Three perspectives on Bochner's theorem: from Herglotz representation to Pontryagin duality

Ikhan Choi

Abstract

Bochner's theorem states that the set of continuous positive definite functions on a locally compact abelian group is the image of finite Borel measures on its dual group under the Fourier-Stieltjes transform. This thesis approaches Bochner's theorem from three different viewpoints; complex analysis, probability theory, and representation theory. Special cases of Bochner's theorem will be discussed in the first two chapters via the Herglotz representation theorem and the Lévy continuity theorem. In the rest of the thesis, we prove Bochner's theorem in two ways and also prove the Pontryagin duality theorem as an application in the representation theory of locally compact abelian groups.

Contents

1 Introduction

1.1 A brief history of Bochner's theorem

Bochner's theorem originates from questions about Fourier coefficients and the Fourier transforms of measures. It describes a necessary and sufficient condition for a sequence or a function to be Fourier coefficients or a Fourier transform of a measure. More precisely, the results like the following theorems are examples of *Bochner-type theorems*:

Theorem 1.1. *A function $c : \mathbb{Z} \to \mathbb{C}$ is positive definite if and only if there is a unique finite regular Borel measure μ on $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ such that*

$$c(k) = \int_0^{2\pi} e^{-ik\theta}\, d\mu(\theta)$$

for all $k \in \mathbb{Z}$.

Theorem 1.2. *A continuous function $\varphi : \mathbb{R} \to \mathbb{C}$ is positive definite if and only if there is a unique finite regular Borel measure μ on $\mathbb{R}$ such that*

$$\varphi(t) = \int e^{itx}\, d\mu(x)$$

for all $t \in \mathbb{R}$.

The concept of positive definite functions first appeared in the problem in the complex function theory, called the Carathéodory coefficient problem. It asks the condition for the power series coefficients to form an analytic function that maps the open unit disk into the right half plane. Carathéodory [2] showed in 1907 that one such necessary and sufficient condition would be that the points whose coordinates are given by the power series coefficient of such functions lie in the convex hull of a particular curve. Toeplitz reformulated in 1911 the geometric condition of Carathéodory into algebraic terms — namely the positive definiteness of a sequence in his short article [13]. The Herglotz representation theorem is the most comprehensive result that contains the above two theorems, and relates the probability measure on the circle group $\mathbb{T}$ to the aforementioned positive definite sequences. This result by Herglotz [5] is considered as the first prototype of Bochner's theorem.

Mathias [7] defined and studied the basic properties of the positive definite functions on $\mathbb{R}$ in 1923. Around 1925, the Fourier transform of a measure on $\mathbb{R}$ began to be studied actively by probabilists such as Lévy in order to study the weak convergence of probability measures. Recall that a probability distribution of a real-valued random variable is defined as a probability measure on $\mathbb{R}$. The Fourier transform of a probability measure with reversed sign on the phase term is called the characteristic function of the probability measure. According to the Lévy continuity theorem, the pointwise convergence of characteristic functions implies the weak convergence of a sequence of probability measures. In the celebrated paper [1] published in 1932, Bochner proved that a function on $\mathbb{R}$ is a Fourier transform of a finite measure if and only if it is positive definite and continuous, which gave the theorem his name. See [12] for the further survey about the history of positive definite functions.

Fourier analysis was then extended to abstract groups, and the Banach algebra approaches emerged in the 1940s. For locally compact abelian groups, Weil, Povzner, and Raikov almost simultaneously generalized Bochner's theorem. We introduce a proof of the Pontryagin duality theorem as an application of Bochner's theorem, which states the bidual group is isomorphic to the original group. The original proof by Pontryagin and van Kampen in [11] and [14] uses a different method.

In Chapter 2, we state and prove the Carathéodory coefficient problem and the Toeplitz theorem. Then, we prove the Herglotz representation theorem, and Theorem 1.1, the Bochner theorem on the additive group $\mathbb{Z}$, will be proved as its corollary. We also provide a geometric description of the space of positive definite sequences. In Chapter 3, we review the theory of weak convergence of probability measures on $\mathbb{R}$, including the Lévy-Prokhorov metric and the Prokhorov theorem, and prove Bochner's theorem using the Lévy continuity theorem. Then, we move to general locally compact abelian groups in Chapter 4, and suggest two different methods to prove Bochner's theorem: one direct proof by Fourier transform, another by using the Gelfand-Naimark-Segal construction. We end by proving the Pontryagin duality theorem using Bochner'stheorem, which is one of the most famous applications of Bochner's theorem.

1.2 Positive definite functions

This section discusses the basic properties and examples of positive definite functions. They will be used frequently throughout the whole thesis.

Definition 1.1. Let G be a group. A function $f : G \to \mathbb{C}$ is called *positive definite* if for each positive integer n a non-negativity condition

$$\sum_{k,l=1}^{n} f(x_l^{-1}x_k)\xi_k\overline{\xi}_l \geq 0$$

is satisfied for every n-tuple $(x_1, \cdots, x_n) \in G^n$ and every vector $(\xi_1, \cdots, \xi_n) \in \mathbb{C}^n$.

