be needed later and is easy to prove.
Proposition 6.46. The involution $\mu \mapsto \mu^*$ on $\mathcal{M}^1(G)$ yields an involution $f \mapsto f^*$ on $L^1(G)$, with
\[ f^*(s) = \Delta(s^{-1})\overline{f(s^{-1})}, \]
where $\Delta$ is the modular function of $G$. If $G$ is unimodular, then we have the simpler formula
\[ f^*(s) = \overline{f(s^{-1})}. \]

It is also easy to see that the convolution $f \ast g$ of two functions is given by the following equivalent equations:
\begin{align*}
    f \ast g(s) &= \int f(t)g(t^{-1}s)\,d\lambda(t) \\
    &= \int f(st)g(t^{-1})\,d\lambda(t) \\
    &= \int f(t^{-1})g(ts)\Delta(t^{-1})\,d\lambda(t) \\
    &= \int f(st^{-1})g(t)\Delta(t^{-1})\,d\lambda(t).
\end{align*}

We go from the first to the second equation using left-invariance by changing $t$ to $st$. We go from the first to the third equation by changing $t$ to $t^{-1}$ and using Proposition 6.27. We go from the second to the fourth equation by changing $t$ to $t^{-1}$ and using Proposition 6.27.

Folland gives the following tips to remember how to arrange the variables:

1. The variable of integration $t$ appears at $t$ in one factor and as $t^{-1}$ in the other.
2. The two occurrences of the variable of integration are adjacent to each other, not separated by the variable $s$.

When $G$ is unimodular, the factor $\Delta(t^{-1})$ disappears. If $G$ is abelian, it is customary to use an additive notation, and we have
\[ f \ast g(x) = \int f(y)g(x - y)\,d\lambda(y) = \int f(x - y)g(y)\,d\lambda(y). \]

If $G = \mathbb{R}$, then the function $x \mapsto f(x - y)$ is the function $f$ translated along the $x$-axis by the amount $y$ (and similarly the function $x \mapsto g(x - y)$ is the function $g$ translated along the $x$-axis by the amount $y$). Thus we can think of $f \ast g(x)$ as a continuous superposition of translates of $g$, or as a continuous superposition of translates of $f$. We can interpret these continuous superpositions as moving weighted averages. For example, $f \ast g(x) = \int f(y)g(x - y)\,d\lambda(y)$ is the weighted average of $f$ (on the whole line) with respect to the weight function $w(y) = g(x - y)$. In particular, if $g(x) = 0$ for all $x$ such that $|x| > a$, then
6.11. CONVOLUTION

\( g(x - y) = 0 \) for all \( x \) such that \( |x - y| > a \), and then \( f \ast g(x) \) is a weighted average of \( f \) on the interval \([x - a, x + a]\). In particular, if \( g \) is given by

\[
g(x) = \begin{cases} 
\frac{1}{2a} & \text{if } -a < x < a; \\
0 & \text{if } |x| \geq a,
\end{cases}
\]

then

\[
f \ast g(x) = \frac{1}{2a} \int_{x-a}^{x+a} f(y) \, dy.
\]

If \( a \) is very small, then \( f \ast g(x) \) is approximately \( f(x) \). This corresponds to letting \( g \) be an approximation to the Dirac function.

With its involution operation, \( L_1(G) \) is a normed Banach algebra (with involution), but generally without a multiplicative unit. It is called the \( L_1 \) group algebra of \( G \). When \( G \) is discrete or finite, \( L_1(G) \) is isomorphic to the algebra \( \mathbb{C}[G] \) consisting of all finite formal linear combinations of the form

\[
\sum_{s \in G} a_s s, \quad a_s \in \mathbb{C},
\]

where \( a_s = 0 \) for all but finitely many \( s \in G \), with the mutiplication given by

\[
 \left( \sum_{s_1 \in G} a_{s_1} s_1 \right) \left( \sum_{s_2 \in G} b_{s_2} s_2 \right) = \sum_{s \in G} \left( \sum_{t u = s} a_t b_u \right) s = \sum_{s \in S} \left( \sum_{t \in G} a_t b_{t^{-1} s} \right) s,
\]

where the last expression is the discrete convolution \( a \ast b \) of \( a = (a_{s_1})_{s_1 \in G} \) and \( b = (b_{s_2})_{s_2 \in G} \), with

\[
(a \ast b)_s = \sum_{t \in G} a_t b_{t^{-1} s}.
\]

In particular, if \( G = \mathbb{Z} \) and if \( a = \sum_{i=0}^{m} a_i i \) and \( b = \sum_{j=0}^{n} b_j j \), with no negative elements from \( \mathbb{Z} \), we have

\[
(a \ast b)_k = \sum_{i=0}^{k} a_i b_{k-i}, \quad 0 \leq k \leq m + n.
\]

We can view \( a = \sum_{i=0}^{m} a_i i \) as the polynomial \( a(X) = \sum_{i=0}^{m} a_i X^i \) and \( b = \sum_{j=0}^{n} b_j j \) as the polynomial \( b(X) = \sum_{j=0}^{n} b_j X^j \).

Convolution can be extended from \( L_1 \) to the spaces \( L^2 \) and \( L^\infty \).

**Proposition 6.47.** Let \( G \) be locally compact group equipped with a left Haar measure \( \lambda \). For any \( f \in L^1(G) \) and for any \( g \in L^p(G) \) with \( p = 1, 2, \infty \), the following facts hold:

1. The integral \( \int f(t)g(t^{-1} s) \, d\lambda(t) \) (defining \( f \ast g \)) converges absolutely for almost all \( s \), and we have \( f \ast g \in L^p(G) \) and \( \| f \ast g \|_p \leq \| f \|_1 \| g \|_p \).

2. If \( G \) is unimodular, then (1) hold with \( f \ast g \) replaced by \( g \ast f \).
(3) If \( G \) is not unimodular but \( f \) has compact support, then \( g \ast f \in L^p(G) \).

(4) When \( p = \infty \), the function \( f \ast g \) is continuous and under the conditions of (2) or (3), so is \( g \ast f \).

Proposition 6.47 is proven in Folland [31] (Chapter 2, Section 2.5).

Observe that Proposition 6.47 does not say anything when \( f \in L^2(G) \). The following proposition takes care of this case.

**Proposition 6.48.** Let \( G \) be locally compact group equipped with a left Haar measure \( \lambda \). For any \( f, g \in L^2(G) \), we have \( f \ast g \in C_0(G; \mathbb{C}) \), and

\[
\| f \ast g \|_{\infty} \leq \| f \|_2 \| g \|_2.
\]

Proposition 6.48 is proven in Folland [31] (Chapter 2, Section 2.5).

One of the applications of convolution is regularization. In order to prove that a complex measure can be approximated by complex measures with compact support we need to define the convolution of a complex measure \( \mu \) and of a function \( g \). The process of deriving a formula for \( \mu \ast g \) is very similar to the process used to derive a formula for \( f \ast g \), so we proceed quickly.

Let \( \mu \) be a complex measure in \( M^1(G) \) and let \( g \) be a function in \( L^1(G) \). For any \( h \in C_0(G; \mathbb{C}) \), we view \( g \) as the complex measure \( gd\lambda \), and we have

\[
\Phi(h) = \int \int h(ts)g(s)\,d\lambda(s)\,d\mu(t) = \int \left( \int h(ts)g(s)\,d\lambda(s) \right) \,d\mu(t)
\]

\[
= \int \left( \int h(s)g(t^{-1}s)\,d\lambda(s) \right) \,d\mu(t)
\]

\[
= \int \left( \int h(s)g(t^{-1}s)\,d\mu(t) \right) \,d\lambda(s)
\]

\[
= \int \left( \int g(t^{-1}s)\,d\mu(t) \right) h(s) \,d\lambda(s).
\]

Fubini’s theorem implies that the integral \( \int g(t^{-1}s)\,d\mu(t) \) is defined for almost all \( s \). This leads to the following definition.

**Definition 6.22.** Let \( G \) be a locally compact group equipped with a left Haar measure \( \lambda \). For any complex measure \( \mu \in M^1(G) \) and any function \( g \in L^1(G) \), the function \( \mu \ast g \) called the convolution function of \( \mu \) and \( g \) is the function defined for all almost all \( s \) by

\[
(\mu \ast g)(s) = \int g(t^{-1}s)\,d\mu(t).
\]

It satisfies the inequality \( \| \mu \ast g \|_1 \leq \| \mu \| \| g \|_1 \). The complex measure \( \mu \ast g \,d\lambda \) is the convolution of \( \mu \) and \( g \). We have

\[
\int \int h(ts)g(s)\,d\lambda(s)\,d\mu(t) = \int ((\mu \ast g)(s)) \,h(s) \,d\lambda(s) \quad \text{for all } h \in C_0(G; \mathbb{C}).
\]
The convolution of a complex measure $\mu \in \mathcal{M}^1(G)$ and any function $g \in L^2(G)$ is also defined by the same formula, and $\|\mu \ast g\|_2 \leq \|\mu\| \|g\|_2$. Observe that

$$(\delta_s \ast f)(t) = f(s^{-1}t) = (\lambda_s f)(t),$$

so

$$\delta_s \ast f = \lambda_s f.$$  

Similarly, let $f$ be a function in $L^1(G)$, and let $\mu$ be a complex measure in $\mathcal{M}^1(G)$. For any $h \in C_0(G; \mathbb{C})$, we view $f$ as the complex measure $fd\lambda$, and using Proposition 6.22 and Fubini’s theorem, we have

$$\Phi(h) = \int \int h(ts)f(t)d\lambda(t)d\mu(s) = \int \left( \int h(ts)f(t)d\lambda(t) \right) d\mu(s)$$

$$= \int \Delta(s^{-1}) \left( \int h(t)f(ts^{-1})d\lambda(t) \right) d\mu(s)$$

$$= \int \left( \int f(ts^{-1})\Delta(s^{-1)}d\mu(s) \right) h(t)d\lambda(t).$$

Consequently, the convolution $f \ast \mu$ of a function $f \in L^1(G)$ and a complex measure $\mu \in \mathcal{M}^1(G)$ is defined as follows.

**Definition 6.23.** Let $G$ be a locally compact group equipped with a left Haar measure $\lambda$. For any function $f \in L^1(G)$ and any complex measure $\mu \in \mathcal{M}^1(G)$, the function $f \ast \mu$ called the convolution function of $f$ and $\mu$ is the function defined for all almost all $s$ by

$$(f \ast \mu)(s) = \int f(st^{-1})\Delta(t^{-1})d\mu(t).$$

The complex measure $(f \ast \mu)d\lambda$ is the convolution of $f$ and $\mu$. The same definition applies if $f \in L^2(G)$. The inequality $\|f \ast \mu\|_p \leq \|\mu\| \|f\|_p$ holds for $p = 1, 2$.

Observe that

$$(f \ast \delta_s)(t) = \Delta(s^{-1})f(ts^{-1}) = \Delta(s^{-1})(\rho_{s^{-1}}f)(t),$$

so

$$f \ast \delta_s = \Delta(s^{-1})\rho_{s^{-1}}f.$$ 

### 6.12 Regularization

Given a function $g: G \to \mathbb{C}$, typically continuous, for some “well chosen” function $f$, the convolution $f \ast g$ might be more regular than $g$, for example, $f \ast g$ could become a polynomial function, or a sum of trigonometric functions, a $C^2$ functions, etc. In many cases, there exists a sequence $(f_n)$ of functions such that each $f_n \ast g$ is more “regular” than $g$, and the sequence
$(f_n * g)$ converges uniformly to $g$, as least on every compact subset of $G$. It is argued in Folland that such sequences always exist; see Folland [31] (Chapter 2, Section 2.5).

Sequences of functions $f_n$ as above can be thought of as approximations of the infamous Dirac $\delta$ function, in the sense that the sequence of integrals $\int |f_n| \, d\lambda$ is bounded, $\int f_n \, d\lambda$ tends to 1, and that for each open subset $V$ containing 1, the integral $\int_{G-V} |f_n| \, d\lambda$ tends to zero. In fact, Lang calls such sequences Dirac sequences; see Lang [52] (Chapter VIII, Section 3).

There are various formulations of the regularization theorem. Here is a version due to Dieudonné; see [22] (Chapter XIV, Section 11).

**Proposition 6.49.** Let $G$ be locally compact group (with identity element $e$) equipped with a left Haar measure $\lambda$. Let $(f_n)$ be a sequence of functions $f_n \in L^1(G)$ whose supports are contained in a fixed compact subset $K$, and which satisfy the following conditions:

1. The sequence of integrals $\int |f_n| \, d\lambda$ is bounded.
2. The sequence of integrals $\int f_n \, d\lambda$ tends to 1.
3. For each open subset $V$ containing $e$, the sequence of integrals $\int_{G-V} |f_n| \, d\lambda$ tends to zero.

Then the following properties hold:

(i) For every bounded continuous function $g$ on $G$, the sequence $(f_n * g)$ converges uniformly to $g$ on every compact subset of $G$.

(ii) If $p = 1$ or 2, and if $g \in L^p(G)$, the sequence of norms $\|f_n * g - g\|_p$ tends to 0 as $n$ goes to infinity.

**Proof.** We follow Dieudonné’s proof [22] (Chapter XIV, Section 11, Theorem 14.11.1).

(i) For every $x \in G$ and for every compact neighborhood $V$ of $e$, we have from the definitions

\[ g(x) - (f_n * g)(x) = g(x) \left( 1 - \int_V f_n(s) \, d\lambda(s) \right) \\
+ \int_V f_n(s)(g(x) - g(s^{-1}x)) \, d\lambda(s) \\
- \int_{G-V} f_n(s)g(s^{-1}x) \, d\lambda(s). \]

1The Dirac $\delta$ “function” is characterized by the following two properties: (1) $\delta(x) = 0$ for all $x \in \mathbb{R} - \{0\}$, $\delta(0) = +\infty$; (2) $\int \delta(x) \, dx = 1$. There is no such function. To make sense of it, one has to view $\delta$ as a distribution.
Let $L$ be any compact subset of $G$, and let $V_0$ be a compact neighborhood of $e$. By continuity of the group operations, $V_0^{-1}L$ is a compact, so the restriction of $g$ to $V_0^{-1}L$ is uniformly continuous. Hence for every $\epsilon > 0$, there is a compact neighborhood $V \subseteq V_0$ such that

$$ |g(x) - g(s^{-1}x)| \leq \epsilon \quad \text{for all } x \in L \text{ and all } s \in V. $$

Pick $n_0$ such that

$$ \int_{G-V} |f_n(s)| \, d\lambda(s) \leq \epsilon, \quad \left| 1 - \int_{V} f_n(s) \, d\lambda(s) \right| \leq \epsilon, $$

for all $n \geq n_0$. Then we have

$$ \left| 1 - \int_{V} f_n(s) \, d\lambda(s) \right| \leq \left| 1 - \int_{V} f_n(s) \, d\lambda(s) \right| + \int_{G-V} |f_n(s)| \, d\lambda(s) \leq 2\epsilon, $$

and therefore, since $g$ is a bounded function, for all $x \in L$,

$$ \left| g(x) \left( 1 - \int_{V} f_n(s) \, d\lambda(s) \right) \right| \leq 2\|g\| \epsilon $$

$$ \left| \int_{V} f_n(s)(g(x) - g(s^{-1}x)) \, d\lambda(s) \right| \leq c\epsilon $$

$$ \left| \int_{G-V} f_n(s)g(s^{-1}x) \, d\lambda(s) \right| \leq \|g\| \epsilon, $$

where

$$ c = \sup_n \int |f_n(s)| \, d\lambda(s). $$

Consequently,

$$ |g(x) - (f_n * g)(x)| \leq (c + 3\|g\|)\epsilon, $$

which proves the uniform convergence on the compact $L$.

(ii) By Theorem 5.13, since $\mathcal{K}_C(G)$ is dense in $\mathcal{L}^p(G)$ for $p = 1, 2$, for every $\epsilon > 0$ there is some $h \in \mathcal{K}_C(G)$ such that $\|g - h\|_p \leq \epsilon$. By Proposition 6.47(1), we have

$$ \|(f_n * g) - (f_n * h)\|_p \leq \|f_n\|_1 \|g - h\|_p \leq c\epsilon, $$

since by Property (a) the $\|f_n\|_1$ are bounded by some constant $c > 0$. Therefore we are reduced to proving (ii) for functions in $g \in \mathcal{K}_C(G)$. If $S = \text{supp}(g)$, then it is easy to see that $\text{supp}(f * g) \subseteq KS$, where $K$ is a compact set such that $\text{supp}(f_n) \subseteq K$ for all $n$, which exists by hypothesis. By (i), the sequence $(f_n * g)$ converges uniformly to $g$ on $S \cup KS$, and vanishes (as does $g$) outside this set. To conclude that $\lim_{n \to \infty} \|(f_n * g) - g\|_p = 0$, we use Proposition 4.23(2), which says that

$$ \left| \int_A f \, d\lambda \right| \leq \int_A |f| \, d\lambda \leq \|f\|_\infty \lambda(A). $$

Here $A = S \cup KS$, a compact subset, so $\lambda(A) < \infty$. Alternatively, see Dieudonné [22] (Chapter XIII, Section 12, Theorem 13.12.2.2).
A sequence \((f_n)\) of functions satisfying the conditions of Proposition 6.49 is called a regularizing sequence.

A neat application of Proposition 6.49 is a quick proof of the Weierstrass theorem on the uniform approximation of continuous functions on \([-1/2, 1/2]\) by polynomials.

**Example 6.9.** Let \(G = \mathbb{R}\), with the Lebesgue measure, and let \(g\) be a continuous function with support in \([-1/2, 1/2]\). Consider the Landau functions \(f_n\) given by

\[
f_n(x) = \begin{cases} \frac{1}{a_n} (1 - x^2)^n & \text{if } |x| \leq 1 \\ 0 & \text{if } |x| > 1, \end{cases}
\]

where \(a_n = \int_{-1}^{1} (1 - x^2)^n \, dx\).

Since \(1 - x^2 \geq 1 - |x|\) for \(1 \leq x \leq -1\), we have

\[
a_n \geq 2 \int_0^1 (1 - x)^n \, dx = \frac{2}{n + 1},
\]

which implies that

\[
f_n(x) \leq (n + 1)(1 - x^2)^n
\]

for all \(x \in [-1, +1]\). Thus \(f_n(x)\) tends to 0 uniformly on every compact interval not containing 0, which implies Property (3) of a regularizing sequence. Since by construction \(\int |f_n(x)| \, dx = \int f_n(x) \, dx = 1\), \((f_n)\) is a regularizing sequence. We have

\[
f_n * g(x) = \int_{-1/2}^{1/2} f_n(x - y)g(y) \, dy = \int_{-1/2}^{1/2} (1 - (x - y)^2)^n g(y) \, dy
\]

and by using the binomial formula we see that \((1 - (x - y)^2)^n\) is a polynomial in \(y\),

\[
(1 - (x - y)^2)^n = \sum_{j=0}^{2n} u_j(x) y^j,
\]

for some polynomials \(u_j(x)\) in \(x\), and so

\[
f_n * g(x) = \int_{-1/2}^{1/2} (1 - (x - y)^2)^n g(y) \, dy = \int_{-1/2}^{1/2} \sum_{j=0}^{2n} u_j(x) y^j g(y) \, dy = \sum_{j=0}^{2n} u_j(x) \int_{-1/2}^{1/2} y^j g(y) \, dy,
\]

a polynomial in \(x\). Proposition 6.49 shows that the sequence of polynomials \((f_n * g)\) converges uniformly to \(g\) on the compact interval \([-1/2, +1/2]\), giving another proof of the Weierstrass approximation theorem.

In this example, a continuous function \(g\), which could be much more complicated than a polynomial, and in particular, could lack derivatives of order \(\geq 1\), becomes a polynomial when convolved with \(f_n\). This is a perfect example of regularization.
Here is another example from Lang [52] (Chapter VIII, Section 3), the Cesàro summation of Fourier series of continuous functions on the unit circle \( \mathbb{T} = \mathbb{U}(1) = \{ e^{i\theta} \mid -\pi \leq \theta < \pi \} \).

Functions on the unit circle \( \mathbb{T} \) are equivalent to periodic functions on \( \mathbb{R} \) as defined next.

**Definition 6.24.** A function \( f : \mathbb{R} \to \mathbb{C} \) is *periodic with period* \( T \) (for some \( T \in \mathbb{R} \), with \( T > 0 \)), if \( f(x + T) = f(x) \) for all \( x \in \mathbb{R} \).

Obviously, a periodic function is completely defined by its restriction to the interval \( [-T, T) \). In most cases, the periods \( T = 1 \) or \( T = 2\pi \) are considered, and which is picked is a matter of taste. We pick \( T = 2\pi \). Then we have the following two transformations.

Given a periodic function \( f : \mathbb{R} \to \mathbb{C} \) (with period \( 2\pi \)), let \( f_T : \mathbb{T} \to \mathbb{C} \) be the function given by

\[
f_T(e^{i\theta}) = f(\theta), \quad -\pi \leq \theta < \pi.
\]

Given a function \( g : \mathbb{T} \to \mathbb{C} \), let \( g_R : \mathbb{R} \to \mathbb{C} \) be the periodic function (with period \( 2\pi \)) given by

\[
g_R(\theta) = g(e^{i\theta}), \quad \theta \in \mathbb{R}.
\]

Observe that because the map \( \theta \mapsto e^{i\theta} \) is a bijection between \( [-\pi, \pi) \) and \( \mathbb{T} \), we have

\[
(f_T)_R = f, \quad (g_R)_T = g,
\]

which shows that there is a bijection between the space of periodic function \( f : \mathbb{R} \to \mathbb{C} \) (with period \( 2\pi \)), and the space of functions \( g : \mathbb{T} \to \mathbb{C} \). This bijection restricts to the space of periodic \( L^p \) functions that are integrable over \( [-\pi, \pi] \), and the space \( L^p(\mathbb{T}) \), for \( p = 1, 2, \infty \).

The map \( \sigma : \mathbb{R} \to \mathbb{T} \) given by

\[
\sigma(\theta) = e^{i\theta}
\]

is clearly a surjective group homomorphism (with \( \mathbb{R} \) under addition, and \( \mathbb{T} \) under multiplication). Since \( e^{i\theta} = 1 \) iff \( \theta = k2\pi \) with \( k \in \mathbb{Z} \), we see that the kernel of \( \sigma \) is \( 2\pi \mathbb{Z} \), so by the first isomorphism theorem \( \mathbb{R}/(2\pi \mathbb{Z}) \) is isomorphic to \( \mathbb{T} \). This isomorphism allows to view a complex number of unit length as \( e^{i\theta} \), with \( \theta \) defined modulo \( 2\pi \), which is often more convenient than picking a representative of the equivalence class of \( \theta \) (mod \( 2\pi \)) in \( [-\pi, \pi] \).

The identification between \( \mathbb{R}/(2\pi \mathbb{Z}) \) and \( \mathbb{T} \), and between the space of functions defined on \( \mathbb{T} \) and the space of periodic function on \( \mathbb{R} \) is often implicit, and in what follows, we take the view that functions on \( \mathbb{T} \) are periodic (with period \( 2\pi \)). The reader should be cautioned that other authors use the period 1, so the factor \( 1/(2\pi) \) showing up in our formulae is missing in the other version (assuming period 1).

**Example 6.10.** From now on, we will use the normalized Haar measure \( dx/2\pi \) on \( \mathbb{T} \) so that \( \mathbb{T} \) has measure 1. With this normalized measure, the most important results come out cleaner (without an extra factor \( 1/2\pi \)). With this measure, the integral of \( f \in L^1(\mathbb{T}) \) is

\[
\int_{-\pi}^{\pi} f(\theta) \frac{dx(\theta)}{2\pi},
\]
and the convolution of two functions \( f, g \in L^1(\mathbb{T}) \) is
\[
(f * g)(\theta) = \int_{-\pi}^{\pi} f(\theta - \varphi)g(\varphi) \frac{d\varphi}{2\pi} = \int_{-\pi}^{\pi} f(\varphi)g(\theta - \varphi) \frac{dx(\varphi)}{2\pi}.
\]

Let \( g \) be a continuous periodic function (of period \( 2\pi \)) (equivalently, a function on \( \mathbb{T} = \mathbb{U}(1) \)). The \( n \)th partial sum \( S_{n,g} \) of the Fourier series for \( g \) is given by
\[
S_{n,g}(x) = \sum_{k=-n}^{n} c_ke^{ikx}, \quad \text{with} \quad c_k = \int_{-\pi}^{\pi} g(t)e^{-ikt} \frac{dt}{2\pi},
\]
where \( c_k \) is called the \( k \)th Fourier coefficient of \( g \). Let \( A_{n,g} \) be the average of these partial sums, that is,
\[
A_{n,g} = \frac{1}{n}(S_{0,g} + \cdots + S_{n-1,g}).
\]
The average sums \( A_{n,g} \) are known as Cesàro sums (or Cesàro means). Since \( g \) is continuous and bounded, \( g \in L^2(\mathbb{T}) \), and although the partial sums \( S_{n,g} \) converge to \( g \) in the \( \| \cdot \|_2 \)-norm (see Theorem 7.2(3)), they may not converge pointwise to \( g \); see Stein and Shakarchi [77] (Chapter 3, Subsection 2.2), and Rudin [66] (Chapters 4 and 5). On the other hand Fejér’s theorem asserts that the sequence \( (A_{n,g}) \) of average sums converges uniformly to \( g \) (see Stein and Shakarchi [77], Chapter 2, Section 5, Theorem 5.2).

This can be shown to be a consequence of Proposition 6.49 by defining the following regularizing functions \( D_n \) and \( K_n \):
\[
D_n(x) = \sum_{k=-n}^{n} e^{ikx},
\]
\[
K_n(x) = \sum_{m=0}^{n-1} \sum_{k=-m}^{m} e^{ikx} = \frac{1}{n}(D_0 + \cdots + D_{n-1}).
\]
We leave it as an exercise to prove that
\[
D_n(X) = \frac{\sin((2m + 1)x/2)}{\sin(x/2)}
\]
\[
K_n(x) = \frac{1}{n} \left( \frac{\sin(nx/2)}{\sin(x/2)} \right)^2.
\]
The functions \( D_n \) are known as Dirichlet kernels, and the functions \( K_n \) are Fejér kernels; see Stein and Shakarchi [77] (Chapter 2). It is then easy to check that \( (K_n) \) is a regularizing sequence, that \( D_n * g = S_{n,g} \), and that \( K_n * g = A_{n,g} \) (see immediately after Proposition 7.1). By Proposition 6.49, the sequence \( (A_{n,g}) = (K_n * g) \) of averages of the partial sums of the Fourier series of \( g \) converge uniformly to \( g \), which is Fejér’s theorem (see Stein and Shakarchi [77], Chapter 2, Section 5, Theorem 5.2).
Again, we have a very good example of regularization. After convolving a continuous periodic function $g$ which could be much more complicated than a sum of complex exponentials, after convolving with $D_n$, we obtain a function $D_n * g$ which is a sum of complex exponentials.

It is possible to generalize regularizing sequences to families of functions parametrized by a continuous parameter, often called kernels.

**Example 6.11.** As in the previous example, we use the normalized Haar measure $dx/2\pi$ on $\mathbb{T}$ so that $\mathbb{T}$ has measure 1. The Poisson kernel on the unit disk is the family of functions $P_r(\theta)$, parametrized by $r \in [0,1)$, and given by

$$P_r(\theta) = \sum_{n=-\infty}^{n=\infty} r^{|n|} e^{i n \theta}.$$ 

To sum this series, using the formula for the sum of a geometric series, observe that

$$\sum_{n=-\infty}^{n=\infty} r^{|n|} e^{i n \theta} = \sum_{n=0}^{n=\infty} (re^{i\theta})^n + \sum_{p=1}^{p=\infty} (re^{-i\theta})^p$$ 

$$= \frac{1}{1 - re^{i\theta}} + \frac{re^{-i\theta}}{1 - re^{-i\theta}}$$ 

$$= \frac{1 - r^2}{1 - 2r \cos \theta + r^2}.$$ 

Thus

$$P_r(\theta) = \frac{1 - r^2}{1 - 2r \cos \theta + r^2}.$$ 

Instead of being a sequence of functions indexed by natural numbers, the family $(P_r)$ is a family of functions indexed by the continuous parameter $r \in [0,1)$, but it possesses properties analogous to the properties of regularizing functions, and Proposition 6.49 can be adapted to show that for a bounded periodic continuous function $g$ on $\mathbb{R}$, the functions $(P_r * g)(\theta)$ converge to $g$ uniformly as $r$ tends to 1.

The functions $P_r$ are harmonic for the Laplacian given in polar coordinates by

$$\Delta u = \frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2},$$

which means that

$$\Delta P_r = 0.$$ 

If we write $u(r, \theta) = (P_r * g)(\theta)$, then it can be shown that

$$\Delta u(r, \theta) = 0,$$
that is, \( u(r, \theta) \) is harmonic. As a consequence of Proposition 6.49 (suitably generalized), when \( r \) tends to 1, we have \( u(1, \theta) = g(\theta) \). This shows that \( u(r, \theta) \) is a harmonic function solution of a boundary value problem, namely that

\[
\Delta u(r, \theta) = 0,
\]

with \( u(1, \theta) = g(\theta) \) on the boundary, for a prescribed periodic function \( g \).

Other kernels exist for solving partial differential equations, such as the heat equation for the Laplace operator; see Section 7.6. We refer the interested reader to Lang [52] (Chapter VIII, Section 3), Stein and Shakarchi [77], and Folland [30], for more on this topic.

Regularization can also be used to prove various approximation results involving complex measures. Because we are dealing with locally compact spaces that may not be metrizable, we need the general machinery of filters to define convergence; see Section 1.5. We begin with the following general result from Bourbaki [6] (Chapter VIII, Section 2, No. 7, Lemma 4).

**Proposition 6.50.** Let \( X \) be a locally compact space, \( a \in X \) a given point, \( M \) a subset of \( \mathcal{M}^1(X, \mathcal{A}) \), and \( \mathcal{F} \) a filter on \( M \). Suppose the following properties hold:

1. For every compact subset \( K \) of \( X \), the set of numbers \( \{ |\mu|(K) \mid \mu \in M \} \) is bounded.
2. For every compact subset \( K \) of \( X - \{a\} \), we have \( \lim_{\mu, \mathcal{F}} |\mu|(K) = 0 \).
3. There is some compact neighborhood \( V \) of \( a \) such that \( \lim_{\mu, \mathcal{F}} |\mu|(V) = 1 \).

Then the filter \( \mathcal{F} \) converges to the Dirac measure \( \delta_a \) in \( \mathcal{M}^1(X, \mathcal{A}) \) (in the norm topology).

**Corollary 6.51.** Let \( X \) be a locally compact space, \( a \in X \) a given point, \( M \) a subset of \( \mathcal{M}^1(X, \mathcal{A}) \), and \( \mathcal{F} \) a filter on \( M \). Suppose the properties of Proposition 6.50 hold and that there is a compact subset \( K_0 \) of \( X \) such that the complex measures in \( M \) have support in \( K_0 \). Then the filter \( \mathcal{F} \) converges to the Dirac measure \( \delta_a \) in the space \( \mathcal{M}^1_c(X, \mathcal{A}) \) of complex measures with compact support (in the norm topology).

Proposition 6.50 can be used to prove the following regularization result for complex measures from Bourbaki [6] (Chapter VIII, Section 4, No. 7, Proposition 19).

**Proposition 6.52.** Let \( G \) be a locally compact group equipped with a left Haar measure \( \lambda \). Let \( \mathcal{B} \) be a filter basis of neighborhoods of the identity \( e \), consisting of compact neighborhoods, and for every \( V \in \mathcal{B} \), let \( f_V \) be a positive continuous functions with compact support contained in \( V \), such that \( \int f_V \, d\lambda = 1 \). For any complex regular Borel measure \( \mu \in \mathcal{M}^1(G) \), we have the family of measures \( (\mu * f_V) \, d\lambda \), which has the structure of a filter base on \( \mathcal{M}^1(G) \), by considering the subsets \( S_V = \{ (\mu * f_W) \, d\lambda \mid W \in \mathcal{B}, W \subseteq V \} \) (corresponding to the filter of sections of \( \mathcal{B} \)), and the filter base of subsets \( S_V \) converges to \( \mu \) (in the norm topology).
Since the measures \((\mu * f_V) d\lambda\) have compact support, Proposition 6.52 shows how the complex regular Borel measure \(\mu\) can be approximated (in a suitable sense) by measures with compact support (and continuous density). Proposition 6.52 also has the following useful corollary from Bourbaki [6] (Chapter VIII, Section 4, No. 7, Corollary of Proposition 19). A slightly different version of this proposition is given in Folland [31] (Chapter 2, Proposition 2.42).

**Proposition 6.53.** Let \(G\) be a locally compact group equipped with a left Haar measure \(\lambda\). Let \(B\) be a filter basis of neighborhoods of the identity \(e\), consisting of compact neighborhoods, and for every \(V \in B\), let \(f_V\) be a positive continuous functions with compact support contained in \(V\), such that \(\int f_V d\lambda = 1\). For any function \(g \in L^p(G), p = 1, 2\), we have the family of functions \((g * f_V)\), which has the structure of a filter base on \(L^p(G)\), by considering the subsets \(S_V = \{(g * f_W) \mid W \in B, W \subseteq V\}\) (corresponding to the filter of sections of \(B\)), and the filter base of subsets \(S_V\) converges to \(g\) (in the norm topology \(\|\cdot\|_p\)).

In particular, Proposition 6.53 implies that if \(\mu\) and \(\nu\) are two complex regular Borel measures, the identity of Definition 6.19 characterizing the convolution \(\mu * \nu\) of \(\mu\) and \(\nu\),

\[
\int f d(\mu * \nu) = \int \int f(st) d\mu(s) d\nu(t)
\]

holds not only for all functions \(f \in C_0(G, \mathbb{C})\), but also for all bounded continuous functions \(f \in C_b(G; \mathbb{C})\). This fact will be needed in Chapter 9.

Families of functions \(f_V\) as in Propositions 6.52 and 6.52 are easily constructed using continuous bump functions (since \(G\) is locally compact; see Proposition 15.39).
Chapter 7

The Fourier Transform and the Fourier Cotransform on $\mathbb{T}^n$, $\mathbb{Z}^n$, $\mathbb{R}^n$

Historically, trigonometric series were first used by D’Alembert (1747) to solve the equation of a vibrating string, elaborated by Euler a year later, and then solved in a different way essentially using Fourier series by D. Bernoulli (1753). However it was Fourier who introduced and developed Fourier series in order to solve the heat equation, in a sequence of works on heat diffusion, starting in 1807, and culminating with his famous book, *Théorie analytique de la chaleur*, published in 1822.

Originally, the theory of Fourier series is meant to deal with periodic functions on the circle $\mathbb{T}$, say functions with period $2\pi$. Remarkably (but we must apologize for the oversimplification), the theory of Fourier series is captured by the following two equations:

$$f(\theta) = \sum_{m \in \mathbb{Z}} c_m e^{im\theta}. \quad (1)$$

$$c_m = \int_{-\pi}^{\pi} f(\theta) e^{-im\theta} \, d\theta. \quad (2)$$

Equation (1) involves a series, and Equation (2) involves an integral. There are two ways of interpreting these equations.

The first way consists of starting with a convergent series as given by the right-hand side of (1) (of course $c_n \in \mathbb{C}$), and to ask what kind of function is obtained. A second question is the following: Are the coefficients in (1) computable in terms of the formulae given by (2)?

Such questions were considered by Riemann, and then Cantor and Lebesgue. Since they deal with the notion of integral, it is not surprising that they motivated the invention of the Riemann integral, and then the Lebesgue integral.

The second way is to start with a periodic function $f$, apply Equation (2) to obtain the $c_m$, called *Fourier coefficients*, and then to consider Equation (1). Does the series $\sum_{m \in \mathbb{Z}} c_m e^{im\theta}$ (called *Fourier series*) converge at all? Does it converge to $f$?
Observe that the expression \( f(\theta) = \sum_{m \in \mathbb{Z}} c_m e^{im\theta} \) may be interpreted as a countably infinite superposition of elementary periodic functions, intuitively representing simple wave functions, the functions \( \theta \mapsto e^{im\theta} \). We can think of \( m \) as the frequency of this wave function.

The above questions were first considered by Fourier. Fourier boldly claimed that every function can be represented by a Fourier series. Of course, this is false, and for several reasons. First, one need to define what is an integrable function, and there are plenty of nonintegrable functions. Second, it depends on the kind of convergence that we are dealing with. The \( n \)th partial sum \( S_n,f \) of the Fourier series \( \sum_{m \in \mathbb{Z}} c_m e^{im\theta} \) for \( f \) (where the \( c_m \) are given by Equation (2)) is given by

\[
S_{n,f}(\theta) = \sum_{k=-n}^{n} c_k e^{ik\theta}.
\]

The most common type of convergence is pointwise convergence, which means that for every \( \theta \), we have \( \lim_{n \to \infty} |f(\theta) - S_{n,f}(\theta)| = 0 \). Even if \( f \) is a continuous function, there are examples of Fourier series that do not converge pointwise for \( \theta = 0 \) (du Bois-Reymond). There is even a function in \( L^1(\mathbb{T}) \) whose Fourier series diverges for all \( \theta \) (Kolmogoroff). The convergence of Fourier series is a subtle matter.

But Fourier was almost right. If we consider a function \( f \) in \( L^2(\mathbb{T}) \), a famous and deep theorem of Carleson states that its Fourier series converges to \( f \) almost everywhere in the \( L^2 \)-norm. Other ways to ensure the convergence of the Fourier series of a function is to either restrict the class of functions being considered (Dirichlet, Jordan), or to use different kinds of summation (Abel, Cesàro). Abel summation leads to the Poisson kernel, and Cesàro summation leads to the Féjér kernel; see Example 6.10, Section 7.1, and Stein and Shakarchi [77] (Chapter 2).

In Section 7.1, as a motivation for Fourier analysis on \( \mathbb{T} \), we solve the wave equation for a vibrating string. We are led immediately to the problem of Fourier inversion.

Given a periodic function \( f \), the problem of determining when \( f \) can be reconstructed as the Fourier series (Equation (1)) given by its Fourier coefficients \( c_m \) (Equation (2)) is called the problem of Fourier inversion. To discuss this problem, it is useful to adopt a more general point of view of the correspondence between functions and Fourier coefficients, and Fourier coefficients and Fourier series.

Given a function \( f \in L^1(\mathbb{T}) \), Equation (2) yields the \( \mathbb{Z} \)-indexed sequence \( (c_m)_{m \in \mathbb{Z}} \) of Fourier coefficients of \( f \), with

\[
c_m = \int_{-\pi}^{\pi} f(\theta) e^{-im\theta} \frac{d\theta}{2\pi},
\]

which we call the Fourier transform of \( f \), and denote by \( \hat{f} \), or \( \mathcal{F}(f) \). We can view the Fourier transform \( \mathcal{F}(f) \) of \( f \) as a function \( \mathcal{F}(f) : \mathbb{Z} \to \mathbb{C} \) with domain \( \mathbb{Z} \).
On the other hand, given a \( \mathbb{Z} \)-indexed sequence \( c = (c_m)_{m \in \mathbb{Z}} \) of complex numbers \( c_m \), we can define the Fourier series \( \mathcal{F}(c) \) associated with \( c \), or Fourier cotransform of \( c \), given by

\[
\mathcal{F}(c)(\theta) = \sum_{m \in \mathbb{Z}} c_m e^{im\theta}.
\]

This time, \( \mathcal{F}(c) \) is a function \( \mathcal{F}(c) : \mathbb{T} \to \mathbb{C} \) with domain \( \mathbb{T} \). Fourier inversion can be stated as the equation

\[
f(\theta) = ((\mathcal{F} \circ \mathcal{F})(f))(\theta).
\]

Of course, there is an issue of convergence. Namely, in general, \( \hat{f} = \mathcal{F}(f) \) does not belong to \( L^1(\mathbb{T}) \). There are special cases for which Fourier inversion holds. One case is if \( \mathcal{F}(f) \in l^1(\mathbb{Z}) \), which means that the sum \( \sum_{m \in \mathbb{Z}} |c_m| \) is finite. Another case is if \( f \in L^2(\mathbb{T}) \). In fact, Plancherel’s theorem asserts that the map \( f \mapsto \hat{f} \) is an isometric isomorphism between \( L^2(\mathbb{T}) \) and \( l^2(\mathbb{Z}) \).

In Section 7.2, we generalize the results of Section 7.1 to \( \mathbb{T}^n \) and \( \mathbb{Z}^n \). In addition to the definition of the Fourier transform on \( \mathbb{T}^n \), we define the Fourier cotransform on \( \mathbb{T}^n \), and in addition to the definition of the cotransform on \( \mathbb{Z}^n \), we define the Fourier transform on \( \mathbb{Z}^n \). We also generalize the Poisson kernel to \( \mathbb{T}^n \), and prove generalizations of the results of Section 7.1 on spectral synthesis and Abel summation. Plancherel’s theorem asserts that the map \( f \mapsto \hat{f} \) is an isometric isomorphism between \( L^2(\mathbb{T}^n) \) and \( l^2(\mathbb{Z}^n) \). The inverse is the Fourier cotransform.

In Section 7.3, we discuss the Fourier transform of functions defined on the entire real line \( \mathbb{R} \) that are not necessarily periodic. Because \( \mathbb{R} \) is not compact, \( L^1(\mathbb{R}) \) and \( L^2(\mathbb{R}) \) are incomparable (with respect to inclusion), and the theory of the Fourier transform on \( \mathbb{R} \) is more delicate than the Fourier theory on \( \mathbb{T} \).

In Section 7.4, we consider a classical problem in signal processing, which is to reconstruct a function \( f : \mathbb{R} \to \mathbb{C} \) which is band-limited, which means that its Fourier transform \( \hat{f} \) vanishes outside some interval \([−\Omega, \Omega]\). Then \( f \) can be completely reconstructed by sampling at the points \( t_n = n\pi/\Omega \), for \( n \in \mathbb{N} \). We obtain the sampling theorem (Theorem 7.19).

The results of Section 7.3 are generalized to \( \mathbb{R}^n \) in Section 7.5.

In Section 7.6 we define a class \( \mathcal{S}(\mathbb{R}^n) \) of smooth functions that decay quickly when \( ||x|| \) goes to infinity called the Schwartz space. The space \( \mathcal{S}(\mathbb{R}^n) \) is not a normed vector space, but its topology can be defined by a countable family of semi-norms. It is a metrizable space that is complete, called a Fréchet space. The Schwartz space is closed under the Fourier transform and cotransform, and is generally well-behaved. Fourier inversion holds, and taking the Fourier transform of a derivative is just multiplication of the Fourier transform by a variable.

This last property can be exploited to solve certain partial differential equations by converting them to ordinary differential equations by applying the Fourier transform. We illustrate this method by solving the steady-state heat equation in the upper half-plane.
In Section 7.7, we discuss the Poisson summation formula, which is a way of finding the Fourier coefficients of the periodic function obtained from a nonperiodic function by applying the process of periodization.

In Section 7.8, we return to the issue of pointwise convergence of Fourier series on \( \mathbb{T} \). We give examples of functions for which the Fourier series does not converge pointwise, or worse. We show that for the class of functions of bounded variation, there is a pointwise convergence theorem due to Dirichlet and Jordan.

In section 7.9, we show that roughly, a function \( f \) and its Fourier transform \( \hat{f} \) can’t be both highly localized. This can be stated precisely in terms of the dispersion of \( f \) about the point \( a \) given by

\[
\Delta_a f = \int (x-a)^2 |f(x)|^2 \, dx \int |f(x)|^2 \, dx.
\]

The Heisenberg inequality states that if \( f \) is a function in \( L^2(\mathbb{R}) \), then for all \( a, b \in \mathbb{R} \), we have

\[
(\Delta_a f)(\Delta_b \hat{f}) \geq \frac{1}{4}.
\]

We briefly discuss the interpretation of this inequality in quantum mechanics, called the Heisenberg uncertainty principle.

In the last section, Section 7.10, we give a brief summary of Fourier’s captivating life.

## 7.1 Fourier Analysis on \( \mathbb{T} \)

We begin this chapter with a preview of Fourier analysis on one of the simplest abelian locally compact groups, namely \( \mathbb{T} \). As in Example 6.10, we continue to use the normalized Haar measure \( dx/2\pi \) on \( \mathbb{T} \) so that \( \mathbb{T} \) has measure 1.

The solution of the wave equation for a vibrating string provides an excellent motivation for using Fourier series on \( \mathbb{T} \).

Consider a homogeneous string in the \((x,y)\)-plane, stretched along the \( x \)-axis between \( x = 0 \) and \( x = \pi \). The constant \( \pi \) is chosen for mathematical convenience; we could use any constant \( L > 0 \), but by a change of units, we may assume that it is equal to \( \pi \). If the string is set to vibrate, its displacement \( u(x,t) \) is then a function of \( x \) and \( t \). We assume that its endpoints are fixed, so that we have the initial conditions

\[
u(0,t) = u(\pi,t) = 0 \quad \text{for all } t.
\]

We also assume that the initial position and velocity of the string are given by two functions \( f \) and \( g \) defined on \([0, \pi]\) (with \( f(0) = f(\pi) = 0 \)), so that

\[
u(x,0) = f(x), \quad \frac{\partial u}{\partial t}(x,0) = g(x).
\]
We extend the functions \( f \) and \( g \) to \([-\pi, \pi]\) by making them odd, namely, we set \( f(-x) = -f(x) \) and \( g(-x) = -g(x) \) for \( x \in [0, \pi] \), and then we extend \( f \) and \( g \) to \( \mathbb{R} \) by making them periodic of period \( 2\pi \) (so, \( f(x + 2\pi k) = f(x) \) and \( g(x + 2\pi k) = g(x) \), for all \( k \in \mathbb{Z} \) and \( x \in [-\pi, \pi] \)).

Using some physics, it can be shown that \( u \) is a solution of the one-dimensional wave-equation,

\[
\frac{\partial^2 u}{\partial x^2} = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2},
\]

for some constant \( c \). Again, by a change of units, we may assume that \( c = 1 \), so the wave equation becomes

\[
\frac{\partial^2 u}{\partial x^2} = \frac{\partial^2 u}{\partial t^2}.
\]

Equation (*) can be solved by two methods:

1. Using traveling waves.
2. Using standing waves.

The method of traveling waves was used by d’Alembert, and the method of standing waves by D. Bernoulli; see Stein and Shakarchi [77] (Chapter 1).

The method of standing waves leads immediately to Fourier series. In this method, we use the technique of separation of variables, which means that we express the solution \( u(x, t) \) as the product \( u(x, t) = \varphi(x)\psi(t) \), where \( \varphi(x) \) and \( \psi(t) \) are functions of the two independent variables \( x \) and \( t \). Equation (*) yields the equation

\[
\varphi(x)\psi''(t) = \varphi''(x)\psi(t),
\]

which can be written as

\[
\frac{\psi''(t)}{\psi(t)} = \frac{\varphi''(x)}{\varphi(x)}.
\]

Since the left-hand side depends only on \( t \), and the right-hand side depends only on \( x \), the above equation can hold only if both sides are equal to the same constant, say \( \lambda \), so we deduce that

\[
\varphi''(x) - \lambda \varphi(x) = 0
\]

\[
\psi''(t) - \lambda \psi(t) = 0.
\]

These equations have well-known solutions. If \( \lambda \geq 0 \), we obtain a solution which is not physically possible since the displacement of the string is unbounded, so we must have \( \lambda < 0 \), say \( \lambda = -m^2 \). The solution is given by

\[
\varphi(x) = C \cos mx + D \sin mx
\]
\[
\psi(t) = A \cos mt + B \sin mt.
\]
Since $\varphi(0) = \varphi(\pi) = 0$, we get $C = 0$, and if $D \neq 0$, then $m$ must be an integer. If $m = 0$, then $\varphi(x) = 0$ for all $x$, and if $m \leq -1$, we can rename the constants and reduce to the case $m \geq 1$ (since cos is even and sin is odd). Finally, we arrive at the solution

$$u_m(x, t) = (A_m \cos mt + B_m \sin mt) \sin mx.$$  \hfill (**) 

Since the wave equation is linear, any linear combination of the functions in (**) is also a solution, so we are led to the fact that a general solution $u(x, t)$ of the wave equation is a superposition of the solutions $u_m$, that is,

$$u(x, t) = \sum_{m=1}^{\infty} (A_m \cos mt + B_m \sin mt) \sin mx.$$ 

There is obviously an issue of convergence, but we will not worry about this yet. The last step is to impose the boundary conditions

$$u(x, 0) = f(x), \quad \frac{\partial u}{\partial t}(x, 0) = g(x).$$

which yield the equations

$$f(x) = \sum_{m=1}^{\infty} A_m \sin mx,$$

$$g(x) = \sum_{m=1}^{\infty} mB_m \sin mx.$$ 

Thus we arrived at the following question: Given a “reasonable” periodic function $f: \mathbb{T} \to \mathbb{C}$ (say $f \in L^1(\mathbb{T})$), can we find some coefficients $c_m \in \mathbb{C}$ such that

$$f(\theta) = \sum_{m \in \mathbb{Z}} c_m e^{im\theta},$$

where the series on the right-hand side is the Fourier series associated with $c_m$?

This is the basic problem that motivated Fourier in his quest for solving the heat equation on various domains.

The integer $m \geq 1$ is the frequency of the wave component $(A_m \cos mt + B_m \sin mt) \sin mx$, which is called a harmonic or tone. The general solution is thus a superposition of harmonics. The case $m = 1$ corresponds to the first harmonic or fundamental tone. If the vibrating strings is the string of a violin, then the principal harmonic is the sound of lowest pitch.

If $f \in L^1(\mathbb{T})$, then we can compute the Fourier coefficients $c_m$ by the formula

$$c_m = \int_{-\pi}^{\pi} f(t) e^{-imt} \frac{dx(t)}{2\pi},$$
and then the question is whether the Fourier series \( \sum_{m \in \mathbb{Z}} c_m e^{im\theta} \) converges to \( f \) (and in what sense).

Recall that the \( n \)th partial sum \( S_{n,f} \) of the Fourier series \( \sum_{m \in \mathbb{Z}} c_m e^{im\theta} \) for \( f \) is given by

\[
S_{n,f}(\theta) = \sum_{k=-n}^{n} c_k e^{ik\theta},
\]

and the average \( A_{n,f} \) of these partial sums is given by

\[
A_{n,f} = \frac{1}{n} (S_{0,f} + \cdots + S_{n-1,f}).
\]

It would be desirable that the partial sums \( S_{n,f} \) converge pointwise to \( f \), but in general, this is not the case, even for continuous functions. We will see that if \( f \in L^2(\mathbb{T}) \), then \( S_{n,f} \) converge to \( f \) in the \( L^2 \)-sense (Proposition 7.2), but the convergence may fail to be pointwise.

In Example 6.10, we discussed Cesàro sums and Féjer’s theorem. We showed that the average sums \( A_{n,f} \) converge uniformly to \( f \), if \( f \) is continuous. We now discuss Abel’s sums and Poisson kernels, which yield another kind of convergence.

Recall from Definition 6.11 that the Poisson kernel on the unit disk is the family of functions \( P_r(\theta) \), parametrized by \( r \in [0,1) \), and given by

\[
P_r(\theta) = \sum_{n=-\infty}^{n=\infty} r^{|n|} e^{in\theta} = \frac{1 - r^2}{1 - 2r \cos \theta + r^2}.
\]

We have the following result using the Poisson kernel which gives a preview of Fourier analysis.

**Proposition 7.1.** For any \( r \in [0,1) \), if \( f \in L^1(\mathbb{T}) \) and if \( P_r \) is the Poisson kernel, then for all \( \theta \in \mathbb{R}/2\pi\mathbb{Z} \), we have

\[
(P_r * f)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m r^{|m|} e^{im\theta},
\]

where \( c_m \) is the \( m \)th Fourier coefficient of \( f \),

\[
c_m = \int_{-\pi}^{\pi} f(t)e^{-imt} \frac{dx(t)}{2\pi}.
\]
Proof. For $0 \leq r < 1$, the series defining $P_r$ is absolutely convergent, so

$$(P_r * f)(\theta) = \int_{-\pi}^{\pi} P_r(\theta - \varphi) f(\varphi) \frac{dx(\varphi)}{2\pi}$$

$$= \int_{-\pi}^{\pi} \sum_{m=-\infty}^{m=\infty} r^{|m|} e^{im(\theta - \varphi)} f(\varphi) \frac{dx(\varphi)}{2\pi}$$

$$= \sum_{m=-\infty}^{m=\infty} \int_{-\pi}^{\pi} r^{|m|} e^{im(\theta - \varphi)} f(\varphi) \frac{dx(\varphi)}{2\pi}$$

$$= \sum_{m=-\infty}^{m=\infty} r^{|m|} e^{im\theta} \int_{-\pi}^{\pi} f(\varphi) e^{-im\varphi} \frac{dx(\varphi)}{2\pi}$$

$$= \sum_{m=-\infty}^{m=\infty} c_m r^{|m|} e^{im\theta},$$

as claimed. \qed

Recall that the functions $D_n$ and $K_n$ are defined as

$$D_n(x) = \sum_{k=-n}^{n} e^{ikx}$$

$$K_n(x) = \sum_{m=0}^{n-1} \sum_{k=-m}^{m} e^{ikx} = \frac{1}{n} (D_0 + \cdots + D_{n-1}).$$

It can be shown that

$$D_n(X) = \frac{\sin((2m + 1)x/2)}{\sin(x/2)}$$

$$K_n(x) = \frac{1}{n} \left( \frac{\sin(nx/2)}{\sin(x/2)} \right)^2.$$

The functions $D_n$ are known as Dirichlet kernels, and the functions $K_n$ are Fejér kernels.

Observe that for $r = 1$, the partial sum $\sum_{m=-n}^{m=n} c_m r^{|m|} e^{im\theta}$ is the partial sum $S_{n,f}$ of the Fourier series for $f$, and the partial sum $\sum_{m=-n}^{m=n} r^{|m|} e^{im\theta}$ of the Poisson kernel is the Dirichlet kernel $D_n$. A slight modification of the proof of Proposition 7.1 shows that

$$D_n * f = S_{n,f},$$

and this immediately implies that

$$K_n * f = A_{n,f}.$$
Recall that for any $p \geq 1$, the space $l^p(\mathbb{Z})$ is the set of sequences $x = (x_n)_{n \in \mathbb{Z}}$ with $x_n \in \mathbb{C}$ such that $\sum_{n \in \mathbb{Z}} |x_n|^p < \infty$. Also, if $1 \leq p < q$, then $l^p(\mathbb{Z}) \subseteq l^q(\mathbb{Z})$.

Indeed, since the sequence $|x_m|^p$ converges, for some $M > 0$ we have $|x_m| < 1$ for all $|m| \geq M$, and since if $q > p$ we have $|x_m|^q \leq |x_m|^p$ (because $|x_m|^p - |x_m|^q = |x_m|^p(1 - |x_m|^{q-p}) \geq 0$ since $|x_m| < 1$), thus $\sum_{|m| \geq M} |x_m|^q \leq \sum_{|m| \geq M} |x_m|^p$, and since $\sum_{n \in \mathbb{Z}} |x_n|^p < \infty$, we also have $\sum_{n \in \mathbb{Z}} |x_n|^q < \infty$.

In general, given a function $f \in L^1(\mathbb{T})$, the Fourier series $\sum_{m \in \mathbb{Z}} c_m e^{im\theta}$ does not converge pointwise. However, if $0 \leq r < 1$, then the series $f_r(\theta) = (P_r * f)(\theta) = \sum_{m \in \mathbb{Z}} c_m r^{|m|} e^{im\theta}$ converges absolutely and uniformly. The following results shows that if $r$ tend to 1, then $f_r$ is an approximation of $f$ that tends to $f$ (in a technical sense). Since $\mathbb{T}$ is compact, we have $L^2(\mathbb{T}) \subseteq L^1(\mathbb{T})$. Then if $f \in L^2(\mathbb{T})$, the partial sums of the Fourier series of $f$ converge to $f$ in the $L^2$-norm.

**Theorem 7.2.** *(Spectral synthesis)*

1. If $f \in L^p(\mathbb{T})$ for $p = 1, 2$, and if $r \in [0, 1)$, for all $\theta \in \mathbb{R}/2\pi\mathbb{Z}$, write

$$f_r(\theta) = (P_r * f)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m r^{|m|} e^{im\theta},$$

with

$$c_m = \int_{-\pi}^{\pi} f(t)e^{-imt} \frac{dt}{2\pi}.$$

Then $\lim_{r \to 1} \|f - f_r\|_p = 0$.

2. If $f \in C(\mathbb{T})$, then $\lim_{r \to 1} \|f - f_r\|_\infty = 0$.

3. If $f \in L^2(\mathbb{T})$, then

$$\lim_{n \to \infty} \left\| f - \sum_{m=-n}^{m=n} c_m e^{im\theta} \right\|_2 = 0.$$

Furthermore, we have the Parseval theorem:

$$\|f\|_2^2 = \sum_{m=-\infty}^{m=\infty} |c_m|^2.$$

The above implies that $c = (c_m)_{m \in \mathbb{Z}} \in l^2(\mathbb{Z})$.

Theorem 7.2 is proven in Malliavin [55] (Chapter 3, Section 2.2.5). The function

$$f_r(\theta) = (P_r * f)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m r^{|m|} e^{im\theta}$$
is known as the \( r \)th Abel mean of the Fourier series
\[
\sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}
\]
of \( f \). The Fourier series does not always converge pointwise, but the \( r \)th Abel mean \( f_r \) converges uniformly for all \( r < 1 \) \((r \geq 0)\).

The results of Theorem 7.2 are examples of spectral synthesis, namely, the reconstruction of a function from its Fourier coefficients. Facts (1) and (2) are not very practical because they require first summing the series \( f_r \). Fact (2) for continuous functions is better because it shows uniform convergence. Fact (3) is very satisfactory since it shows convergence of the partial sums of the Fourier series in the \( L^2 \)-sense, but convergence pointwise generally fails.

Remark: Lennart Carleson showed in 1966 that for any function \( f \in L^2(\mathbb{T}) \), the partial sums of the Fourier series of \( f \) converge pointwise almost everywhere to \( f \), putting a close to a problem that had been open for fifty years.

The following result shows that if the sequence \( c = (c_m)_{m \in \mathbb{Z}} \) of Fourier coefficients of \( f \) is well-behaved, then \( f \) can be reconstructed from \( c \).

**Theorem 7.3.** \((\text{Fourier inversion formula})\) Let \( f \in L^1(\mathbb{T}) \). If \( c = (c_m)_{m \in \mathbb{Z}} \in l^1(\mathbb{Z}) \), that is, if the series \( \sum_{m=-\infty}^{m=\infty} |c_m| \) converges, where \( c_m \) is the Fourier coefficient
\[
c_m = \int_{-\pi}^{\pi} f(t) e^{-imt} \frac{dx(t)}{2\pi},
\]
then
\[
f(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}
\]
for all almost all \( \theta \in \mathbb{R}/2\pi\mathbb{Z} \). Furthermore, if \( f \) is continuous, then equality holds everywhere.

**Proof.** Write
\[
\varphi = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}
\]
and recall that
\[
f_r(\theta) = \sum_{m=-\infty}^{m=\infty} c_m r^{|m|} e^{im\theta}.
\]
Since the series \( \sum_{m=-\infty}^{m=\infty} |c_m| \) converges, the series defining \( \varphi \) converges absolutely, so \( \varphi \) is continuous. We claim that
\[
\lim_{r \to 1} \| \varphi - f_r \|_\infty = 0.
\]
We have
\[ \|\varphi - f_r\|_\infty \leq \sum_{m=-\infty}^{m=\infty} |c_m|(1 - r^{|m|}). \]

Given \( \epsilon > 0 \), we can find \( p \) so that \( \sum_{|m|>p} |c_m| \leq \epsilon/2 \). Then \( \sum_{|m|\leq p} |c_m|(1 - r^{|m|}) \) is the sum of \( 2p + 1 \) terms that tend to 0 as \( r \) tends to 1, so for \( r \) close enough to 1 so that \( \sum_{|m|>p} |c_m|(1 - r^{|m|}) < \epsilon/2 \), we have
\[ \sum_{|m|>p} |c_m|(1 - r^{|m|}) + \sum_{|m|\leq p} |c_m|(1 - r^{|m|}) < \epsilon/2 + \epsilon/2 = \epsilon, \]
which shows that \( \lim_{r \to 1} \|\varphi - f_r\|_\infty = 0 \).

Since \( \|\varphi - f_r\|_1 \leq 2\pi \|\varphi - f_r\|_\infty \), we also have
\[ \lim_{r \to 1} \|\varphi - f_r\|_1 = 0. \]

By Theorem 7.2(1),
\[ \lim_{r \to 1} \|f - f_r\|_1 = 0, \]
and since
\[ \|f - \varphi\|_1 \leq \|f - f_r\|_1 + \|f_r - \varphi\|_1, \]
we deduce that
\[ \|f - \varphi\|_1 = 0, \]
which means that \( f = \varphi \) almost everywhere. If \( f \) is continuous, since \( \varphi \) is also continuous \( f - \varphi = h \) is continuous. But if \( h \neq 0 \), then \( h \) is nonzero on some interval, which contradicts the fact that \( f = \varphi \) almost everywhere.

**Definition 7.1.** Given any function \( f \in L^1(\mathbb{T}) \), the function \( \mathcal{F}(f) : \mathbb{Z} \to \mathbb{C} \) given by \( \mathcal{F}(f)(m) = c_m \), where \( c_m \) is the Fourier coefficient
\[ c_m = \int_{-\pi}^{\pi} f(t)e^{-imt} \frac{dx(t)}{2\pi}, \]
is called the Fourier transform of \( f \). We identify the sequence \( \mathcal{F}(f) \) with the sequence \( (c_m)_{m \in \mathbb{Z}} \), which is also denoted by \( \widehat{f} \).

Theorem 7.2(3) (Parseval’s theorem) implies that if \( f \in L^2(\mathbb{T}) \), then \( \widehat{f} \in l^2(\mathbb{Z}) \). However, if \( f \in L^1(\mathbb{T}) \), then it may not be the case that \( \widehat{f} \in l^1(\mathbb{Z}) \).

Theorem 7.3 says that if \( f \in L^1(\mathbb{T}) \) and if \( \widehat{f} \in l^1(\mathbb{Z}) \), then \( f \) can be reconstructed by its Fourier series \( \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta} \). Theorem 7.2(3) says that if \( f \in L^2(\mathbb{T}) \) then the partial sums of the Fourier series of \( f \) converge to \( f \) in the \( L^2 \)-norm. In fact, we have a stronger result.
Theorem 7.4. (Plancherel) The map $f \mapsto \hat{f}$ is an isometric isomorphism of the Hilbert spaces $L^2(\mathbb{T})$ and $l^2(\mathbb{Z})$.

Theorem 7.4 is classical theorem of Hilbert theory. Its proof can be found in Rudin [66] (Chapter 4), or Malliavin [55] (Chapter 3, Section 2.2.5).

Definition 7.2. Given a sequence $c = (c_m)_{m \in \mathbb{Z}} \in l^1(\mathbb{Z})$, we define the Fourier cotransform $\mathcal{F}(c)$ of $c$ as the function $\mathcal{F}(c) : \mathbb{T} \to \mathbb{C}$ defined on $\mathbb{T}$ given by

$$\mathcal{F}(c)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta} = \sum_{m=-\infty}^{m=\infty} c_m (e^{i\theta})^m,$$

the Fourier series associated with $c$ (with $\theta \in \mathbb{R}/2\pi \mathbb{Z}$). Given a function $f \in L^1(\mathbb{T})$, if $\hat{f}$ is the Fourier transform of $f$, then the Fourier cotransform $\mathcal{F}(\hat{f}) = \sum_{m=-\infty}^{m=\infty} \hat{f}_m e^{im\theta}$ of $\hat{f}$ is called the Fourier series of $f$.

Note that $e^{im\theta}$ is used instead of the term $e^{-im\theta}$ occurring in the Fourier transform.

Since $c \in l^1(\mathbb{Z}) \subseteq l^2(\mathbb{Z})$, the series $\mathcal{F}(c)(\theta)$ converges uniformly. Theorem 7.3 and Theorem 7.4 assert that $\mathcal{F}$ and $\overline{\mathcal{F}}$ are mutual inverses between $L^2(\mathbb{T})$ and $l^2(\mathbb{Z})$. However, in general, for an arbitrary function $f \in L^1(\mathbb{T})$, there is no guarantee that $\hat{f} \in l^1(\mathbb{Z})$, so the Fourier series $\mathcal{F}(\hat{f}) = \sum_{m=-\infty}^{m=\infty} \hat{f}_m e^{im\theta}$ may not converge to $f$ in the $L^1$ sense or pointwise.

Remark: The maps $e^{i\theta} \mapsto e^{im\theta} = (e^{i\theta})^m$, for $m \in \mathbb{Z}$ and $\theta \in \mathbb{R}/(2\pi \mathbb{Z})$, are continuous homomorphisms of the group $\mathbb{T} = \mathbb{U}(1)$ into itself. In fact, it can be shown that they are the only ones of this kind. They are called the characters of $\mathbb{T}$. Obviously the set of characters of $\mathbb{T}$ is in bijection with $\mathbb{Z}$. Thus the Fourier transform $\mathcal{F}(f)$ of a function $f \in L^2(\mathbb{T})$, a sequence of complex numbers indexed by $\mathbb{Z}$, can be viewed as a function of the characters of $\mathbb{T}$.

The characters of $\mathbb{Z}$ are the group homomorphisms $\varphi : \mathbb{Z} \to \mathbb{T}$. Since $\mathbb{Z}$ is generated by 1, a homomorphism satisfies the equation

$$\varphi(m) = (\varphi(1))^m, \quad m \in \mathbb{Z},$$

so it is uniquely determined by picking $\varphi(1) = e^{i\theta} \in \mathbb{T}$ (with $\theta \in \mathbb{R}/(2\pi \mathbb{Z})$), and is of the form $\varphi(m) = (e^{i\theta})^m = e^{im\theta}$ for all $m \in \mathbb{Z}$. Thus the set of characters of $\mathbb{Z}$ is in bijection with $\mathbb{T}$. Then the Fourier cotransform $\mathcal{F}(c)$ of a “function” $c \in l^2(\mathbb{Z})$ ($\mathcal{F}(c)$ is the Fourier series associated with $c$) can also be viewed as a function on the characters of $\mathbb{Z}$, namely a function on $\mathbb{T}$. This fact generalizes to an arbitrary abelian locally compact group, and is the key to the definition of the Fourier transform on such a group.

Sometimes it is more convenient to express the Fourier series $\mathcal{F}(c)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta}$ in terms of $\cos m\theta$ and $\sin m\theta$ instead of the complex exponentials $e^{im\theta}$. Since

$$e^{im\theta} = \cos m\theta + i \sin m\theta,$$
the Fourier series $F(c)(\theta)$ can be expressed as

$$
F(c)(\theta) = \sum_{m=-\infty}^{\infty} c_m e^{im\theta}
$$

$$
= \sum_{m=-\infty}^{-1} c_m (\cos m\theta + i \sin m\theta) + \sum_{m=0}^{\infty} c_m (\cos m\theta + i \sin m\theta)
$$

$$
= c_0 + \sum_{m=1}^{\infty} c_m (\cos m\theta - i \sin m\theta) + \sum_{m=1}^{\infty} c_m (\cos m\theta + i \sin m\theta)
$$

$$
= c_0 + \sum_{m=1}^{\infty} (c_m + c_{-m}) \cos m\theta + i(c_m - c_{-m}) \sin m\theta).
$$

Therefore, if we let

$$
a_0 = 2c_0, \quad a_m = c_m + c_{-m}, \quad b_m = i(c_m - c_{-m}), \quad m \geq 1,
$$

then we have

$$
F(c)(\theta) = \frac{1}{2} a_0 + \sum_{m=1}^{\infty} (a_m \cos m\theta + b_m \sin m\theta).
$$

Equation (†) makes it very clear that the function $F(c)(\theta)$ can be viewed as the countably infinite superposition of the basic periodic functions $\cos m\theta$ and $\sin m\theta$, often called harmonics. The number $m \in \mathbb{N} - \{0\}$ is called a frequency.

Conversely, if

$$
F(c)(\theta) = \frac{1}{2} a_0 + \sum_{m=1}^{\infty} (a_m \cos m\theta + b_m \sin m\theta),
$$

then if we let

$$
c_0 = \frac{1}{2} a_0, \quad c_m = \frac{1}{2} (a_m - ib_m), \quad c_{-m} = \frac{1}{2} (a_m + ib_m), \quad m \geq 1,
$$

then

$$
F(c)(\theta) = \sum_{m=-\infty}^{\infty} c_m e^{im\theta}.
$$

From

$$
c_m = \int_{-\pi}^{\pi} f(t) e^{-imt} \frac{dx(t)}{2\pi},
$$

and

$$
a_0 = 2c_0, \quad a_m = c_m + c_{-m}, \quad b_m = i(c_m - c_{-m}), \quad m \geq 1,
$$
for \( m \geq 1 \), we get
\[
a_m = c_m + c_{-m} = \int_{-\pi}^{\pi} f(t)(e^{-imt} + e^{imt}) \frac{dx(t)}{2\pi} = 2 \int_{-\pi}^{\pi} f(t) \cos mt \frac{dx(t)}{2\pi}
\]
and
\[
a_b = i(c_m - c_{-m}) = \int_{-\pi}^{\pi} f(t)i(e^{-imt} - e^{imt}) \frac{dx(t)}{2\pi} = 2 \int_{-\pi}^{\pi} f(t) \sin mt \frac{dx(t)}{2\pi},
\]
that is, for \( m \geq 1 \), we have
\[
a_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos mt \, dt
\]
\[
b_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin mt \, dt.
\]
We also have
\[
a_0 = 2c_0 = 2 \int_{-\pi}^{\pi} f(t) \frac{dx(t)}{2\pi} = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \, dt.
\]
Therefore we can combine the above equations and we obtain
\[
a_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos mt \, dt \quad (m \geq 0)
\]
\[
b_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin mt \, dt \quad (m \geq 1).
\]
Note that the equation for \( a_m \) also holds for \( m = 0 \). This is the reason for the term \((1/2)a_0\) in equation (†). The numbers \( a_m \) and \( b_m \) are also called the Fourier coefficients of \( f \). If the function \( f \) is real-valued, then the coefficients \( a_m \) and \( b_m \) are real.

Observe that
\[
c_0 = \frac{1}{2} a_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \, dt
\]
is the mean value of \( f \) over the interval \([-\pi, \pi]\).

Here are a few examples of Fourier transforms. Many more examples can be found in Folland [30].

**Example 7.1.** Let \( f: [-\pi, \pi] \to \mathbb{R} \) be the periodic function given by
\[
f(\theta) = |\theta|, \quad -\pi \leq \theta \leq \pi.
\]
The graph of \( f(\theta) \) is shown in Figure 7.1.

Let us compute the coefficients \( a_m \) and \( b_m \). Since the function \( f \) is even, we have \( b_m = 0 \) for all \( m \geq 1 \), and
\[
a_m = \frac{1}{\pi} \int_{-\pi}^{\pi} f(\theta) \cos m\theta \, d\theta = \frac{2}{\pi} \int_0^{\pi} f(\theta) \cos m\theta \, d\theta.
\]
Thus for \( m = 0 \) we have
\[
a_0 = \frac{2}{\pi} \int_{0}^{\pi} \theta \, d\theta = \frac{1}{\pi} \left[ \theta^2 \right]_0^\pi = \pi,
\]
and for \( m \geq 1 \), integrating by parts we have
\[
a_m = \frac{2}{\pi} \left[ \theta \sin \frac{m\pi}{m} \right]_0^{\pi} - \frac{2}{\pi} \int_0^{\pi} \frac{\sin m\theta}{m} \, d\theta
= \frac{2}{\pi} \left[ \frac{\cos m\theta}{m^2} \right]_0^{\pi} = \frac{2}{\pi} \frac{(-1)^m - 1}{m^2}
\]
since \( \sin m\pi = 0 \) and \( \cos m\pi = (-1)^m \). Now \( (-1)^m - 1 = -2 \) when \( m \) is odd and 0 when \( m \) is even, so we find that the Fourier series for \( f \) is given by
\[
\frac{\pi}{2} - \frac{4}{\pi} \sum_{m \text{ odd}} \frac{1}{m^2} \cos m\theta = \frac{\pi}{2} - \frac{4}{\pi} \sum_{k=1}^{\infty} \frac{\cos(2k - 1)\theta}{(2k - 1)^2}.
\]

If we plot the graphs of the partial sums for a few terms (say five terms), we see that they provide a very good approximation to \( f \). See Figures 7.2 and 7.3. The series converges uniformly to \( f \) due to the presence of the term \( 1/(2k - 1)^2 \).

**Example 7.2.** Let \( f : [-\pi, \pi] \to \mathbb{R} \) be the periodic function given by
\[
f(\theta) = \theta, \quad -\pi < \theta \leq \pi.
\]
The graph of \( f(\theta) \) is shown in Figure 7.4.

This time, let us compute the coefficients \( c_m \). We have
\[
c_0 = \frac{1}{2\pi} \int_{-\pi}^{\pi} \theta \, d\theta = 0,
\]
Figure 7.2: Let $S_M = \frac{\pi}{2} - \frac{4}{\pi} \sum_{k=1}^{M} \frac{\cos(2k-1)\theta}{(2k-1)^2}$. Figure a. is the graph of $S_1$; Figure b. is the graph of $S_2$; Figure c. is the graph of $S_3$, and Figure d. is the graph of $S_4$.

and for $m \neq 0$, by integrating by parts, we have

$$c_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} \theta e^{-im\theta} d\theta = \frac{1}{2\pi} \left[ \theta e^{-im\theta} \right]_{-\pi}^{\pi} - \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-im\theta} d\theta$$

$$= \frac{1}{2\pi} \left[ e^{-im\theta} \left( \frac{\theta}{-im} + \frac{1}{m^2} \right) \right]_{-\pi}^{\pi}$$

$$= \frac{(-1)^{m+1}}{im}$$

since $e^{-im\pi} = e^{im\pi} = (-1)^m$. Hence the Fourier series for $f$ is

$$\sum_{m \neq 0} \frac{(-1)^{m+1}}{im} e^{im\theta}.$$

Since $(-1)^m = (-1)^{-m}$, the $m$th and the $(-m)$th term can be combined to give

$$(-1)^{m+1} \left( \frac{e^{im\theta}}{im} + \frac{e^{-im\theta}}{-im} \right) = \frac{2(-1)^{m+1}}{m} \sin m\theta,$$
and we obtain the Fourier series

\[ 2 \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \sin m\theta. \]

This time, if we plot the graphs of the partial sums, we see that they approximate the function \( f \), but the quality of the approximation is inferior to that of Example 7.1. See Figures 7.5 and 7.6.

This is due to the fact that the function of Example 7.1 is continuous, but the function of Example 7.2 has jump discontinuities. The other reason why the quality of approximation is not as good as in Example 7.1 is that the terms of the series in Example 7.1 tend to zero faster than the terms of the series in Example 7.2. Thus, in Example 7.2, the influence of the higher order terms is much more significant in Example 7.2. The point is that the rougher the function is, the more difficult it is to approximate it by smooth functions such as \( \cos m\theta \) and \( \sin m\theta \). In fact, it is not obvious that the series

\[ 2 \sum_{m=1}^{\infty} \frac{(-1)^{m+1}}{m} \sin m\theta \]
Figure 7.5: Let $S_M = 2 \sum_{m=1}^{M} (-1)^{m+1} \sin m\theta$. Figure a. is the graph of $S_3$ Figure b. is the graph of $S_5$; Figure c. is the graph of $S_{14}$, and Figure d. is the graph of $S_{40}$.

converges. It does, according to a theorem of Dirichlet (see Section 7.8).

A phenomenon that shows up in Example 7.2 is the Gibbs phenomenon. Even for a partial sum of 40 terms, we observe some spikes near the discontinuities. These spikes tend to zero in width, but not in height; see Folland [30] (Chapter 2, Section 2.6).

7.2 The Fourier Transform and the Fourier Cotransform on $\mathbb{T}^n$ and $\mathbb{Z}^n$

In Section 7.1 we introduced the Fourier transform on $\mathbb{T}$, and the Fourier cotransform on $\mathbb{Z}$. In this section we briefly present the generalization to $\mathbb{T}^n = \mathbb{T} \times \cdots \times \mathbb{T}$, called the n-torus, and to $\mathbb{Z}^n$. As in Section 7.1, we use the normalized Haar measure $dx/2\pi$ on $\mathbb{T}$ (with $dx$ the Lebesgue measure on $\mathbb{R}$), so that $\mathbb{T}$ has measure 1.

Recall that given any function $f \in L^1(\mathbb{T})$, the function $\mathcal{F}(f): \mathbb{Z} \rightarrow \mathbb{C}$ given by $\mathcal{F}(f)(m) = c_m$, where $c_m$ is the Fourier coefficient

$$c_m = \int_{-\pi}^{\pi} f(t) e^{-imt} \frac{dx(t)}{2\pi},$$

is called the Fourier transform of $f$. We identify the sequence $\mathcal{F}(f)$ with the sequence $(c_m)_{m \in \mathbb{Z}}$, which is also denoted by $\hat{f}$. 

$\text{Figure 7.5}$
7.2. THE FOURIER TRANSFORM AND COTRANSFORM ON $\mathbb{T}^n$ AND $\mathbb{Z}^n$}

Recall that for any $p \geq 1$, the space $l^p(\mathbb{Z})$ is the set of sequences $x = (x_n)_{n \in \mathbb{Z}}$ with $x_n \in \mathbb{C}$ such that $\sum_{n \in \mathbb{Z}} |x_n|^p < \infty$.

Given a sequence $c = (c_m)_{m \in \mathbb{Z}} \in l^1(\mathbb{Z})$, we define the Fourier cotransform $\mathcal{F}(c)$ of $c$ as the function $\mathcal{F}(c): \mathbb{T} \to \mathbb{C}$ defined on $\mathbb{T}$ given by

$$\mathcal{F}(c)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{im\theta},$$

the Fourier series associated with $c$ (with $\theta \in \mathbb{R}/2\pi\mathbb{Z}$). Note that $e^{im\theta}$ is used instead of the term $e^{-im\theta}$ occurring in the Fourier transform.

For symmetry reasons, it seems natural to define a Fourier cotransform on $\mathbb{T}$ and a Fourier transform on $\mathbb{Z}$. The Fourier cotransform $\mathcal{F}(f)$ of a function $f \in L^1(\mathbb{T})$ is the $\mathbb{Z}$-indexed sequence $\mathcal{F}(f): \mathbb{Z} \to \mathbb{C}$ given by

$$\mathcal{F}(f)(m) = \int_{-\pi}^{\pi} f(t) e^{imt} \frac{dt}{2\pi},$$

and the Fourier transform $\mathcal{F}(c)$ of a sequence $c = (c_m)_{m \in \mathbb{Z}} \in l^1(\mathbb{Z})$ is the function $\mathcal{F}(c): \mathbb{T} \to \mathbb{C}$ given by

$$\mathcal{F}(c)(\theta) = \sum_{m=-\infty}^{m=\infty} c_m e^{-im\theta},$$

with $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. Observe that if $f \in L^1(\mathbb{T})$, then

$$\mathcal{F}(f)(m) = \mathcal{F}(f)(-m) = \overline{\mathcal{F}(f)(m)},$$
with \( m \in \mathbb{Z} \), and if \( c \in l^1(\mathbb{Z}) \), then
\[
\mathcal{F}(c)(\theta) = \mathcal{F}(c)(-\theta) = \overline{\mathcal{F}(c)(\theta)}.
\]
where \( \theta \in \mathbb{R}/2\pi \mathbb{Z} \). Thus only one of the two transforms is really needed, but it is convenient to use both (especially in stating Fourier inversion).

**Remark:** Note a certain asymmetry in the measure chosen on \( T \) and \( Z \). The measure on \( T \) is \( dx/2\pi \), so that \( T \) has measure 1, and the measure on \( Z \) is the counting measure.

The main results are:

1. The spectral synthesis, Theorem 7.2.
2. The Fourier inversion formula, Theorem 7.3. This result can be expressed as follows. If \( f \in L^1(T) \) and if \( \widehat{f} = \mathcal{F}(f) \in l^1(Z) \), then
\[
f(\theta) = (\mathcal{F} \circ \mathcal{F})(f)(\theta) = \sum_{m \in Z} \widehat{f}_m e^{im\theta}.
\]
3. Plancherel’s theorem, Theorem 7.4. This theorem asserts that \( \mathcal{F} \) and \( \overline{\mathcal{F}} \) are inverse isometric isomorphisms between the Hilbert spaces \( L^2(T) \) and \( l^2(Z) \).

All three results stated above generalize to \( T^n \) and \( Z^n \). First we need a bit of notation.

**Definition 7.3.** A *multi-index* is a sequence \( \alpha = (\alpha_1, \ldots, \alpha_n) \) of natural numbers \( \alpha_i \in \mathbb{N} \). Define \( |\alpha|, \alpha!, \partial^\alpha, \text{ and } x^\alpha \) by
\[
|\alpha| = \sum_{i=1}^n \alpha_i, \quad \alpha! = \alpha_1! \times \cdots \times \alpha_n!, \quad \partial^\alpha = \left(\frac{\partial}{\partial x_1}\right)^{\alpha_1} \cdots \left(\frac{\partial}{\partial x_n}\right)^{\alpha_n}, \quad x^\alpha = \prod_{i=1}^n x_i^{\alpha_i}.
\]

We now generalize the Poisson kernel and the Fourier transform (and cotransform) to \( T^n \) and \( Z^n \).

Observe that a function \( z: Z^n \to \mathbb{C} \) can be viewed as a \( Z^n \)-indexed sequence \( z = (z_m)_{m \in Z^n} \), with \( z_m \in \mathbb{C} \). Let \( l^p(Z^n) \) \((p \geq 1)\) be the Hilbert space
\[
l^p(Z^n) = \left\{ z = (z_m)_{m \in Z^n}, \ z_m \in \mathbb{C} \mid \sum_{m \in Z^n} |z_m|^p < \infty \right\}.
\]
As in the case \( n = 1 \), if \( 1 \leq p < q \), then \( l^p(Z^n) \subseteq l^q(Z^n) \).

We denote the product measure on \( T^n \) by \( dx_n/(2\pi)^n = (1/(2\pi)^n) \, dx \otimes \cdots \otimes dx \), where \( dx \) is the Lebesgue measure on \( \mathbb{R} \). With this measure, \( T^n \) has measure 1.
Definition 7.4. The Poisson kernel $P_r(\theta)$ on $\mathbb{T}^n$ (with $\theta = (\theta_1, \ldots, \theta_n) \in \mathbb{R}^n/2\pi\mathbb{Z}^n$) is the family of functions $P_r(\theta)$, parametrized by $r \in [0, 1)$, given by

$$P_r(\theta) = \prod_{k=1}^{n} P_r(\theta_k),$$

with

$$P_r(\theta_k) = \sum_{n=-\infty}^{\infty} r^{|n|} e^{inxk}.$$

Definition 7.5. For any function $f \in \mathcal{L}^1(\mathbb{T}^n)$, the Fourier transform $\hat{f} = \mathcal{F}(f)$ of $f$ is the $\mathbb{Z}^n$-indexed sequence $\mathcal{F}(f) : \mathbb{Z}^n \to \mathbb{C}$ given by

$$\hat{f}(m) = \mathcal{F}(f)(m) = \int_{\mathbb{T}^n} f(\theta) e^{-im \cdot \theta} \frac{dx_n(\theta)}{(2\pi)^n},$$

and the Fourier cotransform $\overline{\mathcal{F}}(f)$ of $f$ is the $\mathbb{Z}^n$-indexed sequence $\overline{\mathcal{F}}(f) : \mathbb{Z}^n \to \mathbb{C}$ given by

$$\overline{\mathcal{F}} f(m) = \int_{\mathbb{T}^n} f(\theta) e^{im \cdot \theta} \frac{dx_n(\theta)}{(2\pi)^n},$$

with $m \in \mathbb{Z}^n$, $\theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n$, and with $m \cdot \theta = \sum_{k=1}^{n} m_k \theta_k$, the inner product of the vectors $m = (m_1, \ldots, m_n)$ and $\theta = (\theta_1, \ldots, \theta_n)$.

For any $c \in l^1(\mathbb{Z}^n)$, the Fourier transform $\mathcal{F}(c)$ of $c$ is the function $\mathcal{F}(c) : \mathbb{T}^n \to \mathbb{C}$ defined on $\mathbb{T}^n$ given by

$$\mathcal{F}(c)(\theta) = \sum_{m \in \mathbb{Z}^n} c_m e^{-im \cdot \theta},$$

and the Fourier cotransform $\overline{\mathcal{F}}(c)$ of $c$ is the function $\overline{\mathcal{F}}(c) : \mathbb{T}^n \to \mathbb{C}$ defined on $\mathbb{T}^n$ given by

$$\overline{\mathcal{F}}(c)(\theta) = \sum_{m \in \mathbb{Z}^n} c_m e^{im \cdot \theta},$$

with $\theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n$.

Remark: The Fourier cotransform is also called the inverse Fourier transform by some authors, including Hewitt and Ross.

It can be shown that $|\hat{f}(m)|$ tends to zero when $|m|$ tends to infinity. This is a special case of Proposition 9.17.

One of the main reasons why the Fourier transform is useful is that it converts a convolution into a product. Remember that the convolution of two functions $f, g \in \mathcal{L}^1(\mathbb{T}^n)$ is

$$(f * g)(\theta) = \int_{\mathbb{T}^n} f(\theta - \varphi) g(\varphi) \frac{dx_n(\varphi)}{(2\pi)^n} = \int_{\mathbb{T}^n} f(\varphi) g(\theta - \varphi) \frac{dx_n(\varphi)}{(2\pi)^n}.$$
Also, the inner product of two functions \( f, g \in \mathbb{L}^2(\mathbb{T}^n) \) is given by
\[
\langle f, g \rangle = \int_{\mathbb{T}^n} f(\theta) \overline{g(\theta)} \frac{dx_n(\theta)}{(2\pi)^n},
\]
and the inner product of two sequences \( x, y \in \mathbb{l}^2(\mathbb{Z}^n) \) is given by
\[
\langle x, y \rangle = \sum_{m \in \mathbb{Z}^n} x_m \overline{y_m}.
\]

**Proposition 7.5.** For any two functions \( f, g \in \mathbb{L}^1(\mathbb{T}^n) \), we have
\[
\mathcal{F}(f * g) = \mathcal{F}(f) \mathcal{F}(g), \quad \mathcal{F}(f * g) = \mathcal{F}(f) \overline{\mathcal{F}(g)}.
\]
The equation \( \mathcal{F}(f * g) = \mathcal{F}(f) \mathcal{F}(g) \) can also be written as \( \widehat{f * g} = \widehat{f} \overline{\widehat{g}} \).

Proposition 7.5 actually holds in the more general framework of locally compact abelian groups, and a proof is given in Proposition 9.4 (see also Proposition 9.17).

It is not hard to adapt the proof of Proposition 7.1 to prove that for any \( f \in \mathbb{L}^1(\mathbb{T}^n) \), for all \( \theta \in \mathbb{R}^n / 2\pi \mathbb{Z}^n \), we have
\[
(f * P_r)(\theta) = \sum_{m \in \mathbb{Z}^n} \hat{f}(m) r^{\|m\|_1} e^{im \cdot \theta},
\]
where \( \|m\|_1 = |m_1| + \cdots + |m_n| \). As a consequence we have the following result.

**Theorem 7.6.** (Spectral synthesis)

(1) If \( f \in \mathbb{L}^p(\mathbb{T}^n) \) for \( p = 1, 2 \) and if \( r \in [0, 1) \), for any \( \theta \in \mathbb{R}^n / 2\pi \mathbb{Z}^n \), write
\[
f_r(\theta) = (P_r * f)(\theta) = \sum_{m \in \mathbb{Z}^n} \hat{f}(m) r^{\|m\|_1} e^{im \cdot \theta}.
\]

Then \( \lim_{r \to 1} \|f - f_r\|_p = 0 \).

(2) If \( f \in \mathbb{C}(\mathbb{T}^n) \), then \( \lim_{r \to 1} \|f - f_r\|_\infty = 0 \).

For any \( p \in \mathbb{N} \), let
\[
S_p = \{ m \in \mathbb{Z}^n \mid |m_k| \leq p, \ k = 1, \ldots, n \}.
\]

**Theorem 7.7.** Let \( f \in \mathbb{L}^2(\mathbb{T}^n) \). Then we have the equality (Parseval)
\[
\|f\|_2^2 = \sum_{m \in \mathbb{Z}^n} |\hat{f}(m)|^2.
\]
Define \( s_p(\theta) \) by
\[
s_p(\theta) = \sum_{m \in S_p} \hat{f}(m) e^{im \cdot \theta}.
\]
Then we have
\[
\lim_{p \to \infty} \|f - s_p\|_2 = 0.
\]
7.3. THE FOURIER TRANSFORM AND THE FOURIER COTRANSFORM ON $\mathbb{R}$

Plancherel’s theorem holds.

**Theorem 7.8.** *(Plancherel)* The map $f \mapsto \hat{f}$ is an isometric isomorphism of the Hilbert spaces $L^2(\mathbb{T}^n)$ and $l^2(\mathbb{Z}^n)$. The inverse is the Fourier cotransform.

The Fourier inversion formula is generalized as follows.

**Theorem 7.9.** *(Fourier inversion formula)* Let $f \in L^1(\mathbb{T}^n)$. If $\hat{f} \in l^1(\mathbb{Z}^n)$ then

$$f(\theta) = \sum_{m \in \mathbb{Z}^n} \hat{f}_m e^{im \cdot \theta} = (\mathcal{F}(\hat{f}))(\theta),$$

for all almost all $\theta \in \mathbb{R}^n/2\pi \mathbb{Z}^n$. Furthermore, if $f$ is continuous, then equality holds everywhere.

We now turn to the Fourier transform on $\mathbb{R}$.

### 7.3 The Fourier Transform and the Fourier Cotransform on $\mathbb{R}$

In this section we discuss the Fourier transform of functions defined on the entire real line $\mathbb{R}$ that are not necessarily periodic. Because $\mathbb{R}$ is not compact, $L^1(\mathbb{R})$ and $L^2(\mathbb{R})$ are incomparable (with respect to inclusion), and the theory of the Fourier transform on $\mathbb{R}$ is more delicate than the Fourier theory on $\mathbb{T}$. In particular, although Plancherel’s theorem holds (Theorem 7.8), its proof is more complicated.

**Definition 7.6.** For any function $f \in L^1(\mathbb{R})$, the Fourier transform $\hat{f} = \mathcal{F}(f)$ of $f$ is the function $\mathcal{F}(f) : \mathbb{R} \to \mathbb{C}$ defined on $\mathbb{R}$ given by

$$\hat{f}(x) = \mathcal{F}(f)(x) = \int_{\mathbb{R}} f(y) e^{-iyx} \frac{dy}{\sqrt{2\pi}},$$

and the Fourier cotransform $\overline{F}(f)$ of $f$ is the function $\overline{F}(f) : \mathbb{R} \to \mathbb{C}$ defined on $\mathbb{R}$ given by

$$\overline{F}f(x) = \int_{\mathbb{R}} f(y) e^{iyx} \frac{dy}{\sqrt{2\pi}},$$

where $dx$ is the Lebesgue measure on $\mathbb{R}$.

**Remark:** The Fourier cotransform is also called the *inverse Fourier transform* by some authors, including Hewitt and Ross.

The formula for $\mathcal{F}(f)$ (and $\overline{F}(f)$) is reminiscent of the formula

$$\mathcal{F}(c)(x) = \sum_{m \in \mathbb{Z}} c_m e^{-imx},$$
where \((c_m)_{m \in \mathbb{Z}}\) is a sequence, except that the infinite sum is replaced by an integral. The integer \(m\) is replaced by the real number \(y\), the coefficient \(c_m\) is replaced by the value \(f(y)\) of the function \(f\) at \(y\), and the exponential \(e^{-im\theta}\) is replaced by \(e^{-i\theta y}\). Thus we can view \(F(f)(x)\) as a continuous superposition of the basic periodic functions \(y \mapsto e^{-iyx}\). However, this time, \(F(f)(x)\) is not necessarily periodic. We can still think of \(y\) as a frequency. In fact, in signal analysis, the domain of the Fourier transform is called the frequency domain.

The reader might be puzzled by the presence of the scale factor \(1/\sqrt{2\pi}\). The reason why it is included is that it makes certain formulae more symmetric, for example, the Fourier inversion formula and the Plancherel isomorphism. The deep reason for its need has to do with the fact that the domain of a Fourier transform \(\hat{f}\) is not actually \(\mathbb{R}\), but an isomorphic copy \(\hat{\mathbb{R}}\) of \(\mathbb{R}\), with a certain measure which is not necessarily identical to the measure on \(\mathbb{R}\). In order for certain results to hold, such as Fourier inversion, if \(\mathbb{R}\) is given the Lebesgue measure \(dx\), then \(\hat{\mathbb{R}}\) should be given the measure \(dx/2\pi\). Some authors use this normalization. Following Rudin [66, 67], a more symmetric normalization is to use the same scale factor \(1/\sqrt{2\pi}\) for both \(\mathbb{R}\) and \(\hat{\mathbb{R}}\). Another approach is to incorporate the factor \(2\pi\) in the exponential; that is, to use \(e^{-2\pi i\theta x}\) instead of \(e^{-i\theta x}\). In this case, the Lebesgue measure can be used for both \(\mathbb{R}\) and \(\hat{\mathbb{R}}\); see Folland [32, 31]. All this will be elucidated in Chapter 9. A consequence of using the measure \(dx/\sqrt{2\pi}\) is that the convolution of two functions \(f, g \in L^1(\mathbb{R})\) is

\[
(f * g)(x) = \int_{\mathbb{R}} f(x - y) g(y) \frac{dx(y)}{\sqrt{2\pi}} = \int_{\mathbb{R}} f(y) g(x - y) \frac{dx(y)}{\sqrt{2\pi}},
\]

and the inner product of two functions \(f, g \in L^2(\mathbb{R})\) is given by

\[
\langle f, g \rangle = \int_{\mathbb{R}} f(x) \overline{g(x)} \frac{dx}{\sqrt{2\pi}}.
\]

It is immediately verified that \(F(f)(x) = F(f)(-x) = \overline{F(f)(x)}\).

We will now state the most important results about the Fourier theory for \(\mathbb{R}\) without proof. Proofs of these results can be found in Folland [32], Rudin [66, 67], Stein and Shakarchi [77], and Malliavin [55]. The most important results will be proven in Chapter 9.

First, following Stein and Shakarchi [77] (Chapter 5 Section 1), observe that there is a nice class Mod(\(\mathbb{R}\)) of continuous functions \(f\) such that Mod(\(\mathbb{R}\)) \(\subseteq L^1(\mathbb{R})\), and such that the Fourier transform \(\hat{f} = F(f)\) of \(f\) is well-defined.

**Definition 7.7.** A continuous function \(f: \mathbb{R} \to \mathbb{C}\) is of *moderate decrease* is there is some \(A > 0\) such that

\[
|f(x)| \leq \frac{A}{1 + x^2} \quad \text{for all } x \in \mathbb{R}.
\]

The set of functions of moderate decrease is denoted by Mod(\(\mathbb{R}\)).
7.3. THE FOURIER TRANSFORM AND THE FOURIER COTRANSFORM ON $\mathbb{R}$ 243

It is shown in Stein and Shakarchi [77] (Chapter 5 Section 1) that $\text{Mod(}\mathbb{R}\text{)}$ is a vector space contained in $L^1(\mathbb{R})$, and that the Fourier transform $\hat{f} = F(f)$ of every function $f \in \text{Mod}(\mathbb{R})$ is well-defined. However, the Fourier transform $\hat{f}$ may not be of moderate decrease.

Let us give a few examples of Fourier transforms.

Example 7.3.

(1) Let $f$ be the characteristic function $\chi_{[-a,a]}$ of the interval $[-a,a]$, with $a > 0$. Then we have

$$F(f)(x) = \int_{\mathbb{R}} \chi_{[-a,a]}(y) e^{-iyx} \frac{dy}{\sqrt{2\pi}} = \int_{-a}^{a} e^{-iyx} \frac{dy}{\sqrt{2\pi}} = \frac{e^{-iax} - e^{iax}}{-ix\sqrt{2\pi}} = \frac{2}{\sqrt{2\pi}} \frac{\sin ax}{x}.$$ 

Therefore,

$$F(f)(x) = \frac{2}{\sqrt{2\pi}} \frac{\sin ax}{x}.$$ 

We also have

$$\overline{F}(f)(x) = \frac{2}{\sqrt{2\pi}} \frac{\sin ax}{x},$$

because

$$F(f)(x) = \int_{\mathbb{R}} \chi_{[-a,a]}(y) e^{iyx} \frac{dy}{\sqrt{2\pi}} = \int_{-a}^{a} e^{iyx} \frac{dy}{\sqrt{2\pi}} = \frac{e^{iax} - e^{-iax}}{ix\sqrt{2\pi}} = \frac{2}{\sqrt{2\pi}} \frac{\sin ax}{x}.$$ 

Remark: The function sinc is defined by

$$\text{sinc}(x) = \begin{cases} \frac{\sin \pi x}{\pi x} & \text{if } x \neq 0 \\ 1 & \text{if } x = 0. \end{cases}$$

Since clearly $\chi_{[-a,a]} \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, by Plancherel theorem (Theorem 7.16), $\chi_{[-a,a]} = \frac{2}{\sqrt{2\pi}} \frac{\sin ax}{x} \in L^2(\mathbb{R})$, so sinc $\in L^2(\mathbb{R})$. This can also be shown directly by showing that $(\sin \pi x/\pi x)^2$ is continuous and bounded near zero. However, the function sinc is not in $L^1(\mathbb{R})$, because

$$\int_{-\infty}^{\infty} \left| \frac{\sin \pi x}{\pi x} \right| dx = \infty.$$ 

As a consequence, its Fourier transform undefined. However, by Plancherel’s theorem (Theorem 7.16) the Fourier transform has a unique extension to $L^2(\mathbb{R})$, so the (extended) Fourier transform of sinc is $(1/\sqrt{2\pi})\chi_{[-\pi,\pi]}$.

The function sinc plays a crucial role in the sampling theorem, which gives a nice expression for a function $f \in L^2(\mathbb{R})$ which is band-limited, which means that $\hat{f}(x) = 0$ for all $x$ such that $|x| > a$. 

(2) Let $f$ be the function given by

$$f(x) = \frac{y}{x^2 + y^2};$$

with $y > 0$ fixed, and let $g$ be the function given by

$$g(x) = \frac{\pi}{\sqrt{2\pi}} e^{-y|x|}.$$

Then we have

$$\mathcal{F}(f)(x) = g(x)$$

$$\mathcal{F}(g)(x) = f(x).$$

The second formula is proven as follows. Using the fact that $y > 0$, we have

$$\mathcal{F}(g)(x) = \int_{\mathbb{R}} g(t) e^{itx} \frac{dt}{\sqrt{2\pi}}$$

$$= \int_{-\infty}^{0} \frac{\pi}{\sqrt{2\pi}} e^{yt} e^{itx} \frac{dt}{\sqrt{2\pi}} + \int_{0}^{\infty} \frac{\pi}{\sqrt{2\pi}} e^{-yt} e^{itx} \frac{dt}{\sqrt{2\pi}}$$

$$= \int_{-\infty}^{0} \frac{\pi}{\sqrt{2\pi}} e^{i(x-iy)t} \frac{dt}{\sqrt{2\pi}} + \int_{0}^{\infty} \frac{\pi}{\sqrt{2\pi}} e^{i(x+iy)t} \frac{dt}{\sqrt{2\pi}}$$

$$= \left[ \frac{\pi}{2\pi i (x - iy)} \right]_{-\infty}^{0} + \left[ \frac{\pi}{2\pi i (x + iy)} \right]_{0}^{\infty}$$

$$= \frac{1}{2i(x - iy)} - \frac{1}{2i(x + iy)}$$

$$= \frac{y}{x^2 + y^2}.$$

The first formula is harder to prove directly, but it follows from Fourier inversion (see Theorem 7.14).

We now return to the general case of functions in $L^1(\mathbb{R})$.

**Proposition 7.10.** *(Riemann–Lebesgue)* For any function $f \in L^1(\mathbb{R})$, the Fourier transform $\hat{f}$ (and the Fourier cotransform $\mathcal{F}(f)$) is continuous and tends to zero at infinity; that is, $\hat{f} \in \mathcal{C}_0(\mathbb{R}; \mathbb{C})$. Furthermore

$$\|\hat{f}\|_{\infty} \leq \|f\|_1.$$

As for the Fourier transform on $\mathbb{T}$, the Fourier transform converts a convolution into a product. The following proposition is a special case of Proposition 9.17 and Proposition 9.25, Parts (3) and (4).
Proposition 7.11. For any two functions \( f, g \in L^1(\mathbb{R}) \), the following properties hold:

1. \( \widehat{f * g} = \hat{f} \hat{g} \).
2. \( (\lambda_y f)(x) = e^{-iyx} \hat{f}(x) \).
3. \( (e^{iyx} f)(x) = \lambda_y(\hat{f})(x) \).
4. If \( \alpha > 0 \) and \( h(x) = f(x/\alpha) \), then \( \hat{h}(x) = \alpha \hat{f}(\alpha x) \).

For spectral synthesis in \( L^1(\mathbb{R}) \), the Poisson kernel is replaced by a function \( G_\mu \) defined using the following result.

Proposition 7.12. For any \( \mu > 0 \), we have

\[
e^{-\frac{x^2}{2\mu}} = \frac{1}{\sqrt{\mu}} \int e^{-\frac{y^2}{2\mu}} e^{i xy} \frac{dx(y)}{\sqrt{2\pi}}.
\]

Let \( \varphi \) be the function given by

\[
\varphi(x) = e^{-\frac{x^2}{2}}.
\]

Then \( \hat{\varphi} = \varphi \), and \( \varphi(0) = \int \varphi(x) dx \).

For a proof of Proposition 7.12, see Rudin [67] (Chapter 7, Lemma 7.6) or Folland [32] (Chapter 8, Section 8.3, Proposition 8.24).

Definition 7.8. The function \( \varphi \) given by

\[
\varphi(x) = e^{-\frac{x^2}{2}}
\]

is called a Gauss kernel or Weierstrass kernel.

For any \( \mu > 0 \), let \( G_\mu \) be the following function:

\[
G_\mu(x) = \frac{1}{\sqrt{\mu}} e^{-\frac{x^2}{2\mu}}.
\]

In view of Proposition 7.12 (replacing \( \mu \) by \( 1/\mu \)) we have

\[
G_\mu(x) = \frac{1}{\sqrt{\mu}} e^{-\frac{x^2}{2\mu}} = \int e^{-\frac{y^2}{2\mu}} e^{i xy} \frac{dx(y)}{\sqrt{2\pi}}.
\]

Here is our first result about spectral synthesis analogous to Theorem 7.2(1)-(2).
Proposition 7.13. (Spectral synthesis) Let \( f \in L^1(\mathbb{R}) \), let \( \hat{f} \) be its Fourier transform, and for any \( \mu > 0 \), let
\[
g_\mu(x) = (f * G_\mu)(x) = \int e^{iyx} \hat{f}(y) e^{-\frac{\mu y^2}{2}} \frac{dy}{\sqrt{2\pi}}.
\]
If \( f \in L^1(\mathbb{R}) \), then\[
\lim_{\mu \to 0} \| f - g_\mu \|_1 = 0,
\]
and if \( f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \), then\[
\lim_{\mu \to 0} \| f - g_\mu \|_2 = 0.
\]

Proposition 7.13 is proven in Malliavin [55] (Chapter 3, Section 2.4, Theorem 2.4.5).

In general, given a function \( f \in L^1(\mathbb{R}) \), the Fourier series
\[
\int e^{iyx} \hat{f}(y) \frac{dy}{\sqrt{2\pi}} = (\mathcal{F}(\hat{f}))(x)
\]
does not converge. However, for \( \mu > 0 \), the function \( g_\mu(x) = (f * G_\mu)(x) \) is well-defined and when \( \mu \) tends to 0, the function \( g_\mu \) is an approximation of \( f \) that tends to \( f \).

Now comes our first Fourier inversion theorem analogous to Theorem 7.3.

Theorem 7.14. (Fourier inversion formula) Let \( f \in L^1(\mathbb{R}) \). If \( \hat{f} \in L^1(\mathbb{R}) \), then
\[
f(x) = \int e^{iyx} \hat{f}(y) \frac{dy}{\sqrt{2\pi}} = (\mathcal{F}(\hat{f}))(x),
\]
almost everywhere. If \( f \) is continuous, the equation holds for all \( x \in \mathbb{R} \).

Theorem 7.14 is proven in Rudin [66] (Chapter 9, Theorem 9.11) Folland [32] (Chapter 8, Section 8.3, Theorem 8.26) and Malliavin [55] (Chapter 3, Section 2.4).

Proposition 7.15. If \( f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \), then
\[
\| f \|_2 = \| \hat{f} \|_2.
\]

Proposition 7.15 is proven in Malliavin [55] (Chapter 3, Section 4.2).

Here is the version of Plancherel’s theorem for \( L^2(\mathbb{R}) \).

Theorem 7.16. (Plancherel) If \( f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \), then \( \hat{f} \in L^2(\mathbb{R}) \). The Fourier transform defined on \( L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \) has a unique extension \( \mathcal{F} \) to \( L^2(\mathbb{R}) \) which is an isometric isomorphism of the Hilbert space \( L^2(\mathbb{R}) \) whose inverse is the (extension of) Fourier cotransform \( \mathcal{F}^{-1} \).
Theorem 7.16 proven in Rudin [66] (Chapter 9, Theorem 9.13) Folland [32] (Chapter 8, Section 8.3, Theorem 8.29) and Malliavin [55] (Chapter 3, Section 2.4)

Although Theorem 7.16 says that the Fourier transform $F$ extends to an isometric isomorphism of the Hilbert space $L^2(\mathbb{R})$, this result is useless in practice because for a function $f \in L^2(\mathbb{R})$ not in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, the extension of $F$ to $f$ is given by a limit.

The Fourier inversion formula also holds in the following situation.

**Proposition 7.17.** (Fourier inversion formula, II) Let $f \in L^2(\mathbb{R})$. If $\hat{f} \in L^1(\mathbb{R})$, then

$$f(x) = \int e^{ixy} \hat{f}(y) \frac{dy}{\sqrt{2\pi}} = (\mathcal{F}(\hat{f}))(x),$$

almost everywhere. If $f$ is continuous, the equation holds for all $x \in \mathbb{R}$.

Proposition 7.17 is proven in Rudin [66] (Chapter 9, Theorem 9.14).

**Definition 7.9.** Let $B(\mathbb{R}) = \{f \in L^1(\mathbb{R}) \mid \hat{f} \in L^1(\mathbb{R})\}$.

**Proposition 7.18.** The space $B(\mathbb{R})$ is dense in $L^1(\mathbb{R})$, $L^2(\mathbb{R})$, and $C_0(\mathbb{R}; \mathbb{C})$.

Proposition 7.18 is proven in Malliavin [55] (Chapter 3, Section 2.4).

## 7.4 The Sampling Theorem

In signal analysis, a function $f : \mathbb{R} \to \mathbb{C}$ represents a physical signal, and a common problem is to try to reconstruct this signal by sampling it, which means to compute its values at some sequence $t_1 < t_2 < \ldots$ of times. A basic issue is to determine how much information can be gained this way.

It turns out that if the signal $f$ is band-limited, which means that its Fourier transform $\hat{f}$ vanishes outside some interval $[-\Omega, \Omega]$, then $f$ can be completely reconstructed by sampling at the points $t_n = n\pi/\Omega$, for $n \in \mathbb{N}$.

**Theorem 7.19.** (Sampling theorem) Suppose that $f \in L^2(\mathbb{R})$ and that there is some $\Omega > 0$ such that $\hat{f}(x) = 0$ for all $|x| \geq \Omega$. Then $f$ is completely determined by its values at the points $t_n = n\pi/\Omega$, $n \in \mathbb{N}$. In fact, we have

$$f(t) = \sum_{n=0}^{\infty} f\left(\frac{n\pi}{\Omega}\right) \frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi}.$$

**Proof.** We follow Folland [30] (Chapter 7, Section 7.3). We can extend $\hat{f}$ to a periodic function of period $2\Omega$, and expand it as a Fourier series over the interval $[-\Omega, \Omega]$. For reason of later convenience, we use the index $-n$ instead of $n$, so we write

$$\hat{f}(t) = \sum_{n \in \mathbb{Z}} c_{-n} e^{-i\pi t/\Omega}, \quad (|t| \leq \Omega).$$
It is easy to show that the Fourier coefficients $c_{-n}$ are given by

$$c_{-n} = \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} \hat{f}(t) e^{i\pi nt/\Omega} dt = \frac{1}{2\Omega} \int_{-\infty}^{\infty} \hat{f}(t) e^{i\pi nt/\Omega} dt = \frac{\sqrt{2\pi}}{2\Omega} f\left(\frac{n\pi}{\Omega}\right),$$

where we used Fourier inversion (Theorem 7.14) and the fact that $\hat{f}(t) = 0$ for $|t| \geq \Omega$. Using these two facts again, we have

$$f(t) = \int_{-\Omega}^{\Omega} \hat{f}(\omega) e^{i\omega t} \frac{d\omega}{\sqrt{2\pi}} = \int_{-\Omega}^{\Omega} \sum_{n=-\infty}^{\infty} \frac{\sqrt{2\pi}}{2\Omega} f\left(\frac{n\pi}{\Omega}\right) e^{-i\pi n\omega/\Omega} e^{i\omega t} \frac{d\omega}{\sqrt{2\pi}}$$

$$= \frac{1}{2\Omega} \int_{-\Omega}^{\Omega} \sum_{n=-\infty}^{\infty} f\left(\frac{n\pi}{\Omega}\right) e^{i(\Omega t - n\pi)\omega/\Omega} d\omega$$

$$= \frac{1}{2\Omega} \sum_{n=-\infty}^{\infty} f\left(\frac{n\pi}{\Omega}\right) \left[ \frac{e^{i(\Omega t - n\pi)\omega/\Omega}}{i(\Omega t - n\pi)/\Omega} \right]_{\omega=-\Omega}^{\omega=\Omega}$$

$$= \sum_{n=-\infty}^{\infty} f\left(\frac{n\pi}{\Omega}\right) \sin(\Omega t - n\pi) \frac{\Omega t - n\pi}{\Omega t - n\pi}. $$

Since $\hat{f} \in L^2(\mathbb{R})$ the above manipulations are legitimate.

Observe that variants of the function $\text{sinc}(x) = \frac{\sin \pi x}{\pi x}$ show up.

Theorem 7.19 is due independently to E.T. Whittaker and Shannon (a similar result was published by Kotelnikov).

It worth noting that the functions

$$s_n(t) = \frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi}$$

form an orthonormal Hilbert basis for the Hilbert space of functions $f$ in $L^2(\mathbb{R})$ such that $\hat{f} = 0$ a.e. outside $(-\Omega, \Omega)$. This is because the computations in the proof of the sampling theorem show that $s_n$ is the Fourier cotransform (inverse Fourier transform) of the function

$$\hat{s}_n(x) = \begin{cases} \pi / \Omega & \text{if } |t| < \Omega \\ 0 & \text{otherwise.} \end{cases}$$

By Plancherel theorem and the fact that the functions $t \mapsto e^{-i\pi nt/\Omega}$ constitute an orthonormal Hilbert basis for $L^2(-\Omega, \Omega)$, we deduce that the $s_n$ form a Hilbert basis.

From a practical point of view, the expansion of $f$ given by the sampling theorem has the disadvantage that it generally does not converge very rapidly. By oversampling, that is, evaluating $f$ at a more closely spaced sequence of points $n\pi/\lambda\Omega$, with $\lambda > 1$, we can replace
7.5. THE FOURIER TRANSFORM AND THE FOURIER COTRANSFORM ON $\mathbb{R}^n$ 249

The functions $s_n$ by functions $g_\lambda(t - n\pi/\lambda\Omega)$ that vanish like $1/t^2$ when $t$ goes to infinity. The function $g_\lambda$ is given by

$$g_\lambda(t) = \frac{\cos \Omega t - \cos \lambda t}{\pi (\lambda - 1) t^2};$$

see Folland [30] (Chapter 7, Section 7.3, Exercise 8), and Stein and Shakarchi [77] (Chapter 5, Problem 20).

There is also a dual version of the sampling theorem for functions $f \in L^2(\mathbb{R})$ that vanish outside an interval $[-L, L]$. Then the Fourier transform $\hat{f}$ of $f$ is determined by sampling at the points $\omega = n\pi/L$, and $\hat{f}$ is given by the formula

$$\hat{f}(t) = \sum_{n=0}^{\infty} \hat{f}\left(\frac{n\pi}{\Omega}\right) \frac{\sin(\Omega t - n\pi)}{\Omega t - n\pi},$$

which is obtained from the formula of Theorem 7.19 by replacing $f$ with $\hat{f}$.

7.5 The Fourier Transform and the Fourier Cotransform on $\mathbb{R}^n$

The generalization of the results of Section 7.3 to $\mathbb{R}^n$ is straightforward.

**Definition 7.10.** For any function $f \in L^1(\mathbb{R}^n)$, the Fourier transform $\hat{f} = \mathcal{F}(f)$ of $f$ is the function $\mathcal{F}(f): \mathbb{R}^n \to \mathbb{C}$ defined on $\mathbb{R}^n$ given by

$$\hat{f}(x) = \mathcal{F}(f)(x) = \int_{\mathbb{R}^n} f(y) e^{-iy \cdot x} \frac{dx_n(y)}{(2\pi)^{n/2}},$$

and the Fourier cotransform $\overline{f} = \mathcal{F}(f)$ of $f$ is the function $\overline{\mathcal{F}(f)}: \mathbb{R}^n \to \mathbb{C}$ defined on $\mathbb{R}^n$ given by

$$\overline{\mathcal{F}} f(x) = \int_{\mathbb{R}^n} f(y) e^{iy \cdot x} \frac{dx_n(y)}{(2\pi)^{n/2}},$$

where $dx_n$ is the Lebesgue measure on $\mathbb{R}^n$, and $x \cdot y = \sum_{k=1}^{n} x_k y_k$ is the inner product of $x, y \in \mathbb{R}^n$.

**Remark:** The Fourier cotransform is also called the inverse Fourier transform by some authors, including Hewitt and Ross.

Again, we are using Rudin’s normalization scale factor $1/(2\pi)^{n/2}$, so we are really using the measure $dx_n/(2\pi)^{n/2}$. In particular, the convolution of two functions $f, g \in L^1(\mathbb{R}^n)$ is

$$(f * g)(x) = \int_{\mathbb{R}} f(x - y)g(y) \frac{dx_n(y)}{(2\pi)^{n/2}} = \int_{\mathbb{R}} f(y)g(x - y) \frac{dx_n(y)}{(2\pi)^{n/2}},$$
and the inner product of two functions \( f, g \in L^2(\mathbb{R}^n) \) is given by
\[
\langle f, g \rangle = \int_{\mathbb{R}^n} f(x)\overline{g(x)} \frac{dx_n}{(2\pi)^{n/2}}.
\]

It is immediately verified that \( \mathcal{F}(f)(x) = \mathcal{F}(f)(-x) = \overline{\mathcal{F}(f)(x)} \).

**Proposition 7.20.** (Riemann–Lebesgue) For any function \( f \in L^1(\mathbb{R}^n) \), the Fourier transform \( \hat{f} \) (and the Fourier cotransform \( \mathcal{F}(f) \)) is continuous and tends to zero at infinity; that is, \( \hat{f} \in C_0(\mathbb{R}^n; \mathbb{C}) \). Furthermore
\[
\|\hat{f}\|_\infty \leq \|f\|_1.
\]

As for the Fourier transform on \( \mathbb{T}^n \), the Fourier transform converts a convolution into a product.

**Proposition 7.21.** For any two functions \( f, g \in L^1(\mathbb{R}^n) \), the following properties hold for all \( x, y \in \mathbb{R}^n \):

1. \( \hat{f} \ast \hat{g} = \hat{f \ast g} \).
2. \( (\lambda y) \hat{f}(x) = e^{-iy \cdot x} \hat{f}(x) \).
3. \( (e^{iy \cdot x}) \hat{f}(x) = \lambda y \hat{f}(x) \).
4. If \( \alpha > 0 \) and \( h(x) = f(x/\alpha) \), then \( \hat{h}(x) = \alpha^n \hat{f}(\alpha x) \).

Proposition 7.21 is proven in Rudin [67] (Chapter 7, Theorem 7.2).

For any \( \mu > 0 \), let \( G_\mu \) be the following function:
\[
G_\mu(x) = \frac{1}{\mu^{n/2}} e^{-\frac{\|x\|^2}{2\mu}},
\]
where \( \|x\|^2 = x_1^2 + \cdots + x_n^2 \). Using Proposition 7.12, it is easy to see that
\[
G_\mu(x) = \int e^{-\frac{\|y\|^2}{2}} e^{ix \cdot y} \frac{dx_n(y)}{(2\pi)^{n/2}}.
\]

**Proposition 7.22.** (Spectral synthesis) Let \( f \in L^1(\mathbb{R}^n) \), let \( \hat{f} \) be its Fourier transform, and for any \( \mu > 0 \), let
\[
g_\mu(x) = (f \ast G_\mu)(x) = \int e^{iy \cdot x} \hat{f}(y) e^{-\frac{\|y\|^2}{2}} \frac{dx_n(y)}{(2\pi)^{n/2}}.
\]
If \( f \in L^1(\mathbb{R}^n) \), then
\[
\lim_{\mu \to 0} \|f - g_\mu\|_1 = 0,
\]
and if \( f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n) \), then
\[
\lim_{\mu \to 0} \|f - g_\mu\|_2 = 0.
\]
Proposition 7.13 is proven in Malliavin [55] (Chapter 3, Section 2.4).

**Theorem 7.23. (Fourier inversion formula)** Let \( f \in L^1(\mathbb{R}^n) \). If \( \hat{f} \in L^1(\mathbb{R}^n) \), then

\[
f(x) = \int e^{iy \cdot x} \hat{f}(y) \frac{dx_n(y)}{(2\pi)^{n/2}} = (\mathcal{F}(\hat{f}))(x),
\]

almost everywhere. If \( f \) is continuous, the equation holds for all \( x \in \mathbb{R}^n \).

Theorem 7.14 is proven in Rudin [67] (Chapter 7, Theorem 7.7) Folland [32] (Chapter 8, Section 8.3, Theorem 8.26) and Malliavin [55] (Chapter 3, Section 2.4).

**Definition 7.11.** Let \( B(\mathbb{R}^n) = \{ f \in L^1(\mathbb{R}^n) \mid \hat{f} \in L^1(\mathbb{R}^n) \} \).

**Proposition 7.24.** The space \( B(\mathbb{R}^n) \) is dense in \( L^1(\mathbb{R}^n) \), \( L^2(\mathbb{R}^n) \), and \( C_0(\mathbb{R}^n; \mathbb{C}) \).

Proposition 7.18 is proven in Malliavin [55] (Chapter 3, Section 2.4)

Here is the version of Plancherel’s theorem for \( L^2(\mathbb{R}^n) \).

**Theorem 7.25. (Plancherel)** If \( f \in L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n) \), then \( \hat{f} \in L^2(\mathbb{R}^n) \). The Fourier transform defined on \( L^1(\mathbb{R}^n) \cap L^2(\mathbb{R}^n) \) has a unique extension \( \mathcal{F} \) to \( L^2(\mathbb{R}^n) \) which is an isometric isomorphism of the Hilbert space \( L^2(\mathbb{R}^n) \) whose inverse is the (extension of) Fourier cotransform \( \mathcal{F} \).

Theorem 7.16 proven in Rudin [67] (Chapter 7, Theorem 7.9), Folland [32] (Chapter 8, Section 8.3, Theorem 8.29) and Malliavin [55] (Chapter 3, Section 2.4)

### 7.6 The Schwartz Space

It turns out that \( L^1(\mathbb{R}^n) \) contains an important subspace \( S(\mathbb{R}^n) \) of rapidly decreasing smooth functions, and that the Fourier transform is an isomorphism of this space, whose inverse is the Fourier cotransform. Functions in the space \( S(\mathbb{R}^n) \) and all their derivatives vanish at infinity faster than any power of \( \|x\| \), where \( \|x\| \) is the Euclidean norm on \( \mathbb{R}^n \). Technically, we introduce the following family of norms.

**Definition 7.12.** A continuous function \( f \in C(\mathbb{R}^n, \mathbb{C}) \) is rapidly decreasing if for every integer \( m \geq 0 \), there is some \( C > 0 \) such that \( (1 + \|x\|^2)^m |f(x)| \) remains bounded for all \( x \) such that \( \|x\| \geq C \). Let \( C_{0,0}(\mathbb{R}^n) \) be the set of rapidly decreasing functions. For every \( m \in \mathbb{N} \), define the norm \( \|f\|_{m,0} \) by

\[
\|f\|_{m,0} = \sup_{x \in \mathbb{R}^n} (1 + \|x\|^2)^m |f(x)|.
\]
Observe that Definition 7.12 immediately implies that $C_{0,0}(\mathbb{R}^n)$ is a subspace of $C_0(\mathbb{R}; \mathbb{C})$. Also, since $(1 + ||x||^2)^m|f(x)| \leq (1 + ||x||^2)^{m+1}|f(x)|/(1 + ||x||^2)$, if $(1 + ||x||^2)^{m+1}|f(x)|$ is bounded for all $x$ such that $||x||$ is large enough, we see that

$$\lim_{||x|| \to \infty} (1 + ||x||^2)^m|f(x)| = 0, \quad \text{for all } m \in \mathbb{N}. \quad (*)$$

Conversely, Condition $(*)$ implies that $(1 + ||x||^2)^{m+1}|f(x)|$ is bounded for all $x$ such that $||x||$ is large enough. Therefore, $(*)$ is equivalent to the condition used in Definition 7.12. In view of all this, we have

$$C_{0,0}(\mathbb{R}^n) = \{ f \in C_0(\mathbb{R}^n; \mathbb{C}) \mid ||f||_{m,0} < \infty, \text{ for all } m \geq 0 \}.$$  

**Definition 7.13.** The **Schwartz space** $\mathcal{S}(\mathbb{R}^n)$ consists of all smooth functions (that is, differentiable at all orders) given by

$$\mathcal{S}(\mathbb{R}^n) = \{ f \in C^\infty(\mathbb{R}^n) \cap C_{0,0}(\mathbb{R}^n) \mid \partial^\alpha f \in C_{0,0}(\mathbb{R}^n) \quad \text{for every multi-index } \alpha \}. \quad \text{For all } m, p \in \mathbb{N}, \text{ define the norm } ||f||_{m,p} \text{ by}$$

$$||f||_{m,p} = \sup_{|\alpha| \leq p} ||\partial^\alpha f||_{m,0} = \sup_{x \in \mathbb{R}^n, |\alpha| \leq p} (1 + ||x||^2)^m|\partial^\alpha f(x)|.$$  

Definition 7.13 is due to Laurent Schwartz. Observe that when $p = 0$, the norm $||f||_{m,p}$ is just the norm $||f||_{m,0}$ introduced in Definition 7.12, and that by definition,

$$\mathcal{S}(\mathbb{R}^n) = \{ f \in C_0(\mathbb{R}^n; \mathbb{C}) \mid ||f||_{m,p} < \infty, \quad \text{for all } m, p \in \mathbb{N} \}.$$  

Functions such as $x^k e^{-x^2}$ belong to $\mathcal{S}(\mathbb{R})$. The functions $e^{-c||x||^{2m}}$ and $e^{-c(1+||x||^2)^\alpha}$ with $c > 0$ and $\alpha > 0$ belong to $\mathcal{S}(\mathbb{R}^n)$.  

**Remark:** Altought it decreases very fast at infinity, the function $x \mapsto e^{-y|x|}$ (with $y > 0$ fixed) does not belong to $\mathcal{S}(\mathbb{R})$, because it is not differentiable at $x = 0$.

The space $\mathcal{D}(\mathbb{R}^n)$ of smooth functions with compact support is obviously a subspace of $\mathcal{S}(\mathbb{R}^n)$. In Section 4.9, the space $\mathcal{D}(\mathbb{R}^n)$ was denoted $\mathcal{K}_c^\infty(\mathbb{R}^n)$, but the notation $\mathcal{D}(\mathbb{R}^n)$ is more common.

Since the Schwartz space is a subspace of $C_0(\mathbb{R}^n; \mathbb{C})$, we can make it a normed vector space by giving it the norm $|| \cdot ||_\infty$. Unfortunately, with this norm it is not complete. We can give it a topology induced by the family of norms $|| \cdot ||_{m,p}$; according to the standard process for defining a topology in terms of a family of semi-norms described in Section 1.6. However, because this topology is Hausdorff and the family of norms is countable, $\mathcal{S}(\mathbb{R}^n)$ is actually a metric space; better, a complete metric space.

Let us now use the family of semi-norms $|| \cdot ||_{m,p}$ to define a topology on the Schwartz space $\mathcal{S}(\mathbb{R}^n)$. The semi-norms $|| \cdot ||_{m,p}$ are actually norms, so by Proposition 1.24, the space $\mathcal{S}(\mathbb{R}^n)$ is Hausdorff.
Definition 7.14. The vector space $S(\mathbb{R}^n)$ endowed with the topology induced by the countable family of norms $\| \cdot \|_{m,p}$ is a Hausdorff space called the topological Schwartz space.

We usually omit the word topological in topological Schwartz space. The value of the topology defined above is that $S(\mathbb{R}^n)$ is complete.

Theorem 7.26. The topological Schwartz space $S(\mathbb{R}^n)$ is a Fréchet space; that is, it is complete for the metric given by

$$d(x, y) = \sum_{m=0, p=0}^{\infty} \frac{1}{2^{m+p}} \frac{\|y - x\|_{m,p}}{1 + \|y - x\|_{m,p}}.$$ 

The space $D(\mathbb{R}^n)$ of smooth functions with compact support is dense in $S(\mathbb{R}^n)$.

Theorem 7.26 is proven in Rudin [67] (Chapter 7, Theorem 7.4 and Theorem 7.10).

The following result is proven in Malliavin [55] using the technique of regularization by some suitable convolution (Chapter 3, Section 3.2, Proposition 3.2.4).

Proposition 7.27. The space $D(\mathbb{R}^n)$ of smooth functions with compact support is dense in $L^1(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$ (with the Lebesgue measure). As a corollary, the Schwartz space $S(\mathbb{R}^n)$ is dense in $L^1(\mathbb{R}^n)$ and $L^2(\mathbb{R}^n)$.

The Fourier theory of $S(\mathbb{R}^n)$ is particularly nice because the Fourier transform is a map from $S(\mathbb{R}^n)$ to itself. The following results can be shown.

Theorem 7.28. If $f$ is any function in $S(\mathbb{R}^n)$, then the following properties hold:

1. We have $\hat{f} = \mathcal{F}(f) \in L^1(\mathbb{R}^n)$ and Fourier inversion holds:

$$f(x) = \int \hat{f}(y) e^{iy \cdot x} \frac{dx_n(y)}{(2\pi)^{n/2}} = (\mathcal{F}(\hat{f}))(x).$$

2. Actually, $\hat{f} \in S(\mathbb{R}^n)$ and there exist constants $c_{r,s}$ such that

$$\|\hat{f}\|_{r,s} \leq c_{r,s} \|f\|_{m+s,r}, \quad m > n.$$

3. The map $f \mapsto \hat{f} = \mathcal{F}(f)$ is an (algebra) isomorphism and a homeomorphism from $S(\mathbb{R}^n)$ to itself whose inverse is $\mathcal{F}$.

4. $x_k \hat{f}(x) = i \frac{\partial}{\partial x_k} \hat{f}(x)$.
Equation (5) is a small miracle since it says that the Fourier transform of a derivative acts as multiplication of the Fourier transform by $ix_k$, and it can be used to solve certain partial differential equations. Several examples of this technique are presented in Folland [30] and Stein and Shakarchi [77]. We give an example involving the heat equation.

Consider a region of the plane. Given an initial heat distribution, we are interested in finding the temperature $u(x, y, t)$ of the point $(x, y)$ at time $t$. Using Newton’s law of cooling, it can be shown that $u$ satisfies the partial differential equation called the time-dependent heat equation

$$
\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \frac{\sigma}{\kappa} \frac{\partial u}{\partial t};
$$

see Stein and Shakarchi [77] (Chapter 1) or Folland [30] (Appendix 1). After a long period of time, there is no more heat exchange, so that the system reaches a thermal equilibrium, and then $\frac{\partial u}{\partial t} = 0$. In this case, $u$ depends only on $x$ and $y$, and the time-dependent equation reduces to the steady-state heat equation

$$
\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0. \quad (1)
$$

The expression on the left-hand side of (1) is the Laplacian of $u$. Suppose our domain is the upper half plane

$$
\mathbb{R}^2_+ = \{(x, y) \in \mathbb{R}^2 \mid y > 0\}.
$$

We would like to find a (the?) solution $u(x, y)$ of the above equation, given that

$$
u(x, 0) = f(x) \quad (2)
$$
on the boundary, where $f$ is some given function (in $S(\mathbb{R})$).

The method for finding the solution $u$ proceeds in two steps.

Step 1. The first trick is to apply the Fourier transform with respect to $x$ to both (1) and (2). In view of Equation (5) of Theorem 7.28, we get the two equations

$$-x^2 \hat{u}(x, y) + \frac{\partial^2 \hat{u}}{\partial y^2}(x, y) = 0
$$

$$
\hat{u}(x, 0) = \hat{f}(x).
$$
Observe that we now have a much simpler problem, namely an *ordinary differential equation* with respect to $y$ in the unknown $\hat{u}(x,y)$. The solution of the first equation is well-known:

$$\hat{u}(x,y) = C_1(x)e^{\left|x\right|y} + C_2(x)e^{-\left|x\right|y},$$

with

$$C_1(x) + C_2(x) = \hat{f}(x).$$

Since the first term has exponential increase, it has to be discarded (because we are seeking solutions in $S(\mathbb{R})$), so we must have $C_1 = 0$, and we get

$$\hat{u}(x,y) = \hat{f}(x)e^{-\left|x\right|y}. \quad (\dagger)$$

*Step 2.* The second trick is that if we can find the Fourier cotransform (inverse Fourier transform) $x \mapsto \mathcal{P}_y(x)$ of $x \mapsto e^{-\left|x\right|y}$, since $\mathcal{F}(\mathcal{P}_y)(x) = e^{-\left|x\right|y}$, we have

$$\hat{u}(x,y) = \hat{f}(x)e^{-\left|x\right|y} = \mathcal{F}(f)(x)\mathcal{F}(\mathcal{P}_y)(x) = \mathcal{F}(f*\mathcal{P}_y)(x),$$

and then by Fourier inversion we can express the solution $u$ as the convolution

$$u(x,y) = (f * \mathcal{P}_y)(x).$$

But we showed in Example 7.3(2) that the Fourier cotransform (inverse Fourier transform) of

$$g(x) = \frac{\pi}{\sqrt{2\pi}} e^{-y|y|}$$

(with $y > 0$) is

$$f(x) = \frac{y}{x^2 + y^2},$$

so the Fourier cotransform (inverse Fourier transform) of $x \mapsto e^{-|x|y}$ is

$$\mathcal{P}_y(x) = \frac{\sqrt{2\pi}}{\pi} \frac{y}{x^2 + y^2},$$

Therefore we obtain the solution

$$u(x,y) = (f * \mathcal{P}_y)(x).$$

Explicitly, we have

$$u(x,y) = \int_\mathbb{R} \frac{yf(x-t)}{\pi(x^2 + y^2)} \, dt,$$

which is called the *Poisson integral formula*, and the function

$$\mathcal{P}_y(x) = \frac{\sqrt{2\pi}}{\pi} \frac{y}{x^2 + y^2}.$$
is called the Poisson kernel for the upper half plane (there are variants of $\mathcal{P}_y(x)$ with different constants).

To be honest, we still need to check carefully that $u(x,y) = (f * \mathcal{P}_y)(x)$ is indeed a solution of the problem; for details, see Stein and Shakarchi [77] (Chapter 5).

Various other problems involving the wave equation or the heat equation can be solved using the above method; see Stein and Shakarchi [77] and Folland [30].

The reader may have noticed that in order to apply the above method we had to assume that the solutions belong to $\mathcal{S}(\mathbb{R})$, which is a heavy restriction. It turns out that if we use certain kinds of generalized functions, called distributions, then we can apply a more general version of Theorem 7.28, and obtain more general solutions for various partial differential equations.

### 7.7 The Poisson Summation Formula

Given a function $f \in \mathcal{S}(\mathbb{R})$, it is sometimes desirable to make a periodic function from $f$. One way to do this is to define the function $F_1$ as follows.

**Definition 7.15.** Given a function $f \in \mathcal{S}(\mathbb{R})$, the function $F_1 : \mathbb{R} \to \mathbb{C}$ is given by

$$F_1(x) = \sum_{n \in \mathbb{N}} f(x + 2\pi n).$$

Since $f \in \mathcal{S}(\mathbb{R})$, the series converges absolutely and uniformly on every compact subset of $\mathbb{R}$, so $F_1$ is continuous. It is also clear that

$$F_1(x) = F_1(x + 2\pi n), \quad n \in \mathbb{Z},$$

so $F_1$ is indeed periodic. We call $F_1$ the *periodization* of $f$.

There is another way to make $f$ periodic, which is to use the sequence of numbers $(\widehat{f}(n))_{n \in \mathbb{Z}}$ and to make the Fourier series from it,

$$F_2(x) = \sum_{n \in \mathbb{Z}} \widehat{f}(n)e^{inx}.$$

Again, since $f \in \mathcal{S}(\mathbb{R})$, the sum converges absolutely and uniformly since $\widehat{f} \in \mathcal{S}(\mathbb{R})$, so $F_2$ is continuous. The remarkable fact is that $F_1 = F_2$.

**Theorem 7.29.** *(Poisson summation formula)* For any function $f \in \mathcal{S}(\mathbb{R})$, we have

$$\sum_{n \in \mathbb{Z}} f(x + 2\pi n) = \sum_{n \in \mathbb{Z}} \widehat{f}(n)e^{inx}.$$

In other words, the Fourier coefficients of $F_1(x) = \sum_{n \in \mathbb{Z}} f(x + 2\pi n)$ are the numbers $\widehat{f}(n)$. In particular,

$$\sum_{n \in \mathbb{Z}} f(2\pi n) = \sum_{n \in \mathbb{Z}} \widehat{f}(n).$$
The proof of Theorem 7.29 can be found in Stein and Shakarchi [77] (Chapter 5, Theorem 3.1). It consists in computing the Fourier coefficients of $F_1$.

Theorem 7.29 also holds when both $f$ and $\hat{f}$ are of moderate decrease (recall Definition 7.7).

**Remark:** There is a relationship between the Poisson kernel on the unit disk, 
\[ P_r(\theta) = \sum_{n=\infty}^{-\infty} r^{|n|} e^{in\theta} = \frac{1 - r^2}{1 - 2r \cos \theta + r^2}, \]
and the Poisson kernel on the upper-half plane, 
\[ P_y(x) = \frac{\sqrt{2\pi}}{\pi} \frac{y}{x^2 + y^2}, \]
obtained by applying the Poisson summation formula to $f(x) = P_y(x)$ and $\hat{f}(\theta) = e^{-|\theta|y}$.

We get 
\[ \sum_{n \in \mathbb{Z}} P_y(x + 2\pi n) = \sum_{n \in \mathbb{Z}} e^{-|n|y} e^{inx} = \sum_{n \in \mathbb{Z}} (e^{-y})^{|n|} e^{inx} = P_{e^{-y}}(x). \]

In summary, (for $y > 0$), we have 
\[ P_{e^{-y}}(x) = \sum_{n \in \mathbb{Z}} P_y(x + 2\pi n). \]

### 7.8 Pointwise Convergence of Fourier Series on $\mathbb{T}$

In this section, we come back to periodic functions on $\mathbb{R}$, that is, to functions in $L^1(\mathbb{T})$. By Theorem 7.2, if $f \in L^2(\mathbb{T})$, then 
\[ \lim_{n \to \infty} \left\| f - \sum_{m=-n}^{m=n} c_m e^{im\theta} \right\|_2 = 0, \]
where $c_m$ is the $m$th Fourier coefficient of $f$, 
\[ c_m = \int_{-\pi}^{\pi} f(t) e^{-imt} \frac{dx(t)}{2\pi}. \]

Thus the partial sums 
\[ S_{m,f}(\theta) = \sum_{k=-m}^{k=m} c_k e^{ik\theta} \]
converge to $f$ in the $L^2$-norm. However, even if $f$ is continuous, the partial sums $S_{m,f}$ may not converge to $f$ pointwise.
The first example of a function whose Fourier series diverges at 0 was given by du Bois–Reymond in 1873. This is a fairly complicated example involving a piecewise monotone function that oscillates indefinitely near 0. Simpler examples were given later by Fejér and Lebesgue.

Fejér’s example makes use of the functions
\[ F_n(x) = \sin x + \frac{1}{2} \sin 2x + \cdots + \frac{1}{n} \sin nx, \]
which are uniformly bounded. The series
\[ \sum_{n=1}^{\infty} \frac{1}{n^2} \sin(3^{n^2}x)F_{3^{n^2}}(x) \]
defines a continuous function \( f \), but it can be shown that its Fourier series diverges for \( x = 0 \). See also the example in Stein and Shakarchi [77] (Chapter 3, Section 2.2).

In 1926, Kolmogoroff gave an example of a function \( f \in L^1(\mathbb{T}) \) whose Fourier series diverges for all \( x \).

Later it was found that a systematic method for producing functions with a “bad” Fourier series was to use the Banach–Steinhaus theorem.

\textbf{Definition 7.16.} For any fixed \( \theta \in \mathbb{R}/2\pi\mathbb{Z} \) and for any continuous function \( f \in \mathcal{C}(\mathbb{T}; \mathbb{C}) \), let
\[ S^*(f, \theta) = \sup_{m \in \mathbb{N}} |S_{m,f}(\theta)|. \]
Also recall the following definition about Borel sets.

\textbf{Definition 7.17.} Let \( X \) be a topological space. Countable unions of closed subsets of \( X \) are called \( F_\sigma \)-sets, and countable intersections of open sets of \( X \) are called \( G_\delta \)-sets.

The following result is proven in Rudin [66] (Chapter 5, page 102). The proof uses the Banach–Steinhaus theorem.

\textbf{Proposition 7.30.} For every \( \theta \in \mathbb{R}/2\pi\mathbb{Z} \), there is a subset \( E_\theta \subseteq \mathcal{C}(\mathbb{T}; \mathbb{C}) \) of continuous functions which is a dense \( G_\delta \) set in \( \mathcal{C}(\mathbb{T}; \mathbb{C}) \) such that \( S^*(f, \theta) = \infty \) for all \( f \in E_\theta \). Consequently, the Fourier series of every \( f \in E_\theta \) diverges at \( \theta \).

Using Baire’s theorem, a stronger result can be obtained, as shown in Rudin [66] (Chapter 5, Theorem 5.12).

\textbf{Proposition 7.31.} There is a set \( E \subseteq \mathcal{C}(\mathbb{T}; \mathbb{C}) \) of continuous functions which is a dense \( G_\delta \) set in \( \mathcal{C}(\mathbb{T}; \mathbb{C}) \) and which has the following property: For every \( f \in E \), the set
\[ Q_f = \{ \theta \in \mathbb{R}/2\pi\mathbb{Z} \mid S^*(f, \theta) = \infty \} \]
is a dense \( G_\delta \) set in \( \mathbb{R}/2\pi\mathbb{Z} \).
As a consequence, the Fourier series of every continuous function \( f \in E \) diverges for infinitely many points. In fact, \( E \) are the \( Q_f \) are uncountable; see Rudin [66] (Chapter 5, Theorem 5.13).

We just saw that in general, the partial sums \( S_{m,f} \) do not behave well, so if we want to approximate a continuous function on \( \mathbb{T} \), we should not count on the partial sums to do the job. We saw in Example 6.10 that the Cesàro means,

\[
A_{n,f} = \frac{1}{n} (S_{0,f} + \cdots + S_{n-1,f}),
\]

have a much better behavior, since they converge uniformly to \( f \).

We are led to the conclusion that in order to obtain positive results for pointwise convergence of the partial sums \( S_{m,f} \), we must restrict the class of functions that we are considering. Dirichlet was the first to obtain a significant result. In 1829, he proved that the partial sums \( S_{m,f} \) converge pointwise to \((f(x^+)+f(x^-))/2\), for every piecewise continuous and piecewise monotone function \( f \). Here \( f(x^+) \) is the limit when \( y \) tends to \( x \) from above, and \( f(x^-) \) is the limit when \( y \) tends to \( x \) from below (see Definition 1.13). His paper is not only significant because of its results but because it raises the standards of rigor in mathematical exposition to a new level. Dirichlet’s full paper is reproduced in Kahane and Lemarié–Rieusset [45]. As Kahane comments, “Dirichlet’s style is superb and incredibly modern.”

Later on, it was realized that what is really needed for this pointwise convergence result to hold is that the functions have bounded variation. Camille Jordan’s, whose mathematical interests were group theory, algebra, and its relations to geometry, introduced the notion of function of bounded variation in 1881, and published one paper generalizing Dirichlet’s paper to this class of functions.

Let \( f : [a, b] \to \mathbb{R} \) be a function. The intuition is that the variation of \( f \) over \([a, b]\) is the total distance travelled from time \( a \) to time \( b \). If \( f' \) exits and is continuous, then the variation is \( \int_a^b |f'(t)| \, dt \). Otherwise, we approximate the curve by a piecewise affine function. For this, we subdivide the interval \([a, b]\) into smaller intervals, \([t_{j-1}, t_j]\) and approximate \( f \) on this subinterval by the line segment from \((t_{j-1}, f(t_{j-1}))\) to \((t_j, f(t_j))\).

More precisely, consider a function \( f : [a, b] \to \mathbb{C} \). It will be more convenient to assume that \( a = -\infty \) and to view \( b \) as the variable, say \( x \). We consider subdivisions of the interval \((-\infty, x]\) using finite sequences \( x_0 < x_1 < \cdots < x_n = x \).

**Definition 7.18.** Let \( f : [a, b] \to \mathbb{C} \) be a function. The total variation function \( T_f \) of \( f \) is the function given by

\[
T_f(x) = \sup \left\{ \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| \mid -\infty < x_0 < x_1 < \cdots < x_n = x, \, n \in \mathbb{N} - \{0\} \right\},
\]
where the supremum is taken over all finite subdivisions \( x_0 < x_1 < \cdots < x_n = x \). If \([a, b]\) is a finite interval \((a \leq b)\), then the total variation of \( f \) on \([a, b]\) is the quantity

\[
V(f, a, b) = \sup \left\{ \sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| \mid a = x_0 < x_1 < \cdots < x_n = b, \; n \in \mathbb{N} - \{0\} \right\}.
\]

The set \( BV \) of functions of bounded variation is the set of functions \( f : \mathbb{R} \to \mathbb{C} \) such that \( \lim_{x \to -\infty} |T_f(x)| < \infty \). The set \( BV([a, b]) \) of functions of bounded variation over \([a, b]\) is the set of functions \( f : [a, b] \to \mathbb{C} \) such that \( V(f, a, b) \) is finite.

If \( f : [a, b] \to \mathbb{C} \) is a function in \( BV \), since \( a \) can be chosen as a subdivision point we see immediately that

\[
V(f, a, b) = T_f(b) - T_f(a),
\]

so \( f \in BV([a, b]) \). We also see that \( T_f(x) \) is an increasing function. The limits \( \lim_{x \to -\infty} f(x) \) and \( \lim_{x \to +\infty} f(x) \) also exist, as a corollary of Proposition 7.34 below.

\section*{Example 7.4.}

1. If \( f : \mathbb{R} \to \mathbb{R} \) is bounded and increasing, then \( f \in BV \). In fact, \( T_f(x) = f(x) - f(-\infty) \).

2. The space \( BV \) is a complex vector space.

3. If \( f \) is differentiable on \( \mathbb{R} \) and if \( f' \) is bounded, then \( f \in BV([a, b]) \) for every finite interval \([a, b]\) (by the mean value theorem).

4. If \( f(x) = \sin x \), then \( f \in BV([a, b]) \) for every finite interval \([a, b]\), but \( f \notin BV \) (it oscillates forever).

5. If \( f(x) = x \sin(x^{-1}) \) for \( x \neq 0 \) and \( f(0) = 0 \), then \( f \notin BV([a, b]) \) for \( a \leq 0 < b \) or \( a < 0 \leq b \).

Here are some of the main properties of functions of bounded variation; proofs can be found in Folland [32] (Chapter 3, Section 3.5).

\section*{Proposition 7.32.} Let \( f : \mathbb{R} \to \mathbb{C} \) be a function. If \( f \in BV \), then \( \lim_{x \to -\infty} T_f(x) = 0 \).

\section*{Proof.} By definition of \( T_f(x) \) as a least upper bound, for every \( \epsilon > 0 \), there is some subdivision \( x_0 < x_1 < \cdots < x_n = x \) such that

\[
\sum_{j=1}^{n} |f(x_j) - f(x_{j-1})| \geq T_f(x) - \epsilon.
\]

By definition of \( V(f, x_0, x) \), this implies that

\[
T_f(x) - T_f(x_0) \geq T_f(x) - \epsilon,
\]

so \( T_f(x_0) \leq \epsilon \). Since \( T_f \) is increasing, we also have \( T_f(y) \leq \epsilon \) for all \( y \leq x_0 \). Since \( \epsilon \) is arbitrary, we must have \( \lim_{x \to -\infty} f(x) = 0 \).

\qed
**Proposition 7.33.** Let \( f : \mathbb{R} \to \mathbb{R} \) be a function. If \( f \in BV \), then \( T_f + f \) and \( T_f - f \) are increasing functions.

**Proposition 7.34.** Let \( f : \mathbb{R} \to \mathbb{C} \) be a function.

1. We have \( f \in BV \) iff both the real part and the imaginary part of \( f \) belong to \( BV \).

2. If \( f : \mathbb{R} \to \mathbb{R} \), then \( f \in BV \) iff \( f \) can be written as the difference of two bounded increasing functions; these can be chosen as \((T_f + f)/2\) and \((T_f - f)/2\). This is called a Jordan decomposition.

3. If \( f : \mathbb{R} \to \mathbb{C} \) and if \( f \in BV \), then the left limit \( f(x-) \) and the right limit \( f(x+) \) exist for all \( x \in \mathbb{R} \), including \( x = -\infty \) and \( x = +\infty \).

4. If \( f : \mathbb{R} \to \mathbb{C} \) and if \( f \in BV \), then \( f \) has at most countably many discontinuities.

5. If \( f \in BV \) and if we let \( g(x) = f(x+) \), then \( f' \) and \( g' \) exist almost everywhere and are equal almost everywhere.

**Remark:** The space \( NBV \) consists of the functions \( f : \mathbb{R} \to \mathbb{C} \) in \( BV \) which are right continuous and such that \( f(-\infty) = 0 \). There is a relationship between the space \( NBV \) and the complex Borel measures on \( \mathbb{R} \). If \( \mu \) is a complex Borel measure, then the function \( F(x) = \mu((-\infty, x]) \) is in \( NBV \), and conversely, given any function \( f \in NBV \), there is a unique complex measure \( \mu_f \) such that \( f(x) = \mu_f((-\infty, x]) \); see Folland [32] (Chapter 3, Section 3.5, Theorem 3.29).

We now return to Fourier series on \( \mathbb{T} \), and state the following theorem essentially due to Jordan which generalizes an historically famous result of Dirichlet.

**Theorem 7.35.** For any \( f \in L^1(\mathbb{T}) \), if \( f \in BV([\pi, \pi]) \), then

\[
\lim_{m \to \infty} S_m,f(x) = \frac{f(x+) + f(x-)}{2}
\]

for all \( x \in [-\pi, \pi] \). In particular, \( \lim_{m \to \infty} S_m,f(x) = f(x) \) whenever \( f \) is continuous at \( x \).

Theorem 7.35 is proven in Folland [32] (Chapter 8, Section 8.5, Theorem 8.43).

Other convergence theorems (some about pointwise convergence) are discussed in Folland [30] and Stein and Shakarchi [77].

The behavior of \( S_m,f \) at a jump discontinuity \( x = m \), with \( m \in \mathbb{Z} \), (a point \( x \) where \( f(x) \neq f(x-) \) or \( f(x) \neq f(x+) \)) is a little strange. It turns out that near an integer value of \( x \), the function \( S_m,f \) contains spikes that overshoot or undershoot the function \( f \), and when \( m \) tends to infinity, the width of the spikes tends to zero but the height does not. This behavior is known as Gibbs phenomenon. For example, the function

\[
\varphi(x) = 2\pi \left( \frac{1}{2} - (x - [x]) \right)
\]
(where \(\lfloor x \rfloor\) is the greatest integer \(\leq x\)) is periodic (of period \(2\pi\)) and exhibits the Gibbs phenomenon. One easily computes the Fourier coefficients, which are

\[
c_0 = 0, \quad c_m = \frac{1}{im}, \quad m \neq 0.
\]

Then we get

\[
S_{m,\varphi}(x) = \sum_{k=1}^{m} \frac{2\sin(kx)}{k}.
\]

For more details, see Folland [32] (Chapter 8, Section 8.5).

### 7.9 The Heisenberg Uncertainly Principle

A fundamental fact about Fourier series is that it is impossible for a nonzero function \(f \in L^2(\mathbb{R})\) that both \(f\) and its Fourier transform \(\hat{f}\) vanish outside of some finite interval. This can be shown easily using some elementary complex analysis; see Folland [30] (Chapter 7).

There is an even stronger limitation. Roughly, \(f\) and \(\hat{f}\) can’t be both highly localized. A precise way to state this fact is to define the notion of dispersion.

**Definition 7.19.** For any function \(f \in L^2(\mathbb{R})\), the dispersion of \(f\) about the point \(a\) is given by

\[
\Delta_a f = \left( \int (x - a)^2 |f(x)|^2 \, dx \right)^{1/2} \left( \int |f(x)|^2 \, dx \right)^{-1/2} = \frac{\int (x - a)^2 |f(x)|^2 \, dx}{\int |f(x)|^2 \, dx}.
\]

Then we have the following theorem.

**Theorem 7.36.** (Heisenberg inequality) Let \(f\) be a function in \(L^2(\mathbb{R})\). Then for all \(a, b \in \mathbb{R}\), we have

\[
(\Delta_a f)(\Delta_b \hat{f}) \geq \frac{1}{4}.
\]

Theorem 7.36 is proven in Stein and Shakarchi [77] in the case where \(f \in \mathcal{S}(\mathbb{R})\) (Chapter 5, Section 4), and in Folland [30] (Chapter 7), in a more general situation.

Theorem 7.36 has an interpretation in quantum mechanics (we apologize to those who are familiar with quantum mechanics for the vagueness of our comments). In quantum mechanics, among other things, one studies the motion of particles. For this, we need to know the position and the momentum of the particle, but these are not known exactly, but instead in terms of probabilities. For simplicity, assume that we are dealing with an electron that travels along the real line. There is a function \(\psi\), called a state function (or wave function), which we assume to be in \(\mathcal{S}(\mathbb{R})\), normalized so that

\[
\int |\psi(t)|^2 \, dt = 1,
\]
such that the probability that the electron is located in the interval \([a, b]\) is

\[
\int_a^b |\psi(t)|^2 \, dt.
\]

The *expectation* of where the particle might be is the best guess of the position of the particle, and it is given by

\[
\bar{x} = \int\! t |\psi(t)|^2 \, dt.
\]

The *uncertainty* attached to our expectation, or *variance*, is given by the quantity

\[
\Delta_x \psi = \int\! (t - \bar{x})^2 |\psi(t)|^2 \, dt.
\]

It is easy to see that the expectation \(\bar{x}\) is the choice of \(a\) that minimizes the variance \(\int\! (t - a)^2 |\psi(t)|^2 \, dt\).

Now in quantum mechanics, the momentum \(\xi\) of the particle is determined by the Fourier transform \(\hat{\psi}\) of \(\psi\), in the sense that the probability that the electron has momentum \(\xi\) in the interval \([a, b]\) is

\[
\int_a^b |\hat{\psi}(t)|^2 \, dt.
\]

As above, we also have the expectation

\[
\bar{\xi} = \int\! t |\hat{\psi}(t)|^2 \, dt,
\]

and the variance

\[
\Delta_{\bar{\xi}} \hat{\psi} = \int\! (t - \bar{\xi})^2 |\hat{\psi}(t)|^2 \, dt.
\]

Theorem 7.36 states that

\[
(\Delta_x \psi)(\Delta_{\bar{\xi}} \hat{\psi}) \geq \frac{1}{4},
\]

which is the *Heisenberg uncertainty principle*. Intuitively, it says that the more certain we are about the position of the particle, the less certain we are about its momentum, and vice versa. Actually, we have ignored units of measurements, and in fact Planck’s constant \(\hbar\) should be inserted, so the physically correct statement of Heisenberg uncertainty principle is that

\[
(\Delta_x \psi)(\Delta_{\bar{\xi}} \hat{\psi}) \geq \frac{\hbar}{4}.
\]

For more details, see Stein and Shakarchi [77] (Chapter 5, Section 4), and Folland [30] (Chapter 7), and for even more, any text on quantum mechanics.
7.10 Fourier’s Life; a Brief Summary

Joseph Fourier was born on March 21, 1768, in Auxerre, a town in northern Burgundy, France, and died in 1830. Because he was 21 during the French revolution (1789), he had a particularly exciting life. In this section we give a very condensed summary of his life, based on the wonderful account in Chapter 1 of Kahane [45].

Fourier’s family was poor. At age 10, Fourier had already lost his mother and his father. The organist of the cathedral had noticed that Fourier was exceptionally gifted, so he arranged to have him attend the military college in Auxerre. Teaching was provided by Benedictine monks. Fourier fell in love with mathematics through the writings of Bézout and Clairaut. He worked very hard, and completed his studies early at age 14. It was arranged that he stayed in the college, in preparation for starting teaching there at age 16. He already sent some papers on locating the roots of algebraic equations to the Institut, which were noticed by Legendre. Legendre requested that Fourier join the army in the artillery (the most scientific branch), but his request was denied because Fourier was not a “noble” (he was of humble extraction). So in 1787 he entered the benedictine abbey of Saint-Benoit in Fleury in preparation for becoming a monk.

Fourier stayed in Fleury until 1789, where he taught mathematics. He was going to become a monk on November 5, 1798, but the revolution had taken place and put a hold on new religious positions on November 2. Fourier never became a monk!

Between 1789 and 1793, Fourier continues working on mathematics, but also gets involved in the revolution. He is involved in the supply of food and weapons to Orléans, and being a good politician, does a very good job at that.

In 1794, he is sent to Paris where a new school called “École Normale” has been created. There he meets other mathematicians such as Laplace and Monge. The “École Polytechnique” is created in 1794, and Fourier teaches there between 1795 and 1798.

Apparently, Fourier is noticed by Napoleon, and he follows Napoleon for the expedition to Egypt. There, Napoleon creates a replica of the “Institut de France,” headed by Monge, and with Fourier as “perpetual secretary.” So Fourier becomes an archeologist.

Napoleon goes back to France where he proclaims himself emperor. Still in Egypt, Fourier negotiates the retreat of the French defeated by the British. Fourier returns to France in 1801. At his return, Napoleon charges Fourier with the important administrative position of “préfet” (sort of superintendent) of the department of Isère. France was divided in 90 departments (districts), and the main city in Isère is Grenoble. One might think that this would signal the end of Fourier’s mathematical life, but not at all. Fourier was also an astute politician, and a good administrator, so he excelled at everything he did. He started working on his theory of heat propagation.

In 1807, he submitted a paper on this subject to the Institut. Lagrange, Laplace, Lacroix, and Monge were the referees. Lagrange felt that the paper was not rigorous enough, and the paper was rejected. The topic of heat propagation was then proposed for a competition.
Fourier reworked his paper which was submitted in 1811, and this time, the same referees awarded him the price. However, the commentaries, although they praised the originality of the work, especially the heat equation, pointed out some lack of rigor.

Fourier continued to work on a major manuscript on the analytic theory of heat, but this manuscript was not published until 1822.

In the meantime, Napoleon abdicated in 1815. Life is hectic. Fourier is opposed to Napoleon III. Although he is promoted as préfet of the Rhone, he resigned from this position and returns to Paris. There, with the help of a former student, he finds the position of director of the bureau of statistics! He is elected at the Academy of sciences in 1817.

Fourier continues working on his book on the analytic theory of heat, but also does some work in statistics. In 1822, he finally publishes his book, *Théorie analytique de la chaleur*, which includes all of his work on the subject, starting with his work between 1807 and 1811, and then 1816, 1821.

Laplace, Monge, Liouville, Dirichlet, Navier, Sturm had great respect for Fourier. However, Poisson and Cauchy, who were his rivals, were not his friends. The obituaries by Arago and Cousin did not do justice to Fourier’s work. It is sad that the collected works of Fourier were never gathered and published. Darboux collected some of Fourier’s papers, but ignored all his work on what is now called linear programming, saying that Fourier attributed an exaggerated importance to this type of work. After all, Fourier adapted Gaussian elimination to linear inequalities (Fourier–Motzkin elimination).

However, Fourier’s work had a tremendous influence in mathematics, physics, and engineering, so even if he did not get the recognition that he deserved from his peers, the public voted with his feet.

We must conclude with a famous note of Jacobi to Legendre, sent on July 2, 1831, after Fourier’ death.

Fourier deeply believed that the main goal of mathematics was to provide a clear explanation of natural phenomena. His is book, he writes”

“L’étude approfondie de la nature est la source la plus féconde des découvertes mathématiques.”

Jacobi (1804-1851) complains to Legendre that Poisson included in a report that Fourier made the reproach to Abel and Jacobi that they did not work enough on the theory of heat, but instead on number theory. Jacobi says:

“... mais un philosophe comme lui aurait dû savoir que le but unique de la science, c’est l’honneur de l’esprit humain, et que, sous ce titre, une question de nombres vaut autant qu’une question du système du monde.”

Roughly translated: But such a philosopher should have known that the unique goal of science is the honor of the human spirit, and that, as such, a question about numbers is as worthy as a question about the system of the world.
A very complete account of the mathematical history of Fourier series and its influence on mathematics can be found in the captivating book by Kahane and Lemarié–Rieusset [45].
Chapter 8

Normed Algebras and Spectral Theory

Let $G$ be a locally compact abelian group. In order to define the notion of Fourier transform on $L^1(G)$, one needs to figure out what is its domain. The answer is that the domain of the Fourier transform on $L^1(G)$ is the group $\hat{G}$ of (unitary) characters of $G$, the homomorphisms $\chi: G \to \mathbb{C}$ such that $|\chi(g)| = 1$ for all $g \in G$. Then one has to give $\hat{G}$ a topology that makes it into a locally compact group. Doing this is not obvious, but it turns out that as a topological space, $\hat{G}$ is homeomorphic to the space $X(L^1(G))$ of characters of the algebra $L^1(G)$, the set of algebra homomorphisms $\chi: L^1(G) \to \mathbb{C}$. Here $L^1(G)$ is the commutative algebra whose multiplication operation is convolution, and it can be shown that $X(L^1(G))$ is locally compact.

There is an even deeper connection between the space $X(L^1(G))$ of algebra characters and the Fourier transform. Indeed the Fourier cotransform on the commutative algebra $L^1(G)$ is the Gelfand transform on $L^1(G)$. For any $f \in L^1(G)$, the Gelfand transform $\mathcal{G}_f$ is a function defined on the set $X(L^1(G))$ of characters of $L^1(G)$ by

$$\mathcal{G}_f(\zeta) = \zeta(f), \quad \zeta \in X(L^1(G)).$$

In general, the algebra $L^1(G)$ is a complete normed algebra (a Banach algebra), but it does not have a multiplicative unit. However, the space $\mathcal{M}_{reg,\mathbb{C}}(G)$ of complex regular Borel measures on $G$, simply denoted $\mathcal{M}^1(G)$, is a unital Banach algebra, with the norm $\|\mu\| = |\mu|(G)$ defined in Definition 5.9, with the convolution of measures as multiplication (see Definition 6.19), and with the Dirac measure $\delta_1$ as multiplicative unit. Furthermore, $L^1(G)$ can be identified with a subalgebra of $\mathcal{M}^1(G)$, using the embedding $f \mapsto fd\lambda$ given by Proposition 5.31.

Therefore we are led to the study of algebras and normed algebras, in particular to complete normed algebras, called Banach algebras. If an algebra $A$ is commutative, then Gelfand had the idea to realize $A$ as a set of complex-valued functions on the set of characters.
Let $X(A)$ be the set of functions from $X(A)$ to $\mathbb{C}$. The map $G: A \to C^{X(A)}$, called Gelfand transform, is defined as follows: for every $a \in A$,

$$G_a(\chi) = \chi(a), \quad \chi \in X(A).$$

If $A$ is a commutative Banach algebra with multiplicative identity element $e$, then the range of the Gelfand transform $G_a$, the set $\{\chi(a) | \chi \in X(A)\}$, is equal to the spectrum $\sigma(a)$ of $a$, namely the set of complex numbers $\lambda$ such that $\lambda e - a$ is not invertible in $A$. The spectrum $\sigma(a)$ of $a$ is a generalization of the notion of eigenvalue of a linear map.

The study of algebras and normed algebras focuses on three concepts:

(1) The notion of spectrum $\sigma(a)$ of an element $a$ of an algebra $A$.

(2) If $A$ is a commutative algebra, the notion of character, and the space $X(A)$ of characters of $A$.

(3) If $A$ is a commutative algebra, the notion of Gelfand transform, $G: A \to C(X(A); \mathbb{C})$.

In Section 8.1, we define algebras and algebras with a multiplicative unit, called unital algebras. We also define normed algebras (without a multiplicative unit), and Cauchy-complete algebras, called Banach algebras. In Section 8.2, we show that every non-unital $K$-algebra $A$ can be embedded into a unital $K$-algebra $\tilde{A}$. We also define the notion of quotient of a normed algebra by an ideal.

If $E$ is an infinite-dimensional vector space, and if $f: E \to E$ is a linear map, the definition of an eigenvalue $\lambda$ of a linear map $f$ used in finite-dimension in terms of the existence of nonzero vector $u$ such that $f(u) = \lambda u$ no longer works, because a non-invertible linear map may still be injective. Consequently, there may be some complex number $\lambda$ such that $\lambda e - f$ is not invertible, yet there is no nonzero vector $u$ such that $f(u) = \lambda u$. This suggests defining a spectral value as a complex number $\lambda$ such that $\lambda e - f$ is not invertible. Then it is easy step to generalize this definition to any unital algebra $A$ with multiplicative identity $e$. Given any $a \in A$, viewed as a sort of generalized linear map, a number $\lambda \in \mathbb{C}$ is a spectral value for $a$ if $\lambda e - a$ is not invertible.

The notion of spectrum is defined and investigated in Section 8.3. The complement $\mathbb{C} - \sigma(a)$ of $\sigma(a)$ is called the resolvent set of $a$, and for any fixed $a \in A$, the function $R(a, \lambda)$ (defined on the set $\mathbb{C} - \sigma(a)$) given by $R(a, \lambda) = (\lambda e - a)^{-1}$ is called the resolvent of $a$. In general, there is no guarantee that $\sigma(a)$ is nonempty, but if $A$ is a unital Banach algebra, then $\sigma(a) \neq \emptyset$.

Next we define the notion of character of a commutative unital algebra $A$. This is a homomorphism $\chi: A \to \mathbb{C}$. The space of characters on $A$ is denoted by $X(A)$. A first connection between the notion of spectrum and the notion of character is that for any $a \in A$ and any $\chi \in X(A)$, we have $\chi(a) \in \sigma(a)$. 
The third key notion, the Gelfand transform, is then defined. The map $G: A \rightarrow \mathbb{C}^X(A)$, called \textit{Gelfand transform}, is defined as follows: for every $a \in A$,

$$G_a(\chi) = \chi(a), \quad \chi \in X(A).$$

The function $G_a$ (or $G(a)$) is called the \textit{Gelfand transform} of $a$. If we give $X(A)$ the topology of pointwise convergence, then $G$ becomes a continuous map from $A$ to the space $\mathcal{C}(X(A); \mathbb{C})$ of continuous functions on $X(A)$.

In order to obtain sharper results about spectra and characters, we consider unital Banach algebras in Section 8.4.

Theorem 8.10 states that if $A$ is a unital Banach algebra, then for any $a \in A$, the spectrum $\sigma(a)$ is nonempty, compact, and contained in the closed ball of radius $\|a\|$. Furthermore, the map $\lambda \mapsto R(a, \lambda)$ is holomorphic.

We prove the Gelfand–Mazur theorem (Theorem 8.11), which says that if a unital Banach algebra $A$ is a (possibly noncommutative) field, then $A$ is isometrically isomorphic to $\mathbb{C}$.

Certain notions defined for complex matrices can be generalized to algebras and normed algebras. In particular, if $A$ is a normed algebra, for any $a \in A$, the number $\rho(a) = \inf_n \|a^n\|^{1/n}$ converges and is called the \textit{spectral radius} of $a$. Then if $A$ is a unital Banach algebra, we have

$$\rho(a) = \sup\{\lambda \mid \lambda \in \sigma(a)\}.$$

If $A$ is a unital Banach algebra, then $X(A)$ is compact and Hausdorff. If $A$ is a nonunital Banach algebra, then $X(A)$ is locally compact.

Properties of the Gelfand transform holding for unital Banach algebras are proven in Section 8.5.

Let $A$ be a commutative unital Banach algebra.

(1) For every $a \in A$, the range of $G_a$ is equal to the spectrum $\sigma(a)$ of $a$; that is,

$$G_a(X(A)) = \{\chi(a) \mid \chi \in X(A)\} = \sigma(a).$$

(2) The Gelfand transform $G: A \rightarrow \mathcal{C}(X(A); \mathbb{C})$ is a continuous homomorphism such that $\|G_a\|_\infty = \rho(a) \leq \|a\|$, and $G_e$ is the constant function 1.

(3) An element $a \in A$ is invertible iff $G_a$ does not vanish on $X(A)$.

If $A$ is a commutative nonunital Banach algebra, then the Gelfand transform is a homomorphism $G: A \rightarrow \mathcal{C}_0(X(A); \mathbb{C})$.

We can characterize when the Gelfand transform is an isometry and when it is injective in terms of its radical. Given a commutative algebra $A$, the \textit{radical} of $A$, rad $A$, is the intersection of all maximal ideals in $A$. 
Let $A$ be a commutative unital Banach algebra. We have $\text{Ker} \mathcal{G} = \text{rad} A$, and the Gelfand transform $\mathcal{G}: A \to C(\mathcal{X}(A); \mathbb{C})$ is injective iff the radical of $A$ is trivial; that is, $\text{rad} A = (0)$.

In Section 8.6, we consider special algebras equipped with an involution, in particular, $C^*$-algebras. If $A$ is an algebra, an involution is a bijection $a \mapsto a^*$ that satisfies the equations of the conjugate-transpose $A^* = (A^\top)$ on complex matrices.

If $A$ is a normed algebra, then an involutive normed algebra is an algebra with an involution $a \mapsto a^*$ satisfying the extra axiom
\[ \|a\| = \|a^*\|, \quad \text{for all } a \in A. \] (i)

A $C^*$-algebra is a Banach algebra satisfying the axiom
\[ \|a\|^2 = \|a^*a\|, \quad \text{for all } a \in A. \] (C*)

A $C^*$-algebra automatically satisfies axiom (i). The normed algebras $\mathcal{M}^1(G)$ and $L^1(G)$ are involutive algebras, but in general, not $C^*$-algebras. The main example of a $C^*$-algebra is the algebra $\mathcal{L}(H)$ of continuous linear maps on a complex Hilbert space $H$.

If $A$ is an involutive algebra, an element $a \in A$ is hermitian if $a = a^*$, normal if $aa^* = a^*a$, unitary if $aa^* = a^*a = e$. As in the case of matrices, if $A$ is a unital $C^*$-algebra, for every $a \in A$, if $a$ is hermitian then $\sigma(a) \subseteq \mathbb{R}$, and if $a$ is unitary then $\sigma(a) \subseteq \mathbb{T} = U(1)$.

In Section 8.7, we consider characters and the Gelfand transform in a $C^*$-algebra. Let $A$ be a commutative unital $C^*$-algebra. Then for any character $\chi \in \mathcal{X}(A)$, we have
\[ \chi(a^*) = \overline{\chi(a)}, \quad \text{for all } a \in A, \] or equivalently $\chi(a) = \overline{\chi(a^*)}$. We say that the characters of $A$ are hermitian.

The main theorem of the theory of commutative unital $C^*$-algebras, due to Gelfand and Naimark, states that every commutative unital $C^*$-algebra can be viewed as the algebra of continuous functions on a compact space, namely its space of characters $\mathcal{X}(A)$.

More precisely, let $A$ be a commutative unital $C^*$-algebra. Then the Gelfand transform $\mathcal{G}: A \to C(\mathcal{X}(A); \mathbb{C})$ is an isometric isomorphism between $A$ and $C(\mathcal{X}(A); \mathbb{C})$ (and so $\|\mathcal{G}_a\| = \|a\| = \rho(a)$ for all $a \in A$). Furthermore the Gelfand maps $\mathcal{G}_a$ are hermitian.

The Gelfand–Naimark theorem is used to prove the Plancherel–Godement theorem (see Section 10.5, Theorem 10.37), and some representation theory results in harmonic analysis; see Dieudonné [20].

The spectral theory of $C^*$-algebras is the key machinery used to develop generalizations of the spectral theorems for normal matrices to bounded (and unbounded) operators of various kinds on a Hilbert space. A condensed presentation of these spectral theorems is given in
Folland [31] (Chapter 1, Section 1.4). An extensive treatment of spectral theorems is given in Rudin [67], and in Lax [54].

In Section 8.8, given an involutive Banach algebra $A$, we construct a $C^*$-algebra $\text{St}(A)$ and an involutive homomorphism $j: A \to \text{St}(A)$ that satisfy a universal mapping condition with respect to homomorphisms of $A$ into a $C^*$-algebra. For every involutive homomorphism $\varphi: A \to B$ of $A$ into a $C^*$-algebra $B$, there is a unique involutive homomorphism $\overline{\varphi}: \text{St}(A) \to B$ such that

$$\varphi = \overline{\varphi} \circ j,$$

as shown in the following commutative diagram:

$$
\begin{array}{ccc}
A & \xrightarrow{j} & \text{St}(A) \\
& \varphi \searrow & \downarrow \overline{\varphi} \\
& & B.
\end{array}
$$

It is also possible to characterize the set of characters $\mathcal{X}(\text{St}(A))$ of $\text{St}(A)$. Let $A$ be an involutive Banach algebra. Then the map $\mathcal{X}(j)$ is a homeomorphism of the set of characters $\mathcal{X}(\text{St}(A))$ onto the subspace $H$ of hermitian characters in $\mathcal{X}(A)$; that is, the characters $\chi: A \to \mathbb{C}$ such that $\chi(a) = \overline{\chi(a^*)}$ for all $a \in A$.

The above result applies to the involutive Banach algebra $L^1(G)$ associated with a locally compact group $G$. In general, $L^1(G)$ is not a $C^*$-algebra. Thus we can form the enveloping $C^*$-algebra $\text{St}(L^1(G))$ of $L^1(G)$, denoted $\text{St}(G)$. Remarkably, the canonical map $j$ is injective. As a consequence, it can be shown that there is a homeomorphism between $\mathcal{X}(\text{St}(G))$ and $\mathcal{X}(L^1(G))$.

More generally, the material of this chapter (spectra, algebra characters, the Gelfand transform), are covered in Bourbaki [8], Dieudonné [22], Folland [31], Lang [52], Lax [54], Rudin [67], and Schwartz [70], the most complete presentations being Rudin [67] and Bourbaki [8].

### 8.1 Normed Algebras, Banach Algebras

Before defining normed algebras, let us recall the definition of an algebra over a field $K$. Since the algebras that we will be dealing with do not always have a multiplicative unit, for example $L^1(G)$ with the convolution as product (where $G$ is a locally compact group equipped with a left Haar measure), we use the following definition.

**Definition 8.1.** Given a field $K$, a $K$-algebra is a $K$-vector space $A$ together with a bilinear operation $\ast: A \times A \to A$, called *multiplication*, which is associative; that is,

$$(a \ast b) \ast c = a \ast (b \ast c) \quad \text{for all} \ a, b, c \in A.$$
A $K$-algebra $A$ is **unital** if there is a multiplicative identity element $1 \neq 0$ so that

$$1 \ast a = a \ast 1 = a,$$

for all $a \in A$.

A $K$-algebra $A$ is **commutative** if

$$a \ast b = b \ast a,$$

for all $a, b \in A$.

If $A$ is a unital algebra, an element $a \in A$ is **invertible** if there is some $b \in A$ such that

$$a \ast b = b \ast a = 1.$$  

Given two $K$-algebras $A$ and $B$, a $K$-algebra homomorphism $h: A \to B$ is a linear map such that

$$h(a \ast_A b) = h(a) \ast_B h(b)$$

for all $a, b \in A$.

If $A$ and $B$ are unital, we also require that $h(1_A) = 1_B$.

For example, the ring $M_n(K)$ of all $n \times n$ matrices over a field $K$ is a unital $K$-algebra with multiplicative identity element $1 = I_n$.

There is an obvious notion of **subalgebra** and **ideal** of a $K$-algebra:

**Definition 8.2.** A subalgebra $B$ of a $K$-algebra $A$ is linear subspace closed under multiplication; that is, $x \ast y \in B$ for all $x, y \in B$. If $A$ is unital, we require that $1 \in B$. An ideal $\mathfrak{A} \subseteq A$ is a linear subspace of $A$ that is also a two-sided ideal with respect to multiplication in $A$, which means that $x \ast a \ast y \in \mathfrak{A}$ for all $a \in \mathfrak{A}$ and all $x, y \in A$. An ideal $\mathfrak{A}$ is a proper ideal if $\mathfrak{A} \neq A$. An ideal $\mathfrak{m}$ is a maximal ideal if it is proper and if there is no proper ideal $\mathfrak{n}$ such that $\mathfrak{m} \subseteq \mathfrak{n}$ and $\mathfrak{m} \neq \mathfrak{n}$.

If $A$ is nonunital, then an ideal is a subalgebra, but the converse is false in general. If $A$ is unital with identity element $1$, then an ideal $\mathfrak{A}$ is proper iff $1 \notin \mathfrak{A}$, iff $\mathfrak{A}$ contains no invertible element.

*From now on we assume that the field $K$ is the field $\mathbb{C}$ of complex numbers.*

**Definition 8.3.** A **normed algebra** is a $\mathbb{C}$-algebra $A$ endowed with a norm $\| \|: A \to \mathbb{R}_+$ satisfying the inequality

$$\|x \ast y\| \leq \|x\| \|y\|$$

for all $x, y \in A$.

If $A$ is unital with identity $1$, then we require that

$$\|1\| = 1.$$  

If the underlying normed vector space of $A$ is complete, then we say that $A$ is a **Banach algebra**.
The inequality \( \|x \star y\| \leq \|x\| \|y\| \) shows that the multiplication operation \( \star \) is continuous. It is a generalization of the inequality characterizing matrix norms. Intuitively, a normed algebra can be viewed as a sort of generalized space of matrices.

For simplicity of notation we write \( xy \) instead of \( x \star y \), unless confusion arises. We also use the notation \( e \) for the multiplicative identity of \( A \) instead of \( 1 \).

**Example 8.1.**

(1) If \( E \) is a normed vector space, then the space \( \mathcal{L}(E) \) of continuous linear maps \( f : E \to E \) is a unital algebra under composition, with \( \text{id}_E \) as identity element. It is a normed algebra under the operator norm,
\[
\|f\| = \sup\{\|f(x)\| \mid \|x\| = 1\}.
\]

If \( E \) is a Banach space (that is, a complete normed vector space), then \( \mathcal{L}(E) \) is a Banach algebra.

(2) Let \( X \) be a topological space. Then the space \( \mathcal{C}_b(X; \mathbb{C}) \) of bounded continuous functions on \( X \) is a commutative unital Banach algebra, with the norm \( \|\cdot\|_\infty \); functions are multiplied pointwise; that is, \( (fg)(x) = f(x)g(x) \) for all \( x \in X \). The multiplicative identity is the constant function \( 1 \). If \( X \) is a (Hausdorff) locally compact space, then the space \( \mathcal{C}_0(X; \mathbb{C}) \) is a commutative Banach algebra and an ideal in \( \mathcal{C}_b(X; \mathbb{C}) \), but it not unital unless \( X \) is compact. The space of \( \mathcal{K}_C(X) \) of continuous functions with compact support is a commutative normed algebra and an ideal in \( \mathcal{C}_b(X; \mathbb{C}) \), but it is not complete (and not unital unless \( X \) is compact).

(3) Let \( G \) be a locally compact group. The space \( \mathcal{M}^{1}_{\text{reg}, \mathbb{C}}(G) \) of complex regular Borel measures on \( G \), simply denoted \( \mathcal{M}^{1}(G) \), is a unital Banach algebra, with the norm \( \|\mu\| = |\mu|(G) \) defined in Definition 5.9, with the convolution of measures as multiplication (see Definition 6.19), and with the Dirac measure \( \delta_1 \) as multiplicative unit (see Section 6.11).

(4) Let \( G \) be a locally compact group equipped with a left Haar measure \( \lambda \). The space \( L^1(G) \) (with the \( L^1 \)-norm) can be identified with a subalgebra of \( \mathcal{M}^1(G) \), using the embedding \( f \mapsto fd\lambda \) given by Proposition 5.31. The space \( L^1(G) \) is a Banach algebra with the convolution of functions as multiplication, but it is not unital unless \( G \) is discrete.

(5) As a special case of (4), let \( G = \mathbb{Z} \), in which case \( L^1(G) \) is the set of all sequences \( x = (x_m)_{m \in \mathbb{Z}} \) with \( x_m \in \mathbb{C} \), such that \( \sum_{m \in \mathbb{Z}} |x_m| < \infty \). This space is also denoted \( l^1(\mathbb{Z}) \). The convolution product \( x \star y \) of \( x = (x_m) \) and \( y = (y_m) \) is given by
\[
(x \star y)_m = \sum_{p \in \mathbb{Z}} x_p y_{m-p}.
\]
and the norm by \( \|x\| = \sum_{m \in \mathbb{Z}} |x_m| \). This is a commutative unital Banach algebra with identity element \( e_0 \) such that \( e_0(0) = 1 \) and \( e_0(m) = 0 \) for all \( m \neq 0 \).

Define \( \delta^m \) by \( \delta^m(m) = 1 \) and \( \delta^m(k) = 0 \) if \( k \neq m \). It is easy to see that \( \delta^0 = e_0 \), and \( \delta^m \ast \delta^n = \delta^{m+n} \), so \( (\delta^m)^{-1} = \delta^{-m} \). Then we see that for any \( x = (x_m)_{m \in \mathbb{Z}} \in l^1(\mathbb{Z}) \), we have \( x = \sum_{m \in \mathbb{Z}} x_m \delta^m \), which shows that \( l^1(\mathbb{Z}) \) is generated by \( \delta^1 \) and \( \delta^{-1} = (\delta^1)^{-1} \).

(6) For any \( c = (c_m) \in l^1(\mathbb{Z}) \), let \( \varphi_c : \mathbb{T} \to \mathbb{C} \) be the function given by
\[
\varphi_c(e^{i\theta}) = \sum_{m \in \mathbb{Z}} c_m e^{im\theta}, \quad \theta \in \mathbb{R}/(2\pi \mathbb{Z}).
\]
The series defining \( \varphi_c \) is absolutely and uniformly convergent. Thus we deduce that
\[
c_m = \frac{1}{2\pi} \int_{-\pi}^{\pi} \varphi_c(e^{i\theta}) e^{-im\theta} \, d\theta,
\]
the \( m \)th Fourier coefficient of \( \varphi_c \); so the map \( \varphi : c \mapsto \varphi_c \) is an injection from \( l^1(\mathbb{Z}) \) into the commutative unital Banach algebra of absolutely and uniformly convergent Fourier series (under convolution of functions in \( L^1(\mathbb{T}) \)). The identity is the constant function 1.

(7) Let \( G \) be a compact group equipped with a Haar measure. In this case \( L^2(G) \subset L^1(G) \), and both are unital Banach algebras (\( L^2(G) \) with the \( L^2 \)-norm and \( L^1(G) \) with the \( L^1 \)-norm). The Banach algebra \( L^2(G) \) is also a Hilbert space. This yields even more structure on \( L^2(G) \) which turns out to be a Hilbert algebra, as discussed later. The significance of this fact is that there is a structure theorem about separable Hilbert algebras (they can be expressed as the Hilbert sum of special kinds of ideals), and this structure theorem is a key ingredient in the proof of the Peter–Weyl theorem, the fundamental theorem about the the structure of \( L^2(G) \), and of the irreducible unitary representations of the compact group \( G \).

(8) Let \( C^n([0,1]) \) be the algebra of functions \( f : [0,1] \to \mathbb{C} \) having a continuous derivative \( f^{(k)} \) for \( k = 1, \ldots, n \), under pointwise addition and multiplication. If we let
\[
\|f\| = \sum_{k=0}^{n} \frac{1}{k!} \sup_{0 \leq t \leq 1} |f^{(k)}(t)|,
\]
then we can check that this is indeed a norm, and that \( \|fg\| \leq \|f\| \|g\| \). With this norm, \( C^n([0,1]) \) is a commutative unital Banach algebra.

The following simple proposition is left as an exercise.

**Proposition 8.1.** Let \( A \) be a normed algebra. The closure of a subalgebra of \( A \) is a subalgebra of \( A \), and the closure of any ideal of \( A \) is an ideal in \( A \).
The following proposition implies that the set of invertible elements in a unital Banach algebra is open.

**Proposition 8.2.** Let $A$ be a unital Banach algebra.

1. For any element $a \in A$, if $\|a\| < 1$, then $e - a$ is invertible, the series $\sum_{n=0}^{\infty} a^n$ converges absolutely,

$$ (e - a)^{-1} = \sum_{n=0}^{\infty} a^n = e + a + a^2 + \cdots + a^n + \cdots, $$

and

$$ \|(e - a)^{-1} - e - a\| \leq \frac{\|a\|^2}{1 - \|a\|}. $$

2. Let $a \in A$ be an invertible element. Then for any $h \in A$, if $\|h\| < (1/2) \|a^{-1}\|^{-1}$, then $a + h$ is invertible, and

$$ \|(a + h)^{-1} - a^{-1} + a^{-1}ha^{-1}\| \leq 2 \|a^{-1}\|^3 \|h\|^2. $$

**Proof.** (1) Since $\|xy\| \leq \|x\| \|y\|$, we have $\|a^n\| \leq \|a\|^n$, and so

$$ 1 + \|a\| + \cdots + \|a^n\| \leq 1 + \|a\| + \cdots + \|a\|^n = \frac{1 - \|a\|^{n+1}}{1 - \|a\|}. $$

Since $\|a\| < 1$, the geometric series converges, and so does the series $(\sum_n \|a^n\|)$. Let

$$ S_n = e + a + \cdots + a^n. $$

For all $n \geq m$, we have

$$ \|S_n - S_m\| = \|a^{m+1} + \cdots + a^n\| \leq \|a^{m+1}\| + \cdots + \|a^n\|. $$

Since the sequence $\sum_{k=m}^{n} \|a^k\|$ converges, it is a Cauchy sequence, thus for every $\epsilon > 0$, we can find $p > 0$ so that for all $m \geq n \geq p$ we have

$$ \|a^{m+1}\| + \cdots + \|a^n\| \leq \epsilon, $$

thus the sequence $(S_m)$ is also a Cauchy sequence. Since $A$ is complete, this sequence has a limit, say $b$. Since

$$ S_n(e - a) = (e - a)S_n = e - a^{n+1}, $$

and since $\lim_{n \to 0} a^{n+1} = 0$ (because $\|a^{n+1}\| \leq \|a\|^{n+1}$ and $\|a\| < 1$), by continuity of multiplication, we get

$$ b(e - a) = (e - a)b = e, $$

which means that $b$ is the inverse of $e - a$. 

---

8.1. NORMED ALGEBRAS, BANACH ALGEBRAS

275
Since \( b = \sum_{n=0}^{\infty} a^n \), we have
\[
\|b - e - a\| = \| a^2 + \cdots + a^n + \cdots \| \leq \sum_{n=2}^{\infty} \|a\|^n = \frac{\|a\|^2}{1 - \|a\|}.
\]

(2) Since \( a \) is invertible,
\[
a + h = a(e + a^{-1}h),
\]
and since \( \|h\| < (1/2) \|a^{-1}\|^{-1} \), we get
\[
\|a^{-1}h\| \leq \|a^{-1}\| \|h\| \leq \|a^{-1}\| (1/2) \|a^{-1}\|^{-1} = 1/2.
\]
Applying (1) to \(-a^{-1}h\), since \( \|a^{-1}h\| = \|a^{-1}h\| < 1/2 \), we deduce that \( e + a^{-1}h \) is invertible. Therefore \( a + h = a(e + a^{-1}h) \) is invertible. We also have \( (a + h)^{-1} = (a(e + a^{-1}h))^{-1} = (e + a^{-1}h)^{-1}a^{-1} \), and so
\[
(a + h)^{-1} - a^{-1} + a^{-1}ha^{-1} = ((e + a^{-1}h)^{-1}e + a^{-1}h)a^{-1} = ((e - (-a^{-1}h))^{-1}e - (-a^{-1}h))a^{-1},
\]
which by (1) and the fact that \( \|a^{-1}h\| < 1/2 \) yields
\[
\|(a + h)^{-1} - a^{-1} + a^{-1}ha^{-1}\| = \|(e - (-a^{-1}h))^{-1}e - (-a^{-1}h))a^{-1}\|
\leq \frac{\|a^{-1}h\|^2 \|a^{-1}\|}{1 - \|a^{-1}h\|}
\leq 2 \|a^{-1}h\|^2 \|a^{-1}\|
\leq 2 \|a^{-1}\|^3 \|h\|^2,
\]
as claimed. \(\square\)

As a corollary of Proposition 8.2, we obtain the following result.

**Proposition 8.3.** Let \( A \) be a unital Banach algebra. The set \( G(A) \) of invertible elements of \( A \) is open and contains the open ball of center \( e \) given by \( \{a \in A \mid \|a - e\| < 1\} \). The inversion map \( \iota: G(A) \to G(A) \) given by \( \iota(a) = (a^{-1})^{-1} \) is differentiable and its derivative is given by \( d\iota(a)(h) = -a^{-1}ha^{-1} \) for all \( a \in G(A) \) and all \( h \in A \). Consequently, \( \iota \) is continuous.

### 8.2 Two Algebra Constructions

Let \( K \) be an arbitrary field. Every non-unital \( K \)-algebra \( A \) can be embedded into a unital \( K \)-algebra \( \tilde{A} \) as follows.

Consider the vector space \( \tilde{A} = K \times A \), and define a multiplication by
\[
(\lambda, a)(\mu, b) = (\lambda\mu, \lambda b + \mu a + ab).
\]

(\text{\textit{*}})
8.2. TWO ALGEBRA CONSTRUCTIONS

It is easily verified that $\tilde{A}$ with this multiplication is a $K$-algebra, and that $e = (1, 0)$ is a multiplicative unit. The algebra $A$ is embedded into $\tilde{A}$ using the map $a \mapsto (0, a)$, and it is immediately verified that $A$ is an ideal in $\tilde{A}$ (in fact, a maximal ideal). Since $(\lambda, a)(0, b) = (0, \lambda b + ab)$, and $e = (1, 0)$, no element in $A$ is invertible.

Since $(\lambda, a) = \lambda(1, 0) + (0, a)$, we can write $(\lambda, a) = \lambda e + a$. With this notation the multiplication of $(\lambda, a)$ and $(\mu, b)$ becomes

$$(\lambda e + a)(\mu e + b) = \lambda \mu e + \lambda b + \mu a + ab,$$

which shows that the formula $(\ast)$ defining multiplication is not so strange after all.

If $A$ already has a multiplicative identity $\epsilon$, we immediately check that $e - \epsilon = (1, -\epsilon)$ has the following properties:

$$(e - \epsilon)^2 = e - \epsilon, \quad (e - \epsilon)A = A(e - \epsilon) = \{(0, 0)\}.$$

Then $K(e - \epsilon)$ is a unital algebra with identity $e - \epsilon$ (in fact, a field). Let $(K(e - \epsilon)) \times A$ be the product algebra of $K(e - \epsilon)$ and $A$ under componentwise addition and multiplication. It is a unital algebra with identity $(e - \epsilon, \epsilon)$. To avoid confusion between the two multiplications, for the rest of this paragraph, let use denote multiplication in $\tilde{A}$ by $\star$. Define the map $\varphi: \tilde{A} \to (K(e - \epsilon)) \times A$ by

$$\varphi(\lambda, a) = (\lambda(e - \epsilon), \lambda e + a).$$

We check immediately that $\varphi$ is a linear map. We have

$$\varphi(\lambda, a)\varphi(\mu, b) = (\lambda(e - \epsilon), \lambda e + a)(\mu(e - \epsilon), \mu e + b)$$
$$= (\lambda \mu(e - \epsilon), (\lambda e + a)(\mu e + b))$$
$$= (\lambda \mu(e - \epsilon), \lambda \mu e + \lambda b + \mu a + ab),$$

and

$$\varphi((\lambda, a) \star (\mu, b)) = \varphi(\lambda \mu, \lambda b + \mu a + ab)$$
$$= (\lambda \mu(e - \epsilon), \lambda \mu e + \lambda b + \mu a + ab),$$

so

$$\varphi((\lambda, a) \star (\mu, b)) = \varphi(\lambda, a)\varphi(\mu, b),$$

which shows that $\varphi$ is an algebra homomorphism. The map $\varphi$ is obviously injective, and it is surjective because

$$\varphi(\lambda, a - \lambda \epsilon) = (\lambda(e - \epsilon), \lambda e + a - \lambda \epsilon) = (\lambda(e - \epsilon), a).$$

In summary we have the following result.

**Proposition 8.4.** If $A$ is a unital $K$-algebra, then we have an algebra isomorphism $\varphi: \tilde{A} \to (K(e - \epsilon)) \times A$. 

If $A$ is a normed algebra, it is also easy to check that the map
\[ \|(\lambda, a)\| = |\lambda| + \|a\| \]
makes $\tilde{A}$ into a unital normed algebra, and if $A$ is a Banach algebra, then so is $\tilde{A}$.

Later on we will encounter special kinds of Banach algebras called $C^*$-algebras. In this case, in order to make $\tilde{A}$ into a $C^*$-algebra, a different norm is needed. It turns out that
\[ \|(\lambda, a)\| = \sup \{|\lambda b + ab| : b \in A, \|b\| \leq 1\} \]
makes $\tilde{A}$ into a $C^*$-algebra, and agrees with the original norm on $A$. It is the only norm that does so; see Folland [31] (Chapter 1, Section 4), or Bourbaki [8] (Chapter 1, Section 6, No. 3).

In many concrete cases, the general construction of $\tilde{A}$ is not needed. For example, since $L^1(G)$ is a subalgebra of $M^1(G)$, we can use $L^1(G) \oplus \mathbb{C}\delta_1$ as the completion of $L^1(G)$ into a unital algebra.

The following construction will be used in Section 8.4. If $A$ is a normed algebra and if $m$ is a (topologically) closed ideal in $A$, then the quotient $A/m$ is an algebra which can be made into a normed algebra as follows. If $\pi: A \to A/m$ is the quotient map, then for any $a \in A$, let
\[ \|\pi(a)\| = \inf \{\|a + z\| : z \in m\}. \]

**Proposition 8.5.** Let $A$ be a normed algebra and let $m$ be a closed ideal in $A$. The map $\|\|: A/m \to \mathbb{R}_+$ given by
\[ \|\pi(a)\| = \inf \{\|a + z\| : z \in m\}, \quad a \in A, \]
is a norm on $A/m$. If $A$ is a Banach algebra, then so is $A/m$, and if $A$ is unital, then so is $A/m$.

**Proof sketch.** It is proven in Schwartz [70] (Chapter II, Section 2) that the map $\|\|: A/m \to \mathbb{R}_+$ is a norm if and only if $m$ is closed.

First, observe that by definition of $\|\pi(a)\|$, we have $\|\pi(a)\| \leq \|a\|$, so $\pi$ is continuous. It is also easy to check that $\pi$ is a homomorphism. Let us check the triangle inequality and the fact that if $\|\pi(a)\| = 0$, then $a \in m$, which means that $\pi(a) = 0$.

By definition of a greatest lower bound, given any $\epsilon > 0$, for any two elements $x, y \in A$, there exist some $u, v \in m$ such that
\[ \|x + u\| \leq \|x\| + \epsilon/2 \]
\[ \|y + v\| \leq \|y\| + \epsilon/2. \]
But $\pi(x + y + u + v) = \pi(x + y)$, so
\[ \|\pi(x + y)\| \leq \|x + y + u + v\| \leq \|x + u\| + \|y + v\| \leq \|\pi(x)\| + \|\pi(y)\| + \epsilon. \]
Since $\epsilon$ is arbitrary, we get
\[
\|\pi(x + y)\| = \|\pi(x) + \pi(y)\| \leq \|\pi(x)\| + \|\pi(y)\|.
\]

We leave the verification that
\[
\|\lambda\pi(x)\| = |\lambda| \|\pi(x)\|
\]
as an exercise.

Assume that $\|\pi(a)\| = 0$. We want to prove that $a \in \mathfrak{m}$. For every $n > 0$, there is some $u_n \in \mathfrak{m}$ such that $\|a + u_n\| < \|\pi(a)\| + 1/n = 1/n$. Consequently the sequence $(a + u_n)$ converges to 0, which implies that the sequence $(u_n)$ converges to $-a$. Since $u_n \in \mathfrak{m}$ for all $n \geq 1$ and since $\mathfrak{m}$ is closed, $-a \in \mathfrak{m}$. But $\mathfrak{m}$ is a vector space so $a \in \mathfrak{m}$, that is, $\pi(a) = 0$, as claimed.

We also have
\[
\|(x + u)(y + v)\| \leq \|x + u\| \|y + v\| \leq (\|\pi(x)\| + \epsilon/2)(\|\pi(y)\| + \epsilon/2)
\]
Since $\epsilon$ is arbitrary, by definition of $\|\pi(x)\pi(y)\| = \|\pi(xy)\|$, since $\pi(xy) = \pi((x+u)(y+v)) = \pi(xy + xv + uy + uv)$ (because $\mathfrak{m}$ is an ideal), we deduce that
\[
\|\pi(x)\pi(y)\| \leq \|\pi(x)\| \|\pi(y)\|.
\]

If $e$ is the unit of $A$, then $\pi(e)$ is the unit of $A/\mathfrak{m}$. By definition $\|\pi(e)\| \leq \|e\| = 1$. We also have
\[
\|\pi(e)\| = \|(\pi(e))^2\| \leq (\|\pi(e)\|)^2,
\]
which implies that $\|\pi(e)\| \geq 1$, so $\|\pi(e)\| = 1$.

It remains to prove that if $A$ is complete, then $A/\mathfrak{m}$ is also complete. Let $(\pi(a_n))$ be a Cauchy sequence in $A/\mathfrak{m}$. Taking a subsequence we may assume that
\[
\|\pi(a_n) - \pi(a_{n-1})\| < \frac{1}{2^n} \quad \text{for all } n \geq 1.
\]

We construct inductively a sequence $(x_n)$ in $A$ such that $\pi(x_n) = \pi(a_n)$ and
\[
\|x_n - x_{n-1}\| < \frac{1}{2^n} \quad \text{for all } n \geq 1.
\]

Initially $x_1 = a_1$, and if $x_1, \ldots, x_n$ have been defined, since
\[
\|\pi(a_{n+1}) - \pi(a_n)\| = \|\pi(a_{n+1} - a_n)\| < \frac{1}{2^{n+1}}
\]
there is some $y \in \mathfrak{m}$ such that
\[
\|a_{n+1} - a_n + y\| < \frac{1}{2^{n+1}}.
\]
Then we let \( x_{n+1} = x_n + a_{n+1} - a_n + y \), which works since
\[
\pi(x_n + a_{n+1} - a_n + y) = \pi(x_n) + \pi(a_{n+1}) - \pi(a_n) + \pi(y) = \pi(a_{n+1}) - \pi(a_n) = \pi(a_{n+1}).
\]
The sequence \((x_n)\) is a Cauchy sequence in \(A\), and since \(A\) is complete, this sequence converges to a limit \(a \in A\). But \(\pi\) is continuous so the sequence \((\pi(a_n)) = (\pi(x_n))\) converges to \(\pi(a)\), as desired. 

8.3 Spectrum, Characters, Gelfand Transform, I; For an Algebra

Let \(E\) be a finite-dimensional vector space over the field \(\mathbb{C}\), and let \(f : E \to E\) be a linear map. Recall that some \(\lambda \in \mathbb{C}\) is an eigenvalue of \(f\) if there is some nonzero vector \(u \in E\), called an eigenvector associated with \(\lambda\), such that
\[
f(u) = \lambda u. \tag{*}
\]

Equation \((*)\) holds iff
\[
(\lambda \text{id} - f)(u) = 0
\]
iff \(\lambda \text{id} - f\) is not injective. But since \(E\) is finite-dimensional, a linear map is injective iff it is invertible, thus \(\lambda \text{id} - f\) is not injective iff \(\lambda \text{id} - f\) is not invertible. Therefore, \(\lambda \in \mathbb{C}\) is an eigenvalue for \(f\) iff \(\lambda \text{id} - f\) is not invertible, which we call the second definition of an eigenvalue. In turn, \(\lambda \text{id} - f\) is not invertible iff \(\det(\lambda \text{id} - f) = 0\), thus the eigenvalues of \(f\) are precisely the zeros of the polynomial \(\det(\lambda \text{id} - f) = 0\), viewed as a polynomial in \(\lambda\).

If \(E\) is infinite-dimensional, the situation is more complicated because an injective linear map may not be surjective, so \(\lambda \text{id} - f\) could be noninvertible, yet injective for some \(\lambda\), in which case there are no nonzero eigenvectors associated with \(\lambda\). Here is an example illustrating this situation.

Example 8.2. Let \(H\) be the Hilbert space \(L^2([0,1])\), and let \(E = \mathcal{L}(H)\), the Banach space of continuous linear maps from \(H\) to \(H\). Let \(T : H \to H\) be the operator in \(E\) given by
\[
T(h)(x) = xh(x), \quad \text{for all } h \in H \text{ and all } x \in [0,1].
\]

For any \(\lambda \in [0,1]\), observe that
\[
(\lambda \text{id}_E - T)(h)(x) = \lambda h(x) - xh(x) = (\lambda - x)h(x),
\]
so \((\lambda \text{id}_E - T)(h)(\lambda) = 0\). This shows that for every \(h \in H\), the function \((\lambda \text{id}_E - T)(h)\) vanishes at \(\lambda\). Thus \(\lambda \text{id}_E - T\) is not surjective, because the constant function 1 belongs to \(H\), but it is not in the range of \(\lambda \text{id}_E - T\) since it does not vanish anywhere. Therefore \(\lambda \text{id}_E - T\) is not invertible, so \(\lambda\) satisfies the second definition for being an eigenvalue of \(T\).
However, there is no nonzero eigenvector $h \in H$ associated with $\lambda$. Indeed, such a function $h \in H$ would satisfy the equation

$$(\lambda \text{id}_E - T)(h)(x) = 0 \quad \text{for all } x \in [0, 1],$$

that is,

$$(\lambda - x)h(x) = 0 \quad \text{for all } x \in [0, 1],$$

which implies that $h(x) = 0$ for all $x \in [0, 1]$, except for $x = \lambda$. This function is equal to the zero function almost everywhere, so in $H = L^2([0, 1])$, it is the zero function. This shows that $\lambda \text{id}_E - T$ is injective for all $\lambda \in [0, 1]$, but we also showed earlier that $\lambda \text{id}_E - T$ is not surjective.

In summary, the definition of an eigenvalue $\lambda$ of a linear map $f$ used in finite-dimension in terms of the existence of nonzero vector $u$ such that $f(u) = \lambda u$ no longer works in infinite dimension. However, if we redefine an eigenvalue of $T$ to be a complex number $\lambda$ such that $\lambda \text{id}_E - T$ is not invertible, then every $\lambda \in [0, 1]$ is an eigenvalue of $T$, even though $T$ has no eigenvectors.

Example 8.2 suggests a definition of the notion of eigenvalue for a linear map $f$ defined on an infinite-dimensional space $E$: it is a number $\lambda \in \mathbb{C}$ such that $\lambda \text{id} - f$ is not invertible. From this, it is an easy step to generalize this definition to any unital algebra $A$. Given any $a \in A$, viewed as a sort of generalized linear map, a number $\lambda \in \mathbb{C}$ is a spectral value for $a$ if $\lambda e - a$ is not invertible.

**Definition 8.4.** Let $A$ be complex unital algebra with multiplicative unit $e$ ($e \neq 0$). For any $a \in A$, the spectrum $\sigma(a)$ of $a$ is the set of all $\lambda \in \mathbb{C}$ such that $\lambda e - a$ is not invertible. The complement $\mathbb{C} - \sigma(a)$ of $\sigma(a)$ is called the resolvent set of $a$. For any fixed $a \in A$, the function $R(a, \lambda)$ with values in $A$ defined on the set $\mathbb{C} - \sigma(a)$ and given by

$$R(a, \lambda) = (\lambda e - a)^{-1}$$

is called the resolvant of $a$.

If different algebras are involved, to avoid confusion, for any $a \in A$ we write $\sigma_A(a)$ for $\sigma(a)$. Note that there is no guarantee that the spectrum is nonempty for any $a \in A$. However, if $A$ is a unital Banach algebra, we will see that $\sigma(a)$ is nonempty for all $a \in A$. More is true: each $\sigma(a)$ is compact.

Let us mention some simple properties of the spectrum, most of which are proven in Bourbaki [8] (Chapter 1, Section 1, No. 2).

1. For all $\lambda \in \mathbb{C}$, we have $\sigma(\lambda e) = \{\lambda\}$, and $\sigma(a + \lambda e) = \sigma(a) + \lambda$.

2. An element $a \in A$ is invertible iff $0 \notin \sigma(a)$.
3. Let $P(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_n x^n$ be a polynomial of degree $n \geq 1$, with $\alpha_n \neq 0$. For any $a \in A$, let

$$P(a) = \alpha_0 e + \alpha_1 a + \cdots + \alpha_n a^n.$$  

Then we have

$$P(\sigma(a)) = \sigma(P(a)),$$

with $P(\sigma(a)) = \{P(\lambda) \mid \lambda \in \sigma(a)\}$.

4. If $a \in A$ is nilpotent (that is, $a^n = 0$ for some $n \in \mathbb{N}$), then $\sigma(a) = \{0\}$.

5. Let $A$ and $B$ be two unital algebras, and let $\varphi: A \to B$ be a homomorphism between them; recall that $\varphi(e_A) = e_B$. Then for any $a \in A$, we have $\sigma_B(\varphi(a)) \subseteq \sigma_A(a)$.

**Remark:** If $A$ is not a unital algebra, for any $a \in A$ we define the spectrum $\sigma'(a)$ of $a$ as the spectrum of $a$ in $\tilde{A}$. Since no element of $A$ is invertible, we always have $0 \in \sigma'(a)$. If $A$ is already a unital algebra with multiplicative identity $e$, then we saw in the previous section that $\tilde{A}$ is isomorphic to the product algebra $(K(e - e)) \times A$. If $(A_1, e_1)$ and $(A_2, e_2)$ are unital algebras, then for any $(a_1, a_2) \in A_1 \times A_2$ we have

$$\sigma((a_1, a_2)) = \sigma(a_1) \cup \sigma(a_2),$$

because $\lambda(e_1, e_2) - (a_1, a_2) = (\lambda e_1 - a_1, \lambda e_2 - a_2)$ is not invertible iff either $\lambda e_1 - a_1$ is not invertible or $\lambda e_2 - a_2$ is not invertible. Therefore, if $A$ is a unital algebra,

$$\sigma'(a) = \sigma(a) \cup \{0\}.$$  

In general, if $A$ is a unital algebra, then $\sigma(ab) \neq \sigma(ba)$. For example, let $H$ be a Hilbert space with a countable orthonormal basis $(e_1, e_2, \ldots, e_n, \ldots)$. Let $f: H \to H$ and $g: H \to H$ be the continuous linear maps defined by

$$f(e_n) = e_{n+1}, \quad n \geq 1,$$

$$g(e_{n+1}) = e_n, \quad n \geq 1,$$

$$g(e_1) = 0.$$  

Then $g \circ f = \text{id}_H$, but $(f \circ g)(e_{n+1}) = e_{n+1}$ and $(f \circ g)(e_1) = 0$, so $\sigma(g \circ f) = \{1\}$, and $\sigma(f \circ g) = \{0, 1\}$.

However, for any algebra (unital or not),

$$\sigma'(ab) = \sigma'(ba).$$

The above equation follows from the following proposition, which generalizes a well-known property of matrices; namely that if $A$ is a $m \times n$ matrix and $B$ is a $n \times m$ matrix, then $AB$ and $BA$ have the same nonzero eigenvalues.
Proposition 8.6. Let $A$ be a unital algebra. For any two elements $a, b \in A$ and any nonzero scalar $\lambda \in \mathbb{C}$ (actually, any field $K$), if $ab - \lambda e$ is invertible, then $ba - \lambda e$ is invertible.

Proof. Let $u$ be the inverse of $ab - \lambda e$. We have

$$e = (ab - \lambda e)u = abu - \lambda u,$$

so $abu = \lambda u + e$, and then

$$(ba - \lambda e)(bua - e) = b(abu)a - ba - \lambda bua + \lambda e$$

$$= b(\lambda u + e)a - ba - \lambda bua + \lambda e$$

$$= \lambda e.$$  

We also have

$$e = u(ab - \lambda e) = uab - \lambda u,$$

so $uab = \lambda u + e$, and then

$$(bua - e)(ba - \lambda e) = b(uab)a - ba - \lambda bua + \lambda e$$

$$= b(\lambda u + e)a - ba - \lambda bua + \lambda e$$

$$= \lambda e.$$  

Since $\lambda \neq 0$, the above shows that $\lambda^{-1}(bua - e)$ is the inverse of $ba - \lambda e$.  

The notion of character of an algebra plays a crucial role in harmonic analysis, as a technical tool to generalize the Fourier transform. Thus we introduce it right away.

Definition 8.5. Let $A$ be a complex, commutative, unital algebra with multiplicative identity $e$. A character of $A$ is any algebra homomorphism $\chi: A \to \mathbb{C}$. Thus it is a linear form such that

$$\chi(ab) = \chi(a)\chi(b),$$

for all $a, b \in A$, and

$$\chi(e) = 1.$$  

The set of characters of $A$ is denoted by $\mathcal{X}(A)$.

Note that even though $A$ is commutative, we do not denote its multiplication by $+$, to avoid confusion with addition in $A$.

Remark: Definition 8.5 still makes sense if $A$ is a noncommutative unital algebra. However, certain results about characters no longer hold. The main problem is that if $A$ is noncommutative, although the Gelfand transform $\mathcal{G}: A \to C(\mathcal{X}(A); \mathbb{C})$ (see Definition 8.7) is an algebra homomorphism, since the algebra $C(\mathcal{X}(A); \mathbb{C})$ (under pointwise multiplication) is commutative, the Gelfand transform can’t be injective.
Proposition 8.7. Let $A$ be a commutative unital algebra with multiplicative identity $e$. For any character $\chi$, the condition $\chi(e) = 1$ is equivalent to the condition that $\chi$ is not identically $0$. If so, $\chi$ is surjective.

Proof. Obviously, if $\chi(e) = 1$, then $\chi$ is not identically $0$. Conversely, if $\chi(a) \neq 0$ for some $a \in A$ then 
$$\chi(a) = \chi(ae) = \chi(a)\chi(e),$$
so $\chi(a)(1 - \chi(e)) = 0$, and since $\chi(a) \neq 0$, then $\chi(e) = 1$. Since $\chi$ is linear, $\chi(\lambda e) = \lambda \chi(e) = \lambda$, so $\chi$ is surjective.

Proposition 8.8. Let $A$ be a unital commutative algebra. For any character $\chi$, if $a \in A$ is invertible, then $\chi(a) \neq 0$.

Proof. We have 
$$1 = \chi(e) = \chi(aa^{-1}) = \chi(a)\chi(a^{-1}),$$
which implies that $\chi(a) \neq 0$.

Definition 8.6. Let $\varphi: A \rightarrow B$ be a homomorphism of commutative unital algebras. Then $\varphi$ induces a map $X(\varphi): X(B) \rightarrow X(A)$ given by 
$$X(\varphi)(\chi) = \chi \circ \varphi, \quad \chi \in X(B).$$

It is immediately verified that $X(\psi \circ \varphi) = X(\varphi) \circ X(\psi)$ and $X(id_A) = id_{X(A)}$.

It is easy to show that if $\varphi: A \rightarrow B$ is surjective, then $X(\varphi)$ is a bijection of $X(B)$ onto the set of characters of $A$ that vanish on Ker $\varphi$.

The following proposition characterizes $X(A)$ in terms of ideals of $A$, and shows a connection between spectra and characters.

Proposition 8.9. Let $A$ be a unital commutative algebra.

(1) If $\mathcal{Y}$ denotes the set of ideals of codimension 1 in $A$, then the map $\chi \mapsto \text{Ker} \chi$ is a bijection between $X(A)$ and $\mathcal{Y}$.

(2) If $a \in A$ and if $\chi \in X(A)$, then $\chi(a) \in \sigma(a)$.

Proof. (1) Since $\mathbb{C}$ has dimension 1 and a character $\chi$ is surjective onto $\mathbb{C}$, the kernel of $\chi$ has codimension 1. If $m \in \mathcal{Y}$ is an ideal of codimension 1, then $A/m$ is an algebra of dimension 1, so it is isomorphic to $\mathbb{C}$. But since $A/m$ has an identity element $\epsilon$, a homomorphism $\varphi$ from $A/m$ to $\mathbb{C}$ is uniquely determined by its value on $\epsilon$, so there is a unique isomorphism from $A/m$ to $\mathbb{C}$. If $\pi: A \rightarrow A/m$ is the canonical projection, then the homomorphism $\varphi \circ \pi$ from $A$ to $\mathbb{C}$ is the unique character with kernel $m$.

(2) Since $\chi$ is a homomorphism, we have 
$$\chi(\chi(a)e - a) = \chi(a)\chi(e) - \chi(a) = \chi(a)1 - \chi(a) = 0.$$ 
This implies that $\chi(a)e - a$ is not invertible, since otherwise by Proposition 8.8 we would have $\chi(\chi(a)e - a) \neq 0$. 


Remark: When $A$ is a commutative nonunital algebra, a character is still a homomorphism $\chi: A \to \mathbb{C}$, but this time, the trivial map with constant value 0 is a character. We denote the set of characters of $A$ by $X(A)$, and we let $X(A) = X'(A) - \{0\}$. If $A$ is already unital, then a character $\chi \in X'(A)$ (which is a homomorphism not restricted to map $e$ to 1) is nonzero iff $\chi(e) = 1$, which shows that $X(A) = X'(A) - \{0\}$ is equal to the set of characters of the unital algebra $A$. It is also easy to see that if $A$ is nonunital, then there is a bijection between the set of characters $X(\bar{A})$ and the set of characters $X'(A)$.

The third fundamental concept in the theory of algebras is due to Gelfand. The idea is to realize a commutative algebra $A$ as a set of complex-valued functions on the set of characters $X(A)$ of $A$. Let $\mathbb{C}^{X(A)}$ be the set of functions from $X(A)$ to $\mathbb{C}$.

**Definition 8.7.** Let $A$ be a commutative unital algebra. The map $G: A \to \mathbb{C}^{X(A)}$, called Gelfand transform, is defined as follows: for every $a \in A$,

$$G_a(\chi) = \chi(a), \quad \chi \in X(A).$$

The function $G_a$ (or $G(a)$) is called the Gelfand transform of $a$.

If necessary to avoid ambiguities, we write $G_A$ instead of $G$. Note that $G_a(\chi)$ is just evaluation of $\chi$ on $a$.

The set $\mathbb{C}^{X(A)}$ of functions from $X(A)$ to $\mathbb{C}$ is a commutative unital algebra under pointwise multiplication. Observe that map $G: A \to \mathbb{C}^{X(A)}$ is a homomorphism. Indeed we have

$$G_{ab}(\chi) = \chi(ab) = \chi(a)\chi(b) = G_a(\chi)G_b(\chi),$$

so $G_{ab} = G_aG_b$. We also have $G_e(\chi) = \chi(e) = 1$, so $G_e$ is the multiplicative unit in $\mathbb{C}^{X(A)}$.

Observe that the Gelfand transform is not necessarily injective. For example, if $a \in A$ is a nonzero nilpotent, then $\sigma(a) = \{0\}$, so $\chi(a) = 0$ for all characters $\chi$, which means that $G_a = 0$, even though $a \neq 0$.

Since $X(A)$ consists of functions from $A$ to $\mathbb{C}$, we can give it the topology of pointwise convergence; see Definition 1.5. Recall that in this topology, a subset of functions $f: A \to \mathbb{C}$ is open if it is the union of subsets $U_\Omega$ of functions for which there is a finite subset $\Omega$ of $A$ and some open intervals $(-r_a, r_a)$ (with $r_a > 0$) for all $a \in \Omega$, such that $f(a) \in (-r_a, r_a)$ for all $a \in \Omega$ (and $f(a)$ is arbitrary for all $a \in A - \Omega$). Because $\mathbb{C}$ is Hausdorff, the topology of pointwise convergence is Hausdorff. It is also easy to see that a sequence $(f_n)$ of functions $f_n: A \to \mathbb{C}$ converges to a function $f$ in this topology iff it converges pointwise to $f$ (that is, for every $a \in A$, the sequence $(f_n(a))$ converges to $f(a)$); see Folland [32] (Chapter 4, Proposition 4.2). Then, by definition of the topology of pointwise convergence, each Gelfand map $G_a: \chi \mapsto \chi(a)$ is continuous. In fact, the topology of pointwise convergence on $X(A)$ is the weakest (coarsest) topology for which the Gelfand maps $G_a$ are continuous. This topology is also called the weak topology on $X(A)$. In summary, the Gelfand transform $G$ maps $A$ to the space $\mathcal{C}(X(A); \mathbb{C})$ of continuous functions on $X(A)$.

In order to obtain sharper results about spectra and characters, we now consider unital Banach algebras.
SECTION 8. NORMED ALGEBRAS AND SPECTRAL THEORY

8.4 Spectrum, Characters, II; For a Banach Algebra

If $A$ is a unital Banach algebra, then we have a more precise characterization of the spectrum of an element $a \in A$. Part (3) of Theorem 8.10 is a nontrivial and deep fact.

**Theorem 8.10.** Let $A$ be a unital Banach algebra. The following properties hold.

1. For every $a \in A$, the spectrum $\sigma(a)$ is a compact subset of $\mathbb{C}$ contained in the closed ball of radius $\|a\|$ (thus $|\lambda| \leq \|a\|$ for all $\lambda \in \sigma(a)$).

2. For any fixed $a \in A$, the resolvent $R(a, \lambda) = (\lambda e - a)^{-1}$ is a holomorphic function $R(a, \lambda) : (\mathbb{C} - \sigma(a)) \to A$ (which means that $(d/d\lambda)(R(a, \lambda))$ exists for all $\lambda \in \mathbb{C} - \sigma(a)$), and tends to zero at infinity.

3. For every $a \in A$, the spectrum $\sigma(a)$ is nonempty.

**Proof sketch.** Theorem 8.10 is proven in Dieudonné [22] (Chapter XV, Section 2), Rudin [67] (Chapter 10, Theorem 10.13), Bourbaki [8] (Chapter 1, Section 2, No. 5), and Folland [31] (Chapter 1, Section 1).

If $|\lambda| > \|a\|$, then by Proposition 8.2 the element $e - \lambda^{-1}a$ is invertible since $\|\lambda^{-1}a\| = |\lambda^{-1}||a| < 1$. Since $\lambda \neq 0$, the element $\lambda e - a$ is also invertible, which shows that $\lambda \notin \sigma(a)$. Therefore, if $\lambda \in \sigma(a)$, then $|\lambda| \leq \|a\|$.

Define the map $g$ by $g(\lambda) = \lambda e - a$. Then $g$ is continuous and invertible on $\Omega = \mathbb{C} - \sigma(a)$, so by Proposition 8.3, since inversion is continuous, $\Omega = g^{-1}(G(A))$ is open, where $G(A)$ is the set of invertible elements in $A$. Therefore $\sigma(a)$ is closed.

Replace $a$ by $\lambda e - a$ and $h$ by $(\mu - \lambda)e$ in Proposition 8.2(2). If $\lambda \in \Omega$ and $\mu$ is close enough to $\lambda$ we have

$$\|R(a, \mu) - R(a, \lambda) + (\mu - \lambda)R(a, \lambda)\|^2 \leq 2\|R(a, \lambda)\|^3|\mu - \lambda|^2,$$

which shows that

$$\frac{d}{d\lambda} R(a, \lambda) = -R(a, \lambda)^2.$$

Therefore $R(a, \lambda)$ is analytic on $\Omega$.

For $\lambda$ such that $|\lambda| > \|a\|$ we know that $e - \lambda^{-1}a$ is invertible, and by Proposition 8.2(1) we get

$$R(a, \lambda) = (\lambda e - a)^{-1} = \lambda^{-1}(e - \lambda^{-1}a)^{-1} = \sum_{n=0}^{\infty} \lambda^{-(n+1)}a^n,$$

and when $|\lambda|$ tends to infinity, since $|\lambda^{-1}|$ tends to zero, the above series tends to 0.
As explained in Rudin [67] (Chapter 3, pages 82-85), Cauchy’s theorem and Liouville’s theorem generalize to holomorphic functions from an open subset of \( \mathbb{C} \) to a Banach space (even a Fréchet space). If \( \sigma(a) \) was empty, then \( R(a, \lambda) \) would be a holomorphic entire function which is bounded, so by Liouville’s theorem it would be constant. Since \( R(a, \lambda) \) tends to zero at infinity, this constant would be zero, which is absurd (see also Rudin [67], Chapter 10, Theorem 10.13).

Remark: It is easy to show that

\[ R(a, \mu) - R(a, \lambda) = (\lambda - \mu) R(a, \lambda) R(a, \mu) \]

for all \( (\lambda, \mu) \in (\mathbb{C} - \sigma(a)) \times (\mathbb{C} - \sigma(a)) \). This shows that \( R(\lambda, a) \) and \( R(\mu, a) \) commute.

As a corollary of Theorem 8.10 we have the following theorem.

**Theorem 8.11.** (Gelfand–Mazur) Let \( A \) be a unital Banach algebra. If \( A \) is a (possibly noncommutative) field, then \( A \) is isometrically isomorphic to \( \mathbb{C} \).

**Proof.** For any \( a \in A \), since \( \sigma(a) \) is nonempty there is some \( \lambda \) such that \( \lambda e - a \) is not invertible. Since \( A \) is a field, every nonzero element is invertible, and we must have \( \lambda e - a = 0 \), so \( a = \lambda e \). But if \( \lambda_1 \neq \lambda_2 \), then at most one of \( \lambda_1 e - a \) and \( \lambda_2 e - a \) is zero, so there is a unique \( \lambda(a) \) such that \( a = \lambda(a)e \), and so the map \( a \mapsto \lambda(a) \) is an isomorphism. We have \( |\lambda(a)| = \|\lambda(a)e\| = \lambda(a) \), so this map is an isometry. \( \square \)

**Proposition 8.12.** Let \( A \) be a unital Banach algebra. For any invertible element \( a \in A \), if \( \|a\| = \|a^{-1}\| = 1 \), then \( \sigma(a) \subseteq \mathbb{U}(1) \).

**Proof.** We know by Theorem 8.10(1) that \( \sigma(a) \subseteq \{z \in \mathbb{C} \mid |z| \leq 1\} \), and similarly \( \sigma(a^{-1}) \subseteq \{z \in \mathbb{C} \mid |z| \leq 1\} \). However if \( a \) is invertible it is easy to show that \( \sigma(a) = (\sigma(a))^{-1} \) (or see Dieudonné [22] (Chapter XV, Section 2), so \( \sigma(a) \subseteq \mathbb{U}(1) \)). \( \square \)

Proposition 8.12 generalizes the fact that the eigenvalues of a unitary matrix belong to \( \mathbb{U}(1) \).

We can improve the bound on the radius of the smallest closed disc containing \( \sigma(a) \) by introducing the spectral radius of \( a \).

**Proposition 8.13.** Let \( A \) be a normed algebra. For any \( a \in A \), the sequence \( (\|a^n\|^{1/n}) \) converges and its limit is \( \inf_n \|a^n\|^{1/n} \).

Proposition is proven in Dieudonné [22] (Chapter XV, Section 2) and Rudin [67] (Chapter 10, Theorem 10.13).

**Definition 8.8.** Let \( A \) be a normed algebra. For any \( a \in A \), the number \( \rho(a) = \inf_n \|a^n\|^{1/n} \) is called the spectral radius of \( a \).
By definition, \( \rho(a) \leq \|a\| \). It is easy to see that \( \rho(a^n) = (\rho(a))^n \).

**Proposition 8.14.** Let \( A \) be a unital Banach algebra. For any \( a \in A \), the spectral radius \( \rho(a) \) of \( a \) is equal to the radius of the smallest closed disc containing the spectrum \( \sigma(a) \) of \( a \), that is, \( \rho(a) = \sup\{ |\lambda| \mid \lambda \in \sigma(a) \} \).

A proof of Proposition 8.14 is given in Dieudonné [22] (Chapter XV, Section 2), Rudin [67] (Chapter 10, Theorem 10.13), and Folland [31] (Chapter 1, Section 1.1).

Proposition 8.14 is a generalization of a well-known fact about the spectral radius of a matrix, which is the largest modulus of the eigenvalues of \( A \).

If \( A \) and \( B \) are two unital Banach algebras, and if \( A \) is a subalgebra of \( B \), for any element \( a \in A \), if \( a \) has an inverse \( a^{-1} \) in \( A \), then \( a^{-1} \in B \) so \( a \) is also invertible in \( B \), but \( a \) could have an inverse \( a^{-1} \in B \) such that \( a^{-1} \notin A \). The following proposition addresses this situation.

**Proposition 8.15.** Let \( A \) and \( B \) be two unital Banach algebras, with \( A \) a closed subalgebra of \( B \). For any \( a \in A \), we have \( \sigma_B(a) \subseteq \sigma_A(a) \). Every boundary point of \( \sigma_A(a) \) belongs to \( \sigma_B(a) \). Hence if \( \sigma_A(a) \) has empty interior, then \( \sigma_B(a) = \sigma_A(a) \).

**Proof.** Since for any \( a \in A \), if the element \( \lambda e - a \) is not invertible in \( B \) then it is not invertible in \( A \), we have \( \sigma_B(a) \subseteq \sigma_A(a) \).

For the second statement, we need to show that if \( \lambda_0 \) belongs to the boundary of \( \sigma_A(a) \), then \( \lambda_0 e - a \) is not invertible in \( B \). Since \( \lambda_0 \) belongs to the boundary of \( \sigma_A(a) \) there is a sequence \( (\lambda_n) \) with \( \lambda_n \in \mathbb{C} - \sigma_A(a) \) converging to \( \lambda_0 \). For every \( n \), the inverse \( (\lambda_ne - a)^{-1} \) exists in \( A \), and thus \( (\lambda_ne - a)^{-1} \in B \). If \( \lambda_0 \notin \sigma_B(a) \), then \( (\lambda_0 e - a)^{-1} \in B \), and by Theorem 8.10, since the resolvent \( R(a, \lambda) \) is continuous, the sequence \( ((\lambda_ne - a)^{-1}) \) would converge to \( (\lambda_0 e - a)^{-1} \). Since \( A \) is closed in \( B \), and since \( (\lambda_ne - a)^{-1} \in A \), we would have \( (\lambda_0 e - a)^{-1} \in A \), contradicting the hypothesis that \( \lambda_0 \in \sigma_A(a) \).

Let us now turn to the characters of a commutative unital Banach algebra.

**Theorem 8.16.** Let \( A \) be a commutative unital Banach algebra.

1. Every character \( \chi \in \mathcal{X}(A) \) is a continuous map of norm \( \leq 1 \).
2. The space \( \mathcal{X}(A) \) is compact and Hausdorff.

**Proof sketch.** Part (1) of Theorem 8.16 is easy. From Proposition 8.9, we have \( \chi(a) \in \sigma(a) \) for all \( a \in A \), and by Proposition 8.14, \( |\chi(a)| \leq \rho(a) \leq \|a\| \), which implies (1).

Part (2) is proven in Bourbaki [8] (Chapter 1, Section 3, No. 1) and Rudin [67] (Chapter 11, Theorem 11.9). By Part (1), \( \mathcal{X}(A) \) is a subset of the closed unit ball \( B \) in \( A' \), where \( A' \) is the dual of \( A \), that is, the space of continuous linear forms on \( A \).

**Definition 8.9.** We define the weak\(^\ast\)-topology on \( A' \) as the topology of pointwise convergence on \( A \).
It can be shown that $X(A)$ is closed in $A'$ in the weak $^*$-topology on $A'$, and the restriction of this topology to $X(A)$ is the topology of pointwise convergence. By the Banach–Alaoglu theorem, $B$ is weak $^*$-compact (see Rudin [67] (Chapter 3, Theorem 3.15, or Folland [32] (Chapter 5, Theorem 5.18)), which implies that $X(A)$ is compact in the topology of pointwise convergence.

**Remark:** If the commutative unital Banach algebra $A$ is also separable, then $X(A)$ is metrizable; see Dieudonné [22] (Chapter XV, Section 3, Theorem 15.3.2).

**Proposition 8.17.** If $A$ is a nonunital commutative Banach algebra, then $X(A)$ is compact, and $X(A)$ is locally compact.

For a proof, see Bourbaki [8] (Chapter 1, Section 3, No. 1) or Folland [31] (Chapter 1, Section 1.3, Theorem 1.30).

**Theorem 8.18.** Let $A$ be a commutative unital Banach algebra. The map $\chi \mapsto \text{Ker} \chi$ is a bijection from $X(A)$ to the set of maximal ideals in $A$.

**Proof.** Let $m$ be a maximal ideal in $A$. By Proposition 8.1, this ideal must be closed, since otherwise the closure of $m$ would be an ideal strictly containing $m$, and such an ideal is proper because $m$ is contained in the set of noninvertible elements of $A$, which is closed, contradicting the maximality of $m$. By Proposition 8.5, the quotient algebra $A/m$ is a unital Banach algebra. But since $m$ is a maximal ideal, $A/m$ is a field. By the Gelfand–Mazur theorem (Theorem 8.11), the Banach algebra $A/m$ is isomorphic to $\mathbb{C}$, which implies that $m$ has codimension 1. Obviously an ideal of codimension 1 is maximal. By Proposition 8.9, the map map $\chi \mapsto \text{Ker} \chi$ is a bijection from $X(A)$ to the set of maximal ideals in $A$. \qed

As a corollary of Theorem 8.18 we have the following result proven in Bourbaki [8] (Chapter 1, Section 3, No. 2) showing that every compact space $E$ is realized as the set of characters of some commutative unital Banach algebra of functions. Recall that for every $a \in E$, we have the linear functional (evaluation at $a$) given by $\delta_a(f) = f(a)$ for all $f \in K_C(E)$, called (with an abuse of language) a Dirac measure (see Example 5.1).

**Proposition 8.19.** Let $E$ be a compact space. The map $a \mapsto \delta_a$ is a homeomorphism from $E$ to the set of characters $X(K_C(E))$ of the unital Banach algebra $K_C(E)$. If $E$ is a locally compact space, then the map $a \mapsto \delta_a$ is a homeomorphism from $E$ to the set of characters $X(C_0(E; \mathbb{C}))$ of the unital Banach algebra $C_0(E; \mathbb{C})$ of continuous functions that tend to zero at infinity.

Proposition 8.19 implies that if $E$ is compact then the set of characters of $K_C(E)$ is the set of Dirac measures. Similarly, if $E$ is locally compact then the set of characters of $C_0(E; \mathbb{C})$ is also the set of Dirac measures. In both cases, we have a bijection between the set of characters and $E$ itself.
Using the fact that the unital commutative Banach algebra $l^1(\mathbb{Z})$ of Example 8.1(5) is generated by $\delta^1$ and $(\delta^1)^{-1}$, it can be shown that $X(l^1(\mathbb{Z})) = T$; see Folland [31] (Chapter 1, Section 1.2).

If $A$ is the unital commutative Banach algebra of absolutely convergent Fourier series of Example 8.1(6), it can be shown that again $X(A) = T$; see Folland [31] (Chapter 1, Section 1.2).

8.5 Gelfand Transform, II; For a Banach Algebra

If $A$ is a commutative unital algebra, we already know that for each $a \in A$ the Gelfand transform $G_a$ is a continuous map $G_a: X(A) \to \mathbb{C}$ (where $X(A)$ is given the topology of pointwise convergence). If $A$ is a Banach algebra, then we have the following sharper result.

**Theorem 8.20.** Let $A$ be a commutative unital Banach algebra.

1. For every $a \in A$, the range of $G_a$ is equal to the spectrum $\sigma(a)$ of $a$; that is,
   \[ G_a(X(A)) = \{ \chi(a) \mid \chi \in X(A) \} = \sigma(a). \]

2. The Gelfand transform $G: A \to C(X(A); \mathbb{C})$ is a continuous homomorphism such that $\|G_a\|_\infty = \rho(a) \leq \|a\|$, and $G_e$ is the constant function 1.

3. An element $a \in A$ is invertible iff $G_a$ does not vanish on $X(A)$.

**Proof.** (1) We already know from Proposition 8.9 that $\chi(a) \in \sigma(a)$ for every $a \in A$, so we just have to prove that for every $\lambda \in \sigma(a)$, there is some $\chi \in X(A)$ such that $\chi(a) = \lambda$. If $\lambda \in \sigma(a)$, then $\lambda e - a$ is not invertible. Since $\lambda e - a$ is not invertible, the ideal $A(\lambda e - a)$ generated by $\lambda e - a$ is distinct from $A$. Using Zorn’s lemma, it is a standard argument to show that this ideal is contained in a maximal ideal $m$. By Theorem 8.18, there is some character $\chi$ such that $m = \text{Ker} \chi$, so $\chi$ vanishes on $A(\lambda e - a)$, and in particular on $\lambda e - a$, which shows that $\chi(a) = \lambda$.

(2) We already showed that $G$ is a homomorphism and that $G_e = 1$. Since by (1) we have $G_a(X(A)) = \sigma(a)$ and since by Proposition 8.14, we have $\rho(a) = \sup\{|\lambda| \mid \lambda \in \sigma(a)\}$, we get
   \[ \|G_a\|_\infty = \sup |\sigma(a)| = \rho(a). \]
   We already know that $\rho(a) \leq \|a\|$, so $\|G_a\| \leq \|a\|$, which shows that $G$ is continuous.

(3) We know that $a$ is invertible iff $0 \notin \sigma(a)$, and by (1), this is equivalent to the fact that $G_a$ does not vanish on $X(A)$.

**Remark:** If $A$ is a commutative nonunital Banach algebra, then the Gelfand transform is a homomorphism $G: A \to C_0(X(A); \mathbb{C})$; see Bourbaki [8] (Chapter 1, Section 3, No. 3) or Folland [31] (Chapter 1, Section 1.3, Theorem 1.30).

As a corollary of Theorem 8.20, we have the following result.
Proposition 8.21. Let $A$ be a commutative unital Banach algebra. For any fixed $a \in A$, the Gelfand transform $G_a$ is a homeomorphism from $X(A)$ to $\sigma(a)$ in the following two cases:

1. The algebra $A$ is generated by $a$ and $e$.

2. The algebra $A$ is generated by $a$ and $a^{-1}$ (assuming that $a$ is invertible).

Proof. By Theorem 8.20, the map $G_a$ is continuous and surjective onto $\sigma(a)$. Since $X(A)$ and $\sigma(a)$ are compact Hausdorff spaces, it suffices to show that this map is injective. But any character $\chi: A \to \mathbb{C}$ is uniquely determined by $\chi(a)$, which is trivial in (1), and in (2) follows from the fact that $\chi(a^{-1}) = (\chi(a))^{-1}$. If $G_a(\chi_1) = G_a(\chi_2)$, then $\chi_1(a) = \chi_2(a)$, and since $\chi_1$ is completely determined by $\chi_1(a)$ and similarly $\chi_2$ is completely determined by $\chi_2(a)$, we have $\chi_1 = \chi_2$, and $G_a$ is injective.

In particular, since $l^1(\mathbb{Z})$ is generated by $\delta^1$ and $\delta^{-1} = (\delta^1)^{-1}$, Proposition 8.21 shows that $X(l^1(\mathbb{Z}))$ is homeomorphic to the spectrum of $\delta^1$ (recall Example 8.1(5)). As we said earlier, this spectrum is equal to $T$, so $X(l^1(\mathbb{Z})) \cong T$.

Proposition 8.21 allows us to figure out what the Gelfand transform on $l^1(\mathbb{Z})$ is. Indeed, for every spectral value $e^{i\theta} \in T$, there is a unique character $\chi_\theta$ such that $\chi_\theta(\delta^1) = e^{i\theta}$. Since every $c \in l^1(\mathbb{Z})$ is written uniquely as $c = \sum_{m \in \mathbb{Z}} c_m(\delta^1)_m$, we have

$$\chi_\theta(c) = \sum_{m \in \mathbb{Z}} c_m \chi_\theta((\delta^1)_m) = \sum_{m \in \mathbb{Z}} c_m (\chi(\delta^1))_m = \sum_{m \in \mathbb{Z}} c_m e^{im\theta},$$

the Fourier series associated with $c = (c_m)_{m \in \mathbb{Z}} \in l^1(\mathbb{Z})$. Since there is a bijection between $X(l^1(\mathbb{Z}))$ and $T$, we can identify $\chi_\theta$ and $e^{i\theta}$, and we see that the Gelfand transform from $l^1(\mathbb{Z})$ to $C(X(l^1(\mathbb{Z})))$ is given by

$$G_e(e^{i\theta}) = \chi_\theta(c) = \sum_{m \in \mathbb{Z}} c_m e^{im\theta}.$$

As a corollary of the above characterization of the Gelfand transform on $l^1(\mathbb{Z})$ and Theorem 8.20 we obtain the following nontrivial theorem of Wiener.

Proposition 8.22. For any $c = (c_m)_{m \in \mathbb{Z}} \in l^1(\mathbb{Z})$, if the Fourier series $f(e^{i\theta}) = \sum_{m \in \mathbb{Z}} c_m e^{im\theta}$ does not vanish (does not take the value 0 for any $\theta \in \mathbb{R}/(2\pi\mathbb{Z})$), then $1/f$ is given by the absolutely convergent Fourier series $\sum_{m \in \mathbb{Z}} b_m e^{im\theta}$, with $b = c^{-1}$.

Proof. By Theorem 8.20(3), the element $c \in l^1(\mathbb{Z})$ is invertible iff the Fourier series $f = G_e$ does not vanish. In this case, $G_{c^{-1}}$ is the inverse $1/f$ of $f$. 

The Gelfand transform on $l^1(\mathbb{Z})$ turns out to be the Fourier cotransform on $l^1(\mathbb{Z})$. More generally, if $G$ is a commutative locally compact group equipped with a Haar measure $\lambda$, the Gelfand transform can be viewed as the Fourier cotransform on the commutative Banach
algebra $L^1(G)$. For any $f \in L^1(G)$, the Gelfand transform $G_f$ is a function defined on the set $X(L^1(G))$ of characters of $L^1(G)$ by

$$G_f(\zeta) = \zeta(f), \quad \zeta \in X(L^1(G)).$$

However, it turns out that there is a homeomorphism between $X(L^1(G))$ and the dual group $\hat{G}$ of $G$, which is the group of continuous homomorphisms $\chi: G \to U(1)$; for example, see Folland [31] (Chapter 4, Section 1, Theorem 4.2). This homeomorphism from the dual group $\hat{G}$ to $X(L^1(G))$ is given by the map $\chi \mapsto \zeta$, with

$$\zeta(\chi) = \int \chi(s)f(s)\,d\lambda(s), \quad \chi \in \hat{G}.$$

Consequently we can view the Gelfand transform $G_f$ of $f$ as a map $\mathcal{F}(f)$ defined on $\hat{G}$ instead of $X(L^1(G))$, namely

$$\mathcal{F}(f)(\chi) = \int \chi(s)f(s)\,d\lambda(s), \quad \chi \in \hat{G}.$$ 

The map $\mathcal{F}(f)$ is the Fourier cotransform of $f$. For technical reasons, this map is denoted as $\mathcal{F}(f)$ instead of $F(f)$. The Fourier transform of $f$ is the map $F(f)$ defined on $\hat{G}$ by

$$F(f)(\chi) = \int \overline{\chi(s)}f(s)\,d\lambda(s), \quad \chi \in \hat{G}.$$ 

Most authors define the Fourier transform with the conjugate term $\overline{\chi(s)}$ under the integral, but this convention is not universally adopted. As Folland and Bourbaki, this is the convention that we adopt. The theory of Fourier transforms on a commutative locally compact group will be discussed thoroughly in Chapter 9.

Finally, we can characterize when the Gelfand transform is an isometry and when it is injective.

**Proposition 8.23.** Let $A$ be a commutative unital Banach algebra. The Gelfand transform $G: A \to C(X(A); \mathbb{C})$ is an isometry iff $\|a^2\| = \|a\|^2$ for all $a \in A$.

**Proof.** First we prove that $\|G_a\| = \|a\|$ iff $\|a^2\|^k = \|a\|^{2k}$ for all $k \geq 1$.

Since by Theorem 8.20 we have $\|G_a\| = \rho(a)$, and since

$$\rho(a) = \lim_{k \to \infty} \left\| a^{2k} \right\|^{1/2k},$$

if $\left\| a^{2k} \right\| = \|a\|^{2k}$ for all $k \geq 1$, then $\|G_a\| = \|a\|$.

Conversely, assume that $\|G_a\| = \|a\|$. Then

$$\left\| a^{2k} \right\| \leq \|a\|^{2k} = \|G_a\|^{2k} = \rho(a)^{2k} = \rho(a^{2k}) \leq \left\| a^{2k} \right\|,$$
which shows that \( \|a^{2k}\| = \|a\|^{2k} \) for all \( k \geq 1 \). Now if \( \|a^2\| = \|a\|^2 \), then by induction \( \|a^{2k}\| = \|a\|^{2k} \) for all \( k \geq 1 \), so \( \|G_a\| = \|a\| \). Conversely, we already proved that if \( \|G_a\| = \|a\| \) then \( \|a^{2k}\| = \|a\|^{2k} \) for all \( k \geq 1 \); in particular, \( \|a^2\| = \|a\|^2 \).

**Definition 8.10.** Given a commutative algebra \( A \), the *radical of \( A \), \( \text{rad} A \), is the intersection of all maximal ideals in \( A \).

**Proposition 8.24.** Let \( A \) be a commutative unital Banach algebra. We have \( \text{Ker } G = \text{rad } A \), so the following statements are equivalent.

1. The Gelfand transform \( G : A \to \mathcal{C}(\mathcal{X}(A); \mathbb{C}) \) is injective.
2. The radical of \( A \) is trivial; that is, \( \text{rad } A = (0) \).
3. The set \( \{ a \in A : \sigma(a) = \{0\} \} \) is reduced to \( \{0\} \).
4. The set \( \{ a \in A : \rho(a) = 0 \} \) is reduced to \( \{0\} \).

**Proof.** The Gelfand transform \( G : A \to \mathcal{C}(\mathcal{X}(A); \mathbb{C}) \) is injective iff its kernel is \( (0) \). We have \( a \in \text{Ker } G \) iff \( G_a = 0 \), which means that \( \chi(a) = 0 \) for all \( \chi \in \mathcal{X}(A) \), that is, \( a \in \text{Ker } \chi \) for all \( \chi \in \mathcal{X}(A) \). Since \( \text{Ker } \chi \) is a maximal ideal this shows that \( a \in \text{rad } A \). Thus \( \text{Ker } G \subseteq \text{rad } A \). Since every maximal ideal is the kernel of some character, \( \text{rad } A \subseteq \text{Ker } G \), so we have \( \text{Ker } G = \text{rad } A \).

\( \text{Ker } G = \text{rad } A \) implies that (1) and (2) are equivalent.

Since \( a \in \text{Ker } G \) iff \( \sigma(a) = G_a(\mathcal{X}(A)) = \{0\} \), we see that \( G \) is not injective iff there is some nonzero \( a \in A \) such that \( \sigma(a) = \{0\} \), so (1) and (3) are indeed equivalent.

Since \( \sigma(a) = \{0\} \) iff \( \rho(a) = 0 \), (3) and (4) are equivalent. \( \square \)

We will show later on that if \( G \) is a locally compact group, then \( L^1(G) \) has a trivial radical.

### 8.6 Banach Algebras with Involution; \( C^* \)-Algebras

If \( A \) is a complex matrix, then we have its conjugate-transpose \( A^\ast = (A^\top)^\top = (A)^\top \), and we know that it satisfies various identities such as

\[
(A + B)^\ast = A^\ast + B^\ast, \quad (AB)^\ast = B^\ast A^\ast, \quad (A^\ast)^\ast = A,
\]

and so on. It turns out to be fruitful to define algebras and normed algebras having an operation \( a \mapsto a^\ast \) satisfying the most useful laws of conjugate-transposition.
**Definition 8.11.** Let $A$ be an algebra over $\mathbb{C}$ (not necessarily unital). An *involution* on $A$ is a bijection $a \mapsto a^*$ satisfying the following axioms:

\[
\begin{align*}
(a^*)^* &= a, \\
(a + b)^* &= a^* + b^*, \\
(\lambda a)^* &= \bar{\lambda} a^*, \\
(ab)^* &= b^* a^*,
\end{align*}
\]

for all $a, b \in A$ and all $\lambda \in \mathbb{C}$. The element $a^*$ is called the *adjoint* of $a$. If $a = a^*$, then $a$ is called *hermitian* (or *self-adjoint*). An algebra with an involution is called an *involutive algebra*.

If $A$ is a normed algebra, then an *involutive normed algebra* is an algebra with an involution $a \mapsto a^*$ satisfying the extra axiom

\[
\|a\| = \|a^*\|,
\]

for all $a \in A$. \hfill (i)

A *$C^*$-algebra* is a Banach algebra with an involution $a \mapsto a^*$ satisfying the axiom

\[
\|a\|^2 = \|a^* a\|,
\]

for all $a \in A$. \hfill ($C^*$)

**Remark:** Oddly, Rudin uses the term $B^*$-algebra instead of $C^*$-algebra; see Rudin [67], Definition 11.17. The term $C^*$-algebra seems to be used predominantly.

Here are a few immediate consequences of the axioms.

1. We have $0^* = 0$.

   This is because $0^* = (0 + 0)^* = 0^* + 0^*$, so $0^* = 0$.

2. If $A$ is unital with identity $e$, then $e^* = e$.

   This is because using the axioms, for all $a \in A$, we have

   \[
e^* a = e^* (a^*)^* = (ea^*)^* = (a^*)^* = a,
\]

   and similarly

   \[
e^* a = e^* (a^*)^* = (a^* e)^* = (a^*)^* = a,
\]

   so $e^*$ is a multiplicative identity element, and by uniqueness of such an element, $e^* = e$.

3. If $a \in A$ is invertible, then so is $a^*$ and $(a^*)^{-1} = (a^{-1})^*$.

   We have

   \[
   (a^{-1})^* a^* = (aa^{-1})^* = e^* = e,
   \]

   and

   \[
   a^* (a^{-1})^* = (a^{-1} a)^* = e^* = e,
   \]

   so $(a^*)^{-1} = (a^{-1})^*$. 
4. For any $a \in A$, $a$ is invertible iff $a^*$ is invertible.

Since $(a^*)^* = a$, by applying (3) to $a^*$ we see that if $a^*$ is invertible then $a$ is invertible, and (3) gives the converse.

5. For all $a \in A$ and all $\lambda \in \mathbb{C}$, we have $\lambda \in \sigma(a)$ iff $\lambda e - a^* = \overline{\lambda} e - a^*$ is invertible.

This is because, by (4), $\lambda e - a$ is invertible iff $\lambda e - a^* = \overline{\lambda} e - a^*$ is invertible.

6. If $A$ is a $C^*$-algebra, then equation (i) holds. Therefore a $C^*$-algebra is an involutive algebra.

We already know that $0^* = 0$. For any $a \neq 0$,

$$\|a\|^2 = \|a^* a\| \leq \|a^*\| \|a\|,$$

so $\|a\| \leq \|a^*\|$. Since $(a^*)^* = a$, we also get $\|a^*\| \leq \|a\|$, so $\|a^*\| = \|a\|$.

7. If $A$ is a $C^*$-algebra then $\|aa^*\| = \|a\|^2$.

Using the fact that $(a^*)^* = a$, and $\|a\| = \|a^*\|$, substituting $a^*$ for $a$ in $\|a^* a\| = \|a\|^2$, we get $\|aa^*\| = \|a^*\|^2 = \|a\|^2$.

8. If $A$ is a unital $C^*$-algebra, then $\|e\| = 1$.

We have

$$\|e\|^2 = \|e^* e\| = \|e\|. $$

This implies that $\|e\| = 0, 1$, but since $e \neq 0$, we must have $\|e\| = 1$.

**Example 8.3.** The examples below are among the examples listed in Example 8.1.

1. If $E$ is a complex Hilbert space, then the space $\mathcal{L}(E)$ of continuous linear maps $f : E \to E$ is a $C^*$-algebra, with involution $h \mapsto h^*$.

2. Let $X$ be a topological space. Then the space $\mathcal{C}(X; \mathbb{C})$ of bounded continuous functions on $X$ is a commutative unital involutive Banach algebra with involution $f \mapsto f^\ast$. Let $X$ be a compact topological space. Then the space $\mathcal{C}(X; \mathbb{C})$ of continuous functions on $X$ is a commutative $C^*$ unital Banach algebra with involution $f \mapsto f^\ast$. If $X$ is a locally compact space, then $\mathcal{C}_0(X; \mathbb{C})$ is a nonunital $C^*$-algebra with involution $f \mapsto f^\ast$.

3. Let $G$ be a locally compact group. The space $\mathcal{M}_{\text{reg}, \mathbb{C}}(G)$ of complex regular Borel measures on $G$, simply denoted $\mathcal{M}(G)$, is a unital Banach algebra, with the norm $\|\mu\| = |\mu|(G)$ defined in Definition 5.9, with the convolution as multiplication, and with the Dirac measure $\delta_1$ as multiplicative unit. By Proposition 6.45, the map $\mu \mapsto \mu^* = \overline{\mu}$ is an involution that makes $\mathcal{M}(G)$ into an involutive algebra. In general, it is not a $C^*$-algebra.
(4) Let $G$ be a locally compact group equipped with a left Haar measure $\lambda$. The space $L^1(G)$ (with the $L^1$-norm) can be identified with a subalgebra of $M^1(G)$, using the embedding $f \mapsto f d\lambda$ given by Proposition 5.31. The space $L^1(G)$ is a Banach algebra, but it is not unital unless $G$ is discrete. If $G$ is unimodular, then the map $f \mapsto f^*$ where $f^*(s) = \overline{f(s^{-1})}$ is an involution. If $G$ is not unimodular, then we define $f^*$ by $f^*(s) = \Delta(s^{-1})f(s^{-1})$, and then the map $f \mapsto f^*$ is an involution such that the map $f \mapsto f d\lambda$ is an embedding of the involutive Banach algebra $L^1(G)$ into the unital involutive Banach algebra $M^1(G)$. In general, $L^1(G)$ is not a $C^*$-algebra.

(5) As a special case of (4), let $G = \mathbb{Z}$, in which case $L^1(G)$ is the set of all sequences $x = (x_m)_{m \in \mathbb{Z}}$ with $x_m \in \mathbb{C}$, such that $\sum_{m \in \mathbb{Z}} |x_m| < \infty$. This space is also denoted $l^1(\mathbb{Z})$. The convolution product $x \ast y$ of $x = (x_m)$ and $y = (y_m)$ is given by

$$(x \ast y)_m = \sum_{p \in \mathbb{Z}} x_p y_{m-p},$$

and the norm by $\|x\| = \sum_{m \in \mathbb{Z}} |x_m|$. This is a unital Banach algebra with identity element $e_0$ such that $e_0(0) = 1$ and $e_0(m) = 0$ for all $m \neq 0$. The map $x \mapsto x^*$ where $$x_m^* = \overline{x}_{-m}.$$ is an involution. The involutive algebra $l^1(\mathbb{Z})$ is not a $C^*$-algebra.

**Definition 8.12.** Let $A$ be an involutive algebra. An element $a \in A$ is hermitian if $a^* = a$.

Observe for any $a \in A$, the elements $x_1 = (a + a^*)/2$ and $x_2 = (a - a^*)/(2i)$ are also hermitian, and so are $aa^*$ and $a^*a$.

**Proposition 8.25.** Let $A$ be an involutive algebra. Every $a \in A$ can be written as $a = x_1 + ix_2$ for two unique hermitian elements $x_1, x_2$.

**Proof.** The elements $x_1 = (a + a^*)/2$ and $x_2 = (a - a^*)/(2i)$ are hermitian, and we have $a = x_1 + ix_2$. Conversely, if $a = x_1 + ix_2$ with $x_1, x_2$ hermitian, then $a^* = x_1^* - ix_2^* = x_1 - ix_2$, so $x_1 = (a + a^*)/2$ and $x_2 = (a - a^*)/(2i)$ are uniquely determined hermitian elements. \(\square\)

**Definition 8.13.** Let $A$ be an involutive algebra. An element $a \in A$ is normal if $a^*a = aa^*$. If $A$ is unital with identity element $e$, then $a \in A$ is unitary if $aa^* = a^*a = e$, that is, if $a$ is invertible and if $a^{-1} = a^*$.

If $A$ is a unital involutive algebra, then the unitary elements form a subgroup of $A$. If $a$ is unitary, namely $a^{-1} = a^*$, then

$$(a^{-1})^* = (a^*)^* = a = (a^{-1})^{-1}$$

so $a^{-1}$ is unitary. And if $a^{-1} = a^*$ and $b^{-1} = b^*$, then

$$(ab)^* = b^*a^* = b^{-1}a^{-1} = (ab)^{-1},$$

so $ab$ is unitary.

As in the case of matrices, we have the following result about spectra.
8.6. BANACH ALGEBRAS WITH INVOLUTION; C*-ALGEBRAS

Proposition 8.26. Let A be a unital C*-algebra. For every a ∈ A, if a is hermitian then σ(a) ⊆ ℝ, and if a is unitary then σ(a) ⊆ T = U(1).

Proof. Assume that a* = a. If α + iβ ∈ σ(a), with α, β ∈ ℝ, then for every real number λ, we have α + i(β + λ) ∈ σ(a + iλe), since (α + i(β + λ))e − (a + iλe) = (α + iβ)e − a. By Theorem 8.10, we have |µ| ≤ ∥b∥ for every µ ∈ σ(b), and since λ is real λ = λ, so we have

\[ \sigma^2 + (\beta + \lambda)^2 \leq \| a + i\lambda e \|^2 \]
\[ = \| (a + i\lambda e)^*(a + i\lambda e) \| \]
\[ = \| a^*a + i\lambda a^* - i\lambda^2 e \| \]
\[ = \| a^*a + \lambda^2 e \| \]
\[ \leq \| a^*a \| + \lambda^2 \]
\[ = \| a \|^2 + \lambda^2, \]

which yields

\[ 2\beta \lambda \leq \| a \|^2 - \alpha^2 - \beta^2. \]

Since the above holds for all λ ∈ ℝ, by picking λ of the same sign as β and |λ| large enough we would violate the above inequality, so we must have β = 0.

If aa* = a*a = e, then

\[ \| a \|^2 = \| a^*a \| = \| e \| = 1, \]

hence \( \| a \| = 1 \). Similarly, since a⁻¹ is also unitary, we have \( \| a^{-1} \| = 1 \). By Proposition 8.12, we conclude that \( \sigma(a) \subseteq U(1) \).

Remark: The first part of Proposition 8.26 also holds for a nonunital C*-algebra A. For every a ∈ A, if a is hermitian then \( \sigma'(a) \subseteq \mathbb{R} \); see Bourbaki [8] (Chapter I, §6, No. 3, Proposition 3).

Definition 8.14. Let A and B be two involutive algebras. A map \( \varphi : A \rightarrow B \) is an involutive homomorphism if it is a homomorphism of algebras such that \( \varphi(a^*) = \varphi(a)^* \) for all \( a \in A \). As usual, if A and B are unital we require that \( \varphi(e_A) = e_B \). We say that A is an involutive subalgebra of B if it is a subalgebra of B and if A is closed under the involution \( a \mapsto a^* \); that is, if \( a \in A \), then \( a^* \in A \).

The following result is proven in L. Schwartz [70] (Chapter II, Section 14, page 374).

Proposition 8.27. Let A be a C*-algebra (not necessarily commutative). For every a ∈ A, if a is normal, then we have \( \| a^2 \|^2 = \| a \|^4 \). As a consequence, \( \rho(a) = \| a \| \). In particular, if A is commutative then the above facts hold.
Proof. Using the fact that \( \|bb^*\| = \|b\|^2 \), we get
\[
\|aa^*(aa^*)^*\| = \|aa^*\|^2 = \|a\|^4.
\]
We also have \( \|aa^*(aa^*)^*\| = \|aa^*(a^*)a^*\| = \|aa^*a^*\| \), and since \( a \) is normal, \( aa^* = a^*a \), so
\[
\|aa^*a^*\| = \|a^2(a^*)^2\|.
\]
But \( (a^2)^* = (aa)^* = a^*a^* = (a^*)^2 \), so
\[
\|a^2(a^*)^2\| = \|a^2(2a^*)^2\| = \|a^2\|^2.
\]
Consequently \( \|a^2\|^2 = \|a\|^4 \), and so \( \|a^2\| = \|a\|^2 \). By induction, we get \( \|a^{2k}\| = \|a\|^{2k} \). Since
\[
\rho(a) = \lim_{k \to \infty} \|a^{2k}\|^{1/2k},
\]
we conclude that \( \rho(a) = \|a\| \). \( \square \)

**Proposition 8.28.** Let \( A \) be a unital involutive Banach algebra and let \( B \) be a unital C*-algebra. If \( \varphi: A \to B \) is an involutive homomorphism, then \( \|\varphi(a)\| \leq \|a\| \) for all \( a \in A \). Thus \( \varphi \) is continuous.

*Proof. Strongly.* We know from Proposition 8.27 that \( \rho(b) = \|b\| \) for every hermitian \( b \in A \) (since a hermitian element is obviously normal). We also know that \( \sigma_B(\varphi(a)) \subseteq \sigma_A(a) \), so by definition of \( \rho(a) \) (see Definition 8.8), we have
\[
\rho(\varphi(a)) \leq \rho(a) \leq \|a\|.
\]
Since \( \rho(b) = \|b\| \) for every hermitian \( b \in A \), and since \( \varphi(a^*) \) is hermitian (because \( \varphi(a^*) = \varphi((a^*)^*) = \varphi(a^*a) \)), we get
\[
\|\varphi(a)\|^2 = \|\varphi(a)^*\varphi(a)\| = \|\varphi(a^*)\varphi(a)\| = \|\varphi(a^*)\| = \rho(\varphi(a^*)) \leq \|a^*a\| = \|a\|^2,
\]
which implies \( \|\varphi(a)\| \leq \|a\| \). \( \square \)

Proposition 8.28 also holds if \( A \) and \( B \) are not unital. This is because \( \sigma'_B(\varphi(a)) \subseteq \sigma'_A(a) \); see Bourbaki [8] (Chapter I, §6, No. 3, Proposition 1).

Proposition 8.15 is sharpened as follows.

**Proposition 8.29.** Let \( A \) and \( B \) be two unital C*-algebra with \( A \) a closed involutive subalgebra of \( B \). If \( a \in A \) is invertible in \( B \), then \( a^{-1} \in A \), and \( \sigma_B(a) = \sigma_A(a) \) for all \( a \in A \).

*Proof.* If \( a \in A \) is hermitian, we know from Proposition 8.26 that \( \sigma_A(a) \subseteq \mathbb{R} \). Thus all points of \( \sigma_A(a) \) are boundary points, so by Proposition 8.15 we have \( \sigma_B(a) = \sigma_A(a) \).

Consider any \( a \in A \) such that \( a^{-1} \in B \). Then \( a^* \) is also invertible in \( B \), so \( aa^* \) is invertible in \( B \) (similarly \( a^*a \) is invertible in \( B \)). Since \( a \in A \) and \( A \) is an involutive subalgebra, \( a^* \in A \), then \( aa^* \in A \), and since \( aa^* \) is hermitian, by the fact we just proved above, \( (aa^*)^{-1} \in A \). This implies that \( aa^*(aa^*)^{-1} = e_A \), so \( a \) has a right inverse in \( A \). A similar argument applied to \( a^*a \) shows that \( (a^*a)^{-1} \in A \), so \( (a^*a)^{-1}a^*a = e_A \) and \( a \) has a left inverse in \( A \). Therefore \( a^{-1} \in A \). This argument applied to \( a^*a - a \) (with \( a \in A \)) shows that \( \sigma_B(a) = \sigma_A(a) \). \( \square \)
Let $A$ be an involutive algebra. For any linear form $f : A \to \mathbb{C}$, let $f^*$ be the map given by

$$f^*(a) = \overline{f(a^*)}, \quad a \in A.$$  

We have

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

and

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

so $f^*$ is also a linear form. We verify immediately that

$$(f^*)^* = f, \quad (f + g)^* = f^* + g^*, \quad (\lambda f)^* = \overline{\lambda} f^*.$$  

**Definition 8.15.** Let $A$ be an involutive algebra. For any linear form $f : A \to \mathbb{C}$, the linear form $f^*$ given by

$$f^*(a) = \overline{f(a^*)}, \quad a \in A$$

is called the adjoint of $f$. We say that $f$ is hermitian (or self-adjoint) if $f^* = f$.

### 8.7 Characters and Gelfand Transform in a $C^*$-Algebra

Interestingly, the characters of a commutative unital $C^*$-algebra are hermitian.

**Proposition 8.30.** Let $A$ be a commutative unital $C^*$-algebra. Then for any character $\chi \in \mathcal{X}(A)$, we have

$$\chi(a^*) = \overline{\chi(a)}, \quad \text{for all} \quad a \in A,$$

or equivalently $\chi(a) = \overline{\chi(a^*)}$, which shows that the characters are hermitian. Consequently $\mathcal{G}_{a^*} = \overline{\mathcal{G}_a}$ for all $a \in A$, where $\mathcal{G}_a$ is the Gelfand transform of $a$.

**Proof.** First, assume that $a \in A$ is hermitian. By Proposition 8.26 we have $\sigma(a) \subseteq \mathbb{R}$, and since by Proposition 8.9 we have $\chi(a) \in \sigma(a)$ for any $\chi \in \mathcal{X}(A)$, we have $\chi(a) \in \mathbb{R}$, and since $a$ is hermitian $a^* = a$, so

$$\chi(a^*) = \chi(a) = \overline{\chi(a)}.$$ 

Any arbitrary $a \in A$ can be written as $a = x_1 + ix_2$ for two unique hermitian elements $x_1, x_2 \in A$, and by the above fact

$$\chi(x_1) = \overline{\chi(x_1)}, \quad \chi(x_2) = \overline{\chi(x_2)}$$

and since $a = x_1 + ix_2$ and $a^* = x_1^* - ix_2 = x_1 - ix_2$ (because $x_1$ and $x_2$ are hermitian),

$$\chi(a^*) = \chi(x_1 - ix_2) = \chi(x_1) - i\chi(x_2) = \overline{\chi(x_1) - i\chi(x_1)} = \overline{\chi(x_1) + i\chi(x_2)} = \overline{\chi(a)}.$$  

Since by definition $\mathcal{G}_a(\chi) = \chi(a)$, we have

$$\mathcal{G}_{a^*}(\chi) = \chi(a^*) = \overline{\chi(a)} = \overline{\mathcal{G}_a(\chi)},$$

which means that $\mathcal{G}_{a^*} = \overline{\mathcal{G}_a}$, as claimed.  

Remark: Proposition 8.30 also holds if $A$ is a commutative nonunital $C^*$-algebra; see Bourbaki [8] (Chapter I, §6, No. 4, Theorem 1).

If $A$ is a commutative unital $C^*$-algebra, we can make the following addition to Proposition 8.21.

**Proposition 8.31.** Let $A$ be a commutative unital $C^*$-algebra. For any fixed $a \in A$, if the algebra $A$ is generated by $a$, $a^*$, and $e$, then the Gelfand transform $\mathcal{G}_a$ is a homeomorphism from $X(A)$ to $\sigma(a)$.

**Proof.** By Theorem 8.20, the map $\mathcal{G}_a$ is continuous and surjective onto $\sigma(a)$. Since $X(A)$ and $\sigma(a)$ are compact Hausdorff spaces, it suffices to show that this map is injective. But any character $\chi: A \to \mathbb{C}$ is uniquely determined by $\chi(a)$, since by Proposition 8.30, we have $\chi(a^*) = \overline{\chi(a)}$. If $\mathcal{G}_a(\chi_1) = \mathcal{G}_a(\chi_2)$, then $\chi_1(a) = \chi_2(a)$, and since $\chi_1$ is completely determined by $\chi_1(a)$ and similarly $\chi_2$ is completely determined by $\chi_2(a)$, we have $\chi_1 = \chi_2$, and $\mathcal{G}_a$ is injective. \qed

As an application of Proposition 8.31, let $H$ be a Hilbert space, and let $T$ be a bounded normal operator on $H$ (that is, $TT^* = T^*T$). Let $\mathcal{A}_T$ be the subalgebra of $\mathcal{L}(H)$ generated by $T, T^*$ and $I$. Since $T$ and $T^*$ commute, $\mathcal{A}_T$ is a commutative unital $C^*$-algebra, and $X(\mathcal{A}_T)$ is homeomorphic to the spectrum $\sigma(T)$ of $T$ in $\mathcal{L}(H)$. This is the first step in obtaining a spectral theorem for a bounded normal operator on a Hilbert space; see Folland [31] (Chapter 1, Section 1.4).

We are now ready to prove the main theorem of the theory of commutative unital $C^*$-algebras due to Gelfand and Naimark, namely that every commutative unital $C^*$-algebra can be viewed as the algebra of continuous functions on a compact space, namely its space of characters $X(A)$. The proof makes use of the version of the Stone–Weierstrass theorem for complex-valued functions that we now recall.

**Theorem 8.32.** (Stone–Weierstrass) Let $X$ be a compact space, and let $\mathcal{C}(X; \mathbb{C})$ be the algebra of continuous functions on $X$. Let $B$ be a subalgebra of $\mathcal{C}(X; \mathbb{C})$ satisfying the following properties:

1. The algebra $B$ contains the constant functions.

2. The algebra $B$ separates the points of $X$, which means that for any points $x, y \in X$, if $x \neq y$ then there is some function $f \in B$ such that $f(x) \neq f(y)$.

3. The algebra $B$ is stable under conjugation; that is, for any $f \in B$, the function $\overline{f}$ also belongs to $B$ (where $\overline{f}(x) = \overline{f(x)}$ for all $x \in X$).

Then $B$ is dense in $\mathcal{C}(X; \mathbb{C})$ (with respect to the $\| \cdot \|_\infty$ norm); that is, for every function $f \in \mathcal{C}(X; \mathbb{C})$, there is a sequence $(f_n)$ with $f_n \in B$ that converges uniformly to $f$. 
Theorem 8.32 is a cornerstone of analysis. Its proof can be found in many books, including Schwartz [70], Folland [32] (Chapter 4, Theorem 4.15), and Rudin [67] (Chapter 5, Theorem 5.7).

**Theorem 8.33.** *(Gelfand–Naimark)* Let $A$ be a commutative unital $C^*$-algebra. Then the Gelfand transform $\mathcal{G}: A \rightarrow \mathcal{C}(X(A); \mathbb{C})$ is an isometric isomorphism between $A$ and $\mathcal{C}(X(A); \mathbb{C})$ (and so $\|G_a\| = \|a\| = \rho(a)$ for all $a \in A$). Furthermore the Gelfand maps $\mathcal{G}_a$ are hermitian, which means that $\mathcal{G}_{\overline{a}} = \overline{\mathcal{G}_a}$, for all $a \in A$.

**Proof.** Since $A$ is commutative, by Proposition 8.27, we have $\|a^2\| = \|a\|^2$ for all $a \in A$. Since $A$ is a unital Banach algebra, by Proposition 8.23 the Gelfand transform $\mathcal{G}: A \rightarrow \mathcal{C}(X(A); \mathbb{C})$ is an isometry. In particular it is injective. This implies that the image $\mathcal{G}(A)$ of $A$ is closed in $\mathcal{C}(X(A); \mathbb{C})$. Indeed, for any Cauchy sequence $(\mathcal{G}_{a_n})$ in $\mathcal{C}(X(A); \mathbb{C})$, since $\|a_m - a_n\| = \|\mathcal{G}_{a_m} - \mathcal{G}_{a_n}\|$, the sequence $(a_n)$ is a Cauchy sequence in $A$, and since $A$ is a Banach space the sequence $(a_n)$ has a limit $a \in A$. Since the Gelfand transform is continuous, $\mathcal{G}_a$ is the limit of the sequence $(\mathcal{G}_{a_n})$. Therefore $\mathcal{G}(A)$ is closed in $\mathcal{C}(X(A); \mathbb{C})$. It remains to prove that $\mathcal{G}(A) = \mathcal{C}(X(A); \mathbb{C})$.

For this, we check that the hypotheses of the Stone–Weierstrass theorem (Theorem 8.32) are satisfied. Since $A$ is algebra and $\mathcal{G}$ is a homomorphism, $B = \mathcal{G}(A)$ is a subalgebra of $\mathcal{C}(X(A); \mathbb{C})$.

1. The algebra $B$ contains all the constant functions, since $\mathcal{G}_{\lambda e}$ is the constant function $\lambda$.

2. The algebra $B$ separates points. Indeed, if $\chi_1$ and $\chi_2$ are two distinct characters, then $\chi_1(a) \neq \chi_2(a)$ for some $a \in A$, and then $\mathcal{G}_a(\chi_1) = \chi_1(a) \neq \chi_2(a) = \mathcal{G}_a(\chi_2)$, so $\mathcal{G}_a$ separates $\chi_1$ and $\chi_2$.

3. By Proposition 8.30, we have $\mathcal{G}_{a^*} = \overline{\mathcal{G}_a}$ for all $a \in A$, so $B$ is stable under conjugation.

By the Stone–Weierstrass theorem, $B$ is dense in $\mathcal{C}(X(A); \mathbb{C})$. But $B$ is closed in $\mathcal{C}(X(A); \mathbb{C})$, so $\mathcal{G}(A) = B = \mathcal{C}(X(A); \mathbb{C})$, proving that $\mathcal{G}$ is an isomorphism. 

The Gelfand–Naimark theorem is used to prove the Plancherel–Godement theorem (see Section 10.5, Theorem 10.37), and some representation theory results in harmonic analysis; see Dieudonné [20].

If $A$ is a nonunital commutative $C^*$-algebra, then there is a version of the Gelfand–Naimark theorem in which the algebra $\mathcal{C}(X(A); \mathbb{C})$ is replaced by the algebra $\mathcal{C}_0(X(A); \mathbb{C})$ of continuous functions that tend to zero at infinity. Thus there is an isometric isomorphism $\mathcal{G}: A \rightarrow \mathcal{C}_0(X(A); \mathbb{C})$; see Bourbaki [8] (Chapter 1, Section 6, No. 4) and Folland [31] (Chapter 1, Section 1.3).

There is also a version of the Gelfand–Naimark theorem for noncommutative $C^*$-algebras. Roughly speaking, a $C^*$-algebra is isometricaly isomorphic to a $C^*$-subalgebra of the algebra of bounded operators on some Hilbert space; see Rudin [67].
The spectral theory of $C^*$-algebras is the key machinery used to develop generalizations of the spectral theorems for normal matrices to bounded (and unbounded) operators of various kinds on a Hilbert space. A condensed presentation of these spectral theorems is given in Folland [31] (Chapter 1, Section 1.4). An extensive treatment of spectral theorems is given in Rudin [67], and in Lax [54].

8.8 The Enveloping $C^*$-Algebra of an Involutive Banach Algebra

If $A$ is an involutive Banach algebra, there is a $C^*$-algebra $\text{St}(A)$ and an involutive homomorphism $j: A \to \text{St}(A)$ that satisfy a universal mapping condition with respect to homomorphisms of $A$ into a $C^*$-algebra. To construct $\text{St}(A)$, first we establish the following result.

**Proposition 8.34.** Let $A$ be an involutive algebra (over $\mathbb{C}$) and let $p$ be a semi-norm on $A$. The following conditions are equivalent:

1. We have $p(ab) \leq p(a)p(b)$, $p(a^*) = p(a)$, $(p(a))^2 = p(a^*a)$, for all $a, b \in A$.
2. The set $\mathfrak{N}$ of all $a \in A$ such that $p(a) = 0$ is a self-adjoint ($\mathfrak{N}^* = \mathfrak{N}$) two-sided ideal of $A$, and the norm induced on $A/\mathfrak{N}$ makes $A/\mathfrak{N}$ an involutive normed algebra whose completion is a $C^*$-algebra.
3. There is a homomorphism $\varphi$ of involutive algebras of $A$ into a $C^*$-algebra such that $p(a) = \|\varphi(a)\|$ for all $a \in A$.

The proof of Proposition 8.34 is quite simple and is given in Bourbaki [8] (Chapter I, §6, No. 6, Lemma 1).

**Definition 8.16.** Let $A$ be an involutive algebra (over $\mathbb{C}$). A semi-norm $p$ satisfying the conditions of Proposition 8.34 is called a **stellar semi-norm** on $A$.

Let us now assume that $A$ is an involutive Banach algebra. Let $S$ be the set of stellar semi-norms on $A$. By Proposition 8.28, we have $p(a) \leq \|a\|$ for all $a \in A$ and all $p \in S$. Consequently the function $a \mapsto \|a\|_*$ given by

$$\|a\|_* = \sup_{p \in S} p(a)$$

is the largest stellar semi-norm in $A$.

**Definition 8.17.** Let $A$ be an involutive Banach algebra, and let $\mathfrak{N}$ be the set of all $a \in A$ such that $\|a\|_* = 0$. The $C^*$-algebra obtained by completing the normed involutive algebra $A/\mathfrak{N}$ for the norm induced by $\|\|_*$ is called the **enveloping $C^*$-algebra** of $A$, and is denoted $\text{St}(A)$. The map from $A$ to the completion of $A/\mathfrak{N}$ induced by the canonical map from $A$ to $A/\mathfrak{N}$ is denoted by $j$. 
The enveloping $C^*$-algebra $\text{St}(A)$ has the following universal mapping property.

**Theorem 8.35.** Let $A$ be an involutive Banach algebra. For every involutive homomorphism $\varphi: A \to B$ of $A$ into a $C^*$-algebra $B$, there is a unique involutive homomorphism $\varphi: \text{St}(A) \to B$ such that

$$\varphi = \overline{\varphi} \circ j,$$

as shown in the following commutative diagram:

$$
\begin{array}{ccc}
A & \xrightarrow{j} & \text{St}(A) \\
\varphi \downarrow & & \downarrow \overline{\varphi} \\
B & & 
\end{array}
$$

If $A$ is commutative, then $\text{St}(A)$ is commutative, and if $A$ is unital, then $\text{St}(A)$ is unital.

**Proof sketch.** The map $a \mapsto ||\varphi(a)||$ is a stellar semi-norm on $A$. Thus $||\varphi(a)|| \leq ||a||_*$ for all $a \in A$, and it is easy to check that the homomorphism obtained by the quotient operation is continuous from $A/\mathfrak{N}$ to $B$, and extends uniquely to $\text{St}(A)$. The uniqueness of $\varphi$ is standard. \qed

It is also possible to characterize the set of characters $\text{X}(\text{St}(A))$ of $\text{St}(A)$. Recall that given the homomorphism $j: A \to \text{St}(A)$, the homomorphism $\text{X}(j): \text{X}(\text{St}(A)) \to \text{X}(A)$ is given by

$$\text{X}(j)(\chi) = \chi \circ j, \quad \chi \in \text{X}(\text{St}(A)).$$

**Proposition 8.36.** Let $A$ be an involutive Banach algebra. The map $\text{X}(j)$ is a homeomorphism of the set of characters $\text{X}(\text{St}(A))$ onto the subspace $H$ of hermitian characters in $\text{X}(A)$; that is, the characters $\chi: A \to \mathbb{C}$ such that $\chi(a) = \overline{\chi(a^*)}$ for all $a \in A$.

Proposition 8.36 is proven in Bourbaki [8] (Chapter I, §6, No. 6, corollary).

It is easy to show that if $a \in \text{rad } A$ (the radical of $A$), then $j(a) = 0$.

If $G$ is a locally compact group, then $L^1(G)$ is an involutive Banach algebra, but in general it is not a $C^*$-algebra. Thus we can form the enveloping $C^*$-algebra $\text{St}(L^1(G))$ of $L^1(G)$, denoted $\text{St}(G)$.

**Definition 8.18.** If $G$ is a locally compact group, then the enveloping $C^*$-algebra $\text{St}(L^1(G))$ of $L^1(G)$ is denoted $\text{St}(G)$.

Remarkably, the canonical map $j$ is injective.

**Proposition 8.37.** Let $G$ be a locally compact group. The canonical map $j$ from $L^1(G)$ to its enveloping $C^*$-algebra $\text{St}(G)$ is injective.
Proposition 8.37 is proved in Bourbaki [8] (Chapter I, §6, No. 7, Proposition 12).

As a corollary, the following result is obtained; see Bourbaki [8] (Chapter 1, Section 7, No. 7).

**Proposition 8.38.** If $G$ is a locally compact group, then the radical of $L^1(G)$ is the zero ideal.

We will see in Section 9.1 that every character of $L^1(G)$ is hermitian. By Proposition 8.36, there is a homeomorphism between $X(St(G))$ and $X(L^1(G))$ (see Proposition 9.7).
Chapter 9

Harmonic Analysis on Locally Compact Abelian Groups

In this chapter we generalize the various Fourier transforms and cotransforms defined in Chapter 7 to an arbitrary locally compact abelian group.

The fact that the Fourier transform \( \hat{f}: \mathbb{R} \rightarrow \mathbb{C} \) of a function \( f \in L^1(\mathbb{R}) \), given by

\[
\hat{f}(x) = \int e^{-iyx} f(y) \frac{dy}{\sqrt{2\pi}}
\]

is also a function defined on \( \mathbb{R} \), is an accident (perhaps a convenient accident).

On the other hand, given a function \( f \in L^1(\mathbb{T}) \) (a periodic function of period \( 2\pi \)), its Fourier transform \( \hat{f} = F(f) \) is the sequence \( \hat{f} = (c_m)_{m \in \mathbb{Z}} \) of Fourier coefficients

\[
c_m = \int_{-\pi}^{\pi} e^{-imx} f(x) \frac{dx}{2\pi}.
\]

We can view \( \hat{f} = (c_m)_{m \in \mathbb{Z}} \) as a function \( \hat{f}: \mathbb{Z} \rightarrow \mathbb{C} \). The domain of the Fourier transform \( \hat{f} \) is \( \mathbb{Z} \), which is completely different from \( \mathbb{T} \).

If we now consider functions in \( l^1(\mathbb{Z}) \), which are sequences \( c = (c_m)_{m \in \mathbb{Z}} \) with \( c_m \in \mathbb{C} \) such that \( \sum_{m \in \mathbb{Z}} |c_m| < \infty \), then we can define the Fourier transform \( \hat{c} = F(c) \) of \( c \) as the function \( F(c): \mathbb{T} \rightarrow \mathbb{C} \) defined on \( \mathbb{T} \) given by

\[
F(c)(e^{i\theta}) = \sum_{m \in \mathbb{Z}} c_m e^{-im\theta}.
\]

The domain of this Fourier transform is \( \mathbb{T} \), which is completely different from \( \mathbb{Z} \). Because \( \sum_{m \in \mathbb{Z}} |c_m| < \infty \), the above series is absolutely convergent, so \( F(c) \in L^1(\mathbb{T}) \).

Observe an asymmetry. If \( f \in L^1(\mathbb{T}) \), then the Fourier transform \( \hat{f} = F(f) = (c_m)_{m \in \mathbb{Z}} \) may not belong to \( l^1(\mathbb{Z}) \). The same problem arises for the Fourier transform on \( L^1(\mathbb{R}) \). In
general, \( \hat{f} \notin L^1(\mathbb{R}) \). This is the problem of Fourier inversion. We would like to know when it is possible to recover a function \( f \) from its Fourier transform \( \hat{f} \). This is a difficult problem.

Remarkably, the Fourier transform on \( L^2 \) is better behaved. This is the content of the Plancherel theorem which shows that Fourier inversion is possible.

The question remains, what should be the domain of a Fourier transform?

Observe that the examples that we considered involve the Fourier transform of the \( L^1 \)-functions on the abelian groups, \( \mathbb{R}, \mathbb{T}, \) and \( \mathbb{Z} \). These groups are locally compact.

In the case of a commutative locally compact group \( G \) (equipped with a Haar measure \( \lambda \)), it turns out that a good solution is to define the domain of the Fourier transform as the dual group \( \hat{G} \) of \( G \), which is a certain group of homomorphisms \( \chi: G \to \mathbb{C} \), namely the continuous unitary homomorphisms \( \chi: G \to \mathbb{U}(1) \).

The group \( \hat{G} \) of characters of \( G \) is defined in Section 9.1. Having defined a group structure on \( \hat{G} \), the next goal is to make \( \hat{G} \) into a topological group which is locally compact. Since \( \hat{G} \) consists of continuous functions from \( G \) to \( \mathbb{U}(1) \), we can give \( \hat{G} \) the compact-open topology, but proving that the resulting space is locally compact is nontrivial. This can be done by proving that the spaces \( \hat{G} \) and \( \text{X}(L^1(G)) \) are homeomorphic; see Theorem 9.5. Since by Proposition 8.17, the space \( \text{X}(L^1(G)) \) is locally compact, we obtain the fact that \( \hat{G} \) is locally compact.

Actually, in Proposition 9.4, we prove that if \( G \) is a locally compact group (not necessarily abelian) with a left Haar measure \( \lambda \), then for every character \( \chi \in \hat{G} \), the map \( \zeta_{\chi} \) given by

\[
\zeta_{\chi}(\mu) = \int \chi(a) \, d\mu(a) \quad \text{for all } \mu \in \mathcal{M}^1(G)
\]

is a hermitian character (an algebra homomorphism such that \( \zeta_{\chi}(\overline{\mu}) = \overline{\zeta_{\chi}(\mu)} \)) of the algebra \( \mathcal{M}^1(G) \). By restriction to \( L^1(G) \), the map \( \zeta_{\chi} \) given by

\[
\zeta_{\chi}(f) = \int \chi(a) f(a) \, d\lambda(a) \quad \text{for all } f \in L^1(G)
\]

is a hermitian character of \( L^1(G) \) not equal to the zero function. Then the map \( j: \hat{G} \to \text{X}(L^1(G)) \) given by

\[
j(\chi)(f) = \zeta_{\chi}(f) = \int \chi(a) f(a) \, d\lambda(a), \quad \chi \in \hat{G}, \ f \in L^1(G),
\]

is a homeomorphism of \( \hat{G} \) onto \( \text{X}(L^1(G)) \).

We also prove that, the spaces \( \hat{G}, \text{X}(L^1(G)), \) and \( \text{X}(\text{St}(G)) \), are homeomorphic.

Next we determine the characters of the groups \( \mathbb{Z}, \mathbb{T}, \mathbb{Z}/p\mathbb{Z}, \) and \( \mathbb{R} \). As a corollary, we obtain the isomorphisms

\[
\hat{\mathbb{R}}^n \cong \mathbb{R}^n, \quad \hat{\mathbb{T}}^n \cong \mathbb{Z}^n, \quad \hat{\mathbb{Z}}^n \cong \mathbb{T}^n.
\]
We prove that if $G$ is a finite locally compact abelian group, then $\hat{G}$ is isomorphic to $G$.

If $G$ is a compact abelian group of Haar measure 1, then its characters form an orthonormal set in $L^2(G)$.

We conclude by showing that there is a natural injection of $G$ into its double dual $\hat{\hat{G}}$.

Given any $a \in G$, define the map $\eta_a : \hat{G} \to \mathbb{C}$ by

$$\eta_a(\chi) = \chi(a), \quad \text{evaluation at } a.$$ 

The map $\eta : G \to \hat{G}$ given by $\eta(a) = \eta_a$ is a continuous homomorphism from $G$ to its double dual $\hat{G}$.

Actually, $\eta$ is an isomorphism, but this is much harder to prove (this is the Pontrjagin duality theorem).

Section 9.2 is devoted to the definition of the Fourier transform and the Fourier cotransform on an arbitrary locally compact abelian group $G$ equipped with a Haar measure $\lambda$. For every function $f \in L^1(G)$,

1. The Fourier transform of $f$ is the function $\mathcal{F}(f) : \hat{G} \to \mathbb{C}$ given by

$$\mathcal{F}(f)(\chi) = \int \overline{\chi(a)} f(a) \, d\lambda(a), \quad \chi \in \hat{G}.$$

2. The Fourier cotransform of $f$ is the function $\overline{\mathcal{F}(f)} : \hat{G} \to \mathbb{C}$ given by

$$\overline{\mathcal{F}(f)}(\chi) = \int \chi(a) f(a) \, d\lambda(a), \quad \chi \in \hat{G}.$$

These transforms are not independent. In fact, each one can be obtained from the other. For all $f \in L^1(G)$, and all $\chi \in \hat{G}$, we have

$$\overline{\mathcal{F}(f)}(\chi) = \mathcal{F}(f)(\chi^{-1}) = \mathcal{F}(\overline{f})(\chi) = \overline{\mathcal{F}(\overline{f})(\chi)}.$$

We show that modulo the isomorphism $j : \hat{G} \to X(L^1(G))$, the Fourier cotransform $\overline{\mathcal{F}(f)}$ is the Gelfand transform $G_f$ from $L^1(G)$ to $X(L^1(G))$.

Actually, it is possible to define the Fourier transform and the Fourier cotransform on the algebra $\mathcal{M}^1(G)$; see Definition 9.4.

We prove the main properties of the Fourier transform and of the Fourier cotransform. In particular, the Fourier transform $\mathcal{F}$ and the Fourier cotransform $\overline{\mathcal{F}}$ are injective involutive homomorphisms from the involutive Banach algebra $L^1(G)$ to the involutive Banach algebra $C_0(\hat{G}; \mathbb{C})$ of continuous functions on $\hat{G}$ that tend to zero at infinity. In particular for any two functions $f, g \in L^1(G)$, we have

$$\mathcal{F}(f \ast g) = \mathcal{F}(f) \mathcal{F}(g), \quad \overline{\mathcal{F}(f \ast g)} = \overline{\mathcal{F}(f)} \overline{\mathcal{F}(g)},$$
and
\[ \mathcal{F}(f^*) = (\mathcal{F}(f))^*, \quad \mathcal{F}(f^*) = (\mathcal{F}(f))^*. \]

In Section 9.3, we discuss thoroughly the Fourier transform \( \mathcal{F} \) on \( L^2(G) \) and the Fourier cotransform \( \mathcal{F}^* \) on \( L^2(\hat{G}) \), for a finite abelian group. In this case, it is possible to work out directly Fourier inversion, the Plancherel theorem, and the convolution rule.

In Section 9.4, we make a brief excursion into number theory. If we consider the multiplicative group \( G = (\mathbb{Z}/m\mathbb{Z})^* \) of units of the group \( \mathbb{Z}/m\mathbb{Z} \), then its characters are the Dirichlet characters. Then can be extended to \( \mathbb{Z} \) and are called Dirichlet characters modulo \( m \). It turns out the Fourier inversion formula for functions over \( (\mathbb{Z}/m\mathbb{Z})^* \) is one of the steps in the proof of Dirichlet’s famous theorem on arithmetic progressions of integers \( mk + \ell \) with \( \gcd(\ell, m) = 1 \) and \( k \in \mathbb{N} \), that says that such a sequence contains infinitely many primes. We briefly discuss this fascinating result.

If \( G \) is a finite abelian group, it is possible to formulate the Fourier transform \( \mathcal{F} \) on \( L^2(G) \) and the Fourier cotransform \( \mathcal{F}^* \) on \( L^2(\hat{G}) \) in terms of matrices. This is achieved in Section 9.5. The details are a bit technical due to the appearance of various dual spaces. If we denote the vector space of functions from \( G \) to \( \mathbb{C} \) as \( [G \to \mathbb{C}] \) (which is equal to \( L^1(G) \) and \( L^2(G) \)), then it turns out that the key is to extend \( \mathcal{F} \) to a bilinear form on \( [G \to \mathbb{C}]^* \), and to extend \( \mathcal{F}^* \) to a bilinear form on \( [G \to \mathbb{C}]^{**} \). The matrices associated with these bilinear forms are called Fourier matrices, and they are mutual inverses.

In Section 9.6 we consider the special case of the group \( G = \mathbb{Z}/n\mathbb{Z} \). We obtain the discrete Fourier transform and the discrete Fourier cotransform or inverse discrete Fourier transform. The Fourier matrix \( F \) is particularly interesting, as it is a Vandermonde matrix determined by the primitive \( n \)th root of unity \( \omega = e^{-2\pi i/n} \) and its powers. Convolution of two sequence \( f \) and \( g \) in \( \mathbb{C}^n \) can be expressed as \( H(f)g \), where \( H(f) \) is a circulant matrix. The matrix \( H(f) \) has the remarkable property that its eigenvector are the columns of the matrix \( F \), with corresponding eigenvalues the entries in the vector \( n\hat{f} \) (where \( \hat{f} \) is the discrete Fourier transform of \( f \)). As a consequence, we obtain another proof of the convolution rule.

Some properties of the Fourier transform (and cotransform) are presented in Section 9.7.

Section 9.8 discusses Plancherel’s Theorem and Fourier inversion.

Let \( G \) be a locally compact abelian group equipped with a Haar measure \( \lambda \). In general, given a function \( f \in L^1(G) \), its Fourier transform \( \mathcal{F}(f) \) does not belong to \( L^1(\hat{G}) \).

Plancherel’s theorem (Theorem 9.26) asserts that there is a Haar measure \( \hat{\lambda} \) on the dual group \( \hat{G} \) such that the map \( f \mapsto \mathcal{F}(f) \) sends \( L^1(G) \cap L^2(G) \) into \( L^2(\hat{G}) \), and has a unique extension which is an isometry from \( L^2(G) \) to \( L^2(\hat{G}) \).

One should realize that Theorem 9.26 does not say that the Fourier transform \( \mathcal{F} \) (or the Fourier cotransform \( \mathcal{F}^* \)) is defined on \( L^2(G) \), because in general the integral will not converge for \( f \) outside of \( L^1(G) \cap L^2(G) \). What is happening is more subtle. It is always possible by using a limit process to define the Fourier transform of any \( f \in L^2(G) \), and this extension of \( \mathcal{F} \) to \( L^2(G) \) is an isometry.
Plancherel’s theorem has an interesting corollary when $G$ is compact and abelian. If $G$ is a compact abelian group endowed with a Haar measure $\lambda$ normalized so that $G$ has measure $\lambda(G) = 1$, then $\hat{G}$ is a Hilbert basis for $L^2(G)$ (it is orthonormal and dense in $L^2(G)$).

The Pontryagin duality theorem is presented in Section 9.9. This is the most important and most beautiful theorems of the theory of locally compact abelian groups.

Let $G$ be a locally compact abelian group endowed with a Haar measure $\lambda$, let $\hat{G}$ be its dual group endowed with the associated Haar measure $\hat{\lambda}$ (see Definition 9.12), and let $\hat{\hat{G}}$ be its double dual endowed with the associated measure $\hat{\hat{\lambda}}$. The Pontryagin duality theorem asserts two facts:

1. The map $\eta: G \to \hat{\hat{G}}$ is an isomorphism and a homeomorphism between the topological groups $G$ and $\hat{\hat{G}}$.

2. If we identify $G$ and $\hat{\hat{G}}$ using the isomorphism $\eta$, then the extension $\mathcal{F}: L^2(G) \to L^2(\hat{G})$ of the Fourier transform to $L^2(G)$ and the extension $\mathcal{F}^*: L^2(\hat{G}) \to L^2(G)$ of the Fourier cotransform to $L^2(\hat{G})$ are mutual inverses. In particular, Fourier inversion holds; that is,

$$f = (\mathcal{F} \circ \mathcal{F}^*)(f) \circ \eta, \quad \text{for all } f \in L^2(G).$$

As a corollary of the Pontryagin duality theorem we can show that Fourier inversion holds for an interesting class of functions. We define $B(G)$ as the set of functions

$$B(G) = \{ f \in L^1(G) \mid \mathcal{F}(f) \in L^1(\hat{G}) \}.$$

The restriction of $\mathcal{F}$ to $B(G)$ is a bijection from $B(G)$ to $B(\hat{G})$, whose inverse is the restriction of $\mathcal{F}^*$ to $B(\hat{G})$.

Another corollary is that for any locally compact abelian group $G$, the group $G$ is discrete if and only if $\hat{G}$ is compact (and by duality, $G$ is compact if and only if $\hat{\hat{G}}$ is discrete).

The dual group $\hat{G}$ was first defined by Pontryagin (1934) and van Kampen (1935). Versions of the duality theorem were also first proven by Pontryagin and van Kampen. The first proof of the general version of Pontryagin duality appears to have been published by André Weil [84] (Chapter VI, Section 28). The definition of the Fourier transform on an arbitrary locally compact group is due to Weil [84] (Chapter VI, Section 30). In this same section, Weil proves versions of Planchel’s theorem and of the Pontryagin duality theorem.

Our exposition relies heavily on Folland [31] and Bourbaki [8], which is even more abstract than Folland. A more elementary presentation (dealing with $\sigma$-compact, metrizable, locally compact, abelian groups) can be found in Deitmar [17], which constitutes a very good warm up for the more general treatment given in this chapter.
9.1 Characters and The Dual Group

The dual of a commutative locally compact group $G$ is defined in terms of certain homomorphisms $\chi: G \to U(1)$ called characters. Even though $G$ is commutative, we use a multiplicative notation for the group operation.

**Definition 9.1.** Let $G$ be a commutative locally compact group. A character\(^1\) is a continuous homomorphism $\chi: G \to U(1)$, that is, we have

$$\chi(ab) = \chi(a)\chi(b), \quad \text{for all } a, b \in G,$$

$$|\chi(a)| = 1, \quad \text{for all } a \in G.$$

The set of characters of $G$ is denoted by $\hat{G}$.

The characters of $G$ satisfy the following properties.

**Proposition 9.1.** Let $G$ be a commutative locally compact group. The following properties hold:

(1) For every character $\chi \in \hat{G}$, we have $\chi(e) = 1$.

(2) For every character $\chi \in \hat{G}$, for every $a \in G$,

$$\chi(a^{-1}) = (\chi(a))^{-1}.$$

(3) For every character $\chi \in \hat{G}$, for every $a \in G$,

$$\chi(a^{-1}) = \overline{\chi(a)}.$$

**Proof.** Since $\chi$ is a homomorphism, we have $\chi(e) = 1$, because $\chi(e) = \chi(ee) = \chi(e)\chi(e)$, and since $|\chi(e)| = 1$, we deduce that $\chi(e) = 1$. We also have

$$1 = \chi(aa^{-1}) = \chi(a)\chi(a^{-1})$$

and

$$1 = \chi(a^{-1}a) = \chi(a^{-1})\chi(a),$$

so $\chi(a^{-1}) = (\chi(a))^{-1}$, and since $\chi(a) \in \mathbb{C}$ and $|\chi(a)| = 1$, we have $(\chi(a))^{-1} = \overline{\chi(a)}$, so we get

$$\chi(a^{-1}) = \overline{\chi(a)}, \quad \text{for all } a \in G.$$

\(^1\)Sometimes, to emphasize that their range is $U(1)$, they are called unitary characters.
Our next goal is to make the set of characters into a commutative locally compact group. The first step is to define a multiplication operation on characters. We proceed as follows.

Given two characters $\chi_1, \chi_2 \in \hat{G}$, we define $\chi_1 \chi_2$ by

$$(\chi_1 \chi_2)(a) = \chi_1(a) \chi_2(a), \quad a \in G.$$ 

Since $G$ is commutative, we have

$$(\chi_1 \chi_2)(ab) = \chi_1(ab) \chi_2(ab) = \chi_1(a) \chi_1(b) \chi_2(a) \chi_2(b) = \chi_1(a) \chi_2(a) \chi_1(b) \chi_2(b) = (\chi_1 \chi_2)(a)(\chi_1 \chi_2)(b)$$

so $\chi_1 \chi_2$ is a homomorphism. Since $|\chi_1(a)| = |\chi_2(a)| = 1$, we also have

$$|(\chi_1 \chi_2)(a)| = |\chi_1(a) \chi_2(a)| = |\chi_1(a)||\chi_2(a)| = 1.$$ 

Since $\chi_1$ and $\chi_2$ are continuous, and multiplication on $\mathbb{C}$ is continuous, the map $\chi_1 \chi_2$ is continuous. Therefore $\chi_1 \chi_2$ is a character.

If we define $\overline{\chi}$ by

$$\overline{\chi}(a) = \overline{\chi(a)}, \quad a \in A,$$

then we have

$$\overline{\chi(ab)} = \overline{\chi(ab)} = \overline{\chi(a) \chi(b)} = \overline{\chi(a)} \overline{\chi(b)} = \overline{\chi(a) \chi(b)},$$

and

$$|\overline{\chi}(a)| = |\overline{\chi(a)}| = |\chi(a)| = 1,$$

and since $\chi$ is continuous, and conjugation on $\mathbb{C}$ is continuous, $\overline{\chi}$ is obviously continuous. Thus $\overline{\chi}$ is a character. Finally, since $|\chi(a)| = 1$,

$$(\chi \overline{\chi})(a) = \chi(a) \overline{\chi(a)} = 1$$

and

$$(\overline{\chi} \chi)(a) = \overline{\chi(a)} \chi(a) = 1.$$ 

In summary, we proved the following result.

**Proposition 9.2.** The set $\hat{G}$ of characters with the multiplication operation $(\chi_1, \chi_2) \mapsto \chi_1 \chi_2$ defined above is a commutative group with the constant function from $G$ to $\mathbb{U}(1)$ with value 1 as identity. The inverse operation is $\chi \mapsto \overline{\chi}$.

The next step is to give $\hat{G}$ a topology that will make it a locally compact group. Although it is far from obvious why this works, since $\hat{G}$ consists of continuous functions from $G$ to
$U(1)$, we give it the compact-open topology (see Definition 1.11). A subbasis for this topology consists of the sets

$$S(K, U) = \{ f \mid f \in \hat{G}, f(K) \subseteq U \},$$

where $K$ is any compact subset of $G$ and $U$ is any open subset of $U(1)$; see Definition 1.11. The group operations (multiplication and inversion) are continuous in this topology, although this not immediately obvious.

Since $\hat{G}$ is a group, it suffices to show that for every open subset $S(K, U)$ containing the constant function 1, there is some subset $S(K', V)$ such that $S(K', V)S(K', V) \subseteq S(K, U)$, and that there is some subset $S(K', V)$ such that $(S(K', V))^{-1} \subseteq S(K, U)$. Since $U(1)$ is a topological group, there is some open subset $V$ of $U$ such that $VV \subseteq U$, and then $1 \in S(K, V)$ and $S(K, V)S(K, V) \subseteq S(K, U)$. Since inversion in $G$ is continuous, there is also an open subset $V$ of $U(1)$ such that $V^{-1} \subseteq U$, and then $1 \in S(K, V^{-1})$ and $(S(K, V^{-1})^{-1} \subseteq S(K, U)$.

Since $U(1)$ is Hausdorff, the compact-open topology is Hausdorff. Indeed, if $\chi_1, \chi_2 \in \hat{G}$ and if $\chi_1 \neq \chi_2$, then there is some $a \in G$ such that $\chi_1(a) \neq \chi_2(a)$, and since $U(1)$ is Hausdorff there exists two disjoint open subsets $U_1, U_2$ with $\chi_1(a) \in U_1$ and $\chi_2(a) \in U_2$. Then $S(\{a\}, U_2) \cap S(\{a\}, U_1) = \emptyset$, with $\chi_1 \in S(\{a\}, U_1)$ and $\chi_2 \in S(\{a\}, U_2)$.

In summary, we proved the following result.

**Proposition 9.3.** Let $G$ be a locally compact abelian group. The set $\hat{G}$ of characters of $G$ with the multiplication $(\chi_1, \chi_2) \mapsto \chi_1\chi_2$ given by

$$(\chi_1\chi_2)(a) = \chi_1(a)\chi_2(a), \quad a \in G,$$

and endowed with the compact-open topology, is an abelian topological group.

**Definition 9.2.** Let $G$ be a locally compact abelian group. The topological abelian group $\hat{G}$ of characters of $G$ is called the **dual** (or **Pontrjagin dual**) of $G$.

Proving directly that $\hat{G}$ is locally compact is not so easy. André Weil gives a clever proof in [84]. Another way to proceed is to use the fact that $G$ is endowed with a Haar measure $\lambda$, and to prove that $\hat{G}$ is homeomorphic to $X(L^1(G))$, a remarkable result showing that the groups characters of $\hat{G}$ and the algebra characters of $L^1(G)$ are in some sense equivalent.

The proof that $\hat{G}$ and $X(L^1(G))$ are homeomorphic is quite technical. Folland [31] (Chapter 4) gives the main idea, but does not prove that the map is injective, nor that its inverse is continuous. Bourbaki [8] (Chapter 2, Section 1) gives a complete, but terse proof. The details of the proof are not illuminating so we will only indicate its main ideas.

---

[2] Recall that a compact subset $K$ of $G$ is a subset such that every open cover of $K$ by open subsets in $G$ contains a finite subfamily covering $K$. Obviously every finite subset is compact.
The crucial step is to show that every group character \( \chi \in \hat{G} \) induces an algebra character \( \zeta_\chi \in \mathcal{X}(L^1(G)) \); namely, for every \( f \in L^1(G) \),

\[
\zeta_\chi(f) = \int \chi(a)f(a) \, d\lambda(a).
\]

Actually, every group character \( \chi \in \hat{G} \) induces an algebra character \( \zeta_\chi \in \mathcal{X}(M^1(G)) \); namely, the map given by

\[
\zeta_\chi(\mu) = \int \chi(a) \, d\mu(a)
\]

for every complex measure \( \mu \in M^1(G) \) is a character of the algebra \( M^1(G) \). This more general fact will be needed.

Recall from Example 8.3(3) that \( M^1(G) \) is a unital Banach algebra under convolution, with involution given by \( \mu^* = \overline{\mu} \). The identity element of \( M^1(G) \) is the Dirac measure \( \delta_1 \). The algebra \( L^1(G) \) is also a Banach algebra under convolution, with involution given by \( f \mapsto \overline{f} \), but unless \( G \) is discrete, it is nonunital. However, \( L^1(G) \) is embedded in \( M^1(G) \) as a closed Banach involutive subalgebra (via \( f \mapsto f \, d\lambda \)), so we can consider the unital Banach involutive subalgebra \( L^1(G) \oplus \mathbb{C}\delta_1 \). Any character \( \chi: L^1(G) \to \mathbb{C} \) extends uniquely to a character \( \chi': (L^1(G) \oplus \mathbb{C}\delta_1) \to \mathbb{C} \) by letting

\[
\chi'(f \, d\lambda + \alpha \delta_1) = \chi(f) + \alpha.
\]

**Proposition 9.4.** Let \( G \) be any locally compact group (not necessarily abelian) with a left Haar measure \( \lambda \). For every character \( \chi \in \hat{G} \), the map \( \zeta_\chi \) given by

\[
\zeta_\chi(\mu) = \int \chi(a) \, d\mu(a) \quad \text{for all } \mu \in M^1(G)
\]

is a hermitian character (an algebra homomorphism such that \( \zeta_\chi(\overline{\mu}) = \overline{\zeta_\chi(\mu)} \)) of the algebra \( M^1(G) \). By restriction to \( L^1(G) \), the map \( \zeta_\chi \) given by

\[
\zeta_\chi(f) = \int \chi(a)f(a) \, d\lambda(a) \quad \text{for all } f \in L^1(G)
\]

is a hermitian character of \( L^1(G) \) not equal to the zero function.

**Proof.** Let \( \mu, \nu \) be any two complex measures in \( M^1(G) \). The function \( \chi: G \to \mathbb{C} \) is continu-
uous and bounded, so by the corollary of Proposition 6.53, we have
\[ \zeta_{\chi}(\mu \ast \nu) = \int \chi(a) d(\mu \ast \nu)(a) \]
\[ = \int \chi(ab) d\mu(a) d\nu(b) \]
\[ = \int \chi(a)\chi(b) d\mu(a) d\nu(b) \]
\[ = \left( \int \chi(a) d\mu(a) \right) \left( \int \chi(b) d\nu(b) \right) \]
\[ = \zeta_{\chi}(\mu) \zeta_{\chi}(\nu). \]

Recall from Proposition 5.23,
\[ \int \varphi d\bar{\mu} = \int \overline{\varphi(s)} d\mu(s), \]
and by Proposition 6.44,
\[ \int \varphi d\bar{\mu} = \int \varphi d\mu, \]
with \( \check{\varphi}(a) = \varphi(a^{-1}) \) for all \( a \in G \). Thus we have
\[ \zeta_{\chi}(\bar{\mu}) = \int \chi(a) d\bar{\mu} \]
\[ = \int \overline{\chi(a)} d\bar{\mu} \]
\[ = \int \chi(a^{-1}) d\bar{\mu} \]
\[ = \int \chi(a) d\mu(a) = \zeta_{\chi}(\mu). \]
Thus \( \zeta_{\chi} \) is a Hermitian character.

By restriction to \( L^1(G) \), we obtain an algebra homomorphism, with
\[ \zeta_{\chi}(f) = \int f(a)\chi(a) d\lambda(a). \]

We need to prove that \( \zeta_{\chi} \) is not the zero function. The proof is not trivial. One method is to observe that this is a special case of the fundamental fact that there is a bijection between the set of unitary representations of the group \( G \) and the set of nondegenerate representations of the algebra \( L^1(G) \). This connection will be discussed in Section 11.3. It is the method followed by Folland [31] (Chapter 3, Theorem 3.9, and Chapter 4, Section 4.1). The characters \( \chi: G \to U(1) \) are indeed unitary representations of \( G \). The other method
used by Bourbaki [8] (Chapter 2, Section, No 1) is to use Corollary 6.51. We can choose a filter where the measures in $M^1(G)$ are of the form $f d\lambda$ with $f \in K_C(G)$, so that $\zeta(\chi)$ tends to $\chi(e) = 1 \neq 0$.

The following deep result is obtained.

**Theorem 9.5.** Let $G$ be a locally compact abelian group (equipped with a Haar measure $\lambda$). The map $j: \hat{G} \to X(L^1(G))$ given by

$$j(\chi)(f) = \zeta(\chi) = \int \chi(a) f(a) d\lambda(a), \quad \chi \in \hat{G}, f \in L^1(G),$$

is a homeomorphism of $\hat{G}$ onto $X(L^1(G))$.

**Proof sketch.** We follow Bourbaki [8] (Chapter 2, Section 1, No 1). In this proof, we view every function in $L^1(G)$ as a measure in $M^1(G)$. To show that $j$ is injective, for any $\chi \in \hat{G}$, any $f \in L^1(G)$, and any $a \in G$, we have

$$\zeta(\delta_a * f) = \chi(a) \zeta(\chi) \neq 0,$$

because $\zeta(\delta_a) = \int \chi(b) d\delta_a(b) = \chi(a)$. Since there is some function $f \in L^1(G)$ such that $\zeta(\chi) \neq 0$, if $\zeta$ is the zero function, then the equation

$$\zeta(\delta_a * f) = \chi(a) \zeta(\chi), \quad \text{for all } a \in G,$$

shows that $\chi = 0$. Therefore $j$ is injective.

To prove surjectivity, we use the following trick. For any $\zeta \in X(L^1(G))$ other than the zero function, pick some function $f \in L^1(G)$ such that $\zeta(f) \neq 0$. Define $\chi: G \to \mathbb{C}$ by

$$\chi(a) = \zeta(\delta_a * f)/\zeta(f).$$

The goal is to show that $\chi$ is a character of $G$ such that $\zeta = \zeta(\chi)$.

It can be shown that the map $a \mapsto \delta_a * f$ is continuous, so $\chi$ is continuous. We also have

$$|\chi(a)| \leq \|\delta_a * f\|_{\infty} / |\zeta(f)| = \|f\|_{\infty} / |\zeta(f)|.$$

Therefore, $\chi$ is continuous and bounded. The next step is the most technical part of the proof. It can be shown that there is a filter base $B$ of $e \in G$ consisting of compact subsets, such that, for every $V \in B$, there is a continuous function $g_V$ which is positive, zero outside of $V$, and with $\int g_V d\lambda = 1$, and such that

$$\delta_a * f = \lim \delta_a * g_V * f.$$

Then, since $\zeta(\delta_a * g_V * f) = \zeta(\delta_a * g_V) \zeta(f)$, and by definition $\chi(a) = \zeta(\delta_a * f)/\zeta(f)$, we get

$$\chi(a) = \lim \zeta(\delta_a * g_V).$$
and for any $h \in L^1(G)$,
\[
\zeta(\delta_a \ast h) = \lim \zeta(\delta_a \ast g_V \ast h) = (\lim \zeta(\delta_a \ast g_V))\zeta(h) = \chi(a)\zeta(h).
\]
Using the above equation, for all $a, b \in G$, we have
\[
\chi(ab) = \zeta(\delta_a \ast \delta_b \ast f)/\zeta(f) = \zeta(\delta_a)\zeta(\delta_b \ast f)/\zeta(f) = \chi(a)\zeta(\delta_b \ast f)/\zeta(f) = \chi(a)\chi(b),
\]
which proves that $\chi$ is a character of $G$. Since for any $f \in L^1(G)$ we have $(\delta_a \ast f)(s) = f(a^{-1}s)$, (see just after Definition 6.22), we have
\[
(g \ast f)(s) = \int g(a)f(a^{-1}s)\,d\lambda(s) = \int (\delta_a \ast f)(s)g(a)\,d\lambda(a).
\]
(Alternatively, see Bourbaki [6] (Chapter VIII, Section 1, Proposition 7). Using this fact, we get
\[
\zeta(g)\zeta(f) = \zeta(g \ast f) = \int \zeta(\delta_a \ast f)g(a)\,d\lambda(a) = \zeta(f)\int \chi(a)g(a)\,d\lambda(a) = \xi_\chi(g)\zeta(f).
\]
If we pick $f$ such that $\zeta(f) \neq 0$, we deduce that $\zeta = \zeta_\chi$, establishing the fact that $j$ is surjective. Since $j$ is injective and surjective, it is a bijection. It remains to prove that $j$ is a homeomorphism. We skip this proof, referring the reader to Bourbaki [8] (Chapter 2, Section 1, No. 1).

As a corollary of Theorem 9.5, since by Proposition 8.17 the space $X(L^1(G))$ is locally compact, we have the following fact.

**Corollary 9.6.** Let $G$ be a locally compact abelian group. The group $\hat{G}$ of characters of $G$ with the compact-open topology is locally compact.

Theorem 9.5, Proposition 9.4, and Proposition 8.36, also imply the following result.

**Proposition 9.7.** Let $G$ be a locally compact abelian group. The characters of the algebra $L^1(G)$ are hermitian. There is a homeomorphism between the set of characters $X(\text{St}(G))$ of the enveloping $C^*$-algebra $\text{St}(G)$ of $L^1(G)$ and the set of characters $X(L^1(G))$ of $L^1(G)$.

In summary, the spaces $\hat{G}$, $X(L^1(G))$, and $X(\text{St}(G))$, are homeomorphic.

It is instructive to figure out the duals of various familiar locally compact abelian groups.
Proposition 9.8. The locally compact abelian groups, \( \mathbb{Z}, \mathbb{T}, \mathbb{Z}/n\mathbb{Z}, \) and \( \mathbb{R} \), have the following characters and dual groups.

1. For \( \mathbb{Z} \), the homomorphisms \( m \mapsto e^{im\theta} = (e^{i\theta})^m \), for any fixed \( \theta \in \mathbb{R}/2\pi \mathbb{Z} \), and any \( m \in \mathbb{Z} \). Therefore, the dual group \( \hat{\mathbb{Z}} \) of \( \mathbb{Z} \) is isomorphic to \( \mathbb{T} \).

2. For \( \mathbb{T} \), the homomorphisms \( e^{i\theta} \mapsto e^{im\theta} = (e^{i\theta})^m \), for any fixed \( m \in \mathbb{Z} \), and any \( \theta \in \mathbb{R}/2\pi \mathbb{Z} \). Therefore, the dual group \( \hat{\mathbb{T}} \) of \( \mathbb{T} \) is isomorphic to \( \mathbb{Z} \).

3. For \( \mathbb{Z}/n\mathbb{Z} \), the homomorphisms \( m \mapsto e^{2\pi imk/n} = (e^{2\pi ik/n})^m \), for any fixed \( k \in \mathbb{Z}/n\mathbb{Z} \), and any \( m \in \mathbb{Z}/n\mathbb{Z} \). Therefore, the dual group \( \mathbb{Z}/n\mathbb{Z} \) of \( \mathbb{Z}/n\mathbb{Z} \) is isomorphic to \( \mathbb{Z}/n\mathbb{Z} \) (itself).

4. For \( \mathbb{R} \), the homomorphisms \( x \mapsto e^{iyx} = (e^{iy})^x \), for any fixed \( y \in \mathbb{R} \), and all \( x \in \mathbb{R} \). Therefore, the dual group \( \hat{\mathbb{R}} \) of \( \mathbb{R} \) is isomorphic to \( \mathbb{R} \) (itself).

Proof. (1) Since \( \mathbb{Z} \) is a cyclic group generated by 1, every homomorphism \( \varphi: \mathbb{Z} \to \mathbb{U}(1) \) satisfies the equation

\[
\varphi(m) = (\varphi(1))^m, \quad \text{for all } m \in \mathbb{Z}.
\]

Thus \( \varphi \) is uniquely determined by picking \( \varphi(1) = e^{i\theta} \) in \( \mathbb{U}(1) \).

The characters of \( \mathbb{T} \) are easily obtained from the characters of \( \mathbb{R} \), so we consider (4) next.

(4) Folland has a particularly nice proof of (4). Any homomorphism \( \varphi: \mathbb{R} \to \mathbb{U}(1) \) satisfies \( \varphi(0) = 1 \), and since \( \varphi \) is continuous, there is some \( a > 0 \) such that \( \int_0^a \varphi(t) \, dt \neq 0 \). Let \( A = \int_0^a \varphi(t) \, dt \). Since \( \varphi(x + t) = \varphi(x)\varphi(t) \), we have

\[
A\varphi(x) = \left( \int_0^a \varphi(t) \, dt \right) \varphi(x) = \int_0^a \varphi(t)\varphi(x) \, dt = \int_0^a \varphi(t) + x \, dt = \int_x^{a+x} \varphi(u) \, du.
\]

It follows that \( \varphi \) is differentiable, and we have

\[
\varphi'(x) = A^{-1}(\varphi(a + x) - \varphi(x)) = A^{-1}(\varphi(a)\varphi(x) - \varphi(x)) = A^{-1}(\varphi(a) - 1)\varphi(x).
\]

If we let \( c = A^{-1}(\varphi(a) - 1) \), then (using the fact that \( \varphi(0) = 1 \)) we deduce that

\[
\varphi(x) = e^{cx}.
\]

Since \( |\varphi(x)| = 1 \), \( c \) must be a pure imaginary number of the form \( c = iy \), with \( y \in \mathbb{R} \), so \( \varphi(x) = e^{iyx} \).

(2) Recall that we have the surjective homomorphism \( \sigma: \mathbb{R} \to \mathbb{T} \) given by \( \sigma(\theta) = e^{i\theta} \), and that the kernel of \( \sigma \) is \( 2\pi \mathbb{Z} \). Therefore, there is a bijection between the continuous homomorphisms \( \varphi: \mathbb{T} \to \mathbb{U}(1) \) and the continuous homomorphisms \( \psi: \mathbb{R} \to \mathbb{U}(1) \) such that \( \text{Ker } \psi = 2\pi \mathbb{Z} \). By (4) the homomorphisms \( \psi \) are of the form \( \psi(\theta) = (e^{iy})^{\theta} \) for some \( y \in \mathbb{R} \), and for \( \psi \) to have kernel \( 2\pi \mathbb{Z} \), it must be the case that \( \theta \equiv 0 \pmod{2\pi} \) implies that \( y\theta \equiv 0 \pmod{2\pi} \), so \( y = m \in \mathbb{Z} \), and \( \varphi(e^{i\theta}) = (e^{im})^{\theta} \).
(3) We have the canonical surjective homomorphism \( pr: \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \) whose kernel is \( \text{Ker} \, pr = \mathbb{Z}/n\mathbb{Z} \). It follows that there is a bijection between the homomorphisms \( \varphi: \mathbb{Z}/n\mathbb{Z} \to \mathbb{U}(1) \) and the homomorphisms \( \psi: \mathbb{Z} \to \mathbb{U}(1) \) such that \( \text{Ker} \, \psi = \mathbb{Z}/n\mathbb{Z} \). By (1) the homomorphisms \( \psi \) are of the form \( \psi(m) = e^{im\theta} = (e^{i\theta})^m \), and for \( \psi \) to have kernel \( \mathbb{Z}/n\mathbb{Z} \), it must be the case that \( m \equiv 0 \pmod{n} \) implies that \( \theta m \equiv 0 \pmod{2\pi} \), so for \( m = dn \) with \( d \in \mathbb{Z} \), we must have \( \theta dn \equiv 0 \pmod{2\pi} \), which implies that \( \theta = 2\pi k/n \) with \( k \in \mathbb{Z}/n\mathbb{Z} \), and then \( \psi(m) = (e^{2\pi ik/n})^m \).

Given a locally compact abelian group \( G \) and its dual \( \widehat{G} \), there is a canonical pairing \( \langle -, - \rangle: G \times \widehat{G} \to \mathbb{T} \) given by evaluation,

\[
\langle a, \chi \rangle = \chi(a), \quad a \in G, \ \chi \in \widehat{G}.
\]

In practice, the dual group \( \widehat{G} \) is only determined up to isomorphism. What this means is that if \( G_1 \) and \( G_2 \) are isomorphic to \( \widehat{G} \), the two pairings are usually different. This issue comes up with the group \( \mathbb{R} \). We figured out that one of the groups, \( \mathbb{R}_1 \), isomorphic to \( \mathbb{R} \) consists of the characters \( \chi^1(x) = (e^{iy})^x \), with \( y \in \mathbb{R} \). The pairing is given by

\[
\langle x, \chi^1 \rangle_1 = (e^{iy})^x.
\]

However, another isomorphic copy \( \mathbb{R}_2 \) of \( \mathbb{R} \) is often used, in which the characters are given by

\[
\chi^2(x) = (e^{2\pi iy})^x,
\]

with \( x \in \mathbb{R} \), in which case the pairing is

\[
\langle x, \chi^2 \rangle_2 = (e^{2\pi iy})^x.
\]

A similar problem comes up with the groups \( \widehat{\mathbb{Z}} \) and \( \widehat{\mathbb{T}} \). The characters of \( \mathbb{Z} \) can also be viewed as the homomorphisms \( m \mapsto e^{2\pi i m\theta} = (e^{2\pi i \theta})^m \), with \( \theta \in \mathbb{R}/\mathbb{Z} \), and the characters of \( \mathbb{T} \) can be viewed as the homomorphisms \( e^{2\pi i \theta} \mapsto e^{2\pi i m\theta} = (e^{2\pi i \theta})^m \), with \( \theta \in \mathbb{R}/\mathbb{Z} \).

This subtle issue is related to the choice of the normalization of the Haar measures on \( G \) and \( \widehat{G} \) and will come up when we consider the inverse Fourier transform.

Products behave well with respect to characters.

**Proposition 9.9.** If \( G_1, \ldots, G_n \) are locally compact abelian groups, then

\[
(G_1 \times \cdots \times G_n) \widehat{\times} \cong \widehat{G_1 \times \cdots \times G_n}.
\]

**Proof.** Every tuple of characters \( (\chi_1, \ldots, \chi_n) \) with \( \chi_i \in \widehat{G_i} \) induces a character on \( G_1 \times \cdots \times G_n \), by

\[
(\chi_1, \ldots, \chi_n)(a_1, \ldots, a_n) = \chi_1(a_1) \cdots \chi_n(a_n)
\]
for all \( a_i \in G_i, \ i = 1, \ldots, n \). Conversely, every character \( \chi : G_1 \times \cdots \times G_n \to \mathbb{T} \) can be written as
\[
(\chi_1, \ldots, \chi_n)(a_1, \ldots, a_n) = \chi_1(a_1) \cdots \chi_n(a_n),
\]
with \( \chi_i \) given by
\[
\chi_i(a_i) = \chi(1, \ldots, 1, a_i, 1, \ldots, 1).
\]
This proves the isomorphism of the proposition. \( \square \)

Propositions 9.8 and 9.9 imply the following facts.

**Corollary 9.10.** We have the following isomorphisms:
\[
\hat{\mathbb{R}}^n \cong \mathbb{R}^n, \quad \hat{\mathbb{T}}^n \cong \mathbb{Z}^n, \quad \hat{\mathbb{Z}}^n \cong \mathbb{T}^n.
\]
In particular, the characters in \( \hat{\mathbb{R}}^n \) are the homomorphisms from \( \mathbb{R}^n \) to \( \mathbb{T} \) given by
\[
x \mapsto e^{iy \cdot x}, \quad x, y \in \mathbb{R}^n,
\]
where \( y \cdot x \) is the Euclidean product in \( \mathbb{R}^n \); that is, \( y \cdot x = \sum_{k=1}^{n} y_k x_k \).

The characters in \( \hat{\mathbb{T}}^n \) are the homomorphisms from \( \mathbb{T}^n \) to \( \mathbb{T} \) given by
\[
(e^{i\theta_1}, \ldots, e^{i\theta_n}) \mapsto e^{im \cdot \theta}, \quad m \in \mathbb{Z}^n, \ \theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n,
\]
and the characters in \( \hat{\mathbb{Z}}^n \) are the homomorphisms from \( \mathbb{Z}^n \) to \( \mathbb{T} \) given by
\[
m \mapsto e^{im \cdot \theta}, \quad \theta \in \mathbb{R}^n/2\pi\mathbb{Z}^n, \ m \in \mathbb{Z}^n.
\]

As a corollary of Proposition 9.9, since by the structure theorem for finitely generated abelian groups, every finite abelian group is isomorphic to a product of cyclic groups \( \mathbb{Z}/p\mathbb{Z} \), by Proposition 9.8 and Proposition 9.9, we see that every finite abelian group is isomorphic to its dual.

**Proposition 9.11.** If \( G \) is a finite locally compact abelian group, then \( G \) is isomorphic to its dual \( \hat{G} \).

This fact can also be shown more directly, and there is no canonical isomorphism; see Apostol [2] (Chapter 6, Theorem 6.8).

If the abelian group is compact, then \( L^2(G) \) is a subspace of \( L^1(G) \), and a Hilbert space, with the hermitian inner product given by
\[
\langle f, g \rangle = \int f(s)\overline{g(s)} \, d\lambda(s), \quad \text{for all } f, g \in L^2(G).
\]
If we assume that \( \int_G 1 \, d\lambda = 1 \), then it is remarkable that the set of characters is an orthonormal set in \( L^2(G) \).
Proposition 9.12. Let \( G \) be a compact abelian group with a Haar measure normalized so that \( G \) has measure 1. Then for any character \( \chi \in \hat{G} \), we have \( \langle \chi, \chi \rangle = 1 \), and for any two distinct characters \( \chi_1, \chi_2 \in \hat{G} \), we have \( \langle \chi_1, \chi_2 \rangle = 0 \); that is, the characters form an orthonormal set in \( L^2(G) \).

Proof. We have
\[
\langle \chi, \chi \rangle = \int \chi(s)\overline{\chi}(s) \, d\lambda(s) = \int 1 \, d\lambda(s) = 1.
\]

If \( \chi_1 \neq \chi_2 \), then there is some \( a \in G \) such that \( \chi_1(a) \neq \chi_2(a) \), which is equivalent to \( (\chi_1\chi_2)(a) \neq 1 \). Then, using properties of characters and left invariance of the Haar measure, we have
\[
\langle \chi_1, \chi_2 \rangle = \int \chi_1(s)\overline{\chi_2}(s) \, d\lambda(s)
= \int (\chi_1\chi_2^{-1})(s) \, d\lambda(s)
= \int (\chi_1\chi_2^{-1})(aa^{-1}s) \, d\lambda(s)
= (\chi_1\chi_2^{-1})(a) \int (\chi_1\chi_2^{-1})(a^{-1}s) \, d\lambda(s)
= (\chi_1\chi_2^{-1})(a)\langle \chi_1, \chi_2 \rangle.
\]

Since \( (\chi_1\chi_2^{-1})(a) \neq 1 \), we conclude that \( \langle \chi_1, \chi_2 \rangle = 0 \). \( \square \)

Proposition 9.12 implies the following fact.

Proposition 9.13. For any locally compact abelian group \( G \), for any character \( \chi \in \hat{G} \), if \( \chi \neq 1 \), that is, \( \chi \) is not the trivial character with constant value 1, then \( \int \chi(s) \, d\lambda(s) = 0 \).

Proof. Since \( \chi \neq 1 \), \( \chi \) is orthogonal to the trivial character 1, so
\[
0 = \langle \chi, 1 \rangle = \int f \overline{1} \, d\lambda = \int f \, d\lambda.
\]

If \( G \) is a compact abelian group, it is remarkable that the orthonormal set of characters \( \hat{G} \) is a Hilbert basis of \( L^2(G) \). This means that the set \( \hat{G} \) is dense in \( L^2(G) \), so for every function \( f \in L^2(G) \) there is a sequence of linear combinations of characters converging to \( f \) in the \( L^2 \)-norm. The proof is nontrivial, and relies on Plancherel’s theorem; see Section 9.8.
(it is also a corollary of the Peter–Weyl theorem; see Dieudonné [19] (Chapter XXI, Section 3)).

In particular, for \( G = \mathbb{T} \), it turns out that the characters are the functions \( e^{i\theta} \mapsto e^{im\theta} \), so we obtain a proof of the fact that \((e^{im\theta})_{m \in \mathbb{Z}}\) is an orthonormal system and that every functions in \( L^1(\mathbb{T}) \) (a periodic function) is given by a Fourier series.

In the case of a finite locally compact group \( G \), we have \( L^1(G) = L^2(G) \), the functions in this space are just finite sequences \( x = (x_a)_{a \in G} \) of complex numbers indexed by \( G \), and we can figure out explicitly what is integration, the inner product, and convolution.

**Example 9.1.** Let \( G \) be a finite locally compact abelian group. Since \( G \) must be Hausdorff, it must have the discrete topology (because every singleton set must be closed, and since \( G \) is finite, by closure under finite unions of closed sets, every subset is closed). The Haar measure \( \lambda \) is just the counting measure, and the integral \( \int x \, d\lambda \) is the sum \( \sum_{a \in G} x_a \).

Here we have a choice; if \( |G| = n \), then we can normalize the Haar measure so that \( G \) has measure 1, or assume that it has measure \( n \). Let us adopt the first choice, \( \lambda(G) = 1 \), which implies that

\[
\int x_a \, d\lambda(a) = \frac{1}{|G|} \sum_{a \in G} x_a.
\]

The inner product \( \langle x, y \rangle \) of \( x, y \in L^2(G) \) is

\[
\langle x, y \rangle = \int x_a \overline{y_a} \, d\lambda(a) = \frac{1}{|G|} \sum_{a \in G} x_a \overline{y_a}.
\]

The convolution \( x \ast y \) of \( x, y \in L^2(G) \) is given by

\[
(x \ast y)_a = \frac{1}{|G|} \sum_{b \in G} x_b y_{b^{-1}a} = \frac{1}{|G|} \sum_{b, c \in G} x_b y_c.
\]

In the special case where \( G = \mathbb{Z}/n\mathbb{Z} \), if \( x = (x_0, \ldots, x_{n-1}) \) and \( y = (y_0, \ldots, y_{n-1}) \), for \( k = 0, \ldots, n-1 \), we have

\[
(x \ast y)_k = \frac{1}{n} \sum_{i+j \equiv k \pmod{n}} x_i y_j.
\]

For example, if \( G = \mathbb{Z}/3\mathbb{Z} \), the convolution of \( x = (x_0, x_1, x_2) \) and \( y = (y_0, y_1, y_2) \) is the sequence

\[
x \ast y = \frac{1}{3} (x_0y_0 + x_1y_2 + x_2y_1, x_0y_1 + x_1y_0 + x_2y_2, x_0y_2 + x_1y_1 + x_2y_0).
\]

Observe that

\[
\begin{pmatrix}
x_0 & x_2 & x_1 \\
x_1 & x_0 & x_2 \\
x_2 & x_1 & x_0
\end{pmatrix}
\begin{pmatrix}
y_0 \\
y_1 \\
y_2
\end{pmatrix}
=
\begin{pmatrix}
x_0y_0 + x_1y_2 + x_2y_1 \\
x_0y_1 + x_1y_0 + x_2y_2 \\
x_0y_2 + x_1y_1 + x_2y_0
\end{pmatrix}.
\]
The matrix
\[
\begin{pmatrix}
x_0 & x_2 & x_1 \\
x_1 & x_0 & x_2 \\
x_2 & x_1 & x_0
\end{pmatrix}
\]
is called a circulant matrix. It is obtained from a given column vector by repeatedly making cyclic permutations on the coordinates. The formula giving the convolution of two vectors in terms of a circulant matrix holds for any \( n \).

Since the vector space \( L^2(G) \) has dimension \(|G|\) and since by Proposition 9.11 the dual group \( \hat{G} \) is isomorphic to \( G \), by Proposition 9.12, the set of \( \hat{G} \) of characters of \( G \) is an orthonormal basis of \( L^2(G) \).

Another remarkable property of the dual group is that there is a natural homomorphism from \( G \) to its double dual \( \hat{\hat{G}} \). In fact, a fundamental (and famous) theorem due to Pontrjagin asserts that this map is an isomorphism (and a homeomorphism), which will allow us to define an inverse of the Fourier transform (considering functions in \( L^2(G) \)).

**Proposition 9.14.** Let \( G \) be a locally compact abelian group. Given any \( a \in G \), define the map \( \eta_a : \hat{G} \to \mathbb{C} \) by
\[
\eta_a(\chi) = \chi(a), \quad \text{evaluation at } a.
\]
The map \( \eta : G \to \hat{\hat{G}} \) given by \( \eta(a) = \eta_a \) is a continuous homomorphism from \( G \) to its double dual \( \hat{\hat{G}} \).

**Proof.** First let us check that \( \eta_a \) is a character of the group \( \hat{G} \). For any \( a \in G \), for any two characters \( \chi_1, \chi_2 \in \hat{G} \), we have
\[
\eta_a(\chi_1 \chi_2) = \chi_1(a) \chi_2(a) = \eta_a(\chi_1) \eta_a(\chi_2),
\]
so \( \eta_a \) is a homomorphism. Since the characters have range \( \mathbb{U}(1) \) and \( \eta_a(\chi) = \chi(a) \), we see that \( \eta_a : \hat{G} \to \mathbb{U}(1) \). Since \( G \) is locally compact, the map \( (a, \chi) \mapsto \chi(a) \) from \( G \times \hat{G} \) to \( \mathbb{C} \) is continuous, and this implies that \( \eta_a \) is continuous; for details, see Bourbaki [8] (Chapter 2, Section 1, No. 1).

Let us now check that \( \eta \) is a homomorphism. For all \( a, b \in G \) and all \( \chi \in \hat{G} \), we have
\[
\eta_{ab}(\chi) = \chi(ab) = \chi(a) \chi(b) = \eta_a(\chi) \eta_b(\chi),
\]
showing that \( \eta \) is a homomorphism. The fact that \( \eta \) is continuous is a consequence of the continuity of the map \( (a, \chi) \mapsto \chi(a) \); for details, see Bourbaki [8] (Chapter 2, Section 1, No. 1).

Neither the injectivity nor the surjectivity of the map \( \eta \) is easy to prove.
9.2 The Fourier Transform and the Fourier Cotransform

Given a locally compact abelian group $G$ equipped with a Haar measure $\lambda$, the Fourier transform $\mathcal{F}(f)$ of a function $f : G \to \mathbb{C}$ is a complex-valued function $\mathcal{F}(f) : \hat{G} \to \mathbb{C}$ defined not on $G$, but on its dual group $\hat{G}$ of characters, by the formula

$$\mathcal{F}(f)(\chi) = \int \overline{\chi(a)} f(a) \, d\lambda(a), \quad \chi \in \hat{G}. \quad (*)$$

The first issue is to determine when this integral converges. Since $|\chi|$ is a bounded continuous function $\chi : G \to \mathbb{T}$, by Proposition 4.35, the integral in $(*)$ is well-defined if $f \in L^1(G)$. If $G$ is not compact, in general, $L^1(G)$ is not a subspace of $L^2(G)$ and $L^2(G)$ is not a subspace of $L^1(G)$, so in general, the integral in $(*)$ does not converge if $f \in L^2(G)$.

The second issue is to determine the class of functions on $\hat{G}$ to which $\mathcal{F}(f)$ belongs.

We will see that if $f \in L^1(G)$, then $\mathcal{F}(f) \in C_0(\hat{G}; \mathbb{C})$, so in general $\mathcal{F}(f) \notin L^1(\hat{G})$. Remarkably, if $f \in L^1(G) \cap L^2(G)$, then $\mathcal{F}(f) \in L^2(\hat{G})$. In fact, because $L^1(G) \cap L^2(G)$ is dense in $L^2(G)$, the Fourier transform $\mathcal{F}$ has a unique isometric extension from $L^2(G)$ to $L^2(\hat{G})$. This is Plancherel’s theorem.

If $\mathcal{F}(f) \in L^1(\hat{G})$ or if $\mathcal{F}(f) \in L^2(\hat{G})$, for some function $f : G \to \mathbb{C}$, then the question of finding an inverse of the Fourier transform arises. Is there a transform $\mathcal{G}$ defined on $L^1(\hat{G})$ or $L^2(\hat{G})$, or some subspace of them, such that

$$f = \mathcal{G}(\mathcal{F}(f))?$$

We call this equation a Fourier inversion formula. As stated, the problem does not make sense because $\mathcal{G}$, being defined on functions in $L^1(\hat{G})$ or $L^2(\hat{G})$, is a function defined on $\hat{\hat{G}}$, and not $G$. However, we showed in Proposition 9.14 that there is a homomorphism $\eta : G \to \hat{\hat{G}}$, and by Pontrjagin duality theorem, this map is an isomorphism, so the correct way to state Fourier inversion formula is so say that

$$f = (\mathcal{G} \circ \mathcal{F})(f) \circ \eta.$$

A bit less formally, since $\eta$ is an isomorphism, by identifying $G$ and $\hat{\hat{G}}$, we can drop $\eta$ from the above formula. Then, amazingly, the inverse $\mathcal{G}$ of $\mathcal{F}$ almost looks like $\mathcal{F}$, except that there is no conjugation on the character; that is, for every function $F \in L^1(\hat{G})$, for every character $\zeta \in \hat{\hat{G}}$, we have

$$\mathcal{G}(F)(\zeta) = \int \zeta(\chi) F(\chi) \, d\hat{\lambda}(\chi), \quad \zeta \in \hat{\hat{G}}. \quad (**)$$
In the above formula, \( \hat{\lambda} \) is a Haar measure on the dual group \( \hat{G} \) (suitably normalized), and \( \chi \) is any element in \( \hat{G} \) (so, \( \chi \) is a character of \( G \)).

Following Bourbaki, it seems fair to define simultaneously two notions of Fourier transforms.

**Definition 9.3.** Let \( G \) be a locally compact group equipped with a Haar measure \( \lambda \). For every function \( f \in L^1(G) \),

1. The *Fourier transform* of \( f \) is the function \( F(f) : \hat{G} \to \mathbb{C} \) given by
   \[
   F(f)(\chi) = \int \chi(a)f(a)\,d\lambda(a), \quad \chi \in \hat{G}.
   \]
2. The *Fourier cotransform* of \( f \) is the function \( \overline{F(f)} : \hat{G} \to \mathbb{C} \) given by
   \[
   \overline{F(f)}(\chi) = \int \chi(a)f(a)\,d\lambda(a), \quad \chi \in \hat{G}.
   \]

These transforms are not independent. In fact, each one can be obtained from the other.

**Proposition 9.15.** The Fourier transform \( F \) and the Fourier cotransform \( \overline{F} \) are related as follows: for all \( f \in L^1(G) \), and all \( \chi \in \hat{G} \),

\[
F(f)(\chi) = F(f)(\chi^{-1}) = F(\hat{f})(\chi) = \overline{F(\overline{f})}(\chi).
\]

**Proof.** We have

\[
\overline{F(f)}(\chi) = \int \chi(a)f(a)\,d\lambda(a)
= \int \overline{\chi(a)}f(a)\,d\lambda(a)
= \int \overline{\chi^{-1}(a)}f(a)\,d\lambda(a) = F(f)(\chi^{-1}),
\]

\[
\int \overline{\chi(a)f(a)}\,d\lambda(a) = \int \overline{\chi(a)f(a)}\,d\lambda(a) = F(\overline{f})(\chi),
\]

and

\[
\int \overline{\chi^{-1}(a)}f(a)\,d\lambda(a) = \int \overline{\chi(a^{-1})}f(a)\,d\lambda(a)
= \int \overline{\chi(a)f(a^{-1})}\,d\lambda(a) = F(\hat{f})(\chi),
\]

so \( F(f)(\chi) = F(f)(\chi^{-1}) = F(\hat{f})(\chi) = \overline{F(\overline{f})}(\chi) \).
With these definitions, the Fourier inversion formula is
\[ f = (\mathcal{F} \circ \overline{\mathcal{F}})(f) \circ \eta. \]
In the above formula, $\overline{\mathcal{F}}$ is the Fourier cotransform on functions defined on the dual group $\hat{G}$. If we replace $\overline{\mathcal{F}}$ by $\mathcal{F}$, then we don’t quite the right formula. Instead we obtain
\[ \hat{f} = (\mathcal{F} \circ \mathcal{F})(f) \circ \eta. \]
Of course, in the above formula the leftmost occurrence of $\mathcal{F}$ is the Fourier transform on functions defined on the dual group $\hat{G}$. But if we “apply $\mathcal{F}$ four times,” we get
\[ f = (\mathcal{F} \circ \mathcal{F} \circ \mathcal{F} \circ \mathcal{F})(f) \circ \eta. \]
This looks a little silly to us and seems another justification for considering $\overline{\mathcal{F}}$ on an equal footing with $\mathcal{F}$.

Actually, in view of the isomorphism $j : \hat{G} \to X(L^1(G))$ given by Theorem 9.5, the Fourier cotransform $\overline{\mathcal{F}}$ can be viewed as the Gelfand transform from $L^1(G)$ to $X(L^1(G))$. For any $f \in L^1(G)$, the Gelfand transform $\mathcal{G}_f$ of $f$ is given by
\[ \mathcal{G}_f(\zeta) = \zeta(f), \quad \zeta \in X(L^1(G)). \]
By Theorem 9.5, the map $j : \hat{G} \to X(L^1(G))$ given by
\[ j(\chi)(f) = \zeta_\chi(f) = \int \chi(a)f(a) \, d\lambda(a), \quad \chi \in \hat{G}, \ f \in L^1(G), \]
is a homeomorphism of $\hat{G}$ onto $X(L^1(G))$. But
\[ \overline{\mathcal{F}}(f)(\chi) = \int \chi(a)f(a) \, d\lambda(a), \]
so
\[ j(\chi)(f) = \zeta_\chi(f) = \overline{\mathcal{F}}(f)(\chi), \]
and since $\mathcal{G}_f(\zeta_\chi) = \zeta_\chi(f)$ and $\zeta_\chi = j(\chi)$, we see that
\[ \mathcal{G}_f(j(\chi)) = \mathcal{G}_f(\zeta_\chi) = \overline{\mathcal{F}}(f)(\chi). \tag{†} \]

In summary, we proved the following result.

**Proposition 9.16.** Modulo the isomorphism $j : \hat{G} \to X(L^1(G))$, the Fourier cotransform $\overline{\mathcal{F}}(f)$ is the Gelfand transform $\mathcal{G}_f$ from $L^1(G)$ to $X(L^1(G))$.

This is good news because this shows that we can apply results known for the Gelfand transform to the Fourier cotransform, and thus to the Fourier transform.

The Fourier inversion formula turns out to hold in the following cases:
(1) If we identify $G$ and $\hat{G}$ using the Gelfand isomorphism theorem (and give suitably normalized Haar measures to $G$ and $\hat{G}$), then there is a unique extension of the Fourier transform $F$ to $L^2(G)$ and a unique extension of the Fourier cotransform $\overline{F}$ to $L^2(\hat{G})$, so that they are mutual inverses.

(2) If $A(G)$ is the subspace of $L^1(G)$ spanned by all functions of the form $f \ast g$ with $f, g \in L^1(G) \cap L^2(G)$, then $A(G)$ is an ideal of $L^1(G)$ contained in $L^1(G) \cap L^2(G)$; if $f \in A(G)$, then $F \in L^1(\hat{G})$, and the Fourier inversion formula holds.

(3) If $B(G) = \{ f \in L^1(G) \mid F(f) \in L^1(\hat{G}) \}$, then the restriction of $F$ to $B(G)$ is a bijection onto $B(\hat{G})$, and its inverse is the restriction of $\overline{F}$ to $B(\hat{G})$.

Thus it appears that the $L^2$ theory has the best behavior with respect to the Fourier transform. This had been observed for $G = \mathbb{R}$ long ago.

We now define the Fourier transform and the Fourier cotransform. Actually, these definitions apply to complex measures $\mu \in \mathcal{M}^1(G)$.

**Definition 9.4.** Let $G$ be a locally compact abelian group equipped with a Haar measure $\lambda$. For every complex measure $\mu \in \mathcal{M}^1(G)$, the Fourier transform $F(\mu)$ and the Fourier cotransform $\overline{F}(\mu)$ of $\mu$ are the functions $F(\mu) : \hat{G} \rightarrow \mathbb{C}$ and $\overline{F}(\mu) : \hat{G} \rightarrow \mathbb{C}$ defined on the group $\hat{G}$ by

$$F(\mu)(\chi) = \int \overline{\chi(a)} \, d\mu(a)$$

$$\overline{F}(\mu)(\chi) = \int \chi(a) \, d\mu(a),$$

for all $\chi \in \hat{G}$.

For every function $f \in L^1(G)$, the Fourier transform $F(f)$ and the Fourier cotransform $\overline{F}(f)$ of $f$ are the functions $F(f) : \hat{G} \rightarrow \mathbb{C}$ and $\overline{F}(f) : \hat{G} \rightarrow \mathbb{C}$ defined on the group $\hat{G}$ by

$$F(f)(\chi) = \int \overline{\chi(a)} f(a) \, d\lambda(a)$$

$$\overline{F}(f)(\chi) = \int \chi(a) f(a) \, d\lambda(a),$$

for all $\chi \in \hat{G}$.

**Remark:** The Fourier cotransform is also called the inverse Fourier transform by some authors, including Hewitt and Ross.

As in the case of functions,

$$\overline{F}(\mu)(\chi) = F(\mu)(\chi^{-1}) = F(\hat{\mu})(\chi) = \overline{F(\pi)(\chi)}.$$
9.2. THE FOURIER TRANSFORM AND THE FOURIER COTRANSFORM

Here is our first result. In particular, it gives the fundamental property of the Fourier transform (and cotransform), which is to convert a convolution into a (pointwise) product of functions. Recall that for a function $f: G \to \mathbb{C}$, we have $f^*(a) = \overline{f(a^{-1})}$.

**Proposition 9.17.**

1. The Fourier transform $\mathcal{F}$ and the Fourier cotransform $\mathcal{F}^*$ are involutive homomorphisms from the unital involutive Banach algebra $\mathcal{M}^1(G)$ to the unital involutive Banach algebra $\mathcal{C}_0(\hat{G}; \mathbb{C})$ of continuous bounded functions on $\hat{G}$. In particular for any two complex measures $\mu, \nu \in \mathcal{M}^1(G)$, we have

$$\mathcal{F}(\mu * \nu) = \mathcal{F}(\mu) \mathcal{F}(\nu), \quad \mathcal{F}(\mu * \nu) = \mathcal{F}(\mu) \mathcal{F}(\nu),$$

and

$$\mathcal{F}(\overline{\mu}) = (\mathcal{F}(\mu))^*, \quad \mathcal{F}(\overline{\mu}) = (\mathcal{F}(\mu))^*.$$

2. The Fourier transform $\mathcal{F}$ and the Fourier cotransform $\mathcal{F}^*$ are injective involutive homomorphisms from the involutive Banach algebra $L^1(G)$ to the involutive Banach algebra $\mathcal{C}_0(\hat{G}; \mathbb{C})$ of continuous functions on $\hat{G}$ that tend to zero at infinity. In particular for any two functions $f, g \in L^1(G)$, we have

$$\mathcal{F}(f * g) = \mathcal{F}(f) \mathcal{F}(g), \quad \mathcal{F}(f * g) = \mathcal{F}(f) \mathcal{F}(g),$$

and

$$\mathcal{F}(f^*) = (\mathcal{F}(f))^*, \quad \mathcal{F}(f^*) = (\mathcal{F}(f))^*.$$

**Proof.** (1) Observe that for any character $\chi \in \hat{G}$, we have $\mathcal{F}(\mu)(\chi) = \zeta(\chi(\mu))$, as defined in Proposition 9.4, and by this proposition, $\zeta$ is a character of the algebra $\mathcal{M}^1(G)$, so $\mathcal{F}$ is a homomorphism. Since $\mathcal{F}(\mu)(\chi) = \mathcal{F}(\mu)(\chi^{-1})$, the Fourier transform $\mathcal{F}$ is also a homomorphism. Since $|\chi(a)| = 1$ for all $a \in G$, we have

$$|\mathcal{F}(\mu)(\chi)| = \left| \int \overline{\chi(a)} d\mu(a) \right| \leq \|\mu\|,$$

and the same argument applies to $\mathcal{F}(\mu)$. Therefore $\mathcal{F}(\mu)$ and $\mathcal{F}(\mu)$ are bounded functions on $\hat{G}$.

Finally, by the definition of the compact-open topology on $\hat{G}$, if $\chi$ tends to $\chi_0$ in $\hat{G}$, then the function $\chi$ defined on $G$ tends to $\chi_0$ uniformly on every compact set, while being bounded by 1. It follows that $\mathcal{F}(\mu)(\chi)$ tends to $\mathcal{F}(\mu)(\chi_0)$. Thus $\mathcal{F}(\mu)$ is continuous. A similar reasoning shows that $\mathcal{F}(\mu)$ is continuous.

(2) Since $L^1(G)$ is a subalgebra of $\mathcal{M}^1(G)$, we see immediately that $\mathcal{F}$ and $\mathcal{F}$ are homomorphisms. Since $\hat{G}$ and $\mathcal{X}(L^1(G))$ are homeomorphic, $\mathcal{F}(f)$ can be identified with the Gelfand transform $\mathcal{G}_f$, by the remark following Theorem 8.20, the Gelfand transform $\mathcal{G}$ maps
L^1(G) into C_0(X(L^1(G)) \cong C_0(\hat{G})$, so $\mathcal{F}$ maps $L^1(G)$ to $C_0(\hat{G})$. Since $\mathcal{F}(\mu)(\chi) = \mathcal{F}(\mu)(\chi^{-1})$, the Fourier transform $\mathcal{F}$ also maps $L^1(G)$ to $C_0(\hat{G})$.

By Proposition 8.24, the Gelfand transform $\mathcal{G}$ is injective because it can be shown that $L^1(G)$ has radical $(0)$; see Proposition 8.38. As a consequence $\overline{\mathcal{F}}$ is injective, and since $\mathcal{F}(\mu)(\chi) = \mathcal{F}(\mu)(\chi^{-1})$, the Fourier transform is also injective.

**Remark:** If $G = \mathbb{R}^n$, the fact that the Fourier transform $\mathcal{F}(f)$ is a continuous function that tends to zero at infinity is known as the Riemann–Lebesgue lemma; see Folland [32] (Chapter 8, Theorem 8.22).

**Example 9.2.**

(1) Let $G = \mathbb{R}$, and equip $G$ with the Lebesgue measure $dx$. Recall from Proposition 9.8 that $\hat{\mathbb{R}}$ is isomorphic to $\mathbb{R}$. If we choose the characters to be the maps $x \mapsto e^{iyx}$ for some $y \in \mathbb{R}$, then it turns out that for the Fourier inversion formula to come out right, $\hat{\mathbb{R}}$ needs to be equipped with $dx/2\pi$. The Fourier transform on $\mathbb{R}$ is

$$\hat{f}(x) = \mathcal{F}(f)(x) = \int f(y)e^{-ixy} \, dy,$$

and on $\hat{\mathbb{R}}$ it is

$$\mathcal{F}(f)(x) = \int f(y)e^{-ixy} \frac{dy}{2\pi}.$$

The Fourier cotransform is obtained by changing the sign $-$ in the exponent to a $+$ sign. Then the inversion formula is

$$f(x) = \int \hat{f}(y)e^{ixy} \frac{dy}{2\pi}.$$

This is choice made in Malliavin [55].

Another choice that works is to equip both $\mathbb{R}$ and $\hat{\mathbb{R}}$ with the normalized Lebesgue measure $dx/\sqrt{2\pi}$, as in Rudin [66, 67], and in Chapter 7.

With the choice of characters $x \mapsto e^{2\pi ixy}$, the Lebesgue measure is self-dual; see Folland [32]. The Fourier transform on $\mathbb{R}$ and $\hat{\mathbb{R}}$ is

$$\hat{f}(x) = \mathcal{F}(f)(x) = \int f(y)e^{-2\pi ixy} \, dy.$$

The Fourier cotransform is obtained by changing the sign $-$ in the exponent to a $+$ sign, and the inversion formula is

$$f(x) = \int \hat{f}(y)e^{2\pi ixy} \, dy.$$

The trick to pick the right normalization factor is that the Fourier transform $\hat{g}$ of the function $g(x) = e^{-\pi x^2}$ should be $g$ itself; $\hat{g} = g$; see Folland [32] (Chapter 8, Proposition 8.24).
(2) Let $G = \mathbb{T}$, equipped with the Haar measure $d\nu_1/2\pi$ inherited from $\mathbb{R}$, by viewing $\mathbb{T}$ as $\mathbb{R}/2\pi\mathbb{Z}$, so that $\mathbb{T}$ has measure 1; see Example 6.4. The characters of $\mathbb{T}$ are the maps $e^{i\theta} \mapsto e^{im\theta}$, with $m \in \mathbb{Z}$. We equip $\mathbb{Z}$ with the counting measure, and the characters are the maps $m \mapsto e^{im\theta}$, with $\theta \in \mathbb{R}/2\pi\mathbb{Z}$. The Fourier transform $\hat{f} = \mathcal{F}(f)$ of a function $f \in L^1(\mathbb{T})$ is the $\mathbb{Z}$-indexed sequence whose $m$th element $\hat{f}_m$, the $m$th Fourier coefficient of $f$, is given by

$$\hat{f}_m = \mathcal{F}(f)(m) = \int_{-\pi}^{\pi} f(\theta)e^{-im\theta} d\theta/2\pi.$$ 

The Fourier transform $\mathcal{F}(c)$ of a sequence $c = (c_m)_{m \in \mathbb{Z}} \in L^1(\mathbb{Z}) = l^1(\mathbb{Z})$ is the function

$$f(e^{i\theta}) = \mathcal{F}(c)(e^{i\theta}) = \sum_{m \in \mathbb{Z}} c_m e^{-im\theta},$$

traditionally called the Fourier series associated with $c = (c_m)_{m \in \mathbb{Z}}$. The Fourier co-transform is obtained by changing the sign $-$ in the exponent to a $+$ sign. The inversion formula is

$$f(e^{i\theta}) = \sum_{m \in \mathbb{Z}} \hat{f}(m)e^{im\theta} d\theta.$$ 

(3) Consider the group $G = (\mathbb{R}_+^*, *)$, the group of positive reals under multiplication. This group comes up in the Mellin transform. Since we have the isomorphism $\varphi: \mathbb{R} \to \mathbb{R}_+^*$ given by $\varphi(x) = e^x$, where the group operation on $\mathbb{R}$ is addition, it is easy to figure out the characters of $\mathbb{R}_+^*$, which are homomorphisms $\chi: \mathbb{R}_+^* \to \mathbb{U}(1)$. Indeed, there is a bijection between the characters $\zeta \in \hat{\mathbb{R}}$, which are homomorphisms $\zeta: \mathbb{R} \to \mathbb{U}(1)$, and the characters $\chi \in \hat{\mathbb{R}}_+^*$, with $\chi: \mathbb{R}_+^* \to \mathbb{U}(1)$, given by

$$\zeta = \chi \circ \varphi.$$ 

By Proposition 9.8, every character $\zeta \in \hat{\mathbb{R}}$ is of the form $\zeta(x) = e^{ixx}$ for some $x \in \mathbb{R}$, and since $\chi(y) = \zeta(\varphi^{-1}(y)) = \zeta(\log y)$ for any $y > 0$, we get

$$\chi(y) = \zeta(\log y) = e^{ix\log y} = (e^{\log y})^x = y^x.$$ 

Therefore, the characters of $\mathbb{R}_+^*$ are of the form

$$y \mapsto y^x, \quad y \in \mathbb{R}_+^*,$$

for some fixed $x \in \mathbb{R}$, which shows that

$$\hat{\mathbb{R}}_+^* \cong \mathbb{R}.$$ 

We can also easily determine the characters of $\hat{\mathbb{R}}_+^*$.
We know from Proposition 9.8 that every $\zeta \in \hat{\mathbb{R}}$ is of the form $\zeta(z) = e^{ixz}$ for some $x \in \mathbb{R}$. If we let $y = e^x$, then $y > 0$, and

$$\zeta(z) = e^{ixz} = y^iz.$$

Therefore, every character of $\hat{\mathbb{R}}^*_+$ is of the form

$$z \mapsto y^iz, \quad z \in \mathbb{R},$$

for some fixed $y \in \mathbb{R}^*_+$. Thus

$$\hat{\mathbb{R}}^*_+ \cong \mathbb{R}^*_+,$$

which confirms the general theory. If we pick the left-invariant Haar measure $dy/y$ on $\mathbb{R}^*_+$ (see Example 6.3), then the Fourier transform on $L^1(\mathbb{R}^*_+)$ is given by

$$\mathcal{F}(f)(x) = \int_0^\infty y^{-ix}f(y) \frac{dy}{y},$$

for any $f \in L^1(\mathbb{R}^*_+)$ and all $x \in \mathbb{R}$. This is one of the formulations of the Mellin transform, often denoted $\mathcal{M}(f)$.

If we give $\hat{\mathbb{R}}^*_+$ the Haar measure $dx/2\pi$, then the Fourier cotransform on $L^1(\hat{\mathbb{R}}^*_+) = L^1(\mathbb{R})$ is given by

$$\mathcal{F}(g)(y) = \frac{1}{2\pi} \int_{\mathbb{R}} y^{ix}g(x) \, dx,$$

for any $g \in L^1(\mathbb{R})$ and all $y \in \mathbb{R}^*_+$. This is one of the formulations of the inverse Mellin transform, often denoted $\mathcal{M}^{-1}(g)$. The inversion formula is

$$f(y) = \frac{1}{2\pi} \int_{\mathbb{R}} y^{ix}\mathcal{M}(f)(x) \, dx, \quad y \in \mathbb{R}^*_+.$$

Observe that on $L^1(\mathbb{R}^*_+)$, convolution is given by

$$(f * g)(y) = \int_0^\infty f(z)g\left(\frac{y}{z}\right) \frac{dz}{z},$$

and by the general theory,

$$\mathcal{M}(f * g) = \mathcal{M}(f)\mathcal{M}(g).$$
9.3 The Fourier Transform on a Finite Abelian Group

Let $G$ be a finite locally compact abelian group, and as Example 9.1, assume that the Haar measure is normalized so that $\lambda(G) = 1$. It is possible and very instructive to work out explicitly the Fourier transform on $L^2(G)$ and the Fourier cotransform on $L^2(\hat{G})$. We will also prove directly that Fourier inversion holds.

Recall that the inner product $\langle x, y \rangle$ of $x, y \in L^2(G)$ is given by

$$\langle x, y \rangle = \frac{1}{|G|} \sum_{a \in G} x_a \overline{y_a}.$$ 

Then the Fourier transform and the Fourier cotransform of $x = (x_a)_{a \in G} \in L^2(G)$ are given by

$$\mathcal{F}(x)(\chi) = \frac{1}{|G|} \sum_{a \in G} x_a \chi(a)$$
$$\mathcal{F}(x)(\chi) = \frac{1}{|G|} \sum_{a \in G} x_a \chi(a).$$

where $\chi: G \to \mathbb{T}$ is a character of $G$.

Observe that

$$\mathcal{F}(x)(\chi) = \frac{1}{|G|} \sum_{a \in G} x_a \overline{\chi(a)} = \langle x, \chi \rangle.$$

Recall that $\hat{G} \cong G$ by Proposition 9.11, so we can write $\hat{G} = \{\chi_1, \ldots, \chi_n\}$, with $n = |G|$. We may call $\hat{x}(\chi_i) = \mathcal{F}(x)(\chi_i) = \langle x, \chi_i \rangle$ the $i$th Fourier coefficient of $x$.

**Proposition 9.18.** (Fourier inversion formula) Let $G$ be a finite abelian group of order $n$, and let $\hat{G} = \{\chi_1, \ldots, \chi_n\}$. For every $x \in L^2(G)$, we have

$$x = \sum_{j=1}^{n} \mathcal{F}(x)(\chi_j) \chi_j = \sum_{j=1}^{n} \hat{x}(\chi_j) \chi_j,$$

with

$$\mathcal{F}(x)(\chi_j) = \langle x, \chi_j \rangle = \frac{1}{|G|} \sum_{a \in G} x_a \chi_j(a).$$

**Proof.** By Proposition 9.12 the characters $\{\chi_1, \ldots, \chi_n\}$ are orthonormal, and since $L^2(G)$ is a vector space of dimension $n$, they form a basis. Consequently we can write

$$x = \sum_{j=1}^{n} c_j \chi_j$$
for some $c_j \in \mathbb{C}$. Taking the inner product with $\chi_j$, we obtain
\[ \langle x, \chi_j \rangle = c_j \langle \chi_j, \chi_j \rangle = c_j, \]
which concludes the proof.

Let us equip $\hat{G}$ with the counting measure $\hat{\lambda}$ normalized so that $\hat{\lambda}(\hat{G}) = |\hat{G}| = |G|$. Then the integral of a function $H \in L^2(\hat{G})$ is given by
\[ \int H(\chi) \, d\hat{\lambda}(\chi) = \sum_{\chi \in \hat{G}} H_{\chi}, \]
the inner product of $F, H \in L^2(\hat{G})$ is given by
\[ \langle F, H \rangle = \sum_{\chi \in \hat{G}} F_{\chi} \overline{H_{\chi}}, \]
and the Fourier transform and the Fourier cotransform of $H = (H_{\chi})_{\chi \in \hat{G}} \in L^2(\hat{G})$ are given by
\[ \mathcal{F}(H)(\zeta) = \sum_{\chi \in \hat{G}} H_{\chi} \overline{\zeta(\chi)} \]
\[ \mathcal{F}(H)(\zeta) = \sum_{\chi \in \hat{G}} H_{\chi} \zeta(\chi), \]
where $\zeta : \hat{G} \to \mathbb{T}$ is a character of $\hat{G}$. Observe that the factor $1/|G|$ is missing.

Recall that $\eta_a \in \hat{G} \subseteq L^2(\hat{G})$ is a character on $\hat{G}$ such that $\eta_a(\chi) = \chi(a)$.

**Proposition 9.19.** *(Fourier inversion)* For any $x \in L^2(G)$ and any $a \in G$, we have
\[ (\mathcal{F} \circ \mathcal{F})(x)(\eta_a) = x_a, \]
that is,
\[ x = (\mathcal{F} \circ \mathcal{F})(x) \circ \eta. \]

**Proof.** Let us compute $(\mathcal{F} \circ \mathcal{F})(x)(\eta_a) = \mathcal{F}(\mathcal{F}(x))(\eta_a)$, for any $x = (x_b)_{b \in G}$ and any $a \in G$. 
Since \( \eta_a(\chi) = \chi(a) \), we have
\[
(F \circ F)(x)(\eta_a) = F(F(x))(\eta_a)
\]
\[
= \sum_{\chi \in \hat{G}} F(x)_\chi \eta_a(\chi)
\]
\[
= \frac{1}{|G|} \sum_{\chi \in \hat{G}} \sum_{b \in G} x_b \overline{\chi(b)} \chi(a)
\]
\[
= \frac{1}{|G|} \sum_{b \in G} x_b \sum_{\chi \in \hat{G}} \chi(a) \overline{\eta_b(\chi)}
\]
\[
= \frac{1}{|G|} \sum_{b \in G} x_b \langle \eta_a, \eta_b \rangle.
\]
But \( \eta_a, \eta_b \in \hat{G} \subseteq L^2(\hat{G}) \), so by Proposition 9.12 (applied to \( \hat{G} \)), all terms \( \langle \eta_a, \eta_b \rangle \) are zero if \( a \neq b \), and \( \langle \eta_a, \eta_a \rangle = |G| \), because the Haar measure on \( \hat{G} \) is normalized so that \( \lambda(\hat{G}) = |\hat{G}| = |G| \), so the factor \( 1/|G| \) is missing. Therefore,
\[
(F \circ F)(x)(\eta_a) = \frac{1}{|G|} \sum_{b \in G} x_b \langle \eta_a, \eta_b \rangle = \frac{1}{|G|} x_a \langle \eta_a, \eta_a \rangle = \frac{1}{|G|} x_a |G| = x_a,
\]
which proves that
\[
x = (F \circ F)(x) \circ \eta,
\]
namely, the Fourier inversion formula. \( \square \)

Observe that in order for the inversion formula to be correct, the normalization factor of the Haar measure \( \lambda \) on \( G \) and the normalization factor of the Haar measure \( \hat{\lambda} \) on \( \hat{G} \) have to be chosen carefully. In our case, we chose \( \lambda(G) = 1 \) and \( \hat{\lambda}(\hat{G}) = |G| \). Instead we could have chosen \( \lambda(G) = |G| \) and \( \hat{\lambda}(\hat{G}) = 1 \). The reader should check that the self-dual choice \( \lambda(G) = \hat{\lambda}(\hat{G}) = \sqrt{|G|} \) also works.

The reader should also check that if we use \( \mathcal{F} \) instead of \( F \) on \( L^2(\hat{G}) \), then we get
\[
\tilde{x} = (\mathcal{F} \circ \mathcal{F})(x) \circ \eta,
\]
where \( \tilde{x}(a) = x_{a^{-1}} \).

**Proposition 9.20.** (Plancherel theorem) For all \( x, y \in L^2(G) \), we have
\[
\langle x, y \rangle = \langle F(x), F(y) \rangle.
\]
As a consequence, \( L^2(G) \) and \( L^2(\hat{G}) \) are isometric.
CHAPTER 9. ANALYSIS ON LOCALLY COMPACT ABELIAN GROUPS

Proof. We have

\[
\langle F(x), F(y) \rangle = \sum_{\chi \in \hat{G}} F(x) \overline{F(y)} \chi
\]

\[
= \frac{1}{|G|^2} \sum_{a \in G} \sum_{b \in G} \sum_{\chi \in \hat{G}} x_a \overline{\chi(a)} y_b \chi(b)
\]

\[
= \frac{1}{|G|^2} \sum_{a \in G} \sum_{b \in G} x_a y_b \sum_{\chi \in \hat{G}} \chi(b) \overline{\chi(a)}
\]

\[
= \frac{1}{|G|^2} \sum_{a \in G} \sum_{b \in G} x_a y_b \sum_{\chi \in \hat{G}} \eta_b(\chi) \overline{\eta_a(\chi)}
\]

\[
= \frac{1}{|G|^2} \sum_{a \in G} \sum_{b \in G} x_a y_b \langle \eta_b, \eta_a \rangle
\]

\[
= \frac{1}{|G|^2} \sum_{a \in G} x_a y_a |G|
\]

\[
= \langle x, y \rangle.
\]

Again, we used the fact that in $L^2(\hat{G})$, the measure $\hat{\lambda}$ is normalized so that $\hat{\lambda}(\hat{G}) = |G|$, so $\langle \eta_a, \eta_a \rangle = |G|$ (and $\langle \eta_a, \eta_b \rangle = 0$ whenever $a \neq b$).

\[\square\]

Proposition 9.20 is a special case of Plancherel theorem.

**Proposition 9.21.** (Convolution rule) For all $x, y \in L^2(G)$, we have

\[F(x * y) = F(x)F(y).\]

Proof. Recall from Example 9.1 that the convolution $x * y$ of $x, y \in L^2(G)$ is given by

\[(x * y)_a = \frac{1}{|G|} \sum_{b \in G} x_b y_{b^{-1}a} = \frac{1}{|G|} \sum_{b, c \in G} y_{d^{-1}a}.
\]

Thus we have

\[F(x * y)(\chi) = \frac{1}{|G|} \sum_{c \in G} (x * y)_c \overline{\chi(c)}
\]

\[= \frac{1}{|G|} \sum_{c \in G} \frac{1}{|G|} \sum_{d \in G} x_d y_{d^{-1}c} \overline{\chi(c)}.
\]
If we replace \( c \) by \( dc \), since \( G \) is a group the sum does not change, and we get

\[
\frac{1}{|G|} \sum_{c \in G} \frac{1}{|G|} \sum_{d \in G} x_d y_{d^{-1}} c \chi(c) = \frac{1}{|G|} \sum_{dc \in G} \frac{1}{|G|} \sum_{d \in G} x_d y_{d^{-1}} dc \chi(dc)
\]

\[
= \frac{1}{|G|} \sum_{dc \in G} x_d \frac{1}{|G|} \sum_{c \in G} y_c \chi(d) \chi(c)
\]

\[
= \frac{1}{|G|} \sum_{d \in G} x_d \chi(d) \frac{1}{|G|} \sum_{c \in G} y_c \chi(c)
\]

\[
= \mathcal{F}(x)(\chi)\mathcal{F}(y)(\chi).
\]

Therefore, we proved that

\[
\mathcal{F}(x * y) = \mathcal{F}(x)\mathcal{F}(y).
\]

### 9.4 Dirichlet Characters

Let \( G = (\mathbb{Z}/m\mathbb{Z})^* \) equipped with the counting measure. The group \((\mathbb{Z}/m\mathbb{Z})^*\) is the multiplicative group of units in \(\mathbb{Z}/m\mathbb{Z}\), that is, those elements \( a \in \mathbb{Z}/m\mathbb{Z} \) such that there is some \( b \in \mathbb{Z}/m\mathbb{Z} \) and \( ab = ba = 1 \) \((\text{mod } m)\). It is well known that \( a \in \mathbb{Z}/m\mathbb{Z} \) is a unit if and only if \( \gcd(a, m) = 1 \), and the group \((\mathbb{Z}/m\mathbb{Z})^*\) has \( \varphi(m) \) elements, where \( \varphi \) is the Euler phi-function \((\varphi(m) \) is the number of integers \( a \), with \( 1 \leq a \leq m \), such that \( \gcd(a, m) = 1 \)). The group \((\mathbb{Z}/m\mathbb{Z})^*\) is not always cyclic, (for example, if \( m = 2^k, k \geq 3 \)), but Gauss determined when this happens. However, when \((\mathbb{Z}/m\mathbb{Z})^*\) is cyclic, finding a generator for it is computationally hard.

The characters \( \chi: (\mathbb{Z}/m\mathbb{Z})^* \to \mathbb{C} \) are called Dirichlet characters. Every such character can be extended to a function \( \chi: \mathbb{Z} \to \mathbb{C} \) given by

\[
\chi(n) = \begin{cases} 
\chi(n) & \text{if } \gcd(m, n) = 1 \\
0 & \text{otherwise.}
\end{cases}
\]

It is immediately verified that such functions are multiplicative, which means that

\[
\chi(rs) = \chi(r)\chi(s) \quad \text{for all } r, s \in \mathbb{Z}.
\]

They are also periodic with period \( m \) \((\chi(n + m) = \chi(n) \text{ for all } n \in \mathbb{Z})\). These functions are called Dirichlet characters modulo \( m \). The trivial Dirichlet character is the Dirichlet character such that \( \chi_0(m) = 1 \) iff \( m \) and \( n \) are relatively prime, and \( 0 \) otherwise.

For every \( \ell \in \mathbb{N} - \{0\} \), define the function \( \delta_\ell: (\mathbb{Z}/m\mathbb{Z})^* \to \mathbb{C} \) given by

\[
\delta_\ell(n) = \begin{cases} 
1 & \text{if } n \equiv \ell \pmod{m} \\
0 & \text{otherwise.}
\end{cases}
\]
By the Fourier inversion formula (Proposition 9.18) we can write

$$\delta_\ell(n) = \sum_{\chi \in \hat{G}} F(\delta_\ell)(\chi) \chi(n),$$

with

$$F(\delta_\ell)(\chi) = \frac{1}{\varphi(m)} \sum_{k \in G} \delta_\ell(k) \chi(k) = \frac{1}{\varphi(m)} \chi(\ell).$$

Therefore,

$$\delta_\ell(n) = \frac{1}{\varphi(m)} \sum_{\chi \in \hat{G}} \chi(n) \overline{\chi(\ell)}.$$

As the characters, the functions $\delta_\ell$ can be extended to $\mathbb{Z}$, by setting $\delta_\ell(n) = 0$ if $m$ and $n$ are not relatively prime. The above shows that

$$\delta_\ell(n) = \frac{1}{\varphi(m)} \sum_{\chi} \chi(n) \overline{\chi(\ell)}, \quad n \in \mathbb{Z},$$

where the sum is over the Dirichlet characters modulo $m$.

The above result is one of the steps in the proof of Dirichlet’s theorem on arithmetic progressions of integers $mk + \ell$ with $\gcd(\ell, m) = 1$ and $k \in \mathbb{N}$, that says that such a sequence contains infinitely many primes.

Dirichlet’s theorem is a consequence of the fact that the sum

$$\sum_{\substack{p \equiv \ell \pmod{m} \\ p \text{ prime}}} \frac{1}{p},$$

is infinite. This follows from the fact the the limit of the sum

$$\sum_{\substack{p \equiv \ell \pmod{m} \\ p \text{ prime}}} \frac{1}{p^s}$$

tends to infinity when $s > 1$ tends to 1. To simplify notation, write

$$\sum_{\substack{p \equiv \ell \\ p \text{ prime}}} \frac{1}{p^s},$$

where it is understood that $p$ is prime and congruent to $\ell$ modulo $m$.

We can write

$$\sum_{\substack{p \equiv \ell \\ p \text{ prime}}} \frac{1}{p^s} = \sum_{p \text{ prime}} \frac{\delta_\ell(p)}{p^s} = \frac{1}{\varphi(m)} \sum_{\chi} \overline{\chi(\ell)} \sum_{p \text{ prime}} \frac{\chi(p)}{p^s}.$$
We can divide the above sum into two parts depending on whether or not \( \chi \) is trivial, and we get

\[
\sum_{p \equiv \ell} \frac{1}{p^s} = \frac{1}{\varphi(m)} \sum_p \frac{\chi_0(p)}{p^s} + \frac{1}{\varphi(m)} \sum_{\chi \neq \chi_0} \chi(\ell) \sum_p \frac{\chi(p)}{p^s},
\]

where the sums on the right-hand side are over all primes \( p \). Since there are only finitely many primes dividing \( m \), the sum

\[
\sum_{p \nmid m} \frac{1}{p^s}
\]

tends to infinity when \( s > 1 \) tends to 1, because the series

\[
\sum_p \frac{1}{p}
\]

diverges. This is a classical result of number theory going back to Euler (1737); see Stein and Shakarchi [77] (Chapter 8), or Apostol [2] (Chapter 1, Section 1.6).

If we could prove that the sum

\[
\sum_{p \text{ prime}} \frac{\chi(p)}{p^s}
\]
remains bounded when \( s > 1 \) tends to 1, we would be done, because then

\[
\sum_{p \equiv \ell} \frac{1}{p^s}
\]

tends to infinity as \( s > 1 \) tends to 1.

The above suggests studying the behavior of the functions \( L(s, \chi) \) given by

\[
L(s, \chi) = \sum_{n=1}^{\infty} \frac{\chi(n)}{n^s}, \quad s > 1,
\]

where \( \chi \) is a Dirichlet character, called \( L \)-functions. The series \( L(s, \chi) \) is absolutely convergent for \( s > 1 \). Actually, if \( \chi \neq \chi_0 \), the series \( L(s, \chi) \) converges for \( s > 0 \) and is continuously differentiable for \( s > 0 \); see Stein and Shakarchi [77] (Chapter 8). There is also a remarkable product formula: For \( s > 1 \), we have

\[
L(s, \chi) = \prod_{p \text{ prime}} \frac{1}{1 - \chi(p)p^{-s}}.
\]
The above is a generalization of Euler’s formula for expressing the zeta function
\[ \zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s} \]
as the infinite product
\[ \prod_{p \text{ prime}} \frac{1}{1 - p^{-s}}. \]

In a tour de force, Dirichlet proved that \( L(1, \chi) \) is finite and that \( L(1, \chi) \neq 0 \), if \( \chi \) is not the trivial character. By taking the logarithm of both sides of the product formula and using some properties of the log function one obtains
\[ \log L(s, \chi) = \sum_{p \text{ prime}} \frac{\chi(p)}{p^s} + O(1), \]
so if \( L(1, \chi) \) is nonzero, then the sum
\[ \sum_{p \text{ prime}} \frac{\chi(p)}{p^s} \]
remains bounded when \( s > 1 \) tends to 1, and the proof of Dirichlet’s theorem is completed.

The above sketch lacks rigor on several fronts, and a rigorous proof involves a lot of rather difficult technical details, the hardest step being the proof that \( L(1, \chi) \neq 0 \), if \( \chi \) is not the trivial character. We refer the interested reader to Stein and Shakarchi [77] (Chapter 8), Apostol [2] (Chapters 6 and 7), or Serre [75] (Chapter VI).

### 9.5 Fourier transform and Cotransform in Terms of Matrices

In this section we formulate the Fourier transform (and cotransform) in terms of matrices. Write \( n = |G| \), and denote by \([G \to \mathbb{C}]\) the vector space of all functions \( x: G \to \mathbb{C} \). Since \( G \) is finite, we have \( L^1(G) = L^2(G) = [G \to \mathbb{C}] \). Of course, the space \([G \to \mathbb{C}]\) is isomorphic to \( \mathbb{C}^n \), but it is better to stick with \([G \to \mathbb{C}]\). Our goal is to extend the Fourier transform \( \mathcal{F} \) on \( L^2(G) = [G \to \mathbb{C}] \) (defined on \( \hat{G} \)) to a bilinear form on \([G \to \mathbb{C}]^*\), and to extend the Fourier transform \( \mathcal{F}^* \) on \( L^2(\hat{G}) = [\hat{G} \to \mathbb{C}] \) (defined on \( \hat{\hat{G}} \)) to a bilinear form on \([G \to \mathbb{C}]^{**}\). We will also prove the Fourier inversion theorem for these extensions.

For any \( a \in G \), we denote by \( e_a \in [G \to \mathbb{C}] \) the map given by
\[ e_a(b) = \begin{cases} 1 & \text{if } b = a \\ 0 & \text{if } b \neq a. \end{cases} \]
If we order \( G \) as \( (a_1, \ldots, a_n) \), then \( (e_{a_1}, \ldots, e_{a_n}) \) is a basis of \([G \to \mathbb{C}]\). Viewed as an element of \( \mathbb{C}^n \), the vector \( e_{a_i} \) corresponds to the canonical \( i \)th basis vector \( e_i \). For any function \( x \in [G \to \mathbb{C}] \), we can write
\[
x = \sum_{i=1}^{n} x(a_i) e_{a_i}.
\]
For simplicity of notation, we write \( x_i \) (or \( x_{a_i} \)) instead of \( x(a_i) \). To show that Fourier inversion holds, we need to view a function \( x \in [G \to \mathbb{C}] \) as a linear form in \([G \to \mathbb{C}]^*\).

**Definition 9.5.** Let \( G \) be a finite abelian group, and write \( G = \{a_1, \ldots, a_n\} \). Every vector \( x \in [G \to \mathbb{C}] \) determines uniquely the linear form \( \tilde{x} \in [G \to \mathbb{C}]^* \) defined on the basis \((e_{a_1}, \ldots, e_{a_n})\) of \([G \to \mathbb{C}]\) by
\[
\tilde{x}(e_{a_j}) = x(e_{a_j}).
\]
Since \( G \) is finite, we know that \( \hat{G} \) is isomorphic to \( G \), and we order \( \hat{G} \) as \( (\chi_1, \ldots, \chi_n) \). Every character \( \chi_i \in \hat{G} \) is a function \( \chi_i : G \to \mathbb{C} \), so \( \chi_i \in [G \to \mathbb{C}] \), and the orthogonality conditions of Proposition 9.12 imply that \( (\chi_1, \ldots, \chi_n) \) are linearly independent.

As above, every character \( \chi_i \) determines uniquely the linear form \( \tilde{\chi}_i \in [G \to \mathbb{C}]^* \) defined on the basis \((e_{a_1}, \ldots, e_{a_n})\) by
\[
\tilde{\chi}_i(e_{a_j}) = \chi_i(a_j).
\]
If there is a linear dependence
\[
\sum_{j=1}^{n} \alpha_j \tilde{\chi}_j = 0,
\]
by applying the above linear form to the \( n \) vector \( e_{a_j} \), we obtain the \( n \) equations
\[
\sum_{j=1}^{n} \alpha_j \tilde{\chi}_j(e_{a_j}) = \sum_{j=1}^{n} \alpha_j \chi_j(a_j) = 0, \quad j = 1, \ldots, n,
\]
which are equivalent to
\[
\sum_{j=1}^{n} \alpha_j (\chi_j(a_1), \ldots, \chi_j(a_n)) = 0,
\]
that is
\[
\sum_{j=1}^{n} \alpha_j \chi_j = 0,
\]
and since \( (\chi_1, \ldots, \chi_n) \) are linearly independent, we must have \( \alpha_1 = \cdots = \alpha_n = 0 \), and so the linear forms \( (\tilde{\chi}_1, \ldots, \tilde{\chi}_n) \) are also linearly independent.

Our first goal is to extend the Fourier transform \( \mathcal{F} \) on \( L^2(G) = [G \to \mathbb{C}] \) to a bilinear form on \([G \to \mathbb{C}]^*\). Given any \( x \in [G \to \mathbb{C}] \), its Fourier transform is the map \( \mathcal{F}(f) : \hat{G} \to \mathbb{C} \) given by
\[
\mathcal{F}(x)(\chi_i) = \frac{1}{n} \sum_{j=1}^{n} x(a_j) \overline{\chi_i(a_j)} = \frac{1}{n} \sum_{j=1}^{n} \tilde{x}(e_{a_j}) \overline{\chi_i(e_{a_j})}.
\]
Definition 9.6. Let $G$ be a finite abelian group, and write $G = \{a_1, \ldots, a_n\}$. For all $f, \gamma \in [G \to \mathbb{C}]^*$ we define the Fourier transform $F$ as the bilinear form on $[G \to \mathbb{C}]^*$ given by

$$F(f)(\gamma) = \frac{1}{n} \sum_{j=1}^{n} f(e_{a_j})\overline{\gamma(e_{a_j})}.$$ 

If $f = \widetilde{x}$ and $\gamma = \widetilde{\chi}_i$, the right-hand side of the above definition is equal to the right-hand side of (*), so this definition extends the definition of the Fourier transform on $[G \to \mathbb{C}]$ and $\hat{G}$.

If we write $f = \sum_{i=1}^{n} x_ie_{a_i}^*$, then we have

$$F(f)(\gamma) = \sum_{i=1}^{n} x_i e_{a_i}^* F(e_{a_i}^*)(\gamma) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} x_i e_{a_i}^* (e_{a_j}) \overline{\gamma(e_{a_j})} = \frac{1}{n} \sum_{j=1}^{n} x_j \overline{\gamma(e_{a_j})}.$$ 

Then we get

$$F(f) \left( \sum_{i=1}^{n} w_i \widetilde{\chi}_i \right) = \sum_{i=1}^{n} w_i F(f)(\widetilde{\chi}_i)$$

$$= \frac{1}{n} \sum_{i=1}^{n} w_i \sum_{j=1}^{n} x_j \overline{\chi}_i(e_{a_j})$$

$$= \frac{1}{n} \sum_{i=1}^{n} w_i \sum_{j=1}^{n} x_j \overline{\chi}_i(a_j)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \left( \sum_{j=1}^{n} \overline{\chi}_i(a_j) x_j \right) w_i.$$ 

Definition 9.7. Let $G$ be a finite abelian group of order $n$, and write $G = \{a_1, \ldots, a_n\}$. Define the $n \times n$ matrix $F = (F_{ij})$, called the Fourier matrix of $G$ with respect to the characters $\{\chi_1, \ldots, \chi_n\}$ of $G$, by

$$F_{ij} = \overline{\chi_i(a_j)}.$$ 

For example, if $n = 3$, we have the matrix

$$F = \begin{pmatrix} \overline{\chi_1(a_1)} & \overline{\chi_1(a_2)} & \overline{\chi_1(a_3)} \\ \overline{\chi_2(a_1)} & \overline{\chi_2(a_2)} & \overline{\chi_2(a_3)} \\ \overline{\chi_3(a_1)} & \overline{\chi_3(a_2)} & \overline{\chi_3(a_3)} \end{pmatrix}.$$ 

We obtained the following result.
9.5. FOURIER TRANSFORM AND COTRANSFORM IN TERMS OF MATRICES

Proposition 9.22. If \( f \in [G \to \mathbb{C}]^* \) is expressed over the basis \( (e^*_a, \ldots, e^*_n) \) as \( f = \sum_{j=1}^n x_j e^*_a \), and if \( \gamma \in [G \to \mathbb{C}]^* \) is expressed over the basis \( (\chi_1, \ldots, \chi_n) \) as \( \gamma = \sum_{i=1}^n w_i \chi_i \), then the value \( F(f)(\gamma) \) of the bilinear form \( F \) is given by

\[
F(f)(\gamma) = \frac{1}{n} w^\top Fx = \frac{1}{n} x^\top F^\top w.
\]

As a linear map from \([G \to \mathbb{C}]^* \) to \( \mathbb{C} \), the matrix of \( F(f) \) over the basis \( (\chi_1, \ldots, \chi_n) \) is the row vector \( \frac{1}{n} x^\top F^\top \). As an element of \([G \to \mathbb{C}]^{**} \cong [G \to \mathbb{C}] \), the linear form \( F(f) \) is represented by the column vector \( \frac{1}{n} Fx \) over the dual basis \( (\chi_1^*, \ldots, \chi_n^*) \).

Our next goal is to extend the Fourier cotransform \( \mathcal{F} \) on \( L^2(\hat{G}) = [\hat{G} \to \mathbb{C}] \) to a bilinear form on \([G \to \mathbb{C}]^{**} \). Given any function \( \xi \in [\hat{G} \to \mathbb{C}] \), the Fourier cotransform \( \mathcal{F}(\xi) \) of \( \xi \) is the map \( \mathcal{F}(\xi) : \hat{G} \to \mathbb{C} \) given by

\[
\mathcal{F}(\xi)(\zeta) = \sum_{j=1}^n \xi(\chi_j) \zeta(\chi_j), \quad \zeta \in \hat{G},
\]

with \( \hat{G} = \{\chi_1, \ldots, \chi_n\} \).