A function f is positive definite if and only if bilinear forms defined by matrices $(f(x_l^{-1}x_k))_{k,l=1}^{n}$ for each positive integer n are Hermitian, and positive *semi*-definite, regardless of any choices of $(x_1, \cdots, x_n) \in G^n$. We give some several properties and examples of positive definite functions:

Proposition 1.3. *Let G be a group with identity e, and let $(f_m)_{m=1}^{\infty}$ be a sequence of positive functions on G. Then,*

(a) $\overline{f}_1$ *is positive definite. Indeed,* $\overline{f_1(x)} = f_1(x^{-1})$.

(b) af_1 *is positive definite for* $a \geq 0$.

(c) $f_1 + f_2$ *is positive definite.*

(d) $f_1 f_2$ *is positive definite.*

(e) $|f_1(x)| \leq f_1(e)$ *for all* $x \in G$.

(f) *If the pointwise limit* $f = \lim_{m\to\infty} f_m$ *exists, then* f *is positive definite.*

(g) *Let G be a topological group. If* f_1 *is continuous at the e, then it is both-sided uniformly continuous.*

Proof. (a) Note $0 \leq \overline{\xi} f(e)\xi$ implies $f(e) \in \mathbb{R}$. Since

$$0 \leq \begin{pmatrix} 1 & \overline{\xi} \end{pmatrix} \begin{pmatrix} f(e) & f(x^{-1}) \\ f(x) & f(e) \end{pmatrix} \begin{pmatrix} 1 \\ \xi \end{pmatrix} = f(x^{-1})\xi + f(e)(1 + |\xi|^2) + f(x)\overline{\xi},$$

we have

$$\begin{aligned} 0 &= \operatorname{Im}(f(x^{-1})\xi + f(x)\overline{\xi}) \\ &= (\operatorname{Re} f(x^{-1}) - \operatorname{Re} f(x)) \operatorname{Im} \xi + (\operatorname{Im} f(x^{-1}) + \operatorname{Im} f(x)) \operatorname{Re} \xi \end{aligned}$$

for all $\xi \in \mathbb{C}$, so $\overline{f}(x) = f(x^{-1})$.

(b) and (c) are clear from definition.

(d) It follows from the Schur product theorem, which states that the Hadamard product(componentwise product) of two positive semi-definite matrices is also positive semi-definite.

(e) Let $f_1 = f$ and write

$$0 \le \begin{pmatrix} 1 & \overline{\xi} \end{pmatrix} \begin{pmatrix} f(e) & f(x^{-1}) \\ f(x) & f(e) \end{pmatrix} \begin{pmatrix} 1 \\ \xi \end{pmatrix} = f(e)(1 + |\xi|^2) + 2\,\mathrm{Re}(f(x)\overline{\xi}).$$

Taking $\xi = f(x)/|f(x)|$ if $f(x) \neq 0$, we obtain $|f(x)| \le f(e)$.

(f) The defining property of positive definite functions is conditioned by finitely many algebraic operations for each fixed n, $(x_1, \cdots, x_n)$, and $(\xi_1, \cdots, \xi_n)$, so the positive definiteness is preserved by pointwise limit.

(g) Let $f = f_1$ and write

$$\begin{aligned} 0 &\le \begin{pmatrix} 1 & \overline{\xi} & \overline{\eta} \end{pmatrix} \begin{pmatrix} f(e) & f(x^{-1}) & f(h^{-1}x^{-1}) \\ f(x) & f(e) & f(h^{-1}) \\ f(xh) & f(h) & f(e) \end{pmatrix} \begin{pmatrix} 1 \\ \xi \\ \eta \end{pmatrix} \\ &= f(e)(1 + |\xi|^2 + |\eta|^2) + 2\,\mathrm{Re}(f(x)\overline{\xi} + f(xh)\overline{\eta} + f(h)\xi\overline{\eta}). \end{aligned}$$

If $\eta = -\xi$, then

$$0 \le f(e) + 2(f(e) - \mathrm{Re}\, f(h))|\xi|^2 + 2\,\mathrm{Re}((f(x) - f(xh))\overline{\xi}).$$

Taking

$$\xi = \frac{1}{\varepsilon} \cdot \frac{f(xh) - f(x)}{|f(xh) - f(x)|}$$

for $\varepsilon > 0$ if $f(x) \neq f(xh)$, we obtain an inequality

$$|f(xh) - f(x)| \le \frac{\varepsilon}{2} f(e) + \frac{1}{\varepsilon}(f(e) - \mathrm{Re}\, f(h)),$$

so that we have

$$\limsup_{h \to e} \sup_{x \in G} |f(xh) - f(x)| \le \frac{\varepsilon}{2} f(e).$$

Since ε can be taken arbitrarily, f is right uniformly continuous. The left uniform continuity is shown in the same manner. □

Example 1.1. Let $G = \mathbb{R}$. Then, $f(x) := \cos x$ is positive definite since

$$\begin{aligned} \sum_{k,l=1}^{n} \cos(x_k - x_l)\xi_k\overline{\xi}_l &= \sum_{k,l=1}^{n} (\cos x_k \cos x_l + \sin x_k \sin x_l)\xi_k\overline{\xi}_l \\ &= \left|\sum_{k=1}^{n} \xi_k \cos x_k\right|^2 + \left|\sum_{k=1}^{n} \xi_k \sin x_k\right|^2 \ge 0. \end{aligned}$$

2 Bochner's theorem on $\mathbb{Z}$: complex analysis

Bochner's theorem is about the correspondence between positive definite functions and probability Borel measures. On the additive group $\mathbb{Z}$, positive definite functions become sequences, and the domain of probability measures is the one-dimensional torus $\mathbb{T} := \mathbb{R}/2\pi\mathbb{Z}$.