Since every character in \( \hat{G} \) is a function in \([G \to \mathbb{C}] \), every function \( \xi \in [\hat{G} \to \mathbb{C}] \) can be viewed as a function in \([G \to \mathbb{C}] \to \mathbb{C} \). Similarly, every character \( \zeta \in \hat{G} \) is a function in \([G \to \mathbb{C}] \to \mathbb{C} \), and we know that \( \hat{G} \) is isomorphic to \( G \).

Recall that every function \( x \in [G \to \mathbb{C}] \) defines uniquely a linear form \( \bar{x} \in [G \to \mathbb{C}] \). Similarly, every function \( \xi \in [[G \to \mathbb{C}] \to \mathbb{C}] \) can be extended uniquely to the linear form \( \tilde{\xi} \in [G \to \mathbb{C}]^{**} \) defined on the basis \( (\overline{\chi}_1, \ldots, \overline{\chi}_n) \) of \([G \to \mathbb{C}]^* \) by

\[
\tilde{\xi}(\overline{\chi}_i) = \xi(\chi_i).
\]

The vector space \([\hat{G} \to \mathbb{C}] \) is isomorphic to \([G \to \mathbb{C}]^{**} \). Using the basis \( (\overline{\chi}_1^*, \ldots, \overline{\chi}_n^*) \) in \([G \to \mathbb{C}]^{**} \), any \( \omega \in [G \to \mathbb{C}]^{**} \) can be expressed as \( \omega = \sum_{i=1}^n y_i \overline{\chi}_i^* \).

Since \( \mathcal{F}(\xi)(\zeta) \) is defined as

\[
\mathcal{F}(\xi)(\zeta) = \sum_{j=1}^n \xi(\chi_j) \zeta(\chi_j),
\]

we have

\[
\mathcal{F}(\xi)(\zeta) = \sum_{j=1}^n \xi(\chi_j) \zeta(\chi_j) = \sum_{j=1}^n \tilde{\xi}(\overline{\chi}_j) \tilde{\zeta}(\overline{\chi}_j).
\]
Definition 9.8. Let $G$ be a finite abelian group, write $G = \{a_1, \ldots, a_n\}$, and let $\{\chi_1, \ldots, \chi_n\}$ be the characters of $G$. For all $\omega, \gamma \in [G \to \mathbb{C}]^\ast$, we define the Fourier cotransform $\mathcal{F}$ as the bilinear form on $[G \to \mathbb{C}]^{\ast\ast}$ given by

$$\mathcal{F}(\omega)(\gamma) = \sum_{j=1}^{n} \omega(\tilde{\chi}_j)\gamma(\tilde{\chi}_j).$$

If $\omega = \tilde{\xi}$ and $\gamma = \tilde{\zeta}$, then the right-hand side of the above equation is equal to $\mathcal{F}(\xi)(\eta)$, so this definition extends the Fourier cotransform on $[\hat{G} \to \mathbb{C}]$ and $\hat{\hat{G}}$.

If we write

$$\omega = \sum_{i=1}^{n} y_i\tilde{\chi}_i^\ast,$$

then we have

$$\mathcal{F}(\omega)(\gamma) = \sum_{i=1}^{n} y_i\mathcal{F}(\tilde{\chi}_i^\ast)(\gamma) = \sum_{i=1}^{n} \sum_{j=1}^{n} y_i\tilde{\chi}_i^\ast(\tilde{\chi}_j)\gamma(\tilde{\chi}_j) = \sum_{j=1}^{n} y_j\gamma(\tilde{\chi}_j).$$

We also have a natural isomorphism $\eta$ from $[G \to \mathbb{C}]$ to $[G \to \mathbb{C}]^{\ast\ast}$, defined such that the linear form $\eta_u \in [G \to \mathbb{C}]^{\ast\ast}$ is given by

$$\eta_u(\chi) = \chi(u) \quad \chi \in [G \to \mathbb{C}]^\ast, \ u \in [G \to \mathbb{C}].$$

Observe that

$$e_{a_i}^{\ast\ast} = \eta_{e_{a_i}},$$

because on the basis $(e_{a_1}^\ast, \ldots, e_{a_n}^\ast)$ of $[G \to \mathbb{C}]^\ast$, we have

$$e_{a_i}^{\ast\ast}(e_{a_j}^\ast) = \delta_{ij} = e_{a_j}^\ast(e_{a_i}) = \eta_{e_{a_i}}(e_{a_j}^\ast).$$

Then $(\eta_{e_{a_1}}, \ldots, \eta_{e_{a_n}}) = (e_{a_1}^{\ast\ast}, \ldots, e_{a_n}^{\ast\ast})$ is a basis of $[G \to \mathbb{C}]^{\ast\ast}$, and if $\omega = \sum_{i=1}^{n} y_i\tilde{\chi}_i^\ast$, then $\mathcal{F}(\omega)$ is given by

$$\mathcal{F}(\omega)\left(\sum_{i=1}^{n} z_i\eta_{e_{a_i}}\right) = \sum_{i=1}^{n} z_i\mathcal{F}(\omega)(\eta_{e_{a_i}})$$

$$= \sum_{i=1}^{n} z_i \sum_{j=1}^{n} y_j\eta_{e_{a_i}}(\tilde{\chi}_j)$$

$$= \sum_{i=1}^{n} z_i \sum_{j=1}^{n} y_j\tilde{\chi}_j(\tilde{\chi}_j)$$

$$= \sum_{i=1}^{n} \left(\sum_{j=1}^{n} y_j\chi_j(a_i)\right)z_i$$

$$= y^\top \mathcal{F}z = z^\top F^\ast y.$$

In summary, we proved the following result.
Proposition 9.23. If \( \omega \in [G \to \mathbb{C}]^{**} \) is expressed over the basis \((\widetilde{\chi}_1^*, \ldots, \widetilde{\chi}_n^*)\) as \( \omega = \sum_{i=1}^{n} y_i \widetilde{\chi}_i^* \), and if \( \zeta \in [G \to \mathbb{C}]^{**} \) is expressed over the basis \((e_{a_1}^{**}, \ldots, e_{a_n}^{**})\) as \( \zeta = \sum_{i=1}^{n} z_i e_{a_i}^{**} \), then the value \( F(\omega)(\zeta) \) of the bilinear form \( F \) is given by
\[
F(\omega)(\zeta) = y^T \mathbf{F} z = z^T \mathbf{F}^* y.
\]

As a linear map from \([G \to \mathbb{C}]^{**} \cong [G \to \mathbb{C}] \) to \( \mathbb{C} \), the matrix of \( F(\omega) \) over the basis \((e_{a_1}^{**}, \ldots, e_{a_n}^{**})\) is the row vector \( y^T \mathbf{F} \). As an element of \([G \to \mathbb{C}]^{***} \cong [G \to \mathbb{C}]^* \), the linear form \( F(\omega) \) is represented by the column vector \( \mathbf{F}^* y \) over the dual basis \((e_{a_1}^{***}, \ldots, e_{a_n}^{***})\).

The orthogonality conditions of Proposition 9.12 imply that
\[
\frac{1}{n} \mathbf{F} \mathbf{F}^* = \frac{1}{n} \mathbf{F}^* \mathbf{F} = I,
\]
so \( \mathbf{F}^* \) is the inverse of \( \frac{1}{n} \mathbf{F} \).

Now, if \( f \in [G \to \mathbb{C}]^* \) with \( f = \sum_{j=1}^{n} x_j e_{a_j}^* \), then \( F(f) \in [G \to \mathbb{C}]^{**} \), and since \( F(f) \) is expressed over the basis \((\widetilde{\chi}_1^*, \ldots, \widetilde{\chi}_n^*)\) and \( F \) is also defined over this basis, with \( \omega = F(f) \), we know that \( \omega \) is represented by the column vector \( y = Fx \) over the basis \((\widetilde{\chi}_1^*, \ldots, \widetilde{\chi}_n^*)\), and if \( \zeta = \sum_{i=1}^{n} z_i e_{a_i}^{**} = \sum_{i=1}^{n} z_i \eta_{e_{a_i}} \), then \( \mathcal{F}(\omega)(\zeta) \) is given by
\[
\mathcal{F}(\omega)(\zeta) = z^T \mathbf{F}^* y,
\]
so we get
\[
(\mathcal{F}(F(f)))(\zeta) = z^T \frac{1}{n} \mathbf{F}^* F x = z^T x = f \left( \sum_{i=1}^{n} z_i e_{a_i} \right).
\]
But
\[
\zeta = \sum_{i=1}^{n} z_i \eta_{e_{a_i}} = \eta \sum_{i=1}^{n} z_i e_{a_i},
\]
so the above equation shows that
\[
(F \circ F)(f) \circ \eta \sum_{i=1}^{n} z_i e_{a_i} = f \left( \sum_{i=1}^{n} z_i e_{a_i} \right),
\]
that is,
\[
f = (F \circ F)(f) \circ \eta.
\]

In summary, we proved the following theorem.

Theorem 9.24. Let \( G \) be a finite abelian group of order \( n \). The Fourier transform \( \mathcal{F} \) defined on \([G \to \mathbb{C}]^* \) in Definition 9.6 and the Fourier cotransform \( \mathcal{F} \) defined on \([G \to \mathbb{C}]^{**} \) in Definition 9.8 satisfy the Fourier inversion equation
\[
f = (F \circ F)(f) \circ \eta,
\]
for all \( f \in [G \to \mathbb{C}]^* \).
A nice feature of Definition 9.6 and Definition 9.8 is that they are intrinsic, that is, independent the choice of bases. A slight disadvantage is that the cotransform is defined on the double dual $[G \to \mathbb{C}]^{**}$ of $[G \to \mathbb{C}]$. But $[G \to \mathbb{C}]$ and $[G \to \mathbb{C}]^{**}$ are isomorphic (in fact, canonically), so it should be possible to define directly a Fourier cotransform on $[G \to \mathbb{C}]$. The slight complication is that we would like Fourier inversion to hold, and for this it appears that we need to use a noncanonical isomorphism, namely the isomorphism $\theta : [G \to \mathbb{C}] \to [G \to \mathbb{C}]^{**}$ defined as follows: if $c = \sum_{i=1}^{n} y_i e_{a_i}$, then

$$\theta_c = \sum_{i=1}^{n} y_i \tilde{\chi}_i^*.$$

**Definition 9.9.** For any vector $c = \sum_{i=1}^{n} y_i e_{a_i} \in [G \to \mathbb{C}]$, the Fourier cotransform $\mathcal{F}_2(c)$ of $c$ is the linear form on $[G \to \mathbb{C}]$ defined such that for all $a \in [G \to \mathbb{C}]$, we have

$$\mathcal{F}_2(c)(a) = \sum_{j=1}^{n} y_j \tilde{\chi}_j(a).$$

Observe that as long as we express the vector $c \in [G \to \mathbb{C}]$ over the basis $(e_{a_1}, \ldots, e_{a_n})$, the map $\mathcal{F}_2$ is bilinear. We may think of the components of $c$ as the Fourier coefficients of some form $f \in [G \to \mathbb{C}]^*$, and

$$\mathcal{F}_2(c) = \sum_{j=1}^{n} y_j \tilde{\chi}_j$$

is the “Fourier series” associated with $c$ (a linear form on $[G \to \mathbb{C}]$).

We can then repeat our familiar computation to prove that if $a = \sum_{j=1}^{n} z_j e_{a_j}$, then

$$\mathcal{F}_2(c)(a) = y^T \mathcal{F} z = z^T F^* y.$$

As a linear map from $[G \to \mathbb{C}]$ to $\mathbb{C}$, the matrix of $\mathcal{F}_2(c)$ over the basis $(e_{a_1}, \ldots, e_{a_n})$ is the row vector $y^T \mathcal{F}$, and as an element of $[G \to \mathbb{C}]^*$, the linear form $\mathcal{F}_2(c)$ is represented by the column vector $F^* y$.

In order to be able to compose $\mathcal{F}$ and $\mathcal{F}_2$, we need to convert $\mathcal{F}(f) \in [G \to \mathbb{C}]^{**}$, the result of applying $\mathcal{F}$ to $f \in [G \to \mathbb{C}]^*$, to a vector in $[G \to \mathbb{C}]$, and we can do this by applying the isomorphism $\theta^{-1} : [G \to \mathbb{C}]^{**} \to [G \to \mathbb{C}]$. Then Fourier inversion becomes the identity

$$f = (\mathcal{F}_2 \circ \theta^{-1} \circ \mathcal{F})(f),$$

for all $f \in [G \to \mathbb{C}]^*$, which is another way of stating Proposition 9.18.

Since the vector spaces $[G \to \mathbb{C}]$, $[G \to \mathbb{C}]^*$, $[G \to \mathbb{C}]^{**}$, and $[G \to \mathbb{C}]^{***}$, are all isomorphic (and isomorphic to $\mathbb{C}^n$, where $n = |G|$), if we are just interested in transformations on sequences of complex numbers of length $n$ indexed by the elements of the group $G$, namely elements of $[G \to \mathbb{C}]$, we can formulate versions of the Fourier transform and of the Fourier cotransform in terms of the Fourier matrix defined in Definition 9.7.
Definition 9.10. Let $G$ be a finite abelian group, $G = \{a_1, \ldots, a_n\}$, and let $\{\chi_1, \ldots, \chi_n\}$ be the characters of $G$. If $F = (\chi_i(a_j))$ is the Fourier matrix of $G$ (as in Definition 9.7), then for every sequence $x \in [G \to \mathbb{C}]$, the sequence 

$$\hat{x} = F(x) = \frac{1}{n}Fx$$

is called the Fourier transform of $x$, and given any sequence $\xi \in [G \to \mathbb{C}]$, the sequence

$$\overline{F}(x) = F^*\xi$$

is called the inverse Fourier transform or Fourier cotransform of $\xi$.

Recall that

$$\frac{1}{n}FF^* = \frac{1}{n}F^*F = I,$$

so the two transforms are mutual inverses.

Recall that we proved earlier that

$$\mathcal{F}(x * y) = \mathcal{F}(x)\mathcal{F}(y),$$

where $x * y$ is the convolution of $x$ and $y$, given by

$$(x * y)_a = \frac{1}{|G|} \sum_{b \in G} x_b y_{b^{-1}a} = \frac{1}{|G|} \sum_{b,c \in G} x_b y_c.$$

In matrix terms, $\mathcal{F}(x)\mathcal{F}(y) = \hat{x}\hat{y}$ is the vector whose $a$th entry $(\hat{x}\hat{y})_a$ is the product of the $a$th entry $\hat{x}_a$ of the vector $\hat{x}$ by the $a$th entry $\hat{y}_a$ of the vector $\hat{y}$. In matrix terms, it can be expressed as

$$\text{diag}(\hat{x})\hat{y},$$

where $\text{diag}(\hat{x})$ is the diagonal matrix whose diagonal entries are the entries in the vector $\hat{x}$.

Other aspects of harmonic analysis on finite abelian groups can be found in Terras [79]. In the next section, we consider the special case where $G = \mathbb{Z}/m\mathbb{Z}$.

9.6 The Discrete Fourier Transform (on $\mathbb{Z}/n\mathbb{Z}$)

If $G = \mathbb{Z}/n\mathbb{Z}$, then we know from Proposition 9.8(3) that the characters of $\mathbb{Z}/n\mathbb{Z}$ are the $n$ homomorphisms $\chi_k$ given by

$$m \mapsto e^{2\pi imk/n}, \quad k = 0, 1, \ldots, n - 1, \quad m \in \mathbb{Z}/n\mathbb{Z}.$$

Observe that the characters are index by $0, 1, \ldots, n - 1$ rather than $1, 2, \ldots, n$, but this is actually more convenient in what follows. The complex numbers

$$\{1, e^{2\pi i/n}, e^{2\pi i2/n}, \ldots, e^{2\pi in/n}, \ldots, e^{2\pi i(n-1)/n}\}$$
(with \(0 \leq m \leq n-1\)) are the \(n\)th roots of unity (because obviously, \((e^{2\pi ik/n})^n = e^{2\pi i k} = 1\)). They form a subgroup of \(\mathbb{C}\) denoted \(\mu_n(\mathbb{C})\) isomorphic to \(\mathbb{Z}/n\mathbb{Z}\) under the isomorphism \(k \mapsto e^{2\pi ik/n}\), for \(k \in \mathbb{Z}/n\mathbb{Z}\).

The Fourier matrix \(F_n = (\chi_k(m))_{0 \leq k \leq n-1}^{0 \leq m \leq n-1}\) is given by

\[
F_n = (e^{-2\pi im/n})_{0 \leq k \leq n-1}^{0 \leq m \leq n-1}.
\]

The first row \((k = 0)\) consists of 1s, the second row \((k = 1)\) consists of the consecutive inverse powers \(\zeta^{-m}\) of \(\zeta = e^{2\pi i/n}\) (a primitive \(n\)th root of unity) for \(m = 0, \ldots, n-1\),

\[
\begin{pmatrix}
1 & e^{-2\pi i/n} & e^{-2\pi i2/n} & \ldots & e^{-2\pi im/n} & e^{-2\pi i(n-1)/n},
\end{pmatrix}
\]

and the \((k+1)\)th row \((0 \leq k \leq n-1)\) consists of the \(k\)th powers of the entries in the second row,

\[
\begin{pmatrix}
1 & (e^{-2\pi i/n})^k & (e^{-2\pi i2/n})^k & \ldots & (e^{-2\pi im/n})^k & (e^{-2\pi i(n-1)/n})^k,
\end{pmatrix}
\]

Observe that \(F_n\) is symmetric. It is also a Vandermonde matrix for the roots of unity

\[
\{1, e^{-2\pi i/n}, e^{-2\pi i2/n}, \ldots, e^{-2\pi im/n}, \ldots, e^{-2\pi i(n-1)/n}\},
\]

namely, with \(\omega = \zeta^{-1} = e^{-2\pi i/n}\) (also a primitive \(n\)th root of unity), we have

\[
F_n = \begin{pmatrix}
1 & 1 & \ldots & 1 & 1 \\
1 & \omega & \ldots & \omega^{n-2} & \omega^{n-1} \\
1 & \omega^2 & \ldots & \omega^{2(n-2)} & \omega^{2(n-1)} \\
1 & \omega^3 & \ldots & \omega^{3(n-2)} & \omega^{3(n-1)} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
1 & \omega^{n-2} & \ldots & \omega^{(n-2)^2} & \omega^{(n-2)(n-1)} \\
1 & \omega^{n-1} & \ldots & \omega^{(n-1)(n-2)} & \omega^{(n-1)^2}
\end{pmatrix}.
\]

For example, if \(n = 5\), we have

\[
F_5 = \begin{pmatrix}
1 & 1 & 1 & 1 & 1 \\
1 & e^{-2\pi i/5} & e^{-2\pi i2/5} & e^{-2\pi i3/5} & e^{-2\pi i4/5} \\
1 & e^{-2\pi i2/5} & e^{-2\pi i4/5} & e^{-2\pi i6/5} & e^{-2\pi i8/5} \\
1 & e^{-2\pi i3/5} & e^{-2\pi i6/5} & e^{-2\pi i9/5} & e^{-2\pi i12/5} \\
1 & e^{-2\pi i4/5} & e^{-2\pi i8/5} & e^{-2\pi i12/5} & e^{-2\pi i16/5}
\end{pmatrix}
\]

\[
= \begin{pmatrix}
1 & 1 & 1 & 1 & 1 \\
1 & e^{-2\pi i/5} & e^{-2\pi i2/5} & e^{-2\pi i3/5} & e^{-2\pi i4/5} \\
1 & e^{-2\pi i2/5} & e^{-2\pi i4/5} & e^{-2\pi i5/5} & e^{-2\pi i3/5} \\
1 & e^{-2\pi i3/5} & e^{-2\pi i5/5} & e^{-2\pi i4/5} & e^{-2\pi i2/5} \\
1 & e^{-2\pi i4/5} & e^{-2\pi i3/5} & e^{-2\pi i2/5} & e^{-2\pi i5/5}
\end{pmatrix}.
\]
Given a sequence \( x = (x_0, \ldots, x_{n-1}) \in \mathbb{C}^n \), its Fourier transform, also called discrete Fourier transform, is
\[
\hat{x} = \frac{1}{n} F_n x.
\]
We can think of \( c = \hat{x} \) as the sequence of Fourier coefficients of \( x \). Similarly, given a sequence \( c = (c_0, \ldots, c_{n-1}) \in \mathbb{C}^n \), its discrete inverse Fourier transform (or discrete Fourier cotransform) is
\[
\overline{F}(c) = F^* c = \overline{F}_n c.
\]
Every sequence \( c = (c_0, \ldots, c_{n-1}) \in \mathbb{C}^n \) of “Fourier coefficients” determines a periodic function \( f_c : \mathbb{R} \to \mathbb{C} \) (of period \( 2\pi \)) known as discrete Fourier series, or phase polynomial, defined such that
\[
f_c(\theta) = c_0 + c_1 e^{i\theta} + \cdots + c_{n-1} e^{i(n-1)\theta} = \sum_{k=0}^{n-1} c_k e^{ik\theta}.
\]
Then, given any sequence \( f = (f_0, \ldots, f_{n-1}) \) of data points, it is desirable to find the “Fourier coefficients” \( c = (c_0, \ldots, c_{n-1}) \) of the discrete Fourier series \( f_c \) such that
\[
f_c(2\pi k/n) = f_k,
\]
for every \( k, 0 \leq k \leq n - 1 \).

The problem amounts to solving the linear system of \( n \) equations
\[
c_0 + c_1 e^{2\pi ik/n} + \cdots + c_{n-1} e^{2\pi i(n-1)k/n} = f_k, \quad k = 0, \ldots, n - 1,
\]
which is just the system
\[
\overline{F}_n c = f.
\]
Since
\[
\frac{1}{n} F_n \overline{F}_n = \frac{1}{n} \overline{F}_n F_n = I,
\]
we see that \( c \) is given by
\[
c = \frac{1}{n} F_n f = \hat{f},
\]
the discrete Fourier transform of \( f \), so
\[
c_k = \frac{1}{n} \sum_{j=0}^{n-1} f_j e^{-2\pi jk/n}, \quad k = 0, \ldots, n - 1.
\]

Note the analogy with the case of \( \mathbb{T} \) and \( \mathbb{Z} \), where the Fourier cotransform \( \overline{F}(c) \) of the sequence \( (c_m)_{m \in \mathbb{Z}} \) is given by
\[
f(\theta) = \sum_{k=-\infty}^{\infty} c_k e^{ik\theta},
\]
and the Fourier coefficients of the function $f : \mathbb{T} \to \mathbb{C}$ are given by the formulae

$$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x)e^{-ikx} dx.$$ 

In $\mathbb{Z}/n\mathbb{Z}$, the convolution of two sequences $f = (f_0, \ldots, f_{n-1})$ and $g = (g_0, \ldots, g_{n-1})$ is given by

$$(f * g)_k = \frac{1}{n} \sum_{i,j \in \mathbb{Z}/n\mathbb{Z}} f_i g_j, \quad k = 0, \ldots, n - 1.$$ 

It is remarkable that the convolution $f * g$ can be expressed in matrix form as

$$f * g = \frac{1}{n} H(f)g$$

for some matrix $H(f)$. The matrix $H(f)$ is a circulant matrix.

We define the circular shift matrix $S_n$ (of order $n$) as the matrix

$$S_n = \begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}$$

consisting of cyclic permutations of its first column.

For any sequence $f = (f_0, \ldots, f_{n-1}) \in \mathbb{C}^n$, we define the circulant matrix $H(f)$ as

$$H(f) = \sum_{j=0}^{n-1} f_j S_n^j,$$

where $S_n^0 = I_n$, as usual. For example, the circulant matrix associated with the sequence $f = (a, b, c, d)$ is

$$\begin{pmatrix} a & d & c & b \\ b & a & d & c \\ c & b & a & d \\ d & c & b & a \end{pmatrix}$$

It is not hard to prove that the convolution $f * g$ of two sequences $f = (f_0, \ldots, f_{n-1})$ and $g = (g_0, \ldots, g_{n-1})$ is given by

$$f * g = \frac{1}{n} H(f)g,$$
viewing $f$ and $g$ as column vectors.

Then, the miracle (which is not too hard to prove!) is that we have

$$H(f)F_n = F_n \text{diag}(n\hat{f}),$$

(\dag)

where $\text{diag}(n\hat{f})$ is the diagonal matrix whose diagonal entries are the elements of the vector $n\hat{f}$, which means that the columns of the Fourier matrix $F_n$ are the eigenvectors of the circulant matrix $H(f)$, and that the eigenvalue associated with the $k$th eigenvector is $(n\hat{f})_k$, that is, $n$ times the $k$th component of the Fourier transform $\hat{f}$ of $f$ (counting from $0$).

To prove (\dag), we first prove that the eigenvectors $u_k$ of the circular shift matrix $S_n$ (indexing from $0$ to $n-1$) are the columns of $F_n$, where the column of rank $k$ whose entries are

$$1 \ (e^{2\pi i/n})^k \ (e^{2\pi i 2/n})^k \ \ldots \ (e^{2\pi i m/n})^k \ (e^{2\pi i (n-1)/n})^k$$

is associated with the eigenvalue $e^{-2\pi ik/n}$.

Indeed, applying $S_n$ to $u_k$, the last entry $e^{2\pi i (n-1)/n} = e^{-2\pi ik/n}$ in $u_k$ becomes the first entry in $S_n u_k$, so the corresponding eigenvalue is $e^{-2\pi ik/n}$. For example, if $n = 4$, since $(1, e^{2\pi i 4}, e^{2\pi i 2/4}, e^{2\pi i 3/4}) = (1, i, -1, -i)$ and $(1, e^{-2\pi i 4}, e^{-2\pi i 2/4}, e^{-2\pi i 3/4}) = (1, -i, 1, i)$, we have

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

Since $H(f) = \sum_{j=0}^{n-1} f_j S_n^j$, the eigenvectors remain the same, and it is easy to see that the $k$th eigenvalue of $H(f)$ (indexing from $0$) is $\sum_{j=0}^{n-1} f_j e^{-2\pi ijk/n} = n\hat{f}_k$.

If we recall that $F_n F_n = F_n F_n = n I_n$, multiplying the equation $H(f)F_n = F_n \text{diag}(\hat{f})$ both on the left and on the right by $F_n$, we get

$$F_n H(f)(n I_n) = (n I_n) \text{diag}(\hat{f}) F_n,$$

that is,

$$F_n H(f) = \text{diag}(n\hat{f}) F_n.$$

If we apply both sides to any sequence $g \in \mathbb{C}^n$, we get

$$F_n H(f)g = \text{diag}(n\hat{f}) F_n g.$$

Since $\hat{f} = \frac{1}{n} F_n f$, $\hat{g} = \frac{1}{n} F_n g$, $f \ast g = \frac{1}{n} H(f)g$, and $f \ast g = \frac{1}{n} F_n(f \ast g)$, multiplying both sides by $1/n^2$, the above equation yields

$$\frac{1}{n} F_n \frac{1}{n} H(f)g = \text{diag}(\hat{f}) \frac{1}{n} F_n g,$$

which means that

$$\hat{f} \ast \hat{g} = \text{diag}(\hat{f}) \hat{g} = \hat{f} \hat{g}.$$

Therefore, we have given another proof of the convolution rule.
9.7 Some Properties of the Fourier Transform

Before moving on to the Plancherel theorem, let us state some useful properties of the Fourier transform.

**Proposition 9.25.** Let $G$ be locally compact abelian group. The Fourier transforms and the Fourier cotransforms satisfy the following equations: For all $f \in L^1(G)$, and all $\chi \in \widehat{G}$,

1. $F(f)(\chi) = F(f)(\chi^{-1}) = F(\hat{f})(\chi) = \overline{F(f)(\chi)}$.

   These equations also hold with $f$ replaced by a complex measure $\mu \in M^1(G)$. We also have $F(\delta_a)(\chi) = \chi(a)$, $F(\delta_a)(\chi) = \chi(a)$.

2. $\|F(f)\|_{\infty} = \|F(f)\|_{\infty} \leq \|f\|_1$.

   These equations also hold with $f$ replaced by a complex measure $\mu \in M^1(G)$.

3. $F(\lambda_a(f))(\chi) = \overline{\chi(a)} F(f)(\chi)$

   $\overline{\lambda_a(f)}(\chi) = \chi(a) F(f)(\chi)$.

   These equations also hold with $f$ replaced by a complex measure $\mu \in M^1(G)$.

4. For all $f \in L^1(G)$, and all $\chi, \xi \in \widehat{G}$,

   $F(\xi f)(\chi) = F(f)(\xi^{-1} \chi) = \lambda_{\xi}(F(f))(\chi)$

   $\overline{F(\xi f)}(\chi) = \overline{F(f)(\xi^{-1} \chi)} = \lambda_{\xi}(\overline{F(f)})(\chi)$.

   These equations also hold with $f$ replaced by a complex measure $\mu \in M^1(G)$.

**Proof.** (1) We already proved Equations (1).

(3) We have

\[
F(\lambda_a(f))(\chi) = \int \overline{\chi(b)} f(a^{-1}b) d\lambda(b)
= \int \overline{\chi(ab)} f(b) d\lambda(b)
= \int \overline{\chi(a)} \chi(b) f(b) d\lambda(b)
= \chi(a) \int \overline{\chi(b)} f(b) d\lambda(b) = \chi(a) F(f)(\chi).
\]
The second equation is proven in a similar way.

(4) We have

\[ \mathcal{F}(\xi f)(\chi) = \int \overline{\chi(a)} \xi(a) f(a) \, d\lambda(a) \]

\[ = \int (\xi^{-1} \chi)(a) f(a) \, d\lambda(a) \]

\[ = \mathcal{F}(f)(\xi^{-1} \chi) = \lambda_\xi(\mathcal{F}(f))(\chi). \]

We leave the proof of the other equations as exercises. \( \square \)

**Example 9.3.** If \( G = \mathbb{R} \), since the characters are of the form \( x \mapsto e^{ixy} \) (with \( y \in \mathbb{R} \)), then, with a slight abuse of notation, Equations (3) yields the well-known formula

\[ \mathcal{F}(\lambda_a(f))(x) = \mathcal{F}(f(x - a)) = e^{-iax} \mathcal{F}(f)(x), \]

and Equations (4) yields

\[ \mathcal{F}(e^{iax} f(x)) = \mathcal{F}(f)(x - a); \]

See Rudin [66] (Chapter 9) or Folland [32] (Chapter 8); recall that \( \lambda_a(f)(x) = f(x - a) \).

If \( G = \mathbb{T} \), since the characters are of the form \( e^{i\theta} \mapsto e^{im\theta} \) (with \( m \in \mathbb{Z} \)), then, with a slight abuse of notation, Equations (3) yields the formula

\[ \mathcal{F}(\lambda_{e^{i\theta}}(f))(m) = e^{-im\varphi} \mathcal{F}(f)(m), \]

and Equations (4) yields

\[ \mathcal{F}(e^{im\varphi} f(e^{i\theta}))(m) = \mathcal{F}(f)(m - n). \]

Recall that \( f : \mathbb{T} \to \mathbb{C} \), and that \( \mathcal{F}(f) \) is a \( \mathbb{Z} \)-indexed sequence of complex numbers, namely the Fourier coefficients of \( f \).

If \( G = \mathbb{Z} \), since the characters are of the form \( m \mapsto e^{im\theta} \) with \( \theta \in \mathbb{R}/2\pi\mathbb{Z} \), Equations (3) yields the formula

\[ \mathcal{F}(\lambda_n(f))(\theta) = e^{-in\varphi} \mathcal{F}(f)(\theta), \]

and Equations (4) yields

\[ \mathcal{F}(e^{im\varphi} f(m))(\theta) = \mathcal{F}(f)(\theta - \varphi). \]

Recall that \( f : \mathbb{Z} \to \mathbb{C} \) is a \( \mathbb{Z} \)-indexed sequence, and that \( \mathcal{F}(f)(\theta) = \sum_{m \in \mathbb{Z}} f(m)e^{-im\theta} \) is a Fourier series.
9.8 Plancherel’s Theorem and Fourier Inversion

Let $G$ be a locally compact abelian group equipped with a Haar measure $\lambda$. In general, given a function $f \in L^1(G)$, its Fourier transform $\mathcal{F}(f)$ does not belong to $L^1(\hat{G})$.

Plancherel’s theorem (Theorem 9.26) asserts that there is a Haar measure $\hat{\lambda}$ on the dual group $\hat{G}$ such that the map $f \mapsto \mathcal{F}(f)$ sends $L^1(G) \cap L^2(G)$ into $L^2(\hat{G})$, and has a unique extension which is an isometry from $L^2(G)$ to $L^2(\hat{G})$.

We will follow Bourbaki’s proof [8] (Chapter 2, Section 1, No. 3). The crucial step is to define a subspace $A(G)$ of $L^1(G) \cap L^2(G)$ which is dense in both $L^1(G)$ and $L^2(G)$, and to show that there is a Haar measure $\nu$ on $\hat{G}$ such that

$$\int_{\hat{G}} |\mathcal{F}(f)|^2 \, d\nu = \int_G |f|^2 \, d\lambda$$

for all $f \in A(G)$. There are many technical details so we will focus on the main ideas.

**Definition 9.11.** Let $A(G)$ be the subspace of $L^1(G)$ spanned by the set of functions of the form $f \ast g$, with $f, g \in L^1(G) \cap L^2(G)$.

It is immediately verified that $A(G)$ is an ideal of $L^1(G)$ contained in $L^1(G) \cap L^2(G)$.

The first step is to show that there is a filter base $\mathcal{B}$ defined on $A(G) \cap \mathcal{K}_c(G)$, where the functions in this filter base approximate the Dirac measure, so that

1. $\delta_e = \lim_{\mathcal{B}} \varphi d\lambda$, for all subsets in the filter $\mathcal{B}$ and for all $\varphi$ in this subset.
2. $\lim_{\mathcal{B}} \mathcal{F}(\varphi) = 1$, and $\|\mathcal{F}(\varphi)\|_{\infty} \leq 1$, for all subsets in the filter $\mathcal{B}$ and for all $\varphi$ in this subset.
3. $\lim_{\mathcal{B}} \varphi \ast f = f$, for all $f \in L^p(G)$, $p = 1, 2$.

Condition (3) implies that $A(G)$ is dense in both $L^1(G)$ and $L^2(G)$, and it can also be shown that $\mathcal{F}(G(A))$ is dense in $\mathcal{C}_0(\hat{G})$.

The second step is show that there is a Haar measure $\nu$ on $\hat{G}$ such that

$$\int_{\hat{G}} |\mathcal{F}(f)|^2 \, d\nu = \int_G |f|^2 \, d\lambda \quad (*)$$

for all $f \in A(G)$. This goes as follows. It can be shown that for every $f \in A(G)$, there is a unique positive measure $\mu_f$ on $\hat{G}$ such that

$$(g \ast f)(e) = \int_{\hat{G}} \mathcal{F}(g) \, d\mu_f, \quad \text{for all } g \in A(G).$$
Then for every $f \in A(G)$, define $\Omega_f$ as the open subset of $\hat{G}$ given by

$$\Omega_f = \{ \chi \in \hat{G} \mid \mathcal{F}(f)(\chi) \neq 0 \}.$$ 

It can be shown that the subsets $\Omega_f$ form an open cover of $\hat{G}$ when $f$ ranges over $A$. For every $\Omega_f$, let $\nu_f$ be the positive measure on $\Omega_f$ associated with the Radon functional given by $\Phi_{\nu_f} = 1/|\mathcal{F}(f)| \Phi_{\mu_f}$, where $\Phi_{\mu_f}$ is the Radon functional associated with the measure $\mu_f$.

It can be shown that the local measures $\nu_f$ patch to a global Haar measure $\nu$ on $\hat{G}$, and that (*) is satisfied.

**Definition 9.12.** Let $G$ be a locally compact abelian group. For every Haar measure $\lambda$ on $G$, the Haar measure $\hat{\lambda} = \nu$ given by the previous construction is called the measure associated with $\lambda$, or dual measure.

The following theorem is obtained.

**Theorem 9.26.** (Plancherel) Let $G$ be a locally compact abelian group equipped with a Haar measure $\lambda$. There is a Haar measure $\hat{\lambda}$ on the dual $\hat{G}$ such that for any $f \in L^1(G) \cap L^2(G)$, we have $\mathcal{F}(f) \in L^2(\hat{G})$. Furthermore, the map $f \mapsto \mathcal{F}(f)$ from $L^1(G) \cap L^2(G)$ to $L^2(\hat{G})$ has a unique extension which is an isometry from $L^2(G)$ to $L^2(\hat{G})$.

**Proof sketch.** By Property (3) of the filter base, $A(G)$ is dense in $L^2(G)$. By (*), the Fourier transform is an isometry from $A(G) \subseteq L^2(G)$ to a subspace of $L^2(\hat{G})$. Therefore it has a unique extension $\Phi$ to $L^2(G)$. To finish the proof, it suffices to show that $\Phi(A(G)) = L^2(\hat{G})$ and that $\Phi(f) = \mathcal{F}(f)$ for all $f \in L^1(G) \cap L^2(G)$.

To prove that $\Phi(A(G)) = L^2(\hat{G})$, since $\mathcal{F}(A(G))$ is closed in $L^2(\hat{G})$, it suffices to show that $\mathcal{F}(A(G))$ is dense in $L^2(\hat{G})$.

Assume that some $h \in L^2(\hat{G})$ is orthogonal to $\mathcal{F}(A(G))$. For all $f, g \in A(G)$, we have $\mathcal{F}(f)\mathcal{F}(g) = \mathcal{F}(f * g) \in \mathcal{F}(A(G))$. Then

$$\langle h \mathcal{F}(f), \mathcal{F}(g) \rangle = \int h(\chi) \mathcal{F}(f)(\chi) \overline{\mathcal{F}(g)(\chi)} \, d\hat{\lambda}(\chi)$$

$$= \int h(\chi) \overline{\mathcal{F}(f)(\chi)} \mathcal{F}(g)(\chi) \, d\hat{\lambda}(\chi)$$

$$= \int h(\chi) \overline{\mathcal{F}(f^*)(\chi)} \mathcal{F}(g)(\chi) \, d\hat{\lambda}(\chi)$$

$$= \int h(\chi) \overline{\mathcal{F}(f^* * g)(\chi)} \, d\hat{\lambda}(\chi)$$

$$= \langle h, \mathcal{F}(f^* * g) \rangle = 0,$$

because $f^* * g \in A(G)$ and $h$ is orthogonal to $\mathcal{F}(A(G))$. The above shows that $h \mathcal{F}(f)$ is orthogonal to $\mathcal{F}(A(G))$. Then it can be shown that this implies that $h = 0$, which
shows that \( \Phi(A(G)) \) is dense in \( L^2(\hat{G}) \). The last step is to show that \( \Phi(f) = F(f) \) for all \( f \in L^1(G) \cap L^2(G) \). The details are a bit involved, so we refer the reader to Bourbaki [8] (Chapter 2, Section 1, No. 3, Theorem 1). Another proof of Plancherel’s theorem is given in Folland [31] (Chapter 4, Section 2, Theorem 4.25). It follows similar lines but uses a class of functions \( B^1 \) different from \( A(G) \).

The unique extension of \( F \) is also denoted \( F \). It is easy to see that \( F \) also has a unique extension to \( L^2(G) \), and that it is an isometry.

On should realize that Theorem 9.26 does not say that the Fourier transform \( F \) (or the Fourier cotransform \( \overline{F} \)) is defined on \( L^2(G) \), because in general the integral will not converge for \( f \) outside of \( L^1(G) \cap L^2(G) \). What is happening is more subtle. It is always possible by using a limit process to define the Fourier transform of any \( f \in L^2(G) \), and this extension of \( F \) to \( L^2(G) \) is an isometry. This is still quite remarkable because there is no such result for \( L^1(G) \). Since the extension of \( F \) to \( L^2(G) \) is an isometry, it has an inverse, but it is far from obvious that this inverse has any relation to the Fourier cotransform on \( L^2(\hat{G}) \). In fact it does, but this requires proving Gelfand’s duality theorem, that \( G \) and its double dual \( \hat{\hat{G}} \) are isomorphic.

As a corollary of Theorem 9.26 and Proposition 9.12, we can prove the fact announced just after Proposition 9.12.

**Proposition 9.27.** If \( G \) is a compact abelian group endowed with a Haar measure \( \lambda \) normalized so that \( G \) has measure \( \lambda(G) = 1 \), then \( \hat{G} \) is a Hilbert basis for \( L^2(G) \) (it is orthonormal and dense in \( L^2(G) \)).

**Proof.** By a well known fact of Hilbert space theory, it suffices to show that there is no nonzero function \( f \in L^2(G) \) orthogonal to every character \( \chi \in \hat{G} \). Assume \( f \in L^2(G) \) is orthogonal to every character \( \chi \in \hat{G} \). This means that \( \int f(a) \overline{\chi(a)} \, d\lambda(a) = 0 \), and since

\[
\int f(a) \overline{\chi(a)} \, d\lambda(a) = F(f)(\chi),
\]

we get

\[
F(f)(\chi) = 0, \quad \text{for all } \chi \in \hat{G}.
\]

Since Plancherel’s theorem asserts that \( F \) is an isometry, it is injective, so \( f = 0 \), establishing our result. \( \square \)

We now turn to the issue of Fourier inversion. If \( f, g \in L^2(G) \), the convolution \( f \ast g \) is given by

\[
(f \ast g)(s) = \int f(t)g(t^{-1}s) \, d\lambda(t)
\]

so for \( s = e \),

\[
(f \ast g)(e) = \int f(t)g(t^{-1}) \, d\lambda(s) = \int f(t)^\ast \overline{g(t)} \, d\lambda(s) = \int f(t)g^\ast(t) \, d\lambda(s).
\]
We also have
\[ \langle f, g \rangle = \int f(t)g(t) \, d\lambda(t), \]
so we deduce
\[ (f * g^*)(e) = \langle f, g \rangle. \]
But Plancherel’s theorem implies that
\[ \langle f, g \rangle = \langle F(f), F(g) \rangle = \int F(f)(\chi)\overline{F(g)(\chi)} \, d\hat{\lambda}(\chi), \]
and we conclude that
\[ (f * g^*)(e) = \int F(f)(\chi)\overline{F(g)(\chi)} \, d\hat{\lambda}(\chi). \]

**Proposition 9.28.** (Fourier inversion for \( A(G) \)) If \( f \in A(G) \), then \( F(f) \in L^1(\hat{G}) \), and
\[ f(a) = \int_{\hat{G}} \chi(a)F(f)(\chi) \, d\hat{\lambda}(\chi) \quad \text{for all } a \in G. \]
Equivalenty,
\[ f = (F \circ F)(f) \circ \eta, \quad \text{for all } f \in A(G), \]
where \( \eta: G \to \hat{G} \) is the canonical injection.

Proposition 9.28 is proven by using the formula
\[ (f * g^*)(e) = \int F(f)(\chi)\overline{F(g)(\chi)} \, d\hat{\lambda}(\chi), \]
and using some functions \( g \) that approximate the Dirac measure \( \delta_e \); see Bourbaki [8] (Chapter 2, Section 1, No. 4). Unlike the situation in Plancherel’s theorem, there is no need to extend \( F \) (and \( F \) on \( L^1(\hat{G}) \)).

It is also shown in Bourbaki that the inversion formula holds for all \( f \in L^2(G) \) such that \( F(f) \in L^1(\hat{G}) \).

In order to proceed any further, we need Pontrjagin’s duality theorem asserting that \( \eta \) is an isomorphism.

### 9.9 Pontrjagin Duality and Fourier Inversion

The Pontrjagin duality theorem is one of the most important and most beautiful theorems of the theory of locally compact abelian groups. Recall that we have a canonical map \( \eta: G \to \hat{G} \) given by
\[ \eta(a)(\chi) = \chi(a), \quad a \in G, \chi \in \hat{G}, \]
which is a homomorphism.
Theorem 9.29. (Pontrjagin duality theorem) Let $G$ be a locally compact abelian group endowed with a Haar measure $\lambda$, let $\hat{G}$ be its dual group endowed with the associated Haar measure $\hat{\lambda}$ (see Definition 9.12), and let $\hat{\hat{G}}$ be its double dual endowed with the associated measure $\hat{\hat{\lambda}}$. The map $\eta: G \to \hat{\hat{G}}$ is an isomorphism and a homeomorphism between the topological groups $G$ and $\hat{\hat{G}}$. If we identify $G$ and $\hat{\hat{G}}$ using the isomorphism $\eta$, then the extension $\mathcal{F}: L^2(G) \to L^2(\hat{\hat{G}})$ of the Fourier transform to $L^2(G)$ and the extension $\hat{\mathcal{F}}: L^2(\hat{\hat{G}}) \to L^2(G)$ of the Fourier cotransform to $L^2(G)$ are mutual inverses. In particular, Fourier inversion holds; that is,

$$f = (\hat{\mathcal{F}} \circ \mathcal{F})(f) \circ \eta, \quad \text{for all } f \in L^2(G).$$

Proof idea. The proof of Theorem 9.29 is too technical to be presented in full detail here. A proof can found in Bourbaki [8] (Chapter 2, Section 1, No. 5, Theorem 2), and Folland [31] (Chapter 4, Section 4.3, Theorem 4.31).

The first part of the proof of Pontrjagin duality establishes the fact that $\eta$ is injective and a homeomorphism onto its image, which is closed in $\hat{\hat{G}}$. To prove that $\eta$ is injective and that $\eta^{-1}$ is continuous, it suffices to prove that for every neighborhood $U$ of $e$ in $G$ there is some neighborhood $W$ of $e$ in $\hat{\hat{G}}$ such that $\eta^{-1}(W) \subseteq U$.

We can find a compact symmetric neighborhood $V$ of $e$ in $G$ such that $V^2 \subseteq U$, and some positive function $f \in \mathcal{K}_c(G)$ whose support is contained in $V$. If we let $g = f^* * f$, then we see that $g \in A(G)$, $\text{supp}(g) \subseteq U$, and $g(e) > 0$. Since $\hat{\hat{G}}$ has the compact open topology, which is equivalent to the topology of pointwise convergence on $L^1(\hat{G})$, there is a neighborhood $W$ of the identity $\hat{e}$ in $\hat{\hat{G}}$ such that, $\zeta \in W$ implies that

$$\left| \int \zeta(\chi) \mathcal{F}(g)(\chi) \, d\hat{\lambda}(\chi) - \int \hat{\mathcal{F}}(g)(\chi) \, d\hat{\hat{\lambda}}(\chi) \right| < \frac{1}{2} g(e).$$

By Proposition 9.28, the above is equivalent to

$$|\mathcal{F}(g)(\zeta) - \mathcal{F}(g)(\hat{e})| < \frac{1}{2} g(e), \quad \zeta \in W,$n

so if $a \in \eta^{-1}(W)$, since $g \in A(G)$, by Proposition 9.28 we have $g = (\mathcal{F} \circ \mathcal{F})(g) \circ \eta$, and we obtain

$$|g(a) - g(e)| < \frac{1}{2} g(e).$$

Therefore, $g(a) \neq 0$, and since $\text{supp}(g) \subseteq U$, we have $a \in U$, which shows that $\eta^{-1}(W) \subseteq U$, as desired.

The second part is to prove that $\eta$ is surjective. Bourbaki’s proof uses the following fact which shows the existence of certain kinds of bump functions on $L^1(G)$.

Proposition 9.30. Given any closed subset $P$ of $\hat{\hat{G}}$ and any $\chi \in \hat{\hat{G}}$, if $\chi \notin P$, then there exists some $f \in L^1(G)$ such that $\mathcal{F}(f)(\chi) = 1$ and $\mathcal{F}(f)$ vanishes on $P$. 
Assume that there is some \( \zeta \in \hat{\hat{G}} \) such that \( \zeta \not\in \eta(G) \). By Proposition 9.30, there is some nonzero function \( f \in L^1(\hat{G}) \) such that \( \mathcal{F}(f) \) vanishes on \( \eta(G) \), that is (since \( \eta_\chi(\chi) = \chi(a) \)),

\[
\mathcal{F}(f)(\eta_a) = \int f(\chi)\overline{\eta_a(\chi)}\,d\hat{\lambda}(\chi) = \int f(\chi)\overline{\chi(a)}\,d\hat{\lambda}(\chi) = 0, \quad \text{for all} \ a \in G.
\]

Using Fubini’s theorem, for any \( g \in L^1(G) \), we have

\[
\int f(\chi)\mathcal{F}(g)(\chi)\,d\hat{\lambda}(\chi) = \int f(\chi) \left( \int g(a)\overline{\chi(a)}\,d\lambda(a) \right)\,d\hat{\lambda}(\chi)
= \int \left( \int f(\chi)\overline{\chi(a)}\,d\hat{\lambda}(\chi) \right) g(a)\,d\lambda(a) = 0.
\]

Since \( \mathcal{F}(A(G)) \) is dense in \( L^1(G) \), we must have \( f \equiv 0 \), a contradiction. Therefore, \( \eta \) is surjective, thus an isomorphism. Then using Proposition 9.28, we can show that \( \mathcal{F} \) (defined on \( L^2(\hat{G}) \)) is an isometry between \( L^2(\hat{G}) \) and \( L^2(\hat{\hat{G}}) \).

From now on, we identify \( G \) and \( \hat{\hat{G}} \), unless specified otherwise. We will show that there is another class of functions for which \( \mathcal{F} \) and \( \mathcal{F} \) are mutual inverses. For this we need the following result.

**Proposition 9.31.** For every \( f \in L^1(G) \) and every \( H \in L^1(\hat{G}) \), we have

\[
\int_G f(a)\mathcal{F}(H)(a)\,d\lambda(a) = \int_{\hat{G}} \mathcal{F}(f)(\chi)H(\chi)\,d\hat{\lambda}(\chi).
\]

The proof of Proposition 9.31 is a simple application of Fubini’s theorem.

**Definition 9.13.** Let \( G \) be a locally compact abelian group. We define \( B(G) \) as the set of functions

\[
B(G) = \{ f \in L^1(G) \mid \mathcal{F}(f) \in L^1(\hat{G}) \}.
\]

**Theorem 9.32.** The restriction of \( \mathcal{F} \) to \( B(G) \) is a bijection from \( B(G) \) to \( B(\hat{G}) \), whose inverse is the restriction of \( \mathcal{F} \) to \( B(\hat{G}) \).

**Proof.** If \( f \in B(G) \), then \( \mathcal{F}(f) \in L^1(\hat{G}) \cap C_0(\hat{G}) \subseteq L^1(\hat{G}) \cap L^2(\hat{G}) \). Let \( h = (\mathcal{F} \circ \mathcal{F})(f) \in L^2(\hat{G}) \). For every \( g \in \mathcal{K}_c(G) \), by Theorem 9.29, \( \mathcal{F} \) is an isometry, \( \mathcal{F} \) and \( \mathcal{F} \) are mutual
inverses, so that $F(h) = F(f)$, and Proposition 9.31, we have

$$
\int_G h(a)F(g)(a) \, d\lambda = \langle F(g), h \rangle \\
= \langle g, \overline{F(h)} \rangle \\
= \langle g, \overline{F(h)} \rangle \\
= \int_{\hat{G}} F(h)(\chi)g(\chi) \, d\hat{\lambda}(\chi) \\
= \int_{\hat{G}} F(f)(\chi)g(\chi) \, d\hat{\lambda}(\chi) \\
= \int_G f(a)F(g)(a) \, d\lambda(a).
$$

It follows that

$$
\int_G h(a)F(g)(a) \, d\lambda = \int_G f(a)F(g)(a) \, d\lambda(a) \quad \text{for all } g \in K_C(G),
$$

and we deduce that $h = f \in L^1(G)$. Since $h = (\overline{F} \circ F)(f)$, we conclude that $F(f) \in B(\hat{G})$.

We also proved that $\overline{F} \circ F$ is the identity on $B(G)$. By exchanging the roles of $G$ and $\hat{G}$, we can show that $F \circ \overline{F}$ is the identity on $B(\hat{G})$. The theorem follows immediately.

Note that unlike the situation in Theorem 9.29, in Theorem 9.32, there is no need to extend $F$ (and $\overline{F}$ on $B(\hat{G})$).

**Remark:** The space $B(G)$ is an algebra for the pointwise product and the convolution product. It is also easy to see that $F(fg) = F(f) \ast F(g)$, in addition to $F(f \ast g) = F(f)F(g)$.

The property $F(fg) = F(f) \ast F(g)$ is also satisfied by $L^2(G)$.

**Proposition 9.33.** For any two functions $f, g \in L^2(G)$, we have

$$
F(fg) = F(f) \ast F(g).
$$

Proposition 9.33 is proven in Bourbaki [8] (Chapter 2, Section 1, No. 6, Theorem 3), and Folland [31] (Chapter 4, Section 4.3, Proposition 4.36).

The following proposition explains a phenomenon that we have already observed for $G = \mathbb{T}$ and $G = \mathbb{Z}$.

**Proposition 9.34.** For any locally compact abelian group $G$, the group $G$ is discrete if and only if $\hat{G}$ is compact (and by duality, $G$ is compact if and only if $\hat{G}$ is discrete). Furthermore, if $G$ is compact and endowed with the Haar measure $\lambda$ normalized so that $\lambda(G) = 1$, then the associated measure $\hat{\lambda}$ on $\hat{G}$ is the counting measure. If $G$ is discrete and endowed with the counting measure, then the associated measure $\hat{\lambda}$ on $\hat{G}$ is normalized so that $\hat{\lambda}(\hat{G}) = 1$. 
Proof. Assume that $G$ is discrete. If so $L^1(G)$ is unital with identity $δ_1$. By Theorem 8.16, the algebra $X(L^1(G))$ is compact, and since $\hat{G}$ is homeomorphic to $X(L^1(G))$, we deduce that $\hat{G}$ is compact.

Assume now that $\hat{G}$ is compact. Since the characters are uniformly continuous, there is an open subset $V$ containing the identity in $\hat{G}$ such that: for all $χ ∈ V$, for all $a ∈ G$, we have $|χ(a) - 1| ≤ 1$. Since $a ∈ G$ is arbitrary, we can replace it by $a^n$ for any $n ∈ Z$, and since $χ(a^n) = χ(a)^n$, we get

$$|χ(a)^n - 1| ≤ 1,$$

for all $a ∈ G$ and all $n ∈ Z$.

Since $χ(a)$ is a complex number of unit length, say $χ(a) = \cos θ + i \sin θ$, with $0 ≤ θ < 2π$, we have $χ(a)^n = \cos nθ + i \sin nθ$, and

$$|\cos nθ - 1 + i \sin nθ|^2 = (\cos nθ - 1)^2 + \sin^2 nθ = \cos^2 nθ - 2 \cos nθ + 1 + \sin^2 nθ = 2(1 - \cos nθ).$$

Unless $θ = 0$, we can find some $n$ so that $\cos nθ < 0$, and we get a contradiction to the inequality $|χ(a)^n - 1| ≤ 1$. Therefore, $χ(a) = 1$ for all $a ∈ G$, which implies that $V = \{e\}$ (since the other characters are not the constant character 1). Thus we proved that $\{e\}$ is an open subset of $\hat{G}$, so every singleton subset $\{χ\}$ is open, which means that $\hat{G}$ is discrete. The second part of the proposition is proved in Folland [31] (Chapter 4, Section 2, Proposition 4.24).

We haven’t discussed functions of positive type yet. They play an important role in the theory of unitary representations of a locally compact group. A function $φ ∈ L^∞(G)$ is of positive type if

$$\int (f^* * f)φ dλ ≥ 0,$$

for all $f ∈ L^1(G)$.

Let $P_+(G)$ be the space of functions of positive type. There is a connection with the dual group $\hat{G}$. Indeed, for any measure $μ ∈ M(\hat{G})$, define $φ_μ$ by

$$φ_μ(a) = \int_{\hat{G}} χ(a)dμ(χ).$$

Then a theorem of Bochner states that for any function of positive type $φ ∈ P_+(G)$, there is a unique positive measure $μ ∈ M(\hat{G})$ such that $φ = φ_μ$. We will return to positive functions in Chapter 11 (Section 11.4) and Chapter 14, and refer the reader to Folland for a discussion of this topic; see [31] Chapters 3 and Chapter 4, Theorem 4.18.
Chapter 10
Hilbert Algebras

In order to generalize harmonic analysis to compact groups, we need to introduce group representations. However, it turns out that in order to prove the main theorem of the subject, the Peter–Weyl theorem, one needs the notion of representation of algebras, because there is a bijection between the set of unitary representations of a locally compact group $G$ and the set of nondegenerate representations of the involutive algebra $L^1(G)$. When $G$ is compact, $L^2(G)$ is actually a Hilbert algebra, and there is a beautiful structure theorem for Hilbert algebras which says that such an algebra splits as a Hilbert sum of minimal left ideals, and this result can be used to prove the Peter–Weyl theorem.

The purpose of this chapter is to define the notion of Hilbert algebra and to develop the machinery needed to prove three fundamental theorems (Theorem 10.27, Theorem 10.28, and Theorem 10.30) about complete Hilbert algebras. We also state two important theorems about commutative Hilbert algebras; the Plancherel–Godement theorem and the Bochner-Godement theorem. These theorems will be needed later when we discuss Gelfand pairs. We mostly follow Dieudonné [22] (Chapter XV, Sections 15.5-15.9), occasionally borrowing from Folland [31] (Chapters 1 and 3). This is a rather technical chapter and we do not give all proofs, relying on the above references for details.

We begin with the definition of the notion of Hilbert algebra and to develop the machinery needed to prove three fundamental theorems (Theorem 10.27, Theorem 10.28, and Theorem 10.30) about complete Hilbert algebras. We also state two important theorems about commutative Hilbert algebras; the Plancherel–Godement theorem and the Bochner-Godement theorem. These theorems will be needed later when we discuss Gelfand pairs. We mostly follow Dieudonné [22] (Chapter XV, Sections 15.5-15.9), occasionally borrowing from Folland [31] (Chapters 1 and 3). This is a rather technical chapter and we do not give all proofs, relying on the above references for details.

In Section 10.2 we define positive linear forms and positive Hilbert forms on an involutive algebra $A$. Positive Hilbert forms are positive hermitian forms which may fail to be positive definite but satisfy a kind of adjunction property, namely positive hermitian form $g$ satisfying the condition
\[ g(xy, z) = g(y, x^*z) \quad \text{for all } x, y, z \in A. \]
They can be used to define topologically cyclic representations. A good method for producing positive Hilbert forms is to use positive linear forms, which satisfy the property $f(s^*s) \geq 0$,
for all \( s \in A \), a kind of positive semidefinite property. Then the map \( g \) given by \( g(x, y) = f(y^*x) \) is a positive Hilbert form.

In Section 10.3 we introduce bitraces and Hilbert algebras. A bitrace is a positive Hilbert form \( g: A \times A \to \mathbb{C} \) such that
\[
g(t^*, s^*) = g(s, t), \quad \text{for all } s, t \in A.
\]

The most important concept of this section is the notion of Hilbert algebra. An involutive algebra \( A \) is a Hilbert algebra if its underlying vector space is a hermitian space whose hermitian inner product \( \langle -,- \rangle \) is a bitrace satisfying two extra conditions (U) and (N) (see Definition 10.14). Specifically, the conditions for being a bitrace hold
\[
\langle y^*, x^* \rangle = \langle x, y \rangle
\]
and the following two conditions hold: for every \( x \in A \), there is some \( M_x \geq 0 \) such that
\[
\langle xy, xy \rangle \leq M_x \langle y, y \rangle, \quad \text{for all } y \in A,
\]
and
\[
\text{the subspace spanned by the set } \{ xy \mid x, y \in A \} \text{ is dense in } A.
\]

In general the map \((x, y) \mapsto xy\) is not continuous, so a Hilbert algebra is not a normed algebra in the sense of Definition 8.3. However, if the Hilbert algebra \( A \) is complete, then it can be shown that the map \((x, y) \mapsto xy\) is continuous, and thus that \( A \) is a normable algebra which is a Banach space (see Proposition 10.11).

An important example of a complete Hilbert algebra is the algebra \( L_2(H) \) of Hilbert–Schmidt operators on a separable Hilbert space \( H \). Another very important example of a complete Hilbert algebra is \( L^2(G) \), where \( G \) is a compact, metrizable group.

Section 10.4 which is about complete separable Hilbert algebras contains the most important results of this chapter regarding the structure of such algebras. To understand the structure of complete separable Hilbert algebras we need to study minimal left ideals and the irreducible self-adjoint idempotents, which generate them. Recall that an element \( e \) of an algebra \( A \) is idempotent if \( e^2 = e \), and self-adjoint if \( e = e^* \).

Roughly speaking, the master decomposition theorem (Theorem 10.27) states that given a complete separable Hilbert algebra \( A \), there is an irredundant list \((l_k)_{k \in J}\) of the minimal left ideals of \( A \), and \( A \) is the Hilbert sum of two-sided ideals \( a_k \),
\[
A = \bigoplus_{j \in J} a_k,
\]
where each \( a_k \) is the Hilbert sum obtained by picking a certain number of copies of the minimal left ideal \( l_k \) of \( A \),
\[
a_k = \bigoplus_{j \in l_k} l_j.
\]
with \( l'_j \) isomorphic to \( l_k \).

Each two-sided ideal \( a_k \) contains no closed two-sided ideal other than \((0)\) and \( a_k \). They are said to be topologically simple.

Theorem 10.28 gives the structure of a topologically simple Hilbert algebra. Theorem 10.28 implies that in the Hilbert sum

\[ A = \bigoplus_{k \in J} a_k \]

given by Theorem 10.27, the Hilbert algebra \( a_k \), which is a building block of the decomposition, is either isomorphic to the algebra \( L_2(l_k) \) of Hilbert–Schmidt operators on \( l_k \), or to the finite-dimensional algebra \( \text{End}_C(l_k) \) of all endomorphisms of the vector space \( l_k \). If \( G \) is a metrizable compact group and \( A = L^2(G) \), then every \( a_k \) in the Hilbert sum for \( A \) is isomorphic to the finite-dimensional algebra \( \text{End}_C(l_k) \).

The master decomposition for a nondegenerate continuous representation \( V : A \to \mathcal{L}(H) \) (Theorem 10.30) states that the Hilbert space \( H \) is a Hilbert sum \( H = \bigoplus_{k \in J} H_k \) of subspaces invariant under \( V \), and that the restriction \( V_k \) of \( V \) to \( a_k \) can be considered as a representation of \( a_k \) in \( H_k \). Furthermore, each representation \( V_k \) is the Hilbert sum of irreducible representations, each equivalent to a representation \( U_{l_k} \) canonically associated with a minimal ideal \( l_k \) of \( a_k \).

Section 10.5 discusses the Plancherel–Godement theorem and the Bochner–Godement theorem without proofs. These theorems apply to a commutative Hilbert algebra (not necessarily complete) arising from the quotient of a commutative Hilbert algebra by a left ideal induced by a bitrace satisfying two additional conditions. Therefore we go back to positive Hilbert forms to describe the construction of a certain representation.

The idea is that if \( g \) is a positive Hilbert form on \( A \), it almost defines an inner product, but in general it fails to be positive definite because there may be nonzero elements \( s \in A \) such that \( g(s, s) = 0 \). However, if we take the quotient of \( A \) by the set \( n = \{ s \in A \mid g(s, s) = 0 \} \), which is a left ideal because \( g \) is a positive Hilbert form, then we can define an inner product on the quotient vector space \( A/n \). If \( g \) is a bitrace, then \( A/n \) is an involutive algebra (see Proposition 10.31).

If a positive Hilbert form \( g \) satisfies the analog of Condition (U) of Definition 10.14, namely, for every \( s \in A \), there is some \( M_s \geq 0 \) such that

\[ g(st, st) \leq M_s g(t, t), \quad \text{for all } t \in A, \] (U)

and if the hermitian space \( A/n_g \) is separable, where \( n_g = \{ s \in A \mid g(s, s) = 0 \} \), then \( g \) defines a unitary representation \( U_g : A \to \mathcal{L}(H_g) \), where \( H_g \) is the Hilbert space which is the completion of \( A/n_g \) (see Proposition 10.33).

In general, the representation \( U_g : A \to \mathcal{L}(H_g) \) given by Proposition 10.33 may be degenerate. It is nondegenerate if and only if the following condition holds:

the subspace spanned by the set \( \{ \pi_g(st) \mid s, t \in A \} \) is dense in \( A/n_g \). (N)
If $A$ is a commutative Hilbert algebra and if Property (U) holds, then the representation $U_g: A \to \mathcal{L}(H_g)$ is nondegenerate and the image of $A$ under $U_g$ is a commutative subalgebra of the $C^*$-algebra $\mathcal{L}(H_g)$. Let $\mathcal{A}_g$ be the closure of $U_g(A)$ in $\mathcal{L}(H_g)$, so that $\mathcal{A}_g$ is a commutative $C^*$-algebra.

Roughly speaking, the Plancherel–Godement theorem (Theorem 10.37) states that if $A$ is a commutative involutive algebra, if $g$ is a bitrace on $A$ satisfying Conditions (U) and (N), and if the hermitian space $A/\mathfrak{n}_g$ and the $C^*$-algebra $\mathcal{A}_g \subseteq \mathcal{L}(H_g)$ are separable, then $g$ is obtained from a positive measure by a process of integration from hermitian characters.

## 10.1 Representations of Algebras with Involution

Let $A$ be an algebra with an involution (not necessarily a normed algebra, nor a unital algebra).

If $H$ is a complex vector space, recall that a map $\langle -,-\rangle: H \times H \to \mathbb{C}$ is a hermitian form if it satisfies the following properties for all $x, y, x_1, x_2, y_1, y_2 \in H$ and all $\lambda \in \mathbb{C}$: it is sesquilinear, which means that

$$
\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle,
\langle x, y_1 + y_2 \rangle = \langle x, y_1 \rangle + \langle x, y_2 \rangle,
\langle \lambda x, y \rangle = \lambda \langle x, y \rangle,
\langle x, \lambda y \rangle = \overline{\lambda} \langle x, y \rangle,
$$

and satisfies the hermitian property,

$$
\langle y, x \rangle = \overline{\langle x, y \rangle}.
$$

The hermitian property implies that $\langle x, x \rangle \in \mathbb{R}$ for all $x \in H$.

A hermitian form $\langle -,-\rangle: H \times H \to \mathbb{C}$ is positive if

$$
\langle x, x \rangle \geq 0 \quad \text{for all } x \in H.
$$

A positive hermitian form satisfies the Cauchy-Schwarz inequality:

$$
|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle, \quad \text{for all } x, y \in H.
$$

A positive hermitian form is positive definite if for all $x \in H$,

$$
\langle x, x \rangle = 0 \quad \text{implies that } x = 0,
$$

or equivalently,

$$
\langle x, x \rangle > 0 \quad \text{for all } x \neq 0.
$$
A positive definite hermitian form on $H$ is often called a hermitian inner product on $H$, and $H$ is called a hermitian space (sometimes a pre-Hilbert space).

If $H$ is a hermitian space with a hermitian inner product $\langle -,- \rangle$, then the map $x \mapsto \|x\| = \sqrt{\langle x,x \rangle}$ is a norm on $H$. We say that $H$ is a Hilbert space if $H$ is complete for the norm $\|\|$ (every Cauchy sequence converges).

Let $H$ be a Hilbert space. An orthonormal family $(a_\alpha)_{\alpha \in \Lambda}$ of vectors $a_\alpha \in H$ (which means that $\langle a_\alpha,a_\beta \rangle = 0$ for all $\alpha \neq \beta$, and $\langle a_\alpha,a_\alpha \rangle = 1$, for all $\alpha, \beta \in \Lambda$) is a Hilbert basis of $H$ if the subspace spanned by $(a_\alpha)_{\alpha \in \Lambda}$ (the set of all finite linear combinations of vectors in $(a_\alpha)_{\alpha \in \Lambda}$) is dense in $H$. Every Hilbert space admits a Hilbert basis, and the cardinality of the index set $\Lambda$ is the same for any two Hilbert bases; see Rudin [66] (Chapter 4) and Schwartz [69] (Chapter XXIII).

A Hilbert space is separable if it has a countable Hilbert basis.

**Definition 10.1.** Given an algebra $A$ as above, and a Hilbert space $H$, a representation of $A$ in $H$\footnote{Technically, we are defining unitary representations, but since we shall not discuss other types of representations, we shall suppress the word “unitary.”} is an algebra homomorphism $U: A \to \mathcal{L}(H)$ from $A$ to the involutive algebra $\mathcal{L}(H)$ of continuous linear maps from $H$ to itself, which means that $U$ satisfies the conditions

$$
U(s + t) = U(s) + U(t) \\
U(\lambda s) = \lambda U(s) \\
U(st) = U(s) \circ U(t) \\
U(s^*) = (U(s))^*,
$$

for all $s, t \in A$ and all $\lambda \in \mathbb{C}$. If $A$ is unital with identity element $e$, we require that

$$
U(e) = \text{id}_H.
$$

The Hilbert space $H$ is called the representation space. The representation $U$ is faithful if the homomorphism $s \mapsto U(s)$ is injective, which means that $U(s)(x) = 0$ for all $x \in H$ implies that $U(s) = 0$.

**Remark:** Folland [31] uses the terminology $*$-representation for a representation of an involutive algebra. When different representations $U: A \to \mathcal{L}(H)$ of the same algebra $A$ arise, it is sometimes convenient to denote the representation space by $H_U$. Although a representation of an algebra $A$ consists of a homomorphism $U$ and of a Hilbert space $H$, by abuse of language, we often refer to a representation as $U$.

Definition 10.1 implies that if $s$ is self-adjoint $(s^* = s)$, then $U(s)$ is self-adjoint. Observe that $U(s)$ is not necessarily invertible. Also, if $A$ is a normed algebra, then the map $U: A \to \mathcal{L}(H)$ is not necessarily continuous. However, if $A$ is a unital Banach algebra with involution,
then by Proposition 10.7, the map $U: A \to \mathcal{L}(H)$ is continuous. To simplify notation, we often write $U_s$ instead of $U(s)$.

Recall that if $(H_1, \langle -, -, \rangle_1)$ and $(H_2, \langle -, -, \rangle_2)$ are two Hilbert spaces, a Hilbert space isomorphism is a continuous linear map $T: H_1 \to H_2$ whose inverse is also continuous, and $T$ is an isometry, which means that

$$\langle T(x), T(y) \rangle_2 = \langle x, y \rangle_1 \quad \text{for all } x, y \in H.$$

If $H_1 = H_2$, then a Hilbert space automorphism $T: H \to H$ is an invertible element in $\mathcal{L}(H)$ such that

$$TT^* = T^*T = \text{id},$$

where $T^*$ is the adjoint of $T$ (defined by the property that

$$\langle Tx, y \rangle = \langle x, T^*y \rangle \quad \text{for all } x, y \in H.$$

Such maps are called unitary.

**Definition 10.2.** If $A$ is an algebra (as above) and if $H_1$ and $H_2$ are two Hilbert spaces, two representations $U_1: A \to \mathcal{L}(H_1)$ and $U_2: A \to \mathcal{L}(H_2)$ are equivalent if there a Hilbert space isomorphism $T: H_1 \to H_2$ such that

$$U_2(s) = TU_1(s)T^{-1} \quad \text{for all } s \in A,$$

as illustrated by the following diagram:

$$\begin{array}{ccc}
H_2 & \xrightarrow{T^{-1}} & H_1 \\
\downarrow{v_2(s)} & & \downarrow{v_1(s)} \\
H_2 & \xrightarrow{T} & H_1.
\end{array}$$

It is often useful to make a new representation from old ones using the process of constructing a Hilbert sum. We begin with the simplest case involving two Hilbert spaces. Later we generalize this construction to an arbitrary family of Hilbert spaces.

Let $H_1$ and $H_2$ be two Hilbert spaces, and let $U_1: A \to \mathcal{L}(H_1)$ and $U_2: A \to \mathcal{L}(H_2)$ be two representations. The *Hilbert sum* $H$ of $H_1$ and $H_2$, which is the direct sum $H_1 \oplus H_2$ of $H_1$ and $H_2$ with the hermitian product given by

$$\langle x_1 + x_2, y_1 + y_2 \rangle = \langle x_1, y_1 \rangle_1 + \langle x_2, y_2 \rangle_2,$$

for all $x_1, y_1 \in H_1$ and all $x_2, y_2 \in H_2$, We define the representation $U: A \to H$ by

$$U(s)(x_1 + x_2) = U_1(s)(x_1) + U_2(s)(x_2),$$

for all $x_1 \in H_1$ and all $x_2 \in H_2$. It is immediately verified that $U(s) \in \mathcal{L}(H)$ for all $s \in A$, and that $U$ is a representation of $A$. 


Definition 10.3. The representation $U$ constructed as above from two representations $U_1: A \to \mathcal{L}(H_1)$ and $U_2: A \to \mathcal{L}(H_2)$ is called the Hilbert sum of $U_1$ and $U_2$.

Two key notions of representation theory are invariant subspaces and (topologically) irreducible representations.

Definition 10.4. Let $U: A \to \mathcal{L}(H)$ be a representation of an algebra $A$. A subspace $E$ of $H$ is invariant (or stable) under the representation $U$ if $U(s)(x) \in E$ for all $s \in A$ and all $x \in E$. The representation $U_E: A \to \mathcal{L}(E)$ given by $U_E(s)(x) = U(s)(x)$ for all $x \in E$, is called a subrepresentation of $A$ in $E$.

Observe a small abuse of language: if $E$ is not a closed subspace of $H$, then $E$ is not a Hilbert space, and so $U_E: A \to \mathcal{L}(E)$ is not a representation. Thus the notion of subrepresentation should be defined for closed invariant subspaces of $H$. However, Proposition 10.1 shows that the closure $\overline{E}$ of an invariant subspace $E$ is invariant, so we can define the subrepresentation $U_E: A \to \mathcal{L}(\overline{E})$.

The following facts hold.

Proposition 10.1. Let $U: A \to \mathcal{L}(H)$ be a representation of an algebra $A$.

1. If the subspace $E$ of $H$ is invariant under $U$, then its closure $\overline{E}$ is also invariant under $U$.

2. Let $E$ be a closed subspace of $H$ invariant under $U$. If $E^\perp$ is the orthogonal complement of $E$ in $H$, then $E^\perp$ is invariant under $U$. If $U_1(s)$ and $U_2(s)$ are the restrictions of $U(s)$ to $E$ and $E^\perp$, then the representation $U$ is the Hilbert sum of the representations $U_1$ and $U_2$.

Proof. Part (1) is easy to prove and follows from the continuity of $U(s)$; see Dieudonné [23], (Chapter III, Section 11). For Part (2), let $x \in E$ and $y \in E^\perp$. For any $s \in A$ we have

$$\langle x, U(s)(y) \rangle = \langle (U(s))^*(x), y \rangle = \langle U(s^*)(x), y \rangle = 0,$$

since $E$ is invariant under $U$, so $U(s^*)(x) \in E$, and since $E^\perp$ is the orthogonal complement of $E$ and $y \in E^\perp$. Then $U(s)(y)$ is orthogonal to all $x \in E$, which means that $U(s)(y) \in E^\perp$, so $E^\perp$ is invariant under $U$. The last property is obvious because as $E$ is closed, $H$ is the (algebraic) direct sum $H = E \oplus E^\perp$. \qed

Definition 10.5. Given a Hilbert space $H$, an orthogonal projector is a continuous linear map which is an orthogonal projection onto a closed subspace of $H$.

The following result is not hard to prove; see Dieudonné [22], (Chapter XV, Section 5).

Proposition 10.2. Let $H$ be a Hilbert space. A continuous linear $P$ map on $H$ is an orthogonal projector iff it is idempotent ($P^2 = P \circ P = P$) and hermitian ($P^* = P$).
Here is a convenient way to characterize when a closed subspace is invariant.

**Proposition 10.3.** Let \( U: A \to \mathcal{L}(H) \) be a representation of an algebra \( A \). A closed subspace \( E \) of \( H \) is invariant under \( U \) iff \( P_E U(s) = U(s) P_E \) for all \( s \in A \), where \( P_E: H \to E \) is the orthogonal projection of \( H \) onto \( E \).

The proof of Proposition 10.3 is identical to the proof for group representations; see Proposition 11.6 and its proof. Proposition 10.3 is also proven in Dieudonné [22], (Chapter XV, Section 5, Proposition 15.5.3).

We now generalize the construction of Hilbert sum of two representations to any family of Hilbert spaces. First we need to define the Hilbert sum of any arbitrary family of Hilbert spaces.

Let \((H_\alpha, \langle -,- \rangle_\alpha)\) be a family of Hilbert spaces indexed by some index set \( \Lambda \). In most applications, \( \Lambda = \mathbb{N} \), so for simplicity the reader may assume this. We define the set \( H \) as the set of all sequences \((x_\alpha)_{\alpha \in \Lambda}\) with \( x_\alpha \in H_\alpha \), such that \( \sum_{\alpha \in \Lambda} \|x_\alpha\|^2 < \infty \) (this implies that only countably many elements \( x_\alpha \) are nonzero). We define a vector space structure on \( H \) by defining
\[
(x_\alpha) + (y_\alpha) = (x_\alpha + y_\alpha)
\]
\[
\lambda(x_\alpha) = (\lambda x_\alpha),
\]
with \( x_\alpha, y_\alpha \in H_\alpha \). It is easy to check that these operations make \( H \) into a vector space. We define the inner product \( \langle -, - \rangle \) on \( H \) by
\[
\langle (x_\alpha), (y_\alpha) \rangle = \sum_{\alpha \in \Lambda} \langle x_\alpha, y_\alpha \rangle_\alpha.
\]
It can be verified that \( \langle -, - \rangle \) is a Hermitian product on \( H \). It can also be shown that \( H \) is complete, so it is a Hilbert space. For details, see Dieudonné [23], (Chapter VI, Section 4) and Schwartz [69] (Chapter XXIII).

**Definition 10.6.** Let \((H_\alpha, \langle -,- \rangle_\alpha)_{\alpha \in \Lambda}\) be a family of Hilbert spaces indexed by some index set \( \Lambda \). The space \( H \) constructed as above is called the **Hilbert sum** of the sequence of Hilbert spaces \((H_\alpha)\), and is denoted by
\[
H = \bigoplus_{\alpha \in \Lambda} H_\alpha.
\]

We define continuous injections \( j_\alpha: H_\alpha \to H \) such that \( j_\alpha(x_\alpha) = (0, \ldots, 0, x_\alpha, 0, \ldots) \), with the \( \alpha \)-th term being \( x_\alpha \). Each \( j_\alpha \) is an isomorphism of \( H_\alpha \) onto a closed subspace of \( H \) denoted \( H'_\alpha \). By definition of the inner product on \( H \), \( \langle j_\alpha(x_\alpha), j_\beta(x_\beta) \rangle = 0 \) for all \( \alpha \neq \beta \), all \( x_\alpha \in H_\alpha \), and all \( x_\beta \in H_\beta \). Then for any \( x = (x_\alpha) \in H \), the sequence \( (j_\alpha(x_\alpha)) \in H \) converges to \( x \) and \( x = \sum_{\alpha \in \Lambda} j_\alpha(x_\alpha) \) (the convergence is not necessarily uniform). It follows that the direct sum \( \bigoplus_{\alpha \in \Lambda} H'_\alpha \) is dense in \( H \) (recall that \( \bigoplus_{\alpha \in \Lambda} H'_\alpha \) consists of all sequences
(x_\alpha)_{\alpha \in \Lambda} such that x_\alpha = 0 for all but finitely many indices \alpha). We often identify H_\alpha and H'_\alpha. The above construction defines what we might call an external Hilbert sum.

Unfortunately, the notation

\[ H = \bigoplus_{\alpha \in \Lambda} H_\alpha \]

for the Hilbert sum of a family \((H_\alpha, \langle -,- \rangle_\alpha)_{\alpha \in \Lambda}\) of Hilbert spaces clashes with the notion of algebraic direct sum

\[ \bigoplus_{\alpha \in \Lambda} H_\alpha \]

of vector spaces. The second definition refers to the subspace of sequences \((x_\alpha)_{\alpha \in \Lambda}\) such that \(x_\alpha = 0\) for all but finitely many indices \(\alpha\). If we temporarily denote the Hilbert sum by

\[ H = \bigoplus_{\alpha \in \Lambda} H_\alpha, \]

then we see that if the index set \(\Lambda\) is finite, then the two notions agree. But if \(\Lambda\) is infinite, then the algebraic direct sum is a proper subspace of the Hilbert sum, because the Hilbert sum consists of sequences \((x_\alpha)_{\alpha \in \Lambda}\) such that \(\sum_{\alpha \in \Lambda} \|x_\alpha\|^2 < \infty\), which may contain a countably infinite number of nonzero \(x_\alpha\). The direct sum is a dense subspace of the Hilbert sum. It is usually clear from the context which sum of spaces is intended (typically, when we refer to Hilbert spaces, we mean a Hilbert sum), so we will not use the heavier notation \(\bigoplus^H\) for Hilbert sums.

We also have the following proposition proven in Dieudonné [23], (Chapter VI, Section 4) and Schwartz [69] (Chapter XXIII), which gives the definition of an internal Hilbert sum (in the sense that the \(H_\alpha\) are subspaces of an already given space \(H\)).

**Proposition 10.4.** Let \(H\) be a Hilbert space, and let \((H_\alpha)_{\alpha \in \Lambda}\) be a family of closed subspaces of \(H\) satisfying the following conditions:

1. For all \(\alpha \neq \beta\), the subspaces \(H_\alpha\) and \(H_\beta\) are orthogonal.
2. The direct sum \(\bigoplus_{\alpha \in \Lambda} H_\alpha\) is dense in \(H\).

If \(E\) is the Hilbert sum of the sequence \((H_\alpha)\), then there is a unique isomorphism of \(H\) onto \(E\), which, on each \(H_\alpha\), coincides with the injection \(j_\alpha\) of \(H_\alpha\) into \(E\).

If \(\Lambda\) is finite, (2) is equivalent to the fact that \(H\) is the direct sum of the \(H_\alpha\) (because, if \(E\) is any closed subspace of a Hilbert space, then \(H = E \oplus E^\perp\) as a direct sum, so by picking \(E = H_{\alpha_1}\), we see that the direct sum \(\bigoplus_{j=2}^n H_{\alpha_j}\) is dense in \(E^\perp\), and we finish by induction).

Assume a Hilbert space \(H\) is the Hilbert sum of a sequence \((H_\alpha)_{\alpha \in \Lambda}\) of subspaces invariant under a representation \(U\) of \(A\). For every \(s \in A\), let \(U_\alpha(s)\) be the restriction of \(U\) to \(H_\alpha\), so that the map \(s \mapsto U_\alpha(s)\) is a representation of \(A\) in \(H_\alpha\). By abuse of language, we say that
$U$ is the Hilbert sum of the representations $U_\alpha$. For each $s \in A$ and each $x = \sum \alpha x_\alpha \in H$, where $x_\alpha \in H_\alpha$, we have

$$U(s)(x) = \sum_\alpha U_\alpha(s)(x_\alpha),$$

and

$$\sum_\alpha \|U_\alpha(s)(x_\alpha)\|^2 = \|U(s)(x)\|^2.$$

The notion of a topologically irreducible representation is similar in spirit to the notion of prime number. Namely, a topologically irreducible representation cannot be decomposed into simpler representations. It is one of the most important concepts in representation theory.

**Definition 10.7.** A representation $U : A \to L(H)$ of $A$ in $H$ is topologically irreducible if $H \neq (0)$ and if there is no closed subspace $E$ of $H$ other than $\{0\}$ and $H$ which is invariant under $U$.

**Proposition 10.5.** Let $U : A \to L(H)$ be a representation of $A$ in $H$, let $E$ be the closure of the subspace spanned by the set

$$\{U(s)(x) \mid s \in A, x \in H\},$$

and let $E'$ be the set

$$E' = \{x \in H \mid U(s)(x) = 0, \text{ for all } s \in A\}.$$

Then $E$ and $E'$ are invariant under $U$, and $E'$ is the orthogonal complement of $E$ in $H$ (that is, $E' = E^\perp$ and $H = E \oplus E'$).

**Proof.** Since $U(rs) = U(r)U(s)$ for all $r, s \in A$, it is clear that $E$ and $E'$ are invariant under $U$. Let us prove that $E^\perp \subseteq E'$, where $E^\perp$ is the orthogonal complement of $E$ in $H$. We already know from Proposition 10.1 that $E^\perp$ is invariant under $U$ so for any $x \in E^\perp$ we have $U(s)(x) \in E^\perp$ for all $s \in A$. But by definition of $E$, we have $U(s)(x) \in E$, so $U(s)(x) \in E \cap E^\perp = (0)$, which means that $U(s)(x) = 0$ for all $s \in A$, that is, $x \in E'$. Therefore $E^\perp \subseteq E'$.

Next we prove that $E' \subseteq E^\perp$. If $x \in E'$, for any $s \in A$ and any $y \in H$ we have

$$\langle x, U(s)(y) \rangle = \langle U(s^*)(x), y \rangle = 0,$$

since $x \in E'$ means that $U(s)(x) = 0$ for all $s \in A$, and since $s$ and $y$ are arbitrary, by definition of $E$, this means that $x$ is orthogonal to $E$, that is, $x \in E^\perp$, and thus $E' \subseteq E^\perp$.

In conclusion, $E' = E^\perp$.

**Definition 10.8.** Let $U : A \to L(H)$ be a representation of $A$ in $H$. The subspace $E$ defined in Proposition 10.5 is called the essential subspace for $U$. If $E' = (0)$, then we say that the representation is nondegenerate.
We see immediately that a representation $U$ of $A$ in $H$ is nondegenerate if the subspace spanned by set $\{U(s)(x) \mid s \in A, x \in H\}$ is dense in $H$. If $A$ is unital, since $U(e) = \text{id}$, this is always the case.

**Definition 10.9.** Let $U : A \to \mathcal{L}(H)$ be a representation of $A$ in $H$. A vector $x_0 \in H$ is called a totalizer or totalizing vector (or cyclic vector) for the representation $U$ if the subspace of $H$ spanned by the set $\{U(s)(x_0) \mid s \in A\}$ is dense in $H$. Equivalently if $M_{x_0}$ denotes the closure of the set $\{U(s)(x_0) \mid s \in A\}$, called the cyclic subspace generated by $x_0$, which is invariant under $U$, then $x_0$ is a totalizer (a cyclic vector) if $M_{x_0} = H$. A representation which admits a totalizer is said to be topologically cyclic.

The following fact follows immediately from the definitions: A representation $U$ is topologically irreducible if every nonzero vector $x_0 \in H$ is a totalizer. The importance of totalizers stems from the following result.

**Proposition 10.6.** Assume that the algebra $A$ is unital, and let $U : A \to \mathcal{L}(H)$ be a representation of $A$ in $H \neq (0)$ (which must be nondegenerate). Then $H$ is the Hilbert sum of a sequence $(H_\alpha)_{\alpha \in \Lambda}$ of closed subspaces $H_\alpha \neq (0)$ of $H$ invariant under $U$, and such that the restriction of $U$ to each $H_\alpha$ is topologically cyclic. If $H$ is separable, the family $\Lambda$ is countable (possibly finite).

Proposition 10.6 is proven in the separable case in Dieudonné [22] (Chapter XV, Section 5), and in general, using Zorn’s lemma; see Folland [31] (Chapter 3, Proposition 3.3).

The following result will be needed.

**Proposition 10.7.** If $A$ is a unital Banach algebra with involution, then every representation $U : A \to \mathcal{L}(H)$ satisfies the condition $\|U(s)\| \leq \|s\|$, which implies that $U$ is a continuous mapping from $A$ to $\mathcal{L}(H)$.

**Proof.** Recall that $\mathcal{L}(H)$ is a $C^*$-algebra, so by Proposition 8.27 $\rho(T) = \|T\|$ for all $T \in \mathcal{L}(H)$, so we have

$$\|U(s)\|^2 = \|U(s)^*U(s)\| = \rho(U(s)^*U(s)).$$

Since $U(s)^*U(s) = U(s^*s)$, by Property (5) just after Definition 8.4 and Proposition 8.14, we have

$$\rho(U(s)^*U(s)) = \rho(U(s^*s)) \leq \rho(s^*s) \leq \|s^*s\| \|s\| = \|s\|^2,$$

which proves our result. 

\[\square\]

## 10.2 Positive Linear Forms and Positive Hilbert Forms

The Peter–Weyl theorem can be obtained from a structure theorem about certain kinds of algebras with a hermitian inner product satisfying special conditions. Such inner products are bitraces, which are special kinds of positive Hilbert forms. A good method for producing positive Hilbert forms is to use positive linear forms.
Definition 10.10. Let $A$ be an involutive algebra (not necessarily unital and not necessarily normed). A linear form $f : A \to \mathbb{C}$ is **positive** if

$$f(s^*s) \geq 0 \quad \text{for all } s \in A.$$

Positive linear forms arise from representations as follows.

Definition 10.11. Let $U : A \to \mathcal{L}(H)$ be a representation of $A$ in $H$. For any $x_0 \in H$, define the map \( f_{x_0} : A \to \mathbb{C} \) by

$$f_{x_0}(s) = \langle U(s)(x_0), x_0 \rangle, \quad s \in A.$$

Proposition 10.8. The map $f_{x_0} : A \to \mathbb{C}$ is a positive linear form.

**Proof.** It is clear that $f_{x_0}$ is a linear form, and since $U(s^*s) = U(s^*)U(s)$, we have

$$f_{x_0}(s^*s) = \langle (U(s^*)U(s))(x_0), x_0 \rangle = \langle U(s)(x_0), U(s)(x_0) \rangle = \|U(s)(x_0)\|^2 \geq 0.$$  \hfill \Box

A positive linear form also defines a positive hermitian form as follows.

Proposition 10.9. Given any positive linear form $f$ on an involutive algebra $A$, let $g : A \times A \to \mathbb{C}$ be the map given by

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

Then $g$ is a positive hermitian form, and the following properties hold:

1. For all $x, y \in A$, we have
   $$f(x^*y) = f(y^*x).$$

2. For all $x, y \in A$, we have
   $$|f(y^*x)|^2 \leq f(x^*)f(y^*y).$$

3. If $A$ is unital, then
   $$f(x^*) = \overline{f(x)}, \quad |f(x)|^2 \leq f(e)f(x^*x).$$

**Proof.** To prove (1), since $f$ is linear, we have

\[
g(x + y, x + y) = f((x + y)^*(x + y)) = f(x^*x + x^*y + y^*x + y^*y) = f(x^*x) + f(x^*y) + f(y^*x) + f(y^*y) = g(x, x) + g(y, x) + g(x, y) + g(y, y).
\]

Since $g(x + y, x + y), g(x, x)$, and $g(y, y)$ are real (and nonnegative), we must have

$$\Re(g(y, y)) = -\Im(g(x, x)).$$
If we replace \( x \) by \( ix \), this becomes
\[
\mathcal{R}(g(y, x)) = \mathcal{R}(g(x, y)).
\]
Therefore,
\[
g(y, x) = f(x^*y) = \overline{f(y^*x)} = \overline{g(x, y)},
\]
as claimed. Part (2) is the Cauchy–Schwartz inequality, and (3) follows from (1) and (2) by replacing \( y \) by \( e \) and the fact that \( e^* = e \).

The hermitian form \( g \) obtained from the positive linear form \( f \) is not arbitrary since it satisfies the condition
\[
g(xy, z) = f(z^*xy) = f((x^*z)^*y) = g(y, x^*z)
\]
for all \( x, y, z \in A \).

This motivates the following definition.

**Definition 10.12.** A positive Hilbert form on an involutive algebra \( A \) is a positive hermitian form \( g \) satisfying the condition
\[
g(xy, z) = g(y, x^*z)
\]
for all \( x, y, z \in A \).

If \( A \) is unital, then every positive Hilbert \( g \) form comes from a positive linear form \( f \): define \( f \) by \( f(s) = g(s, e) \). Indeed, the positive Hilbert form \( g' \) induced by \( f \) is given by
\[
g'(s, t) = f(t^*s) = g(t^*s, e) = g(s, t),
\]
by setting \( x = t^*, y = s, \) and \( z = e \) in the equation of Definition 10.12.

Given a representation \( U : A \to \mathcal{L}(H) \), observe that the positive Hilbert form \( g_{x_0} \) associated with the positive linear form \( f_{x_0} \) is given by
\[
g_{x_0}(s, t) = \langle U(s)(x_0), U(t)(x_0) \rangle,
\]
for all \( s, t \in A \).

Remarkably, every topologically cyclic representation arises from a positive Hilbert form, but we won’t need this fact until Section 10.5, so we postpone discussing this matter.

### 10.3 Traces, Bitraces, Hilbert Algebras

If \( A \) is an involutive algebra and if \( f \) is a positive linear form on \( A \), in general \( f(st) \neq f(ts) \).

**Definition 10.13.** Let \( A \) be an involutive algebra. A trace on \( A \) is a positive linear form \( f : A \to \mathbb{C} \) such that
\[
f(st) = f(ts) \quad \text{for all } s, t \in A.
\]

A bitrace is a positive Hilbert form \( g : A \times A \to \mathbb{C} \) such that
\[
g(t^*, s^*) = g(s, t), \quad \text{for all } s, t \in A.
\]
For example, if $H$ is a finite-dimensional vector space of dimension $n$ with a hermitian inner product, and if $A = \mathcal{L}(H)$, the algebra of linear maps from $H$ to itself, for any orthonormal basis $(e_1, \ldots, e_n)$ of $H$, then for any linear map $T \in \mathcal{L}(H)$, the linear form
\[
\text{Tr}(T) = \sum_{i=1}^{n} \langle T(e_i), e_i \rangle
\]
is a trace. In fact, $\text{Tr}(T) = \sum_{i=1}^{n} a_{ii}$, the trace of the matrix $(a_{ij})$ representing $T$ over the basis $(e_1, \ldots, e_n)$. If $H$ is an infinite-dimensional Hilbert space, it can be shown that there exists no trace on $\mathcal{L}(H)$.

**Proposition 10.10.** Let $f : A \to \mathbb{C}$ be a positive linear form on an involutive algebra $A$. If $f(ss^*) = f(s^*s)$, for all $s \in A$,
then $f$ is a trace.

**Proof.** By replacing $s$ by $s + t$, using Proposition 10.9(1), we have
\[
f((s + t)(s + t)^*) = f(ss^* + st^* + ts^* + tt^*)
\]
\[
= f(ss^*) + f(st^*) + f(ts^*) + f(tt^*)
\]
\[
= f(s^*s) + f(st^*) + \overline{f(st^*)} + f(t^*t),
\]
and similarly
\[
f((s + t)^*(s + t)) = f(s^*s) + f(t^*s) + f(t^*t)
\]
and since $f((s + t)(s + t)^*) = f((s + t)^*(s + t))$, we get
\[
f(st^*) + \overline{f(st^*)} = f(t^*s) + \overline{f(t^*s)},
\]
that is
\[
\Re(f(t^*s)) = \Re(f(st^*)).
\]
If we replace $s$ by $is$, we get
\[
\Im(f(t^*s)) = \Im(f(st^*)),
\]
so that
\[
f(t^*s) = f(st^*),
\]
for all $s, t \in A$, as claimed. \(\Box\)

Observe that if the positive Hilbert form $g$ arises from a positive linear form $f$, as $g(s, t) = f(t^*s)$, then expressing that $g$ is bitrace says that
\[
f(t^*s) = g(s, t) = g(t^*, s^*) = f((s^*)^*t^*) = f(st^*),
\]
namely that $f$ is a trace.

One of the most important example of a bitrace arises when $G$ is a compact group. In this case, the inner product on the involutive Banach algebra $L^2(G)$ is a bitrace. In fact, this bitrace satisfies two more properties that makes $L^2(G)$ into a Hilbert algebra, defined next.
10.3. TRACES, BITRACES, HILBERT ALGEBRAS

Definition 10.14. An involutive algebra $A$ is a Hilbert algebra if its underlying vector space is a hermitian space whose hermitian inner product $\langle -, - \rangle$ is a bitrace satisfying two extra conditions (U) and (N). Specifically, the conditions for being a bitrace hold

\[
\langle y^*, x^* \rangle = \langle x, y \rangle \tag{1}
\]
\[
\langle xy, z \rangle = \langle y, x^* z \rangle, \tag{2}
\]
and the following two conditions hold: for every $x \in A$, there is some $M_x \geq 0$ such that

\[
\langle xy, xy \rangle \leq M_x \langle y, y \rangle, \quad \text{for all } y \in A, \tag{U}
\]

and

the subspace spanned by the set $\{xy \mid x, y \in A\}$ is dense in $A$. \tag{N}

From Conditions (1) and (2) and the hermitian property, we get

\[
\langle yx, z \rangle = \langle z^*, x^* y^* \rangle = \langle x^* y^*, z^* \rangle = \langle y^*, xz^* \rangle = \langle xz^*, y^* \rangle = \langle y, zx^* \rangle,
\]

so

\[
\langle yx, z \rangle = \langle y, zx^* \rangle. \tag{2'}
\]

By (1) and (U), for every $x \in A$, we also have

\[
\langle yx, yx \rangle = \langle x^* y^*, x^* y^* \rangle \leq M_{x^*} \langle y^*, y^* \rangle = M_{x^*} \langle y, y \rangle,
\]

namely

\[
\langle yx, yx \rangle \leq M_{x^*} \langle y, y \rangle, \quad \text{for all } y \in A. \tag{U'}
\]

The inequality (U) says that, for $x$ fixed, the linear map $y \mapsto xy$ is continuous, and the inequality (U') says that, for $y$ fixed, the linear map $x \mapsto xy$ is continuous.

Beware that in general, this does not imply that the map $(x, y) \mapsto xy$ is continuous. Thus, in general, a Hilbert algebra is not a normed algebra in the sense of Definition 8.3, because in a normed algebra, the map $(x, y) \mapsto xy$ is continuous. However, if the Hilbert algebra $A$ is complete, then using Baire’s theorem, it can be shown that the map $(x, y) \mapsto xy$ is continuous; see Dieudonné [22] (Chapter XII, Section 16, Problem 8(c)). As a consequence, we can show that a complete Hilbert algebra is normable, and thus a Banach algebra.

Proposition 10.11. If $A$ is a complete Hilbert algebra, then there is a norm $\| \|_b$ equivalent to the norm $x \mapsto \|x\| = \sqrt{\langle x, x \rangle}$ induced by the inner product $\langle -, - \rangle$ on $A$, such that $A$ is a Banach algebra with the norm $\| \|_b$.

Proof. First, assume that $A$ is not unital. Since $A$ is complete, the bilinear map $(x, y) \mapsto xy$ is continuous, so there is some constant $c > 0$ such that $\|xy\| \leq c \|x\| \|y\|$ for all $x, y \in A$. Since $c \|xy\| \leq c \|x\| c \|y\|$, if we let $\|x\|_b = c \|x\|$, we obtain a norm equivalent to $\| \|$ such that

\[
\|xy\|_b \leq \|x\|_b \|y\|_b,
\]
and with this norm, $A$ is a Banach algebra.

If $A$ has a multiplicative unit $e$, then the norm $\| \cdot \|_b$ also needs to satisfy the condition $\|e\|_b = 1$, so we need a different construction. For every $x \in A$, let $L_x : A \to A$ be the linear map given by

$$L_x(y) = xy, \quad y \in A.$$ 

Since the bilinear map $(x, y) \mapsto xy$ is continuous, the linear map $L_x$ is continuous. We check immediately that

$$L_{x+y} = L_x + L_y,$$
$$L_{\alpha x} = \alpha L_x,$$
$$L_{xy} = L_x \circ L_y.$$

Therefore, the map $L : A \to \mathcal{L}(A)$ given by $x \mapsto L_x$ is an algebra homomorphism. The homomorphism $L$ is injective, because if $L_x = 0$, then $L_x(e) = xe = x = 0$. We claim that there is a constant $c > 0$ such that $\|L_x\| \leq c \|x\|$. Recall that for the operator norm $\|L_x\|$, we have

$$\|L_x\| = \sup\{\|L_x(y)\| : \|y\| = 1\} = \sup\{\|xy\| : \|y\| = 1\},$$

and since the bilinear map $(x, y) \mapsto xy$ is continuous

$$\|L_x\| = \sup\{\|xy\| : \|y\| = 1\} \leq \sup \{c \|x\| \|y\| : \|y\| = 1\} = c \|x\|.$$ 

On the other hand, since $\|L_x\|$ is the operator norm,

$$\|x\| = \|xe\| = \|L_x(e)\| \leq \|L_x\| \|e\|,$$

Therefore, if we let $\|x\|_b = \|L_x\|$, we obtain a norm on $A$ equivalent to the norm $\| \cdot \|$, and $A$ is a normed algebra with the norm $\| \cdot \|_b$, since $\|L_{xy}\| = \|L_x \circ L_y\| \leq \|L_x\| \|L_y\|$.

**Remark:** The proof of Proposition 10.11 shows that if $A$ is an algebra whose topology is defined by a norm $\| \cdot \|$ and if the bilinear map $(x, y) \mapsto xy$ is continuous, then there is a norm $\| \cdot \|_b$ equivalent to the norm $\| \cdot \|$, such that $A$ is a normed algebra with the norm $\| \cdot \|_b$.

The following result will be needed to prove that if $A$ is a Hilbert algebra and if $x \in A$ with $x \neq 0$, then $Ax \neq (0)$.

**Proposition 10.12.** Let $A$ be a Hilbert algebra. For every $x \in A$, if $x^*x = 0$, then $x = 0$.

**Proof.** Since $x^*x = 0$, from Property (2) we have

$$\langle xy, xy \rangle = \langle x^*xy, y \rangle = 0,$$

hence $xy = 0$ for all $y \in A$ (since the hermitian product is positive definite). By (2'), we get

$$\langle x, yz \rangle = \langle xz^*, y \rangle = 0$$

for all $y, z \in A$, and since by (N) the subspace spanned by the set $\{xy : x, y \in A\}$ is dense in $A$, we conclude that $x = 0$ (since the hermitian product is positive definite). $\square$
10.3. TRACES, BITRACES, HILBERT ALGEBRAS

Example 10.1. Let $H$ be a separable Hilbert space. Recall that this means that $H$ has a countable Hilbert basis, that is, a countable orthonormal basis $(a_i)_{i \geq 1}$ such that the subspace spanned by $(a_i)_{i \geq 1}$ is dense in $H$ (for every vector $x \in H$, there is some sequence $(x_n)$, with $x_n$ a linear combination $\sum_{k \in I_n} \lambda_k a_k$ where $I_n$ a finite set, and $x_n$ converges to $x$).

A linear map $u \in \mathcal{L}(H)$ is a Hilbert–Schmidt operator if the series $\sum_{n=1}^{\infty} \|u(a_n)\|^2$ converges, that is, $\sum_{n=1}^{\infty} \|u(a_n)\|^2 < \infty$. It can be shown that the quantity $\sum_{n=1}^{\infty} \|u(a_n)\|^2$ is independent of the Hilbert basis $(a_n)$. Then we define the map $u \mapsto \|u\|_{HS}$ on the set $\mathcal{L}_2(H)$ of Hilbert–Schmidt operators by

$$\|u\|^2_{HS} = \sum_{n=1}^{\infty} \|u(a_n)\|^2.$$ 

We have $\|u\|_{HS} = \|u^*\|_{HS}$. For any two Hilbert–Schmidt operators $u, v \in \mathcal{L}_2(H)$, we have

$$\|u \circ v\|_{HS} \leq \|u\|_{\infty} \|v\|_{HS}.$$ 

It can also be shown that if $u \in \mathcal{L}_2(H)$, then

$$\|u\|_{\infty} \leq \|u\|_{HS}.$$ 

Then it can be shown that with the norm $\|\|_{HS}$, the space $\mathcal{L}_2(H)$ of Hilbert–Schmidt operators is an involutive Banach algebra under composition, with the involution given by $u \mapsto u^*$ (where $u^*$ is the adjoint of $u$). The space $\mathcal{L}_2(H)$ is a self-adjoint two-sided ideal in the involutive Banach algebra $\mathcal{L}(H)$, but in general, it is not closed in $\mathcal{L}(H)$.

The space $\mathcal{L}_2(H)$ contains the continuous linear maps of finite rank. The algebra $\mathcal{L}_2(H)$ is not unital because the identity map is not a Hilbert–Schmidt operator, and it is not a $C^*$-algebra.

If $E$ and $F$ are two normed vector spaces, a linear map $u: E \to F$ is a compact operator if the closure of $f(B)$ is compact for every bounded subset $B$ of $E$. A compact operator is continuous. Every Hilbert–Schmidt operator is a compact operator, but the converse is false. The above facts are proven in Dieudonné [22] (Chapter XV, Section 4).

For any two Hilbert–Schmidt operators $u, v \in \mathcal{L}_2(H)$, it can be shown that the quantity

$$g(u, v) = \sum_{n \geq 1} \langle u(a_n), v(a_n) \rangle$$

is defined and independent of the Hilbert basis $(a_n)$. Then it can be shown that $g$ is a hermitian inner product which is a bitrace such that $g(u, u) = \|u\|_{HS}^2$; see Dieudonné [22] (Chapter XV, Section 7).

If $H$ is a finite-dimensional Hilbert space of dimension $n$, then for every linear map $u: H \to H$, for every orthonormal basis $(a_1, \ldots, a_n)$, the quantity $\|u\|^2_{HS} = \sum_{i=1}^{n} \|u(a_i)\|^2$ is defined. Since $\|u(a_i)\|^2 = \langle u(a_i), u(a_i) \rangle = \langle (u^* \circ u)(a_i), a_i \rangle$, we have

$$\|u\|^2_{HS} = \sum_{i=1}^{n} \langle (u^* \circ u)(a_i), a_i \rangle = \text{tr}(u^* \circ u) = \text{tr}(u \circ u^*).$$
which is just the \textit{Frobenius norm} (also called \textit{Hilbert–Schmidt norm}). Then
\[
g(u,v) = \sum_{i=1}^{n} \langle u(a_n), v(a_n) \rangle = \sum_{i=1}^{n} \langle (v^* \circ u)(a_n), a_n \rangle = \text{tr}(v^* \circ u)
\]
is the corresponding inner product, denoted by $\langle u,v \rangle_{\text{HS}}$.

If $\mathcal{L}_2(H)$ is the involutive algebra of Hilbert–Schmidt operator of Example 10.1, then it can be shown that the bitrace $g$ satisfies the properties (U) and (N), so it is a Hilbert algebra; see Dieudonné [22] (Chapter XV, Section 7).

Recall from Proposition 15.43 that if a topological space is metrizable and compact, then it is separable (which means that it contains a countable dense subset).

\textbf{Proposition 10.13.} If $G$ is a compact metrizable group, then the involutive (complex) Banach algebra $\mathbb{L}^2(G)$ is a separable Hilbert algebra.

\textit{Proof sketch.} Proposition 10.13 is proven in Dieudonné [22] (Chapter XXI, Section 2). Condition (1) follows from the definition of the inner product on $\mathbb{L}^2(G)$. Since by Proposition 6.48, $\|f * g\|_\infty \leq \|f\|_2 \|g\|_2$ (where $\|\|_2$ is the $\mathbb{L}^2$ semi-norm on $\mathbb{L}^2(G)$), we also have $\|f * g\|_2 \leq \|f\|_2 \|g\|_2$, and Condition (U) follows. Condition (N) is a corollary of regularization (see Section 6.12). Finally, Condition (2), namely
\[
\langle f * g, h \rangle = \langle g, f^* * h \rangle
\]
is easily shown if $f$ is real and continuous. In this case, $f^* = \hat{f}$. We obtain the formula in general by continuity and using the fact that $\mathcal{K}_R(G)$ is dense in $\mathbb{L}^2(G)$.

In the next section we consider the special case in which a Hilbert algebra is complete.

\section{Complete Separable Hilbert Algebras}

We now consider the case where the Hilbert algebra $A$ is complete, which means that it is a Hilbert space. Although this is not obvious, as said in the previous section, it can be shown that the map $(x, y) \mapsto xy$ from $A \times A$ to $A$ is continuous. Consequently, by Proposition 10.11, a complete Hilbert algebra is normable, and thus a Banach algebra.

Our main goal is to show that every separable complete Hilbert algebra is the (countable) Hilbert sum of two-sided ideals $\mathfrak{a}_k$, where each ideal $\mathfrak{a}_k$ is the (countable) Hilbert sum of minimal left ideals all isomorphic to a common left ideal (Theorem 10.27). This is a beautiful and powerful result which is one of the main steps in proving the Peter–Weyl theorem.

Following Dieudonné, we only deal with the case where the Hilbert algebras $A$ are separable, because then only countable Hilbert sums are needed. This implies that when we consider the Hilbert algebra $A = \mathbb{L}^2(G)$, the group $G$ is metrizable and compact. This is
not a serious restriction, because every Lie group, being a second-countable manifold, is metrizable. The reader who wishes to see an exposition of the Peter–Weyl theorem in the general case of an arbitrary compact group is invited to consult Folland [31] (Chapter 5). Hilbert sums indexed by arbitrary (possibly uncountable) index sets arise.

We are led to the study of minimal left ideals, and to irreducible self-adjoint idempotents, which generate them. Recall that an element $e$ of an algebra $A$ is idempotent if $e^2 = e$, self-adjoint if $e = e^*$. Complete proofs are provided in Dieudonné [22] (Chapter XV, Section 8). There are many tedious technical details so to make it easier on the reader, we decided to only state most results (except the most important ones) without proof.

First observe that every closed self-adjoint subalgebra $B$ (which means that $B^* = B$) of a complete Hilbert algebra $A$ is a complete Hilbert algebra. Certain representations play a crucial role.

**Proposition 10.14.** Let $A$ be a complete Hilbert algebra. For any closed left ideal $b$ in $A$, let $U_b$ be the map from $A$ to $\mathcal{L}(b)$ given by

$$U_b(x)(y) = xy, \quad x \in A, \ y \in b.$$ 

Observe that $U_b(x)$ is left multiplication by $x \in A$. Then $U_b$ is a representation of $A$ in $b$.

**Proof.** For every $x \in A$, the linear map $U_b(x)$ from $b$ to itself is continuous because of Condition (U). The verification that $U_b(x_1x_2) = U_b(x_1) \circ U_b(x_2)$ is immediate. For all $y, z \in b$, by Condition (2), we have

$$\langle U_b(x)^*(y), z \rangle = \langle y, U_b(x)(z) \rangle = \langle y, xz \rangle = \langle x^*y, z \rangle = \langle U_b(x^*)(y), z \rangle,$$

which implies that $U_b(x)^* = U_b(x^*)$. If $A$ has the unit element $e$, then $U_b(e)$ is the identity transformation. \qed

**Definition 10.15.** Let $A$ be a complete Hilbert algebra. For any closed left ideal $b$ in $A$, let $U_b : A \to \mathcal{L}(b)$ be the representation given by

$$U_b(x)(y) = xy, \quad x \in A, \ y \in b.$$ 

The representation $U_A : A \to \mathcal{L}(A)$ is called the regular representation of $A$.

When $b = A$, we usually write $U(x)$ instead of $U_A(x)$. By Proposition 10.12, the representation $U_A$ is faithful. Moreover, since the map $(x, y) \mapsto xy$ from $A \times A$ to $A$ is continuous, the map $x \mapsto U_A(x)$ from $A$ to $\mathcal{L}(A)$ is also continuous.
Let $A$ be a complete Hilbert algebra. For each left ideal $b$ of $A$, the set $b^* = \{ x^* \mid x \in b \}$ is a right ideal in $A$. We now start a fairly long chain of definitions and results leading to our main result (Theorem 10.27).

In the sequel, $A$ denotes a complete Hilbert algebra. Two keys concepts are the notion of an irreducible self-adjoint idempotent and of a minimal left ideal.

**Proposition 10.15.** For every left ideal $l$ of $A$, the orthogonal complement $l^\perp$ of the closure of $l$ is a left ideal.

This proposition gives useful properties of idempotents.

**Proposition 10.16.** Let $e \neq 0$ be an idempotent element in $A$ ($e^2 = e$). Then the following properties hold:

1. $\|e\| \geq 1$;
2. $e^*$ is idempotent;
3. The left ideal $Ae$ is equal to the set $\{ x \in A \mid x = xe \}$, and is closed in $A$.

The next proposition gives orthogonality properties of self-adjoint idempotents.

**Proposition 10.17.** If $e_1$ and $e_2$ are self-adjoint idempotents in $A$, then the following properties are equivalent:

1. $\langle e_1, e_2 \rangle = 0$;
2. $e_1 e_2 = 0$;
3. $e_2 e_1 = 0$.

The following proposition shows that there are plenty nonzero self-adjoint idempotents.

**Proposition 10.18.** Every left ideal $l \neq (0)$ in $A$ contains a nonzero self-adjoint idempotent.

The proof uses the representation $U_A$.

**Definition 10.16.** A self-adjoint idempotent $e \neq 0$ is reducible if there exist two orthogonal nonzero self-adjoint idempotents $e_1$ and $e_2$ such that $e = e_1 + e_2$. If a self-adjoint idempotent $e \neq 0$ is not reducible, we say that it is irreducible.

By Proposition 10.17, if $e = e_1 + e_2$ is reducible, then $ee_1 = e_1 e = e_1$ and $ee_2 = e_2 e = e_2$.

Proposition 10.19 shows that irreducible self-adjoint idempotents are the building blocks for self-adjoint idempotents.

**Proposition 10.19.** Every self-adjoint idempotent $e \neq 0$ is the sum of a finite number of irreducible self-adjoint idempotents in $Ae$. Every left ideal $l \neq (0)$ contains an irreducible self-adjoint idempotent.
**Definition 10.17.** A left ideal \( I \) is *minimal* if \( I \neq (0) \) and if there exists no nonzero left ideal \( I' \neq I \) such that \( I' \subseteq I \). A similar definition applies to minimal right ideals.

Here is the first significant result which gives the structure of minimal left ideals in terms of irreducible self-adjoint idempotents.

**Theorem 10.20.** A left ideal \( I \) in \( A \) is minimal if and only if it is of the form \( I = Ae \), where \( e \neq 0 \) and \( e \) is an irreducible self-adjoint idempotent.

Theorem 10.20 and Proposition 10.17 imply the following second significant result which shows that the minimal left ideals are the building blocks of left ideals.

**Theorem 10.21.** Every left ideal in \( A \) contains a minimal left ideal. Every minimal left ideal is closed.

The next two results are technical lemmas that are needed to prove Theorem 10.24. They build special kinds of orthogonal systems from self-adjoint idempotents.

**Proposition 10.22.** The following properties hold.

1. If \( e \) and \( e' \) are two orthogonal self-adjoint idempotents, then the left ideals \( Ae \) and \( Ae' \) are orthogonal.

2. Let \( (e_i)_{1 \leq i \leq n} \) be a finite family of pairwise orthogonal, self-adjoint idempotents. Then, for every \( x \in A \), the element \( x - \sum_{i=1}^{n} xe_i \) is orthogonal to each \( Ae_j \) \((1 \leq j \leq n)\).

**Proposition 10.23.** For every \( x \in A \) there exists a finite or countably infinite sequence \( (e_n) \) of pairwise orthogonal irreducible self-adjoint idempotents belonging to the closure \( I \) of the ideal \( Ax \), such that \( x = \sum_n xe_n \) (this series being convergent in \( A \)), and \( \|x\|^2 = \sum_n \|xe_n\|^2 \).

Form now, on, we assume that the complete Hilbert algebra \( A \) is separable. We have the following important result.

**Theorem 10.24.** Suppose \( A \) is separable. Then every closed left ideal \( b \) is the Hilbert sum of a finite or countably infinite sequence of minimal left ideals \( I_n = Ae_n \), where \( e_n \) is an irreducible self-adjoint idempotent. For every \( x \in b \), we have \( x = \sum_n xe_n \), and for all \( x, y \in b \) we have \( \langle x, y \rangle = \sum_n \langle xe_n, ye_n \rangle \).

**Proof.** We follow Dieudonné’s proof from [22] (Chapter XV, Section 8, Theorem 15.8.11). By Proposition 10.22 and by the properties of Hilbert sums (see Proposition 10.4), the vector \( xe_n \) is the orthogonal projection of \( x \) onto \( Ae_n \), so the second and the third assertions follow.

To prove the first assertion, let \( (x_n)_{n \geq 1} \) be a dense sequence in \( b \), which exists since \( A \) is separable. We define inductively, for each \( n \), a finite or countably infinite sequence \( (e_{n,i})_{i \in I_n} \) of irreducible self-adjoint idempotents as follows. For \( n = 1 \), by Proposition 10.23 applied to \( x_1 \in b \), since \( Ax_1 \subseteq b \), there is a finite or countably infinite sequence \( (e_{1,i})_{i \in I_1} \) of pairwise orthogonal irreducible self-adjoint idempotents belonging to \( b \) such that \( x_1 = \sum_{i \in I_1} x_1 e_{1,i} \).
CHAPTER 10. HILBERT ALGEBRAS

Suppose that the $e_{m,i}$ have been defined for all $m \leq n$ in such a way that they are pairwise orthogonal and belong to $b$, and are such that the $x_m$ with $m \leq n$ belong to the closure $a_n \subseteq b$ of the left ideal which is the sum of the ideals $Ae_{m,i}$ for all $m \leq n$ and all $i \in I_m$,

$$a_n = \bigoplus_{m \leq n, i \in I_m} Ae_{m,i}.$$  

Let $x'_{n+1}$ be the orthogonal projection of $x_{n+1}$ onto $a_n^\perp \cap b$. Using Proposition 10.23 applied to $x'_{n+1} \in a_n^\perp \cap b$, since $Ax'_{n+1} \subseteq a_n^\perp \cap b$, there is a finite or countably infinite sequence $(e_{n+1,i})_{i \in I_{n+1}}$ of pairwise orthogonal irreducible self-adjoint idempotents which belong to $a_n^\perp \cap b$ and are such that $x'_{n+1} = \sum_{i \in I_{n+1}} x'_{n+1} e_{n+1,i}$. Using Proposition 10.22 and the properties of Hilbert sums, it is easy to check that the double family $(e_{n,i})$ arranged in a single sequence has the desired properties. \hfill \Box

Theorem 10.24 applies in particular when $b = A$. In this case, we get a decomposition of $A$ as a Hilbert sum of minimal left ideals. However, in general, there are infinitely many such decompositions. More precisely we have the following result.

**Proposition 10.25.** Suppose $A$ is separable, and let $l$ be a minimal left ideal of $A$. Then there exists a decomposition of $A$ as a Hilbert sum of minimal left ideals $l_n$ such that $l_1 = l$.

The following technical result is needed in the proof of the main theorem (Theorem 10.27).

**Theorem 10.26.** Let $e$ and $e'$ be two irreducible self-adjoint idempotents, and let $l = Ae$, and $l' = Ae'$ be the corresponding minimal left ideals. The following properties hold.

1. Every homomorphism of the $A$-module $l$ into the $A$-module $l'$ is a map $f_a$ of the form $f_a(x) = xa$ ($x \in l$), for some $a \in eAe' = eA\cap Ae'$; it is either zero or bijective, and the map $a \mapsto f_a$ is an isomorphism of the complex vector space $eAe'$ onto $\text{Hom}_A(l,l')$ such that $f_{ab} = f_a \circ f_b$.

2. The $\mathbb{C}$-algebra $eAe$, isomorphic to $\text{End}_A(l) = \text{Hom}_A(l,l)$ (the space of endomorphisms of $l$), is a field, equal to $\mathbb{C}e$ (and therefore isomorphic to $\mathbb{C}$).

3. If $l$ and $l'$ are not isomorphic as $A$-modules, then $e$ an $e'$ (and consequently $l$ and $l'$) are orthogonal, and $ll' = l'l = (0)$. If $l$ and $l'$ are isomorphic as $A$-modules, then $eAe'$ is a complex vector space of dimension 1, and $ll' = l'$.

4. If $x \in A$, then $lx$ is a left ideal which is either (0) or isomorphic (as $A$-module) to $l$.

**Proof.** We follow Dieudonné's proof from [22] (Chapter XV, Section 8, Theorem 15.8.12). (1) Let $g : l \to l'$ be an $A$-module homomorphism, and let $a = g(e)$. For every $x \in l = Ae$, since $e^2 = e$, we have

$$g(x) = g(xe) = xg(e) = xa,$$

...
so \( g = f_a \). Since \( a \in g(1) \subseteq \mathfrak{l}' \), \( \mathfrak{l}' = Ae' \) and \((e')^2 = e'\), we have \( a = ae' \). On the other hand,

\[ a = g(e) = g(e^2) = eg(e) = ea, \]

so \( a = eae' \in eAe' \).

By definition of \( eAe' \), we have \( eAe' \subseteq eA \cap Ae' \). Conversely, if \( y \in eA \cap Ae' \), as \( e^2 = e \) and \((e')^2 = e'\), we have \( y = ye' \) and \( y = ey \), so that \( y = ye' \in eAe' \). Therefore, \( eAe' = eA \cap Ae' \).

The image \( g(1) \) of \( g \) is a left ideal contained in \( \mathfrak{l}' \), and since \( \mathfrak{l}' \) is a minimal left ideal, either \( g(1) = (0) \) or \( g(1) = \mathfrak{l}' \). Likewise, the kernel \( \text{Ker} g \) of \( g \) is a left ideal contained in \( \mathfrak{l} \), and since \( \mathfrak{l} \) is a minimal left ideal, either \( \text{Ker} g = (0) \) or \( \text{Ker} g = \mathfrak{l} \). If \( \text{Ker} g = \mathfrak{l} \), then \( g(1) = (0) \) and \( g \) is the zero map. If \( \text{Ker} g = (0) \), then \( g(1) \neq (0) \), so we must have \( g(1) = \mathfrak{l}' \), and \( g \) is bijective.

If \( g = f_a = 0 \), then \( f_a(e) = ea = 0 \); but \( a \in eAe' \), so \( ea = a \), and consequently \( a = 0 \), which shows that the map \( a \mapsto f_a \) is an isomorphism.

(2) By Proposition 10.16(3), the \( \mathbb{C} \)-algebra \( eAe \) is a closed subalgebra of \( A \). In (1), we saw that every element of \( \text{End}_A(\mathfrak{l}) \) is either zero or invertible, so \( \text{End}_A(\mathfrak{l}) \) is a (possibly noncomutative) field, and since \( eAe \) isomorphic to \( \text{End}_A(\mathfrak{l}) \), it is also a field. Clearly \( e \) is a unit in \( eAe \), and since \( A \) is a complete Hilbert algebra, it is a Banach algebra, and since \( eAe \) is closed in \( A \), it is also a Banach algebra. By the Gelfand–Mazur theorem (Theorem 8.11), \( eAe = \mathbb{C}e \cong \mathbb{C} \).

(3) If \( \mathfrak{l} \) and \( \mathfrak{l}' \) are not isomorphic, by (1) we have \( eAe' = (0) \), and in particular, \( ee' = 0 \). By Proposition 10.17, \( e \) and \( e' \) are orthogonal. By Proposition 10.22, the left ideals \( \mathfrak{l} \) and \( \mathfrak{l}' \) are orthogonal, and \( \mathfrak{l}' = Ae' Ae' = A(0) = (0) \). Similarly \( e' Ae = (0) \), so \( \mathfrak{l}' = (0) \). If \( \mathfrak{l} \) and \( \mathfrak{l}' \) are isomorphic, and if \( g: \mathfrak{l} \to \mathfrak{l}' \) is an isomorphism from \( \mathfrak{l} \) to \( \mathfrak{l}' \), then every homomorphism \( h: \mathfrak{l} \to \mathfrak{l}' \) is of the form \( h = g \circ u \), where \( u \in \text{End}_A(\mathfrak{l}) \), which means that \( \text{End}_A(\mathfrak{l}) \) and \( \text{Hom}_A(\mathfrak{l}, \mathfrak{l}') \) are isomorphic. By (2), the space \( \text{End}_A(\mathfrak{l}) \) is one-dimensional, so \( \text{Hom}_A(\mathfrak{l}, \mathfrak{l}') \) is also one-dimensional, and by (1), the space \( \text{Hom}_A(\mathfrak{l}, \mathfrak{l}') \) is isomorphic to \( eAe' \), so we deduce that \( eAe' \) is complex vector space of dimension 1. The ideal \( \mathfrak{l}' \) is obviously contained in \( \mathfrak{l}' \), and contains \( eAe' \neq (0) \). Since \( \mathfrak{l}' \) is a minimal left ideal, we must have \( \mathfrak{l}' = \mathfrak{l}' \).

(4) Since \( \mathfrak{l}x \) is the image of \( \mathfrak{l} \) under the homomorphism \( \varphi: \mathfrak{l} \to A \) given by \( \varphi(y) = yx \), it is a left ideal isomorphic to \( \mathfrak{l}/\text{Ker} \varphi \). But since \( \text{Ker} \varphi \) is a left ideal contained in the minimal left ideal \( \mathfrak{l} \), either \( \text{Ker} \varphi = (0) \) or \( \text{Ker} \varphi = \mathfrak{l} \), which means that \( \varphi \) is either injective or zero, and so \( \mathfrak{l}x \) is a left ideal which is either \((0)\) or isomorphic (as \( A \)-module) to \( \mathfrak{l} \).

Finally, here is the main theorem of this section.

**Theorem 10.27.** (Master decomposition theorem) Let \( A \) be a complete, separable, Hilbert algebra. The following properties hold.

(1) There exists a finite or countably infinite sequence \((\mathfrak{l}_k)_{k \in J}\) of minimal left ideals, no pair of which are isomorphic, such that every minimal left ideal of \( A \) is isomorphic (as an \( A \)-module) to some \( \mathfrak{l}_k \).
(2) For each index \( k \in J \), the closure of the sum of all the minimal left ideals of \( A \) which are isomorphic to \( l_k \) is a self-adjoint two-sided ideal \( a_k \). Every minimal left ideal of the Hilbert algebra \( a_k \) is a minimal left ideal of \( A \), isomorphic to \( l_k \), and the algebra \( a_k \) contains no closed two-sided ideals other than \((0)\) and \( a_k \).

(3) Each of the algebras \( a_k \) is a Hilbert sum

\[
a_k = \bigoplus_{j \in I_k} l'_j,
\]

with \( I_k \) finite or countably infinite, and where each \( l'_j \) is a minimal left ideal isomorphic to \( l_k \). The algebra \( A \) is the Hilbert sum

\[
A = \bigoplus_{k \in J} a_k,
\]

and \( a_h a_k = (0) \) for all \( h \neq k \).

Proof. We reproduce Dieudonné’s proof from [22] (Chapter XV, Section 8, Theorem 15.8.13). By Theorem 10.24, we obtain \( A \) as the Hilbert sum

\[
A = \bigoplus_{n \in L} l'_n
\]

of minimal left ideals \( l'_n \) of \( A \), where \( L \) is finite or countably infinite. We define inductively the index set \( J \) and the sequence \((l_k)_{k \in J}\) of minimal left ideals \( l_k \), no pair of which are isomorphic, as follows. Start with \( l_1 = l'_1 \). Having defined \( l_1, \ldots, l_k \), let \( l_{k+1} \) be the equal to \( l'_m \), where \( m \) is the smallest integer such that \( l'_m \) is not isomorphic to any of the ideals \( l_1, \ldots, l_k \). If all the \( l'_n \) are isomorphic to one of the ideals \( l_1, \ldots, l_k \), then stop. Let \( J \) be the finite or countably infinite sequence of indices \( k \) so obtained, and for every \( k \in J \), let \( I_k \) be the sequence of integers \( n \) such that \( l'_n \) is isomorphic to \( l_k \). If \( J \) is infinite, then each \( I_k \) is finite. Otherwise, \( I_1, \ldots, I_{k-1} \) are finite, and \( I_k \) is finite or countably infinite. Define \( a_k \) as the Hilbert sum

\[
a_k = \bigoplus_{j \in I_k} l'_j.
\]

By construction, it is clear that each \( a_k \) is a left ideal and that \( H \) is the Hilbert sum

\[
A = \bigoplus_{k \in J} a_k.
\]

Let \( l \) be any minimal left ideal in \( A \). Then \( l \) must be isomorphic to one of the \( l_k \), for otherwise by Theorem 10.26(3), it would be orthogonal to all of the \( l'_n \), and hence orthogonal to \( A \) itself, a contradiction. The same argument shows that \( l \) is orthogonal to all the \( a_h \) with \( h \neq k \). Hence, since \( a_k \) is the orthogonal complement of the Hilbert sum

\[
\bigoplus_{h \in J - \{k\}} a_h,
\]
we must have \( l \subseteq a_k \). This implies that \( a_k \) is the closure of the sum of all the minimal left ideals of \( A \) which are isomorphic to \( l_k \), and therefore \( a_k \) is independent of the decomposition of \( A \) as the Hilbert sum of the \( l'_n \) from which we started. Moreover, for every \( x \in A \), and every \( n \in I_k \), by Theorem 10.26(4), we know that \( l'_n x \) is a left ideal which is either \((0)\) or isomorphic to \( l'_n \), hence contained in \( a_k \). This proves that \( a_k \) is a two-sided ideal. If \( l'_n = A e'_n \), where \( e'_n \) is an irreducible self-adjoint idempotent, then \((l'_n)^* = e'_n A \), hence \( a_k^* = a_k \).

Let \( l'' \) be a minimal left ideal of the Hilbert algebra \( a_k \). By Theorem 10.20, we have \( l'' = a_k e'' \), where \( e'' \) is a self-adjoint idempotent, and \( e'_n e'' \) cannot vanish for all \( n \in I_k \), otherwise \( l'' \) would be orthogonal to all of the \( l'_n \) with \( n \in I_k \), and therefore to the closure of their sum, namely to \( a_k \), which is impossible because \( l'' \neq (0) \). Hence there exists at least one index \( n \in I_k \) such that \( l'_n l'' \neq (0) \); since \( l'_n l'' \) is a left ideal in \( a_k \), we must have \( l'_n l'' = l'' \). By Theorem 10.26(3), \( l'' \) is a minimal left ideal of \( A \) necessarily isomorphic to \( l'_n \), and therefore to \( l_k \).

If \( b \) is a nonzero two-sided ideal of the algebra \( a_k \), by Theorem 10.21, it contains at least one minimal left ideal \( l'' \) of this algebra, hence also contains all the \( l'' l'_n \) \((n \in I_k)\). But by Theorem 10.26(3), we have \( l'' l'_n = l'_n \), and therefore \( b \) contains the sum of all the \( l'_n \) \((n \in I_k)\). If \( b \) is closed, it follows that \( b = a_k \). Finally, \( a_h a_k \subseteq a_h \cap a_k = (0) \) if \( h \neq k \), because \( a_h \) and \( a_k \) are two-sided ideals.

The proof of the theorem shows that if \( J \) is infinite, then all the index sets \( I_k \) are finite.

Roughly speaking, Theorem 10.27 says that if \( A \) is a complete separable Hilbert algebra, then there is an irredundant list \((l_k)_{k \in J}\) of the minimal left ideals of \( A \), and \( A \) is the Hilbert sum of two-sided ideals \( a_k \), where each \( a_k \) is the Hilbert sum obtained by picking a certain number of copies of the minimal left ideal \( l_k \) of \( A \).

The ideals \( a_k \) have properties analogous to those of simple algebras.

**Definition 10.18.** A complete Hilbert algebra \( A \) is topologically simple if it contains no closed two-sided ideal other than \((0)\) and \( A \).

Theorem 10.27 shows that the complete Hilbert algebras \( a_k \) are topologically simple. Theorem 10.27 also shows that the study of complete separable Hilbert algebras reduces to the study of the topologically simple ones. We have the following theorem that gives us a sharper image of the structure of topologically simple complete separable Hilbert algebras.

**Theorem 10.28.** (Structure of a topologically simple Hilbert algebra) Let \( A \) be topologically simple, complete, separable, Hilbert algebra. For any minimal left ideal \( l \) of \( A \), the representation \( U_l : A \to \mathcal{L}(l) \) of \( A \) in the Hilbert space \( l \) is faithful.

If \( A \) is infinite-dimensional, then so is \( l \). The image of \( A \) under \( U_l \) is the algebra \( \mathcal{L}_2(l) \) of Hilbert–Schmidt operators on \( l \), and there exists a constant \( c > 0 \) such that

\[
  c(x, y)_A = \langle U_l(x), U_l(y) \rangle_{HS} \quad \text{for all } x, y \in A.
\]  

(*)
The inner product on the right-hand side is the inner product defined in Example 10.1.

If $A$ is finite-dimensional, then the image of $A$ under $U_l$ is the algebra $\text{End}_\mathbb{C}(l)$ of all endomorphisms of the vector space $l$, and $(\ast)$ remains valid (the inner product on the right-hand side is also the inner product defined in Example 10.1). In fact,

$$A = \bigoplus_{j=1}^n l_j, \quad l_j = Ae_j = \bigoplus_{i=1}^n e_i Ae_j,$$

where $e_1, \ldots, e_n$ are irreducible self-adjoint idempotents in $A$, the $l_j = Ae_j$ are isomorphic minimal left ideals, each space $e_i Ae_j$ is one-dimensional, and $e_1 + \cdots + e_n$ is the unit of $A$.

Proof. We follow Dieudonné’s proof from [22] (Chapter XV, Section 8, Theorem 15.8.14). By Proposition 10.25 and by Theorem 10.27, we may assume that $A$ is the Hilbert sum of a finite or countably infinite sequence of minimal left ideals $l_n = Ae_n$, where $l = l_1$ and all the $l_n$ are isomorphic. For any $x \neq 0$, we have $x \in Ae_n$ for some $n$, and by Theorem 10.26(2), the space $e_n Ae_n$ is one-dimensional, which implies that $Ax \neq (0)$. This also follows from Proposition 10.12. Indeed, if $x \neq 0$ and if $Ax = (0)$, then for $x^* \in A$ we have $x^* x = 0$, but by Proposition 10.12 this implies that $x = 0$, a contradiction.

First we prove that the representation $U_l$ is faithful. If $U_l(x) = 0$ for some $x \neq 0$, that is, if $x l = (0)$, then we should have $(Ax) l = (0)$. Since the ideal $Ax$ is nonzero, by Proposition 10.21, the ideal $Ax$ contains a minimal left ideal $l'$, and by Theorem 10.27, the left ideal $l'$ is isomorphic to $l$; hence $l'l = (0)$, contrary to Theorem 10.26(3). Therefore the representation $U_l$ is faithful.

Next, we claim that $l$ is the Hilbert sum of the subspaces $e_n Ae_1$,

$$l = \bigoplus_n e_n Ae_1,$$

where the spaces $e_n Ae_1$ are one-dimensional.

Let $P_n = U_l(e_n)$. By Theorem 10.26(3), since the $l_i$ are isomorphic, the subspace $e_n Ae_1$ is one-dimensional, and we shall show that $P_n$ is the orthogonal projection of $l = Ae_1$ onto $e_n Ae_1$. Since $e_n$ is a self-adjoint idempotent and since an arbitrary element of $Ae_1$ is of the form $xe_1$ for some $x \in A$, and $U_l(e_n)(xe_1) = e_n xe_1$, we have

$$\langle xe_1 - e_n xe_1, e_n ye_1 \rangle = \langle e_n xe_1 - e_n^2 xe_1, ye_1 \rangle = \langle e_n xe_1 - e_n xe_1, ye_1 \rangle = 0.$$
to all the subspaces \( e_n A e_1 \), then \( P_n(x e_1) = 0 \) for all \( n \), so that \( e_n x e_1 = 0 \) for all \( n \), and thus \( y = x e_1 \) belongs to the right annihilator of \( A \). Proposition 10.12 implies that this right annihilator is equal to \((0)\), because \( y^* \in A \), so \( y^* y = 0 \), and by Proposition 10.12 we must have \( y = x e_1 = 0 \). A similar proof shows that \( I_j \) is the Hilbert sum

\[
I_j = \bigoplus_k e_k A e_j,
\]

where the spaces \( e_k A e_j \) are one-dimensional.

The above fact shows that the sequence \((I_k)\) is finite iff \( I \) (and thus each \((I_k)\)) is finite-dimensional over \( C \), or equivalently iff \( A \) is finite-dimensional. If \( A \) is finite-dimensional, then

\[
A = \bigoplus_{j=1}^n I_j, \quad I_j = A e_j = \bigoplus_{i=1}^n e_i A e_j,
\]

where \( e_1, \ldots, e_n \) are irreducible self-adjoint idempotents in \( A \), the \( I_j = A e_j \) are isomorphic minimal left ideals, and each space \( e_i A e_j \) is one-dimensional. By Proposition 10.17, since the subspaces \( A e_j \) are pairwise orthogonal, \( e_j e_j = 0 \) whenever \( i \neq j \), so \( e_1 + \cdots + e_n \) is a unit for \( e_i A e_j \), and thus for \( A \).

Let \((a_n)\) be an orthonormal basis of \( I \) such that \( a_n \in e_n A e_1 \) for each \( n \). Since the \( e_n \) are self-adjoint idempotents, we have

\[
a_n e_1 = a_n, \quad e_n a_n = a_n, \quad a_n a_n^* \in e_n A e_n, \quad a_n^* a_n \in e_1 A e_1,
\]

and by Theorem 10.26(2), we must have \( a_n a_n^* = \alpha_n e_n \) for some nonzero \( \alpha_n \in C \). Similarly, \( a_n^* a_n \in e_1 A e_1 \), and we must have \( a_n^* a_n = \alpha_n' e_1 \) for some nonzero \( \alpha_n' \in C \).

We claim that \( \alpha_n = \alpha_n' \) for all \( n \).

On the one hand,

\[
a_n a_n^* a_n a_n^* = \alpha_n e_n \alpha_n e_n = \alpha_n^2 e_n^2 = \alpha_n^2 e_n,
\]

and on the other hand, since \( a_n e_1 = a_n \), we have

\[
a_n a_n^* a_n a_n^* = a_n \alpha_n' e_1 a_n^* = \alpha_n' a_n e_1 a_n^* = \alpha_n' a_n a_n^* = \alpha_n' \alpha_n e_n.
\]

Consequently, \( \alpha_n^2 e_n = \alpha_n' \alpha_n e_n \), and since \( \alpha_n', \alpha_n \neq 0 \), we conclude that \( \alpha_n = \alpha_n' \) for all \( n \).

We also have

\[
\langle e_n, e_n \rangle = \langle e_1, e_1 \rangle, \quad \text{for all } n \geq 1.
\]

Indeed, we have

\[
1 = \langle a_n, a_n \rangle = \langle a_n, e_n a_n \rangle = \langle a_n a_n^*, e_n \rangle = \alpha_n \langle e_n, e_n \rangle,
\]

and

\[
1 = \langle a_n, a_n \rangle = \langle a_n, a_n e_1 \rangle = \langle a_n^* a_n, e_1 \rangle = \alpha_n \langle e_1, e_1 \rangle,
\]
and since $\alpha_n \neq 0$, we obtain $\langle c_n, e_n \rangle = \langle e_1, e_1 \rangle$ for all $n \geq 1$.

Let $c = 1/\langle e_1, e_1 \rangle = \alpha_n$. Then $a_n a_n^* = ce_n$, and for all $x, y \in A$, we have

$$\langle xa_n, ya_n \rangle = \langle y^* x, a_n a_n^* \rangle = \langle y^* x, ce_n \rangle = c \langle xe_n, ye_n \rangle.$$  

By Theorem 10.24, the series with general term $\langle xe_n, ye_n \rangle$ is absolutely convergent with sum $\langle x, y \rangle$, and since $\langle xa_n, ya_n \rangle = c \langle xe_n, ye_n \rangle$, if $A$ is infinite-dimensional, then $\sum_n \langle xa_n, xa_n \rangle = \sum_n \|U_1(x)(a_n)\|^2$ converges, so $U_1(x)$ is a Hilbert–Schmidt operator, and Equation (*) holds. Since $A$ is a Hilbert space, so is its image under $U_1$, and to show that this image is the whole of the Hilbert space $\mathcal{L}_2(\mathfrak{l})$ (see Example 10.1), it suffices to show that $U_1(A)$ is dense in $\mathcal{L}_2(\mathfrak{l})$.

Now for all $m, n$ with $m \neq n$, by Theorem 10.26(3), we have

$$(e_m A e_n)(e_n A e_1) = e_m (A e_n)(A e_1) = e_m A e_1,$$

and since $e_n A e_1 = C a_n$, $a_n \in e_n A e_1$, and $a_m \in e_m A e_1$, it follows that there exists $e_{mn} \in e_m A e_n$ such that $e_{mn} a_n = a_m$ (which implies that $e_{mn} = c^{-1} a_m a_n^*$, since $a_n a_n^* = ce_n$ and $e_{mn} \in e_m A e_n$), and clearly, $e_{mn} a_p = 0$ if $p \neq n$. We conclude from this that $E_{mn} = U_1(e_{mn})$ is the continuous endomorphism of the Hilbert space $\mathfrak{l}$ such that $E_{mn}(a_n) = a_m$ and $E_{mn}(a_p) = 0$ if $p \neq n$. Our assertion now follows from the fact that it is not hard to show that the finite linear combinations of the $E_{mn}$ are dense in $\mathcal{L}_2(\mathfrak{l})$.

We already took care of the case where $A$ is finite-dimensional.  

Theorem 10.28 implies that in the Hilbert sum

$$A = \bigoplus_{k \in J} a_k$$

given by Theorem 10.27, the Hilbert algebra $a_k$, which is a building block of the decomposition, is either isomorphic to the algebra $\mathcal{L}_2(\mathfrak{l}_k)$ of Hilbert–Schmidt operators on $\mathfrak{l}_k$, or to the finite-dimensional algebra $\text{End}_\mathbb{C}(\mathfrak{l}_k)$ of all endomorphisms of the vector space $\mathfrak{l}_k$. If $G$ is a metrizable compact group and $A = L^2(G)$, then every $a_k$ in the Hilbert sum for $A$ is isomorphic to the finite-dimensional algebra $\text{End}_\mathbb{C}(\mathfrak{l}_k)$.

Observe that in all cases, the spaces $U_1(\mathfrak{l}_k)$ consist of endomorphisms of the form $U(x) \circ P_n$ (where the $P_n$ are orthogonal projections), and therefore consist of endomorphisms of rank 1 of $\mathfrak{l}$.

The following result gives a sufficient condition for a Hilbert algebra satisfying the hypotheses of Theorem 10.28 to be finite. Recall that the center $Z(A)$ of an algebra $A$ is the set $Z(A) = \{ x \in A \mid xy = yx \text{ for all } y \in A \}$.

**Proposition 10.29.** Let $A$ be topologically simple, complete, separable, Hilbert algebra. If there is an element $c \neq 0$ in the center of $A$, then $A$ is finite-dimensional. In that case, the center of $A$ is $\mathbb{C}1$, where $1$ is the identity element of $A$.  

10.4. COMPLETE SEPARABLE HILBERT ALGEBRAS

Proof. If $c$ belongs to the center of $A$, then for any irreducible self-adjoint idempotent $e$ in $A$ (which exists by Theorem 10.20 and Theorem 10.21), $U(c)$ is an endomorphism of the $A$-module $I = Ae$, and by Proposition 10.26(2), it is a map of the form $x \mapsto \alpha x$ with $\alpha \in \mathbb{C}$. However, such a map cannot be a Hilbert-Schmidt operator on an infinite-dimensional space unless it is the zero map. Therefore $A$ is finite-dimensional. \hfill \square

Theorem 10.27 will be applied to the complete, separable, Hilbert algebra $L^2(G)$ ($G$ metrizable and compact), which is then a Hilbert sum of two-sided ideals $a_k$. It turns out that Proposition 10.29 applies to the Hilbert algebras $a_k$, so they are finite-dimensional. This will yield the first part of the Peter–Weyl theorem.

We conclude with a result about representations of complete Hilbert algebras that makes use of Theorem 10.27.

**Theorem 10.30.** (Master decomposition for nondegenerate representations) Let $A$ be a separable, complete, Hilbert algebra and let $a_{k \in J}$ be the topologically simple Hilbert algebras which are the Hilbert summands as in Theorem 10.27(3). For every $k \in J$, let $l_k$ be a minimal left ideal of $a_k$. Let $V : A \to L(H)$ be a nondegenerate representation of $A$ in a separable Hilbert space $H$, such that $V : A \to L(H)$ is continuous. The following properties hold.

1. The Hilbert space $H$ is the Hilbert sum of subspaces $H_k$ ($k \in J$) invariant under $V$, such that if $V_k$ is the restriction of $V$ to $H_k$, we have $V_k(s) = 0$ for all $s \in a_h$ and for all $h$ with $h \neq k$; thus $V_k$ can be considered as a representation of $a_k$ in $H_k$.

2. The representation $V_k$ is the Hilbert sum of a finite or countably infinite sequence of irreducible representations, each equivalent to the representation $U_{lk}$ of $a_k$ as in Theorem 10.28.

Proof. We follow Dieudonné’s proof from [22] (Chapter XV, Section 8, Theorem 15.8.16). Let $H_k$ be the closure of the subspace spanned by the set of vectors

$$\{V(s_k)(x) \mid s_k \in a_k, x \in H\}.$$

Since every $s \in A$ can be written as $s = \sum_k s_k$ for some $s_k \in a_k$ and since $V$ is continuous, we have

$$V(s)(x) = \sum_k V(s_k)(x),$$

which implies that $H$ is the closure of the sum of the $H_k$. Also, if $h \neq k$ and if $s_h \in a_h, s_k \in a_k$, we have

$$\langle V(s_h)(x), V(s_k)(y) \rangle = \langle V^*(s_k)V(s_h)(x), y \rangle = \langle V(s_k s_h)(x), y \rangle = \langle V(0)(x), y \rangle = \langle 0, y \rangle = 0,$$

because $a_k$ is self-adjoint and $a_k a_h = (0)$ since $h \neq k$, so $s_k^* s_h = 0$. Thus (1) holds.

Now assume that $A$ is topologically simple and finite-dimensional over $\mathbb{C}$, and thus has a unit element. The case where $A$ is infinite-dimensional can also be handled but it is a bit
more complicated; see Dieudonné’s [22] (Chapter XV, Section 8, Problem 1). By Proposition 10.6, we may assume that $A$ has a cyclic vector $x_0$. The subspace of $H$ spanned by the set of vectors $\{ V(s)(x_0) \mid s \in A \}$ is finite-dimensional, and thus closed and dense in $H$, so it is equal to $H$. As a consequence, we can argue by induction on the dimension of $H$.

Since $A$ is the sum of a finite number of minimal left ideals, there is at least one minimal ideal, say $l$, such that the subspace $E = V(l)(x_0) \subseteq H$ is nonzero. The surjection $\varphi: l \to E$ given by $\varphi(s) = V(s)(x_0)$ is then an $A$-module homomorphism, and since its kernel is a left ideal $l'$ contained in $l$ and distinct from $l$, since $l$ is minimal, we must have $l' = (0)$. Hence $\varphi: l \to E$ is a linear isomorphism and $E$ is a subspace of $H$ invariant under $V$. By Theorem 10.28, the representation $U_l$ is faithful and it is an isomorphism between $A$ and $\text{End}_C(l)$, with $l$ finite-dimensional, so if $l'$ is a proper linear subspace of $l$ invariant under $U_l$, as $U_l$ remains a faithful representation in $l'$ (because any linear map on $l'$ can be extended by zero to a linear map on $l$), $A$ would be isomorphic to $\text{End}_C(l')$ whose dimension is strictly smaller than the dimension of $\text{End}_C(0)$, a contradiction. Therefore, the representation $U_l$ is irreducible, and $\varphi: l \to E$ establishes an equivalence between $U_l$ and the restriction of $V$ to $E$. Since the orthogonal complement $E^\perp$ of $E$ in $H$ is also invariant under $V$ and has dimension strictly smaller than the dimension of $H$, we can apply the induction hypothesis to complete the proof.

Theorem 10.30 will be used to prove another part of the Peter–Weyl theorem. In fact, we will only use Part (2) of Theorem 10.30 when $a_k$ is finite-dimensional, and we gave a proof in this case.

### 10.5 The Plancherel–Godement Theorem

The main goal of this section is to state the Plancherel–Godement theorem, which will be used later in discussing harmonic analysis on the space induced by a Gelfand pair. A related theorem is the Bochner–Godement theorem. These theorems apply to a commutative Hilbert algebra (not necessarily complete) arising from the quotient of a commutative Hilbert algebra by a left ideal induced by a bitrace satisfying two additional conditions. Therefore we go back to positive Hilbert forms to describe the construction of a certain representation. The first step is the following proposition.

The idea is that if $g$ is a positive Hilbert form on $A$, it almost defines an inner product, but in general it fails to be positive definite because there may be nonzero elements $s \in A$ such that $g(s, s) = 0$. However, if we take the quotient of $A$ by the set $n = \{ s \in A \mid g(s, s) = 0 \}$, which is a left ideal because $g$ is a positive Hilbert form, then we can define an inner product on the quotient vector space $A/n$. If $g$ is a bitrace, then $A/n$ is an involutive algebra.

**Proposition 10.31.** Let $g$ be a positive Hilbert form on an involutive algebra $A$. The set
\[ n = \{ s \in A \mid g(s, s) = 0 \} \]
is a left ideal in $A$, and

$$n = \{ s \in A \mid g(s, t) = 0 \text{ for all } t \in A \}.$$  

If $\pi : A \to A/n$ is the quotient map, then there exists a hermitian inner product $\langle -, - \rangle$ on $A/n$ such that

$$\langle \pi(s), \pi(t) \rangle = g(s, t) \quad \text{for all } s, t \in A.$$  

Furthermore, if $g$ is a bitrace, then $n$ is a self-adjoint two-sided ideal, so $A/n$ is an algebra. The involution $s \mapsto s^*$ on $A$ induces an involution on $A/n$ given by $(\pi(s))^* = \pi(s^*)$, and the hermitian inner product $\langle -, - \rangle$ on $A/n$ is a bitrace on $A/n$.

Proof. By the Cauchy–Schwarz inequality

$$|g(s, t)|^2 \leq g(s, s)g(t, t),$$

we see that

$$n = \{ s \in A \mid g(s, t) = 0 \text{ for all } t \in A \}. \tag{*}$$

Since $g$ is a positive Hilbert form, we have

$$g(xy, z) = g(y, x^* z) \quad \text{for all } x, y \in A,$$

so if $s \in n$, that is, $g(s, s) = 0$, then by (*),

$$g(ts, ts) = g(s, t^* st) = 0,$$

so $n$ is a left ideal. Since $g$ is hermitian, $g(t, s) = \overline{g(s, t)}$, and we also have

$$n = \{ s \in A \mid g(t, s) = 0 \text{ for all } t \in A \}.$$  

Consequently, if $s - s' \in n$ and if $t - t' \in n$ (that is, $\pi(s) = \pi(s')$ and $\pi(t) = \pi(t')$), since

$$g(s, t - t') + g(s - s', t') = g(s, t) - g(s, t') + g(s, t') - g(s', t') = g(s, t) - g(s', t'),$$

and since $s, t' \in A, t - t', s - s' \in n$, we have $g(s, t - t') = g(s - s', t') = 0$, and thus $g(s, t) - g(s', t') = 0$, that is,

$$g(s, t) = g(s', t').$$

We can now define the function $\langle -, - \rangle$ on $A/n$ by

$$\langle \pi(s), \pi(t) \rangle = g(s, t) \quad \text{for all } s, t \in A,$$

and it is well-defined since $\pi(s) = \pi(s')$ and $\pi(t) = \pi(t')$ imply that $g(s, t) = g(s', t')$. It is immediately verified that $\langle -, - \rangle$ is a hermitian inner product on $A/n$.

If $g$ is a bitrace, then

$$g(y^*, x^*) = g(x, y) \quad \text{for all } x, y \in A,$$
so
\[ g(s^*, s^*) = g(s, s) \quad \text{for all } s \in A, \]
which shows that \( n^* = n \). If \( s \in n \) then \( s^* \in n \), and since \((st)^* = t^*s^*\) and \( n \) is a left ideal,
\[ g(st, st) = g((st)^* (st)^*) = g(t^*s^*, t^*s^*) = 0, \]
which proves that \( n \) is also a right ideal. It follows that \( A/n \) is an algebra. The verification that the map given by \((\pi(s))^* = \pi(s^*)\) is an involution on \( A/n \) is immediate because \( \pi \) is a homomorphism. For example,
\[ (\pi(st))^* = \pi((st)^*) = \pi(t^*s^*) = \pi(t^*)\pi(s^*) = (\pi(t))^*(\pi(s))^*. \]
Finally, since by definition
\[ \langle \pi(s), \pi(t) \rangle = g(s, t), \]
we get
\[ \langle (\pi(t))^*, (\pi(s))^* \rangle = \langle \pi(t^*), \pi(s^*) \rangle = g(t^*, s^*) = g(s, t) = \langle \pi(s), \pi(t) \rangle, \]
which shows that \( \langle -, - \rangle \) is a bitrace on \( A/n \).

We are now going to show that if we add one more condition to a positive Hilbert form that insures that certain linear maps on \( A/n \) are continuous, then we can define a representation of \( A \) into a Hilbert space which is the completion of \( A/n \).

However, let us first observe that if \( g \) arises from a positive linear form \( f_{x_0} \) induced by a topologically cyclic representation \( U: A \to \mathcal{L}(H) \) with cyclic vector \( x_0 \in H \), where
\[ f_{x_0}(s) = \langle U(s)(x_0), x_0 \rangle, \quad s \in A, \]
then the hermitian space \( H_0 = \{ U(s)(x_0) \mid s \in A \} \), which is dense in \( H \), is determined by \( g_{x_0} \), and in fact, the representation \( U \) is determined by \( g_{x_0} \).

Indeed, since \( g(s, t) = g_{x_0}(s, t) = f_{x_0}(t^*s) \), we have
\[ g_{x_0}(s, t) = \langle U(s)(x_0), U(t)(x_0) \rangle, \quad (*) \]
and we see that
\[ g_{x_0}(s, s) = \| U(s)(x_0) \|^2, \]

hence \( n \) is the kernel of the linear map \( h: A \to H_0 \) given by
\[ h(s) = U(s)(x_0). \]

Equation (*) shows that the quotient map \( \widehat{h}: A/n \to H_0 \) is a bijective isometry. This shows that, \( H_0 \), and therefore the Hilbert space \( H \) (since \( H_0 \) is dense in \( H \)), is determined by \( g_{x_0} \) up to isomorphism.
Since $h = \hat{h} \circ \pi$, for every $s \in A$, the map $U(s)$ is completely determined by $\hat{h}$ on $H_0$, because

$$U(s)(U(t)(x_0)) = U(st)(x_0) = h(st) = \hat{h}(\pi(st)).$$

Since $H_0$ is dense in $H$, the continuous map $U(s)$ extends uniquely to $H$. Therefore, the representation $U$ is determined by $g_{x_0}$.

We also have the following uniqueness result up to equivalence.

**Proposition 10.32.** If $U_1: A \rightarrow \mathcal{L}(H_1)$ and $U_2: A \rightarrow \mathcal{L}(H_2)$ are two topologically cyclic representations with respective cyclic vectors $x_0$ and $x'_0$, and if $f_{x_0} = f_{x'_0}$, that is, $\langle U_1(s)(x_0), x_0 \rangle = \langle U_2(s)(x'_0), x'_0 \rangle$ for all $s \in A$, then the representations $U_1$ and $U_2$ are equivalent.

**Proof.** We follow Dieudonné [22] (Chapter XV, Section 6, Theorem 15.6.7). For all $s, t \in A$ we have

$$\langle U_1(s)(x_0), U_1(t)(x_0) \rangle = \langle U_1(t^* s)(x_0), x_0 \rangle = \langle U_2(t^* s)(x'_0), x'_0 \rangle = \langle U_2(s)(x'_0), U_2(t)(x'_0) \rangle;$$

that is,

$$\langle U_1(s)(x_0), U_1(t)(x_0) \rangle = \langle U_2(s)(x'_0), U_2(t)(x'_0) \rangle, \quad \text{for all } s, t \in A. \quad (*)$$

Since the vectors $U_1(t)(x_0)$ (resp. $U_2(t)(x'_0)$) form a dense subset $H_1^0$ (resp. $H_2^0$) of $H_1$ (resp. $H_2$), we deduce that $U_1(s)(x_0) = 0$ iff $U_2(s)(x'_0) = 0$ for all $s \in A$. As a consequence, for every $z \in H_1^0$ and for all $s, s' \in A$ such that $U_1(s)(x_0) = U_1(t)(x_0) = z$, since $U_1(s-t)(x_0) = 0$ iff $U_2(s-t)(x'_0) = 0$, we have $U_2(s)(x'_0) = U_2(t)(x'_0)$, which means that the vector $U_2(s)(x'_0)$ has the same value $z' \in H_2^0$ for all $s \in A$ such that $U_1(s)(x_0) = z$, so we can define the map $T: H_1^0 \rightarrow H_2^0$ by $T(z) = z'$, or equivalently,

$$T(U_1(s)(x_0)) = U_2(s)(x'_0), \quad s \in A.$$

It is immediately verified that $T$ is a surjective linear map, and by $(*)$, it is an isometry of the hermitian space $H_1^0$ onto the hermitian space $H_2^0$. But then this isomorphism extends uniquely to an isomorphism, also denoted $T$, between the Hilbert spaces $H_1$ and $H_2$. It remains to show that $T$ induces an equivalence of the representations $U_1$ and $U_2$. Since $H_1^0$ is dense in $H_1$ and $H_2^0$ is dense in $H_2$, it suffices to prove that

$$T(U_1(s)(z)) = U_2(s)(T(z))$$

for all $z$ of the form $z = U_1(t)(x_0)$ ($t \in A$). Since

$$U_1(s)(U_1(t)(x_0)) = U_1(st)(x_0),$$
by definition of $T$, we have

$$T(U_1(s)(z)) = T(U_1(s)(U_1(t)(x_0)))$$

$$= T(U_1(st)(x_0))$$

$$= U_2(st)(x'_0)$$

$$= U_2(s)(U_2(t)(x'_0)) = U_2(s)(T(z)),$$

as claimed.

We now go back to our positive Hilbert form $g$, and we assume that it satisfies the analog of Condition (U) of Definition 10.14: for every $s \in A$, there is some $M_s \geq 0$ such that

$$g(st, st) \leq M_s g(t, t), \quad \text{for all } t \in A, \quad (U)$$

If $g$ arises from a representation $U : A \to \mathcal{L}(H)$ as

$$g(s, t) = \langle U(s)(x_0), U(t)(x_0) \rangle, \quad s, t \in A,$$

then it is easy to see that $g$ satisfies Property (U).

If the positive Hilbert form $g$ in Proposition 10.31 satisfies Condition (U), then we check immediately that the inner product $\langle -, - \rangle$ on $A/\mathfrak{n}$ given by

$$\langle \pi(s), \pi(t) \rangle = g(s, t)$$

also satisfies Property (U).

The following proposition shows how to contruct a representation of $A$ from $g$.

**Proposition 10.33.** Let $g$ be a positive Hilbert form on an involutive algebra $A$, satisfying Condition (U), let $\mathfrak{n}_g$ be the left ideal given by

$$\mathfrak{n}_g = \{ s \in A \mid g(s, s) = 0 \},$$

and suppose that the hermitian space $A/\mathfrak{n}_g$ constructed in Proposition 10.31 is separable. If so, let $H_g$ be the Hilbert space which is the completion of $A/\mathfrak{n}_g$, so that $A/\mathfrak{n}_g$ can be identified with a dense subspace $H_0$ of the separable Hilbert space $H_g$. If $\pi_g : A \to A/\mathfrak{n}_g$ denotes the quotient map, for every $s \in A$, the linear map $U_g(s) : A/\mathfrak{n}_g \to A/\mathfrak{n}_g$ given by

$$U_g(s)(\pi_g(t)) = \pi_g(s)\pi_g(t) = \pi_g(st)$$

(left multiplication by $\pi_g(s)$) extends to a continuous linear map $U_g(s) : H_g \to H_g$, and the map $s \mapsto U_g(s)$ is a representation of $A$ in $H_g$. If $g$ is a bitrace, then $A/\mathfrak{n}_g$ is an involutive algebra, and the inner product $\langle -, - \rangle_g$ on $A/\mathfrak{n}_g$ given by

$$\langle \pi_g(s), \pi_g(t) \rangle_g = g(s, t)$$

satisfies Property (U).
10.5. THE PLANCHEREL–GODEMENT THEOREM

Proof. If \( \pi_g(t) = \pi_g(t') \), then \( \pi_g(st) = \pi_g(st') \) because \( n_g \) is a left ideal. Hence for any fixed \( s \in A \), the endomorphism of \( H_0 \) given by \( \pi_g(t) \mapsto \pi_g(st) \) is well-defined. The definition of the inner product \( \langle \cdot, \cdot \rangle_g \) on \( A/n_g \) and Condition (U) ensure that this map is continuous. Since \( H_0 \) is dense in \( H_g \) and the map \( \pi_g(t) \mapsto U_g(s)(\pi_g(t)) = \pi_g(s)\pi_g(t) = \pi_g(st) \) is continuous, it extends to a continuous map \( U_g(s) : H_g \to H_g \).

Since \( \pi_g((ss')t) = \pi_g(s(st')) \), we have \( U_g(ss') = U_g(s) \circ U_g(s') \). Since \( g \) is a positive Hilbert form, we have

\[
g(xy, z) = g(y, x^*z) \quad \text{for all } x, y, z \in A
\]

and since \( g \) is hermitian, we have

\[
\langle U_g(s^*)(\pi_g(t)), \pi_g(t') \rangle_g = g(s^*t, t') = g(t, st') = \frac{\langle U_g(s)(\pi_g(t')), \pi_g(t) \rangle_g}{\langle \pi_g(t), U_g(s)(\pi_g(t')) \rangle_g} = \langle \pi_g(t), U_g(s)(\pi_g(t')) \rangle_g,
\]

which shows that \( U_g(s^*) = (U_g(s))^* \). If \( A \) has a unit element \( e \) then \( U_g(e) = \text{id} \). Therefore \( U_g \) is a representation of \( A \) in \( H_g \).

In general, the representation \( U_g : A \to \mathcal{L}(H_g) \) given by Proposition 10.33 may be degenerate. It is nondegenerate if and only if the following condition holds:

\[
\text{the subspace spanned by the set } \{ \pi_g(st) \mid s, t \in A \} \text{ is dense in } A/n_g. \quad (N)
\]

Observe that the above condition is the generalization of Condition (N) of Definition 10.14 to \( A/n_g \). In Definition 10.14, the bitrace \( g \) is an inner product so \( n_g = (0) \). If \( A \) has a unit element, Condition (N) holds trivially.

If the positive Hilbert form \( g \) in Proposition 10.31 satisfies Condition (N), then we check immediately that the inner product \( \langle \cdot, \cdot \rangle \) on \( A/n \) given by

\[
\langle \pi(s), \pi(t) \rangle = g(s, t)
\]

also satisfies Property (N).

It can be shown that if \( A \) is a unital Banach algebra with involution, then of course Condition (N) is automatically satisfied, but Condition (U) also holds. This follows from the following result proven in Dieudonné [22] (Chapter XV, Section 7, Theorem 15.6.11).

**Proposition 10.34.** Let \( A \) be a unital Banach algebra with involution and unit element \( e \neq 0 \). For any positive linear form \( f \), the following properties hold.

1. \( f \) is continuous and \( \|f\| = f(e) \).
(2) \(|f(y^*xy)| \leq \|x\| f(y^*y)\), for all \(x, y \in A\).

Recall that since \(A\) is unital, every positive Hilbert form \(g\) arises from the positive linear form \(f\) given by \(f(s) = g(s, e)\). Then for every \(s \in A\),

\[
g(st, st) = f((st)^*st) = f(t^*s^*st) \leq \|s^*s\| f(t^*t) = \|s^*s\| g(t, t),
\]

which is Condition (U) with \(M_s = \|s^*s\|\).

The proof of Proposition 10.34 makes use of the following result of independent interest also proven in Dieudonné [22] (Chapter XV, Section 7, Theorem 15.6.11.1).

**Proposition 10.35.** Let \(A\) be a unital Banach algebra with involution and unit element \(e \neq 0\). If \(x \in A\) is self-adjoint and \(\|x\| < 1\), then there exists a self-adjoint element \(y \in A\) such that \(y^2 = e + x\).

After these preliminaries, we assume that \(A\) is a commutative (but not necessarily complete) involutive algebra equipped with a bitrace \(g\) satisfying Conditions (U) and (N). We also assume that the hermitian space \(A/\mathfrak{n}_g\) is separable, as in Proposition 10.33. Then, by Proposition 10.33, the bitrace \(g\) induces a representation \(U_g : A \to \mathcal{L}(H_g)\), where the separable Hilbert space \(H_g\) is the completion of \(A/\mathfrak{n}_g\), so that \(A/\mathfrak{n}_g\) is dense in the Hilbert space \(H_g\). Since Property (N) holds, the representation \(U_g\) is nondegenerate. The image of \(A\) under \(U_g\) is a commutative subalgebra of the \(C^*\)-algebra \(\mathcal{L}(H_g)\). Let \(\mathcal{A}_g\) be the closure of \(U_g(A)\) in \(\mathcal{L}(H_g)\), so that \(\mathcal{A}_g\) is a commutative \(C^*\)-algebra (and thus, consists of normal operators). In general, \(\mathcal{A}_g\) is not separable, but we assume it is separable.

Here is a fairly general situation where a positive Hilbert form satisfies conditions (U) and (N).

**Proposition 10.36.** Let \(A\) be a unital separable Banach algebra with involution and unit element \(e \neq 0\). For any positive linear form \(f\) on \(A\), let \(g\) be the corresponding positive Hilbert form given by \(g(x, y) = f(y^*x)\) for all \(x, y \in A\). Then \(g\) satisfies Property (U), and the hermitian space \(A/\mathfrak{n}_g\) and the \(C^*\)-algebra \(\mathcal{A}_g \subseteq \mathcal{L}(H_g)\) are separable.

**Proof.** By Proposition 10.34, the positive Hilbert form \(g\) induced by the positive linear form \(f\) satisfies Property (U). Again, by Proposition 10.34, since \(f\) is continuous we get

\[
\|\pi_g(x)\|^2 = f(x^*x) \leq \|x\| \|x^*x\| \leq \|f\| \|x\|^2,
\]

which shows that the image under \(\pi_g\) of a countable dense set in \(A\) is dense in \(A/\mathfrak{n}_g\), so \(A/\mathfrak{n}_g\) is separable. Since by Proposition 10.7, the map \(U_g\) is continuous, it sends a countable dense set into a countable dense set in \(\mathcal{A}_g\), so \(\mathcal{A}_g\) is separable. \(\square\)

A particular example of a trace \(f\) on \(A\), which gives rise to a bitrace \(g\) (such that \(g(x, y) = f(y^*x)\)) is provided by the hermitian characters of \(A\). These are the characters \(\chi \in \mathcal{X}(A)\) such that

\[
\chi(x^*) = \overline{\chi(x)} \quad \text{for all} \ x \in A.
\]
10.5. THE PLANCHEREL–GODEMENT THEOREM

We have $\chi(x^*x) = \chi(x^*)\chi(x) = |\chi(x)|^2$, and

$$g(x, y) = \chi(y^*x) = \chi(y^*)\chi(x) = \overline{\chi(y)}\chi(x).$$

Condition (U) holds because

$$g(st, st) = \overline{\chi(st)}\chi(st) = \overline{\chi(s)}\overline{\chi(t)}\chi(s)\chi(t) = |\chi(s)|^2|\chi(t)|^2 = |\chi(s)|^2g(t, t).$$

The ideal $n_g$ is the kernel of $\chi$, so it is a hyperplane in $A$, thus $A/n_g$ is isomorphic to $\mathbb{C}$, and Condition (N) follows immediately from the fact that $\chi(x) \neq 0$ implies that $\chi(x^2) = (\chi(x))^2 \neq 0$. It can be shown that the corresponding representation is irreducible.

Let $H(A)$ denote the set of hermitian characters in $X(A)$. The set $H(A)$ is a subset of the product space $\mathbb{C}^A$ and is closed under the product topology. We give $H(A)$ the topology induced by the product topology (the topology of pointwise convergence). When $A$ is a unital commutative Banach algebra with involution, the space $H(A)$ is a compact subspace of $X(A)$, since $X(A)$ is compact, by Theorem 8.16. If $A$ is also separable, then it can be shown that $X(A)$ is metrizable; see Dieudonné [22] (Chapter XV, Section 3, Theorem 15.3.2). In general, $H(A) \neq X(A)$, but if $A$ is a unital $C^*$-algebra, then $H(A) = X(A)$, by Proposition 8.30.

The following theorem shows that the bitraces discussed in this section all arise from a positive measure by a process of integration, from hermitian characters.

**Theorem 10.37.** (Plancherel-Godement theorem) Let $A$ be a commutative involutive algebra, and let $g$ be a bitrace on $A$ satisfying Conditions (U) and (N), and such that the hermitian space $A/n_g$ and the $C^*$-algebra $A_g \subseteq \mathcal{L}(H_g)$ are separable.

(I) We can define canonically: (1) a subspace $S_g$ of $H(A)$ whose closure in $\mathbb{C}^A$ is either $S_g$ or $S_g \cup \{0\}$ and is metrizable and compact (so that $S_g$ is locally compact, metrizable, and separable); (2) a (positive) Radon measure $m_g$ on $S_g$, with the following properties:

(i) For each $x \in A$, define the function $\widehat{x}: S_g \to \mathbb{C}$ by $\widehat{x}(\chi) = \chi(x)$. Then $\widehat{x} \in \mathcal{L}^2_{m_g}(S_g)$, and we have

$$g(x, y) = \int_{S_g} \chi(xy^*) \, d m_g(\chi) = \int_{S_g} \widehat{x}(\chi) \overline{\widehat{y}(\chi)} \, d m_g(\chi) \quad \text{for all } x, y \in A. \quad (\dagger)$$

(ii) As $x$ runs through $A$, the set of functions $\widehat{x}$ is contained in $\mathcal{C}_0(S_g; \mathbb{C})$, and is dense in this Banach space.

(iii) The map $x \mapsto \widehat{x}$ factors as $T_0 \circ \pi_g$, where the map $T_0: A/n_g \to \mathcal{C}_0(S_g; \mathbb{C})$ extends to an isomorphism $T: H_g \to \mathcal{L}^2_{m_g}(S_g)$, such that for all $x \in A$, we have $U_g(x) = T^{-1}M(\widehat{x})T$, where $M(\widehat{x})$ is multiplication by the class of $\widehat{x}$ in $\mathcal{L}^2_{m_g}(S_g)$, where $U_g$ is the representation $U_g: A \to \mathcal{L}(H_g)$. 

(iv) We have
\[ \| \lambda \text{id}_{H_0} + U_g(x) \| = \| \lambda \text{id}_S + \hat{x} \| \]
for all \( \lambda \in \mathbb{C} \) and all \( x \in A \).

(II) Conversely, let \( S \) be a subspace of \( \mathcal{H}(A) \) such that \( S \cup \{0\} \) is compact and metrizable, and let \( m \) be a positive measure on \( S \), such that for all \( x \in A \), the function \( \hat{x}: S \to \mathbb{C} \) given by \( \hat{x}(\chi) = \chi(x) \) belongs to \( \mathcal{L}^2_m(S) \cap C_0(S; \mathbb{C}) \).

Then the map \( g' \) given by
\[ g'(x, y) = \int_S \hat{x}(\chi) \overline{y}(\chi) \, dm(\chi) \quad \text{for all } x, y \in A \]
is a bitrace on \( A \) satisfying Conditions (U) and (N), such that \( A/\mathfrak{n}_{g'} \) and \( \mathcal{A}_{g'} \) are separable, and we have \( S_{g'} = S \) and \( m_{g'} = m \).

Theorem 10.37 is proven in Dieudonné [22] (Chapter XV, Section 9, Theorem 15.9.2). The proof is long and very technical. Among other results, it uses the Gelfand–Naimark theorem (Theorem 8.33).

We should mention that Dieudonné uses a theory of integration in which positive Radon functionals are used instead of Borel measures. However, the version in Dieudonné also states that the support of the Radon functional is the whole of \( S_g \), and by Proposition 15.45, since \( S_g \) is locally compact, metrizable and separable, it is \( \sigma \)-compact, so there is no problem in obtaining the theorem for Radon measures measures by using Radon–Riesz II (Theorem 5.10).

If the bitrace \( g \) arises from a positive linear form \( f \), which is a trace since \( A \) is commutative, the formula (†) leads us to ask whether we also have
\[ f(x) = \int_{S_g} \hat{x}(\chi) \, dm_g(\chi). \]

The Bochner–Godement theorem provides a partial answer to this question.

**Theorem 10.38.** (Bochner–Godement theorem) Let \( A \) be a commutative involutive algebra \( A \).

(1) Let \( f \) be a positive linear form such that the bitrace \( g \) given by \( g(x, y) = f(y^*x) \) satisfies the hypotheses of the Plancherel–Godement theorem (Theorem 10.37). Then if the formula
\[ f(x) = \int_{S_g} \hat{x}(\chi) \, dm_g(\chi) \quad \text{(BG)} \]
holds and if the (positive) Radon measure \( m_g \) is bounded, then \( f \) satisfies the following condition:
\[ \text{There is some } M > 0 \text{ such that } |f(x)|^2 \leq M f(x^*x) \text{ for all } x \in A. \quad \text{(B)} \]
(2) Conversely, let $f$ be positive linear form on $A$ which satisfies Condition (B), and suppose that the induced bitrace $g$ given by $g(x, y) = f(y^* x)$ satisfies Condition (U), and is such that the hermitian space $A/n_g$ and the $C^*$-algebra $A_g$ are separable. Then $g$ also satisfies Condition (N), the (positive) Radon measure $m_g$ is bounded, and formula (BG) holds.

Theorem 10.38 is proven in Dieudonné [22] (Chapter XV, Section 9, Theorem 15.9.4). The proof is shorter than the proof of the Plancherel–Godement theorem, and also uses the Gelfand–Naimark theorem.

If $A$ is unital, then by Proposition 10.9(3), Condition (B) is satisfied with $M = f(e)$. Also, by Proposition 10.36, if $A$ is a unital separable Banach algebra with involution, then the hermitian space $A/n_g$ and the $C^*$-algebra $A_g$ are separable. Therefore, if $A$ is commutative unital separable Banach algebra with involution, then the Bochner–Godement theorem Part 2 applies to any positive linear form on $A$.

Here is another situation in which the Bochner–Godement theorem Part 2 applies. Let $U: A \to \mathcal{L}(H)$ be a representation of a commutative involutive algebra $A$ into a separable Hilbert space $H$. For any $x_0 \in H$, let $f_{x_0}$ be the positive linear form given by

$$f_{x_0}(s) = \langle U(s)(x_0), x_0 \rangle.$$

We claim that $f_{x_0}$ satisfies Condition (B). Indeed, by Cauchy–Schwarz, we have

$$|f_{x_0}(s)|^2 \leq \|U(s)(x_0)\|^2 \|x_0\|^2 = \|x_0\|^2 f_{x_0}(s^* s).$$

We know that the bitrace $g$ induced by $f_{x_0}$ (given by $g(x, y) = f_{x_0}(y^* x)$) satisfies Condition (U), and because $H$ is separable, the hermitian space $A/n_g$ and the $C^*$-algebra $A_g$ are separable. Therefore the Bochner-Godement theorem applies to the positive linear form $f_{x_0}$, and Equation (BG) shows that $f_{x_0}$ is determined by the hermitian characters of $X(A)$.

As shown just after Proposition 10.31, if the topologically cyclic representation $U: A \to \mathcal{L}(H)$ has $x_0$ has cyclic vector, then it is completely determined by $f_{x_0}$. Also, by Proposition 10.6, if $A$ is unital, then every (nondegenerate) representation of $A$ in a separable Hilbert space is the countable Hilbert sum of topologically cyclic representations. Therefore, if $A$ is commutative unital algebra, then every representation $U: A \to \mathcal{L}(H)$ of $A$ in a separable Hilbert space $H$ is completely determined by the hermitian characters of $X(A)$. 
Chapter 11

Unitary Representations of Locally Compact Groups

In this chapter we discuss representations of locally compact groups. For simplicity, we begin with finite-dimensional representations, which are group homomorphisms $U: G \to \text{GL}(V)$, where $V$ is a finite-dimensional complex vector space (see Section 11.1). Next we consider unitary representations, which are certain kinds of continuous homomorphisms $U: G \to U(H)$, where $H$ is a Hilbert space (typically separable), and $U(H)$ is the group of unitary operators on $H$, that is, the continuous linear maps $f: H \to H$ that have a continuous inverse, and preserve the inner product; that is,

$$\langle f(x), f(y) \rangle = \langle x, y \rangle \quad \text{for all } x, y \in H.$$

Then a unitary operator is a continuous linear map $f: H \to H$ such that $f^{-1} = f^*$, where $f^*$ is the adjoint of $f$, the unique continuous linear map determined by the equation

$$\langle f^*(x), y \rangle = \langle x, f(y) \rangle \quad \text{for all } x, y \in H.$$

The basic theory of unitary representations is discussed in Section 11.2.

There are two main results in this chapter.

The main first main result (first shown by Naimark) is that every unitary representation $U: G \to U(H)$ of a locally compact group $G$ defines a nondegenerate unitary representation $U_{\text{ext}}: \text{L}^1(G) \to \mathcal{L}(H)$ of the involutive Banach algebra $\text{L}^1(G)$, and that conversely, for every nondegenerate unitary representation $V: \text{L}^1(G) \to \mathcal{L}(H)$ of $\text{L}^1(G)$, there is a unique unitary representation $U: G \to U(H)$ of the group $G$ such that $V = U_{\text{ext}}$ (see Section 11.3, Theorem 11.9 and Theorem 11.10).

The bijection $U \mapsto U_{\text{ext}}$ between unitary representations of a locally compact group $G$ and nondegenerate unitary representations of the algebra $\text{L}^1(G)$ is a basic tool that allows the transfer of results about representations of algebras to representations of groups, and vice-versa. It will play a crucial role in the proof of the Peter–Weyl theorem.
The second main result (Gelfand and Raikov, Godement) is that there is one-to-one correspondence between unitary cyclic representations of a locally compact group $G$ and certain bounded continuous functions on $G$ called functions of positive type.

Let $U: G \to \text{U}(H)$ be a unitary representation of the locally compact group $G$ in a Hilbert space $H$, let $x_0$ be any vector in $H$, and define the map $p = \psi_{U,x_0}$ by

$$p(s) = \psi_{U,x_0}(s) = \langle U(s)(x_0), x_0 \rangle, \quad s \in G.$$  

It turns out that the function $p$ is continuous and bounded, and that it satisfies the following property:

$$\int (f^* \ast f)(s) p(s) \, d\lambda(s) \geq 0 \quad \text{for all } f \in K_C(G).$$

Such functions are called functions of positive type. Remarkably, every continuous function $p$ of positive type determines a unitary topologically cyclic representation $U$ with a cyclic vector $x_0$, such that $p = \psi_{U,x_0}$ (see Theorem 11.11). The connection between cyclic unitary representations and functions of positive type is discussed in Section 11.4.

In Section 11.5 we present the Gelfand–Raikov theorem without proof (see Theorem 11.15). Informally, this theorem says that there is vast supply of irreducible unitary representations for any locally compact group. This is far from obvious a priori. For example, $\text{SL}(2, \mathbb{R})$ does not have finite-dimensional unitary representations, and it is not that easy to find irreducible unitary representations.

Section 11.6 is devoted to measures of positive type, which generalize functions of positive type. A complex or $\sigma$-Radon measure $\mu$ is of positive type if

$$\int (f^* \ast f)(s) \, d\mu(s) = \int \int \overline{f(t)f(ts)} \, d\lambda(t) \, d\mu(s) \geq 0, \quad \text{for all } f \in K_C(G).$$

The Dirac measure $\delta_e$ is a measure of positive type, and more generally, if $\nu$ is a complex measure, then the measure $\overline{\nu} \ast \nu$ is of positive type. The main point is that a measure $\mu$ of positive type defines a unitary representation $U_\mu$ of $G$ in a separable Hilbert space $H$ (see Theorem 11.19). This construction will be used in Section 14.6 to define the Plancherel transform.

Basically all the material of this chapter is presented in a more condensed form in Dixmier’s classical book Dixmier [24] (see also the English translation published by the AMS).

### 11.1 Finite-Dimensional Group Representations

For simplicity, we begin with finite-dimensional representations. A good and easy-going introduction to representations of Lie groups can be found in Hall [38].
Definition 11.1. Given a locally compact group $G$ and a vector space $V$ of dimension $n$, a linear representation of $G$ in $V$ of dimension (or degree) $n$ is a group homomorphism $U: G \to \text{GL}(V)$ such that the map $g \mapsto U(g)(u)$ is continuous for every $u \in V$, where $\text{GL}(V)$ denotes the group of invertible linear maps from $V$ to itself. The space $V$, called the representation space, may be a real or a complex vector space. If $V$ has a Hermitian (resp. Euclidean) inner product $\langle -, - \rangle$, we say that $U: G \to \text{GL}(V)$ is a unitary representation iff

$$\langle U(g)(u), U(g)(v) \rangle = \langle u, v \rangle,$$

for all $g \in G$ and all $u, v \in V$.

Thus, a linear representation of $G$ is a map $U: G \to \text{GL}(V)$ satisfying the properties:

$$U(gh) = U(g)U(h)$$

$$U(g^{-1}) = U(g)^{-1}$$

$$U(1) = \text{id}_V.$$ 

If $U$ is a unitary representation, then we also have

$$(U(g))^{-1} = (U(g))^*.$$ 

If $G$ is a finite group, the continuity requirement is omitted.

Note that a major difference with the notion of representation of an algebra, is that for a group representation $U: G \to \text{GL}(V)$, the linear map $U(g)$ must be invertible for every $g \in G$. For an algebra representation $U: A \to \mathcal{L}(H)$ (where $H$ is a Hilbert space), the linear maps $U(s)$ are generally not invertible.

For simplicity of language, we usually abbreviate linear representation as representation. The representation space $V$ is also called a $G$-module, since the representation $U: G \to \text{GL}(V)$ is equivalent to the left action $\cdot: G \times V \to V$, with $g \cdot v = U(g)(v)$. The representation such that $U(g) = \text{id}_V$ for all $g \in G$ is called the trivial representation.

It should be noted that because $V$ is finite-dimensional, the condition that for every $u \in V$, the map $g \mapsto U(g)(u)$ is continuous, is actually equivalent to the fact that the map $g \mapsto U(g)$ from $G$ to $\mathcal{L}(V)$ equipped with any norm, is continuous.

Indeed, for any basis of $V$, the fact that the map $g \mapsto U(g)(u)$ is continuous implies that the matrix $(U_{ij}(g))$ representing $U(g)$ consists of continuous functions on $G$.

As an example, we describe a class of representations of $G = \text{SL}(2, \mathbb{C})$, the group of complex matrices with determinant +1,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad ad - bc = 1.$$ 

Recall that $\mathcal{P}^C_k(2)$ denotes the vector space of complex homogeneous polynomials of degree $k$ in two variables $(z_1, z_2)$. A complex homogeneous polynomials of degree $k$ in two variables
$(z_1, z_2)$ is an expression of the form $P(z_1, z_2) = \sum_{i=0}^{k} c_i z_1^i z_2^{k-i}$, with $c_i \in \mathbb{C}$. For every matrix $A \in \text{SL}(2, \mathbb{C})$, with
\[ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \]
for every homogeneous polynomial $Q \in P_k^C(2)$, we define $U_k(A)(Q(z_1, z_2))$ by
\[ U_k(A)(Q(z_1, z_2)) = Q(dz_1 - bz_2, -cz_1 + az_2). \]
If we think of the homogeneous polynomial $Q(z_1, z_2)$ as a function $Q(z_1 z_2)$ of the vector $(z_1, z_2)$, then
\[ U_k(A) \left( \frac{z_1}{z_2} \right) = QA^{-1} \left( \frac{z_1}{z_2} \right) = Q \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}. \]
The expression above makes it clear that
\[ U_k(AB) = U_k(A)U_k(B) \]
for any two matrices $A, B \in \text{SL}(2, \mathbb{C})$, so $U_k$ is indeed a representation of $\text{SL}(2, \mathbb{C})$ into $P_k^C(2)$.

One might wonder why we considered $\text{SL}(2, \mathbb{C})$ rather than $\text{SL}(2, \mathbb{R})$. This is because it can be shown that $\text{SL}(2, \mathbb{R})$ has no nontrivial unitary (finite-dimensional) representations! For more on representations of $\text{SL}(2, \mathbb{R})$, see Dieudonné [18] (Chapter 14).

Given any basis $(e_1, \ldots, e_n)$ of $V$, each $U(g)$ is represented by an $n \times n$ matrix $U(g) = (U_{ij}(g))$. We may think of the scalar functions $g \mapsto U_{ij}(g)$ as special functions on $G$. As explained in Dieudonné [18] (see also Vilenkin [82]), essentially all special functions (Legendre polynomials, ultraspherical polynomials, Bessel functions etc.) arise in this way by choosing some suitable $G$ and $V$.

There is a natural and useful notion of equivalence of representations:

**Definition 11.2.** Given any two representations $U_1 : G \to \text{GL}(V_1)$ and $U_2 : G \to \text{GL}(V_2)$, a $G$-map (or morphism of representations) $\varphi : U_1 \to U_2$ is a linear map $\varphi : V_1 \to V_2$ so that the following diagram commutes for every $g \in G$:

\[ V_1 \xrightarrow{U_1(g)} V_1 \\
\varphi \downarrow \quad \quad \downarrow \varphi \\
V_2 \xrightarrow{U_2(g)} V_2, \]

i.e.
\[ \varphi(U_1(g)(v)) = U_2(g)(\varphi(v)), \quad v \in V_1. \]

The space of all $G$-maps between two representations as above is denoted $\text{Hom}_G(U_1, U_2)$. Two representations $U_1 : G \to \text{GL}(V_1)$ and $U_2 : G \to \text{GL}(V_2)$ are equivalent iff $\varphi : V_1 \to V_2$
is an invertible linear map (which implies that \( \dim V_1 = \dim V_2 \)). In terms of matrices, the representations \( U_1: G \to \text{GL}(V_1) \) and \( U_2: G \to \text{GL}(V_2) \) are equivalent iff there is some invertible \( n \times n \) matrix, \( P \), so that

\[
U_2(g) = PU_1(g)P^{-1}, \quad g \in G.
\]

If \( W \subseteq V \) is a subspace of \( V \), then in some cases, a representation \( U: G \to \text{GL}(V) \) yields a representation \( U: G \to \text{GL}(W) \). This is interesting because under certain conditions on \( G \) (e.g., \( G \) compact) every representation may be decomposed into a “sum” of so-called irreducible representations (defined below), and thus the study of all representations of \( G \) boils down to the study of irreducible representations of \( G \); for instance, see Knapp [49] (Chapter 4, Corollary 4.7), or Bröcker and tom Dieck [14] (Chapter 2, Proposition 1.9).

**Definition 11.3.** Let \( U: G \to \text{GL}(V) \) be a representation of \( G \). If \( W \subseteq V \) is a subspace of \( V \), then we say that \( W \) is invariant (or stable) under \( U \) iff \( U(g)(w) \in W \), for all \( g \in G \) and all \( w \in W \). If \( W \) is invariant under \( U \), then we have a homomorphism, \( U: G \to \text{GL}(W) \), called a subrepresentation of \( G \). A representation \( U: G \to \text{GL}(V) \) with \( V \neq (0) \) is irreducible iff it only has the two subrepresentations \( U: G \to \text{GL}(W) \) corresponding to \( W = (0) \) or \( W = V \).

It can be shown that the representations \( U_k \) of \( \text{SL}(2, \mathbb{C}) \) defined earlier are irreducible, and that every representation of \( \text{SL}(2, \mathbb{C}) \) is equivalent to one of the \( U_k \)'s (see Bröcker and tom Dieck [14], Chapter 2, Section 5). The representations \( U_k \) are also representations of \( \text{SU}(2) \). Again, they are irreducible representations of \( \text{SU}(2) \), and they constitute all of them (up to equivalence). The reader should consult Hall [38] for more examples of representations of Lie groups.

An easy but crucial lemma about irreducible representations is “Schur’s Lemma.”

**Lemma 11.1.** (Schur’s Lemma) Let \( U_1: G \to \text{GL}(V) \) and \( U_2: G \to \text{GL}(W) \) be any two real or complex representations of a group \( G \). If \( U_1 \) and \( U_2 \) are irreducible, then the following properties hold:

(i) Every \( G \)-map \( \varphi: U_1 \to U_2 \) is either the zero map or an isomorphism.

(ii) If \( U_1 \) is a complex representation, then every \( G \)-map \( \varphi: U_1 \to U_1 \) is of the form \( \varphi = \lambda \text{id} \), for some \( \lambda \in \mathbb{C} \).

**Proof.** (i) Observe that the kernel \( \text{Ker} \varphi \subseteq V \) of \( \varphi \) is invariant under \( U_1 \). Indeed, for every \( v \in \text{Ker} \varphi \) and every \( g \in G \), we have

\[
\varphi(U_1(g)(v)) = U_2(g)(\varphi(v)) = U_2(g)(0) = 0,
\]

so \( U_1(g)(v) \in \text{Ker} \varphi \). Thus, \( U_1: G \to \text{GL}(\text{Ker} \varphi) \) is a subrepresentation of \( U_1 \), and as \( U_1 \) is irreducible, either \( \text{Ker} \varphi = (0) \) or \( \text{Ker} \varphi = V \). In the second case, \( \varphi = 0 \). If \( \text{Ker} \varphi = (0) \),
then $\varphi$ is injective. However, $\varphi(V) \subseteq W$ is invariant under $U_2$, since for every $v \in V$ and every $g \in G$,

$$U_2(g)(\varphi(v)) = \varphi(U_1(g)(v)) \in \varphi(V),$$

and as $\varphi(V) \neq (0)$ (as $V \neq (0)$ since $U_1$ is irreducible) and $U_2$ is irreducible, we must have $\varphi(V) = W$; that is, $\varphi$ is an isomorphism.

(ii) Since $V$ is a complex vector space, the linear map $\varphi$ has some eigenvalue $\lambda \in \mathbb{C}$. Let $E_\lambda \subseteq V$ be the eigenspace associated with $\lambda$. The subspace $E_\lambda$ is invariant under $U_1$, since for every $u \in E_\lambda$ and every $g \in G$, we have

$$\varphi(U_1(g)(u)) = U_1(g)(\varphi(u)) = U_1(g)(\lambda u) = \lambda U_1(g)(u),$$

so $U_1 : G \to \text{GL}(E_\lambda)$ is a subrepresentation of $U_1$, and as $U_1$ is irreducible and $E_\lambda \neq (0)$, we must have $E_\lambda = V$.

An interesting corollary of Schur’s Lemma is the following fact:

**Proposition 11.2.** Every complex irreducible representation $U : G \to \text{GL}(V)$ of a commutative group $G$ is one-dimensional.