In this chapter, we will establish the following one-to-one correspondences:

$$\begin{tikzcd}
 & \left\{\begin{array}{c}\text{Points in the closed convex hull of}\\ \text{the curve } (e^{-i\theta}, e^{-i2\theta}, \cdots) \text{ in } \mathbb{C}^{\mathbb{N}}\end{array}\right\} \arrow[d, leftrightarrow, "2.1"] & \\
\left\{\begin{array}{c}\text{Positive definite}\\ \text{sequences } (c_k)_{k\in\mathbb{Z}}\\ \text{with } c_0 = 1\end{array}\right\} \arrow[r, leftrightarrow, "2.2"] & \left\{\text{Carathéodory functions}\right\} \arrow[r, leftrightarrow, "2.3"] & \left\{\begin{array}{c}\text{Probability Borel}\\ \text{measures on } \mathbb{T}\end{array}\right\}.
\end{tikzcd}$$

The vertical, left, and right arrows in the above diagram are discussed in Section 2.1, 2.2, and 2.3 respectively, and the definition of each term will be given throughout this chapter. Bochner's theorem on the additive group $\mathbb{Z}$ will be finally deduced as a corollary of the two horizontal correspondences in the above diagram.

2.1 The Carathéodory coefficient problem

We are going to investigate the origin of positive definiteness that occurs in the context of complex analysis. The concept of positive definiteness of functions was originally inspired by the "Carathéodory coefficient problem" in early complex analysis. The problem asks the condition on the power series coefficients for an analytic function defined on the open unit disk to have a positive real part. In other words, the Carathéodory coefficient problem describes the power series coefficients of some special functions precisely defined as follows:

Definition 2.1 (Carathéodory functions). The *Carathéodory class* is the set of all analytic functions f that map the open unit disk into the region of positive real part, with normalization condition $f(0) = 1$. A function in the Carathéodory class will be often called a *Carathéodory function*.

Example 2.1 (Möbius transforms). Typical examples of functions in the Carathéodory class are given by the family of functions

$$f_\theta(z) = \frac{e^{i\theta} + z}{e^{i\theta} - z} = 1 + \sum_{k=1}^{\infty} 2e^{-ik\theta} z^k$$

parametrized by $\theta \in [0, 2\pi)$. We can check that they are exactly the Möbius transformations that map the unit disk to the right half space having normalization $f(0) = 1$. This family of examples play a crucial role in the representation problem of functions in the Carathéodory class.

Example 2.2 (Convex combinations). Note that the Carathéodory class is convex; if f_0 and f_1 belong to the Carathéodory class, then the real part of the image of the function

$$f_t(z) = (1-t)f_0(z) + tf_1(z)$$

is also positive for $0 < t < 1$ and $f_t(0) = (1-t) + t = 1$, so f_t also belongs to the Carathéodory class.

Example 2.3 (Positive harmonic functions). Let f be in the Carathéodory class. By definition, the real part $\operatorname{Re} f : \mathbb{D} \to \mathbb{R}$ is a positive harmonic function such that $f(0) = 1$. Conversely, since there is a unique harmonic conjugate up to constant, we can recover f from its real part by letting $\operatorname{Im} f(0) = 0$. In other words, there is a one-to-one correspondence between the Carathódory class and the positive harmonic functions on the open unit disk that have value one at zero.

Carathéodory's result intuitively tells us that every function in the Carathéodory class can be constructed by convex combinations of Möbius transforms f_θ. As a result, they can be viewed as "extreme points" in the Carathéodory class. We will discuss the extreme points after the proof of the Carathéodory theorem.

Before the discussion, we develop a lemma as a preparation for the interplay between complex analysis and Fourier analysis.

Lemma 2.1 (Fourier coefficients of analytic functions)**.** *Let f be an analytic function on the open unit disk $\mathbb{D}$ with $f(0) \in \mathbb{R}$ with*

$$f(z) = c_0 + \sum_{k=1}^{\infty} 2c_k z^k,$$

the power series expansion of f at $z = 0$. Then, for $0 \leq r < 1$ and $k \in \mathbb{Z}$ we have

$$c_k r^{|k|} = \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} f(re^{i\theta}) e^{-ik\theta}\, d\theta,$$

where we use the notation $c_{-k} := \overline{c}_k$.