**Proof.** Since $G$ is abelian, we claim that for every $g \in G$, the map $\tau_g : V \to V$ given by $\tau_g(v) = U(g)(v)$ for all $v \in V$ is a $G$-map. This amounts to checking that the following diagram commutes

$$
\begin{array}{ccc}
V & \xrightarrow{U(g_1)} & V \\
\downarrow{\tau_g} & & \downarrow{\tau_g} \\
V & \xrightarrow{U(g_1)} & V \\
\end{array}
$$

for all $g, g_1 \in G$. This is equivalent to checking that

$$\tau_g(U(g_1)(v)) = U(g)(U(g_1)(v)) = U(gg_1)(v) = U(g_1)(\tau_g(v)) = U(g_1)(U(g)(v)) = U(g_1g)(v)$$

for all $v \in V$, that is, $U(gg_1)(v) = U(g_1g)(v)$, which holds since $G$ is commutative (so $gg_1 = g_1g$).

By Schur’s Lemma (Lemma 11.1 (ii)), $\tau_g = \lambda_g \text{id}$ for some $\lambda_g \in \mathbb{C}$. It follows that any subspace of $V$ is invariant. If the representation is irreducible, we must have $\dim(V) = 1$ since otherwise $V$ would contain a one-dimensional invariant subspace, contradicting the assumption that $U$ is irreducible.

## 11.2 Unitary Group Representations

We now generalize representations to allow the representing space to be a Hilbert space (typically separable).
**Definition 11.4.** Given a locally compact group $G$ and a Hilbert space $H$, a unitary representation of $G$ in $H$ is a group homomorphism $U : G \to \mathbf{U}(H)$, where $\mathbf{U}(H)$ is the group of unitary operators on $H$, such that:

1. The map $g \mapsto U(g)(u)$ is continuous for every $u \in H$.

2. The linear maps $U(g)$ are invertible, continuous, have a continuous inverse, and are isometries of $H$, that is,

   \[ \langle U(g)(u), U(g)(v) \rangle = \langle u, v \rangle \quad \text{for all } g \in G \text{ and all } u, v \in H. \]

   The above condition is equivalent to: the linear maps $U(g)$ are invertible, continuous, and

   \[ (U(g))^{-1} = (U(g))^* \quad \text{for all } g \in G. \]

**Remark:** Sometimes, a unitary representation as in Definition 11.4 is called a continuous unitary representation. Note that if $H$ is infinite-dimensional, the map $g \mapsto U(g)$ is not necessarily continuous. For a counter-example involving the regular representation of an infinite compact group $G$ in $L^2(G)$, see Dieudonné [19] (Chapter XXI, Section 1, Problem 3). However, the left action $U^\alpha : G \times H \to H$ associated with $U$ given by

\[ U^\alpha(s, x) = U(s)(x), \quad \text{for all } s \in G \text{ and all } x \in H \]

is continuous. Indeed, since $U(s)$ is a unitary map, we have $\|U(s)(w)\| = \|w\|$ for all $w \in H$, so for all $s, t \in G$ and all $x, y \in H$, we have

\[
\|U^\alpha(s, x) - U^\alpha(t, y)\| \leq \|U(s)(x) - U(s)(y)\| + \|U(s)(y) - U(t)(y)\| \\
= \|U(s)(x - y)\| + \|U(s)(y) - U(t)(y)\| \\
= \|x - y\| + \|U(s)(y) - U(t)(y)\|,
\]

and since by hypothesis, for any fixed $y \in H$, the map $s \mapsto U(s)(y)$ is continuous, we see that the action $U^\alpha$ is continuous. Conversely, if the action $U^\alpha : G \times H \to H$ is continuous, then obviously the map $s \mapsto U(s)(y)$ is continuous, so $U$ is a unitary representation.

The notion of morphism of unitary representations and of equivalence is adapted as follows.

**Definition 11.5.** Given any two unitary representations $U_1 : G \to \mathbf{U}(H_1)$ and $U_2 : G \to \mathbf{U}(H_2)$, a $G$-map (or morphism of representations) $\varphi : U_1 \to U_2$ is a continuous linear isometry $\varphi : H_1 \to H_2$, which means that

\[ \langle \varphi(u), \varphi(v) \rangle_2 = \langle u, v \rangle_1, \quad \text{for all } u, v \in H_1, \]
so that the following diagram commutes for every $g \in G$:

$$
\begin{array}{ccc}
H_1 & \xrightarrow{U_1(g)} & H_1 \\
\downarrow{\varphi} & & \downarrow{\varphi} \\
H_2 & \xrightarrow{U_2(g)} & H_2,
\end{array}
$$

i.e.

$$
\varphi(U_1(g)(v)) = U_2(g)(\varphi(v)), \quad v \in H_1.
$$

The space of all $G$-maps between two representations as above is denoted $\text{Hom}_G(U_1, U_2)$. Two unitary representations $U_1 : G \to \mathbf{U}(H_1)$ and $U_2 : G \to \mathbf{U}(H_2)$ are equivalent iff $\varphi : H_1 \to H_2$ is an invertible linear isometry whose inverse is also continuous; thus $U_2(g) = \varphi \circ U_1(g) \circ \varphi^{-1}$, for all $g \in G$.

Given a unitary representation $U : G \to \mathbf{U}(H)$, the definition of an invariant subspace $W \subseteq H$ is the same as in Definition 11.3. If $W \subseteq H$ is invariant under $U$, we say that the subrepresentation $U : G \to \mathbf{U}(W)$ is closed if $W$ is closed in $H$. As in the case of unitary representations of algebras, the notion of subrepresentation is only well defined for closed invariant subspaces of $H$. However, by Proposition 11.4, since the closure $\overline{W}$ of an invariant subspace $W$ is closed, the notion of subrepresentation of $G$ in $\overline{W}$ is well defined.

In the definition of an irreducible unitary representation $U : G \to \mathbf{U}(H)$, we require that the only closed subrepresentations $U : G \to \mathbf{U}(W)$ of the representation $U : G \to \mathbf{U}(H)$ correspond to $W = (0)$ or $W = H$.

If $U : G \to \mathbf{U}(H)$ is an irreducible unitary representation, then the subspace spanned by the set $\{U(s)(u) \mid s \in G, u \in H\}$ is dense in $H$. As for representations of algebras, we can define topologically cyclic representations and cyclic vectors.

**Definition 11.6.** Let $U : G \to \mathbf{U}(H)$ be a unitary representation of $G$ in $H$. A vector $x_0 \in H$ is called a totalizer, or totalizing vector, or cyclic vector for the representation $U$ if the subspace of $H$ spanned by the set $\{U(s)(x_0) \mid s \in G\}$ is dense in $H$. Equivalently if $\mathcal{M}_{x_0}$ denotes the closure of the set $\{U(s)(x_0) \mid s \in G\}$, called the cyclic subspace generated by $x_0$, which is invariant under $U$, then $x_0$ is a totalizer (a cyclic vector) if $\mathcal{M}_{x_0} = H$. A representation which admits a totalizer is said to be topologically cyclic.

The importance of totalizers stems from the following result.

**Proposition 11.3.** Let $U : G \to \mathbf{U}(H)$ be a unitary representation of $G$ in $H$. Then $H$ is the Hilbert sum of a sequence $(H_\alpha)_{\alpha \in \Lambda}$ of closed subspaces $H_\alpha \neq (0)$ of $H$ invariant under $U$, and such that the restriction of $U$ to each $H_\alpha$ is topologically cyclic. If $H$ is separable, the family $\Lambda$ is countable (possibly finite).

Proposition 11.3 is proven in the separable case in Dieudonné [22] (Chapter XV, Section 5), and in general, using Zorn’s lemma; see Folland [31] (Chapter 3, Proposition 3.3).
11.2. UNITARY GROUP REPRESENTATIONS

Hilbert sums of unitary representations of a locally compact group are defined just as in the case of an algebra; see just after Proposition 10.4. We also have the following version of Proposition 10.1 for group representations.

**Proposition 11.4.** Let \( U : G \to \mathbf{U}(H) \) be a unitary representation of \( G \) in \( H \).

1. If the subspace \( E \) of \( H \) is invariant under \( U \), then its closure \( \overline{E} \) is also invariant under \( U \).

2. Let \( E \) be a closed subspace of \( H \) invariant under \( U \). If \( E^\perp \) is the orthogonal complement of \( E \) in \( H \), then \( E^\perp \) is invariant under \( U \). If \( U_1(s) \) and \( U_2(s) \) are the restrictions of \( U(s) \) to \( E \) and \( E^\perp \), then \( H = E \oplus E^\perp \) (the algebraic direct sum), and the representation \( U \) is the Hilbert sum of the representations \( U_1 \) and \( U_2 \).

**Proof.** Part (1) is easy to prove and follows from the continuity of \( U(s) \); see Dieudonné [23], (Chapter III, Section 11). For Part (2), let \( x \in E \) and \( y \in E^\perp \). For any \( s \in G \) we have
\[
\langle x, U(s)(y) \rangle = \langle (U(s))^*(x), y \rangle = \langle (U(s))^{-1}(x), y \rangle = \langle (s^{-1})(x), y \rangle = 0,
\]
since \( E \) is invariant under \( U \), so \( U(s^{-1})(x) \in E \), and since \( E^\perp \) is the orthogonal complement of \( E \) and \( y \in E^\perp \). Then \( U(s)(y) \) is orthogonal to all \( x \in E \), which means that \( U(s)(y) \in E^\perp \), so \( E^\perp \) is invariant under \( U \). The last property is obvious.

One should realize that Property (2) or Proposition 11.4 fails for nonunitary representations. For example, the map
\[
U : x \mapsto \begin{pmatrix} 1 & x \\ 0 & 1 \end{pmatrix}
\]
is a representation of \( \mathbb{R} \) in \( \mathbb{C}^2 \), but the only invariant subspace is the one-dimensional subspace spanned by \((1, 0)\). The problem is that because \( \mathbb{R} \) is not compact, there is no way to define an inner product on \( \mathbb{C}^2 \) invariant under \( U \).

However, using the Haar measure, Theorem 6.35 shows that if \( H \) is a finite-dimensional hermitian space, then there is an inner-product on \( H \) for which the linear maps \( U(s) \) are unitary.

**Theorem 11.5.** (Complete Reducibility) Let \( U : G \to \mathbf{U}(H) \) be a linear representation of a compact group \( G \) in a finite-dimensional hermitian space \( H \) of dimension \( n \geq 1 \). There is a hermitian inner product \( \langle - , - \rangle \) on \( H \) such that \( U : G \to \mathbf{U}(H) \) is a unitary representation of \( G \) in the hermitian space \( (H, \langle - , - \rangle) \). The representation \( U \) is the direct sum of a finite number of irreducible unitary representations.

**Proof.** As we noted before, the representation \( U : G \to \mathbf{U}(H) \) is a continuous linear map \( g \mapsto U(g) \) from \( G \) to \( \mathcal{L}(H) \) equipped with any norm. Theorem 6.35 yields an inner product on \( H \) which is invariant under \( U \).
We proceed by complete induction on the dimension \( n \geq 1 \) of \( H \). When \( n = 1 \), the representation is automatically irreducible. If \( n > 1 \) and the representation is not irreducible, then it has some invariant subspace \( H_1 \) of dimension \( n_1 \) with \( 1 \leq n_1 < n \). By Proposition 11.4, the orthogonal complement \( H_2 = H_1^\perp \) of \( H_1 \) is also invariant under \( U \), and its dimension \( n_2 \) satisfies \( n_2 \geq 1 \) an \( n_1 + n_2 = n \), with \( n > 1 \) and \( 1 \leq n_1 < n \), so we also have \( 1 \leq n_2 < n \). We can apply the induction hypothesis to the subrepresentations \( U: G \to U(H_1) \) and \( U: G \to U(H_2) \), with \( H = H_1 \oplus H_2 \), and we obtain a collection of irreducible representations of \( G \) whose direct sum is \( U \).

Theorem 11.5 is very significant because it shows that the study of arbitrary finite-dimensional representations of a compact group \( G \) reduces to the study of the irreducible unitary (finite-dimensional) representations of \( G \).

The analog of Proposition 10.3 holds for unitary group representations.

**Proposition 11.6.** Let \( U: G \to U(H) \) be a unitary representation of \( G \) in \( H \). A closed subspace \( E \) of \( H \) is invariant under \( U \) iff \( P_E U(g) = U(g) P_E \) for all \( g \in G \), in other words, \( P_E \in \text{Hom}_G(U,U) \), where \( P_E: H \to E \) is the orthogonal projection of \( H \) onto \( E \).

**Proof.** Assume that \( P_E \in \text{Hom}_G(U,U) \), so that the following diagram commutes:

\[
\begin{array}{ccc}
H & \xrightarrow{U(g)} & H \\
\downarrow{P_E} & & \downarrow{P_E} \\
H & \xrightarrow{U(g)} & H.
\end{array}
\]

For any \( x \in E \), since \( P_E \) is the orthogonal projection of \( H \) onto \( E \), we have \( P_E(x) = x \), so

\[ P_E(U(g)(x)) = U(g)(P_E(x)) = U(g)(x), \]

which shows that \( U(g)(x) \in E \), and thus that \( E \) is invariant under \( U \).

Conversely, assume that \( E \) is invariant under \( U \). Since \( E \) is closed, by a well-known result of Hilbert space theory, we have \( H = E \oplus E^\perp \), an algebraic direct sum. For any \( x \in E \), since \( E \) is invariant under \( U \), we have \( U(g)(x) \in E \) for all \( g \in G \), and since \( P_E \) is a projection onto \( E \), we have

\[ U(g)(P_E(x)) = U(g)(x) = P_E(U(g)(x)) \quad \text{for all } x \in E. \]

By Proposition 11.4, the subspace \( E^\perp \) is also invariant under \( U \). For any \( x \in E^\perp \), we have \( U(g)(x) \in E^\perp \), so \( P_E(x) = P_E(U(g)(x)) = 0 \), and we have

\[ U(g)(P_E(x)) = 0 = P_E(U(g)(x)) \quad \text{for all } x \in E^\perp. \]

Since \( U = E \oplus E^\perp \), we have

\[ U(g)(P_E(x)) = 0 = P_E(U(g)(x)) \quad \text{for all } x \in H, \]

namely, \( P_E \in \text{Hom}_G(U,U) \). \qed
Proposition yields a method for proving that a unitary representation \( U : G \to U(H) \) is irreducible. Indeed, if \( U \) is reducible, then there is some \( G \)-map \( \varphi \in \text{Hom}_G(U,U) \) which is not invertible. Thus, if every \( G \)-map in \( \text{Hom}_G(U,U) \) is invertible, then \( U \) must be irreducible.

Actually, there is a generalization of Schur’s lemma to (complex) unitary representations, which says that if a unitary representation \( U : G \to U(H) \) is irreducible, then every \( G \)-map in \( \text{Hom}_G(U,U) \) is of the form \( \alpha \text{id}_H \), for some \( \alpha \in \mathbb{C} \).

The proof is much more difficult because a linear operator on an infinite-dimensional vector space may not have eigenvectors! It uses some results from the spectral theory of algebras.

Since we won’t need it, we state the Schur’s lemma for unitary representations without proof, and refer the interested reader to Folland [31] for a proof (Chapter 3, Proposition 3.5).

**Theorem 11.7. (Schur’s lemma for unitary representations)** The following properties hold.

1. A (complex) unitary representation \( U : G \to U(H) \) is irreducible iff every \( G \)-map in \( \text{Hom}_G(U,U) \) is of the form \( \alpha \text{id}_H \), for some \( \alpha \in \mathbb{C} \).
2. Let \( U_1 : G \to U(H) \) and \( U_2 : G \to U(H) \) be two complex unitary representations. If \( U_1 \) and \( U_2 \) are equivalent, then \( \text{Hom}_G(U_1,U_2) \) is one-dimensional; otherwise we have \( \text{Hom}_G(U_1,U_2) = (0) \).

### 11.3 Unitary Representations of \( G \) and \( L^1(G) \)

In this section, we discuss the crucial fact that every unitary representation \( U : G \to U(H) \) of a locally compact group \( G \) defines a nondegenerate unitary representation \( U_{\text{ext}} : L^1(G) \to L(H) \) of the involutive Banach algebra \( L^1(G) \), and that conversely, for every nondegenerate unitary representation \( \rho : L^1(G) \to L(H) \) of \( L^1(G) \), there is a unique unitary representation \( U : G \to U(H) \) of the group \( G \) such that \( \rho = U_{\text{ext}} \).

Dieudonné [19] (Chapter XXI, Section 1) proves the above results under the simplifying assumption that \( G \) is metrizable, separable, and unimodular (and of course locally compact). One of the reasons is that Dieudonné only shows the existence of the Haar measure for a metrizable, separable, locally compact group. We prove it for an arbitrary locally compact group.

The bijection holds for any locally compact group, not necessarily unimodular, and is proven in Folland [31] (Chapter 3). Since the technical details are not particularly illuminating, we will give an outline of the constructions and proofs, using the simplifying assumption that \( G \) is metrizable. This includes the case of Lie groups. The involution \( f \mapsto f^* \) in \( L^1(G) \) is given by \( f^*(s) = \Delta(s^{-1})f(s^{-1}) \), but not much simplification is afforded if we assume that \( G \) is unimodular.

First we show that unitary representations of \( L^1(G) \) are continuous.
Proposition 11.8. Let $G$ be a locally compact group and let $V : \mathcal{L}^1(G) \to \mathcal{L}(H)$ be a unitary representation. We have
\[ \|V(f)\| \leq \|f\|_1 \quad \text{for all } f \in \mathcal{L}^1(G), \]
and thus $V$ is continuous.

Proof. If $G$ is discrete, this follows by Proposition 10.7. Otherwise, we can extend $V$ to a unitary representation of the unital subalgebra $\mathcal{L}^1(G) \oplus \mathbb{C}\delta_e$ of $\mathcal{M}^1(G)$ by setting $V(fd\lambda + \alpha \delta_e) = V(f) + \alpha \text{id}_H$, and then we apply Proposition 10.7 to this representation.

We begin by explaining how to construct a nondegenerate unitary representation of the algebra $\mathcal{M}^1(G)$ of complex regular Borel measures (see Definition 5.21) from a continuous unitary representation $U : G \to U(H)$ of $G$. Recall that if $\lambda$ is a left Haar measure on $G$, then we have an embedding of $\mathcal{L}^1(G)$ into $\mathcal{M}^1(G)$ given by $f \mapsto fd\lambda$.

We need to define a map $\hat{U} : \mathcal{M}^1(G) \to \mathcal{L}(H)$ which is an algebra homomorphism. Pick any complex regular Borel measure $\mu \in \mathcal{M}^1(G)$. We need to define $\hat{U}(\mu)$ as a continuous linear map from $H$ to itself. An obvious candidate is
\[ \hat{U}(\mu)(x) = \int U(s)(x) \, d\mu(s), \quad x \in H \]
but the right-hand side is an integral over the vector-valued function $s \mapsto U(s)(x)$ from $G$ to $H$, and $\mu$ is generally not a positive measure, so the theory of integration that we have presented does not apply. The theory does extend to complex measures (see Schwartz [72]), but we do not know whether this type of integral has the properties needed to obtain the desired results, so instead we will resort to a so-called weak integral. The idea is to use the duality between the Hilbert space $H$ and its dual $H'$, the space of continuous linear forms. So, for $x \in H$ fixed, we define the semilinear form $\Phi_{\mu,x} : H \to \mathbb{C}$ given by
\[ \Phi_{\mu,x}(y) = \int \langle U(s)(x), y \rangle \, d\mu(s), \quad y \in H; \]
this form is semilinear because $\Phi_{\mu,x}(y_1 + y_2) = \Phi_{\mu,x}(y_1) + \Phi_{\mu,x}(y_2)$, but $\Phi_{\mu,x}(\lambda y) = \overline{\lambda} \Phi_{\mu,x}(y)$. The function $s \mapsto \langle U(s)(x), y \rangle$ is continuous and bounded because $\|U(s)(x)\| = \|x\|$ since $U(s)$ is unitary, so it is $\mu$-integrable (recall that $|\mu|(X)$ is finite). Using the Cauchy-Schwarz inequality we also have
\[ |\Phi_{\mu,x}(y)| = \left| \int \langle U(s)(x), y \rangle \, d\mu(s) \right| \leq \|\mu\| \|x\| \|y\|, \]
so the linear form $\Phi_{\mu,x}$ is continuous. By the Riesz representation theorem for Hilbert spaces, there is a unique vector $\tilde{U}(\mu)(x) \in H$ such that
\[ \langle \tilde{U}(\mu)(x), y \rangle = \Phi_{\mu,x}(y) \quad \text{for all } y \in H. \]
If we let \( y = \tilde{U}(\mu)(x) \) in the inequality

\[
|\langle \tilde{U}(\mu)(x), y \rangle| \leq \|\mu\| \|x\| \|y\|
\]

we get

\[
\|\tilde{U}(\mu)(x)\| \leq \|\mu\| \|x\|
\]

and so

\[
\|\tilde{U}(\mu)\| \leq \|\mu\|. \quad \text{(C)}
\]

This shows that \( \tilde{U}(\mu) \) is a continuous linear map (in \( \mathcal{L}(H) \), we use the operator norm induced by the Hermitian norm on \( H \)).

**Definition 11.7.** Given any complex regular Borel measure \( \mu \in \mathcal{M}^1(G) \), for every \( x \in H \), let \( \Phi_{\mu,x} : H \to \mathbb{C} \) be the continuous linear form given by

\[
\Phi_{\mu,x}(y) = \int \langle U(s)(x), y \rangle \, d\mu(s), \quad y \in H.
\]

The unique vector \( \tilde{U}(\mu)(x) \in H \) such that

\[
\langle \tilde{U}(\mu)(x), y \rangle = \Phi_{\mu,x}(y) = \int \langle U(s)(x), y \rangle \, d\mu(s) \quad \text{for all } y \in H
\]

is called the **weak integral** of the function \( s \mapsto U(s)(x) \) from \( G \) to \( H \) with respect to \( \mu \), and is denoted by

\[
\int U(s)(x) \, d\mu(s) = \tilde{U}(\mu)(x).
\]

Observe that

\[
\tilde{U}(\delta_s) = U(s) \quad \text{for all } s \in G,
\]

where \( \delta_s \) is the Dirac measure at \( s \). Also, when \( \mu = f \, d\lambda \) with \( f \in \mathcal{L}^1(G) \), we have

\[
\langle \tilde{U}(f \, d\lambda)(x), y \rangle = \int \langle U(s)(x), y \rangle f(s) \, d\lambda(s) \quad \text{for all } y \in H.
\]

For simplicity of notation, we also write \( \tilde{U}(f) \) instead of \( \tilde{U}(f \, d\lambda) \).

The next step is to show that the map \( \mu \mapsto \tilde{U}(\mu) \) is a representation of the unital involutive Banach algebra \( \mathcal{M}^1(G) \).

**Theorem 11.9.** Let \( G \) be a metrizable and locally compact group, and let \( U : G \to U(H) \) be a unitary representation of \( G \) in \( H \). The map \( \tilde{U} : \mathcal{M}^1(G) \to \mathcal{L}(H) \) defined above is a unitary representation of the unital involutive Banach algebra \( \mathcal{M}^1(G) \). The restriction \( \text{U}_{\text{ext}} : \mathcal{L}^1(G) \to \mathcal{L}(H) \) of \( \tilde{U} \) to the involutive Banach algebra \( \mathcal{L}^1(G) \) is nondegenerate. The theorem also holds for any arbitrary locally compact group \( G \).
Proof. Theorem 11.9 is proven in Dieudonné [19] (Chapter XXI, Section 1, Theorem 21.1.6). Let us verify that \( \widetilde{U}(\mu \ast \nu) = \widetilde{U}(\mu) \circ \widetilde{U}(\nu) \). For all \( x, y \in H \) and all \( s \in G \), we have

\[
\langle \widetilde{U}(\mu \ast \nu)(x), y \rangle = \int \langle U(s)(x), y \rangle d(\mu \ast \nu) = \int \int \langle U(st)(x), y \rangle d\mu(s) d\nu(t) \\
= \int \int \langle U(t)(x), U(s^*)(y) \rangle d\nu(t) d\mu(s) = \int \langle \widetilde{U}(\nu)(x), U(s^*)(y) \rangle d\mu(s) \\
= \int \langle U(s)(\widetilde{U}(\nu)(x)), y \rangle d\mu(s) \\
= \langle \widetilde{U}(\mu)(\widetilde{U}(\nu)(x)), y \rangle,
\]

which proves that \( \widetilde{U}(\mu \ast \nu) = \widetilde{U}(\mu) \circ \widetilde{U}(\nu) \).

Next recall that

\[
\int \varphi(s) d\overline{\mu}(s) = \int \overline{\varphi(s)} d\mu(s)
\]

and

\[
\int \varphi(s) d\tilde{\mu}(s) = \int \varphi(s^{-1}) d\mu(s),
\]

see Proposition 5.23 and Proposition 6.44. Then using the fact that since \( U \) is a unitary representation we have \( (U(s))^* = U(s^{-1}) \), we have

\[
\langle (U(\mu))^*(x), y \rangle = \langle x, U(\mu)(y) \rangle = \langle U(\mu)(y), x \rangle \\
= \int \langle U(s)(y), x \rangle d\mu(s) \\
= \int \overline{\langle U(s)(y), x \rangle} d\overline{\mu}(s) = \int \overline{\langle x, U(s)(y) \rangle} d\overline{\mu}(s) \\
= \int \langle (U(s))^*(x), y \rangle d\overline{\mu}(s) = \int \langle U(s^{-1})(x), y \rangle d\overline{\mu}(s) \\
= \int \langle U(s)(x), y \rangle d\overline{\mu}(s) = \langle U(\overline{\mu})(x), y \rangle,
\]

which proves that \( (U(\mu))^* = U(\overline{\mu}) \).

To prove that the restriction of \( \widetilde{U} \) to \( L^1(G) \) is nondegenerate, since \( G \) is metrizable, we can find a neighborhood base of \( e \) (the identity element of \( G \)) consisting of a sequence \( (V_n) \) of open neighborhoods of \( e \) such that \( V_{n+1} \subseteq V_n \) for all \( n \).\(^1\) Fix \( s \in G \). Using Proposition 15.39, for every \( n \geq 1 \) we can define a positive function \( u_n \in K_\infty(G) \) of support contained in \( sV_n \), such that \( \int u_n \, d\lambda = 1 \). For every \( x \in H \) and every \( \epsilon > 0 \), there is some \( n > 0 \) such that

\[
\|U(t)(x) - U(s)(x)\| \leq \epsilon, \quad \text{for all } t \in sV_n.
\]

\(^1\)This is where we assumption that \( G \) is metrizable is used. Otherwise, we may have to use an uncountable family.
For all $y \in H$, since $\int u_n \, d\lambda = 1$, we have

$$\langle \tilde{U}(u_n \, d\lambda)(x) - U(s)(x), y \rangle = \int \langle U(t)(x) - U(s)(x), y \rangle u_n(t) \, d\lambda(t),$$

so if we choose $y = \tilde{U}(u_n \, d\lambda)(x) - U(s)(x)$, we get

$$\| \tilde{U}(u_n \, d\lambda)(x) - U(s)(x) \|^2 = \int \langle U(t)(x) - U(s)(x), \tilde{U}(u_n \, d\lambda)(x) - U(s)(x) \rangle u_n(t) \, d\lambda(t).$$

Using the Cauchy–Schwarz inequality, we have

$$\| \tilde{U}(u_n \, d\lambda)(x) - U(s)(x) \|^2 = \int \langle U(t)(x) - U(s)(x), \tilde{U}(u_n \, d\lambda)(x) - U(s)(x) \rangle u_n(t) \, d\lambda(t) \leq \int \| U(t)(x) - U(s)(x) \| \| \tilde{U}(u_n \, d\lambda)(x) - U(s)(x) \| u_n(t) \, d\lambda(t) \leq \int \| U(t)(x) - U(s)(x) \| \| \tilde{U}(u_n \, d\lambda)(x) - U(s)(x) \| u_n(t) \, d\lambda(t) \leq \| \tilde{U}(u_n \, d\lambda)(x) - U(s)(x) \| \epsilon \int u_n(t) \, d\lambda(t) = \| \tilde{U}(u_n \, d\lambda)(x) - U(s)(x) \| \epsilon,$$

so we deduce that

$$\| \tilde{U}(u_n \, d\lambda)(x) - U(s)(x) \| \leq \epsilon.$$

If there was some $x \neq 0$ such that $\tilde{U}(f \, d\lambda)(x) = 0$ for all $f \in L^1(G)$, then for $f = u_n$ we would have $U(s)(x) = 0$ for all $s \in G$, which is absurd for $s = e$. Therefore the restriction of $\tilde{U}$ to $L^1(G)$ is nondegenerate.

If $G$ is not metrizable, we have to use a more general neighborhood base and a filter argument.

For simplicity of notation, we write $U \text{ext}(f)$ instead of $U \text{ext}(f \, d\lambda)$. The following converse holds.

**Theorem 11.10.** Let $G$ be a metrizable and locally compact group. For every nondegenerate unitary representation $V : L^1(G) \to L(H)$ of $L^1(G)$, there is a unique unitary representation $U : G \to U(H)$ of the group $G$ such that $V = U \text{ext}$. Consequently, the map $U \mapsto U \text{ext}$ is a bijection between the set of unitary representations of the group $G$ and the set of non-degenerate unitary representations of the involutive Banach algebra $L^1(G)$. Furthermore, a closed subspace $E$ of $H$ is invariant under the linear map $U(s)$ for every $s \in G$ if and only if it is invariant under the linear map $V(f)$ for every $f \in L^1(G)$ (in fact, since $K_R(G)$ is dense in $L^1(G)$, for every $f \in K_R(G)$). Consequently, the map $U \mapsto U \text{ext}$ is a bijection between the set of irreducible unitary representations of $G$ and the set of nondegenerate topologically irreducible unitary representations of $L^1(G)$. If $H = \bigoplus_n H_n$ is a Hilbert sum,
then there is a bijection between the Hilbert sum $U = \bigoplus_n U_n$ of the unitary representations $U_n: G \to \mathbf{H}(H_n)$ and the Hilbert sum $U_{\text{ext}} = \bigoplus_n (U_n)_{\text{ext}}$ of the unitary nondegenerate representations $(U_n)_{\text{ext}}: L^1(G) \to \mathcal{L}(H_n)$. The theorem also holds for any arbitrary locally compact group $G$.

Proof. Theorem 11.10 is proven in Dieudonné [19] (Chapter XXI, Section 1, Theorem 21.1.7). Folland [31] (Chapter, Theorem 3.11) gives a different proof that applies to any locally compact group.

The proof of Theorem 11.9 shows that $U_{\text{ext}}(s)(x)$ is the limit of a sequence $\tilde{U}(u_n d\lambda)(x)$, with $u_n \in \mathcal{K}_\mathbb{R}(G)$, which shows that the map $U \mapsto U_{\text{ext}}$ is injective. It also shows that if the closed subspace $E$ is invariant under the $\tilde{U}(f)$ with $f \in \mathcal{K}_\mathbb{R}(G)$, then it is invariant under the maps $U(s)$ for all $s \in G$. The converse is immediate by definition of $\tilde{U}(\mu)$.

Given a nondegenerate unitary representation $V: L^1(G) \to \mathcal{L}(H)$ of $L^1(G)$, we need to construct a unitary representation $U: G \to \mathbf{U}(H)$ of the group $G$ such that $V = U_{\text{ext}}$.

The idea is that $U(s)(y)$ is the limit of a sequence $V(u_n)(y)$ for a sequence $(u_n)$ of functions that tends to the Dirac delta function at $s$. To make this rigorous, we proceed as follows.

Consider the subspace $E$ of $H$ spanned by the set $\{V(f)(x) \mid f \in \mathcal{L}^1(G), x \in H\}$. Since $V$ is nondegenerate, $E$ is dense in $H$. Pick $s \in G$ and define the open neighborhoods $(V_n)$ of $e$ and the sequence $(u_n)$ of functions $u_n \in \mathcal{K}_\mathbb{R}(G)$ of support contained in $sV_n$, as in the proof of Theorem 11.9. By Proposition 6.49, for every $f \in \mathcal{L}^1(G)$, we have

$$\lim_{n \to \infty} \|u_n \ast f - \delta_s \ast f\|_1 = 0.$$ 

Note that in Proposition 6.49, the functions $f_n$ are defined such that $\int_{G-V} |f_n| d\lambda$ tends to zero as $n$ goes to infinity, for every open subset $V$ containing $e$. Since the $sV_n$ are centered on $s$ rather than $e$, and the function $u_n$ has support contained in $sV_n$, the function $\delta_{s^{-1}} \ast u_n$ has support contained in $V_n$, so Proposition 6.49 shows that $\lim_{n \to \infty} \|\delta_{s^{-1}} \ast u_n \ast f - f\|_1 = 0$, and since

$$\|u_n \ast f - \delta_s \ast f\|_1 = \|\delta_s \ast \delta_{s^{-1}} \ast u_n \ast f - \delta_s \ast f\|_1 \leq \|\delta_s\|_1 \|\delta_{s^{-1}} \ast u_n \ast f - f\|_1,$$

we deduce that $\lim_{n \to \infty} \|u_n \ast f - \delta_s \ast f\|_1 = 0$. Alternatively, we can prove this by going back to the proof of Proposition 6.49.

By Proposition 11.8, we have

$$\|V(f)\| \leq \|f\|_1 \quad \text{for all } f \in \mathcal{L}^1(G). \quad (\ast)$$

Applying $V$ to $u_n \ast f - \delta_s \ast f$, using the above inequality, we get

$$\lim_{n \to \infty} \|V(u_n) \circ V(f) - V(\delta_s \ast f)\| = 0.$$
The above proves that for every linear combination \( y = \sum_{k} V(f_k)(x_k) \in E \), with \( f_k \in \mathcal{L}^1(G) \) and \( x_k \in H \), the sequence \((V(u_n)(y))\) has a limit in \( H \) equal to \( \sum_{k} V(\delta_s \ast f_k)(x_k) \), because

\[
\left\| V(u_n) \left( \sum_{k} V(f_k)(x_k) \right) - \sum_{k} V(\delta_s \ast f_k)(x_k) \right\| \leq \sum_{k} \|V(u_n)(V(f_k)(x_k)) - V(\delta_s \ast f_k)(x_k)\| \\
\leq \sum_{k} \|V(u_n) \circ V(f_k) - V(\delta_s \ast f_k)\| \|x_k\|,
\]

and since

\[
\lim_{n \to \infty} \|V(u_n) \circ V(f_k) - V(\delta_s \ast f_k)\| = 0,
\]

we also have

\[
\lim_{n \to \infty} \left\| V(u_n) \left( \sum_{k} V(f_k)(x_k) \right) - \sum_{k} V(\delta_s \ast f_k)(x_k) \right\| = 0.
\]

Therefore, for \( y = \sum_{k} V(f_k)(x_k) \), we define \( U(s)(y) \) by

\[
U(s)(y) = U(s) \left( \sum_{k} V(f_k)(x_k) \right) = \sum_{k} V(\delta_s \ast f_k)(x_k).
\]

We obtain a linear map \( U(s) \) from \( E \) to \( H \) such that

\[
U(s) \circ V(f) = V(\delta_s \ast f), \quad \text{for all } f \in \mathcal{L}^1(G),
\]

which shows that \( U(s) \) maps \( E \) into itself. By \((*)\), we have \( \|V(u_n)\| \leq \|u_n\|_1 = 1 \), so

\[
\|U(s)(y)\| \leq \|y\| \quad \text{for all } y \in E \text{ and all } s \in G,
\]

and \( U(s) \) extends uniquely to a continuous map on \( H \), also denoted \( U(s) \). What just did also shows that

\[
\|U(s)\| \leq 1 \quad \text{for all } s \in G. \quad \text{(††)}
\]

It remains to prove that the map \( s \mapsto U(s) \) is a unitary representation of \( G \), and that \( V = U_{\text{ext}} \).

By definition, for any \( y = \sum_{k} V(f_k)(x_k) \in E \),

\[
U(s)(y) = \sum_{k} V(\delta_s \ast f_k)(x_k),
\]

and the above expression is continuous in \( s \).

For all \( s, t \in G \) and all \( f \in \mathcal{L}^1(G) \), using \((\dagger)\) we have

\[
U(st) \circ V(f) = V(\delta_{st} \ast f) \\
= V(\delta_s \ast (\delta_t \ast f)) \\
= U(s) \circ V(\delta_t \ast f) \\
= U(s) \circ U(t) \circ V(f),
\]

where \( V(\delta_t \ast f) = V(\delta_t \ast f) \) by \((\dagger)\).
which implies that \( U(st)(y) = U(s)(U(t)(y)) \) for all \( y \in E \), and then by continuity \( U(st) = U(s) \circ U(t) \) in \( L(H) \). By (†), we also have \( U(e) = id_H \).

By (††), we have \( \|U(s)(x)\| \leq \|x\| \) for all \( s \in G \) and all \( x \in H \), so \( \|U(s^{-1})(x)\| \leq \|x\| \), and then \( \|x\| = \|U(s^{-1})(U(s)(x))\| \leq \|U(s)(x)\| \), so \( \|U(s)(x)\| = \|x\| \), and since \( U(s) \) is linear, by the polarization identity for a hermitian inner product, \( U(s) \) is a continuous isometry. Therefore, \( U \) is a unitary representation of \( G \) in \( H \).

To prove that \( V = U_{ext} \), we use the fact that by Theorem 4.49, the dual \( L^1(G)' \) of \( L^1(G) \) is isomorphic to \( L^\infty(G) \). This means that for every continuous form \( \Phi \in L^1(G)' \), there is a unique function \( h \in L^\infty(G) \) such that

\[
\Phi(f) = \int f(s)h(s) \, d\lambda(s) \quad \text{for all } f \in L^1(G).
\]

With some abuse of notation, we write \( \Phi(f) = (h, f) = \int f(s)h(s) \, d\lambda(s) \).

We use the following trick (see Dieudonné [19] (Chapter XXI, Section 1, Theorem 21.1.7). For all \( f, g \in L^1(G) \) and all \( h \in L^\infty(G) \),

\[
(h, f \ast g) = \int f(s)(h, \delta_s \ast g) \, d\lambda(s). \tag{**}
\]

Indeed, using the fact that \( (\delta_s \ast g)(t) = g(s^{-1}t) \) and Fubini’s theorem, we have

\[
(h, f \ast g) = \int (f \ast g)(t)h(t) \, d\lambda(t)
= \int \int f(s)g(s^{-1}t) \, d\lambda(s)h(t) \, d\lambda(t)
= \int \int f(s)(\delta_s \ast g)(t) \, d\lambda(s)h(t) \, d\lambda(t)
= \int f(s)\int (\delta_s \ast g)(t)h(t) \, d\lambda(t) \, d\lambda(s)
= \int f(s)(h, \delta_s \ast g) \, d\lambda(s).
\]

For any fixed pair \( x, y \in H \), the map \( f \mapsto \langle V(f)(x), y \rangle \) is a continuous linear form on \( L^1(G) \), so there is a unique \( h \in L^\infty(G) \) such that \( (h, f) = \langle V(f)(x), y \rangle \), for all \( f \in L^1(G) \), and we get

\[
\langle V(f)(V(g)(x)), y \rangle = \langle V(f \ast g)(x), y \rangle \quad \text{by definition of } \tilde{U}(f).
\]
This proves that \( \langle \tilde{U}(f)(z), y \rangle = \langle V(f)(z), y \rangle \) for all \( z \in E \) and all \( y \in H \), and since \( E \) is dense in \( H \), since \( U_{\text{ext}} \) is the restriction of \( \tilde{U} \) to \( L^1(G) \), we conclude that \( U_{\text{ext}}(f) = V(f) \).

If \( G \) is not metrizable, we have to use a more general neighborhood base and a filter argument.

As an application of Theorem 11.9, we obtain an injective representation of \( L^1(G) \) into \( L^2(G) \) which will be needed in the proof of the Peter–Weyl theorem. It is shown in Dieudonné [22] (Chapter XIV, Section 9, Theorem 14.9.2) that for every \( s \in G \), for any \( f \in L^2(G) \), we have \( \delta_s * f = \lambda_s(f) \in L^2(G) \), and that \( \|\delta_s * f\|_2 = \|f\| \). Consequently, the map \( f \mapsto \delta_s * f = \lambda_s(f) \), denoted \( R(s) \), is a unitary operator on \( L^2(G) \). Furthermore, by Theorem 14.10.6.3 of Dieudonné [22] (Chapter XIV, Section 10), the map \( s \mapsto R(s) \) is a unitary representation of \( G \) in \( L^2(G) \).

**Definition 11.8.** The representation \( R : G \to U(L^2(G)) \) given by

\[
(R(s)(f))(t) = \lambda_s(f)(t) = f(s^{-1}t), \quad f \in L^2(G), \ s, t \in G,
\]

is called the left regular representation of \( G \) in \( L^2(G) \).

By Theorem 11.9, we obtain a representation \( R_{\text{ext}} \) of \( L^1(G) \) in \( L^2(G) \) (a homomorphism from \( L^1(G) \) to \( L(L^2(G)) \)) given by \( R(f)(g) = f * g \), with \( f \in L^1(G) \) and \( g \in L^2(G) \). Using Proposition 6.49, it can be shown that \( R_{\text{ext}} \) is injective, because if \( f * g \) is zero almost everywhere for all \( g \in L^2(G) \), then \( f = 0 \) almost everywhere.

**Definition 11.9.** The representation \( R_{\text{ext}} : L^1(G) \to L(L^2(G)) \) given by

\[
(R_{\text{ext}}(f))(g) = f * g, \quad f \in L^1(G), \ g \in L^2(G),
\]

is called the left regular representation of \( L^1(G) \) in \( L^2(G) \).

### 11.4 Functions of Positive Type and Unitary Representations

There is deep and fruitful connection between topologically cyclic unitary representations \( U : G \to U(H) \) and certain kinds of continuous functions \( p \in C(G; \mathbb{C}) \) called functions of positive type.

Let \( U : G \to U(H) \) be a unitary representation of the locally compact group \( G \) in a Hilbert space \( H \), let \( x_0 \) be any vector in \( H \), and define the map \( p = \psi_{U,x_0} \) by

\[
p(s) = \psi_{U,x_0}(s) = \langle U(s)(x_0), x_0 \rangle, \quad s \in G.
\]

Note that this definition is analogous to Definition 10.11 which involves representations of algebras, but here we are dealing with a group representation. By its very definition the
function \( \psi_{U,x_0} \) is continuous, but it is also bounded, because \( U(s) \) is unitary for every \( s \in G \), so \( \|U(s)(x_0)\| = \|x_0\| \), which implies by Cauchy–Schwarz that

\[
|\psi_{U,x_0}(s)| = |\langle U(s)(x_0), x_0 \rangle| \leq \|U(s)(x_0)\| \|x_0\| \|x_0\|^2 = \psi_{U,x_0}(e),
\]

for all \( s \in G \). Consequently, \( p = \psi_{U,x_0} \in \mathcal{L}^\infty(G; \mathbb{C}) \), so we obtain a continuous linear form \( \omega: \mathcal{L}^1(G; \mathbb{C}) \to \mathbb{C} \) given by

\[
\omega(f) = \int f(s)p(s) \, d\lambda(s) = \int f(s)\langle U(s)(x_0), x_0 \rangle \, d\lambda(s), \quad \text{for all } f \in \mathcal{L}^1(G; \mathbb{C}).
\]

We recognize above the weak integral \( U_{ext}(f) \), so we have

\[
\omega(f) = \int f(s)U(U(s)(x_0), x_0) \, d\lambda(s) = \langle U_{ext}(f)(x_0), x_0 \rangle.
\]

The term on the right-hand side is exactly the term of Definition 10.11, so by Proposition 10.8, \( \omega \) is a positive linear form, which means that \( \omega(f^* * f) \geq 0 \) for all \( f \in \mathcal{L}^1(G; \mathbb{C}) \), that is,

\[
\int (f^* * f)(s)p(s) \, d\lambda(s) \geq 0 \quad \text{for all } f \in \mathcal{L}^1(G; \mathbb{C}).
\]

But \( f^*(s) = \Delta(s^{-1})\overline{f(s)} \), so by changing \( t \) to \( t^{-1} \), by Fubini, and the left invariance of the Haar measure,

\[
\int (f^* * f)(s)p(s) \, d\lambda(s) = \int \int \Delta(t^{-1})\overline{f(t^{-1})}f(t^{-1}s)p(s) \, d\lambda(t) \, d\lambda(s)
= \int \int \overline{f(t)}f(ts)p(s) \, d\lambda(t) \, d\lambda(s)
= \int \int \overline{f(t)}f(ts)p(s) \, d\lambda(s) \, d\lambda(t)
= \int \int p(t^{-1}s)\overline{f(t)}f(s) \, d\lambda(s) \, d\lambda(t).
\]

Observe that we also have

\[
\psi_{U,x_0}(s^{-1}) = \overline{\psi_{U,x_0}(s)},
\]

because

\[
\psi_{U,x_0}(s^{-1}) = \langle U(s^{-1})(x_0), x_0 \rangle = \langle (U(s)^*)(x_0), x_0 \rangle = \langle x_0, U(s)(x_0) \rangle = \overline{\psi_{U,x_0}(s)}.
\]

**Definition 11.10.** If \( G \) is a locally compact group, then a continuous function \( p \in \mathcal{C}(G; \mathbb{C}) \) is of positive type if

\[
\int (f^* * f)(s)p(s) \, d\lambda(s) \geq 0 \quad \text{for all } f \in \mathcal{K}(G),
\]

or equivalently if

\[
\int \int p(t^{-1}s)\overline{f(t)}f(s) \, d\lambda(s) \, d\lambda(t) \geq 0 \quad \text{for all } f \in \mathcal{K}(G).
\]
Every constant function with a nonnegative value is obviously of positive type. For every \( f \in L^2(G; \mathbb{C}) \), we have the left regular representation of \( G \) in \( L^2(G) \) with \((R(s))(f) = \lambda_s f\) (see Definition 11.8), and we have
\[
\langle (R(s))(\bar{f}), \bar{f} \rangle = \int f(s^{-1}t)f(t) d\lambda(t) = \int \bar{f}(t^{-1}s)f(t) d\lambda(t) = (f * \bar{f})(s),
\]
so \( f * \bar{f} = \psi_{R, \bar{f}} \) is of positive type.

We showed that the functions of the form \( \psi_{U, x_0} \) are of positive type. Remarkably, every continuous function \( p \) of positive type determines a unitary topologically cyclic representation \( U \) with a cyclic vector \( x_0 \), such that \( p = \psi_{U, x_0} \).

**Theorem 11.11.** Let \( G \) be a metrizable, separable, locally compact group. For any continuous function \( p \in C(G; \mathbb{C}) \), the following properties are equivalent:

(a) There is a unitary representation \( U: G \to U(H) \) of \( G \) in a separable Hilbert space \( H \) and a vector \( x_0 \in H \) such that \( p = \psi_{U, x_0} \).

(b) The function \( p \) is of positive type, that is,
\[
\int \int p(t^{-1}s)f(s) d\lambda(s) d\lambda(t) \geq 0 \quad \text{for all } f \in K_C(G).
\]

(c) The function \( p \) is bounded by \( p(e) \geq 0, \tilde{p} = p \), and for every complex measure \( \mu \in \mathcal{M}(G) \), we have
\[
\int p(s) d(\tilde{\mu} * \mu)(s) = \int \int p(t^{-1}s) d\tilde{\mu}(t) d\mu(t) \geq 0.
\]

If \( p \) satisfies the above conditions, then there exists a topologically cyclic unitary representation \( V_1 \) of \( G \) in a separable Hilbert space \( H_1 \) and a cyclic vector \( x_1 \) such that \( p = \psi_{V_1, x_1} \). The topologically cyclic representation is unique up to equivalence, in the sense that if \( V_2 \) is another topologically cyclic representation in a separable Hilbert \( H_2 \) and if \( x_2 \) is a cyclic vector for \( V_2 \) such that \( p = \psi_{V_2, x_2} \), then there is an isomorphism \( T: H_1 \to H_2 \) such that \( T(x_1) = x_2 \) and \( V_2 = TV_1T^{-1} \).

**Proof.** We follow Dieudonné [20] (Chapter XXII, Section 1, Theorem 22.1.3). We already proved that (a) implies (b). Let us prove that (b) implies (c).

By Proposition 6.44, we have \( \int \varphi(t) d\mu(t) = \int \varphi(t^{-1}) d\mu(t) \). By the definition of the convolution of measures,
\[
\int p(z) d(\tilde{\mu} * \mu)(z) = \int \int p(ts) d\tilde{\mu}(t) d\mu(s) = \int \int p(t^{-1}s) d\tilde{\mu}(t) d\mu(s).
\]

For every complex measure \( \mu \), the union of all the open sets \( A \) of measure zero (that is, \( \mu(A) = 0 \)) has measure zero, so there is a largest open set of measure zero.
Definition 11.11. The \textit{support} \( \text{supp}(\mu) \) of the measure \( \mu \) is the complement of the largest open set of measure zero.

The support of the measure \( \mu \) has the property that for every \( x \in \text{supp}(\mu) \), for every neighborhood \( V \) of \( x \), there is a continuous function \( f \) with compact support contained in \( V \) such that \( \int f(s) \, d\mu(s) \neq 0 \); see Dieudonné [22] (Chapter XIII, Section 19).

Let us first assume that \( \mu \) has compact support. In this case, for any \( f \in \mathcal{K}_c(G) \), we have \( \mu * f \in \mathcal{K}_c(G) \) (see Dieudonné [22] (Chapter XIV, Section 14, 14.5.4 and 14.9.2), and it follows that

\[
0 \leq \iint p(t^{-1}s)(\mu * f)(t)(\mu * f)(s) \, d\lambda(s) \, d\lambda(t),
\]

and since

\[
(\mu * f)(t) = \int f(x^{-1}t) \, d\mu(x),
\]

we have

\[
\begin{align*}
\iint p(t^{-1}s)(\mu * f)(t)(\mu * f)(s) \, d\lambda(s) \, d\lambda(t) \\
= \iint p(t^{-1}s) \int f(x^{-1}t) \, d\mu(x) \int f(y^{-1}s) \, d\mu(y) \, d\lambda(s) \, d\lambda(t) \\
= \iint p(t^{-1}s) \int f(x^{-1}t) \, d\tilde{\mu}(x) \int f(y^{-1}s) \, d\tilde{\mu}(y) \, d\lambda(s) \, d\lambda(t) \\
= \int \int \left( \int \int p(t^{-1}xy^{-1}s)f(t) \, d\lambda(t) \, f(s) \, d\lambda(s) \right) \, d\tilde{\mu}(x) \, d\tilde{\mu}(y) \\
= \int \int \left( \int \int \tilde{f}(t)f(s)p(t^{-1}x(s^{-1}y)^{-1}) \, d\lambda(t) \, d\lambda(s) \right) \, d\tilde{\mu}(x) \, d\tilde{\mu}(y)
\end{align*}
\]

Define the group \( G \times G \) as the Cartesian product \( G \times G \) with the multiplication

\[(s_1, t_1)(s_2, t_2) = (s_1s_2, t_1t_2).\]

Then the the convolution of the functions \((t, s) \mapsto F(t, s) = \tilde{f}(t)f(s) \) and \((x, y) \mapsto \Pi(x, y) = p(xy^{-1}) \) is given by

\[(x, y) \mapsto \iint F(t, s)\Pi((t^{-1}, s^{-1})(x, y)) \, d\lambda(t) \, d\lambda(s) = \iint F(t, s)\Pi(t^{-1}x, s^{-1}y) \, d\lambda(t) \, d\lambda(s)
= \iint \tilde{f}(t)f(s)p(t^{-1}xy^{-1}s) \, d\lambda(t) \, d\lambda(s).\]

This suggests using the regularization method (Proposition 6.49). Let \( (V_n) \) be a fundamental system of compact neighborhoods of \( e \) such that \( V_{n+1} \subseteq V_n \) for all \( n \), and let \( f_n \) be a
continuous function $f_n \geq 0$ with compact support contained in $V_n$ and such that $\int f_n \, d\lambda = 1$. Since $f_n$ is real, $\overline{f_n} = f_n$. Then if we let $F_n(t,s) = f_n(t) f_n(s)$, we have $
int F_n(t,s) \, d\lambda(t) \, d\lambda(s) = \nint f_n(t) f_n(s) \, d\lambda(t) \, d\lambda(s) = 1$, and by Proposition 6.49, the sequence of functions
\[(F_n * \mathcal{P})(x,y)\]
converges uniformly to the function $(x,y) \mapsto p(xy^{-1})$ on every compact subset. By passing to the limit, using Proposition 13.19.3 of Dieudonné [22] (Chapter XIII, Section 19) which says that on a compact subset we can interchange the integral and the limit, we obtain the inequality
\[\nint p(xy^{-1}) \, d\tilde{\mu}(x) \, d\tilde{\mu}(y) = \nint p(xy) \, d\tilde{\mu}(x) \, d\mu(y) = \int p(s) d(\tilde{\mu} * \mu)(s) \geq 0.\]

Now we show that $p$ is bounded by $p(e)$. For any finite subset $\{s_1, \ldots, s_n\}$ of $G$ and for any complex numbers $\xi_1, \ldots, \xi_n$, the linear functional $\alpha$ on $\mathcal{K}_C(G)$ given by
\[\alpha(f) = \sum_{j=1}^{n} \xi_j f(s_j)\]
is continuous, so by Radon–Riesz III (Theorem 5.29), there is a unique complex measure $\mu$ corresponding to $\alpha$, called an atomic measure. For the measure $\mu$, the inequality in (c) becomes
\[\sum_{j,k} p(s_j^{-1} s_k) \xi_j \xi_k \geq 0.\]
This means that the sesquilinear form $\Phi$ defined by $\Phi(x,y) = \sum_{i,j=1}^{n} p(s_i^{-1} s_j) x_i \overline{y_j}$ must satisfy the property $\Phi(x,x) \geq 0$ for all $x \in \mathbb{C}^n$. Since
\[\Phi(x+y, x+y) = \Phi(x,x) + \Phi(x,y) + \Phi(y,x) + \Phi(y,y),\]
we see that $\Phi(x,y) + \Phi(y,x)$ must be real. By replacing $x$ by $ix$, we see that $i\Phi(x,y) - i\Phi(y,x)$ must be real, so we must have
\[\Phi(y,x) = \overline{\Phi(x,y)}.\]
Therefore, the matrix $(p(s_j^{-1} s_k))$ is hermitian positive semidefinite. In particular, when $n = 2$ and with the set $\{e, s\}$, the matrix
\[
\begin{pmatrix}
p(e) & p(s) \\
p(s^{-1}) & p(e)
\end{pmatrix}
\]
must be hermitian positive semidefinite, which implies that $p(e) \geq 0,$
\[p(s^{-1}) = \overline{p(s)},\]
so $\overline{p} = p,$ and
\[p(e)^2 - p(s)p(s^{-1}) = p(e)^2 - p(s)p(s) \geq 0,\]
and thus
\[ |p(s)| \leq p(e), \quad \text{for all } s \in G, \]
namely, \( p \) is bounded (by \( p(e) \)).

Let us now consider an arbitrary complex measure \( \mu \). By Proposition 15.45, since \( G \) is locally compact and metrizable, there is a sequence \((K_n)\) of compact subsets of \( G \) such that \( K_n \subseteq K_{n+1} \) and \( G = \bigcup_n K_n \). Then it can be shown that \( \lim_{n \to \infty} |\mu|(G - K_n) = 0 \) (see Dieudonné [22], Chapter XIII, Section 8, Proposition 13.8.7). By Radon–Riesz III, the continuous linear functional \( f \mapsto \int \chi_{K_n} f \, d\mu \) (with \( f \in \mathcal{K}(G) \)) corresponds to a measure \( \mu_n \) of compact support \( K_n \). Then it can be shown that \( \lim_{n \to \infty} \|\mu - \mu_n\| = 0 \), and also \( \lim_{n \to \infty} \|\mu - \mu_n\| = 0 \) (see Dieudonné [22], Chapter XIV, Section 6, Proposition 14.6.2). Since \( \mu_n \) has compact support, by our previous result \( \int p(s) d(\mu_n)(s) \geq 0 \), but \( p \) is bounded and by the dominated convergence theorem, \( \int p(s) d(\tilde{\mu}_n \ast \mu_n)(s) \) tends to \( \int p(s) d(\tilde{\mu} \ast \mu)(s) \), and thus \( \int p(s) d(\tilde{\mu} \ast \mu)(s) \geq 0 \).

Finally, we prove that (c) implies (a). Consider the linear form \( \varphi_p \) defined on the unital involutive Banach algebra \( \mathcal{M}^1(G) \) given by
\[
\varphi_p(\mu) = \int p(s) \, d\mu(s), \quad \mu \in \mathcal{M}^1(G),
\]
where \( p \) is a bounded continuous function satisfying (c), and with the involution of \( \mathcal{M}^1(G) \) being given by \( \mu^* = \tilde{\mu} \). Observe that \( \varphi_p(\tilde{\mu} \ast \mu) = \int p(s) d(\tilde{\mu} \ast \mu)(s) \geq 0 \), by (c). Therefore, \( \varphi_p \) is a positive linear form on \( \mathcal{M}^1(G) \), according to Definition 10.10. Recall that \( L^1(G) \oplus \mathbb{C} \delta_e \) is also a unital involutive Banach algebra, and the restriction of \( \varphi_p \) to \( L^1(G) \oplus \mathbb{C} \delta_e \) is also a positive linear form. By Proposition 10.34(1), the linear form \( \varphi_p \) is continuous.

Recall from Proposition 10.9 that \( \varphi_p \) induces a positive Hilbert form \( \gamma \) given by \( \gamma(\mu, \nu) = \varphi_p(\nu^* \ast \mu) \). Then we are almost in the position of applying Proposition 10.33 to obtain a representation of the algebra \( L^1(G) \oplus \mathbb{C} \delta_e \), but it is not clear that Condition (U) is satisfied so we proceed directly.

Proposition 10.31 applies to the positive Hilbert form \( \gamma \). To simplify notation, write \( A = L^1(G) \oplus \mathbb{C} \delta_e \). If
\[
\mathfrak{n} = \{ \mu \in A \mid \gamma(\mu, \mu) = \varphi_p(\mu^* \ast \mu) = 0 \},
\]
then \( \mathfrak{n} \) is a left ideal in \( A \) and \( A/\mathfrak{n} = H_0 \) is a hermitian space with the inner product given by
\[
\langle \pi(\mu), \pi(\nu) \rangle = \gamma(\mu, \nu) = \varphi_p(\nu^* \ast \mu),
\]
where \( \pi: A \to A/\mathfrak{n} = H_0 \) is the quotient map. Observe that by Proposition 10.34, \( \varphi_p \) is a continuous linear form such that \( \|\varphi_p\| = \varphi_p(e) \), so we have
\[
\|\pi(\mu)\|^2 = \gamma(\mu, \mu) = \varphi_p(\mu^* \ast \mu) \leq \varphi_p(e) \|\mu^* \ast \mu\| \leq \varphi_p(e) \|\mu\|^2,
\]
so \( \pi \) is continuous. Since \( A \) is separable, so is \( H_0 \). The completion \( \overline{H} \) of \( H_0 \) is a separable Hilbert space. As in the proof of Proposition 10.33, the endomorphism \( V(\mu) \) given by
\[
V(\mu)(\pi(\nu)) = \pi(\mu \ast \nu) = \pi(\mu) \pi(\nu)
\]
 extends to a continuous map $V(\mu): H \to H$ which is a representation of $A$ (left multiplication). Since $A$ has a unit element $\delta_e$, we see that

$$V(\mu)(\pi(\delta_e)) = \pi(\mu \ast \delta_e) = \pi(\mu),$$

so $x_0 = \pi(\delta_e)$ is a cyclic vector for $V$. Since

$$V(\mu)(x_0) = \pi(\mu),$$

we have

$$\langle V(\mu)(x_0), x_0 \rangle = \langle \pi(\mu), \pi(\delta_e) \rangle = \gamma(\mu, \delta_e) = \varphi_p(\delta_e^* \ast \mu) = \varphi_p(\mu).$$

We also claim that the representation $V$ is nondegenerate. It suffices to prove that the set of elements of the form $f \ast g$ with $f, g \in L^1(G)$ is dense in $L^1(G)$ (property (N)). But this follows immediately by regularization (Proposition 6.49).

We can now apply Theorem 11.10, to obtain a unitary representation $U: G \to \mathbf{U}(H)$, topologically cyclic, and such that $U_{\text{ext}} = V$. We know from (†) in the proof of Theorem 11.10 that $U$ is given by

$$U(s) \circ V(\nu) = V(\delta_s \ast \nu),$$

and since $V(\mu)(x_0) = \pi(\mu)$, this means that

$$U(s)(\pi(\nu)) = \pi(\delta_s \ast \nu).$$

In particular, $U(s)(x_0) = \pi(\delta_s)$. Since $H_U = \{U(s)(x_0) \mid s \in G\}$ is invariant under $U(s)$ for every $s \in G$, by Theorem 11.10, the closed subset $H_U$ is also invariant under $V(\mu)$ for all $\mu \in A$, but $H_0 = \{V(\mu)(x_0) \mid \mu \in A\}$ and $x_0 \in H_U$, so we must have $H_U = H_0$, and $\{U(s)(x_0) \mid s \in G\}$ is dense in $H$. Therefore, $x_0$ is a cyclic vector for $U$, which means that the set $\{\pi(\delta_s) \mid s \in G\}$ is dense in $H$.

We have

$$\langle U(s)(x_0), x_0 \rangle = \langle \pi(\delta_s), \pi(\delta_e) \rangle = \gamma(\delta_s, \delta_e) = \varphi_p(\delta_e^* \ast \delta_s) = \varphi_p(\delta_s) = \int p(t) \, d\delta_s(t) = p(s),$$

so $p = \psi_{U,x_0}$, as desired. The uniqueness of $U$ up to equivalence follows from Proposition 10.32.

We conclude this section with a discussion of the Gelfand–Raikov theorem.

### 11.5 The Gelfand–Raikov Theorem

We will not prove the Gelfand–Raikov theorem but we will prove several technical propositions needed for its proof that are of independent interest.
Proposition 11.12. Let \( p \) be a function of positive type on \( G \). For all \( s, t \in G \), we have
\[
|p(s) - p(t)|^2 \leq 2p(e)(p(e) - \Re(p(s^{-1}t))).
\]

Proof. By Theorem 11.11, we may assume that there is cyclic unitary representation \( U \) and a cyclic vector \( x_0 \) such that
\[
p(s) = \langle U(s)(x_0), x_0 \rangle.
\]
This immediately implies that \( p(e) = \langle U(e)(x_0), x_0 \rangle = \langle x_0, x_0 \rangle = \|x_0\|^2 \). By Cauchy–Schwarz and the fact that \( p(s^{-1}) = \overline{p(s)} \), we have
\[
|p(s) - p(t)|^2 = |\langle (U(s) - U(t))(x_0), x_0 \rangle|^2
\leq \|x_0\|^2 \|(U(s) - U(t))(x_0)\|^2
\leq p(e)(\|U(s)(x_0)\|^2 + \|U(t)(x_0)\|^2 - 2\Re(\langle U(s)(x_0), U(t)(x_0) \rangle))
\leq p(e)(2\|x_0\|^2 - 2\Re(\langle U(t^{-1}s)(x_0), x_0 \rangle))
\leq 2p(e)(p(e) - \Re(p(t^{-1}s)))) = 2p(e)(p(e) - \Re(p(s^{-1}t))))
= 2p(e)(p(e) - \Re(p(s^{-1}t)))),
\]
as claimed. \( \square \)

Let \( p = \psi_{U,x_0} \) be a function of positive type given by a cyclic unitary representation \( U: G \to \text{U}(H) \) with cyclic vector \( x_0 \). If \( (a_n) \) is a Hilbert basis of \( H \) (recall that the Hilbert space \( H \) is separable), the we can write
\[
U(s)(x_0) = \sum_n p_n(s)a_n,
\]
where each function \( p_n \) is continuous and we have
\[
\sum_n |p_n(s)|^2 = \|x_0\|^2, \quad \text{for all } s \in G.
\]
We deduce that
\[
p(s^{-1}t) = \langle U(t)(x_0), U(s)(x_0) \rangle = \sum_n \overline{p_n(s)}p_n(t), \quad (\ast)
\]
with \( \sum_n |p_n(s)p_n(t)| \leq \|x_0\|^2 \).

Proposition 11.13. The product \( pq \) of two functions \( p \) and \( q \) of positive type on \( G \) is a function of positive type.

Proof. Using a Hilbert basis as above, (\ast), and a corollary of the dominated convergence theorem (Proposition 4.36), for every \( f \in K_C(G) \), we have
\[
\int\int p(s^{-1}t)q(s^{-1}t)f(s)f(t) d\lambda(s) d\lambda(t) = \sum_n \int\int q(s^{-1}t)\overline{p_n(s)f(s)p_n(t)f(t)} d\lambda(s) d\lambda(t),
\]
but since \( q \) is also of positive type, we have

\[
\int \int q(s^{-1}t)p_n(s)f(s)p_n(t)f(t) d\lambda(s) d\lambda(t) \geq 0, \quad \text{for all } n,
\]

so

\[
\int \int p(s^{-1}t)q(s^{-1}t)f(s)f(t) d\lambda(s) d\lambda(t) \geq 0,
\]

that is, \( pq \) is of positive type. \( \square \)

Two subsets of the set \( \mathcal{P} \) of continuous functions of positive type on \( G \) come up in the proof of the Gelfand–Raikov theorem and are of particular interest:

\[
\mathcal{P}_1 = \{ f \in \mathcal{P} \mid f(e) = 1 \}
\]
\[
\mathcal{P}_0 = \{ f \in \mathcal{P} \mid 0 \leq f(e) \leq 1 \}.
\]

**Remark:** The above notation is from Folland [31] (Chapter 3). Unfortunately, Dieudonné denotes \( \mathcal{P}_1 \) as \( \mathcal{P}_0 \).

The sets \( \mathcal{P}_0 \) and \( \mathcal{P}_1 \) are convex and bounded (\( \mathcal{P} \) itself is a convex cone). Recall the definition of an extreme point. Given a nonempty convex set \( S \), a point \( a \) of the boundary of \( S \) is extreme (or extremal) if \( S - \{a\} \) is still convex. Equivalently, there does not exist two distinct points \( x, y \in S \) such that \( a = (1 - \lambda)x + \lambda y \), with \( 0 < \lambda < 1 \).

Let \( \mathcal{E}(\mathcal{P}_0) \) (resp. \( \mathcal{E}(\mathcal{P}_1) \)) be the set of extreme points of \( \mathcal{P}_0 \) (resp. \( \mathcal{P}_1 \)). The following results are shown in Folland [31] (Chapter 3, Theorem 3.25 and Lemma 3.26).

**Theorem 11.14.** If \( p \in \mathcal{P}_1 \), then the cyclic unitary representation \( U \) associated with \( p \) given by Theorem 11.11 is irreducible iff \( p \in \mathcal{E}(\mathcal{P}_1) \). We have \( \mathcal{E}(\mathcal{P}_0) = \mathcal{E}(\mathcal{P}_1) \cup \{0\} \).

In order to state the next results we need to define the weak \( * \)-topology on \( L^\infty(G) \). Recall from Theorem 4.49 that \( L^\infty(G) \) is isomorphic to the dual \( (L^1(G))' \) of \( L^1(G) \) under the pairing \( (-,-): L^1(G) \times L^\infty(G) \to \mathbb{C} \) given by

\[
(f, g) = \int f(s)g(s) d\lambda(s).
\]

Every function \( g \in L^\infty(G) \) defines the continuous linear form in \( (L^1(G))' \) given by \( f \mapsto (f, g) \), for every \( f \in L^1(G) \), and every linear form in \( (L^1(G))' \) arises in this fashion for a unique function \( g \in L^\infty(G) \).

**Definition 11.12.** The weak \( * \)-topology on \( L^\infty(G) \) is the topology of pointwise convergence on \( (L^1(G))' \). This topology is defined directly on \( L^\infty(G) \) by the family \( (p_f)_{f \in L^1(G)} \) of seminorms indexed by the set of functions \( L^1(G) \), such that for every \( f \in L^1(G) \),

\[
p_f(g) = |(f, g)|, \quad \text{for every } g \in L^\infty(G).
\]

(See Section 1.6 and Dieudonné [22] (Chapter XII, Section 15).)
It is proven in Folland [31] (Chapter 3, Theorem 3.31) that the topology induced on \( P_1 \subseteq L^\infty(G) \) by the weak\(^*\)-topology of \( L^\infty(G) \) coincides with the topology induced on \( P_1 \) by topology of compact convergence in \( \mathbb{C}^G \) (see Definition 1.9). This result is one of the key facts in the proof of the Gelfand–Raikov theorem.

In Dieudonné [20] (Chapter XXII, Section 1, Theorem 2.1.11), it is shown that the topology induced on \( P_1 \subseteq \mathcal{C}(G; \mathbb{C}) \) by the topology of Fréchet space of \( \mathcal{C}(G; \mathbb{C}) \) (see Section 1.6) and the topology induced on \( P_1 \) by the weak\(^*\)-topology of \( L^\infty(G) \) coincide.

An important theorem due to Gelfand and Raikov shows that there is vast supply of irreducible unitary representations for any locally compact group. This is far from obvious a priori. For example, \( \text{SL}(2, \mathbb{R}) \) does not have finite-dimensional unitary representations, and it is not that easy to find irreducible unitary representations.

**Theorem 11.15. (Gelfand–Raikov)** If \( G \) is an locally compact group, then the irreducible unitary representations of \( G \) separate points. This means that for any \( s, t \in F \), if \( s \neq t \), then there is an irreducible representation \( U \) such that \( U(s) \neq U(t) \).

Theorem 11.15 is proven in Folland [31] (Chapter 3, Theorem 3.34).

The notion of function of positive type is closely related to the notion of positive semidefinite function defined below, which came up during the proof of Theorem 11.11.

**Definition 11.13.** A function \( p : G \to \mathbb{C} \) (not necessarily continuous) is positive semidefinite if for all \( s_1, \ldots, s_n \in G \) and all \( \xi_1, \ldots, \xi_n \in \mathbb{C} \), we have

\[
\sum_{j, k=1}^n p(s_j^{-1}s_k)\xi_k\overline{\xi_j} \geq 0.
\]

As we showed during the proof of Theorem 11.11, the matrix \( (p(s_j^{-1}s_k)) \) is hermitian positive semidefinite. We also have \( p(s^{-1}) = \overline{p(s)} \) and \( |p(s)| \leq p(e) \), so \( p \) is bounded, but examples of discontinuous or even nonmeasurable positive semidefinite \( p \) can be given. However, if \( p \) is continuous, then \( p \) is actually a function of positive type. The following result is shown in Folland [31] (Chapter 3, Proposition 3.35).

**Proposition 11.16.** Let \( G \) be a locally compact group. For any bounded continuous function \( p : G \to \mathbb{C} \), the following are equivalent:

1. The function \( p \) is of positive type.
2. The function \( p \) is positive semidefinite.
3. We have \( \int (f^* f)(s)p(s) \, d\lambda(s) \geq 0 \) for all \( f \in \mathcal{K}_\mathcal{C}(G) \).

**Remark:** If \( p \) is of positive type, then \( \int (f^* f)(s)p(s) \, d\lambda(s) \geq 0 \) for all \( f \in L^1(G) \). Indeed, \( \mathcal{K}_\mathcal{C}(G) \) is dense in \( L^1(G) \), and for any sequence \( (f_n) \) with \( f_n \in \mathcal{K}_\mathcal{C}(G) \) converging to \( f \) in
11.6 Measures of Positive Type and Unitary Representations

As in the previous section, we assume that \( G \) is a separable, metrizable, locally compact group.

For any complex or \( \sigma \)-Radon measure \( \mu \), and any function \( f \in \mathcal{K}_C(G) \), we have

\[
\int (f^* f)(s) \, d\mu(s) = \int \Delta(t^{-1}) \bar{f(t^{-1})} f(t^{-1}s) \, d\lambda(t) \, d\mu(s) = \int \bar{f(t)f(ts)} \, d\lambda(t) \, d\mu(s).
\]

This suggests defining a measure of positive type as follows.

**Definition 11.14.** A complex or \( \sigma \)-Radon measure \( \mu \) is of positive type if

\[
\int (f^* f)(s) \, d\mu(s) = \int \bar{f(t)f(ts)} \, d\lambda(t) \, d\mu(s) \geq 0, \quad \text{for all } f \in \mathcal{K}_C(G).
\]

Observe that if \( \mu = p \, d\lambda \) for some \( p \in L^1(G) \), then

\[
\int (f^* f)(s) \, d\mu(s) = \int \bar{f(t)f(ts)}p(s) \, d\lambda(t) \, d\lambda(s) = \int \bar{f(t)f(s)}p(t^{-1}s) \, d\lambda(t) \, d\lambda(s),
\]

which is exactly the expression defining a function of positive type in Definition 11.10. But here, \( p \in L^1(G) \) is not necessarily continuous, so Definition 11.14 yields a generalization of the notion of function of positive type.

**Proposition 11.17.** For every complex measure \( \nu \in \mathcal{M}^1(G) \), the measure \( \mu = \bar{\nu} * \nu \) is of positive type on \( G \).
Proof. For every function $f \in K_C(G)$, we have

$$
\int (f \ast f)(s)d(\tilde{\nu} \ast \nu)(s) = \int \int \int f(tyz) \lambda(t) d\tilde{\nu}(y) d\nu(z)
$$

$$
= \int \int \int f(ty^{-1}z) \lambda(t) d\tilde{\nu}(y) d\nu(z)
$$

$$
= \int \int \int \Delta(t^{-1})f(t^{-1}y^{-1}z) \lambda(t) d\tilde{\nu}(y) d\nu(z)
$$

$$
= \int \int \int \Delta(t^{-1}y)f(t^{-1}z) d\tilde{\nu}(y) d\nu(z) d\lambda(t)
$$

$$
= \Delta(t^{-1}) \int |f(t^{-1}x) d\nu(x)|^2 d\lambda(t) \geq 0,
$$

since the modular function $\Delta$ is strictly positive. \hfill \square

As a corollary of Proposition 11.17, since $\delta_e = \tilde{\delta}_e = \delta_e \ast \delta_e$, we see that the Dirac measure $\delta_e$ is of positive type.

In the special case where $\nu = f d\lambda$, with $f \in L^1(G)$, we see that $f \ast f$ is of positive type. It should be noted that if $f \in L^1(G)$, the function of positive type $f \ast f \in L^1(G)$ may not be bounded. For example, if $G = \mathbb{R}$ and if $f(x) = x^{-1/2}$ for $0 < x < 1$ and $f(x) = 0$ otherwise, then $f \ast f$ is not bounded.

Observe that $f^*(x) = f(x^{-1}) = \sqrt{x}$ if $x > 1$ and $f^*(x) = 0$ otherwise. We have

$$
g(x) = (f \ast f)(x) = \int_{\mathbb{R}} f(t)f(x-t) dt,
$$

and this integral is not zero if

$$
t > 1, \quad x - 1 < t < x.
$$

If $x \leq 1$, then $g(x) = 0$. If $1 < x \leq 2$, then

$$
g(x) = (f \ast f)(x) = \int_{1}^{x} \frac{\sqrt{t}}{\sqrt{x-t}} dt,
$$

and if $x > 2$, then

$$
g(x) = (f \ast f)(x) = \int_{x-1}^{x} \frac{\sqrt{t}}{\sqrt{x-t}} dt.
$$

Let us compute the integral

$$
\int_{x-1}^{x} \frac{\sqrt{t}}{\sqrt{x-t}} dt = \int_{x-1}^{x} \frac{1}{\sqrt{\frac{t}{t-1}}} dt.
$$
If we do the change of variable

\[ u = \frac{x}{t} - 1, \]

we get

\[ t = \frac{x}{u + 1}, \quad dt = \frac{-t^2}{x} du = -\frac{x du}{(u + 1)^2}, \]

so

\[ \int_{x-1}^{x} \frac{dt}{\sqrt{x t - 1}} = \int_{0}^{x-1} \frac{x du}{\sqrt{u(u + 1)^2}}. \]

Next we make the change of variable

\[ u = w^2, \]

so we have

\[ w = \sqrt{u}, \quad du = 2wdw, \]

and we get

\[ \int_{0}^{x-1} \frac{x du}{\sqrt{u(u + 1)^2}} = 2x \int_{0}^{\frac{1}{\sqrt{x - 1}}} wdw \]

But

\[ \int \frac{dw}{(w^2 + 1)^2} = \int \frac{w^2 + 1 - w^2}{(w^2 + 1)^2} dw = \int \frac{dw}{(w^2 + 1)} - \int \frac{w^2 dw}{(w^2 + 1)^2} = \arctan w - \int w \frac{wdw}{(w^2 + 1)^2}, \]

and by integrating the second term by parts, we get

\[ \int \frac{dw}{(w^2 + 1)^2} = \frac{w}{2(w^2 + 1)} + \frac{1}{2} \arctan w. \]

We finally obtain

\[ g(x) = 2x \left[ \frac{w}{2(w^2 + 1)} + \frac{1}{2} \arctan w \right]_{0}^{\frac{1}{\sqrt{x - 1}}} = \sqrt{x - 1} + x \arctan \left( \frac{1}{\sqrt{x - 1}} \right), \quad x > 2. \]

When \( x > 2 \) goes to infinity, the second term remains positive (in fact, goes to infinity, as we can see by using the power series for \( \arctan y \) with \(|y| < 1\)), and the first term goes to infinity.

For the sake of completeness, if \( 1 < x \leq 2 \), we have

\[ \int_{1}^{x} \frac{dt}{\sqrt{x t - 1}} = \int_{0}^{x-1} \frac{x du}{\sqrt{u(u + 1)^2}}, \]
and
\[
\int_0^{x^{-1}} \frac{xu}{\sqrt{u(u+1)^2}} \, du = 2x \int_0^{\sqrt{x^{-1}}} \frac{dw}{(w^2 + 1)^2}.
\]

It follows that for \(1 < x \leq 2\), we have
\[
g(x) = 2x \left[ \frac{w}{2(w^2 + 1)} + \frac{1}{2} \arctan w \right]_0^{\sqrt{x^{-1}}}
= \sqrt{x - 1} + x \arctan \left( \sqrt{x - 1} \right).
\]

Therefore, the function \(g = f^* * f\) is given by
\[
g(x) = \begin{cases} 
0 & \text{if } x \leq 1 \\
\sqrt{x - 1} + x \arctan \left( \sqrt{x - 1} \right) & \text{if } 1 < x \leq 2 \\
\sqrt{x - 1} + x \arctan \left( \frac{1}{\sqrt{x - 1}} \right) & \text{if } x > 2.
\end{cases}
\]

We showed that a continuous function \(p\) of positive type satisfies the property \(p(s^{-1}) = \overline{p(s)}\), equivalently, \(\tilde{p} = p\). This property generalizes to measures of positive type.

**Proposition 11.18.** For every measure \(\mu\) of positive type on \(G\), we have \(\tilde{\mu} = \mu\).

**Proof.** Observe that if we can prove for all \(f, g \in \mathcal{K}_C(G)\) that
\[
\int (f^* * g)(s) \, d\mu(s) = \int (f^* * g)(s) \, d\tilde{\mu}(s), \quad (\ast)
\]
then by regularization (Proposition 6.49), we will have
\[
\int h(s) \, d\mu(s) = \int h(s) \, d\tilde{\mu}(s), \quad \text{for all } h \in \mathcal{K}_C(G),
\]
which proves that \(\tilde{\mu} = \mu\). By polarization, to prove \((\ast)\), it suffices to prove it for \(g = f\), since
\[
4g^* * f = (f + g)^* * (f + g) - (f - g)^* * (f - g) + i(f + ig)^* * (f + ig) - i(f - ig)^* * (f - ig).
\]
If \(f \in \mathcal{K}_C(G)\), with \(\nu = f \, d\lambda\), by Proposition 11.17 we see that \(f^* * f \in \mathcal{K}_C(G)\) is of positive type, so
\[
\overline{(f^* * f)(s^{-1})} = (f^* * f)(s),
\]
and since \(\mu\) is of positive type we have \(\int (f^* * f)(s) \, d\mu(s) \geq 0\). By definition
\[
\int g(s) \, d\tilde{\mu}(s) = \overline{\tilde{g}(s) \, d\mu(s)},
\]
and we obtain
\[
\int (f^* \ast f)(s) \, d\overline{\mu}(s) = \overline{\int (f^* \ast f)(s^{-1}) \, d\mu(s)} = \int (f^* \ast f)(s) \, d\mu(s) = \int (f^* \ast f)(s) \, d\mu(s),
\]
as claimed. \qed

We conclude this section by showing that a measure \( \mu \) of positive type defines a unitary representation \( U_\mu \) of \( G \). This construction will be used in Section 14.6 to define the Plancherel transform.

The vector space \( \mathcal{K}_C(G) \) is a nonunital algebra under convolution with involution \( f \mapsto f^* \), with \( f^*(s) = \Delta(s^{-1})\overline{f}(s) \). Because \( \mu \) is of positive type, the linear form \( \varphi_\mu : \mathcal{K}_C(G) \to \mathbb{C} \) given by
\[
\varphi_\mu(f) = \int f(s) \, d\mu(s)
\]
is a positive linear form in the sense of Definition 10.10. As in Section 11.4, the set
\[
\mathfrak{n} = \{ f \in \mathcal{K}_C(G) \mid \varphi_\mu(f^* \ast f) = 0 \}
\]
is a left ideal in \( \mathcal{K}_C(G) \), and \( H_0 = \mathcal{K}_C(G)/\mathfrak{n} \) is a hermitian space with the hermitian inner product
\[
\langle \pi(f), \pi(g) \rangle_\mu = \varphi_\mu(g^* \ast f) = \int (g^* \ast f)(s) \, d\mu(s),
\]
where \( \pi : \mathcal{K}_C(G) \to \mathcal{K}_C(G)/\mathfrak{n} \) is the quotient map. Since
\[
\int (g^* \ast f)(s) \, d\mu(s) = \int \int \Delta(t^{-1})g(t^{-1})f(ts) \, d\lambda(t) \, d\mu(s) = \int \int \overline{g(t)}f(ts) \, d\lambda(t) \, d\mu(s),
\]
we have
\[
\langle \pi(f), \pi(g) \rangle_\mu = \varphi_\mu(g^* \ast f) = \int \int \overline{g(t)}f(ts) \, d\lambda(t) \, d\mu(s). \tag{\dagger_2}
\]
We claim that \( H_0 = \mathcal{K}_C(G)/\mathfrak{n} \) is separable. If \( (f_n) \) is a sequence of functions in \( \mathcal{K}_C(G) \) converging uniformly to a function \( f \in \mathcal{K}_C(G) \), with the supports of the \( f_n \) remaining within some fixed compact subset, then \( f_n^* \ast f_n \) converges uniformly to \( f^* \ast f \), the supports of the \( f_n^* \ast f_n \) remaining with some fixed compact set, thus \( \| \pi(f_n) - \pi(f) \|_\mu \) tends to zero as \( n \) tends to infinity. This shows that \( H_0 \) is separable, and we let \( H \) (or \( H_\mu \)) denote the separable Hilbert space which is its completion.
Instead of first defining a nondegenerate representation \( V \) of the algebra \( \mathcal{K}_C(G) \) and then the unitary representation \( U \) of \( G \) such that \( U_{\text{ext}} = V \), we define \( U_\mu : G \rightarrow \text{GL}(H_0) \) as follows:

\[
U_\mu(s)(\pi(f)) = \pi(\delta_s * f), \quad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_C(G). \tag{*_{U_\mu}}
\]

Recall \( (\delta_s * f)(t) = f(s^{-1}t) \), and that if \( f \in \mathcal{K}_C(G) \), then \( \delta_s * f \in \mathcal{K}_C(G) \).

**Theorem 11.19.** For any measure \( \mu \) of positive type, with the notation as above, if \( U_\mu : G \rightarrow \text{GL}(H_0) \) is the map defined by

\[
U_\mu(s)(\pi(f)) = \pi(\delta_s * f), \quad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_C(G),
\]

then each linear map \( U_\mu(s) \) from \( H_0 \) to itself is continuous and unitary, thus the map \( U_\mu(s) \) extends to a unitary map of \( H \), and we obtain a homomorphism \( U_\mu : G \rightarrow \text{U}(H) \). For each \( x \in H \), the map \( s \mapsto U_\mu(s)(x) \) is continuous, therefore, \( U_\mu : G \rightarrow \text{U}(H) \) is a unitary representation of \( G \) in \( H \).

**Proof.** By \((\dagger)_2\), we have

\[
\|U_\mu(s)(\pi(f))\|_\mu^2 = \langle U_\mu(s)(\pi(f)), U_\mu(s)(\pi(f)) \rangle_\mu = \langle \pi(\delta_s * f), \pi(\delta_s * f) \rangle_\mu
\]

\[
= \int \int \overline{f(s^{-1}t)} f(s^{-1}tu) d\lambda(t) d\mu(u)
\]

\[
= \int \int \overline{f(t)} f(tu) d\lambda(t) d\mu(u)
\]

\[
= \langle \pi(f), \pi(f) \rangle_\mu = \|\pi(f)\|_\mu^2.
\]

Thus \( U_\mu(s) \) is unitary and continuous.

For any \( f \in \mathcal{K}_C(G) \), and \( s \in G \), and any sequence \((s_n)\) in \( G \) converging to \( s \in G \), the sequence \((\lambda_{s_n}f)\) converges uniformly to \( \lambda_s f \) (recall that \((\lambda_s f)(t) = f(s^{-1}t))\), the support of each \( \lambda_s f \) remaining in a fixed compact set, so as before, the sequence \((\delta_{s_n} * f)\) converges to \( \delta_s * f \), and since \( U_\mu(s_n)(\pi(f)) = \pi(\delta_{s_n} * f) = \pi(\lambda_{s_n} f) \), the sequence \((U_\mu(s_n)(\pi(f)))\) converges to \( U_\mu(s)(\pi(f)) \in H \). Since \( H_0 \) is dense in \( H \), and since the set of maps \( \{U_\mu(s) \mid s \in G\} \) from \( H \) to itself is equicontinuous (see Proposition 1.13 or Dieudonné [22], Chapter XII, Section 15, Theorem 12.15.7.1), and as a consequence, each map \( s \mapsto U_\mu(s)(x) \) is continuous (see Proposition 1.12 or Dieudonné [23], Chapter VII, Section 5, Theorem 7.5.5).

**Remark:** According to Definition 11.7, the algebra representation \((U_\mu)_{\text{ext}} : L^1(G) \rightarrow \mathcal{L}(H)\) is defined such that for all \( f \in L^1(G) \) and all \( g \in \mathcal{K}_C(G) \), the vector \((U_\mu)_{\text{ext}}(f)(\pi(g))\) is uniquely determined by the equation

\[
\langle (U_\mu)_{\text{ext}}(f)(\pi(g)), \pi(h) \rangle_\mu = \int f(s)(U_\mu(s)(\pi(g)), \pi(h))_\mu d\lambda(s) \quad \text{for all } h \in \mathcal{K}_C(G),
\]
and since 
\[ U_\mu(s)(\pi(g)) = \pi(\delta_s \ast g) \]
and, by \((\dagger)_2\),
\[ \langle U_\mu(s)(\pi(g)), \pi(h) \rangle_\mu = \langle \pi(\delta_s \ast g), \pi(h) \rangle_\mu = \int \int \overline{h(u)}(\delta_s \ast g)(ut) \, d\lambda(u) \, d\mu(t) \]
\[ = \int \int g(s^{-1}ut)\overline{h(u)} \, d\lambda(u) \, d\mu(t), \]
we obtain
\[ \langle (U_\mu)_{\text{ext}}(f)(\pi(g)), \pi(h) \rangle_\mu = \int \int \int f(s)g(s^{-1}ut)\overline{h(u)} \, d\lambda(u) \, d\mu(t) \, d\lambda(s) \]
\[ = \int \int \int f(s)g(s^{-1}ut)\overline{h(u)} \, d\lambda(s) \, d\lambda(u) \, d\mu(t), \]
as in Dieudonné [20] (Chapter XXII, Section 7, no. 22.7.2.1), except that in Dieudonné [20], \(u^{-1}\) occurs instead of \(u\). But Dieudonné assumes that \(G\) is unimodular, so this does not make any difference. To deal with the case where \(G\) is not unimodular, we need to replace \(\overline{f}\) by \(f^* = \Delta^{-1}\overline{f}\), as we did, following Folland [31].

If \(f \in \mathcal{K}_C(G)\), then for all \(g \in \mathcal{K}_C(G)\) we have the simpler expression
\[ (U_\mu)_{\text{ext}}(f)(\pi(g)) = \pi(f \ast g), \]
as in Proposition 10.33.

Representation theory is a vast area of mathematics and we will only give a few references. A classic on the general theory is Kirillov [46]. Kirillov’s survey [47] gives an excellent panorama of the field. Another encyclopedic source that covers a lot of the general theory is Hewitt and Ross [43]. Another good source for the general theory is Folland [31]. Bröcker and tom Dieck [14], Dieudonné [19], and Knapp [49] cover the representation theory of compact groups in great depth. Fulton and Harris [34], Humphreys [44], Knapp [48], Taylor [78], Varadarajan [80, 81], and Vilenkin [82], cover the representation theory of Lie groups.

We are now ready to prove the famous Peter–Weyl theorem.
Chapter 12

Analysis on Compact Groups and
Representations

Chapter 9 is devoted to harmonic analysis on abelian locally compact (not necessarily compact) groups. In this chapter, we consider the case of a compact not necessarily abelian group \( G \). Noncommutativity causes trouble. In particular, the characters no longer form a group. Irreducible representations of the group \( G \) become a substitute for the group characters of abelian groups. Fortunately, compactness also has a positive influence.

The structure of the algebra \( L^2(G) \) is described by a Hilbert sum of finite-dimensional matrix algebras which are representation spaces of irreducible unitary representations of \( G \) (in fact, up to equivalence, all of them). These results constitute a deep and beautiful theorem due to Peter and Weyl, and most of this chapter is devoted to its proof.

If the (metrizable) group \( G \) is compact, then some remarkable things happen:

(1) The involutive algebra (under convolution) \( L^2(G) \) is a complete Hilbert algebra, and as a consequence of Theorem 10.27, the algebra \( L^2(G) \) is a finite or countably infinite Hilbert sum \( \bigoplus_{\rho \in R} a_\rho \) of topologically simple algebras, but because \( G \) is compact, each \( a_\rho \) is isomorphic to a finite-dimensional matrix algebra \( M_{n_\rho}(\mathbb{C}) \). The elements of \( a_\rho \) are continuous functions on \( G \). This is the first half of the first part of a theorem due to Peter and Weyl (1927); see Theorem 12.2.

Since each minimal two-sided ideal \( a_\rho \) is finite-dimensional, it can be expressed as a finite direct sum

\[
a_\rho = \bigoplus_{1 \leq j \leq n_\rho} a_\rho \ast m_j,
\]

of orthogonal minimal left ideals, where the \( m_j \) are self-adjoint irreducible idempotents. We can pick a Hilbert basis \((a_j)_{1 \leq j \leq n_\rho}\) in \( l_1 = a_\rho \ast m_1 \), such that \( a_j \in m_j \ast a_\rho \ast m_1 \), and then it turns out that there is some \( \gamma > 0 \) such that

\[
a_j \ast \bar{a}_j = \gamma m_j, \quad \bar{a}_j \ast a_j = \gamma m_1, \quad 1 \leq j \leq n_\rho.
\]
In fact, we will show that $\gamma = n_\rho^{-1}$. Finally, for all $j, k$ with $1 \leq j, k \leq n_\rho$, let

$$m_{jk} = \gamma^{-1} a_j * \tilde{a}_k,$$

which we also denote by $m^{(\rho)}_{jk}$. We have $m_{jj} = m_j$.

Then the family

$$\left( \frac{1}{\sqrt{n_\rho}} m^{(\rho)}_{ij} \right)_{1 \leq i, j \leq n_\rho, \rho \in \mathbb{R}}$$

is a Hilbert basis of $L^2(G)$.

Furthermore, for every $s \in G$, if we define the $n_\rho \times n_\rho$ matrix $M_\rho(s)$ by

$$M_\rho(s) = \left( \frac{1}{n_\rho} m_{ij}(s) \right),$$

then these matrices are invertible and satisfy the equations

$$M_\rho(st) = M_\rho(s)M_\rho(t) \quad \text{and} \quad M_\rho(s^{-1}) = (M_\rho(s))^*.$$

Thus, the map $s \mapsto M_\rho(s)$ is a continuous unitary representation $M_\rho : G \to U(\mathbb{C}^{n_\rho})$ of $G$ in $\mathbb{C}^{n_\rho}$.

The above results are parts of Theorem 12.6, which constitutes the second half of the part of the Peter–Weyl theorem dealing with the structure of $L^2(G)$ as Hilbert sum of finite-dimensional matrix algebras. But already, representations show their nose.

The unit of every two-sided ideal $a_\rho$ is $u_\rho = \sum_{j=1}^{n_\rho} m_{jj}$, and we show that the center of the Hilbert algebra $L^2(G)$ is the Hilbert sum of the one-dimensional spaces $\mathbb{C}u_\rho$. The above results are shown in Section 12.1.

(2) Besides characters of groups and characters of algebras, there is one more kind of characters, namely, characters of finite-dimensional representations. For every $\rho \in \mathbb{R}$, define the character $\chi_\rho$ of $G$ associated with the ideal $a_\rho$ as the function given by

$$\chi_\rho(s) = \frac{1}{n_\rho} u_\rho(s) = \text{tr}(M_\rho(s)), \quad \text{for all } s \in G.$$

The character $\chi_{\rho_0}$ associated with $a_{\rho_0}$ is the constant function $\chi_{\rho_0}(s) = 1$ for all $s \in G$, called the trivial character of $G$. One of the main property of the characters is that the family of characters $(\chi_\rho)_{\rho \in \mathbb{R}}$ forms a Hilbert basis of the center of $L^2(G)$; see Proposition 12.9. Other properties of the characters $\chi_\rho$ are shown in Section 12.2. In particular, if $G$ is compact and abelian, then the characters are continuous homomorphisms of $G$ to $U(1)$, and they form a Hilbert basis for $L^2(G)$. 
(3) The second part of the Peter–Weyl theorem (Theorem 12.15) deals with unitary representations and is discussed in Section 12.3. This theorem asserts the following facts. Let $V: G \to \text{U}(H)$ be a unitary representation of $G$ in a separable Hilbert space $H$. Then $H$ is a Hilbert sum of subspaces $E_\rho$ invariant under $V$, and each nontrivial $E_\rho$ is the Hilbert sum of invariant subspaces corresponding to irreducible representations of $G$. More precisely:

1. For every $\rho \in \mathcal{R}$, there is an orthogonal projection of $H$ onto a closed subspace $E_\rho$ (which may be reduced to $(0)$), and $H$ is the Hilbert sum of the $E_\rho \neq (0)$.
2. Every subspace $E_\rho \neq (0)$ is invariant under $V$, and the restriction $V_\rho$ of $V$ to $E_\rho$ is a finite or countably infinite Hilbert sum of irreducible representations, all equivalent to $M_\rho$.

In particular, all the representations $M_\rho: G \to \text{U}(\mathbb{C}^{n_\rho})$ occurring in Peter–Weyl I are irreducible, and the index set $\mathcal{R}$ corresponds to the set of equivalence classes of unitary representations of $G$. Now because $G$ is compact, there is a normalized Haar measure $\lambda_G$ on $G$ such that $\lambda_G(G) = 1$, and it can be shown that for any finite-dimensional representation $V: G \to \text{GL}(H)$, there is an inner product on $H$ such that $V$ becomes a unitary representation. Then we define the character $\chi_V$ of the representation $V$ by

$$\chi_V(s) = \text{tr}(V(s)), \quad s \in G.$$ 

Theorem 12.17 shows that two finite-dimensional unitary representations $V_1: G \to \text{U}(H_2)$ and $V_2: G \to \text{U}(H_2)$ of $G$ are equivalent if and only if $\chi_{V_1} = \chi_{V_2}$. This confirms the importance of the characters; they determine the equivalence classes of finite-dimensional representations of a metrizable compact group.

The Fourier transform and the Fourier cotransform can also be generalized, but they involve the unitary irreducible representations of $G$, which are usually very difficult to determine, so they are not so useful in practice.

In Section 12.4, we define a notion of Fourier transform and Fourier cotransform for a (metrizable) compact group $G$. Since for a nonabelian compact group, the set of characters is not a group, the definition of the spaces $L^p(\hat{G})$ is more complicated. The Fourier transform $\mathcal{F}f$ of a function $f \in L^1(G)$ is now a function with domain $\mathcal{R}$, the set of equivalence classes of unitary representations of $G$, such that for every $\rho \in \mathcal{R}$,

$$\mathcal{F}(f)(\rho) = \int f(t)(M_\rho(t))^* \, d\lambda(t).$$

The Fourier transform defined above is the natural generalization of the definition of the Fourier transform when $G$ is an abelian compact group (Definition 9.3),

$$\mathcal{F}(f)(\chi) = \int f(s)\overline{\chi(s)} \, d\lambda(s) = \int f(s)\chi(s^{-1}) \, d\lambda(s);$$
the character \( \chi \) is replaced by the irreducible representation \( M_\rho \).

The definition of \( \mathcal{F}(f)(\rho) \) implies that \( \mathcal{F}(f)(\rho) \) is a linear map from \( \mathbb{C}^{n_\rho} \) to itself (since \( (M_\rho(t))^* \) is a matrix). Thus, \( \mathcal{F}(f) \in \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \). Every element \( F \in \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \) is an \( R \)-indexed sequence \( F = (F(\rho))_{\rho \in R} \) of \( n_\rho \times n_\rho \) matrices \( F(\rho) \). These sequences can be added and rescaled componentwise, so we obtain a vector space.

It is natural to define \( \hat{G} \) as \( R \), but the vector space \( \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \) is too big. Thus, we define some normed vector spaces \( L^p(\hat{G}) \) which are subspaces of \( \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \). For this, we need to define some norms due to von Neumann.

We obtain some Banach spaces \( L^1(\hat{G}) \), \( L^2(\hat{G}) \), and \( L^\infty(\hat{G}) \), and \( L^2(\hat{G}) \) is a Hilbert space. The following result is obtained (Theorem 12.23). Let \( G \) be a compact group.

1. The map \( f \mapsto \mathcal{F}(f) \) is a non norm-increasing injective involutive algebra homomorphism from \( L^1(G) \) into \( L^\infty(\hat{G}) \). In particular, for all \( f, g \in L^1(G) \), for all \( \rho \in R \), we have

\[
(\mathcal{F}(f \ast g))(\rho) = \mathcal{F}(g)(\rho) \circ \mathcal{F}(f)(\rho).
\]

2. For every \( \rho \in R \), the map \( f \mapsto \mathcal{F}(f)(\rho) \) is an algebra representation of \( L^1(G) \) in \( \mathbb{C}^{n_\rho} \).

We also have a version of Plancherel’s theorem (see Theorem 12.26). If \( G \) is a compact group, then the map \( f \mapsto \mathcal{F}(f) \) is an isometric isomorphism between the Hilbert space \( L^2(G) \) and the Hilbert space \( L^2(\hat{G}) \).

We can also define a notion of Fourier co-transform and there are versions of Fourier inversion. For any \( F \in \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \), the Fourier co-transform \( \overline{\mathcal{F}}(F) \) of \( F \) is the function on \( G \) given by

\[
\overline{\mathcal{F}}(F)(s) = \sum_{\rho \in R} n_\rho \text{tr}(F(\rho)M_\rho(s)), \quad s \in G.
\]

Of course, there are convergence issues. It can be shown (Theorem 12.27) that if \( F \in L^1(\hat{G}) \), then the map

\[
s \mapsto (\overline{\mathcal{F}}(F))(s) = \sum_{\rho \in R} n_\rho \text{tr}(F(\rho)M_\rho(s))
\]

converges uniformly to a continuous function \( f \). Furthermore, we have the Fourier inversion formula

\[
(\overline{\mathcal{F}}(\mathcal{F}(f)))(s) = \sum_{\rho \in R} n_\rho \text{tr}(\mathcal{F}(f)(\rho)M_\rho(s)), \quad s \in G.
\]

Also, Fourier inversion holds for \( L^2(G) \). (Theorem 12.29). The Fourier co-transform \( \overline{\mathcal{F}}(F) \) of any \( F \in L^2(\hat{G}) \) converges in the \( \| \cdot \|_2 \)-norm, and for every \( f \in L^2(G) \), we have

\[
f(s) = (\overline{\mathcal{F}}(\mathcal{F}(f)))(s) = \sum_{\rho \in R} n_\rho \text{tr}(\mathcal{F}(f)(\rho)M_\rho(s)), \quad s \in G
\]
in the \( L^2 \)-norm.
12.1 The Peter–Weyl Theorem, I

The theorem below is the first of several theorems describing the structure of the involutive Banach algebra $L^2(G)$, where $G$ is a metrizable compact group. By Proposition 10.13, the Banach algebra $L^2(G)$ is a complete separable Hilbert algebra, so Theorem 10.27 is applicable, and yields most of a deep theorem first proved by Peter and Weyl (1927); see Theorem 12.2.

No matter how it is approached, the proof of the Peter–Weyl theorem (Theorem 12.2) is hard. We follow Dieudonné’s exposition [19] (Sections 1-4). The disadvantage in doing so is that it requires some material on Hilbert algebras from Chapter 10, in particular, Theorem 10.27, whose proof is long. The advantage is that we obtain a sharper and more informative version of the Peter–Weyl theorem.

Since $G$ is compact, it is unimodular, and so it has a Haar measure $\lambda$ which is both left and right invariant. We also assume that $\lambda$ is normalized so that $\lambda(G) = 1$.

When we describe operations on elements of $L^2(G)$, such as $f * g$ for the convolution $[f] *[g]$ of $[f]$ and $[g]$ in $L^2(G)$, we mean the equivalence class $[f * g]$ of $f * g$, where $f, g \in L^2(G)$ are representatives in the equivalence classes $[f], [g] \in L^2(G)$ (where two functions are equivalent iff they are equal almost everywhere). To be perfectly rigorous, we should check that these constructions do not depend on the representatives chosen in these equivalence classes, but we will not inflict such verifications on the reader.

Recall that by Proposition 6.48, if $f, g \in L^2(G)$, then $f * g \in C_0(G; \mathbb{C})$. In particular, $f * g$ is continuous.

**Definition 12.1.** A function $h \in L^2(G)$ is central if its class in $L^2(G)$ belongs to the center of $L^2(G)$. This means that for every $f \in L^2(G)$, we have $f * h = h * f$ almost everywhere.

The following auxiliary result is needed.

**Proposition 12.1.** Let $G$ be a compact group. A continuous function $h \in L^2(G)$ is central if and only if $h(sts^{-1}) = h(t)$ for all $s, t \in G$. The class of every central function $f \in L^2(G)$ belongs to the center of $M^1(G)$.

**Proof.** If $f * h = h * f$ almost everywhere, since $f * h$ and $h * f$ are continuous, we must have $f * h = h * f$ everywhere. Since $G$ is compact, it is unimodular, so have

$$f * h(s) = \int f(t) h(t^{-1}s) \, d\lambda(t) = \int f(t^{-1}) h(ts) \, d\lambda(t)$$

and

$$h * f(s) = \int h(t) f(t^{-1}s) \, d\lambda(t) = \int h(st) f(t^{-1}) \, d\lambda(t),$$

Thus, for every $s \in G$, we have

$$\int_G f(t^{-1})(h(st) - h(ts)) \, d\lambda(t) = 0.$$
The above implies that $h(st) = h(ts)$ for all $t$ in the complement of a set of measure zero (depending on $s$), but since $h$ is continuous, this subset must be empty. It follows that $h(st) = h(ts)$ for all $s, t \in G$, and if we replace $t$ by $ts^{-1}$, we obtain $h(sts^{-1}) = h(t)$ for all $s, t \in G$.

If we let $t = t's$, then the condition $h(sts^{-1}) = h(t)$ is equivalent to $h(st') = h(t's)$. By the formula

$$(\mu * f)(s) = \int f(t^{-1}s) \, d\mu(t)$$

in Definition 6.22, and the formula

$$(f * \mu)(s) = \int f(st^{-1})\Delta(t^{-1}) \, d\mu(t)$$

in Definition 6.23, since $G$ is compact, it is unimodular, so $\Delta(t^{-1}) = 1$, and if $f$ is a central function, then $\mu * f = f * \mu$, which shows that $f$ is in the center of $\mathcal{M}^1(G)$. 

**Theorem 12.2. (Peter–Weyl theorem, I) Let $G$ be a metrizable compact group. The complete Hilbert algebra $L^2(G)$ is the Hilbert sum of a finite or countably infinite family $(a_\rho)_{\rho \in \mathbb{R}}$ of topologically simple algebras $a_\rho$ of finite dimension $n_\rho^2$, where each $a_\rho$ is a minimal two-sided ideal of $L^2(G)$ isomorphic to a matrix algebra $M_{n_\rho}(\mathbb{C})$. The elements of $a_\rho$ are classes of continuous functions on $G$; the unit element of $a_\rho$ is the class of a continuous function $u_\rho$ such that $\delta_\rho = u_\rho$, and the orthogonal projection of $L^2(G)$ onto $a_\rho$ is the map $f \mapsto f * u_\rho = u_\rho * f$, for every $f \in L^2(G)$. Furthermore, for every $f \in L^2(G)$, we have

$$f = \sum_{\rho \in \mathbb{R}} f * u_\rho,$$

where the series on the right-hand side is commutatively convergent.**

**Proof.** We follow Dieudonné’s proof [19] (Chapter XXI, Section 2, Theorem 21.2.3). Since By Proposition 10.13, the Banach algebra $A = L^2(G)$ is a complete separable Hilbert algebra, Theorem 10.27 shows that $L^2(G)$ is the Hilbert sum of a finite or countably infinite family $(a_\rho)_{\rho \in \mathbb{R}}$ of two-sided ideals which are topologically simple Hilbert algebras, and $a_\rho a_k = (0)$ for all $h \neq k$. If we can prove that every $a_\rho$ is finite-dimensional, we will be done because then, by Theorem 10.28, each $a_\rho$ will be a finite Hilbert sum of isomorphic minimal left ideals $I_j$, each generated by a self-adjoint irreducible idempotent $e_j$, and the sum of these idempotents will be the unit $1_\rho$ of the algebra $a_\rho$. If $v$ is a function whose class is $1_\rho$, every element of $a_\rho$ will be the class of a function of the form $f * v$, with $f \in L^2(G)$, which is a continuous function by Proposition 6.48. The other assertions of the theorem follow from Theorem 10.24.

By Proposition 10.29, if we can prove that there is a nonzero element in the center of $a_\rho$, then $a_\rho$ will be finite-dimensional. This is a consequence of the following proposition.

**Proposition 12.3.** For every closed two-sided ideal $b \neq (0)$ in $L^2(G)$, there is some nonzero element $c \in b$ in the center of $L^2(G)$.**
Proof. The proof of Proposition 12.3 makes use of the following result.

**Proposition 12.4.** Let $b$ be a closed subspace (as a vector space) of $L^2(G)$. Then the following conditions are equivalent.

1. $b$ is a left ideal in $L^2(G)$.
2. $b$ is invariant under the regular representation $R_{\text{ext}}$ of $L^1(G)$ in $L^2(G)$ (see Definition 11.9).
3. For every function $f$ whose class is in $b$ and for all $s \in G$, the class of $\delta_s * f = \lambda_s(f)$ is in $b$.

**Proof.** The equivalence of (2) and (3) follows from Theorem 11.10 applied to the regular representation $R_{\text{ext}}$. It is clear that (2) implies (1). On the other hand, Theorem 5.13 implies that $L^2(G)$ is dense in $L^1(G)$, and by Proposition 6.47, the map $f \mapsto f * g$ from $L^1(G)$ to $L^2(G)$ is continuous for every $g \in L^2(G)$, so (1) implies (2).

Proposition 12.4 also applies to right ideals in (1) and to $f \delta_s = \rho_{s^{-1}}(f)$ in (3).

We can now prove Proposition 12.3. First, let us prove that $b$ contains the class of a continuous function $f$ such that $f$ is not the zero function (we need a continuous function, because we want to construct a central function, and such a function must be continuous). Indeed, let $g \in b$ be a function not zero almost everywhere. Then the class of the function $g \tilde{\ast} g$ also belongs to $b$. But $g \tilde{\ast} g$ is continuous (by Proposition 6.48), and since the definition of the convolution of functions implies that

\[(g \tilde{\ast} g)(e) = \int g(s)\tilde{g}(s^{-1}) d\lambda(s) = \int g(s)\tilde{g}(s) d\lambda(s) = \|g\|_2^2 > 0,\]

we can pick $f = g \tilde{\ast} g$ in $b$. Consider the function $h$ given by

\[h(t) = \int_G f(sts^{-1}) d\lambda(s).\]

Since the function $(x, y, z) \mapsto f(xyz)$ is uniformly continuous on $G \times G \times G$, we see that $h$ is continuous, and since $h(e) = f(e) \neq 0$, it is not the zero function. For all $s \in G$, we have

\[h(xtx^{-1}) = \int_G f((sx)t(sx)^{-1}) d\lambda(s) = h(t),\]

since the Haar measure on a compact group is left and right invariant. By Proposition 12.1, the function $h$ belongs to the center of $L^2(G)$. It remains to show that the class of $h$ belongs to $b$. Since $L^2(G) = b \oplus b^\perp$ as a Hilbert sum, and $b^\perp$ is also a two-sided ideal by Proposition 10.15, it suffices to prove that $\langle h, w \rangle = 0$ for all $w \in b^\perp$ (we are abusing notation, $h$ and
$w$ should be equivalence classes). Using the fact that the Haar measure is left and right invariant, and Fubini’s theorem, we have

\[
\langle h, w \rangle = \int \overline{w(t)} \int f(sts^{-1}) d\lambda(s) d\lambda(t)
\]

\[
= \int \left( \int \overline{w(t)} f(sts^{-1}) d\lambda(t) \right) d\lambda(s)
\]

\[
= \int \left( \int \overline{w(s^{-1}ts)} f(t) d\lambda(t) \right) d\lambda(s)
\]

Since $w \in b^\perp$, by Proposition 12.4 and its version for right ideals, the class of $\delta_s * w * \delta_{s^{-1}}$ also belongs to $b^\perp$. Since $G$ is unimodular (see after Definition 6.22 and Definition 6.23), we have

\[
(\delta_s * w * \delta_{s^{-1}})(t) = w(s^{-1}ts),
\]

and since $f \in b$ and $\delta_s * w * \delta_{s^{-1}} \in b^\perp$,

\[
\int \overline{w(s^{-1}ts)} f(t) d\lambda(t) = 0,
\]

which concludes the proof of Proposition 12.3.\qed

This also concludes the proof of Theorem 12.2.\qed

We will identify every element of $a_\rho$ with the unique continuous function belonging to this class.

Our next goal is to get a better understanding of the structure of the algebras $a_\rho$ by decomposing them as finite Hilbert sums of minimal left ideals, and by choosing some Hilbert bases in these ideals.

For every $\rho \in R$, we assume that we have chosen a decomposition of $a_\rho$ as a finite Hilbert sum of $n_\rho$ minimal left ideals $l_j = a_\rho * m_j$,

\[
a_\rho = \bigoplus_{1 \leq j \leq n_\rho} a_\rho * m_j,
\]

where the $a_\rho * m_j$, also denoted $l_\rho^{(j)}$, are pairwise isomorphic and orthogonal, and where $m_j$ is a self-adjoint irreducible idempotent ($1 \leq j \leq n_\rho$), so that the unit of $a_\rho$ is

\[
u_\rho = \sum_{j=1}^{n_\rho} m_j;
\]

see Theorem 10.28.\footnote{Note that the $m_j$ are the $e_j$ used in the proof of Theorem 10.28.} By Proposition 10.17, since $m_j$ and $m_j$ are orthogonal when $i \neq j$, we have $m_i * m_j = 0$ if $i \neq j$. Let $(a_j)_{1 \leq j \leq n_\rho}$ be a Hilbert basis of $l_1 = a_\rho * m_1$, such that $a_j \in m_j * a_\rho * m_1$.\footnote{Note that the $m_j$ are the $e_j$ used in the proof of Theorem 10.28.}
Observe that $a_i \ast \tilde{a}_j = \tilde{a}_j \ast a_i = 0$ whenever $i \neq j$. Since $a_j \in m_j \ast a_{\rho} \ast m_1$, we also have

$$a_j \ast m_1 = a_j, \quad m_j \ast a_j = a_j, \quad 1 \leq j \leq n_{\rho}.$$ 

We have the following proposition, which is in fact part of the proof of Theorem 10.28. To simplify notation, write $a^* = \tilde{a}$.

**Proposition 12.5.** The inner products $\langle m_j, m_j \rangle$ have the same value $\gamma > 0$, and we have

$$a_j \ast a_j^* = \gamma m_j, \quad a_j^* \ast a_j = \gamma m_1, \quad 1 \leq j \leq n_{\rho}.$$ 

**Proof.** Since the $m_j$ are self-adjoint idempotent $(m_j \ast m_j = m_j$ and $m_j^* = m_j)$ and $(a \ast b)^* = b^* \ast a^*$, as $a_j \in m_j \ast a_{\rho} \ast m_1$, we have $a_j \ast a_j^* \in m_j \ast a_{\rho} \ast m_1$. By Theorem 10.26(2), we must have $a_j \ast a_j^* = \lambda_j m_j$, for some $\lambda_j \in \mathbb{C}$, with $\lambda_j \neq 0$. Similarly, $a_j^* \ast a_j \in m_1 \ast a_{\rho} \ast m_1$, so $a_j^* \ast a_j = \lambda_j' m_1$ for some $\lambda_j' \in \mathbb{C}$, with $\lambda_j' \neq 0$. We claim that $\lambda_j = \lambda_j'$.

First, we have

$$a_j \ast a_j^* \ast a_j \ast a_j^* = \lambda_j m_j \ast \lambda_j m_j = \lambda_j^2 m_j \ast m_j = \lambda_j^2 m_j,$$ 

and second

$$a_j \ast a_j^* \ast a_j \ast a_j^* = a_j \ast \lambda_j' m_1 \ast a_j^* = \lambda_j' a_j \ast m_1 \ast a_j^* = \lambda_j' a_j \ast a_j^* = \lambda_j' \lambda_j m_j,$$

since $a_j \ast m_1 = a_j$. Therefore, $\lambda_j^2 = \lambda_j' \lambda_j$, and since $\lambda_j$ and $\lambda_j'$ are nonzero, we deduce that $\lambda_j = \lambda_j'$, for $j = 1, \ldots, n_{\rho}$.

We also have

$$1 = \langle a_j, a_j \rangle = \langle a_j, m_j \ast a_j \rangle = \langle a_j \ast a_j^*, m_j \rangle = \lambda_j \langle m_j, m_j \rangle,$$

and

$$1 = \langle a_j, a_j \rangle = \langle a_j, a_j \ast m_1 \rangle = \langle a_j^* \ast a_j, m_1 \rangle = \lambda_j \langle m_1, m_1 \rangle.$$

Since $\lambda_j \neq 0$, we deduce that

$$\langle m_j, m_j \rangle = \langle m_1, m_1 \rangle, \quad 1 \leq j \leq n_{\rho}.$$ 

and so

$$\lambda_j = \langle m_1, m_1 \rangle^{-1} = \gamma, \quad 1 \leq j \leq n_{\rho}. \quad \square$$

Thus there is some $\gamma > 0$ such that

$$a_j \ast \tilde{a}_j = \gamma m_j, \quad \tilde{a}_j \ast a_j = \gamma m_1, \quad 1 \leq j \leq n_{\rho},$$

**Definition 12.2.** For all $j, k$ with $1 \leq j, k \leq n_{\rho}$, let

$$m_{jk} = \gamma^{-1} a_j \ast \tilde{a}_k.$$
Then we have $m_{jj} = m_j$, and since the $a_k$ are self-adjoint two sided-ideals such that $a_k a_h = (0)$ whenever $h \neq k$, we have $\tilde{a}_k * a_h = 0$ whenever $k \neq h$, and so

$$m_{jk} * a_h = \delta_{kh} a_j. \quad (*)$$

**Remark:** Observe that the $m_{ij}$ are the $e_{mn}$ introduced during the proof of Theorem 10.28.

We will also write $m_{ij}^{(\rho)}$ instead of $m_{ij}$. The following result reveals that some representations are hidden in the Hilbert sum of the $\tilde{a}_\rho$.

**Theorem 12.6.** With the above notation, the following properties hold.

1. For every $i$ with $1 \leq i \leq n_\rho$, the $(m_{ij})_{1 \leq j \leq n_\rho}$ form an orthogonal basis of $l_i$.

2. We have $m_{ji} = \bar{m}_{ij}$ and $m_{ij} * m_{kk} = \delta_{ij} m_{ik}$.

3. We have $(m_{ij}, m_{ij}) = n_\rho$ and $m_{ij}(e) = n_\rho \delta_{ij}$, for all $i, j$ with $1 \leq i, j \leq n_\rho$ (in other words, $\gamma = (n_\rho)^{-1}$). Thus the family of functions

$$\left( \frac{1}{\sqrt{n_\rho}} m_{ij}^{(\rho)} \right)_{1 \leq i, j \leq n_\rho, \rho \in \mathbb{R}}$$

is a Hilbert basis of $L^2(G)$.

4. For every $s \in G$, if we define the $n_\rho \times n_\rho$ matrix $M_\rho(s)$ by

$$M_\rho(s) = \left( \frac{1}{n_\rho} m_{ij}(s) \right),$$

then these matrices are invertible and satisfy the equations

$$M_\rho(s t) = M_\rho(s) M_\rho(t) \quad \text{and} \quad M_\rho(s^{-1}) = (M_\rho(s))^*.$$

Thus, the map $s \mapsto M_\rho(s)$ is a continuous unitary representation $M_\rho : G \to U(\mathbb{C}^{n_\rho})$ of $G$ in $\mathbb{C}^{n_\rho}$, for the standard hermitian inner product $\sum_{j=1}^{n_\rho} \alpha_j \overline{\beta_j}$.

**Proof.** (1) Since $m_{ij} = \gamma^{-1} a_i * \tilde{a}_j = \gamma^{-1} a_i \gamma a_j^*$, it suffices to prove that for any fixed $i$,

$$\langle a_i \gamma a_j^*, a_j \gamma a_k^* \rangle = 0 \quad \text{for all } j \neq k.$$

Since $a_i^* a_i = \gamma m_1$, $m_1^* = m_1$, and $a_k * m_1 = a_k$, we have

$$\langle a_i \gamma a_j^*, a_k \gamma a_k^* \rangle = \langle a_j^*, a_i^* a_i a_k^* \rangle$$

$$= \langle a_j^*, \gamma m_1 a_k^* \rangle$$

$$= \gamma \langle a_k * m_1, a_j \rangle$$

$$= \gamma \langle a_k, a_j \rangle = 0.$$
since the \( a_j \) are pairwise orthogonal.

(2) We have
\[
m_{ji}(s) = \gamma^{-1}(a_j \ast \tilde{\alpha}_i)(s)
\]
\[
= \gamma^{-1} \int a_j(t) \tilde{\alpha}_i(t^{-1} s) d\lambda(t)
\]
\[
= \gamma^{-1} \int a_j(t) a_i(s^{-1} t) d\lambda(t)
\]
\[
= \gamma^{-1} \int a_i(t) a_j(st) d\lambda(t)
\]
\[
= \gamma^{-1} \int a_i(t) \tilde{a}_j(t^{-1} s^{-1}) d\lambda(t)
\]
\[
= \gamma^{-1} (a_i \ast \tilde{\alpha}_j)(s^{-1})
\]
\[
= \tilde{m}_{ij}(s^{-1}) = \tilde{m}_{ij}(s).
\]

Since \( a_i \in m_i \ast a_p \ast m_1 \), we have \( a_i^* \ast a_j \in m_1 \ast a_p \ast m_i \ast m_j \ast a_p \ast m_1 \) whenever \( i \neq j \), since the \( m_i \) are pairwise orthogonal self-adjoint irreducible idempotents, so \( m_i \ast m_j = 0 \) whenever \( i \neq j \). Consequently
\[
a_j^* \ast a_h = \delta_{jh} \gamma m_1,
\]
and thus
\[
m_{ij} \ast m_{hk} = \gamma^{-1} a_i \ast a_j^* \ast \gamma^{-1} a_h \ast a_k^* = \gamma^{-2} a_i \ast \delta_{jh} \gamma m_1 \ast a_k^* = \delta_{jh} \gamma^{-1} a_i \ast a_k^* = \delta_{jh} m_{ik}.
\]

(3) Since \( a_p \) is a Hilbert algebra, by (2) and (2') (see Definition 10.14), we have
\[
\langle m_{ij}, m_{ij} \rangle = \gamma^{-2} \langle a_i \ast \tilde{\alpha}_j, a_i \ast \tilde{\alpha}_j \rangle = \gamma^{-2} \langle \tilde{\alpha}_i \ast a_i, \tilde{\alpha}_j \ast a_j \rangle = \langle m_1, m_1 \rangle.
\]
To compute this value, observe that for every \( k \), by Proposition 12.4, the function \( t \mapsto m_{ik}(st) \) belongs to \( l_k \) for all \( s \in G \). Thus we can write
\[
m_{ik}(st) = \sum_{j=1}^{n_p} c_{ij}(s)m_{jk}(t).
\]

On the other hand, using the fact that \( m_{ji} = \tilde{m}_{ij} \) and (2), we have
\[
m_{jk}(t) = (m_{j1} \ast m_{1k})(t) = \int_G m_{j1}(tx)m_{1k}(x^{-1}) d\lambda(x) = \int_G m_{j1}(tx)\overline{m_{1k}(x)} d\lambda(x),
\]
which yields
\[
m_{jk}(e) = \langle m_{j1}, m_{k1} \rangle,
\]
and if we let $t = e$ in ($*_{2}$), using the orthogonality properties of the $m_{ij}$ and the fact that $m_{jj} = m_{j}$ and $\langle m_{j}, m_{j} \rangle = \langle m_{1}, m_{1} \rangle$, we get

$$m_{ik}(s) = \langle m_{1}, m_{1} \rangle c_{ik}(s). \quad (*_{3})$$

If we let $s = t^{-1}$ and $i = k = 1$ in ($*_{2}$) and $j = k = 1$ in $m_{jk}(e) = \langle m_{j_{1}}, m_{k_{1}} \rangle$, we get

$$\langle m_{1}, m_{1} \rangle = m_{1}(e) = \sum_{j=1}^{n_{\rho}} c_{1j}(s)m_{j_{1}}(s^{-1}) = \sum_{j=1}^{n_{\rho}} c_{1j}(s)m_{j_{1}}(s),$$

and using ($*_{3}$),

$$c_{1j} = \frac{1}{\langle m_{1}, m_{1} \rangle} m_{1j},$$

we obtain

$$\sum_{j=1}^{n_{\rho}} m_{1j}(s)m_{j_{1}}(s) = \langle m_{1}, m_{1} \rangle^{2}.$$}

If we integrate both sides of the above in $G$, since

$$\langle m_{1}, m_{1} \rangle = \langle m_{1,j}, m_{1,j} \rangle = \int m_{1j}(s)m_{j_{1}}(s) \, d\lambda(s),$$

we finally get

$$\langle m_{1}, m_{1} \rangle^{2} = n_{\rho}\langle m_{1}, m_{1} \rangle,$$

so $\langle m_{ij}, m_{ij} \rangle = \langle m_{1}, m_{1} \rangle = n_{\rho}$, which proves (3). The equations in (4) follow immediately from (2), ($*_{2}$), and ($*_{3}$).

We will see later that the representations $M_{\rho}$ are irreducible. The center of $L^{2}(G)$ is characterized as follows.

**Proposition 12.7.** Let $G$ be a metrizable compact group. The center of the Hilbert algebra $L^{2}(G)$ is the Hilbert sum of the one-dimensional spaces $\mathbb{C}u_{\rho}$ (with $\rho \in \mathbb{R}$). In particular, if $G$ is commutative, then every ideal $a_{\rho}$ is one-dimensional ($n_{\rho} = 1$).

**Proof.** Since $u_{\rho}$ is the unit element of $a_{\rho}$ and since $a_{\rho} * a_{\rho'} = (0)$ whenever $\rho \neq \rho'$, we see that $u_{\rho}$ belongs to the center of $L^{2}(G)$. If the class of a function $f \in L^{2}(G)$ belongs to the center of $L^{2}(G)$, since $u_{\rho}$ also belongs to this center, we deduce that the class of $f * u_{\rho} \in a_{\rho}$ belongs to the center $L^{2}(G)$, but since $a_{\rho}$ is topologically simple, complete, separable algebra, by Proposition 10.29, the center of $a_{\rho}$ is one-dimensional, so $f * u_{\rho} = c_{\rho}u_{\rho}$ for some $c_{\rho} \in \mathbb{C}$. Since by Theorem 12.2, we have

$$f = \sum_{\rho \in R} f * u_{\rho}$$

for every $f \in L^{2}(G)$, we must have $f = \sum_{\rho \in R} c_{\rho}u_{\rho}$, as claimed. \(\square\)
Since the group $G$ is compact, for every $f \in L^2(G)$ and every constant function $\alpha$, we have
\[
f \ast \alpha = \alpha \ast f = \alpha \left( \int_G f(s) \, d\lambda(s) \right).
\]
Therefore the (complex) constant functions form a two-sided ideal in $L^2(G)$, and thus must be an ideal of the form $a_{\rho_0}$.  

**Definition 12.3.** The ideal $a_{\rho_0}$ is called the **trivial ideal**.

The corresponding representation $M_{\rho_0}$ is one-dimensional, and $M_{\rho_0}(s) = 1$ for all $s \in G$. In other words, $M_{\rho_0}$ is the trivial representation of $G$. For all $\rho \neq \rho_0$, since the spaces $a_\rho$ and $a_{\rho_0}$ are orthogonal, we have
\[
\int_G m_{ij}^{(\rho)}(s) \, d\lambda(s) = \int_G m_{ij}^{(\rho)}(s) \overline{1} \, d\lambda(s) = 0.
\]
Therefore, for every $\rho \neq \rho_0$, we have
\[
\int_G m_{ij}^{(\rho)}(s) \, d\lambda(s) = 0.
\]

We also have the following results.

**Proposition 12.8.** Let $G$ be a metrizable compact group. With the notation as above, the following properties hold.

1. If $f$ and $g$ are two continuous functions in $C(G; \mathbb{C})$, then we have
\[
f \ast g = \sum_{\rho \in \mathcal{R}} \left( \sum_{1 \leq i, j \leq n_\rho} \frac{1}{n_\rho} \langle g, m_{ij}^{(\rho)} \rangle \langle f \ast m_{ij}^{(\rho)} \rangle \right),
\]
where the family on the right-hand side converges for the topology of uniform convergence.

2. The family of continuous functions $\{m_{ij}^{(\rho)} \mid \rho \in \mathcal{R}, 1 \leq i, j \leq n_\rho\}$ is an orthogonal system that is dense in $C(G; \mathbb{C})$ for the topology of uniform convergence.

**Proof.** (1) From Theorem 12.6 which says that the family of functions
\[
\left( \frac{1}{\sqrt{n_\rho}} m_{ij}^{(\rho)} \right)_{1 \leq i, j \leq n_\rho, \rho \in \mathcal{R}}
\]
is a Hilbert basis of $L^2(G)$, we have
\[
g = \sum_{\rho \in \mathcal{R}} \sum_{1 \leq i, j \leq n_\rho} \frac{1}{n_\rho} \langle g, m_{ij}^{(\rho)} \rangle m_{ij}^{(\rho)}.
\]
Since the map \( u_\rho \mapsto f * u_\rho \) is a continuous map from \( L^2(G) \) to \( \mathcal{C}(G; \mathbb{C}) \), (since by Proposition 6.48, \( \| f * u_\rho \|_\infty \leq \| f \|_2 \| u_\rho \|_2 \)), we can apply convolution to both sides, and we get the equation in (1).

(2) By Proposition 6.49, for every continuous function \( g \in \mathcal{C}(G; \mathbb{C}) \), there is some continuous function \( f \) such that \( \| f \ast g - g \|_\infty \) can be made arbitrarily small. But, for every \( \rho \in \mathbb{R} \), the functions \( f \ast m_{ij}^{(\rho)} \) belong to the two-sided ideal \( a_\rho \), and so they are (complex) linear combinations of the \( m_{hk}^{(\rho)} \) with \( 1 \leq h, k \leq n_\rho \). Therefore, the formula of (1) for \( f \ast g \) shows that \( f \ast g \) can be expressed in terms of the \( m_{hk}^{(\rho)} \), which proves (2).

12.2 Characters of Compact Groups

Besides characters of groups and characters of algebras, there is one more kind of characters, namely, characters of finite-dimensional representations. As in the previous section, we work with metrizable compact groups. Since we have the Peter–Weyl theorem and Theorem 12.6 at our disposal, it will be fairly easy to prove the properties of characters of these groups.

Definition 12.4. Let \( G \) be a metrizable compact group. With the notations of Section 12.2, for every \( \rho \in \mathbb{R} \), define the character \( \chi_\rho \) of \( G \) associated with the ideal \( a_\rho \) as the function given by

\[
\chi_\rho(s) = \frac{1}{n_\rho} u_\rho(s) = \frac{1}{n_\rho} \sum_{j=1}^{n_\rho} m_{jj}^{(\rho)}(s) = \text{tr}(M_\rho(s)), \quad \text{for all } s \in G.
\]

The character \( \chi_{\rho_0} \) associated with \( a_{\rho_0} \) is the constant function \( \chi_{\rho_0}(s) = 1 \) for all \( s \in G \), called the trivial character of \( G \).

The properties stated in the following proposition are immediate consequences of Theorem 12.2 and Theorem 12.6.

Proposition 12.9. The following properties hold.

1. Every character \( \chi_\rho \) is a continuous central function, which means that

\[
\chi_\rho(sts^{-1}) = \chi_\rho(s) \quad \text{for all } s, t \in G.
\]

2. We have

\[
\chi_\rho(s^{-1}) = \overline{\chi_\rho(s)} \quad \text{for all } s \in G.
\]

3. We have

\[
\chi_\rho \ast \chi_{\rho'} = 0 \quad \text{whenever } \rho \neq \rho', \quad \text{and} \quad \chi_\rho \ast \chi_\rho = \frac{1}{n_\rho} \chi_\rho.
\]

4. The family of characters \( (\chi_\rho)_{\rho \in \mathbb{R}} \) forms a Hilbert basis of the center of \( L^2(G) \), which means that:
(a) We have
\[ \langle \chi_\rho, \chi_{\rho'} \rangle = \int \chi_\rho(s) \overline{\chi_{\rho'}(s)} d\lambda(s) = 0 \quad \text{whenever } \rho \neq \rho' \]
\[ \langle \chi_\rho, \chi_\rho \rangle = \int |\chi_\rho(s)|^2 d\lambda(s) = 1. \]

(b) For every central function \( f \in L^2(G) \),
\[ f = \sum_{\rho \in R} \langle f, \chi_\rho \rangle \chi_\rho = \sum_{\rho \in R} n_\rho (f \ast \chi_\rho). \]

(c) We have
\[ \int \chi_\rho(s) d\lambda(s) = 0 \quad \text{for all } \rho \neq \rho_0. \]

(d) For all \( s \in G \), we have
\[ \chi_\rho(s) = \text{tr}(M_\rho(s)), \]
and
\[ \chi_\rho(e) = n_\rho. \]

Observe that, unlike the characters of a locally compact abelian group \( G \), which take their values in \( U(1) \cong T \), the characters \( \chi_\rho \) of a compact not necessarily abelian group \( G \) take their values in \( \mathbb{C} \). For instance \( \chi_\rho(e) = n_\rho \), and in general, \( n_\rho > 1 \).

The next proposition is needed to prove Theorem 12.11.

**Proposition 12.10.** For any two continuous central functions \( f, g \) in \( C(G; \mathbb{C}) \), we have
\[ f \ast g = \sum_{\rho \in R} \langle g, \chi_\rho \rangle (f \ast \chi_\rho), \]
where the family on the right-hand side converges in the topology of uniform convergence.

**Proof.** This follows from the fact shown in Proposition 12.9(4) that the family of characters \( (\chi_\rho)_{\rho \in R} \) is a Hilbert basis of the center of \( L^2(G) \), and the fact that \( \|f \ast g\|_\infty \leq \|f\|_2 \|g\|_2 \).

The next theorem will require an auxiliary proposition.

**Theorem 12.11.** The family of continuous central functions \( (\chi_\rho)_{\rho \in R} \) constitutes an orthonormal system which is dense in the space of continuous central functions in \( C(G; \mathbb{C}) \) for the topology of uniform convergence.

**Proof.** For every continuous central function \( f \), since \( \chi_\rho = \frac{1}{n_\rho} u_\rho \), by Proposition 12.7, the function \( f \ast \chi_\rho \) is a scalar multiple of \( \chi_\rho \). In view of the formula of Proposition 12.10, it suffices to show that for every continuous central function \( g \), there exists a continuous central function \( f \) such that \( \|f \ast g - g\|_\infty \) is arbitrarily small. Recall that for any element \( t \in G \) the inner automorphism \( C_t \) is defined by \( C_t(s) = tst^{-1} \), for all \( s \in G \). The following result is needed.
Proposition 12.12. The following properties hold.

(1) Let $G$ be a metrizable topological group, and let $K$ be a compact subset of $G$. For every neighborhood $U$ of the identity element $e$ of $G$, there is a neighborhood $V \subseteq U$ of $e$ such that, $tVt^{-1} \subseteq U$ for all $t \in K$.

(2) In a metrizable compact group $G$, there exists a neighborhood base of neighborhoods of $e$ invariant under all inner automorphisms of $G$. For such a neighborhood $T$, there exists a continuous central function $h \geq 0$, of support contained in $T$, such that $\int_G h(s) d\lambda(s) = 1$.

Proof. (1) Let $U_0$ be a neighborhood of $e$ such that $U_0^3 \subseteq U_0$. For every $s \in G$, there is a neighborhood $V_s$ of $e$ in $G$ such that $sV_s^{-1} \subseteq U_0$; by continuity, there exists a neighborhood $W_s$ of $s$ such that for all $t \in W_s$, we also have $tV_s t^{-1} \subseteq U$. Since $K$ is compact, there exists a finite number of $m$ elements $s_j \in K$ such that $W_{s_1} \cup \cdots \cup W_{s_m}$ covers $K$. If we let $V = \bigcap_{j=1}^m V_{s_j}$, we have $tVt^{-1} \subseteq U$ for all $t \in K$.

(2) Apply (1) with $K = G$. Then

$$T = \bigcup_{t \in G} tVt^{-1}$$

is a neighborhood of $e$ contained in $U$, obviously invariant under the inner automorphisms of $G$. To define $h$, start with a continuous function $f \geq 0$ of support contained in $T$, such that $f(e) > 0$. Let

$$h(t) = c \int_G f(sts^{-1}) d\lambda(s),$$

where the constant $c > 0$ is chosen in a suitable way, and then the proof that $h$ works is the same as in the proof of Theorem 12.2. This finishes the proof of Proposition 12.12. □

In view of Proposition 12.12(2), Proposition 6.49(i) shows that for every continuous central function $g$, there is some continuous central function $h$ such that $\|g * h - g\|_\infty$ can be made arbitrarily small, which finishes the proof of Theorem 12.11. □

The following result shows certain independence results.

Theorem 12.13. The following properties hold.

(1) For every $s \in G$, if $s \neq e$, then there is some $\rho \in R$ such that $\chi_\rho(s) \neq \chi_\rho(e)$.

(2) We have

$$\bigcap_{\rho \in R} N_\rho = \{e\},$$

where $N_\rho$ is the kernel of the (group) homomorphism $s \mapsto M_\rho(s)$.
12.2. CHARACTERS OF COMPACT GROUPS

Proof. (1) If there is some $s \neq e$ such that $\chi_\rho(s) = \chi_\rho(e)$ for all $\rho \in \mathbb{R}$, then by Theorem 12.11 we would have $f(s) = f(e)$ for all continuous central functions, but this contradicts Proposition 12.12(2), since we can find a continuous central function $h$ with $h(e) \neq 0$ whose support $T$ does not contain $s \neq 0$.

(2) If $s \in N_\rho$, then since $\chi_\rho(s) = \text{tr}(M_\rho(s))$, we have $\chi_\rho(s) = \chi_\rho(e)$, and (1) implies that

$$ \bigcap_{\rho \in \mathbb{R}} N_\rho = \{e\}, $$

as claimed. \qed

The following is a product formula for the characters.

**Proposition 12.14.** For every character $\chi$ of $G$ we have

$$ \chi(s)\chi(t) = \chi(e) \int_G \chi(usu^{-1}t) d\lambda(u). $$

Proof. By definition, and since $M_\rho(st) = M_\rho(s)M_\rho(t)$ for all $s, t \in G$, we have

$$ \chi_\rho(usu^{-1}t) = \frac{1}{n_\rho} \sum_i m_{ii}^{(\rho)}(usu^{-1}t) = \frac{1}{n_\rho} \sum_{i,j,h,k} m_{ij}(u)m_{jh}(s)m_{hk}(u^{-1})m_{ki}(t). $$

Using Theorem 12.6 (2) and (3), it follows that

$$ \int_G \chi_\rho(usu^{-1}t) d\lambda(u) = \frac{1}{n_\rho^4} \sum_{i,j,h,k} m_{jh}(s)m_{ki}(t) \int m_{ij}(u)m_{hk}(u^{-1}) d\lambda(u) $$

$$ = \frac{1}{n_\rho^3} \sum_{i,j,h,k} \delta_{jk}\delta_{ik}m_{jh}(s)m_{ki}(t) $$

$$ = \frac{1}{n_\rho^3} \sum_{i,j} m_{jj}(s)m_{ii}(t) $$

$$ = \frac{1}{n_\rho} \chi_\rho(s)\chi_\rho(t), $$

as claimed. \qed

It is clear that the function which maps the class of a function $f \in L^2(G)$ to the class of its complex conjugate $\overline{f}$ is a semilinear bijection of $L^2(G)$, and an automorphism of its ring structure.

**Definition 12.5.** The above automorphism of $L^2(G)$ maps every ideal $a_\rho$ into a minimal two-sided ideal that we denote by $a_\overline{\rho}$. 
If (as usual), given a complex matrix $X = (x_{ij})$, we denote by $\overline{X}$ the matrix $(\overline{x}_{ij})$, then we have

$$M_\rho(s) = \overline{M_\rho(s)} \quad \text{for all } s \in G.$$  

The corresponding characters are such that

$$\chi_\rho = \overline{\chi_\rho}.$$  

Thus the equation $a_\rho = a_\rho$ is equivalent to saying that the character $\chi_\rho$ only takes real values.

Let us now consider the special cases where either $G$ is compact and abelian, or $G$ is finite.

**Example 12.1.** We will consider the case where $G$ is a compact (metrizable) abelian group, but before doing this, let $G$ be a not necessarily abelian compact group. Let $f \in L^2(G)$ be a function not zero everywhere such that for every $s \in G$,

$$f(st) = f(s)f(t) \quad \text{for almost all } t \in G.$$  

Then

$$f(s^{-1}t) = f(s^{-1})f(t),$$

so the subgroup $\mathbb{C}f$ of $L^2(G)$ is invariant under the map $g \mapsto \lambda_s(g)$ for every $s \in G$, and by Proposition 12.4, this subgroup if a closed minimal left ideal of dimension 1. This is only possible if this left ideal is one of the $a_\rho$ for which $n_\rho = 1$, and then $f$ is equal to the character $\chi_\rho$ almost everywhere. Such characters are called *abelian characters*. The above reasoning shows that these are the only continuous homomorphisms of $G$ into $\mathbb{C}^\ast$. Since the image of $G$ by such a character $\chi_\rho$ is a compact subgroup of $\mathbb{C}^\ast$, it must be contained in $U(1)$.

If $G$ is compact and abelian, then every character is obviously abelian, since the algebras $a_\rho$ are commutative. Then by Theorem 12.6(3), the characters of $G$ form a Hilbert basis of $L^2(G)$, and every continuous function is a uniform limit of linear combinations of characters (by Theorem 12.11). We have determined the characters of several compact abelian groups, such as $\mathbb{Z}/n\mathbb{Z}$ and $\mathbb{T}^n$, in Proposition 9.8 and Corollary 9.10.

**Example 12.2.** Let $G$ be a finite (not necessarily abelian) group of order $|G| = g$. In this case, the algebras $L^1(G)$ and $L^2(G)$ are the same and equal to the algebra $\mathbb{C}[G] = [G \to \mathbb{C}]$ of formal linear combinations $\sum_{s \in G} x_s s$ with $x_s \in \mathbb{C}$, with the convolution

$$\frac{1}{g} \left( \sum_{s_1 \in G} x_{s_1} s_1 \right) * \left( \sum_{s_2 \in G} y_{s_2} s_2 \right) = \frac{1}{g} \sum_{s \in G} \left( \sum_{s_1 \cdot s_2 = s} x_{s_1} y_{s_2} \right) s = \frac{1}{g} \sum_{s \in G} \left( \sum_{t \in G} x_{ts^{-1}} y_{t^{-1}s} \right) s.$$  

Recall that two elements $a, b \in G$ are *conjugate* if $b = sas^{-1}$ for some $s \in G$. Conjugation is an equivalence relation in $G$, and its classes, called the *conjugacy classes* of $G$, are the sets

$$C_a = \{sas^{-1} \mid s \in G\}.$$
Since $G$ is finite, it has finite number $r$ of conjugacy classes $C_1, \ldots, C_r$, and we assume that $C_1 = \{e\}$.

The central functions (also called class functions) are constant on the conjugacy classes. Also, since $G$ is finite, by Theorem 12.11, every central function is a linear combination of characters, and since they are linearly independent, the number $r$ of conjugacy classes is equal to the number of characters, and to the dimension of the center of the algebra $Z(\mathbb{C}[G])$ of $\mathbb{C}[G]$.

Let $R = \{\rho_1, \ldots, \rho_r\}$, and write $\chi_{ij}$ for the value of the character $\chi_i$ on the conjugacy class $C_j$ $(1 \leq i, j \leq r)$. If $g$ is the order of the group $G$ and $h_j$ is the number of elements in the conjugacy class $C_j$, then the orthogonality relations in Proposition 12.9(4) become

\[
\frac{1}{g} \sum_{k=1}^{r} h_k \chi_{ik} \overline{\chi_{jk}} = \delta_{ij} \quad (1 \leq i, j \leq r).
\]

In other words, the matrix

\[
\left( \frac{h_k}{g} \chi_{ik} \right)_{1 \leq i, k \leq r}
\]

is unitary. We get more identities by expressing that the transpose of the above matrix is unitary, namely:

\[
\sum_{i=1}^{r} \chi_{ik} \overline{\chi_{il}} = 0 \quad \text{if } k \neq l
\]

and

\[
\sum_{i=1}^{r} |\chi_{ik}|^2 = \frac{g}{h_k}.
\]

These formulae can also be written as

\[
\sum_{\rho \in R} \chi_{\rho}(s) \chi_{\rho}(t^{-1}) = 0 \quad (\dagger_1)
\]

if $s$ and $t$ are not conjugate in $G$, and

\[
\sum_{\rho \in R} |\chi_{\rho}(s)|^2 = \frac{g}{h_k}, \quad \text{if } s \in C_k. \quad (\dagger_2)
\]

Since $e$ is not conjugate to any other element in $G$, if we let $t = e$ in $(\dagger_1)$, using the fact that $\chi_{\rho}(e) = n_{\rho}$, we obtain

\[
\sum_{\rho \in R} n_{\rho} \chi_{\rho}(s) = 0 \quad \text{if } s \neq e.
\]

If we let $s = e$ in $(\dagger_2)$, we obtain the relation

\[
\sum_{\rho \in R} n_{\rho}^2 = g. \quad (\dagger_3)
\]

The above equation confirms that $L^2(G)$ is the direct sum of the $\mathfrak{a}_\rho$.

For much more about the representation of finite groups, see Serre [76].
12.3 The Peter–Weyl Theorem, II

In this section we prove the second part of the Peter–Weyl theorem which has to do with unitary representations. In particular, we prove the important result that the unitary representations $s \mapsto M_{\rho}(s)$ of $G$ discussed in Theorem 12.6 are irreducible (in fact, all of them, up to equivalence).

Given a unitary representation $V: G \to U(H)$, recall from Definition 11.7 specialized to measures of the form $f d\lambda$, where $f \in L^1(G)$, that for every $x \in H$, the unique vector $	ilde{V}(f d\lambda)(x) \in H$ such that

$$\langle \tilde{V}(f d\lambda)(x), y \rangle = \int \langle V(s)(x), y \rangle f(s) d\lambda(s)$$

for all $y \in H$

is called the weak integral of the function $s \mapsto V(s)(x)$ from $G$ to $H$ with respect to $f d\lambda$, and is denoted by

$$\int V(s)(x) f(x) d\lambda(s).$$

We write $V_{\text{ext}}(f)$ instead of $\tilde{V}(f d\lambda)$. We know from by Proposition 11.8 that $\|V_{\text{ext}}(f)\| \leq \|f\|_1$.

Theorem 12.15. (Peter–Weyl theorem, II) Let $G$ be a metrizable compact group, and let $V: G \to U(H)$ be a unitary representation of $G$ in a separable Hilbert space $H$.

1. For every $\rho \in \mathcal{R}$, the map $V_{\text{ext}}(u_{\rho})$ is an orthogonal projection of $H$ onto a closed subspace $E_{\rho}$ (which may be reduced to $(0)$), and $H$ is the Hilbert sum of the $E_{\rho} \neq (0)$.

2. Every subspace $E_{\rho} \neq (0)$ is invariant under $V$, and the restriction $V_{\rho}$ of $V$ to $E_{\rho}$ is a finite or countably infinite Hilbert sum of irreducible representations, all equivalent to $M_{\rho}$.

Proof. (1) By the Peter–Weyl theorem (Theorem 12.2), the $u_{\rho}$ are self-adjoint idempotents, that is, $u_{\rho} * u_{\rho} = u_{\rho}$ and $u_{\rho}^* = \overline{u}_\rho = u_{\rho}$, and since $V_{\text{ext}}: L^1(G) \to \mathcal{L}(H)$ is an algebra homomorphism, we have

$$V_{\text{ext}}(u_{\rho}) = V_{\text{ext}}(u_{\rho} * u_{\rho}) = V_{\text{ext}}(u_{\rho}) \circ V_{\text{ext}}(u_{\rho}),$$

and

$$V_{\text{ext}}(u_{\rho}) = V_{\text{ext}}(u_{\rho}^*) = (V_{\text{ext}}(u_{\rho}))^*,$$

so the continuous linear map $V_{\text{ext}}(u_{\rho})$ is idempotent and hermitian, and by Proposition 10.2, it is an orthogonal projection. Since $u_{\rho} * u_{\rho'} = 0$ if $\rho \neq \rho'$, we have $V_{\text{ext}}(u_{\rho}) \circ V_{\text{ext}}(u_{\rho'}) = 0,$
which implies that the images $E_\rho$ of the projections $V_{\text{ext}}(u_\rho)$ are closed subspaces that are pairwise orthogonal. To prove that $H$ is the Hilbert sum of the family $(E_\rho)_{\rho \in R}$, we need to show that the algebraic direct sum $\bigoplus_{\rho \in R} E_\rho$ is dense in $H$. We know from the proof of Theorem 11.10 that the linear span $E$ of the set $\{V_{\text{ext}}(f)(x) \mid f \in C(G; \mathbb{C}), x \in H\}$ is dense in $H$, and by Proposition 11.8, we have $\|V_{\text{ext}}(f)\| \leq \|f\|_1$. By Proposition 12.8(2), if $f$ is continuous, then for every $\epsilon > 0$, there is a (finite) linear combination $\sum_{i,j,\rho} c_{ij}^{(\rho)} m_{ij}^{(\rho)}$ such that

$$\left\| f - \sum_{i,j,\rho} c_{ij}^{(\rho)} m_{ij}^{(\rho)} \right\|_\infty \leq \epsilon,$$

and since $\|V_{\text{ext}}(f)\| \leq \|f\|_1$, $G$ is compact, and $\lambda(G) = 1$, we have $\|g\|_1 \leq \|g\|_\infty$ for any $g \in L^1(G)$, and this implies that

$$\left\| V_{\text{ext}}(f) - \sum_{i,j,\rho} c_{ij}^{(\rho)} V_{\text{ext}}(m_{ij}^{(\rho)}) \right\|_1 \leq \left\| f - \sum_{i,j,\rho} c_{ij}^{(\rho)} m_{ij}^{(\rho)} \right\|_1 \leq \left\| f - \sum_{i,j,\rho} c_{ij}^{(\rho)} m_{ij}^{(\rho)} \right\|_\infty \leq \epsilon.$$

Since $m_{ij}^{(\rho)} = u_\rho \ast m_{ij}^{(\rho)}$, we have $V_{\text{ext}}(m_{ij}^{(\rho)}) = V_{\text{ext}}(u_\rho) \circ V_{\text{ext}}(m_{ij}^{(\rho)})$, thus the vector $(\sum_{i,j,\rho} c_{ij}^{(\rho)} V_{\text{ext}}(m_{ij}^{(\rho)}))(x)$ belongs to the sum of the $E_\rho$, which proves that $E$ is dense in $H$.

We delete the summands $E_\rho$ such that $E_\rho = (0)$.

2

Step 1. We prove that the subspaces $E_\rho$ which are invariant under $U_{\text{ext}}$ are also invariant under $U$. Since by Proposition 12.7, the $u_\rho$ belong to the center of $L^2(G)$, by Proposition 12.1, the $u_\rho$ belong to the center of $\mathcal{M}^1(G)$, and since

$$V(s)(V_{\text{ext}}(f)(x)) = V_{\text{ext}}(\delta_s \ast f)(x), \quad \text{for all } f \in L^1(G), \text{ and all } x \in H, \quad (\dagger)$$

from the proof of Theorem 11.10, as any $E_\rho \neq (0)$ is the image of $V_{\text{ext}}(u_\rho)$, we have

$$V(s)(V_{\text{ext}}(u_\rho)(x)) = V_{\text{ext}}(\delta_s \ast u_\rho)(x) = V_{\text{ext}}(u_\rho \ast \delta_s)(x) = V_{\text{ext}}(u_\rho)(V_{\text{ext}}(\delta_s)(x)) = V_{\text{ext}}(u_\rho)(V(s)(x)),$$

we conclude that the $E_\rho$ are invariant under $V$.

Step 2. If $V_\rho$ is the restriction of the representation $V$ of $G$ to $E_\rho$ we obtain a Hilbert sum decomposition into irreducible representations for the restriction of the nondegenerate representation $(V_\rho)_{\text{ext}}$ to $a_\rho$ in $E_\rho$.

If $V_\rho$ is the restriction of the representation $V$ of $G$ to $E_\rho$, then $(V_\rho)_{\text{ext}}(u_{\rho'}) = 0$ for all $\rho' \neq \rho$, since $u_\rho \ast u_{\rho'} = 0$. The representation $(V_\rho)_{\text{ext}}$ is a representation of $L^1(G)$ in $E_\rho$, and since $L^2(G) \subseteq L^1(G)$ is the Hilbert sum $L^2(G) = \bigoplus_{\rho \in R} a_\rho$, and the projection of $L^2(G)$
onto $\mathfrak{a}_\rho$ is the map $f \mapsto f * u_\rho$, we see that the restriction of $(V_\rho)_{\text{ext}}$ to $L^2(G)$ can be viewed as a nondegenerate representation of the algebra $\mathfrak{a}_\rho$ in $E_\rho$. Since by Proposition 11.8, the representation $(V_\rho)_{\text{ext}}$ is continuous, by Theorem 10.30, the nondegenerate representation $(V_\rho)_{\text{ext}}$ of $\mathfrak{a}_\rho$ in $E_\rho$ is a finite or countably infinite Hilbert sum of topologically irreducible representations all equivalent to the representation $U_{l_1}$.

*Step 3.* We convert the Hilbert sum decomposition into irreducible representations of the nondegenerate representation $(V_\rho)_{\text{ext}}$ of $\mathfrak{a}_\rho$ (extended to $L^1(G)$ by 0 on the other summands $\mathfrak{a}_{\rho'}$) into a Hilbert sum decomposition of the representation $V_\rho$ of $G$ into irreducible representations, and find explicitly what these representations are.

By Theorem 11.10, the representation $V_\rho$ of $G$ in $E_\rho$ is also a finite or countably infinite Hilbert sum of irreducible representations all equivalent to a representation of $G$ in $l_1$. It remains to figure out what this representation is.

By definition of $U_{l_1}$ in Proposition 10.14, namely

$$U_{l_1}(f)(g) = f * g, \quad f \in \mathfrak{a}_\rho, \ g \in l_1,$$

and of the left regular representation $R_{\text{ext}}$, of $L^1(G)$ in $L^2(G)$, given in Definition 11.9 by

$$(R_{\text{ext}}(f))(g) = f * g, \quad f \in L^1(G), \ g \in L^2(G),$$

we see that $U_{l_1}$ is the restriction of $R_{\text{ext}}$ to $\mathfrak{a}_\rho$. Thus the irreducible representation of $G$ in $l_1$ is the restriction of $R$, the left regular representation of $G$ in $L^2(G)$, to $l_1$.

We can figure out the matrix of $R(s)$ with respect to the basis of $l_1$ consisting of the vectors $(\frac{1}{n_\rho} m_1)_{1 \leq i \leq n_\rho}$. Using Theorem 12.6(4), we have

$$R(s)(m_j) = m_j(s^{-1}t) = \frac{1}{n_\rho} \sum_{i=1}^{n_\rho} m_{ji}(s^{-1}) m_i(t)$$

and we recognize that the matrix of $R(s)$ is

$$(M_\rho(s^{-1}))^T = M_\rho(s) = M_\pi(s),$$

as claimed. \hfill \square

Let us emphasize that Theorem 12.15 proves that every representation $M_\rho$ is irreducible, which is not at all obvious from their definition.

If the compact group $G$ is abelian, then every algebra $\mathfrak{a}_\rho$ is abelian, and since it is simple, it must be one-dimensional. Therefore, every unitary representation of a metrizable compact abelian group is a finite or a countably infinite Hilbert sum of one-dimensional representations.

It is customary to introduce the following terminology.
12.3. THE PETER–WEYL THEOREM, II

Definition 12.6. With the notations of Theorem 12.15, if \( V: G \to U(H) \) is a unitary representation of \( G \) in a separable Hilbert space \( H \), and if \( H = \bigoplus_{\rho \in R} E_{\rho} \) is the Hilbert sum induced by the projections \( V_{\text{ext}}(u_{\rho}) \), whenever \( E_{\rho} \neq (0) \), we say that the irreducible representation \( M_{\rho} \) is contained in the representation \( V \). If \( E_{\rho} \) is finite-dimensional of dimension \( d_{\rho}n_{\rho} > 0 \) we say that \( M_{\rho} \) is contained \( d_{\rho} \) times in \( V \) (or infinitely many times if \( E_{\rho} \) is infinite-dimensional). We also call \( d_{\rho} \) the multiplicity of \( M_{\rho} \) in \( V \). The representations \( M_{\rho} \) such that \( d_{\rho} > 0 \) are called the irreducible components of the representation \( V \).

If we consider the left regular representation \( R \), then the proof of Theorem 12.15 shows that \( E_{\rho} = a_{\rho} \) for all \( \rho \in R \), and that on \( a_{\rho} \), the representation \( R \) splits into \( n_{\rho} \) irreducible representations all equivalent to \( M_{\rho} \). We can view these representation as acting on the columns of \( a_{\rho} \), which span \( n_{\rho} \) minimal left ideals \( l_{j}^{(\rho)} \) of \( a_{\rho} \); that is,

\[
a_{\rho} = \bigoplus_{j=1}^{n_{\rho}} l_{j}^{(\rho)} \quad \text{and} \quad l_{j}^{(\rho)} = \bigoplus_{k=1}^{n_{\rho}} \mathbb{C} m_{kj}^{(\rho)}.
\]

If \( V \) is a finite-dimensional unitary representation, then the trace of the linear map \( V(s) \) plays a crucial role. In fact, it determines this representation up to equivalence.

Proposition 12.16. Let \( G \) be a metrizable compact group. For any unitary representation \( V: G \to U(H) \) of \( G \) in a finite-dimensional hermitian space \( H \) of dimension \( d \), assume that for every \( \rho \in R \), the irreducible representation \( M_{\rho} \) is contained \( d_{\rho}n_{\rho} \) times in \( V \), so that

\[
d = \sum_{\rho \in R} d_{\rho}n_{\rho},
\]

where \( d_{\rho} \neq 0 \) for only finitely many \( \rho \in R \). Then we have

\[
\text{tr}(V(s)) = \sum_{\rho \in R} d_{\rho} \chi_{\rho}(s), \quad \text{for all } s \in G.
\]

Proof. We can write \( H \) as the direct sum of finite-dimensional spaces, and by picking bases, we can express \( V(s) \) as a sum of matrices similar to some of the \( M_{\rho}(s) \). Then the above formula follows from the fact that \( \chi_{\rho}(s) = \text{tr}(M_{\rho}(s)) \) and the fact that \( \text{tr}(UPU^{-1}) = \text{tr}(U) \).

Theorem 12.17. Let \( G \) be a metrizable compact group. Two unitary representations \( V_{1}: G \to U(H_{2}) \) and \( V_{2}: G \to U(H_{2}) \) of \( G \) in finite-dimensional hermitian spaces \( H_{1} \) and \( H_{2} \) of dimensions \( d_{1} \) and \( d_{2} \) are equivalent if and only if

\[
\text{tr}(V_{1}(s)) = \text{tr}(V_{2}(s)) \quad \text{for all } s \in G.
\]

In particular, if \( V_{1} \) an \( V_{2} \) are equivalent, then \( d_{1} = d_{2} \).
Proof. Clearly, if \( V_1 \) and \( V_2 \) are equivalent, the formula of the theorem holds. Conversely, by Proposition 12.16, since

\[
\text{tr}(V_1(s)) = \sum_{\rho \in R_1} d_{\rho} \chi_{\rho}(s) \quad \text{and} \quad \text{tr}(V_2(s)) = \sum_{\rho \in R_2} d_{\rho} \chi_{\rho}(s)
\]

for some finite subsets \( R_1 \) and \( R_2 \) of \( R \), and since by Theorem 12.6(3) the characters are linearly independent, we must have \( R_1 = R_2 \), and \( d_1 = d_2 \). \( \square \)

Theorem 12.17 suggests the following (standard) definition.

**Definition 12.7.** Let \( G \) be a metrizable compact group. For any unitary representation \( V: G \to U(H) \) of \( G \) in a finite-dimensional hermitian space \( H \) of dimension \( d \), we define the character \( \chi_V \) of the representation \( V \) as the map \( \chi_V: G \to \mathbb{C} \) given by

\[
\chi_V(s) = \text{tr}(V(s)) \quad \text{for all} \ s \in G.
\]

Observe that the definition of the character of a representation makes sense even if the representation is not unitary (it only needs to be finite-dimensional).

The characters of a finite-dimensional unitary representation are central functions, and in view of Proposition 12.16, they have many of the properties of the characters \( \chi_{\rho} \).

Theorem 12.17 shows that two finite-dimensional unitary representations \( V_1: G \to U(H_2) \) and \( V_2: G \to U(H_2) \) of \( G \) are equivalent if and only if \( \chi_{V_1} = \chi_{V_2} \). This confirms the importance of the characters; they determine the equivalence classes of finite-dimensional unitary representations of a metrizable compact group.

In view of Theorem 11.5 and the discussion following it, every finite-dimensional representation of \( G \) (not necessarily unitary) can be viewed as a unitary representation for some suitable hermitian inner product, so Proposition 12.16 and Theorem 12.17 also apply to such representations. Consequently, the characters also determine the equivalence classes of all finite-dimensional, not necessarily unitary, representations of a metrizable compact group.

If \( G \) is finite, it may be possible to build a character table for \( G \), and this determines all the irreducible representations of \( G \). In general, this is very difficult. An excellent source on the representation theory of finite groups is Serre [76].

Operations on (finite-dimensional) vector space induce operations on finite-dimensional, not necessarily unitary, representations, which in turn induce operations on their characters.

Given two finite-dimensional representations \( U_1: G \to U(H_1) \) and \( U_2: G \to U(H_2) \), with \( d_1 = \dim(H_1) \) and \( d_2 = \dim(H_2) \), we already defined their direct sum as the representation \( U_1 \oplus U_2 \) of \( G \) in \( H_1 \oplus H_2 \) given by

\[
(U_1 \oplus U_2)(s)(x_1 + x_2) = U_1(s)(x_1) + U_2(s)(x_2), \quad s \in G, \ x_1 \in H_1, \ x_2 \in H_2.
\]
We also define the tensor product $U_1 \otimes U_2$ of $U_1$ and $U_2$ as the representation of $G$ in $H_1 \otimes H_2$ given by

$$(U_1 \otimes U_2)(s) = U_1(s) \otimes U_2(s), \quad s \in G,$$

where $U_1(s) \otimes U_2(s)$ is the tensor product of the linear maps $U_1(s)$ and $U_2(s)$. In terms of matrices, it is a $(d_1d_2) \times (d_1d_2)$ matrix; as a block matrix, it is the $d_1 \times d_1$ matrix of $d_2 \times d_2$ blocks where the $(i, j)$ block is the matrix $U_1(s)_{ij}U_2(s)$ $(1 \leq i, j \leq d_1)$. This matrix is called the Kronecker product of $U_1(s)$ and $U_2(s)$. It is well known that

$$\text{tr}(U_1(s) \otimes U_2(s)) = \text{tr}(U_1(s)) + \text{tr}(U_2(s))$$

$$\text{tr}(U_1(s) \otimes U_2(s)) = \text{tr}(U_1(s))\text{tr}(U_2(s)).$$

In particular, if $U_1 = M_{\rho}$ and $U_2 = M_{\rho''}$, two irreducible representations, then since $\chi_{\rho'} \chi_{\rho''} = \text{tr}(M_{\rho'} \otimes M_{\rho''})$, which is finite-dimensional, Proposition 12.16 implies that

$$\chi_{\rho'} \chi_{\rho''} = \sum_{\rho \in R} c_{\rho', \rho''} \chi_{\rho},$$

where $c_{\rho', \rho''} \geq 0$ is an integer, the number of times that the representations $M_{\rho}$ is contained in $M_{\rho'} \otimes M_{\rho''}$ (this is $d_{\rho}$).

The determination of the $c_{\rho', \rho''}$ is usually very difficult. When $G = SU(2)$, the irreducible representations can be completely determined, and the $c_{\rho', \rho''}$ are known as the Clebsch–Gordan coefficients. They play an important role in physics.

Since the characters are linearly independent, we see that they form a subring of $\mathcal{C}(G; \mathbb{C})$ spanned by the characters, which is a $\mathbb{Z}$-algebra having the trivial character as identity, where the characters form a basis over $\mathbb{Z}$, and whose multiplication table is given as above.

**Remark:** For every $\rho \in R$, the trivial representation is contained in $M_{\rho} \otimes M_{\tau} = M_{\rho} \otimes M_{\bar{\tau}}$. Otherwise, by Proposition 12.9(4)(a,c) and by (⊗), since $c_{\rho', \rho''} \geq 0$, we would have

$$0 = \int \chi_{\rho}(s)\overline{\chi_{\rho}(s)} \, d\lambda(s) = \int |\chi_{\rho}(s)|^2 \, d\lambda(s),$$

which is absurd.

Since any irreducible representation $V$ of $G$ is equivalent to a unique representation $M_{\rho}$, we call $\rho$ the class of $V$ and we write $\rho = \text{cl}(V)$. Any finite-dimensional representation $V$ of $G$ corresponds uniquely to the formal linear combinations $\text{cl}(V) = \sum_{\rho \in R} d_{\rho} \rho$, over those $\rho$ for which $M_{\rho}$ occurs $d_{\rho}$ times. The $\mathbb{Z}$-module $\mathbb{Z}^{(R)}$ of formal linear combinations $\sum_{\rho \in R_1} m_{\rho} \rho$, with $m_{\rho} \in \mathbb{Z}$ and $R_1$ a finite subset of $R$, is isomorphic to the subring of of $\mathcal{C}(G; \mathbb{C})$ spanned by the characters, and we can give it a multiplication operation using formula (⊗). With this multiplication, we have

$$\text{cl}(U_1 \otimes U_2) = \text{cl}(U_1)\text{cl}(U_2).$$

This ring is the ring of linear representations of $G$. It is a substitute for the group of characters $\hat{G}$, when $G$ is abelian.
12.4 The Fourier Transform

Using the notations introduced just after Theorem 12.2, the formula

\[ f = \sum_{\rho \in R} f \ast u_{\rho}, \quad f \in L^2(G) \]

given by this theorem can be written as

\[ f = \sum_{\rho \in R} \left( \sum_{j=1}^{n_\rho} (f \ast m_{jj}^{(\rho)}) \right). \tag{*} \]

But, by definition,

\[ (f \ast m_{jj}^{(\rho)})(s) = \int f(t) m_{jj}^{(\rho)}(t^{-1}s) \, d\lambda(t), \]

so, if we recall that

\[ M_{\rho}(s) = \left( \frac{1}{n_\rho} m_{ij}(s) \right), \]

since \( M_{\rho}(t^{-1}s) \) is finite-dimensional, we get

\[ \sum_{j=1}^{n_\rho} (f \ast m_{jj}^{(\rho)})(s) = n_\rho \operatorname{tr} \left( \int f(t) M_{\rho}(t^{-1}s) \, d\lambda(t) \right). \]

However, we have

\[ M_{\rho}(t^{-1}s) = M_{\rho}(t^{-1}) M_{\rho}(s), \]

and

\[ \int f(t) M_{\rho}(t^{-1}) \, d\lambda(t) = M_{\rho}(\hat{f}), \]

because for every \( x \in \mathbb{C}^{n_\rho} \), the vector \( M_{\rho}(\hat{f})(x) \) is the unique vector \( \Phi(x) \in \mathbb{C}^{n_\rho} \) such that

\[ \langle \Phi(x), y \rangle = \int f(t) \langle M_{\rho}(t^{-1})(x), y \rangle \, d\lambda(t) \quad \text{for all } y \in \mathbb{C}^{n_\rho}, \]

and since \( G \) is unimodular, we get

\[ \langle \Phi(x), y \rangle = \int f(t^{-1}) \langle M_{\rho}(t)(x), y \rangle \, d\lambda(t) \quad \text{for all } y \in \mathbb{C}^{n_\rho}, \]

and by definition, \( \Phi(x) = M_{\rho}(\hat{f})(x) \). Therefore,

\[
\begin{align*}
\operatorname{tr} \left( \int f(t) M_{\rho}(t^{-1}s) \, d\lambda(t) \right) &= \operatorname{tr} \left( \int f(t) M_{\rho}(t^{-1}) M_{\rho}(s) \, d\lambda(t) \right) \\
&= \operatorname{tr} \left( \left( \int f(t) M_{\rho}(t^{-1}) \, d\lambda(t) \right) M_{\rho}(s) \right) \\
&= \operatorname{tr} \left( M_{\rho}(\hat{f}) M_{\rho}(s) \right),
\end{align*}
\]
and we obtain
\[ f(s) = \sum_{\rho \in R} n_{\rho} \operatorname{tr} (M_{\rho}(\hat{f}) M_{\rho}(s)) \quad f \in L^2(G), \ s \in G. \quad (\text{FI}_1) \]

The above suggests the following definition for the generalization of the Fourier transform to compact groups.

**Definition 12.8.** Let \( G \) be a compact group. For any function \( f \in L^1(G) \), the Fourier transform \( \mathcal{F}(f) \) of \( f \) is the map with domain \( R \) given by
\[ \mathcal{F}(f)(\rho) = M_{\rho}(\hat{f}) = \int f(t) M_{\rho}(t^{-1}) \, d\lambda(t) = \int f(t)(M_{\rho}(t))^* \, d\lambda(t), \quad \rho \in R. \]

Observe that \( \mathcal{F}(f)(\rho) \in M_{n_{\rho}}(\mathbb{C}) \cong \operatorname{Hom}_{\mathbb{C}}(\mathbb{C}^{n_{\rho}}, \mathbb{C}^{n_{\rho}}) \). The Fourier transform of Definition 12.8 is the natural generalization of the definition of the Fourier transform when \( G \) is an abelian compact group (Definition 9.3),
\[ \mathcal{F}(f)(\chi) = \int f(s) \overline{\chi(s)} \, d\lambda(s) = \int f(s)(s^{-1}) \, d\lambda(s); \]
the character \( \chi \) is replaced by the irreducible representation \( M_{\rho} \).

**Remark:** The Fourier transform of Definition 12.8 is related to the Fourier transform \( \mathcal{F}_2 \) defined by Hewitt and Ross [43] (Chapter VII, Definition 28.34) as the map
\[ \mathcal{F}_2(f)(\rho) = \int f(t) \overline{M_{\rho}(t)} \, d\lambda(t) = \overline{M_{\rho}(f)}, \quad f \in L^1(G), \ \rho \in R. \]

Let \( D_\rho : \mathbb{C}^{n_{\rho}} \to \mathbb{C}^{n_{\rho}} \) be the semilinear map given by
\[ D_\rho \left( \sum_{k=1}^{n_{\rho}} \alpha_k e_k \right) = \sum_{k=1}^{n_{\rho}} \overline{\alpha_k} e_k, \quad \alpha_k \in \mathbb{C}, \]
where \((e_1, \ldots, e_{n_{\rho}})\) is the canonical basis of \( \mathbb{C}^{n_{\rho}} \). It is immediately verified that
\[ \langle D_\rho x, D_\rho y \rangle = \langle y, x \rangle = \overline{\langle x, y \rangle}, \]
and
\[ D_\rho^2 = \operatorname{id}. \]

Then it is not hard to show (see Hewitt and Ross [43], Chapter IX, Lemma 34.1) that
\[ \mathcal{F}(f)(\rho) = D_\rho \circ (\mathcal{F}_2(f)(\rho))^* \circ D_\rho. \]

The definition of the Fourier transform \( \mathcal{F} \) given in Definition 12.8 is identical to the definition given by Kirillov; see [47] (Section 2.3) and by Folland [31] (Chapter 5, Section...
5.3). It has the advantage that the Fourier cotransform has a simpler formulation, and since for \( p \) with \( 1 \leq p \leq \infty \), the spaces \( L^p(\hat{G}) \) are closed under adjunction and conjugation, all the results proved for the Fourier transform \( \mathcal{F}_2 \) in Hewitt and Ross [43] also hold for the Fourier transform \( \mathcal{F} \).

Equation (FI\(_1\)) can also be written as

\[
\hat{f}(s) = \sum_{\rho \in R} n_\rho \text{tr}(\mathcal{F}(\rho)\rho M_\rho(s)) \quad f \in L^2(G), \ s \in G. \tag{FI}
\]

The Fourier transform \( \mathcal{F}(f) \) is a function with domain \( R \), the set of “equivalence classes” of irreducible representations of \( G \), which plays the analog of \( \hat{G} \), to the space \( \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \), where \( M_{n_\rho}(\mathbb{C}) \) is the algebra of \( n_\rho \times n_\rho \) complex matrices. Every element \( F \in \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \) is an \( R \)-indexed sequence \( F = (F(\rho))_{\rho \in R} \) of \( n_\rho \times n_\rho \) matrices \( F(\rho) \). Sequences in \( \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \) are added and multiplied by a scalar componentwise. Thus \( \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \) is a (complex) vector space. Given, \( F \in \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \), the adjoint \( F^* \) of \( F \) is defined componentwise by \( F^* = (F^*_\rho)_{\rho \in R} \).

Thus it is natural to define \( \hat{G} \) as \( R \). Note the analogy to the situation where \( G = \mathbb{T} \) and \( \hat{G} = \mathbb{T} = \mathbb{Z} \), except that, \( L^1(\mathbb{T}) = l^1(\mathbb{Z}) \) consists of \( \mathbb{Z} \)-indexed sequences of complex numbers, but the \( F(\rho) \) are matrices. By analogy with the case \( G = \mathbb{T} \) and \( \mathbb{T} = \mathbb{Z} \), where the numbers \( \mathcal{F}(f)(m) = \hat{f}(m) \) are the Fourier coefficients of \( f \in L^1(\mathbb{T}) \), the endomorphisms \( \mathcal{F}(\rho)(\cdot) \in \text{Hom}_\mathbb{C}(\mathbb{C}^{n_\rho}, \mathbb{C}^{n_\rho}) \), represented by matrices in \( M_{n_\rho}(\mathbb{C}) \), can be viewed as generalized Fourier coefficients of \( f \in L^1(G) \), where \( G \) is a compact group.

The equation (FI) is a kind of Fourier inversion formula.

We can define the Fourier cotransform \( \overline{\mathcal{F}} \), defined on \( \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \), with input in \( G \), by

\[
\overline{\mathcal{F}}(F)(s) = \sum_{\rho \in R} n_\rho \text{tr}(F(\rho)\rho M_\rho(s)), \quad F \in \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}), \ s \in G. \tag{FC}
\]

Of course, there is an issue of convergence with (FC). The space \( \prod_{\rho \in R} M_{n_\rho}(\mathbb{C}) \) is just too big, so following Hewitt and Ross [43] (Chapter VII, Section 28.24), we define normed subspaces \( L^p(\hat{G}) \) as follows.

First, we need to define some norms on \( n \times n \) matrices introduced by von Neumann.

**Definition 12.9.** Let \( A \in M_n(\mathbb{C}) \) be any complex \( n \times n \) matrix, and let \( (\sigma_1, \ldots, \sigma_n) \) be the sequence of nonnegative square roots of the eigenvalues of \( A^*A \) listed in any order (the positive square roots are the singular values of \( A \)). For any \( p, 1 \leq p < \infty \), define the **von Neumann norm** \( ||A||_{\psi_p} \) of \( A \) by

\[
||A||_{\psi_p} = \left( \sum_{k=1}^n \sigma_k^p \right)^{1/p},
\]
and $\|A\|_{\infty}$ by
\[
\|A\|_{\infty} = \max_{1 \leq k \leq n} \sigma_k.
\]

It is not obvious that the functions defined in Definition 12.9 are matrix norms, but this is proven in Hewitt and Ross, see [43] (Appendix D, Theorem D40).

Since $(\sigma_1^2, \ldots, \sigma_n^2)$ are the eigenvalues of $A^*A$, we see that
\[
\|A\|_{2} = \sum_{k=1}^{n} \sigma_k^2 = \text{tr}(A^*A) = \|A\|_{\text{HS}}^2,
\]
where $\|A\|_{\text{HS}}$ is a Hilbert–Schmidt norm, also known as Frobenius norm, of $A$ (see Definition 16.6). Also,
\[
\|A\|_{\infty} = \max_{1 \leq k \leq n} \sigma_k = \|A\|_{2},
\]
where $\|A\|_{2}$ is the operator norm induced by the 2-norm; see Definition 16.7 and Proposition 16.8.

Next we use the norms of Definition 12.9 define norms on $\prod_{\rho \in R} M_{n_{\rho}}(\mathbb{C})$.

**Definition 12.10.** For any $F \in \prod_{\rho \in R} M_{n_{\rho}}(\mathbb{C})$, for any sequence $(a_{\rho})_{\rho \in R}$ of reals $a_{\rho} \geq 1$, if $1 \leq p < \infty$, define $\|F\|_p$ by
\[
\|F\|_p = \left( \sum_{\rho \in R} a_{\rho} \|F(\rho)\|_{\varphi_p}^p \right)^{1/p},
\]
and for $p = \infty$, let
\[
\|F\|_{\infty} = \sup_{\rho \in R} \|F(\rho)\|_{\varphi_{\infty}},
\]
where $\|F(\rho)\|_{\varphi_p}$ is the von Neumann $p$-norm of the matrix $F(\rho)$.

Following Hewitt and Ross [43] (Chapter VII, Section 28.24), we make the following definitions.

**Definition 12.11.** Denote $\prod_{\rho \in R} M_{n_{\rho}}(\mathbb{C})$ by $\mathcal{E}(\hat{G})$. Let $\mathcal{E}(\hat{G})_{0,0}$ be the subspace of $\mathcal{E}(\hat{G})$ consisting of all sequences $F = (F(\rho))_{\rho \in R}$ such that the set $\{\rho \in R \mid F(\rho) \neq 0\}$ is finite, and let $\mathcal{E}(\hat{G})_0$ be the subspace of $\mathcal{E}(\hat{G})$ consisting of all sequences $F = (F(\rho))_{\rho \in R}$ such that the set $\{\rho \in R \mid \|F(\rho)\|_{\varphi_{\infty}} \geq \epsilon\}$ is finite for all $\epsilon > 0$.

For any $p$ with $1 \leq p \leq \infty$, we define $L^p(R) = L^p(\hat{G})$ as
\[
L^p(\hat{G}) = \left\{ F \in \prod_{\rho \in R} M_{n_{\rho}}(\mathbb{C}) \mid \|F\|_p < \infty \right\} = \left\{ F \in \mathcal{E}(\hat{G}) \mid \|F\|_p < \infty \right\}. 
\]
The following results are shown in Hewitt and Ross [43] (Theorem 28.25 and Theorem 28.26).

**Proposition 12.18.** Let $G$ be a compact group. For any sequence $(a_\rho)_{\rho \in \mathbb{R}}$ of reals $a_\rho \geq 1$, for any $p$ such that $1 \leq p \leq \infty$, the space $L^p(\hat{G})$ is a Banach space. For any $F \in L^p(\hat{G})$, we have $F^* \in L^p(\hat{G})$ and $\|F^*\|_p = \|F\|_p$. The space $L^\infty(\hat{G})$ is a Banach algebra under componentwise multiplication, and $\|FF^*\|_\infty = \|F\|_\infty^2$ for any $F \in L^\infty(\hat{G})$.

The following results is shown in Hewitt and Ross [43] (Theorem 28.27).

**Proposition 12.19.** Let $G$ be a compact group, and let $(a_\rho)_{\rho \in \mathbb{R}}$ be any sequence of reals $a_\rho \geq 1$. With the norm $\|\|_\infty$, the space $E(\hat{G})_0$ is a closed two-sided ideal of $L^\infty(\hat{G})$. For any $p$ such that $1 \leq p < \infty$, the space $E(\hat{G})_{0,0}$ is a dense two-sided ideal of $E(\hat{G})_0$, and a dense two-sided ideal of $L^p(\hat{G})$. Both $E(\hat{G})_{0,0}$ and $E(\hat{G})_0$ are closed under adjunction ($F \mapsto F^*$).

It is also possible to define an inner product on $L^p(\hat{G})$ based on the following proposition shown in Hewitt and Ross [43] (Lemma 28.28).

**Proposition 12.20.** Let $G$ be a compact group, and let $(a_\rho)_{\rho \in \mathbb{R}}$ be any sequence of reals $a_\rho \geq 1$. For any $p$, $1 \leq p \leq \infty$, for all $E, F \in L^p(\hat{G})$, the following facts hold:

1. The number
   \[ \langle E, F \rangle = \sum_{\rho \in \mathbb{R}} a_\rho \text{tr}(F^*_\rho E_\rho) \]
   is well defined (the series converges absolutely).

2. We have
   \[ \langle F, E \rangle = \langle E, F \rangle. \]

3. If $q$ is defined such that \( \frac{1}{p} + \frac{1}{q} = 1 \), then
   \[ |\langle E, F \rangle| \leq \|E\|_p \|F\|_q. \]

Then we have the following theorem shown in Hewitt and Ross [43] (Theorem 28.30).

**Theorem 12.21.** Let $G$ be a compact group, and let $(a_\rho)_{\rho \in \mathbb{R}}$ be any sequence of reals $a_\rho \geq 1$. The space $L^2(\hat{G})$ is a Hilbert space with the inner product
\[ \langle E, F \rangle = \sum_{\rho \in \mathbb{R}} a_\rho \text{tr}(F^*_\rho E_\rho), \]
and we have
\[ \|E\|_2^2 = \langle E, E \rangle. \]
12.4. THE FOURIER TRANSFORM

We also have the following result shown in Hewitt and Ross [43] (Theorem 28.32).

Proposition 12.22. Let \( G \) be a compact group, and let \((a_\rho)_{\rho \in \mathbb{R}}\) be any sequence of reals \(a_\rho \geq 1\).

1. For any \( p \) such that \( 1 \leq p \leq \infty \), if \( q \) is given by such that \( \frac{1}{p} + \frac{1}{q} = 1 \), for any \( E \in \mathcal{L}^p(\hat{G}) \) and \( F \in \mathcal{L}^q(\hat{G}) \), we have \( EF \in \mathcal{L}^1(\hat{G}) \), and

\[
\|EF\|_1 \leq \|E\|_p \|F\|_q.
\]

2. For any \( p, q \) such that \( 1 \leq p < q \leq \infty \), we have

\[
\mathcal{L}^p(\hat{G}) \subseteq \mathcal{L}^q(\hat{G})
\]

and for every \( E \in \mathcal{L}^p(\hat{G}) \),

\[
\|E\|_q \leq \|E\|_p.
\]

3. For any \( p \) such that \( 1 \leq p \leq \infty \), for all \( E, F \in \mathcal{L}^p(\hat{G}) \), we have \( EF \in \mathcal{L}^p(\hat{G}) \), and

\[
\|EF\|_p \leq \|E\|_p \|F\|_p.
\]

We now have the following results about the Fourier transform on a compact group, generalizing similar results about the Fourier transform on \( \mathbb{T} \). From now on, we assume that the sequence \((a_\rho)_{\rho \in \mathbb{R}}\) of reals \(a_\rho \geq 1\) is the sequence of positive integers \((n_\rho)_{\rho \in \mathbb{R}}\).

Theorem 12.23. Let \( G \) be a compact group.

1. If we define the multiplication on \( \mathcal{L}^\infty(\hat{G}) \) as \((F_1 \cdot F_2)(\rho) = F_2(\rho) \circ F_1(\rho)\), then the map \( f \mapsto \mathcal{F}(f) \) is a non norm-increasing injective involutive algebra homomorphism from \( \mathcal{L}^1(G) \) into \( \mathcal{L}^\infty(\hat{G}) \). In particular, for all \( f, g \in \mathcal{L}^1(G) \), for all \( \rho \in R \), we have

\[
(\mathcal{F}(f * g))(\rho) = \mathcal{F}(g)(\rho) \circ \mathcal{F}(f)(\rho).
\]

2. For every \( \rho \in R \), the map \( f \mapsto \mathcal{F}(f)(\rho) \) is an algebra representation of \( \mathcal{L}^1(G) \) in \( \mathbb{C}^{n_\rho} \).

Proof sketch. Theorem 12.23 is proven in Hewitt and Ross [43] (Theorem 28.36); see also Folland [31] (Section 5.3, Equations 5.17, 5.18). It is instructive to prove that

\[
(\mathcal{F}(f * g))(\rho) = \mathcal{F}(g)(\rho) \circ \mathcal{F}(f)(\rho).
\]

By definition, \((\mathcal{F}(f * g)(\rho))(x)\) is the unique vector such that

\[
\langle (\mathcal{F}(f * g)(\rho))(x), y \rangle = \int \langle M^*_\rho(s)(x), y \rangle (f * g)(s) d\lambda(s) \quad \text{for all } x, y \in \mathbb{C}^{n_\rho},
\]
and, using the fact \((\mathcal{F}(g)(\rho))(\mathbf{z})\) is the unique vector such that
\[
\langle (\mathcal{F}(g)(\rho))(\mathbf{z}), \mathbf{y} \rangle = \int \langle M^*_\rho(s)(\mathbf{z}), y \rangle g(s) \, d\lambda(s) \quad \text{for all } y, \mathbf{z} \in \mathbb{C}^{n_\rho}
\]
and \((\mathcal{F}(f)(\rho))(x)\) is the unique vector such that
\[
\langle (\mathcal{F}(f)(\rho))(x), y \rangle = \int \langle M^*_\rho(s)(x), y \rangle f(s) \, d\lambda(s) \quad \text{for all } x, y \in \mathbb{C}^{n_\rho},
\]
we have
\[
\int \langle M^*_\rho(s)(x), y \rangle (f * g)(s) \, d\lambda(s) = \int \langle M^*_\rho(s)(x), y \rangle \int f(t) g(t^{-1} s) \, d\lambda(t) \, d\lambda(s)
\]
\[
= \int \langle M^*_\rho(t^{-1} s) M^*_\rho(t)(x), y \rangle \int f(t) g(t^{-1} s) \, d\lambda(t) \, d\lambda(s)
\]
\[
= \int \left( \int \langle M^*_\rho(t^{-1} s) M^*_\rho(t)(x), y \rangle g(t^{-1} s) \, d\lambda(s) \right) f(t) \, d\lambda(t)
\]
\[
= \int \left( \int \langle M^*_\rho(s) M^*_\rho(t)(x), y \rangle g(s) \, d\lambda(s) \right) f(t) \, d\lambda(t)
\]
\[
= \int \langle (\mathcal{F}(g)(\rho))(M^*_\rho(t)(x)), y \rangle f(t) \, d\lambda(t)
\]
\[
= \int \langle (\mathcal{F}(f)(\rho))(\mathcal{F}(g)(\rho))^*(y) \rangle f(t) \, d\lambda(t)
\]
\[
= \langle (\mathcal{F}(f)(\rho))(\mathcal{F}(g)(\rho))^*(y) \rangle
\]
\[
= \langle (\mathcal{F}(g)(\rho))(\mathcal{F}(f)(\rho))(x), y \rangle,
\]
which proves that
\[
(\mathcal{F}(f * g)(\rho)) = \mathcal{F}(g)(\rho) \circ \mathcal{F}(f)(\rho),
\]
as claimed. □

**Remark:** Notice that the order of \(f\) and \(g\) is switched on the right-hand side. If we use the Fourier transform \(\mathcal{F}_2\) instead of the Fourier transform \(\mathcal{F}\), then we get
\[
(\mathcal{F}_2(f * g))(\rho) = \mathcal{F}_2(f)(\rho) \circ \mathcal{F}_2(g)(\rho),
\]
as in Hewitt and Ross [43].

**Definition 12.12.** Let \(G\) be a compact group. For any \(\rho \in \mathcal{R}\), let \(\mathcal{T}_\rho\) be the space of functions from \(G\) to \(\mathbb{C}\) spanned by the set of functions
\[
s \mapsto \langle M_\rho(s)(x), y \rangle, \quad x, y \in \mathbb{C}^{n_\rho},
\]
called *matrix coefficients*. Let \(\mathcal{T}(G)\) be the space of functions spanned by the set
\[
\bigcup_{\rho \in \mathcal{R}} \mathcal{T}_\rho(G).
\]
12.4. THE FOURIER TRANSFORM

Since the $M_\rho$ are representations, we have $\mathcal{T}(G) \subseteq \mathcal{C}(G; \mathbb{C})$.

**Theorem 12.24.** Let $G$ be a compact group.

1. For every $\rho \in R$, we have
   \[ \{ \mathcal{F}(f)(\rho) | f \in \mathcal{T}_\rho(G) \} = \text{Hom}_\mathbb{C}(\mathbb{C}^{n_\rho}, \mathbb{C}^{n_\rho}) \].

2. We have
   \[ \{ \mathcal{F}(f) | f \in \mathcal{T}(G) \} = \mathcal{E}_{0,0}(G) \].


**Theorem 12.25.** Let $G$ be a compact group. The map $f \mapsto \mathcal{F}(f)$ is a non norm-increasing involutive isomorphism of $L^1(G)$ onto a dense subalgebra of $\mathcal{E}_0(\widehat{G}) \subseteq L^\infty(\widehat{G})$.

   Theorem 12.25 is proven in Hewitt and Ross [43] (Theorem 28.40). It is a version of the Riemann–Lebesgue lemma for compact groups; indeed, since $\widehat{G} = R$ is discrete, by definition of $\mathcal{E}_0(\widehat{G})$, we can view the functions in $\mathcal{E}_0(\widehat{G})$ as continuous functions that tend to zero at infinity. See Proposition 9.17 in the case of abelian locally compact groups.

Finally, we have the following version of Plancherel’s theorem.

**Theorem 12.26.** (Plancherel) Let $G$ be a compact group. The map $f \mapsto \mathcal{F}(f)$ is an isometric isomorphism between the Hilbert space $L^2(G)$ and the Hilbert space $L^2(\widehat{G})$. If we pick any orthonormal basis $(e^\rho_1, \ldots, e^\rho_{n_\rho})$ in $\mathbb{C}^{n_\rho}$, then for every $f \in L^2(G)$, we have

\[ f = \sum_{\rho \in R} n_\rho \sum_{j,k=1}^{n_\rho} \langle (\mathcal{F}(f)(\rho))(e^\rho_k), e^\rho_j \rangle u^\rho_{jk}, \]

where $u^\rho_{jk}$ is the function on $G$ given by

\[ u^\rho_{jk}(s) = \langle M_\rho(s)(e^\rho_k), e^\rho_j \rangle, \quad s \in G, \ 1 \leq j, k \leq n_\rho. \]

The functions $u^\rho_{jk}$ are called the coordinate functions for $M_\rho$ and the basis $(e^\rho_1, \ldots, e^\rho_{n_\rho})$.

Theorem 12.26 is proven in Hewitt and Ross [43] (Theorem 28.43).

We now return to the Fourier cotransform.

**Definition 12.13.** Let $G$ be a compact group. For any $F \in \mathcal{E}(\widehat{G}) = \prod_{\rho \in R} M_{n_\rho}(\mathbb{C})$, the Fourier cotransform $\mathcal{F}(F)$ of $F$ is the function on $G$ given by

\[ \mathcal{F}(F)(s) = \sum_{\rho \in R} n_\rho \text{tr}(F(\rho)M_\rho(s)), \quad s \in G. \]
Remark: Since Hewitt and Ross use the Fourier transform $\mathcal{F}_2$, related to the Fourier transform $\mathcal{F}$ by the equation
\[
\mathcal{F}_2(f)(\rho) = D_{\rho} \circ (\mathcal{F}_2(f)(\rho))^* \circ D_{\rho},
\]
the definition of the Fourier cotransform (called inverse Fourier transform) given by Hewitt and Ross (Chapter IX, Section 34.47) is
\[
\mathcal{F}_2(F)(s) = \sum_{\rho \in R} n_{\rho} \text{tr}(D_{\rho} F(\rho)^* D_{\rho} M_{\rho}(s)), \quad s \in G.
\]

Following Hewitt and Ross, it is natural to make the following definition (see [43], Definition 34.4).