Proof. Suppose $k > 0$ first. The Cauchy integral formula writes

$$2c_k k! = \frac{\partial^k f}{\partial z^k}(0) = \frac{k!}{2\pi i} \int_{|z|=r} \frac{f(z)}{z^{k+1}}\, dz = \frac{k!}{2\pi i} \int_0^{2\pi} \frac{f(re^{i\theta})}{(re^{i\theta})^{k+1}}\, ire^{i\theta}\, d\theta,$$

and it implies

$$2c_k r^k = \frac{1}{2\pi} \int_0^{2\pi} f(re^{i\theta}) e^{-ik\theta}\, d\theta.$$

Since $f(z) z^k$ is analytic, the Cauchy theorem is applied to have

$$0 = \frac{1}{2\pi i} \int_{|z|=r} f(z) z^k\, dz = \frac{1}{2\pi} \int_0^{2\pi} f(re^{i\theta}) r^k e^{ik\theta}\, d\theta,$$

and it implies

$$0 = \frac{1}{2\pi} \int_0^{2\pi} \overline{f(re^{i\theta})} e^{-ik\theta}\, d\theta.$$

By combining the above equations, we obtain the formula. For $k = 0$, applying the Cauchy theorem for f, we have

$$c_0 = f(0) = \frac{1}{2\pi i} \int_{|z|=r} \frac{f(z)}{z}\, dz = \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} f(re^{i\theta})\, d\theta.$$

For $k < 0$, we can obtain the same formula by taking complex conjugation on the case $k > 0$.

Alternatively, we can show the same result using the orthogonal relation of complex exponential functions. Easy computation shows the identity

$$\begin{aligned}
\operatorname{Re} f(re^{i\theta}) &= \frac{1}{2}[f(re^{i\theta}) + \overline{f(re^{i\theta})}] \\
&= \frac{1}{2}\left[\left(1 + \sum_{k=1}^{\infty} 2c_k (re^{i\theta})^k\right) + \overline{\left(1 + \sum_{k=1}^{\infty} 2c_k (re^{i\theta})^k\right)}\right] \\
&= \frac{1}{2}\left[\left(1 + \sum_{k=1}^{\infty} 2c_k r^k e^{ik\theta}\right) + \left(1 + \sum_{k=1}^{\infty} 2\overline{c_k} r^k e^{-ik\theta}\right)\right] \\
&= \sum_{k=-\infty}^{\infty} c_k r^{|k|} e^{ik\theta}.
\end{aligned}$$

From the uniform convergence of the power series on the compact set $\{z : |z| \le (r+1)/2\}$ and the orthogonality

$$\frac{1}{2\pi}\int_0^{2\pi} e^{-ik\theta}e^{il\theta}\,d\theta = \begin{cases} 1 & \text{if } k = l \\ 0 & \text{if } k \neq l \end{cases},$$

it follows that

$$\frac{1}{2\pi}\int_0^{2\pi} \operatorname{Re} f(re^{i\theta})e^{-ik\theta}\,d\theta = \sum_{l=-\infty}^{\infty} c_l r^{|l|}\frac{1}{2\pi}\int_0^{2\pi} e^{il\theta}e^{-ik\theta}\,d\theta = c_k r^{|k|}. \qquad \square$$

Now, we prove the theorem. The original paper of Carathéodory deals with the functions analytic on a neighborhood of the closed unit disk, but the same idea can be extended well to the functions that may have harsh behavior on the boundary. Furthermore, by loosening the regularity requirements at the boundary, we can establish the exact description of Carathéodory functions in terms of their coefficients.

Theorem 2.2 (Carathéodory)**.** *Let f be an analytic function on the open unit disk with the power series expansion*

$$f(z) = 1 + \sum_{k=1}^{\infty} 2c_k z^k.$$

Then, f belongs to the Carathéodory class if and only if for each n the point $(c_1, \cdots, c_n) \in \mathbb{C}^n$ belongs to the convex hull of the curve $(e^{-i\theta}, \cdots, e^{-in\theta}) \in \mathbb{C}^n$ parametrized by $\theta \in [0, 2\pi)$.

Proof. ($\Leftarrow$) Denote by K_n the convex hull of the curve $\theta \mapsto (e^{-i\theta}, \cdots, e^{-in\theta}) \in \mathbb{C}^n$. Suppose first that $(c_1, \cdots, c_n) \in K_n$. For each n, there exists a finite sequence of pairs $(\lambda_{n,j}, \theta_{n,j})_j$ having the following convex combination

$$(c_1, \cdots, c_n) = \sum_j \lambda_{n,j}(e^{-i\theta_{n,j}}, \cdots, e^{-in\theta_{n,j}})$$

with coefficients $\lambda_{n,j} \ge 0$ such that $\sum_j \lambda_{n,j} = 1$. Define

$$f_n(z) := \sum_j \lambda_{n,j}\frac{e^{i\theta_{n,j}} + z}{e^{i\theta_{n,j}} - z},$$

which has positive real part on $|z| < 1$ because $\operatorname{Re}(e^{i\theta_{n,j}} + z)/(e^{i\theta_{n,j}} - z) > 0$ for $|z| < 1$. Then,

$$\begin{aligned} f_n(z) &= \sum_j \lambda_{n,j}(1 + \sum_{k=1}^{\infty} 2e^{-ik\theta_{n,j}}z^k) \\ &= 1 + \sum_{k=1}^{n} 2c_k z^k + \sum_{k=n+1}^{\infty}\left(\sum_j 2\lambda_{n,j}e^{-ik\theta_{n,j}}\right)z^k \end{aligned}$$

implies

$$\begin{aligned} |f_n(z) - f(z)| &= \left|\sum_{k=n+1}^{\infty}\left(\sum_j 2\lambda_{n,j}e^{-ik\theta_{n,j}}\right)z^k - \sum_{k=n+1}^{\infty} 2c_k z^k\right| \\ &\le \sum_{k=n+1}^{\infty}\left|\left(\sum_j 2\lambda_{n,j}e^{-ik\theta_{n,j}}\right) - 2c_k\right||z|^k \\ &\le \sum_{k=n+1}^{\infty} 4|z|^k \end{aligned}$$

converges to zero for $|z| < 1$. Therefore, f has a non-negative real part on the open unit disk. The non-negativity can be strengthened to positivity by the open mapping theorem so that f belongs to the Carathéodory class.