**Definition 12.14.** Let $G$ be a compact group. The subspace $\mathfrak{R}(G)$ of $L^1(G)$ is defined by
\[
\mathfrak{R}(G) = \{ f \in L^1(G) \mid \|\mathcal{F}(f)\|_1 < \infty \} = \left\{ f \in L^1(G) \mid \sum_{\rho \in R} n_{\rho} \|\mathcal{F}(f)(\rho)\|_{\varphi_1} < \infty \right\}.
\]
The subspace $\mathfrak{R}(G)$ is called the space of *absolutely convergent Fourier series*. We define $\|f\|_{\varphi_1}$ by
\[
\|f\|_{\varphi_1} = \|\mathcal{F}(f)\|_1.
\]
For any function $f \in L^1(G)$, the formal expression
\[
(\mathcal{F}(\mathcal{F}(f)))(s) = \sum_{\rho \in R} n_{\rho} \text{tr}(\mathcal{F}(g)(\rho) M_{\rho}(s))
\]
is called the *Fourier series* of $f$.

Observe that Definition 12.14 is the generalization of the case $G = \mathbb{T}$ and $\hat{\mathbb{T}} = \mathbb{Z}$ where for every $f \in L^1(\mathbb{T})$ we define the Fourier series of $f$ as the map
\[
\theta \mapsto \sum_{m \in \mathbb{Z}} \hat{f}(m)e^{im\theta} = (\mathcal{F}(\hat{f}))(\theta).
\]
Here the character $\theta \mapsto e^{im\theta}$ is replaced by the irreducible representation $M_{\rho}$, and the trace function is needed to convert the matrix $\mathcal{F}(f)(\rho) M_{\rho}(s)$ to a number (and the dimensions $n_{\rho}$ must be accounted for).

The following results are shown in Hewitt and Ross [43] (Theorem 34.5).

**Theorem 12.27.** Let $G$ be a compact group.

1. If $F \in L^1(\hat{G})$, then the map
\[
s \mapsto \sum_{\rho \in R} n_{\rho} \text{tr}(|F(\rho) M_{\rho}(s)|)
\]
is uniformly convergent.
(2) The map
\[ s \mapsto (\mathcal{F}(F))(s) = \sum_{\rho \in \mathbb{R}} n_\rho \text{tr}(F(\rho)M_\rho(s)) \]
converges uniformly to a continuous function \( f \). Furthermore, we have the Fourier inversion formula
\[ (\mathcal{F}(\mathcal{F}(f)))(s) = \sum_{\rho \in \mathbb{R}} n_\rho \text{tr}(F(f)(\rho)M_\rho(s)), \quad s \in G, \]
where \((\mathcal{F}(\mathcal{F}(f)))(s)\) is the Fourier series of \( f \), so \( f \in \mathcal{R}(G) \).

(3) We have
\[ \|f\|_\infty \leq \|f\|_{\varphi_1} = \|\mathcal{F}(f)\|_1, \]
where \( \|f\|_\infty \) is the sup norm on \( \mathcal{C}(G; \mathbb{C}) \).

The Fourier series of \( f \) is not necessarily convergent, but we have the following results; see Hewitt and Ross [43] (Corollary 34.6 and Corollary 34.7).

**Theorem 12.28.** Let \( G \) be a compact group.

1. For any function \( f \in \mathcal{R}(G) \), the Fourier series of \( f \) converges uniformly and
\[ f = (\mathcal{F}(\mathcal{F}(f)))(s) = \sum_{\rho \in \mathbb{R}} n_\rho \text{tr}(F(g)(\rho)M_\rho(s)) \]
for almost all \( s \in G \). We have
\[ \|f\|_\infty \leq \|f\|_{\varphi_1}. \]

2. The map \( f \mapsto \mathcal{F}(f) \) is a norm-preserving linear isomorphism from \( \mathcal{R}(G) \) onto \( L^1(\hat{G}) \), so \( \mathcal{R}(G) \) is a Banach space.

For the record, in view of Theorem 12.26 and (FI), we have the following result (see also Hewitt and Ross [43], Chapter IX, Section 34.17(a)).

**Theorem 12.29.** (Fourier inversion for \( L^2(G) \)) Let \( G \) be a compact group. The Fourier cotransform \( \mathcal{F}(F) \) of any \( F \in L^2(\hat{G}) \) converges in the \( \| \|_2 \)-norm, and for every \( f \in L^2(G) \), we have
\[ f(s) = (\mathcal{F}(\mathcal{F}(f)))(s) = \sum_{\rho \in \mathbb{R}} n_\rho \text{tr}(F(f)(\rho)M_\rho(s)), \quad s \in G \]
in the \( L^2 \)-norm.
There is another way to express (FI) in terms of the characters of $G$. Indeed, we have
\[
\text{tr} \left( \int f(t) M_\rho(t^{-1}s) \, d\lambda(t) \right) = \int f(t) \text{tr} \left( M_\rho(t^{-1}s) \right) \, d\lambda(t) = (f * \chi_\rho)(s).
\]
Therefore, we have
\[
f = \sum_{\rho \in R} n_\rho f \ast \chi_\rho, \quad f \in L^2(G).
\] (FI')

See also Hewitt and Ross [43] (Chapter XII, Theorem 27.41). This generalizes Proposition 12.9(4)(b) when $f$ is a central function.

If $G$ is a finite group, then $\hat{G} = \{ \rho_1, \ldots, \rho_r \}$, where $r$ is the number of conjugacy classes of $G$. If we give $G$ the counting measure normalized so that $G$ has measure 1, then the Fourier transform of any function $f \in L^2(G)$ is given by
\[
\mathcal{F}(f)(\rho) = \frac{1}{|G|} \sum_{s \in G} f(s)(M_\rho(s))^*,
\]
where $M_\rho_1, \ldots, M_\rho_r$ are the irreducible representations of $G$ (up to equivalence), for every $F \in L^2(\hat{G})$, the Fourier cotransform of $F$ is given by
\[
\mathcal{F}(F)(s) = \sum_{k=1}^r n_{\rho_k} \text{tr}(F(\rho_k) M_{\rho_k}(s)), \quad s \in G,
\]
and the Fourier inversion formula is given by
\[
f = \sum_{k=1}^r n_{\rho_k} \text{tr}((\mathcal{F}(f))(\rho_k) M_{\rho_k}(s)).
\]
The fact that $\mathcal{F}$ is an isometry is expressed by the equation
\[
\langle f_1, f_2 \rangle = \frac{1}{|G|} \sum_{s \in G} f_1(s) \overline{f_2(s)} = \langle \mathcal{F}(f_1), \mathcal{F}(f_2) \rangle = \sum_{k=1}^r n_{\rho_k} \text{tr}((\mathcal{F}(f_2))^* \mathcal{F}(f_1)),
\]
for all $f_1, f_2 \in L^2(G)$.

For all $f, g \in L^2(G)$, the convolution $f \ast g$ is given by
\[
(f \ast g)(s) = \frac{1}{|G|} \sum_{s_1 s_2 = s} f(s_1) g(s_2) = \frac{1}{|G|} \sum_{t \in G} f(t) g(t^{-1}s),
\]
and we can write explicitly the equation
\[
(\mathcal{F}(f \ast g))(\rho) = (\mathcal{F}(g))(\rho) \circ (\mathcal{F}(f))(\rho).
\]
We leave it to the diligence of reader to check that it holds.
Chapter 13

Induced Representations

If \( G \) is a locally compact group and if \( H \) is a closed subgroup of \( G \), under certain conditions, it is possible to construct a Hilbert space \( \mathcal{H} \) and a unitary representation \( \Pi: G \to U(\mathcal{H}) \) of \( G \) in \( \mathcal{H} \) from a unitary representation \( U: H \to U(E) \) of \( H \) in a (separable) Hilbert space \( E \). The representation \( \Pi \) is called an induced representation. In particular, this construction can be used to define unitary representations of the group \( \text{SL}(2, \mathbb{R}) \), which would be hard to find if we did not have this method.

There are two approaches for the construction of the Hilbert space \( \mathcal{H} \):

1. The Hilbert space \( \mathcal{H} \) is a set of functions from \( X = G/H \) to \( E \).

2. The Hilbert space \( \mathcal{H} \) is a set of functions from \( G \) to \( E \).

In the first approach, we will construct unitary representations of \( G \) in \( \mathcal{H} \) using certain functions \( \alpha: G \times (G/H) \to \text{GL}(E) \) called cocycles. In the second approach, the construction of the Hilbert space \( \mathcal{H} \) is more complicated, but the definition of the operators \( \Pi_s \) is simpler.

The general construction (in the first approach) consists of seven steps, where the first four are purely algebraic and we do not deal with unitary representations, but instead linear representations (group homomorphisms into \( \text{GL}(E) \), where \( E \) is just a vector space with no additional structure):

1. Let \( G \) be a group acting (on the left) on a set \( X \) and let \( E \) be a vector space. In Section 13.1, we define the notion of equilinear action of \( G \) on \( X \times E \), which is an action of the form

\[
s \cdot (x, z) = (s \cdot x, \alpha(s, x)(z)), \quad s \in G, x \in X, z \in E,
\]

where \( \alpha(s, x) \) is a linear automorphism of \( E \) satisfying the conditions

(a) For all \( x \in X \)

\[
\alpha(e, x) = \text{id}_E.
\]
(b) For all \( x \in X \) and all \( s, t \in G \),
\[
\alpha(st, x) = \alpha(s, t \cdot x) \circ \alpha(t, x).
\]

A map \( \alpha: G \times X \to \text{GL}(E) \) satisfying Conditions (a) and (b) is called a \textit{cocycle of } \( G \) \textit{with values in } \text{GL}(E). Conversely, an action of \( G \) on \( X \) and a cocycle \( \alpha: G \times X \to \text{GL}(E) \) determines an equilinear action of \( G \) on \( X \times E \). Then we show that an equilinear action of \( G \) on \( X \times E \) induces a homomorphism \( \Pi: G \to \text{GL}(E^X) \), where \( E^X \) is the vector space of all functions from \( X \) to \( E \). More precisely, for every function \( f: X \to E \), for every \( s \in G, \Pi_s(f): X \to E \) is function given by
\[
(\Pi_s(f))(x) = \alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x)), \quad \text{for every } x \in X.
\]

(2) In Section 13.2, we specialize the construction to the homogeneous space \( X = G/H \) of left cosets. Then \( G \) acts on \( G/H \) on the left by
\[
s \cdot (gH) = sgH.
\]
By choosing a set of representatives in the cosets of \( G/H \), a cocycle \( \alpha: G \times X \to \text{GL}(E) \) determines a homomorphism \( \sigma: H \to \text{GL}(E) \) and a map \( \beta: X \to \text{GL}(E) \). Conversely, a homomorphism \( \sigma: H \to \text{GL}(E) \) and a map \( \beta: X \to \text{GL}(E) \) determine a cocycle \( \alpha: G \times X \to \text{GL}(E) \). This step is the most important application of Step 1, and \( E \) is an arbitrary vector space.

(3) For a given homomorphism \( \sigma: H \to \text{GL}(E) \), the homomorphisms \( \Pi: G \to \text{GL}(E^X) \) corresponding to cocycles associated with different maps \( \beta \) are equivalent.

(4) In Section 13.3, we show that a cocycle \( \alpha: G \times X \to \text{GL}(E) \) determines a bijection between \( E^X \) and a subspace \( L^\alpha \) of the set \( E^G \) of maps from \( G \) to \( E \). As a consequence, the homomorphism \( \Pi: G \to \text{GL}(E^X) \) corresponding to a cocycle \( \alpha \) determines a homomorphism \( \Pi_{L^\alpha}: G \to \text{GL}(L^\alpha) \). This completes the purely algebraic construction. The next steps use topology and analysis to construct \textit{unitary} representations.

(5) In Section 13.4, we assume that \( G \) is a locally compact group and \( H \) is a closed subgroup of \( G \), in which case \( G/H \) is also locally compact. Let \( \mu \) be a positive measure on \( X = G/H \), and assume that \( E \) is a separable Hilbert space. Then we define a Hilbert space \( L^2_\mu(X; E) \) consisting of measurable functions from \( X \) to \( E \).

(6) In Section 13.5, given a unitary representation \( U \) of \( H \) in \( E \), we assume that the measure \( \mu \) on \( X = G/H \) is \( G \)-invariant, and that the cocycle \( \alpha \) satisfies the conditions:

(i) The linear automorphisms \( \alpha(s, x) \) of \( E \) are unitary operators of \( E \) for all \( s \in G \) and all \( x \in G/H \), and \( \alpha(h, x_0) = U(h) \) for all \( h \in H \) (where \( x_0 \) denotes the coset \( H \)).
(ii) For every $s \in G$, for every $f \in \mathcal{L}^2_\mu(X;E)$, the map $x \mapsto \alpha(s,x)(f(x))$ from $X$ to $E$ is $\mu$-measurable.

(iii) For every $f \in \mathcal{L}^2_\mu(X;E)$, the map $s \mapsto [\Pi(s)(f)]$ from $G$ to $\mathcal{L}^2_\mu(X;E)$ is continuous.

Then the homomorphism $s \mapsto \Pi_s([f]) = [\Pi_s(f)]$ is a unitary representation of $G$ in $\mathcal{L}^2_\mu(X;E) = \mathcal{H}$.

(7) In Sections 13.6 and 13.7, we generalize the previous construction to certain measure called quasi-invariant. If the measure $\mu$ on $G/H$ is quasi-invariant and another technical condition is satisfied, then the homomorphism $s \mapsto \Pi_s([f]) = [\Pi_s(f)]$ is a unitary representation of $G$ in $\mathcal{L}^2_\mu(X;E)$. Quasi-invariant measures on $G/H$ always exist and can be constructed using rho-functions. We conclude by showing how to construct unitary representations of $\text{SL}(2,\mathbb{R})$ using induced representations. One example involves the action of $\text{SL}(2,\mathbb{R})$ on the projective line $\mathbb{RP}^1$, and the other example involves the action of $\text{SL}(2,\mathbb{R})$ on the upper half plane.

In Section 13.8, we consider a compact (metrizable) group $G$ and a closed subgroup $H$ of $G$, and our goal is to determine the canonical (unitary) representation of $G$ in $\mathcal{L}^2_\mu(G/H;\mathbb{C})$ induced by the trivial representation of $H$ in $E = \mathbb{C}$ (see Definition 13.10), where $\mu$ is the $G$-invariant measure on $G/H$ induced by a Haar measure $\lambda$ on $G$. For simplicity of notation we write $\mathcal{L}^2_\mu(G/H)$ instead of $\mathcal{L}^2_\mu(G/H;\mathbb{C})$. To do this it is necessary to understand what is the restriction of the representation $M_{\rho}: G \to \text{U}(\mathbb{C}^{n_\rho})$ to $H$, with $\rho \in R(G)$.

In Proposition 13.18, we show that the space $\mathcal{L}^2_\mu(G/H)$ is the Hilbert sum of subspaces $L_{\rho} \subseteq a_{\rho}$. If the trivial representation $\sigma_0$ of $H$ is contained $d = (\rho : \sigma_0) \geq 1$ times in the restriction of $M_{\rho}$ to $H$, then $L_{\rho}$ is the direct sum of the first $d$ columns of $M_{\rho}$,

$$L_{\rho} = \bigoplus_{j=1}^{d} t_{(j)}^{(\rho)} \quad \text{and} \quad t_{(j)}^{(\rho)} = \bigoplus_{k=1}^{n_\rho} \mathbb{C}m_{kj}^{(\rho)}.$$

If $d = 0$, then $L_{\rho} = (0)$.

Then we consider the space $H\backslash G$ of right cosets $Hs$ of $G$ ($s \in G$). If $\pi: G \to H\backslash G$ is the quotient map $\pi(s) = Hs$, the fact that the Haar measure $\lambda$ on a compact group is left and right invariant implies immediately that there is a $G$-invariant measure $\mu'$ on $H\backslash G$. We show in Proposition 13.19 that the space $\mathcal{L}^2_{\mu'}(H\backslash G)$ is the Hilbert sum of subspaces $\mathcal{T}_{\rho} \subseteq a_{\rho}$. If the trivial representation $\sigma_0$ of $H$ is contained $d = (\rho : \sigma_0) \geq 1$ times in the restriction of $M_{\rho}$ to $H$, then $\mathcal{T}_{\rho}$ is the direct sum of the first $d$ rows of $M_{\rho}$; that is,

$$\mathcal{T}_{\rho} = \bigoplus_{i=1}^{d} \bigoplus_{j=1}^{n_\rho} \mathbb{C}m_{ij}^{(\rho)}.$$


CHAPTER 13. INDUCED REPRESENTATIONS

In preparation for Chapter 14, we consider the intersection \( L^2_\mu(G/H) \cap L^2_\mu(H\backslash G) \). This is a closed involutive subalgebra of \( L^2(G) \), thus a complete Hilbert algebra. We can view a function \( g \in L^2_\mu(G/H) \cap L^2_\mu(H\backslash G) \) as a function \( g \in L^2(G) \) such that
\[
g(tst') = g(s) \quad \text{for all } t, t' \in H \text{ and all } s \in G.
\]
We can also think of the functions \( g \in L^2_\mu(G/H) \cap L^2_\mu(H\backslash G) \) as functions defined on the double classes \( HsH \) of \( G \) with respect to \( H \).

We denote the algebra of functions in \( L^2(G) \) satisfying \((*=H\backslash G/H)\) as \( L^2(H\backslash G/H) \). Then we show in Proposition 13.20 that the algebra \( L^2(H\backslash G/H) \) is the Hilbert sum of the minimal two-sided ideals
\[
a_{\rho,\sigma_0} = L\rho \cap \overline{L}_\rho = \bigoplus_{i=1}^d \bigoplus_{j=1}^d \mathbb{C} m_{ij}^{(\rho)}.
\]
Each \( a_{\rho,\sigma_0} \) is a matrix algebra of dimension \( d^2 \) having the family \( (m_{ij}^{(\rho)})_{1\leq i,j\leq d} \) as a basis.

Again, in preparation for Chapter 14 on Gelfand pairs, we show in Proposition 13.21 that the algebra \( L^2(H\backslash G/H) \) is commutative if and only if \((\rho:\sigma_0) \leq 1\) for all \( \rho \in R(G) \). If so, then for every \( \rho \in R(G) \) such that \((\rho:\sigma_0) = 1\), the ideal \( a_{\rho,\sigma_0} \) is one-dimensional and is spanned by the function
\[
\omega_\rho(s) = \frac{1}{n_\rho} m_{11}^{(\rho)}(s),
\]
which is continuous and of positive type. Thus
\[
L^2(H\backslash G/H) = \bigoplus_{\rho | (\rho:\sigma_0) = 1} \mathbb{C} \omega_\rho.
\]
The function \( \omega_\rho \) also satisfies the following equations:
\[
\omega_\rho(tst') = \omega_\rho(s), \quad \text{for all } s \in G \text{ and all } t, t' \in H
\]
\[
\omega_\rho(e) = 1.
\]
The function \( \omega_\rho \) is called a \((\text{zonal}) \text{ spherical function}\). Such functions are crucial in generalizing the notion of Fourier transform to a homogeneous space \( G/H \). In Chapter 14, we will see how to achieve this when \( G \) is not compact (but \( H \) is compact). The key point is to consider pairs \((G, H)\) for which \( L^2(H\backslash G/H) \) is commutative. Actually, we can’t quite work with \( L^2(H\backslash G/H) \) because this space is not closed under convolution, but we will be able to work with another commutative algebra \( L^1(H\backslash G/H) \).

13.1 Cocycles and Induced Representations

As a warm up and as an example of the second approach, we consider the case where \( G \) is compact, and \( U \) is a unitary representation of \( H \) in a finite-dimensional vector space \( E \).
Consider the Hilbert space $L^2(G; E)$ consisting of all functions $f: G \to E$ such that for any orthonormal basis $(e_1, \ldots, e_n)$ of $E$, $f = f_1 e_1 + \cdots + f_n e_n$, where the $f_i$ are functions in $L^2(G)$; equivalently, $L^2(G; E)$ is the finite Hilbert sum $L^2(G; E) = \bigoplus_{i=1}^n L^2(G)e_i$. The inner product of two functions $f = \sum_{i=1}^n f_i e_i$ and $g = \sum_{i=1}^n g_i e_i$ is

$$
\langle f, g \rangle = \sum_{i=1}^n \int_G f_i(s)\overline{g_i(s)} \, d\lambda(s),
$$

where $\lambda$ is a Haar measure on $G$. This construction will be generalized in Section 13.4 to an infinite-dimensional Hilbert space. Consider the subspace $\mathcal{H}$ of $L^2(G; E)$ consisting of all functions $f$ such that

$$
f(sh) = U(h^{-1})(f(s)), \quad \text{for all } s \in G \text{ and all } h \in H. \quad (*)
$$

It is easy to check that $\mathcal{H}$ is closed in $L^2(G; E)$, so it is a Hilbert space. For any $f \in \mathcal{H}$, as before, let $\lambda_s f$ be the function given by

$$(\lambda_s f)(t) = f(s^{-1}t), \quad s, t \in G.
$$

For $s \in G$ fixed, the map $f \mapsto \lambda_s f$ is obviously linear. Observe that by $(*)$, for all $s, t \in G$, all $h \in H$, and all $f \in \mathcal{H}$, we have

$$(\lambda_t f)(sh) = f(t^{-1}sh) = U(h^{-1})(f(t^{-1}s)) = U(h^{-1})(\lambda_t f)(s),
$$

so $\lambda_t f \in \mathcal{H}$. For all $s, t, t' \in G$, we also have

$$(\lambda_{tt'} f)(s) = f((tt')^{-1}s) = f(t^{-1}t^{-1}s) = (\lambda_{t'} f)(t^{-1}s) = \lambda_{t'}((\lambda_t f))(s).
$$

If we define the map $\Pi: G \to GL(\mathcal{H})$ by

$$
\Pi_s(f) = \lambda_s f, \quad s \in G, \ f \in \mathcal{H},
$$

then we see that $\Pi$ is a linear representation of $G$ in $\mathcal{H}$. Since the Haar measure is left and right invariant, the maps $\lambda_t f$ are unitary ($f \in \mathcal{H}$), so $\Pi$ is a unitary representation of $G$ in $\mathcal{H}$, called the representation induced by $U$.

It is easy to see that if we replace $U$ by an equivalent representation $h \mapsto PU(h)P^{-1}$, where $P$ is a unitary transformation $U: E \to E'$, then, the corresponding induced representation is $s \mapsto f_P \Pi_s f_P^{-1}$, a unitary representation equivalent to $\Pi$, where $f_P$ is the linear map from $\mathcal{H}$ to $\mathcal{H'}$ given by $f_P(f) = P \circ f$. Therefore, the above construction defines a class of unitary representations of $G$ induced by a class of unitary representations of $H$.

Let us now consider a more general situation. Our first construction is purely algebraic and does not assume that the group $G$ or the vector space $E$ have any topology.
Chapter 13. Induced Representations

Definition 13.1. If we have a left group action \( \cdot : G \times X \rightarrow X \) of a group \( G \) on a set \( X \), for any vector space \( E \), a left action \( \cdot : G \times (X \times E) \rightarrow X \times E \) is equilinear if there is some function \( \alpha: G \times X \rightarrow \text{GL}(E) \) such that
\[
s \cdot (x, z) = (s \cdot x, \alpha(s, x)(z)), \quad \text{for all } s \in G, \text{ all } x \in X, \text{ and all } z \in E.
\]

If we have an equilinear action \( \cdot : G \times (X \times E) \rightarrow X \times E \), then the condition for being a left action are
\[
e \cdot (x, z) = (x, z) \\
(st) \cdot (x, z) = s \cdot (t \cdot (x, z)),
\]
which translate to
\[
(e \cdot x, \alpha(e, x)(z)) = (x, z) \\
((st) \cdot x, \alpha(st, x)(z)) = s \cdot (t \cdot x, \alpha(t, x)(z)) \\
= (s \cdot (t \cdot x), \alpha(s, t \cdot x)(\alpha(t, x)(z)),
\]
so we must have
\[
\alpha(e, x)(z) = z \\
\alpha(st, x)(z) = (\alpha(s, t \cdot x) \circ \alpha(t, x))(z),
\]
for all \( s, t \in G, \) all \( x \in X, \) and all \( z \in E \). By reversing the above computations, we see that if a function \( \alpha: G \times X \rightarrow \text{GL}(E) \) satisfies the above two conditions, then the map given by
\[
s \cdot (x, z) = (s \cdot x, \alpha(s, x)(z)), \quad \text{for all } s \in G, \text{ all } x \in X, \text{ and all } z \in E
\]
is an equilinear action. In summary, we proved the following proposition.

Proposition 13.1. Given a left group action \( \cdot : G \times X \rightarrow X \) and a vector space \( E \), for any function \( \alpha: G \times X \rightarrow \text{GL}(E) \), the map \( \cdot : G \times (X \times E) \rightarrow X \times E \) given by
\[
s \cdot (x, z) = (s \cdot x, \alpha(s, x)(z)), \quad \text{for all } s \in G, \text{ all } x \in X, \text{ and all } z \in E
\]
is an equilinear action if and only if the following two conditions hold:

(a) For all \( x \in X \)
\[
\alpha(e, x) = \text{id}_E.
\]

(b) For all \( x \in X \) and all \( s, t \in G \),
\[
\alpha(st, x) = \alpha(s, t \cdot x) \circ \alpha(t, x).
\]

In view of Proposition 13.1, we make the following definition.
13.1. COCYCLES AND INDUCED REPRESENTATIONS

Definition 13.2. Let $G$ be a left action of a group $G$ on a set $X$, and let $E$ be a vector space. Let $\alpha: G \times X \to \mathbf{GL}(E)$ be a function and assume that the following conditions hold:

(a) For all $x \in X$
$$\alpha(e, x) = \text{id}_E.$$

(b) For all $x \in X$ and all $s, t \in G$,
$$\alpha(st, x) = \alpha(s, t \cdot x) \circ \alpha(t, x).$$

A map $\alpha: G \times X \to \mathbf{GL}(E)$ satisfying Conditions (a) and (b) is called a cocycle of $G$ with values in $\mathbf{GL}(E)$.

The point of equilinear actions is that they yield homomorphisms $\Pi: G \to \mathbf{GL}(E^X)$, that is, linear representations of $G$ in the vector space $[X \to E] = E^X$.

Definition 13.3. Let $G$ be a left action of a group $G$ on a set $X$, and let $E$ be a vector space. For every equilinear action $\cdot: G \times (X \times E) \to X \times E$ defined by a cocycle $\alpha: G \times X \to \mathbf{GL}(E)$, for every function $f : X \to E$, for every $s \in G$, let $\Pi^a_s(f) : X \to E$ be the function given by

$$(\Pi^a_s(f))(x) = \alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x)), \quad \text{for every } x \in X.$$  

The above equation defines a map $\Pi^a_s : E^X \to E^X$. The map $\Pi^a : G \to \mathbf{GL}(E^X)$ given by $s \mapsto \Pi^a_s$ is the (linear) representation of $G$ in $E^X$ induced by the cocycle $\alpha$. For simplicity of notation, we write $\Pi$ instead of $\Pi^a$.

Proposition 13.2. Let $G$ be a left action of a group $G$ on a set $X$, and let $E$ be a vector space. For every equilinear action $\cdot: G \times (X \times E) \to X \times E$ defined by a cocycle $\alpha: G \times X \to \mathbf{GL}(E)$, for every $s \in G$, the map $\Pi_s: E^X \to E^X$ is linear, and the map $\Pi: G \to \mathbf{GL}(E^X)$ given by $s \mapsto \Pi_s$ is a homomorphism, that is, a linear representation of $G$ in the vector space $E^X$.

Proof. Since $\alpha(s, s^{-1} \cdot x)$ is a linear automorphism of $E$, we have

$$(\Pi_s(f_1 + f_2))(x) = \alpha(s, s^{-1} \cdot x)((f_1 + f_2)(s^{-1} \cdot x))$$
$$= \alpha(s, s^{-1} \cdot x)(f_1(s^{-1} \cdot x) + f_2(s^{-1} \cdot x))$$
$$= \alpha(s, s^{-1} \cdot x)(f_1(s^{-1} \cdot x)) + \alpha(s, s^{-1} \cdot x)(f_2(s^{-1} \cdot x))$$
$$= (\Pi_s(f_1))(x) + (\Pi_s(f_2))(x),$$

and for every $\lambda \in \mathbb{C}$,

$$(\Pi_s(\lambda f))(x) = \alpha(s, s^{-1} \cdot x)((\lambda f)(s^{-1} \cdot x))$$
$$= \alpha(s, s^{-1} \cdot x)(\lambda f(s^{-1} \cdot x))$$
$$= \lambda \alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x))$$
$$= \lambda (\Pi_s(f))(x),$$
so the map \( f \mapsto \Pi_s(f) \) from \( E^X \) to itself is linear.

For all \( s, t \in G \), we have
\[
g(y) = (\Pi_t(f))(y) = \alpha(t, t^{-1} \cdot x)(f(t^{-1} \cdot y)),
\]
so
\[
(\Pi_s(\Pi_t(f)))(x) = (\Pi_s(g))(x)
= \alpha(s, s^{-1} \cdot x)(g(s^{-1} \cdot x))
= \left(\alpha(s, s^{-1} \cdot x) \circ \alpha(t, t^{-1} \cdot (s^{-1} \cdot x))\right)(f(t^{-1} \cdot (s^{-1} \cdot x))),
\]
and we also have
\[
(\Pi_{st}(f))(x) = \alpha(st, (st)^{-1} \cdot x)(f((st)^{-1} \cdot x))
= \left(\alpha(s, t \cdot ((t^{-1}s)^{-1} \cdot x)) \circ \alpha(t, (t^{-1}s)^{-1} \cdot x)\right)(f(t^{-1}s^{-1} \cdot x))
= \left(\alpha(s, s^{-1} \cdot x) \circ \alpha(t, t^{-1} \cdot (s^{-1} \cdot x))\right)(f(t^{-1} \cdot (s^{-1} \cdot x)))
= (\Pi_s(\Pi_t(f)))(x),
\]
which proves that \( \Pi_{st}(f) = (\Pi_s \circ \Pi_t)(f) \), that is, \( \Pi \) is a homomorphism. \( \square \)

If we let \( t = s^{-1} \) in (b) of Definition 13.2, we obtain
\[
\alpha(s^{-1}, x) = (\alpha(s, s^{-1} \cdot x))^{-1},
\]
so \( \Pi_s(f) \) can also be written as
\[
(\Pi_s(f))(x) = (\alpha(s^{-1}, x))^{-1}(f(s^{-1} \cdot x)).
\]

# 13.2 Cocycles on a Homogeneous Space \( X = G/H \)

We now consider the special case where \( X = G/H \) is the homogeneous space of left cosets for some subgroup \( H \) of \( G \), and the left action of \( G \) acts on \( G/H \) given by
\[
s \cdot (gH) = sgH.
\]

For every coset \( x \in G/H \), let \( r_x \in G \) be a representative of \( x \), so that every element \( g \) of \( x = r_xH \) is written uniquely as \( g = r_xh \), with \( h \in H \). Denote the coset \( H \) by \( x_0 \) and pick \( r_{x_0} = e \). If we denote the quotient map by \( \pi: G \to G/H \), then picking a set of representatives \( (r_x)_{x \in G/H} \) in the cosets of \( G/H \) is equivalent to picking a section of \( \pi \), that is, a map \( r: G/H \to G \) such that \( \pi \circ r = \text{id}_{G/H} \).

For any \( s \in G \), the representative of \( s \cdot x = s \cdot r_xH = sr_xH \) is denoted by \( r_{sx} \). Since for every coset \( x \in H \) we have \( x = r_x \cdot x_0 \) (the class \( r_xH \), which is \( x \)), Condition (b) yields
\[
\alpha(sr_x, x_0) = \alpha(s, r_x \cdot x_0) \circ \alpha(r_x, x_0) = \alpha(s, x) \circ \alpha(r_x, x_0),
\]
13.2. COCYCLES ON A HOMOGENEOUS SPACE $X = G/H$

and so

$$\alpha(s, x) = \alpha(sr_x, x_0) \circ (\alpha(r_x, x_0))^{-1}. \quad (*)$$

Equation $(*)$ shows that the automorphisms $\alpha(s, x_0)$ of $E$ determine the $\alpha(s, x)$ for all $x \in X$.

Denote $\alpha(s, x_0)$ by $\alpha(s)$. Conditions (a) and (b) imply that

$$\alpha(e) = \text{id}_E$$

$$\alpha(sh, x_0) = \alpha(s, h \cdot x_0) \circ \alpha(h, x_0),$$

for all $s \in G$ and all $h \in H$, and since $h \cdot x_0 = x_0$, we get

$$\alpha(sh) = \alpha(s) \circ \alpha(h) \quad \text{for all } s \in G \text{ and all } h \in H. \quad (\ast_2)$$

Now $sr_x$ belongs to the coset $sr_xH = s \cdot r_xH = s \cdot x = r_{s,x}H$, so there is a unique element of $H$, denoted $h(s, x)$, such that

$$sr_x = r_{s,x}h(s, x),$$

and by $(\ast_2)$,

$$\alpha(sr_x) = \alpha(r_{s,x}h(s, x)) = \alpha(r_h) \circ \alpha(h(s, x)),$$

so $(\ast_1)$ can be written as

$$\alpha(s, x) = \alpha(r_h) \circ \alpha(h(s, x)) \circ (\alpha(r_x))^{-1}. \quad (\ast_3)$$

**Definition 13.4.** For all $s \in G$, all $h \in H$, and all $x \in X$, define $\alpha(s), \sigma(h), \beta(x)$ and $h(s, x)$ by

$$\alpha(s) = \alpha(s, x_0)$$

$$\sigma(h) = \alpha(h)$$

$$\beta(x) = \alpha(r_x)$$

$$h(s, x) = r_{s,x}^{-1}sr_x \in H.$$

Then $(\ast_3)$ becomes

$$\alpha(s, x) = \beta(s \cdot x) \circ \sigma(h(s, x)) \circ (\beta(x))^{-1}, \quad (\ast_4)$$

and $(\ast_2)$ implies that

$$\sigma(h_1h_2) = \sigma(h_1) \circ \sigma(h_2) \quad \text{for all } h_1, h_2 \in H, \quad (\ast_5)$$

which shows that $\sigma: H \to \text{GL}(E)$ is a homomorphism. Thus the restriction of the cocycle $\alpha$ to $H \times \{x_0\}$ is a representation of $H$ in $E$.

Conversely, let $\sigma: H \to \text{GL}(E)$ be any homomorphism, and let $\beta: X \to \text{GL}(E)$ be any function. Then using the equation

$$h(s, x) = r_{s,x}^{-1}sr_x$$
of Definition 13.4 and \(\ast_4\), we define a function \(\alpha: G \times X \to \text{GL}(E)\) by
\[
\alpha(s, x) = \beta(s \cdot x) \circ \sigma(h(s, x)) \circ (\beta(x))^{-1}.
\]
Observe that for all \(s, t \in G\) and all \(x \in X\), we have
\[
h(st, x) = h(s, t \cdot x)h(t, x),
\]
which shows that \(\alpha\) is a cocycle. In summary, we obtained the following result.

**Proposition 13.3.** Let \(G\) be a group, \(H\) be a subgroup of \(G\), and \(E\) be a vector space. Choose a set \((r_x)_{x \in G/H}\) of representatives for the cosets of \(X = G/H\) as explained above, with \(x_0 = H\) and \(r_{x_0} = e\). Every cocycle \(\alpha: G \times X \to \text{GL}(E)\) determines a homomorphism \(\sigma: H \to \text{GL}(E)\) with \(\sigma(h) = \alpha(h, x_0)\) for all \(h \in H\), a map \(\beta: X \to \text{GL}(E)\) given by \(\beta(x) = \alpha(r_x)\) for all \(x \in X\), and a map \(h(s, x)\) given by \(h(s, x) = r_{sx}^{-1}sr_x \in H\), such that
\[
\alpha(s, x) = \beta(s \cdot x) \circ \sigma(h(s, x)) \circ (\beta(x))^{-1}.
\]
Conversely, given a homomorphism \(\sigma: H \to \text{GL}(E)\) and a map \(\beta: X \to \text{GL}(E)\), if we set
\[
h(s, x) = r_{sx}^{-1}sr_x
\]
and
\[
\alpha(s, x) = \beta(s \cdot x) \circ \sigma(h(s, x)) \circ (\beta(x))^{-1},
\]
then \(\alpha: G \times X \to \text{GL}(E)\) is a cocycle.
In view of Proposition 13.3, given a homomorphism $\sigma: H \to \text{GL}(E)$ and a map $\beta: X \to \text{GL}(E)$, if $\alpha$ is the cocycle associated with $\sigma$ and $\beta$, we say that the representation $\Pi^\alpha$ of $G$ in $E^X$ defined by $\alpha$ is the representation induced by $\sigma$ and $\beta$.

Remarkably, for a given homomorphism $\sigma: H \to \text{GL}(E)$, the representations $\Pi_1: G \to \text{GL}(E^X)$ and $\Pi_2: G \to \text{GL}(E^X)$ corresponding to the cocycles $\alpha_1$ and $\alpha_2$ associated with two maps $\beta_1$ and $\beta_2$ are equivalent, in the sense that there is an automorphism $\gamma$ of $E^X$ such that

$$\Pi_2 = \gamma \circ \Pi_1 \circ \gamma^{-1}.$$ 

This is proven as follows.

**Proposition 13.4.** Let $G$ be a group, $H$ be a subgroup of $G$, and $E$ be a vector space. Choose a set $(r_x)_{x \in G/H}$ of representatives for the cosets of $X = G/H$ as explained above, with $x_0 = H$ and $r_{x_0} = e$. Let $\sigma: H \to \text{GL}(E)$ be a homomorphism, let $\beta: X \to \text{GL}(E)$ be a map, and let $\alpha$ be the cocycle determined by $\sigma$ and $\beta$ as in Proposition 13.3, and let $\Pi: G \to \text{GL}(E^X)$ be the corresponding representation. If $c(x) = \beta(x)^{-1}$ for all $x \in X$, then define the automorphism $\gamma$ of $E^X$ by

$$(\gamma(f))(x) = c(x)(f(x)), \quad f \in E^X, \ x \in X.$$ 

Then the representation

$$\Pi' = \gamma \circ \Pi \circ \gamma^{-1}$$

is associated with the cocycle $\alpha'$ given by

$$\alpha'(s, x) = \sigma(h(s, x)), $$

with

$$h(s, x) = r^{-1}_{s \cdot x} r_x.$$ 

Thus, the representation $\Pi$ induced by $\sigma$ and $\beta$ is equivalent to the representation induced by $\sigma$ and $\beta'$, with $\beta'(x) = \text{id}_E$ for all $x \in X$.

**Proof.** Consider any map $c: X \to \text{GL}(E)$ such that $c(x_0) = \text{id}_E$. Define the automorphism $\gamma$ of $E^X$ by

$$(\gamma(f))(x) = c(x)(f(x)), \quad f \in E^X, \ x \in X,$$

and let

$$\Pi' = \gamma \circ \Pi \circ \gamma^{-1}.$$ 

Since $\gamma$ is an automorphism of $E^X$, the map $\Pi'$ is a linear representation of $G$ in $E^X$. Clearly, the inverse of $\gamma$ is given by

$$(\gamma^{-1}(f))(x) = c(x)^{-1}(f(x)), \quad f \in E^X, \ x \in X,$$

and since for any $g \in E^X$, we have

$$(\Pi_s(g))(x) = \alpha(s, s^{-1} \cdot x)(g(s^{-1} \cdot x)),$$
with \( g = \gamma^{-1}(f) \), we obtain
\[
(\Pi_s(\gamma^{-1})(f))(x) = \alpha(s, s^{-1} \cdot x)(\gamma^{-1}(f)(s^{-1} \cdot x))
\]
\[
= \alpha(s, s^{-1} \cdot x)(c(s^{-1} \cdot x)^{-1}(f(s^{-1} \cdot x))),
\]
and so
\[
c(x)((\Pi_s(\gamma^{-1}))(x)) = c(x)(\alpha(s, s^{-1} \cdot x)(c(s^{-1} \cdot x)^{-1}(f(s^{-1} \cdot x))));
\]
that is,
\[
(\Pi'_s(f))(x) = (c(x) \circ \alpha(s, s^{-1} \cdot x) \circ c(s^{-1} \cdot x)^{-1})(f(s^{-1} \cdot x)),
\]
which shows that \( \Pi' \) is obtained from \( \alpha' \) as \( \Pi \) is obtained from \( \alpha \), with
\[
\alpha'(s, x) = c(s \cdot x) \circ \alpha(s, x) \circ c(x)^{-1}.
\]
If we write \( \alpha'(s) = \alpha'(s, x_0) \) and \( \beta'(x) = \alpha'(r_x) \) as before, then the hypothesis \( c(x_0) = \text{id}_E \) implies that
\[
\alpha'(h) = \alpha'(h, x_0) = c(h \cdot x_0) \circ \alpha(h, x_0) \circ c(x_0)^{-1} = c(x_0) \circ \sigma(h) \circ c(x_0)^{-1} = \sigma(h)
\]
for all \( h \in H \), and
\[
\beta'(x) = \alpha'(r_x, x_0)
\]
\[
= c(r_x \cdot x_0) \circ \alpha(r_x, x_0) \circ c(x_0)^{-1}
\]
\[
= c(x) \circ \alpha(r_x, x_0)
\]
\[
= c(x) \circ \beta(x);
\]
that is,
\[
\beta'(x) = c(x) \circ \beta(x).
\]
Since \( \beta(x_0) = \text{id}_E \), we can pick \( c(x) = \beta(x)^{-1} \), and then
\[
\beta'(x) = \text{id}_E(x) \quad \text{for all } x \in X,
\]
and since
\[
\alpha(s, x) = \beta(s \cdot x) \circ \sigma(h(s, x)) \circ (\beta(x))^{-1},
\]
from \((\ast_7)\) we obtain
\[
\alpha'(s, x) = \sigma(h(s, x)).
\]
Therefore, \( \Pi \) is equivalent to \( \Pi' \) with \( \beta'(x) = \text{id}_E \) for all \( x \in X \).

It is also easy to check that if \( \sigma \) is replaced by an equivalent representation \( \sigma' \) of \( H \) in \( E^X \), then the corresponding representations \( \Pi \) and \( \Pi' \) of \( G \) in \( E^X \) are equivalent.

Therefore, the process for making a representation \( \Pi \) of \( G \) in \( E^X \) from a representation \( \sigma \) of \( H \) in \( E^X \) and a function \( \beta: X \to \text{GL}(E) \) defines a class of representations of \( G \) in \( E^X \).
Furthermore, there is a special representation associated with \( \sigma \) and the constant function \( \beta \) given by \( \beta(x) = \text{id}_E \), for all \( x \in X \).

From a theoretical point of view, a cocycle \( \alpha \) is equivalent to a pair \((\sigma, \beta)\) as in Proposition 13.3, but from a practical point of view, it may be very hard (if not impossible) to find constructively a set \( \{r_x\}_{x \in G/H} \) of representatives for the cosets of \( G/H \). Thus we use cocycles \( \alpha \) that agree with a given representation \( \sigma: H \to \text{GL}(E) \), in the sense that \( \alpha(h, x_0) = \sigma(h) \) for all \( h \in H \).

### 13.3 Converting Induced Representations of \( G \) From \( E^X \) to \( E^G \)

We can also show that a cocycle \( \alpha: G \times X \to \text{GL}(E) \) defines an isomorphism \( \tau \) between the space \( E^X \) and a subspace \( L^\alpha \) of the space \( E^G \).

**Definition 13.5.** Let \( G \) be a group, \( H \) be a subgroup of \( G \), and \( E \) be a vector space, and write \( X = G/H \). Given any cocycle \( \alpha: G \times X \to \text{GL}(E) \), for any function \( f: X \to E \), the function \( f^\alpha: G \to E \) is given by

\[
f^\alpha(s) = \alpha(s^{-1}, s \cdot x_0)(f(s \cdot x_0)) = (\alpha(s, x_0))^{-1}(f(s \cdot x_0)), \quad \text{for all } s \in G,
\]

with \( x_0 = H \).

**Proposition 13.5.** With the hypotheses of Definition 13.5, the function \( f^\alpha \) satisfies the equation

\[
f^\alpha(sh) = \sigma(h^{-1})(f^\alpha(s)), \quad \text{for all } h \in H \text{ and all } s \in G.
\]

**Proof.** By (b) and since \( h \cdot x_0 = x_0 \), we have

\[
f^\alpha(sh) = \alpha((sh)^{-1}, (sh) \cdot x_0)(f((sh) \cdot x_0))
\]

\[
= \alpha(h^{-1}s^{-1}, (sh) \cdot x_0)(f((sh) \cdot x_0))
\]

\[
= (\alpha(h^{-1}, s^{-1} \cdot (s \cdot (h \cdot x_0))) \circ \alpha(s^{-1}, s \cdot (h \cdot x_0)))(f(s \cdot (h \cdot x_0)))
\]

\[
= (\alpha(h^{-1}, x_0) \circ \alpha(s^{-1}, s \cdot x_0))(f(s \cdot x_0))
\]

\[
= \sigma(h^{-1})(f^\alpha(s)),
\]

establishing the proposition. \(\square\)

**Definition 13.6.** Let \( G \) be a group, \( H \) be a subgroup of \( G \), and \( E \) be a vector space, and write \( X = G/H \). Given any cocycle \( \alpha: G \times X \to \text{GL}(E) \), let \( L^\alpha \) be the subspace of \( E^G \) consisting of all functions \( g: G \to E \) such that

\[
g(sh) = \sigma(h^{-1})(g(s)), \quad \text{for all } s \in G \text{ and all } h \in H,
\]

where \( \sigma(h) = \alpha(h, x_0) \), for all \( h \in H \) (with \( x_0 = H \)).
Proof. Note that the function \( \tau \) is unique function \( f : E \rightarrow X \) such that \( g = f^\alpha \). Therefore, the map \( \tau : E^X \rightarrow L^\alpha \) given by \( \tau(f) = f^\alpha \) is an isomorphism.

\[
\alpha(sh, x_0)(g(sh)) = (\alpha(s, h \cdot x_0) \circ \alpha(h, x_0))(g(sh)) \\
= (\alpha(s, x_0) \circ \alpha(h, x_0))(\sigma(h^{-1})(g(s))) \\
= (\alpha(s, x_0) \circ \sigma(h) \circ \sigma(h^{-1}))(g(s)) \\
= \alpha(s, x_0)(g(s)).
\]

Therefore, we have a well-defined function \( f : X \rightarrow E \) given by

\[
f(x) = f(s \cdot x_0) = \alpha(s, x_0)(g(s)),
\]

and by definition of \( f^\alpha \), we have

\[
f^\alpha(s) = (\alpha(s, x_0))^{-1}(f(s \cdot x_0)) \\
= (\alpha(s, x_0))^{-1}(\alpha(s, x_0)(g(s))) \\
= g(s),
\]

that is, \( f^\alpha = g \), which shows that \( \tau \) is surjective.

Since \( \alpha(s, x) \) is an automorphism and since the map \( s \mapsto s \cdot x_0 \) from \( G \) to \( G/H \) is surjective, for any two functions \( f_1, f_2 \in E^X \), if \( f_1^\alpha = f_2^\alpha \), then

\[
\alpha(s^{-1}, s \cdot x_0)(f_1(s \cdot x_0)) = \alpha(s^{-1}, s \cdot x_0)(f_2(s \cdot x_0))
\]

for all \( s \in G \), so \( f_1 = f_2 \), which shows that \( \tau \) is injective. \( \square \)

Observe that the proof of Proposition 13.6 shows that if \( g \in L^\alpha \), then

\[
(\tau^{-1}(g))(s \cdot x_0) = \alpha(s, x_0)(g(s)).
\]

For any cocycle \( \alpha : G \times X \rightarrow \text{GL}(E) \), we can use the isomorphism \( \tau : E^X \rightarrow L^\alpha \) to convert the representation \( \Pi : G \rightarrow \text{GL}(E^X) \) defined by \( \alpha \) into the equivalent representation \( \Pi_{L^\alpha} \) given by \( \Pi_{L^\alpha}(s) = \tau \circ \Pi(s) \circ \tau^{-1} \).

Proposition 13.7. For every cocycle \( \alpha : G \times X \rightarrow \text{GL}(E) \), if \( \Pi : G \rightarrow \text{GL}(E^X) \) is the representation defined by \( \alpha \), then the equivalent representation \( \Pi_{L^\alpha} : G \rightarrow \text{GL}(L^\alpha) \) defined by \( \Pi_{L^\alpha}(s) = \tau \circ \Pi(s) \circ \tau^{-1} \) is given by

\[
((\Pi_{L^\alpha})(g))(t) = g(s^{-1}t) \quad \text{for all } g \in L^\alpha \text{ and all } s, t \in G.
\]
13.4. CONSTRUCTION OF THE HILBERT SPACE $L^2_{\mu}(X; E)$

Proof. For any $g \in L^\alpha$, since

$$(\tau^{-1}(g))(u \cdot x_0) = \alpha(u, x_0)(g(u))$$

and

$$(\Pi_s(f))(x) = \alpha(s, s^{-1} \cdot x_0)(f(s^{-1} \cdot x)),$$

with $x = t \cdot x_0$, we have

$$(\Pi_s(f))(t \cdot x_0) = \alpha(s, s^{-1} \cdot x_0)(f(s^{-1} \cdot (t \cdot x_0))),$$

and by setting $f = \tau^{-1}(g)$, we get

$$(\Pi_s(\tau^{-1}(g)))(t \cdot x_0) = \alpha(s, s^{-1} \cdot (t \cdot x_0))((\tau^{-1}(g))(s^{-1} \cdot (t \cdot x_0)))$$

$$= \alpha(s, (s^{-1}t) \cdot x_0)((\tau^{-1}(g))(s^{-1}t) \cdot x_0)$$

$$= \alpha(s, (s^{-1}t) \cdot x_0)(\alpha(s^{-1}t, x_0)(g(s^{-1}t)))$$

$$= \alpha(ss^{-1}t, x_0)(g(s^{-1}t))$$

$$= \alpha(t, x_0)(g(s^{-1}t)).$$

Since

$$(\tau(h))(t) = (\alpha(t, x_0))^{-1}(h(t \cdot x_0)),$$

for any $h \in E^X$, with $h = \Pi_s(\tau^{-1}(g))$, we obtain

$$\tau(\Pi_s(\tau^{-1}(g)))(t) = (\alpha(t, x_0))^{-1}(\alpha(t, x_0)(g(s^{-1}t))) = g(s^{-1}t),$$

as claimed. \qed

Remark: Observe that $L^\alpha$ only depends on $\sigma$, so we may write $L^{\sigma}$ instead of $L^\alpha$, and $\Pi_{L^\alpha}$ depends only on $\sigma$, so we may also write $\Pi_{L^{\sigma}}$ instead of $\Pi_{L^\alpha}$.

We have concluded our discussion of algebraic methods for constructing representations of $G$ from representations of a subgroup $H$ of $G$.

13.4 Construction of the Hilbert Space $L^2_{\mu}(X; E)$

We now assume that $G$ is a locally compact group and that $H$ is a closed subgroup of $G$. By Proposition 6.6(1), the space $X = G/H$ is also locally compact. If $G$ is separable, then so is $G/H$, and if $G$ is metrizable, then so $G/H$; see Dieudonné [22] (Chapter XII, Sections 10 and 11).

Given a unitary representation $U: H \to U(E)$ of $H$ we would like to construct a unitary representation $\Pi: G \to U(H)$ of $G$. This is possible under certain conditions on $H$ and $G$ and on measures on $X = G/H$. The first step is to construct a Hilbert space $H$ that will be the representation space of a unitary representation of $G$. There are two approaches:
1. The Hilbert space $H$ is a set of functions from $X = G/H$ to $E$. 

2. The Hilbert space $H$ is a set of functions from $G$ to $E$, analogous to the space $L^{\alpha}$ of Section 13.3.

The second step is to define the operators $\Pi_s$ (for $s \in G$) so that they are unitary operators of $H$. This involves defining an inner product in $H$ that makes the operators $\Pi_s$ unitary. In the first approach that makes use of cocycles, the definition of the inner product on $H$ is straightforward. To ensure that the operators $\Pi_s$ are unitary, a Borel measure $\mu$ on $X = G/H$ is needed, and the cocycles must satisfy some additional conditions with respect to the measure $\mu$. The case where the measure $\mu$ is $G$-invariant is simpler than the case where $\mu$ is only quasi-invariant.

In the second approach, the definition of the Hilbert space $H$ is more complicated and requires a completion. We will sketch two variants of this method at the end of Section 13.7.

A good candidate for the first approach is a subspace $L^{\mu^2}(X; E)$ of the vector space $E^X$, where $\mu$ is positive Borel measure on $G/H$. In the special case where $H$ is compact, given a cocycle $\alpha$ on $G \times X$ satisfying some suitable conditions, the space $L^{\alpha}$ will be a subspace of $L^{\lambda^2}(G; E) \subseteq E^G$, where $\lambda$ is a left-invariant Haar measure on $G$.

Whether $\mu$ is $G$-invariant is an issue that will come up later, but for the time being we can ignore it.

Let $E$ be a separable Hilbert space, and let $(a_n)$ be a Hilbert basis of $E$. Every function $f: X \to E$ can be written uniquely as $f = \sum_n f_n a_n$, where $f_n: X \to \mathbb{C}$, and such that the series $\sum_n |f_n(x)|^2$ converges for all $x \in X$. By definition, we let

$$\|f(x)\|^2 = \sum_n |f_n(x)|^2.$$  

We claim that a function $f: X \to E$ is $\mu$-measurable iff all the $f_n$ are $\mu$-measurable.

If $f$ is $\mu$-measurable, and since $f_n(x) = \langle f(x), a_n \rangle$, the $f_n$ are $\mu$-measurable. Conversely, if the the $f_n$ are $\mu$-measurable, then Egoroff’s theorem implies that $f$ is $\mu$-measurable; see Dieudonné [22] (Chapter XIII, Theorem 13.9.10).

**Definition 13.7.** Let $G$ be a locally compact group, let $H$ be a closed subgroup of $G$, let $\mu$ be a positive Borel measure on $X = G/H$, and let $E$ be a separable Hilbert space. For any Hilbert basis $(a_n)$ of $E$, let $L^{\mu^2}(X; E)$ be the space of all $\mu$-measurable functions $f: X \to E$ with $f = \sum_n f_n a_n$, such that the function $x \mapsto \sum_n |f_n(x)|^2 = \|f(x)\|^2$ is $\mu$-integrable.

It is easy to see that if $f = \sum_n f_n a_n$, then $f_n \in L^{\mu}(X; \mathbb{C})$, and

$$\int_{G/H} \|f\|^2 \, d\mu = \sum_n \int_{G/H} |f_n|^2 \, d\mu;$$
see Dieudonné [22] (Chapter XIII, Sections 8 and 9). Then given two functions \( f = \sum_n f_n a_n \)
and \( g = \sum_n g_n a_n \) in \( L^2_\mu(X; E) \), the function \( x \mapsto \langle f(x), g(x) \rangle \) is integrable and
\[
\int_{G/H} \langle f(x), g(x) \rangle \, d\mu(x) = \sum_n \int_{G/H} f_n(x) \overline{g_n(x)} \, d\mu(x).
\]

**Definition 13.8.** We say that a function \( f \in L^2_\mu(X; E) \) is *negligeable* if the function \( x \mapsto \| f(x) \|^2 \) is zero almost everywhere.

The quotient of the space \( L^2_\mu(X; E) \) by the subspace of negligable functions is denoted by \( L^2_\mu(X; E) \). It is a hermitian space under the inner product
\[
\langle [f], [g] \rangle = \int_{G/H} \langle f(x), g(x) \rangle \, d\mu(x),
\]
and we have the norm \( \|\cdot\|_2 \) given by
\[
\|f\|_2 = \sqrt{\langle [f], [f] \rangle}.
\]

Actually, it turns out that the hermitian space \( L^2_\mu(X; E) \) is complete, that is, it is a Hilbert space. In fact, it is a separable Hilbert space.

**Proposition 13.8.** Let \( G \) be a locally compact group, let \( H \) be a closed subgroup of \( G \), let \( \mu \) be a positive Borel measure on \( X = G/H \), and let \( E \) be a separable Hilbert space. The space \( L^2_\mu(X; E) \) is a separable Hilbert space.

**Proof.** Let \( (f^{(m)}) \) be a Cauchy sequence in \( L^2_\mu(X; E) \), with \( f^{(m)} = \sum_n f^{(m)}_n a_n \). For every \( \epsilon > 0 \), there is some \( m_0 \) such that for all \( p, q \geq m_0 \), we have
\[
\sum_n \int_{G/H} |f^{(p)}_n - f^{(q)}_n|^2 \, d\mu \leq \epsilon, \quad (\ast_1)
\]
and this implies that for every \( n \), the sequence \( (f^{(m)}_n)_{m \geq 1} \) is a Cauchy sequence in \( L^2(X; \mathbb{C}) \).

Therefore, each sequence \( (f^{(m)}_n)_{m \geq 1} \) has a limit \( g_n \in L^2(X; \mathbb{C}) \), since \( L^2(X; \mathbb{C}) \) is complete by Fischer–Riesz. For every integer \( N > 0 \), if we let \( q \) tend to \( +\infty \) in \( (\ast_1) \), we see that
\[
\sum_{n=1}^N \|g_n - f^{(p)}_n\|^2 \leq \epsilon, \quad (\ast_2)
\]
so
\[
\sum_{n=1}^N \|g_n\|^2 \leq \sum_{n=1}^N \|g_n - f^{(p)}_n\|^2 + \sum_{n=1}^N \|f^{(p)}_n\|^2 \leq \epsilon + \|f^{(p)}\|^2.
\]
which proves that the series \( \sum_{n=1}^{\infty} \|g_n\|^2 \) converges. It follows that \( g = \sum_n g_n a_n \in L^2_\mu(X; E) \), and by \((*)_2\)
\[
\|g - f^{(p)}\|^2 = \sum_n \|g_n - f_n^{(p)}\|^2 \leq \epsilon
\]
for all \( p \geq m_0 \), and so \( g \) is the limit of the sequence \((f^{(m)})\) in \( L^2_\mu(X; E) \).

If \( D \) is a countable dense subset of \( L^2_\mu(X; \mathbb{C}) \), then we can check that the set of functions
\[
f = \sum f_n a_n \quad \text{such that} \quad f_n \in D \quad \text{for all} \quad n \quad \text{and} \quad f_n = 0 \quad \text{but all for finitely many values of} \quad n
\]
is dense in \( L^2_\mu(X; E) \).

13.5 Induced Representations, I; \( G/H \) has a \( G \)-Invariant Measure

We will now assume that the positive Borel measure \( \mu \) on \( X = G/H \) is \( G \)-invariant. Recall from Section 6.10 (Definition 6.16) that
\[
(\lambda_s(\mu))(A) = \mu(s^{-1} \cdot A),
\]
for every Borel subset \( A \) of \( X \), so \( \mu \) is \( G \)-invariant if for every Borel subset \( A \) of \( X \),
\[
\mu(s^{-1} \cdot A) = \mu(A) \quad \text{for all} \quad s \in G.
\]
In this case,
\[
\int_{G/H} f(s \cdot x) \, d\mu(x) = \int_{G/H} f(x) \, d\mu(x), \quad \text{for all} \quad s \in G.
\]

Let \( E \) be a separable Hilbert space, and let \( U : H \to U(E) \) be a unitary representation of \( H \).

**Theorem 13.9.** Let \( G \) be a locally compact group, \( H \) be a closed subgroup of \( G \), \( E \) be a separable Hilbert space, and \( U : H \to U(E) \) be a unitary representation of \( H \). If \( X = G/H \) admits a \( G \)-invariant \( \sigma \)-Radon measure \( \mu \), and for any cocycle \( \alpha : G \times X \to U(E) \), if the following conditions hold

1. We have \( \alpha(h, x_0) = U(h) \) for all \( h \in H \);
2. For every \( s \in G \), for every \( f \in L^2_\mu(X; E) \), the map \( x \mapsto \alpha(x, s)(f(x)) \) from \( X \) to \( E \) is \( \mu \)-measurable;
3. For every \( f \in L^2_\mu(X; E) \), the map \( s \mapsto \Pi_s(f) \) is a continuous map from \( G \) to \( L^2_\mu(X; E) \), where \( \Pi \) is the homomorphism \( \Pi : G \to GL(E^X) \) induced by the cocycle \( \alpha \);

then the homomorphism \( \Pi : G \to U(L^2_\mu(X; E)) \) induced by the cocycle \( \alpha \) given by
\[
(\Pi_s(f))(x) = (\alpha(s^{-1}, x))^{-1}(f(s^{-1} \cdot x)), \quad f \in L^2_\mu(X; E), \ x \in X,
\]
(see Definition 13.3) is a unitary representation of \( G \).
13.5. \textit{Induced Representations, I}; \textit{G/H Has a G-Invariant Measure} 491

Proof. We simply have to prove that
\[ \|\Pi_s(f)\|_2 = \|f\|_2, \]
for all \( f \in \mathcal{L}_\mu^2(X; E) \) and all \( s \in G \),
which implies that \( \Pi_s(f) \in \mathcal{L}_\mu^2(X; E) \), and the other conditions imply that the homomorphism \( \Pi: G \to \text{GL}(\mathcal{L}_\mu^2(X; E)) \) induced by \( \alpha \) is a unitary representation of \( G \). Since by hypothesis \( \alpha(s, s^{-1} \cdot x) \) is a unitary operator, we have
\[ \|\Pi_s(f)\|_2 = \|\alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x))\|_E = \|f(s^{-1} \cdot x)\|_E, \]
and since \( \mu \) is \( G \)-invariant,
\[ \int_{G/H} \|f(s^{-1} \cdot x)\|^2 \, d\mu(x) = \int_{G/H} \|f(x)\|^2 \, d\mu(x), \]
and so \( \|\Pi_s(f)\|_2 = \|f\|_2 \).
\[ \square \]

\textbf{Definition 13.9.} The unitary representation \( \Pi: G \to \text{U}(\mathcal{L}_\mu^2(X; E)) \) induced by the cocycle \( \alpha \) (and the unitary representation \( U: H \to \text{U}(E) \)) is denoted \( \text{Ind}_{G/H} \alpha \), or by abuse of notation even \( \text{Ind}_{G/H}^G U \).

\textbf{Remark:} To be very precise, the representing space \( \mathcal{L}_\mu^2(X; E) \) of this representation should be specified, for example as in \( \text{Ind}_{H,L_\mu^2(X; E)}^G \alpha \), because there are variants of this construction that use a different representation space.

If \( U \) is the trivial representation of \( H \) in \( E \), and if we choose \( \alpha(s, x) = \text{id}_E \) for all \( (s, x) \in G \times (G/H) \), then it can be verified that the hypotheses of Theorem 13.9 are satisfied. To verify Condition (3), we use the fact that the family of maps \( f \mapsto \Pi_s(f) \ (s \in G) \) is equicontinuous; see Proposition 1.13. Then we use Proposition 1.12; for details, see Dieudonné [20], (Chapter XXII, Section 3). In this case, the subspace \( L^\alpha \) corresponding to \( \mathcal{L}_\mu^2(X; E) \) consists of all functions of the form \( f \circ \pi \) with \( f \in \mathcal{L}_\mu^2(X; E) \), where \( \pi: G \to G/H \) is the projection map.

If \( H \) is a (closed) compact subgroup of \( G \), then by Proposition 6.42, the space \( G/H \) has \( G \)-invariant measures (unique up to a scalar). This is a special case of particular interest.

Under some mild additional conditions, induced unitary representations of \( G \) in \( \mathcal{L}_\mu^2(X; E) \) can converted to unitary representations of \( G \) in a closed subspace of \( \mathcal{L}_\lambda^2(G; E) \) (where \( \lambda \) is a left Haar measure on \( G \)).

Suppose that the cocycle \( \alpha \) has the property that the map
\[ s \mapsto f^\alpha(s) = \alpha(s^{-1}, s \cdot x_0)(f(s \cdot x_0)) \]
from \( G \) to \( E \) is \( \lambda \)-measurable for every \( f \in \mathcal{L}_\mu^2(X; E) \). If so, using Proposition 13.5 we have
\[ \|f^\alpha(sh)\| = \|f^\alpha(s)\| = \|f(s \cdot x_0)\| \]
for all \( s \in G \) and all \( h \in H \), and since by Proposition 6.42 we have
\[
\int_{G/H} f \, d\mu = \int_G (f \circ \pi) \, d\lambda,
\]
we conclude that \( f^\alpha \in \mathcal{L}^2_\lambda(G; E) \).

Conversely, if \( g \in \mathcal{L}^2_\lambda(G; E) \) satisfies the property
\[
g(sh) = U(h^{-1})(g(s)) \quad \text{for all } s \in G \text{ and all } h \in H,
\]
then if the map \( s \mapsto \alpha(s, x_0)(g(x)) \) from \( G \) to \( E \) is \( \lambda \)-measurable, then as in Proposition 13.6 we can write this map as \( f \circ \pi \) for some \( f \in L^2_\mu(X; E) \), and we have \( g = f^\alpha \).

In this case, up to equivalence, we can consider the unitary representation \( \text{Ind}_{H, F}^G \alpha \) induced by \( \alpha \) as a unitary representation of \( G \) in the closed subspace \( F \) of \( L^2_\lambda(G; E) \) spanned by the functions \( g \in \mathcal{L}^2_\lambda(G; E) \) satisfying property \((*_U)\), and then for all \( s \in G \),
\[
(\text{Ind}_{H, F}^G \alpha)_s(g) = \lambda_s g, \quad \text{for all } g \in F,
\]
equivalently, for all \( s, t \in G \),
\[
((\text{Ind}_{H, F}^G \alpha)_s(g))(t) = g(s^{-1}t), \quad \text{for all } g \in F.
\]
Notice the analogy with Proposition 13.7.

Note that \( \text{Ind}_{H, F}^G \alpha \) depends only on \( U \), so we usually write \( \text{Ind}_{H, F}^G U \) instead of \( \text{Ind}_{H, F}^G \alpha \).

If \( E = \mathbb{C} \), then \( \text{Ind}_{H, F}^G U \) is a subrepresentation of the regular representation of \( G \) in \( L^2(G) \).

**Definition 13.10.** If we choose \( U \) to be the trivial representation of \( H \) in \( E \), then the functions \( g \in \mathcal{L}^2_\lambda(G; E) \) satisfying Condition \((*_U)\) are constant on the classes \( sU \), so we can identify \( F \) with \( L^2_\mu(X; E) \). In this case we say that the induced representation \( \text{Ind}_{H}^G U \) of \( G \) in \( L^2_\mu(X; E) \) is the **canonical representation** of \( G \) corresponding to the compact subgroup \( H \) and to its trivial representation in \( E \).

If \( H = (e) \) and \( E = \mathbb{C} \), then the induced representation is the regular representation of \( G \) in \( L^2(G) \).

Going back to the case where \( H \) is an arbitrary closed subgroup of \( G \), and where there is a \( G \)-invariant measure on \( G/H \), there is another method, not using cocyles, for defining a unitary induced representation of \( G \) from a unitary representation \( U : H \to U(E) \). We can define a Hilbert space \( \mathcal{H} \) such that formula \((\text{Ind}_G)\) defines a unitary induced representation \( \text{Ind}_{H, \mathcal{H}}^G U \) of \( G \) in \( \mathcal{H} \). This method is described in Folland [31] (Chapter 6, Section 1), and we briefly describe it.
Given a unitary representation $U: H \to \mathbb{U}(E)$, let $\mathcal{H}_0$ be the following set of functions:

$$\mathcal{H}_0 = \{ f \in C(G, E) \mid \pi(\text{supp}(f)) \text{ is compact and } f(sh) = U(h^{-1})(f(s)) \text{ for all } s \in G \text{ and all } h \in H \}.$$ 

The problem is that it is not obvious that $\mathcal{H}_0$ is nonempty! However, the following result proven in Folland [31] (Chapter 6, Proposition 6.1) shows that this is not the case.

**Proposition 13.10.** If $\varphi: G \to E$ is a continuous function with compact support, then the function $f_\varphi$ from $G$ to $E$ given by

$$f_\varphi(s) = \int_H U(h)(\varphi(hs)) d\lambda_H(h)$$

belongs to $\mathcal{H}_0$ and is uniformly continuous on $G$. Moreover, every element of $\mathcal{H}_0$ is of the form $f_\varphi$ for some $\varphi \in K(G, E)$.

The group $G$ acts on the left on $\mathcal{H}_0$ by $f \mapsto \lambda_s f$. In order to act by unitary maps, we need to define an inner product on $\mathcal{H}_0$ with respect to which these left translations are isometries. Since $G/H$ has $G$-invariant measures, this is easy to achieve. If $f, g \in \mathcal{H}_0$, then the map $s \mapsto \langle f(s), g(s) \rangle_E$ depends only on the coset $sH$, so we can define the inner product $\langle f, g \rangle$ by

$$\langle f, g \rangle = \int_{G/H} \langle f(s), g(s) \rangle_E d\mu(sH).$$

This is clearly a positive hermitian form, and it is positive definite because $\mu(A) > 0$ for every nonempty open set $A$. This inner product is invariant under the left translations $\lambda_s$ because $\mu$ is $G$-invariant. Therefore, with respect to this inner product, the maps $f \mapsto \lambda_s f$ are unitary. If $\mathcal{H}$ is the Hilbert space which is the completion of $\mathcal{H}_0$, then the maps $f \mapsto \lambda_s f$ extend to unitary operators on $\mathcal{H}$. It follows from Proposition 13.10 that the map $s \mapsto \lambda_s f$ from $G$ to $\mathcal{H}$ are continuous for every $f \in \mathcal{H}_0$. Therefore, they define a unitary representation of $G$ in $\mathcal{H}$ given by

$$(\text{Ind}_{H/H}^G U)_s(f) = \lambda_s(f), \quad f \in \mathcal{H}.$$ 

This unitary representation has the advantage that it depends only on $U$, but one should not neglect the fact that the construction involving cocycles allows more flexibility. The Hilbert space $\mathcal{H}$ is also more complicated than the Hilbert space $L^2_\mu(X; E)$.

When $G/H$ admits no $G$-invariant measure, then we need to use a weaker notion of invariance. It turns out that the notion of (strong) quasi-invariance does the job.

### 13.6 Quasi-Invariant Measures on $G/H$

The notion of quasi-invariance was first introduced by Mackey and Bruhat in the early 1950’s. It also occurs in Bourbaki [6] (Chapter VII, §2, No. 5). We follow the exposition in Folland [31] (Chapter 2, Section 2.6, and Chapter 6, Section 1).
The idea is that if $\mu$ is a measure on $G/H$, then it is (strongly) quasi-invariant if there is a continuous function $\varphi: G \times (G/H) \to (0, \infty)$ such that

$$\int_{G/H} g(s \cdot x) d\mu(x) = \int_{G/H} \varphi(s, x) g(x) d\mu(x), \quad \text{for all } g \in \mathcal{K}_C(G/H) \text{ and all } s \in G. \quad (\text{qi}_\varphi)$$

The key to quasi-invariance is the existence of certain functions from $G$ to $(0, \infty)$ called rho-functions.

**Definition 13.11.** A function $\rho: G \to (0, \infty)$ is a rho-function for the pair $(G, H)$ if it is a continuous function such that

$$\rho(sh) = \frac{\Delta_H(h)}{\Delta_G(h)} \rho(s), \quad s \in G, h \in H, \quad (\ast)$$

where $\Delta_G$ is the modular function on $G$ and $\Delta_H$ is the modular function on $H$.

**Proposition 13.11.** If $G$ is any locally compact group and $H$ is any closed subgroup of $G$, then $(G, H)$ admits rho-functions.

Proposition 13.11 is proven in Folland [31] (Chapter 2, Proposition 2.54). One first proves ([31] (Chapter 2, Lemma 2.53) that there is a continuous function $\varphi: G \to (0, \infty)$ such that the following properties hold:

(i) $\{y \in G \mid \varphi(y) > 0\} \cap sH \neq \emptyset$ for all $s \in G$.

(ii) supp($\varphi$) $\cap KH$ is compact for every compact subset $K$ of $G$.

Then define $\rho$ by

$$\rho(s) = \int_H \frac{\Delta_G(h)}{\Delta_H(h)} \varphi(sh) d\lambda_H(h).$$

It is not hard to check that the above function is a rho-function.

Recall from Definition 6.18 the definition of the projection map $P: \mathcal{K}_C(G) \to \mathcal{K}_C(G/H)$ defined as follow: for every $f \in \mathcal{K}_C(G)$, for every $s \in G$, let

$$(P(f))(sH) = \int_H f(sh) d\lambda_H(h).$$

By Proposition 6.39, the map $P$ is surjective.

The next proposition is proven in Folland [31] (Chapter 2, Lemma 2.55).

**Proposition 13.12.** For any function $f \in \mathcal{K}_C(G)$, if $P(f) = 0$ then $\int f \rho \, d\lambda = 0$, for any rho-function $\rho$.

The proof of Proposition 13.12 is very similar to the argument given in the proof of Theorem 6.41. Then we have our first main theorem. Recall that $\pi: G \to G/H$ denotes the quotient map.
13.6. QUASI-ININVARIANT MEASURES ON $G/H$

**Theorem 13.13.** Let $G$ be any locally compact group and $H$ be any closed subgroup of $G$. For every rho-function $\rho$ for $(G,H)$, there is a unique quasi-invariant measure $\mu$ on $G/H$ such that

$$
\int_{G/H} P(f)(x) \, d\mu(x) = \int_G f(s) \rho(s) \, d\lambda(s), \quad \text{for all } f \in K_C(G).
$$

(qi)

Furthermore, if we let $\varrho : G \times (G/H) \to (0, \infty)$ be the continuous function given by

$$
\varrho(s, \pi(t)) = \frac{\rho(s^{-1}t)}{\rho(t)} \quad s, t \in G,
$$

then for every $g \in K_C(G/H)$, we have

$$
\int_{G/H} g(s \cdot x) \, d\mu(x) = \int_{G/H} \varrho(s, x) g(x) \, d\mu(x), \quad \text{for all } s \in G.
$$

(qi)\rho

**Proof.** Theorem 13.13 is proven in Folland [31] (Chapter 2, Theorem 2.56). For any $f \in K_C(G)$, since $P$ is surjective and since by Proposition 13.12, if $P(f) = P(g)$, then $\int_G f \rho \, d\lambda = \int_G g \rho \, d\lambda$, the map $\Phi$ given by $\Phi(P(f)) = \int_G f \rho \, d\lambda$ is a well-defined positive linear functional on $K_C(G/H)$. By Radon–Riesz I, it defines a unique $\sigma$-Radon measure $\mu$ on $G/H$ satisfying (qi).

The equation

$$
\rho(sh) = \frac{\Delta_H(h)}{\Delta_G(h)} \rho(s), \quad s \in G, h \in H,
$$

satisfied by a rho-function shows that the ratio $\rho(st)/\rho(t)$ depends only on the coset $\pi(t) = tH$, because

$$
\frac{\rho(st)}{\rho(th)} = \frac{\Delta_H(h)}{\Delta_G(h)} \frac{\rho(st)}{\rho(th)} = \frac{\Delta_H(h)}{\Delta_G(h)} \frac{\Delta_G(h)}{\Delta_H(h)} \frac{\rho(st)}{\rho(t)} = \frac{\rho(st)}{\rho(t)},
$$

so we obtain a continuous function $\varrho : G \times (G/H) \to (0, \infty)$ given by

$$
\varrho(s, \pi(t)) = \frac{\rho(s^{-1}t)}{\rho(t)} \quad s, t \in G.
$$

We have

$$
\int_{G/H} P(f)(s \cdot x) \, d\mu(x) = \int_G f(st) \rho(t) \, d\lambda(t)
$$

$$
= \int_G f(t) \rho(s^{-1}t) \, d\lambda(t)
$$

$$
= \int_G \varrho(s, \pi(t)) f(t) \rho(t) \, d\lambda(t)
$$

$$
= \int_{G/H} P(\varrho(s, \pi(-)) f)(x) \, d\mu(x)
$$

$$
= \int_{G/H} \varrho(s, x) P(f)(x) \, d\mu(x),
$$

where we used Proposition 6.37(3) to prove the last step, which concludes the proof. 

\[\square\]
CHAPTER 13. INDUCED REPRESENTATIONS

Remark: The map \( x \mapsto \varrho(s, x) \) is the Radon–Nikodym derivative of \( \lambda_s(\mu) \) with respect to \( \mu \).

The following converse of Theorem 13.13 is proven in Folland [31] (Chapter 2, Theorem 2.59).

**Theorem 13.14.** Let \( G \) be any locally compact group and \( H \) be any closed subgroup of \( G \). Every quasi-invariant measure \( \mu \) on \( G/H \) arises from a rho-function as in (qi). Furthermore, any two such measures \( \mu \) and \( \mu' \) are strongly equivalent, which means that there is a continuous function \( \varphi: G/H \to (0, \infty) \) such that \( \int_{G/H} g(x) \, d\mu'(x) = \int_{G/H} \varphi(x) g(x) \, d\mu(x) \) for all \( g \in K_c(G/H) \).

The following proposition shows that \( \varrho \) behaves like a cocycle.

**Proposition 13.15.** Let \( G \) be any locally compact group and \( H \) be any closed subgroup of \( G \). For any quasi-invariant measure \( \mu \) on \( G/H \) associated with the continuous function \( \varrho: G \times (G/H) \to (0, \infty) \), we have

\[
\varrho(st, x) = \varrho(s, t \cdot x) \varrho(t, x), \quad \text{for all } s, t \in G \text{ and all } x \in G/H \tag{*_{\varrho}}
\]

Proof. Using (qi), for every function \( g \in K_c(G/H) \), we have

\[
\int g(x) \varrho(st, x) \, d\mu(x) = \int g((st) \cdot x) \, d\mu(x) = \int g(s \cdot (t \cdot x)) \, d\mu(x) = \int \varrho(s, t \cdot x) g(t \cdot x) \, d\mu(x) = \int g(s) \varrho(s, t \cdot x) \varrho(t, x) \, d\mu(x)
\]

which proves that

\[
\varrho(st, x) = \varrho(s, t \cdot x) \varrho(t, x),
\]

as claimed. \( \square \)

Remark: Dieudonné denotes \( \varrho(s, x) \) by \( J_s(x) \); see Dieudonné [20] (Chapter XXII, Section 3, No. 22.3.8.1-22.3.8.2).

We now use quasi-invariant measures to generalize the construction of Section 13.5.

### 13.7 Induced Representations, II; \( G/H \) has a Quasi-Invariant Measure

If \( \mu \) is a quasi-invariant measure on \( G/H \), then by making a simple modification to Condition (1) of Theorem 13.9 we obtain the following result.
**Theorem 13.16.** Let $G$ be a locally compact group, $H$ be a closed subgroup of $G$, $E$ be a separable Hilbert space, and $U : H \to U(E)$ be a unitary representation of $H$. For any quasi-invariant measure $\mu$ on $X = G/H$ associated with the continuous function $\varrho : G \times (G/H) \to (0, \infty)$, for any cocycle $\alpha : G \times X \to U(E)$, if the following conditions hold

1. The map $\varrho(s^{-1}, x)^{1/2}\alpha(s, x)$ is a unitary map of $E$ for all $s \in G$ and all $x \in X$, such that $\varrho(h^{-1}, x_0)^{1/2}\alpha(h, x_0) = U(h)$ for all $h \in H$;
2. For every $s \in G$, for every $f \in L^2_\mu(X; E)$, the map $x \mapsto \alpha(x, s)(f(x))$ from $X$ to $E$ is $\mu$-measurable;
3. For every $f \in L^2_\mu(X; E)$, the map $s \mapsto \Pi_s(f)$ is a continuous map from $G$ to $L^2_\mu(X; E)$, where $\Pi$ is the homomorphism $\Pi : G \to GL(E^X)$ induced by the cocycle $\alpha$;

then the homomorphism $\Pi : G \to U(L^2_\mu(X; E))$ induced by the cocycle $\alpha$ given by

$$(\Pi_s(f))(x) = (\alpha(s^{-1}, x))^{-1}(f(s^{-1} \cdot x)), \quad f \in L^2_\mu(X; E), \ x \in X,$$

(see Definition 13.3) is a unitary representation of $G$.

**Proof.** We simply have to prove that

$$\|\Pi_s(f)\|_2 = \|f\|_2, \quad \text{for all } f \in L^2_\mu(X; E) \text{ and all } s \in G,$$

which implies that $\Pi_s(f) \in L^2_\mu(X; E)$, and the other conditions imply that the homomorphism $\Pi : G \to GL(L^2_\mu(X; E))$ induced by $\alpha$ is a unitary representation of $G$. Since by hypothesis $\varrho(s^{-1}, x)^{1/2}\alpha(s, x)$ is a unitary operator, using (qi), we have

$$\|\Pi_s(f)\|_2^2 = \int \|\alpha(s, s^{-1} \cdot x)(f(s^{-1} \cdot x))\|^2 d\mu(x)$$

$$= \int \varrho(s^{-1}, x) \|\alpha(s, x)(f(x))\|^2 d\mu(x)$$

$$= \int \|\varrho(s^{-1}, x)^{1/2}\alpha(s, x)(f(x))\|^2 d\mu(x)$$

$$= \int \|f(x)\|^2 d\mu(x),$$

and so $\|\Pi_s(f)\|_2 = \|f\|_2$. \qed

As an application of Theorem 13.16, we can pick

$$\alpha(s, x) = (\varrho(s^{-1}, x))^{-1/2+r} id_E,$$

with $r \in \mathbb{R}$. By Proposition 13.15, the function $\alpha$ is a cocycle. Condition (2) is satisfied because $\varrho$ is measurable (in fact, continuous). The maps $\varrho(s^{-1}, x)^{1/2}\alpha(s, x) = (\varrho(s^{-1}, x))^{r} id_E$ are unitary, since they are multiplication by a complex number of modulus 1. It remains
CHAPTER 13. INDUCED REPRESENTATIONS

to check Condition (3). This verification is performed in Dieudonné [20] (Chapter XXII, Section 3, No. 22.3.8.3).

Remark: It is possible to modify the construction of the Hilbert space $\mathcal{H}$ and the inner product described at the end of Section 13.5 to deal with the situation where $G/H$ has no $G$-invariant measure. This can be done in two ways as explained in Folland [31] (Chapter 6, Section 6.1). These constructions yield induced unitary representations of $G$ from a unitary representation $U : H \to U(E)$ of $H$ and do not involve cocycles.

In the first construction, the space $\mathcal{H}_0$ and the inner product are defined as before, namely

$$\mathcal{H}_0 = \{ f \in \mathcal{C}(G, E) \mid \pi(\text{supp}(f)) \text{ is compact and} \quad f(sh) = U(h^{-1})(f(s)) \quad \text{for all } s \in G \text{ and all } h \in H \},$$

and

$$\langle f, g \rangle = \int_{G/H} \langle f(s), g(s) \rangle_E \, d\mu(sH).$$

The Hilbert space $\mathcal{H}$ is the completion of $\mathcal{H}_0$. The new ingredient is that to make the operators $\Pi_s$ unitary, we use a quasi-invariant measure $\mu$ associated with a continuous function $\varrho : G \times (G/H) \to (0, \infty)$. We define $\Pi^\mu_s$ by

$$(\Pi^\mu_s(f))(t) = \varrho(s, tH)^{1/2}f(s^{-1}t), \quad f \in \mathcal{H}, \ s, t \in G.$$ 

Then, as in the proof of Theorem 13.16, we check that the operators $\Pi^\mu_s$ are unitary with respect to the inner product on $\mathcal{H}$ defined above, and we obtain a unitary representation $\Pi^\mu : G \to U(\mathcal{H})$, also denoted $\text{Ind}^{G,\mu}_{H,\mathcal{H}} U$ (for short $\text{Ind}^{G}_{H} U$). This representation depends on $\mu$, but it can be shown that if $\mu'$ is another quasi-invariant measure on $G/H$ associated with $\varrho' : G \times (G/H) \to (0, \infty)$, then the automorphism $f \mapsto (\rho'/\rho)^{1/2}f$ is a unitary equivalence of $\rho^\mu$ and $\rho'^\mu$, where $\rho$ and $\rho'$ are the rho-functions associated with $\mu$ and $\mu'$ (see Theorem 13.14).

The second construction, due to Blattner, does not make use of quasi-invariant measures, but instead modifies the definition of the space $\mathcal{H}_0$. In this sense, it is more intrinsic. Define the space $\mathcal{H}^0$ as

$$\mathcal{H}^0 = \{ f \in \mathcal{C}(G, E) \mid \pi(\text{supp}(f)) \text{ is compact and} \quad f(sh) = \left(\frac{\Delta_H(h)}{\Delta_G(h)}\right)^{1/2} U(h^{-1})(f(s)) \quad \text{for all } s \in G \text{ and all } h \in H \}.$$

Again, it is not obvious that $\mathcal{H}^0$ is not empty, but Proposition 13.10 can be modified (by adding the factor $\left(\frac{\Delta_G(h)}{\Delta_H(h)}\right)^{1/2}$ under the integral) to show that $\mathcal{H}^0$ is nonempty.

Next we need to define an inner product on $\mathcal{H}^0$ so that the operators $\Pi_s$ become unitary. The construction of such an inner product is a bit eccentric. For every $f \in \mathcal{H}^0$, the map
13.7. INDUCED REPRESENTATIONS, II; QUASI-ININVARIANT MEASURES

$s \mapsto \|f(s)\|_E^2$ is almost a rho-function, except that it is not strictly positive. However, it is still possible to prove that the map

$$P(\varphi) \mapsto \int_G \varphi(s) \|f(s)\|_E^2 d\lambda(s), \quad \varphi \in \mathcal{K}_C(G)$$

is a positive Radon functional on $\mathcal{K}_C(G/H)$, so by Radon–Riesz I, there is $\sigma$-Radon measure $\mu_f$ on $G/H$ such that

$$\int_{G/H} P(\varphi) \ d\mu_f = \int_G \varphi(s) \|f(s)\|_E^2 d\lambda(s),$$

for all $\varphi \in \mathcal{K}_C(G)$. Furthermore, the support of $\mu_f$ is contained in $\pi(supp(f))$, hence compact. Therefore, $\mu_f(G/H)$ is finite. Then, given $f, g \in \mathcal{H}^0$, define the complex measure $\mu_{f,g}$ by polarization as

$$\mu_{f,g} = \frac{1}{4}(\mu_{f+g} - \mu_{f-g} + i\mu_{f+ig} - i\mu_{f-ig}),$$

so that

$$\int_{G/H} P(\varphi) \ d\mu_{f,g} = \int_G \varphi(s) \langle f(s), g(s) \rangle_E d\lambda(s), \quad \varphi \in \mathcal{K}_C(G).$$

The inner product on $\mathcal{H}^0$ is defined as

$$\langle f, g \rangle = \mu_{f,g}(G/H).$$

It can be verified that we obtain a hermitian inner product, and we let $\mathcal{H}'$ be the Hilbert space completion of $\mathcal{H}^0$. Finally, it can be verified that the operators $\Pi_s$ are unitary with respect to the inner product on $\mathcal{H}'$ defined above, where

$$(\Pi_s(f))(t) = f(s^{-1}t), \quad f \in \mathcal{H}, \ s, t \in G.$$  

Therefore we obtain a unitary representation $\Pi' : G \to U(\mathcal{H}')$, also denoted $\text{Ind}_{G/H}^{\mathcal{H}'} U$ (for short $\text{Ind}_{G/H}^G U$).

It can also be shown that for any quasi-invariant measure $\mu$ on $G/H$, the representations $\Pi^\mu : G \to U(\mathcal{H})$ and $\Pi' : G \to U(\mathcal{H}')$ are equivalent (the isomorphism between $\mathcal{H}$ and $\mathcal{H}'$ is given by $f \mapsto \rho^{1/2} f$).

There is also an interpretation of the representations $\Pi^\mu$ and $\Pi'$ as sections of homogeneous hermitian vector bundles over $G/H$, but this would lead us to far; see Folland [31] (Chapter 6) or Kirillov [46] (Section 13).

We will now give several examples of the application of Theorem 13.16 to the group $\text{SL}(2, \mathbb{R})$. It turns out that the group $\text{SL}(2, \mathbb{R})$ has no finite-dimensional unitary representations except the trivial one, and Theorem 13.16 can be used to produce nontrivial unitary representations.
Example 13.1. Let $G = \text{SL}(2, \mathbb{R})$ and $H = S_1$ be the subgroup

$$S_1 = \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \big| a, b \in \mathbb{R}, a \neq 0 \right\},$$

and let $E = \mathbb{C}$. We claim that the homogeneous space $\text{SL}(2, \mathbb{R})/S_1$ is homeomorphic to $\mathbb{RP}^1(\mathbb{R}) = \mathbb{RP}^1$, the real projective line. Indeed, there is an action of $\text{SL}(2, \mathbb{R})$ on $\mathbb{RP}^1$ viewed as $\mathbb{R} \cup \{ \infty \}$ given by

$$s \cdot z = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d}, \quad z \in \mathbb{RP}^1,$$

with the convention that when $z = -d/c$, then the result is $\infty$, and when $z = \infty$, then the result if $a/c$. It is easy to check that this action is transitive and that the stabilizer of $\infty$ is the subgroup $S_1$. We give $\mathbb{RP}^1$ the measure $\mu$ which is the Lebesgue measure extended so that $\{\infty\}$ has measure zero. Since

$$\left( \begin{array}{cc} a & b \\ c & d \end{array} \right)^{-1} = \left( \begin{array}{cc} d & -b \\ -c & a \end{array} \right),$$

(recall that $ad - bc = 1$), and since the derivative of the function

$$x \mapsto \frac{dx - b}{-cx + a}$$

is

$$\frac{d(-cx + a) - (dx - b)(-c)}{(-cx + a)^2} = \frac{1}{(cx - a)^2},$$

we see that for any function $f \in L^2_\mu(\mathbb{RP}^1; \mathbb{C})$

$$\int_{-\infty}^{+\infty} f(s \cdot z) \, d\mu(z) = \int_{-\infty}^{+\infty} f\left( \frac{az + b}{cz + d} \right) \, d\mu(z) = \int_{-\infty}^{+\infty} \frac{1}{(cx - a)^2} f(x) \, d\mu(x).$$

It follows that $\mu$ is quasi-invariant with

$$\rho(s, x) = \frac{1}{(cx - a)^2}, \quad \text{where} \quad s = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The method of Theorem 13.16 with

$$\alpha(s, x) = (\rho(s^{-1}, x))^{-1/2 + (r/2)i} \text{id}_\mathbb{C}$$

where $r \in \mathbb{R}$, and with

$$(\Pi_s(f))(x) = (\alpha(s^{-1}, x))^{-1} (f(s^{-1} \cdot x)).$$
13.7. INDUCED REPRESENTATIONS, II; QUASI-ININVARIANT MEASURES

yields the unitary representations of $\text{SL}(2, \mathbb{R})$ in $L^2_{\mu}(\mathbb{RP}^1; \mathbb{C})$ given by

$$
\Pi_s(f)(x) = |cx-a|^{-1+i\xi} f\left(\frac{b-dx}{cx-a}\right), \quad f \in L^2_{\mu}(\mathbb{RP}^1; \mathbb{C}), \quad \text{where } s = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.
$$

It is also easy to check that the cocycles

$$
\alpha(s, x) = \left(\frac{1}{(cx-a)^2}\right)^{-1/2+(r/2)i} \text{sign}(cx-a) \text{id}_{\mathbb{C}}
$$

with $r \in \mathbb{R}$ also work, and we get the representations of $\text{SL}(2, \mathbb{R})$ in $L^2_{\mu}(\mathbb{RP}^1; \mathbb{C})$ given by

$$
\Pi_s(f)(x) = |cx-a|^{-1+i\xi} \text{sign}(cx-a) f\left(\frac{b-dx}{cx-a}\right), \quad f \in L^2_{\mu}(\mathbb{RP}^1; \mathbb{C}), \quad \text{where } s = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.
$$

It can be shown that all these representations are irreducible and pairwise inequivalent for $r > 0$. These representations constitute the principal series of irreducible unitary representations of $\text{SL}(2, \mathbb{R})$.

Example 13.2. Let $G = \text{SL}(2, \mathbb{R})$ and $H = \text{SO}(2)$ be the subgroup

$$
\text{SO}(2) = \left\{ \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \bigg| 0 \leq \theta \leq 2\pi \right\},
$$

and let $E = \mathbb{C}$. We claim that the homogeneous space $\text{SL}(2, \mathbb{R})/\text{SO}(2)$ is homeomorphic to the upper half plane $P = \{ z = x + iy \in \mathbb{C} \mid y > 0 \}$. Indeed, there is an action of $\text{SL}(2, \mathbb{R})$ on $P$ given by

$$
s \cdot z = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az+b}{cz+d}, \quad z = x + iy \in P,
$$

It is easy to check that this action is transitive and that the stabilizer of $z = i$ is $\text{SO}(2)$. Since the group $\text{SO}(2)$ is compact, the space $P = \text{SL}(2, \mathbb{R})/\text{SO}(2)$ admits $\text{SL}(2, \mathbb{R})$-invariant measures. In fact, the measure $\mu$ corresponding to the positive Radon functional

$$
h \mapsto \int_P h(x+i y) \frac{dx \, dy}{y^2} = \int_{y>0} \int_{x=-\infty}^{+\infty} h(x+i y) \frac{dx \, dy}{y^2}, \quad h \in K_{\mathbb{C}}(P)
$$

is such a measure.

We will need a method for picking a representative in every coset of $\text{SL}(2, \mathbb{R})/\text{SO}(2)$ that corresponds in a one-to-one fashion to an element $z = x + iy \in P$. For this, we use the fact that every matrix $s \in \text{SL}(2, \mathbb{R})$ can be uniquely factored as

$$
\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} \lambda & \mu \\ 0 & \lambda^{-1} \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} \lambda \cos \theta + \mu \sin \theta & -\lambda \sin \theta + \mu \cos \theta \\ \lambda^{-1} \sin \theta & \lambda^{-1} \cos \theta \end{pmatrix},
$$

with $\lambda, \mu \in \mathbb{R}$, $\lambda > 0$, and $0 \leq \theta < 2\pi$. 


Indeed, if there is such a decomposition, then
\[ c = \lambda^{-1} \sin \theta, \quad d = \lambda^{-1} \cos \theta, \]
so
\[ \sin \theta = \lambda c, \quad \cos \theta = \lambda d, \]
and since
\[
\begin{pmatrix}
\lambda & \mu \\
0 & \lambda^{-1}
\end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} = \begin{pmatrix} a \cos \theta - b \sin \theta & a \sin \theta + b \cos \theta \\ c \cos \theta - d \sin \theta & c \sin \theta + d \cos \theta \end{pmatrix},
\]
we see that
\[ \lambda^{-1} = c \sin \theta + d \cos \theta = \lambda(c^2 + d^2), \]
and since \( ad - bc = 1 \), we have \( c^2 + d^2 \neq 0 \), so we can pick
\[ \lambda = \frac{1}{\sqrt{c^2 + d^2}}, \]
and then \( \theta \in [0, 2\pi) \) is uniquely determined by
\[ \cos \theta = \frac{d}{\sqrt{c^2 + d^2}}, \quad \sin \theta = \frac{c}{\sqrt{c^2 + d^2}}, \]
and \( \mu \) is determined by
\[ \mu = a \sin \theta + b \cos \theta = \lambda(ac + bd) = \frac{ac + db}{\sqrt{c^2 + d^2}}. \]
Observe that the group
\[ S_0 = \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \mid a, b \in \mathbb{R}, a > 0 \right\} \]
is a subgroup of the group \( S_1 \) of Example 13.1.

Given any \( z = x + iy \in P \), there is a unique coset \( s\text{SO}(2) \subseteq \text{SL}(2, \mathbb{R}) \) (where \( s \in \text{SL}(2, \mathbb{R}) \)) that maps \( i \) to \( z \), and in view of the above factorization of matrices in \( \text{SL}(2, \mathbb{R}) \), we can pick as a representative of this coset \( s\text{SO}(2) \) the matrix \( r_z \in S_0 \) such that
\[ r_z \cdot i = z = x + iy, \]
namely
\[ r_z = \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix}. \]
We now determine \( h(s, z) \) such that \( sr_z = r_{s,z}h(s, z) \), with \( h(s, z) \in \text{SO}(2) \) and
\[ s = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \]
as in Section 13.2 (see Propositions 13.3 and 13.4). Since the imaginary part of \( s \cdot z = (az + b)/(cz + d) \) is \( y/|cz + d|^2 \), we have

\[
rs_z = \begin{pmatrix} \sqrt{y}/|cz + d| & * \\ 0 & |cz + d|/\sqrt{y} \end{pmatrix},
\]

so the equation \( sr_z = r_{s,z} h(s, z) \) with

\[
h(s, z) = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}
\]

namely

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix} = \begin{pmatrix} \sqrt{y}/|cz + d| & * \\ 0 & |cz + d|/\sqrt{y} \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}
\]

yields in particular

\[
c\sqrt{y} = |cz + d|\sin \theta, \quad cx + d = |cz + d|\cos \theta.
\]

Therefore,

\[
\cos \theta + i\sin \theta = \frac{cx + d + cy}{|cz + d|} = \frac{cz + d}{|cz + d|},
\]

equivalently

\[
e^{i\theta} = \frac{cz + d}{|cz + d|}.
\]

The group \( \text{SO}(2) \) is abelian, and since its unitary representations in \( \mathbb{C} \) are characters, by Proposition 9.8, they are of the form

\[
h \mapsto \sigma_n(h) = e^{ni\theta}, \quad n \in \mathbb{Z},
\]

with

\[
h = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}.
\]

By the method of Section 13.2, since \( e^{i\theta} = \frac{cz + d}{|cz + d|} \), we can pick the cocycle \( \alpha \) to be

\[
\alpha(s, z) = \sigma_n(h(s, z)) = \frac{(cz + d)^n}{|cz + d|^n} \text{id}_\mathbb{C},
\]

and then

\[
(\Pi_s(f))(z) = (\alpha(s^{-1}, x))^{-1}(f(s^{-1} \cdot z)), \quad f \in L^2_h(P; \mathbb{C}), \quad z \in P.
\]

Since

\[
s^{-1} = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix},
\]
we get
\[ \alpha(s^{-1}, z) = \frac{(cz - a)^n}{|cz - a|^n} \text{id}_\mathbb{C}, \]
and thus
\[ (\Pi_s(f))(z) = (cz - a)^{-n} g \left( \frac{b - dz}{cz - a} \right), \quad f \in L^2_P, \quad z \in P, \quad s = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \]
and for \( n \in \mathbb{Z} \).

These representations are not always irreducible. A way to see this is to construct an equivalent representation by using the cocycles \( \alpha'(s, z) = c(s \cdot z) \circ \alpha(s, z) \circ c(z)^{-1} \), as in Section 13.2, with
\[ c(z) = c(x + iy) = y^{-n/2}. \]
The image of \( L^2_\mu(P; \mathbb{C}) \) under the map \( f \mapsto cf \) is the space \( E_n \) of functions \( g: P \to \mathbb{C} \) such that the map \( z \mapsto y^n g(z)^2 \) is \( \mu \)-integrable. One can then show that the equivalent unitary representation is given by
\[ (\Pi_s(f))(z) = (cz - a)^{-n} g \left( \frac{b - dz}{cz - a} \right), \quad g \in E_n, \quad z \in P, \quad s = \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \]

It can be shown that for \( n > 1 \), the space \( H^2_n \) of holomorphic functions on \( P \) is nonempty and invariant under \( \Pi \). Furthermore, these representations in \( H^2_n \) are irreducible. The complex conjugates of these representations are also irreducible, and not equivalent to the previous ones; see Dieudonné [20] (Chapter XXII, Section 3).

These irreducible unitary representations of \( \text{SL}(2, \mathbb{R}) \) constitute the discrete series. There are other irreducible unitary representations of \( \text{SL}(2, \mathbb{R}) \) called the complementary series.

For comprehensive treatments of the irreducible representations of \( \text{SL}(2, \mathbb{R}) \) and other semisimple Lie groups, see Knapp [48], Vilenkin [82], and Taylor [78].

### 13.8 Partial Traces, Induced Representations of Compact Groups

In this section we consider a compact (metrizable) group \( G \) and a closed subgroup \( H \) of \( G \), and our goal is to determine the canonical (unitary) representation of \( G \) in \( L^2_\mu(G/H; \mathbb{C}) \) induced by the trivial representation of \( H \) in \( E = \mathbb{C} \) (see Definition 13.10), where \( \mu \) is the \( G \)-invariant measure on \( G/H \) induced by a Haar measure \( \lambda \) on \( G \). For simplicity of notation we write \( L^2_\mu(G/H) \) instead of \( L^2_\mu(G/H; \mathbb{C}) \) To do this it is necessary to understand what is the restriction of the representation \( M_\rho: G \to \text{U}(\mathbb{C}^n) \) to \( H \), with \( \rho \in R(G) \).

We will denote the set of equivalence classes of the irreducible representations of \( G \) given by the Peter-Weyl theorem I (Theorem 12.2) by \( \rho \in R(G) \), the corresponding representations
by $M_\rho: G \to U(\mathbb{C}^{n_\rho})$, and the identity element of $a_\rho$ by $u_\rho = \frac{1}{n_\rho} \chi_\rho$, where $\chi_\rho$ is the character associated with $\rho$. Similarly, we will denote the set of equivalence classes of the irreducible representations of $H$ given by the Peter-Weyl theorem I by $\sigma \in R(H)$, the corresponding representations by $M_\sigma: H \to U(\mathbb{C}^{n_\sigma})$, and the identity element of $a_\sigma$ by $u_\sigma = \frac{1}{n_\sigma} \chi_\sigma$, where $\chi_\sigma$ is the character associated with $\sigma$. The Haar measure on $G$ is denoted by $\lambda_G$ and the Haar measure on $H$ is denoted by $\lambda_H$.

Consider the restriction $V: H \to U(\mathbb{C}^{n_\rho})$ of the representation $M_\rho: G \to U(\mathbb{C}^{n_\rho})$ to $H$. Recall that for any function $f \in L^2(H)$, $V_{\text{ext}}(f)$ is the weak integral of the function $t \mapsto V(t)(x)$ with respect to $fd\lambda_H (t \in H)$. We will write $M_\rho(f)$ for $V_{\text{ext}}(f)$. By the Peter–Weyl theorem II (Theorem 12.15), for every $\sigma \in R(H)$, the map

$$M_\rho(u_\sigma) = \frac{1}{n_\sigma} \int_H M_\rho(t) \overline{\chi_\sigma(t)} \, d\lambda_H(t)$$

is the orthogonal projection of $\mathbb{C}^{n_\rho}$ onto a closed subspace $E_\sigma$ of $\mathbb{C}^{n_\rho}$, and we have a Hilbert sum

$$\mathbb{C}^{n_\rho} = \bigoplus_{\sigma \in R(H)} E_\sigma.$$ 

Furthermore, for each subspace $E_\sigma \neq (0)$, each irreducible representation $M_\sigma$ of $H$ is contained a certain number of times in the restriction of $M_\rho$ to $H$, which we denote $d = (\rho : \sigma)$, so $E_\sigma$ is a finite Hilbert sum

$$E_\sigma = \bigoplus_{k=1}^d F_k,$$

of subspaces $F_1, F_2, \ldots, F_d$ of dimension $n_\sigma$, invariant under $M_\rho(t)$ for every $t \in H$, and such that the restriction of $M_\rho$ to each $F_j$ is equivalent to the irreducible representation $M_\sigma$. Thus $E_\rho$ has dimension $p = dn_\sigma$.

We can pick an orthonormal basis of $\mathbb{C}^{n_\rho}$ consisting of the union of orthonormal bases of each of the $F_j$ and of a basis of the orthogonal complement $F'$ of $E_\sigma$ in $\mathbb{C}^{n_\rho}$, so that for the basis of $E_\sigma$ consisting of the first $p = dn_\sigma$ vectors of this basis, the matrix of the restriction of $M_\rho(t)$ to $E_\sigma$ is a block diagonal matrix of the form

$$
\begin{pmatrix}
M_\rho(t) & 0 & \cdots & 0 \\
0 & M_\rho(t) & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & M_\sigma(t)
\end{pmatrix}
$$

for every $t \in H$.

Consider the endomorphism $M_\rho(u_\sigma)M_\rho(s)M_\rho(u_\sigma)$ of $\mathbb{C}^{n_\rho}$, with $s \in G$, which can be viewed as an endomorphism $P_{\rho,\sigma}(s)$ of $E_\sigma$ since $M_\rho(u_\sigma)$ is the orthogonal projection of $\mathbb{C}^{n_\rho}$.
CHAPTER 13. INDUCED REPRESENTATIONS

with image $E_\rho$, whose matrix with respect to the above basis is

$$
\frac{1}{n_\rho} \begin{pmatrix}
m^{(\rho)}_{11}(s) & m^{(\rho)}_{12}(s) & \cdots & m^{(\rho)}_{1p}(s) \\

m^{(\rho)}_{21}(s) & m^{(\rho)}_{22}(s) & \cdots & m^{(\rho)}_{2p}(s) \\

\vdots & \vdots & \ddots & \vdots \\
m^{(\rho)}_{p1}(s) & m^{(\rho)}_{p2}(s) & \cdots & m^{(\rho)}_{pp}(s)
\end{pmatrix}.
$$

For any $t \in H$, the matrix of $P_{\rho,\sigma}(t)$ is the block diagonal matrix shown above, because the subspaces $F_k$ are invariant under $M_\rho(t)$ for $t \in H$. Since

$$M_\rho(tst') = M_\rho(t)M_\rho(s)M_\rho(t'),$$

and

$$M_\rho(t)M_\rho(u_\sigma) = M_\rho(u_\sigma)M_\rho(t) = \begin{pmatrix} P_{\rho,\sigma}(t) & 0 \\ 0 & 0 \end{pmatrix},$$

for all $s \in G$ and all $t, t' \in H$, we obtain the equation

$$P_{\rho,\sigma}(tst') = P_{\rho,\sigma}(t)P_{\rho,\sigma}(s)P_{\rho,\sigma}(t'), \quad \text{for all } s \in G \text{ and all } t, t' \in H. \quad (*)$$

**Definition 13.12.** The partial trace of $\rho$ with respect to $\sigma$ is the function

$$s \mapsto \theta_{\rho,\sigma}(s) = \text{tr}(M_\rho(u_\sigma)M_\rho(s)M_\rho(u_\sigma)) = \text{tr}(M_\rho(u_\sigma)M_\rho(s)),$$

which can also be expressed as

$$\theta_{\rho,\sigma}(s) = \frac{1}{n_\rho}(m^{(\rho)}_{11}(s) + \cdots + m^{(\rho)}_{pp}(s)).$$

The function $\theta_{\rho,\sigma}$ is continuous, no longer central, and not identically zero if $\sigma$ is contained in the restriction of $M_\rho$ to $H$. This function depends on $\rho$ and $\sigma$, and we have

$$\theta_{\rho,\sigma}(t) = (\rho : \sigma)\chi_\sigma(t) \quad \text{for all } t \in H$$

and

$$\chi_\rho(s) = \sum_\sigma \theta_{\rho,\sigma}(s) \quad \text{for all } s \in G.$$

It will be shown in Section 14.1 (see Example 14.5) that the partial traces for which $p = 1$ are the spherical functions when $(G, H)$ is a Gelfand pair.

We have the following proposition.

**Proposition 13.17.** The following properties hold.

1. We have

$$\theta_{\rho,\sigma}(tst^{-1}) = \theta_{\rho,\sigma}(s), \quad \text{for all } s \in G \text{ and all } t \in H.$$
13.8. PARTIAL TRACES, INDUCED REPRESENTATIONS OF COMPACT GROUPS 507

(2) When $\rho$ ranges over $R(G)$ and $\sigma$ ranges over $R(H)$, the partial traces $\theta_{\rho,\sigma}$ are pairwise orthogonal. In particular, $\theta_{\rho,\sigma}$ and $\theta_{\rho',\sigma'}$ can only be proportional if $\rho' = \rho$ and $\sigma' = \sigma$.

(3) The partial traces $\theta_{\rho,\sigma}$ are continuous functions of positive type.

Proof. (1) This equation follows immediately from $(\ast)$ and the commutativity of the trace.
(2) This follows from the equation

$$\theta_{\rho,\sigma}(s) = \frac{1}{n_{\rho}}(m_{11}^{(\rho)}(s) + \cdots + m_{pp}^{(\rho)}(s))$$

and the orthogonality properties of the $m_{ij}^{(\rho)}$; see Theorem 12.6.

(3) This follows from the properties $m_{ij}^{(\rho)} = \overline{m_{ij}^{(\rho)}} = m_{ij}^{(\rho)} * m_{ij}^{(\rho)}$ of Theorem 12.6, the fact that $f * \overline{f}$ is of positive type for every $f \in L^2(G)$, and the equation

$$\theta_{\rho,\sigma}(s) = \frac{1}{n_{\rho}}(m_{11}^{(\rho)}(s) + \cdots + m_{pp}^{(\rho)}(s)).$$

Since $G$ and $H$ are compact, $G/H$ has a $G$-invariant measure $\mu$ induced by a Haar measure on $G$. We now try to understand what the canonical unitary representation of $G$ in $L^2_{\mu}(G/H)$ induced by the trivial representation of $H$ in $E = \mathbb{C}$ looks like. With the notations as above, we have $n_{\sigma_0} = 1$, and $p = d$.

First, let us observe that a function $g \in L^2_{\mu}(G/H)$ can be viewed as a function $\overline{g} \in L^2(G)$ such that

$$g(st) = g(s) \quad \text{for all } t \in H \text{ and all } s \in G. \quad (\ast_{G/H})$$

Since $(g * \delta_t)(s) = g(st)$, the above condition is equivalent to

$$g * \delta_t = g \quad \text{for all } t \in H, \quad (\ast'_{G/H})$$

and thus for any measure $\nu \in M^1(G)$, the function $\nu * g \in L^2_{\mu}(G/H)$ also satisfies the equation

$$(\nu * g) * \delta_t = \nu * g,$$

so we deduce that $L^2_{\mu}(G/H)$ is a closed left ideal in $M^1(G)$, which implies that $L^2_{\mu}(G/H)$ is a closed left ideal in $L^2(G)$. In particular, for every $\rho \in R(G)$, the projection $g \mapsto u_{\rho} * g$ of $L^2(G)$ onto the ideal $a_{\rho}$ maps $L^2_{\mu}(G/H)$ onto itself, so $L^2_{\mu}(G/H)$ is the Hilbert sum of the subspaces

$$L_{\rho} = L^2_{\mu}(G/H) \cap a_{\rho}.$$

It remains to determine what the $L_{\rho}$ are. By Peter–Weyl II (Theorem 12.15), $\mathbb{C}^{n_{\rho}}$ is a Hilbert sum

$$\mathbb{C}^{n_{\rho}} = \bigoplus_{k=1}^{r} F_{\sigma_k},$$
of subspaces $F_{\sigma_k}$ invariant under the $M_\rho(t)$ for all $t \in H$, the subrepresentation corresponding to $F_{\sigma_k}$ being equivalent to the irreducible representation $M_{\sigma_k}$ of $H$. If we pick for an orthonormal basis of $\mathbb{C}^{n_\rho}$ the union of orthonormal bases of the $F_{\sigma_k}$, then for any $t \in H$ we have

$$M_\rho(t) = \begin{pmatrix} M_{\sigma_1}(t) & 0 & \cdots & 0 \\ 0 & M_{\sigma_1}(t) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & M_{\sigma_r}(t) \end{pmatrix}. $$

Proposition 13.18. The space $L^2_\mu(G/H)$ is the Hilbert sum of subspaces $L_\rho \subseteq a_\rho$. If the trivial representation $\sigma_0$ of $H$ is contained $d = (\rho : \sigma_0) \geq 1$ times in the restriction of $M_\rho$ to $H$, then $L_\rho$ is the direct sum of the first $d$ columns of $M_\rho$,

$$L_\rho = \bigoplus_{j=1}^d t_j^{(\rho)} \quad \text{and} \quad t_j^{(\rho)} = \bigoplus_{k=1}^{n_\rho} \mathbb{C} m_{kj}^{(\rho)}.$$

If $d = 0$, then $L_\rho = (0)$. The subrepresentation in $L_\rho$ of the canonical representation of $G$ in $L^2_\mu(G/H)$ induced by the trivial representation of $H$ in $\mathbb{C}$ is the Hilbert sum of $d = (\rho : \sigma_0)$ irreducible representations equivalent to $M_\pi$.

Proof. Since any function $g \in L^2(G/H) \cap a_\rho$ can be written as $g = \sum_{i,j} c_{ij} m_{ij}^{(\rho)}$, the equation $(*)_{G/H}$ and Theorem 12.6(4) yields

$$\sum_{i,j,k} c_{ij} m_{ik}^{(\rho)}(s) m_{kj}^{(\rho)}(t) = \sum_{i,j} c_{ij} m_{ij}^{(\rho)}(s),$$

with $1 \leq i, j, k \leq n_\rho$, all $s \in G$ and all $t \in H$, and since the $n_\rho^2$ functions $m_{ij}^{(\rho)}$ are linearly independent, we get

$$\sum_{j=1}^{n_\rho} c_{ij} m_{kj}^{(\rho)}(t) = c_{ik}, \quad \text{for all } i, k \text{ with } 1 \leq i, k \leq n_\rho \text{ and for all } t \in H. \quad (\star_1)$$

Suppose that the trivial representation $\sigma_0$ of $H$ is contained $d \geq 1$ times in the restriction of $M_\rho$ to $H$, which means that the first $d$ matrices $M_{\sigma_k}(t)$ are just one-dimensional matrices equal to 1, the other matrices being at least two-dimensional. We then have $m_{kj}^{(\rho)}(t) = \delta_{kj}$ for $k \leq d$ or $j \leq d$, hence $(\star_1)$ is trivially verified for $k \leq d$, and we are left with the equations

$$\sum_{j=d+1}^{n_\rho} c_{ij} m_{kj}^{(\rho)}(t) = c_{ik}, \quad k > d \text{ and } t \in H. \quad (\star_2)$$

Consider one of the matrices $M_{\sigma_k}(t)$ and assume it corresponds to the lines of index $k$ such that $k' \leq k \leq k''$. Then we have $m_{kj}^{(\rho)} = 0$ for $k' \leq k \leq k''$ and all $j$ except those for which
k'_h ≤ j ≤ k''_h; in addition, since σ ≠ σ_0, by the fact stated just after Definition 12.3, we have
$$\int_H m_{kj}^{(\rho)}(t) d\lambda_H(t) = 0$$
for all these indices. Integrating both sides of (*_2), we see that c_{ik} = 0 for all indices i and all k > d.

Therefore \( L_\rho \) is the subspace of \( a_\rho \), of dimension \( dn_\rho \), spanned by the \( m_{ij}^{(\rho)} \) such that \( j ≤ d \), equivalently, the direct sum of the first \( d \) columns of \( M_\rho \),

$$L_\rho = \bigoplus_{j=1}^{d} \tau_j^{(\rho)} \quad \text{and} \quad \tau_j^{(\rho)} = \bigoplus_{k=1}^{n_\rho} \mathbb{C} m_{kj}^{(\rho)}.$$

If \( d = 0 \), then the above reasoning shows that \( L_\rho = (0) \).

The canonical representation of \( G \) in \( L^2_\mu(G/H) \) induced by the trivial representation of \( H \) in \( \mathbb{C} \) is a subrepresentation of the regular representation of \( G \). We know from the discussion just after Definition 12.6 that on \( a_\rho \), the regular representation \( R \) splits into \( n_\rho \) irreducible representations all equivalent to \( M_\rho \), and can view these representation as acting on the columns of \( M_\rho \), the left ideals \( \tau_j^{(\rho)} \). Therefore, the subrepresentation in \( L_\rho \) of the canonical representation of \( G \) induced by the trivial representation of \( H \) in \( \mathbb{C} \) is the Hilbert sum of \((\rho : \sigma_0)\) irreducible representations equivalent to \( M_\rho \).

We can also consider the space \( H \setminus G \) of right cosets \( Hs \) of \( G \) \( (s ∈ G) \). If \( π: G → H \setminus G \) is the quotient map \( π(s) = Hs \), the fact that the Haar measure \( \lambda \) on a compact group is left and right invariant implies immediately that there is a \( G \)-invariant measure \( \mu' \) on \( H \setminus G \) such that
$$\int_{G/H} g(x) d\mu'(x) = \int_G (g \circ π) d\lambda,$$
and
$$\int_{G/H} g(x \cdot s) d\mu'(x) = \int_{G/H} g(x) d\mu'(x) \quad \text{for all} \ s ∈ G,$$
with
$$(Ht) \cdot s = Hts, \quad s, t ∈ G.$$

Every function \( g ∈ \mathcal{L}^2_\mu(H \setminus G) \) can be viewed as a function \( g ∈ \mathcal{L}^2(G) \) such that
$$g(ts) = g(s) \quad \text{for all} \ t ∈ H \text{ and all} \ s ∈ G. \quad \text{(*_H\setminus G)}$$

Since \((δ_t * g)(s) = g(t^{-1}s)\), the above condition is equivalent to
$$δ_t * g = g \quad \text{for all} \ t ∈ H. \quad \text{(**_H\setminus G)}$$
The space $L^2_{\mu'}(H \setminus G)$ is the image of the space $L^2_{\mu'}(G/H)$ under the isomorphism $g \mapsto \tilde{g}$ (here we use the fact that $G$ is unimodular). Therefore $L^2_{\mu'}(H \setminus G)$ is a closed right ideal in $L^2(G)$, and it is the Hilbert sum of the images $\tilde{L}_\rho$ of the $L_\rho$; since by Theorem 12.6(2) we have $m_{ji} = \tilde{m}_{ij}$, we deduce that $\tilde{L}_\rho$ is the direct sum of the first $d$ rows of $M_\rho$ (with $d = (\rho : \sigma_0)$).

Let us record this fact.

**Proposition 13.19.** The space $L^2_{\mu'}(H \setminus G)$ is the Hilbert sum of subspaces $\tilde{L}_\rho \subseteq a_\rho$. If the trivial representation $\sigma_0$ of $H$ is contained $d = (\rho : \sigma_0) \geq 1$ times in the restriction of $M_\rho$ to $H$, then $\tilde{L}_\rho$ is the direct sum of the first $d$ rows of $M_\rho$; that is,

$$\tilde{L}_\rho = \bigoplus_{i=1}^d \bigoplus_{j=1}^{n_\rho} \mathbb{C}m_{ij}^{(\rho)}.$$

Let us now consider the intersection $L^2_{\mu'}(G/H) \cap L^2_{\mu'}(H \setminus G)$. This is a closed involutive subalgebra of $L^2(G)$, thus a complete Hilbert algebra. We can view a function $g \in L^2_{\mu'}(G/H) \cap L^2_{\mu'}(H \setminus G)$ as a function $g \in L^2(G)$ such that $g(tst') = g(s)$ for all $t, t' \in H$ and all $s \in G$.  

(*$_{H \setminus G/H}$)

of equivalently

$$\delta_t \ast g \ast \delta_{t'} = g \quad \text{for all } t, t' \in H. \quad (*'_{H \setminus G/H})$$

We can also think of the functions $g \in L^2_{\mu'}(G/H) \cap L^2_{\mu'}(H \setminus G)$ as functions defined on the double classes $HsH$ of $G$ with respect to $H$.

We denote the algebra of functions in $L^2(G)$ satisfying (*$_{H \setminus G/H}$) as $L^2(H \setminus G/H)$. The following proposition follows immediately from the previous two propositions.

**Proposition 13.20.** The algebra $L^2(H \setminus G/H)$ is the Hilbert sum of the minimal two-sided ideals

$$a_{\rho, \sigma_0} = L_\rho \cap \tilde{L}_\rho = \bigoplus_{i=1}^d \bigoplus_{j=1}^d \mathbb{C}m_{ij}^{(\rho)}.$$

Each $a_{\rho, \sigma_0}$ is a matrix algebra of dimension $d^2$ having the family $(m_{ij}^{(\rho)})_{1 \leq i, j \leq d}$ as a basis. The center of $a_{\rho, \sigma_0}$ is the one-dimensional subspace

$$\mathbb{C}(m_{11}^{(\rho)} + \cdots + m_{dd}^{(\rho)}) = \mathbb{C}n_\rho \theta_{\rho, \sigma_0},$$

and $u_{\rho, \sigma_0} = m_{11}^{(\rho)} + \cdots + m_{dd}^{(\rho)}$ is the unit of $a_{\rho, \sigma_0}$. The map $g \mapsto u_{\rho, \sigma_0} \ast g = g \ast u_{\rho, \sigma_0}$ is the orthogonal projection of $L^2(H \setminus G/H)$ onto $a_{\rho, \sigma_0}$. 


The subspace
\[ t^{(\rho)}_{1, \sigma_0} = t^{(\rho)}_1 \cap a_{\rho, \sigma_0} = \mathbb{C}m^{(\rho)}_{11} \oplus \cdots \oplus \mathbb{C}m^{(\rho)}_{d_1} \]
is a minimal left ideal of \( L^2(H\backslash G/H) \). By Theorem 10.30, this ideal defines the irreducible representation \( W_\rho : L^2(H\backslash G/H) \to t^{(\rho)}_{1, \sigma_0} \) of the algebra \( L^2(H\backslash G/H) \) in \( t^{(\rho)}_{1, \sigma_0} \) (of dimension \( d \)), given by
\[ (W_\rho(f))(g) = f \ast g, \quad f \in L^2(H\backslash G/H), \; g \in t^{(\rho)}_{1, \sigma_0}. \]
From Theorem 10.30(2), up to equivalence, we obtain all irreducible representations of \( L^2(H\backslash G/H) \) in \( t^{(\rho)}_{1, \sigma_0} \) in this fashion. We can describe the representation \( W_\rho \) explicitly as follows. For every \( g \in L^2(H\backslash G/H) \), we can write
\[ g \ast u_{\rho, \sigma_0} = \sum_{1 \leq i,j \leq d} c_{ij}(g)m^{(\rho)}_{ij} \in a_{\rho, \sigma_0}, \]
and \( W_\rho(g) = (c_{ij}(g)) \), a \( d \times d \) matrix.

The above facts imply the following proposition.

**Proposition 13.21.** The algebra \( L^2(H\backslash G/H) \) is commutative if and only if \((\rho : \sigma_0) \leq 1\) for all \( \rho \in R(G) \). If so, then for every \( \rho \in R(G) \) such that \((\rho : \sigma_0) = 1\), the ideal \( a_{\rho, \sigma_0} \) is one-dimensional and is spanned by the function
\[ \omega_\rho(s) = \theta_{\rho, \sigma_0} = \frac{1}{n_\rho}m^{(\rho)}_{11}(s), \]
which is continuous and of positive type. Thus
\[ L^2(H\backslash G/H) = \bigoplus_{\rho|\rho(\sigma_0) = 1} \mathbb{C}\omega_\rho. \]
The orthogonal projection of \( L^2(H\backslash G/H) \) onto \( \mathbb{C}\omega_\rho \) is given by
\[ g \mapsto \omega_\rho \ast g = g \ast \omega_\rho. \]
The function \( \omega_\rho \) also satisfies the following equations:
\[ \omega_\rho(tst') = \omega_\rho(s), \quad \text{for all } s \in G \text{ and all } t,t' \in H \]
\[ \omega_\rho(e) = 1. \]

The function \( \omega_\rho \) is called a (zonal) spherical function.

The irreducible unitary representation \( W_\rho \) is one-dimensional, which implies that for every \( g \in L^2(H\backslash G/H) \) (since \( g \ast \omega_\rho \) is continuous), we have
\[ g \ast \omega_\rho = \zeta(g)\omega_\rho, \]
where \( \zeta \) is must be a character of \( L^2(H\backslash G/H) \) with values in \( \mathbb{U}(1) \) (because \( \omega_\rho \ast \omega_\rho = \omega_\rho \) and \( g \ast \omega_\rho = \omega_\rho \ast g \)). Since \( \zeta \) is an algebra homomorphism and \( \zeta(g) \in \mathbb{U}(1) \), we conclude that \( \zeta \) is a hermitian character of \( L^2(H\backslash G/H) \).

The left ideal \( L_\rho \) is minimal in \( L^2(H\backslash G/H) \), and by Proposition 12.4, it is spanned by the elements of the form \( \lambda_s\omega_\rho = \delta_s \ast \omega_\rho \), for all \( s \in G \).
Chapter 14

Harmonic Analysis on Gelfand Pairs

This chapter is the culmination of all of the theories discussed in this book. We are able to present a very general version of the Fourier transform on a homogeneous space $G/K$, where $(G, K)$ is a Gelfand pair. This chapter presents material discussed in Dieudonné [20] (Chapter XXII, Sections 6-10).

We saw in Section 13.8 that if $G$ is a compact group and if $H$ is a closed subgroup of $G$, then the algebra $L^2(H \backslash G/H)$ is commutative if and only if $(\rho: \sigma_0) \leq 1$ for all $\rho \in R(G)$ (where $\sigma_0$ is the class of the trivial representation of $H$). If so, then for every $\rho \in R(G)$ such that $(\rho: \sigma_0) = 1$, the ideal $a_{\rho, \sigma_0}$ is one-dimensional and is spanned by the function

$$\omega_{\rho}(s) = \theta_{\rho, \sigma_0} = \frac{1}{n_{\rho}} m_{11}^{(\rho)}(s),$$

which is continuous and of positive type. The function $\omega_{\rho}$ is called a (zonal) spherical function.

The goal of this chapter is to generalize the above results for a compact group to a locally compact (metrizable and separable) unimodular group $G$, and to a compact subgroup $K$ of $G$.

The first difficulty is that if $G$ is not compact, then $L^2(G)$ is not closed under convolution (in general, $L^2(G)$ is not contained in $L^1(G)$). So we have to work with $\mathcal{K}(G)$ instead (recall that $\mathcal{K}(G)$ is the subset of $\mathcal{C}(G)$ consisting of the continuous functions with compact support $f: G \to \mathbb{C}$).

There is a bijection between the space $\mathcal{C}(G/K)$ of continuous functions $f: G/K \to \mathbb{C}$ and the subspace of continuous functions $g: G \to \mathbb{C}$ such that

$$g(st) = g(s), \quad \text{for all } s \in G \text{ and all } t \in K.$$

We also have a bijection between the space $\mathcal{C}(K \backslash G)$ of continuous functions $f: K \backslash G \to \mathbb{C}$ and the subspace of continuous functions $g: G \to \mathbb{C}$ such that

$$g(ts) = g(s), \quad \text{for all } s \in G \text{ and all } t \in K.$$
Let $\mathcal{C}(K\backslash G/K) = \mathcal{C}(G/K) \cap \mathcal{C}(K\backslash G)$, which consists of the continuous functions $f : G \to \mathbb{C}$ which are constant on double cosets $KsK$ ($s \in G$), and let $\mathcal{K}(K\backslash G/K)$ be the subspace of $\mathcal{C}(K\backslash G/K)$ consisting of the continuous functions with compact support. The space $\mathcal{K}(K\backslash G/K)$ is an involutive subalgebra of $\mathcal{K}$, and thus of $L^1(G)$.

The key ingredient is the Banach algebra $L^1(K\backslash G/K)$, the closure of $\mathcal{K}(K\backslash G/K)$ in $L^1(G)$. Gelfand’s remarkable discovery is that much of the harmonic analysis on abelian locally compact groups and compact groups can be generalized to a pair $(G, K)$ where $G$ is a noncommutative locally compact unimodular (metrizable and separable) group, and $K$ is a compact subgroup of $G$, if the algebra $L^1(K\backslash G/K)$ is commutative. In this case, $(G, K)$ is called a Gelfand pair.

Fortunately, there is a sufficient criterion for a pair $(G, K)$ to be a Gelfand pair involving an involutive isomorphism $\sigma : G \to G$ such that $K$ is the set of fixed points of $\sigma$ (see Theorem 14.2). This criterion is reminiscent of Élie Cartan’s notion of symmetric space (see Helgason [42] or Gallier and Quaintance [36]), and indeed, many kinds of symmetric spaces are Gelfand pairs. The proof that a pair satisfying this criterion is a Gelfand pair is given in Section 14.1. The conditions of this criterion are flexible enough to apply to three broad classes of pairs $(G, K)$; see Section 14.4.

As seen from the preceding paragraph, an important class of Lie groups that yield Gelfand pairs are the real forms of a complex semi-simple Lie group, and Section 14.2 is devoted to a discussion of these Lie groups. One needs to understand how to find the real forms $g_0$ of a complex Lie algebra $g$, that is, the real Lie algebras $g_0$ such that

$$g = g_0 \oplus ig_0,$$

namely, $g$ is the complexification of $g_0$. Finding such real algebras $g_0$ is equivalent to finding certain semilinear idempotent maps on $g$ called conjugations (see Proposition 14.4). Now, if we happen to have some semi-simple real form $g_u$ such that

$$g = g_u \oplus ig_u,$$

and if $c_u$ is the corresponding conjugation, then it can be shown that all the other real forms $g_0$ of $g$ are given by conjugations $c_0$ that commute with $c_u$. Then all this has to be promoted to Lie groups (essentially by using the exponential map).

The purpose of Section 14.3 is to characterize the characters of the algebra $L^1(K\backslash G/K)$ in terms of certain functions in $\mathcal{C}(K\backslash G/K)$ called spherical functions. Every character $\zeta$ of the commutative Banach algebra $L^1(K\backslash G/K)$ is given by a unique function $\omega \in \mathcal{C}(K\backslash G/K)$ which is bounded and continuous on $G$, with

$$\zeta(f) = (f, \omega) = \int_G f(x)\omega(x) \, d\lambda_G(x), \quad f \in L^1(K\backslash G/K);$$

see Proposition 14.10. A function $\omega \in \mathcal{C}(K\backslash G/K)$ as above is called a spherical function.
Two criteria for a bounded function $\omega \in C(K\backslash G/K)$ (different from the zero function) to be a spherical function are given in Theorem 14.12. In particular, the function $\omega$ is a spherical function on $G$ relative to $K$ iff

$$
\int_K \omega(xty) \, d\lambda_K(t) = \omega(x)\omega(y) \quad \text{for all } x, y \in G.
$$

(s1)

The space of spherical functions on the Gelfand pair $(G, K)$ is denoted $S(G/K)$. The subspace of characters of the commutative involutive Banach algebra $A = L^1(K\backslash G/K) \oplus \mathbb{C}\delta_e$ whose restriction to $L^1(K\backslash G/K)$ is not the zero function is denoted by $X_0(A)$. This subspace is locally compact in the weak $^*$-topology (metrizable and separable).

The map $\omega \mapsto \zeta_\omega = (f, \omega)$ is a homeomorphism of $S(G/K)$ equipped with the induced topology of Fréchet space of $C(G)$ onto $X_0(A)$ equipped with the topology induced by the weak $^*$-topology of the dual $A'$ of $A$. Consequently, $S(G/K)$ is locally compact.

Three examples of Gelfand pairs are discussed in Section 14.4. In the first example, $G$ is a compact Lie group that has an involutive automorphism $\sigma$; this corresponds to symmetric spaces of compact type. In the second example, $G_1$ arises as a real form of a complex, semi-simple, simply-connected Lie group $G$, and $G_1$ has finite center. This corresponds to a symmetric space of noncompact type. The third example is a certain kind of semi-direct product; a typical illustration of this case is the group of rigid motions $\text{SE}(n, \mathbb{R})$.

The Fourier transform and the Fourier cotransform for a Gelfand pair are introduced in Section 14.5. For every function $f \in L^1(K\backslash G/K)$, the Fourier cotransform $\overline{F}f$ of $f$ is the function $\overline{F}f : S(G/K) \to \mathbb{C}$ given by

$$(\overline{F}f)(\omega) = (f, \omega) = \int_G f(x)\omega(x) \, d\lambda_G(x), \quad \omega \in S(G/K),$$

and the Fourier transform $Ff$ of $f$ is the function $Ff : S(G/K) \to \mathbb{C}$ given by

$$(Ff)(\omega) = (\tilde{f}, \omega) = \int_G f(x^{-1})\omega(x) \, d\lambda_G(x) = \int_G f(x)\omega(x^{-1}) \, d\lambda_G(x), \quad \omega \in S(G/K).$$

Using the space $S(G/K)$ of spherical functions as the domain of $F$ and $\overline{F}$ instead of characters yields a simultaneous generalization of the case where $G$ is commutative and the case where $G$ is compact. On $L^1(K\backslash G/K)$, we have the familiar equations

$$F(f \ast g) = (Ff)(Fg), \quad \overline{F}(f \ast g) = (\overline{F}f)(\overline{F}g).$$

In Section 14.6 we generalize Fourier inversion. For this, we use the construction of certain positive $\sigma$-Radon measures from measures of positive type (recall Definition 11.14) using the Plancherel transform.

When $G$ is not compact, the spherical functions in $S(G/K)$ are not necessarily of positive type. The subset of $S(G/K)$ consisting of the spherical functions of positive type is denoted by $Z(G/K)$. 

CHAPTER 14. HARMONIC ANALYSIS ON GELFAND PAIRS

Theorem 14.20 states that given a measure \( \mu \) of positive type on \( G \), there is a unique (positive) Radon measure \( \mu^\triangle \) on \( Z(G/K) \) such that for every function such that for every function \( f \in \mathcal{K}(K/G/K) \), the Fourier cotransform \( \mathcal{F}f \) belong to \( L^2_{\mu^\triangle}(Z(G/K); \mathbb{C}) \), and for any two functions \( f, g \in \mathcal{K}(K/G/K) \), we have

\[
\int_{G} (g^* \ast f) d\mu = \int_{Z(G/K)} (\mathcal{F}f)(\omega)(\mathcal{F}g)(\omega) d\mu^\triangle(\omega).
\]

The measure \( \mu^\triangle \) is called the Plancherel transform of \( \mu \).

In particular, if \( \mu = \delta_e \), the Dirac measure, we obtain a measure \( m_Z = \delta_e^\triangle \), called the canonical measure on \( Z(G/K) \), and then the linear maps \( f \mapsto \mathcal{F}f \) and \( f \mapsto \mathcal{F}f \), with \( f, g \in \mathcal{K}(K/G/K) \), are isometries, and by Theorem 14.20, these maps extend to isomorphisms from the Hilbert space \( L^2(K\backslash G/K) \) onto the Hilbert space \( L^2_{m_Z}(Z(G/K)) \). This is a generalization of the Plancherel theorem (Theorem 9.26).

Another type of Fourier inversion formula is given by Proposition 14.25. The map \( p \mapsto (p \lambda_G)^\triangle \) is a bijection between the space \( \mathcal{P}_+(K\backslash G/K) \) of functions in \( \mathcal{C}(K\backslash G/K) \) which are of positive type onto the space \( \mathcal{M}_1^+(Z(G/K)) \) of bounded positive measures on \( Z(G/K) \).

Section 14.7 discusses an extension of the Plancherel transform to the space \( \mathcal{P}(G) \) which is the complex vector space spanned by the union of the complex measures and the Radon measures. In Proposition 14.31 we obtain a Fourier inversion formula which yields the inversion formula of the Pontrjagin duality theorem, Theorem 9.29, as a special case.

Finally, in Section 14.8, we show that functions of positive type induce irreducible representations; see Theorem 14.32. We also state a theorem of Stone characterizing the unitary representations of \( \mathbb{R} \) in a separable Hilbert space (Theorem 14.37).

14.1 Gelfand Pairs

Recall that there is a bijection between the space \( \mathcal{C}(G/K) \) of continuous functions \( f : G/K \to \mathbb{C} \) and the subspace of continuous functions \( g : G \to \mathbb{C} \) such that

\[
g(st) = g(s), \quad \text{for all } s \in G \text{ and all } t \in K,
\]

equivalently

\[
g \ast \delta_t = g, \quad \text{for all } t \in K,
\]

given by the map \( f \mapsto f \circ \pi \), where \( \pi : G \to G/K \) is the projection map. Observe that \( (g \ast \delta_t)(s) = g(st^{-1}) \) (since \( G \) is unimodular, the term \( \Delta(t^{-1}) \) is equal to 1), so the condition \( g \ast \delta_t = g \) is equivalent to \( g(st^{-1}) = g(s) \) for all \( t \in K \), but this is equivalent to \( g(st) = g(s) \) for all \( t \in K \) since \( K \) is a group. From now on, we identify \( \mathcal{C}(G/K) \) with the subspace of \( \mathcal{C}(G) \) satisfying the above equivalent properties.
We also have a bijection between the space $\mathcal{C}(K\backslash G)$ of continuous functions $f : K\backslash G \to \mathbb{C}$ and the subspace of continuous functions $g : G \to \mathbb{C}$ such that

$$g(ts) = g(s), \quad \text{for all } s \in G \text{ and all } t \in K,$$

equivalently
$$\delta_t * g = g, \quad \text{for all } t \in K,$$

and we identify $\mathcal{C}(K\backslash G)$ with the subspace of $\mathcal{C}(G)$ satisfying the above equivalent properties. Observe that $(\delta_t * g)(s) = g(t^{-1}s)$, so $\delta_t * g = g$ is equivalent to $g(t^{-1}s) = g(s)$ for all $t \in K$, which is equivalent to $g(ts) = g(s)$ for all $t \in K$.

Let $\mathcal{C}(K\backslash G/K) = \mathcal{C}(G/K) \cap \mathcal{C}(K\backslash G)$, which consists of the continuous functions $f : G \to \mathbb{C}$ which are constant on double cosets $KsK$ ($s \in G$).

Since $K$ is compact, for every compact subset $A$ of $G/K$, the subset $\pi^{-1}(A)$ is compact in $G$. The map $f \mapsto f \circ \pi$ is thus a bijection between the subspace $\mathcal{K}(G/K)$ of $\mathcal{C}(G/K)$ onto a subspace of $\mathcal{K}(G)$. This subspace is denoted by $\mathcal{K}(G) \cap \mathcal{C}(G/K)$. Similarly, there is a bijection between the space $\mathcal{K}(K\backslash G/K)$ and the space $\mathcal{K}(G) \cap \mathcal{C}(K\backslash G/K)$, and a bijection between the space $\mathcal{K}(K\backslash G/K)$ and the space $\mathcal{K}(G) \cap \mathcal{C}(K\backslash G/K)$.

We saw earlier that $\mathcal{K}(G)$ is an involutive subalgebra under convolution of the algebra $L^1(G)$ (see Example 8.3(4)). It follows that $\mathcal{K}(G/K)$ is left ideal in $\mathcal{K}(G)$ and that $\mathcal{K}(K\backslash G)$ is right ideal in $\mathcal{K}(G)$, and the involution $f \mapsto \overline{f}$ maps $\mathcal{K}(G/K)$ onto $\mathcal{K}(K\backslash G)$. As a consequence, $\mathcal{K}(K\backslash G/K)$ is an involutive subalgebra of $\mathcal{K}(G)$ (and so of $L^1(G)$).

Let $\lambda_K$ be the Haar measure on $K$ normalized so that $\lambda_K(K) = 1$, and since $G$ is assumed to be unimodular, let $\lambda_G$ be a left and right-invariant Haar measure on $G$. In order to study the characters of the Banach algebra $L^1(K\backslash G/K)$, which is the closure of $\mathcal{K}(K\backslash G/K)$ in $L^1(G)$, we need to project $\mathcal{C}(G)$ onto $\mathcal{C}(K\backslash G/K)$.

**Definition 14.1.** We define a projection map from $\mathcal{C}(G)$ onto $\mathcal{C}(K\backslash G/K)$ by

$$f^\sharp(s) = \int_K \int_K f(tst') \, d\lambda_K(t) \, d\lambda_K(t'), \quad s \in G,$$

for any $f \in \mathcal{C}(G)$.

It is easily checked that if $f \in \mathcal{C}(G)$, then $f^\sharp(t_1st'_1) = f^\sharp(s)$ for all $t_1, t'_1 \in K$, so $f^\sharp \in \mathcal{C}(K\backslash G/K)$. As a consequence, since $\lambda(K) = 1$,

$$f^{\sharp\sharp}(s) = \int_K \int_K f^\sharp(tst') \, d\lambda_K(t) \, d\lambda_K(t') = \int_K \int_K f^\sharp(s) \, d\lambda_K(t) \, d\lambda_K(t') = f^\sharp(s),$$

and the map $f \mapsto f^\sharp$ is indeed a projection. It is also easy to check that for all $f \in \mathcal{C}(K\backslash G/K)$ and for all $g \in \mathcal{C}(G)$, we have

$$(fg)^\sharp = fg^\sharp.$$
**Proposition 14.1.** The restriction of the projection \( f \mapsto f^\sharp \) to \( \mathcal{K}(G) \) maps \( \mathcal{K}(G) \) onto \( \mathcal{K}(K \backslash G/K) \), and we have

\[
(f * g)^\sharp = f * g^\sharp, \quad (g * f)^\sharp = g^\sharp * f,
\]

for all \( f \in \mathcal{K}(K \backslash G/K) \), and all \( g \in \mathcal{C}(G) \).

**Proof.** We have

\[
(f * g)(x) = \int_G f(s)g(s^{-1}x) \, d\lambda_G(s),
\]

and using the left invariance of \( \lambda_G \) and the fact that \( f(ts) = f(s) \) and \( \lambda_K(K) = 1 \), we have

\[
(f * g)^\sharp(x) = \int_K \int_K \int_G f(s)g(s^{-1}txt') \, d\lambda_G(s) \, d\lambda_K(t) \, d\lambda_K(t')
\]

\[
= \int_K \int_K \int_G f(ts)g(s^{-1}xt') \, d\lambda_G(s) \, d\lambda_K(t) \, d\lambda_K(t')
\]

\[
= \int_K \int_K \int_G f(s)g(s^{-1}xt') \, d\lambda_G(s) \, d\lambda_K(t) \, d\lambda_K(t')
\]

\[
= \int_K \int_G f(s)g(s^{-1}xt') \, d\lambda_G(s) \, d\lambda_K(t').
\]

We also have

\[
g^\sharp(x) = \int_K \int_K g(txt') \, d\lambda_K(t) \, d\lambda_K(t'),
\]

and using the right invariance of \( \lambda_G \) and the fact that \( f(st) = f(s) \) and \( \lambda_K(K) = 1 \), we have

\[
(f * g^\sharp)(x) = \int_G \int_K \int_K f(s)g(ts^{-1}xt') \, d\lambda_K(t) \, d\lambda_K(t') \, d\lambda_G(s)
\]

\[
= \int_G \int_K \int_K f(st)g(s^{-1}xt') \, d\lambda_K(t) \, d\lambda_K(t') \, d\lambda_G(s)
\]

\[
= \int_G \int_K \int_K f(s)g(s^{-1}xt') \, d\lambda_K(t) \, d\lambda_K(t') \, d\lambda_G(s)
\]

\[
= \int_G \int_K f(s)g(s^{-1}xt') \, d\lambda_K(t') \, d\lambda_G(s),
\]

and by Fubini, we can interchange the integrations, which shows that \( (f * g)^\sharp = f * g^\sharp \). The proof that \( (g * f)^\sharp = g^\sharp * f \) is analogous.

A Gelfand pair \((G, K)\) is defined as follows.

**Definition 14.2.** Let \( G \) be a locally compact (metrizable and separable) unimodular group, and let \( K \) be a compact subgroup \( K \). We say that the pair \((G, K)\) is a **Gelfand pair** if the algebra \( \mathcal{K}(K \backslash G/K) \) is commutative (under convolution).
14.1. GELFAND PAIRS

Obviously, if $G$ is abelian, then $(G, K)$ is a Gelfand pair. The following theorem, which is one of the key theorems of the theory of Gelfand pairs, gives a sufficient criterion for a pair $(G, K)$ to be a Gelfand pair. Using this criterion, we will show later that there are lots of Gelfand pairs.

**Theorem 14.2.** (Gelfand) Let $G$ be locally compact (metrizable and separable) unimodular group, and let $\sigma : G \to G$ be an involutive isomorphism of $G$ $(\sigma \circ \sigma = \text{id}_G)$ having the following properties:

1. The subgroup $K = \{s \in G \mid \sigma(s) = s\}$ of elements of $G$ left fixed by $\sigma$ is compact.
2. Every $x \in G$ can be written (possibly not uniquely) as $x = yz$, with $y \in K$, $z \in G$, and $\sigma(z) = z^{-1}$.

Then $(G, K)$ is Gelfand pair.

**Proof.** Since $\sigma^2 = \text{id}_G$, we have $\text{mod}(\sigma)^2 = \text{mod(id)}_G = 1$, and since $\text{mod}(\sigma) > 0$, we get $\text{mod}(\sigma) = 1$ (see Definition 6.15 and Proposition 6.31). It follows that $\sigma$ leaves any Haar measure $\lambda_G$ of $G$ invariant. For every $f \in \mathcal{K}(G)$, let $f^\sigma$ be the function given by

$$f^\sigma(s) = f(\sigma(s)) = f(\sigma^{-1}(s)), \quad s \in G.$$ 

We check immediately that the map $\widehat{\sigma} : \mathcal{K}(G) \to \mathcal{K}(G)$ given by $\widehat{\sigma}(f) = f^\sigma$ is an involutive automorphism of the vector space $\mathcal{K}(G)$, and since $\sigma$ leaves any Haar measure on $G$ invariant, it is an automorphism of the algebra $\mathcal{K}(G)$ (under convolution). Since $\sigma(t) = t$ for all $t \in K$, if $f(x) = f(txt')$ for all $t, t' \in K$, then $f(\sigma(txt')) = f(t\sigma(x)t') = f(\sigma(x))$, so if $f \in \mathcal{K}(K\backslash G/K)$ then $f^\sigma \in \mathcal{K}(K\backslash G/K)$, which means that the automorphism $\widehat{\sigma}$ leaves the algebra $\mathcal{K}(K\backslash G/K)$ invariant, and its restriction to $\mathcal{K}(K\backslash G/K)$ is an automorphism of this subalgebra. If we can prove that

$$f^\sigma * g^\sigma = g^\sigma * f^\sigma \quad \text{for all } f, g \in \mathcal{K}(K\backslash G/K),$$

then we will have proven that $\mathcal{K}(K\backslash G/K)$ is commutative.

The trick is that for any function $f \in \mathcal{K}(K\backslash G/K)$, we have $f^\sigma = \tilde{f}$. Every $x \in G$ can be written as $x = yz$ with $y \in K$ and $\sigma(z) = z^{-1}$, which yields

$$\sigma(x) = \sigma(yz) = \sigma(y)\sigma(z) = yz^{-1} = y(z^{-1}y^{-1})y,$$

and for every $f \in \mathcal{K}(K\backslash G/K)$, we have

$$f(\sigma(x)) = f(y(z^{-1}y^{-1})y) = f(z^{-1}y^{-1}) = f(x^{-1}) = \tilde{f}(x);$$

that is, $f^\sigma = \tilde{f}$. But $L^1(G)$ is an involutive algebra, so for any two function $f, g \in \mathcal{K}(G)$, we have

$$(\tilde{f} \ast \tilde{g}) = g \ast f$$

because $G$ is unimodular so the Haar measure is invariant under the map $s \mapsto s^{-1}$. Then for $f, g \in \mathcal{K}(K\backslash G/K)$, we have

$$f^\sigma \ast g^\sigma = \tilde{f} \ast \tilde{g} = (g \ast f)^\sigma = (g \ast f) = g^\sigma \ast f^\sigma,$$

as claimed. □
From now on we assume that \((G, K)\) is a Gelfand pair.

**Definition 14.3.** The closure of \(\mathcal{K}(K \backslash G / K)\) in \(L^1(G)\) is denoted \(L^1(K \backslash G / K)\). Then \(L^1(K \backslash G / K)\) is an involutive, commutative, Banach subalgebra of \(L^1(G)\). Similarly, denote by \(L^2(K \backslash G / K)\) the closure of \(\mathcal{K}(K \backslash G / K)\) in \(L^2(G)\).

The following results are proven in Dieudonné [20] (Chapter XXII, Section 6).

**Proposition 14.3.** The projection \(f \mapsto f^\sharp\) of \(\mathcal{K}(G)\) onto \(\mathcal{K}(K \backslash G / K)\) extends to a continuous projection of \(L^1(G)\) onto \(L^1(K \backslash G / K)\). For any \(f \in L^1(G)\), we have \(\|f^\sharp\|_1 \leq \|f\|_1\), and the class \([f^\sharp]\) is the class of the function \(f^\sharp\) equal almost everywhere to the function

\[
s(t) \mapsto \int_K \int_K f(tst') \, d\lambda_K(t) \, d\lambda_K(t').
\]

Similarly, the projection \(f \mapsto f^\sharp\) of \(\mathcal{K}(G)\) onto \(\mathcal{K}(K \backslash G / K)\) extends to a continuous projection of \(L^2(G)\) onto \(L^2(K \backslash G / K)\). For any \(f \in L^2(G)\), we have \(\|f^\sharp\|_2 \leq \|f\|_2\), and the class \([f^\sharp]\) is the class of the function \(f^\sharp\) equal almost everywhere to the function

\[
s(t) \mapsto \int_K \int_K f(tst') \, d\lambda_K(t) \, d\lambda_K(t').
\]

**Definition 14.4.** We denote by \(L^1(G / K)\) and \(L^2(G / K)\) the subspaces of \(L^1(G)\) and \(L^2(G)\) consisting of the functions \(f\) such that for almost all \(s \in G\), we have

\[
f(st) = f(s) \quad \text{for all} \quad t \in K.
\]

Similarly, we denote by \(L^1(K \backslash G)\) and \(L^2(K \backslash G)\) the subspaces of \(L^1(G)\) and \(L^2(G)\) consisting of the functions \(f\) such that for almost all \(s \in G\), we have

\[
f(ts) = f(s) \quad \text{for all} \quad t \in K.
\]

Let \(L^1(K \backslash G / K) = L^1(G / K) \cap L^1(K \backslash G)\) and \(L^2(K \backslash G / K) = L^2(G / K) \cap L^2(K \backslash G)\).

If \(f \in L^1(G)\) (resp. \(f \in L^2(G)\)), then \(f^\sharp \in L^1(K \backslash G / K)\) (resp. \(f^\sharp \in L^2(K \backslash G / K)\)), and \(L^1(K \backslash G / K)\) (resp. \(L^2(K \backslash G / K)\)) is the canonical image in \(L^1(G)\) (resp. \(L^2(G)\)) of \(L^1(K \backslash G / K)\) (resp. \(L^2(K \backslash G)\)).

We also obtain an alternative description of \(L^1(K \backslash G / K)\) (resp. \(L^2(K \backslash G / K)\)) in terms of \(L^1(G / K)\) and \(L^1(K \backslash G)\) (resp. \(L^2(G / K)\) and \(L^2(K \backslash G)\)).

If we denote by \(L^1(G / K)\), \(L^1(K \backslash G)\) (resp. \(L^2(G / K)\), \(L^2(K \backslash G)\)) the canonical images in \(L^1(G)\) (resp. \(L^2(G)\)) of \(L^1(G / K)\), \(L^1(K \backslash G)\) (resp. \(L^2(G / K)\), \(L^2(K \backslash G)\)), then we have \(L^1(K \backslash G / K) = L^1(G / K) \cap L^1(K \backslash G)\) (resp. \(L^2(K \backslash G / K) = L^2(G / K) \cap L^2(K \backslash G)\)).

In order to present some of the examples of Gelfand pairs involving Lie groups, we need to discuss some material about semi-simple Lie groups.
14.2 Real Forms of a Complex Semi-Simple Lie Group

First we need to review the process of “complexifying” a real vector space \( V \). We can define the complexification \( V_C \) of the real vector space \( V \) as the complex vector space whose carrier is the tensor product \( \mathbb{C} \otimes_{\mathbb{R}} V \), but more directly as \( V \times V \), with the addition operation

\[(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2), \quad u_1, u_2, v_1, v_2 \in V,\]

and the scalar product given by

\[(a + ib)(u, v) = (au - bv, av + bu), \quad a, b \in \mathbb{R}, u, v \in V.\]

Observe that

\[(0, v) = i(v, 0),\]

so we can write

\[(u, v) = (u, 0) + i(v, 0),\]

and if \( j: V \to V_C \) is the injection given by \( j(u) = (u, 0) \), for all \( u \in V \), then we have an isomorphism

\[V_C \cong V \oplus iV.\]

By abuse of notation, we usually write \( V_C = V \oplus iV \). Using the above isomorphism, the scalar multiplication of a vector \( u + iv \in V_C \) by a complex number \( a + ib \) is given by

\[(a + ib)(u + iv) = au - bv + i(av + bu).\]

The map \( c_V: V_C \to V_C \) given by

\[c_V(u + iv) = u - iv, \quad u, v \in V,\]

is semi-linear, and an involution; that is, \( c_V^2 = \text{id}_{V_C} \). The map \( c_V \) is called the conjugation of \( V_C \) associated with \( V \). Observe that

\[V = \{w \in V_C \mid c_V(w) = w\}, \quad iV = \{w \in V_C \mid c_V(w) = -w\}.\]

To simplify notation we usually write \( c \) instead of \( c_V \).

**Definition 14.5.** If \( V \) is a complex vector space, then we denote by \( V|_{\mathbb{R}} \) the vector space whose underlying abelian group is \( V \), but with the scalar multiplication restricted to \( \mathbb{R} \).

If \( \mathfrak{g} \) is a real Lie algebra, then its complexification \( \mathfrak{g}_C \) is complex Lie algebra whose carrier is the complex vector space \( \mathfrak{g} + i\mathfrak{g} \), with the Lie bracket given by

\[[u + iv, x + iy]_C = [u, x] - [v, y] + i([u, y] + [v, x])].\]
If $c$ is the conjugation of $g_C$ associated with $g$, then
\[
    c([u + iv, x + iy]_C) = c([u, x] - [v, y] + i([u, y] + [v, x]))
    = c([u, x]) - c([v, y]) - i(c([u, y]) + c([v, x]))
    = [u, x] - [v, y] - i([u, y] + [v, x])
    = [u - iv, x - iy]_C = [c(u + iv), c(x + iy)]_C.
\]
This shows that $c$ is an automorphism of the real Lie algebra $g_C|_R$.

Definition 14.6. Given a complex Lie algebra $g$, a real Lie algebra $g_0$ such that $g = g_0 \oplus ig_0$ is called a real form of $g$. The complex Lie algebra $g$ is the complexification of $g_0$.

The following proposition shows that finding a real form of a complex Lie algebra $g$ is equivalent to finding an automorphism $c$ of the real Lie algebra $g|_R$ satisfying the properties stated in the following proposition.

Proposition 14.4. Let $g$ be a complex Lie algebra, $g_0$ be a real Lie algebra, and assume that $g = g_0 \oplus ig_0$; that is, $g$ is the complexification of $g_0$. Then the conjugation $c_0 : g \rightarrow g$ given by $c_0(u + iv) = u - iv$ ($u, v \in g_0$) has the following properties:

(a) The map $c_0$ is semi-linear, which means that $c_0(x + y) = c_0(x) + c_0(y)$, and $c_0(ix) = -ic_0(x)$, for all $x, y \in g$.

(b) The map $c_0$ is idempotent; that is, $c_0^2 = \text{id}_g$.

(c) We have
\[
    c_0([x, y]) = [c_0(x), c_0(y)], \quad \text{for all } x, y \in g.
\]

Conversely, if a map $c : g \rightarrow g$ has Properties (a), (b), (c), then if we consider the linear automorphism $c : g|_R \rightarrow g|_R$ of $g$ viewed as a real vector space, and if we let
\[
    g_1 = \{x \in g \mid c(x) = x\},
\]
the subspace $g_1$ is a real Lie subalgebra of $g|_R$, and we have
\[
    g = g_1 \oplus ig_1;
\]
that is, $g$ is the complexification of $g_1$.

Proof. We already proved the first part of the proposition. Conversely, since $c$ is an involutive automorphism of $g|_R$, we know by linear algebra that
\[
    g = g_1 \oplus g_{-1}
\]
where \( g_1 \) and \( g_{-1} \) are the real eigenspaces of \( c \) given by
\[
g_1 = \{ x \in g \mid c(x) = x \}, \quad g_{-1} = \{ x \in g \mid c(x) = -x \}.
\]
Since \( c \) is semi-linear, for every \( x \in g_1 \), we have
\[
c(ix) = -ic(x) = -ix,
\]
so \( ix \in g_{-1} \), which shows that
\[
i g_1 \subseteq g_{-1} \tag{\ast}
\]
For every \( x \in g_{-1} \), we have
\[
c(ix) = -ic(x) = ix
\]
so \( ix \in g_1 \), which shows that
\[
i g_{-1} \subseteq g_1 \tag{\ast\ast}
\]
But the above inclusion implies that
\[
g_{-1} \subseteq i g_1 \tag{\ast\ast\ast}
\]
By (\ast) and (\ast\ast\ast), we get
\[
g_{-1} = i g_1,
\]
and so
\[
g = g_1 \oplus i g_1,
\]
as claimed. Since \( c \) is the identity on \( g_1 \) and satisfies (c), we conclude that \( g_1 \) is a (real) subalgebra of \( g \).

**Definition 14.7.** Given a complex Lie algebra \( g \), a map \( c_0: g \to g \) satisfying Conditions (a), (b), (c) of Proposition 14.4 is called a *conjugation*.

If \( g_u \) is a real semi-simple Lie algebra and if \( g = g_u \oplus i g_u \) is the complex Lie algebra which is its complexification, then it is possible to determine all the real forms \( g_0 \) of \( g \) (up to isomorphism). The conjugation \( c_u: g \to g \) associated with \( g_u \) is given by
\[
c_u(x + iy) = x - iy, \quad x, y \in g_u.
\]
It can be proven that in order to find all conjugations \( c_0: g \to g \) of \( g \), it suffices to consider conjugations that commute with \( g_u \); see Dieudonné [19] (Chapter XXI, no. 21.18.3).

Let \( c_0: g \to g \) be a conjugation of \( g \) that commutes with \( c_u \). In this case, for any \( x \in g_u \), since
\[
c_0(x) = c_0(c_u(x)) = c_u(c_0(x)),
\]
we see that \( c_0(x) \in g_u \), so \( g_u \) is invariant under \( c_0 \), and similarly, for any \( x \in g_u \), since
\[
c_0(ix) = -c_0(c_u(ix)) = -c_u(c_0(ix)),
\]
so \( i\mathfrak{g}_u \) is also invariant under \( c_0 \).

Since the restriction of \( c_0 \) to \( \mathfrak{g}_u \) is an involutive automorphism of \( \mathfrak{g}_u \), we know by linear algebra that

\[
\mathfrak{g}_u = E_1 \oplus E_{-1},
\]

where \( E_1 \) and \( E_{-1} \) are the real eigenspaces of \( c_0 \) given by

\[
E_1 = \{ x \in \mathfrak{g}_u \mid c_0(x) = x \}, \quad E_{-1} = \{ x \in \mathfrak{g}_u \mid c_0(x) = -x \}.
\]

It is customary to denote \( E_1 \) by \( \mathfrak{k}_0 \) and \( E_{-1} \) by \( \mathfrak{i}p_0 \), where both are real vector spaces, so that

\[
\mathfrak{g}_u = \mathfrak{k}_0 \oplus \mathfrak{i}p_0,
\]

and \( i\mathfrak{g}_u = i\mathfrak{k}_0 \oplus \mathfrak{p}_0 \). Since \( c_0 \) is semi-linear, for any \( ix \in i\mathfrak{p}_0 \), we have

\[
-ix = c_0(ix) = -ic_0(x),
\]

so \( c_0(x) = x \) if \( x \in \mathfrak{p}_0 \), and for \( ix \in i\mathfrak{k}_0 \), we have

\[
c_0(ix) = -ic_0(x) = -ix,
\]

since \( c_0(x) = x \) for \( x \in \mathfrak{k}_0 \). Consequently the (real) Lie algebra \( \mathfrak{g}_0 \) of fixed points of \( c_0 \) is

\[
\mathfrak{g}_0 = \mathfrak{k}_0 \oplus \mathfrak{p}_0,
\]

and

\[
\mathfrak{g} = \mathfrak{g}_0 + i\mathfrak{g}_0 = \mathfrak{k}_0 \oplus \mathfrak{p}_0 \oplus i\mathfrak{k}_0 \oplus i\mathfrak{p}_0 = \mathfrak{k}_0 \oplus i\mathfrak{p}_0 \oplus i\mathfrak{k}_0 \oplus \mathfrak{p}_0 = \mathfrak{g}_u \oplus i\mathfrak{g}_u,
\]

so \( \mathfrak{g}_0 \) is a semi-simple Lie algebra.

**Definition 14.8.** The decomposition

\[
\mathfrak{g}_0 = \mathfrak{k}_0 \oplus \mathfrak{p}_0
\]

is called a *Cartan decomposition* of the real Lie algebra \( \mathfrak{g}_0 \) (with respect to the conjugation \( c_0 \)).

**Example 14.1.** Consider the real Lie algebra \( \mathfrak{g}_u = \mathfrak{su}(n) \) of \( n \times n \) complex skew-hermitian matrices with zero trace,

\[
\mathfrak{su}(n) = \{ X \in \mathfrak{m}_n(\mathbb{C}) \mid X^* = -X, \, \text{tr}(X) = 0 \}.
\]

We claim that the complexification \( \mathfrak{g} = \mathfrak{su}(n)_\mathbb{C} \) of \( \mathfrak{su}(n) \) is the complex Lie algebra \( \mathfrak{sl}(n, \mathbb{C}) \) of \( n \times n \) complex matrices with zero trace,

\[
\mathfrak{sl}(n, \mathbb{C}) = \{ X \in \mathfrak{m}_n(\mathbb{C}) \mid \text{tr}(X) = 0 \}.
\]
First observe that $\mathfrak{isu}(n)$ is the (real) vector space (not a Lie algebra) of hermitian matrices with zero trace,

$$\mathfrak{isu}(n) = \{ X \in M_n(\mathbb{C}) \mid X^* = X, \, \text{tr}(X) = 0 \}.$$ 

Indeed, if $X^* = -X$, then $(iX)^* = -iX^* = iX$, and if \(\text{tr}(X) = 0\), then $\text{tr}(iX) = i\text{tr}(X) = 0$, so

$$\mathfrak{isu}(n) \subseteq \{ X \in M_n(\mathbb{C}) \mid X^* = X, \, \text{tr}(X) = 0 \}.$$

Conversely, if $X^* = X$ then $(iX)^* = -iX^* = -iX$, and if $\text{tr}(X) = 0$, then $\text{tr}(iX) = i\text{tr}(X) = 0$, so

$$i\{ X \in M_n(\mathbb{C}) \mid X^* = X, \, \text{tr}(X) = 0 \} \subseteq \mathfrak{su}(n).$$

But the above equation implies that

$$\{ X \in M_n(\mathbb{C}) \mid X^* = X, \, \text{tr}(X) = 0 \} \subseteq \mathfrak{isu}(n),$$

so

$$\mathfrak{isu}(n) = \{ X \in M_n(\mathbb{C}) \mid X^* = X, \, \text{tr}(X) = 0 \}.$$

Every complex matrix $X \in M_n(\mathbb{C})$ can be written as

$$X = \frac{1}{2}(X + X^*) + \frac{1}{2}(X - X^*),$$

and we have

$$\frac{1}{2}(X + X^*)^* = \frac{1}{2}(X^* + X^*) = \frac{1}{2}(X^* + X) = \frac{1}{2}(X + X^*),$$

and

$$\frac{1}{2}(X - X^*)^* = \frac{1}{2}(X^* - X^*) = \frac{1}{2}(X^* - X) = -\frac{1}{2}(X - X^*).$$

Also, if $\text{tr}(X) = 0$, then

$$\text{tr}\left(\frac{1}{2}(X + X^*)\right) = \frac{1}{2}(\text{tr}(X) + \text{tr}(X^*)) = \frac{1}{2}(\text{tr}(X) + \text{tr}(X)) = 0,$$

and

$$\text{tr}\left(\frac{1}{2}(X - X^*)\right) = \frac{1}{2}(\text{tr}(X) - \text{tr}(X^*)) = \frac{1}{2}(\text{tr}(X) - \text{tr}(X)) = 0.$$

Thus if $X \in \mathfrak{sl}(n, \mathbb{C})$, then $\frac{1}{2}(X + X^*) \in \mathfrak{su}(n)$ and $\frac{1}{2}(X - X^*) \in \mathfrak{isu}(n)$, which proves that

$$\mathfrak{sl}(n, \mathbb{C}) = \mathfrak{su}(n) \oplus \mathfrak{isu}(n).$$

The sum is a direct sum because the only matrix such that $X^* = -X$ and $X^* = X$ is the matrix $X = 0$.

Since $\mathfrak{su}(n) = \{ X \in \mathfrak{sl}(n, \mathbb{C}) \mid X^* = -X \}$, the conjugation $c_u$ of $\mathfrak{sl}(n, \mathbb{C})$ associated with $\mathfrak{su}(n)$ is the map given by $c_u(X) = -X^*$.
We now consider three types of conjugations on $g = \mathfrak{sl}(n, \mathbb{C})$ which lead to interesting real forms of $\mathfrak{sl}(n, \mathbb{C})$.

**Example 14.2.** Consider the conjugation $c_0$ of $\mathfrak{sl}(n, \mathbb{C})$ given by $c_0(X) = \overline{X}$. Obviously, $c_0$ commutes with $c_a$. The restriction of $c_0$ to $\mathfrak{g}_u = \mathfrak{su}(n)$ is also $c_0$, and we obtain

$$\mathfrak{e}_0 = \mathfrak{so}(n) = \{X \in \mathfrak{su}(n) \mid \overline{X} = X\}$$

$$i\mathfrak{p}_0 = \{X \in \mathfrak{su}(n) \mid \overline{X} = -X\},$$

where $\mathfrak{so}(n)$ is the Lie algebra of $n \times n$ real skew-symmetric matrices.

But for any $X = (x_{jk}) \in \mathfrak{su}(n)$, if $x_{jk} = a_{jk} + ib_{jk}$, with $a_{jk}, b_{jk} \in \mathbb{R}$, since $X^* = -X$, we have $x_{kj} = -a_{jk} + ib_{kj}$, so $a_{kk} = 0$. If we also have $\overline{X} = -X$, then $a_{jk} + ib_{jk} = -a_{jk} + ib_{jk}$, so $a_{jk} = 0$ for all $j, k$, which means that $X = (ib_{jk})$, with $(b_{jk})$ a real symmetric matrix. Thus

$$\mathfrak{p}_0 = \mathfrak{s}(n) = \{X \in \mathfrak{M}_n(\mathbb{R}) \mid X^\top = X\},$$

the vector space of $n \times n$ real symmetric matrices (not a Lie algebra). We obtain

$$\mathfrak{g}_0 = \mathfrak{e}_0 \oplus \mathfrak{p}_0 = \mathfrak{so}(n) \oplus \mathfrak{s}(n) = \mathfrak{sl}(n, \mathbb{R}),$$

with

$$\mathfrak{sl}(n, \mathbb{R}) = \{X \in \mathfrak{M}_n(\mathbb{R}) \mid \text{tr}(X) = 0\}.$$

It turns out that a Gelfand pair arises from two Lie groups $G_0$ and $K_0$ whose Lie algebras are $\mathfrak{sl}(n, \mathbb{R})$ and $\mathfrak{sl}(n, \mathbb{R}) \cap \mathfrak{su}(n) = \mathfrak{so}(n)$. To describe these groups, first we need to consider the complex simply-connected Lie group $G = \mathbf{SL}(n, \mathbb{C})$ whose complex Lie algebra is $\mathfrak{g} = \mathfrak{sl}(n, \mathbb{C})$. A complex Lie group $G$ of (complex) dimension $n$ can also be viewed as a real Lie group of (real) dimension $2n$ denoted $G|_{\mathbb{R}}$, by viewing a holomorphic chart $\varphi : U \to \mathbb{C}^n$ of $G$ as a real smooth function $\varphi : U \to \mathbb{R}^{2n}$. More generally, a complex manifold $M$ of (complex) dimension $n$ can be viewed as a real manifold of dimension $2n$ denoted $M|_{\mathbb{R}}$.

Using the correspondence between simply-connected real Lie groups and real Lie algebras, there is a unique automorphism $\sigma_0 : \mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}} \to \mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}}$ such that $d(\sigma_0)_I = c_0$, where $c_0 : \mathfrak{sl}(n, \mathbb{C})|_{\mathbb{R}} \to \mathfrak{sl}(n, \mathbb{C})|_{\mathbb{R}}$ is the map $c_0(X) = \overline{X}$, and $\sigma_0$ is also given by

$$\sigma_0(X) = \overline{X}.$$

The real Lie group $G_0$ is the set of fixed points of $\mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}}$ under the automorphism $\sigma_0$, given by

$$G_0 = \{X \in \mathbf{SL}(n, \mathbb{C})|_{\mathbb{R}} \mid \overline{X} = X\} = \mathbf{SL}(n, \mathbb{R}).$$

Note that the Lie algebra of $G_0 = \mathbf{SL}(n, \mathbb{R})$ is $\mathfrak{g}_0 = \mathfrak{sl}(n, \mathbb{R})$. The simply-connected real Lie group $G_u$ whose Lie algebra is $\mathfrak{su}(n)$ is $G_u = \mathbf{SU}(n)$. We define $K_0$ by

$$K_0 = G_0 \cap G_u = \mathbf{SL}(n, \mathbb{R}) \cap \mathbf{SU}(n) = \mathbf{SO}(n).$$
Note that the Lie algebra of the compact Lie group \( K_0 = \SO(n) \) is \( \mathfrak{k}_0 = \mathfrak{so}(n) \).

The real Lie group \( G_0 = \SL(n, \mathbb{R}) \) is semi-simple and the real Lie group \( K_0 = \SO(n) \) is semi-simple for \( n \geq 3 \). These groups are connected but not-simply connected. For \( n = 2 \), the universal cover of \( \SL(2, \mathbb{R}) \) is \( \mathbb{R}^3 \) and the universal cover of \( \SO(2) \) is \( \mathbb{R} \). For \( n \geq 3 \), the universal cover \( \tilde{G}_0 = \tilde{\SL}(n, \mathbb{R}) \) of \( G_0 = \SL(n, \mathbb{R}) \) is not a matrix group and the universal cover of \( K_0 = \SO(n) \) is \( \tilde{K}_0 = \Spin(n) \). The real semi-simple (connected) Lie group \( G_0 = \SL(n, \mathbb{R}) \) is called a real form of the complex semi-simple (simply-connected) Lie group \( G = \SL(n, \mathbb{C}) \).

We will show later that the pair \( (G_0, K_0) = (\SL(n, \mathbb{R}), \SO(n)) \) is a Gelfand pair.

Example 14.3. Again, consider the real Lie algebra \( \mathfrak{g}_u = \mathfrak{su}(n) \) of \( n \times n \) complex skew-hermitian matrices with zero trace. We showed in Example 14.2 that the complexification \( \mathfrak{su}(n)_c \) of \( \mathfrak{su}(n) \) is the complex Lie algebra \( \mathfrak{g} = \mathfrak{sl}(n, \mathbb{C}) \) of \( n \times n \) complex matrices with zero trace. This time, let \( c_0 : \mathfrak{sl}(n, \mathbb{C}) \to \mathfrak{sl}(n, \mathbb{C}) \) be the conjugation given by

\[
c_0(X) = -I_{p,n-p}X^*I_{p,n-p},
\]

where

\[
I_{p,n-p} = \begin{pmatrix} I_p & 0_{p,n-p} \\ 0_{n-p,p} & -I_{n-p} \end{pmatrix},
\]

with \( 1 \leq p \leq n-1 \). Obviously \( I_{p,n-p}^* = I_{p,n-p} \) and \( I_{p,n-p}^2 = I_n \), and \( c_0 \) commutes with \( c_u \) (given by \( c_u(X) = -X^* \)). Since matrices in \( \mathfrak{su}(n) \) satisfy the property \( X^* = -X \), the restriction of \( c_0 \) to \( \mathfrak{su}(n) \) is given by \( c_0(X) = I_{p,n-p}X^*I_{p,n-p} \). If we write

\[
X = \begin{pmatrix} U & B \\ A & V \end{pmatrix},
\]

where \( U \in M_p(\mathbb{C}), V \in M_{n-p}(\mathbb{C}), A \in M_{n-p,p}(\mathbb{C}), B \in M_{p,n-p}(\mathbb{C}) \), then

\[
I_{p,n-p}X^*I_{p,n-p} = \begin{pmatrix} I_p & 0_{p,n-p} \\ 0_{n-p,p} & -I_{n-p} \end{pmatrix} \begin{pmatrix} U & B \\ A & V \end{pmatrix} \begin{pmatrix} I_p & 0_{p,n-p} \\ 0_{n-p,p} & -I_{n-p} \end{pmatrix} = \begin{pmatrix} U & -B \\ -A & V \end{pmatrix}.
\]

Therefore

\[
\mathfrak{k}_0 = \{ X \in \mathfrak{su}(n) \mid I_{p,n-p}X^*I_{p,n-p} = X \}
\]

\[
= \left\{ \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \mid U^* = -U, \ V^* = -V, \ \text{tr}(U) + \text{tr}(V) = 0 \right\}
\]

\[
= \left\{ \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \mid U \in \mathfrak{u}(p), \ V \in \mathfrak{u}(n-p), \ \text{tr}(U) + \text{tr}(V) = 0 \right\},
\]

and

\[
i\mathfrak{p}_0 = \{ X \in \mathfrak{su}(n) \mid -I_{p,n-p}X^*I_{p,n-p} = X \}
\]

\[
= \left\{ \begin{pmatrix} 0 & -A^* \\ A & 0 \end{pmatrix} \mid A \in M_{n-p,p}(\mathbb{C}) \right\}.
\]
so

\[ p_0 = \left\{ \begin{pmatrix} 0 & -iA^* \\ iA & 0 \end{pmatrix} \mid A \in M_{n-p,p}(\mathbb{C}) \right\} = \left\{ \begin{pmatrix} 0 & (iA)^* \\ iA & 0 \end{pmatrix} \mid A \in M_{n-p,p}(\mathbb{C}) \right\} = \left\{ \begin{pmatrix} 0 & A^* \\ A & 0 \end{pmatrix} \mid A \in M_{n-p,p}(\mathbb{C}) \right\}. \]

Consequently the Lie algebra \( g_0 \) corresponding to the conjugation \( c_0 \) is given by

\[ g_0 = \mathfrak{k}_0 \oplus p_0 = \left\{ \begin{pmatrix} U & A^* \\ A & V \end{pmatrix} \mid U \in \mathfrak{u}(p), V \in \mathfrak{u}(n-p), A \in M_{n-p,p}(\mathbb{C}), \text{tr}(U) + \text{tr}(V) = 0 \right\}. \]

Thus \( g_0 = \mathfrak{su}(p, n-p) \), the Lie algebra of the real Lie group \( \text{SU}(p, n-p) \) of \( n \times n \) complex matrices \( X \) such that

\[ X^*I_{p,n-p}X = I_{p,n-p}, \quad \det(X) = 1. \]

We can also write every matrix \( X \in \mathfrak{t}_0 \) as

\[ X = \begin{pmatrix} U_1 & 0 \\ 0 & 0 \end{pmatrix} + i \begin{pmatrix} 2I_p & 0 \\ 0 & -\alpha p_{n-p} \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ 0 & V_1 \end{pmatrix}, \quad U_1 \in \mathfrak{su}(p), V_1 \in \mathfrak{su}(n-p), \alpha \in \mathbb{R}, \]

so

\[ \mathfrak{t}_0 \cong \mathfrak{su}(p) \oplus i\mathbb{R} \oplus \mathfrak{su}(n-p), \]

where all the summands are ideals.

The derivative \( d(\sigma_0)_I \) of the automorphism \( \sigma_0: \text{SL}(n, \mathbb{C})|_\mathbb{R} \rightarrow \text{SL}(n, \mathbb{C})|_\mathbb{R} \) given by

\[ \sigma_0(X) = I_{p,n-p}(X^*)^{-1}I_{p,n-p} \]

is the map \( c_0: \text{sl}(n, \mathbb{C})|_\mathbb{R} \rightarrow \text{sl}(n, \mathbb{C})|_\mathbb{C} \) also given by \( c_0(X) = -I_{p,n-p}X^*I_{p,n-p} \). The real Lie group \( G_0 \) is the set of fixed points of \( \text{SL}(n, \mathbb{C})|_\mathbb{R} \) under the automorphism \( \sigma_0 \), given by

\[ G_0 = \{ X \in \text{SL}(n, \mathbb{C})|_\mathbb{R} \mid X = I_{p,n-p}(X^*)^{-1}I_{p,n-p} \} = \{ X \in \text{SL}(n, \mathbb{C})|_\mathbb{R} \mid X^*I_{p,n-p}X = I_{p,n-p} \} = \text{SU}(p, n-p). \]

If we write

\[ X = \begin{pmatrix} U & B \\ A & V \end{pmatrix} \]

for any \( X \in \text{SU}(p, n-p) \), it is easy to check that if \( X \in \text{SU}(p, n-p) \cap \text{SU}(n) \), then

\[ U^*U - A^*A = I_p \]
\[ V^*V - B^*B = I_{n-p} \]
\[ U^*U + A^*A = I_p \]
\[ V^*V + B^*B = I_{n-p} \]
\[ U^*B - A^*V = 0 \]
\[ U^*B + A^*V = 0, \]
14.2. REAL FORMS OF A COMPLEX SEMI-SIMPLE LIE GROUP

which implies that $A = 0$ and $B = 0$. Therefore,

$$K_0 = G_0 \cap G_u = SU(p, n - p) \cap SU(n)$$

$$= \left\{ \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \bigg| U \in U(p), \ V \in U(n-p), \ \det(U) \det(V) = 1 \right\}. $$

This group is usually denoted $S(U(p) \times U(n-p))$. For any $X \in S(U(p) \times U(n-p))$ we can write

$$X = \begin{pmatrix} U_1 & 0 \\ 0 & I_{n-p} \end{pmatrix} \begin{pmatrix} \lambda & 0 & 0 & 0 \\ 0 & I_{p-1} & 0 & 0 \\ 0 & 0 & \lambda^{-1} & 0 \\ 0 & 0 & 0 & I_{n-p-1} \end{pmatrix} \begin{pmatrix} I_p & 0 \\ 0 & V_1 \end{pmatrix},$$

with $U_1 \in SU(p), V_1 \in SU(n-p)$ and $\lambda \in \mathbb{C}$ with $|\lambda| = 1$. Therefore

$$K_0 = S(U(p) \times U(n-p)) \cong SU(p) \times U(1) \times SU(n-p).$$

The Lie algebra of the real compact Lie group $K_0 = SU(p) \times U(1) \times SU(n-p)$ is $\mathfrak{k}_0 = \mathfrak{su}(p) \oplus i\mathbb{R} \oplus \mathfrak{su}(n-p)$. The real Lie groups $G_0 = SU(p, n-p)$ and $K_0 = SU(p) \times U(1) \times SU(n-p)$ are semi-simple and connected but not simply-connected. The universal cover of $K_0$ is $\tilde{K}_0 = SU(p) \times \mathbb{R} \times SU(n-p)$, and the universal cover of $G_0 = SU(p, n-p)$ is $G_0 = \text{Spin}(p, n-p)$. The real semi-simple (connected) Lie group $G_0 = SU(p, n-q)$ is called a real form of the complex semi-simple (simply-connected) Lie group $G = SL(n, \mathbb{C})$. We will show later that the pair $(G_0, K_0) = (SU(p, n-p), SU(p) \times U(1) \times SU(n-p))$ is a Gelfand pair.

**Example 14.4.** Again, consider the real Lie algebra $\mathfrak{g}_u = \mathfrak{su}(n)$ of $n \times n$ complex skew-hermitian matrices with zero trace and its complexification $\mathfrak{g} = \mathfrak{sl}(n, \mathbb{C})$. This time, assume $n = 2m$, and let $c_0 : \mathfrak{sl}(n, \mathbb{C}) \to \mathfrak{sl}(n, \mathbb{C})$ be the conjugation given by

$$c_0(X) = -J_m \overline{X} J_m,$$

with

$$J_m = \begin{pmatrix} 0_m & I_m \\ -I_m & 0_m \end{pmatrix}.$$ 

Since $J_m^2 = -I_{2m}$, $\overline{J_m} = J_m$, and $J_m^\top = -J_m$, the conjugation $c_0$ commutes with $c_u$. The automorphism $\sigma_0 : SL(2m, \mathbb{C})_{\mathbb{R}} \to SL(2m, \mathbb{C})_{\mathbb{R}}$ such that $(d\sigma_0)_I = c_0$ is also $\sigma_0(X) = -J_m \overline{X} J_m.$

If we write

$$X = \begin{pmatrix} U & V \\ A & B \end{pmatrix} \in SL(2m, \mathbb{C}),$$

where $U, V, A, B \in M_m(\mathbb{C})$, then we have

$$-J_m \overline{X} J_m = \begin{pmatrix} 0 & -I_m \\ I_m & 0 \end{pmatrix} \begin{pmatrix} \overline{U} & \overline{V} \\ \overline{A} & \overline{B} \end{pmatrix} \begin{pmatrix} 0 & I_m \\ -I_m & 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 & -I_m \\ I_m & 0 \end{pmatrix} \begin{pmatrix} -\overline{V} & \overline{U} \\ -\overline{B} & \overline{A} \end{pmatrix} = \begin{pmatrix} \overline{B} & -\overline{A} \\ -\overline{V} & \overline{U} \end{pmatrix},$$
so \( X = -J_m \bar{X} J_m \) iff

\[
U = \bar{B}, \quad V = -\bar{A}, \quad A = -\bar{V}, \quad B = U,
\]

which simplifies to

\[
B = U, \quad A = -\bar{V}.
\]

Therefore, \( X \) is of the form

\[
X = \begin{pmatrix} U & V \\ -\bar{V} & \bar{U} \end{pmatrix},
\]

and the real Lie group \( G_0 = \{ X \in \text{SL}(2m, \mathbb{C})|_\mathbb{R} \mid X = -J_m \bar{X} J_m \} \) is given by

\[
G_0 = \left\{ \left( \begin{array}{cc} U & V \\ -\bar{V} & \bar{U} \end{array} \right) \in \text{SL}(2m, \mathbb{C}) \right\}.
\]

**Definition 14.9.** The real Lie group \( \text{SU}^*(2m) \) is defined by

\[
\text{SU}^*(2m) = \left\{ \left( \begin{array}{cc} U & V \\ -\bar{V} & \bar{U} \end{array} \right) \in \text{SL}(2m, \mathbb{C}) \mid U, V \in M_m(\mathbb{C}) \right\}.
\]

The notation \( \text{SU}^*(2m) \) for this real Lie group is introduced in Helgason [42], Chapter X, Section 2. We will show later that the (real) Lie group \( \text{SU}^*(2m) \) can also be defined as the quaternionic (real) Lie group \( \text{SL}(m, \mathbb{H}) \).

Since \( X \) is invertible, for any vector \( z = (x, y) \in \mathbb{C}^{2m} \), if \( Xz = 0 \) then \( z = 0 \), so in particular, for \( y = 0 \), since

\[
Xz = \begin{pmatrix} U & V \\ -\bar{V} & \bar{U} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} Ux +Vy \\ -\bar{V}x + \bar{U}y \end{pmatrix},
\]

we have \( Xz = 0 \) iff \( Ux + Vy = 0 \) and \( \bar{V}x + \bar{U}y = 0 \). So with \( y = 0 \), if \( Ux = 0 \) then \( x = 0 \), and with \( x = 0 \), if \( Vy = 0 \) then \( y = 0 \) which implies that \( U \) and \( V \) are invertible. Therefore, the group \( G_0 \), the set of fixed points of \( \sigma_0 \), is also given by

\[
G_0 = \left\{ X = \begin{pmatrix} U & V \\ -\bar{V} & \bar{U} \end{pmatrix} \mid U, V \in \text{GL}(m, \mathbb{C})|_\mathbb{R}, \det(X) = 1 \right\}.
\]

Since the conjugation \( c_0 \) on \( \mathfrak{sl}(2m, \mathbb{C}) \) has the same expression as the conjugation \( \sigma_0 \) on \( \mathfrak{SL}(2m, \mathbb{C})|_\mathbb{R} \), the same computation as above shows that

\[
\mathfrak{su}^*(2m) = \left\{ \left( \begin{array}{cc} U & V \\ -\bar{V} & \bar{U} \end{array} \right) \mid U, V \in M_m(\mathbb{C}), \text{tr}(U) + \text{tr}(\bar{U}) = 0 \right\}.
\]

We also have

\[
\mathfrak{k}_0 = \{ X \in \mathfrak{su}(2m, \mathbb{C}) \mid X = -J_m \bar{X} J_m \}
\]

\[
= \left\{ \left( \begin{array}{cc} U & V \\ -\bar{V} & \bar{U} \end{array} \right) \mid U \in \mathfrak{u}(m, \mathbb{C}), V \in M_m(\mathbb{C}), V^\top = V \right\}
\]

\[
= \left\{ \left( \begin{array}{cc} U & V \\ -\bar{V} & -U^\top \end{array} \right) \mid U \in \mathfrak{u}(m, \mathbb{C}), V \in M_m(\mathbb{C}), V^\top = V \right\},
\]
and
\[
P_0 = \{ X \in \mathfrak{su}(2m, \mathbb{C}) \mid X = J_m X J_m \}
= \left\{ \begin{pmatrix} U & V \\ V & -U \end{pmatrix} \middle| U \in \mathfrak{u}(m, \mathbb{C}), \ V \in M_m(\mathbb{C}), \ V^\top = -V, \ \text{tr}(U) = 0 \right\}.
\]

Consequently, we have
\[
P_0 = \left\{ \begin{pmatrix} iU & iV \\ iV & -iU \end{pmatrix} \middle| U \in \mathfrak{u}(m, \mathbb{C}), \ V \in M_m(\mathbb{C}), \ V^\top = -V, \ \text{tr}(U) = 0 \right\}
= \left\{ \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \middle| A \in iu(m, \mathbb{C}), \ B \in M_m(\mathbb{C}), \ B^\top = -B, \ \text{tr}(A) = 0 \right\}
= \left\{ \begin{pmatrix} A & B \\ -B & A \end{pmatrix} \middle| A, B \in M_m(\mathbb{C}), \ A^* = A, \ B^\top = -B, \ \text{tr}(A) = 0 \right\}
= \left\{ \begin{pmatrix} A & B \\ -B & A^\top \end{pmatrix} \middle| A, B \in M_m(\mathbb{C}), \ A^* = A, \ B^\top = -B, \ \text{tr}(A) = 0 \right\}.
\]

We immediately check that
\[
\mathfrak{su}^*(2m) = \mathfrak{k}_0 \oplus P_0.
\]

The (real) Lie group $\mathbf{SU}^*(2m)$ is also the (real) Lie group $\mathbf{SL}(m, \mathbb{H})$, where $\mathbb{H}$ is the skew-field of quaternions. To see this, it is convenient to view the real algebra $\mathbb{H}$ as
\[
\mathbb{H} = \mathbb{C} \oplus \mathbb{C}j.
\]
using the fact that every quaternion can be written uniquely as
\[
q = a + xi + yj + zk = u + xi + (y + zi)j, \quad a, x, y, z \in \mathbb{R},
\]
since $ij = k$. Since the conjugate $\overline{q}$ of $q$ is $\overline{q} = a - xi - yj - zk = a - xi - (y + zi)j$, if we write $q = \alpha + \beta j$ with $\alpha, \beta \in \mathbb{C}$, then
\[
\overline{q} = \overline{\alpha} - \beta j.
\]
Also, since $ij = -ji$, for any $\alpha = a + bi \in \mathbb{C}$, we have
\[
j\alpha = j(a + bi) = ja + bji = aj - bij = (a - bi)j = \overline{\alpha}j.
\]
In summary, we have
\[
\overline{\alpha + j\beta} = \overline{\alpha} - \beta j, \quad j\alpha = \overline{\alpha}j. \tag{conj}
\]
Every $m \times m$ matrix $X = (\alpha_{kl} + \beta_{kl}j) \in M_m(\mathbb{H})$ can be written uniquely as
\[
X = U + V j.
\]
with \( U = (\alpha_{k\ell}) \in M_m(\mathbb{C}) \) and \( V = (\beta_{k\ell}) \in M_m(\mathbb{C}) \). Then in view of the equations (conj), if \( X = (\alpha_{k\ell} + \beta_{k\ell}j) \in M_m(\mathbb{H}) \) and if we define

\[
X = (\alpha_{k\ell} + \beta_{k\ell}j), \quad \text{and} \quad X^* = X^T,
\]

then we have

\[
X^* = (U + Vj)^* = U^* - V^Tj.
\]

We also have

\[
jV = Vj.
\]

In summary, we have

\[
(U + Vj)^* = U^* - V^Tj, \quad jV = Vj.
\]  

\[(\text{conj'})\]

The (real) Lie group \( \text{GL}(m, \mathbb{H}) \) is the group of matrices \( X \in M_m(\mathbb{H}) \) such that there is some \( Y \in M_m(\mathbb{H}) \) with

\[
XY = YX = I_m.
\]

Observe that since \( j^2 = -1 \), we have

\[
(U_1 + V_1j)(U_2 + V_2j) = U_1V_1 + U_1V_2j + V_1jU_2 + V_1jV_2j
\]

\[
= U_1V_1 + U_1V_2j + V_1U_2j + V_1V_2j^2
\]

\[
= U_1V_1 - V_1U_2 + (U_1V_2 + V_1U_2)j.
\]

Since

\[
\begin{pmatrix} U_1 & V_1 \\ -\overline{V_2} & \overline{U_2} \end{pmatrix} \begin{pmatrix} U_2 & V_2 \\ -\overline{V_2} & \overline{U_2} \end{pmatrix} = \begin{pmatrix} U_1U_2 - V_1\overline{V_2} & U_1V_2 + V_1\overline{U_2} \\ -\overline{U_1V_2} - \overline{V_1U_2} & \overline{U_1U_2} - \overline{V_1V_2} \end{pmatrix},
\]

the map \( \varphi: M_m(\mathbb{H}) \to M_{2m}(\mathbb{C}) \) given by

\[
\varphi(U + Vj) = \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix}
\]

is an injective \( \mathbb{R} \)-algebra homomorphism. Observe that this is the matrix in the definition of \( G_0 \).