(⇒) Conversely, suppose that f is in the Carathéodory class. Let $(\gamma_1, \cdots, \gamma_n)$ be any point on the surface ∂K_n of K_n and S any supporting hyperplane of K_n tangent at $(\gamma_1, \cdots, \gamma_n)$. Let $(u_1, \cdots, u_n)$ be the outward unit normal vector of the supporting hyperplane S. Note that this unit normal vector is uniquely determined with respect to the induced real inner product structure on $2n$-dimensional space $\mathbb{C}^n$ described by

$$\langle (z_1, \cdots, z_n), (w_1, \cdots, w_n) \rangle = \sum_{k=1}^{n} (\operatorname{Re} z_k \operatorname{Re} w_k + \operatorname{Im} z_k \operatorname{Im} w_k) = \operatorname{Re} \sum_{k=1}^{n} z_k \overline{w}_k.$$

Then, $\sum_{k=1}^{n} |u_k|^2 = 1$ and further that the maximum

$$M := \max_{(x_1, \cdots, x_n) \in K_n} \operatorname{Re} \sum_{k=1}^{n} x_k \overline{u}_k > 0$$

is attained at $(\gamma_1, \cdots, \gamma_n)$. Our goal is to verify the bound

$$\operatorname{Re} \sum_{k=1}^{n} c_k \overline{u}_k \leq M,$$

which implies that $(c_1, \cdots, c_n)$ is contained in every half space tangent to K_n so that we finally obtain $(c_1, \cdots, c_n) \in K_n$.

Since for any $\theta \in [0, 2\pi)$ the point $(e^{-i\theta}, \cdots, e^{-in\theta})$ is in K_n so that

$$\operatorname{Re} \sum_{k=1}^{n} e^{-ik\theta} \overline{u}_k \leq M,$$

we have for arbitrarily small $\varepsilon > 0$ that

$$\operatorname{Re} \sum_{k=1}^{n} \frac{1}{r^k} e^{-ik\theta} \overline{u}_k \leq M + \varepsilon$$

for any $0 < r < 1$ sufficiently close to 1, thus we can write

$$\begin{aligned}
\operatorname{Re} \sum_{k=1}^{n} c_k \overline{u}_k &= \operatorname{Re} \sum_{k=1}^{n} \frac{1}{2\pi r^k} \int_0^{2\pi} \operatorname{Re} f(re^{i\theta}) e^{-ik\theta} \overline{u}_k \, d\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} f(re^{i\theta}) \operatorname{Re} \sum_{k=1}^{n} \frac{1}{r^k} e^{-ik\theta} \overline{u}_k \, d\theta \\
&\leq \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} f(re^{i\theta}) \, d\theta \cdot (M + \varepsilon) \\
&= M + \varepsilon
\end{aligned}$$

thanks to the positivity of $\operatorname{Re} f$, and by limiting $r \to 1$ from left we get the bound

$$\operatorname{Re} \sum_{k=1}^{n} c_k \overline{u}_k \leq M. \qquad \square$$

Here we introduce an infinite-dimensional version of this theorem.

Proposition 2.3. *Consider a sequence space $\mathbb{C}^{\mathbb{N}}$, endowed with the standard product topology. Then, the condition addressed in Caracthéodory's theorem is equivalent to the following: the point $(c_1, c_2, \cdots) \in \mathbb{C}^{\mathbb{N}}$ belongs to the closed convex hull of the curve $(e^{-i\theta}, e^{-i2\theta}, \cdots) \in \mathbb{C}^{\mathbb{N}}$ parametrized by $\theta \in [0, 2\pi)$.*

Furthermore, the curve $(e^{-i\theta}, e^{-i2\theta}, \cdots) \in \mathbb{C}^{\mathbb{N}}$ is the set of extreme points of its closed convex hull.

Proof. Denote by K_n the convex hull of the curve $\theta \mapsto (e^{-i\theta}, \cdots, e^{-in\theta}) \in \mathbb{C}^n$, and by K the closed convex hull of the curve $\theta \mapsto (e^{-i\theta}, e^{-i2\theta}, \cdots) \in \mathbb{C}^{\mathbb{N}}$. If we assume the Carathéodory coefficient condition is true, then since for each n we have a convex combination

$$(c_1, \cdots, c_n) = \sum_j \lambda_{n,j}(e^{-i\theta_{n,j}}, \cdots, e^{-in\theta_{n,j}})$$

with coefficients such that $\lambda_{n,j} \geq 0$ and $\sum_j \lambda_{n,j} = 1$, the sequence

$$\begin{aligned}(c_1, \cdots, c_n, &\sum_j \lambda_{n,j} e^{-i(n+1)\theta_{n,j}}, \sum_j \lambda_{n,j} e^{-i(n+2)\theta_{n,j}} \cdots) \\ &= \sum_j \lambda_{n,j}(e^{-i\theta_{n,j}}, \cdots, e^{-in\theta_{n,j}}, e^{-i(n+1)\theta_{n,j}}, e^{-i(n+2)\theta_{n,j}}, \cdots)\end{aligned}$$

indexed by n is contained in K and converges to the point $(c_1, c_2, \cdots)$ in the product topology as $n \to \infty$, so we arrive at the desired result. For the opposite direction, let $(c_1, c_2, \cdots) \in K$. By definition of K we have an expression

$$c_k = \lim_{m \to \infty} \sum_{j=1}^m \lambda_{m,j} e^{-ik\theta_{m,j}}$$

with $\lambda_{m,j} \geq 0$ and $\sum_{j=1}^m \lambda_{m,j} = 1$, for each k. Then,

$$(c_1, \cdots, c_n) = \lim_{m \to \infty} \sum_{j=1}^m \lambda_{m,j}(e^{-i\theta_{m,j}}, \cdots, e^{-in\theta_{m,j}})$$

belongs to K_n because K_n is closed.

We can also prove the proposition about extreme points using the Krein-Milman theorem and its converse. See Proposition 1.5 in [10] for the proof of the converse theorem of the Krein-Milman theorem. We will give an alternative proof without functional analysis in Section 2.3. □

2.2 Toeplitz's algebraic condition

Toeplitz discovered that the coefficient condition addressed in the Carathéodory's paper can be equivalently described in terms of an algebraic condition that the Hermitian matrices

$$H_n := (c_{k-l})_{k,l=1}^n = \begin{pmatrix} c_0 & c_{-1} & c_{-2} & \cdots & c_{-n+1} \\ c_1 & c_0 & c_{-1} & \cdots & c_{-n+2} \\ c_2 & c_1 & c_0 & \cdots & c_{-n+3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ c_{n-1} & c_{n-2} & c_{n-3} & \cdots & c_0 \end{pmatrix}$$

of size $n \times n$ always have non-negative determinants for any n. This algebraic condition is equivalent to the H_n being all positive semi-definite matrices. The principal minors of a positive semi-definite matrix are positive semi-definite, and a Hermitian matrix such that every leading principal minor has non-negative determinant is positive semi-definite. Threrfore, the bilateral sequence $(c_k)_{k=-\infty}^{\infty}$ is a positive definite function when we consider it as a complex-valued function on $\mathbb{Z}$ that maps an integer k to c_k if and only if it is a positive definite *sequence* in the following sense:

Definition 2.2. A bilateral complex sequence $(c_k)_{k=-\infty}^{\infty}$ is said to be *positive definite* if

$$\sum_{k,l=1}^n c_{k-l}\xi_k\overline{\xi}_l \geq 0$$

for each n and $(\xi_1, \cdots, \xi_n) \in \mathbb{C}^n$.

Theorem 2.4 (Carathéodory-Toeplitz). *Let f be an analytic function on the open unit disk with the power series expansion*

$$f(z) = 1 + \sum_{k=1}^{\infty} 2c_k z^k.$$

Then, f belongs to the Carathéodory class if and only if the sequence $(c_k)_{k=-\infty}^{\infty}$ is positive definite, where we use the notations $c_0 = 1$ and $c_{-k} = \overline{c_k}$.

Proof. ($\Rightarrow$) If f is in the Carathéodory class, then because

$$c_{k-l} r^{|k-l|} = \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} f(re^{i\theta}) e^{-i(k-l)\theta} \, d\theta,$$

we have

$$\sum_{k,l=1}^{n} c_{k-l} \xi_k \overline{\xi}_l = \lim_{r \uparrow 1} \frac{1}{2\pi} \int_0^{2\pi} \operatorname{Re} f(re^{i\theta}) \left| \sum_{k=1}^{n} e^{-ik\theta} \xi_k \right|^2 d\theta \geq 0$$

for each n.

($\Leftarrow$) Conversely, assume that the coefficient sequence $(c_k)_{k=-\infty}^{\infty}$ is positive definite. Put $\xi_k = z^{k-1}$ and $z = re^{i\theta}$ to write

$$\begin{aligned}
0 &\leq \sum_{k,l=1}^{n+1} c_{k-l} z^{k-1} (\overline{z})^{l-1} \\
&= \sum_{k,l=0}^{n} c_{k-l} r^{k+l} e^{i(k-l)\theta} \\
&= \sum_{k,l=0}^{n} c_{k-l} r^{|k-l|} r^{2\min\{k,l\}} e^{i(k-l)\theta} \\
&= \sum_{k=-n}^{n} c_k r^{|k|} e^{ik\theta} \sum_{l=0}^{n-|k|} r^{2l} \\
&= \sum_{k=-n}^{n} c_k r^{|k|} e^{ik\theta} \frac{1 - r^{2(n-|k|+1)}}{1 - r^2} \\
&= \frac{1}{1-r^2} \sum_{k=-n}^{n} c_k r^{|k|} e^{ik\theta} - \frac{r^{n+2}}{1-r^2} \sum_{k=-n}^{n} c_k r^{n-|k|} e^{ik\theta}.
\end{aligned}$$