What we did previously shows that this homomorphism restricts to an injective homomorphism \( \varphi: \text{GL}(m, \mathbb{H}) \to \text{GL}(2m, \mathbb{C}) \), which allows us to view the group \( \text{GL}(m, \mathbb{H}) \) as

\[
\varphi(\text{GL}(m, \mathbb{H})) = \left\{ X = \begin{pmatrix} U & V \\ -\overline{V} & \overline{U} \end{pmatrix} \mid X \in \text{GL}(2m, \mathbb{C}), U, V \in \text{GL}(m, \mathbb{C}) \right\}.
\]

**Definition 14.10.** We define the (real) Lie the group \( \text{SL}(m, \mathbb{H}) \) as

\[
\text{SL}(m, \mathbb{H}) = \{ X \in \varphi(\text{GL}(m, \mathbb{H})) \mid \det(X) = 1 \}.
\]
Therefore, we conclude that
\[ G_0 = SU^*(2m) = SL(m, \mathbb{H}). \]
This is a real, semi-simple, simply-connected Lie group.

The Lie algebras \( \mathfrak{gl}(n, \mathbb{H}) \) of \( GL(m, \mathbb{H}) \) and \( \mathfrak{sl}(n, \mathbb{H}) \) of \( SL(m, \mathbb{H}) \) are given by
\[
\mathfrak{gl}(n, \mathbb{H}) = \left\{ \begin{pmatrix} U & V \\ -V & U \end{pmatrix} \right\}, \quad U, V \in M_m(\mathbb{C})
\]
and
\[
\mathfrak{sl}(n, \mathbb{H}) = \left\{ \begin{pmatrix} U & V \\ -V & U \end{pmatrix} \right\}, \quad U, V \in M_m(\mathbb{C}), \quad \text{tr}(U) + \text{tr}(U) = 0.
\]

We can also identify
\[ K_0 = G_0 \cap SU(2m, \mathbb{C}) = SU^*(2m) \cap SU(2m, \mathbb{C}) = SL(m, \mathbb{H}) \cap SU(2m, \mathbb{C}). \]

There are two descriptions for the Lie group \( K_0 \); one as a quaternionic group; another in terms of the complex symplectic group \( \text{Sp}(m, \mathbb{C}) \).

For the quaternionic description, define \( U(m, \mathbb{H}) \) as the (real) Lie group
\[ U(m, \mathbb{H}) = \{ X = U + Vj \in GL(m, \mathbb{H}) \mid X^*X = XX^* = I_m \}. \]

It is easy to see that only the first equation is needed. Using (conj’), this is equivalent to
\[
(U + Vj)^*(U + Vj) = (U^* - V^Tj)(U + Vj)
\]
\[ = U^*U + U^*Vj - V^TU - V^TVj^2 \]
\[ = U^*U + U^*V - V^TU - V^TVj^2 \]
\[ = U^*U + V^TV + (U^*V - V^TU)j = I_m, \]
so we get
\[ U^*U + V^TV = I_m, \quad U^*V - V^TU = 0. \]

On the other hand, for
\[ X = \begin{pmatrix} U & V \\ -V & U \end{pmatrix}, \]
we have \( X^*X = I_{2m} \) iff
\[
\begin{pmatrix} U^* & -V^T \\ V^* & U^T \end{pmatrix} \begin{pmatrix} U & V \\ -V & U \end{pmatrix} = \begin{pmatrix} U^*U + V^TV & U^*V - V^TU \\ V^*U - U^TV & V^*V + U^TU \end{pmatrix} = \begin{pmatrix} I_m & 0 \\ 0 & I_m \end{pmatrix},
\]
and these are equivalent to the same conditions as above,
\[ U^*U + V^TV = I_m, \quad U^*V - V^TU = 0. \]

We conclude that
\[ U(m, \mathbb{H}) = GL(m, \mathbb{H}) \cap U(2m, \mathbb{C}). \]
Definition 14.11. The (real) Lie group $SU(m, \mathbb{H})$ is defined as

$$SU(m, \mathbb{H}) = U(m, \mathbb{H}) \cap SU(2m, \mathbb{C}).$$

Observe that $SU(m, \mathbb{H}) = SL(m, \mathbb{H}) \cap SU(2m, \mathbb{C}).$

In summary, we see that

$$K_0 = SU^*(2m) \cap SU(2m, \mathbb{C}) = SL(m, \mathbb{H}) \cap SU(2m, \mathbb{C}) = U(m, \mathbb{H}) \cap SU(2m, \mathbb{C})$$

$$= SU(m, \mathbb{H}).$$

This is a real, semi-simple, simply-connected Lie group.

The Lie algebra $u(n, \mathbb{H})$ of $U(n, \mathbb{H})$ consists of the space of matrices

$$u(n, \mathbb{H}) = \{X \in M_n(\mathbb{H}) \mid X^* = -X\}.$$

If we write $X = U + VJ$ (with $U, V \in M_n(\mathbb{C})$), then $X^* = -X$ is equivalent to

$$(U + VJ)^* = U^* - V^Tj = -U - Vj,$$

that is,

$$U^* = -U, \quad V^T = V.$$

Thus we can also write

$$u(n, \mathbb{H}) = \left\{ \begin{pmatrix} U & V \\ -V & U \end{pmatrix} \mid U, V \in M_m(\mathbb{C}), U^* = -U, V^T = V \right\}.$$

Similarly,

$$su(n, \mathbb{H}) = \left\{ \begin{pmatrix} U & V \\ -V & U \end{pmatrix} \mid U, V \in M_m(\mathbb{C}), U^* = -U, V^T = V, \text{tr}(U) + \text{tr}(U) = 0 \right\}.$$

The real Lie group $SU(m, \mathbb{H})$ has another description in terms of the complex symplectic group $Sp(m, \mathbb{C})$. Since

$$SU(m, \mathbb{H}) = \{X \in M_{2m}(\mathbb{C}) \mid X^*X = I_{2m}, \ X = -J_mXJ_m, \ \det(X) = 1\},$$

first from $X^*X = I_{2m}$ we get

$$I_{2m} = X^*X = -X^*J_mXJ_m,$$

since $J_m^2 = -I_{2m}$ we get

$$X^*J_mX = J_m,$$

and since $J_m = J_m$, by conjugating both sides we get

$$X^\top J_mX = J_m.$$
14.2. REAL FORMS OF A COMPLEX SEMI-SIMPLE LIE GROUP

**Definition 14.12.** The complex symplectic group $\text{Sp}(m, \mathbb{C})$ is defined as

$$\text{Sp}(m, \mathbb{C}) = \{X \in M_{2m}(\mathbb{C}) \mid X^T J_m X = J_m\}.$$ 

It can be shown that the Lie algebra $\mathfrak{sp}(m, \mathbb{C})$ of the Lie group $\text{Sp}(m, \mathbb{C})$ is given by

$$\mathfrak{sp}(m, \mathbb{C}) = \left\{ \begin{pmatrix} U & V_1 \\ V_2 & -U^T \end{pmatrix} \mid U, V_1, V_2 \in M_m(\mathbb{C}), V_1^T = V_1, V_2^T = V_2 \right\};$$

see Helgason [42] (Chapter X, Section 2, page 447).

The preceding argument showed that

$$\text{SU}(m, \mathbb{H}) = \text{SL}(m, \mathbb{H}) \cap \text{SU}(2m, \mathbb{C}) \subseteq \text{Sp}(m, \mathbb{C}) \cap \text{U}(2m, \mathbb{C}).$$

It can also be shown that the group $\text{Sp}(m, \mathbb{C})$ is connected; see Helgason [42] (Chapter X, Section 2, Lemma 2.4) or Knapp [49] (Chapter I, Proposition 1.145). Since $J_m$ is invertible ($J_m^{-1} = -J_m$), the equation $X^T J_m X = J_m$ shows that $\det(X) = \pm 1$. Since $\det(I) = 1$ and $\text{Sp}(m, \mathbb{C})$ is connected, we deduce that for every $X \in \text{Sp}(m, \mathbb{C})$, we have $\det(X) = 1$. (There are also purely algebraic proofs of this property using the fact that the symplectic transvections generate $\text{Sp}(m, \mathbb{C})$, and that they have determinant 1.) Thus if $X \in \text{Sp}(m, \mathbb{C}) \cap \text{U}(2m, \mathbb{C})$, then in fact $X \in \text{Sp}(m, \mathbb{C}) \cap \text{SU}(2m, \mathbb{C})$, and from $X^T J_m X = J_m$ and $X^* X = I_{2m}$, by reversing the above argument, we deduce that $X = -J_m X J_m$, and so

$$\text{Sp}(m, \mathbb{C}) \cap \text{U}(2m, \mathbb{C}) \subseteq \text{SL}(m, \mathbb{H}) \cap \text{SU}(2m, \mathbb{C}) = \text{SU}(m, \mathbb{H}).$$

We just showed that

$$\text{SU}(m, \mathbb{H}) = \text{U}(m, \mathbb{H}) \cap \text{SU}(2m, \mathbb{C}) = \text{SL}(m, \mathbb{H}) \cap \text{SU}(2m, \mathbb{C}) = \text{Sp}(m, \mathbb{C}) \cap \text{U}(2m, \mathbb{C}).$$

**Definition 14.13.** We define the real Lie group $\text{Sp}(m)$ as

$$\text{Sp}(m) = \text{Sp}(m, \mathbb{C}) \cap \text{U}(2m, \mathbb{C});$$

see Helgason [42] (Chapter X, Section 2).

We have

$$\text{Sp}(m) = \text{SU}(m, \mathbb{H}),$$

and the Lie algebra $\mathfrak{sp}(m)$ of $\text{Sp}(m)$ is equal to $\mathfrak{sp}(m, \mathbb{C}) \cap u(2m, \mathbb{C})$, so it is given by

$$\mathfrak{sp}(m) = \left\{ \begin{pmatrix} U & V \\ -V^* & -U^T \end{pmatrix} \mid U \in u(m, \mathbb{C}), V \in M_m(\mathbb{C}), V^T = V \right\};$$

$$= \left\{ \begin{pmatrix} U & V \\ -V & -U^T \end{pmatrix} \mid U \in u(m, \mathbb{C}), V \in M_m(\mathbb{C}), V^T = V \right\}.$$ 

The real semi-simple simply-connected Lie group $G_0 = \text{SL}(m, \mathbb{H}) = \text{SU}^*(2m)$ is another real form of the complex semi-simple (simply-connected) Lie group $G = \text{SL}(2m, \mathbb{C})$. We will show later that the pair $(G_0, K_0) = (\text{SL}(m, \mathbb{H}), \text{SU}(m, \mathbb{H})) = (\text{SU}^*(2m), \text{Sp}(m))$ is a Gelfand pair.
It can be shown that up to isomorphism, \( \text{SL}(n, \mathbb{R}) \), \( \text{SU}(p, n - p) \), and \( \text{SL}(m, \mathbb{H}) = \text{SU}^*(2m) \) (when \( n = 2m \)), are the only real forms of \( \text{SL}(n, \mathbb{C}) \).

A general method to find real forms of a complex, semi-simple, simply-connected Lie group \( G \) with complex semi-simple Lie algebra \( \mathfrak{g} \) goes as follows. Suppose we have real, compact, semi-simple, simply-connected Lie group \( G_u \) with Lie algebra \( \mathfrak{g}_u \), such that \( \mathfrak{g} = \mathfrak{g}_u \oplus i\mathfrak{g}_u \) is the complexification of the real Lie algebra \( \mathfrak{g}_u \). In our previous examples, \( \mathfrak{g} = \mathfrak{sl}(n, \mathbb{C}), \mathfrak{g}_u = \mathfrak{su}(n, \mathbb{C}), \) and \( G_u = \text{SU}(n, \mathbb{C}) \). By a famous theorem of Hermann Weyl, such a real semi-simple Lie algebra \( \mathfrak{g}_u \) and such a compact semi-simple Lie group \( G_u \) always exist; see see Dieudonné [19] (Chapter XXI, no. 21.20.7).

Also assume that we have a conjugation \( c_0 : \mathfrak{g} \to \mathfrak{g} \) that commutes with the conjugation \( c_u \) associated with \( \mathfrak{g}_u \). Using the correspondence between (real) simply-connected Lie groups and (real) Lie algebras, there is a unique involutive automorphism \( \sigma_0 : G|_\mathbb{R} \to G|_\mathbb{R} \) such that \( d(\sigma_0)_{|_{\mathbb{R}}} = c_0 \), where \( c_0 : \mathfrak{g}|_\mathbb{R} \to \mathfrak{g}|_\mathbb{R} \). If \( P \) is the closed (complex) submanifold of \( G \) which is the image of \( i\mathfrak{g}_u \) by the exponential map \( \exp_G : \mathfrak{g} \to G \) \( (ix \mapsto \exp(ix), x \in \mathfrak{g}_u) \), then the following result can be shown.

**Proposition 14.5.** The exponential map \( \exp_G \) is a diffeomorphism of \( i\mathfrak{g}_u \) onto \( P \), and the map \( (x, y) \mapsto xy \) is a diffeomorphism of \( G_u \times P \) onto \( G \) (as complex manifolds).

The automorphism \( \sigma_0 \) of \( G|_\mathbb{R} \) leaves \( G_u \) and \( P|_\mathbb{R} \) invariant, since \( c_0 \) leaves \( \mathfrak{g}_u \) and \( i\mathfrak{g}_u \) invariant.

If \( G = \text{SL}(n, \mathbb{C}), \mathfrak{g} = \mathfrak{sl}(n, \mathbb{C}), \mathfrak{g}_u = \mathfrak{su}(n, \mathbb{C}), \) and \( \mathfrak{g}_u = \mathfrak{su}(n, \mathbb{C}), \) then

\[
    i\mathfrak{g}_u = i\mathfrak{su}(n, \mathbb{C}) = \{ X \in M_n(\mathbb{C}) \mid X^* = X \},
\]

and \( P = \exp_G(i\mathfrak{g}_u) \) is the manifold of hermitian positive definite matrices with determinant \(+1\). The factorization \( \text{SL}(2, \mathbb{C}) = \text{SU}(n, \mathbb{C})P \) is the polar form for matrices in \( \text{SL}(2, \mathbb{C}) \).

Let \( G_0 \) be the real Lie subgroup of \( G|_\mathbb{R} \) given by

\[
    G_0 = \{ s \in G|_\mathbb{R} \mid \sigma_0(s) = s \}.
\]

The group \( G_0 \) is a real semi-simple Lie group because its Lie algebra is \( \mathfrak{g}_0 \). Since \( \sigma_0 \) leaves \( G_u \) invariant, the group

\[
    K_0 = G_0 \cap G_u = \{ s \in G_u \mid \sigma_0(s) = s \}
\]

is a real compact subgroup of \( G_u \) consisting of the fixed points of \( G_u \) under \( \sigma_0 \). The group \( G_0 \) also contains the image \( P_0 \subseteq i\mathfrak{g}_u \) under the exponential map \( v \mapsto \exp_{G_0}(v) \), and since \( \exp_{G_0}(c_0(v)) = \sigma_0(\exp_{G_0}(v)) \), the manifold \( P_0 \) is the set of fixed points of \( P|_\mathbb{R} \) under \( \sigma_0 \).

**Proposition 14.6.** The map \( v \mapsto \exp_{G_0}(v) \) is a diffeomorphism of \( p_0 \) onto the closed manifold \( P_0 \), and the map \( (y, z) \mapsto yz \) from \( K_0 \times P_0 \) to \( G_0 \) is a diffeomorphism.
If $G = \text{SL}(n, \mathbb{C})$, $g = \mathfrak{sl}(n, \mathbb{C})$, $G_u = \text{SU}(n)$, and $g_u = \mathfrak{su}(n)$, and $c_0(X) = X$, as Example 14.2, we have

$$G_0 = \text{SL}(n, \mathbb{R}), \quad K_0 = \text{SO}(n), \quad \mathfrak{k}_0 = \mathfrak{so}(n), \quad p_0 = \mathfrak{s}(n) = \{ X \in M_n(\mathbb{R}) \mid X^\top = X \},$$

and $P_0 = \exp_{G_0}(p_0)$ is the manifold of symmetric positive definite matrices with determinant $+1$. The factorization $\text{SL}(2, \mathbb{R}) = \text{SO}(n) P_0$ is the polar form for matrices in $\text{SL}(2, \mathbb{R})$.

Thus $G_0$ is a real noncompact semi-simple connected Lie group, diffeomorphic to the product of a compact semi-simple Lie group $K_0$ and some $\mathbb{R}^N$. It can be shown that the group $K_0$ is a maximal compact subgroup of $G_0$, is connected, but in general not semi-simple nor simply-connected.

**Definition 14.14.** The real, semi-simple, simply-connected Lie group $G_u$ is called the **compact form** of the complex semi-simple Lie group $G$, and the real, semi-simple, connected Lie group $G_0$ is a **real form** of $G$.

The pair $(G_0, K_0)$ is a Gelfand pair, but not with respect to the involution $\sigma_0$, because $\sigma_0$ is the identity on $P_0$. However, there is also a unique involutive automorphism $\sigma_u : G|_{\mathbb{R}} \rightarrow G|_{\mathbb{R}}$ such that $d(\sigma_u)_e = c_u$, and this involution makes $(G_0, K_0)$ a Gelfand pair.

Since $\mathfrak{k}_0 \subseteq \mathfrak{g}_u$ and since $\mathfrak{g}_u$ is the set of fixed points of $c_u$, the Lie group $K_0$ is a set of fixed points of $\sigma_u$. Similarly, since $p_0 \subseteq i\mathfrak{g}_u$, and since $c_u(ix) = -ix$ for all $x \in \mathfrak{g}_u$, we see that $\sigma_u(\exp_{G_0}(ix)) = \exp_{G_0}(c_u(ix)) = \exp_{G_0}(-ix) = (\exp_{G_0}(ix))^{-1}$, for all $ix \in p_0$, so $\sigma_u(s) = s^{-1}$ for all $s \in P_0$ (since $P_0 = \exp_{G_0}(p_0)$).

**Proposition 14.7.** We have $G_0 = K_0 P_0$, $\sigma_u(s) = s$ for all $s \in K_0$, and $\sigma_u(s) = s^{-1}$ for all $s \in P_0$. The group $K_0$ is the set of fixed points of $\sigma_u$, $K_0 = \{ s \in G_0 \mid \sigma_u(s) = s \}$.

**Proof.** Since every $s \in G_0 = K_0 P_0$ can be written as $s = xy$ with $x \in K_0$ and $y \in P_0$, we have $\sigma_u(s) = s$ iff $\sigma_u(xy) = xy$, but then

$$xy = \sigma_u(xy) = \sigma_u(x)\sigma_u(y) = xy^{-1},$$

so we deduce that

$$y^2 = e.$$ 

However, $y \in P_0 = \exp_{G_0}(p_0)$, so for $y = \exp_{G_0}(w)$ we have $y^2 = \exp_{G_0}(2w) = e$, and since $\exp_{G_0}$ is a diffeomorphism on $p_0$, we must have $w = 0$, and thus $y = e$. 

In summary, since $K_0$ is compact, we proved that $(G_0, K_0)$ is a Gelfand pair with involution $\sigma_u$.

The real semi-simple connected Lie group $G_0$ is called a real form of the complex semi-simple Lie group $G$ because its real Lie algebra $\mathfrak{g}_0$ is a real form of the complex Lie algebra $\mathfrak{g}$ of $G$. 
There are other real semi-simple connected Lie groups having \( \mathfrak{g}_0 \) as Lie algebra, and they can all be found (up to isomorphism) as follows; see Dieudonné [19] (Chapter XXI, no. 21.18.8-21.18.12).

**Proposition 14.8.** Let \( \tilde{G}_0 \) be the universal cover of \( G_0 \), let \( \pi: \tilde{G}_0 \to G_0 \) be the covering map, and let \( \tilde{K}_0 = \pi^{-1}(K_0) \). Then \( \tilde{K}_0 \) is isomorphic to the universal cover of the compact Lie group \( K_0 \), the exponential map \( \exp_{\tilde{G}_0} \) is a diffeomorphism of \( \mathfrak{p}_0 \) onto a closed submanifold \( \tilde{P}_0 \) of \( \tilde{G}_0 \) such that \( \tilde{K}_0 \cap \tilde{P}_0 = \{e\} \), and the map \( (x,y) \mapsto xy \) is a diffeomorphism of \( \tilde{K}_0 \times \tilde{P}_0 \) onto \( \tilde{G}_0 \). The center \( Z \) of \( \tilde{G}_0 \) is a discrete subgroup contained in the center of \( \tilde{K}_0 \).

If \( K_0 \) is not semi-simple, then \( Z \) is not equal to the center of \( \tilde{K}_0 \) and \( \tilde{K}_0 \) is not compact.

**Theorem 14.9.** Every real semi-simple connected Lie group \( G_1 \) having \( \mathfrak{g}_0 \) as real Lie algebra is of the form \( G_1 = \tilde{G}_0/D \), where \( D \) is a (discrete) subgroup of the center \( Z \) of \( \tilde{G}_0 \). The center \( C_1 \) of \( G_1 \) is given by \( C_1 = Z/D \). The Lie group \( K_1 = \tilde{K}_0/D \) is a connected subgroup of \( G_1 \) which contains \( C_1 \) (in general, \( C_1 \) is not equal to the center of \( G_1 \)), and whose Lie algebra is \( \mathfrak{t}_0 \). The Lie group \( K_1 \) is compact iff \( C_1 \) is finite. The exponential map \( \exp_{G_1} \) is a diffeomorphism of \( \mathfrak{p}_0 \) onto a closed submanifold \( P_1 \) of \( G_1 \) such that \( K_1 \cap P_1 = \{e\} \), and the map \( (x,y) \mapsto xy \) is a diffeomorphism of \( K_1 \times P_1 \) onto \( G_1 \).

**Definition 14.15.** The factorization \( G_1 = K_1 P_1 \) is called a Cartan decomposition of \( G_1 \).

The Cartan decomposition is a generalization of the polar form for invertible matrices.

It can also be shown that \( K_1 \) is isomorphic to the product of a compact Lie group with some \( \mathbb{R}^m \). Thus \( G_1 \) is diffeomorphic to the product of a compact Lie group with some \( \mathbb{R}^M \) (in fact, this compact group is maximal in \( G_1 \)).

The reasoning involving the conjugation \( \sigma_u \) can be used to show that the conjugation \( \sigma_u \) on \( \tilde{G}_0 \) such that \( (d\sigma_u)_e = c_u \) induces a conjugation on \( G_1 = \tilde{G}_0/D \) by passing to the quotient, and because \( Z \) is contained in \( \tilde{K}_0 \), that

\[
K_1 = \{ s \in G_1 \mid \sigma_u(s) = s \},
\]

and that \( \sigma_u(s) = s^{-1} \) for all \( s \in P_1 \). Since \( G_1 = K_1 P_1 \), if the center \( C_1 \) of \( G_1 \) is finite, then \( K_1 \) is compact, and then \( (G_1, K_1) \) is a Gelfand pair with involution \( \sigma_u \).

### 14.3 Spherical Functions

Our next goal is to characterize the characters of the algebra \( L^1(K\backslash G/K) \) in terms of certain functions in \( \mathcal{C}(K\backslash G/K) \) called spherical functions.

**Proposition 14.10.** Every character \( \zeta \) of the commutative Banach algebra \( L^1(K\backslash G/K) \) is given by a unique function \( \omega \in \mathcal{C}(K\backslash G/K) \) which is bounded and continuous on \( G \), with

\[
\zeta(f) = (f, \omega) = \int_G f(x)\omega(x) \, d\lambda_G(x), \quad f \in L^1(K\backslash G/K).
\]
Furthermore, \( \omega \) is uniformly continuous for every left-invariant metric on \( G \), and \( |\omega(s)| \leq \omega(e) = 1 \) for all \( s \in G \).

**Proof.** Every character \( \zeta \) of the commutative subalgebra \( L^1(K\backslash G/K) \) of \( \mathcal{M}^1(G) \) can be extended to a character \( \zeta' \) of the unital commutative Banach algebra \( L^1(K\backslash G/K) \oplus \mathbb{C} \delta_e \) by setting \( \zeta'(f + \lambda \delta_e) = \zeta(f) + \lambda \). By Theorem 8.16, we have \( |\zeta(f)| \leq \|f\|_1 \), for all \( f \in L^1(K\backslash G/K) \). As a consequence, the map \( \Phi: L^1(G) \to \mathbb{C} \) given by

\[
\Phi(f) = \zeta(f^2), \quad f \in L^1(G)
\]

is a linear form of norm \( \leq 1 \) because by Proposition 14.3, \( \|f^2\|_1 \leq \|f\|_1 \), and by Theorem 4.49, there is a unique function \( \omega_0 \in \mathcal{L}^\infty(G) \) with \( \|\omega_0\|_\infty \leq 1 \), such that

\[
\Phi(f) = \zeta(f^2) = \int_{G} f(x)\omega_0(x) \, d\lambda_G(x).
\]

The problem is that \( \omega_0 \in \mathcal{L}^\infty(G) \) is in the wrong space, because we need it to be in \( \mathcal{C}(K\backslash G/K) \) (and to be bounded by 1). To remedy this problem, we define another function \( \omega \) in terms of \( \omega_0 \), which we regularize by integrating against some function \( f_0 \in \mathcal{K}(K\backslash G/K) \). In the end, we will see that \( \omega_0 \) and \( \omega \) are equal almost everywhere, but this will take some work.

First, observe that for all \( t, t' \in K \) and all \( f \in \mathcal{K}(G) \), we have

\[
\int_{K} f(t x t') \omega_0(x) \, d\lambda_G(x) = \int_{K} f(x)\omega_0(x) \, d\lambda_G(x).
\]  

(*)

To prove this, if we let \( h_{t, t'} \) be the function given by \( h_{t, t'}(x) = f(t x t') \), then we have \( h_{t, t'}^2 = f^2 \), and so \( \zeta(f^2) = \zeta(h_{t, t'}^2) \). Indeed, using the left an right invariance of the Haar measure \( \lambda_K \), we have

\[
h_{t, t'}^2(x) = \int_{K} h_{t, t'}(t_1 x t_2) \, d\lambda_K(t_1) \, d\lambda_K(t_2)
\]

\[
= \int_{K} \int_{K} f(t_1 x t_2 t') \, d\lambda_K(t_1) \, d\lambda_K(t_2)
\]

\[
= \int_{K} \int_{K} f(t_1 x t_2) \, d\lambda_K(t_1) \, d\lambda_K(t_2) = f^2(x).
\]

By definition, the character \( \zeta \) is not identically zero, so there is some function \( f_0 \in \mathcal{K}(K\backslash G/K) \) such that \( \zeta(f_0) \neq 0 \). Since \( \zeta \) is a character of \( \mathcal{L}^1(K\backslash G/K) \), for every function \( g \in \mathcal{K}(K\backslash G/K) \), since \( f_0 \in \mathcal{K}(K\backslash G/K) \) we also have \( g * f_0 \in \mathcal{K}(K\backslash G/K) \), so \( (g * f_0)^2 = g * f_0 \), and using Fubini’s theorem and the fact that \( \mathcal{L}^1(K\backslash G/K) \) is commutative, we have

\[
\zeta(g) = \zeta(f_0)^{-1}\zeta((g * f_0)^2) = \zeta(f_0)^{-1}\zeta(g * f_0)
\]

\[
= \zeta(f_0)^{-1} \int_{G \times G} f_0(s^{-1} x) g(s)\omega_0(s) \, d\lambda_G(s) \, d\lambda_G(x)
\]

\[
= \int_{G} g(s)\omega(s) \, d\lambda_G(s),
\]
with
\[ \omega(s) = \zeta(f_0)^{-1} \int_G f_0(s^{-1}x) \omega_0(x) \, d\lambda_G(x) = \zeta(f_0)^{-1} \int_G f_0(x) \omega_0(sx) \, d\lambda_G(x). \]

It follows by Theorem 14.10.6(ii) of Dieudonné [22] (Chapter XIV, Section 10) that \( \omega \) is bounded in \( G \) and uniformly continuous for every left-invariant distance on \( G \). Observe that for the integral on the right-hand side of
\[ \zeta(g) = \int_G g(s) \omega(s) \, d\lambda_G(s) \]
to make sense, we need \( g \in \mathcal{K}(K\backslash G/K) \), since we only know that \( \omega_0 \in \mathcal{L}^\infty(G) \). We will extend the above equation to functions in \( L^1(K\backslash G/K) \) by density.

Next we need to prove that \( \omega \in \mathcal{C}(K\backslash G/K) \). For all \( t, t' \in K \), since \( f_0 \in \mathcal{K}(K\backslash G/K) \), we have
\[ \omega(tst') = \zeta(f_0)^{-1} \int_G f_0(t'^{-1}s^{-1}t^{-1}x) \omega_0(x) \, d\lambda_G(x) \]
\[ = \zeta(f_0)^{-1} \int_G f_0(s^{-1}t^{-1}x) \omega_0(x) \, d\lambda_G(x) \]
\[ = \zeta(f_0)^{-1} \int_G f_0(s^{-1}x) \omega_0(x) \, d\lambda_G(x) = \omega(x), \]
where we used (\( \ast \)) with \( f(x) = f_0(s^{-1}x) \) in the last step. Since \( \mathcal{K}(K\backslash G/K) \) is dense in \( L^1(K\backslash G/K) \), we proved that
\[ \zeta(f) = (f, \omega) = \int_G f(x) \omega(x) \, d\lambda_G(x), \quad f \in L^1(K\backslash G/K). \]

We still need to prove that \( |\omega(s)| \leq 1 \) for all \( s \in G \). Since we know that this is true of \( \omega_0 \), we prove that \([\omega] = [\omega_0] \). By Theorem 5.13 and Theorem 4.49, it suffices to prove that
\[ \int_G f(x) \omega(x) \, d\lambda(x) = \int_G f(x) \omega_0(x) \, d\lambda(x) = \zeta(f^\sharp) \quad \text{for all } f \in \mathcal{K}(G). \]

Observe that since \( \omega^\sharp = \omega \) (because \( \omega \in \mathcal{C}(K\backslash G/K) \)), and
\[ \zeta(g) = \int_G g(s) \omega(s) \, d\lambda_G(s), \quad g \in L^1(K\backslash G/K), \]
if we can prove that
\[ (f^\sharp, \psi) = \int_G f^\sharp(x) \psi(x) \, d\lambda_G(x) = \int_G f(x) \psi^\sharp(x) \, d\lambda_G(x) = (f, \psi^\sharp), \quad (**) \]
for all \( f \in \mathcal{K}(G) \) and all \( \psi \in \mathcal{C}(G) \), we will be done, because
\[
\int_G f(x)\omega_0(x) \, d\lambda(x) = \zeta(f^\sharp) = \int_G f^\sharp(s)\omega(s) \, d\lambda_G(s)
\]
\[
= \int_G f(s)\omega^\sharp(s) \, d\lambda_G(s)
\]
\[
= \int_G f(s)\omega(s) \, d\lambda_G(s),
\]
as claimed. To prove (**), using Fubini, and the fact that \( G \) is unimodular, we have
\[
\int_G f^\sharp(x)\psi(x) \, d\lambda_G(x) = \int_G \int_K \int_K f(tst')\psi(s) \, d\lambda_K(t) \, d\lambda_K(t') \, d\lambda_G(s)
\]
\[
= \int_G \int_K \int_K f(s)\psi(t^{-1}st') \, d\lambda_K(t) \, d\lambda_K(t') \, d\lambda_G(s)
\]
\[
= \int_G \int_K \int_K f(s)\psi(tst') \, d\lambda_K(t) \, d\lambda_K(t') \, d\lambda_G(s)
\]
\[
= \int_G f(x)\psi^\sharp(x) \, d\lambda_G(x).
\]

Now, since \( \omega = \omega_0 \) almost everywhere and since \( \omega \) is continuous, we have proven that there is unique function \( \omega \in \mathcal{C}(K \setminus G/K) \) satisfying the condition of the proposition, and that \( |\omega(s)| \leq 1 \), since \( \|\omega_0\|_\infty \leq 1 \). If we let \( s = e \) in
\[
\omega(s) = \zeta(f_0)^{-1} \int_G f_0(s^{-1}x)\omega_0(x) \, d\lambda_G(x),
\]
since
\[
\zeta(f_0^\sharp) = \int_G f_0(x)\omega_0(x) \, d\lambda_G(x)
\]
and \( f_0^\sharp = f_0 \) because \( f_0 \in \mathcal{K}(K \setminus G/K) \), we get
\[
\omega(e) = \zeta(f_0)^{-1} \int_G f_0(x)\omega_0(x) \, d\lambda_G(x)
\]
\[
= \zeta(f_0)^{-1}\zeta(f_0^\sharp) = \zeta(f_0)^{-1}\zeta(f_0) = 1,
\]
as claimed. \( \square \)

**Remark:** The above proof shows that
\[
\omega(s) = (f_0,\omega)^{-1} \int_G f_0(s^{-1}x)\omega(x) \, d\lambda_G(x),
\]
with \( (f_0,\omega) = \int_G f_0(s)\omega(s) \, d\lambda_G(s), \) for any function \( f_0 \in \mathcal{K}(K \setminus G/K) \) such that \( (f_0,\omega) \neq 0 \).
Definition 14.16. A bounded function \( \omega \in C(K\backslash G/K) \) is a spherical (or zonal spherical) function on \( G \) relative to \( K \), if the function
\[
f \mapsto (f, \omega) = \int_G f(s) \omega(s) \, d\lambda_G(s), \quad f \in L^1(K\backslash G/K)
\]
is a character of \( L^1(K\backslash G/K) \), which means that the map \( f \mapsto (f, \omega) \) is linear in \( f \in L^1(K\backslash G/K) \) and that
\[
(f \ast g, \omega) = (f, \omega)(g, \omega), \quad \text{for all } f, g \in L^1(K\backslash G/K).
\]

Proposition 14.10 shows that if \( \omega \) is a spherical function, then \( \omega(e) = 1 \) and \( |\omega(s)| \leq 1 \) for all \( s \in G \).

Proposition 14.11. If \( \omega \) is a spherical function, then \( \overline{\omega} \) and \( \hat{\omega} \) are also spherical functions.

Proof. For every \( f \in \mathcal{K}(K\backslash G/K) \), we have
\[
\int f(s) \overline{\omega}(s) \, d\lambda_G(s) = \overline{\int f(s) \omega(s) \, d\lambda_G(s)}
\]
and for all \( f, g \in \mathcal{K}(K\backslash G/K) \) we have
\[
\overline{f \ast g} = \overline{f} \ast \overline{g},
\]
which proves that \( \overline{\omega} \) induces a character, because
\[
\int (f \ast g)(s) \overline{\omega}(s) \, d\lambda_G(s) = \overline{\int (f \ast g)(s) \omega(s) \, d\lambda_G(s)} = \overline{\int \overline{f}(s) \omega(s) \, d\lambda_G(s) \int \overline{g}(s) \omega(s) \, d\lambda_G(s)}
\]
\[
= \int f(s) \overline{\omega}(s) \, d\lambda_G(s) \int \overline{g}(s) \omega(s) \, d\lambda_G(s)
\]
\[
= \int f(s) \overline{\omega}(s) \, d\lambda_G(s) \int \overline{g}(s) \omega(s) \, d\lambda_G(s).
\]

Since \( G \) is unimodular, we also have
\[
\int f(s) \hat{\omega}(s) \, d\lambda_G(s) = \int f(s) \omega(s^{-1}) \, d\lambda_G(s)
\]
\[
= \int f(s^{-1}) \omega(s) \, d\lambda_G(s)
\]
\[
= \int \overline{f}(s) \omega(s) \, d\lambda_G(s),
\]
and if \( f, g \in \mathcal{K}(K\backslash G/K) \), we have
\[
(f \ast g) = \hat{g} \ast \hat{f} = \hat{f} \ast \hat{g},
\]
so \( \hat{\omega} \) induces a character. \( \square \)
Observe that Proposition 14.11 shows that a spherical function is uniformly continuous for every left-invariant as well as every right-invariant distance on $G$. In general, a spherical function does not have compact support.

The next theorem gives criteria for a bounded function in $\mathcal{C}(K\backslash G/K)$ (different from the zero function) to be a spherical function.

**Theorem 14.12.** Let $\omega$ be a bounded function in $\mathcal{C}(K\backslash G/K)$ not equal to the zero function. The following properties are equivalent:

1. The function $\omega$ is a spherical function on $G$ relative to $K$.

2. We have
   $$\int_K \omega(xty) d\lambda_K(t) = \omega(x)\omega(y) \quad \text{for all } x, y \in G.$$  \hfill (s$_1$)

3. We have $\omega(e) = 1$, and for every function $f \in \mathcal{K}(K\backslash G/K)$, there is some $\lambda_f \in \mathbb{C}$ such that
   $$f \ast \omega = \lambda_f \omega.$$ \hfill (s$_2$)

In fact
$$\lambda_f = (\hat{f}, \omega) = \int_G \hat{f}(x)\omega(x) d\lambda_G(x).$$

Similarly, $\omega \ast f = \lambda_f \omega$, for the same $\lambda_f$.

**Proof.** First we prove that (3) implies (1). Assume that $f \ast \omega = \lambda_f \omega$ for every $f \in \mathcal{K}(K\backslash G/K)$. Since $\omega(e) = 1$, we have
$$\lambda_f = (\tilde{f} \ast \omega)(e) = \int_G f(s)\omega(s) d\lambda_G(s) = (f, \omega).$$

Therefore,
$$\lambda_f = (\hat{f}, \omega).$$

For all $f, g \in \mathcal{K}(K\backslash G/K)$, we have $(f \ast g) \ast \omega = (\tilde{g} \ast \tilde{f}) \ast \omega = \tilde{g} \ast (\hat{f} \ast \omega)$, which implies that
$$\lambda_{\tilde{g} \ast \hat{f}} = \lambda_{\tilde{g}} \lambda_{\hat{f}},$$
and so
$$(f \ast g, \omega) = (f, \omega)(g, \omega),$$
which shows that the map $f \mapsto (f, \omega)$ is a character. The proof is similar if $\omega \ast f = \lambda_f \omega$.

We now prove that (1) implies (3). We claim that for all $f \in \mathcal{K}(K\backslash G/K)$, we have
$$f \ast \omega = (\tilde{f}, \omega) \omega = \int_G f(s^{-1})\omega(s) d\lambda_G(s).$$
Since $f \ast \omega \in \mathcal{C}(K\backslash G/K)$, in view of (**)\namely
\[(h^\sharp, \psi) = (h, \psi^\sharp), \quad \text{for all } h \in \mathcal{K}(G) \text{ and all } \psi \in \mathcal{K}(G),\]
and since $\omega^\sharp = \omega$, it suffices to prove that
\[(g, f \ast \omega) = (\tilde{f}, \omega)(g, \omega), \quad \text{for all } g \in \mathcal{K}(K\backslash G/K),\]
because $(f \ast \omega)^\sharp = f \ast \omega^\sharp = f \ast \omega$, so
\[(h, f \ast \omega) = (h, (f \ast \omega^\sharp)) = (h^\sharp, f \ast \omega) = (\tilde{f}, \omega)(h^\sharp, \omega) = (\tilde{f}, \omega)(h, \omega^\sharp) = (\tilde{f}, \omega)(h, \omega) = (h, (\tilde{f}, \omega) \omega)\]
for all $h \in \mathcal{K}(G)$, and thus $f \ast \omega = (\tilde{f}, \omega) \omega$. In that last step, we used the fact that the map $f_1, f_2 \mapsto (f_1, f_2)$ is bilinear, so for every $\lambda \in \mathbb{C}$ we have $\lambda(f_1, f_2) = (\lambda f_1, f_2) = (f_1, \lambda f_2)$.

Using Fubini and the fact that $G$ is unimodular, we have
\[(g, f \ast \omega) = (\tilde{f} \ast g, \omega),\]
namely
\[
\int_G \int_G g(s) f(t) \omega(t^{-1}s) d\lambda_G(t) d\lambda_G(s) = \int_G \int_G g(ts) f(t) \omega(s) d\lambda_G(t) d\lambda_G(s)
= \int_G \int_G f(t^{-1}) g(t^{-1}s) \omega(s) d\lambda_G(t) d\lambda_G(s).
\]
Since $\omega$ is spherical function,
\[(g, f \ast \omega) = (\tilde{f} \ast g, \omega) = (\tilde{f}, \omega)(g, \omega).\]
Since $\tilde{\omega}$ is also spherical (by Proposition 14.11), and since $G$ is unimodular, we have
\[\omega \ast f = \overline{(\tilde{f} \ast \tilde{\omega})} = \overline{(\tilde{f}, \tilde{\omega})} \omega = (\tilde{f}, \omega) \omega.\]

Let us now prove that (2) and (3) are equivalent. For any $\omega \in \mathcal{C}(K\backslash G/K)$, define the function $h$ by
\[h(x, y) = \int_K \omega(xty) d\lambda_K(t), \quad x, y \in G.\]
A simple adaptation of the proof of Proposition 6.20 shows the map $x \mapsto h(x, y)$ is continuous. For all $t' \in K$, since $\omega(t'xy) = \omega(xty)$, we have $h(t'x, y) = h(x, y)$, and due to the invariance
of \( \lambda_K \), we have \( h(xt', y) = h(x, y) \). It follows that the function \( x \mapsto h(x, y) \) is in \( \mathcal{C}(K\backslash G/K) \).

Let us show that for every function \( f \in \mathcal{C}(K\backslash G/K) \), we have

\[
\int_G \hat{f}(x) h(x, y) \, d\lambda_G(x) = (f * \omega)(y). \tag{†}
\]

Since \( f * \omega \in \mathcal{C}(K\backslash G/K) \) and \( G \) is unimodular, by Fubini, we have

\[
\int_G \hat{f}(x) h(x, y) \, d\lambda_G(x) = \int_G \int_K \hat{f}(x) \omega(xty) \, d\lambda_K(t) \, d\lambda_G(x) \\
= \int_G \int_K f(x) \omega(x^{-1}ty) \, d\lambda_G(x) \, d\lambda_K(t) \\
= \int_K (f * \omega)(ty) \, d\lambda_K(t) = \int_K (f * \omega)(y) \, d\lambda_K(t) = (f * \omega)(y).
\]

If (2) holds, equation (s1), namely

\[
\int_K \omega(xty) \, d\lambda_K(t) = \omega(x)\omega(y),
\]

implies that \( h(x, y) = \omega(x)\omega(y) \), so by (†) we have

\[
(f * \omega)(y) = \int_G \hat{f}(x) h(x, y) \, d\lambda_G(x) = \int_G \hat{f}(x) \omega(x)\omega(y) \, d\lambda_G(x) \\
= \omega(y) \int_G \hat{f}(x) \omega(x) \, d\lambda_G(x) = (\hat{f}, \omega)\omega(y),
\]

which proves (3).

Conversely, if (3) holds, we saw in proving that (3) implies (1) that \( f * \omega = (\hat{f}, \omega)\omega \). We prove that this implies that \( h(x, y) = \omega(x)\omega(y) \), which is (s1). For any function \( g \in \mathcal{C}(G) \), by (**) using the fact that \( h, \omega \in \mathcal{C}(K\backslash G/K) \) and (†), we have

\[
(g, h(-, y)) = (g, h(-, y)^g) = (g^\sharp, h(-, y)) = (\tilde{g} \ast \omega)(y) \\
= (g^\sharp, \omega)\omega(y) = (g, \omega^\sharp)\omega(y) = (g, \omega)\omega(y) = (g, \omega(y)\omega),
\]

and so \( h(x, y) = \omega(x)\omega(y) \), as claimed. \( \square \)

**Remark:** If a bounded continuous function \( \omega \) on \( G \) not equal to the zero function satisfies the equation

\[
\int_K \omega(xty) \, d\lambda_K(t) = \omega(x)\omega(y)
\]

for all \( x, y \in G \), then it must belong to \( \mathcal{C}(K\backslash G/K) \), and thus is a spherical function. Indeed, since \( \lambda_K \) is left and right invariant, for any \( t' \in K \), we have

\[
\omega(xt')\omega(y) = \omega(x)\omega(y) = \omega(x)\omega(t'y)
\]

for all \( x, y \in G \) and all \( t' \in K \), which shows that \( \omega(xt') = \omega(x) \) and \( \omega(t'y) = \omega(y) \).
Definition 14.17. Let $S(G/K)$, or simply $S$, denote the space of spherical functions on $G$ relative to $K$. This is a subspace of $C(K\backslash G/K) \cap L^\infty(G)$.

Let $A = L^1(K\backslash G/K) \oplus \mathbb{C}\delta_e$, a commutative, involutive, unital Banach algebra. In the degenerate case where $\delta_e \in L^1(K\backslash G/K)$, the group $G$ is a discrete group and $\delta_e$ invariant by translation by the elements of $K$. This implies that $K = \{e\}$, and that $G$ is commutative and discrete. Otherwise, there is exactly one character of the algebra $A = L^1(K\backslash G/K) \oplus \mathbb{C}\delta_e$ whose restriction to $L^1(K\backslash G/K)$ is the zero function.

Definition 14.18. The subspace of characters of the commutative involutive Banach algebra $A = L^1(K\backslash G/K) \oplus \mathbb{C}\delta_e$ whose restriction to $L^1(K\backslash G/K)$ is not the zero function is denoted by $X_0(A)$. This subspace is locally compact (metrizable and separable).

For every spherical function $\omega \in S(G/K)$, let $\zeta_\omega \in X_0(A)$ be the character given by

$$\zeta_\omega(f) = (f, \omega) = \int_G f(x)\omega(x) \, d\lambda_G(x), \quad f \in L^1(K\backslash G/K).$$

The map $\omega \mapsto \zeta_\omega$ is a bijection between $S(G/K)$ and $X_0(A)$. In fact, we have the following stronger result.

Theorem 14.13. The following properties hold:

1. The map $\omega \mapsto \zeta_\omega$ is a homeomorphism of $S(G/K)$ equipped with the induced topology of Fréchet space of $C(G)$ (see Definition 1.28 and Proposition 1.27) onto $X_0(A)$ equipped with the topology induced by the weak*-topology of the dual $A'$ of $A$ (see Definition 8.9). Consequently, $S(G/K)$ is locally compact.

2. Every compact subset $L$ of $S(G/K)$ is equicontinuous.

3. The map $(x, \omega) \mapsto \omega(x)$ from $G \times S(G/K)$ to $\mathbb{C}$ is continuous.

Proof. The proof of (1) is given in Dieudonné [20] (Chapter XXII, Section 6, no. 22.6.9). This proof is very technical and makes use of the following results proven in Dieudonné [20] (Chapter XXII, Section 1, no. 22.1.11.2 and 22.1.11.2).

Proposition 14.14. For every subset $B$ of $C(G)$ consisting of uniformly bounded functions, the topology induced by the weak*-topology on $L^\infty(G)$ (see Definition 11.12) is coarser than the topology induced by the topology of $C(G)$ as a Fréchet space.

Using Proposition 14.14, the following result can be shown.

Proposition 14.15. Let $B$ be a subset of $C(G)$ consisting of uniformly bounded functions and having the following property: for every $p_0 \in B$, for every compact subset $K$ of $G$, for every $\epsilon > 0$, there is some neighborhood $U$ of $p_0$ in $B$ for the weak*-topology on $L^\infty(G)$ and some compact neighborhood $W$ of $e$ in $G$, such that for every function $p \in U$, we have

$$|(a^{-1}\chi_W \ast p)(s) - p(s)| \leq \epsilon \quad \text{for all } s \in K,$$
14.4. EXAMPLES OF GELFAND PAIRS

where \( a = \lambda_G(W) \). Then the topology on \( B \) induced by the weak\(^*\)-topology on \( L^\infty(G) \) is identical to the topology induced by the topology of \( \mathcal{C}(G) \) as a Fréchet space.

Proposition 14.15 follows from the following result also proven in Dieudonné [20] (Chapter XXII, Section 1, no. 22.1.11.5).

**Proposition 14.16.** For every function \( f \in L^1(G) \) and for every bounded subset \( B \) of the Banach space \( L^\infty(G) \), the map \( g \mapsto f \ast g \) is a continuous map from \( B \) equipped with the weak\(^*\)-topology on \( L^\infty(G) \) to the Fréchet space \( \mathcal{C}(G) \).

The proof of Proposition 14.16 uses the trick that

\[
(f \ast g)(s) = \int_G f(t)g(t^{-1}s) \, d\lambda_G(t) = \int_G f(st)g(t^{-1}) \, d\lambda_G(t) = \int_G (\lambda_{s^{-1}}f)(t)\tilde{g}(t) \, d\lambda_G(t) = (\tilde{g}, \lambda_{s^{-1}}f).
\]

(2) Pick any \( x_0 \in G \). For every compact neighborhood \( V_0 \) of \( x_0 \), by definition of the Fréchet topology on \( \mathcal{C}(G) \), the restriction map \( f \mapsto f|_{V_0} \) from \( \mathcal{C}(G) \) to \( \mathcal{C}(V; \mathbb{C}) \) is continuous, thus the image \( L_0 \) of \( L \) under this map is compact. By Ascoli III (Theorem 1.14, Dieudonné’s version), \( L_0 \) is equicontinuous. Consequently, for every \( \epsilon > 0 \) there is a neighborhood \( V \subseteq V_0 \) of \( x_0 \) such that \( |\omega(x) - \omega(x_0)| \leq \epsilon \) for all \( x \in V \) and all \( \omega \in L \).

(3) Let \((x_0, \omega_0)\) be an element of \( G \times S(G/K) \). By (2), for every \( \epsilon > 0 \), there is a compact neighborhood \( V \) of \( x_0 \) in \( G \) and a compact neighborhood \( W \) of \( \omega_0 \) in \( S(G/K) \), such that \( |\omega(x) - \omega(x_0)| \leq \epsilon \) for all \( x \in V \) and all \( \omega \in W \). By definition of the Fréchet topology on \( \mathcal{C}(G) \), there is a neighborhood \( U \subseteq W \) of \( \omega_0 \) in \( S(G/K) \) such that \( |\omega(x) - \omega_0(x)| \leq \epsilon \) for all \( x \in V \) and all \( \omega \in U \). We deduce that

\[
|\omega(x) - \omega_0(x_0)| \leq |\omega(x) - \omega(x_0)| + |\omega(x_0) - \omega_0(x_0)| \leq 2\epsilon
\]

for all \( x \in V \) and all \( \omega \in U \).

A theory of spherical functions for Lie groups, in particular symmetric spaces, not based on Gelfand pairs but instead on certain invariant differential operators is discussed in Helgason [41] (Chapter 4).

### 14.4 Examples of Gelfand Pairs

There are three important cases for which Gelfand’s theorem applies.

1. Let \( G \) be a compact, connected real Lie group, and let \( \sigma \) be an involutive automorphism of \( G \) (\( \sigma \neq \text{id}_G \)); see Dieudonné [19] (Chapter XXI, no. 21.18.13). Let \( G^\sigma \) be the closed (thus compact) Lie subgroup of \( G \) consisting of the fixed points of \( \sigma \),

\[
G^\sigma = \{ s \in G \mid \sigma(s) = s \}.
\]

A theory of spherical functions for Lie groups, in particular symmetric spaces, not based on Gelfand pairs but instead on certain invariant differential operators is discussed in Helgason [41] (Chapter 4).
The derivative $\theta = d\sigma_x$ of $\sigma$ is an involution of the Lie algebra $\mathfrak{g}$ of $G$. Then as in Section 14.2, we know by linear algebra that

$$\mathfrak{g} = \mathfrak{g}_1 \oplus \mathfrak{g}_{-1}$$

where $\mathfrak{g}_1$ and $\mathfrak{g}_{-1}$ are the real eigenspaces of $\theta$ given by

$$\mathfrak{g}_1 = \{ x \in \mathfrak{g} \mid \theta(x) = x \}, \quad \mathfrak{g}_{-1} = \{ x \in \mathfrak{g} \mid \theta(x) = -x \}.$$

Let $G_0^\sigma$ be the connected component of $e$ in $G^\sigma$. For any closed subgroup $K$ (thus compact) such that $G_0^\sigma \subseteq K \subseteq G^\sigma$, using some differential geometry, it can be shown that $\mathfrak{g}_1$ is the Lie algebra of $K$; see O’Neill [63] of Gallier and Quaintance [36] (Proposition 22.33). Let $P = \exp(\mathfrak{g}_{-1})$. Then $\sigma(s) = s$ for all $s \in K$ and $\sigma(s) = s^{-1}$ for all $s \in P$ (since $\theta(x) = -x$ for all $x \in \mathfrak{g}_{-1}$). It remains to prove that $G = KP$. For this, again we use some differential geometry.

Since $K$ is compact, $G/K$ has some $G$-invariant metric. In fact, $G/K$ is a naturally reductive homogeneous space. Since $G$ is compact, $G/K$ is compact, and by Hopf–Rinov, it is geodesically complete. But since $G/K$ is naturally reductive, the tangent space $T_e(G/K) \cong \mathfrak{g}_{-1}$, and every geodesic $\gamma_x$ through $e$ with initial velocity $x \in \mathfrak{g}_{-1}$ is given by

$$\gamma_x(t) = \pi(\exp_G(tx)),$$

see Proposition 22.27 in Gallier and Quaintance [36]. Consequently, $\pi(P) = G/K$, or equivalently $G = PK$. But since $P = \exp(\mathfrak{g}_{-1})$, we see that $P$ is closed under the map $s \mapsto s^{-1}$, and since $G = PK$, for every $s \in G$, we have $s = xy$ with $x \in P$ and $y \in K$, so $s^{-1} = y^{-1}x^{-1} \in KP$, and since this holds for any $s \in G$, we have $G = KP$. Therefore, $(G, K)$ is a Gelfand pair for the involution $\sigma$.

If the compact Lie group $G$ is also semi-simple, then its Killing form is negative definite, so $G/K$ is a symmetric space of compact type.

(2) Let $G$ be a real, compact, semi-simple, simply-connected Lie group, $\mathfrak{g}_u$ be its Lie algebra, $\mathfrak{g} = \mathfrak{g}_u \oplus i\mathfrak{g}_u$ be the complexification of $\mathfrak{g}_u$, and let $G$ be the complex, semi-simple, simply-connected Lie group with Lie algebra $\mathfrak{g}$. For any conjugation $c_0 : \mathfrak{g} \to \mathfrak{g}$ that commutes with the conjugation $c_u$ associated with $\mathfrak{g}_u$, let $\mathfrak{g}_0$ be the real form of $\mathfrak{g}$ induced by $c_0$. For any real form $G_1$ of $G$ with Lie algebra $\mathfrak{g}_0$, let $\sigma_u$ be the involution of $G_1$ such that $(d\sigma_u)_e = c_u$, as explained in Section 14.2. If the center $C_1$ of $G_1$ is finite, then $K_1 = \{ s \in G_1 \mid \sigma_u(s) = s \}$ is a compact subgroup of $G_1$ such that $(G_1, K_1)$ is Gelfand pair. The space $G_1/K_1$ is a symmetric space of non-compact type.

A typical example is given by $G = SL(n, \mathbb{C})$, $G_u = SU(n)$, $G_1 = SL(n, \mathbb{R})$, and $K_1 = SO(n, \mathbb{R})$; the maps $c_0$, $c_u$, and $\sigma_0$, are given by $c_0(X) = \overline{X}$, $c_u(X) = -X^*$, $\sigma_0(X) = \overline{X}$. 
14.4. EXAMPLES OF GELFAND PAIRS

(3) The group \( G \) is unimodular and contains

(a) A closed commutative, normal subgroup \( A \) such that \( s^2 = e \) implies \( s = e \) for all \( s \in A \), and

(b) A compact subgroup \( K \) such that the mapping \( (t, s) \mapsto ts \) from \( K \times A \) to \( G \) is a homeomorphism (this implies that \( G \) is a semi-direct product of \( K \) and \( A \)). A typical example is \( G = SE(n, \mathbb{R}) \).

Let \( \sigma \) be given by \( \sigma(ts) = ts^{-1} \), for all \( t \in K \) and all \( s \in A \). Obviously \( \sigma \) is continuous, and we have

\[
\sigma^2(ts) = \sigma(ts^{-1}) = ts,
\]

so \( \sigma^2 = \text{id}_G \). For \( t, t \in K \) and \( s, s' \in A \), since \( A \) is a normal subgroup of \( G \) and \( s \in A \), we have \( t^{-1}s^{-1}t' \in A \), and since \( s' \in A \), so we also have \( t^{-1}st's' \in A \), and we have

\[
\sigma(tst's') = \sigma((tt')(t^{-1}st's')) = tt'(s'^{-1}t^{-1}s^{-1}t'),
\]

and

\[
\sigma(ts)\sigma(t's') = ts^{-1}t's'^{-1} = tt'(t^{-1}s^{-1}t's'^{-1}).
\]

Since from above \( t^{-1}s^{-1}t' \in A \), and since \( A \) is abelian, \( t^{-1}s^{-1}t's'^{-1} = s'^{-1}t^{-1}s^{-1}t' \), and so \( \sigma(tst's') = \sigma(ts)\sigma(t's') \). Thus, \( \sigma \) is an involutive automorphism of \( G \). By definition, \( K = \{ t \in G \mid \sigma(t) = t \} \), and \( \sigma(s) = s^{-1} \) for \( s \in A \). Therefore, \( (G, K) \) is a Gelfand pair.

**Example 14.5.** Assume the group \( G \) is compact. If \( (G, K) \) is a Gelfand pair, then the closure \( L^2(K\backslash G/K) \) of \( \mathcal{K}(K\backslash G/K) \) in \( L^2(G) \) is commutative, which corresponds to the situation considered in Proposition 13.21. Furthermore, the restriction of every character of \( L^1(K\backslash G/K) \) to \( L^2(K\backslash G/K) \) is a (continuous) character of the algebra \( L^2(K\backslash G/K) \). We know that the direct sum

\[
B = \bigoplus_{(p,\sigma_0) = 1} \mathbb{C}\omega_p
\]

is a dense algebra in \( L^2(K\backslash G/K) \), and it is easy to see that the only homomorphisms from \( B \) to \( \mathbb{C} \) different from the zero function are the maps of the form \( f \mapsto (f, \overline{\omega_p}) \), which shows that the spherical functions of \( G \) relative to \( K \) are the partial traces \( \omega_p \) for all \( p \in R(G) \) such that \( (p : \sigma_0) = 1 \) (see Definition 13.12 and Proposition 13.21). Since \( \omega_p \in L^1(G) \), the set of functions \( f \in L^\infty(G) \) such that \( |(f - \omega_p, \overline{\omega_p})| \leq \frac{1}{2} \) is a neighborhood of \( \omega_p \) in the weak *-topology of \( L^\infty(G) \). Since \( (\omega_p, \overline{\omega_p}) = 1 \) and \( (\omega_{p'}, \overline{\omega_p}) = 0 \) when \( p \neq p' \), we conclude that the space \( S(G/K) \) is discrete.

The Grassmannians constitute a very good example. Let \( G = SO(n) \) (with \( n \geq 2 \)), let

\[
I_{k,n-k} = \begin{pmatrix} I_k & 0 \\ 0 & -I_{n-k} \end{pmatrix},
\]

where \( I_k \) is the \( k \times k \)-identity matrix, and let \( \sigma \) be given by

\[
\sigma(P) = I_{k,n-k}PI_{k,n-k}, \quad P \in SO(n).
\]
It is clear that $\sigma$ is an involutive automorphism of $G$. Let us find the set $G^\sigma$ of fixed points of $\sigma$. If we write

$$ P = \begin{pmatrix} Q & U \\ V & R \end{pmatrix}, \quad Q \in M_{k,k}(\mathbb{R}), \ U \in M_{k,n-k}(\mathbb{R}), \ V \in M_{n-k,k}(\mathbb{R}), \ R \in M_{n-k,n-k}(\mathbb{R}), $$

then $P = I_{k,n-k} P I_{k,n-k}$ iff

$$ \begin{pmatrix} Q & U \\ V & R \end{pmatrix} = \begin{pmatrix} I_k & 0 \\ 0 & -I_{n-k} \end{pmatrix} \begin{pmatrix} Q & U \\ V & R \end{pmatrix} \begin{pmatrix} I_k & 0 \\ 0 & -I_{n-k} \end{pmatrix} $$

iff

$$ \begin{pmatrix} Q & U \\ V & R \end{pmatrix} = \begin{pmatrix} Q & -U \\ V & R \end{pmatrix}, $$

so $U = 0$, $V = 0$, $Q \in O(k)$ and $R \in O(n-k)$. Since $P \in SO(n)$, we conclude that $\det(Q) \det(R) = 1$, so

$$ G^\sigma = \left\{ \begin{pmatrix} Q & 0 \\ 0 & R \end{pmatrix} \bigg| \ Q \in O(k), \ R \in O(n-k), \ det(R) \det(S) = 1 \right\}; $$

that is,

$$ G^\sigma = S(O(k) \times O(n-k)). $$

We also have

$$ G^\sigma_0 = SO(k) \times SO(n-k). $$

For $K = G^\sigma$, the homogeneous space

$$ G/K = SO(n)/(S(O(k) \times O(n-k))) $$

is the Grassmannian $G(k,n)$ of $k$-subspaces in $\mathbb{R}^n$. For $K = SO(k) \times SO(n-k)$, the homogeneous space

$$ G/K = SO(n)/(SO(k) \times SO(n-k)) $$

is the Grassmannian $G^0(k,n)$ of oriented $k$-subspaces in $\mathbb{R}^n$. In particular, for $k = 1$, $G^0(1,n-1) = S^{n-1}$ and $G(1,n-1) = \mathbb{RP}^{n-1}$.

**Example 14.6.** Let $G$ be a real semi-simple connected noncompact Lie group with finite center and let $K$ be a maximal compact subgroup of $K$, so that $(G,K)$ is a Gelfand pair, as in (2) above.

We showed that $G = KP$ where $P$ is closed manifold in $G$, and $P$ is closed under the map $s \mapsto s^{-1}$, but in general is not a group. However, it is known that there is closed solvable subgroup $S$ of $G$ such that $G = KS$, and that the map $(x,y) \mapsto xy$ from $K \times S$ to $G$ is a diffeomorphism. This is a corollary of the Iwasawa decomposition, which is a generalization of the $QR$-decomposition for invertible matrices; see Dieudonné [19] (Chapter
Similarly, we can write \( x = st \) and the equations just after (\ref{14.12}). Then we can extend \( \alpha \) to \( G \) as follows:

\[
\alpha(st) = \alpha(s) \quad \text{for all } s \in S \text{ and all } t \in K
\]

We claim that the following properties hold

\[
\begin{align*}
\alpha(xt) &= \alpha(x) \quad \text{for all } x \in G \text{ and all } t \in K, \\
\alpha(sx) &= \alpha(s)\alpha(x) \quad \text{for all } s \in S \text{ and all } x \in G.
\end{align*}
\]

Suppose we have a continuous homomorphism \( \alpha: S \to \mathbb{C}^* \) (called an exponential of \( S \)). Then we can extend \( \alpha \) to \( G \) as follows:

\[
\begin{align*}
\alpha(st) &= \alpha(s) \quad \text{for all } s \in S \text{ and all } t \in K
\end{align*}
\]

We claim that the following properties hold

\[
\begin{align*}
\alpha(xt) &= \alpha(x) \quad \text{for all } x \in G \text{ and all } t \in K, \\
\alpha(sx) &= \alpha(s)\alpha(x) \quad \text{for all } s \in S \text{ and all } x \in G.
\end{align*}
\]

Since for \( x \in G \) we can write \( x = st' \) with \( s \in S \) and \( t' \in K \), by (\ref{14.12}) we have

\[
\alpha(xt) = \alpha(stt') = \alpha(s) = \alpha(st') = \alpha(x).
\]

Similarly, we can write \( x = s't \) with \( s' \in S \) and \( t \in K \), so by (\ref{14.12}), we have

\[
\alpha(sx) = \alpha(ss'x) = \alpha(ss') = \alpha(s)\alpha(s') = \alpha(s)\alpha(s)t = \alpha(s)\alpha(x).
\]

Define the function \( \omega: G \to \mathbb{C} \) by

\[
\omega(x) = \int_K \alpha(tx)\,d\lambda_K(t).
\]

The function \( \omega \) is continuous, and we we claim that if it is bounded, then it is a spherical function. By the remark just after Proposition 14.12, is suffices to prove that the equation

\[
\int_K \omega(xt'y)\,d\lambda_K(t') = \omega(x)\omega(y) \quad \text{for all } x, y \in G
\]

holds. The left-hand side of this equation is

\[
\int_K \omega(xt'y)\,d\lambda_K(t') = \int_K \int_K \alpha(txt'y)\,d\lambda_K(t)\,d\lambda_K(t').
\]

We can also write for every fixed \( x \in G \) and all \( t \in K \), \( tx = s(t)u(t) \), with \( s(t) \in S \) and \( u(t) \in K \), where \( s \) and \( u \) are continuous in \( t \in K \), and using Fubini, the invariance of \( \lambda_K \), and the equations just after (\ref{14.12}),

\[
\begin{align*}
\int_K \int_K \alpha(txt'y)\,d\lambda_K(t)\,d\lambda_K(t') &= \int_K \int_K \alpha(s(t)u(t)t'y)\,d\lambda_K(t)\,d\lambda_K(t') \\
&= \int_K \int_K \alpha(s(t)t'y)\,d\lambda_K(t)\,d\lambda_K(t') \\
&= \int_K \int_K \alpha(s(t))\alpha(t'y)\,d\lambda_K(t)\,d\lambda_K(t') \\
&= \int_K \int_K \alpha(s(t)u(t))\alpha(t'y)\,d\lambda_K(t)\,d\lambda_K(t') \\
&= \int_K \int_K \alpha(tx)\alpha(t'y)\,d\lambda_K(t)\,d\lambda_K(t') = \omega(x)\omega(y).
\end{align*}
\]
CHAPTER 14. HARMONIC ANALYSIS ON GELFAND PAIRS

Harish–Chandra has shown that all continuous solutions of the functional equation \((s_1)\) are given by \((\dagger)\). Such functions may be called generalized spherical functions. He also determined explicitly the exponentials \(\alpha : S \to \mathbb{C}^*\) of \(S\), by a very deep study of the Lie algebra of \(G\). One also knows exactly when the generalized spherical functions are bounded, and thus the spherical functions are completely known.

Consider the groups \(G = \text{SL}(2, \mathbb{R})\), \(K = \text{SO}(2)\), and

\[ S = S_0 = \left\{ \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} \mid a, b \in \mathbb{R}, a > 0 \right\}, \]

from Example 13.2. We know that \(G = \text{SL}(2, \mathbb{R})\) acts transitively on the upper half plane \(P = \{ z = x + iy \in \mathbb{C} \mid y > 0 \}\) by the action given by

\[ X \cdot z = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot z = \frac{az + b}{cz + d}, \quad z = x + iy \in P, \]

and that the stabilizer of \(i\) is \(\text{SO}(2)\). Given any \(z = x + iy \in P\), there is a unique coset \(X\text{SO}(2)\subseteq \text{SL}(2, \mathbb{R})\) (where \(X \in \text{SL}(2, \mathbb{R})\)) that maps \(i\) to \(z\), and in view of the unique factorization of matrix \(X\) in \(\text{SL}(2, \mathbb{R})\) as \(X = st\) with \(s \in S_0\) and \(t \in K\), we can pick as a representative of this coset \(X\text{SO}(2)\) the matrix \(s_z \in S_0\) such that \(s_z \cdot i = z = x + iy\), namely

\[ s_z = \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix}. \]

The functions in \(\mathcal{C}(K \backslash G / K)\) can be determined as follows. Since every coset \(XK\) (with \(X \in \text{SL}(2, \mathbb{R})\)) corresponds to a unique point \(z \in P\) in the upper half plane, and since there is a unique matrix \(s_z \in S_0\) such that \(s_z \cdot i = z = x + iy\), the coset \(XK\) is uniquely represented by the matrix \(s_z \in S_0\). It follows that the double coset \(KXK = Ks_zK\) is uniquely determined by the set of matrices \(Ks_z\), and geometrically, this set of matrices corresponds to the orbit in \(P\) of the (left) action of \(K = \text{SO}(2)\) on the point \(z\). Although this is not obvious, such an orbit is a circle centered on the \(y\)-axis. To show this, we will prove that it suffices to prove that the orbit of a point \(ir\) on the imaginary axis \((r > 0)\) under the action of \(K\) is a circle of center \(iv\) and radius \(\rho\), with

\[ v = \frac{1}{2} \left( r + \frac{1}{r} \right), \quad \rho = \frac{1}{2} \left| \frac{1}{r} - r \right|. \]

If so, the equation of this circle is \(x^2 + (y - v)^2 = \rho^2\), that is,

\[ x^2 + y^2 - 2yv = \rho^2 - v^2. \]

But

\[ v^2 - 1 = \frac{1}{4}r^2 + \frac{1}{2} + \frac{1}{4r^2} - 1 = \frac{1}{4}r^2 - \frac{1}{2} + \frac{1}{4r^2} = \left( \frac{1}{2} \left( \frac{1}{r} - r \right) \right)^2 = \rho^2, \]
so $\rho^2 - v^2 = -1$, and the equation of our circle is

$$x^2 + y^2 + 1 = 2yv. \quad (\ast_1)$$

We also showed that $\rho = \sqrt{v^2 - 1}$.

We can now prove that the orbit of any point $z = x + iy$ in the upper half plane under the action of $K$ is a circle centered on the $y$-axis, which is also the orbit of a point $ir$ with $0 < r \leq 1$ under the action of $K$. Since $y > 0$, the point $z = x + iy$ is on the circle given by $(\ast_1)$ iff

$$2v = \frac{x^2 + y^2 + 1}{y}.$$ 

Observe that the right-hand side is $\geq 2$, so $v \geq 1$. We have to find $r > 0$ such that

$$r + \frac{1}{r} = 2v,$$

that is

$$r^2 - 2vr + 1 = 0,$$

and the zeros of this equation are

$$r = v \pm \sqrt{v^2 - 1}.$$ 

If $r = v - \sqrt{v^2 - 1}$, then $0 < r \leq 1$. Therefore, we found that $z \in P$ is on the circle of center $iv$ and radius $\rho$, the orbit of $ir$ by the action of $K$, with

$$v = \frac{x^2 + y^2 + 1}{2y}, \quad \rho = \sqrt{v^2 - 1}, \quad r = v - \sqrt{v^2 - 1} = v - \rho.$$ 

It remains to prove that the orbit of the point $ir$ with $0 < r \leq 1$ under the action of $K$ is the circle of center $iv$ and radius $\rho$, with

$$v = \frac{1}{2} \left( r + \frac{1}{r} \right), \quad \rho = \sqrt{v^2 - 1}.$$ 

Since a rotation in $K = \text{SO}(2)$ is of the form

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

the orbit of $ir$ consists of the points $x_\theta + iy_\theta = t_\theta \cdot ir$, with

$$x_\theta + iy_\theta = \frac{ir \cos \theta - \sin \theta}{ir \sin \theta + \cos \theta}, \quad (\ast_2)$$
From (*2), we have
\[
x_\theta + iy_\theta = \frac{ir\cos \theta - \sin \theta}{ir\sin \theta + \cos \theta} = \frac{ir\cos \theta - \sin \theta - ir\sin \theta + \cos \theta}{ir\sin \theta + \cos \theta - ir\sin \theta + \cos \theta} = \frac{(r^2 - 1)\sin \theta \cos \theta + ir}{r^2 \sin^2 \theta + \cos^2 \theta} = \frac{(r^2 - 1)\sin \theta \cos \theta + ir}{(1 - r^2)(\cos^2 \theta + r^2)} = \frac{(r^2 - 1) \sin 2\theta + ir}{(1 - r^2)(2 \cos^2 \theta - 1) + r^2 + \frac{1 - r^2}{2}} = \frac{-(1 - r^2) \sin 2\theta + i2r}{(1 - r^2) \cos 2\theta + r^2 + 1}.
\]

Let us compute \(x_\theta + iy_\theta - iv\). We have
\[
x_\theta + iy_\theta - iv = \frac{-(1 - r^2) \sin 2\theta + i2r - i(r^2 + 1)}{(1 - r^2) \cos 2\theta + r^2 + 1} = \frac{-2r(1 - r^2) \sin 2\theta + i4r^2 - i(r^2 + 1)((1 - r^2) \cos 2\theta + r^2 + 1)}{2r((1 - r^2) \cos 2\theta + r^2 + 1)} = \frac{(1 - r^2)(-2r \sin 2\theta - i(1 - r^2 + (r^2 + 1) \cos 2\theta))}{2r((1 - r^2) \cos 2\theta + r^2 + 1)}.
\]

The numerator of \(|x_\theta + i(y_\theta - v)|^2\) is
\[
N = (1 - r^2)^2(4r^2 \sin^2 2\theta + (1 - r^2)^2 + 2(1 - r^2)(r^2 + 1) \cos 2\theta + (r^2 + 1)^2 \cos^2 2\theta) = (1 - r^2)^2(4r^2(1 - \cos^2 2\theta) + (1 - r^2)^2 + 2(1 - r^2)(r^2 + 1) \cos 2\theta + (r^2 + 1)^2 \cos^2 2\theta) = (1 - r^2)^2((1 - r^2) \cos 2\theta + 2(1 - r^2)(r^2 + 1) \cos 2\theta + (1 + r^2)^2) = (1 - r^2)^2((1 - r^2) \cos 2\theta + (1 + r^2))^2.
\]

The denominator of \(|x_\theta + i(y_\theta - v)|^2\) is
\[
4r^2((1 - r^2) \cos 2\theta + (1 + r^2))^2.
\]

Therefore
\[
|x_\theta + i(y_\theta - v)|^2 = \frac{(1 - r^2)^2}{4r^2} = \left(\frac{1}{2} \left(\frac{1}{r} - r\right)\right)^2 = \rho^2 = v^2 - 1,
\]
or equivalently
\[
x_\theta^2 + (y_\theta - v)^2 = v^2 - 1,
\]
which confirms our claim that the point \(x_\theta + iy_\theta\) is on the circle of center \(v\) and radius \(\rho = \sqrt{v^2 - 1}\).
Therefore the matrices of the double class $Ks_zK$ are exactly the matrices $s_\theta t$ where $t \in K = \text{SO}(2)$ and $s_\theta \in S_0$ is the matrix

$$s_\theta = \begin{pmatrix} \sqrt{y_\theta} & x_\theta/\sqrt{y_\theta} \\ 0 & 1/\sqrt{y_\theta} \end{pmatrix}$$

corresponding to the point $x_\theta + iy_\theta = t_\theta \cdot ir$ in $(*)_2$. The matrices $s_\theta$ satisfy the property

$$\text{tr}(s_\theta^\top s_\theta) = \frac{x^2_\theta + y^2_\theta + 1}{y_\theta} = 2v,$$

which characterizes them. Now one has $(s_\theta t)^\top (s_\theta t) = t^\top (s_\theta^\top s_\theta) t$, and as $t^\top = t^{-1}$, we get

$$\text{tr}((s_\theta t)^\top (s_\theta t)) = \text{tr}(s_\theta^\top s_\theta).$$

We deduce that the matrices

$$X = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in G = \text{SL}(2,\mathbb{R}), \quad \text{with} \quad X = s_\theta t,$$

which form the double class $Ks_zK$, are exactly those for which

$$\text{tr}(X^\top X) = a^2 + b^2 + c^2 + d^2 = 2v,$$

with $v \geq 1$. Hence the functions $f \in C(K\backslash G/K)$ are those which may be written as $f(1/2(a^2 + b^2 + c^2 + d^2))$, where $f$ is a continuous function defined on the interval $[1, +\infty)$.

We can also determine the exponentials $\alpha: S \to \mathbb{C}^*$ of $S$. Given any group $G$, recall that for any two elements $a, b \in G$, the element $a^{-1}b^{-1}ab$ denoted by $[a, b]$ is the commutator of $a$ and $b$, and that the subgroup of $G$ generated by the commutators is called the commutator subgroup of $G$ and is denoted by $[G, G]$. If $\alpha: S \to \mathbb{C}^*$ is a (continuous) homomorphism, then obviously $\alpha$ has the value 1 on the commutator subgroup $[S, S]$ of $S$ (since $\alpha(a^{-1}b^{-1}ab) = \alpha(a)^{-1}\alpha(b)^{-1}\alpha(a)\alpha(b) = \alpha(a)^{-1}\alpha(b)^{-1}\alpha(a)\alpha(b) = 1$). But since

$$\begin{pmatrix} a_1 & b_1 \\ 0 & a_1^{-1} \end{pmatrix} \begin{pmatrix} a_2 & b_2 \\ 0 & a_2^{-1} \end{pmatrix} = \begin{pmatrix} a_1a_2 & a_1b_2 + b_1a_2^{-1} \\ 0 & a_1^{-1}a_2^{-1} \end{pmatrix},$$

for any two matrices in $S$, we see that

$$[S, S] = \left\{ \begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix} \bigg| \ b \in \mathbb{R} \right\}. $$

Furthermore, every matrix $X \in S$ can be factored as

$$X = \begin{pmatrix} a & b \\ 0 & a^{-1} \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \begin{pmatrix} 1 & a^{-1}b \\ 0 & 1 \end{pmatrix}.$$
with $a > 0$, so the homomorphism $\alpha$ is determined by its restriction to the subgroup

$$\left\{ \begin{pmatrix} a & 0 \\ 0 & a^{-1} \end{pmatrix} \bigg| a > 0 \right\} \cong \mathbb{R}^*_+,$$

and thus it corresponds to a continuous homomorphism from $\mathbb{R}^*_+$ to $\mathbb{C}^*$, which is well known to be of the form $t \mapsto t^\lambda = e^{\lambda \log t}$, for any $\lambda \in \mathbb{C}$. We showed earlier that

$$x + iy = \frac{ir \cos \theta - \sin \theta}{ir \sin \theta + \cos \theta} = \frac{(r^2 - 1) \sin \theta \cos \theta + ir}{r^2 \sin^2 \theta + \cos^2 \theta},$$

so we get

$$\frac{1}{y} = r \sin^2 \theta + \frac{1}{r} \cos^2 \theta.$$

Using the fact that

$$v^2 - 1 = \left( \frac{1}{2} \left( \frac{1}{r} - r \right) \right)^2,$$

we get

$$\frac{1}{y} = r \sin^2 \theta + \frac{1}{r} \cos^2 \theta
= v + r \sin^2 \theta - \frac{r}{2} + \frac{1}{r} \cos^2 \theta - \frac{1}{2r} = v + \frac{r}{2} (2 \sin^2 \theta - 1) + \frac{1}{2r} (2 \cos^2 \theta - 1)
= v + \frac{r}{2} (1 - 2 \cos^2 \theta) + \frac{1}{2r} (2 \cos^2 \theta - 1)
= v + \frac{1}{2} \left( \frac{1}{r} - r \right) \cos 2\theta
= v + \sqrt{v^2 - 1} \cos 2\theta.$$

Since

$$X = \begin{pmatrix} \sqrt{y} & x/\sqrt{y} \\ 0 & 1/\sqrt{y} \end{pmatrix},$$

the above reasoning show that

$$\alpha(XU) = \alpha(X) = y^{-\rho},$$

where $\rho \in \mathbb{C}$ is any complex number, so finally, the generalized spherical function given by

$$P_\rho(v) = \frac{1}{2\pi} \int_0^{2\pi} (v + \sqrt{v^2 - 1} \cos \varphi)^\rho d\varphi, \quad v \geq 1, \: \rho \in \mathbb{C}.$$
14.4. EXAMPLES OF GELFAND PAIRS

In order for the function \( P_\rho \) to be bounded when \( v \geq 1 \), some conditions must be imposed on \( \rho \). If \( \rho \) is purely imaginary, then \( P_\rho \) is bounded, but if \( \rho \) is a positive real, then it is not bounded.

One can check that the functional equation \((s_1)\) becomes

\[
\frac{1}{2\pi} \int_0^{2\pi} P_\rho(\cosh t \cosh u + \sinh t \sinh u \cos \varphi) \, d\varphi = P_\rho(\cosh t)P_\rho(\cosh u),
\]

for all \( t, u \in \mathbb{R} \).

**Example 14.7.** Let now consider Case (3) above, where \( G \) is a unimodular group containing an abelian normal subgroup \( A \) and a compact subgroup \( K \) such that the map \((t, s) \mapsto ts\) is a homeomorphism from \( K \times A \) to \( G \). Let \( \alpha : A \to \mathbb{C}^* \) be a continuous homomorphism (an exponential of \( A \)). By analogy with Example 14.6, define the function \( \omega : G \to \mathbb{C} \) by

\[
\omega(x) = \int_K \alpha(usu^{-1}) \, d\lambda_K(u), \quad x = ts, \ t \in K, \ s \in A.
\]

The function \( \omega \) is continuous, and we claim that if \( \omega \) is bounded, then it is a spherical function for \((G,K)\). For this, we verify that the functional equation \((s_1)\) holds.

Let \( x = t_1s_1, y = t_2s_2 \), with \( t_1, t_2 \in K, s_1, s_2 \in A \). For \( v \in K \), we may write

\[
xvy = t_1s_1vt_2s_2 = (t_1vt_2)((v_2t_2)^{-1}s_1(v_2t_2)s_2),
\]

with \( t_1vt_2 \in K \) and \((v_2t_2)^{-1}s_1(v_2t_2)s_2 \in A \), because \( A \) is normal, so \((v_2t_2)^{-1}s_1(v_2t_2) \in A \), and \((v_2t_2)^{-1}s_1(v_2t_2)s_2 \in A \). Consequently, since \( \alpha \) is a homomorphism of \( A \), the subgroup \( A \) is a normal subgroup, and by Fubini, we have

\[
\int_K \omega(xvy) \, d\lambda_K(v) = \int_K \int_K \alpha((v_2t_2)^{-1}s_1(v_2t_2)s_2u^{-1}) \, d\lambda_K(u) \, d\lambda_K(v)
\]

\[
= \int_K \int_K \alpha((v_2t_2u^{-1})^{-1}s_1(v_2t_2u^{-1})) \, d\lambda_K(u) \, d\lambda_K(v)
\]

\[
= \int_K \int_K \alpha((v_2t_2u^{-1})^{-1}s_1(v_2t_2u^{-1})) \alpha(us_2u^{-1}) \, d\lambda_K(u) \, d\lambda_K(v)
\]

\[
= \int_K \alpha(us_2u^{-1}) \int_K \alpha((v_2t_2u^{-1})^{-1}s_1(v_2t_2u^{-1})) \, d\lambda_K(v) \, d\lambda_K(u).
\]

But since \( K \) is unimodular, we have

\[
\int_K \alpha((v_2t_2u^{-1})^{-1}s_1(v_2t_2u^{-1})) \, d\lambda_K(v) = \int_K \alpha(v^{-1}s_1v) \, d\lambda_K(v) = \int_K \alpha(vs_1v^{-1}) \, d\lambda_K(v) = \omega(x),
\]
and thus
\[
\int_K \omega(xvy) \, d\lambda_K(v) = \int_K \alpha(us_2u^{-1}) \left( \int_K \alpha((vt_2u^{-1})^{-1}s_1(vt_2u^{-1})) \, d\lambda_K(v) \, d\lambda_K(u) \right)
\]
\[
= \int_K \alpha(us_2u^{-1})\omega(x) \, d\lambda_K(u) = \omega(x) \omega(y),
\]
as claimed.

Conversely, it can be shown that all spherical functions are given by (†). A proof is sketched in Dieudonné [18] (Chapter 16).

Consider the example \( G = \text{SE}(2, \mathbb{R}) \) of the group of rigid motions of \( \mathbb{R}^2 \), consisting of the matrices

\[
\text{SE}(2, \mathbb{R}) = \left\{ \begin{pmatrix} \cos \theta & -\sin \theta & a \\ \sin \theta & \cos \theta & b \\ 0 & 0 & 1 \end{pmatrix} \mid a, b \in \mathbb{R}, \ 0 \leq \theta \leq 2\pi \right\}.
\]

Here \( K = \text{SO}(2) \) and \( A = \mathbb{R}^2 \), the group of translations of the plane.

For any matrix \( s \in \text{SE}(2, \mathbb{R}) \) and any matrices \( t_1, t_2 \in \text{SO}(2) \), if we write

\[
s = \begin{pmatrix} R & x \\ 0 & 1 \end{pmatrix}, \quad t_1 = \begin{pmatrix} Q_1 & 0 \\ 0 & 1 \end{pmatrix}, \quad t_2 = \begin{pmatrix} Q_2 & 0 \\ 0 & 1 \end{pmatrix}, \quad Q_1, Q_2, R \in \text{SO}(2), \ x \in \mathbb{R}^2,
\]

then we have

\[
t_1st_2 = \begin{pmatrix} Q_1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} R & x \\ 0 & 1 \end{pmatrix} \begin{pmatrix} Q_2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} Q_1R & Q_1x \end{pmatrix} \begin{pmatrix} Q_2 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} Q_1RQ_2 & Q_1x \\ 0 & 1 \end{pmatrix}.
\]

If we pick \( Q_1 = I \) and \( Q_2 = R^\top \), then we see that the matrix

\[
\begin{pmatrix} I & x \\ 0 & 1 \end{pmatrix}
\]

belong to the class \( KsK \). For any other vector \( y \) such that \( \|y\| = \|x\| = r^2 \ (r \geq 0) \), since \( \text{SO}(2) \) acts transitively on \( S^1 \), there is some \( Q \in \text{SO}(2) \) such that \( Qx = y \), so the matrix

\[
\begin{pmatrix} I & y \\ 0 & 1 \end{pmatrix}
\]

also belongs to double class \( KsK \). Therefore, every double class \( KsK \) corresponds bijectively to some \( r \in \mathbb{R} \) with \( r \geq 0 \). We can also view such a double class \( KsK \) as any vector \((a, b)\) for which \( a^2 + b^2 = r^2 \) for a fixed \( r \geq 0 \) in \( \mathbb{R} \). The functions in \( C(K \backslash G/K) \) are those of the form \( \psi(r) \), where \( \psi: [0, +\infty) \to \mathbb{C} \) is any continuous function. We showed earlier that the continuous homomorphisms \( \alpha: \mathbb{R}^2 \to \mathbb{C}^* \) are of the form

\[
\alpha(a, b) = e^{\lambda a + \mu b}, \quad \lambda, \mu \in \mathbb{C}.
\]
For any $u \in K$ and any $s \in A$, we have
\[
usu^{-1} = \begin{pmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
0 & 0 & a \\
0 & 0 & b \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
\cos \theta & \sin \theta & 0 \\
-\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
0 & 0 & a \\
0 & 0 & b \\
0 & 0 & 1
\end{pmatrix}
= \begin{pmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
0 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 1
\end{pmatrix}
= \begin{pmatrix}
\cos \theta & -\sin \theta & 0 \\
\sin \theta & \cos \theta & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
0 & 0 & \cos \theta a + \sin \theta b \\
0 & 0 & \sin \theta a - \sin \theta b \\
0 & 0 & 1
\end{pmatrix}.
\]

For $a = r \cos \varphi$ and $b = r \sin \varphi$, so that $a^2 + b^2 = r^2$, we have
\[
usu^{-1} = \begin{pmatrix}
0 & 0 & r \cos(\theta + \varphi) \\
0 & 0 & r \sin(\theta + \varphi) \\
0 & 0 & 1
\end{pmatrix}.
\]

Consequently, according to (†), for
\[
x = ts = t \begin{pmatrix}
0 & 0 & r \cos \varphi \\
0 & 0 & r \sin \varphi \\
0 & 0 & 1
\end{pmatrix}, \quad t \in K.
\]
we have
\[
\omega(x) = \int_{K} \alpha(usu^{-1}) d\lambda_{K}(u) = \frac{1}{2\pi} \int_{0}^{2\pi} e^{r(\lambda \cos(\theta + \varphi) + \mu \sin(\theta + \varphi))} d\theta = \frac{1}{2\pi} \int_{0}^{2\pi} e^{r(\lambda \cos \theta + \mu \sin \theta)} d\theta.
\]

It follows that the generalized spherical functions are the continuous functions on $[0, +\infty)$ given by
\[
\psi(r) = \frac{1}{2\pi} \int_{0}^{2\pi} e^{r(\lambda \cos \theta + \mu \sin \theta)} d\theta,
\]
for any $\lambda, \mu \in \mathbb{C}$. For $\lambda$ and $\mu$ imaginary, these functions are bounded, hence they really are spherical functions. In the special case where $\lambda = 0$ and $\mu = i$, the function $\psi$ is the Bessel function $J_0$. By a change of variable, if both $\lambda$ and $\mu$ are imaginary, the function $\psi$ becomes $J_0$.

The irreducible unitary representations of $\text{SE}(2, \mathbb{R})$ are completely determined and can be expressed by means of Bessel functions; see Vilenkin [82].

In the general case where $G = \text{SE}(n, \mathbb{R})$, with $K = \text{SO}(n)$ and $A = \mathbb{R}^n$, with
\[
\text{SE}(n, \mathbb{R}) = \left\{ \begin{pmatrix} Q & w \\ 0 & 1 \end{pmatrix} \mid Q \in \text{SO}(n), w \in \mathbb{R}^n \right\},
\]
one can also determine the generalized spherical functions, and they are now expressed in terms of the Bessel functions $J_{(n-2)/2}$; see Vilenkin [82].
In general, if $G$ is a connected unimodular Lie group and $K$ is a compact subgroup of $G$ such that $(G, K)$ is a Gelfand pair, it can be shown that the spherical functions are not only continuous but also smooth. The proof is not difficult but not that informative, so omit it. This proof can be found in Dieudonné [18] (Chapter 12).

It is also possible to figure out how a differential operator on $G$ that is invariant by left translations by $G$ and invariant by right translations by $K$ operates on spherical functions. In this case, the spherical functions are eigenfunctions of all such differential operators. If $G$ is also semi-simple, then more can be said (there are elliptic operators, in particular, the Casimir operator), but will not go into this right now. These topics are discussed in Dieudonné [21] (Chapter XXIII, Sections 36 and 37).

14.5 The Fourier Transform

Again, let $(G, K)$ be a Gelfand pair. Recall from Definition 14.18 that every spherical function $\omega \in S(G/K)$ defines the character $\zeta_\omega \in X_0(A)$ given by

$$
\zeta_\omega(f) = (f, \omega) = \int_G f(x) \omega(x) \, d\lambda_G(x), \quad f \in L^1(K \backslash G/K),
$$

where $A = L^1(K \backslash G/K) \oplus \mathbb{C}\delta_e$, a commutative, involutive, unital Banach algebra. By Theorem 14.13 the map $\omega \mapsto \zeta_\omega$ is a homeomorphism of $S(G/K)$ equipped with the induced topology of Fréchet space of $\mathcal{C}(G)$, onto $X_0(A)$ equipped with the topology induced by the weak *-topology of the dual $A'$ of $A$.

It follows that the restriction of the Gelfand map to $X_0(A)$ of an element $f \in L^1(K \backslash G/K)$ can be identified with the function $\mathcal{F}f : S(G/K) \to \mathbb{C}$ given by

$$
(\mathcal{F}f)(\omega) = (f, \omega) = \int_G f(x) \omega(x) \, d\lambda_G(x).
$$

The above can be viewed as the Fourier cotransform of $f$. Thus we are led to the following definition.

**Definition 14.19.** Let $(G, K)$ be a Gelfand pair (recall that $G$ is unimodular). For every function $f \in \mathcal{L}^1(K \backslash G/K)$, the Fourier cotransform $\mathcal{F}f$ of $f$ is the function $\mathcal{F}f : S(G/K) \to \mathbb{C}$ given by

$$
(\mathcal{F}f)(\omega) = (f, \omega) = \int_G f(x) \omega(x) \, d\lambda_G(x), \quad \omega \in S(G/K),
$$

and the Fourier transform $\mathcal{F}f$ of $f$ is the function $\mathcal{F}f : S(G/K) \to \mathbb{C}$ given by

$$
(\mathcal{F}f)(\omega) = (\tilde{f}, \omega) = \int_G f(x^{-1}) \omega(x) \, d\lambda_G(x)
$$

$$
= (f, \tilde{\omega}) = \int_G f(x) \omega(x^{-1}) \, d\lambda_G(x), \quad \omega \in S(G/K).$$
Observe that
\[ \mathcal{F} f = \overline{f}. \]
It is also clear that \( \mathcal{F} f \) and \( \overline{f} \) depend only on the equivalence class \([f] \in L^1(K \backslash G/K)\), so the Fourier transform and the Fourier cotransform are also defined on \( L^1(K \backslash G/K) \).

Definition 14.19 also applies to arbitrary functions \( f \in \mathcal{L}^1(G) \). By (***) of Section 14.3, namely
\[ (f^\sharp, \psi) = \int_G f^\sharp(x) \psi(x) \, d\lambda_G(x) = \int_G f(x) \psi^\sharp(x) \, d\lambda_G(x) = (f, \psi^\sharp), \]
for all \( f \in \mathcal{K}(G) \) and all \( \psi \in \mathcal{C}(G) \), since \( \omega^\sharp = \omega \), we obtain
\[ \mathcal{F} f = \mathcal{F}(f^\sharp), \quad \overline{f} = \overline{f}\]
for all \( f \in \mathcal{L}^1(G) \).

The Fourier transform and the Fourier cotransform are also defined on \( L^1(G) \).

As a consequence of the properties of the Gelfand transform, we have the following results.

**Proposition 14.17.** Let \((G, K)\) be a Gelfand pair. For every function \( f \in \mathcal{L}^1(G) \), the Fourier transform \( \mathcal{F} f \) and the Fourier cotransform \( \overline{f} \) are continuous functions that tend to zero at infinity. For all \( f, g \in \mathcal{L}^1(K \backslash G/K) \), we have
\[ \mathcal{F}(f \ast g) = (\mathcal{F} f)(\mathcal{F} g), \quad \overline{f} \ast g = (\overline{f})(\mathcal{F} g). \]
For all \( f \in \mathcal{L}^1(G) \), we have
\[ \| \mathcal{F} f \| \leq \| f \|_1, \quad \| \overline{f} \| \leq \| f \|_1. \]
Therefore \( \mathcal{F} \) and \( \overline{f} \) are continuous linear maps from \( L^1(G) \) to \( C_0(S(G/K); \mathbb{C}) \).

Beware that the equations in (**) generally fail if \( f, g \in \mathcal{L}^1(G) \). Also, in general, even if \( f \in \mathcal{K}_C(G) \), the functions \( \mathcal{F} f \) and \( \overline{f} \) do not have compact support. However, we have the following properties.

**Proposition 14.18.** Let \( f \) and \( g \) be two functions in \( \mathcal{L}^1(G) \). If either \( f \in \mathcal{L}^1(G/K) \) or \( g \in \mathcal{L}^1(K \backslash G) \), then
\[ \mathcal{F}(f \ast g) = (\mathcal{F} f) \ast (\mathcal{F} g), \quad \text{and} \quad \overline{f} \ast g = (\overline{f})(\mathcal{F} g). \]

**Proof.** We prove the first equation assuming that \( f \in \mathcal{L}^1(G/K) \), the proof of the other equations being similar. By left-invariance, we have
\[
\mathcal{F}(f \ast g) = \int_G \int_G \omega(x^{-1}) f(s) g(s^{-1} x) \, d\lambda_G(s) \, d\lambda_G(x)
= \int_G \int_G \omega(x^{-1}s^{-1}) f(s) g(x) \, d\lambda_G(s) \, d\lambda_G(x),
\]
and by right-invariance, since \( f(st^{-1}) = f(s) \) for all \( t \in K \), \( \lambda_K(K) = 1 \), and by \((s_1)\) of Theorem 14.12, we have

\[
\mathcal{F}(f * g) = \int_G \omega(x^{-1}s^{-1})f(s)g(x) \, d\lambda_G(s) \, d\lambda_G(x) \\
= \int_G \int_G \omega(x^{-1}ts^{-1})f(st^{-1})g(x) \, d\lambda_G(s) \, d\lambda_G(x) \\
= \int_G \int_G \omega(x^{-1}ts^{-1})f(s)g(x) \, d\lambda_G(s) \, d\lambda_G(x) \\
= \int_G \int_K \omega(x^{-1}ts^{-1}) \, d\lambda_K(t) f(s)g(x) \, d\lambda_G(s) \, d\lambda_G(x) \\
= \int_G \int_G \omega(x^{-1}) \, d\lambda_G(s) \, \left( \int_G \omega(x^{-1})g(x) \, d\lambda_G(x) \right) = (\mathcal{F}f) * (\mathcal{F}g),
\]

as claimed.

In general, \( \mathcal{F}(f * g) \neq (\mathcal{F}f) * (\mathcal{F}g) \) if \( f \in \mathcal{L}^1(G/K) \) and \( g \in \mathcal{L}^1(K \setminus G) \).

The following properties also hold.

**Proposition 14.19.** Let \((G, K)\) be a Gelfand pair. For every function \( f \in \mathcal{L}^1(K \setminus G) \), every \( s \in G \), and every \( \omega \in S(G/K) \), we have

\[
\mathcal{F}(\omega^{-1}s f)(\omega) = \omega(s^{-1}) \mathcal{F}(f)(\omega), \quad \mathcal{F}(\omega^{-1}sf)(\omega) = \omega(s) \mathcal{F}(f)(\omega),
\]

and for every function \( f \in \mathcal{L}^1(G/K) \), every \( s \in G \), and every \( \omega \in S(G/K) \), we have

\[
\mathcal{F}(\omega^{-1}sf)(\omega) = \omega(s) \mathcal{F}(f)(\omega), \quad \mathcal{F}(\omega^{-1}sf)(\omega) = \omega(s^{-1}) \mathcal{F}(f)(\omega).
\]

**Proof.** We prove that \( \mathcal{F}(\omega^{-1}sf)(\omega) = \omega(s) \mathcal{F}(f)(\omega) \), the proof for the other formulae being similar. Since \( f(tx) = f(x) \) for all \( t \in K \) and almost all \( x \in G \), and since \( \lambda_G \) is left-invariant, we have

\[
\mathcal{F}(\omega^{-1}sf)(\omega) = \int_G f((s^{-1}x)\omega(x) \, d\lambda_G(x) = \int_G f(x) \omega(sx) \, d\lambda_G(x) \\
= \int_G f(tx) \omega(stx) \, d\lambda_G(x) = \int_G f(x) \omega(stx) \, d\lambda_G(x).
\]

Then, since the rightmost integral above is independent of \( t \) because \( \mathcal{F}(\omega^{-1}sf)(\omega) \) is independent of \( t \), by \((s_1)\) (from Theorem 14.12) and Fubini, we have

\[
\int_G f(x) \omega(stx) \, d\lambda_G(x) = \int_G \int_K f(x) \omega(stx) \, d\lambda_G(x) \, d\lambda_K(t) \\
= \int_G f(x) \int_K \omega(stx) \, d\lambda_K(t) \, d\lambda_G(x) \\
= \omega(s) \int_G f(x) \omega(x) \, d\lambda_G(x) = \omega(x) \mathcal{F}(f)(\omega),
\]
14.6. THE PLANCHEREL TRANSFORM

as claimed.

In the next section we try to generalize Fourier inversion.

14.6 The Plancherel Transform

As in the previous section, let $(G, K)$ be a Gelfand pair. If $G$ is compact, then by Example 14.5 and Proposition 13.21, the spherical functions in $\mathbf{S}(G/K)$ are of positive type (recall Definition 11.10). However, when $G$ is not compact, the spherical functions in $\mathbf{S}(G/K)$ are not necessarily of positive type. The subspace of spherical functions of positive type is deeply related to the measures of positive type (recall Definition 11.14) and is the domain of certain positive measures that yield a kind of Fourier inversion.

Definition 14.20. The subset of $\mathbf{S}(G/K)$ consisting of the spherical functions of positive type is denoted by $\mathbf{Z}(G/K)$. This space is equipped with the induced topology of Fréchet space of $\mathcal{C}(G)$.

In view of Theorem 11.11, the space $\mathbf{Z}(G/K)$ is closed in $\mathbf{S}(G/K)$, and thus it is locally compact.

Given a measure of positive type $\mu$ on $G$ (see Definition 11.14), recall from Section 11.6 that the linear map $\varphi_\mu: \mathcal{K}_C(G) \to \mathbb{C}$ given by

$$\varphi_\mu(f) = \int f(s) \, d\mu(s)$$

is a positive linear form in the sense of Definition 10.10. As in Section 11.4, the set

$$n = \{ f \in \mathcal{K}_C(G) \mid \varphi_\mu(f^* \ast f) = 0 \}$$

is a left ideal in $\mathcal{K}_C(G)$, and $H_0 = \mathcal{K}_C(G)/n$ is a hermitian space with the hermitian inner product

$$\langle \pi(f), \pi(g) \rangle_\mu = \varphi_\mu(g^* \ast f) = \int (g^* \ast f)(s) \, d\mu(s), \quad (\dagger_1)$$

where $\pi: \mathcal{K}_C(G) \to \mathcal{K}_C(G)/n$ is the quotient map. Since

$$\int (g^* \ast f)(s) \, d\mu(s) = \iint \overline{g(t^{-1})} f(t^{-1}s) \, d\lambda(t) \, d\mu(s)$$
$$= \iint \overline{g(t)} f(ts) \, d\lambda(t) \, d\mu(s),$$

we have

$$\langle \pi(f), \pi(g) \rangle_\mu = \varphi_\mu(g^* \ast f) = \int \int \overline{g(t)} f(ts) \, d\lambda(t) \, d\mu(s). \quad (\dagger_2)$$
The hermitian space $H_0 = \mathcal{K}_c(G)/\mathfrak{n}$ is separable, and we let $H_\mu$ be the Hilbert space which is the completion of $H_0$. By Theorem 11.19, the measure of positive type $\mu$ defines a unitary representation $U_\mu: G \to \mathbf{U}(H_\mu)$, where $U_\mu(s)$ is the extension of the map $U_\mu(s) \in \mathbf{GL}(H_0)$ defined by

$$U_\mu(s)(\pi(f)) = \pi(\delta_s \ast f), \quad \text{for all } s \in G \text{ and all } f \in \mathcal{K}_c(G).$$

By Theorem 11.9, the unitary representation $U_\mu: G \to \mathbf{U}(H_\mu)$ extends to a non-degenerate algebra representation $(U_\mu)_{\text{ext}}: L^1(G) \to \mathcal{L}(H_\mu)$.

The map $\gamma$ defined on $\mathcal{K}(G) \times \mathcal{K}(G)$ by

$$\gamma(g, h) = \varphi_\mu(h^* \ast g)$$

satisfies the Conditions (U) and (N) of Section 10.5, and thus the restriction of $\gamma$ to the involutive and commutative subalgebra $\mathcal{K}(K\backslash G/K)$ is a bitrace (see Definition 10.13), which also satisfies Condition (U). Actually, this bitrace also satisfies Condition (N). This can be shown using a regularization argument that we omit. For details, see Dieudonné [20] (Chapter XXII, Section 7). Then $\mathcal{K}(K\backslash G/K)$ and $\pi(\mathcal{K}(K\backslash G/K))$ are commutative Hilbert algebras.

Let $\mathcal{H}_\mu$ be the closure of $\pi(\mathcal{K}(K\backslash G/K))$ in $H_\mu$. Then the map $f \mapsto U_\mu(f)|_{\mathcal{H}_\mu}$ is a representation of $\mathcal{K}(K\backslash G/K)$ in the separable Hilbert space $\mathcal{H}_\mu$ that we denote $V_\mu$. Thus we have

$$V_\mu(f)(\pi(g)) = \pi(f \ast g) \quad f, g \in \mathcal{K}(K\backslash G/K),$$

and we have

$$\|V_\mu(f)\| \leq \|f\|_1.$$

This means that $V_\mu$ is a continuous algebra homomorphism from $\mathcal{K}(K\backslash G/K)$ (with the topology induced by the topology of $L^1(K\backslash G/K)$) to the algebra $\mathcal{L}(\mathcal{H}_\mu)$ of continuous linear operators of $\mathcal{H}_\mu$. Since $L^1(G)$ is a separable Banach algebra, we deduce that the closure $\mathcal{A}_\mu$ of $V_\mu(K\backslash G/K)$ is a $C^*$ commutative separable subalgebra of $\mathcal{L}(\mathcal{H}_\mu)$.

Thus we see that the bitrace obtained by restricting the bitrace $\gamma$ to $\mathcal{K}(K\backslash G/K)$ satisfies all the hypotheses of the Plancherel–Godement theorem (Theorem 10.37). The Plancherel–Godement theorem yields the following result.

**Theorem 14.20.** (Plancherel Transform Theorem) Let $(G, K)$ be a Gelfand pair. For every measure of positive type $\mu$ on $G$, there is a unique (positive) Radon measure $\mu^\triangle$ defined on the locally compact space $Z(G/K)$ of spherical functions of positive type, such that for every function $f \in \mathcal{K}(K\backslash G/K)$, the Fourier cotransform $\overline{\mathcal{F}}f$ belong to $L^2_{\mu^\triangle}(Z(G/K); \mathbb{C})$, and for any two functions $f, g \in \mathcal{K}(K\backslash G/K)$, we have

$$\int_G (g^* \ast f) d\mu = \int_{Z(G/K)} (\overline{\mathcal{F}}f)(\omega)(\overline{\mathcal{F}}g)(\omega) d\mu^\triangle(\omega).$$

The map $\Phi: f \mapsto [\overline{\mathcal{F}}f]$ from $\mathcal{K}(K\backslash G/K)$ to $L^2_{\mu^\triangle}(Z(G/K); \mathbb{C})$ factors as

$$\Phi = T_0 \circ \pi,$$
with \( T_0 : \pi(K(K\setminus G/K)) \to L^2_{\mu}(\mathbb{Z}(G/K); \mathbb{C}) \), and \( T_0 \) extends to an isomorphism between the Hilbert space \( \mathcal{H}_{\mu} \) and the Hilbert space \( L^2_{\mu}(\mathbb{Z}(G/K); \mathbb{C}) \).

The proof of Theorem 14.20 requires some knowledge of the proof of the Plancherel–Godement theorem and is omitted. The reader is referred to Dieudonné [20] (Chapter XXII, Section 7, Theorem 22.7.4).

**Definition 14.21.** Let \((G, K)\) be a Gelfand pair. For every complex measure \( \mu \) of positive type on \( G \), the (positive) Radon measure \( \mu^\triangle \) on \( \mathbb{Z}(G/K) \) given by Theorem 14.20 is called the Plancherel transform of \( \mu \).

It is also useful to define the projection of a complex measure \( \mu \in \mathbb{C}\mathcal{M}^1(G) \) onto the subspace \( \mathbb{C}\mathcal{M}^1(K\setminus G/K) \) of complex measures invariant by left and right translations by elements of \( K \).

First, assume that \( \mu \) is a positive finite measure. We define the linear functional \( \Phi^\sharp_\mu \) by

\[
\Phi^\sharp_\mu(f) = \int_G f^\sharp d\mu, \quad f \in \mathcal{K}(G).
\]

Since \( \mu \) is a positive measure, the functional \( \Phi^\sharp_\mu \) is positive, so by Radon–Riesz I (Theorem 5.8), there is a unique \( \sigma \)-Radon measure \( \mu^\sharp \) such that

\[
\int f d\mu^\sharp = \int f^\sharp d\mu, \quad f \in \mathcal{K}(G).
\]

Going back to an arbitrary complex measure \( \mu \) and expressing it as \( \mu = \mu_1^+ - \mu_1^- + i(\mu_2^+ - \mu_2^-) \), where the four measures on the right-hand side are positive, we obtain a complex measure \( \mu^\sharp \) such that

\[
\int f d\mu^\sharp = \int f^\sharp d\mu, \quad f \in \mathcal{K}(G).
\]

Consequently, we see that

\[
\lambda_t \mu^\sharp = \rho_t \mu^\sharp = \mu^\sharp, \quad \text{for all } t \in K.
\]

Conversely, the above equations imply that \( \mu = \mu^\sharp \). Thus the map \( \mu \mapsto \mu^\sharp \) is a projection of \( \mathbb{C}\mathcal{M}^1(G) \) onto the subspace \( \mathbb{C}\mathcal{M}^1(K\setminus G/K) \).

The following result is not hard to prove.

**Proposition 14.21.** If \( \mu \) is a measure of positive type on \( G \), then for every \( f \in \mathcal{K}(G) \), we have

\[
\int_G (f^* \ast f)^\sharp d\mu \geq 0.
\]
Proposition 14.21 is proven in Dieudonné [20] (Chapter XXII, Section 7, Lemma 22.7.4.3).

Using Proposition 14.21, we see that if \( \mu \) is of positive type, then so is \( \mu^\triangle \). Also, \( \mu^\sharp = 0 \) means that \( \mu \) vanishes on the subspace \( \mathcal{K}(K \backslash G/K) \) of \( \mathcal{K}(G) \). Thus by the uniqueness clause in Theorem 14.20, we have the following result.

**Proposition 14.22.** If \( \mu \) is a measure of positive type on \( G \), then \( (\mu^\#)^\triangle = \mu^\triangle \). For any two measures of positive type \( \mu \) and \( \nu \), we have \( \mu^\triangle = \nu^\triangle \) iff \( \mu^\# = \nu^\# \).

By Theorem 11.11(3), if \( p \) is a function of positive type, then \( \check{p} = p \), so \( \check{\omega} = \omega \) for all \( \omega \in Z(G/K) \), and for every \( f \in \mathcal{K}(K \backslash G/K) \), we have

\[
(\mathcal{F} f)(\omega) = \int_G f(x) \omega(x) \, d\lambda_G(x) = \int_G f(x) \overline{\omega(x^{-1})} \, d\lambda_G(x) = \int_G \overline{f(x)} \omega(x^{-1}) \, d\lambda_G(x) = (\overline{\mathcal{F} f})(\omega).
\]

Therefore,

\[
(\mathcal{F} f)(\omega) = \overline{(\mathcal{F} f)(\omega)}.
\]

**Example 14.8.** One of the main examples of Plancherel transform is the Dirac measure \( \mu = \delta_e \). As a corollary of Proposition 11.17, the Dirac measure \( \delta_e \) is of positive type.

**Definition 14.22.** The Plancherel transform \( \delta_e^\triangle \) of the Dirac measure \( \delta_e \) is called the canonical measure on \( Z(G/K) \) and is denoted \( m_Z \).

Because \( G \) is unimodular, we have

\[
\int_G g^* \ast f \, d\delta_e = (g^* \ast f)(e) = \int_G f(s^{-1})g(s^{-1}) \, d\lambda_G(s) = \int_G f(s)g(s) \, d\lambda_G(s),
\]

and for any two functions \( f, g \in \mathcal{K}(K \backslash G/K) \), by Theorem 14.20, we have

\[
\int_G f(s)g(s) \, d\lambda_G(s) = \int_G g^* \ast f \, d\delta_e = \int_{Z(G/K)} (\mathcal{F} f)(\omega)(\mathcal{F} g)(\omega) \, dm_Z(\omega).
\]

Write \( f = f_1 + if_2 \) and \( g = g_1 + ig_2 \), where \( f_1, f_2, g_1, g_2 \) are all real-valued. Then we have

\[
\int_{Z(G/K)} (\mathcal{F} f)(\omega)(\mathcal{F} g)(\omega) \, dm_Z(\omega) = \int_{Z(G/K)} (\mathcal{F} f_1)(\omega)(\mathcal{F} g_1)(\omega) \, dm_Z(\omega) \\
+ \int_{Z(G/K)} (\mathcal{F} f_2)(\omega)(\mathcal{F} g_2)(\omega) \, dm_Z(\omega) \\
- i \int_{Z(G/K)} (\mathcal{F} f_1)(\omega)(\mathcal{F} g_2)(\omega) \, dm_Z(\omega) \\
+ i \int_{Z(G/K)} (\mathcal{F} g_1)(\omega)(\mathcal{F} f_2)(\omega) \, dm_Z(\omega).
\]
Since \((\overline{\mathcal{F}} f)(\omega) = (\overline{\mathcal{F} f})(\omega)\), and since \(f_1, f_2, g_1, g_2\) are real-valued, we have

\[
\int_G f_1 g_1 d\lambda_G = \int_{Z(G/K)} (\overline{\mathcal{F}} f_1)(\omega)(\overline{\mathcal{F} g_1})(\omega) \, dm_Z(\omega) = \int_{Z(G/K)} (\mathcal{F} f_1)(\omega)(\overline{\mathcal{F} g_1})(\omega) \, dm_Z(\omega)
\]

\[
\int_G f_2 g_2 d\lambda_G = \int_{Z(G/K)} (\overline{\mathcal{F}} f_2)(\omega)(\overline{\mathcal{F} g_2})(\omega) \, dm_Z(\omega) = \int_{Z(G/K)} (\mathcal{F} f_2)(\omega)(\overline{\mathcal{F} g_2})(\omega) \, dm_Z(\omega)
\]

\[
\int_G f_1 g_2 d\lambda_G = \int_{Z(G/K)} (\overline{\mathcal{F}} f_1)(\omega)(\overline{\mathcal{F} g_2})(\omega) \, dm_Z(\omega) = \int_{Z(G/K)} (\mathcal{F} f_1)(\omega)(\overline{\mathcal{F} g_2})(\omega) \, dm_Z(\omega)
\]

\[
\int_G g_1 f_2 d\lambda_G = \int_{Z(G/K)} (\overline{\mathcal{F}} g_1)(\omega)(\overline{\mathcal{F} f_2})(\omega) \, dm_Z(\omega) = \int_{Z(G/K)} (\mathcal{F} g_1)(\omega)(\overline{\mathcal{F} f_2})(\omega) \, dm_Z(\omega),
\]

because the left integrals are real, and thus

\[
\int_G f(s)\overline{g(s)} \, d\lambda_G(s) = \int_{Z(G/K)} (\overline{\mathcal{F}} f)(\omega)(\overline{\mathcal{F} g})(\omega) \, dm_Z(\omega)
\]

\[
= \int_{Z(G/K)} (\mathcal{F} f)(\omega)(\overline{\mathcal{F} g})(\omega) \, dm_Z(\omega).
\]

Observe that the above integrals are inner products. As a consequence, we have the following result.

**Proposition 14.23.** For any two functions \(f, g \in \mathcal{K}(K\backslash G/K)\), we have

\[
\int_G f(s)\overline{g(s)} \, d\lambda_G(s) = \int_{Z(G/K)} (\overline{\mathcal{F}} f)(\omega)(\overline{\mathcal{F} g})(\omega) \, dm_Z(\omega)
\]

\[
= \int_{Z(G/K)} (\mathcal{F} f)(\omega)(\overline{\mathcal{F} g})(\omega) \, dm_Z(\omega).
\]

The linear maps \(f \mapsto \mathcal{F} f\) and \(f \mapsto \overline{\mathcal{F} f}\), with \(f, g \in \mathcal{K}(K\backslash G/K)\), are isometries, and by Theorem 14.20, these maps extend to isomorphisms from the Hilbert space \(L^2(K\backslash G/K)\) onto the Hilbert space \(L^2_{m_{Z}}(Z(G/K))\).

We can further extend these maps to linear maps of \(L^2(G)\) onto \(L^2_{m_{Z}}(Z(G/K))\), by setting \(\mathcal{F}([f]) = \mathcal{F}([f])\) and \(\overline{\mathcal{F}}([f]) = \overline{\mathcal{F}}([f])\). The equation \(\overline{\mathcal{F}}([f]) = \overline{\mathcal{F}([f])}\) holds. By abuse of notation, we write \(\mathcal{F} f\) (resp. \(\overline{\mathcal{F}} f\)) for any function in the class \(\mathcal{F}([f])\) (resp, \(\overline{\mathcal{F}}([f])\)). With these notation, \((*)_1\) and \((*)_2\) hold for \(f, g \in L^2(K\backslash G/K)\).

Proposition 14.23 is a generalization of the Plancherel theorem (Theorem 9.26), as we will see in Example 14.9.

**Example 14.9.** Another important example is the case where \(G\) is commutative and \(K = \{e\}\). In this case, the functional equation \((s_1)\) characterizing spherical functions reduces to

\[
\omega(xy) = \omega(x)\omega(y),
\]
so \( \omega: G \rightarrow \mathbb{C}^* \) is a continuous homomorphism such that \( \omega(e) = 1 \) and \( |\omega(x)| \leq 1 \) for all \( x \in G \). Since \( \omega(x^{-1}) = \omega(x)^{-1} \), we conclude that \( |\omega(x)| = 1 \) for all \( x \in G \), which means that \( \omega \) is a group character. The function \( \omega \) is of positive type. This is because \( \omega(y^{-1}x) = \omega(y)\omega(x) \), so for every function \( f \in \mathcal{K}(G) \), have

\[
\int_G (f^* \ast f) \omega \, d\lambda_G = \int_G \int_G f^*(y) f(y^{-1}x) \omega(x) \, d\lambda_G(y) \, d\lambda_G(x) \\
= \int_G \int_G \overline{f(y^{-1})} f(y^{-1}x) \omega(x) \, d\lambda_G(y) \, d\lambda_G(x) \\
= \int_G \int_G \overline{f(y)} f(x) \omega(y^{-1}x) \, d\lambda_G(y) \, d\lambda_G(x) \\
= \int_G \int_G \overline{f(y)} \omega(y) f(x) \omega(x) \, d\lambda_G(y) \, d\lambda_G(x) \\
= \left| \int_G f(x) \omega(x) \, d\lambda_G(x) \right|^2,
\]

so

\[
\int_G (f^* \ast f) \omega \, d\lambda_G = \left| \int_G f(x) \omega(x) \, d\lambda_G(x) \right|^2 \geq 0.
\]

Consequently, \( S(G/\{e\}) = Z(G/\{e\}) \) is the space of group characters of \( G \), and this space is homeomorphic to \( X_0(A) \), where \( A = L^1(G) \oplus \mathbb{C} \delta_e \).

The topological space \( \widehat{G} = S(G/\{e\}) \) of group characters is a group, and it can be shown that the topology of \( S(G/\{e\}) \) is compatible with the group structure; see Dieudonné [20] (Chapter XXII, Section 10, Lemma 22.10.2). Therefore, \( \widehat{G} \) is a commutative topological group which is locally compact, metrizable, and separable. Given a (left) Haar measure \( \lambda_G \) on \( G \), it can be shown that the Plancherel transform \( \lambda_{\widehat{G}} = m_Z = \delta_e^\Delta \) is a (left) Haar measure on \( \widehat{G} \); see Dieudonné [20] (Chapter XXII, Section 10, Lemma 22.10.5). The Haar measure \( \lambda_{\widehat{G}} \) on \( \widehat{G} \) and the Haar measure \( \lambda_G \) on \( G \) are said to be associated. If \( \lambda_G \) is replaced by \( a \lambda_G \) with \( a > 0 \), then \( \lambda_{\widehat{G}} \) is replaced by \( a^{-1} \lambda_{\widehat{G}} \).

Proposition 14.23 shows that the equation

\[
\int_G f(s) g(s) \, d\lambda_G(s) = \int_{\widehat{G}} (\mathcal{F}f)(\omega)(\mathcal{F}g)(\omega) \, d\lambda_{\widehat{G}}(\omega)
\]

holds, and that \( \mathcal{F} \) has an extension which is an isometry from \( L^2(G) \) to \( L^2(\widehat{G}) \), providing another proof of the Plancherel theorem, Theorem 9.26.

We now return to an arbitrary locally compact group \( G \) (metrizable, separable, and unimodular). For any function \( \omega \in Z(G/K) \) of positive type, the measure \( \omega \, d\lambda_G \) is a measure of positive type. It can be shown that

\[
(\omega \, d\lambda_G)^\Delta = \delta_\omega,
\]
14.6. THE PLANCHEREL TRANSFORM

the Dirac measure at \( \omega \); Dieudonné [20] (Chapter XXII, Section 7, Lemma 22.7.6.1). The result is a sort of Fourier inversion formula.

**Proposition 14.24.** If \( \mu \) is a measure of positive type on \( G \), then for every function \( f \in \mathcal{K}(K \backslash G/K) \), if the Fourier cotransform \( \mathcal{F} f \) belongs to \( L^1_{\mu^\Delta}(Z(G/K)) \), then

\[
\int_G f \, d\mu = \int_{Z(G/K)} (\mathcal{F} f)(\omega) \, d\mu^\Delta(\omega).
\]

Proposition 14.24 is proven in Dieudonné [20] (Chapter XXII, Section 7, Lemma 22.7.8).

In general, given a function \( f \in \mathcal{K}(K \backslash G/K) \), by Theorem 14.20, \( \mathcal{F} f \notin L^1_{\mu^\Delta}(Z(G/K)) \). However, if \( g \) is another function in \( \mathcal{K}(K \backslash G/K) \), then \( \mathcal{F}(f \ast g) = (\mathcal{F} f)(\mathcal{F} g) \in L^1_{\mu^\Delta}(Z(G/K)) \). Thus we can use the technique of regularization. If \( (g_n) \) is a sequence of positive functions in \( \mathcal{K}(K \backslash G/K) \) having a compact support that tends to \( \{e\} \), and such that \( \int g_n \, d\lambda_G = 1 \), then \( \int (f \ast g_n) \, d\mu \) tends to \( \int f \, d\mu \), and we have

\[
\int f \, d\mu = \lim_{n \to \infty} \int (f \ast g_n) \, d\mu = \lim_{n \to \infty} \int_{Z(G/K)} (\mathcal{F} f)(\omega)(\mathcal{F} g_n)(\omega) \, d\mu^\Delta(\omega).
\]

In particular, if we apply the above formula to \( \mu = \delta_e \), if we compute \( \int_G (\lambda_s f) \, d\delta_e = f(s^{-1}) \), using Proposition 14.19, we find that for every \( f \in \mathcal{K}(K \backslash G/K) \), we have

\[
f(s) = \lim_{n \to \infty} \int_{Z(G/K)} \omega(s^{-1})(\mathcal{F} f)(\omega)(\mathcal{F} g_n)(\omega) \, d\mu^\Delta(\omega).
\]

Summation processes for the inversion of the Fourier cotransform are usually used when \( G \) is abelian and \( K = \{e\} \).

We now take a closer look at the space \( \mathcal{P}_+(K \backslash G/K) \) of functions in \( \mathcal{C}(K \backslash G/K) \) which are of positive type.

**Proposition 14.25.** The map \( p \mapsto (p \lambda_G)^\Delta \) is a bijection between the space \( \mathcal{P}_+(K \backslash G/K) \) of functions in \( \mathcal{C}(K \backslash G/K) \) which are of positive type onto the space \( \mathcal{M}^1_+(Z(G/K)) \) of bounded positive measures on \( Z(G/K) \). The inverse \( \mathcal{F}' \) of the above map \( (p \mapsto (p \lambda_G)^\Delta) \) is given by

\[
(\mathcal{F}' \mu)(x) = \int_{Z(G/K)} \omega(x) \, d\mu(\omega), \quad \mu \in \mathcal{M}^1_+(Z(G/K)).
\]

For every \( f \in L^1(G) \), we have

\[
\int_G f(x)(\mathcal{F}' \mu)(x) \, d\lambda_G(x) = \int_{Z(G/K)} (\mathcal{F} f)(\omega) \, d\mu(\omega).
\]
Proposition 14.25 is proven in Dieudonné [20] (Chapter XXII, Section 7, Lemma 22.7.10). The proof uses the Bochner–Godement theorem (Theorem 10.38).

In the special case where \( G \) is commutative and \( K = \{e\} \), the space \( \mathcal{P}_+(G) \) is the set of all continuous functions of positive type on \( G \), and Proposition 14.25 implies that every function \( p \) of positive type can be written uniquely as

\[
p(x) = \int_{\hat{G}} \omega(x) \, d\mu(\omega)
\]

for some positive measure \( \mu \) on \( \hat{G} \), a result known as Bochner’s theorem; see also Folland [31] Chapter 4, Theorem 4.18.

We conclude with three remarks.

1. If the Haar measure \( \lambda_G \) on \( G \) is replaced by \( a\lambda_G \) with \( a > 0 \), then the space \( \mathcal{S}(G/K) \) of spherical functions is unchanged. The Fourier transform and the Fourier cotransform are multiplied by \( a \), and the Plancherel transform (see Definition 14.21) is multiplied by \( a^{-1} \).

2. If \((G,K)\) is a Gelfand pair and if \( G \) is compact, then by Example 14.5, Proposition 13.17, and Proposition 13.21, all spherical functions are of positive type. However, if \( G \) is not compact, this is generally false. For instance, the functions \( P_\rho \) of Example 14.6 do not satisfy the property \( \overline{\omega} = \omega \) unless \( \Re \rho = -1/2 \). It can be shown that these functions are of positive type if \( \Re \rho = -1/2 \).

3. If \( G \) is compact, then we saw that the space \( \mathcal{S}(G/K) \) is discrete and in bijection with the subset of \( R(G) \) (of irreducible representations of \( G \)) consisting of those \( \rho \in R(G) \) such that \((\rho : \sigma_0) = 1\). We can view the Fourier transform \( \mathcal{F}f \) of a function \( f \in L^1(G) \) as the family

\[
\left( c_\rho = \frac{1}{n_\rho} \langle f, m_1^{(\rho)} \rangle \right)_{\rho \in \mathcal{S}(G/K)}.
\]

The Fourier transform and the Plancherel measures are discussed from a different point of view for symmetric spaces in Helgason [41] (Chapter 4) and Helgason [40] (Chapter 3, especially Section 12).

### 14.7 Extension of the Plancherel Transform; \( P(G) \) and \( P'(Z) \)

The purpose of this section is to extend the Plancherel transform to a bigger set of measures and to define the notion of Fourier transform of a measure. Let \( X \) be a locally compact space. Recall that the space of \( \sigma \)-Radon measures on \( X \) is denoted by \( \mathcal{M}_\sigma^+(X) \) and the
space of Radon measures on $X$ is denoted $\mathcal{M}^+(X)$. The space of complex measures on $X$ is denoted by $\mathbb{C}\mathcal{M}^1(X)$, and called the space of \textit{bounded measures}. The space of regular complex Borel measures is denoted by $\mathcal{M}^1(X)$. The space $\mathbb{C}\mathcal{M}^1(X)$ contains the space of positive bounded Borel measures, and of course, $\mathcal{M}^1(X) \subseteq \mathbb{C}\mathcal{M}^1(X)$.

Let $G$ be a locally compact, metrizable, separable, and unimodular group.

\textbf{Definition 14.23.} The complex vector space spanned by the union of the complex measures and the $\sigma$-Radon measures is denoted by $\mathcal{M}_C(G)$. Let $\mathcal{P}_+(G)$ be the set of measures of positive type, and let $\mathcal{P}(G)$ be the complex subspace of $\mathcal{M}_C(G)$ spanned by $\mathcal{P}_+(G)$, which consists of all combinations $\mu_1 - \mu_2 + i\mu_3 - i\mu_4$, where the $\mu_i$ belong to $\mathcal{P}_+(G)$.

As a general rule, the subscript $+$ indicates that we are dealing with functions or measures of positive type, and the suppression of the subscript $+$ that we are considering the vector space spanned by that set.

It is easy to check that if $\mu \in \mathcal{P}(G)$, then $\mu^\sharp \in \mathcal{P}(G)$. The image of $\mathcal{P}(G)$ by the map $\mu \mapsto \mu^\sharp$ is $\mathcal{P}(G) \cap \mathcal{M}_C(K\backslash G/K)$.

Let $(G, K)$ be a Gelfand pair. The Plancherel transform $\mu \mapsto \mu^\triangle$ is a map from $\mathcal{P}_+(G)$ to $\mathcal{M}^+(Z(G/K))$, the space of positive measures on $Z(G/K)$. We have $(\mu + \nu)^\triangle = \mu^\triangle + \nu^\triangle$ and $(c\mu)^\triangle = c\mu^\triangle$, for any $c > 0$. From this, it is easy to show that for any combination $\mu = \mu_1 - \mu_2 + i\mu_3 - i\mu_4$ of measures $\mu_i \in \mathcal{P}_+(G)$, the sum $\mu^\triangle_1 - \mu^\triangle_2 + i\mu^\triangle_3 - i\mu^\triangle_4$ is a measure on $Z(G/K)$ that depends only on $\mu$ and not on its decomposition.

Consequently, we obtain a $\mathbb{C}$-linear map $\mu \mapsto \mu^\triangle$ from $\mathcal{P}(G)$ to $\mathcal{M}_C(Z(G/K))$ that we also call \textit{Plancherel transform}.

Since $\mathcal{F}g = \mathcal{F}\hat{g}$ and $\mathcal{F}g = \mathcal{F}\tilde{g}$, we have $\overline{\mathcal{F}g} = \mathcal{F}\overline{g}$ and $\mathcal{F}\tilde{g} = \mathcal{F}g$, and so

$$\overline{\mathcal{F}g^*} = \overline{\mathcal{F}\overline{g}} = \mathcal{F}\tilde{g} = \mathcal{F}g.$$ 

From Theorem 14.20 with $g$ replaced by $g^*$, we deduce that

$$\int_G (g*f) \ d\mu = \int_{Z(G/K)} (\mathcal{F}f)(\omega)(\mathcal{F}g)(\omega) \ d\mu^\triangle(\omega) = \int_G (f*g) \ d\mu, \quad \text{for all } f, g \in \mathcal{K}(K\backslash G/K).$$ 

By regularization, it can be shown that for any two measures $\mu, \nu \in \mathcal{P}(G) \cap \mathcal{M}_C(K\backslash G/K)$, if $\mu^\triangle = \nu^\triangle$, then $\mu = \nu$. But $\mu = \mu^\sharp$ and $\nu = \nu^\sharp$, so it follows that the kernel of the Plancherel transform is the subspace of measures $\mu \in \mathcal{P}(G)$ such that $\mu^\sharp = 0$.

\textbf{Definition 14.24.} The image of $\mathcal{P}(G)$ (or $\mathcal{P}(G) \cap \mathcal{M}_C(K\backslash G/K)$) by the Plancherel transform is denoted by $\mathcal{P}'(Z)$. Let $\mathcal{P}(G)$ be the complex vector space spanned by the functions of positive type on $G$.

The space $\mathcal{P}(G)$ is a subspace of both $L^\infty(G)$ and $\mathcal{P}(G)$, by viewing $f \in \mathcal{P}(G)$ as $f \, d\lambda_G$. By Proposition 14.25, the image of $\mathcal{P}(G)$ by the Plancherel transform is $\mathbb{C}\mathcal{M}^1(Z(G/K))$, the space of bounded measures on $Z(G/K)$. Consequently we obtain the following result.
Chapter 14. Harmonic Analysis on Gelfand Pairs

Proposition 14.26. Let $\mathcal{P}(K\backslash G/K)$ be the subspace of $\mathcal{P}(G)$ consisting of the functions invariant by left and right translations by elements of $K$. Then the map $f \mapsto (f\lambda_G)^\Delta$ is a linear bijection between $\mathcal{P}(K\backslash G/K)$ and $\mathbb{C}\mathcal{M}^1(Z(G/K))$. The inverse map is denoted by $\mathcal{F}'$ and is given by the formula

$$(\mathcal{F}' \mu')(x) = \int_{Z(G/K)} \omega(x) d\mu'(\omega), \quad \mu' \in \mathcal{M}^1(Z(G/K)).$$

We also have

$$\|\mathcal{F}' \mu'\| \leq \|\mu'\|,$$

since $|\omega(x)| \leq 1$ for all $x \in G$ and all $\omega \in Z(G/K)$.

Thus the linear map $\mathcal{F}'$ from $\mathbb{C}\mathcal{M}^1(Z(G/K))$ to $\mathcal{P}(K\backslash G/K)$ is continuous for the topology of the Banach space $\mathcal{C}_b(G)$. However, in general, $\mathcal{P}(K\backslash G/K)$ is not closed in $\mathcal{C}_b(G)$.

Unfortunately, no convenient characterizations of the spaces $\mathcal{P}(G)$ and $\mathcal{P}'(Z)$ are known, besides their definition. There are necessary or sufficient conditions, but no necessary and sufficient conditions known. For example, a necessary condition for a measure $\mu'$ on $Z(G/K)$ to belong to $\mathcal{P}'(Z)$ is that the Fourier cotransforms $\mathcal{F}f$ of functions $f \in K(K\backslash G/K)$ belong to $L^2_{\mu'}(Z(G/K))$, but there are counter-examples showing that this condition is not sufficient.

Also, we showed that $\mathbb{C}\mathcal{M}^1(Z(G/K)) \subseteq \mathcal{P}'(Z)$, but there are unbounded measures in $\mathcal{P}'(Z)$ (for example, the Haar measure on $\hat{G}$, when $G$ is a commutative noncompact locally compact group).

The following result holds.

Proposition 14.27. For every measure $\mu' \in \mathcal{P}'(Z)$, and every function $g' \in L^2_{\mu'}(Z(G/K))$, we have $g' d\mu' \in \mathcal{P}'(Z)$. If $\nu \in \mathcal{P}(G)$ is a measure such that $\nu^\Delta = g' d\mu'$, then for every $f \in K(K\backslash G/K)$ we have

$$\int_G f d\nu = \int_{Z(G/K)} (\mathcal{F}f)(\omega) g'(\omega) d\mu'(\omega).$$

Proposition 14.27 is proven in Dieudonné [20] (Chapter XXII, Section 8, Lemma 22.8.3).

In particular, if we take $\mu'$ to be the canonical measure $m_Z$, then we obtain the following corollary.

Proposition 14.28. The space $L^2_{m_Z}(Z(G/K))$ (viewed as a subspace of $\mathcal{M}_C(Z(G/K))$) is contained in $\mathcal{P}'(Z)$. The restriction to $L^2(K\backslash G/K) \subseteq \mathcal{P}(G)$ of the Plancherel transform is identical to the extension of the Fourier transform $\mathcal{F}$ from $L^2(K\backslash G/K)$ to $L^2_{m_Z}(Z(G/K))$ given by Proposition 14.23. In particular, for any $f \in L^2(K\backslash G/K)$, we have $(f d\lambda_G)^\Delta = (\mathcal{F}f) dm_Z$. 
**14.7. EXTENSION OF THE PLANCHEREL TRANSFORM; \( P(G) \) AND \( P'(Z) \)**

Proposition 14.28 is proven in Dieudonné [20] (Chapter XXII, Section 8, Lemma 22.8.4).

**Remark:** It is possible that there is some positive measure \( \mu' \in P'(Z) \), yet there are positive measures \( \nu' \) with \( 0 \leq \nu' \leq \mu' \), and \( \nu' \notin P'(Z) \).

Here are more results about the space \( C\mathcal{M}^1(G) \) of bounded measures on \( G \).

**Proposition 14.29.** The space \( C\mathcal{M}^1(G) \) of bounded measures on \( G \) is contained in \( P(G) \).

For every \( \mu \in C\mathcal{M}^1(G) \), we have

\[
\mu^\triangle = (F\mu) \, dm_Z,
\]

where \( F\mu \) is a continuous bounded function on \( S(G/K) \) given by

\[
(F\mu)(\omega) = \int_G \omega(x^{-1}) \, d\mu(x).
\]

Furthermore, for every function \( f \in L^1(K\backslash G/K) \) (resp. \( f \in L^2(K\backslash G/K) \)), we have \( \mu \ast f, f \ast \mu \in L^1(G) \) (resp. \( L^2(G) \)), and

\[
(F(\mu \ast f)) = (F\mu)(Ff) = (Ff)(F\mu)
\]

almost everywhere w.r.t. \( m_Z \).

Proposition 14.29 is proven in Dieudonné [20] (Chapter XXII, Section 8, Lemma 22.8.4.3). The following definition generalizes Definition 9.4.

**Definition 14.25.** If \( \mu \in C\mathcal{M}^1(G) \) is a bounded measure, then the function \( F\mu \) (defined on \( S(G/K) \)) given by

\[
(F\mu)(\omega) = \int_G \omega(x^{-1}) \, d\mu(x)
\]

is called the *Fourier transform* of \( \mu \). We define the *Fourier cotransform* \( \tilde{F}\mu \) of \( \mu \) as \( \tilde{F}\mu \); that is,

\[
(F\mu)(\omega) = \int_G \omega(x^{-1}) \, d\mu(x) = \int_G \omega(x) \, d\tilde{\mu}(x) = \int_G \omega(x) \, d\mu(x).
\]

For every function \( f \in L^1(G) \), we have \( Ff = F(f \, d\lambda_G) \), which justifies the terminology. We have

\[
(F\delta_e)(\omega) = \omega(x^{-1}), \quad (\tilde{F}\delta_e)(\omega) = \omega(x).
\]

It is clear that \( \|F\mu\| \leq \|\mu\| \), so \( F \) is a continuous linear map from the Banach space \( C\mathcal{M}^1(G) \) to the Banach space \( C_b(S(G/K)) \) of continuous bounded functions on \( S(G/K) \). However, in general, the bounded function \( F\mu \) does not tend to zero at infinity, as shown by \( \mu = \delta_e \), for which \( F\delta_e \) is the constant 1.

As a corollary of Proposition 14.28 and Proposition 14.29, we have

\[
L^1(G) \cap L^2(G) = C\mathcal{M}^1(G) \cap L^2(G),
\]

and the class of the Fourier transform \( Ff \) of a function \( f \in L^1(G) \cap L^2(G) \) is identical to the class \( F[f] \) as in Definition 14.19.
Proposition 14.30. For any two functions \( f, g \in L^2(K \backslash G/K) \), the bounded continuous function \( f * g \) belongs to \( \mathcal{P}(K \backslash G/K) \) and we have
\[
((f * g) d\lambda_G) = (\mathcal{F}f)(\mathcal{F}g) \, dm_Z.
\]

Proposition 14.30 is proven in Dieudonné [20] (Chapter XXII, Section 8, Lemma 22.8.8). Note that \( \mathcal{F}(f * g) \) is integrable for the measure \( m_Z \), but in general, \( f * g \) is not integrable for the measure \( m_Z \).

Definition 14.26. Define the spaces \( \mathcal{P}^1(K \backslash G/K) \) and \( \mathcal{P}^2(K \backslash G/K) \) by
\[
\mathcal{P}^1(K \backslash G/K) = \mathcal{P}(K \backslash G/K) \cap L^1(K \backslash G/K),
\quad \mathcal{P}^2(K \backslash G/K) = \mathcal{P}(K \backslash G/K) \cap L^2(K \backslash G/K).
\]

Since for any function \( f \in L^\infty(G) \cap L^1(G) \) we have \( |f(x)|^2 \leq \|f\|_\infty |f(x)| \) almost everywhere, which implies that \( f \in L^2(G) \). Since the functions in \( \mathcal{P}(G) \) are bounded, we have the inclusions
\[
\mathcal{P}^1(K \backslash G/K) \subseteq \mathcal{P}^2(K \backslash G/K) \subseteq \mathcal{P}(K \backslash G/K).
\]

Proposition 14.31. The image of \( \mathcal{P}^2(K \backslash G/K) \) (as a subspace of \( \mathbb{C}\mathcal{M}^1(K \backslash G/K) \)) under the Plancherel transform is the subspace \( L^1_{m_Z}(Z(G/K)) \cap L^2_{m_Z}(Z(G/K)) \) of \( \mathbb{C}\mathcal{M}^1(Z(G/K)) \). For every function \( f \in \mathcal{P}^2(K \backslash G/K) \), we have the Fourier inversion formula
\[
f = \mathcal{F}'((\mathcal{F}f) \, dm_Z),
\]
where \( \mathcal{F}' \) is defined in Proposition 14.26. Moreover, if we also have \( f \in \mathcal{P}^1(K \backslash G/K) \), then we have the Fourier inversion formula
\[
f(x) = \int_{Z(G/K)} \omega(x) \left( \int_{G} f(y) \omega(y^{-1}) \, d\lambda_G(y) \right) \, dm_Z(\omega).
\]

Proposition 14.31 is proven in Dieudonné [20] (Chapter XXII, Section 8, Lemma 22.8.10).

One should be cautious that in general, \( f \) is not integrable, so we can’t use the formula of Definition 14.19 to define \( \mathcal{F}f \). If \( f \in \mathcal{P}^1(K \backslash G/K) \), then we have the formula above, but the two integrals cannot be replaced by the double integral
\[
\int \int_{G \times Z(G/K)} \omega(x) \omega(y^{-1}) f(y) \, d\lambda_G(y) \, dm_Z(\omega),
\]
because this integral is not defined in general.

Remark: The spaces \( \mathcal{P}(K \backslash G/K), \mathcal{P}^1(K \backslash G/K), \mathcal{P}^2(K \backslash G/K) \) are generally not closed in \( C_b(G) \), but \( \mathcal{P}^1(K \backslash G/K) \) is dense in \( L^1(K \backslash G/K) \), and \( \mathcal{P}^2(K \backslash G/K) \) is dense in \( L^2(K \backslash G/K) \).
In the special case where $G$ is commutative and $K = \{e\}$, we know that $S(G/\{e\}) = \hat{G}$, and by Pontrjagin duality (Theorem 9.29), $\hat{G}$ and $G$ can be identified, and then the transform $\mathcal{F}'$ from $\mathcal{C}_M^1(\hat{G})$ to $\mathcal{P}(G)$ defined in Proposition 14.26 is identified with the Fourier cotransform $\mathcal{F}$ on $\mathcal{M}_1^1(\hat{G})$. The Haar measure $\hat{m}^\wedge$ is also identified with the Haar measure $\lambda^_G$. Then Proposition 14.31 yields the Fourier inversion formula

$$f = \mathcal{F}(\mathcal{F}f),$$

for every function $f \in \mathcal{P}^2(G)$, and since $\mathcal{P}^2(G)$ is dense in $L^2(G)$, the above formula actually holds for all $f \in L^2(G)$. This gives another proof of the inversion formula of the Pontrjagin duality theorem, Theorem 9.29.

By Proposition 14.28 and Proposition 14.29, for any $f \in L^2(G)$ and any $\mu \in \mathcal{C}_M^1(G)$, we have

$$(f d\lambda^G)^\wedge = (\mathcal{F}f) d\lambda^\hat{G}, \quad \mu^\wedge = (\mathcal{F}\mu) d\lambda^\hat{G}.$$

### 14.8 Spherical Functions of Positive Type and Irreducible Representations

Let $(G, K)$ be a Gelfand pair (with $G$ a locally compact, metrizable, separable, unimodular group). From Theorem 11.11, every spherical function of positive type $\omega \in \mathcal{Z}(G/K)$ induces a cyclic unitary representation $U_\omega : G \to \mathbb{U}(H_\omega)$ of $G$ in a separable Hilbert space $H_\omega$. Recall that the map $\varphi_\omega$ given by

$$\varphi_\omega(\mu) = \int_G \omega(s) d\mu, \quad \mu \in \mathcal{M}_1^1(G)$$

is a positive linear form, and so is its restriction to the unital involutive subalgebra $A = L^1(G) \oplus \mathbb{C}\delta_e$. If

$$n = \{\mu \in A \mid \varphi_\omega(\mu^* \ast \mu) = 0\},$$

then $n$ is a left ideal in $A$, and $H_0 = A/n$ is a hermitian space with the inner product

$$\langle \pi(\mu), \pi(\nu) \rangle = \varphi_\omega(\nu^* \ast \mu) = \int_G \omega(s) d(\nu^* \ast \mu)(s),$$

where $\pi : A \to A/n$ is the quotient map (and $\mu^* = \overline{\mu}$). If $H_\omega$ is the separable Hilbert space which is the completion of $H_0 = A/n$, then the unitary representation $U_\omega : G \to \mathbb{U}(H_\omega)$ is completely determined by

$$U_\omega(s)(\pi(\mu)) = \pi(\delta_s \ast \mu), \quad \mu \in A, s \in G.$$

The unique unitary non-degenerate (algebra) representation $(U_\omega)_{\text{ext}} : A \to \mathcal{L}(H_\omega)$ extending $U_\omega$ is completely determined by

$$((U_\omega)_{\text{ext}}(\mu))(\pi(\nu)) = \pi(\mu \ast \nu), \quad \mu, \nu \in A.$$

The vector $x_0 = \pi(\delta_e)$ is a cyclic vector for both representations.
Theorem 14.32. Let \((G, K)\) be a Gelfand pair.

(1) For every spherical function of positive type \(\omega \in \mathbb{Z}(G/K)\), the cyclic unitary representation \(U_\omega: G \to \mathbf{U}(H_\omega)\) (with cyclic vector \(x_0\)) is irreducible, and its restriction to the compact group \(K\) contains the trivial representation of \(K\), which means that

\[ F = \{ x \in H_\omega \mid U_\omega(t)(x) = x \text{ for all } t \in K \} \neq \{0\}. \]

In fact, the cyclic vector \(x_0\) belongs to \(F\).

(2) Conversely, every unitary representation \(U: G \to \mathbf{U}(H)\) whose restriction to \(K\) contains the trivial representation of \(K\) is equivalent to one of the representations \(U_\omega\) with \(\omega \in \mathbb{Z}(G/K)\), and the multiplicity of the trivial representation of \(K\) in \(U\) is 1.

Theorem 14.32 is proven in Dieudonné [20] (Chapter XXII, Section 9, Lemma 22.9.2). To prove that \(U_\omega\) is irreducible, it suffices to show that \(P = \text{ext}(\chi_K \lambda_G)\) is the orthogonal projection of \(H_\omega\) onto the one-dimensional subspace \(\mathbb{C}x_0\). Indeed, if \(F\) is a closed subspace of \(H_\omega\) invariant under \(U_\omega\), and if \(F\) is not orthogonal to \(x_0\), then \(P(F) \subseteq F\) and \(x_0 \in F\), so \(F = H_\omega\) since \(x_0\) is a cyclic vector. On the other hand, if \(F\) is orthogonal to \(x_0\), then its orthogonal complement \(F^\perp\) is also invariant under \(U_\omega\) and contains \(x_0\), and since \(x_0\) is a cyclic vector \(F^\perp = H_\omega\), and thus \(F = \{0\}\).

The proof of Theorem 14.32 makes use of the following proposition.

Proposition 14.33. For every irreducible representation \(V: A \to \mathcal{L}(H)\) of an involutive commutative algebra \(A\) in a separable Hilbert space \(H\), if \(V(A)\) is separable, then \(V\) is a representation in a one-dimensional subspace (of \(H\)).

Proposition 14.33 is proven in Dieudonné [20] (Chapter XXII, Section 9, Lemma 22.9.2.2).

Unlike the case where \(G\) is compact, there may not be any closed subspace \(F\) of \(L^2(G/K)\), invariant under the canonical representation (see Definition 13.10), and such that the subrepresentation of the canonical representation to \(F\) is equivalent to some representation of the form \(U_\omega\). However, we have the following results.

Proposition 14.34. Given a linear map \(f: E \to E\), if \(f\) has rank 1, which means that \(\dim(f(E)) = 1\), then there is a linear form \(\varphi \in E^*\) and some nonzero vector \(u \in E\) such that

\[ f(x) = \varphi(x)u, \quad \text{for all } x \in E. \]

Proof. This fact is immediately obtained by picking a basis \((e_\alpha)_{\alpha \in I}\) in \(E\). Since \(f\) has rank 1, we can pick a nonzero vector \(u \in f(E)\), and then \(f(e_\alpha) = \lambda_\alpha u\) for some \(\lambda_\alpha \in \mathbb{C}\), so we can let \(\varphi\) be the linear form given by \(\varphi(e_\alpha) = \lambda_\alpha\). If \(u\) is replaced by \(cu\) with \(c \neq 0\), then \(\varphi\) is replaced by \(c^{-1}\varphi\). \(\Box\)
We define the trace of \( f \) as
\[
\text{tr}(f) = \varphi(u),
\]
which is independent of the choice of \( u \). If \( g : E \to E \) is any other linear map, then it is easy to see that \( f \circ g \) and \( g \circ f \) have rank 1, and that \( \text{tr}(f \circ g) = \text{tr}(g \circ f) \).

**Proposition 14.35.** The following properties hold.

1. For every function \( f \in \mathcal{L}^1(G/K) \), the linear map \( (U_\omega)_{\text{ext}}(f) \in \mathcal{L}(H_\omega) \) has rank 1, and we have
\[
((U_\omega)_{\text{ext}}(f))(z) = \langle z, x_0 \rangle ((U_\omega)_{\text{ext}}(f))(x_0), \quad z \in H_\omega,
\]
where \( x_0 \) is the cyclic vector \( x_0 = \pi(\delta_e) \). The trace of \( (U_\omega)_{\text{ext}}(f) \) is given by
\[
\text{tr}((U_\omega)_{\text{ext}}(f)) = \mathcal{F}(\omega).
\]

2. For any two functions \( f, g \in \mathcal{L}^1(G/K) \cap \mathcal{L}^2(G/K) \), we have
\[
\text{tr}((U_\omega)_{\text{ext}}(f)(U_\omega)_{\text{ext}}(g)^*) = \mathcal{F}(g^* f)(\omega),
\]
an integrable function for the canonical measure \( m_Z \) on \( Z(G/K) \), and
\[
\int_G f(s) \overline{g(s)} d\lambda_G(s) = \int_{Z(G/K)} \text{tr}((U_\omega)_{\text{ext}}(f)(U_\omega)_{\text{ext}}(g)^*) dm_Z(\omega).
\]

3. For every continuous and bounded function \( f \in \mathcal{L}^1(G/K) \), such that for all \( s \in G \), the function \( \mathcal{F}(\delta_s * f) \) is integrable for \( m_Z \), we have
\[
f(s) = \int_{Z(G/K)} \mathcal{F}(\delta_s * f)(\omega) dm_Z(\omega) = \int_{Z(G/K)} \text{tr}(U_\omega(s)(U_\omega)_{\text{ext}}(f)) dm_Z(\omega).
\]

Proposition 14.35 is proven in Dieudonné [20] (Chapter XXII, Section 9, Lemma 22.9.4).

**Remark:** Using Proposition 14.18 it can be shown that if \( g, h \in \mathcal{K}(G/K) \), then \( f = g * h \) satisfies the hypothesis of Proposition 14.35(3), namely, \( \mathcal{F}(\delta_s * f) \) is integrable for \( m_Z \). Indeed,
\[
\mathcal{F}(\delta_s * f) = \mathcal{F}(\delta_s * (g * h)) = (\mathcal{F}(\delta_s * g))(\mathcal{F}h)
\]
with both factors in \( \mathcal{L}^2_{m_Z}(Z(G/K)) \), so \( \mathcal{F}(\delta_s * f) \) is \( m_Z \)-integrable. We also have \( f = g * h \in \mathcal{K}(G) \).

In the special case where \( G \) is a commutative locally compact metrizable and separable group, we have the following results about unitary representations.
Proposition 14.36. Let $G$ be a commutative locally compact metrizable and separable group. Every unitary cyclic representation $U: G \to \mathbb{U}(H)$ of $G$ in a separable Hilbert space $H$ is equivalent to a representation $M: G \to \mathbb{U}(L^2_\mu(\hat{G}))$, where $\mu$ is a positive bounded measure on $\hat{G}$ (the dual of $G$), and for every $s \in G$, the linear operator $M(s)$ is defined so that for every $g \in L^2_\mu(\hat{G})$, $M(s)(g)$ is the class of the function in $L^2_\mu(\hat{G})$ given by

$$
\chi \mapsto \chi(s)g(\chi), \quad \chi \in \hat{G}.
$$

The proof of Proposition 14.36 is proven in Dieudonné [20] (Chapter XXII, Section 9, Lemma 22.15.1). The proof uses Bochner’s theorem (see Proposition 14.25).

If $G = \mathbb{R}$, there is a more precise result due to Stone.

Theorem 14.37. (Stone) Every unitary representation of the (additive) group $\mathbb{R}$ in a separable Hilbert space $H$ is of the form

$$
t \mapsto e^{itA},
$$

where $A$ is a self-adjoint operator of $H$, not necessarily bounded. Conversely, for every self-adjoint not necessarily bounded operator $A$ of $H$, the map $t \mapsto e^{itA}$ is a unitary representation of $\mathbb{R}$ in $H$.

Theorem 14.37 is proven in Dieudonné [20] (Chapter XXII, Section 9, Lemma 22.15.3).
Chapter 15

Topology

15.1 Metric Spaces and Normed Vector Spaces

This chapter contains a review of basic topological concepts. First metric spaces are defined. Next normed vector spaces are defined. Closed and open sets are defined, and their basic properties are stated. The general concept of a topological space is defined. The closure and the interior of a subset are defined. The subspace topology and the product topology are defined. Continuous maps and homeomorphisms are defined. Limits of sequences are defined. Continuous linear maps and multilinear maps are defined and studied briefly.

Most spaces considered in this book have a topological structure given by a metric or a norm, and we first review these notions. We begin with metric spaces. Recall that $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$.

**Definition 15.1.** A **metric space** is a set $E$ together with a function $d: E \times E \to \mathbb{R}_+$, called a **metric**, or **distance**, assigning a nonnegative real number $d(x, y)$ to any two points $x, y \in E$, and satisfying the following conditions for all $x, y, z \in E$:

1. **(D1)** $d(x, y) = d(y, x)$.
   (symmetry)
2. **(D2)** $d(x, y) \geq 0$, and $d(x, y) = 0$ iff $x = y$.
   (positivity)
3. **(D3)** $d(x, z) \leq d(x, y) + d(y, z)$.
   (triangle inequality)

Geometrically, Condition (D3) expresses the fact that in a triangle with vertices $x, y, z$, the length of any side is bounded by the sum of the lengths of the other two sides. From (D3), we immediately get

$$|d(x, y) - d(y, z)| \leq d(x, z).$$

Let us give some examples of metric spaces. Recall that the **absolute value** $|x|$ of a real number $x \in \mathbb{R}$ is defined such that $|x| = x$ if $x \geq 0$, $|x| = -x$ if $x < 0$, and for a complex number $x = a + ib$, by $|x| = \sqrt{a^2 + b^2}$. 

579
Example 15.1.

1. Let $E = \mathbb{R}$, and $d(x, y) = |x - y|$, the absolute value of $x - y$. This is the so-called natural metric on $\mathbb{R}$.

2. Let $E = \mathbb{R}^n$ (or $E = \mathbb{C}^n$). We have the Euclidean metric

$$d_2(x, y) = (|x_1 - y_1|^2 + \cdots + |x_n - y_n|^2)^{\frac{1}{2}},$$

the distance between the points $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$.

3. For every set $E$, we can define the discrete metric, defined such that $d(x, y) = 1$ iff $x \neq y$, and $d(x, x) = 0$.

4. For any $a, b \in \mathbb{R}$ such that $a < b$, we define the following sets:

$$[a, b] = \{ x \in \mathbb{R} \mid a \leq x \leq b \}, \quad \text{(closed interval)}$$

$$(a, b) = \{ x \in \mathbb{R} \mid a < x < b \}, \quad \text{(open interval)}$$

$$[a, b) = \{ x \in \mathbb{R} \mid a \leq x < b \}, \quad \text{(interval closed on the left, open on the right)}$$

$$(a, b] = \{ x \in \mathbb{R} \mid a < x \leq b \}, \quad \text{(interval open on the left, closed on the right)}$$

Let $E = [a, b]$, and $d(x, y) = |x - y|$. Then, $([a, b], d)$ is a metric space.

We will need to define the notion of proximity in order to define convergence of limits and continuity of functions. For this we introduce some standard “small neighborhoods.”

**Definition 15.2.** Given a metric space $E$ with metric $d$, for every $a \in E$, for every $\rho \in \mathbb{R}$, with $\rho > 0$, the set

$$B(a, \rho) = \{ x \in E \mid d(a, x) \leq \rho \}$$

is called the **closed ball of center $a$ and radius $\rho$**, the set

$$B_0(a, \rho) = \{ x \in E \mid d(a, x) < \rho \}$$

is called the **open ball of center $a$ and radius $\rho$**, and the set

$$S(a, \rho) = \{ x \in E \mid d(a, x) = \rho \}$$

is called the **sphere of center $a$ and radius $\rho$**. It should be noted that $\rho$ is finite (i.e., not $+\infty$). A subset $X$ of a metric space $E$ is **bounded** if there is a closed ball $B(a, \rho)$ such that $X \subseteq B(a, \rho)$.

Clearly, $B(a, \rho) = B_0(a, \rho) \cup S(a, \rho)$. 
Example 15.2.

1. In $E = \mathbb{R}$ with the distance $|x - y|$, an open ball of center $a$ and radius $\rho$ is the open interval $(a - \rho, a + \rho)$.

2. In $E = \mathbb{R}^2$ with the Euclidean metric, an open ball of center $a$ and radius $\rho$ is the set of points inside the disk of center $a$ and radius $\rho$, excluding the boundary points on the circle.

3. In $E = \mathbb{R}^3$ with the Euclidean metric, an open ball of center $a$ and radius $\rho$ is the set of points inside the sphere of center $a$ and radius $\rho$, excluding the boundary points on the sphere.

One should be aware that intuition can be misleading in forming a geometric image of a closed (or open) ball. For example, if $d$ is the discrete metric, a closed ball of center $a$ and radius $\rho < 1$ consists only of its center $a$, and a closed ball of center $a$ and radius $\rho \geq 1$ consists of the entire space!

If $E = [a, b]$, and $d(x, y) = |x - y|$, as in Example 15.1, an open ball $B_0(a, \rho)$, with $\rho < b - a$, is in fact the interval $[a, a + \rho)$, which is closed on the left.

We now consider a very important special case of metric spaces, normed vector spaces. Normed vector spaces have already been defined in Chapter 16 (Definition 16.1) but for the reader’s convenience we repeat the definition.

**Definition 15.3.** Let $E$ be a vector space over a field $K$, where $K$ is either the field $\mathbb{R}$ of reals, or the field $\mathbb{C}$ of complex numbers. A norm on $E$ is a function $\| \| : E \to \mathbb{R}_+$, assigning a nonnegative real number $\|u\|$ to any vector $u \in E$, and satisfying the following conditions for all $x, y, z \in E$:

(N1) $\|x\| \geq 0$, and $\|x\| = 0$ iff $x = 0$. (positivity)

(N2) $\|\lambda x\| = |\lambda| \|x\|$. (scaling)

(N3) $\|x + y\| \leq \|x\| + \|y\|$. (triangle inequality)

A vector space $E$ together with a norm $\| \|$ is called a **normed vector space**. A function $\| \| : E \to \mathbb{R}_+$ satisfying only properties (N2) and (N3) is called a **semi-norm**.

From (N3), we easily get

$$\|x\| - \|y\| \leq \|x - y\|.$$

Given a normed vector space $E$, if we define $d$ such that

$$d(x, y) = \|x - y\|,$$
it is easily seen that \( d \) is a metric. Thus, every normed vector space is immediately a metric space. Note that the metric associated with a norm is invariant under translation, that is, 

\[
d(x + u, y + u) = d(x, y).
\]

For this reason, we can restrict ourselves to open or closed balls of center 0.

If \( \| \cdot \| \) : \( E \to \mathbb{R}_+ \) is a semi-norm, then \( \| x \| = 0 \) does not necessarily imply that \( x = 0 \). However by setting \( \lambda = 0 \) and \( x = 0 \) in (N2), we see that \( \| 0 \| = 0 \). If we let \( \mathcal{N} = \{ x \in E \mid \| x \| = 0 \} \), then \( \mathcal{N} \) is a subspace of \( E \). Indeed, \( 0 \in \mathcal{N} \), and if \( \| x \| = \| y \| = 0 \), then by (N2) and (N3) we have

\[
\| \lambda x + \mu y \| \leq \| \lambda x \| + \| \mu y \| = |\lambda| \| x \| + |\mu| \| y \| = 0 + 0 = 0,
\]

so \( \lambda x + \mu y \in \mathcal{N} \). We can form the quotient space \( E/\mathcal{N} \), and then it is easy to see that the semi-norm \( \| \cdot \| \) induces a norm on \( X/\mathcal{N} \).

Natural examples of semi-norms arise in integration theory; see Chapter 4.

Examples of normed vector spaces were given in Example 16.1. We repeat the most important examples.

**Example 15.3.** Let \( E = \mathbb{R}^n \) (or \( E = \mathbb{C}^n \)). There are three standard norms. For every \((x_1, \ldots, x_n) \in E\), we have the norm \( \| x \|_1 \), defined such that,

\[
\| x \|_1 = |x_1| + \cdots + |x_n|,
\]

we have the *Euclidean norm* \( \| x \|_2 \), defined such that,

\[
\| x \|_2 = (|x_1|^2 + \cdots + |x_n|^2)^{1/2},
\]

and the *sup-norm* \( \| x \|_\infty \), defined such that,

\[
\| x \|_\infty = \max\{|x_i| \mid 1 \leq i \leq n\}.
\]

More generally, we define the \( \ell_p \)-norm (for \( p \geq 1 \)) by

\[
\| x \|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}.
\]

We proved in Proposition 16.1 that the \( \ell_p \)-norms are indeed norms. The closed unit balls centered at \((0,0)\) for \( \| \cdot \|_1 \), \( \| \cdot \|_2 \), and \( \| \cdot \|_\infty \), along with the containment relationships, are shown in Figures 15.1 and 15.2. Figures 15.3 and 15.4 illustrate the situation in \( \mathbb{R}^3 \).

In a normed vector space we define a closed ball or an open ball of radius \( \rho \) as a closed ball or an open ball of center 0. We may use the notation \( B(\rho) \) and \( B_0(\rho) \).

We will now define the crucial notions of open sets and closed sets, and of a topological space.
Figure 15.1: Figure (a) shows the diamond shaped closed ball associated with $\| \|_1$. Figure (b) shows the closed unit disk associated with $\| \|_2$, while Figure (c) illustrates the closed unit ball associated with $\| \|_\infty$.

Figure 15.2: The relationship between the closed unit balls centered at $(0,0)$.

**Definition 15.4.** Let $(E,d)$ be a metric space. A subset $U \subseteq E$ is an **open set** in $E$ if either $U =\emptyset$, or for every $a \in U$, there is some open ball $B_0(a, \rho)$ such that, $B_0(a, \rho) \subseteq U$.\(^1\) A subset $F \subseteq E$ is a **closed set** in $E$ if its complement $E - F$ is open in $E$. See Figure 15.5.

The set $E$ itself is open, since for every $a \in E$, every open ball of center $a$ is contained in $E$. In $E = \mathbb{R}^n$, given $n$ intervals $[a_i, b_i]$, with $a_i < b_i$, it is easy to show that the open $n$-cube

$$\{(x_1, \ldots, x_n) \in E \mid a_i < x_i < b_i, \ 1 \leq i \leq n\}$$

is an open set. In fact, it is possible to find a metric for which such open $n$-cubes are open balls! Similarly, we can define the closed $n$-cube

$$\{(x_1, \ldots, x_n) \in E \mid a_i \leq x_i \leq b_i, \ 1 \leq i \leq n\}$$

\(^1\)Recall that $\rho > 0$. 
CHAPTER 15. TOPOLOGY

Figure 15.3: Figure (a) shows the octahedral shaped closed ball associated with \( \parallel \parallel_1 \). Figure (b) shows the closed spherical associated with \( \parallel \parallel_2 \), while Figure (c) illustrates the closed unit ball associated with \( \parallel \parallel_\infty \).

which is a closed set.

The open sets satisfy some important properties that lead to the definition of a topological space.

**Proposition 15.1.** Given a metric space \( E \) with metric \( d \), the family \( \mathcal{O} \) of all open sets defined in Definition 15.4 satisfies the following properties:

(O1) For every finite family \((U_i)_{1 \leq i \leq n}\) of sets \( U_i \in \mathcal{O} \), we have \( U_1 \cap \cdots \cap U_n \in \mathcal{O} \), i.e., \( \mathcal{O} \) is closed under finite intersections.

(O2) For every arbitrary family \((U_i)_{i \in I}\) of sets \( U_i \in \mathcal{O} \), we have \( \bigcup_{i \in I} U_i \in \mathcal{O} \), i.e., \( \mathcal{O} \) is closed under arbitrary unions.

(O3) \( \emptyset \in \mathcal{O} \), and \( E \in \mathcal{O} \), i.e., \( \emptyset \) and \( E \) belong to \( \mathcal{O} \).

Furthermore, for any two distinct points \( a \neq b \) in \( E \), there exist two open sets \( U_a \) and \( U_b \) such that, \( a \in U_a \), \( b \in U_b \), and \( U_a \cap U_b = \emptyset \).
15.1. METRIC SPACES AND NORMED VECTOR SPACES

Figure 15.4: The relationship between the closed unit balls centered at (0, 0, 0).

Figure 15.5: An open set \( U \) in \( E = \mathbb{R}^2 \) under the standard Euclidean metric. Any point in the peach set \( U \) is surrounded by a small raspberry open set which lies within \( U \).

Proof. It is straightforward. For the last point, letting \( \rho = d(a, b)/3 \) (in fact \( \rho = d(a, b)/2 \) works too), we can pick \( U_a = B_0(a, \rho) \) and \( U_b = B_0(b, \rho) \). By the triangle inequality, we must have \( U_a \cap U_b = \emptyset \).

The above proposition leads to the very general concept of a topological space.

One should be careful that, in general, the family of open sets is not closed under infinite intersections. For example, in \( \mathbb{R} \) under the metric \(|x - y|\), letting \( U_n = (-1/n, +1/n) \), each \( U_n \) is open, but \( \bigcap_n U_n = \{0\} \), which is not open.

Later on, given any nonempty subset \( A \) of a metric space \((E, d)\), we will need to know that certain special sets containing \( A \) are open.

**Definition 15.5.** Let \((E, d)\) be a metric space. For any nonempty subset \( A \) of \( E \) and any \( x \in E \), let

\[
d(x, A) = \inf_{a \in A} d(x, a).
\]
Proposition 15.2. Let \((E,d)\) be a metric space. For any nonempty subset \(A\) of \(E\) and for any two points \(x, y \in E\), we have

\[ |d(x, A) - d(y, A)| \leq d(x, y). \]

Proof. For all \(a \in A\) we have

\[ d(x, a) \leq d(x, y) + d(y, a), \]

which implies

\[
\begin{align*}
    d(x, A) &= \inf_{a \in A} d(x, a) \\
           &\leq \inf_{a \in A} (d(x, y) + d(y, a)) \\
           &= d(x, y) + \inf_{a \in A} d(y, a) \\
           &= d(x, y) + d(y, A).
\end{align*}
\]

By symmetry, we also obtain \(d(y, A) \leq d(x, y) + d(x, A)\), and thus

\[ |d(x, A) - d(y, A)| \leq d(x, y), \]

as claimed. \(\square\)

Definition 15.6. Let \((E,d)\) be a metric space. For any nonempty subset \(A\) of \(E\), and any \(r > 0\), let

\[ V_r(A) = \{ x \in E \mid d(x, A) < r \}. \]

Proposition 15.3. Let \((E,d)\) be a metric space. For any nonempty subset \(A\) of \(E\), and any \(r > 0\), the set \(V_r(A)\) is an open set containing \(A\).

Proof. For any \(y \in E\) such that \(d(x, y) < r - d(x, A)\), by Proposition 15.2 we have

\[ d(y, A) \leq d(x, A) + d(x, y) \leq d(x, A) + r - d(x, A) = r, \]

so \(V_r(A)\) contains the open ball \(B_0(x, r - d(x, A))\), which means that it is open. Obviously, \(A \subseteq V_r(A)\). \(\square\)

15.2 Topological Spaces

Motivated by Proposition 15.1, a topological space is defined in terms of a family of sets satisfying the properties of open sets stated in that proposition.

Definition 15.7. Given a set \(E\), a topology on \(E\) (or a topological structure on \(E\)), is defined as a family \(\mathcal{O}\) of subsets of \(E\) called open sets, and satisfying the following three properties:
(1) For every finite family \((U_i)_{1 \leq i \leq n}\) of sets \(U_i \in \mathcal{O}\), we have \(U_1 \cap \cdots \cap U_n \in \mathcal{O}\), i.e., \(\mathcal{O}\) is closed under finite intersections.

(2) For every arbitrary family \((U_i)_{i \in I}\) of sets \(U_i \in \mathcal{O}\), we have \(\bigcup_{i \in I} U_i \in \mathcal{O}\), i.e., \(\mathcal{O}\) is closed under arbitrary unions.

(3) \(\emptyset \in \mathcal{O}\), and \(E \in \mathcal{O}\), i.e., \(\emptyset\) and \(E\) belong to \(\mathcal{O}\).

A set \(E\) together with a topology \(\mathcal{O}\) on \(E\) is called a topological space. Given a topological space \((E, \mathcal{O})\), a subset \(F\) of \(E\) is a closed set if \(F = E - U\) for some open set \(U \in \mathcal{O}\), i.e., \(F\) is the complement of some open set.

It is possible that an open set is also a closed set. For example, \(\emptyset\) and \(E\) are both open and closed. When a topological space contains a proper nonempty subset \(U\) which is both open and closed, the space \(E\) is said to be disconnected.

Definition 15.8. A topological space \((E, \mathcal{O})\) is said to satisfy the Hausdorff separation axiom (or \(T_2\)-separation axiom) if for any two distinct points \(a \neq b\) in \(E\), there exist two open sets \(U_a\) and \(U_b\) such that, \(a \in U_a\), \(b \in U_b\), and \(U_a \cap U_b = \emptyset\). When the \(T_2\)-separation axiom is satisfied, we also say that \((E, \mathcal{O})\) is a Hausdorff space.

As shown by Proposition 15.1, any metric space is a topological Hausdorff space, the family of open sets being in fact the family of arbitrary unions of open balls. Similarly, any normed vector space is a topological Hausdorff space, the family of open sets being the family of arbitrary unions of open balls. The topology \(\mathcal{O}\) consisting of all subsets of \(E\) is called the discrete topology.

Remark: Most (if not all) spaces used in analysis are Hausdorff spaces. Intuitively, the Hausdorff separation axiom says that there are enough “small” open sets. Without this axiom, some counter-intuitive behaviors may arise. For example, a sequence may have more than one limit point (or a compact set may not be closed). Nevertheless, non-Hausdorff topological spaces arise naturally in algebraic geometry. But even there, some substitute for separation is used.

One of the reasons why topological spaces are important is that the definition of a topology only involves a certain family \(\mathcal{O}\) of sets, and not how such family is generated from a metric or a norm. For example, different metrics or different norms can define the same family of open sets. Many topological properties only depend on the family \(\mathcal{O}\) and not on the specific metric or norm. But the fact that a topology is definable from a metric or a norm is important, because it usually implies nice properties of a space. All our examples will be spaces whose topology is defined by a metric or a norm.

Definition 15.9. A topological space \((E, \mathcal{O})\) is metrizable if there is a distance on \(E\) defining the topology \(\mathcal{O}\).
Note that in a metric space \((E, d)\), the metric \(d\) is explicitly given. However, in general, the topology of a metrizable space \((E, \mathcal{O})\) is not specified by an explicitly given metric, but some metric defining the topology \(\mathcal{O}\) exists. Obviously, a metrizable topological space must be Hausdorff. Actually, a stronger separation property holds, a metrizable space is normal; see Definition 15.30.

Remark: By taking complements we can state properties of the closed sets dual to those of Definition 15.7. Thus, \(\emptyset\) and \(E\) are closed sets, and the closed sets are closed under finite unions and arbitrary intersections.

It is also worth noting that the Hausdorff separation axiom implies that for every \(a \in E\), the set \(\{a\}\) is closed. Indeed, if \(x \in E - \{a\}\), then \(x \neq a\), and so there exist open sets \(U_a\) and \(U_x\) such that \(a \in U_a, x \in U_x\), and \(U_a \cap U_x = \emptyset\). See Figure 15.6. Thus, for every \(x \in E - \{a\}\), there is an open set \(U_x\) containing \(x\) and contained in \(E - \{a\}\), showing by (O3) that \(E - \{a\}\) is open, and thus that the set \(\{a\}\) is closed.

![Figure 15.6: A schematic illustration of the Hausdorff separation property](image)

Given a topological space \((E, \mathcal{O})\), given any subset \(A\) of \(E\), since \(E \in \mathcal{O}\) and \(E\) is a closed set, the family \(\mathcal{C}_A = \{F \mid A \subseteq F, F\text{ a closed set}\}\) of closed sets containing \(A\) is nonempty, and since any arbitrary intersection of closed sets is a closed set, the intersection \(\bigcap \mathcal{C}_A\) of the sets in the family \(\mathcal{C}_A\) is the smallest closed set containing \(A\). By a similar reasoning, the union of all the open subsets contained in \(A\) is the largest open set contained in \(A\).

Definition 15.10. Given a topological space \((E, \mathcal{O})\), given any subset \(A\) of \(E\), the smallest closed set containing \(A\) is denoted by \(\overline{A}\), and is called the closure, or adherence of \(A\). See Figure 15.7. A subset \(A\) of \(E\) is dense in \(E\) if \(\overline{A} = E\). The largest open set contained in \(A\) is denoted by \(\overset{\circ}{A}\), and is called the interior of \(A\). See Figure 15.8. The set \(\text{Fr} A = \overline{A} \cap E - \overset{\circ}{A}\) is called the boundary (or frontier) of \(A\). We also denote the boundary of \(A\) by \(\partial A\). See Figure 15.9.
Figure 15.7: The topological space \((E, \mathcal{O})\) is \(\mathbb{R}^2\) with topology induced by the Euclidean metric. The subset \(A\) is the section \(B_0(1)\) in the first and fourth quadrants bound by the lines \(y = x\) and \(y = -x\). The closure of \(A\) is obtained by the intersection of \(A\) with the closed unit ball.

Figure 15.8: The topological space \((E, \mathcal{O})\) is \(\mathbb{R}^2\) with topology induced by the Euclidean metric. The subset \(A\) is the section \(B_0(1)\) in the first and fourth quadrants bound by the lines \(y = x\) and \(y = -x\). The interior of \(A\) is obtained by the covering \(A\) with small open balls.
Figure 15.9: The topological space \((E, \mathcal{O})\) is \(\mathbb{R}^2\) with topology induced by the Euclidean metric. The subset \(A\) is the section \(B_0(1)\) in the first and fourth quadrants bound by the lines \(y = x\) and \(y = -x\). The boundary of \(A\) is \(\overline{A} - \overset{\circ}{A}\).

**Remark:** The notation \(\overline{A}\) for the closure of a subset \(A\) of \(E\) is somewhat unfortunate, since \(\overline{A}\) is often used to denote the set complement of \(A\) in \(E\). Still, we prefer it to more cumbersome notations such as \(\text{clo}(A)\), and we denote the complement of \(A\) in \(E\) by \(E - A\) (or sometimes, \(A^c\)).

By definition, it is clear that a subset \(A\) of \(E\) is closed iff \(\overline{A} = \overline{A}\). The set \(\mathbb{Q}\) of rationals is dense in \(\mathbb{R}\). It is easily shown that \(\overline{A} = \overset{\circ}{A} \cup \partial A\) and \(\overset{\circ}{A} \cap \partial A = \emptyset\). Another useful characterization of \(\overline{A}\) is given by the following proposition.

**Proposition 15.4.** Given a topological space \((E, \mathcal{O})\), given any subset \(A\) of \(E\), the closure \(\overline{A}\) of \(A\) is the set of all points \(x \in E\) such that for every open set \(U\) containing \(x\), then \(U \cap A \neq \emptyset\). See Figure 15.10.

**Proof.** If \(A = \emptyset\), since \(\emptyset\) is closed, the proposition holds trivially. Thus, assume that \(A \neq \emptyset\). First assume that \(x \in \overline{A}\). Let \(U\) be any open set such that \(x \in U\). If \(U \cap A = \emptyset\), since \(U\) is open, then \(E - U\) is a closed set containing \(A\), and since \(\overline{A}\) is the intersection of all closed sets containing \(A\), we must have \(x \in E - U\), which is impossible. Conversely, assume that \(x \in E\) is a point such that for every open set \(U\) containing \(x\), then \(U \cap A \neq \emptyset\). Let \(F\) be any closed subset containing \(A\). If \(x \notin F\), since \(F\) is closed, then \(U = E - F\) is an open set such that \(x \in U\), and \(U \cap A = \emptyset\), a contradiction. Thus, we have \(x \in F\) for every closed set containing \(A\), that is, \(x \in \overline{A}\). \(\square\)

Often it is necessary to consider a subset \(A\) of a topological space \(E\), and to view the subset \(A\) as a topological space. The following proposition shows how to define a topology on a subset.

**Proposition 15.5.** Given a topological space \((E, \mathcal{O})\), given any subset \(A\) of \(E\), let

\[
U = \{U \cap A \mid U \in \mathcal{O}\}
\]

be the family of all subsets of \(A\) obtained as the intersection of any open set in \(\mathcal{O}\) with \(A\). The following properties hold.
15.2. TOPOLOGICAL SPACES

Figure 15.10: The topological space $(E, \mathcal{O})$ is $\mathbb{R}^2$ with topology induced by the Euclidean metric. The purple subset $A$ is illustrated with three red points, each in its closure since the open ball centered at each point has nontrivial intersection with $A$.

(1) The space $(A, \mathcal{U})$ is a topological space.

(2) If $E$ is a metric space with metric $d$, then the restriction $d_A: A \times A \to \mathbb{R}_+$ of the metric $d$ to $A$ defines a metric space. Furthermore, the topology induced by the metric $d_A$ agrees with the topology defined by $\mathcal{U}$, as above.

Proof. Left as an exercise. \qed

Proposition 15.5 suggests the following definition.

Definition 15.11. Given a topological space $(E, \mathcal{O})$, given any subset $A$ of $E$, the subspace topology on $A$ induced by $\mathcal{O}$ is the family $\mathcal{U}$ of open sets defined such that

$$\mathcal{U} = \{U \cap A \mid U \in \mathcal{O}\}$$

is the family of all subsets of $A$ obtained as the intersection of any open set in $\mathcal{O}$ with $A$. We say that $(A, \mathcal{U})$ has the subspace topology. If $(E, d)$ is a metric space, the restriction $d_A: A \times A \to \mathbb{R}_+$ of the metric $d$ to $A$ is called the subspace metric.

For example, if $E = \mathbb{R}^n$ and $d$ is the Euclidean metric, we obtain the subspace topology on the closed $n$-cube

$$\{(x_1, \ldots, x_n) \in E \mid a_i \leq x_i \leq b_i, \ 1 \leq i \leq n\}.$$
Figure 15.11: An example of an open set in the subspace topology for \( \{(x, y, z) \in \mathbb{R}^3 \mid -1 \leq x \leq 1, -1 \leq y \leq 1, -1 \leq z \leq 1\} \). The open set is the corner region \( ABCD \) and is obtained by intersecting with the cube \( B_0((1, 1, 1), 1) \).

One should realize that every open set \( U \in \mathcal{O} \) which is entirely contained in \( A \) is also in the family \( \mathcal{U} \), but \( \mathcal{U} \) may contain open sets that are not in \( \mathcal{O} \). For example, if \( E = \mathbb{R} \) with \( |x - y| \), and \( A = [a, b] \), then sets of the form \([a, c)\), with \( a < c < b \) belong to \( \mathcal{U} \), but they are not open sets for \( \mathbb{R} \) under \( |x - y| \). However, there is agreement in the following situation.

**Proposition 15.6.** Given a topological space \((E, \mathcal{O})\), given any subset \( A \) of \( E \), if \( \mathcal{U} \) is the subspace topology, then the following properties hold.

1. If \( A \) is an open set \( A \in \mathcal{O} \), then every open set \( U \in \mathcal{U} \) is an open set \( U \in \mathcal{O} \).

2. If \( A \) is a closed set in \( E \), then every closed set w.r.t. the subspace topology is a closed set w.r.t. \( \mathcal{O} \).

**Proof.** Left as an exercise.

The concept of product topology is also useful. We have the following proposition.
Proposition 15.7. Given $n$ topological spaces $(E_i, \mathcal{O}_i)$, let $\mathcal{B}$ be the family of subsets of $E_1 \times \cdots \times E_n$ defined as follows:

$$\mathcal{B} = \{ U_1 \times \cdots \times U_n \mid U_i \in \mathcal{O}_i, 1 \leq i \leq n \},$$

and let $\mathcal{P}$ be the family consisting of arbitrary unions of sets in $\mathcal{B}$, including $\emptyset$. Then $\mathcal{P}$ is a topology on $E_1 \times \cdots \times E_n$.

Proof. Left as an exercise. $\square$

Definition 15.12. Given $n$ topological spaces $(E_i, \mathcal{O}_i)$, the product topology on $E_1 \times \cdots \times E_n$ is the family $\mathcal{P}$ of subsets of $E_1 \times \cdots \times E_n$ defined as follows: if

$$\mathcal{B} = \{ U_1 \times \cdots \times U_n \mid U_i \in \mathcal{O}_i, 1 \leq i \leq n \},$$

then $\mathcal{P}$ is the family consisting of arbitrary unions of sets in $\mathcal{B}$, including $\emptyset$. See Figure 15.12.

![Figure 15.12: Examples of open sets in the product topology for $\mathbb{R}^2$ and $\mathbb{R}^3$ induced by the Euclidean metric.](image)

If each $(E_i, d_{E_i})$ is a metric space, there are three natural metrics that can be defined on $E_1 \times \cdots \times E_n$:

- $d_1((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = d_{E_1}(x_1, y_1) + \cdots + d_{E_n}(x_n, y_n)$,
- $d_2((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \left( (d_{E_1}(x_1, y_1))^2 + \cdots + (d_{E_n}(x_n, y_n))^2 \right)^{\frac{1}{2}}$,
- $d_{\infty}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) = \max\{d_{E_1}(x_1, y_1), \ldots, d_{E_n}(x_n, y_n)\}$.

It is easy to show that

$$d_{\infty}((x_1, \ldots, x_n), (y_1, \ldots, y_n)) \leq d_2((x_1, \ldots, x_n), (y_1, \ldots, y_n)) \leq d_1((x_1, \ldots, x_n), (y_1, \ldots, y_n))$$

$$\leq nd_{\infty}((x_1, \ldots, x_n), (y_1, \ldots, y_n)),$$
so these distances define the same topology, which is the product topology.

If each \((E_i, \| \cdot \|_{E_i})\) is a normed vector space, there are three natural norms that can be defined on \(E_1 \times \cdots \times E_n\):

\[
\| (x_1, \ldots, x_n) \|_1 = \| x_1 \|_{E_1} + \cdots + \| x_n \|_{E_n},
\]

\[
\| (x_1, \ldots, x_n) \|_2 = \left( \| x_1 \|_{E_1}^2 + \cdots + \| x_n \|_{E_n}^2 \right)^{\frac{1}{2}},
\]

\[
\| (x_1, \ldots, x_n) \|_{\infty} = \max \{ \| x_1 \|_{E_1}, \ldots, \| x_n \|_{E_n} \}.
\]

It is easy to show that

\[
\| (x_1, \ldots, x_n) \|_{\infty} \leq \| (x_1, \ldots, x_n) \|_2 \leq \| (x_1, \ldots, x_n) \|_1 \leq n \| (x_1, \ldots, x_n) \|_{\infty},
\]

so these norms define the same topology, which is the product topology. It can also be verified that when \(E_i = \mathbb{R}\), with the standard topology induced by \(|x - y|\), the topology product on \(\mathbb{R}^n\) is the standard topology induced by the Euclidean norm.

**Definition 15.13.** Two metrics \(d\) and \(d'\) on a space \(E\) are equivalent if they induce the same topology \(\mathcal{O}\) on \(E\) (i.e., they define the same family \(\mathcal{O}\) of open sets). Similarly, two norms \(\| \cdot \|\) and \(\| \cdot \|'\) on a space \(E\) are equivalent if they induce the same topology \(\mathcal{O}\) on \(E\).

Given a topological space \((E, \mathcal{O})\), it is often useful, as in Proposition 15.7, to define the topology \(\mathcal{O}\) in terms of a subfamily \(\mathcal{B}\) of subsets of \(E\).

**Definition 15.14.** We say that a family \(\mathcal{B}\) of subsets of \(E\) is a basis for the topology \(\mathcal{O}\), if \(\mathcal{B}\) is a subset of \(\mathcal{O}\), and if every open set \(U\) in \(\mathcal{O}\) can be obtained as some union (possibly infinite) of sets in \(\mathcal{B}\) (agreeing that the empty union is the empty set).

For example, given any metric space \((E, d)\), \(\mathcal{B} = \{ B_0(a, \rho) \mid a \in E, \rho > 0 \}\). In particular, if \(d = \| \cdot \|_2\), the open intervals form a basis for \(\mathbb{R}\), while the open disks form a basis for \(\mathbb{R}^2\). The open rectangles also form a basis for \(\mathbb{R}^2\) with the standard topology. See Figure 15.13.

It is immediately verified that if a family \(\mathcal{B} = (U_i)_{i \in I}\) is a basis for the topology of \((E, \mathcal{O})\), then \(E = \bigcup_{i \in I} U_i\), and the intersection of any two sets \(U_i, U_j \in \mathcal{B}\) is the union of some sets in the family \(\mathcal{B}\) (again, agreeing that the empty union is the empty set). Conversely, a family \(\mathcal{B}\) with these properties is the basis of the topology obtained by forming arbitrary unions of sets in \(\mathcal{B}\).

**Definition 15.15.** A subbasis for \(\mathcal{O}\) is a family \(\mathcal{S}\) of subsets of \(E\), such that the family \(\mathcal{B}\) of all finite intersections of sets in \(\mathcal{S}\) (including \(E\) itself, in case of the empty intersection) is a basis of \(\mathcal{O}\). See Figure 15.13.

The following proposition gives useful criteria for determining whether a family of open subsets is a basis of a topological space.
15.2. **TOPOLOGICAL SPACES**

Figure 15.13: Figure (i.) shows that the set of infinite open intervals forms a subbasis for \( \mathbb{R} \). Figure (ii.) shows that the infinite open strips form a subbasis for \( \mathbb{R}^2 \).

**Proposition 15.8.** Given a topological space \((E, \mathcal{O})\) and a family \(\mathcal{B}\) of open subsets in \(\mathcal{O}\) the following properties hold:

1. The family \(\mathcal{B}\) is a basis for the topology \(\mathcal{O}\) iff for every open set \(U \in \mathcal{O}\) and every \(x \in U\), there is some \(B \in \mathcal{B}\) such that \(x \in B\) and \(B \subseteq U\). See Figure 15.14.

2. The family \(\mathcal{B}\) is a basis for the topology \(\mathcal{O}\) iff
   
   (a) For every \(x \in E\), there is some \(B \in \mathcal{B}\) such that \(x \in B\).
   
   (b) For any two open subsets, \(B_1, B_2 \in \mathcal{B}\), for every \(x \in E\), if \(x \in B_1 \cap B_2\), then there is some \(B_3 \in \mathcal{B}\) such that \(x \in B_3\) and \(B_3 \subseteq B_1 \cap B_2\). See Figure 15.15.

Figure 15.14: Given an open subset \(U\) of \(\mathbb{R}^2\) and \(x \in U\), there exists an open ball \(B\) containing \(x\) with \(B \subseteq U\). There also exists an open rectangle \(B_1\) containing \(x\) with \(B_1 \subseteq U\).

We now consider the fundamental property of continuity.
15.3 Continuous Functions, Limits

Definition 15.16. Let \((E, \mathcal{O}_E)\) and \((F, \mathcal{O}_F)\) be topological spaces, and let \(f: E \to F\) be a function. For every \(a \in E\), we say that \(f\) is continuous at \(a\), if for every open set \(V \in \mathcal{O}_F\) containing \(f(a)\), there is some open set \(U \in \mathcal{O}_E\) containing \(a\), such that, \(f(U) \subseteq V\). See Figure 15.16. We say that \(f\) is continuous if it is continuous at every \(a \in E\).

![Figure 15.16: A schematic illustration of Definition 15.16.](image)

It is easy to see that Definition 15.16 is equivalent to the following statements.

Proposition 15.9. Let \((E, \mathcal{O}_E)\) and \((F, \mathcal{O}_F)\) be topological spaces, and let \(f: E \to F\) be a function. For every \(a \in E\), the function \(f\) is continuous at \(a \in E\) iff for every neighborhood \(N\) of \(f(a)\) in \(F\), then \(f^{-1}(N)\) is a neighborhood of \(a\). The function \(f\) is continuous on \(E\) iff \(f^{-1}(V)\) is an open set in \(\mathcal{O}_E\) for every open set \(V \in \mathcal{O}_F\).
Figure 15.17: A schematic illustration of the neighborhood condition.

If $E$ and $F$ are metric spaces defined by metrics $d_E$ and $d_F$, we can show easily that $f$ is continuous at $a$ iff 
for every $\epsilon > 0$, there is some $\eta > 0$, such that, for every $x \in E$,

$$
\text{if } d_E(a, x) \leq \eta, \text{ then } d_F(f(a), f(x)) \leq \epsilon.
$$

Similarly, if $E$ and $F$ are normed vector spaces defined by norms $\| \cdot \|_E$ and $\| \cdot \|_F$, we can show easily that $f$ is continuous at $a$ iff 
for every $\epsilon > 0$, there is some $\eta > 0$, such that, for every $x \in E$,

$$
\text{if } \|x - a\|_E \leq \eta, \text{ then } \|f(x) - f(a)\|_F \leq \epsilon.
$$

It is worth noting that continuity is a topological notion, in the sense that equivalent metrics (or equivalent norms) define exactly the same notion of continuity.

**Definition 15.17.** If $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ are topological spaces, and $f: E \to F$ is a function, for every nonempty subset $A \subseteq E$ of $E$, we say that $f$ is **continuous on** $A$ if the restriction of $f$ to $A$ is continuous with respect to $(A, \mathcal{U})$ and $(F, \mathcal{O}_F)$, where $\mathcal{U}$ is the subspace topology induced by $\mathcal{O}_E$ on $A$.

Given a product $E_1 \times \cdots \times E_n$ of topological spaces, as usual, we let $\pi_i: E_1 \times \cdots \times E_n \to E_i$ be the projection function such that, $\pi_i(x_1, \ldots, x_n) = x_i$. It is immediately verified that each $\pi_i$ is continuous.

Given a topological space $(E, \mathcal{O})$, we say that a point $a \in E$ is **isolated** if $\{a\}$ is an open set in $\mathcal{O}$. Then if $(E, \mathcal{O}_E)$ and $(F, \mathcal{O}_F)$ are topological spaces, any function $f: E \to F$ is continuous at every isolated point $a \in E$. In the discrete topology, every point is isolated.

In a nontrivial normed vector space $(E, \| \cdot \|)$ (with $E \neq \{0\}$), no point is isolated. To show this, we show that every open ball $B_0(u, \rho)$ contains some vectors different from $u$. 
Indeed, since $E$ is nontrivial, there is some $v \in E$ such that $v \neq 0$, and thus $\lambda = \|v\| > 0$ (by (N1)). Let

$$w = u + \frac{\rho}{\lambda + 1} v.$$ 

Since $v \neq 0$ and $\rho > 0$, we have $w \neq u$. Then,

$$\|w - u\| = \left\| \frac{\rho}{\lambda + 1} v \right\| = \frac{\rho \lambda}{\lambda + 1} < \rho,$$

which shows that $\|w - u\| < \rho$, for $w \neq u$.

The following proposition is easily shown.