For $r = |z| < 1$ the first term tends to

$$\lim_{n \to \infty} \frac{1}{1-r^2} \sum_{k=-n}^{n} c_k r^{|k|} e^{ik\theta} = \frac{1}{1-|z|^2} \operatorname{Re} f(z),$$

and $|c_k| \leq c_0 = 1$ implies the second term vanishes as

$$\left| \frac{r^{n+2}}{1-r^2} \sum_{k=-n}^{n} c_k r^{n-|k|} e^{ik\theta} \right| \leq \frac{r^{n+2}}{1-r^2} (2n+1) \to 0$$

as $n \to \infty$. It proves $\operatorname{Re} f(z) \geq 0$ for $|z| < 1$, and we obtain $\operatorname{Re} f(z) > 0$ by the open mapping theorem. □

2.3 Proof by the Herglotz representation theorem

Herglotz [5] proved another equivalent condition for the Carathéodory class in 1911, which states the correspondence between the Carathéodory class and probability Borel measure on the unit circle.

Nowadays it is considered as the first Bochner-type theorem. The Carathéodory theorem states that the function f in the Carathéodory class is a limit point of the set of convex combinations of Möbius transforms $z \mapsto (e^{i\theta}+z)/(e^{i\theta}-z)$. Herglotz's theorem, which we now also often call as the Herglotz representation theorem, states that in fact f can be directly represented by the integral of the Möbius transforms with respect to a certain probability measure.

The essential difficulty lies in the construction of a measure, and here we resolve this by applying either Helly's selection theorem or the Riesz-Markov-Kakutani representation theorem. Suppose the function f is analytic on a neighborhood of the closed unit disk $\overline{\mathbb{D}}$. In this case, by appropriately manipulating the identities for $r=1$ in Lemma 2.1, or by using the Cauchy integral formula along the unit circle, we can get

$$f(z) = \frac{1}{2\pi}\int_0^{2\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z} \operatorname{Re} f(e^{i\theta})\, d\theta.$$

Based on this representation of f, we will try to approximate the measure $d\mu$ with the absolutely continuous measures $(2\pi)^{-1}\operatorname{Re} f(re^{i\theta})\,d\theta$ by limiting $r \uparrow 1$. More precisely, we will use the following lemma:

Lemma 2.5. *Let f be an analytic function on the open unit disk. For $|z|<1$,*

$$f(z) = \lim_{r\uparrow 1}\frac{1}{2\pi}\int_0^{2\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z} \operatorname{Re} f(re^{i\theta})\, d\theta.$$

Proof. By the uniform convergence of the power series on the closed disk $\{z : |z| \le (r+1)/2\}$ for each fixed $r<1$, we have

$$\begin{aligned}
\lim_{r\uparrow 1}\frac{1}{2\pi}\int_0^{2\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z} \operatorname{Re} f(re^{i\theta})\, d\theta &= \lim_{r\uparrow 1}\frac{1}{2\pi}\int_0^{2\pi}\left(1+\sum_{k=1}^{\infty} 2e^{-ik\theta}z^k\right)\operatorname{Re} f(re^{i\theta})\,d\theta \\
&= 1+\lim_{r\uparrow 1}\sum_{k=1}^{\infty} 2\left(\frac{1}{2\pi}\int_0^{2\pi} e^{-ik\theta}\operatorname{Re} f(re^{-i\theta})\,d\theta\right)z^k \\
&= 1+\lim_{r\uparrow 1}\sum_{k=1}^{\infty} 2c_k r^k z^k \\
&= \lim_{r\uparrow} f(rz) = f(z).
\end{aligned}$$

□

Theorem 2.6 (The Herglotz representation theorem)**.** *Let f be a complex-valued function defined on the open unit disk. Then, f belongs to the Carathéodory class if and only if f is represented as the following Stieltjes integral*

$$f(z) = \int_0^{2\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z}\, d\mu(\theta),$$

where μ is a probability Borel measure on $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$.

First proof: using Helly's selection theorem. ($\Leftarrow$) Take a probability Borel measure μ on $\mathbb{T}$. Then, we can check the function defined by

$$f(z) := \int_0^{2\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z}\, d\mu(\theta)$$

is analytic on the open unit disk easily by using Morera's theorem and Fubini's theorem. Recall that $z \mapsto (e^{i\theta}+z)/(e^{i\theta}-z)$ has positive real part since it is a conformal mapping that maps the open unit disk onto the right half plane. The function f belongs to the Carathéodory class by the open mapping theorem since

$$\operatorname{Re} f(z) = \int_0^{2\pi} \operatorname{Re}\frac{e^{i\theta}+z}{e^{i\theta}-z}\, d\mu(\theta) \ge 0.$$