**Proposition 15.10.** Given topological spaces $(E, \mathcal{O}_E)$, $(F, \mathcal{O}_F)$, and $(G, \mathcal{O}_G)$, and two functions $f : E \to F$ and $g : F \to G$, if $f$ is continuous at $a \in E$ and $g$ is continuous at $f(a) \in F$, then $g \circ f : E \to G$ is continuous at $a \in E$. Given $n$ topological spaces $(F_i, \mathcal{O}_i)$, for every function $f : E \to F_1 \times \cdots \times F_n$, then $f$ is continuous at $a \in E$ iff every $f_i : E \to F_i$ is continuous at $a$, where $f_i = \pi_i \circ f$.

One can also show that in a metric space $(E, d)$, the distance $d : E \times E \to \mathbb{R}$ is continuous, where $E \times E$ has the product topology. By the triangle inequality, we have

$$d(x, y) \leq d(x, x_0) + d(x_0, y_0) + d(y_0, y) = d(x_0, y_0) + d(x_0, x) + d(y_0, y)$$

and

$$d(x_0, y_0) \leq d(x_0, x) + d(x, y) + d(y, y_0) = d(x, y) + d(x_0, x) + d(y_0, y).$$

Consequently,

$$|d(x, y) - d(x_0, y_0)| \leq d(x_0, x) + d(y_0, y),$$

which proves that $d$ is continuous at $(x_0, y_0)$. In fact this shows that $d$ is uniformly continuous; see Definition 15.37.

Given any nonempty subset $A$ of $E$, by Proposition 15.2, the map $x \mapsto d(x, A)$ is continuous (in fact, uniformly continuous).

Similarly, for a normed vector space $(E, \| \|)$, the norm $\| \| : E \to \mathbb{R}$ is (uniformly) continuous.

Given a function $f : E_1 \times \cdots \times E_n \to F$, we can fix $n - 1$ of the arguments, say $a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n$, and view $f$ as a function of the remaining argument,

$$x_i \mapsto f(a_1, \ldots, a_{i-1}, x_i, a_{i+1}, \ldots, a_n),$$

where $x_i \in E_i$. If $f$ is continuous, it is clear that each $f_i$ is continuous.
One should be careful that the converse is false! For example, consider the function
\[ f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \]
defined such that,
\[ f(x, y) = \frac{xy}{x^2 + y^2} \quad \text{if} \ (x, y) \neq (0, 0), \quad \text{and} \quad f(0, 0) = 0. \]
The function \( f \) is continuous on \( \mathbb{R} \times \mathbb{R} - \{(0, 0)\} \), but on the line \( y = mx \), with \( m \neq 0 \), we have \( f(x, y) = \frac{m}{1 + m^2} \neq 0 \), and thus, on this line, \( f(x, y) \) does not approach 0 when \( (x, y) \) approaches \( (0, 0) \). See Figure 15.18.

The following proposition is useful for showing that real-valued functions are continuous.

**Proposition 15.11.** If \( E \) is a topological space, and \((\mathbb{R}, |x - y|)\) the reals under the standard topology, for any two functions \( f : E \to \mathbb{R} \) and \( g : E \to \mathbb{R} \), for any \( a \in E \), for any \( \lambda \in \mathbb{R} \), if \( f \) and \( g \) are continuous at \( a \), then \( f + g, \lambda f, f \cdot g \), are continuous at \( a \), and \( f/g \) is continuous at \( a \) if \( g(a) \neq 0 \).

**Proof.** Left as an exercise. \( \square \)

Using Proposition 15.11, we can show easily that every real polynomial function is continuous.

The notion of isomorphism of topological spaces is defined as follows.

**Definition 15.18.** Let \((E, O_E)\) and \((F, O_F)\) be topological spaces, and let \( f : E \to F \) be a function. We say that \( f \) is a homeomorphism between \( E \) and \( F \) if \( f \) is bijective, and both \( f : E \to F \) and \( f^{-1} : F \to E \) are continuous.
One should be careful that a bijective continuous function \( f: E \to F \) is not necessarily a homeomorphism. For example, if \( E = \mathbb{R} \) with the discrete topology, and \( F = \mathbb{R} \) with the standard topology, the identity is not a homeomorphism. Another interesting example involving a parametric curve is given below. Let \( L: \mathbb{R} \to \mathbb{R}^2 \) be the function, defined such that,

\[
L_1(t) = \frac{t(1+t^2)}{1+t^4},
L_2(t) = \frac{t(1-t^2)}{1+t^4}.
\]

If we think of \((x(t), y(t)) = (L_1(t), L_2(t))\) as a geometric point in \( \mathbb{R}^2 \), the set of points \((x(t), y(t))\) obtained by letting \( t \) vary in \( \mathbb{R} \) from \(-\infty \) to \(+\infty \), defines a curve having the shape of a “figure eight,” with self-intersection at the origin, called the “lemniscate of Bernoulli.” See Figure 15.19. The map \( L \) is continuous, and in fact bijective, but its inverse \( L^{-1} \) is not continuous. Indeed, when we approach the origin on the branch of the curve in the upper left quadrant (i.e., points such that, \( x \leq 0, y \geq 0 \)), then \( t \) goes to \(-\infty \), and when we approach the origin on the branch of the curve in the lower right quadrant (i.e., points such that, \( x \geq 0, y \leq 0 \)), then \( t \) goes to \(+\infty \).

![Figure 15.19: The lemniscate of Bernoulli.](image)

We also review the concept of limit of a sequence. Given any set \( E \), a sequence is any function \( x: \mathbb{N} \to E \), usually denoted by \((x_n)_{n \in \mathbb{N}}\), or \((x_n)_{n \geq 0}\), or even by \((x_n)\).

**Definition 15.19.** Given a topological space \((E, \mathcal{O})\), we say that a sequence \((x_n)_{n \in \mathbb{N}}\) converges to some \( a \in E \) if for every open set \( U \) containing \( a \), there is some \( n_0 \geq 0 \), such that, \( x_n \in U \), for all \( n \geq n_0 \). We also say that \( a \) is a limit of \((x_n)_{n \in \mathbb{N}}\). See Figure 15.20.

When \( E \) is a metric space with metric \( d \), it is easy to show that this is equivalent to the fact that,

for every \( \epsilon > 0 \), there is some \( n_0 \geq 0 \), such that, \( d(x_n, a) \leq \epsilon \), for all \( n \geq n_0 \).

When \( E \) is a normed vector space with norm \( \| \| \), it is easy to show that this is equivalent to the fact that,

for every \( \epsilon > 0 \), there is some \( n_0 \geq 0 \), such that, \( \| x_n - a \| \leq \epsilon \), for all \( n \geq n_0 \).

The following proposition shows the importance of the Hausdorff separation axiom.
15.3. CONTINUOUS FUNCTIONS, LIMITS

Figure 15.20: A schematic illustration of Definition 15.19.

Proposition 15.12. Given a topological space \((E, \mathcal{O})\), if the Hausdorff separation axiom holds, then every sequence has at most one limit.

Proof. Left as an exercise. \qed

It is worth noting that the notion of limit is topological, in the sense that a sequence converge to a limit \(b\) iff it converges to the same limit \(b\) in any equivalent metric (and similarly for equivalent norms).

If \(E\) is a metric space and if \(A\) is a subset of \(E\), there is a convenient way of showing that a point \(x \in E\) belongs to the closure \(\overline{A}\) of \(A\) in terms of sequences.

Proposition 15.13. Given any metric space \((E, d)\), for any subset \(A\) of \(E\) and any point \(x \in E\), we have \(x \in \overline{A}\) iff there is a sequence \((a_n)\) of points \(a_n \in A\) converging to \(x\).

Proof. If the sequence \((a_n)\) of points \(a_n \in A\) converges to \(x\), then for every open subset \(U\) of \(E\) containing \(x\), there is some \(n_0\) such that \(a_n \in U\) for all \(n \geq n_0\), so \(U \cap A \neq \emptyset\), and Proposition 15.4 implies that \(x \in \overline{A}\).

Conversely, assume that \(x \in \overline{A}\). Then for every \(n \geq 1\), consider the open ball \(B_0(x, 1/n)\). By Proposition 15.4, we have \(B_0(x, 1/n) \cap A \neq \emptyset\), so we can pick some \(a_n \in B_0(x, 1/n) \cap A\). This, way, we define a sequence \((a_n)\) of points in \(A\), and by construction \(d(x, a_n) < 1/n\) for all \(n \geq 1\), so the sequence \((a_n)\) converges to \(x\). \qed

We still need one more concept of limit for functions.

Definition 15.20. Let \((E, \mathcal{O}_E)\) and \((F, \mathcal{O}_F)\) be topological spaces, let \(A\) be some nonempty subset of \(E\), and let \(f: A \to F\) be a function. For any \(a \in \overline{A}\) and any \(b \in F\), we say that \(f(x)\) approaches \(b\) as \(x\) approaches \(a\) with values in \(A\) if for every open set \(V \in \mathcal{O}_F\) containing \(b\), there is some open set \(U \in \mathcal{O}_E\) containing \(a\), such that, \(f(U \cap A) \subseteq V\). See Figure 15.21. This is denoted by

\[
\lim_{x \to a, x \in A} f(x) = b.
\]
First, note that by Proposition 15.4, since \( a \in \overline{A} \), for every open set \( U \) containing \( a \), we have \( U \cap A \neq \emptyset \), and the definition is nontrivial. Also, even if \( a \in A \), the value \( f(a) \) of \( f \) at \( a \) plays no role in this definition. When \( E \) and \( F \) are metric spaces with metrics \( d_E \) and \( d_F \), it can be shown easily that the definition can be stated as follows:

For every \( \epsilon > 0 \), there is some \( \eta > 0 \), such that, for every \( x \in A \),

\[
\text{if } d_E(x, a) \leq \eta, \text{ then } d_F(f(x), b) \leq \epsilon.
\]

When \( E \) and \( F \) are normed vector spaces with norms \( \| \cdot \|_E \) and \( \| \cdot \|_F \), it can be shown easily that the definition can be stated as follows:

For every \( \epsilon > 0 \), there is some \( \eta > 0 \), such that, for every \( x \in A \),

\[
\text{if } \|x - a\|_E \leq \eta, \text{ then } \|f(x) - b\|_F \leq \epsilon.
\]

We have the following result relating continuity at a point and the previous notion.

**Proposition 15.14.** Let \((E, \mathcal{O}_E)\) and \((F, \mathcal{O}_F)\) be two topological spaces, and let \( f : E \to F \) be a function. For any \( a \in E \), the function \( f \) is continuous at \( a \) iff \( f(x) \) approaches \( f(a) \) when \( x \) approaches \( a \) (with values in \( E \)).

**Proof.** Left as a trivial exercise. \( \square \)

Another important proposition relating the notion of convergence of a sequence to continuity, is stated without proof.

**Proposition 15.15.** Let \((E, \mathcal{O}_E)\) and \((F, \mathcal{O}_F)\) be two topological spaces, and let \( f : E \to F \) be a function.

(1) If \( f \) is continuous, then for every sequence \((x_n)_{n \in \mathbb{N}}\) in \( E \), if \((x_n)\) converges to \( a \), then \((f(x_n))\) converges to \( f(a) \).
(2) If $E$ is a metric space, and $(f(x_n))$ converges to $f(a)$ whenever $(x_n)$ converges to $a$, for every sequence $(x_n)_{n \in \mathbb{N}}$ in $E$, then $f$ is continuous.

A special case of Definition 15.20 will be used when $E$ and $F$ are (nontrivial) normed vector spaces with norms $\| \cdot \|_E$ and $\| \cdot \|_F$. Let $U$ be any nonempty open subset of $E$. We showed earlier that $E$ has no isolated points and that every set $\{v\}$ is closed, for every $v \in E$. Since $E$ is nontrivial, for every $v \in U$, there is a nontrivial open ball contained in $U$ (an open ball not reduced to its center). Then, for every $v \in U$, $A = U - \{v\}$ is open and nonempty, and clearly, $v \in \overline{A}$. For any $v \in U$, if $f(x)$ approaches $b$ when $x$ approaches $v$ with values in $A = U - \{v\}$, we say that $f(x)$ approaches $b$ when $x$ approaches $v$ with values $\neq v$ in $U$. This is denoted by

$$\lim_{x \to v, x \in U, x \neq v} f(x) = b.$$ 

**Remark:** Variations of the above case show up in the following case: $E = \mathbb{R}$, and $F$ is some arbitrary topological space. Let $A$ be some nonempty subset of $\mathbb{R}$, and let $f: A \to F$ be some function. For any $a \in A$, we say that $f$ is **continuous on the right at** $a$ if

$$\lim_{x \to a, x \in A \cap [a, +\infty)} f(x) = f(a).$$

We can define **continuity on the left at** $a$ in a similar fashion.

Let us consider another variation. Let $A$ be some nonempty subset of $\mathbb{R}$, and let $f: A \to F$ be some function. For any $a \in A$, we say that $f$ has a **discontinuity of the first kind at** $a$ if

$$\lim_{x \to a, x \in A \cap (-\infty, a)} f(x) = f(a_-)$$

and

$$\lim_{x \to a, x \in A \cap (a, +\infty)} f(x) = f(a_+)$$

both exist, and either $f(a_-) \neq f(a)$, or $f(a_+) \neq f(a)$.

Note that it is possible that $f(a_-) = f(a_+)$, but $f$ is still discontinuous at $a$ if this common value differs from $f(a)$. Functions defined on a nonempty subset of $\mathbb{R}$, and that are continuous, except for some points of discontinuity of the first kind, play an important role in analysis.

We now turn to connectivity properties of topological spaces.

### 15.4 Connected Sets

Connectivity properties of topological spaces play a very important role in understanding the topology of surfaces. This section gathers the facts needed to have a good understanding of the classification theorem for compact surfaces (with boundary). The main references are Ahlfors and Sario [1] and Massey [57, 58]. For general background on topology, geometry, and algebraic topology, we also highly recommend Bredon [13] and Fulton [33].
Definition 15.21. A topological space \((E, \mathcal{O})\) is connected if the only subsets of \(E\) that are both open and closed are the empty set and \(E\) itself. Equivalently, \((E, \mathcal{O})\) is connected if \(E\) cannot be written as the union \(E = U \cup V\) of two disjoint nonempty open sets \(U, V\), or if \(E\) cannot be written as the union \(E = U \cup V\) of two disjoint nonempty closed sets. A subset, \(S \subseteq E\), is connected if it is connected in the subspace topology on \(S\) induced by \((E, \mathcal{O})\). See Figure 15.22. A connected open set is called a region and a closed set is a closed region if its interior is a connected (open) set.

Figure 15.22: Figure (i) shows that the union of two disjoint disks in \(\mathbb{R}^2\) is a disconnected set since each circle can be separated by open half regions. Figure (ii) is an example of a connected subset of \(\mathbb{R}^2\) since the two disks can not separated by open sets.

The definition of connectivity is meant to capture the fact that a connected space \(S\) is “one piece.” Given the metric space \((\mathbb{R}^n, \|\|_2)\), the quintessential examples of connected spaces are \(B_0(a, \rho)\) and \(B(a, \rho)\). In particular, the following standard proposition characterizing the connected subsets of \(\mathbb{R}\) can be found in most topology texts (for example, Munkres [62], Schwartz [70]). For the sake of completeness, we give a proof.

Proposition 15.16. A subset of the real line, \(\mathbb{R}\), is connected iff it is an interval, i.e., of the form \([a, b]\), \((a, b]\), where \(a = -\infty\) is possible, \([a, b)\), where \(b = +\infty\) is possible, or \((a, b)\),
where \( a = -\infty \) or \( b = +\infty \) is possible.

**Proof.** Assume that \( A \) is a connected nonempty subset of \( \mathbb{R} \). The cases where \( A = \emptyset \) or \( A \) consists of a single point are trivial. Otherwise, we show that whenever \( a, b \in A, a < b \), then the entire interval \([a, b]\) is a subset of \( A \). Indeed, if this was not the case, there would be some \( c \in (a, b) \) such that \( c \notin A \), and then we could write \( A = ((-\infty, c] \cap A) \cup ((c, +\infty) \cap A) \), where \((\neg \infty, c] \cap A \) and \((c, +\infty) \cap A \) are nonempty and disjoint open subsets of \( A \), contradicting the fact that \( A \) is connected. It follows easily that \( A \) must be an interval.

Conversely, we show that an interval, \( I \), must be connected. Let \( A \) be any nonempty subset of \( I \) which is both open and closed in \( I \). We show that \( I = A \). Fix any \( x \in A \) and consider the set, \( R_x \), of all \( y \) such that \([x, y] \subseteq A \). If the set \( R_x \) is unbounded, then \( R_x = [x, +\infty) \). Otherwise, if this set is bounded, let \( b \) be its least upper bound. We claim that \( b \) is the right boundary of the interval \( I \). Because \( A \) is closed in \( I \), unless \( I \) is open on the right and \( b \) is its right boundary, we must have \( b \in A \). In the first case, \( A \cap [x, b) = I \cap [x, b) = [x, b) \). In the second case, because \( A \) is also open in \( I \), unless \( b \) is the right boundary of the interval \( I \) (closed on the right), there is some open set \((b - \eta, b + \eta) \) contained in \( A \), which implies that \([x, b + \eta/2] \subseteq A \), contradicting the fact that \( b \) is the least upper bound of the set \( R_x \). Thus, \( b \) must be the right boundary of the interval \( I \) (closed on the right). A similar argument applies to the set, \( L_y \), of all \( x \) such that \([x, y] \subseteq A \) and either \( L_y \) is unbounded, or its greatest lower bound \( a \) is the left boundary of \( I \) (open or closed on the left). In all cases, we showed that \( A = I \), and the interval must be connected. \( \square \)

Intuitively, if a space is not connected, it is possible to define a continuous function which is constant on disjoint “connected components” and which takes possibly distinct values on disjoint components. This can be stated in terms of the concept of a locally constant function.

**Definition 15.22.** Given two topological spaces \( X, Y \), a function \( f : X \to Y \) is **locally constant** if for every \( x \in X \), there is an open set \( U \subseteq X \) such that \( x \in U \) and \( f \) is constant on \( U \).

We claim that a locally constant function is continuous. In fact, we will prove that \( f^{-1}(V) \) is open for every subset, \( V \subseteq Y \) (not just for an open set \( V \)). It is enough to show that \( f^{-1}(y) \) is open for every \( y \in Y \), since for every subset \( V \subseteq Y \),

\[
f^{-1}(V) = \bigcup_{y \in V} f^{-1}(y),
\]

and open sets are closed under arbitrary unions. However, either \( f^{-1}(y) = \emptyset \) if \( y \in Y - f(X) \) or \( f \) is constant on \( U = f^{-1}(y) \) if \( y \in f(X) \) (with value \( y \)), and since \( f \) is locally constant, for every \( x \in U \), there is some open set, \( W \subseteq X \), such that \( x \in W \) and \( f \) is constant on \( W \), which implies that \( f(w) = y \) for all \( w \in W \) and thus, that \( W \subseteq U \), showing that \( U \) is a union of open sets and thus, is open. The following proposition shows that a space is connected iff every locally constant function is constant:
Proposition 15.17. A topological space is connected iff every locally constant function is constant. See Figure 15.23.

Figure 15.23: An example of a locally constant, but not constant, real-valued function $f$ over the disconnected set consisting of the disjoint union of the two solid balls. On the pink ball, $f$ is 0, while on the purple ball, $f$ is 1.

Proof. First, assume that $X$ is connected. Let $f : X \to Y$ be a locally constant function to some space $Y$ and assume that $f$ is not constant. Pick any $y \in f(X)$. Since $f$ is not constant, $U_1 = f^{-1}(y) \neq X$, and of course, $U_1 \neq \emptyset$. We proved just before Proposition 15.17 that $f^{-1}(V)$ is open for every subset $V \subseteq Y$, and thus $U_1 = f^{-1}(y) = f^{-1}(\{y\})$ and $U_2 = f^{-1}(Y - \{y\})$ are both open, nonempty, and clearly $X = U_1 \cup U_2$ and $U_1$ and $U_2$ are disjoint. This contradicts the fact that $X$ is connected and $f$ must be constant.

Assume that every locally constant function $f : X \to Y$ is constant. If $X$ is not connected, we can write $X = U_1 \cup U_2$, where both $U_1, U_2$ are open, disjoint, and nonempty. We can define the function, $f : X \to \mathbb{R}$, such that $f(x) = 1$ on $U_1$ and $f(x) = 0$ on $U_2$. Since $U_1$ and $U_2$ are open, the function $f$ is locally constant, and yet not constant, a contradiction. \qed

A characterization on the connected subsets of $\mathbb{R}^n$ is harder and requires the notion of arcwise connectedness. One of the most important properties of connected sets is that they are preserved by continuous maps.

Proposition 15.18. Given any continuous map, $f : E \to F$, if $A \subseteq E$ is connected, then $f(A)$ is connected.

Proof. If $f(A)$ is not connected, then there exist some nonempty open sets, $U, V$, in $F$ such that $f(A) \cap U$ and $f(A) \cap V$ are nonempty and disjoint, and

$$f(A) = (f(A) \cap U) \cup (f(A) \cap V).$$
Then, \( f^{-1}(U) \) and \( f^{-1}(V) \) are nonempty and open since \( f \) is continuous and

\[
A = (A \cap f^{-1}(U)) \cup (A \cap f^{-1}(V)),
\]

with \( A \cap f^{-1}(U) \) and \( A \cap f^{-1}(V) \) nonempty, disjoint, and open in \( A \), contradicting the fact that \( A \) is connected. \( \square \)

An important corollary of Proposition 15.18 is that for every continuous function, \( f : E \to \mathbb{R} \), where \( E \) is a connected space, \( f(E) \) is an interval. Indeed, this follows from Proposition 15.16. Thus, if \( f \) takes the values \( a \) and \( b \) where \( a < b \), then \( f \) takes all values \( c \in [a, b] \). This is a very important property known as the intermediate value theorem.

Even if a topological space is not connected, it turns out that it is the disjoint union of maximal connected subsets and these connected components are closed in \( E \). In order to obtain this result, we need a few lemmas.

**Lemma 15.19.** Given a topological space, \( E \), for any family, \((A_i)_{i \in I}\), of (nonempty) connected subsets of \( E \), if \( A_i \cap A_j \neq \emptyset \) for all \( i, j \in I \), then the union, \( A = \bigcup_{i \in I} A_i \), of the family, \((A_i)_{i \in I}\), is also connected.

**Proof.** Assume that \( \bigcup_{i \in I} A_i \) is not connected. There exists two nonempty open subsets, \( U \) and \( V \), of \( E \) such that \( A \cap U \) and \( A \cap V \) are disjoint and nonempty and such that

\[
A = (A \cap U) \cup (A \cap V).
\]

Now, for every \( i \in I \), we can write

\[
A_i = (A_i \cap U) \cup (A_i \cap V),
\]

where \( A_i \cap U \) and \( A_i \cap V \) are disjoint, since \( A_i \subseteq A \) and \( A \cap U \) and \( A \cap V \) are disjoint. Since \( A_i \) is connected, either \( A_i \cap U = \emptyset \) or \( A_i \cap V = \emptyset \). This implies that either \( A_i \subseteq A \cap U \) or \( A_i \subseteq A \cap V \). However, by assumption, \( A_i \cap A_j \neq \emptyset \), for all \( i, j \in I \), and thus, either both \( A_i \subseteq A \cap U \) and \( A_j \subseteq A \cap U \), or both \( A_i \subseteq A \cap V \) and \( A_j \subseteq A \cap V \), since \( A \cap U \) and \( A \cap V \) are disjoint. Thus, we conclude that either \( A_i \subseteq A \cap U \) for all \( i \in I \), or \( A_i \subseteq A \cap V \) for all \( i \in I \). But this proves that either

\[
A = \bigcup_{i \in I} A_i \subseteq A \cap U,
\]

or

\[
A = \bigcup_{i \in I} A_i \subseteq A \cap V,
\]

contradicting the fact that both \( A \cap U \) and \( A \cap V \) are disjoint and nonempty. Thus, \( A \) must be connected. \( \square \)

In particular, the above lemma applies when the connected sets in a family \((A_i)_{i \in I}\) have a point in common.
Lemma 15.20. If $A$ is a connected subset of a topological space, $E$, then for every subset, $B$, such that $A \subseteq B \subseteq \overline{A}$, where $\overline{A}$ is the closure of $A$ in $E$, the set $B$ is connected.

Proof. If $B$ is not connected, then there are two nonempty open subsets, $U, V$, of $E$ such that $B \cap U$ and $B \cap V$ are disjoint and nonempty, and

$$B = (B \cap U) \cup (B \cap V).$$

Since $A \subseteq B$, the above implies that

$$A = (A \cap U) \cup (A \cap V),$$

and since $A$ is connected, either $A \cap U = \emptyset$, or $A \cap V = \emptyset$. Without loss of generality, assume that $A \cap V = \emptyset$, which implies that $A \subseteq A \cap U \subseteq B \cap U$. However, $B \cap U$ is closed in the subspace topology for $B$ and since $B \subseteq \overline{A}$ and $\overline{A}$ is closed in $E$, the closure of $A$ in $B$ w.r.t. the subspace topology of $B$ is clearly $B \cap \overline{A} = B$, which implies that $B \subseteq B \cap U$ (since the closure is the smallest closed set containing the given set). Thus, $B \cap V = \emptyset$, a contradiction. \qed

In particular, Lemma 15.20 shows that if $A$ is a connected subset, then its closure, $\overline{A}$, is also connected. We are now ready to introduce the connected components of a space.

Definition 15.23. Given a topological space, $(E, \mathcal{O})$, we say that two points, $a, b \in E$, are connected if there is some connected subset, $A$, of $E$ such that $a \in A$ and $b \in A$.

It is immediately verified that the relation “$a$ and $b$ are connected in $E$” is an equivalence relation. Only transitivity is not obvious, but it follows immediately as a special case of Lemma 15.19. Thus, the above equivalence relation defines a partition of $E$ into nonempty disjoint connected components. The following proposition is easily proved using Lemma 15.19 and Lemma 15.20:

Proposition 15.21. Given any topological space, $E$, for any $a \in E$, the connected component containing $a$ is the largest connected set containing $a$. The connected components of $E$ are closed.

The notion of a locally connected space is also useful.

Definition 15.24. A topological space, $(E, \mathcal{O})$, is locally connected if for every $a \in E$, for every neighborhood, $V$, of $a$, there is a connected neighborhood, $U$, of $a$ such that $U \subseteq V$. See Figure 15.24.

As we shall see in a moment, it would be equivalent to require that $E$ has a basis of connected open sets.
15.4. CONNECTED SETS

There are connected spaces that are not locally connected and there are locally connected spaces that are not connected. The two properties are independent. For example, the subspace of \( \mathbb{R}^2 \) \( S = \{(x, \sin(1/x)), | x > 0\} \cup \{(0, y) | -1 \leq y \leq 1\} \) is connected but not locally connected. See Figure 15.25. The subspace \( S \) of \( \mathbb{R} \) consisting \([0, 1] \cup [2, 3]\) is locally connected but not connected.

**Proposition 15.22.** A topological space, \( E \), is locally connected iff for every open subset, \( A \), of \( E \), the connected components of \( A \) are open.

**Proof.** Assume that \( E \) is locally connected. Let \( A \) be any open subset of \( E \) and let \( C \) be one of the connected components of \( A \). For any \( a \in C \subseteq A \), there is some connected neighborhood, \( U \), of \( a \) such that \( U \subseteq A \) and since \( C \) is a connected component of \( A \) containing \( a \), we must have \( U \subseteq C \). This shows that for every \( a \in C \), there is some open subset containing \( a \) contained in \( C \), so \( C \) is open.

Conversely, assume that for every open subset, \( A \), of \( E \), the connected components of \( A \) are open. Then, for every \( a \in E \) and every neighborhood, \( U \), of \( a \), since \( U \) contains some open set \( A \) containing \( a \), the interior, \( \text{int} U \), of \( U \) is an open set containing \( a \) and its connected components are open. In particular, the connected component \( C \) containing \( a \) is a connected open set containing \( a \) and contained in \( U \).

Proposition 15.22 shows that in a locally connected space, the connected open sets form a basis for the topology. It is easily seen that \( \mathbb{R}^n \) is locally connected. Another very important property of surfaces and more generally, manifolds, is to be arcwise connected. The intuition is that any two points can be joined by a continuous arc of curve. This is formalized as follows.

**Definition 15.25.** Given a topological space, \((E, \mathcal{O})\), an arc (or path) is a continuous map, \( \gamma : [a, b] \rightarrow E \), where \([a, b]\) is a closed interval of the real line, \( \mathbb{R} \). The point \( \gamma(a) \) is the initial
Figure 15.25: Let $S$ be the graph of $f(x) = \sin(1/x)$ union the $y$-axis between $-1$ and $1$. This space is connected, but not locally connected.

Point of the arc and the point $\gamma(b)$ is the terminal point of the arc. We say that $\gamma$ is an arc joining $\gamma(a)$ and $\gamma(b)$. See Figure 15.26. An arc is a closed curve if $\gamma(a) = \gamma(b)$. The set $\gamma([a, b])$ is the trace of the arc $\gamma$.

Typically, $a = 0$ and $b = 1$.

One should not confuse an arc, $\gamma: [a, b] \to E$, with its trace. For example, $\gamma$ could be constant, and thus, its trace reduced to a single point.

An arc is a Jordan arc if $\gamma$ is a homeomorphism onto its trace. An arc, $\gamma: [a, b] \to E$, is a Jordan curve if $\gamma(a) = \gamma(b)$ and $\gamma$ is injective on $[a, b]$. Since $[a, b]$ is connected, by Proposition 15.18, the trace $\gamma([a, b])$ of an arc is a connected subset of $E$.

Given two arcs $\gamma: [0, 1] \to E$ and $\delta: [0, 1] \to E$ such that $\gamma(1) = \delta(0)$, we can form a new arc defined as follows:

Definition 15.26. Given two arcs, $\gamma: [0, 1] \to E$ and $\delta: [0, 1] \to E$, such that $\gamma(1) = \delta(0)$, we can form their composition (or product), $\gamma \delta$, defined such that

$$
\gamma \delta(t) = \begin{cases} 
\gamma(2t) & \text{if } 0 \leq t \leq 1/2; \\
\delta(2t - 1) & \text{if } 1/2 \leq t \leq 1.
\end{cases}
$$

The inverse, $\gamma^{-1}$, of the arc, $\gamma$, is the arc defined such that $\gamma^{-1}(t) = \gamma(1 - t)$, for all $t \in [0, 1]$. 
Figure 15.26: Let $E$ be the torus with subspace topology induced from $\mathbb{R}^3$ with red arc $\gamma([a,b])$. The torus is both arcwise connected and locally arcwise connected.

It is trivially verified that Definition 15.26 yields continuous arcs.

**Definition 15.27.** A topological space, $E$, is **arcwise connected** if for any two points, $a, b \in E$, there is an arc, $\gamma : [0,1] \rightarrow E$, joining $a$ and $b$, i.e., such that $\gamma(0) = a$ and $\gamma(1) = b$. A topological space, $E$, is **locally arcwise connected** if for every $a \in E$, for every neighborhood, $V$, of $a$, there is an arcwise connected neighborhood, $U$, of $a$ such that $U \subseteq V$. See Figure 15.26.

The space $\mathbb{R}^n$ is locally arcwise connected, since for any open ball, any two points in this ball are joined by a line segment. Manifolds and surfaces are also locally arcwise connected. Proposition 15.18 also applies to arcwise connectedness (this is a simple exercise). The following theorem is crucial to the theory of manifolds and surfaces:

**Theorem 15.23.** If a topological space, $E$, is arcwise connected, then it is connected. If a topological space, $E$, is connected and locally arcwise connected, then $E$ is arcwise connected.

**Proof.** First, assume that $E$ is arcwise connected. Pick any point, $a$, in $E$. Since $E$ is arcwise connected, for every $b \in E$, there is a path, $\gamma_b : [0,1] \rightarrow E$, from $a$ to $b$ and so,

$$E = \bigcup_{b \in E} \gamma_b([0,1])$$

a union of connected subsets all containing $a$. By Lemma 15.19, $E$ is connected.

Now assume that $E$ is connected and locally arcwise connected. For any point $a \in E$, let $F_a$ be the set of all points, $b$, such that there is an arc, $\gamma_b : [0,1] \rightarrow E$, from $a$ to $b$. Clearly, $F_a$ contains $a$. We show that $F_a$ is both open and closed. For any $b \in F_a$, since $E$ is locally
arcwise connected, there is an arcwise connected neighborhood $U$ containing $b$ (because $E$ is a neighborhood of $b$). Thus, $b$ can be joined to every point $c \in U$ by an arc, and since by the definition of $F_a$, there is an arc from $a$ to $b$, the composition of these two arcs yields an arc from $a$ to $c$, which shows that $c \in F_a$. But then $U \subseteq F_a$ and thus, $F_a$ is open. See Figure 15.27 (i.). Now assume that $b$ is in the complement of $F_a$. As in the previous case, there is some arcwise connected neighborhood $U$ containing $b$. Thus, every point $c \in U$ can be joined to $b$ by an arc. If there was an arc joining $a$ to $c$, we would get an arc from $a$ to $b$, contradicting the fact that $b$ is in the complement of $F_a$. Thus, every point $c \in U$ is in the complement of $F_a$, which shows that $U$ is contained in the complement of $F_a$, and thus, that the complement of $F_a$ is open. See Figure 15.27 (ii.). Consequently, we have shown that $F_a$ is both open and closed and since it is nonempty, we must have $E = F_a$, which shows that $E$ is arcwise connected.

\[
\text{Figure 15.27: Schematic illustrations of the proof techniques that show } F_a \text{ is both open and closed.}
\]

If $E$ is locally arcwise connected, the above argument shows that the connected components of $E$ are arcwise connected.
It is not true that a connected space is arcwise connected. For example, the space consisting of the graph of the function
\[ f(x) = \sin\left(\frac{1}{x}\right), \]
where \( x > 0 \), together with the portion of the \( y \)-axis, for which \(-1 \leq y \leq 1\), is connected, but not arcwise connected. See Figure 15.25.

A trivial modification of the proof of Theorem 15.23 shows that in a normed vector space, \( E \), a connected open set is arcwise connected by polygonal lines (i.e., arcs consisting of line segments). This is because in every open ball, any two points are connected by a line segment. Furthermore, if \( E \) is finite dimensional, these polygonal lines can be forced to be parallel to basis vectors.

We now consider compactness.

15.5 Compact Sets and Locally Compact Spaces

The property of compactness is very important in topology and analysis. We provide a quick review geared towards the study of manifolds, and for details, we refer the reader to Munkres [62], Schwartz [70]. In this section we will need to assume that the topological spaces are Hausdorff spaces. This is not a luxury, as many of the results are false otherwise.

We begin this section by providing the definition of compactness and describing a collection of compact spaces in \( \mathbb{R} \). There are various equivalent ways of defining compactness. For our purposes, the most convenient way involves the notion of open cover.

**Definition 15.28.** Given a topological space \( E \), for any subset \( A \) of \( E \), an open cover \((U_i)_{i \in I}\) of \( A \) is a family of open subsets of \( E \) such that \( A \subseteq \bigcup_{i \in I} U_i \). An open subcover of an open cover \((U_i)_{i \in I}\) of \( A \) is any subfamily \((U_j)_{j \in J}\) which is an open cover of \( A \), with \( J \subseteq I \). An open cover \((U_i)_{i \in I}\) of \( A \) is finite if \( I \) is finite. See Figure 15.28. The topological space \( E \) is compact if it is Hausdorff and for every open cover \((U_i)_{i \in I}\) of \( E \), there is a finite open subcover \((U_j)_{j \in J}\) of \( E \). Given any subset \( A \) of \( E \), we say that \( A \) is compact if it is compact with respect to the subspace topology. We say that \( A \) is relatively compact if its closure \( \overline{A} \) is compact.

It is immediately verified that a subset \( A \) of \( E \) is compact in the subspace topology relative to \( A \) iff for every open cover \((U_i)_{i \in I}\) of \( A \) by open subsets of \( E \), there is a finite open subcover \((U_j)_{j \in J}\) of \( A \). The property that every open cover contains a finite open subcover is often called the Heine-Borel-Lebesgue property. By considering complements, a Hausdorff space is compact iff for every family \((F_i)_{i \in I}\) of closed sets, if \( \bigcap_{i \in I} F_i = \emptyset \), then \( \bigcap_{j \in J} F_j = \emptyset \) for some finite subset \( J \) of \( I \).

Definition 15.28 requires that a compact space be Hausdorff. There are books in which a compact space is not necessarily required to be Hausdorff. Following Schwartz, we prefer calling such a space quasi-compact.
A topological Hausdorff space $E$ is compact iff for every family $(F_i)_{i \in I}$ of closed sets having the finite intersection property, then $\bigcap_{i \in I} F_i \neq \emptyset$.

**Proof.** If $E$ is compact and $(F_i)_{i \in I}$ is a family of closed sets having the finite intersection property, then $\bigcap_{i \in I} F_i$ cannot be empty, since otherwise we would have $\bigcap_{j \in J} F_j = \emptyset$ for some finite subset $J$ of $I$, a contradiction. The converse is equally obvious. \qed

Another useful consequence of compactness is as follows. For any family $(F_i)_{i \in I}$ of closed sets such that $F_{i+1} \subseteq F_i$ for all $i \in I$, if $\bigcap_{i \in I} F_i = \emptyset$, then $F_i = \emptyset$ for some $i \in I$. Indeed, there must be some finite subset $J$ of $I$ such that $\bigcap_{j \in J} F_j = \emptyset$, and since $F_{i+1} \subseteq F_i$ for all $i \in I$, we must have $F_j = \emptyset$ for the smallest $F_j$ in $(F_j)_{j \in I}$. Using this fact, we note that $\mathbb{R}$ is not compact. Indeed, the family of closed sets, $([n, +\infty))_{n \geq 0}$, is decreasing and has an empty intersection.

It is immediately verified that every finite union of compact subsets is compact. Similarly, every finite union of relatively compact subsets is relatively compact (use the fact that $\overline{A \cup B} = \overline{A} \cap \overline{B}$).

---

**Figure 15.28:** An open cover of $S^2$ using two open sets induced by the Euclidean topology of $\mathbb{R}^3$. 

Another equivalent and useful characterization can be given in terms of families having the finite intersection property.

**Definition 15.29.** A family $(F_i)_{i \in I}$ of sets has the finite intersection property if $\bigcap_{j \in J} F_j \neq \emptyset$ for every finite subset $J$ of $I$.

**Proposition 15.24.**
Given a metric space, if we define a *bounded subset* to be a subset that can be enclosed in some closed ball (of finite radius), then any nonbounded subset of a metric space is not compact. However, a closed interval \([a, b]\) of the real line is compact.

**Proposition 15.25.** Every closed interval, \([a, b]\), of the real line is compact.

**Proof.** We proceed by contradiction. Let \((U_i)_{i \in I}\) be any open cover of \([a, b]\) and assume that there is no finite open subcover. Let \(c = (a + b)/2\). If both \([a, c]\) and \([c, b]\) had some finite open subcover, so would \([a, b]\), and thus, either \([a, c]\) does not have any finite subcover, or \([c, b]\) does not have any finite open subcover. Let \([a_1, b_1]\) be such a bad subinterval. The same argument applies and we split \([a_1, b_1]\) into two equal subintervals, one of which must be bad. Thus, having defined \([a_n, b_n]\) of length \((b - a)/2^n\) as an interval having no finite open subcover, splitting \([a_n, b_n]\) into two equal intervals, we know that at least one of the two has no finite open subcover and we denote such a bad interval by \([a_{n+1}, b_{n+1}]\). See Figure 15.29. The sequence \((a_n)\) is nondecreasing and bounded from above by \(b\), and thus, by a fundamental property of the real line, it converges to its least upper bound, \(\alpha\). Similarly, the sequence \((b_n)\) is nonincreasing and bounded from below by \(a\) and thus, it converges to its greatest lowest bound, \(\beta\). However, the common limit \(\alpha = \beta\) of the sequences \((a_n)\) and \((b_n)\) must belong to some open set, \(U_i\), of the open cover and since \(U_i\) is open, it must contain some interval \([c, d]\) containing \(\alpha\). Then, because \(\alpha\) is the common limit of the sequences \((a_n)\) and \((b_n)\), there is some \(N\) such that the intervals \([a_n, b_n]\) are all contained in the interval \([c, d]\) for all \(n \geq N\), which contradicts the fact that none of the intervals \([a_n, b_n]\) has a finite open subcover. Thus, \([a, b]\) is indeed compact. 

The argument of Proposition 15.25 can be adapted to show that in \(\mathbb{R}^m\), every closed set, \([a_1, b_1] \times \cdots \times [a_m, b_m]\), is compact. At every stage, we need to divide into \(2^m\) subpieces instead of 2.

We next discuss some important properties of compact spaces. We begin with two separations axioms which only hold for Hausdorff spaces:

**Proposition 15.26.** Given a topological Hausdorff space, \(E\), for every compact subset, \(A\), and every point, \(b\), not in \(A\), there exist disjoint open sets, \(U\) and \(V\), such that \(A \subseteq U\) and \(b \in V\). See Figure 15.30. As a consequence, every compact subset is closed.

**Proof.** Since \(E\) is Hausdorff, for every \(a \in A\), there are some disjoint open sets, \(U_a\) and \(V_a\), containing \(a\) and \(b\) respectively. Thus, the family, \((U_a)_{a \in A}\), forms an open cover of \(A\). Since \(A\) is compact there is a finite open subcover, \((U_j)_{j \in J}\), of \(A\), where \(J \subseteq A\), and then \(\bigcup_{j \in J} U_j\) is an open set containing \(A\) disjoint from the open set \(\bigcap_{j \in J} V_j\) containing \(b\). This shows that every point, \(b\), in the complement of \(A\) belongs to some open set in this complement and thus, that the complement is open, i.e., that \(A\) is closed. See Figure 15.31.

Actually, the proof of Proposition 15.26 can be used to show the following useful property:
CHAPTER 15. TOPOLOGY

Proposition 15.27. Given a topological Hausdorff space $E$, for every pair of compact disjoint subsets $A$ and $B$, there exist disjoint open sets $U$ and $V$, such that $A \subseteq U$ and $B \subseteq V$.

Proof. We repeat the argument of Proposition 15.26 with $B$ playing the role of $b$ and use Proposition 15.26 to find disjoint open sets $U_a$ containing $a \in A$, and $V_a$ containing $B$. □

The following proposition shows that in a compact topological space, every closed set is compact:

Proposition 15.28. Given a compact topological space $E$, every closed set is compact.

Proof. Since $A$ is closed, $E - A$ is open and from any open cover, $(U_i)_{i \in I}$, of $A$, we can form an open cover of $E$ by adding $E - A$ to $(U_i)_{i \in I}$ and, since $E$ is compact, a finite subcover, $(U_j)_{j \in J} \cup \{E - A\}$, of $E$ can be extracted such that $(U_j)_{j \in J}$ is a finite subcover of $A$. See Figure 15.32. □

Remark: Proposition 15.28 also holds for quasi-compact spaces, i.e., the Hausdorff separation property is not needed.

Putting Proposition 15.27 and Proposition 15.28 together, we note that if $X$ is compact, then for every pair of disjoint closed sets $A$ and $B$, there exist disjoint open sets $U$ and $V$ such that $A \subseteq U$ and $B \subseteq V$. 

Figure 15.29: The first four stages of the nested interval construction utilized in the proof of Proposition 15.25.
15.5. COMPACT SETS AND LOCALLY COMPACT SPACES

Figure 15.30: The compact set of $\mathbb{R}^2$, $A$, is separated by any point in its complement.

Figure 15.31: For the pink compact set $A$, $U$ is the union of the seven disks which cover $A$, while $V$ is the intersection of the seven open sets containing $b$.

**Definition 15.30.** A topological space $E$ is *normal* if every one-point set is closed, and for every pair of disjoint closed sets $A$ and $B$, there exist disjoint open sets $U$ and $V$ such that $A \subseteq U$ and $B \subseteq V$. A topological space $E$ is *regular* if every one-point set is closed, and for every point $a \in E$ and every closed subset $B$ of $E$, if $a \notin B$, then there exist disjoint open sets $U$ and $V$ such that $a \in U$ and $B \subseteq V$.

It is clear that a normal space is regular, and a regular space is Hausdorff. There are examples of Hausdorff spaces that are not regular, and of regular spaces that are not normal.

We just observed that a compact space is normal, and this is worth recording as a proposition.

**Proposition 15.29.** Every (Hausdorff) compact space is normal.

An important property of metrizable spaces is that they are normal.

**Proposition 15.30.** Every metrizable space $E$ is normal.
Figure 15.32: An illustration of the proof of Proposition 15.28. Both $E$ and $A$ are closed squares in $\mathbb{R}^2$. Note that an open cover of $A$, namely the green circles, when combined with the yellow square annulus $E - A$ covers all of the yellow square $E$.

Proof. Assume the topology of $E$ is given by the metric $d$. Since $B$ is closed and $A \cap B = \emptyset$, for every $a \in A$ since $a \notin B = B$, there is some open ball $B_0(a, \epsilon_a)$ of radius $\epsilon_a > 0$ such that $B_0(a, \epsilon_a) \cap B = \emptyset$. Similarly, since $A$ is closed and $A \cap B = \emptyset$, for every $b \in B$ there is some open ball $B_0(b, \epsilon_b)$ of radius $\epsilon_b > 0$ such that $B_0(b, \epsilon_b) \cap A = \emptyset$. Let

$$U = \bigcup_{a \in A} B_0(a, \epsilon_a/2), \quad V = \bigcup_{b \in B} B_0(b, \epsilon_b/2).$$

Then $A$ and $B$ are open sets such that $A \subseteq U$ and $B \subseteq V$, and we claim that $U \cap V = \emptyset$.

If not, then there is some $z \in U \cap V$, which implies that for some $a \in A$ and some $b \in B$, we have

$$z \in B_0(a, \epsilon_a/2) \cap B_0(b, \epsilon_b/2).$$

It follows that

$$d(a, b) \leq d(a, z) + d(z, b) < (\epsilon_a + \epsilon_b)/2.$$ 

If $\epsilon_a \leq \epsilon_b$, then $d(a, b) < \epsilon_b$, so $a \in B_0(b, \epsilon_b)$, contradicting the fact that $B_0(b, \epsilon_b) \cap A = \emptyset$. If $\epsilon_b \leq \epsilon_a$, then $d(a, b) < \epsilon_a$, so $b \in B_0(a, \epsilon_a)$, contradicting the fact that $B_0(a, \epsilon_a) \cap B = \emptyset$. □
Normal spaces have a strong separation property regarding disjoint closed subsets $A$ and $B$. Actually, this separation property can be stated as the existence of a certain continuous function $f: E \to [0,1]$ taking the value 1 on $A$ and the value 0 on $B$. This result is known as Urysohn Lemma. It is an important tool in topology and analysis.

**Theorem 15.31.** (Urysohn Lemma) Let $E$ be a normal space. For any two closed disjoint subsets $A$ and $B$, there is a continuous function $f: E \to [0,1]$ such that $f(x) = 1$ for all $x \in A$ and $f(x) = 0$ for all $x \in B$.

A proof of Theorem 15.31 can be found in Munkres [62] (Chapter 4, Section 33, Theorem 33.1). Theorem 15.31 is one of the ingredients in the Urysohn metrization theorem (Theorem 15.44).

Compact spaces also have the following property.

**Proposition 15.32.** Given a compact topological space, $E$, for every $a \in E$, for every neighborhood, $V$, of $a$, there exists a compact neighborhood, $U$, of $a$ such that $U \subseteq V$. See Figure 15.33.

![Figure 15.33: Let $E$ be the peach square of $\mathbb{R}^2$. Each point of $E$ is contained in a compact neighborhood $U$, in this case the small closed yellow disk.](image)

**Proof.** Since $V$ is a neighborhood of $a$, there is some open subset, $O$, of $V$ containing $a$. Then the complement, $K = E - O$, of $O$ is closed and since $E$ is compact, by Proposition 15.28, $K$ is compact. Now, if we consider the family of all closed sets of the form, $K \cap F$, where $F$ is any closed neighborhood of $a$, since $a \notin K$, this family has an empty intersection and thus, there is a finite number of closed neighborhood, $F_1, \ldots, F_n$, of $a$, such that $K \cap F_1 \cap \cdots \cap F_n = \emptyset$. Then, $U = F_1 \cap \cdots \cap F_n$ is closed and hence by Proposition 15.28, a compact neighborhood of $a$ contained in $O \subseteq V$. See Figure 15.34. □
Figure 15.34: Let $E$ be the peach square of $\mathbb{R}^2$. The compact neighborhood of $a, U$, is the intersection of the closed sets $F_1, F_2, F_3$, each of which are contained in the complement of $K$.

It can be shown that in a normed vector space of finite dimension, a subset is compact iff it is closed and bounded. For $\mathbb{R}^n$ the proof is simple.

In a normed vector space of infinite dimension, there are closed and bounded sets that are not compact!

More could be said about compactness in metric spaces but we will only need the notion of Lebesgue number, which will be discussed a little later. Another crucial property of compactness is that it is preserved under continuity.

**Proposition 15.33.** Let $E$ be a topological space and let $F$ be a topological Hausdorff space. For every compact subset, $A$, of $E$, for every continuous map, $f: E \rightarrow F$, the subspace $f(A)$ is compact.

**Proof.** Let $(U_i)_{i \in I}$ be an open cover of $f(A)$. We claim that $(f^{-1}(U_i))_{i \in I}$ is an open cover of $A$, which is easily checked. Since $A$ is compact, there is a finite open subcover, $(f^{-1}(U_j))_{j \in J}$, of $A$, and thus, $(U_j)_{j \in J}$ is an open subcover of $f(A)$.
As a corollary of Proposition 15.33, if $E$ is compact, $F$ is Hausdorff, and $f: E \to F$ is continuous and bijective, then $f$ is a homeomorphism. Indeed, it is enough to show that $f^{-1}$ is continuous, which is equivalent to showing that $f$ maps closed sets to closed sets. However, closed sets are compact and Proposition 15.33 shows that compact sets are mapped to compact sets, which, by Proposition 15.26, are closed.

Another important corollary of Proposition 15.33 is the following result.

**Proposition 15.34.** If $E$ is a compact nonempty topological space and if $f: E \to \mathbb{R}$ is a continuous function, then there are points $a,b \in E$ such that $f(a)$ is the minimum of $f(E)$ and $f(b)$ is the maximum of $f(E)$.

**Proof.** The set $f(E)$ is a compact subset of $\mathbb{R}$ and thus, a closed and bounded set which contains its greatest lower bound and its least upper bound.

The following property also holds.

**Proposition 15.35.** Let $(E,d)$ be a metric space. For any nonempty subset $A$ of $E$, if $A$ is compact, then for every open subset $U$ such that $A \subseteq U$, there is some $r > 0$ such that $V_r(A) \subseteq U$.

**Proof.** The function $x \mapsto d(x,E-U)$ is continuous and $d(x,E-U) > 0$ for $x \in A$ (since $A \subseteq U$). By Proposition 15.34, there is some $a \in A$ such that

$$d(a,E-U) = \inf_{x \in A} d(x,E-U).$$

But $d(a,E-U) = r > 0$, which implies that $V_r(A) \subseteq U$.

Another useful notion is that of local compactness. Indeed manifolds and surfaces are locally compact.

**Definition 15.31.** A topological space $E$ is locally compact if it is Hausdorff and for every $a \in E$, there is some compact neighborhood $K$ of $a$. See Figure 15.33.

From Proposition 15.32, every compact space is locally compact but the converse is false. For example, $\mathbb{R}$ is locally compact but not compact. In fact it can be shown that a normed vector space of finite dimension is locally compact.

**Proposition 15.36.** Given a locally compact topological space, $E$, for every $a \in E$, for every neighborhood, $N$, of $a$, there exists a compact neighborhood, $U$, of $a$, such that $U \subseteq N$.

**Proof.** For any $a \in E$, there is some compact neighborhood, $V$, of $a$. By Proposition 15.32, every neighborhood of $a$ relative to $V$ contains some compact neighborhood $U$ of $a$ relative to $V$. But every neighborhood of $a$ relative to $V$ is a neighborhood of $a$ relative to $E$ and every neighborhood $N$ of $a$ in $E$ yields a neighborhood, $V \cap N$, of $a$ in $V$ and thus, for every neighborhood, $N$, of $a$, there exists a compact neighborhood, $U$, of $a$ such that $U \subseteq N$. 

When $E$ is a metric space, the subsets $V_r(A)$ defined in Definition 15.6 have the following property.

**Proposition 15.37.** Let $(E,d)$ be a metric space. If $E$ is locally compact, then for any nonempty compact subset $A$ of $E$, there is some $r > 0$ such that $V_r(A)$ is compact.

**Proof.** Since $E$ is locally compact, for every $x \in A$, there is some compact subset $V_x$ whose interior $\overset{\circ}{V_x}$ contains $x$. The family of open subsets $\overset{\circ}{V_x}$ is an open cover $A$, and since $A$ is compact, it has a finite subcover $\{\overset{\circ}{V_{x_1}}, \ldots, \overset{\circ}{V_{x_n}}\}$. Then $U = \overset{\circ}{V_{x_1}} \cup \cdots \cup \overset{\circ}{V_{x_n}}$ is compact (as a finite union of compact subsets), and it contains an open subset containing $A$ (the union of the $\overset{\circ}{V_{x_i}}$). By Proposition 15.35, there is some $r > 0$ such that $V_r(A) \subseteq U$, and thus $\overline{V_r(A)} \subseteq U$. Since $U$ is compact and $\overline{V_r(A)}$ is closed, $\overline{V_r(A)}$ is compact. □

Another very important property of locally compact spaces is the Proposition 15.39 below. This result implies the existence of continuous partitions of unity for a finite open cover of a compact subset. Such partitions of unity are used in proving that Radon functionals correspond to certain Borel measures. First we have the following proposition.

**Proposition 15.38.** Let $E$ be a locally compact (Hausdorff) space. For every compact subset $K$ and every open subset $V$, if $K \subseteq V$, then there is an open set $W$ with compact closure such that $K \subseteq W \subseteq \overline{W} \subseteq V$.

A proof of Proposition 15.38 can be found in Rudin [66] (Chapter 2, Theorem 2.7). The following proposition shows the existence of continuous “bump functions” in a locally compact space. It is sometimes called Urysohn lemma (which is a bit confusing since there is already a Urysohn lemma (Proposition 15.31).

**Proposition 15.39.** Let $E$ be a locally compact (Hausdorff) space. For every compact subset $K$ and every open subset $V$ of $E$, if $K \subseteq V$, there is a continuous function $f: E \to [0,1]$ such that $f(x) = 1$ for all $x \in K$, and such that $\text{supp}(f)$ is compact and $\text{supp}(f) \subseteq V$, where $\text{supp}(f)$ is the closure of the subset $\{x \in E \mid f(x) \neq 0\}$, called the support of $f$.

**Proof.** Theorem 15.39 follows easily from the Urysohn lemma (Theorem 15.31). Since $E$ is locally compact, by Proposition 15.38 we can find some open subset $W$ with compact closure $\overline{W}$ such that $K \subseteq W \subseteq \overline{W} \subseteq V$. Since $\overline{W}$ is compact, it is normal, so we can apply Theorem 15.31 to find a continuous function $f: \overline{W} \to [0,1]$ such that $f(x) = 1$ for all $x \in K$ and $f(x) = 0$ for all $x \in \overline{W} - W$ (the boundary of $W$). Then we extend $f$ to $E$ by setting to 0 outside $W$. Since the support of $f$ is contained in $W$, this function is continuous. □

As a corollary of Proposition 15.39 we obtain the existence of continuous partitions of unity for a finite open cover of a compact subset.
Proposition 15.40. Let $E$ be a locally compact (Hausdorff) space. For any compact subset $K$ of $E$ and any finite open cover $(U_1, \ldots, U_n)$ of $K$ (that is, $K \subseteq \bigcup_{i=1}^n U_i$), there exist $n$ continuous functions $f_i : E \to [0, 1]$ such that $f_i$ has compact support $\text{supp}(f_i) \subseteq U_i$, and

$$\sum_{i=1}^n f_i(x) = 1 \quad \text{for all } x \in K.$$ 

A proof of Proposition 15.40 is not difficult. It can be found in Rudin [66] (Chapter 2, Theorem 2.13) and Lang [52] (Chapter IX, §2). A family $(f_1, \ldots, f_n)$ satisfying the properties of Proposition 15.40 is called a partition of unity on $K$ subordinate to the cover $(U_1, \ldots, U_n)$.

It is much harder to deal with noncompact manifolds than it is to deal with compact manifolds. However, manifolds are locally compact and it turns out that there are various ways of embedding a locally compact Hausdorff space into a compact Hausdorff space. The most economical construction consists in adding just one point. This construction, known as the Alexandroff compactification, is technically useful, and we now describe it and sketch the proof that it achieves its goal.

To help the reader’s intuition, let us consider the case of the plane, $\mathbb{R}^2$. If we view the plane, $\mathbb{R}^2$, as embedded in 3-space, $\mathbb{R}^3$, say as the $xy$ plane of equation $z = 0$, we can consider the sphere, $\Sigma$, of radius 1 centered on the $z$-axis at the point $(0, 0, 1)$ and tangent to the $xOy$ plane at the origin (sphere of equation $x^2 + y^2 + (z - 1)^2 = 1$). If $N$ denotes the north pole on the sphere, i.e., the point of coordinates $(0, 0, 2)$, then any line, $D$, passing through the north pole and not tangent to the sphere (i.e., not parallel to the $xOy$ plane) intersects the $xOy$ plane in a unique point, $M$, and the sphere in a unique point, $P$, other than the north pole, $N$. This, way, we obtain a bijection between the $xOy$ plane and the punctured sphere $\Sigma$, i.e., the sphere with the north pole $N$ deleted. This bijection is called a stereographic projection. See Figure 15.35.

The Alexandroff compactification of the plane puts the north pole back on the sphere, which amounts to adding a single point at infinity $\infty$ to the plane. Intuitively, as we travel away from the origin $O$ towards infinity (in any direction!), we tend towards an ideal point at infinity $\infty$. Imagine that we “bend” the plane so that it gets wrapped around the sphere, according to stereographic projection. See Figure 15.36. A simpler example takes a line and gets a circle as its compactification. The Alexandroff compactification is a generalization of these simple constructions.

Definition 15.32. Let $(E, \mathcal{O})$ be a locally compact space. Let $\omega$ be any point not in $E$, and let $E_\omega = E \cup \{\omega\}$. Define the family, $\mathcal{O}_\omega$, as follows:

$$\mathcal{O}_\omega = \mathcal{O} \cup \{(E - K) \cup \{\omega\} \mid K \text{ compact in } E\}.$$ 

The pair, $(E_\omega, \mathcal{O}_\omega)$, is called the Alexandroff compactification (or one point compactification) of $(E, \mathcal{O})$. See Figure 15.37.
Figure 15.35: The stereographic projections of $x^2 + y^2 + (z - 1)^2 = 1$ onto the $xy$-plane.

The following theorem shows that $(E_\omega, \mathcal{O}_\omega)$ is indeed a topological space, and that it is compact.

**Theorem 15.41.** Let $E$ be a locally compact topological space. The Alexandroff compactification, $E_\omega$, of $E$ is a compact space such that $E$ is a subspace of $E_\omega$ and if $E$ is not compact, then $\overline{E} = E_\omega$.

**Proof.** The verification that $\mathcal{O}_\omega$ is a family of open sets is not difficult but a bit tedious. Details can be found in Munkres [62] or Schwartz [70]. Let us show that $E_\omega$ is compact. For every open cover, $(U_i)_{i \in I}$, of $E_\omega$, since $\omega$ must be covered, there is some $U_{i_0}$ of the form

$$U_{i_0} = (E - K_0) \cup \{\omega\}$$

where $K_0$ is compact in $E$. Consider the family, $(V_i)_{i \in I}$, defined as follows:

$$V_i = U_i \quad \text{if} \quad U_i \in \mathcal{O},$$

$$V_i = E - K \quad \text{if} \quad U_i = (E - K) \cup \{\omega\},$$

where $K$ is compact in $E$. Then, because each $K$ is compact and thus closed in $E$ (since $E$ is Hausdorff), $E - K$ is open, and every $V_i$ is an open subset of $E$. Furthermore, the family, $(V_i)_{i \in I \setminus \{i_0\}}$, is an open cover of $K_0$. Since $K_0$ is compact, there is a finite open subcover, $(V_j)_{j \in J}$, of $K_0$, and thus, $(U_j)_{j \in J \cup \{i_0\}}$ is a finite open cover of $E_\omega$.

Let us show that $E_\omega$ is Hausdorff. Given any two points, $a, b \in E_\omega$, if both $a, b \in E$, since $E$ is Hausdorff and every open set in $\mathcal{O}$ is an open set in $\mathcal{O}_\omega$, there exist disjoint open sets, $U, V$ (in $\mathcal{O}$), such that $a \in U$ and $b \in V$. If $b = \omega$, since $E$ is locally compact, there is some compact set, $K$, containing an open set, $U$, containing $a$ and then, $U$ and $V = (E - K) \cup \{\omega\}$ are disjoint open sets (in $\mathcal{O}_\omega$) such that $a \in U$ and $b \in V$. 
The space $E$ is a subspace of $E_\omega$ because for every open set, $U$, in $O_\omega$, either $U \in O$ and $E \cap U = U$ is open in $E$, or $U = (E - K) \cup \{\omega\}$, where $K$ is compact in $E$, and thus, $U \cap E = E - K$, which is open in $E$, since $K$ is compact in $E$ and thus, closed (since $E$ is Hausdorff). Finally, if $E$ is not compact, for every compact subset, $K$, of $E$, $E - K$ is nonempty and thus, for every open set, $U = (E-K) \cup \{\omega\}$, containing $\omega$, we have $U \cap E \neq \emptyset$, which shows that $\omega \in \overline{E}$ and thus, that $\overline{E} = E_\omega$.

### 15.6 Second-Countable and Separable Spaces

In studying surfaces and manifolds, an important property is the existence of a countable basis for the topology. Indeed this property, among other things, guarantees the existence of triangulations of manifolds, and the fact that a manifold is metrizable.

**Definition 15.33.** A topological space $E$ is called *second-countable* if there is a countable basis for its topology, i.e., if there is a countable family, $(U_i)_{i \geq 0}$, of open sets such that every open set of $E$ is a union of open sets $U_i$.

It is easily seen that $\mathbb{R}^n$ is second-countable and more generally, that every normed vector space of finite dimension is second-countable. More generally, a metric space is second-countable if and only if it is separable, a very useful property that holds for all of the spaces that we will consider in practice.
Figure 15.37: The two types of open sets associated with the Alexandroff compactification of the \( xy \)-plane. The first type of open set does not include \( \omega \), i.e. the north pole, while the second type of open set contains \( \omega \).

**Definition 15.34.** A topological space \( E \) is *separable* if it contains some countable subset \( S \) which is dense in \( X \), that is, \( \overline{S} = E \).

Observe that by Proposition 15.4, a subset \( S \) of \( E \) is dense in \( E \) iff every nonempty open subset of \( E \) contains some element of \( S \).

The (metric) space \( \mathbb{R} \) is separable because \( \mathbb{Q} \) is a countable dense subset of \( \mathbb{R} \). Similarly, \( \mathbb{C} \) is separable. In general, \( \mathbb{Q}^n \) is dense in \( \mathbb{R}^n \), so \( \mathbb{R}^n \) is separable, and similarly, every finite-dimensional normed vector space over \( \mathbb{R} \) (or \( \mathbb{C} \)) is separable. For metric spaces, we have the following useful result.

**Proposition 15.42.** If \( E \) is a metric space, then \( E \) is second-countable if and only if \( E \) is separable.

**Proof.** If \( \mathcal{B} = (B_n) \) is a countable basis for the topology of \( E \), then for any set \( S \) obtained by picking some point \( s_n \) in \( B_n \), since every nonempty open subset \( U \) of \( E \) is the union of some of the \( B_n \), the intersection \( U \cap S \) is nonempty, and so \( S \) is dense in \( E \).

Conversely, assume that there is a countable subset \( S = (s_n) \) of \( E \) which is dense in \( E \). We claim that the countable family \( \mathcal{B} \) of open balls \( B_0(s_n, 1/m) \) \( (m \in \mathbb{N}, m > 0) \) is a basis for the topology of \( E \). For every \( x \in E \) and every \( r > 0 \), there is some \( m > 0 \) such that
1/m < r/2, and some n such that \( s_n \in B_0(x, 1/m) \). It follows that \( x \in B_0(s_n, 1/m) \). For all \( y \in B_0(s_n, 1/m) \), we have
\[
d(x, y) \leq d(x, s_n) + d(s_n, y) \leq 2/m < r,
\]
thus \( B_0(s_n, 1/m) \subseteq B_0(x, r) \), which by Proposition 15.8(a) implies that \( \mathcal{B} \) is a basis for the topology of \( E \).

**Proposition 15.43.** If \( E \) is a compact metric space, then \( E \) is separable.

*Proof.* For every \( n > 0 \), the family of open balls of radius \( 1/n \) forms an open cover of \( E \), and since \( E \) is compact, there is a finite subset \( A_n \) of \( E \) such that \( E = \bigcup_{a_i \in A_n} B_0(a_i, 1/n) \).

It is easy to see that this is equivalent to the condition \( d(x, A_n) < 1/n \) for all \( x \in E \). Let \( A = \bigcup_{n \geq 1} A_n \). Then \( A \) is countable, and for every \( x \in E \), we have
\[
d(x, A) \leq d(x, A_n) < \frac{1}{n}, \text{ for all } n \geq 1,
\]
which implies that \( d(x, A) = 0 \); that is, \( A \) is dense in \( E \).

The following theorem due to Urysohn gives a very useful sufficient condition for a topological space to be metrizable.

**Theorem 15.44.** *(Urysohn metrization theorem)* If a topological space \( E \) is regular and second-countable, then it is metrizable.

The proof of Theorem 15.44 can be found in Munkres [62] (Chapter 4, Theorem 34.1). As a corollary of Theorem 15.44, every (second-countable) manifold, and thus every Lie group, is metrizable.

The following technical result shows that a locally compact metrizable space which is also separable can be expressed as the union of a countable monotonic sequence of compact subsets. This gives us a method for generalizing various properties of compact metric spaces to locally compact metric spaces of the above kind.

**Proposition 15.45.** Let \( E \) be a locally compact metrizable space. The following properties are equivalent:

1. There is a sequence \( (U_n)_{n \geq 0} \) of open subsets such that for all \( n \in \mathbb{N} \), \( U_n \subseteq U_{n+1} \), \( \overline{U_n} \) is compact, \( \overline{U_n} \subseteq U_{n+1} \), and \( E = \bigcup_{n \geq 0} U_n = \bigcup_{n \geq 0} \overline{U_n} \).
2. The space \( E \) is the union of a countable family of compact subsets of \( E \).
3. The space \( E \) is separable.
Proof. We show (1) implies (2), (2) implies (3), and (3) implies (1). Obviously, (1) implies (2) since the $U_n$ are compact.

If (2) holds, then $E = \bigcup_{n \geq 0} K_n$, for some compact subsets $K_n$. By Proposition 15.43, each compact subset $K_n$ is separable, so let $S_n$ be a countable dense subset of $K_n$. Then $S = \bigcup_{n \geq 0} S_n$ is a countable dense subset of $E$, since

$$E = \bigcup_{n \geq 0} K_n \subseteq \bigcup_{n \geq 0} S_n \subseteq S \subseteq E.$$  

Consequently (3) holds.

If (3) holds, let $S = \{s_n\}$ be a countable dense subset of $E$. By Proposition 15.42, the space $E$ has a countable basis $B$ of open sets $O_n$. Since $E$ is locally compact, for every $x \in E$, there is some compact neighborhood $W_x$ containing $x$, and by Proposition 15.8, there some index $n(x)$ such that $x \in O_{n(x)} \subseteq W_x$. Since $W_x$ is a compact neighborhood, we deduce that $O_{n(x)}$ is compact. Consequently, there is a subfamily of $B$ consisting of open subsets $O_i$ such that $O_i$ is compact, which is a countable basis for the topology of $E$, so we may assume that we restrict our attention to this basis. We define the sequence $(U_n)_{n \geq 1}$ of open subsets of $E$ by induction as follows: Set $U_1 = O_1$, and let

$$U_{n+1} = O_{n+1} \cup V_r(U_n),$$

where $r > 0$ is chosen so that $V_r(U_n)$ is compact, which is possible by Proposition 15.37. We immediately check that the $U_n$ satisfy (1) of Proposition 15.45. □

Definition 15.35. Given a topological space $E$, a subset $A$ of $E$ is $\sigma$-compact (or countable at infinity) if $A$ is the union of countably many compact subsets.

Note that Proposition 15.45 implies that a locally compact metrizable space is separable iff it is $\sigma$-compact.

It can also be shown that if $E$ is a locally compact space that has a countable basis, then $E_\omega$ also has a countable basis (and in fact, is metrizable).

We also have the following property.

Proposition 15.46. Given a second-countable topological space $E$, every open cover $(U_i)_{i \in I}$ of $E$ contains some countable subcover.

Proof. Let $(O_n)_{n \geq 0}$ be a countable basis for the topology. Then all sets $O_n$ contained in some $U_i$ can be arranged into a countable subsequence, $(\Omega_m)_{m \geq 0}$, of $(O_n)_{n \geq 0}$ and for every $\Omega_m$, there is some $U_{i_m}$ such that $\Omega_m \subseteq U_{i_m}$. Furthermore, every $U_i$ is some union of sets $\Omega_j$, and thus, every $a \in E$ belongs to some $\Omega_j$, which shows that $(\Omega_m)_{m \geq 0}$ is a countable open subcover of $(U_i)_{i \in I}$. □

As an immediate corollary of Proposition 15.46, a locally connected second-countable space has countably many connected components.
15.7 Sequential Compactness

For a general topological Hausdorff space $E$, the definition of compactness relies on the existence of finite cover. However, when $E$ has a countable basis or is a metric space, we may define the notion of compactness in terms of sequences. To understand how this is done, we need to first define accumulation points.

**Definition 15.36.** Given a topological Hausdorff space, $E$, given any sequence, $(x_n)$, of points in $E$, a point, $l \in E$, is an accumulation point (or cluster point) of the sequence $(x_n)$ if every open set, $U$, containing $l$ contains $x_n$ for infinitely many $n$. See Figure 15.38.

![Figure 15.38](image)

Figure 15.38: The space $E$ is the closed, bounded pink subset of $\mathbb{R}^2$. The sequence $(x_n)$ has two accumulation points, one for the subsequence $(x_{2n+1})$ and one for $(x_{2n})$.

Clearly, if $l$ is a limit of the sequence, $(x_n)$, then it is an accumulation point, since every open set, $U$, containing $a$ contains all $x_n$ except for finitely many $n$.

For second-countable spaces we are able to give another characterization of accumulation points.

**Proposition 15.47.** Given a second-countable topological Hausdorff space, $E$, a point, $l$, is an accumulation point of the sequence, $(x_n)$, iff $l$ is the limit of some subsequence, $(x_{n_k})$, of $(x_n)$.

**Proof.** Clearly, if $l$ is the limit of some subsequence $(x_{n_k})$ of $(x_n)$, it is an accumulation point of $(x_n)$.

Conversely, let $(U_k)_{k \geq 0}$ be the sequence of open sets containing $l$, where each $U_k$ belongs to a countable basis of $E$, and let $V_k = U_1 \cap \cdots \cap U_k$. For every $k \geq 1$, we can find some $n_k > n_{k-1}$ such that $x_{n_k} \in V_k$, since $l$ is an accumulation point of $(x_n)$. Now, since every open set containing $l$ contains some $U_{k_0}$ and since $x_{n_k} \in U_{k_0}$ for all $k \geq 0$, the sequence $(x_{n_k})$ has limit $l$. \qed
Remark: Proposition 15.47 also holds for metric spaces.

As an illustration of Proposition 15.47 let \((x_n)\) be the sequence \((1,-1,1,-1,\ldots)\). This sequence has two accumulation points, namely 1 and \(-1\) since \((x_{2n+1}) = (1)\) and \((x_{2n}) = (-1)\).

In second-countable Hausdorff spaces, compactness can be characterized in terms of accumulation points (this is also true for metric spaces).

**Proposition 15.48.** A second-countable topological Hausdorff space, \(E\), is compact iff every sequence, \((x_n)\), of \(E\) has some accumulation point in \(E\).

**Proof.** Assume that every sequence, \((x_n)\), has some accumulation point. Let \((U_i)_{i \in I}\) be some open cover of \(E\). By Proposition 15.46, there is a countable open subcover, \((O_n)_{n \geq 0}\), for \(E\). Now, if \(E\) is not covered by any finite subcover of \((O_n)_{n \geq 0}\), we can define a sequence, \((x_m)\), by induction as follows:

Let \(x_0\) be arbitrary and for every \(m \geq 1\), let \(x_m\) be some point in \(E\) not in \(O_1 \cup \cdots \cup O_m\), which exists, since \(O_1 \cup \cdots \cup O_m\) is not an open cover of \(E\). We claim that the sequence, \((x_m)\), does not have any accumulation point. Indeed, for every \(l \in E\), since \((O_n)_{n \geq 0}\) is an open cover of \(E\), there is some \(O_m\) such that \(l \in O_m\), and by construction, every \(x_n\) with \(n \geq m + 1\) does not belong to \(O_m\), which means that \(x_n \in O_m\) for only finitely many \(n\) and \(l\) is not an accumulation point. See Figure 15.39.

![Figure 15.39](image)

Figure 15.39: The space \(E\) is the open half plane above the line \(y = -1\). Since \(E\) is not compact, we inductively build a sequence, \((x_n)\) that will have no accumulation point in \(E\). Note the \(y\) coordinate of \(x_n\) approaches infinity.

Conversely, assume that \(E\) is compact, and let \((x_n)\) be any sequence. If \(l \in E\) is not an accumulation point of the sequence, then there is some open set, \(U_l\), such that \(l \in U_l\)
and \( x_n \in U_l \) for only finitely many \( n \). Thus, if \((x_n)\) does not have any accumulation point, the family, \((U_l)_{l \in E}\), is an open cover of \( E \) and since \( E \) is compact, it has some finite open subcover, \((U_l)_{l \in J}\), where \( J \) is a finite subset of \( E \). But every \( U_l \) with \( l \in J \) is such that \( x_n \in U_l \) for only finitely many \( n \), and since \( J \) is finite, \( x_n \in \bigcup_{l \in J} U_l \) for only finitely many \( n \), which contradicts the fact that \((U_l)_{l \in J}\) is an open cover of \( E \), and thus contains all the \( x_n \). Thus, \((x_n)\) has some accumulation point. See Figure 15.40.

Figure 15.40: The space \( E \) the closed triangular region of \( \mathbb{R}^2 \). Given a sequence \((x_n)\) of red points in \( E \), if the sequence has no accumulation points, then each \( l_i \) for \( 1 \leq i \leq 8 \), is not an accumulation point. But as implied by the illustration, \( l_8 \) actually is an accumulation point of \((x_n)\).

Remarks:

1. By combining Propositions 15.47 and 15.48, we have observe that a second-countable Hausdorff space \( E \) is compact iff every sequence \((x_n)\) has a convergent subsequence \((x_{n_k})\). In other words, we say a second-countable Hausdorff space \( E \) is compact iff it is \textit{sequentially compact}.

2. It should be noted that the proof showing that if \( E \) is compact, then every sequence has some accumulation point, holds for any arbitrary compact space (the proof does not use a countable basis for the topology). The converse also holds for metric spaces. We will prove this converse since it is a major property of metric spaces.

Given a metric space in which every sequence has some accumulation point, we first prove the existence of a \textit{Lebesgue number}. 
**Lemma 15.49.** Given a metric space, $E$, if every sequence, $(x_n)$, has an accumulation point, for every open cover, $(U_i)_{i \in I}$, of $E$, there is some $\delta > 0$ (a Lebesgue number for $(U_i)_{i \in I}$) such that, for every open ball, $B_0(a, \epsilon)$, of radius $\epsilon \leq \delta$, there is some open subset, $U_i$, such that $B_0(a, \epsilon) \subseteq U_i$. See Figure 15.41

![Figure 15.41](image)

**Proof.** If there was no $\delta$ with the above property, then, for every natural number, $n$, there would be some open ball, $B_0(a_n, 1/n)$, which is not contained in any open set, $U_i$, of the open cover, $(U_i)_{i \in I}$. However, the sequence, $(a_n)$, has some accumulation point, $a$, and since $(U_i)_{i \in I}$ is an open cover of $E$, there is some $U_i$ such that $a \in U_i$. Since $U_i$ is open, there is some open ball of center $a$ and radius $\epsilon$ contained in $U_i$. Now, since $a$ is an accumulation point of the sequence, $(a_n)$, every open set containing $a$ contains $a_n$ for infinitely many $n$ and thus, there is some $n$ large enough so that

$$1/n \leq \epsilon/2 \quad \text{and} \quad a_n \in B_0(a, \epsilon/2),$$

which implies that

$$B_0(a_n, 1/n) \subseteq B_0(a, \epsilon) \subseteq U_i,$$

a contradiction. \qed

By a previous remark, since the proof of Proposition 15.48 implies that in a compact topological space, every sequence has some accumulation point, by Lemma 15.49, in a compact metric space, every open cover has a Lebesgue number. This fact can be used to prove another important property of compact metric spaces, the uniform continuity theorem.
15.7. **SEQUENTIAL COMPACTNESS**

**Definition 15.37.** Given two metric spaces, \((E, d_E)\) and \((F, d_F)\), a function, \(f : E \to F\), is uniformly continuous if for every \(\epsilon > 0\), there is some \(\eta > 0\), such that, for all \(a, b \in E\),

\[
\text{if } d_E(a, b) \leq \eta \text{ then } d_F(f(a), f(b)) \leq \epsilon.
\]

See Figures 15.42 and 15.43.

![Figure 15.42](image)

**Figure 15.42:** The real valued function \(f(x) = \sqrt{x}\) is uniformly continuous over \((0, \infty)\). Fix \(\epsilon\). If the \(x\) values lie within the rose colored \(\eta\) strip, the \(y\) values always lie within the peach \(\epsilon\) strip.

As we saw earlier, the metric on a metric space is uniformly continuous, and the norm on a normed metric space is uniformly continuous.

The **uniform continuity theorem** can be stated as follows:

**Theorem 15.50.** Given two metric spaces, \((E, d_E)\) and \((F, d_F)\), if \(E\) is compact and if \(f : E \to F\) is a continuous function, then \(f\) is uniformly continuous.

**Proof.** Consider any \(\epsilon > 0\) and let \((B_0(y, \epsilon/2))_{y \in F}\) be the open cover of \(F\) consisting of open balls of radius \(\epsilon/2\). Since \(f\) is continuous, the family,

\[
(f^{-1}(B_0(y, \epsilon/2)))_{y \in F},
\]

is an open cover of \(E\). Since, \(E\) is compact, by Lemma 15.49, there is a Lebesgue number, \(\delta\), such that for every open ball, \(B_0(a, \eta)\), of radius \(\eta \leq \delta\), then \(B_0(a, \eta) \subseteq f^{-1}(B_0(y, \epsilon/2))\), for some \(y \in F\). In particular, for any \(a, b \in E\) such that \(d_E(a, b) \leq \eta = \delta/2\), we have \(a, b \in B_0(a, \delta)\) and thus, \(a, b \in f^{-1}(B_0(y, \epsilon/2))\), which implies that \(f(a), f(b) \in B_0(y, \epsilon/2)\). But then, \(d_F(f(a), f(b)) \leq \epsilon\), as desired. \(\square\)

We now prove another lemma needed to obtain the characterization of compactness in metric spaces in terms of accumulation points.
Figure 15.43: The real valued function \( f(x) = \frac{1}{x} \) is not uniformly continuous over \((0, \infty)\). Fix \( \epsilon \). In order for the \( y \) values to lie within the peach epsilon strip, the widths of the eta strips decrease as \( x \to 0 \).

Lemma 15.51. Given a metric space, \( E \), if every sequence, \((x_n)\), has an accumulation point, then for every \( \epsilon > 0 \), there is a finite open cover, \( B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon) \), of \( E \) by open balls of radius \( \epsilon \).

Proof. Let \( a_0 \) be any point in \( E \). If \( B_0(a_0, \epsilon) = E \), then the lemma is proved. Otherwise, assume that a sequence, \((a_0, a_1, \ldots, a_n)\), has been defined, such that \( B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon) \) does not cover \( E \). Then, there is some \( a_{n+1} \) not in \( B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon) \) and either

\[
B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_{n+1}, \epsilon) = E,
\]

in which case the lemma is proved, or we obtain a sequence, \((a_0, a_1, \ldots, a_{n+1})\), such that \( B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_{n+1}, \epsilon) \) does not cover \( E \). If this process goes on forever, we obtain an infinite sequence, \((a_n)\), such that \( d(a_m, a_n) > \epsilon \) for all \( m \neq n \). Since every sequence in \( E \) has some accumulation point, the sequence, \((a_n)\), has some accumulation point, \( a \). Then, for infinitely many \( n \), we must have \( d(a_n, a) \leq \epsilon/3 \) and thus, for at least two distinct natural numbers, \( p, q \), we must have \( d(a_p, a) \leq \epsilon/3 \) and \( d(a_q, a) \leq \epsilon/3 \), which implies \( d(a_p, a_q) \leq d(a_p, a) + d(a_q, a) \leq 2\epsilon/3 \), contradicting the fact that \( d(a_m, a_n) > \epsilon \) for all \( m \neq n \). See Figure 15.44. Thus, there must be some \( n \) such that

\[
B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon) = E.
\]

Definition 15.38. A metric space \( E \) is said to be precompact (or totally bounded) if for every \( \epsilon > 0 \), there is a finite open cover, \( B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon) \), of \( E \) by open balls of radius \( \epsilon \).

We now obtain the Weierstrass–Bolzano property.
Figure 15.44: Let $E$ be the peach region of $\mathbb{R}^2$. If $E$ is not covered by a finite collection of orange balls with radius $\epsilon$, the points of the sequence $(a_n)$ are separated by a distance of at least $\epsilon$. This contradicts the fact that $a$ is the accumulation point of $a$, as evidenced by the enlargement of the plum disk in Figure (ii).

**Theorem 15.52.** A metric space, $E$, is compact iff every sequence, $(x_n)$, has an accumulation point.

**Proof.** We already observed that the proof of Proposition 15.48 shows that for any compact space (not necessarily metric), every sequence, $(x_n)$, has an accumulation point. Conversely, let $E$ be a metric space, and assume that every sequence, $(x_n)$, has an accumulation point. Given any open cover, $(U_i)_{i \in I}$ for $E$, we must find a finite open subcover of $E$. By Lemma 15.49, there is some $\delta > 0$ (a Lebesgue number for $(U_i)_{i \in I}$) such that, for every open ball, $B_0(a, \epsilon)$, of radius $\epsilon \leq \delta$, there is some open subset, $U_j$, such that $B_0(a, \epsilon) \subseteq U_j$. By Lemma 15.51, for every $\delta > 0$, there is a finite open cover, $B_0(a_0, \delta) \cup \cdots \cup B_0(a_n, \delta)$, of $E$ by open balls of radius $\delta$. But from the previous statement, every open ball, $B_0(a_i, \delta)$, is contained in some open set, $U_{j_i}$, and thus, $\{U_{j_1}, \ldots, U_{j_n}\}$ is an open cover of $E$. \qed

### 15.8 Complete Metric Spaces and Compactness

Another very useful characterization of compact metric spaces is obtained in terms of Cauchy sequences. Such a characterization is quite useful in fractal geometry (and elsewhere). First
recall the definition of a Cauchy sequence and of a complete metric space.

**Definition 15.39.** Given a metric space, \((E,d)\), a sequence, \((x_n)_{n \in \mathbb{N}}\), in \(E\) is a *Cauchy sequence* if the following condition holds: for every \(\epsilon > 0\), there is some \(p \geq 0\), such that, for all \(m, n \geq p\), then \(d(x_m, x_n) \leq \epsilon\).

If every Cauchy sequence in \((E,d)\) converges we say that \((E,d)\) is a *complete metric space*.

First let us show the following proposition:

**Proposition 15.53.** Given a metric space, \(E\), if a Cauchy sequence, \((x_n)\), has some accumulation point, \(a\), then \(a\) is the limit of the sequence, \((x_n)\).

**Proof.** Since \((x_n)\) is a Cauchy sequence, for every \(\epsilon > 0\), there is some \(p \geq 0\), such that, for all \(m, n \geq p\), then \(d(x_m, x_n) \leq \epsilon/2\). Since \(a\) is an accumulation point for \((x_n)\), for infinitely many \(n\), we have \(d(x_n, a) \leq \epsilon/2\), and thus, for at least some \(n \geq p\), we have \(d(x_n, a) \leq \epsilon/2\). Then, for all \(m \geq p\),

\[
d(x_m, a) \leq d(x_m, x_n) + d(x_n, a) \leq \epsilon,
\]

which shows that \(a\) is the limit of the sequence \((x_n)\).

We can now prove the following theorem.

**Theorem 15.54.** A metric space, \(E\), is compact iff it is precompact and complete.

**Proof.** Let \(E\) be compact. For every \(\epsilon > 0\), the family of all open balls of radius \(\epsilon\) is an open cover for \(E\) and since \(E\) is compact, there is a finite subcover, \(B_0(a_0, \epsilon) \cup \cdots \cup B_0(a_n, \epsilon)\), of \(E\) by open balls of radius \(\epsilon\). Thus \(E\) is precompact. Since \(E\) is compact, by Theorem 15.52, every sequence, \((x_n)\), has some accumulation point. Thus every Cauchy sequence, \((x_n)\), has some accumulation point, \(a\), and, by Proposition 15.53, \(a\) is the limit of \((x_n)\). Thus, \(E\) is complete.

Now assume that \(E\) is precompact and complete. We prove that every sequence, \((x_n)\), has an accumulation point. By the other direction of Theorem 15.52, this shows that \(E\) is compact. Given any sequence, \((x_n)\), we construct a Cauchy subsequence, \((y_n)\), of \((x_n)\) as follows: Since \(E\) is precompact, letting \(\epsilon = 1\), there exists a finite cover, \(\mathcal{U}_1\), of \(E\) by open balls of radius \(1\). Thus some open ball, \(B_0^0\), in the cover, \(\mathcal{U}_1\), contains infinitely many elements from the sequence \((x_n)\). Let \(y_0\) be any element of \((x_n)\) in \(B_0^0\). By induction, assume that a sequence of open balls, \((B_i^0)_{1 \leq i \leq m}\), has been defined, such that every ball, \(B_i^0\), has radius \(\frac{1}{2^i}\), contains infinitely many elements from the sequence \((x_n)\) and contains some \(y_i\) from \((x_n)\) such that

\[
d(y_i, y_{i+1}) \leq \frac{1}{2^i},
\]

for all \(i, 0 \leq i \leq m - 1\). See Figure 15.45. Then letting \(\epsilon = \frac{1}{2^{m+1}}\), because \(E\) is precompact, there is some finite cover, \(\mathcal{U}_{m+1}\), of \(E\) by open balls of radius \(\epsilon\) and thus, of the open ball \(B_0^m\).
Thus, some open ball, $B^{m+1}_0$, in the cover, $U_{m+1}$, contains infinitely many elements from the sequence, $(x_n)$, and we let $y_{m+1}$ be any element of $(x_n)$ in $B^{m+1}_0$. Thus, we have defined by induction a sequence, $(y_n)$, which is a subsequence of, $(x_n)$, and such that

$$d(y_i, y_{i+1}) \leq \frac{1}{2^i},$$

for all $i$. However, for all $m, n \geq 1$, we have

$$d(y_m, y_n) \leq d(y_m, y_{m+1}) + \cdots + d(y_{n-1}, y_n) \leq \sum_{i=m}^{n} \frac{1}{2^i} \leq \frac{1}{2^{m-1}},$$

and thus, $(y_n)$ is a Cauchy sequence. Since $E$ is complete, the sequence, $(y_n)$, has a limit, and since it is a subsequence of $(x_n)$, the sequence, $(x_n)$, has some accumulation point. 

---

Figure 15.45: The first three stages of the construction of the Cauchy sequence $(y_n)$, where $E$ is the pink square region of $\mathbb{R}^2$. The original sequence $(x_n)$ is illustrated with plum colored dots. Figure (i.) covers $E$ with ball of radius 1 and shows the selection of $B^0_0$ and $y_0$. Figure (ii.) covers $B^0_0$ with balls of radius 1/2 and selects the yellow ball as $B^1_0$ with point $y_1$. Figure (iii.) covers $B^1_0$ with balls of radius 1/4 and selects the pale peach ball as $B^2_0$ with point $y_2$.

Another useful property of a complete metric space is that a subset is closed iff it is complete. This is shown in the following two propositions.

**Proposition 15.55.** Let $(E, d)$ be a metric space, and let $A$ be a subset of $E$. If $A$ is complete (which means that every Cauchy sequence of elements in $A$ converges to some point of $A$), then $A$ is closed in $E$. 

Proof. Assume \( x \in \overline{A} \). By Proposition 15.13, there is some sequence \((a_n)\) of points \(a_n \in A\) which converges to \(x\). Consequently \((a_n)\) is a Cauchy sequence in \(E\), and thus a Cauchy sequence in \(A\) (since \(a_n \in A\) for all \(n\)). Since \(A\) is complete, the sequence \((a_n)\) has a limit \(a \in A\), but since \(E\) is a metric space it is Hausdorff, so \(a = x\), which shows that \(x \in A\); that is, \(A\) is closed. \(\square\)

**Proposition 15.56.** Let \((E,d)\) be a metric space, and let \(A\) be a subset of \(E\). If \(E\) is complete and if \(A\) is closed in \(E\), then \(A\) is complete.

**Proof.** Let \((a_n)\) be a Cauchy sequence in \(A\). The sequence \((a_n)\) is also a Cauchy sequence in \(E\), and since \(E\) is complete, it has a limit \(x \in E\). But \(a_n \in A\) for all \(n\), so by Proposition 15.13 we must have \(x \in \overline{A}\). Since \(A\) is closed, actually \(x \in A\), which proves that \(A\) is complete. \(\square\)

An arbitrary metric space \((E,d)\) is not necessarily complete, but there is a construction of a metric space \((\hat{E},\hat{d})\) such that \(\hat{E}\) is complete, and there is a continuous (injective) distance-preserving map \(\varphi: E \to \hat{E}\) such that \(\varphi(E)\) is dense in \(\hat{E}\). This is a generalization of the construction of the set \(\mathbb{R}\) of real numbers from the set \(\mathbb{Q}\) of rational numbers in terms of Cauchy sequences. This construction can be immediately adapted to a normed vector space \((E,\|\cdot\|)\) to embed \((E,\|\cdot\|)\) into a complete normed vector space \((\hat{E},\|\cdot\|_E)\) (a Banach space). This construction is used heavily in integration theory, where \(E\) is a set of functions.

### 15.9 Completion of a Metric Space

In order to prove a kind of uniqueness result for the completion \((\hat{E},\hat{d})\) of a metric space \((E,d)\), we need the following result about extending a uniformly continuous function.

Recall that \(E_0\) is dense in \(E\) iff \(\overline{E_0} = E\). Since \(E\) is a metric space, by Proposition 15.13, this means that for every \(x \in E\), there is some sequence \((x_n)\) converging to \(x\), with \(x_n \in E_0\).

**Theorem 15.57.** Let \(E\) and \(F\) be two metric spaces, let \(E_0\) be a dense subspace of \(E\), and let \(f_0: E_0 \to F\) be a continuous function. If \(f_0\) is uniformly continuous and if \(F\) is complete, then there is a unique uniformly continuous function \(f: E \to F\) extending \(f_0\).

**Proof.** We follow Schwartz’s proof; see Schwartz [69] (Chapter XI, Section 3, Theorem 1).

**Step 1.** We begin by constructing a function \(f: E \to F\) extending \(f_0\). Since \(E_0\) is dense in \(E\), for every \(x \in E\), there is some sequence \((x_n)\) converging to \(x\), with \(x_n \in E_0\). Then the sequence \((x_n)\) is a Cauchy sequence in \(E\). We claim that \((f_0(x_n))\) is a Cauchy sequence in \(F\).

**Proof of the claim.** For every \(\epsilon > 0\), since \(f_0\) is uniformly continuous, there is some \(\eta > 0\) such that for all \((y, z) \in E_0\), if \(d(y, z) \leq \eta\), then \(d(f_0(y), f_0(z)) \leq \epsilon\). Since \((x_n)\) is a Cauchy sequence with \(x_n \in E_0\), there is some integer \(p > 0\) such that if \(m, n \geq p\), then \(d(x_m, x_n) \leq \eta\), thus \(d(f_0(x_m), f_0(x_n)) \leq \epsilon\), which proves that \((f_0(x_n))\) is a Cauchy sequence in \(F\). \(\square\)
Since $F$ is complete and $(f_0(x_n))$ is a Cauchy sequence in $F$, the sequence $(f_0(x_n))$ converges to some element of $F$; denote this element by $f(x)$.

**Step 2.** Let us now show that $f(x)$ does not depend on the sequence $(x_n)$ converging to $x$. Suppose that $(x'_n)$ and $(x''_n)$ are two sequences of elements in $E_0$ converging to $x$. Then the mixed sequence
\[
x'_0, x'_n, x'_1, x'_2, \ldots, x'_n, x''_n, \ldots,
\]
also converges to $x$. It follows that the sequence
\[
f_0(x'_0), f_0(x''_0), f_0(x'_1), f_0(x''_1), \ldots, f_0(x'_n), f_0(x''_n), \ldots,
\]
is a Cauchy sequence in $F$, and since $F$ is complete, it converges to some element of $F$, which implies that the sequences $(f_0(x'_n))$ and $(f_0(x''_n))$ converge to the same limit.

As a summary, we have defined a function $f : E \to F$ by
\[
f(x) = \lim_{n \to \infty} f_0(x_n).
\]
for any sequence $(x_n)$ converging to $x$, with $x_n \in E_0$.

**Step 3.** The function $f$ extends $f_0$. Since every element $x \in E_0$ is the limit of the constant sequence $(x_n)$ with $x_n = x$ for all $n \geq 0$, by definition $f(x)$ is the limit of the sequence $(f_0(x_n))$, which is the constant sequence with value $f_0(x)$, so $f(x) = f_0(x)$; that is, $f$ extends $f_0$.

**Step 4.** We now prove that $f$ is uniformly continuous. Since $f_0$ is uniformly continuous, for every $\epsilon > 0$, there is some $\eta > 0$ such that if $a, b \in E_0$ and $d(a, b) \leq \eta$, then $d(f_0(a), f_0(b)) \leq \epsilon$. Consider any two points $x, y \in E$ such that $d(x, y) \leq \eta/2$. We claim that $d(f(x), f(y)) \leq \epsilon$, which shows that $f$ is uniformly continuous.

Let $(x_n)$ be a sequence of points in $E_0$ converging to $x$, and let $(y_n)$ be a sequence of points in $E_0$ converging to $y$. By the triangle inequality,
\[
d(x_n, y_n) \leq d(x_n, x) + d(x, y) + d(y, y_n) = d(x, y) + d(x_n, x) + d(y_n, y),
\]
and since $(x_n)$ converges to $x$ and $(y_n)$ converges to $y$, there is some integer $p > 0$ such that for all $n \geq p$, we have $d(x_n, x) \leq \eta/4$ and $d(y_n, y) \leq \eta/4$, and thus
\[
d(x_n, y_n) \leq d(x, y) + \frac{\eta}{2}.
\]
Since we assumed that $d(x, y) \leq \eta/2$, we get $d(x_n, y_n) \leq \eta$ for all $n \geq p$, and by uniform continuity of $f_0$, we get
\[
d(f_0(x_n), f_0(y_n)) \leq \epsilon
\]
for all $n \geq p$. Since the distance function on $F$ is also continuous, and since $(f_0(x_n))$ converges to $f(x)$ and $(f_0(y_n))$ converges to $f(y)$, we deduce that the sequence $(d(f_0(x_n), f_0(y_n)))$ converges to $d(f(x), f(y))$. This implies that $d(f(x), f(y)) \leq \epsilon$, as desired.
Step 5. It remains to prove that \( f \) is unique. Since \( E_0 \) is dense in \( E \), for every \( x \in E \), there is some sequence \( (x_n) \) converging to \( x \), with \( x_n \in E_0 \). Since \( f \) extends \( f_0 \) and since \( f \) is continuous, we get
\[
f(x) = \lim_{n \to \infty} f_0(x_n),
\]
which only depends on \( f_0 \) and \( x \), and shows that \( f \) is unique.

Remark: It can be shown that the theorem no longer holds if we either omit the hypothesis that \( F \) is complete or omit that \( f_0 \) is uniformly continuous.

For example, if \( E_0 \neq E \) and if we let \( F = E_0 \) and \( f_0 \) be the identity function, it is easy to see that \( f_0 \) cannot be extended to a continuous function from \( E \) to \( E_0 \) (for any \( x \in E - E_0 \), any continuous extension \( f \) of \( f_0 \) would satisfy \( f(x) = x \), which is absurd since \( x \notin E_0 \)).

If \( f_0 \) is continuous but not uniformly continuous, a counter-example can be given by using \( E = \mathbb{R} = \mathbb{R} \cup \{\infty\} \) made into a metric space, \( E_0 = \mathbb{R} \), \( F = \mathbb{R} \), and \( f_0 \) the identity function; for details, see Schwartz [69] (Chapter XI, Section 3, page 134).

**Definition 15.40.** If \((E,d_E)\) and \((F,d_F)\) are two metric spaces, then a function \( f: E \to F \) is distance-preserving, or an isometry, if
\[
d_F(f(x), f(y)) = d_E(x, y), \quad \text{for all } x, y \in E.
\]

Observe that an isometry must be injective, because if \( f(x) = f(y) \), then \( d_F(f(x), f(y)) = 0 \), and since \( d_F(f(x), f(y)) = d_E(x, y) \), we get \( d_E(x, y) = 0 \), but \( d_E(x, y) = 0 \) implies that \( x = y \). Also, an isometry is uniformly continuous (since we can pick \( \eta = \epsilon \) to satisfy the condition of uniform continuity). However, an isometry is not necessarily surjective.

We now give a construction of the completion of a metric space. This construction is just a generalization of the classical construction of \( \mathbb{R} \) from \( \mathbb{Q} \) using Cauchy sequences.

**Theorem 15.58.** Let \((E,d)\) be any metric space. There is a complete metric space \((\hat{E}, \hat{d})\) called a completion of \((E,d)\), and a distance-preserving (uniformly continuous) map \( \varphi: E \to \hat{E} \) such that \( \varphi(E) \) is dense in \( \hat{E} \), and the following extension property holds: for every complete metric space \( F \) and for every uniformly continuous function \( f: E \to F \), there is a unique uniformly continuous function \( \hat{f}: \hat{E} \to F \) such that
\[
f = \hat{f} \circ \varphi,
\]
as illustrated in the following diagram.

\[
\begin{array}{ccc}
E & \xrightarrow{\varphi} & \hat{E} \\
\downarrow{f} & & \downarrow{\hat{f}} \\
& F. & \\
\end{array}
\]

As a consequence, for any two completions \((\hat{E}_1, \hat{d}_1)\) and \((\hat{E}_2, \hat{d}_2)\) of \((E,d)\), there is a unique bijective isometry between \((\hat{E}_1, \hat{d}_1)\) and \((\hat{E}_2, \hat{d}_2)\).
Proof. Consider the set $\mathcal{E}$ of all Cauchy sequences $(x_n)$ in $E$, and define the relation $\sim$ on $\mathcal{E}$ as follows:

$$(x_n) \sim (y_n) \iff \lim_{n \to \infty} d(x_n, y_n) = 0.$$ 

It is easy to check that $\sim$ is an equivalence relation on $\mathcal{E}$, and let $\hat{E} = \mathcal{E}/\sim$ be the quotient set, that is, the set of equivalence classes modulo $\sim$. Our goal is to show that we can endow $\hat{E}$ with a distance that makes it into a complete metric space satisfying the conditions of the theorem. We proceed in several steps.

Step 1. First, let us construct the function $\varphi : E \to \hat{E}$. For every $a \in E$, we have the constant sequence $(a_n)$ such that $a_n = a$ for all $n \geq 0$, which is obviously a Cauchy sequence. Let $\varphi(a) \in \hat{E}$ be the equivalence class $[(a_n)]$ of the constant sequence $(a_n)$ with $a_n = a$ for all $n$. By definition of $\sim$, the equivalence class $\varphi(a)$ is also the equivalence class of all sequences converging to $a$. The map $a \mapsto \varphi(a)$ is injective because a metric space is Hausdorff, so if $a \neq b$, then a sequence converging to $a$ does not converge to $b$. After having defined a distance on $\hat{E}$, we will check that $\varphi$ is an isometry.

Step 2. Let us now define a distance on $\hat{E}$. Let $\alpha = [(a_n)]$ and $\beta = [(b_n)]$ be two equivalence classes of Cauchy sequences in $E$. The triangle inequality implies that

$$d(a_m, b_m) \leq d(a_m, a_n) + d(a_n, b_n) + d(b_n, b_m) = d(a_n, b_n) + d(a_m, a_n) + d(b_m, b_n)$$

and

$$d(a_n, b_n) \leq d(a_n, a_m) + d(a_m, b_m) + d(b_m, b_n) = d(a_m, b_n) + d(a_m, a_n) + d(b_m, b_n),$$

which implies that

$$|d(a_m, b_m) - d(a_n, b_n)| \leq d(a_m, a_n) + d(b_m, b_n).$$

Since $(a_n)$ and $(b_n)$ are Cauchy sequences, it follows that $(d(a_n, b_n))$ is a Cauchy sequence of nonnegative reals. Since $\mathbb{R}$ is complete, the sequence $(d(a_n, b_n))$ has a limit, which we denote by $\tilde{d}(\alpha, \beta)$; that is, we set

$$\tilde{d}(\alpha, \beta) = \lim_{n \to \infty} d(a_n, b_n), \quad \alpha = [(a_n)], \quad \beta = [(b_n)].$$

Step 3. Let us check that $\tilde{d}(\alpha, \beta)$ does not depend on the Cauchy sequences $(a_n)$ and $(b_n)$ chosen in the equivalence classes $\alpha$ and $\beta$.

If $(a_n) \sim (a_n')$ and $(b_n) \sim (b_n')$, then $\lim_{n \to \infty} d(a_n, a_n') = 0$ and $\lim_{n \to \infty} d(b_n, b_n') = 0$, and since

$$d(a_n', b_n') \leq d(a_n', a_n) + d(a_n, b_n) + d(b_n, b_n') = d(a_n, b_n) + d(a_n, a_n') + d(b_n, b_n'),$$

and

$$d(a_n, b_n) \leq d(a_n, a_n') + d(a_n', b_n') + d(b_n', b_n) = d(a_n, b_n') + d(a_n, a_n') + d(b_n, b_n'),$$

we have

$$\tilde{d}(\alpha, \beta) = \lim_{n \to \infty} d(a_n, b_n), \quad \alpha = [(a_n)], \quad \beta = [(b_n)].$$

15.9. COMPLETION OF A METRIC SPACE

Proof. Consider the set $\mathcal{E}$ of all Cauchy sequences $(x_n)$ in $E$, and define the relation $\sim$ on $\mathcal{E}$ as follows:

$$(x_n) \sim (y_n) \iff \lim_{n \to \infty} d(x_n, y_n) = 0.$$ 

It is easy to check that $\sim$ is an equivalence relation on $\mathcal{E}$, and let $\hat{E} = \mathcal{E}/\sim$ be the quotient set, that is, the set of equivalence classes modulo $\sim$. Our goal is to show that we can endow $\hat{E}$ with a distance that makes it into a complete metric space satisfying the conditions of the theorem. We proceed in several steps.

Step 1. First, let us construct the function $\varphi : E \to \hat{E}$. For every $a \in E$, we have the constant sequence $(a_n)$ such that $a_n = a$ for all $n \geq 0$, which is obviously a Cauchy sequence. Let $\varphi(a) \in \hat{E}$ be the equivalence class $[(a_n)]$ of the constant sequence $(a_n)$ with $a_n = a$ for all $n$. By definition of $\sim$, the equivalence class $\varphi(a)$ is also the equivalence class of all sequences converging to $a$. The map $a \mapsto \varphi(a)$ is injective because a metric space is Hausdorff, so if $a \neq b$, then a sequence converging to $a$ does not converge to $b$. After having defined a distance on $\hat{E}$, we will check that $\varphi$ is an isometry.

Step 2. Let us now define a distance on $\hat{E}$. Let $\alpha = [(a_n)]$ and $\beta = [(b_n)]$ be two equivalence classes of Cauchy sequences in $E$. The triangle inequality implies that

$$d(a_m, b_m) \leq d(a_m, a_n) + d(a_n, b_n) + d(b_n, b_m) = d(a_n, b_n) + d(a_m, a_n) + d(b_m, b_n)$$

and

$$d(a_n, b_n) \leq d(a_n, a_m) + d(a_m, b_m) + d(b_m, b_n) = d(a_m, b_n) + d(a_m, a_n) + d(b_m, b_n),$$

which implies that

$$|d(a_m, b_m) - d(a_n, b_n)| \leq d(a_m, a_n) + d(b_m, b_n).$$

Since $(a_n)$ and $(b_n)$ are Cauchy sequences, it follows that $(d(a_n, b_n))$ is a Cauchy sequence of nonnegative reals. Since $\mathbb{R}$ is complete, the sequence $(d(a_n, b_n))$ has a limit, which we denote by $\tilde{d}(\alpha, \beta)$; that is, we set

$$\tilde{d}(\alpha, \beta) = \lim_{n \to \infty} d(a_n, b_n), \quad \alpha = [(a_n)], \quad \beta = [(b_n)].$$

Step 3. Let us check that $\tilde{d}(\alpha, \beta)$ does not depend on the Cauchy sequences $(a_n)$ and $(b_n)$ chosen in the equivalence classes $\alpha$ and $\beta$.

If $(a_n) \sim (a_n')$ and $(b_n) \sim (b_n')$, then $\lim_{n \to \infty} d(a_n, a_n') = 0$ and $\lim_{n \to \infty} d(b_n, b_n') = 0$, and since

$$d(a_n', b_n') \leq d(a_n', a_n) + d(a_n, b_n) + d(b_n, b_n') = d(a_n, b_n) + d(a_n, a_n') + d(b_n, b_n'),$$

and

$$d(a_n, b_n) \leq d(a_n, a_n') + d(a_n', b_n') + d(b_n', b_n) = d(a_n', b_n') + d(a_n, a_n') + d(b_n, b_n'),$$

we have

$$\tilde{d}(\alpha, \beta) = \lim_{n \to \infty} d(a_n, b_n), \quad \alpha = [(a_n)], \quad \beta = [(b_n)].$$
we have

$$|d(a_n, b_n) - d(a'_n, b'_n)| \leq d(a_n, a'_n) + d(b_n, b'_n),$$

so we have \(\lim_{n \to \infty} d(a'_n, b'_n) = \lim_{n \to \infty} d(a_n, b_n) = \tilde{d}(\alpha, \beta)\). Therefore, \(\tilde{d}(\alpha, \beta)\) is indeed well defined.

**Step 4.** Let us check that \(\varphi\) is indeed an isometry.

Given any two elements \(\varphi(a)\) and \(\varphi(b)\) in \(\tilde{E}\), since they are the equivalence classes of the constant sequences \((a_n)\) and \((b_n)\) such that \(a_n = a\) and \(b_n = b\) for all \(n\), the constant sequence \((\hat{d}(a_n, b_n))\) with \(d(a_n, b_n) = d(a, b)\) for all \(n\) converges to \(d(a, b)\), so by definition \(\tilde{d}(\varphi(a), \varphi(b)) = \lim_{n \to \infty} d(a_n, b_n) = d(a, b)\), which shows that \(\varphi\) is an isometry.

**Step 5.** Let us verify that \(\tilde{d}\) is a metric on \(\tilde{E}\). By definition it is obvious that \(\tilde{d}(\alpha, \beta) = \tilde{d}(\beta, \alpha)\). If \(\alpha\) and \(\beta\) are two distinct equivalence classes, then for any Cauchy sequence \((a_n)\) in the equivalence class \(\alpha\) and for any Cauchy sequence \((b_n)\) in the equivalence class \(\beta\), the sequences \((a_n)\) and \((b_n)\) are inequivalent, which means that \(\lim_{n \to \infty} d(a_n, b_n) \neq 0\), that is, \(\tilde{d}(\alpha, \beta) \neq 0\). Obviously, \(\tilde{d}(\alpha, \alpha) = 0\).

For any equivalence classes \(\alpha = [(a_n)], \beta = [(b_n)],\) and \(\gamma = [(c_n)]\), we have the triangle inequality

\[
d(a_n, c_n) \leq d(a_n, b_n) + d(b_n, c_n),
\]

so by continuity of the distance function, by passing to the limit, we obtain

\[
\tilde{d}(\alpha, \gamma) \leq \tilde{d}(\alpha, \beta) + \tilde{d}(\beta, \gamma),
\]

which is the triangle inequality for \(\tilde{d}\). Therefore, \(\tilde{d}\) is a distance on \(\tilde{E}\).

**Step 6.** Let us prove that \(\varphi(\tilde{E})\) is dense in \(\tilde{E}\). For any \(\alpha = [(a_n)]\), let \((x_n)\) be the constant sequence such that \(x_k = a_n\) for all \(k \geq 0\), so that \(\varphi(a_n) = [(x_n)]\). Then we have

\[
\tilde{d}(\alpha, \varphi(a_n)) = \lim_{m \to \infty} d(a_m, a_n) \leq \sup_{p, q \geq n} d(a_p, a_q).
\]

Since \((a_n)\) is a Cauchy sequence, \(\sup_{p, q \geq n} d(a_p, a_q)\) tends to 0 as \(n\) goes to infinity, so

\[
\lim_{n \to \infty} d(\alpha, \varphi(a_n)) = 0,
\]

which means that the sequence \((\varphi(a_n))\) converge to \(\alpha\), and \(\varphi(\tilde{E})\) is indeed dense in \(\tilde{E}\).

**Step 7.** Finally, let us prove that the metric space \(\tilde{E}\) is complete.

Let \((\alpha_n)\) be a Cauchy sequence in \(\tilde{E}\). Since \(\varphi(\tilde{E})\) is dense in \(\tilde{E}\), for every \(n > 0\), there some \(a_n \in E\) such that

\[
\tilde{d}(\alpha_n, \varphi(a_n)) \leq \frac{1}{n}.
\]

Since

\[
\tilde{d}(\varphi(a_m), \varphi(a_n)) \leq \tilde{d}(\varphi(a_m), a_m) + \tilde{d}(a_m, a_n) + \tilde{d}(a_n, \varphi(a_n)) \leq \tilde{d}(a_m, a_n) + \frac{1}{m} + \frac{1}{n},
\]

we have
and since \((\alpha_m)\) is a Cauchy sequence, so is \((\varphi(a_n))\), and as \(\varphi\) is an isometry, the sequence 
\((a_n)\) is a Cauchy sequence in \(E\). Let \(\alpha \in \hat{E}\) be the equivalence class of \((a_n)\). Since
\[
\hat{d}(\alpha, \varphi(a_n)) = \lim_{m \to \infty} d(a_m, a_n)
\]
and \((a_n)\) is a Cauchy sequence, we deduce that the sequence \((\varphi(a_n))\) converges to \(\alpha\), and since 
\(d(\alpha_n, \varphi(a_n)) \leq 1/n\) for all \(n > 0\), the sequence \((\alpha_n)\) also converges to \(\alpha\).

**Step 8.** Let us prove the extension property. Let \(F\) be any complete metric space and let \(f: E \to F\) be any uniformly continuous function. The function \(\varphi: E \to \hat{E}\) is an isometry and a bijection between \(E\) and its image \(\varphi(E)\), so its inverse \(\varphi^{-1}: \varphi(E) \to E\) is also an isometry, and thus is uniformly continuous. If we let \(g = f \circ \varphi^{-1}\), then \(g: \varphi(E) \to F\) is a uniformly continuous function, and \(\varphi(E)\) is dense in \(\hat{E}\), so by Theorem 15.57 there is a unique uniformly continuous function \(\hat{f}: \hat{E} \to F\) extending \(g = f \circ \varphi^{-1}\); see the diagram below:

This means that
\[
\hat{f}|\varphi(E) = f \circ \varphi^{-1},
\]
which implies that
\[
(\hat{f}|\varphi(E)) \circ \varphi = f,
\]
that is, \(f = \hat{f} \circ \varphi\), as illustrated in the diagram below:

If \(h: \hat{E} \to F\) is any other uniformly continuous function such that \(f = h \circ \varphi\), then
\(g = f \circ \varphi^{-1} = h|\varphi(E)\), so \(h\) is a uniformly continuous function extending \(g\), and by Theorem 15.57, we have have \(h = \hat{f}\), so \(\hat{f}\) is indeed unique.

**Step 9.** Uniqueness of the completion \((\hat{E}, \hat{d})\) up to a bijective isometry.

Let \((\hat{E}_1, \hat{d}_1)\) and \((\hat{E}_2, \hat{d}_2)\) be any two completions of \((E, d)\). Then we have two uniformly continuous isometries \(\varphi_1: E \to \hat{E}_1\) and \(\varphi_2: E \to \hat{E}_2\), so by the unique extension property, there exist unique uniformly continuous maps \(\hat{\varphi}_2: \hat{E}_1 \to \hat{E}_2\) and \(\hat{\varphi}_1: \hat{E}_2 \to \hat{E}_1\) such that the following diagrams commute:

\[
\begin{array}{ccc}
E & \xrightarrow{\varphi_1} & \hat{E}_1 \\
\varphi_2 & \downarrow & \hat{\varphi}_2 \\
\hat{E}_2 & \xrightarrow{\hat{\varphi}_2} & \hat{E}_1
\end{array} \quad \begin{array}{ccc}
E & \xrightarrow{\varphi_2} & \hat{E}_2 \\
\varphi_1 & \downarrow & \hat{\varphi}_1 \\
\hat{E}_1 & \xrightarrow{\hat{\varphi}_1} & \hat{E}_1
\end{array}
\]
Consequently we have the following commutative diagrams:

\[
\begin{array}{c}
\begin{array}{c}
E \\ \varphi_1 \\ \varphi_2
\end{array}
\quad \begin{array}{c}
\hat{E}_2 \\ \varphi_1 \\ \varphi_2
\end{array}
\quad \begin{array}{c}
E \\ \varphi_1 \\ \varphi_2
\end{array}
\quad \begin{array}{c}
\hat{E}_1 \\ \varphi_1 \\ \varphi_2
\end{array}
\end{array}
\]
15.10 The Contraction Mapping Theorem

If \((E, d)\) is a nonempty complete metric space, every map, \(f: E \to E\), for which there is some \(k\) such that \(0 \leq k < 1\) and
\[
d(f(x), f(y)) \leq kd(x, y)
\]
for all \(x, y \in E\), has the very important property that it has a unique fixed point, that is, there is a unique, \(a \in E\), such that \(f(a) = a\). A map as above is called a contraction mapping. Furthermore, the fixed point of a contraction mapping can be computed as the limit of a fast converging sequence.

The fixed point property of contraction mappings is used to show some important theorems of analysis, such as the implicit function theorem and the existence of solutions to certain differential equations. It can also be used to show the existence of fractal sets defined in terms of iterated function systems. Since the proof is quite simple, we prove the fixed point property of contraction mappings. First, observe that a contraction mapping is (uniformly) continuous.

**Proposition 15.59.** If \((E, d)\) is a nonempty complete metric space, every contraction mapping, \(f: E \to E\), has a unique fixed point. Furthermore, for every \(x_0 \in E\), defining the sequence, \((x_n)\), such that \(x_{n+1} = f(x_n)\), the sequence, \((x_n)\), converges to the unique fixed point of \(f\).

**Proof.** First we prove that \(f\) has at most one fixed point. Indeed, if \(f(a) = a\) and \(f(b) = b\), since
\[
d(a, b) = d(f(a), f(b)) \leq kd(a, b)
\]
and \(0 \leq k < 1\), we must have \(d(a, b) = 0\), that is, \(a = b\).

Next, we prove that \((x_n)\) is a Cauchy sequence. Observe that
\[
d(x_2, x_1) \leq kd(x_1, x_0),
\]
\[
d(x_3, x_2) \leq kd(x_2, x_1) \leq k^2d(x_1, x_0),
\]
\[
\vdots
\]
\[
d(x_{n+1}, x_n) \leq kd(x_n, x_{n-1}) \leq \cdots \leq k^n d(x_1, x_0).
\]

Thus, we have
\[
d(x_{n+p}, x_n) \leq d(x_{n+p}, x_{n+p-1}) + d(x_{n+p-1}, x_{n+p-2}) + \cdots + d(x_{n+1}, x_n) \\
\leq (k^{p-1} + k^{p-2} + \cdots + k + 1)k^n d(x_1, x_0) \\
\leq \frac{k^n}{1 - k} d(x_1, x_0).
\]
We conclude that \( d(x_{n+p}, x_n) \) converges to 0 when \( n \) goes to infinity, which shows that \((x_n)\) is a Cauchy sequence. Since \( E \) is complete, the sequence \((x_n)\) has a limit, \( a \). Since \( f \) is continuous, the sequence \((f(x_n))\) converges to \( f(a) \). But \( x_{n+1} = f(x_n) \) converges to \( a \) and so \( f(a) = a \), the unique fixed point of \( f \).

Note that no matter how the starting point \( x_0 \) of the sequence \((x_n)\) is chosen, \((x_n)\) converges to the unique fixed point of \( f \). Also, the convergence is fast, since

\[
d(x_n, a) \leq \frac{k^n}{1 - k} d(x_1, x_0).
\]

The Hausdorff distance between compact subsets of a metric space provides a very nice illustration of some of the theorems on complete and compact metric spaces just presented.

**Definition 15.41.** Given a metric space, \((X, d)\), for any subset, \( A \subseteq X \), for any, \( \epsilon \geq 0 \), define the \( \epsilon \)-hull of \( A \) as the set

\[
V_{\epsilon}(A) = \{x \in X, \exists a \in A \mid d(a, x) \leq \epsilon\}.
\]

See Figure 15.46. Given any two nonempty bounded subsets, \( A, B \) of \( X \), define \( D(A, B) \), the *Hausdorff distance between \( A \) and \( B \)*, by

\[
D(A, B) = \inf \{\epsilon \geq 0 \mid A \subseteq V_{\epsilon}(B) \text{ and } B \subseteq V_{\epsilon}(A)\}.
\]

![Figure 15.46: The \( \epsilon \)-hull of a polygonal region \( A \) of \( \mathbb{R}^2 \)](image)

Note that since we are considering nonempty bounded subsets, \( D(A, B) \) is well defined (i.e., not infinite). However, \( D \) is not necessarily a distance function. It is a distance function if we restrict our attention to nonempty compact subsets of \( X \) (actually, it is also a metric on closed and bounded subsets). We let \( K(X) \) denote the set of all nonempty compact subsets of \( X \). The remarkable fact is that \( D \) is a distance on \( K(X) \) and that if \( X \) is complete or compact, then so is \( K(X) \). The following theorem is taken from Edgar [29].
Theorem 15.60. If $(X,d)$ is a metric space, then the Hausdorff distance, $D$, on the set, $\mathcal{K}(X)$, of nonempty compact subsets of $X$ is a distance. If $(X,d)$ is complete, then $(\mathcal{K}(X), D)$ is complete and if $(X,d)$ is compact, then $(\mathcal{K}(X), D)$ is compact.

Proof. Since (nonempty) compact sets are bounded, $D(A,B)$ is well defined. Clearly $D$ is symmetric. Assume that $D(A,B) = 0$. Then for every $\epsilon > 0$, $A \subseteq V_\epsilon(B)$, which means that for every $a \in A$, there is some $b \in B$ such that $d(a,b) \leq \epsilon$, and thus, that $A \subseteq \overline{B}$. Since Proposition 15.26 implies that $B$ is closed, $\overline{B} = B$, and we have $A \subseteq B$. Similarly, $B \subseteq A$, and thus, $A = B$. Clearly, if $A = B$, we have $D(A,B) = 0$. It remains to prove the triangle inequality. Assume that $D(A,B) \leq \epsilon_1$ and that $D(B,C) \leq \epsilon_2$. We must show that $D(A,C) \leq \epsilon_1 + \epsilon_2$. This will be accomplished if we can show that $C \subseteq V_{\epsilon_1+\epsilon_2}(A)$ and $A \subseteq V_{\epsilon_1}(C)$. By assumption and definition of $D$, $B \subseteq V_{\epsilon_1}(A)$ and $C \subseteq V_{\epsilon_2}(B)$. Then

$$V_{\epsilon_2}(B) \subseteq V_{\epsilon_2}(V_{\epsilon_1}(A)),$$

and since a basic application of the triangle inequality implies that

$$V_{\epsilon_2}(V_{\epsilon_1}(A)) \subseteq V_{\epsilon_1+\epsilon_2}(A),$$

we get

$$C \subseteq V_{\epsilon_2}(B) \subseteq V_{\epsilon_1+\epsilon_2}(A).$$

See Figure 15.47.

![Figure 15.47: Let $A$ be the small pink square and $B$ be the small purple triangle in $\mathbb{R}^2$. The periwinkle oval $C$ is contained in $V_{\epsilon_1+\epsilon_2}(A)$.](image)

Similarly, the conditions $(A,B) \leq \epsilon_1$ and $D(B,C) \leq \epsilon_2$ imply that

$$A \subseteq V_{\epsilon_1}(B), \quad B \subseteq V_{\epsilon_2}(C).$$
Hence
\[ A \subseteq V_{\varepsilon_1} (B) \subseteq V_{\varepsilon_1} (V_{\varepsilon_2} (C)) \subseteq V_{\varepsilon_1 + \varepsilon_2} (C), \]
and thus the triangle inequality follows.

Next we need to prove that if \((X, d)\) is complete, then \((\mathcal{K}(X), D)\) is also complete. First we show that if \((A_n)\) is a sequence of nonempty compact sets converging to a nonempty compact set \(A\) in the Hausdorff metric, then
\[ A = \{ x \in X \mid \text{there is a sequence, } (x_n), \text{ with } x_n \in A_n \text{ converging to } x \}. \]

Indeed, if \((x_n)\) is a sequence with \(x_n \in A_n\) converging to \(x\) and \((A_n)\) converges to \(A\) then, for every \(\varepsilon > 0\), there is some \(x_n\) such that \(d(x_n, x) \leq \varepsilon/2\) and there is some \(a_n \in A\) such that \(d(a_n, x_n) \leq \varepsilon/2\) and thus, \(d(a_n, x) \leq \varepsilon\), which shows that \(x \in \overline{A}\). Since \(A\) is compact, it is closed, and \(x \in A\). See Figure 15.48.

![Figure 15.48](image-url)

Figure 15.48: Let \((A_n)\) be the sequence of parallelograms converging to \(A\), the large pale yellow parallelogram. Figure \((ii.\) expands the dashed region and shows why \(d(a_n, x) < \varepsilon\).

Conversely, since \((A_n)\) converges to \(A\), for every \(x \in A\), for every \(n \geq 1\), there is some \(x_n \in A_n\) such that \(d(x_n, x) \leq 1/n\) and the sequence \((x_n)\) converges to \(x\).

Now let \((A_n)\) be a Cauchy sequence in \(\mathcal{K}(X)\). It can be proven that \((A_n)\) converges to the set
\[ A = \{ x \in X \mid \text{there is a sequence, } (x_n), \text{ with } x_n \in A_n \text{ converging to } x \}, \]
and that \(A\) is nonempty and compact. To prove that \(A\) is compact, one proves that it is totally bounded and complete. Details are given in Edgar [29].
Finally we need to prove that if \((X,d)\) is compact, then \((\mathcal{K}(X), D)\) is compact. Since we already know that \((\mathcal{K}(X), D)\) is complete if \((X,d)\) is, it is enough to prove that \((\mathcal{K}(X), D)\) is totally bounded if \((X,d)\) is, which is not hard.

In view of Theorem 15.60 and Theorem 15.59, it is possible to define some nonempty compact subsets of \(X\) in terms of fixed points of contraction maps. This can be done in terms of iterated function systems, yielding a large class of fractals. However, we will omit this topic and instead refer the reader to Edgar [29].

Before considering differentials, we need to look at the continuity of linear maps.

### 15.11 Continuous Linear and Multilinear Maps

If \(E\) and \(F\) are normed vector spaces, we first characterize when a linear map \(f: E \to F\) is continuous.

**Proposition 15.61.** Given two normed vector spaces \(E\) and \(F\), for any linear map \(f: E \to F\), the following conditions are equivalent:

1. The function \(f\) is continuous at 0.
2. There is a constant \(k \geq 0\) such that,
   \[
   \|f(u)\| \leq k, \text{ for every } u \in E \text{ such that } \|u\| \leq 1.
   \]
3. There is a constant \(k \geq 0\) such that,
   \[
   \|f(u)\| \leq k\|u\|, \text{ for every } u \in E.
   \]
4. The function \(f\) is continuous at every point of \(E\).

**Proof.** Assume (1). Then for every \(\epsilon > 0\), there is some \(\eta > 0\) such that, for every \(u \in E\), if \(\|u\| \leq \eta\), then \(\|f(u)\| \leq \epsilon\). Pick \(\epsilon = 1\), so that there is some \(\eta > 0\) such that, if \(\|u\| \leq \eta\), then \(\|f(u)\| \leq 1\). If \(\|u\| \leq 1\), then \(\|\eta u\| \leq \|\eta\|\|u\| \leq \eta\), and so, \(\|f(\eta u)\| \leq 1\), that is, \(\|f(u)\| \leq 1\), which implies \(\|f(u)\| \leq \eta^{-1}\). Thus, (2) holds with \(k = \eta^{-1}\).

Assume that (2) holds. If \(u = 0\), then by linearity, \(f(0) = 0\), and thus \(\|f(0)\| \leq k\|0\|\) holds trivially for all \(k \geq 0\). If \(u \neq 0\), then \(\|u\| > 0\), and since

\[
\frac{u}{\|u\|} = 1,
\]

we have

\[
\left\| f \left( \frac{u}{\|u\|} \right) \right\| \leq k,
\]
which implies that
\[ \|f(u)\| \leq k\|u\|. \]

Thus, (3) holds.

If (3) holds, then for all \( u, v \in E \), we have
\[ \|f(v) - f(u)\| = \|f(v - u)\| \leq k\|v - u\|. \]

If \( k = 0 \), then \( f \) is the zero function, and continuity is obvious. Otherwise, if \( k > 0 \), for every \( \varepsilon > 0 \), if \( \|v - u\| \leq \frac{\varepsilon}{k} \), then \( \|f(v - u)\| \leq \varepsilon \), which shows continuity at every \( u \in E \). Finally, it is obvious that (4) implies (1).

Among other things, Proposition 15.61 shows that a linear map is continuous iff the image of the unit (closed) ball is bounded. Since a continuous linear map satisfies the condition \( \|f(u)\| \leq k\|u\| \) (for some \( k \geq 0 \)), it is also uniformly continuous.

If \( E \) and \( F \) are normed vector spaces, the set of all continuous linear maps \( f : E \to F \) is denoted by \( \mathcal{L}(E; F) \).

Using Proposition 15.61, we can define a norm on \( \mathcal{L}(E; F) \) which makes it into a normed vector space. This definition has already been given in Chapter 16 (Definition 16.7) but for the reader’s convenience, we repeat it here.

**Definition 15.42.** Given two normed vector spaces \( E \) and \( F \), for every continuous linear map \( f : E \to F \), we define the **operator norm** \( \|f\| \) of \( f \) as

\[
\|f\| = \inf \left\{ k \geq 0 \mid \|f(x)\| \leq k\|x\|, \text{ for all } x \in E \right\} \\
= \sup \left\{ \|f(x)\| \mid \|x\| \leq 1 \right\} \\
= \sup \left\{ \|f(x)\| \mid \|x\| = 1 \right\}.
\]

From Definition 15.42, for every continuous linear map \( f \in \mathcal{L}(E; F) \), we have
\[ \|f(x)\| \leq \|f\|\|x\|, \]
for every \( x \in E \). It is easy to verify that \( \mathcal{L}(E; F) \) is a normed vector space under the norm of Definition 15.42. Furthermore, if \( E, F, G \), are normed vector spaces, and \( f : E \to F \) and \( g : F \to G \) are continuous linear maps, we have
\[ \|g \circ f\| \leq \|g\|\|f\|. \]

We can now show that when \( E = \mathbb{R}^n \) or \( E = \mathbb{C}^n \), with any of the norms \( \|\|_1, \|\|_2, \) or \( \|\|_\infty \), then every linear map \( f : E \to F \) is continuous.

**Proposition 15.62.** If \( E = \mathbb{R}^n \) or \( E = \mathbb{C}^n \), with any of the norms \( \|\|_1, \|\|_2, \) or \( \|\|_\infty \), and \( F \) is any normed vector space, then every linear map \( f : E \to F \) is continuous.
Proof. Let \((e_1, \ldots, e_n)\) be the standard basis of \(\mathbb{R}^n\) (a similar proof applies to \(\mathbb{C}^n\)). In view of Proposition 16.2, it is enough to prove the proposition for the norm

\[
\|x\|_\infty = \max\{|x_i| \mid 1 \leq i \leq n\}.
\]

We have,

\[
\|f(v) - f(u)\| = \|f(v - u)\| = \left\| f\left( \sum_{1 \leq i \leq n} (v_i - u_i)e_i \right) \right\| = \left\| \sum_{1 \leq i \leq n} (v_i - u_i)f(e_i) \right\|,
\]

and so,

\[
\|f(v) - f(u)\| \leq \left( \sum_{1 \leq i \leq n} \|f(e_i)\| \right) \max_{1 \leq i \leq n} |v_i - u_i| = \left( \sum_{1 \leq i \leq n} \|f(e_i)\| \right) \|v - u\|_\infty.
\]

By the argument used in Proposition 15.61 to prove that (3) implies (4), \(f\) is continuous. \(\square\)

Actually, we proved in Theorem 16.3 that if \(E\) is a vector space of finite dimension, then any two norms are equivalent, so that they define the same topology. This fact together with Proposition 15.62 prove the following:

**Theorem 15.63.** If \(E\) is a vector space of finite dimension (over \(\mathbb{R}\) or \(\mathbb{C}\)), then all norms are equivalent (define the same topology). Furthermore, for any normed vector space \(F\), every linear map \(f : E \to F\) is continuous.

If \(E\) is a normed vector space of infinite dimension, a linear map \(f : E \to F\) may not be continuous. As an example, let \(E\) be the infinite vector space of all polynomials over \(\mathbb{R}\). Let

\[
\|P(X)\| = \max_{0 \leq x \leq 1} |P(x)|.
\]

We leave as an exercise to show that this is indeed a norm. Let \(F = \mathbb{R}\), and let \(f : E \to F\) be the map defined such that, \(f(P(X)) = P(3)\). It is clear that \(f\) is linear. Consider the sequence of polynomials

\[
P_n(X) = \left( \frac{X}{2} \right)^n.
\]

It is clear that \(\|P_n\| = \left( \frac{1}{2} \right)^n\), and thus, the sequence \(P_n\) has the null polynomial as a limit. However, we have

\[
f(P_n(X)) = P_n(3) = \left( \frac{3}{2} \right)^n,
\]

and the sequence \(f(P_n(X))\) diverges to \(+\infty\). Consequently, in view of Proposition 15.15 (1), \(f\) is not continuous.
We now consider the continuity of multilinear maps. We treat explicitly bilinear maps, the general case being a straightforward extension.

**Proposition 15.64.** Given normed vector spaces $E$, $F$ and $G$, for any bilinear map $f : E \times E \to G$, the following conditions are equivalent:

1. The function $f$ is continuous at $(0, 0)$.
2. There is a constant $k \geq 0$ such that, $\|f(u, v)\| \leq k$, for all $u, v \in E$ such that $\|u\|, \|v\| \leq 1$.
3. There is a constant $k \geq 0$ such that, $\|f(u, v)\| \leq k\|u\|\|v\|$, for all $u, v \in E$.
4. The function $f$ is continuous at every point of $E \times F$.

**Proof.** It is similar to that of Proposition 15.61, with a small subtlety in proving that (3) implies (4), namely that two different $\eta$'s that are not independent are needed. □

In contrast to continuous linear maps, which must be uniformly continuous, nonzero continuous bilinear maps are not uniformly continuous. Let $f : E \times F \to G$ be a continuous bilinear map such that $f(a, b) \neq 0$ for some $a \in E$ and some $b \in F$. Consider the sequences $(u_n)$ and $(v_n)$ (with $n \geq 1$) given by

\[
\begin{aligned}
  u_n &= (x_n, y_n) = (na, nb) \\
  v_n &= (x'_n, y'_n) = \left(\left( n + \frac{1}{n}\right)a, \left( n + \frac{1}{n} \right)b \right).
\end{aligned}
\]

Obviously

\[
\|v_n - u_n\| \leq \frac{1}{n} (\|a\| + \|b\|),
\]

so $\lim_{n \to \infty} \|v_n - u_n\| = 0$. On the other hand

\[
f(x'_n, y'_n) - f(x_n, y_n) = \left(2 + \frac{1}{n^2}\right) f(a, b),
\]

and thus $\lim_{n \to \infty} \|f(x'_n, y'_n) - f(x_n, y_n)\| = 2 \|f(a, b)\| \neq 0$, which shows that $f$ is not uniformly continuous, because if this was the case, this limit would be zero.

If $E$, $F$, and $G$, are normed vector spaces, we denote the set of all continuous bilinear maps $f : E \times F \to G$ by $L_2(E, F; G)$. Using Proposition 15.64, we can define a norm on $L_2(E, F; G)$ which makes it into a normed vector space.
**Definition 15.43.** Given normed vector spaces $E$, $F$, and $G$, for every continuous bilinear map $f : E \times F \to G$, we define the norm $\|f\|$ of $f$ as

$$
\|f\| = \inf \{ k \geq 0 \mid \|f(x, y)\| \leq k\|x\|\|y\|, \text{ for all } x, y \in E \}
= \sup \{\|f(x, y)\| \mid \|x\|, \|y\| \leq 1 \}.
$$

From Definition 15.42, for every continuous bilinear map $f \in \mathcal{L}_2(E, F; G)$, we have

$$
\|f(x, y)\| \leq \|f\|\|x\|\|y\|,
$$
for all $x, y \in E$. It is easy to verify that $\mathcal{L}_2(E, F; G)$ is a normed vector space under the norm of Definition 15.43.

Given a bilinear map $f : E \times F \to G$, for every $u \in E$, we obtain a linear map denoted $fu : F \to G$, defined such that, $fu(v) = f(u, v)$. Furthermore, since

$$
\|f(x, y)\| \leq \|f\|\|x\|\|y\|,
$$
it is clear that $fu$ is continuous. We can then consider the map $\varphi : E \to \mathcal{L}(F; G)$, defined such that, $\varphi(u) = fu$, for any $u \in E$, or equivalently, such that,

$$
\varphi(u)(v) = f(u, v).
$$

Actually, it is easy to show that $\varphi$ is linear and continuous, and that $\|\varphi\| = \|f\|$. Thus, $f \mapsto \varphi$ defines a map from $\mathcal{L}_2(E, F; G)$ to $\mathcal{L}(E; \mathcal{L}(F; G))$. We can also go back from $\mathcal{L}(E; \mathcal{L}(F; G))$ to $\mathcal{L}_2(E, F; G)$. We summarize all this in the following proposition.

**Proposition 15.65.** Let $E, F, G$ be three normed vector spaces. The map $f \mapsto \varphi$, from $\mathcal{L}_2(E, F; G)$ to $\mathcal{L}(E; \mathcal{L}(F; G))$, defined such that, for every $f \in \mathcal{L}_2(E, F; G)$,

$$
\varphi(u)(v) = f(u, v),
$$
is an isomorphism of vector spaces, and furthermore, $\|\varphi\| = \|f\|$.

As a corollary of Proposition 15.65, we get the following proposition which will be useful when we define second-order derivatives.

**Proposition 15.66.** Let $E, F$ be normed vector spaces. The map $\text{app}$ from $\mathcal{L}(E; F) \times E$ to $F$, defined such that, for every $f \in \mathcal{L}(E; F)$, for every $u \in E$,

$$
\text{app}(f, u) = f(u),
$$
is a continuous bilinear map.
Remark: If $E$ and $F$ are nontrivial, it can be shown that $\|\text{app}\| = 1$. It can also be shown that composition

$$\circ: \mathcal{L}(E; F) \times \mathcal{L}(F; G) \to \mathcal{L}(E; G),$$

is bilinear and continuous.

The above propositions and definition generalize to arbitrary $n$-multilinear maps, with $n \geq 2$. Proposition 15.64 extends in the obvious way to any $n$-multilinear map $f: E_1 \times \cdots \times E_n \to F$, but condition (3) becomes:

There is a constant $k \geq 0$ such that,

$$\|f(u_1, \ldots, u_n)\| \leq k\|u_1\| \cdots \|u_n\|, \text{ for all } u_1 \in E_1, \ldots, u_n \in E_n.$$

Definition 15.43 also extends easily to

$$\|f\| = \inf \{k \geq 0 \mid \|f(x_1, \ldots, x_n)\| \leq k\|x_1\| \cdots \|x_n\|, \text{ for all } x_i \in E_i; 1 \leq i \leq n\} = \sup \{\|f(x_1, \ldots, x_n)\| \mid \|x_1\|, \ldots, \|x_n\| \leq 1\}.$$

Proposition 15.65 is also easily extended, and we get an isomorphism between continuous $n$-multilinear maps in $\mathcal{L}_n(E_1, \ldots, E_n; F)$, and continuous linear maps in

$$\mathcal{L}(E_1; \mathcal{L}(E_2; \cdots; \mathcal{L}(E_n; F))).$$

An obvious extension of Proposition 15.66 also holds.

Definition 15.44. A normed vector space $(E, \|\|)$ over $\mathbb{R}$ (or $\mathbb{C}$) which is a complete metric space for the distance $d(u, v) = \|v - u\|$, is called a Banach space.

It can be shown that every normed vector space of finite dimension is a Banach space (is complete). This is because $\mathbb{R}$ (and $\mathbb{C}$) are complete. The following theorem is a key result of the theory of Banach spaces worth proving.

Theorem 15.67. If $E$ and $F$ are normed vector spaces, and if $F$ is a Banach space, then $\mathcal{L}(E; F)$ is a Banach space (with the operator norm).

Proof. Let $(f)_{n \geq 1}$ be a Cauchy sequence of continuous linear maps $f_n: E \to F$. We proceed in several steps.

Step 1. Define the pointwise limit $f: E \to F$ of the sequence $(f_n)_{n \geq 1}$.

Since $(f)_{n \geq 1}$ is a Cauchy sequence, for every $\epsilon > 0$, there is some $N > 0$ such that $\|f_m - f_n\| < \epsilon$ for all $m, n \geq N$. Since $\|\|$ is the operator norm, we deduce that for any $u \in E$, we have

$$\|f_m(u) - f_n(u)\| = \|(f_m - f_n)(u)\| \leq \|f_m - f_n\| \|u\| \leq \epsilon \|u\| \text{ for all } m, n \geq N,$$
that is,
\[ \| f_m(u) - f_n(u) \| \leq \epsilon \| u \| \quad \text{for all } m, n \geq N. \] (\star_1)

If \( u = 0 \), then \( f_m(0) = f_n(0) = 0 \) for all \( m, n \), so the sequence \( (f_n(0)) \) is a Cauchy sequence in \( F \) converging to 0. If \( u \neq 0 \), by replacing \( \epsilon \) by \( \epsilon / \| u \| \), we see that the sequence \( (f_n(u)) \) is a Cauchy sequence in \( F \). Since \( F \) is complete, the sequence \( (f_n(u)) \) has a limit which we denote by \( f(u) \). This defines our candidate limit function \( f \) by
\[ f(u) = \lim_{n \to \infty} f_n(u). \]

It remains to prove that
1. \( f \) is linear.
2. \( f \) is continuous.
3. \( f \) is the limit of \( (f_n) \) for the operator norm.

**Step 2.** The function \( f \) is linear.

Recall that in a normed vector space, addition and multiplication by a fixed scalar are continuous (since \( \| u + v \| \leq \| u \| + \| v \| \) and \( \| \lambda u \| \leq |\lambda| \| u \| \)). Thus by definition of \( f \) and since the \( f_n \) are linear we have
\[
\begin{align*}
    f(u + v) &= \lim_{n \to \infty} f_n(u + v) \quad \text{by definition of } f \\
    &= \lim_{n \to \infty} (f_n(u) + f_n(v)) \quad \text{by linearity of } f_n \\
    &= \lim_{n \to \infty} f_n(u) + \lim_{n \to \infty} f_n(v) \quad \text{since } + \text{ is continuous} \\
    &= f(u) + f(v) \quad \text{by definition of } f.
\end{align*}
\]

Similarly,
\[
\begin{align*}
    f(\lambda u) &= \lim_{n \to \infty} f_n(\lambda u) \quad \text{by definition of } f \\
    &= \lim_{n \to \infty} \lambda f_n(u) \quad \text{by linearity of } f_n \\
    &= \lambda \lim_{n \to \infty} f_n(u) \quad \text{by continuity of scalar multiplication} \\
    &= \lambda f(u) \quad \text{by definition of } f.
\end{align*}
\]

Therefore, \( f \) is linear.

**Step 3.** The function \( f \) is continuous.

Since \( (f_n)_{n \geq 1} \) is a Cauchy sequence, for every \( \epsilon > 0 \), there is some \( N > 0 \) such that \( \| f_m - f_n \| < \epsilon \) for all \( m, n \geq N \). Since \( f_m = f_n + f_m - f_n \), we get \( \| f_m \| \leq \| f_n \| + \| f_m - f_n \| \), which implies that
\[ \| f_m \| \leq \| f_n \| + \epsilon \quad \text{for all } m, n \geq N. \] (\star_2)
CHAPTER 15. TOPOLOGY

Using (\(\ast_2\)), we also have
\[
\|f_m(u)\| \leq \|f_m\| \|u\| \leq (\|f_n\| + \epsilon) \|u\|
\]
for all \(m, n \geq N\), that is,
\[
\|f_m(u)\| \leq (\|f_n\| + \epsilon) \|u\|
\]
for all \(m, n \geq N\). \(\ast_3\)

Hold \(n \geq N\) fixed and let \(m\) tend to \(+\infty\) in \(\ast_3\). Since the norm is continuous, we get
\[
\|f(u)\| \leq (\|f_n\| + \epsilon) \|u\|
\]
for all \(m, n \geq N\).

Step 4. The function \(f\) is the limit of \((f_n)\) for the operator norm.

Recall \(\ast_1\): \[
\|f_m(u) - f_n(u)\| \leq \epsilon \|u\|
\]
for all \(m, n \geq N\). \(\ast_1\)

Hold \(n \geq N\) fixed but this time let \(m\) tend to \(+\infty\) in \(\ast_1\). By continuity of the norm we get
\[
\|f(u) - f_n(u)\| = \|(f - f_n)(u)\| \leq \epsilon \|u\|.
\]

By definition of the operator norm,
\[
\|f - f_n\| = \sup\{\|(f - f_n)(u)\| : \|u\| = 1\} \leq \epsilon
\]
for all \(n \geq N\), which proves that \(f_n\) converges to \(f\) for the operator norm. \(\square\)

As a special case of Theorem 15.67, if we let \(F = \mathbb{R}\) (or \(F = \mathbb{C}\) in the case of complex vector spaces) we see that \(E' = \mathcal{L}(E; \mathbb{R})\) (or \(E' = \mathcal{L}(E; \mathbb{C})\)) is complete (since \(\mathbb{R}\) and \(\mathbb{C}\) are complete). The space \(E'\) of continuous linear forms on \(E\) is called the dual of \(E\). It is a subspace of the algebraic dual \(E^*\) of \(E\) which consists of all linear forms on \(E\), not necessarily continuous.

It can also be shown that if \(E, F\) and \(G\) are normed vector spaces, and if \(G\) is a Banach space, then \(\mathcal{L}_2(E, F; G)\) is a Banach space. The proof is essentially identical.

15.12 Completion of a Normed Vector Space

An easy corollary of Theorem 15.58 and Theorem 15.57 is that every normed vector space can be embedded in a complete normed vector space, that is, a Banach space.

Theorem 15.68. If \((E, \|\|)\) is a normed vector space, then its completion \((\hat{E}, \hat{d})\) as a metric space (where \(E\) is given the metric \(d(x, y) = \|x - y\|\)) can be given a unique vector space structure extending the vector space structure on \(E\), and a norm \(\|\|_{\hat{E}}\), so that \((\hat{E}, \|\|_{\hat{E}})\) is a Banach space, and the metric \(\hat{d}\) is associated with the norm \(\|\|_{\hat{E}}\). Furthermore, the isometry \(\varphi: E \to \hat{E}\) given by Theorem 15.58 is a linear isometry, and \(\varphi(E)\) is dense in \(\hat{E}\).
Proof. The addition operation $+: E \times E \to E$ is uniformly continuous because
\[ \|(u' + v') - (u'' + v'')\| \leq \|u' - u''\| + \|v' - v''\|. \]
It is not hard to show that $\hat{E} \times \hat{E}$ is a complete metric space and that $E \times E$ is dense in $\hat{E} \times \hat{E}$. Then, by Theorem 15.57, the uniformly continuous function $+$ has a unique continuous extension $+: \hat{E} \times \hat{E} \to \hat{E}$.

The map $\cdot: \mathbb{R} \times E \to E$ is not uniformly continuous, but for any fixed $\lambda \in \mathbb{R}$, the map $L_\lambda: E \to E$ given by $L_\lambda(u) = \lambda \cdot u$ is uniformly continuous, so by Theorem 15.57 the function $L_\lambda$ has a unique continuous extension $L_\lambda: \hat{E} \to \hat{E}$, which we use to define the scalar multiplication $\cdot: \mathbb{R} \times \hat{E} \to \hat{E}$. It is easily checked that with the above addition and scalar multiplication, $\hat{E}$ is a vector space.

Since the norm $\|\|$ on $E$ is uniformly continuous, it has a unique continuous extension $\|\|: \hat{E} \to \mathbb{R}_+$. The identities $\|u + v\| \leq \|u\| + \|v\|$ and $\|\lambda u\| \leq |\lambda| \|u\|$ extend to $\hat{E}$ by continuity. The equation
\[ d(u, v) = \|u - v\| \]
also extends to $\hat{E}$ by continuity and yields
\[ \hat{d}(\alpha, \beta) = \|\alpha - \beta\|_{\hat{E}}, \]
which shows that $\|\|_{\hat{E}}$ is indeed a norm, and that the metric $\hat{d}$ is associated to it. Finally, it is easy to verify that the map $\phi$ is linear. The uniqueness of the structure of normed vector space follows from the uniqueness of continuous extensions in Theorem 15.57.

Theorem 15.68 and Theorem 15.57 will be used to show that every Hermitian space can be embedded in a Hilbert space; see Theorem 18.1.

The following version of Theorem 15.57 for normed vector spaces will be needed in the theory of integration.

**Theorem 15.69.** Let $E$ and $F$ be two normed vector spaces, let $E_0$ be a dense subspace of $E$, and let $f_0: E_0 \to F$ be a continuous function. If $f_0$ is uniformly continuous and if $F$ is complete, then there is a unique uniformly continuous function $f: E \to F$ extending $f_0$. Furthermore, if $f_0$ is a continuous linear map, then $f$ is also a linear continuous map, and $\|f\| = \|f_0\|$.

**Proof.** We only need to prove the second statement. Given any two vectors $x, y \in E$, since $E_0$ is dense on $E$ we can pick sequences $(x_n)$ and $(y_n)$ of vectors $x_n, y_n \in E_0$ such that $x = \lim_{n \to \infty} x_n$ and $y = \lim_{n \to \infty} y_n$. Since addition and scalar multiplication are continuous, we get
\[ x + y = \lim_{n \to \infty} (x_n + y_n) \]
\[ \lambda x = \lim_{n \to \infty} (\lambda x_n) \]
for any $\lambda \in \mathbb{R}$ (or $\lambda \in \mathbb{C}$). Since $f(x)$ is defined by
\[ f(x) = \lim_{n \to \infty} f_0(x_n) \]
independently of the sequence $(x_n)$ converging to $x$, and similarly for $f(y)$ and $f(x + y)$, since $f_0$ is linear, we have
\[ f(x + y) = \lim_{n \to \infty} f_0(x_n + y_n) = \lim_{n \to \infty} (f_0(x_n) + f_0(y_n)) = \lim_{n \to \infty} f_0(x_n) + \lim_{n \to \infty} f_0(y_n) = f(x) + f(y). \]
Similarly,
\[ f(\lambda x) = \lim_{n \to \infty} f_0(\lambda x_n) = \lim_{n \to \infty} \lambda f_0(x_n) = \lambda \lim_{n \to \infty} f_0(x_n) = \lambda f(x). \]
Therefore, $f$ is linear. Since the norm is continuous, we have
\[ \|f(x)\| = \left\| \lim_{n \to \infty} f_0(x_n) \right\| = \lim_{n \to \infty} \|f_0(x_n)\|, \]
and since $f_0$ is continuous
\[ \|f_0(x_n)\| \leq \|f_0\| \|x_n\| \quad \text{for all } n \geq 1, \]
so we get
\[ \lim_{n \to \infty} \|f_0(x_n)\| \leq \lim_{n \to \infty} \|f_0\| \|x_n\| \quad \text{for all } n \geq 1, \]
that is,
\[ \|f(x)\| \leq \|f_0\| \|x\|. \]
Since
\[ \|f\| = \sup_{\|x\|=1, x \in E} \|f(x)\|, \]
we deduce that $\|f\| \leq \|f_0\|$. But since $E_0 \subseteq E$ and $f$ agrees with $f_0$ on $E_0$, we also have
\[ \|f_0\| = \sup_{\|x\|=1, x \in E_0} \|f_0(x)\| = \sup_{\|x\|=1, x \in E_0} \|f(x)\| \leq \sup_{\|x\|=1, x \in E} \|f(x)\| = \|f\|, \]
and thus $\|f\| = \|f_0\|$.

\section{Further Readings}

A thorough treatment of general topology can be found in Munkres [62, 61], Dixmier [25], Lang [53, 52], Schwartz [70, 69], and Bredon [13].
Chapter 16

Vector Norms and Matrix Norms

16.1 Normed Vector Spaces

In order to define how close two vectors or two matrices are, and in order to define the convergence of sequences of vectors or matrices, we can use the notion of a norm. Recall that $\mathbb{R}^+ = \{ x \in \mathbb{R} \mid x \geq 0 \}$. Also recall that if $z = a + ib \in \mathbb{C}$ is a complex number, with $a, b \in \mathbb{R}$, then $\bar{z} = a - ib$ and $|z| = \sqrt{z \bar{z}} = \sqrt{a^2 + b^2}$ ($|z|$ is the modulus of $z$).

**Definition 16.1.** Let $E$ be a vector space over a field $K$, where $K$ is either the field $\mathbb{R}$ of reals, or the field $\mathbb{C}$ of complex numbers. A norm on $E$ is a function $\| \| : E \rightarrow \mathbb{R}^+$, assigning a nonnegative real number $\|u\|$ to any vector $u \in E$, and satisfying the following conditions for all $x, y, z \in E$:

(N1) $\|x\| \geq 0$, and $\|x\| = 0$ iff $x = 0$. (positivity)
(N2) $\|\lambda x\| = |\lambda| \|x\|$. (homogeneity (or scaling))
(N3) $\|x + y\| \leq \|x\| + \|y\|$. (triangle inequality)

A vector space $E$ together with a norm $\| \|$ is called a normed vector space.

By (N2), setting $\lambda = -1$, we obtain

$$\|-x\| = \|(-1)x\| = |-1| \|x\| = \|x\|;$$
that is, $\|-x\| = \|x\|$. From (N3), we have

$$\|x\| = \|x - y + y\| \leq \|x - y\| + \|y\|,$$
which implies that

$$\|x\| - \|y\| \leq \|x - y\|.$$

By exchanging $x$ and $y$ and using the fact that by (N2),

$$\|y - x\| = \|-(x - y)\| = \|x - y\|,$$
we also have
\[ \|y\| - \|x\| \leq \|x - y\|. \]
Therefore,
\[ ||x|| - ||y|| \leq \|x - y\|, \quad \text{for all } x, y \in E. \tag{*} \]
Observe that setting \( \lambda = 0 \) in (N2), we deduce that \( \|0\| = 0 \) without assuming (N1). Then, by setting \( y = 0 \) in (*), we obtain
\[ ||x|| \leq \|x\|, \quad \text{for all } x \in E. \]
Therefore, the condition \( \|x\| \geq 0 \) in (N1) follows from (N2) and (N3), and (N1) can be replaced by the weaker condition
\[ (N1') \quad \text{For all } x \in E, \text{ if } \|x\| = 0 \text{ then } x = 0, \]
A function \( \| \| : E \to \mathbb{R} \) satisfying axioms (N2) and (N3) is called a \textit{semi-norm}. From the above discussion, a semi-norm also has the properties
\[ \|x\| \geq 0 \text{ for all } x \in E, \text{ and } \|0\| = 0. \]
However, there may be nonzero vectors \( x \in E \) such that \( \|x\| = 0 \). Let us give some examples of normed vector spaces.

\textbf{Example 16.1.}

1. Let \( E = \mathbb{R} \), and \( \|x\| = |x| \), the absolute value of \( x \).
2. Let \( E = \mathbb{C} \), and \( \|z\| = |z| \), the modulus of \( z \).
3. Let \( E = \mathbb{R}^n \) (or \( E = \mathbb{C}^n \)). There are three standard norms. For every \( (x_1, \ldots, x_n) \in E \), we have the norm \( \|x\|_1 \), defined such that,

\[ \|x\|_1 = |x_1| + \cdots + |x_n|, \]

we have the \textit{Euclidean norm} \( \|x\|_2 \), defined such that,
\[ \|x\|_2 = (|x_1|^2 + \cdots + |x_n|^2)^{\frac{1}{2}}, \]

and the \textit{sup-norm} \( \|x\|_\infty \), defined such that,
\[ \|x\|_\infty = \max\{|x_i| \mid 1 \leq i \leq n\}. \]

More generally, we define the \( \ell_p \)-norm (for \( p \geq 1 \)) by
\[ \|x\|_p = (|x_1|^p + \cdots + |x_n|^p)^{1/p}. \]

There are other norms besides the \( \ell_p \)-norms. Here are some examples.
1. For $E = \mathbb{R}^2$, 
\[ \| (u_1, u_2) \| = |u_1| + 2|u_2|. \]

2. For $E = \mathbb{R}^2$, 
\[ \| (u_1, u_2) \| = \left( (u_1 + u_2)^2 + u_1^2 \right)^{1/2}. \]

3. For $E = \mathbb{C}^2$, 
\[ \| (u_1, u_2) \| = |u_1 + iu_2| + |u_1 - iu_2|. \]

The reader should check that they satisfy all the axioms of a norm.

Some work is required to show the triangle inequality for the $\ell_p$-norm.

**Proposition 16.1.** If $E$ is a finite-dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$, for every real number $p \geq 1$, the $\ell_p$-norm is indeed a norm.

**Proof.** The cases $p = 1$ and $p = \infty$ are easy and left to the reader. If $p > 1$, then let $q > 1$ such that
\[ \frac{1}{p} + \frac{1}{q} = 1. \]

We will make use of the following fact: for all $\alpha, \beta \in \mathbb{R}$, if $\alpha, \beta \geq 0$, then
\[ \alpha \beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}. \] (*)

To prove the above inequality, we use the fact that the exponential function $t \mapsto e^t$ satisfies the following convexity inequality:
\[ e^{\theta x + (1-\theta)y} \leq \theta e^x + (1 - \theta)e^y, \]
for all $x, y \in \mathbb{R}$ and all $\theta$ with $0 \leq \theta \leq 1$.

Since the case $\alpha \beta = 0$ is trivial, let us assume that $\alpha > 0$ and $\beta > 0$. If we replace $\theta$ by $1/p$, $x$ by $p \log \alpha$ and $y$ by $q \log \beta$, then we get
\[ e^{1/p \log \alpha + 1/q \log \beta} \leq \frac{1}{p} e^{p \log \alpha} + \frac{1}{q} e^{q \log \beta}, \]
which simplifies to
\[ \alpha \beta \leq \frac{\alpha^p}{p} + \frac{\beta^q}{q}, \]
as claimed.

We will now prove that for any two vectors $u, v \in E$, we have
\[ \sum_{i=1}^{n} |u_i v_i| \leq \|u\|_p \|v\|_q. \] (**)

Since the above is trivial if \( u = 0 \) or \( v = 0 \), let us assume that \( u \neq 0 \) and \( v \neq 0 \). Then, the inequality (*) with \( \alpha = |u_i|/\|u\|_p \) and \( \beta = |v_i|/\|v\|_q \) yields
\[
\frac{|u_i v_i|}{\|u\|_p \|v\|_q} \leq \frac{|u_i|^p}{p \|u\|_p^p} + \frac{|v_i|^q}{q \|v\|_q^q},
\]
for \( i = 1, \ldots, n \), and by summing up these inequalities, we get
\[
\sum_{i=1}^n |u_i v_i| \leq \|u\|_p \|v\|_q,
\]
as claimed. To finish the proof, we simply have to prove that property (N3) holds, since (N1) and (N2) are clear. Now, for \( i = 1, \ldots, n \), we can write
\[
(|u_i| + |v_i|)^p = |u_i|(|u_i| + |v_i|)^{p-1} + |v_i|(|u_i| + |v_i|)^{p-1},
\]
so that by summing up these equations we get
\[
\sum_{i=1}^n (|u_i| + |v_i|)^p = \sum_{i=1}^n |u_i|(|u_i| + |v_i|)^{p-1} + \sum_{i=1}^n |v_i|(|u_i| + |v_i|)^{p-1},
\]
and using the inequality (**), we get
\[
\sum_{i=1}^n (|u_i| + |v_i|)^p \leq (\|u\|_p + \|v\|_p) \left( \sum_{i=1}^n (|u_i| + |v_i|)^{(p-1)q} \right)^{1/q}.
\]
However, \( 1/p + 1/q = 1 \) implies \( pq = p + q \), that is, \( (p - 1)q = p \), so we have
\[
\sum_{i=1}^n (|u_i| + |v_i|)^p \leq (\|u\|_p + \|v\|_p) \left( \sum_{i=1}^n (|u_i| + |v_i|)^p \right)^{1/q},
\]
which yields
\[
\left( \sum_{i=1}^n (|u_i| + |v_i|)^p \right)^{1/p} \leq \|u\|_p + \|v\|_p.
\]
Since \( |u_i + v_i| \leq |u_i| + |v_i| \), the above implies the triangle inequality \( \|u + v\|_p \leq \|u\|_p + \|v\|_p \), as claimed.

For \( p > 1 \) and \( 1/p + 1/q = 1 \), the inequality
\[
\sum_{i=1}^n |u_i v_i| \leq \left( \sum_{i=1}^n |u_i|^p \right)^{1/p} \left( \sum_{i=1}^n |v_i|^q \right)^{1/q}
\]
is known as Hölder’s inequality. For \( p = 2 \), it is the Cauchy–Schwarz inequality.
Actually, if we define the Hermitian inner product \( \langle -, - \rangle \) on \( \mathbb{C}^n \) by
\[
\langle u, v \rangle = \sum_{i=1}^{n} u_i \overline{v}_i,
\]
where \( u = (u_1, \ldots, u_n) \) and \( v = (v_1, \ldots, v_n) \), then
\[
|\langle u, v \rangle| \leq \sum_{i=1}^{n} |u_i \overline{v}_i| = \sum_{i=1}^{n} |u_i v_i|,
\]
so Hölder’s inequality implies the inequality
\[
|\langle u, v \rangle| \leq \|u\|_p \|v\|_q
\]
also called Hölder’s inequality, which, for \( p = 2 \) is the standard Cauchy–Schwarz inequality. The triangle inequality for the \( \ell_p \)-norm,
\[
\left( \sum_{i=1}^{n} (|u_i + v_i|)^p \right)^{1/p} \leq \left( \sum_{i=1}^{n} |u_i|^p \right)^{1/p} + \left( \sum_{i=1}^{n} |v_i|^q \right)^{1/q},
\]
is known as Minkowski’s inequality.

When we restrict the Hermitian inner product to real vectors, \( u, v \in \mathbb{R}^n \), we get the Euclidean inner product
\[
\langle u, v \rangle = \sum_{i=1}^{n} u_i v_i.
\]
It is very useful to observe that if we represent (as usual) \( u = (u_1, \ldots, u_n) \) and \( v = (v_1, \ldots, v_n) \) (in \( \mathbb{R}^n \)) by column vectors, then their Euclidean inner product is given by
\[
\langle u, v \rangle = u^\top v = v^\top u,
\]
and when \( u, v \in \mathbb{C}^n \), their Hermitian inner product is given by
\[
\langle u, v \rangle = v^* u = \overline{u}^* v.
\]
In particular, when \( u = v \), in the complex case we get
\[
\|u\|_2^2 = u^* u,
\]
and in the real case, this becomes
\[
\|u\|_2^2 = u^\top u.
\]
As convenient as these notations are, we still recommend that you do not abuse them; the notation \( \langle u, v \rangle \) is more intrinsic and still “works” when our vector space is infinite dimensional.

The following proposition is easy to show.
Proposition 16.2. The following inequalities hold for all \( x \in \mathbb{R}^n \) (or \( x \in \mathbb{C}^n \)):

\[
\|x\|_\infty \leq \|x\|_1 \leq n \|x\|_\infty,
\|x\|_\infty \leq \|x\|_2 \leq \sqrt{n} \|x\|_\infty,
\|x\|_2 \leq \|x\|_1 \leq \sqrt{n} \|x\|_2.
\]

Proposition 16.2 is a special case of a very important result: in a finite-dimensional vector space, any two norms are equivalent.

Definition 16.2. Given any (real or complex) vector space \( E \), two norms \( \|\cdot\|_a \) and \( \|\cdot\|_b \) are \textit{equivalent} iff there exists some positive reals \( C_1, C_2 > 0 \), such that

\[
\|u\|_a \leq C_1 \|u\|_b \text{ and } \|u\|_b \leq C_2 \|u\|_a, \text{ for all } u \in E.
\]

Given any norm \( \|\cdot\| \) on a vector space of dimension \( n \), for any basis \( (e_1, \ldots, e_n) \) of \( E \), observe that for any vector \( x = x_1 e_1 + \cdots + x_n e_n \), we have

\[
\|x\| = \|x_1 e_1 + \cdots + x_n e_n\| \leq |x_1| \|e_1\| + \cdots + |x_n| \|e_n\| \leq C(|x_1| + \cdots + |x_n|) = C \|x\|_1,
\]

with \( C = \max_{1 \leq i \leq n} \|e_i\| \) and

\[
\|x\|_1 = \|x_1 e_1 + \cdots + x_n e_n\| = |x_1| + \cdots + |x_n|.
\]

The above implies that

\[
\|u\| - \|v\| \leq \|u - v\| \leq C \|u - v\|_1,
\]

which means that the map \( u \mapsto \|u\| \) is \textit{continuous} with respect to the norm \( \|\cdot\|_1 \).

Let \( S_1^{n-1} \) be the unit sphere with respect to the norm \( \|\cdot\|_1 \), namely

\[
S_1^{n-1} = \{ x \in E \mid \|x\|_1 = 1 \}.
\]

Now, \( S_1^{n-1} \) is a closed and bounded subset of a finite-dimensional vector space, so by Heine–Borel (or equivalently, by Bolzano–Weierstrass), \( S_1^{n-1} \) is compact. On the other hand, it is a well known result of analysis that any continuous real-valued function on a nonempty compact set has a minimum and a maximum, and that they are achieved. Using these facts, we can prove the following important theorem:

Theorem 16.3. If \( E \) is any real or complex vector space of finite dimension, then any two norms on \( E \) are equivalent.

Proof. It is enough to prove that any norm \( \|\cdot\| \) is equivalent to the 1-norm. We already proved that the function \( x \mapsto \|x\| \) is continuous with respect to the norm \( \|\cdot\|_1 \) and we observed that the unit sphere \( S_1^{n-1} \) is compact. Now, we just recalled that because the function \( f : x \mapsto \|x\| \) is continuous and because \( S_1^{n-1} \) is compact, the function \( f \) has a minimum \( m \) and a maximum
16.2. MATRIX NORMS

$M$, and because $\|x\|$ is never zero on $S_1^{n-1}$, we must have $m > 0$. Consequently, we just proved that if $\|x\|_1 = 1$, then

$$0 < m \leq \|x\| \leq M,$$

so for any $x \in E$ with $x \neq 0$, we get

$$m \leq \|x/\|x\|_1\| \leq M,$$

which implies

$$m \|x\|_1 \leq \|x\| \leq M \|x\|_1.$$

Since the above inequality holds trivially if $x = 0$, we just proved that $\|\|$ and $\|\|_1$ are equivalent, as claimed. $\square$

Next, we will consider norms on matrices.

16.2 Matrix Norms

For simplicity of exposition, we will consider the vector spaces $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$ of square $n \times n$ matrices. Most results also hold for the spaces $M_{m,n}(\mathbb{R})$ and $M_{m,n}(\mathbb{C})$ of rectangular $m \times n$ matrices. Since $n \times n$ matrices can be multiplied, the idea behind matrix norms is that they should behave “well” with respect to matrix multiplication.

**Definition 16.3.** A matrix norm $\|\|$ on the space of square $n \times n$ matrices in $M_n(K)$, with $K = \mathbb{R}$ or $K = \mathbb{C}$, is a norm on the vector space $M_n(K)$, with the additional property called submultiplicativity that

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

for all $A, B \in M_n(K)$. A norm on matrices satisfying the above property is often called a submultiplicative matrix norm.

Since $I^2 = I$, from $\|I\| = \|I^2\| \leq \|I\|^2$, we get $\|I\| \geq 1$, for every matrix norm.

Before giving examples of matrix norms, we need to review some basic definitions about matrices. Given any matrix $A = (a_{ij}) \in M_{m,n}(\mathbb{C})$, the conjugate $\overline{A}$ of $A$ is the matrix such that

$$\overline{A}_{ij} = \overline{a}_{ij}, \quad 1 \leq i \leq m, \ 1 \leq j \leq n.$$

The transpose of $A$ is the $n \times m$ matrix $A^\top$ such that

$$A_{ij}^\top = a_{ji}, \quad 1 \leq i \leq m, \ 1 \leq j \leq n.$$

The adjoint of $A$ is the $n \times m$ matrix $A^*$ such that

$$A^* = (A^\top)^\top = (\overline{A})^\top.$$
When $A$ is a real matrix, $A^* = A^\top$. A matrix $A \in M_n(\mathbb{C})$ is *Hermitian* if
\[
A^* = A.
\]
If $A$ is a real matrix ($A \in M_n(\mathbb{R})$), we say that $A$ is *symmetric* if
\[
A^\top = A.
\]
A matrix $A \in M_n(\mathbb{C})$ is *normal* if
\[
AA^* = A^*A,
\]
and if $A$ is a real matrix, it is *normal* if
\[
AA^\top = A^\top A.
\]
A matrix $U \in M_n(\mathbb{C})$ is *unitary* if
\[
UU^* = U^*U = I.
\]
A real matrix $Q \in M_n(\mathbb{R})$ is *orthogonal* if
\[
QQ^\top = Q^\top Q = I.
\]

Given any matrix $A = (a_{ij}) \in M_n(\mathbb{C})$, the *trace* $\text{tr}(A)$ of $A$ is the sum of its diagonal elements
\[
\text{tr}(A) = a_{11} + \cdots + a_{nn}.
\]
It is easy to show that the trace is a linear map, so that
\[
\text{tr}(\lambda A) = \lambda \text{tr}(A)
\]
and
\[
\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B).
\]
Moreover, if $A$ is an $m \times n$ matrix and $B$ is an $n \times m$ matrix, it is not hard to show that
\[
\text{tr}(AB) = \text{tr}(BA).
\]

We also review eigenvalues and eigenvectors. We content ourselves with definition involving matrices.

**Definition 16.4.** Given any square matrix $A \in M_n(\mathbb{C})$, a complex number $\lambda \in \mathbb{C}$ is an *eigenvalue* of $A$ if there is some nonzero vector $u \in \mathbb{C}^n$, such that
\[
Au = \lambda u.
\]
If $\lambda$ is an eigenvalue of $A$, then the nonzero vectors $u \in \mathbb{C}^n$ such that $Au = \lambda u$ are called *eigenvectors of $A$ associated with $\lambda$*; together with the zero vector, these eigenvectors form a subspace of $\mathbb{C}^n$ denoted by $E_\lambda(A)$, and called the *eigenspace associated with $\lambda$*. 
Remark: Note that Definition 16.4 requires an eigenvector to be nonzero. A somewhat unfortunate consequence of this requirement is that the set of eigenvectors is not a subspace, since the zero vector is missing! On the positive side, whenever eigenvectors are involved, there is no need to say that they are nonzero. The fact that eigenvectors are nonzero is implicitly used in all the arguments involving them, so it seems safer (but perhaps not as elegant) to stipulate that eigenvectors should be nonzero.

If $A$ is a square real matrix $A \in M_n(\mathbb{R})$, then we restrict Definition 16.4 to real eigenvalues $\lambda \in \mathbb{R}$ and real eigenvectors. However, it should be noted that although every complex matrix always has at least some complex eigenvalue, a real matrix may not have any real eigenvalues. For example, the matrix

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

has the complex eigenvalues $i$ and $-i$, but no real eigenvalues. Thus, typically, even for real matrices, we consider complex eigenvalues.

Observe that $\lambda \in \mathbb{C}$ is an eigenvalue of $A$

iff $Au = \lambda u$ for some nonzero vector $u \in \mathbb{C}^n$

iff $(\lambda I - A)u = 0$

iff the matrix $\lambda I - A$ defines a linear map which has a nonzero kernel, that is,

iff $\lambda I - A$ not invertible.

However, it is a standard fact of linear algebra that $\lambda I - A$ is not invertible iff

$$\det(\lambda I - A) = 0.$$

Now, $\det(\lambda I - A)$ is a polynomial of degree $n$ in the indeterminate $\lambda$, in fact, of the form

$$\lambda^n - \operatorname{tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A).$$

Thus, we see that the eigenvalues of $A$ are the zeros (also called roots) of the above polynomial. Since every complex polynomial of degree $n$ has exactly $n$ roots, counted with their multiplicity, we have the following definition:

**Definition 16.5.** Given any square $n \times n$ matrix $A \in M_n(\mathbb{C})$, the polynomial

$$\det(\lambda I - A) = \lambda^n - \operatorname{tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A)$$

is called the characteristic polynomial of $A$. The $n$ (not necessarily distinct) roots $\lambda_1, \ldots, \lambda_n$ of the characteristic polynomial are all the eigenvalues of $A$ and constitute the spectrum of $A$. We let

$$\rho(A) = \max_{1 \leq i \leq n} |\lambda_i|$$

be the largest modulus of the eigenvalues of $A$, called the spectral radius of $A$. 
Since the eigenvalue $\lambda_1, \ldots, \lambda_n$ of $A$ are the zeros of the polynomial
$$\det(\lambda I - A) = \lambda^n - \text{tr}(A)\lambda^{n-1} + \cdots + (-1)^n \det(A),$$
we deduce that
$$\text{tr}(A) = \lambda_1 + \cdots + \lambda_n$$
and
$$\det(A) = \lambda_1 \cdots \lambda_n.$$

**Proposition 16.4.** For any matrix norm $\|\cdot\|$ on $M_n(\mathbb{C})$ and for any square $n \times n$ matrix $A \in M_n(\mathbb{C})$, we have
$$\rho(A) \leq \|A\|.$$

**Proof.** Let $\lambda$ be some eigenvalue of $A$ for which $|\lambda|$ is maximum, that is, such that $|\lambda| = \rho(A)$. If $u (\neq 0)$ is any eigenvector associated with $\lambda$ and if $U$ is the $n \times n$ matrix whose columns are all $u$, then $Au = \lambda u$ implies
$$AU = \lambda U,$$
and since
$$|\lambda| \|U\| = \|\lambda U\| = \|AU\| \leq \|A\| \|U\|$$
and $U \neq 0$, we have $\|U\| \neq 0$, and get
$$\rho(A) = |\lambda| \leq \|A\|,$$
as claimed. \qed

Proposition 16.4 also holds for any real matrix norm $\|\cdot\|$ on $M_n(\mathbb{R})$ but the proof is more subtle and requires the notion of induced norm. We prove it after giving Definition 16.7.

Now, it turns out that if $A$ is a real $n \times n$ symmetric matrix, then the eigenvalues of $A$ are all real and there is some orthogonal matrix $Q$ such that
$$A = Q \text{diag}(\lambda_1, \ldots, \lambda_n)Q^\top,$$
where $\text{diag}(\lambda_1, \ldots, \lambda_n)$ denotes the matrix whose only nonzero entries (if any) are its diagonal entries, which are the (real) eigenvalues of $A$. Similarly, if $A$ is a complex $n \times n$ Hermitian matrix, then the eigenvalues of $A$ are all real and there is some unitary matrix $U$ such that
$$A = U \text{diag}(\lambda_1, \ldots, \lambda_n)U^*,$$
where $\text{diag}(\lambda_1, \ldots, \lambda_n)$ denotes the matrix whose only nonzero entries (if any) are its diagonal entries, which are the (real) eigenvalues of $A$.

We now return to matrix norms. We begin with the so-called Frobenius norm, which is just the norm $\|\cdot\|_2$ on $\mathbb{C}^{n^2}$, where the $n \times n$ matrix $A$ is viewed as the vector obtained by concatenating together the rows (or the columns) of $A$. The reader should check that for any $n \times n$ complex matrix $A = (a_{ij})$,
$$\left(\sum_{i,j=1}^n |a_{ij}|^2\right)^{1/2} = \sqrt{\text{tr}(A^*A)} = \sqrt{\text{tr}(AA^*)}.$$
Definition 16.6. The Frobenius norm $\| \cdot \|_F$ is defined so that for every square $n \times n$ matrix $A \in M_n(\mathbb{C})$, 

$$
\|A\|_F = \left( \sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2} = \sqrt{\text{tr}(AA^*)} = \sqrt{\text{tr}(A^*A)}.
$$

The following proposition show that the Frobenius norm is a matrix norm satisfying other nice properties.

Proposition 16.5. The Frobenius norm $\| \cdot \|_F$ on $M_n(\mathbb{C})$ satisfies the following properties:

1. It is a matrix norm; that is, $\|AB\|_F \leq \|A\|_F \|B\|_F$, for all $A,B \in M_n(\mathbb{C})$.

2. It is unitarily invariant, which means that for all unitary matrices $U,V$, we have 

$$
\|A\|_F = \|UA\|_F = \|AV\|_F = \|UAV\|_F.
$$

3. $\sqrt{\rho(A^*A)} \leq \|A\|_F \leq \sqrt{n\rho(A^*A)}$, for all $A \in M_n(\mathbb{C})$.

Proof. (1) The only property that requires a proof is the fact $\|AB\|_F \leq \|A\|_F \|B\|_F$. This follows from the Cauchy–Schwarz inequality:

$$
\|AB\|_F^2 = \sum_{i,j=1}^{n} \left| \sum_{k=1}^{n} a_{ik}b_{kj} \right|^2 
\leq \sum_{i,j=1}^{n} \left( \sum_{h=1}^{n} |a_{ih}|^2 \right) \left( \sum_{k=1}^{n} |b_{kj}|^2 \right) 
= \left( \sum_{i,h=1}^{n} |a_{ih}|^2 \right) \left( \sum_{k,j=1}^{n} |b_{kj}|^2 \right) = \|A\|_F^2 \|B\|_F^2.
$$

(2) We have 

$$
\|A\|_F^2 = \text{tr}(A^*A) = \text{tr}(VV^*A^*A) = \text{tr}(V^*A^*AV) = \|AV\|_F^2,
$$

and 

$$
\|A\|_F^2 = \text{tr}(A^*A) = \text{tr}(A^*U^*UA) = \|UA\|_F^2.
$$

The identity 

$$
\|A\|_F = \|UA\|_F
$$

follows from the previous two.

(3) It is well known that the trace of a matrix is equal to the sum of its eigenvalues. Furthermore, $A^*A$ is symmetric positive semidefinite (which means that its eigenvalues are nonnegative), so $\rho(A^*A)$ is the largest eigenvalue of $A^*A$ and 

$$
\rho(A^*A) \leq \text{tr}(A^*A) \leq n\rho(A^*A),
$$

which yields (3) by taking square roots. \[\square\]
Remark: The Frobenius norm is also known as the Hilbert-Schmidt norm or the Schur norm. So many famous names associated with such a simple thing!

We now give another method for obtaining matrix norms using subordinate norms. First, we need a proposition that shows that in a finite-dimensional space, the linear map induced by a matrix is bounded, and thus continuous.

**Proposition 16.6.** For every norm $\| \cdot \|$ on $\mathbb{C}^n$ (or $\mathbb{R}^n$), for every matrix $A \in M_n(\mathbb{C})$ (or $A \in M_n(\mathbb{R})$), there is a real constant $C_A \geq 0$, such that

$$\|Au\| \leq C_A \|u\|,$$

for every vector $u \in \mathbb{C}^n$ (or $u \in \mathbb{R}^n$ if $A$ is real). 

**Proof.** For every basis $(e_1, \ldots, e_n)$ of $\mathbb{C}^n$ (or $\mathbb{R}^n$), for every vector $u = u_1e_1 + \cdots + u_ne_n$, we have

$$\|Au\| = \|u_1A(e_1) + \cdots + u_nA(e_n)\|$$

$$\leq |u_1|\|A(e_1)\| + \cdots + |u_n|\|A(e_n)\|$$

$$\leq C_1(|u_1| + \cdots + |u_n|) = C_1\|u\|_1,$$

where $C_1 = \max_{1 \leq i \leq n} \|A(e_i)\|$. By Theorem 16.3, the norms $\| \cdot \|$ and $\| \cdot \|_1$ are equivalent, so there is some constant $C_2 > 0$ so that $\|u\|_1 \leq C_2 \|u\|$ for all $u$, which implies that

$$\|Au\| \leq C_A \|u\|,$$

where $C_A = C_1C_2$. \(\square\)

Proposition 16.6 says that every linear map on a finite-dimensional space is bounded. This implies that every linear map on a finite-dimensional space is continuous. Actually, it is not hard to show that a linear map on a normed vector space $E$ is bounded iff it is continuous, regardless of the dimension of $E$.

Proposition 16.6 implies that for every matrix $A \in M_n(\mathbb{C})$ (or $A \in M_n(\mathbb{R})$),

$$\sup_{x \in \mathbb{C}^n \atop x \neq 0} \frac{\|Ax\|}{\|x\|} \leq C_A.$$

Now, since $\|\lambda u\| = |\lambda| \|u\|$, for every nonzero vector $x$, we have

$$\|Ax\| \|x\| = \|x\| \|A(x/\|x\|)\| = \|A(x/\|x\|)\|,$$

which implies that

$$\sup_{x \in \mathbb{C}^n \atop x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{x \in \mathbb{C}^n \atop \|x\|=1} \|Ax\|.$$
Similarly
\[
\sup_{x \in \mathbb{C}^n, x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{x \in \mathbb{C}^n, \|x\|=1} \|Ax\|.
\]
The above considerations justify the following definition.

**Definition 16.7.** If \( \| \| \) is any norm on \( \mathbb{C}^n \), we define the function \( \| \| \) on \( M_n(\mathbb{C}) \) by
\[
\|A\| = \sup_{x \in \mathbb{C}^n, x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{x \in \mathbb{C}^n, \|x\|=1} \|Ax\|.
\]

The function \( A \mapsto \|A\| \) is called the *subordinate matrix norm* or *operator norm* induced by the norm \( \| \| \).

It is easy to check that the function \( A \mapsto \|A\| \) is indeed a norm, and by definition, it satisfies the property
\[
\|Ax\| \leq \|A\| \|x\|, \quad \text{for all } x \in \mathbb{C}^n.
\]
A norm \( \| \| \) on \( M_n(\mathbb{C}) \) satisfying the above property is said to be *subordinate* to the vector norm \( \| \| \) on \( \mathbb{C}^n \). As a consequence of the above inequality, we have
\[
\|ABx\| \leq \|A\| \|Bx\| \leq \|A\| \|B\| \|x\|,
\]
for all \( x \in \mathbb{C}^n \), which implies that
\[
\|AB\| \leq \|A\| \|B\| \quad \text{for all } A, B \in M_n(\mathbb{C}),
\]
showing that \( A \mapsto \|A\| \) is a matrix norm (it is submultiplicative).

Observe that the operator norm is also defined by
\[
\|A\| = \inf \{ \lambda \in \mathbb{R} \mid \|Ax\| \leq \lambda \|x\|, \text{for all } x \in \mathbb{C}^n \}.
\]
Since the function \( x \mapsto \|Ax\| \) is continuous (because \( \|Ay\| - \|Ax\| \leq \|Ay - Ax\| \leq C_A \|x - y\| \)) and the unit sphere \( S^{n-1} = \{ x \in \mathbb{C}^n \mid \|x\| = 1 \} \) is compact, there is some \( x \in \mathbb{C}^n \) such that \( \|x\| = 1 \) and
\[
\|Ax\| = \|A\|.
\]
Equivalently, there is some \( x \in \mathbb{C}^n \) such that \( x \neq 0 \) and
\[
\|Ax\| = \|A\| \|x\|.
\]
The definition of an operator norm also implies that
\[
\|I\| = 1.
\]
The above shows that the Frobenius norm is not a subordinate matrix norm (why?). The notion of subordinate norm can be slightly generalized.
Definition 16.8. If $K = \mathbb{R}$ or $K = \mathbb{C}$, for any norm $\|\|$ on $M_{m,n}(K)$, and for any two norms $\|\|_a$ on $K^n$ and $\|\|_b$ on $K^m$, we say that the norm $\|\|$ is \emph{subordinate} to the norms $\|\|_a$ and $\|\|_b$ if
\[\|Ax\|_b \leq \|A\| \|x\|_a \quad \text{for all } A \in M_{m,n}(K) \text{ and all } x \in K^n.\]

Remark: For any norm $\|\|$ on $\mathbb{C}^n$, we can define the function $\|\|_R$ on $M_n(\mathbb{R})$ by
\[\|A\|_R = \sup_{x \in \mathbb{R}^n, x \neq 0} \frac{\|Ax\|}{\|x\|} = \sup_{x \in \mathbb{R}^n, \|x\|=1} \|Ax\|.
\]
The function $A \mapsto \|A\|_R$ is a matrix norm on $M_n(\mathbb{R})$, and
\[\|A\|_R \leq \|A\|,
\]
for all real matrices $A \in M_n(\mathbb{R})$. However, it is possible to construct vector norms $\|\|$ on $\mathbb{C}^n$ and real matrices $A$ such that $\|A\|_R < \|A\|$. In order to avoid this kind of difficulties, we define subordinate matrix norms over $M_n(\mathbb{C})$. Luckily, it turns out that $\|A\|_R = \|A\|$ for the vector norms, $\|\|_1$, $\|\|_2$, and $\|\|_\infty$.

We now prove Proposition 16.4 for real matrix norms.

Proposition 16.7. For any matrix norm $\|\|$ on $M_n(\mathbb{R})$ and for any square $n \times n$ matrix $A \in M_n(\mathbb{R})$, we have
\[\rho(A) \leq \|A\|.
\]
Proof. We follow the proof in Denis Serre’s book [74]. If $A$ is a real matrix, the problem is that the eigenvectors associated with the eigenvalue of maximum modulus may be complex. We use a trick based on the fact that for every matrix $A$ (real or complex),
\[\rho(A^k) = (\rho(A))^k,\]
which is left as an exercise (it can be shown that if $(\lambda_1, \ldots, \lambda_n)$ are the (not necessarily distinct) eigenvalues of $A$, then $(\lambda_1^k, \ldots, \lambda_n^k)$ are the eigenvalues of $A^k$, for $k \geq 1$).

Pick any complex matrix norm $\|\|_c$ on $\mathbb{C}^n$ (for example, the Frobenius norm, or any subordinate matrix norm induced by a norm on $\mathbb{C}^n$). The restriction of $\|\|_c$ to real matrices is a real norm that we also denote by $\|\|_c$. Now, by Theorem 16.3, since $M_n(\mathbb{R})$ has finite dimension $n^2$, there is some constant $C > 0$ so that
\[\|B\|_c \leq C \|B\|, \quad \text{for all } B \in M_n(\mathbb{R}).\]
Furthermore, for every $k \geq 1$ and for every real $n \times n$ matrix $A$, by Proposition 16.4, $\rho(A^k) \leq \|A^k\|_c$, and because $\|\|$ is a matrix norm, $\|A^k\| \leq \|A\|^k$, so we have
\[(\rho(A))^k = \rho(A^k) \leq \|A^k\|_c \leq C \|A^k\| \leq C \|A\|^k,
\]
for all \( k \geq 1 \). It follows that
\[
\rho(A) \leq C^{1/k} \|A\|, \quad \text{for all} \quad k \geq 1.
\]
However because \( C > 0 \), we have \( \lim_{k \to \infty} C^{1/k} = 1 \) (we have \( \lim_{k \to \infty} \frac{1}{k} \log(C) = 0 \)). Therefore, we conclude that
\[
\rho(A) \leq \|A\|,
\]
as desired. \(\square\)

We now determine explicitly what are the subordinate matrix norms associated with the vector norms \( \|\|_1 \), \( \|\|_2 \), and \( \|\|_\infty \).

**Proposition 16.8.** For every square matrix \( A = (a_{ij}) \in M_n(\mathbb{C}) \), we have
\[
\|A\|_1 = \sup_{x \in \mathbb{C}^n, \|x\|_1 = 1} \|Ax\|_1 = \max_j \sum_{i=1}^n |a_{ij}|
\]
\[
\|A\|_\infty = \sup_{x \in \mathbb{C}^n, \|x\|_\infty = 1} \|Ax\|_\infty = \max_i \sum_{j=1}^n |a_{ij}|
\]
\[
\|A\|_2 = \sup_{x \in \mathbb{C}^n, \|x\|_2 = 1} \|Ax\|_2 = \sqrt{\rho(A^*A)} = \sqrt{\rho(AA^*)}.
\]

Furthermore, \( \|A^*\|_2 = \|A\|_2 \), the norm \( \|\|_2 \) is unitarily invariant, which means that
\[
\|A\|_2 = \|UAV\|_2
\]
for all unitary matrices \( U, V \), and if \( A \) is a normal matrix, then \( \|A\|_2 = \rho(A) \).

**Proof.** For every vector \( u \), we have
\[
\|Au\|_1 = \sum_i \left| \sum_j a_{ij} u_j \right| \leq \sum_j |a_j| \sum_i |a_{ij}| \leq \left( \max_j \sum_i |a_{ij}| \right) \|u\|_1,
\]
which implies that
\[
\|A\|_1 \leq \max_j \sum_{i=1}^n |a_{ij}|.
\]
It remains to show that equality can be achieved. For this let \( j_0 \) be some index such that
\[
\max_j \sum_i |a_{ij}| = \sum_i |a_{ij_0}|,
\]
and let \( u_i = 0 \) for all \( i \neq j_0 \) and \( u_{j_0} = 1 \).
In a similar way, we have

\[ \|Au\|_\infty = \max_i \left| \sum_j a_{ij} u_j \right| \leq \left( \max_i \sum_j |a_{ij}| \right) \|u\|_\infty, \]

which implies that

\[ \|A\|_\infty \leq \max_i \sum_{j=1}^n |a_{ij}|. \]

To achieve equality, let \( i_0 \) be some index such that

\[ \max_i \sum_j |a_{ij}| = \sum_j |a_{i_0j}|. \]

The reader should check that the vector given by

\[ u_j = \begin{cases} \frac{a_{i_0j}}{|a_{i_0j}|} & \text{if } a_{i_0j} \neq 0 \\ 1 & \text{if } a_{i_0j} = 0 \end{cases} \]

works.

We have

\[ \|A\|_2^2 = \sup_{x \in \mathbb{C}^n} \|Ax\|_2^2 = \sup_{x^*x = 1} x^* A^*Ax. \]

Since the matrix \( A^*A \) is symmetric, it has real eigenvalues and it can be diagonalized with respect to an orthogonal matrix. These facts can be used to prove that the function \( x \mapsto x^* A^*Ax \) has a maximum on the sphere \( x^*x = 1 \) equal to the largest eigenvalue of \( A^*A \), namely, \( \rho(A^*A) \). We postpone the proof until we discuss optimizing quadratic functions. Therefore,

\[ \|A\|_2 = \sqrt{\rho(A^*A)}. \]

Let use now prove that \( \rho(A^*A) = \rho(AA^*) \). First, assume that \( \rho(A^*A) > 0 \). In this case, there is some eigenvector \( u \neq 0 \) such that

\[ A^*Au = \rho(A^*A)u, \]

and since \( \rho(A^*A) > 0 \), we must have \( Au \neq 0 \). Since \( Au \neq 0 \),

\[ AA^*(Au) = \rho(A^*A)Au \]

which means that \( \rho(A^*A) \) is an eigenvalue of \( AA^* \), and thus

\[ \rho(A^*A) \leq \rho(AA^*). \]

Because \( (A^*)^* = A \), by replacing \( A \) by \( A^* \), we get

\[ \rho(AA^*) \leq \rho(A^*A), \]
and so \( \rho(A^*A) = \rho(AA^*) \).

If \( \rho(A^*A) = 0 \), then we must have \( \rho(AA^*) = 0 \), since otherwise by the previous reasoning we would have \( \rho(A^*A) = \rho(AA^*) > 0 \). Hence, in all case

\[
\|A\|_2^2 = \rho(A^*A) = \rho(AA^*) = \|A^*\|_2^2.
\]

For any unitary matrices \( U \) and \( V \), it is an easy exercise to prove that \( V^*A^*AV \) and \( A^*A \) have the same eigenvalues, so

\[
\|A\|_2^2 = \rho(A^*A) = \rho(V^*A^*AV) = \|AV\|_2^2,
\]

and also

\[
\|A\|_2^2 = \rho(A^*A) = \rho(A^*U^*UA) = \|UA\|_2^2.
\]

Finally, if \( A \) is a normal matrix \( (AA^* = A^*A) \), it can be shown that there is some unitary matrix \( U \) so that

\[
A = UDU^*,
\]

where \( D = \text{diag}(\lambda_1, \ldots, \lambda_n) \) is a diagonal matrix consisting of the eigenvalues of \( A \), and thus

\[
A^*A = (UDU^*)^*UDU^* = UD^*U^*UDU^* = UD^*DU^*.
\]

However, \( D^*D = \text{diag}(|\lambda_1|^2, \ldots, |\lambda_n|^2) \), which proves that

\[
\rho(A^*A) = \rho(D^*D) = \max_i |\lambda_i|^2 = (\rho(A))^2,
\]

so that \( \|A\|_2 = \rho(A) \).

The norm \( \|A\|_2 \) is often called the spectral norm. Observe that property (3) of proposition 16.5 says that

\[
\|A\|_2 \leq \|A\|_F \leq \sqrt{n} \|A\|_2,
\]

which shows that the Frobenius norm is an upper bound on the spectral norm. The Frobenius norm is much easier to compute than the spectral norm.

The reader will check that the above proof still holds if the matrix \( A \) is real, confirming the fact that \( \|A\|_R = \|A\| \) for the vector norms \( \|\|_1, \|\|_2 \), and \( \|\|_\infty \). It is also easy to verify that the proof goes through for rectangular matrices, with the same formulae. Similarly, the Frobenius norm is also a norm on rectangular matrices. For these norms, whenever \( AB \) makes sense, we have

\[
\|AB\| \leq \|A\| \|B\|.
\]

**Remark:** Let \( (E, \|\|) \) and \( (F, \|\|) \) be two normed vector spaces (for simplicity of notation, we use the same symbol \( \|\| \) for the norms on \( E \) and \( F \); this should not cause any confusion).
Recall that a function $f : E \to F$ is continuous if for every $a \in E$, for every $\epsilon > 0$, there is some $\eta > 0$ such that for all $x \in E$,

$$\text{if } \|x - a\| \leq \eta \text{ then } \|f(x) - f(a)\| \leq \epsilon.$$ 

It is not hard to show that a linear map $f : E \to F$ is continuous iff there is some constant $C \geq 0$ such that

$$\|f(x)\| \leq C \|x\| \text{ for all } x \in E.$$ 

If so, we say that $f$ is bounded (or a linear bounded operator). We let $\mathcal{L}(E; F)$ denote the set of all continuous (equivalently, bounded) linear maps from $E$ to $F$. Then, we can define the operator norm (or subordinate norm) $\|\|$ on $\mathcal{L}(E; F)$ as follows: for every $f \in \mathcal{L}(E; F)$,

$$\|f\| = \sup_{x \in E, \ x \neq 0} \frac{\|f(x)\|}{\|x\|}.$$

or equivalently by

$$\|f\| = \inf \{\lambda \in \mathbb{R} \mid \|f(x)\| \leq \lambda \|x\|, \text{ for all } x \in E\}.$$ 

It is not hard to show that the map $f \mapsto \|f\|$ is a norm on $\mathcal{L}(E; F)$ satisfying the property

$$\|f(x)\| \leq \|f\| \|x\|$$

for all $x \in E$, and that if $f \in \mathcal{L}(E; F)$ and $g \in \mathcal{L}(F; G)$, then

$$\|g \circ f\| \leq \|g\| \|f\|.$$

Operator norms play an important role in functional analysis, especially when the spaces $E$ and $F$ are complete.

The following proposition is needed when to deal with the condition number of a matrix.

**Proposition 16.9.** Let $\|\|$ be any matrix norm and let $B$ be a matrix such that $\|B\| < 1$.

1. If $\|\|$ is a subordinate matrix norm, then the matrix $I + B$ is invertible and

$$\|(I + B)^{-1}\| \leq \frac{1}{1 - \|B\|}.$$

2. If a matrix of the form $I + B$ is singular, then $\|B\| \geq 1$ for every matrix norm (not necessarily subordinate).

**Proof.** (1) Observe that $(I + B)u = 0$ implies $Bu = -u$, so

$$\|u\| = \|Bu\|.$$
Recall that
\[ \|Bu\| \leq \|B\|\|u\| \]
for every subordinate norm. Since \( \|B\| < 1 \), if \( u \neq 0 \), then
\[ \|Bu\| < \|u\| , \]
which contradicts \( \|u\| = \|Bu\| \). Therefore, we must have \( u = 0 \), which proves that \( I + B \) is injective, and thus bijective, i.e., invertible. Then, we have
\[ (I + B)^{-1} + B(I + B)^{-1} = (I + B)(I + B)^{-1} = I , \]
so we get
\[ (I + B)^{-1} = I - B(I + B)^{-1} , \]
which yields
\[ \|(I + B)^{-1}\| \leq 1 + \|B\|\|(I + B)^{-1}\| , \]
and finally,
\[ \|(I + B)^{-1}\| \leq \frac{1}{1 - \|B\|} . \]

(2) If \( I + B \) is singular, then \(-1\) is an eigenvalue of \( B \), and by Proposition 16.4, we get \( \rho(B) \leq \|B\| \), which implies \( 1 \leq \rho(B) \leq \|B\| \). \( \square \)

The following result is needed to deal with the convergence of sequences of powers of matrices.

**Proposition 16.10.** For every matrix \( A \in M_n(\mathbb{C}) \) and for every \( \epsilon > 0 \), there is some subordinate matrix norm \( \|\| \) such that
\[ \|A\| \leq \rho(A) + \epsilon . \]

**Proof.** By the Schur form, there exists some invertible matrix \( U \) and some upper triangular matrix \( T \) such that
\[ A = UTU^{-1} , \]
and say that
\[ T = \begin{pmatrix} \lambda_1 & t_{12} & t_{13} & \cdots & t_{1n} \\ 0 & \lambda_2 & t_{23} & \cdots & t_{2n} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_{n-1} & t_{n-1n} \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix} , \]
where \( \lambda_1, \ldots, \lambda_n \) are the eigenvalues of \( A \). For every \( \delta \neq 0 \), define the diagonal matrix
\[ D_{\delta} = \text{diag}(1, \delta, \delta^2, \ldots, \delta^{n-1}) , \]
and consider the matrix

\[(UD_\delta)^{-1}A(UD_\delta) = D_\delta^{-1}TD_\delta = \begin{pmatrix}
\lambda_1 & \delta t_{12} & \delta^2 t_{13} & \cdots & \delta^{n-1} t_{1n} \\
0 & \lambda_2 & \delta t_{23} & \cdots & \delta^{n-2} t_{2n} \\
\vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & \lambda_{n-1} & \delta t_{n-1n} \\
0 & 0 & \cdots & 0 & \lambda_n
\end{pmatrix}.
\]

Now, define the function \(\|\|: M_n(\mathbb{C}) \to \mathbb{R}\) by

\[\|B\| = \left\|(UD_\delta)^{-1}B(UD_\delta)\right\|_\infty,
\]

for every \(B \in M_n(\mathbb{C})\). Then it is easy to verify that the above function is the matrix norm subordinate to the vector norm

\[v \mapsto \left\|(UD_\delta)^{-1}v\right\|_\infty.
\]

Furthermore, for every \(\epsilon > 0\), we can pick \(\delta\) so that

\[\sum_{j=i+1}^n |\delta^{j-i} t_{ij}| \leq \epsilon, \quad 1 \leq i \leq n - 1,
\]

and by definition of the norm \(\|\|_\infty\), we get

\[\|A\| \leq \rho(A) + \epsilon,
\]

which shows that the norm that we have constructed satisfies the required properties.

Note that equality is generally not possible; consider the matrix

\[A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix},
\]

for which \(\rho(A) = 0 < \|A\|\), since \(A \neq 0\).
Chapter 17

Basics of Groups and Group Actions

This chapter gathers basics of the theory of groups and group actions.

17.1 Groups, Subgroups, Cosets

Definition 17.1. A group is a set $G$ equipped with a binary operation $\cdot : G \times G \rightarrow G$ that associates an element $a \cdot b \in G$ to every pair of elements $a, b \in G$, and having the following properties: $\cdot$ is associative, has an identity element $e \in G$, and every element in $G$ is invertible (w.r.t. $\cdot$). More explicitly, this means that the following equations hold for all $a, b, c \in G$:

$(G1) \quad a \cdot (b \cdot c) = (a \cdot b) \cdot c.$ \hspace{1cm} \text{(associativity)};

$(G2) \quad a \cdot e = e \cdot a = a.$ \hspace{1cm} \text{(identity)};

$(G3) \quad$ For every $a \in G$, there is some $a^{-1} \in G$ such that $a \cdot a^{-1} = a^{-1} \cdot a = e.$ \hspace{1cm} \text{(inverse)}.

A group $G$ is abelian (or commutative) if

$$a \cdot b = b \cdot a \quad \text{for all } a, b \in G.$$  

A set $M$ together with an operation $\cdot : M \times M \rightarrow M$ and an element $e$ satisfying only Conditions $(G1)$ and $(G2)$ is called a monoid. For example, the set $\mathbb{N} = \{0, 1, \ldots, n, \ldots\}$ of natural numbers is a (commutative) monoid under addition. However, it is not a group.

Some examples of groups are given below.

Example 17.1.

1. The set $\mathbb{Z} = \{\ldots, -n, \ldots, -1, 0, 1, \ldots, n, \ldots\}$ of integers is an abelian group under addition, with identity element 0. However, $\mathbb{Z}^* = \mathbb{Z} - \{0\}$ is not a group under multiplication.
2. The set $\mathbb{Q}$ of rational numbers (fractions $p/q$ with $p, q \in \mathbb{Z}$ and $q \neq 0$) is an abelian
group under addition, with identity element 0. The set $\mathbb{Q}^* = \mathbb{Q} - \{0\}$ is also an abelian

group under multiplication, with identity element 1.

3. Given any nonempty set $S$, the set of bijections $f : S \to S$, also called permutations

of $S$, is a group under function composition (i.e., the multiplication of $f$ and $g$ is the

composition $g \circ f$), with identity element the identity function $\text{id}_S$. This group is not

abelian as soon as $S$ has more than two elements. The permutation group of the set

$S = \{1, \ldots, n\}$ is often denoted $\mathfrak{S}_n$ and called the symmetric group on $n$ elements.

4. For any positive integer $p \in \mathbb{N}$, define a relation on $\mathbb{Z}$, denoted $m \equiv n \pmod{p}$, as

follows:

$$m \equiv n \pmod{p} \iff m - n = kp \text{ for some } k \in \mathbb{Z}.$$ 

The reader will easily check that this is an equivalence relation, and, moreover, it is
compatible with respect to addition and multiplication, which means that if $m_1 \equiv n_1
\pmod{p}$ and $m_2 \equiv n_2 \pmod{p}$, then $m_1 + m_2 \equiv n_1 + n_2 \pmod{p}$ and $m_1 m_2 \equiv n_1 n_2
\pmod{p}$. Consequently, we can define an addition operation and a multiplication
operation of the set of equivalence classes $\pmod{p}$:

$$[m] + [n] = [m + n]$$

and

$$[m] \cdot [n] = [mn].$$

The reader will easily check that addition of residue classes $\pmod{p}$ induces an abelian
group structure with $[0]$ as zero. This group is denoted $\mathbb{Z}/p\mathbb{Z}$.

5. The set of $n \times n$ invertible matrices with real (or complex) coefficients is a group under

matrix multiplication, with identity element the identity matrix $I_n$. This group is
called the general linear group and is usually denoted by $\text{GL}(n, \mathbb{R})$ (or $\text{GL}(n, \mathbb{C})$).

6. The set of $n \times n$ invertible matrices $A$ with real (or complex) coefficients such that

$\det(A) = 1$ is a group under matrix multiplication, with identity element the identity

matrix $I_n$. This group is called the special linear group and is usually denoted by

$\text{SL}(n, \mathbb{R})$ (or $\text{SL}(n, \mathbb{C})$).

7. The set of $n \times n$ matrices $Q$ with real coefficients such that

$$QQ^\top = Q^\top Q = I_n$$

is a group under matrix multiplication, with identity element the identity matrix $I_n$;
we have $Q^{-1} = Q^\top$. This group is called the orthogonal group and is usually denoted by $\text{O}(n)$. 
8. The set of $n \times n$ invertible matrices $Q$ with real coefficients such that

$$QQ^\top = Q^\top Q = I_n \quad \text{and} \quad \det(Q) = 1$$

is a group under matrix multiplication, with identity element the identity matrix $I_n$; as in (6), we have $Q^{-1} = Q^\top$. This group is called the special orthogonal group or rotation group and is usually denoted by $\text{SO}(n)$.

The groups in (5)--(8) are nonabelian for $n \geq 2$, except for $\text{SO}(2)$ which is abelian (but $\text{O}(2)$ is not abelian).

It is customary to denote the operation of an abelian group $G$ by $+$, in which case the inverse $a^{-1}$ of an element $a \in G$ is denoted by $-a$.

The identity element of a group is unique. In fact, we can prove a more general fact:

**Proposition 17.1.** If a binary operation $\cdot : M \times M \to M$ is associative and if $e' \in M$ is a left identity and $e'' \in M$ is a right identity, which means that

$$e' \cdot a = a \quad \text{for all} \quad a \in M \quad \text{(G2l)}$$

and

$$a \cdot e'' = a \quad \text{for all} \quad a \in M, \quad \text{(G2r)}$$

then $e' = e''$.

**Proof.** If we let $a = e''$ in equation (G2l), we get

$$e' \cdot e'' = e'',$$

and if we let $a = e'$ in equation (G2r), we get

$$e' \cdot e'' = e',$$

and thus

$$e' = e' \cdot e'' = e'',$$

as claimed. \qed

Proposition 17.1 implies that the identity element of a monoid is unique, and since every group is a monoid, the identity element of a group is unique. Furthermore, every element in a group has a unique inverse. This is a consequence of a slightly more general fact:

**Proposition 17.2.** In a monoid $M$ with identity element $e$, if some element $a \in M$ has some left inverse $a' \in M$ and some right inverse $a'' \in M$, which means that

$$a' \cdot a = e \quad \text{(G3l)}$$

and

$$a \cdot a'' = e, \quad \text{(G3r)}$$

then $a' = a''$. 

Proof. Using (G3l) and the fact that \(e\) is an identity element, we have
\[
(a' \cdot a) \cdot a'' = e \cdot a'' = a''.
\]
Similarly, Using (G3r) and the fact that \(e\) is an identity element, we have
\[
a' \cdot (a \cdot a'') = a' \cdot e = a'.
\]
However, since \(M\) is monoid, the operation \(\cdot\) is associative, so
\[
a' = a' \cdot (a \cdot a'') = (a' \cdot a) \cdot a'' = a'',
\]
as claimed.

Remark: Axioms (G2) and (G3) can be weakened a bit by requiring only (G2r) (the existence of a right identity) and (G3r) (the existence of a right inverse for every element) (or (G2l) and (G3l)). It is a good exercise to prove that the group axioms (G2) and (G3) follow from (G2r) and (G3r).

Definition 17.2. If a group \(G\) has a finite number \(n\) of elements, we say that \(G\) is a group of order \(n\). If \(G\) is infinite, we say that \(G\) has infinite order. The order of a group is usually denoted by \(|G|\) (if \(G\) is finite).

Given a group \(G\), for any two subsets \(R, S \subseteq G\), we let
\[
RS = \{r \cdot s \mid r \in R, s \in S\}.
\]
In particular, for any \(g \in G\), if \(R = \{g\}\), we write
\[
gS = \{g \cdot s \mid s \in S\},
\]
and similarly, if \(S = \{g\}\), we write
\[
Rg = \{r \cdot g \mid r \in R\}.
\]

From now on, we will drop the multiplication sign and write \(g_1g_2\) for \(g_1 \cdot g_2\).

Definition 17.3. Let \(G\) be a group. For any \(g \in G\), define \(L_g\), the left translation by \(g\), by \(L_g(a) = ga\), for all \(a \in G\), and \(R_g\), the right translation by \(g\), by \(R_g(a) = ag\), for all \(a \in G\).

The following simple fact is often used.

Proposition 17.3. Given a group \(G\), the translations \(L_g\) and \(R_g\) are bijections.

Proof. We show this for \(L_g\), the proof for \(R_g\) being similar.

If \(L_g(a) = L_g(b)\), then \(ga = gb\), and multiplying on the left by \(g^{-1}\), we get \(a = b\), so \(L_g\) injective. For any \(b \in G\), we have \(L_g(g^{-1}b) = gg^{-1}b = b\), so \(L_g\) is surjective. Therefore, \(L_g\) is bijective. \(\square\)
Definition 17.4. Given a group $G$, a subset $H$ of $G$ is a subgroup of $G$ iff

1. The identity element $e$ of $G$ also belongs to $H$ ($e \in H$);
2. For all $h_1, h_2 \in H$, we have $h_1 h_2 \in H$;
3. For all $h \in H$, we have $h^{-1} \in H$.

The proof of the following proposition is left as an exercise.

Proposition 17.4. Given a group $G$, a subset $H \subseteq G$ is a subgroup of $G$ iff $H$ is nonempty and whenever $h_1, h_2 \in H$, then $h_1 h_2^{-1} \in H$.

If the group $G$ is finite, then the following criterion can be used.

Proposition 17.5. Given a finite group $G$, a subset $H \subseteq G$ is a subgroup of $G$ iff

1. $e \in H$;
2. $H$ is closed under multiplication.

Proof. We just have to prove that Condition (3) of Definition 17.4 holds. For any $a \in H$, since the left translation $L_a$ is bijective, its restriction to $H$ is injective, and since $H$ is finite, it is also bijective. Since $e \in H$, there is a unique $b \in H$ such that $L_a(b) = ab = e$. However, if $a^{-1}$ is the inverse of $a$ in $G$, we also have $L_a(a^{-1}) = aa^{-1} = e$, and by injectivity of $L_a$, we have $a^{-1} = b \in H$. \qed

Example 17.2.

1. For any integer $n \in \mathbb{Z}$, the set
   $$n\mathbb{Z} = \{nk \mid k \in \mathbb{Z}\}$$
   is a subgroup of the group $\mathbb{Z}$.
2. The set of matrices
   $$\text{GL}^+(n, \mathbb{R}) = \{A \in \text{GL}(n, \mathbb{R}) \mid \det(A) > 0\}$$
   is a subgroup of the group $\text{GL}(n, \mathbb{R})$.
3. The group $\text{SL}(n, \mathbb{R})$ is a subgroup of the group $\text{GL}(n, \mathbb{R})$.
4. The group $\text{O}(n)$ is a subgroup of the group $\text{GL}(n, \mathbb{R})$.
5. The group $\text{SO}(n)$ is a subgroup of the group $\text{O}(n)$, and a subgroup of the group $\text{SL}(n, \mathbb{R})$. 
6. It is not hard to show that every $2 \times 2$ rotation matrix $R \in \text{SO}(2)$ can be written as

$$R = \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix}, \quad \text{with} \quad 0 \leq \theta < 2\pi.$$  

Then $\text{SO}(2)$ can be considered as a subgroup of $\text{SO}(3)$ by viewing the matrix

$$R = \begin{pmatrix} \cos \theta & -\sin \theta & 0 \\ \sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$  

7. The set of $2 \times 2$ upper-triangular matrices of the form

$$\begin{pmatrix} a & b \\ 0 & c \end{pmatrix} \quad a, b, c \in \mathbb{R}, \quad a, c \neq 0$$

is a subgroup of the group $\text{GL}(2, \mathbb{R})$.

8. The set $V$ consisting of the four matrices

$$\begin{pmatrix} \pm 1 & 0 \\ 0 & \pm 1 \end{pmatrix}$$

is a subgroup of the group $\text{GL}(2, \mathbb{R})$ called the *Klein four-group*.

**Definition 17.5.** If $H$ is a subgroup of $G$ and $g \in G$ is any element, the sets of the form $gH$ are called *left cosets of $H$ in $G$* and the sets of the form $Hg$ are called *right cosets of $H$ in $G$*. The left cosets (resp. right cosets) of $H$ induce an equivalence relation $\sim$ defined as follows: For all $g_1, g_2 \in G$,

$$g_1 \sim g_2 \quad \text{iff} \quad g_1H = g_2H$$

(resp. $g_1 \sim g_2$ iff $Hg_1 = Hg_2$). Obviously, $\sim$ is an equivalence relation.

Now, we claim the following fact:

**Proposition 17.6.** Given a group $G$ and any subgroup $H$ of $G$, we have $g_1H = g_2H$ iff $g_2^{-1}g_1H = H$ iff $g_2^{-1}g_1 \in H$, for all $g_1, g_2 \in G$.

*Proof.* If we apply the bijection $L_{g_2^{-1}}$ to both $g_1H$ and $g_2H$ we get $L_{g_2^{-1}}(g_1H) = g_2^{-1}g_1H$ and $L_{g_2^{-1}}(g_2H) = H$, so $g_1H = g_2H$ iff $g_2^{-1}g_1H = H$. If $g_2^{-1}g_1H = H$, since $1 \in H$, we get $g_2^{-1}g_1 \in H$. Conversely, if $g_2^{-1}g_1 \in H$, since $H$ is a group, the left translation $L_{g_2^{-1}g_1}$ is a bijection of $H$, so $g_2^{-1}g_1H = H$. Thus, $g_2^{-1}g_1H = H$ iff $g_2^{-1}g_1 \in H$. \qed
It follows that the equivalence class of an element \( g \in G \) is the coset \( gH \) (resp. \( Hg \)). Since \( L_g \) is a bijection between \( H \) and \( gH \), the cosets \( gH \) all have the same cardinality. The map \( L_{g^{-1}} \circ R_g \) is a bijection between the left coset \( gH \) and the right coset \( Hg \), so they also have the same cardinality. Since the distinct cosets \( gH \) form a partition of \( G \), we obtain the following fact:

**Proposition 17.7.** (Lagrange) For any finite group \( G \) and any subgroup \( H \) of \( G \), the order \( h \) of \( H \) divides the order \( n \) of \( G \).

**Definition 17.6.** Given a finite group \( G \) and a subgroup \( H \) of \( G \), if \( n = |G| \) and \( h = |H| \), then the ratio \( n/h \) is denoted by \((G : H)\) and is called the *index* of \( H \) in \( G \).

The index \((G : H)\) is the number of left (and right) cosets of \( H \) in \( G \). Proposition 17.7 can be stated as

\[ |G| = (G : H)|H| . \]

The set of left cosets of \( H \) in \( G \) (which, in general, is not a group) is denoted \( G/H \). The “points” of \( G/H \) are obtained by “collapsing” all the elements in a coset into a single element.

**Example 17.3.**

1. Let \( n \) be any positive integer, and consider the subgroup \( n\mathbb{Z} \) of \( \mathbb{Z} \) (under addition).
   The coset of 0 is the set \( \{0\} \), and the coset of any nonzero integer \( m \in \mathbb{Z} \) is
   \[ m + n\mathbb{Z} = \{m + nk \mid k \in \mathbb{Z}\} . \]
   By dividing \( m \) by \( n \), we have \( m = nq + r \) for some unique \( r \) such that \( 0 \leq r \leq n-1 \). But then we see that \( r \) is the smallest positive element of the coset \( m + n\mathbb{Z} \). This implies that there is a bijection between the cosets of the subgroup \( n\mathbb{Z} \) of \( \mathbb{Z} \) and the set of residues \( \{0, 1, \ldots, n-1\} \) modulo \( n \), or equivalently a bijection with \( \mathbb{Z}/n\mathbb{Z} \).

2. The cosets of \( \text{SL}(n, \mathbb{R}) \) in \( \text{GL}(n, \mathbb{R}) \) are the sets of matrices
   \[ A\text{SL}(n, \mathbb{R}) = \{AB \mid B \in \text{SL}(n, \mathbb{R})\}, \quad A \in \text{GL}(n, \mathbb{R}) . \]
   Since \( A \) is invertible, \( \det(A) \neq 0 \), and we can write \( A = \left(\det(A)\right)^{1/n}((\det(A))^{-1/n}A) \) if \( \det(A) > 0 \) and \( A = \left(-\det(A)\right)^{1/n}((-\det(A))^{-1/n}A) \) if \( \det(A) < 0 \). But we have \((\det(A))^{-1/n}A \in \text{SL}(n, \mathbb{R})\) if \( \det(A) > 0 \) and \(-(-\det(A))^{-1/n}A \in \text{SL}(n, \mathbb{R})\) if \( \det(A) < 0 \), so the coset \( A\text{SL}(n, \mathbb{R}) \) contains the matrix
   \[ (\det(A))^{1/n}I_n \quad \text{if} \quad \det(A) > 0, \quad -(-\det(A))^{1/n}I_n \quad \text{if} \quad \det(A) < 0 . \]
   It follows that there is a bijection between the cosets of \( \text{SL}(n, \mathbb{R}) \) in \( \text{GL}(n, \mathbb{R}) \) and \( \mathbb{R} \).
3. The cosets of \( \text{SO}(n) \) in \( \text{GL}^+(n, \mathbb{R}) \) are the sets of matrices

\[
A \text{SO}(n) = \{ AQ \mid Q \in \text{SO}(n) \}, \quad A \in \text{GL}^+(n, \mathbb{R}).
\]

It can be shown (using the polar form for matrices) that there is a bijection between the cosets of \( \text{SO}(n) \) in \( \text{GL}^+(n, \mathbb{R}) \) and the set of \( n \times n \) symmetric, positive, definite matrices; these are the symmetric matrices whose eigenvalues are strictly positive.

4. The cosets of \( \text{SO}(2) \) in \( \text{SO}(3) \) are the sets of matrices

\[
Q \text{SO}(2) = \{ QR \mid R \in \text{SO}(2) \}, \quad Q \in \text{SO}(3).
\]

The group \( \text{SO}(3) \) moves the points on the sphere \( S^2 \) in \( \mathbb{R}^3 \), namely for any \( x \in S^2 \),

\[
x \mapsto Qx \quad \text{for any rotation } Q \in \text{SO}(3).
\]

Here,

\[
S^2 = \{(x, y, z) \in \mathbb{R}^3 \mid x^2 + y^2 + z^2 = 1\}.
\]

Let \( N = (0, 0, 1) \) be the north pole on the sphere \( S^2 \). Then it is not hard to show that \( \text{SO}(2) \) is precisely the subgroup of \( \text{SO}(3) \) that leaves \( N \) fixed. As a consequence, all rotations \( QR \) in the coset \( Q \text{SO}(2) \) map \( N \) to the same point \( QN \in S^2 \), and it can be shown that there is a bijection between the cosets of \( \text{SO}(2) \) in \( \text{SO}(3) \) and the points on \( S^2 \). The surjectivity of this map has to do with the fact that the action of \( \text{SO}(3) \) on \( S^2 \) is transitive, which means that for any point \( x \in S^2 \), there is some rotation \( Q \in \text{SO}(3) \) such that \( QN = x \).

It is tempting to define a multiplication operation on left cosets (or right cosets) by setting

\[
(g_1 H)(g_2 H) = (g_1 g_2)H,
\]

but this operation is not well defined in general, unless the subgroup \( H \) possesses a special property. In Example 17.3, it is possible to define multiplication of cosets in (1), but it is not possible in (2) and (3).

The property of the subgroup \( H \) that allows defining a multiplication operation on left cosets is typical of the kernels of group homomorphisms, so we are led to the following definition.

**Definition 17.7.** Given any two groups \( G \) and \( G' \), a function \( \varphi: G \to G' \) is a homomorphism iff

\[
\varphi(g_1 g_2) = \varphi(g_1)\varphi(g_2), \quad \text{for all } g_1, g_2 \in G.
\]

Taking \( g_1 = g_2 = e \) (in \( G \)), we see that

\[
\varphi(e) = e',
\]

and taking \( g_1 = g \) and \( g_2 = g^{-1} \), we see that

\[
\varphi(g^{-1}) = (\varphi(g))^{-1}.
\]
Example 17.4.

1. The map \( \varphi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \) given by \( \varphi(m) = m \mod n \) for all \( m \in \mathbb{Z} \) is a homomorphism.

2. The map \( \text{det} : \text{GL}(n, \mathbb{R}) \to \mathbb{R} \) is a homomorphism because \( \text{det}(AB) = \text{det}(A)\text{det}(B) \) for any two matrices \( A, B \). Similarly, the map \( \text{det} : \text{O}(n) \to \mathbb{R} \) is a homomorphism.

If \( \varphi : G \to G' \) and \( \psi : G' \to G'' \) are group homomorphisms, then \( \psi \circ \varphi : G \to G'' \) is also a homomorphism. If \( \varphi : G \to G' \) is a homomorphism of groups, and if \( H \subseteq G, H' \subseteq G' \) are two subgroups, then it is easily checked that

\[
\text{Im} \, H = \varphi(H) = \{ \varphi(g) \mid g \in H \}
\]

is a subgroup of \( G' \) and

\[
\varphi^{-1}(H') = \{ g \in G \mid \varphi(g) \in H' \}
\]

is a subgroup of \( G \). In particular, when \( H' = \{ e' \} \), we obtain the kernel, \( \text{Ker} \varphi \), of \( \varphi \).

Definition 17.8. If \( \varphi : G \to G' \) is a homomorphism of groups, and if \( H \subseteq G \) is a subgroup of \( G \), then the subgroup of \( G' \),

\[
\text{Im} \, H = \varphi(H) = \{ \varphi(g) \mid g \in H \},
\]

is called the image of \( H \) by \( \varphi \), and the subgroup of \( G \),

\[
\text{Ker} \varphi = \{ g \in G \mid \varphi(g) = e' \},
\]

is called the kernel of \( \varphi \).

Example 17.5.

1. The kernel of the homomorphism \( \varphi : \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z} \) is \( n\mathbb{Z} \).

2. The kernel of the homomorphism \( \text{det} : \text{GL}(n, \mathbb{R}) \to \mathbb{R} \) is \( \text{SL}(n, \mathbb{R}) \). Similarly, the kernel of the homomorphism \( \text{det} : \text{O}(n) \to \mathbb{R} \) is \( \text{SO}(n) \).

The following characterization of the injectivity of a group homomorphism is used all the time.

Proposition 17.8. If \( \varphi : G \to G' \) is a homomorphism of groups, then \( \varphi : G \to G' \) is injective iff \( \text{Ker} \varphi = \{ e \} \). (We also write \( \text{Ker} \varphi = (0) \).)

Proof. Assume \( \varphi \) is injective. Since \( \varphi(e) = e' \), if \( \varphi(g) = e' \), then \( \varphi(g) = \varphi(e) \), and by injectivity of \( \varphi \) we must have \( g = e \), so \( \text{Ker} \varphi = \{ e \} \).

Conversely, assume that \( \text{Ker} \varphi = \{ e \} \). If \( \varphi(g_1) = \varphi(g_2) \), then by multiplication on the left by \( (\varphi(g_1))^{-1} \) we get

\[
e' = (\varphi(g_1))^{-1}\varphi(g_1) = (\varphi(g_1))^{-1}\varphi(g_2),
\]
and since \( \varphi \) is a homomorphism \((\varphi(g_1))^{-1} = \varphi(g_1^{-1})\), so
\[
e' = (\varphi(g_1))^{-1} \varphi(g_2) = \varphi(g_1^{-1}) \varphi(g_2) = \varphi(g_1^{-1}g_2).
\]
This shows that \( g_1^{-1}g_2 \in \text{Ker} \varphi \), but since \( \text{Ker} \varphi = \{e\} \) we have \( g_1^{-1}g_2 = e \), and thus \( g_2 = g_1 \), proving that \( \varphi \) is injective.

**Definition 17.9.** We say that a group homomorphism \( \varphi: G \to G' \) is an **isomorphism** if there is a homomorphism \( \psi: G' \to G \), so that
\[
\psi \circ \varphi = \text{id}_G \quad \text{and} \quad \varphi \circ \psi = \text{id}_{G'}.
\]

If \( \varphi \) is an isomorphism we say that the groups \( G \) and \( G' \) are **isomorphic**. When \( G' = G \), a group isomorphism is called an **automorphism**.

The reasoning used in the proof of Proposition 17.2 shows that if a a group homomorphism \( \varphi: G \to G' \) is an isomorphism, then the homomorphism \( \psi: G' \to G \) satisfying Condition (†) is unique. This homomorphism is denoted \( \varphi^{-1} \).

The left translations \( L_g \) and the right translations \( R_g \) are automorphisms of \( G \).

Suppose \( \varphi: G \to G' \) is a bijective homomorphism, and let \( \varphi^{-1} \) be the inverse of \( \varphi \) (as a function). Then for all \( a, b \in G \), we have
\[
\varphi(\varphi^{-1}(a)\varphi^{-1}(b)) = \varphi(\varphi^{-1}(a))\varphi(\varphi^{-1}(b)) = ab,
\]
and so
\[
\varphi^{-1}(ab) = \varphi^{-1}(a)\varphi^{-1}(b),
\]
which proves that \( \varphi^{-1} \) is a homomorphism. Therefore, we proved the following fact.

**Proposition 17.9.** A bijective group homomorphism \( \varphi: G \to G' \) is an isomorphism.

Observe that the property
\[
gH = Hg, \quad \text{for all } g \in G.
\]
(*)
is equivalent by multiplication on the right by \( g^{-1} \) to
\[
gHg^{-1} = H, \quad \text{for all } g \in G,
\]
and the above is equivalent to
\[
gHg^{-1} \subseteq H, \quad \text{for all } g \in G.
\]
(**)
This is because \( gHg^{-1} \subseteq H \) implies \( H \subseteq g^{-1}Hg \), and this for all \( g \in G \).

**Proposition 17.10.** Let \( \varphi: G \to G' \) be a group homomorphism. Then \( H = \text{Ker} \varphi \) satisfies Property (**) , and thus Property (*).
17.1. GROUPS, SUBGROUPS, COSETS

Proof. We have

\[ \varphi(ghg^{-1}) = \varphi(g)\varphi(h)\varphi(g^{-1}) = \varphi(g)e'\varphi(g)^{-1} = \varphi(g)\varphi(g)^{-1} = e', \]

for all \( h \in H = \text{Ker } \varphi \) and all \( g \in G \). Thus, by definition of \( H = \text{Ker } \varphi \), we have \( gHg^{-1} \subseteq H \). \( \Box \)

Definition 17.10. For any group \( G \), a subgroup \( N \) of \( G \) is a normal subgroup of \( G \) iff

\[ gNg^{-1} = N, \quad \text{for all } g \in G. \]

This is denoted by \( N \triangleleft G \).

Proposition 17.10 shows that the kernel \( \text{Ker } \varphi \) of a homomorphism \( \varphi: G \to G' \) is a normal subgroup of \( G \).

Observe that if \( G \) is abelian, then every subgroup of \( G \) is normal.

Consider Example 17.2. Let \( R \in \text{SO}(2) \) and \( A \in \text{SL}(2, \mathbb{R}) \) be the matrices

\[ R = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}. \]

Then

\[ A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \]

and we have

\[ ARA^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 1 & -1 \end{pmatrix}, \]

and clearly \( ARA^{-1} \notin \text{SO}(2) \). Therefore \( \text{SO}(2) \) is not a normal subgroup of \( \text{SL}(2, \mathbb{R}) \). The same counter-example shows that \( \text{O}(2) \) is not a normal subgroup of \( \text{GL}(2, \mathbb{R}) \).

Let \( R \in \text{SO}(2) \) and \( Q \in \text{SO}(3) \) be the matrices

\[ R = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad Q = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}. \]

Then

\[ Q^{-1} = Q^\top = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \]
and we have

\[
QRQ^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix} \\
= \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}.
\]

Observe that \(QRQ^{-1} \notin \text{SO}(2)\), so \(\text{SO}(2)\) is not a normal subgroup of \(\text{SO}(3)\).

Let \(T\) and \(A \in \text{GL}(2, \mathbb{R})\) be the following matrices

\[
T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.
\]

We have

\[
A^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = A,
\]

and

\[
ATA^{-1} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.
\]

The matrix \(T\) is upper triangular, but \(ATA^{-1}\) is not, so the group of \(2 \times 2\) upper triangular matrices is not a normal subgroup of \(\text{GL}(2, \mathbb{R})\).

Let \(Q \in V\) and \(A \in \text{GL}(2, \mathbb{R})\) be the following matrices

\[
Q = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.
\]

We have

\[
A^{-1} = \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}
\]

and

\[
AQA^{-1} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -1 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & -2 \\ 0 & -1 \end{pmatrix}.
\]

Clearly \(AQA^{-1} \notin V\), which shows that the Klein four group is not a normal subgroup of \(\text{GL}(2, \mathbb{R})\).

The reader should check that the subgroups \(n\mathbb{Z}\), \(\text{GL}^+(n, \mathbb{R})\), \(\text{SL}(n, \mathbb{R})\), and \(\text{SO}(n, \mathbb{R})\) as a subgroup of \(\text{O}(n, \mathbb{R})\), are normal subgroups.

If \(N\) is a normal subgroup of \(G\), the equivalence relation \(\sim\) induced by left cosets (see Definition 17.5) is the same as the equivalence induced by right cosets. Furthermore, this equivalence relation is a congruence, which means that: For all \(g_1, g_2, g'_1, g'_2 \in G\),
(1) If \(g_1 \bar{N} = g_1' \bar{N}\) and \(g_2 \bar{N} = g_2' \bar{N}\), then \(g_1 g_2 \bar{N} = g_1' g_2' \bar{N}\), and

(2) If \(g_1 \bar{N} = g_2 \bar{N}\), then \(g_1^{-1} \bar{N} = g_2^{-1} \bar{N}\).

As a consequence, we can define a group structure on the set \(G/\sim\) of equivalence classes modulo \(\sim\), by setting

\[(g_1 \bar{N})(g_2 \bar{N}) = (g_1 g_2) \bar{N}.\]

**Definition 17.11.** Let \(G\) be a group and \(N\) be a normal subgroup of \(G\). The group obtained by defining the multiplication of (left) cosets by

\[(g_1 \bar{N})(g_2 \bar{N}) = (g_1 g_2) \bar{N}, \quad g_1, g_2 \in G\]

is denoted \(G/N\), and called the **quotient of** \(G\) **by** \(N\). The equivalence class \(g \bar{N}\) of an element \(g \in G\) is also denoted \(g\) (or \([g]\)). The map \(\pi: G \to G/N\) given by

\[\pi(g) = \bar{g} = g \bar{N}\]

is a group homomorphism called the **canonical projection**.

Since the kernel of a homomorphism is a normal subgroup, we obtain the following very useful result.

**Proposition 17.11.** Given a homomorphism of groups \(\varphi: G \to G'\), the groups \(G/\text{Ker} \varphi\) and \(\text{Im} \varphi = \varphi(G)\) are isomorphic.

**Proof.** Since \(\varphi\) is surjective onto its image, we may assume that \(\varphi\) is surjective, so that \(G' = \text{Im} \varphi\). We define a map \(\overline{\varphi}: G/\text{Ker} \varphi \to G'\) as follows:

\[\overline{\varphi}(\bar{g}) = \varphi(g), \quad g \in G.\]

We need to check that the definition of this map does not depend on the representative chosen in the coset \(\bar{g} = g \text{Ker} \varphi\), and that it is a homomorphism. If \(g'\) is another element in the coset \(g\text{ Ker} \varphi\), which means that \(g' = gh\) for some \(h \in \text{Ker} \varphi\), then

\[\varphi(g') = \varphi(gh) = \varphi(g) \varphi(h) = \varphi(g)e' = \varphi(g),\]

since \(\varphi(h) = e'\) as \(h \in \text{Ker} \varphi\). This shows that

\[\overline{\varphi}(\bar{g'}) = \overline{\varphi}(\varphi(g')) = \varphi(g) = \overline{\varphi}(\bar{g}),\]

so the map \(\overline{\varphi}\) is well defined. It is a homomorphism because

\[\overline{\varphi}(\bar{gg'}) = \overline{\varphi}((gg'))\]

\[= \varphi(gg')\]

\[= \varphi(g)\varphi(g')\]

\[= \overline{\varphi}(\bar{g})\overline{\varphi}(\bar{g'}).\]

The map \(\overline{\varphi}\) is injective because \(\overline{\varphi}(\bar{g}) = e'\) iff \(\varphi(g) = e'\) iff \(g \in \text{Ker} \varphi\), iff \(\bar{g} = \bar{e}\). The map \(\overline{\varphi}\) is surjective because \(\varphi\) is surjective. Therefore \(\overline{\varphi}\) is a bijective homomorphism, and thus an isomorphism, as claimed. \(\square\)
Proposition 17.11 is called the \textit{first isomorphism theorem}.

A useful way to construct groups is the \textit{direct product} construction.

\textbf{Definition 17.12.} Given two groups $G$ an $H$, we let $G \times H$ be the Cartesian product of the sets $G$ and $H$ with the multiplication operation \( \cdot \) given by
\[
(g_1, h_1) \cdot (g_2, h_2) = (g_1 g_2, h_1 h_2).
\]
It is immediately verified that $G \times H$ is a group called the \textit{direct product} of $G$ and $H$.

Similarly, given any $n$ groups $G_1, \ldots, G_n$, we can define the direct product $G_1 \times \cdots \times G_n$ in a similar way.

If $G$ is an abelian group and $H_1, \ldots, H_n$ are subgroups of $G$, the situation is simpler. Consider the map
\[
a : H_1 \times \cdots \times H_n \to G
\]
given by
\[
a(h_1, \ldots, h_n) = h_1 + \cdots + h_n,
\]
using $+$ for the operation of the group $G$. It is easy to verify that $a$ is a group homomorphism, so its image is a subgroup of $G$ denoted by $H_1 + \cdots + H_n$, and called the \textit{sum} of the groups $H_i$. The following proposition will be needed.

\textbf{Proposition 17.12.} Given an abelian group $G$, if $H_1$ and $H_2$ are any subgroups of $G$ such that $H_1 \cap H_2 = \{0\}$, then the map $a$ is an isomorphism
\[
a : H_1 \times H_2 \to H_1 + H_2.
\]

\textit{Proof.} The map is surjective by definition, so we just have to check that it is injective. For this, we show that $\ker a = \{(0,0)\}$. We have $a(a_1, a_2) = 0$ iff $a_1 + a_2 = 0$ iff $a_1 = -a_2$. Since $a_1 \in H_1$ and $a_2 \in H_2$, we see that $a_1, a_2 \in H_1 \cap H_2 = \{0\}$, so $a_1 = a_2 = 0$, which proves that $\ker a = \{(0,0)\}$.

Under the conditions of Proposition 17.12, namely $H_1 \cap H_2 = \{0\}$, the group $H_1 + H_2$ is called the \textit{direct sum} of $H_1$ and $H_2$; it is denoted by $H_1 \oplus H_2$, and we have an isomorphism $H_1 \times H_2 \cong H_1 \oplus H_2$.

\section{17.2 Group Actions: Part I, Definition and Examples}

If $X$ is a set (usually some kind of geometric space, for example, the sphere in $\mathbb{R}^3$, the upper half-plane, etc.), the “symmetries” of $X$ are often captured by the action of a group $G$ on $X$. In fact, if $G$ is a Lie group and the action satisfies some simple properties, the set $X$ can be given a manifold structure which makes it a projection (quotient) of $G$, a so-called “homogeneous space.”
17.2. GROUP ACTIONS: PART I, DEFINITION AND EXAMPLES

Definition 17.13. Given a set $X$ and a group $G$, a left action of $G$ on $X$ (for short, an action of $G$ on $X$) is a function $\varphi: G \times X \rightarrow X$, such that:

1. For all $g, h \in G$ and all $x \in X$,
   $$\varphi(g, \varphi(h, x)) = \varphi(gh, x),$$

2. For all $x \in X$,
   $$\varphi(1, x) = x,$$
   where $1 \in G$ is the identity element of $G$.

To alleviate the notation, we usually write $g \cdot x$ or even $gx$ for $\varphi(g, x)$, in which case the above axioms read:

1. For all $g, h \in G$ and all $x \in X$,
   $$g \cdot (h \cdot x) = gh \cdot x,$$

2. For all $x \in X$,
   $$1 \cdot x = x.$$

The set $X$ is called a (left) $G$-set. The action $\varphi$ is faithful or effective iff for every $g$, if $g \cdot x = x$ for all $x \in X$, then $g = 1$. Faithful means that if the action of some element $g$ behaves like the identity, then $g$ must be the identity element. The action $\varphi$ is transitive iff for any two elements $x, y \in X$, there is some $g \in G$ so that $g \cdot x = y$.

Given an action $\varphi: G \times X \rightarrow X$, for every $g \in G$, we have a function $\varphi_g: X \rightarrow X$ defined by

$$\varphi_g(x) = g \cdot x,$$

for all $x \in X$.

Observe that $\varphi_g$ has $\varphi_{g^{-1}}$ as inverse, since

$$\varphi_{g^{-1}}(\varphi_g(x)) = \varphi_{g^{-1}}(g \cdot x) = g^{-1} \cdot (g \cdot x) = (g^{-1}g) \cdot x = 1 \cdot x = x,$$

and similarly, $\varphi_g \circ \varphi_{g^{-1}} = \text{id}$. Therefore, $\varphi_g$ is a bijection of $X$; that is, $\varphi_g$ is a permutation of $X$. Moreover, we check immediately that

$$\varphi_g \circ \varphi_h = \varphi_{gh},$$

so the map $g \mapsto \varphi_g$ is a group homomorphism from $G$ to $\mathfrak{S}_X$, the group of permutations of $X$. With a slight abuse of notation, this group homomorphism $G \rightarrow \mathfrak{S}_X$ is also denoted $\varphi$.

Conversely, it is easy to see that any group homomorphism $\varphi: G \rightarrow \mathfrak{S}_X$ yields a group action $\cdot: G \times X \rightarrow X$, by setting

$$g \cdot x = \varphi(g)(x).$$

Observe that an action $\varphi$ is faithful iff the group homomorphism $\varphi: G \rightarrow \mathfrak{S}_X$ is injective, i.e. iff $\varphi$ has a trivial kernel. Also, we have $g \cdot x = y$ iff $g^{-1} \cdot y = x$, since $(gh) \cdot x = g \cdot (h \cdot x)$ and $1 \cdot x = x$, for all $g, h \in G$ and all $x \in X$. 
Definition 17.14. Given two $G$-sets $X$ and $Y$, a function $f : X \to Y$ is said to be equivariant, or a $G$-map, iff for all $x \in X$ and all $g \in G$, we have

$$f(g \cdot x) = g \cdot f(x).$$

Equivalently, if the $G$-actions are denoted by $\varphi : G \times X \to X$ and $\psi : G \times Y \to Y$, we have the following commutative diagram for all $g \in G$:

$$
\begin{array}{ccc}
X & \xrightarrow{\varphi_g} & X \\
| & f & | \\
Y & \xrightarrow{\psi_g} & Y.
\end{array}
$$

Remark: We can also define a right action $\cdot : X \times G \to X$ of a group $G$ on a set $X$ as a map satisfying the conditions

1. For all $g, h \in G$ and all $x \in X$,

$$(x \cdot g) \cdot h = x \cdot gh,$$

2. For all $x \in X$,

$$x \cdot 1 = x.$$

Every notion defined for left actions is also defined for right actions in the obvious way.

However, one change is necessary. For every $g \in G$, the map $\varphi_g : X \to X$ must be defined as

$$\varphi_g(x) = x \cdot g^{-1},$$

in order for the map $g \mapsto \varphi_g$ from $G$ to $\mathfrak{S}_X$ to be a homomorphism ($\varphi_g \circ \varphi_h = \varphi_{gh}$). Conversely, given a homomorphism $\varphi : G \to \mathfrak{S}_X$, we get a right action $\cdot : X \times G \to X$ by setting

$$x \cdot g = \varphi(g^{-1})(x).$$

Here are some examples of (left) group actions.

Example 17.6. The unit sphere $S^2$ (more generally, $S^{n-1}$).

Recall that for any $n \geq 1$, the (real) unit sphere $S^{n-1}$ is the set of points in $\mathbb{R}^n$ given by

$$S^{n-1} = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_1^2 + \cdots + x_n^2 = 1\}.$$

In particular, $S^2$ is the usual sphere in $\mathbb{R}^3$. Since the group $\text{SO}(3) = \text{SO}(3, \mathbb{R})$ consists of (orientation preserving) linear isometries, i.e., linear maps that are distance preserving (and of determinant +1), and every linear map leaves the origin fixed, we see that any rotation maps $S^2$ into itself.
17.2. GROUP ACTIONS: PART I, DEFINITION AND EXAMPLES

Beware that this would be false if we considered the group of affine isometries $\mathbf{SE}(3)$ of $\mathbb{E}^3$. For example, a screw motion does not map $S^2$ into itself, even though it is distance preserving, because the origin is translated.

Thus, for $X = S^2$ and $G = \text{SO}(3)$, we have an action $\cdot : \text{SO}(3) \times S^2 \to S^2$, given by the matrix multiplication

$$R \cdot x = Rx.$$ 

The verification that the above is indeed an action is trivial. This action is transitive. This is because, for any two points $x, y$ on the sphere $S^2$, there is a rotation whose axis is perpendicular to the plane containing $x, y$ and the center $O$ of the sphere (this plane is not unique when $x$ and $y$ are antipodal, i.e., on a diameter) mapping $x$ to $y$. See Figure 17.1.

![Figure 17.1: The rotation which maps $x$ to $y$.](image)

Similarly, for any $n \geq 1$, let $X = S^{n-1}$ and $G = \text{SO}(n)$ and define the action $\cdot : \text{SO}(n) \times S^{n-1} \to S^{n-1}$ as $R \cdot x = Rx$. It is easy to show that this action is transitive.

Analogously, we can define the (complex) unit sphere $\Sigma^{n-1}$, as the set of points in $\mathbb{C}^n$ given by

$$\Sigma^{n-1} = \{ (z_1, \ldots, z_n) \in \mathbb{C}^n \mid z_1 \overline{z}_1 + \cdots + z_n \overline{z}_n = 1 \}.$$ 

If we write $z_j = x_j + iy_j$, with $x_j, y_j \in \mathbb{R}$, then

$$\Sigma^{n-1} = \{ (x_1, \ldots, x_n, y_1, \ldots, y_n) \in \mathbb{R}^{2n} \mid x_1^2 + \cdots + x_n^2 + y_1^2 + \cdots + y_n^2 = 1 \}.$$ 

Therefore, we can view the complex sphere $\Sigma^{n-1}$ (in $\mathbb{C}^n$) as the real sphere $S^{2n-1}$ (in $\mathbb{R}^{2n}$).

By analogy with the real case, we can define for $X = \Sigma^{n-1}$ and $G = \text{SU}(n)$ an action $\cdot : \text{SU}(n) \times \Sigma^{n-1} \to \Sigma^{n-1}$ of the group $\text{SU}(n)$ of linear maps of $\mathbb{C}^n$ preserving the Hermitian inner product (and the origin, as all linear maps do), and this action is transitive.
One should not confuse the unit sphere $\Sigma_n^{n-1}$ with the hypersurface $S_{C}^{n-1}$, given by

$$S_{C}^{n-1} = \{(z_1, \ldots, z_n) \in \mathbb{C}^n \mid z_1^2 + \cdots + z_n^2 = 1\}.$$ 

For instance, one should check that a line $L$ through the origin intersects $\Sigma_n^{n-1}$ in a circle, whereas it intersects $S_{C}^{n-1}$ in exactly two points! Recall for a fixed $u = (x_1, \ldots, x_n, y_1, \ldots, y_n) \in \mathbb{C}^n$, that $L = \{\gamma u \mid \gamma \in \mathbb{C}\}$. Since $\gamma = \rho(\cos \theta + i \sin \theta)$, we deduce that $L$ is actually the two dimensional subspace through the origin spanned by the orthogonal vectors $(x_1, \ldots, x_n, y_1, \ldots, y_n)$ and $(-y_1, \ldots, -y_n, x_1, \ldots, x_n)$.

**Example 17.7.** The upper half-plane.

The upper half-plane $H$ is the open subset of $\mathbb{R}^2$ consisting of all points $(x, y) \in \mathbb{R}^2$, with $y > 0$. It is convenient to identify $H$ with the set of complex numbers $z \in \mathbb{C}$ such that $\Im z > 0$. Then we can let $X = H$ and $G = \text{SL}(2, \mathbb{R})$ and define an action $:: \text{SL}(2, \mathbb{R}) \times H \to H$ of the group $\text{SL}(2, \mathbb{R})$ on $H$, as follows: For any $z \in H$, for any $A \in \text{SL}(2, \mathbb{R})$, 

$$A \cdot z = \frac{az + b}{cz + d},$$

where

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

with $ad - bc = 1$.

It is easily verified that $A \cdot z$ is indeed always well defined and in $H$ when $z \in H$ (check this). To see why this action is transitive, let $z$ and $w$ be two arbitrary points of $H$ where $z = x + iy$ and $w = u + iv$ with $x, u \in \mathbb{R}$ and $y, v \in \mathbb{R}^+$ (i.e. $y$ and $v$ are positive real numbers). Define $A = \left(\begin{array}{cc} \sqrt{\frac{y}{v}} & \frac{uy - vx}{\sqrt{yv}} \\ 0 & \sqrt{\frac{v}{y}} \end{array}\right)$. Note that $A \in \text{SL}(2, \mathbb{R})$. A routine calculation shows that $A \cdot z = w$.

Before introducing Example 17.8, we need to define the groups of Möbius transformations and the Riemann sphere. Maps of the form

$$z \mapsto \frac{az + b}{cz + d},$$

where $z \in \mathbb{C}$ and $ad - bc = 1$, are called Möbius transformations. Here, $a, b, c, d \in \mathbb{R}$, but in general, we allow $a, b, c, d \in \mathbb{C}$. Actually, these transformations are not necessarily defined everywhere on $\mathbb{C}$, for example, for $z = -d/c$ if $c \neq 0$. To fix this problem, we add a “point at infinity” $\infty$ to $\mathbb{C}$, and define Möbius transformations as functions $\mathbb{C} \cup \{\infty\} \to \mathbb{C} \cup \{\infty\}$. If $c = 0$, the Möbius transformation sends $\infty$ to itself, otherwise, $-d/c \mapsto \infty$ and $\infty \mapsto a/c$.

The space $\mathbb{C} \cup \{\infty\}$ can be viewed as the plane $\mathbb{R}^2$ extended with a point at infinity. Using a stereographic projection from the sphere $S^2$ to the plane (say from the north pole to the
equatorial plane), we see that there is a bijection between the sphere $S^2$ and $\mathbb{C} \cup \{\infty\}$. More precisely, the stereographic projection $\sigma_N$ of the sphere $S^2$ from the north pole $N = (0, 0, 1)$ to the plane $z = 0$ (extended with the point at infinity $\infty$) is given by

$$(x, y, z) \in S^2 - \{(0, 0, 1)\} \mapsto \left(\frac{x}{1-z}, \frac{y}{1-z}\right) = \frac{x + iy}{1-z} \in \mathbb{C}, \quad \text{with} \quad (0, 0, 1) \mapsto \infty.$$ 

The inverse stereographic projection $\sigma^{-1}_N$ is given by

$$(u, v) \mapsto \left(\frac{2u}{u^2 + v^2 + 1}, \frac{2v}{u^2 + v^2 + 1}, \frac{u^2 + v^2 - 1}{u^2 + v^2 + 1}\right), \quad \text{with} \quad \infty \mapsto (0, 0, 1).$$

Intuitively, the inverse stereographic projection “wraps” the equatorial plane around the sphere. See Figure 17.2.

![Figure 17.2: Inverse stereographic projection.](image)

The space $\mathbb{C} \cup \{\infty\}$ is known as the Riemann sphere. We will see shortly that $\mathbb{C} \cup \{\infty\} \cong S^2$ is also the complex projective line $\mathbb{CP}^1$. In summary, Möbius transformations are bijections of the Riemann sphere. It is easy to check that these transformations form a group under composition for all $a, b, c, d \in \mathbb{C}$, with $ad - bc = 1$. This is the Möbius group, denoted $\text{Möb}^+$. The Möbius transformations corresponding to the case $a, b, c, d \in \mathbb{R}$, with $ad - bc = 1$ form a subgroup of $\text{Möb}^+$ denoted $\text{Möb}_R^+$.

The map from $\text{SL}(2, \mathbb{C})$ to $\text{Möb}^+$ that sends $A \in \text{SL}(2, \mathbb{C})$ to the corresponding Möbius transformation is a surjective group homomorphism, and one checks easily that its kernel is $\{-I, I\}$ (where $I$ is the $2 \times 2$ identity matrix). Therefore, the Möbius group $\text{Möb}^+$ is
isomorphic to the quotient group \( SL(2, \mathbb{C})/\{-I, I\} \), denoted \( PSL(2, \mathbb{C}) \). This latter group turns out to be the group of projective transformations of the projective space \( \mathbb{CP}^1 \). The same reasoning shows that the subgroup \( \text{M"ob}^+_R \) is isomorphic to \( SL(2, \mathbb{R})/\{-I, I\} \), denoted \( PSL(2, \mathbb{R}) \).

**Example 17.8.** The Riemann sphere \( \mathbb{C} \cup \{\infty\} \).

Let \( X = \mathbb{C} \cup \{\infty\} \) and \( G = SL(2, \mathbb{C}) \). The group \( SL(2, \mathbb{C}) \) acts on \( \mathbb{C} \cup \{\infty\} \cong S^2 \) the same way that \( SL(2, \mathbb{R}) \) acts on \( H \), namely: For any \( A \in SL(2, \mathbb{C}) \), for any \( z \in \mathbb{C} \cup \{\infty\} \),

\[
A \cdot z = \frac{az + b}{cz + d},
\]
where

\[
A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad \text{with} \quad ad - bc = 1.
\]

This action is transitive, an exercise we leave for the reader.

**Example 17.9.** The unit disk.

One may recall from complex analysis that the scaled (complex) Möbius transformation

\[
z \mapsto \frac{z - i}{z + i}
\]
is a biholomorphic or analytic isomorphism between the upper half plane \( H \) and the open unit disk

\[
D = \{z \in \mathbb{C} \mid |z| < 1\}.
\]

As a consequence, it is possible to define a transitive action of \( SL(2, \mathbb{R}) \) on \( D \). This can be done in a more direct fashion, using a group isomorphic to \( SL(2, \mathbb{R}) \), namely, \( SU(1, 1) \) (a group of complex matrices), but we don’t want to do this right now.

**Example 17.10.** The unit Riemann sphere revisited.

Another interesting action is the action of \( SU(2) \) on the extended plane \( \mathbb{C} \cup \{\infty\} \). Recall that the group \( SU(2) \) consists of all complex matrices of the form

\[
A = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} \quad \alpha, \beta \in \mathbb{C}, \quad \alpha\bar{\alpha} + \beta\bar{\beta} = 1,
\]
Let \( X = \mathbb{C} \cup \{\infty\} \) and \( G = SU(2) \). The action \( \cdot : SU(2) \times (\mathbb{C} \cup \{\infty\}) \to \mathbb{C} \cup \{\infty\} \) is given by

\[
A \cdot w = \frac{\alpha w + \beta}{-\beta w + \alpha}, \quad w \in \mathbb{C} \cup \{\infty\}.
\]
This action is transitive, but the proof of this fact relies on the surjectivity of the group homomorphism

\[
\rho : SU(2) \to SO(3)
\]
defined below, and the stereographic projection $\sigma_N$ from $S^2$ onto $\mathbb{C} \cup \{\infty\}$. In particular, take $z, w \in \mathbb{C} \cup \{\infty\}$, use the inverse stereographic projection to obtain two points on $S^2$, namely $\sigma_N^{-1}(z)$ and $\sigma_N^{-1}(w)$. Then apply the appropriate rotation $R \in \text{SO}(3)$ to map $\sigma_N^{-1}(z)$ onto $\sigma_N^{-1}(w)$. Such a rotation exists by the argument presented in Example 17.6. Since $\rho: \text{SU}(2) \to \text{O}(3)$ is surjective (see below), we know there must exist $A \in \text{SU}(2)$ such that $\rho(A) = R$ and $A \cdot z = w$.

Using the stereographic projection $\sigma_N$ from $S^2$ onto $\mathbb{C} \cup \{\infty\}$ and its inverse $\sigma_N^{-1}$, we can define an action of $\text{SU}(2)$ on $S^2$ by

$$A \cdot (x, y, z) = \sigma_N^{-1}(A \cdot \sigma_N(x, y, z)), \quad (x, y, z) \in S^2.$$  

Although this is not immediately obvious, it turns out that $\text{SU}(2)$ acts on $S^2$ by maps that are restrictions of linear maps to $S^2$, and since these linear maps preserve $S^2$, they are orthogonal transformations. Thus, we obtain a continuous (in fact, smooth) group homomorphism

$$\rho: \text{SU}(2) \to \text{SO}(3).$$  

Since $\text{SU}(2)$ is connected and $\rho$ is continuous, the image of $\text{SU}(2)$ is contained in the connected component of $I$ in $\text{O}(3)$, namely $\text{SO}(3)$, so $\rho$ is a homomorphism

$$\rho: \text{SU}(2) \to \text{SO}(3).$$  

We will see that this homomorphism is surjective and that its kernel is $\{I, -I\}$. The upshot is that we have an isomorphism

$$\text{SO}(3) \cong \text{SU}(2)/\{I, -I\}.$$  

The homomorphism $\rho$ is a way of describing how a unit quaternion (any element of $\text{SU}(2)$) induces a rotation, via the stereographic projection and its inverse. If we write $\alpha = a + ib$ and $\beta = c + id$, a rather tedious computation yields

$$\rho(A) = \begin{pmatrix} a^2 - b^2 - c^2 + d^2 & -2ab - 2cd & -2ac + 2bd \\ 2ab - 2cd & a^2 - b^2 + c^2 - d^2 & -2ad - 2bc \\ 2ac + 2bd & 2ad - 2bc & a^2 + b^2 - c^2 - d^2 \end{pmatrix}.$$  

One can check that $\rho(A)$ is indeed a rotation matrix which represents the rotation whose axis is the line determined by the vector $(d, -c, b)$ and whose angle $\theta \in [-\pi, \pi]$ is determined by

$$\cos \frac{\theta}{2} = |a|.$$  

We can also compute the derivative $d\rho_I: \text{su}(2) \to \text{so}(3)$ of $\rho$ at $I$ as follows. Recall that $\text{su}(2)$ consists of all complex matrices of the form

$$\begin{pmatrix} ib & c + id \\ -c + id & -ib \end{pmatrix}, \quad b, c, d \in \mathbb{R},$$
so pick the following basis for $\mathfrak{su}(2)$,

$$X_1 = \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}, \quad X_2 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 0 & i \\ i & 0 \end{pmatrix},$$

and define the curves in $\text{SU}(2)$ through $I$ given by

$$c_1(t) = \begin{pmatrix} e^{it} & 0 \\ 0 & e^{-it} \end{pmatrix}, \quad c_2(t) = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}, \quad c_3(t) = \begin{pmatrix} \cos t & i \sin t \\ i \sin t & \cos t \end{pmatrix}.$$

It is easy to check that $c_i'(0) = X_i$ for $i = 1, 2, 3$, and that

$$d\rho_I(X_1) = 2 \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad d\rho_I(X_2) = 2 \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad d\rho_I(X_3) = 2 \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}.$$

Thus we have

$$d\rho_I(X_1) = 2E_3, \quad d\rho_I(X_2) = -2E_2, \quad d\rho_I(X_3) = 2E_1,$$

where $(E_1, E_2, E_3)$ is the basis of $\mathfrak{so}(3)$ given by

$$E_1 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

which means that $d\rho_I$ is an isomorphism between the Lie algebras $\mathfrak{su}(2)$ and $\mathfrak{so}(3)$.

Recall that we have the commutative diagram

$$\begin{array}{ccc}
\text{SU}(2) & \overset{\rho}{\longrightarrow} & \text{SO}(3) \\
\downarrow \exp & & \uparrow \exp \\
\mathfrak{su}(2) & \overset{d\rho_I}{\longrightarrow} & \mathfrak{so}(3)
\end{array}$$

Since $d\rho_I$ is surjective and the exponential map $\exp: \mathfrak{so}(3) \to \text{SO}(3)$ is surjective, we conclude that $\rho$ is surjective. (We also know that $\exp: \mathfrak{su}(2) \to \text{SU}(2)$ is surjective.) Observe that $\rho(-A) = \rho(A)$, and it is easy to check that $\text{Ker} \rho = \{I, -I\}$.

**Example 17.11.** The set of $n \times n$ symmetric, positive, definite matrices, $\text{SPD}(n)$.

Let $X = \text{SPD}(n)$ and $G = \text{GL}(n)$. The group $\text{GL}(n) = \text{GL}(n, \mathbb{R})$ acts on $\text{SPD}(n)$ as follows: for all $A \in \text{GL}(n)$ and all $S \in \text{SPD}(n)$,

$$A \cdot S = ASA^T.$$ 

It is easily checked that $ASA^T$ is in $\text{SPD}(n)$ if $S$ is in $\text{SPD}(n)$. First observe that $ASA^T$ is symmetric since

$$(ASA^T)^T = AS^T A^T = ASA^T.$$
Next recall the following characterization of positive definite matrix, namely
\[ y^\top S y > 0, \quad \text{whenever } y \neq 0. \]

We want to show \( x^\top (A^\top S A)x > 0 \) for all \( x \neq 0 \). Since \( A \) is invertible, we have \( x = A^{-1}y \) for some nonzero \( y \), and hence
\[
x^\top (A^\top S A)x = y^\top (A^{-1})^\top A^\top S A A^{-1} y = y^\top S y > 0.
\]

Hence \( A^\top S A \) is positive definite. This action is transitive because every SPD matrix \( S \) can be written as \( S = AA^\top \), for some invertible matrix \( A \) (prove this as an exercise). Given any two SPD matrices \( S_1 = A_1 A_1^\top \) and \( S_2 = A_2 A_2^\top \) with \( A_1 \) and \( A_2 \) invertible, if \( A = A_2 A_1^{-1} \), we have
\[
A \cdot S_1 = A_2 A_1^{-1} S_1 (A_2 A_1^{-1})^\top = A_2 A_1^{-1} S_1 (A_1^\top)^{-1} A_2^\top
= A_2 A_1^{-1} A_1 A_1^\top (A_1^\top)^{-1} A_2^\top = A_2 A_2^\top = S_2.
\]

**Example 17.12.** The projective spaces \( \mathbb{R}P^n \) and \( \mathbb{C}P^n \).

The (real) projective space \( \mathbb{R}P^n \) is the set of all lines through the origin in \( \mathbb{R}^{n+1} \); that is, the set of one-dimensional subspaces of \( \mathbb{R}^{n+1} \) (where \( n \geq 0 \)). Since a one-dimensional subspace \( L \subset \mathbb{R}^{n+1} \) is spanned by any nonzero vector \( u \in L \), we can view \( \mathbb{R}P^n \) as the set of equivalence classes of nonzero vectors in \( \mathbb{R}^{n+1} - \{0\} \) modulo the equivalence relation
\[ u \sim v \iff v = \lambda u, \quad \text{for some } \lambda \in \mathbb{R}, \lambda \neq 0. \]

In terms of this definition, there is a projection \( pr: (\mathbb{R}^{n+1} - \{0\}) \to \mathbb{R}P^n \), given by \( pr(u) = [u]_\sim \), the equivalence class of \( u \) modulo \( \sim \). Write \([u]\) for the line defined by the nonzero vector \( u \). Since every line \( L \) in \( \mathbb{R}^{n+1} \) intersects the sphere \( S^n \) in two antipodal points, we can view \( \mathbb{R}P^n \) as the quotient of the sphere \( S^n \) by identification of antipodal points. See Figures 17.3 and 17.4.

Let \( X = \mathbb{R}P^n \) and \( G = \text{SO}(n+1) \). We define an action of \( \text{SO}(n+1) \) on \( \mathbb{R}P^n \) as follows: For any line \( L = [u] \), for any \( R \in \text{SO}(n+1) \),
\[
R \cdot L = [Ru].
\]

Since \( R \) is linear, the line \([Ru]\) is well defined; that is, does not depend on the choice of \( u \in L \). The reader can show that this action is transitive.

The (complex) projective space \( \mathbb{C}P^n \) is defined analogously as the set of all lines through the origin in \( \mathbb{C}^{n+1} \); that is, the set of one-dimensional subspaces of \( \mathbb{C}^{n+1} \) (where \( n \geq 0 \)). This time, we can view \( \mathbb{C}P^n \) as the set of equivalence classes of vectors in \( \mathbb{C}^{n+1} - \{0\} \) modulo the equivalence relation
\[ u \sim v \iff v = \lambda u, \quad \text{for some } \lambda \neq 0 \in \mathbb{C}. \]
Figure 17.3: Three constructions for $\mathbb{R}P^1 \cong S^1$. Illustration (i.) applies the equivalence relation. Since any line through the origin, excluding the $x$-axis, intersects the line $y = 1$, its equivalence class is represented by its point of intersection on $y = 1$. Hence, $\mathbb{R}P^n$ is the disjoint union of the line $y = 1$ and the point of infinity given by the $x$-axis. Illustration (ii.) represents $\mathbb{R}P^1$ as the quotient of the circle $S^1$ by identification of antipodal points. Illustration (iii.) is a variation which glues the equatorial points of the upper semicircle.

We have the projection $pr: \mathbb{C}^{n+1} - \{0\} \to \mathbb{C}P^n$, given by $pr(u) = [u]_\sim$, the equivalence class of $u$ modulo $\sim$. Again, write $[u]$ for the line defined by the nonzero vector $u$. Let $X = \mathbb{C}P^n$ and $G = \text{SU}(n + 1)$. We define an action of $\text{SU}(n + 1)$ on $\mathbb{C}P^n$ as follows: For any line $L = [u]$, for any $R \in \text{SU}(n + 1)$,$$
 R \cdot L = [Ru].$$Again, this action is well defined and it is transitive. (Check this.)

Before progressing to our final example of group actions, we take a moment to construct $\mathbb{C}P^n$ as a quotient space of $S^{2n+1}$. Recall that $\Sigma^n \subseteq \mathbb{C}^{n+1}$, the unit sphere in $\mathbb{C}^{n+1}$, is defined by$$\Sigma^n = \{(z_1, \ldots, z_{n+1}) \in \mathbb{C}^{n+1} \mid z_1 \bar{z}_1 + \cdots + z_{n+1} \bar{z}_{n+1} = 1\}.$$For any line $L = [u]$, where $u \in \mathbb{C}^{n+1}$ is a nonzero vector, writing $u = (u_1, \ldots, u_{n+1})$, a point $z \in \mathbb{C}^{n+1}$ belongs to $L$ iff $z = \lambda(u_1, \ldots, u_{n+1})$, for some $\lambda \in \mathbb{C}$. Therefore, the intersection
17.2. GROUP ACTIONS: PART I, DEFINITION AND EXAMPLES

Figure 17.4: Three constructions for $\mathbb{R}P^2$. Illustration (i.) applies the equivalence relation. Since any line through the origin which is not contained in the $xy$-plane intersects the plane $z = 1$, its equivalence class is represented by its point of intersection on $z = 1$. Hence, $\mathbb{R}P^2$ is the disjoint union of the plane $z = 1$ and the copy of $\mathbb{R}P^1$ provided by the $xy$-plane. Illustration (ii.) represents $\mathbb{R}P^2$ as the quotient of the sphere $S^2$ by identification of antipodal points. Illustration (iii.) is a variation which glues the antipodal points on boundary of the unit disk, which is represented as as the upper hemisphere.

$L \cap \Sigma^n$ of the line $L$ and the sphere $\Sigma^n$ is given by

$$L \cap \Sigma^n = \{ \lambda(u_1, \ldots, u_{n+1}) \in \mathbb{C}^{n+1} \mid \lambda \in \mathbb{C}, \lambda\overline{\lambda}(u_1\overline{u_1} + \cdots + u_{n+1}\overline{u_{n+1}}) = 1 \},$$

i.e.,

$$L \cap \Sigma^n = \left\{ \lambda(u_1, \ldots, u_{n+1}) \in \mathbb{C}^{n+1} \mid \lambda \in \mathbb{C}, |\lambda| = \frac{1}{\sqrt{|u_1|^2 + \cdots + |u_{n+1}|^2}} \right\}.$$

Thus, we see that there is a bijection between $L \cap \Sigma^n$ and the circle $S^1$; that is, geometrically $L \cap \Sigma^n$ is a circle. Moreover, since any line $L$ through the origin is determined by just one other point, we see that for any two lines $L_1$ and $L_2$ through the origin,

$$L_1 \neq L_2 \iff (L_1 \cap \Sigma^n) \cap (L_2 \cap \Sigma^n) = \emptyset.$$

However, $\Sigma^n$ is the sphere $S^{2n+1}$ in $\mathbb{R}^{2n+2}$. It follows that $\mathbb{C}P^n$ is the quotient of $S^{2n+1}$ by the equivalence relation $\sim$ defined such that

$$y \sim z \iff y, z \in L \cap \Sigma^n, \text{ for some line, } L, \text{ through the origin.}$$
Therefore, we can write
\[ S^{2n+1}/S^1 \cong \mathbb{CP}^n. \]
The case \( n = 1 \) is particularly interesting, as it turns out that
\[ S^3/S^1 \cong S^2. \]
This is the famous Hopf fibration. To show this, proceed as follows: As
\[ S^3 \cong \Sigma^1 = \{(z, z') \in \mathbb{C}^2 \mid |z|^2 + |z'|^2 = 1\}, \]
define a map, \( HF : S^3 \to S^2 \), by
\[ HF((z, z')) = (2zz^\dagger, |z|^2 - |z'|^2). \]
We leave as a homework exercise to prove that this map has range \( S^2 \) and that
\[ HF((z_1, z'_1)) = HF((z_2, z'_2)) \iff (z_1, z'_1) = \lambda(z_2, z'_2), \quad \text{for some } \lambda \text{ with } |\lambda| = 1. \]

In other words, for any point, \( p \in S^2 \), the inverse image \( HF^{-1}(p) \) (also called fibre over \( p \)) is a circle on \( S^3 \). Consequently, \( S^3 \) can be viewed as the union of a family of disjoint circles. This is the Hopf fibration. It is possible to visualize the Hopf fibration using the stereographic projection from \( S^3 \) onto \( \mathbb{R}^3 \). This is a beautiful and puzzling picture. For example, see Berger [4]. Therefore, \( HF \) induces a bijection from \( \mathbb{CP}^1 \) to \( S^2 \), and it is a homeomorphism.

**Example 17.13.** Affine spaces.

Let \( X \) be a set and \( E \) a real vector space. A transitive and faithful action \( \cdot : E \times X \to X \) of the additive group of \( E \) on \( X \) makes \( X \) into an affine space. The intuition is that the members of \( E \) are translations.

Those familiar with affine spaces as in Gallier [35] (Chapter 2) or Berger [4] will point out that if \( X \) is an affine space, then not only is the action of \( E \) on \( X \) transitive, but more is true: For any two points \( a, b \in E \), there is a unique vector \( u \in E \), such that \( u \cdot a = b \). By the way, the action of \( E \) on \( X \) is usually considered to be a right action and is written additively, so \( u \cdot a \) is written \( a + u \) (the result of translating \( a \) by \( u \)). Thus, it would seem that we have to require more of our action. However, this is not necessary because \( E \) (under addition) is abelian. More precisely, we have the proposition

**Proposition 17.13.** If \( G \) is an abelian group acting on a set \( X \) and the action \( \cdot : G \times X \to X \) is transitive and faithful, then for any two elements \( x, y \in X \), there is a unique \( g \in G \) so that \( g \cdot x = y \) (the action is simply transitive).

**Proof.** Since our action is transitive, there is at least some \( g \in G \) so that \( g \cdot x = y \). Assume that we have \( g_1, g_2 \in G \) with
\[ g_1 \cdot x = g_2 \cdot x = y. \]
We shall prove that
\[ g_1 \cdot z = g_2 \cdot z, \quad \text{for all } z \in X. \]

This implies that
\[ g_1 g_2^{-1} \cdot z = z, \quad \text{for all } z \in X. \]

As our action is faithful, $g_1 g_2^{-1} = 1$, and we must have $g_1 = g_2$, which proves our proposition.

Pick any $z \in X$. As our action is transitive, there is some $h \in G$ so that $z = h \cdot x$. Then, we have
\[
g_1 \cdot z = g_1 \cdot (h \cdot x) \\
= (g_1 h) \cdot x \\
= (hg_1) \cdot x \quad \text{(since } G \text{ is abelian)} \\
= h \cdot (g_1 \cdot x) \\
= h \cdot (g_2 \cdot x) \quad \text{(since } g_1 \cdot x = g_2 \cdot x) \\
= (hg_2) \cdot x \\
= (g_2 h) \cdot x \quad \text{(since } G \text{ is abelian)} \\
= g_2 \cdot (h \cdot x) \\
= g_2 \cdot z.
\]

Therefore, $g_1 \cdot z = g_2 \cdot z$ for all $z \in X$, as claimed. \qed

### 17.3 Group Actions: Part II, Stabilizers and Homogeneous Spaces

Now that we have an understanding of how a group $G$ acts on a set $X$, we may use this action to form new topological spaces, namely homogeneous spaces. In the construction of homogeneous spaces, the subset of group elements that leaves some given element $x \in X$ fixed plays an important role.

**Definition 17.15.** Given an action $\cdot \colon G \times X \to X$ of a group $G$ on a set $X$, for any $x \in X$, the group $G_x$ (also denoted $\text{Stab}_G(x)$), called the stabilizer of $x$ or isotropy group at $x$, is given by

\[ G_x = \{ g \in G \mid g \cdot x = x \}. \]

We have to verify that $G_x$ is indeed a subgroup of $G$, but this is easy. Indeed, if $g \cdot x = x$ and $h \cdot x = x$, then we also have $h^{-1} \cdot x = x$ and so, we get $gh^{-1} \cdot x = x$, proving that $G_x$ is a subgroup of $G$. In general, $G_x$ is **not** a normal subgroup.
Observe that
\[ G_{g \cdot x} = gG_x g^{-1}, \]
for all \( g \in G \) and all \( x \in X \). Indeed,
\[
G_{g \cdot x} = \{ h \in G \mid h \cdot (g \cdot x) = g \cdot x \} \\
= \{ h \in G \mid hg \cdot x = g \cdot x \} \\
= \{ h \in G \mid g^{-1}hg \cdot x = x \},
\]
which shows \( g^{-1}G_{g \cdot x}g \subseteq G_x \), or equivalently that \( G_{g \cdot x} \subseteq gG_x g^{-1} \). It remains to show that \( gG_x g^{-1} \subseteq G_{g \cdot x} \). Take an element of \( gG_x g^{-1} \), which has the form \( ghg^{-1} \) with \( h \cdot x = x \). Since \( h \cdot x = x \), we have \( (ghg^{-1}) \cdot gx = gx \), which shows that \( ghg^{-1} \in G_{g \cdot x} \).

Because \( G_{g \cdot x} = gG_x g^{-1} \), the stabilizers of \( x \) and \( g \cdot x \) are conjugate of each other.

When the action of \( G \) on \( X \) is transitive, for any fixed \( x \in G \), the set \( X \) is a quotient (as a set, not as group) of \( G \) by \( G_x \). Indeed, we can define the map, \( \pi_x : G \to X \), by
\[
\pi_x(g) = g \cdot x, \quad \text{for all } g \in G.
\]
Observe that
\[
\pi_x(gG_x) = (gG_x) \cdot x = g \cdot (G_x \cdot x) = g \cdot x = \pi_x(g).
\]
This shows that \( \pi_x : G \to X \) induces a quotient map \( \overline{\pi}_x : G/G_x \to X \), from the set \( G/G_x \) of (left) cosets of \( G_x \) to \( X \), defined by
\[
\overline{\pi}_x(gG_x) = g \cdot x.
\]
Since
\[
\pi_x(g) = \pi_x(h) \iff g \cdot x = h \cdot x \iff g^{-1}h \cdot x = x \iff g^{-1}h \in G_x \iff gG_x = hG_x,
\]
we deduce that \( \overline{\pi}_x : G/G_x \to X \) is injective. However, since our action is transitive, for every \( y \in X \), there is some \( g \in G \) so that \( g \cdot x = y \), and so \( \overline{\pi}_x(gG_x) = g \cdot x = y \); that is, the map \( \pi_x \) is also surjective. Therefore, the map \( \overline{\pi}_x : G/G_x \to X \) is a bijection (of sets, not groups).

The map \( \pi_x : G \to X \) is also surjective. Let us record this important fact as

**Proposition 17.14.** If \( \cdot : G \times X \to X \) is a transitive action of a group \( G \) on a set \( X \), for every fixed \( x \in X \), the surjection \( \pi_x : G \to X \) given by
\[
\pi_x(g) = g \cdot x
\]
induces a bijection
\[
\overline{\pi}_x : G/G_x \to X,
\]
where \( G_x \) is the stabilizer of \( x \). See Figure 17.5.
Figure 17.5: A schematic representation of $G/G_x \cong X$, where $G$ is the gray solid, $X$ is its purple circular base, and $G_x$ is the pink vertical strand. The dotted strands are the fibres $gG_x$.

The map $\pi_x : G \to X$ (corresponding to a fixed $x \in X$) is sometimes called a projection of $G$ onto $X$. Proposition 17.14 shows that for every $y \in X$, the subset $\pi_x^{-1}(y)$ of $G$ (called the fibre above $y$) is equal to some coset $gG_x$ of $G$, and thus is in bijection with the group $G_x$ itself. We can think of $G$ as a moving family of fibres $G_x$ parametrized by $X$. This point of view of viewing a space as a moving family of simpler spaces is typical in (algebraic) geometry, and underlies the notion of (principal) fibre bundle.

Note that if the action $\cdot : G \times X \to X$ is transitive, then the stabilizers $G_x$ and $G_y$ of any two elements $x, y \in X$ are isomorphic, as they as conjugates. Thus, in this case, it is enough to compute one of these stabilizers for a “convenient” $x$.

As the situation of Proposition 17.14 is of particular interest, we make the following definition:

**Definition 17.16.** A set $X$ is said to be a *homogeneous space* if there is a transitive action $\cdot : G \times X \to X$ of some group $G$ on $X$.

We see that all the spaces of Examples 17.6–17.13, are homogeneous spaces. Another example that will play an important role when we deal with Lie groups is the situation where we have a group $G$, a subgroup $H$ of $G$ (not necessarily normal), and where $X = G/H$, the set of left cosets of $G$ modulo $H$. The group $G$ acts on $G/H$ by left multiplication:

$$a \cdot (gH) = (ag)H,$$
where \( a, g \in G \). This action is clearly transitive and one checks that the stabilizer of \( gH \) is \( gHg^{-1} \). If \( G \) is a topological group and \( H \) is a closed subgroup of \( G \) (see later for an explanation), it turns out that \( G/H \) is Hausdorff. If \( G \) is a Lie group, we obtain a manifold.

Even if \( G \) and \( X \) are topological spaces and the action \( \cdot : G \times X \to X \) is continuous, in general, the space \( G/G_x \) under the quotient topology is not homeomorphic to \( X \).

We will give later sufficient conditions that insure that \( X \) is indeed a topological space or even a manifold. In particular, \( X \) will be a manifold when \( G \) is a Lie group.

In general, an action \( \cdot : G \times X \to X \) is not transitive on \( X \), but for every \( x \in X \), it is transitive on the set

\[
O(x) = G \cdot x = \{ g \cdot x \mid g \in G \}.
\]

Such a set is called the orbit of \( x \). The orbits are the equivalence classes of the following equivalence relation:

**Definition 17.17.** Given an action \( \cdot : G \times X \to X \) of some group \( G \) on \( X \), the equivalence relation \( \sim \) on \( X \) is defined so that, for all \( x, y \in X \),

\[
x \sim y \iff y = g \cdot x, \text{ for some } g \in G.
\]

For every \( x \in X \), the equivalence class of \( x \) is the orbit of \( x \), denoted \( O(x) \) or \( G \cdot x \), with

\[
G \cdot x = O(x) = \{ g \cdot x \mid g \in G \}.
\]

The set of orbits is denoted \( X/G \).

We warn the reader that some authors use the notation \( G \setminus X \) for the the set of orbits \( G \cdot x \), because these orbits can be considered as right orbits, by analogy with right cosets \( Hg \) of a subgroup \( H \) of \( G \).

The orbit space \( X/G \) is obtained from \( X \) by an identification (or merging) process: For every orbit, all points in that orbit are merged into a single point. This akin to the process of forming the identification topology. For example, if \( X = S^2 \) and \( G \) is the group consisting of the restrictions of the two linear maps \( I \) and \( -I \) of \( \mathbb{R}^3 \) to \( S^2 \) (where \((-I)(x) = -x \) for all \( x \in \mathbb{R}^3 \)), then

\[
X/G = S^2/\{I, -I\} \cong \mathbb{R}P^2.
\]

See Figure 17.4. More generally, if \( S^n \) is the \( n \)-sphere in \( \mathbb{R}^{n+1} \), then we have a bijection between the orbit space \( S^n/\{I, -I\} \) and \( \mathbb{R}P^n \):

\[
S^n/\{I, -I\} \cong \mathbb{R}P^n.
\]

Many manifolds can be obtained in this fashion, including the torus, the Klein bottle, the Möbius band, etc.
17.3. GROUP ACTIONS: PART II, STABILIZERS AND HOMOGENEOUS SPACES

Since the action of $G$ is transitive on $O(x)$, by Proposition 17.14, we see that for every $x \in X$, we have a bijection

$$O(x) \cong G/G_x.$$ 

As a corollary, if both $X$ and $G$ are finite, for any set $A \subseteq X$ of representatives from every orbit, we have the orbit formula:

$$|X| = \sum_{a \in A} [G: G_a] = \sum_{a \in A} |G|/|G_a|.$$ 

Even if a group action $\cdot: G \times X \to X$ is not transitive, when $X$ is a manifold, we can consider the set of orbits $X/G$, and if the action of $G$ on $X$ satisfies certain conditions, $X/G$ is actually a manifold. Manifolds arising in this fashion are often called orbifolds. In summary, we see that manifolds arise in at least two ways from a group action:

1. As homogeneous spaces $G/G_x$, if the action is transitive.
2. As orbifolds $X/G$ (under certain conditions on the action).

Of course, in both cases, the action must satisfy some additional properties.

For the rest of this section, we reconsider Examples 17.6–17.13 in the context of homogeneous space by determining some stabilizers for those actions.

(a) Consider the action $\cdot: \text{SO}(n) \times S^{n-1} \to S^{n-1}$ of $\text{SO}(n)$ on the sphere $S^{n-1}$ ($n \geq 1$) defined in Example 17.6. Since this action is transitive, we can determine the stabilizer of any convenient element of $S^{n-1}$, say $e_1 = (1,0,\ldots,0)$. In order for any $R \in \text{SO}(n)$ to leave $e_1$ fixed, the first column of $R$ must be $e_1$, so $R$ is an orthogonal matrix of the form

$$R = \begin{pmatrix} \text{ } & U \\ \text{ } & \text{ } \end{pmatrix} \quad \text{with} \quad \det(S) = 1,$$

where $U$ is a $1 \times (n-1)$ row vector. As the rows of $R$ must be unit vectors, we see that $U = 0$ and $S \in \text{SO}(n-1)$. Therefore, the stabilizer of $e_1$ is isomorphic to $\text{SO}(n-1)$, and we deduce the bijection

$$\text{SO}(n)/\text{SO}(n-1) \cong S^{n-1}.$$ 

Strictly speaking, $\text{SO}(n-1)$ is not a subgroup of $\text{SO}(n)$, and in all rigor, we should consider the subgroup $\widetilde{\text{SO}}(n-1)$ of $\text{SO}(n)$ consisting of all matrices of the form

$$\begin{pmatrix} 1 & 0 \\ 0 & S \end{pmatrix}, \quad \text{with} \quad \det(S) = 1,$$

and write

$$\text{SO}(n)/\widetilde{\text{SO}}(n-1) \cong S^{n-1}.$$ 

However, it is common practice to identify $\text{SO}(n-1)$ with $\widetilde{\text{SO}}(n-1)$.
When $n = 2$, as $\text{SO}(1) = \{1\}$, we find that $\text{SO}(2) \cong S^1$, a circle, a fact that we already knew. When $n = 3$, we find that $\text{SO}(3)/\text{SO}(2) \cong S^2$. This says that $\text{SO}(3)$ is somehow the result of glueing circles to the surface of a sphere (in $\mathbb{R}^3$), in such a way that these circles do not intersect. This is hard to visualize!

A similar argument for the complex unit sphere $\Sigma^{n-1}$ shows that

$$\text{SU}(n)/\text{SU}(n-1) \cong \Sigma^{n-1} \cong S^{2n-1}.$$ 

Again, we identify $\text{SU}(n-1)$ with a subgroup of $\text{SU}(n)$, as in the real case. In particular, when $n = 2$, as $\text{SU}(1) = \{1\}$, we find that

$$\text{SU}(2) \cong S^3;$$

that is, the group $\text{SU}(2)$ is topologically the sphere $S^3$! Actually, this is not surprising if we remember that $\text{SU}(2)$ is in fact the group of unit quaternions.

(b) We saw in Example 17.7 that the action $\cdot : \text{SL}(2, \mathbb{R}) \times H \to H$ of the group $\text{SL}(2, \mathbb{R})$ on the upper half plane is transitive. Let us find out what the stabilizer of $z = i$ is. We should have

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

that is, $ai + b = -c + di$, i.e.,

$$(d - a)i = b + c.$$ 

Since $a, b, c, d$ are real, we must have $d = a$ and $b = -c$. Moreover, $ad - bc = 1$, so we get $a^2 + b^2 = 1$. We conclude that a matrix in $\text{SL}(2, \mathbb{R})$ fixes $i$ iff it is of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \quad \text{with} \quad a^2 + b^2 = 1.$$ 

Clearly, these are the rotation matrices in $\text{SO}(2)$, and so the stabilizer of $i$ is $\text{SO}(2)$. We conclude that

$$\text{SL}(2, \mathbb{R})/\text{SO}(2) \cong H.$$ 

This time we can view $\text{SL}(2, \mathbb{R})$ as the result of glueing circles to the upper half plane. This is not so easy to visualize. There is a better way to visualize the topology of $\text{SL}(2, \mathbb{R})$ by making it act on the open disk $D$. We will return to this action in a little while.

(c) Now consider the action of $\text{SL}(2, \mathbb{C})$ on $\mathbb{C} \cup \{\infty\} \cong S^2$ given in Example 17.8. As it is transitive, let us find the stabilizer of $z = 0$. We must have

$$\frac{b}{d} = 0,$$

and as $ad - bc = 1$, we must have $b = 0$ and $ad = 1$. Thus the stabilizer of 0 is the subgroup $\text{SL}(2, \mathbb{C})_0$ of $\text{SL}(2, \mathbb{C})$ consisting of all matrices of the form

$$\begin{pmatrix} a & 0 \\ c & a^{-1} \end{pmatrix}, \quad \text{where} \quad a \in \mathbb{C} \setminus \{0\} \quad \text{and} \quad c \in \mathbb{C}.$$
We get
\[ \text{SL}(2, \mathbb{C})/\text{SL}(2, \mathbb{C})_0 \cong \mathbb{C} \cup \{\infty\} \cong S^2, \]
but this is not very illuminating.

(d) In Example 17.11 we considered the action \( \cdot : \text{GL}(n) \times \text{SPD}(n) \to \text{SPD}(n) \) of \( \text{GL}(n) \) on \( \text{SPD}(n) \), the set of symmetric positive definite matrices. As this action is transitive, let us find the stabilizer of \( I \). For any \( A \in \text{GL}(n) \), the matrix \( A \) stabilizes \( I \) iff
\[ AIA^\top = AA^\top = I. \]
Therefore the stabilizer of \( I \) is \( \text{O}(n) \), and we find that
\[ \text{GL}(n)/\text{O}(n) = \text{SPD}(n). \]

Observe that if \( \text{GL}^+(n) \) denotes the subgroup of \( \text{GL}(n) \) consisting of all matrices with a strictly positive determinant, then we have an action \( \cdot : \text{GL}^+(n) \times \text{SPD}(n) \to \text{SPD}(n) \) of \( \text{GL}^+(n) \) on \( \text{SPD}(n) \). This action is transitive and we find that the stabilizer of \( I \) is \( \text{SO}(n) \); consequently, we get
\[ \text{GL}^+(n)/\text{SO}(n) = \text{SPD}(n). \]

(e) In Example 17.12 we considered the action \( \cdot : \text{SO}(n+1) \times \mathbb{RP}^n \to \mathbb{RP}^n \) of \( \text{SO}(n+1) \) on the (real) projective space \( \mathbb{RP}^n \). As this action is transitive, let us find the stabilizer of the line \( L = \{e_1\} \), where \( e_1 = (1, 0, \ldots, 0) \). For any \( R \in \text{SO}(n+1) \), the line \( L \) is fixed iff either \( R(e_1) = e_1 \) or \( R(e_1) = -e_1 \), since \( e_1 \) and \(-e_1\) define the same line. As \( R \) is orthogonal with \( \det(R) = 1 \), this means that \( R \) is of the form
\[ R = \begin{pmatrix} \alpha & 0 \\ 0 & S \end{pmatrix}, \quad \text{with } \alpha = \pm 1 \quad \text{and} \quad \det(S) = \alpha. \]
But, \( S \) must be orthogonal, so we conclude \( S \in \text{O}(n) \). Therefore, the stabilizer of \( L = \{e_1\} \) is isomorphic to the group \( \text{O}(n) \), and we find that
\[ \text{SO}(n+1)/\text{O}(n) \cong \mathbb{RP}^n. \]

Strictly speaking, \( \text{O}(n) \) is not a subgroup of \( \text{SO}(n+1) \), so the above equation does not make sense. We should write
\[ \text{SO}(n+1)/\tilde{\text{O}}(n) \cong \mathbb{RP}^n, \]
where \( \tilde{\text{O}}(n) \) is the subgroup of \( \text{SO}(n+1) \) consisting of all matrices of the form
\[ \begin{pmatrix} \alpha & 0 \\ 0 & S \end{pmatrix}, \quad \text{with } S \in \text{O}(n), \alpha = \pm 1 \quad \text{and} \quad \det(S) = \alpha. \]
This group is also denoted \( S(\text{O}(1) \times \text{O}(n)) \). However, the common practice is to write \( \text{O}(n) \) instead of \( S(\text{O}(1) \times \text{O}(n)) \).
We should mention that $\mathbb{RP}^3$ and $\text{SO}(3)$ are homeomorphic spaces. This is shown using the quaternions; for example, see Gallier [35], Chapter 8.

A similar argument applies to the action $\cdot : \text{SU}(n+1) \times \mathbb{C}P^n \to \mathbb{C}P^n$ of $\text{SU}(n+1)$ on the (complex) projective space $\mathbb{C}P^n$. We find that

$$\text{SU}(n+1)/\text{U}(n) \cong \mathbb{C}P^n.$$ 

Again, the above is a bit sloppy as $\text{U}(n)$ is not a subgroup of $\text{SU}(n+1)$. To be rigorous, we should use the subgroup $\tilde{\text{U}}(n)$ consisting of all matrices of the form

$$\begin{pmatrix} \alpha & 0 \\ 0 & S \end{pmatrix}, \quad \text{with} \quad S \in \text{U}(n), \quad |\alpha| = 1 \quad \text{and} \quad \det(S) = \bar{\alpha}.$$ 

This group is also denoted $S(\text{U}(1) \times \text{U}(n))$. The common practice is to write $\text{U}(n)$ instead of $S(\text{U}(1) \times \text{U}(n))$. In particular, when $n = 1$, we find that

$$\text{SU}(2)/\text{U}(1) \cong \mathbb{C}P^1.$$ 

But, we know that $\text{SU}(2) \cong S^3$, and clearly $\text{U}(1) \cong S^1$. So, again, we find that $S^3/S^1 \cong \mathbb{C}P^1$ (we know more, namely, $S^3/S^1 \cong S^2 \cong \mathbb{C}P^1$.)

Observe that $\mathbb{C}P^n$ can also be viewed as the orbit space of the action $\cdot : S^1 \times S^{2n+1} \to S^{2n+1}$ given by

$$\lambda \cdot (z_1, \ldots, z_{n+1}) = (\lambda z_1, \ldots, \lambda z_{n+1}),$$

where $S^1 = \text{U}(1)$ (the group of complex numbers of modulus 1) and $S^{2n+1}$ is identified with $\Sigma^n$.

We now return to Case (b) to give a better picture of $\text{SL}(2, \mathbb{R})$. Instead of having $\text{SL}(2, \mathbb{R})$ act on the upper half plane, we define an action of $\text{SL}(2, \mathbb{R})$ on the open unit disk $D$ as we did in Example 17.9. Technically, it is easier to consider the group $\text{SU}(1, 1)$, which is isomorphic to $\text{SL}(2, \mathbb{R})$, and to make $\text{SU}(1, 1)$ act on $D$. The group $\text{SU}(1, 1)$ is the group of $2 \times 2$ complex matrices of the form

$$\begin{pmatrix} a & b \\ \bar{b} & \bar{a} \end{pmatrix}, \quad \text{with} \quad a\bar{a} - b\bar{b} = 1.$$ 

The reader should check that if we let

$$g = \begin{pmatrix} 1 & -i \\ 1 & i \end{pmatrix},$$

then the map from $\text{SL}(2, \mathbb{R})$ to $\text{SU}(1, 1)$ given by

$$A \mapsto gAg^{-1}$$

would be the desired action on $D$. This action shows that $S^2 \cong \mathbb{C}P^1$.
is an isomorphism. Observe that the scaled Möbius transformation associated with \( g \) is

\[ z \mapsto z - i \frac{z}{z + i}, \]

which is the holomorphic isomorphism mapping \( H \) to \( D \) mentioned earlier! We can define a bijection between \( \text{SU}(1, 1) \) and \( S^1 \times D \) given by

\[ \left( \begin{array}{cc} a & b \\ b & a \end{array} \right) \mapsto (a/|a|, b/a). \]

We conclude that \( \text{SL}(2, \mathbb{R}) \cong \text{SU}(1, 1) \) is topologically an open solid torus (i.e., with the surface of the torus removed). It is possible to further classify the elements of \( \text{SL}(2, \mathbb{R}) \) into three categories and to have geometric interpretations of these as certain regions of the torus. For details, the reader should consult Carter, Segal and Macdonald [15] or Duistermatt and Kolk [27] (Chapter 1, Section 1.2).

The group \( \text{SU}(1, 1) \) acts on \( D \) by interpreting any matrix in \( \text{SU}(1, 1) \) as a Möbius transformation; that is,

\[ \left( \begin{array}{cc} a & b \\ b & a \end{array} \right) \mapsto \left( z \mapsto \frac{az + b}{\bar{b}z + \bar{a}} \right). \]

The reader should check that these transformations preserve \( D \).

Both the upper half-plane and the open disk are models of Lobachevsky’s non-Euclidean geometry (where the parallel postulate fails). They are also models of hyperbolic spaces (Riemannian manifolds with constant negative curvature, see Gallot, Hulin and Lafontaine [37], Chapter III). According to Dubrovin, Fomenko, and Novikov [26] (Chapter 2, Section 13.2), the open disk model is due to Poincaré and the upper half-plane model to Klein, although Poincaré was the first to realize that the upper half-plane is a hyperbolic space.

### 17.4 The Grassmann and Stiefel Manifolds

In this section we introduce two very important homogeneous manifolds, the Grassmann manifolds and the Stiefel manifolds. The Grassmann manifolds are generalizations of projective spaces (real and complex), while the Stiefel manifold are generalizations of \( \text{O}(n) \). Both of these manifolds are examples of reductive homogeneous spaces. We begin by defining the Grassmann manifolds \( G(k, n) \).

First consider the real case.

**Definition 17.18.** Given any \( n \geq 1 \), for any \( k \) with \( 0 \leq k \leq n \), the set \( G(k, n) \) of all linear \( k \)-dimensional subspaces of \( \mathbb{R}^n \) (also called \( k \)-planes) is called a Grassmannian (or Grassmann manifold).
Any $k$-dimensional subspace $U$ of $\mathbb{R}^n$ is spanned by $k$ linearly independent vectors $u_1, \ldots, u_k$ in $\mathbb{R}^n$; write $U = \text{span}(u_1, \ldots, u_k)$. We can define an action $\cdot : O(n) \times G(k, n) \to G(k, n)$ as follows: For any $R \in O(n)$, for any $U = \text{span}(u_1, \ldots, u_k)$, let

$$R \cdot U = \text{span}(Ru_1, \ldots, Ru_k).$$

We have to check that the above is well defined. If $U = \text{span}(v_1, \ldots, v_k)$ for any other $k$ linearly independent vectors $v_1, \ldots, v_k$, we have

$$v_i = \sum_{j=1}^{k} a_{ij}u_j, \quad 1 \leq i \leq k,$$

for some $a_{ij} \in \mathbb{R}$, and so

$$Rv_i = \sum_{j=1}^{k} a_{ij}Ru_j, \quad 1 \leq i \leq k,$$

which shows that

$$\text{span}(Ru_1, \ldots, Ru_k) = \text{span}(Rv_1, \ldots, Rv_k);$$

that is, the above action is well defined.

We claim this action is transitive. This is because if $U$ and $V$ are any two $k$-planes, we may assume that $U = \text{span}(u_1, \ldots, u_k)$ and $V = \text{span}(v_1, \ldots, v_k)$, where the $u_i$’s form an orthonormal family and similarly for the $v_i$’s. Then we can extend these families to orthonormal bases $(u_1, \ldots, u_n)$ and $(v_1, \ldots, v_n)$ on $\mathbb{R}^n$, and w.r.t. the orthonormal basis $(u_1, \ldots, u_n)$, the matrix of the linear map sending $u_i$ to $v_i$ is orthogonal. Hence $G(k, n)$ is a homogeneous space.

In order to represent $G(k, n)$ as a quotient space, Proposition 17.14 implies it is enough to find the stabilizer of any $k$-plane. Pick $U = \text{span}(e_1, \ldots, e_k)$, where $(e_1, \ldots, e_n)$ is the canonical basis of $\mathbb{R}^n$ (i.e., $e_i = (0, \ldots, 0, 1, 0, \ldots, 0)$, with the 1 in the $i$th position). Any $R \in O(n)$ stabilizes $U$ iff $R$ maps $e_1, \ldots, e_k$ to $k$ linearly independent vectors in the subspace $U = \text{span}(e_1, \ldots, e_k)$, i.e., $R$ is of the form

$$R = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix},$$

where $S$ is $k \times k$ and $T$ is $(n - k) \times (n - k)$. Moreover, as $R$ is orthogonal, $S$ and $T$ must be orthogonal, that is $S \in O(k)$ and $T \in O(n - k)$. We deduce that the stabilizer of $U$ is isomorphic to $O(k) \times O(n - k)$ and we find that

$$O(n)/(O(k) \times O(n - k)) \cong G(k, n).$$

It turns out that this makes $G(k, n)$ into a smooth manifold of dimension

$$\frac{n(n - 1)}{2} - \frac{k(k - 1)}{2} - \frac{(n - k)(n - k - 1)}{2} = k(n - k).$$
17.4. THE GRASSMANN AND STIEFEL MANIFOLDS

called a Grassmannian.

The restriction of the action of $O(n)$ on $G(k, n)$ to $SO(n)$ yields an action
$\cdot : SO(n) \times G(k, n) \rightarrow G(k, n)$ of $SO(n)$ on $G(k, n)$. Then it is easy to see that this action
is transitive and that the stabilizer of the subspace $U$ is isomorphic to the subgroup
$SO(k) \times O(n - k))$ of $SO(n)$ consisting of the rotations of the form

$$R = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix},$$

with $S \in O(k)$, $T \in O(n - k)$ and $\det(S) \det(T) = 1$. Thus, we also have

$$SO(n)/SO(k) \times O(n - k)) \cong G(k, n).$$

If we recall the projection map of Example 17.12 in Section 17.2, namely $pr : \mathbb{R}^{n+1} - \{0\} \rightarrow \mathbb{R}^n$, by definition, a $k$-plane in $\mathbb{R}^n$ is the image under $pr$ of any $(k + 1)$-plane in $\mathbb{R}^{n+1}$. So, for example, a line in $\mathbb{R}^n$ is the image of a 2-plane in $\mathbb{R}^{n+1}$, and a hyperplane in $\mathbb{R}^n$ is the image of a hyperplane in $\mathbb{R}^{n+1}$. The advantage of this point of view is that the $k$-planes in $\mathbb{R}^n$ are arbitrary; that is, they do not have to go through “the origin” (which does not make sense, anyway!). Then we see that we can interpret the Grassmannian, $G(k + 1, n + 1)$, as a space of “parameters” for the $k$-planes in $\mathbb{R}^n$. For example, $G(2, n + 1)$ parametrizes the lines in $\mathbb{R}^n$. In this viewpoint, $G(k + 1, n + 1)$ is usually denoted $G(k, n)$.

It can be proved (using some exterior algebra) that $G(k, n)$ can be embedded in $\mathbb{R}^{n \choose k} - 1$. Much more is true. For example, $G(k, n)$ is a projective variety, which means that it can be defined as a subset of $\mathbb{R}^{n \choose k} - 1$ equal to the zero locus of a set of homogeneous equations. There is even a set of quadratic equations known as the Plücker equations defining $G(k, n)$. In particular, when $n = 4$ and $k = 2$, we have $G(2, 4) \subseteq \mathbb{R}^5$, and $G(2, 4)$ is defined by a single equation of degree 2. The Grassmannian $G(2, 4) = G(1, 3)$ is known as the Klein quadric. This hypersurface in $\mathbb{R}^5$ parametrizes the lines in $\mathbb{R}^3$.

Complex Grassmannians are defined in a similar way, by replacing $\mathbb{R}$ by $\mathbb{C}$ and $O(n)$ by $U(n)$ throughout. The complex Grassmannian $G_C(k, n)$ is a complex manifold as well as a real manifold, and we have

$$U(n)/(U(k) \times U(n - k)) \cong G_C(k, n).$$

As in the case of the real Grassmannians, the action of $U(n)$ on $G_C(k, n)$ yields an action of $SU(n)$ on $G_C(k, n)$, and we get

$$SU(n)/S(U(k) \times U(n - k)) \cong G_C(k, n),$$

where $S(U(k) \times U(n - k))$ is the subgroup of $SU(n)$ consisting of all matrices $R \in SU(n)$ of the form

$$R = \begin{pmatrix} S & 0 \\ 0 & T \end{pmatrix},$$
with \( S \in U(k) \), \( T \in U(n - k) \) and \( \det(S) \det(T) = 1 \).

Closely related to Grassmannians are the \textit{Stiefel manifolds} \( S(k, n) \). Again we begin with the real case.

**Definition 17.19.** For any \( n \geq 1 \) and any \( k \) with \( 1 \leq k \leq n \), the set \( S(k, n) \) of all orthonormal \( k \)-frames, that is, of \( k \)-tuples of orthonormal vectors \((u_1, \ldots, u_k)\) with \( u_i \in \mathbb{R}^n \), is called a \textit{Stiefel manifold}.

Obviously, \( S(1, n) = S^{n-1} \) and \( S(n, n) = O(n) \), so assume \( k \leq n - 1 \). There is a natural action \( \cdot : SO(n) \times S(k, n) \rightarrow S(k, n) \) of \( SO(n) \) on \( S(k, n) \) given by

\[
R \cdot (u_1, \ldots, u_k) = (Ru_1, \ldots, Ru_k).
\]

This action is transitive, because if \((u_1, \ldots, u_k)\) and \((v_1, \ldots, v_k)\) are any two orthonormal \( k \)-frames, then they can be extended to orthonormal bases (for example, by Gram-Schmidt) \((u_1, \ldots, u_n)\) and \((v_1, \ldots, v_n)\) with the same orientation (since we can pick \( u_n \) and \( v_n \) so that our bases have the same orientation), and there is a unique orthogonal transformation \( R \in SO(n) \) such that \( Ru_i = v_i \) for \( i = 1, \ldots, n \).

In order to apply Proposition 17.14, we need to find the stabilizer of the orthonormal \( k \)-frame \((e_1, \ldots, e_k)\) consisting of the first canonical basis vectors of \( \mathbb{R}^n \). A matrix \( R \in SO(n) \) stabilizes \((e_1, \ldots, e_k)\) iff it is of the form

\[
R = \begin{pmatrix} I_k & 0 \\ 0 & S \end{pmatrix}
\]

where \( S \in SO(n - k) \). Therefore, for \( 1 \leq k \leq n - 1 \), we have

\[
SO(n)/SO(n - k) \cong S(k, n).
\]

This makes \( S(k, n) \) a smooth manifold of dimension

\[
\frac{n(n - 1)}{2} - \frac{(n-k)(n-k-1)}{2} = nk - \frac{k(k + 1)}{2} = k(n - k) + \frac{k(k - 1)}{2}.
\]

**Remark:** It should be noted that we can define another type of Stiefel manifolds, denoted by \( V(k, n) \), using linearly independent \( k \)-tuples \((u_1, \ldots, u_k)\) that do not necessarily form an orthonormal system. In this case, there is an action \( \cdot : GL(n, \mathbb{R}) \times V(k, n) \rightarrow V(k, n) \), and the stabilizer \( H \) of the first \( k \) canonical basis vectors \((e_1, \ldots, e_k)\) is a closed subgroup of \( GL(n, \mathbb{R}) \), but it doesn’t have a simple description (see Warner [83], Chapter 3). We get an isomorphism

\[
V(k, n) \cong GL(n, \mathbb{R})/H.
\]

The version of the Stiefel manifold \( S(k, n) \) using orthonormal frames is sometimes denoted by \( V^0(k, n) \) (Milnor and Stasheff [59] use the notation \( V^0_k(\mathbb{R}^n) \)). Beware that the notation is not standardized. Certain authors use \( V(k, n) \) for what we denote by \( S(k, n) \)!
Complex Stiefel manifolds are defined in a similar way by replacing $\mathbb{R}$ by $\mathbb{C}$ and $\text{SO}(n)$ by $\text{SU}(n)$. For $1 \leq k \leq n - 1$, the complex Stiefel manifold $S_\mathbb{C}(k, n)$ is isomorphic to the quotient

$$\text{SU}(n)/\text{SU}(n - k) \cong S_\mathbb{C}(k, n).$$

If $k = 1$, we have $S_\mathbb{C}(1, n) = S^{2n-1}$, and if $k = n$, we have $S_\mathbb{C}(n, n) = \text{U}(n)$.

The Grassmannians can also be viewed as quotient spaces of the Stiefel manifolds. Every orthonormal $k$-frame $(u_1, \ldots, u_k)$ can be represented by an $n \times k$ matrix $Y$ over the canonical basis of $\mathbb{R}^n$, and such a matrix $Y$ satisfies the equation

$$Y^\top Y = I.$$ 

We have a right action $\cdot : S(k, n) \times \text{O}(k) \to S(k, n)$ given by

$$Y \cdot R = YR,$$

for any $R \in \text{O}(k)$. This action is well defined since

$$(YR)^\top YR = R^\top Y^\top YR = I.$$ 

However, this action is not transitive (unless $k = 1$), but the orbit space $S(k, n)/\text{O}(k)$ is isomorphic to the Grassmannian $G(k, n)$, so we can write

$$G(k, n) \cong S(k, n)/\text{O}(k).$$

Similarly, the complex Grassmannian is isomorphic to the orbit space $S_\mathbb{C}(k, n)/\text{U}(k)$:

$$G_\mathbb{C}(k, n) \cong S_\mathbb{C}(k, n)/\text{U}(k).$$
Chapter 18
Hilbert Spaces

18.1 The Projection Lemma, Duality

If \( E \) is a complex vector space, recall that a map \( \langle -, - \rangle : E \times E \to \mathbb{C} \) is a hermitian form if it satisfies the following properties for all \( x, y, x_1, x_2, y_1, y_2 \in E \) and all \( \lambda \in \mathbb{C} \): it is sesquilinear, which means that

\[
\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle \\
\langle x, y_1 + y_2 \rangle = \langle x, y_1 \rangle + \langle x, y_2 \rangle \\
\langle \lambda x, y \rangle = \lambda \langle x, y \rangle \\
\langle x, \lambda y \rangle = \overline{\lambda} \langle x, y \rangle,
\]

and satisfies the hermitian property,

\[
\langle y, x \rangle = \overline{\langle x, y \rangle}.
\]

The hermitian property implies that \( \langle x, x \rangle \in \mathbb{R} \) for all \( x \in E \).

A hermitian form \( \langle -, - \rangle : E \times E \to \mathbb{C} \) is positive if

\[
\langle x, x \rangle \geq 0 \quad \text{for all } x \in E.
\]

A positive hermitian form satisfies the Cauchy-Schwarz inequality:

\[
|\langle x, y \rangle|^2 \leq \langle x, x \rangle \langle y, y \rangle, \quad \text{for all } x, y \in E.
\]

A positive hermitian form is positive definite if for all \( x \in E \),

\[
\langle x, x \rangle = 0 \quad \text{implies that } x = 0,
\]

or equivalently,

\[
\langle x, x \rangle > 0 \quad \text{for all } x \neq 0.
\]
A positive definite hermitian form on $E$ is often called a *hermitian inner product* on $E$, and $E$ is called a *hermitian space* (sometimes a *pre-Hilbert space*).

Given a Hermitian space $\langle E, \langle-,-\rangle \rangle$, the function $\|\|: E \rightarrow \mathbb{R}$ defined such that $\|u\| = \sqrt{\langle u, u \rangle}$, is a norm on $E$. Thus, $E$ is a normed vector space. If $E$ is also complete, then it is a very interesting space.

In a hermitian space $\langle E, \langle-,-\rangle \rangle$, the inner product $\langle-,-\rangle$ can be recovered from the norm $\|\|$ using the following *polarization identities*: In the complex case,

$$\langle x, y \rangle = \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2 + i \|x + iy\|^2 - i \|x - iy\|^2),$$

and in the real case,

$$\langle x, y \rangle = \frac{1}{2}(\|x + y\|^2 - \|x\|^2 - \|y\|^2).$$

Recall that completeness has to do with the convergence of Cauchy sequences. A normed vector space $\langle E, \|\| \rangle$ is automatically a metric space under the metric $d$ defined such that $d(u, v) = \|v - u\|$ (see Chapter 16 for the definition of a normed vector space and of a metric space, or Lang [52, 53], or Dixmier [25]). Given a metric space $E$ with metric $d$, a sequence $(a_n)_{n \geq 1}$ of elements $a_n \in E$ is a *Cauchy sequence* iff for every $\epsilon > 0$, there is some $N \geq 1$ such that

$$d(a_m, a_n) < \epsilon \quad \text{for all} \quad m, n \geq N.$$ 

We say that $E$ is *complete* iff every Cauchy sequence converges to a limit (which is unique, since a metric space is Hausdorff).

Every finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ is complete. For example, one can show by induction that given any basis $(e_1, \ldots, e_n)$ of $E$, the linear map $h: \mathbb{C}^n \rightarrow E$ defined such that

$$h((z_1, \ldots, z_n)) = z_1 e_1 + \cdots + z_n e_n$$

is a homeomorphism (using the *sup*-norm on $\mathbb{C}^n$). One can also use the fact that any two norms on a finite dimensional vector space over $\mathbb{R}$ or $\mathbb{C}$ are equivalent (see Theorem 16.3, or Lang [53], Dixmier [25], Schwartz [70]).

However, if $E$ has infinite dimension, it may not be complete. When a Hermitian space is complete, a number of the properties that hold for finite dimensional Hermitian spaces also hold for infinite dimensional spaces. For example, any closed subspace has an orthogonal complement, and in particular, a finite dimensional subspace has an orthogonal complement. Hermitian spaces that are also complete play an important role in analysis. Since they were first studied by Hilbert, they are called Hilbert spaces.

**Definition 18.1.** A (complex) Hermitian space $\langle E, \langle-,-\rangle \rangle$ which is a complete normed vector space under the norm $\|\|$ induced by $\langle-,-\rangle$ is called a *Hilbert space*. A real Euclidean space $\langle E, \langle-,-\rangle \rangle$ which is complete under the norm $\|\|$ induced by $\langle-,-\rangle$ is called a *real Hilbert space*.
18.1. THE PROJECTION LEMA, DUALITY

All the results in this section hold for complex Hilbert spaces as well as for real Hilbert spaces. We state all results for the complex case only, since they also apply to the real case, and since the proofs in the complex case need a little more care.

Example 18.1. The space $l^2$ of all countably infinite sequences $x = (x_i)_{i \in \mathbb{N}}$ of complex numbers such that $\sum_{i=0}^{\infty} |x_i|^2 < \infty$ is a Hilbert space. It will be shown later that the map $\langle -, - \rangle : l^2 \times l^2 \rightarrow \mathbb{C}$ defined such that

$$\langle (x_i)_{i \in \mathbb{N}}, (y_i)_{i \in \mathbb{N}} \rangle = \sum_{i=0}^{\infty} x_i \overline{y_i}$$

is well defined, and that $l^2$ is a Hilbert space under $\langle -, - \rangle$. In fact, we will prove a more general result (Proposition 18.13).

Example 18.2. The set $C^\infty[a,b]$ of smooth functions $f : [a,b] \rightarrow \mathbb{C}$ is a Hermitian space under the Hermitian form

$$\langle f, g \rangle = \int_a^b f(x) \overline{g(x)} dx,$$

but it is not a Hilbert space because it is not complete. It is possible to construct its completion $L^2([a,b])$, which turns out to be the space of Lebesgue square-integrable functions on $[a,b]$.

A simple adaptation of the completion theorem for normed vector spaces (Theorem 15.68) shows that every hermitian space has a Hilbert space completion.

Theorem 18.1. If $(E, \langle -, - \rangle)$ is a hermitian space, then the Banach space $(\hat{E}, \| \|_E)$, completion of the normed vector space $(E, \| \|)$ where $\|x\| = \sqrt{\langle x, x \rangle}$ for all $x \in E$, is Hilbert space with the inner product $\langle -, - \rangle_h$ given by

$$\langle x, y \rangle_h = \frac{1}{4}(\|x + y\|^2_E - \|x - y\|^2_E + i \|x + iy\|^2_E - i \|x - iy\|^2_E),$$

and in the real case,

$$\langle x, y \rangle_h = \frac{1}{2}(\|x + y\|^2_E - \|x\|^2_E - \|y\|^2_E)$$

for all $x, y \in \hat{E}$. Furthermore, the linear map $\varphi : E \rightarrow \hat{E}$ given by Theorem 15.68 is inner-product preserving.

Proof. Since $E$ is dense in $E$, $E \times E$ is dense $\hat{E} \times \hat{E}$, and since the map

$$(x, y) \mapsto \frac{1}{4}(\|x + y\|^2 - \|x - y\|^2 + i \|x + iy\|^2 - i \|x - iy\|^2),$$

and in the real case,

$$(x, y) = \frac{1}{2}(\|x + y\|^2 - \|x\|^2 - \|y\|^2),$$
is uniformly continuous, by Theorem 15.57, these maps have unique continuous extensions, and if we define

$$\langle x, y \rangle_h = \frac{1}{4}(\|x + y\|_E^2 - \|x - y\|_E^2 + i \|x + iy\|_E^2 - i \|x - iy\|_E^2),$$

and in the real case,

$$\langle x, y \rangle_h \mapsto \frac{1}{2}(\|x + y\|_E^2 - \|x\|_E^2 - \|y\|_E^2),$$

for all $x, y \in \widehat{E}$, it is easy to check that we obtain positive definite hermitian forms with associated norm $\| \cdot \|_E$, so $\widehat{E}$ a Hilbert space with this inner product. For another proof, see Bourbaki [10].

One of the most important facts about finite-dimensional Hermitian (and Euclidean) spaces is that they have orthonormal bases. This implies that, up to isomorphism, every finite-dimensional Hermitian space is isomorphic to $\mathbb{C}^n$ (for some $n \in \mathbb{N}$) and that the inner product is given by

$$\langle (x_1, \ldots, x_n), (y_1, \ldots, y_n) \rangle = \sum_{i=1}^{n} x_i \overline{y_i}.$$  

Furthermore, every subspace $W$ has an orthogonal complement $W^\perp$, and the inner product induces a natural duality between $E$ and $E^*$, where $E^*$ is the space of linear forms on $E$.

When $E$ is a Hilbert space, $E$ may be infinite dimensional, often of uncountable dimension. Thus, we can’t expect that $E$ always have an orthonormal basis. However, if we modify the notion of basis so that a “Hilbert basis” is an orthogonal family that is also dense in $E$, i.e., every $v \in E$ is the limit of a sequence of finite combinations of vectors from the Hilbert basis, then we can recover most of the “nice” properties of finite-dimensional Hermitian spaces. For instance, if $(u_k)_{k \in K}$ is a Hilbert basis, for every $v \in E$, we can define the Fourier coefficients $c_k = \langle v, u_k \rangle / \|u_k\|$, and then, $v$ is the “sum” of its Fourier series $\sum_{k \in K} c_k u_k$. However, the cardinality of the index set $K$ can be very large, and it is necessary to define what it means for a family of vectors indexed by $K$ to be summable. We will do this in Section 18.2. It turns out that every Hilbert space is isomorphic to a space of the form $l^2(K)$, where $l^2(K)$ is a generalization of the space of Example 18.1 (see Theorem 18.18, usually called the Riesz-Fischer theorem).

Our first goal is to prove that a closed subspace of a Hilbert space has an orthogonal complement. We also show that duality holds if we redefine the dual $E'$ of $E$ to be the space of continuous linear maps on $E$. Our presentation closely follows Bourbaki [10]. We also were inspired by Rudin [66], Lang [52, 53], Schwartz [70, 69], and Dixmier [25]. In fact, we highly recommend Dixmier [25] as a clear and simple text on the basics of topology and analysis. We first prove the so-called projection lemma.

Recall that in a metric space $E$, a subset $X$ of $E$ is closed iff for every convergent sequence $(x_n)$ of points $x_n \in X$, the limit $x = \lim_{n \to \infty} x_n$ also belongs to $X$. The closure $\overline{X}$ of $X$ is
the set of all limits of convergent sequences \((x_n)\) of points \(x_n \in X\). Obviously, \(X \subseteq X\). We say that the subset \(X\) of \(E\) is dense in \(E\) if \(E = \overline{X}\), the closure of \(X\), which means that every \(a \in E\) is the limit of some sequence \((x_n)\) of points \(x_n \in X\). Convex sets will again play a crucial role.

First, we state the following easy “parallelogram inequality”, whose proof is left as an exercise.

**Proposition 18.2.** If \(E\) is a Hermitian space, for any two vectors \(u, v \in E\), we have
\[
\|u + v\|^2 + \|u - v\|^2 = 2(\|u\|^2 + \|v\|^2).
\]

From the above, we get the following proposition:

**Proposition 18.3.** If \(E\) is a Hermitian space, given any \(d, \delta \in \mathbb{R}\) such that \(0 \leq \delta < d\), let
\[
B = \{u \in E \mid \|u\| < d\} \quad \text{and} \quad C = \{u \in E \mid \|u\| \leq d + \delta\}.
\]

For any convex set such \(A\) that \(A \subseteq C - B\), we have
\[
\|v - u\| \leq \sqrt{12d\delta},
\]
for all \(u, v \in A\) (see Figure 18.1).

\begin{figure}[h]
\centering
\includegraphics[width=0.4\textwidth]{figure18.1.png}
\caption{Inequality of Proposition 18.3}
\end{figure}

**Proof.** Since \(A\) is convex, \(\frac{1}{2}(u + v) \in A\) if \(u, v \in A\), and thus, \(\|\frac{1}{2}(u + v)\| \geq d\). From the parallelogram inequality written in the form
\[
\|\frac{1}{2}(u + v)\|^2 + \|\frac{1}{2}(u - v)\|^2 = \frac{1}{2}(\|u\|^2 + \|v\|^2),
\]
since \(\delta < d\), we get
\[
\left\|\frac{1}{2}(u - v)\right\|^2 = \frac{1}{2}(\|u\|^2 + \|v\|^2) - \left\|\frac{1}{2}(u + v)\right\|^2 \leq (d + \delta)^2 - d^2 = 2d\delta + \delta^2 \leq 3d\delta,
\]
from which
\[
\|v - u\| \leq \sqrt{12d\delta}.
\]
\[\square\]
If $X$ is a nonempty subset of a metric space $(E, d)$, for any $a \in E$, recall that we define the distance $d(a, X)$ of $a$ to $X$ as
\[ d(a, X) = \inf_{b \in X} d(a, b). \]

Also, the diameter $\delta(X)$ of $X$ is defined by
\[ \delta(X) = \sup \{ d(a, b) \mid a, b \in X \}. \]

It is possible that $\delta(X) = \infty$. We leave the following standard two facts as an exercise (see Dixmier [25]):

**Proposition 18.4.** Let $E$ be a metric space.

1. For every subset $X \subseteq E$, $\delta(X) = \delta(\overline{X})$.

2. If $E$ is a complete metric space, for every sequence $(F_n)$ of closed nonempty subsets of $E$ such that $F_{n+1} \subseteq F_n$, if $\lim_{n \to \infty} \delta(F_n) = 0$, then $\bigcap_{n=1}^{\infty} F_n$ consists of a single point.

We are now ready to prove the crucial projection lemma.

**Proposition 18.5.** (Projection lemma) Let $E$ be a Hilbert space.

1. For any nonempty convex and closed subset $X \subseteq E$, for any $u \in E$, there is a unique vector $p_X(u) \in X$ such that
\[ \|u - p_X(u)\| = \inf_{v \in X} \|u - v\| = d(u, X). \]

2. The vector $p_X(u)$ is the unique vector $w \in E$ satisfying the following property (see Figure 18.2):
\[ w \in X \quad \text{and} \quad \Re \langle u - w, z - w \rangle \leq 0 \quad \text{for all} \ z \in X. \quad (*) \]

![Figure 18.2: Inequality of Proposition 18.5](image-url)
18.1. THE PROJECTION LEMMA, DUALITY

Proof. (1) Let $d = \inf_{v \in X} \|u - v\| = d(u, X)$. We define a sequence $X_n$ of subsets of $X$ as follows: for every $n \geq 1$,

$$X_n = \left\{ v \in X \mid \|u - v\| \leq d \frac{1}{n} \right\}.$$ 

It is immediately verified that each $X_n$ is nonempty (by definition of $d$), convex, and that $X_{n+1} \subseteq X_n$. Also, by Proposition 18.3, we have

$$\sup \{ \|w - v\| \mid v, w \in X_n \} \leq \sqrt{12d/n},$$ 

and thus, $\bigcap_{n \geq 1} X_n$ contains at most one point. We will prove that $\bigcap_{n \geq 1} X_n$ contains exactly one point, namely, $p_X(u)$. For this, define a sequence $(w_n)_{n \geq 1}$ by picking some $w_n \in X_n$ for every $n \geq 1$. We claim that $(w_n)_{n \geq 1}$ is a Cauchy sequence. Given any $\epsilon > 0$, if we pick $N$ such that

$$N > \frac{12d}{\epsilon^2},$$

since $(X_n)_{n \geq 1}$ is a monotonic decreasing sequence, for all $m, n \geq N$, we have

$$\|w_m - w_n\| \leq \sqrt{12d/N} < \epsilon,$$

as desired. Since $E$ is complete, the sequence $(w_n)_{n \geq 1}$ has a limit $w$, and since $w_n \in X$ and $X$ is closed, we must have $w \in X$. Also observe that

$$\|u - w\| \leq \|u - w_n\| + \|w_n - w\|,$$

and since $w$ is the limit of $(w_n)_{n \geq 1}$ and

$$\|u - w_n\| \leq d + \frac{1}{n},$$

given any $\epsilon > 0$, there is some $n$ large enough so that

$$\frac{1}{n} < \frac{\epsilon}{2} \quad \text{and} \quad \|w_n - w\| \leq \frac{\epsilon}{2},$$

and thus

$$\|u - w\| \leq d + \epsilon.$$

Since the above holds for every $\epsilon > 0$, we have $\|u - w\| = d$. Thus, $w \in X_n$ for all $n \geq 1$, which proves that $\bigcap_{n \geq 1} X_n = \{w\}$. Now, any $z \in X$ such that $\|u - z\| = d(u, X) = d$ also belongs to every $X_n$, and thus $z = w$, proving the uniqueness of $w$, which we denote as $p_X(u)$.

(2) Let $w \in X$. Since $X$ is convex, $z = (1 - \lambda)p_X(u) + \lambda w \in X$ for every $\lambda, 0 \leq \lambda \leq 1$. Then, we have

$$\|u - z\| \geq \|u - p_X(u)\|$$
for all \( \lambda, 0 \leq \lambda \leq 1 \), and since
\[
\|u - z\|^2 = \|u - p_X(u) - \lambda(w - p_X(u))\|^2
= \|u - p_X(u)\|^2 + \lambda^2\|w - p_X(u)\|^2 - 2\lambda\Re\langle u - p_X(u), w - p_X(u) \rangle,
\]
for all \( \lambda, 0 < \lambda \leq 1 \), we get
\[
\Re\langle u - p_X(u), w - p_X(u) \rangle = \frac{1}{2\lambda} \left( \|u - p_X(u)\|^2 - \|u - z\|^2 \right) + \frac{\lambda}{2}\|w - p_X(u)\|^2,
\]
and since this holds for every \( \lambda, 0 < \lambda \leq 1 \) and
\[
\|u - z\| \geq \|u - p_X(u)\|,
\]
we have
\[
\Re\langle u - p_X(u), w - p_X(u) \rangle \leq 0.
\]

Conversely, assume that \( w \in X \) satisfies the condition
\[
\Re\langle u - w, z - w \rangle \leq 0
\]
for all \( z \in X \). For all \( z \in X \), we have
\[
\|u - z\|^2 = \|u - w\|^2 + \|z - w\|^2 - 2\Re\langle u - w, z - w \rangle \geq \|u - w\|^2,
\]
which implies that \( \|u - w\| = d(u, X) = d \), and from (1), that \( w = p_X(u) \).

The vector \( p_X(u) \) is called the projection of \( u \) onto \( X \), and the map \( p_X : E \to X \) is called the projection of \( E \) onto \( X \). In the case of a real Hilbert space, there is an intuitive geometric interpretation of the condition
\[
\langle u - p_X(u), z - p_X(u) \rangle \leq 0
\]
for all \( z \in X \). If we restate the condition as
\[
\langle u - p_X(u), p_X(u) - z \rangle \geq 0
\]
for all \( z \in X \), this says that the absolute value of the measure of the angle between the vectors \( u - p_X(u) \) and \( p_X(u) - z \) is at most \( \pi/2 \). This makes sense, since \( X \) is convex, and points in \( X \) must be on the side opposite to the “tangent space” to \( X \) at \( p_X(u) \), which is orthogonal to \( u - p_X(u) \). Of course, this is only an intuitive description, since the notion of tangent space has not been defined!

The map \( p_X : E \to X \) is continuous, as shown below.

**Proposition 18.6.** Let \( E \) be a Hilbert space. For any nonempty convex and closed subset \( X \subseteq E \), the map \( p_X : E \to X \) is continuous.
Proposition 18.7. 18.1. THE PROJECTION LEMMA, DUALITY

Proof. For any two vectors \( u, v \in E \), let \( x = p_X(u) - u, y = p_X(v) - p_X(u) \), and \( z = v - p_X(v) \). Clearly,

\[
v - u = x + y + z,
\]

and from Proposition 18.5 (2), we also have

\[
\Re \langle x, y \rangle \geq 0 \quad \text{and} \quad \Re \langle z, y \rangle \geq 0,
\]

from which we get

\[
\|v - u\|^2 = \|x + y + z\|^2 = \|x + z + y\|^2 \\
= \|x + z\|^2 + \|y\|^2 + 2\Re \langle x, y \rangle + 2\Re \langle z, y \rangle \\
\geq \|y\|^2 = \|p_X(v) - p_X(u)\|^2.
\]

However, \( \|p_X(v) - p_X(u)\| \leq \|v - u\| \) obviously implies that \( p_X \) is continuous. \( \Box \)

We can now prove the following important proposition.

**Proposition 18.7.** Let \( E \) be a Hilbert space.

1. For any closed subspace \( V \subseteq E \), we have \( E = V \oplus V^\perp \), and the map \( p_V : E \to V \) is linear and continuous.

2. For any \( u \in E \), the projection \( p_V(u) \) is the unique vector \( w \in E \) such that

\[
w \in V \quad \text{and} \quad \langle u - w, z \rangle = 0 \quad \text{for all} \quad z \in V.
\]

Proof. (1) First, we prove that \( u - p_V(u) \in V^\perp \) for all \( u \in E \). For any \( v \in V \), since \( V \) is a subspace, \( z = p_V(u) + \lambda v \in V \) for all \( \lambda \in \mathbb{C} \), and since \( V \) is convex and nonempty (since it is a subspace), and closed by hypothesis, by Proposition 18.5 (2), we have

\[
\Re(\overline{\langle u - p_V(u), v \rangle}) = \Re(\langle u - p_V(u), \lambda v \rangle) = \Re \langle u - p_V(u), z - p_V(u) \rangle \leq 0
\]

for all \( \lambda \in \mathbb{C} \). In particular, the above holds for \( \lambda = \langle u - p_V(u), v \rangle \), which yields

\[
| \langle u - p_V(u), v \rangle | \leq 0,
\]

and thus, \( \langle u - p_V(u), v \rangle = 0 \). As a consequence, \( u - p_V(u) \in V^\perp \) for all \( u \in E \). Since \( u = p_V(u) + u - p_V(u) \) for every \( u \in E \), we have \( E = V + V^\perp \). On the other hand, since \( \langle -, - \rangle \) is positive definite, \( V \cap V^\perp = \{0\} \), and thus \( E = V \oplus V^\perp \).

We already proved in Proposition 18.6 that \( p_V : E \to V \) is continuous. Also, since

\[
p_V(\lambda u + \mu v) - (\lambda p_V(u) + \mu p_V(v)) = p_V(\lambda u + \mu v) - (\lambda u + \mu v) + \lambda(u - p_V(u)) + \mu(v - p_V(v)),
\]

[(Proof continued on the next page...)]
for all \( u, v \in E \), and since the left-hand side term belongs to \( V \), and from what we just showed, the right-hand side term belongs to \( V^\perp \), we have

\[
p_V(\lambda u + \mu v) - (\lambda p_V(u) + \mu p_V(v)) = 0,
\]

showing that \( p_V \) is linear.

(2) This is basically obvious from (1). We proved in (1) that \( u - p_V(u) \in V^\perp \), which is exactly the condition

\[
\langle u - p_V(u), z \rangle = 0
\]

for all \( z \in V \). Conversely, if \( w \in V \) satisfies the condition

\[
\langle u - w, z \rangle = 0
\]

for all \( z \in V \), since \( w \in V \), every vector \( z \in V \) is of the form \( y - w \), with \( y = z + w \in V \), and thus, we have

\[
\langle u - w, y - w \rangle = 0
\]

for all \( y \in V \), which implies the condition of Proposition 18.5 (2):

\[
\Re \langle u - w, y - w \rangle \leq 0
\]

for all \( y \in V \). By Proposition 18.5, \( w = p_V(u) \) is the projection of \( u \) onto \( V \).

Let us illustrate the power of Proposition 18.7 on the following “least squares” problem. Given a real \( m \times n \)-matrix \( A \) and some vector \( b \in \mathbb{R}^m \), we would like to solve the linear system

\[
Ax = b
\]

in the least-squares sense, which means that we would like to find some solution \( x \in \mathbb{R}^n \) that minimizes the Euclidean norm \( \|Ax - b\| \) of the error \( Ax - b \). It is actually not clear that the problem has a solution, but it does! The problem can be restated as follows: Is there some \( x \in \mathbb{R}^n \) such that

\[
\|Ax - b\| = \inf_{y \in \mathbb{R}^n} \|Ay - b\|,
\]

or equivalently, is there some \( z \in \text{Im}(A) \) such that

\[
\|z - b\| = d(b, \text{Im}(A)),
\]

where \( \text{Im}(A) = \{Ay \in \mathbb{R}^m \mid y \in \mathbb{R}^n\} \), the image of the linear map induced by \( A \). Since \( \text{Im}(A) \) is a closed subspace of \( \mathbb{R}^m \), because we are in finite dimension, Proposition 18.7 tells us that there is a unique \( z \in \text{Im}(A) \) such that

\[
\|z - b\| = \inf_{y \in \mathbb{R}^n} \|Ay - b\|,
\]

and thus, the problem always has a solution since \( z \in \text{Im}(A) \), and since there is at least some \( x \in \mathbb{R}^n \) such that \( Ax = z \) (by definition of \( \text{Im}(A) \)). Note that such an \( x \) is not necessarily
unique. Furthermore, Proposition 18.7 also tells us that \( z \in \text{Im}(A) \) is the solution of the equation
\[
\langle z - b, w \rangle = 0 \quad \text{for all } w \in \text{Im}(A),
\]
or equivalently, that \( x \in \mathbb{R}^n \) is the solution of
\[
\langle Ax - b, Ay \rangle = 0 \quad \text{for all } y \in \mathbb{R}^n,
\]
which is equivalent to
\[
\langle A^\top (Ax - b), y \rangle = 0 \quad \text{for all } y \in \mathbb{R}^n,
\]
and thus, since the inner product is positive definite, to \( A^\top (Ax - b) = 0 \), i.e.,
\[
A^\top Ax = A^\top b.
\]

Therefore, the solutions of the original least-squares problem are precisely the solutions of the so-called normal equations
\[
A^\top Ax = A^\top b,
\]
discovered by Gauss and Legendre around 1800. We also proved that the normal equations always have a solution.

Computationally, it is best not to solve the normal equations directly, and instead, to use methods such as the \( QR \)-decomposition (applied to \( A \)) or the SVD-decomposition (in the form of the pseudo-inverse). We will come back to this point later on.

As another corollary of Proposition 18.7, for any continuous nonnull linear map \( h : E \to \mathbb{C} \), the null space
\[
H = \text{Ker } h = \{ u \in E \mid h(u) = 0 \} = h^{-1}(0)
\]
is a closed hyperplane \( H \), and thus, \( H^\perp \) is a subspace of dimension one such that \( E = H \oplus H^\perp \). This suggests defining the dual space of \( E \) as the set of all continuous maps \( h : E \to \mathbb{C} \).

**Remark:** If \( h : E \to \mathbb{C} \) is a linear map which is not continuous, then it can be shown that the hyperplane \( H = \text{Ker } h \) is dense in \( E \)! Thus, \( H^\perp \) is reduced to the trivial subspace \( \{0\} \). This goes against our intuition of what a hyperplane in \( \mathbb{R}^n \) (or \( \mathbb{C}^n \)) is, and warns us not to trust our “physical” intuition too much when dealing with infinite dimensions.

**Definition 18.2.** Given two vector spaces \( E \) and \( F \) over the complex field \( \mathbb{C} \), a function \( f : E \to F \) is *semilinear* if
\[
\begin{align*}
f(u + v) &= f(u) + f(v), \\
f(\lambda u) &= \overline{\lambda} f(u),
\end{align*}
\]
for all \( u, v \in E \) and all \( \lambda \in \mathbb{C} \).
Instead of defining semilinear maps, we can define the vector space $\mathcal{E}$ as the vector space with the same carrier set $E$ whose addition is the same as that of $E$, but whose multiplication by a complex number $\lambda$ is given by

$$(\lambda, u) \mapsto \lambda u.$$ 

Then it is easy to check that a function $f : E \to \mathbb{C}$ is semilinear iff $f : \mathcal{E} \to \mathbb{C}$ is linear.

A fundamental fact about a finite-dimensional hermitian space is that the hermitian inner product induces a bijection (i.e., independent of the choice of bases) between the vector space $E$ and its dual space $E^*$.

Given a Hermitian space $E$, for any vector $u \in E$, let $\varphi^l_u : E \to \mathbb{C}$ be the map defined such that

$$\varphi^l_u(v) = \langle u, v \rangle,$$

for all $v \in E$.

Similarly, for any vector $v \in E$, let $\varphi^r_v : E \to \mathbb{C}$ be the map defined such that

$$\varphi^r_v(u) = \langle u, v \rangle,$$

for all $u \in E$.

Since the Hermitian product is linear in its first argument $u$, the map $\varphi^r_v$ is a linear form in $E^*$, and since it is semilinear in its second argument $v$, the map $\varphi^l_u$ is also a linear form in $E^*$. Thus, we have two maps $\flat^l : E \to E^*$ and $\flat^r : E \to E^*$, defined such that

$$\flat^l(u) = \varphi^l_u, \quad \text{and} \quad \flat^r(v) = \varphi^r_v.$$ 

Actually, $\varphi^l_u = \varphi^r_u$ and $\flat^l = \flat^r$. Indeed, for all $u, v \in E$, we have

$$\flat^l(u)(v) = \varphi^l_u(v)$$

$$= \langle u, v \rangle$$

$$= \langle v, u \rangle$$

$$= \varphi^r_u(v)$$

$$= \flat^r(u)(v).$$

Therefore, we use the notation $\varphi_u$ for both $\varphi^l_u$ and $\varphi^r_u$, and $\flat$ for both $\flat^l$ and $\flat^r$.

**Theorem 18.8.** Let $E$ be a Hermitian space $E$. The map $\flat : E \to E^*$ defined such that

$$\flat(u) = \varphi_u \quad \text{for all} \quad u \in E$$

is semilinear and injective. When $E$ is also of finite dimension, the map $\flat : \mathcal{E} \to E^*$ is a canonical isomorphism.

**Proof.** That $\flat : E \to E^*$ is a semilinear map follows immediately from the fact that $\flat = \flat^r$, and that the Hermitian product is semilinear in its second argument. If $\varphi_u = \varphi_v$, then $\varphi_u(w) = \varphi_v(w)$ for all $w \in E$, which by definition of $\varphi_u$ and $\varphi_v$ means that

$$\langle w, u \rangle = \langle w, v \rangle$$
for all \( w \in E \), which by semilinearity on the right is equivalent to
\[
\langle w, v - u \rangle = 0 \quad \text{for all } w \in E,
\]
which implies that \( u = v \), since the Hermitian product is positive definite. Thus, \( \flat: E \to E^* \) is injective. Finally, when \( E \) is of finite dimension \( n \), \( E^* \) is also of dimension \( n \), and then \( \flat: E \to E^* \) is bijective. Since \( \flat \) is semilinear, the map \( \flat: E \to E^* \) is an isomorphism. \( \square \)

However, if \( E \) is infinite dimensional, the map \( \flat: E \to E^* \) is not surjective, since the linear forms of the form \( u \mapsto \langle u, v \rangle \) (for some fixed vector \( v \in E \)) are continuous (the inner product is continuous), but there are linear forms that are not continuous.

We now show that by redefining the dual space of a Hilbert space as the set of continuous linear forms on \( E \), we recover Theorem 18.8.

**Definition 18.3.** Given a Hilbert space \( E \), we define the dual space \( E' \) of \( E \) as the vector space of all continuous linear forms \( h: E \to \mathbb{C} \). Maps in \( E' \) are also called bounded linear operators, bounded linear functionals, or simply, operators or functionals.

Theorem 18.8 is generalized to Hilbert spaces as follows.

**Theorem 18.9.** *(Riesz representation theorem)* Let \( E \) be a Hilbert space. Then, the map \( \flat: E \to E' \) defined such that
\[
\flat(v) = \varphi_v,
\]
is semilinear, continuous, and bijective.

**Proof.** The proof is basically identical to the proof of Theorem 18.8, except that a different argument is required for the surjectivity of \( \flat: E \to E' \), since \( E \) may not be finite dimensional. For any nonnull linear operator \( h \in E' \), the hyperplane \( H = \ker h = h^{-1}(0) \) is a closed subspace of \( E \), and by Proposition 18.7, \( H^\perp \) is a subspace of dimension one such that \( E = H \oplus H^\perp \). Then, picking any nonnull vector \( w \in H^\perp \), observe that \( H \) is also the kernel of the linear operator \( \varphi_w \), with
\[
\varphi_w(u) = \langle u, w \rangle,
\]
and thus, since any two nonzero linear forms defining the same hyperplane must be proportional, there is some nonzero scalar \( \lambda \in \mathbb{C} \) such that \( h = \lambda \varphi_w \). But then, \( h = \varphi_\lambda w \), proving that \( \flat: E \to E' \) is surjective. \( \square \)

Theorem 18.9 is known as the Riesz representation theorem, or “Little Riesz Theorem.” It shows that the inner product on a Hilbert space induces a natural linear isomorphism between \( E \) and its dual \( E' \).

**Remarks:**

(1) Actually, the map \( \flat: E \to E' \) turns out to be an isometry. To show this, we need to recall the notion of norm of a linear map, which we do not want to do right now.
(2) Many books on quantum mechanics use the so-called Dirac notation to denote objects in the Hilbert space $E$ and operators in its dual space $E'$. In the Dirac notation, an element of $E$ is denoted as $|x\rangle$, and an element of $E'$ is denoted as $\langle t|$. The scalar product is denoted as $\langle t| \cdot |x\rangle$. This uses the isomorphism between $E$ and $E'$, except that the inner product is assumed to be semi-linear on the left, rather than on the right.

Theorem 18.9 allows us to define the adjoint of a linear map, as in the Hermitian case.

**Proposition 18.10.** Given a Hilbert space $E$, for every linear map $f : E \to E$, there is a unique linear map $f^* : E \to E$, such that

$$\langle f^*(u), v \rangle = \langle u, f(v) \rangle$$

for all $u, v \in E$. The map $f^*$ is called the adjoint of $f$.

It is easy to show that if $f$ is continuous, then $f^*$ is also continuous. As in the Hermitian case, given two Hilbert spaces $E$ and $F$, for any linear map $f : E \to F$, such that

$$\langle f(u), v \rangle_2 = \langle u, f^*(v) \rangle_1$$

for all $u \in E$ and all $v \in F$. The linear map $f^*$ is also called the adjoint of $f$.

### 18.2 Total Orthogonal Families (Hilbert Bases), Fourier Coefficients

We conclude our quick tour of Hilbert spaces by showing that the notion of orthogonal basis can be generalized to Hilbert spaces. However, the useful notion is not the usual notion of a basis, but a notion which is an abstraction of the concept of Fourier series. Every element of a Hilbert space is the “sum” of its Fourier series.

**Definition 18.4.** Given a Hilbert space $E$, a family $(u_k)_{k \in K}$ of nonnull vectors is an orthogonal family iff the $u_k$ are pairwise orthogonal, i.e., $\langle u_i, u_j \rangle = 0$ for all $i \neq j$ ($i, j \in K$), and an orthonormal family iff $\langle u_i, u_j \rangle = \delta_{i,j}$, for all $i, j \in K$. A total orthogonal family (or system) or Hilbert basis is an orthogonal family that is dense in $E$. This means that for every $v \in E$, for every $\epsilon > 0$, there is some finite subset $I \subseteq K$ and some family $(\lambda_i)_{i \in I}$ of complex numbers, such that

$$\|v - \sum_{i \in I} \lambda_i u_i\| < \epsilon.$$

Given an orthogonal family $(u_k)_{k \in K}$, for every $v \in E$, for every $k \in K$, the scalar $c_k = \langle v, u_k \rangle / \|u_k\|^2$ is called the $k$-th Fourier coefficient of $v$ over $(u_k)_{k \in K}$.
Remark: The terminology Hilbert basis is misleading, because a Hilbert basis \((u_k)_{k \in K}\) is not necessarily a basis in the algebraic sense. Indeed, in general, \((u_k)_{k \in K}\) does not span \(E\). Intuitively, it takes linear combinations of the \(u_k\)'s with infinitely many nonnull coefficients to span \(E\). Technically, this is achieved in terms of limits. In order to avoid the confusion between bases in the algebraic sense and Hilbert bases, some authors refer to algebraic bases as *Hamel bases* and to total orthogonal families (or Hilbert bases) as *Schauder bases*.

Given an orthogonal family \((u_k)_{k \in K}\), for any finite subset \(I\) of \(K\), we often call sums of the form \(\sum_{i \in I} \lambda_i u_i\) *partial sums of Fourier series*, and if these partial sums converge to a limit denoted as \(\sum_{k \in K} c_k u_k\), we call \(\sum_{k \in K} c_k u_k\) a *Fourier series*.

However, we have to make sense of such sums! Indeed, when \(K\) is unordered or uncountable, the notion of limit or sum has not been defined. This can be done as follows (for more details, see Dixmier [25]):

**Definition 18.5.** Given a normed vector space \(E\) (say, a Hilbert space), for any nonempty index set \(K\), we say that a family \((u_k)_{k \in K}\) of vectors in \(E\) is *summable with sum* \(v \in E\) iff for every \(\epsilon > 0\), there is some finite subset \(I\) of \(K\), such that,

\[
\left\| v - \sum_{j \in J} u_j \right\| < \epsilon
\]

for every finite subset \(J\) with \(I \subseteq J \subseteq K\). We say that the family \((u_k)_{k \in K}\) is *summable* iff there is some \(v \in E\) such that \((u_k)_{k \in K}\) is summable with sum \(v\). A family \((u_k)_{k \in K}\) is a *Cauchy family* iff for every \(\epsilon > 0\), there is a finite subset \(I\) of \(K\), such that,

\[
\left\| \sum_{j \in J} u_j \right\| < \epsilon
\]

for every finite subset \(J\) of \(K\) with \(I \cap J = \emptyset\),

If \((u_k)_{k \in K}\) is summable with sum \(v\), we usually denote \(v\) as \(\sum_{k \in K} u_k\). The following technical proposition will be needed:

**Proposition 18.11.** Let \(E\) be a complete normed vector space (say, a Hilbert space).

1. For any nonempty index set \(K\), a family \((u_k)_{k \in K}\) is summable iff it is a Cauchy family.

2. Given a family \((r_k)_{k \in K}\) of nonnegative reals \(r_k \geq 0\), if there is some real number \(B > 0\) such that \(\sum_{i \in I} r_i < B\) for every finite subset \(I\) of \(K\), then \((r_k)_{k \in K}\) is summable and \(\sum_{k \in K} r_k = r\), where \(r\) is least upper bound of the set of finite sums \(\sum_{i \in I} r_i\) (\(I \subseteq K\)).

**Proof.** (1) If \((u_k)_{k \in K}\) is summable, for every finite subset \(I\) of \(K\), let

\[
u_I = \sum_{i \in I} u_i \quad \text{and} \quad u = \sum_{k \in K} u_k
\]

then \(u_I = \sum_{i \in I} u_i\).
For every $\epsilon > 0$, there is some finite subset $I$ of $K$ such that

$$\|u - u_L\| < \epsilon/2$$

for all finite subsets $L$ such that $I \subseteq L \subseteq K$. For every finite subset $J$ of $K$ such that $I \cap J = \emptyset$, since $I \subseteq I \cup J \subseteq K$ and $I \cup J$ is finite, we have

$$\|u - u_{I \cup J}\| < \epsilon/2 \quad \text{and} \quad \|u - u_I\| < \epsilon/2,$$

and since

$$\|u_{I \cup J} - u_I\| \leq \|u_{I \cup J} - u\| + \|u - u_I\|$$

and $u_{I \cup J} - u_I = u_J$ since $I \cap J = \emptyset$, we get

$$\|u_J\| = \|u_{I \cup J} - u_I\| < \epsilon,$$

which is the condition for $(u_k)_{k \in K}$ to be a Cauchy family.

Conversely, assume that $(u_k)_{k \in K}$ is a Cauchy family. We define inductively a decreasing sequence $(X_n)$ of subsets of $E$, each of diameter at most $1/n$, as follows: For $n = 1$, since $(u_k)_{k \in K}$ is a Cauchy family, there is some finite subset $J_1$ of $K$ such that

$$\|u\| < 1/2$$

for every finite subset $J$ of $K$ with $J_1 \cap J = \emptyset$. We pick some finite subset $J_1$ with the above property, and we let $I_1 = J_1$ and

$$X_1 = \{u_I \mid I_1 \subseteq I \subseteq K, \ I \text{ finite}\}.$$

For $n \geq 1$, there is some finite subset $J_{n+1}$ of $K$ such that

$$\|u_{J_{n+1}}\| < 1/(2n + 2)$$

for every finite subset $J$ of $K$ with $J_{n+1} \cap J = \emptyset$. We pick some finite subset $J_{n+1}$ with the above property, and we let $I_{n+1} = I_n \cup J_{n+1}$ and

$$X_{n+1} = \{u_I \mid I_{n+1} \subseteq I \subseteq K, \ I \text{ finite}\}.$$

Since $I_n \subseteq I_{n+1}$, it is obvious that $X_{n+1} \subseteq X_n$ for all $n \geq 1$. We need to prove that each $X_n$ has diameter at most $1/n$. Since $J_n$ was chosen such that

$$\|u_{J_n}\| < 1/(2n)$$

for every finite subset $J$ of $K$ with $J_n \cap J = \emptyset$, and since $J_n \subseteq I_n$, it is also true that

$$\|u_J\| < 1/(2n)$$
for every finite subset $J$ of $K$ with $I_n \cap J = \emptyset$ (since $I_n \cap J = \emptyset$ and $J_n \subseteq I_n$ implies that $J_n \cap J = \emptyset$). Then, for every two finite subsets $J, L$ such that $I_n \subseteq J, L \subseteq K$, we have

$$\|u_{J-I_n}\| < 1/(2n) \text{ and } \|u_{L-I_n}\| < 1/(2n),$$

and since

$$\|u_J - u_L\| \leq \|u_J - u_{I_n}\| + \|u_{I_n} - u_L\| = \|u_{J-I_n}\| + \|u_{L-I_n}\|,$$

we get

$$\|u_J - u_L\| < 1/n,$$

which proves that $\delta(X_n) \leq 1/n$. Now, if we consider the sequence of closed sets $(\overline{X_n})$, we still have $\overline{X_{n+1}} \subseteq \overline{X_n}$, and by Proposition 18.4, $\delta(\overline{X_n}) = \delta(X_n) \leq 1/n$, which means that $\lim_{n \to \infty} \delta(\overline{X_n}) = 0$, and by Proposition 18.4, $\bigcap_{n=1}^{\infty} \overline{X_n}$ consists of a single element $u$. We claim that $u$ is the sum of the family $(u_k)_{k \in K}$.

For every $\epsilon > 0$, there is some $n \geq 1$ such that $n > 2/\epsilon$, and since $u \in \overline{X_m}$ for all $m \geq 1$, there is some finite subset $J_0$ of $K$ such that $I_n \subseteq J_0$ and

$$\|u - u_{J_0}\| < \epsilon/2,$$

where $I_n$ is the finite subset of $K$ involved in the definition of $X_n$. However, since $\delta(X_n) \leq 1/n$, for every finite subset $J$ of $K$ such that $I_n \subseteq J$, we have

$$\|u_J - u_{J_0}\| \leq 1/n < \epsilon/2,$$

and since

$$\|u - u_J\| \leq \|u - u_{J_0}\| + \|u_{J_0} - u_J\|,$$

we get

$$\|u - u_J\| < \epsilon$$

for every finite subset $J$ of $K$ with $I_n \subseteq J$, which proves that $u$ is the sum of the family $(u_k)_{k \in K}$.

(2) Since every finite sum $\sum_{i \in I} r_i$ is bounded by the uniform bound $B$, the set of these finite sums has a least upper bound $r \leq B$. For every $\epsilon > 0$, since $r$ is the least upper bound of the finite sums $\sum_{i \in I} r_i$ (where $I$ finite, $I \subseteq K$), there is some finite $I \subseteq K$ such that

$$\left| r - \sum_{i \in I} r_i \right| < \epsilon,$$

and since $r_k \geq 0$ for all $k \in K$, we have

$$\sum_{i \in I} r_i \leq \sum_{j \in J} r_j.$$
whenever \( I \subseteq J \), which shows that

\[
\left| r - \sum_{j \in J} r_j \right| \leq \left| r - \sum_{i \in I} r_i \right| < \epsilon
\]

for every finite subset \( J \) such that \( I \subseteq J \subseteq K \), proving that \((r_k)_{k \in K}\) is summable with sum \( \sum_{k \in K} r_k = r \). \( \square \)

**Remark:** The notion of summability implies that the sum of a family \((u_k)_{k \in K}\) is independent of any order on \( K \). In this sense, it is a kind of “commutative summability”. More precisely, it is easy to show that for every bijection \( \varphi: K \to K \) (intuitively, a reordering of \( K \)), the family \((u_k)_{k \in K}\) is summable iff the family \((u_{\varphi(k)})_{k \in K}\) is summable, and if so, they have the same sum.

The following proposition gives some of the main properties of Fourier coefficients. Among other things, at most countably many of the Fourier coefficient may be nonnull, and the partial sums of a Fourier series converge. Given an orthogonal family \((u_k)_{k \in K}\), we let \( U_k = \mathbb{C} u_k \), and \( p_V: E \to U_k \) is the projection of \( E \) onto \( U_k \).

**Proposition 18.12.** Let \( E \) be a Hilbert space, \((u_k)_{k \in K}\) an orthogonal family in \( E \), and \( V \) the closure of the subspace generated by \((u_k)_{k \in K}\). The following properties hold:

1. For every \( v \in E \), for every finite subset \( I \subseteq K \), we have

\[
\sum_{i \in I} |c_i|^2 \leq \|v\|^2;
\]

where the \( c_k \) are the Fourier coefficients of \( v \).

2. For every vector \( v \in E \), if \((c_k)_{k \in K}\) are the Fourier coefficients of \( v \), the following conditions are equivalent:

   (2a) \( v \in V \)

   (2b) The family \((c_k u_k)_{k \in K}\) is summable and \( v = \sum_{k \in K} c_k u_k \).

   (2c) The family \((|c_k|^2)_{k \in K}\) is summable and \( \|v\|^2 = \sum_{k \in K} |c_k|^2 \);

3. The family \((|c_k|^2)_{k \in K}\) is summable, and we have the Bessel inequality:

\[
\sum_{k \in K} |c_k|^2 \leq \|v\|^2.
\]

As a consequence, at most countably many of the \( c_k \) may be nonzero. The family \((c_k u_k)_{k \in K}\) forms a Cauchy family, and thus, the Fourier series \( \sum_{k \in K} c_k u_k \) converges in \( E \) to some vector \( u = \sum_{k \in K} c_k u_k \). Furthermore, \( u = p_V(v) \).


18.2. TOTAL ORTHOGONAL FAMILIES, FOURIER COEFFICIENTS

Proof. (1) Let

\[ u_I = \sum_{i \in I} c_i u_i \]

for any finite subset \( I \) of \( K \). We claim that \( v - u_I \) is orthogonal to \( u_i \) for every \( i \in I \). Indeed,

\[ \langle v - u_I, u_i \rangle = \left\langle v - \sum_{j \in I} c_j u_j, u_i \right\rangle \]

\[ = \langle v, u_i \rangle - \sum_{j \in I} c_j \langle u_j, u_i \rangle \]

\[ = \langle v, u_i \rangle - c_i \| u_i \| ^2 \]

\[ = \langle v, u_i \rangle - \langle v, u_i \rangle = 0, \]

since \( \langle u_j, u_i \rangle = 0 \) for all \( i \neq j \) and \( c_i = \langle v, u_i \rangle / \| u_i \|^2 \). As a consequence, we have

\[ \| v \|^2 = \left\| v - \sum_{i \in I} c_i u_i \right\|^2 \]

\[ = \left\| v - \sum_{i \in I} c_i u_i \right\|^2 + \left\| \sum_{i \in I} c_i u_i \right\|^2 \]

\[ = \left\| v - \sum_{i \in I} c_i u_i \right\|^2 + \sum_{i \in I} |c_i|^2, \]

since the \( u_i \) are pairwise orthogonal, that is,

\[ \| v \|^2 = \left\| v - \sum_{i \in I} c_i u_i \right\|^2 + \sum_{i \in I} |c_i|^2. \]

Thus,

\[ \sum_{i \in I} |c_i|^2 \leq \| v \|^2, \]

as claimed.

(2) We prove the chain of implications \((a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (a)\).

\((a) \Rightarrow (b)\): If \( v \in V \), since \( V \) is the closure of the subspace spanned by \( (u_k)_{k \in K} \), for every \( \epsilon > 0 \), there is some finite subset \( I \) of \( K \) and some family \( (\lambda_i)_{i \in I} \) of complex numbers, such that

\[ \left\| v - \sum_{i \in I} \lambda_i u_i \right\| < \epsilon. \]

Now, for every finite subset \( J \) of \( K \) such that \( I \subseteq J \), we have

\[ \left\| v - \sum_{i \in I} \lambda_i u_i \right\|^2 = \left\| v - \sum_{j \in J} c_j u_j + \sum_{j \in J} c_j u_j - \sum_{i \in I} \lambda_i u_i \right\|^2 \]

\[ = \left\| v - \sum_{j \in J} c_j u_j \right\|^2 + \left\| \sum_{j \in J} c_j u_j - \sum_{i \in I} \lambda_i u_i \right\|^2, \]

which completes the proof.
since $I \subseteq J$ and the $u_j$ (with $j \in J$) are orthogonal to $v - \sum_{j \in J} c_j u_j$ by the argument in (1), which shows that

$$\left\| v - \sum_{j \in J} c_j u_j \right\| \leq \left\| v - \sum_{i \in I} \lambda_i u_i \right\| < \epsilon,$$

and thus, that the family $(c_k u_k)_{k \in K}$ is summable with sum $v$, so that

$$v = \sum_{k \in K} c_k u_k.$$

(b) $\Rightarrow$ (c): If $v = \sum_{k \in K} c_k u_k$, then for every $\epsilon > 0$, there some finite subset $I$ of $K$, such that

$$\left\| v - \sum_{j \in J} c_j u_j \right\| < \sqrt{\epsilon},$$

for every finite subset $J$ of $K$ such that $I \subseteq J$, and since we proved in (1) that

$$\|v\|^2 = \left\| v - \sum_{j \in J} c_j u_j \right\|^2 + \sum_{j \in J} |c_j|^2,$$

we get

$$\|v\|^2 - \sum_{j \in J} |c_j|^2 < \epsilon,$$

which proves that $(|c_k|^2)_{k \in K}$ is summable with sum $\|v\|^2$.

(c) $\Rightarrow$ (a): Finally, if $(|c_k|^2)_{k \in K}$ is summable with sum $\|v\|^2$, for every $\epsilon > 0$, there is some finite subset $I$ of $K$ such that

$$\|v\|^2 - \sum_{j \in J} |c_j|^2 < \epsilon^2$$

for every finite subset $J$ of $K$ such that $I \subseteq J$, and again, using the fact that

$$\|v\|^2 = \left\| v - \sum_{j \in J} c_j u_j \right\|^2 + \sum_{j \in J} |c_j|^2,$$

we get

$$\left\| v - \sum_{j \in J} c_j u_j \right\| < \epsilon,$$

which proves that $(c_k u_k)_{k \in K}$ is summable with sum $\sum_{k \in K} c_k u_k = v$, and $v \in V$.

(3) Since $\sum_{i \in I} |c_i|^2 \leq \|v\|^2$ for every finite subset $I$ of $K$, by Proposition 18.11, the family $(|c_k|^2)_{k \in K}$ is summable. The Bessel inequality

$$\sum_{k \in K} |c_k|^2 \leq \|v\|^2$$
is an obvious consequence of the inequality $\sum_{i \in I} |c_i|^2 \leq \|v\|^2$ (for every finite $I \subseteq K$). Now, for every natural number $n \geq 1$, if $K_n$ is the subset of $K$ consisting of all $c_k$ such that $|c_k| \geq 1/n$, the number of elements in $K_n$ is at most

$$\sum_{k \in K_n} |nc_k|^2 \leq n^2 \sum_{k \in K} |c_k|^2 \leq n^2 \|v\|^2,$$

which is finite, and thus, at most a countable number of the $c_k$ may be nonzero.

Since $(|c_k|^2)_{k \in K}$ is summable with sum $c$, for every $\epsilon > 0$, there is some finite subset $I$ of $K$ such that

$$\sum_{j \in J} |c_j|^2 < \epsilon^2$$

for every finite subset $J$ of $K$ such that $I \cap J = \emptyset$. Since

$$\left\| \sum_{j \in J} c_j u_j \right\|^2 = \sum_{j \in J} |c_j|^2,$$

we get

$$\left\| \sum_{j \in J} c_j u_j \right\| < \epsilon.$$

This proves that $(c_k u_k)_{k \in K}$ is a Cauchy family, which, by Proposition 18.11, implies that $(c_k u_k)_{k \in K}$ is summable, since $E$ is complete. Thus, the Fourier series $\sum_{k \in K} c_k u_k$ is summable, with its sum denoted $u \in V$.

Since $\sum_{k \in K} c_k u_k$ is summable with sum $u$, for every $\epsilon > 0$, there is some finite subset $I_1$ of $K$ such that

$$\left\| u - \sum_{j \in J} c_j u_j \right\| < \epsilon$$

for every finite subset $J$ of $K$ such that $I_1 \subseteq J$. By the triangle inequality, for every finite subset $I$ of $K$,

$$\left\| u - v \right\| \leq \left\| u - \sum_{i \in I} c_i u_i \right\| + \left\| \sum_{i \in I} c_i u_i - v \right\|.$$

By (2), every $w \in V$ is the sum of its Fourier series $\sum_{k \in K} \lambda_k u_k$, and for every $\epsilon > 0$, there is some finite subset $I_2$ of $K$ such that

$$\left\| w - \sum_{j \in J} \lambda_j u_j \right\| < \epsilon$$

for every finite subset $J$ of $K$ such that $I_2 \subseteq J$. By the triangle inequality, for every finite subset $I$ of $K$,

$$\left\| v - \sum_{i \in I} \lambda_i u_i \right\| \leq \left\| v - w \right\| + \left\| w - \sum_{i \in I} \lambda_i u_i \right\|.$$
Letting $I = I_1 \cup I_2$, since we showed in (2) that
\[
\left\| v - \sum_{i \in I} c_i u_i \right\| \leq \left\| v - \sum_{i \in I} \lambda_i u_i \right\|
\]
for every finite subset $I$ of $K$, we get
\[
\| u - v \| \leq \left\| u - \sum_{i \in I} c_i u_i \right\| + \left\| \sum_{i \in I} c_i u_i - v \right\|
\leq \left\| u - \sum_{i \in I} c_i u_i \right\| + \left\| \sum_{i \in I} \lambda_i u_i - v \right\|
\leq \left\| u - \sum_{i \in I} c_i u_i \right\| + \| v - w \| + \left\| w - \sum_{i \in I} \lambda_i u_i \right\|,
\]
and thus
\[
\| u - v \| \leq \| v - w \| + 2\epsilon.
\]
Since this holds for every $\epsilon > 0$, we have
\[
\| u - v \| \leq \| v - w \|
\]
for all $w \in V$, i.e. $\| v - u \| = d(v, V)$, with $u \in V$, which proves that $u = p_V(v)$. \(\square\)

### 18.3 The Hilbert Space $l^2(K)$ and the Riesz-Fischer Theorem

Proposition 18.12 suggests looking at the space of sequences $(z_k)_{k \in K}$ (where $z_k \in \mathbb{C}$) such that $(|z_k|^2)_{k \in K}$ is summable. Indeed, such spaces are Hilbert spaces, and it turns out that every Hilbert space is isomorphic to one of those. Such spaces are the infinite-dimensional version of the spaces $\mathbb{C}^n$ under the usual Euclidean norm.

**Definition 18.6.** Given any nonempty index set $K$, the space $l^2(K)$ is the set of all sequences $(z_k)_{k \in K}$, where $z_k \in \mathbb{C}$, such that $(|z_k|^2)_{k \in K}$ is summable, i.e., $\sum_{k \in K} |z_k|^2 < \infty$.

**Remarks:**

1. When $K$ is a finite set of cardinality $n$, $l^2(K)$ is isomorphic to $\mathbb{C}^n$.

2. When $K = \mathbb{N}$, the space $l^2(\mathbb{N})$ is the space $l^2$ from Example 18.1. It is a Hilbert space, and we now prove this fact for any index set $K$. 
Proposition 18.13. Given any nonempty index set $K$, the space $l^2(K)$ is a Hilbert space under the Hermitian product

$$\langle (x_k)_{k\in K}, (y_k)_{k\in K} \rangle = \sum_{k\in K} x_k \overline{y}_k.$$ 

The subspace consisting of sequences $(z_k)_{k\in K}$ such that $z_k = 0$, except perhaps for finitely many $k$, is a dense subspace of $l^2(K)$.

Proof. First, we need to prove that $l^2(K)$ is a vector space. Assume that $(x_k)_{k\in K}$ and $(y_k)_{k\in K}$ are in $l^2(K)$. This means that $(|x_k|^2)_{k\in K}$ and $(|y_k|^2)_{k\in K}$ are summable, which, in view of Proposition 18.11, is equivalent to the existence of some positive bounds $A$ and $B$ such that $\sum_{i\in I} |x_i|^2 < A$ and $\sum_{i\in I} |y_i|^2 < B$, for every finite subset $I$ of $K$. To prove that $(|x_k + y_k|^2)_{k\in K}$ is summable, it is sufficient to prove that there is some $C > 0$ such that $\sum_{i\in I} |x_i + y_i|^2 < C$ for every finite subset $I$ of $K$. However, the parallelogram inequality implies that

$$\sum_{i\in I} |x_i + y_i|^2 \leq \sum_{i\in I} 2(|x_i|^2 + |y_i|^2) \leq 2(A + B),$$

for every finite subset $I$ of $K$, and we conclude by Proposition 18.11. Similarly, for every $\lambda \in \mathbb{C}$,

$$\sum_{i\in I} |\lambda x_i|^2 \leq \sum_{i\in I} |\lambda|^2 |x_i|^2 \leq |\lambda|^2 A,$$

and $(\lambda_k x_k)_{k\in K}$ is summable. Therefore, $l^2(K)$ is a vector space.

By the Cauchy-Schwarz inequality,

$$\sum_{i\in I} |x_i \overline{y}_i| \leq \sum_{i\in I} |x_i| |y_i| \leq (\sum_{i\in I} |x_i|^2)^{1/2}(\sum_{i\in I} |y_i|^2)^{1/2} \leq \sum_{i\in I} (|x_i|^2 + |y_i|^2)/2 \leq (A + B)/2,$$

for every finite subset $I$ of $K$. Here, we used the fact that

$$4CD \leq (C + D)^2,$$

which is equivalent to

$$(C - D)^2 \geq 0.$$ 

By Proposition 18.11, $(|x_k \overline{y}_k|)_{k\in K}$ is summable. The customary language is that $(x_k \overline{y}_k)_{k\in K}$ is absolutely summable. However, it is a standard fact that this implies that $(x_k \overline{y}_k)_{k\in K}$ is summable (For every $\epsilon > 0$, there is some finite subset $I$ of $K$ such that

$$\sum_{j \in J} |x_j \overline{y}_j| < \epsilon$$

for every finite subset $J$ of $K$ such that $I \cap J = \emptyset$, and thus

$$\sum_{j \in J} |x_j \overline{y}_j| \leq \sum_{i \in J} |x_j \overline{y}_j| < \epsilon,$$
proving that \((x_ky_k)_{k \in K}\) is a Cauchy family, and thus summable). We still have to prove that \(l^2(K)\) is complete.

Consider a sequence \(((\lambda^n_k)_{k \in K})_{n \geq 1}\) of sequences \((\lambda^n_k)_{k \in K} \in l^2(K)\), and assume that it is a Cauchy sequence. This means that for every \(\epsilon > 0\), there is some \(N \geq 1\) such that

\[
\sum_{k \in K} |\lambda^m_k - \lambda^n_k|^2 < \epsilon^2
\]

for all \(m, n \geq N\). For every fixed \(k \in K\), this implies that

\[
|\lambda^m_k - \lambda^n_k| < \epsilon
\]

for all \(m, n \geq N\), which shows that \((\lambda^n_k)_{n \geq 1}\) is a Cauchy sequence in \(\mathbb{C}\). Since \(\mathbb{C}\) is complete, the sequence \((\lambda^n_k)_{n \geq 1}\) has a limit \(\lambda_k \in \mathbb{C}\). We claim that \((\lambda_k)_{k \in K} \in l^2(K)\) and that this is the limit of \(((\lambda^n_k)_{k \in K})_{n \geq 1}\).

Given any \(\epsilon > 0\), the fact that \(((\lambda^n_k)_{k \in K})_{n \geq 1}\) is a Cauchy sequence implies that there is some \(N \geq 1\) such that for every finite subset \(I\) of \(K\), we have

\[
\sum_{i \in I} |\lambda^m_i - \lambda^n_i|^2 < \epsilon/4
\]

for all \(m, n \geq N\). Let \(p = |I|\). Then,

\[
|\lambda^m_i - \lambda^n_i| < \frac{\sqrt{\epsilon}}{2\sqrt{p}}
\]

for every \(i \in I\). Since \(\lambda_i\) is the limit of \((\lambda^n_i)_{n \geq 1}\), we can find some \(n\) large enough so that

\[
|\lambda^n_i - \lambda_i| < \frac{\sqrt{\epsilon}}{2\sqrt{p}}
\]

for every \(i \in I\). Since

\[
|\lambda^m_i - \lambda_i| \leq |\lambda^m_i - \lambda^n_i| + |\lambda^n_i - \lambda_i|
\]

we get

\[
|\lambda^m_i - \lambda_i| < \frac{\sqrt{\epsilon}}{\sqrt{p}}
\]

and thus,

\[
\sum_{i \in I} |\lambda^m_i - \lambda_i|^2 < \epsilon,
\]

for all \(m \geq N\). Since the above holds for every finite subset \(I\) of \(K\), by Proposition 18.11, we get

\[
\sum_{k \in K} |\lambda^m_k - \lambda_k|^2 < \epsilon,
\]
for all $m \geq N$. This proves that $(\lambda^m_k - \lambda_k)_{k \in K} \in l^2(K)$ for all $m \geq N$, and since $l^2(K)$ is a vector space and $(\lambda^m_k)_{k \in K} \in l^2(K)$ for all $m \geq 1$, we get $(\lambda_k)_{k \in K} \in l^2(K)$. However,

$$\sum_{k \in K} |\lambda^m_k - \lambda_k|^2 < \epsilon$$

for all $m \geq N$, means that the sequence $(\lambda^m_k)_{k \in K}$ converges to $(\lambda_k)_{k \in K} \in l^2(K)$. The fact that the subspace consisting of sequences $(z_k)_{k \in K}$ such that $z_k = 0$ except perhaps for finitely many $k$ is a dense subspace of $l^2(K)$ is left as an easy exercise.

**Remark:** The subspace consisting of all sequences $(z_k)_{k \in K}$ such that $z_k = 0$, except perhaps for finitely many $k$, provides an example of a subspace which is not closed in $l^2(K)$. Indeed, this space is strictly contained in $l^2(K)$, since there are countable sequences of nonnull elements in $l^2(K)$ (why?).

We just need two more propositions before being able to prove that every Hilbert space is isomorphic to some $l^2(K)$.

**Proposition 18.14.** Let $E$ be a Hilbert space, and $(u_k)_{k \in K}$ an orthogonal family in $E$. The following properties hold:

1. For every family $(\lambda_k)_{k \in K} \in l^2(K)$, the family $(\lambda_k u_k)_{k \in K}$ is summable. Furthermore, $v = \sum_{k \in K} \lambda_k u_k$ is the only vector such that $c_k = \lambda_k$ for all $k \in K$, where the $c_k$ are the Fourier coefficients of $v$.

2. For any two families $(\lambda_k)_{k \in K} \in l^2(K)$ and $(\mu_k)_{k \in K} \in l^2(K)$, if $v = \sum_{k \in K} \lambda_k u_k$ and $w = \sum_{k \in K} \mu_k u_k$, we have the following equation, also called Parseval identity:

$$\langle v, w \rangle = \sum_{k \in K} \lambda_k \overline{\mu_k}.$$

**Proof.** (1) The fact that $(\lambda_k)_{k \in K} \in l^2(K)$ means that $(|\lambda_k|^2)_{k \in K}$ is summable. The proof given in Proposition 18.12 (3) applies to the family $(|\lambda_k|^2)_{k \in K}$ (instead of $(|c_k|^2)_{k \in K}$), and yields the fact that $(\lambda_k u_k)_{k \in K}$ is summable. Letting $v = \sum_{k \in K} \lambda_k u_k$, recall that $c_k = \langle v, u_k \rangle / \|u_k\|^2$. Pick some $k \in K$. Since $\langle -, - \rangle$ is continuous, for every $\epsilon > 0$, there is some $\eta > 0$ such that

$$|\langle v, u_k \rangle - \langle w, u_k \rangle| < \epsilon \|u_k\|^2$$

whenever

$$\|v - w\| < \eta.$$

However, since for every $\eta > 0$, there is some finite subset $I$ of $K$ such that

$$\left\| v - \sum_{j \in I} \lambda_j u_j \right\| < \eta$$
for every finite subset \(J\) of \(K\) such that \(I \subseteq J\), we can pick \(J = I \cup \{k\}\), and letting \(w = \sum_{j \in J} \lambda_j u_j\), we get

\[
\left| \langle v, u_k \rangle - \left\langle \sum_{j \in J} \lambda_j u_j, u_k \right\rangle \right| < \epsilon \|u_k\|^2.
\]

However,

\[
\langle v, u_k \rangle = c_k \|u_k\|^2 \quad \text{and} \quad \left\langle \sum_{j \in J} \lambda_j u_j, u_k \right\rangle = \lambda_k \|u_k\|^2,
\]

and thus, the above proves that \(|c_k - \lambda_k| < \epsilon\) for every \(\epsilon > 0\), and thus, that \(c_k = \lambda_k\).

(2) Since \(\langle -,-\rangle\) is continuous, for every \(\epsilon > 0\), there are some \(\eta_1 > 0\) and \(\eta_2 > 0\), such that

\[
|\langle x, y \rangle| < \epsilon
\]

whenever \(\|x\| < \eta_1\) and \(\|y\| < \eta_2\). Since \(v = \sum_{k \in K} \lambda_k u_k\) and \(w = \sum_{k \in K} \mu_k u_k\), there is some finite subset \(I_1\) of \(K\) such that

\[
\left\| v - \sum_{j \in J} \lambda_j u_j \right\| < \eta_1
\]

for every finite subset \(J\) of \(K\) such that \(I_1 \subseteq J\), and there is some finite subset \(I_2\) of \(K\) such that

\[
\left\| w - \sum_{j \in J} \mu_j u_j \right\| < \eta_2
\]

for every finite subset \(J\) of \(K\) such that \(I_2 \subseteq J\). Letting \(I = I_1 \cup I_2\), we get

\[
\left| \left\langle v - \sum_{i \in I} \lambda_i u_i, w - \sum_{i \in I} \mu_i u_i \right\rangle \right| < \epsilon.
\]

Furthermore,

\[
\langle v, w \rangle = \langle v - \sum_{i \in I} \lambda_i u_i, w - \sum_{i \in I} \mu_i u_i \rangle + \sum_{i \in I} \lambda_i \mu_i u_i
\]

since the \(u_i\) are orthogonal to \(v - \sum_{i \in I} \lambda_i u_i\) and \(w - \sum_{i \in I} \mu_i u_i\) for all \(i \in I\). This proves that for every \(\epsilon > 0\), there is some finite subset \(I\) of \(K\) such that

\[
\left| \langle v, w \rangle - \sum_{i \in I} \lambda_i \mu_i \right| < \epsilon.
\]
We already know from Proposition 18.13 that \((\lambda_k \mu_k)_{k \in K}\) is summable, and since \(\epsilon > 0\) is arbitrary, we get
\[
\langle v, w \rangle = \sum_{k \in K} \lambda_k \mu_k.
\]

The next proposition states properties characterizing Hilbert bases (total orthogonal families).

**Proposition 18.15.** Let \(E\) be a Hilbert space, and let \((u_k)_{k \in K}\) be an orthogonal family in \(E\). The following properties are equivalent:

1. The family \((u_k)_{k \in K}\) is a total orthogonal family.
2. For every vector \(v \in E\), if \((c_k)_{k \in K}\) are the Fourier coefficients of \(v\), then the family \((c_k u_k)_{k \in K}\) is summable and \(v = \sum_{k \in K} c_k u_k\).
3. For every vector \(v \in E\), we have the Parseval identity:
   \[
   \|v\|^2 = \sum_{k \in K} |c_k|^2.
   \]
4. For every vector \(u \in E\), if \(\langle u, u_k \rangle = 0\) for all \(k \in K\), then \(u = 0\).

**Proof.** The equivalence of (1), (2), and (3), is an immediate consequence of Proposition 18.12 and Proposition 18.14.

(4) If \((u_k)_{k \in K}\) is a total orthogonal family and \(\langle u, u_k \rangle = 0\) for all \(k \in K\), since \(u = \sum_{k \in K} c_k u_k\) where \(c_k = \langle u, u_k \rangle / \|u_k\|^2\), we have \(c_k = 0\) for all \(k \in K\), and \(u = 0\).

Conversely, assume that the closure \(V\) of \((u_k)_{k \in K}\) is different from \(E\). Then, by Proposition 18.7, we have \(E = V \oplus V^\perp\), where \(V^\perp\) is the orthogonal complement of \(V\), and \(V^\perp\) is nontrivial since \(V \neq E\). As a consequence, there is some nonnull vector \(u \in V^\perp\). But then, \(u\) is orthogonal to every vector in \(V\), and in particular,
\[
\langle u, u_k \rangle = 0
\]
for all \(k \in K\), which, by assumption, implies that \(u = 0\), contradicting the fact that \(u \neq 0\). \(\square\)

**Remarks:**
(1) If $E$ is a Hilbert space and $(u_k)_{k \in K}$ is a total orthogonal family in $E$, there is a simpler argument to prove that $u = 0$ if $\langle u, u_k \rangle = 0$ for all $k \in K$, based on the continuity of $\langle - , - \rangle$. The argument is to prove that the assumption implies that $\langle v, u \rangle = 0$ for all $v \in E$. Since $\langle - , - \rangle$ is positive definite, this implies that $u = 0$. By continuity of $\langle - , - \rangle$, for every $\epsilon > 0$, there is some $\eta > 0$ such that for every finite subset $I$ of $K$, for every family $(\lambda_i)_{i \in I}$, for every $v \in E$,
\[
\left| \langle v, u \rangle - \left( \sum_{i \in I} \lambda_i u_i, u \right) \right| < \epsilon
\]
whenever
\[
\left\| v - \sum_{i \in I} \lambda_i u_i \right\| < \eta.
\]

Since $(u_k)_{k \in K}$ is dense in $E$, for every $v \in E$, there is some finite subset $I$ of $K$ and some family $(\lambda_i)_{i \in I}$ such that
\[
\left\| v - \sum_{i \in I} \lambda_i u_i \right\| < \eta,
\]
and since by assumption, $\langle \sum_{i \in I} \lambda_i u_i, u \rangle = 0$, we get
\[
|\langle v, u \rangle| < \epsilon.
\]
Since this holds for every $\epsilon > 0$, we must have $\langle v, u \rangle = 0$.

(2) If $V$ is any nonempty subset of $E$, the kind of argument used in the previous remark can be used to prove that $V^\perp$ is closed (even if $V$ is not), and that $V^{\perp \perp}$ is the closure of $V$.

We will now prove that every Hilbert space has some Hilbert basis. This requires using a fundamental theorem from set theory known as Zorn’s Lemma, which we quickly review.

Given any set $X$ with a partial ordering $\leq$, recall that a nonempty subset $C$ of $X$ is a chain if it is totally ordered (i.e., for all $x, y \in C$, either $x \leq y$ or $y \leq x$). A nonempty subset $Y$ of $X$ is bounded iff there is some $b \in X$ such that $y \leq b$ for all $y \in Y$. Some $m \in X$ is maximal iff for every $x \in X$, $m \leq x$ implies that $x = m$. We can now state Zorn’s Lemma. For more details, see Rudin [66], Lang [51], or Artin [3].

**Proposition 18.16.** Given any nonempty partially ordered set $X$, if every (nonempty) chain in $X$ is bounded, then $X$ has some maximal element.

We can now prove the existence of Hilbert bases. We define a partial order on families $(u_k)_{k \in K}$ as follows: For any two families $(u_k)_{k \in K_1}$ and $(v_k)_{k \in K_2}$, we say that
\[
(u_k)_{k \in K_1} \leq (v_k)_{k \in K_2}
\]
iff $K_1 \subseteq K_2$ and $u_k = v_k$ for all $k \in K_1$. This is clearly a partial order.
Proposition 18.17. Let $E$ be a Hilbert space. Given any orthogonal family $(u_k)_{k \in K}$ in $E$, there is a total orthogonal family $(u_l)_{l \in L}$ containing $(u_k)_{k \in K}$.

Proof. Consider the set $S$ of all orthogonal families greater than or equal to the family $B = (u_k)_{k \in K}$. We claim that every chain in $S$ is bounded. Indeed, if $C = (C_l)_{l \in L}$ is a chain in $S$, where $C_l = (u_{k,l})_{k \in K_l}$, the union family

$$(u_k)_{k \in \bigcup_{l \in L} K_l}, \text{ where } u_k = u_{k,l} \text{ whenever } k \in K_l,$$

is clearly an upper bound for $C$, and it is immediately verified that it is an orthogonal family. By Zorn’s Lemma 18.16, there is a maximal family $(u_l)_{l \in L}$ containing $(u_k)_{k \in K}$. If $(u_l)_{l \in L}$ is not dense in $E$, then its closure $V$ is strictly contained in $E$, and by Proposition 18.7, the orthogonal complement $V^\perp$ of $V$ is nontrivial since $V \neq E$. As a consequence, there is some nonnull vector $u \in V^\perp$. But then, $u$ is orthogonal to every vector in $(u_l)_{l \in L}$, and we can form an orthogonal family strictly greater than $(u_l)_{l \in L}$ by adding $u$ to this family, contradicting the maximality of $(u_l)_{l \in L}$. Therefore, $(u_l)_{l \in L}$ is dense in $E$, and thus, it is a Hilbert basis. \[\square\]

Remark: It is possible to prove that all Hilbert bases for a Hilbert space $E$ have index sets $K$ of the same cardinality. For a proof, see Bourbaki [10].

At last, we can prove that every Hilbert space is isomorphic to some Hilbert space $l^2(K)$ for some suitable $K$.

Theorem 18.18. (Riesz-Fischer) For every Hilbert space $E$, there is some nonempty set $K$ such that $E$ is isomorphic to the Hilbert space $l^2(K)$. More specifically, for any Hilbert basis $(u_k)_{k \in K}$ of $E$, the maps $f: l^2(K) \to E$ and $g: E \to l^2(K)$ defined such that

$$f((\lambda_k)_{k \in K}) = \sum_{k \in K} \lambda_k u_k \quad \text{and} \quad g(u) = (\langle u, u_k \rangle / \|u_k\|^2)_{k \in K} = (c_k)_{k \in K},$$

are bijective linear isometries such that $g \circ f = \text{id}$ and $f \circ g = \text{id}$.

Proof. By Proposition 18.14 (1), the map $f$ is well defined, and it it clearly linear. By Proposition 18.12 (3), the map $g$ is well defined, and it is also clearly linear. By Proposition 18.12 (2b), we have

$$f(g(u)) = u = \sum_{k \in K} c_k u_k,$$

and by Proposition 18.14 (1), we have

$$g(f((\lambda_k)_{k \in K})) = (\lambda_k)_{k \in K},$$

and thus $g \circ f = \text{id}$ and $f \circ g = \text{id}$. By Proposition 18.14 (2), the linear map $g$ is an isometry. Therefore, $f$ is a linear bijection and an isometry between $l^2(K)$ and $E$, with inverse $g$. \[\square\]
Remark: The surjectivity of the map \( g : E \to l^2(K) \) is known as the Riesz-Fischer theorem.

Having done all this hard work, we sketch how these results apply to Fourier series. Again, we refer the readers to Rudin [66] or Lang [52, 53] for a comprehensive exposition.

Let \( C(T) \) denote the set of all periodic continuous functions \( f : [-\pi, \pi] \to \mathbb{C} \) with period \( 2\pi \). There is a Hilbert space \( L^2(T) \) containing \( C(T) \) and such that \( C(T) \) is dense in \( L^2(T) \), whose inner product is given by

\[
\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) \overline{g(x)} \, dx.
\]

The Hilbert space \( L^2(T) \) is the space of Lebesgue square-integrable periodic functions (of period \( 2\pi \)).

It turns out that the family \( (e^{ikx})_{k \in \mathbb{Z}} \) is a total orthogonal family in \( L^2(T) \), because it is already dense in \( C(T) \) (for instance, see Rudin [66]). Then, the Riesz-Fischer theorem says that for every family \( (c_k)_{k \in \mathbb{Z}} \) of complex numbers such that

\[
\sum_{k \in \mathbb{Z}} |c_k|^2 < \infty,
\]

there is a unique function \( f \in L^2(T) \) such that \( f \) is equal to its Fourier series

\[
f(x) = \sum_{k \in \mathbb{Z}} c_ke^{ikx},
\]

where the Fourier coefficients \( c_k \) of \( f \) are given by the formula

\[
c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)e^{-ikt} \, dt.
\]

The Parseval theorem says that

\[
\sum_{k=-\infty}^{+\infty} c_k \overline{d_k} = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)g(t) \, dt
\]

for all \( f, g \in L^2(T) \), where \( c_k \) and \( d_k \) are the Fourier coefficients of \( f \) and \( g \).

Thus, there is an isomorphism between the two Hilbert spaces \( L^2(T) \) and \( l^2(\mathbb{Z}) \), which is the deep reason why the Fourier coefficients “work”. Theorem 18.18 implies that the Fourier series \( \sum_{k \in \mathbb{Z}} c_k e^{ikx} \) of a function \( f \in L^2(T) \) converges to \( f \) in the \( L^2 \)-sense, i.e., in the mean-square sense. This does not necessarily imply that the Fourier series converges to \( f \) pointwise! This is a subtle issue, and for more on this subject, the reader is referred to Lang [52, 53] or Schwartz [72, 73].
We can also consider the set $\mathcal{C}([-1, 1])$ of continuous functions $f: [-1, 1] \to \mathbb{C}$. There is a Hilbert space $L^2([-1, 1])$ containing $\mathcal{C}([-1, 1])$ and such that $\mathcal{C}([-1, 1])$ is dense in $L^2([-1, 1])$, whose inner product is given by

$$\langle f, g \rangle = \int_{-1}^{1} f(x)\overline{g(x)}dx.$$  

The Hilbert space $L^2([-1, 1])$ is the space of Lebesgue square-integrable functions over $[-1, 1]$. The Legendre polynomials $P_n(x)$ form a Hilbert basis of $L^2([-1, 1])$.

Recall that if we let $f_n$ be the function

$$f_n(x) = (x^2 - 1)^n,$$

$P_n(x)$ is defined as follows:

$$P_0(x) = 1, \quad \text{and} \quad P_n(x) = \frac{1}{2^n n!} f_n^{(n)}(x),$$

where $f_n^{(n)}$ is the $n$th derivative of $f_n$. The reason for the leading coefficient is to get $P_n(1) = 1$. It can be shown with much efforts that

$$P_n(x) = \sum_{0 \leq k \leq n/2} (-1)^k \frac{(2(n-k))!}{2^n(n-k)! k! (n-2k)!} x^{n-2k}.$$
Bibliography


751