($\Rightarrow$) Fix a point z in the open unit disk $\mathbb{D}$. Define $f_n(\theta) := (2\pi)^{-1} \operatorname{Re} f((1-n^{-1})e^{i\theta})$ and

$$F_n(\theta) := \int_0^\theta \operatorname{Re} f_n(\psi)\, d\psi$$

for $\theta \in [0, 2\pi]$. Note $F_n(0) = 0$ and $F_n(2\pi) = 1$ for all n. Since $\operatorname{Re} f \geq 0$, F_n is also monotonically increasing. Therefore, the sequence $(F_n)_n$ has a pointwise convergent subsequence $(F_{n_j})_j$ on $[0, 2\pi]$ by Helly's selection theorem. Let

$$F(\theta) := \lim_{\psi \downarrow \theta} \lim_{j \to \infty} F_{n_j}(\psi).$$

Then, F is a distribution function such that $F(0) = 0$ and $F(2\pi) = 1$, and F_{n_j} converges to F at every continuity point θ of F. It means F_{n_j} converges to F weakly as $j \to \infty$, so by the Portmanteau theorem, we get

$$\int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} dF_{n_j}(\theta) \to \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} dF(\theta)$$

as $j \to \infty$ since $\theta \mapsto (e^{i\theta} + z)/(e^{i\theta} - z)$ is continuous and bounded on $\mathbb{T}$. On the other hand,

$$\int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} dF_{n_j}(\theta) = \frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} \operatorname{Re} f((1 - n_j^{-1})e^{i\theta})\, d\theta \to f(z)$$

as $j \to \infty$. Therefore, by comparing both limits, we can conclude that

$$f(z) = \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} dF(\theta) = \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(\theta),$$

where μ is the probability measure on $\mathbb{T}$ defined by the distribution function F as $\mu([0, \theta]) = F(\theta)$. $\square$

Second proof: using the Riesz representation theorem. As we have seen in the first proof that uses Helly's selection theorem, one direction is trivial. Suppose f is a Carathéodory function. Let $g \in C(\mathbb{T})$ be a complex-valued test function. Define a sequence of complex linear functionals l_n on $C(\mathbb{T})$ as

$$l_n[g] := \frac{1}{2\pi} \int_0^{2\pi} g(\theta) \operatorname{Re} f((1 - n^{-1})e^{i\theta})\, d\theta.$$

It is positive and bounded since $\operatorname{Re} f \geq 0$ and $\|l_r\| = l_r[1] = 1$. By the Alaoglu theorem, the sequence has $(l_n)_n$ a subsequence $(l_{n_j})_j$ that converges in the weak* topology of $C(\mathbb{T})^*$. If we let l be the limit, then $l[1] = \lim_{j \to \infty} l_{n_j}[1] = 1$ because $1 \in C(\mathbb{T})$. (Note that it is not valid if the domain space, $\mathbb{T}$ here, is not compact, and we will investigate this problem more carefully in the next chapter.)

By the Riesz-Markov-Kakutani representation theorem, there is a probability Borel measure μ on $\mathbb{T}$ such that

$$l[g] = \frac{1}{2\pi} \int_0^{2\pi} g(\theta)\, d\mu(\theta)$$

for all $g \in C(\mathbb{T})$. Then, for each fixed z in the open unit disk it follows from Lemma 2.5 that

$$\frac{1}{2\pi} \int_0^{2\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(\theta) = l[g_z] = \lim_{j \to \infty} l_{n_j}[g_z] = f(z)$$

since $g_z(\theta) := (e^{i\theta} + z)/(e^{i\theta} - z)$ belongs to $C(\mathbb{T})$. $\square$

As a corollary of Herglotz' theorem, we finally arrive at:

Corollary 2.7 (Bochner's theorem on $\mathbb{Z}$). *A function $c : \mathbb{Z} \to \mathbb{C}$ is positive-definite and $c_0 = 1$ if and only if there is a probability Borel measure μ on $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ such that*

$$c_k = \int_0^{2\pi} e^{-ik\theta}\, d\mu(\theta).$$

Proof. Let μ be a probability Borel measure on $\mathbb{T}$. Then, the sequence defined in the statement is positive definite because

$$\begin{aligned}\sum_{k,l=1}^{n} c_{k-l}\xi_k\overline{\xi}_l &= \sum_{k,l=1}^{n}\int_0^{2\pi} e^{-i(k-l)\theta}\,d\mu(\theta)\;\xi_k\overline{\xi}_l \\ &= \int_0^{2\pi}\left|\sum_{k=1}^{n} e^{-ik\theta}\xi_k\right|^2 d\mu(\theta)\geq 0\end{aligned}$$

for any $(\xi_1,\cdots,\xi_n)\in\mathbb{C}^n$, and $c_0=1$ is clear.

On the other hand, if the sequence $(c_k)_{k=-\infty}^{\infty}$ is positive definite and $c_0=1$, then the function $z\mapsto 1+\sum_{k=1}^{\infty}2c_kz^k$ is in the Carathéodory class. By the Herglotz representation theorem, there is a probability Borel measure μ on $\mathbb{T}$ such that

$$\begin{aligned}1+\sum_{k=1}^{\infty}2c_kz^k &= \int_0^{2\pi}\frac{e^{it}+z}{e^{it}-z}\,d\mu(t) \\ &= \int_0^{2\pi}\left(1+\sum_{k=1}^{\infty}2e^{-ik\theta}z^k\right)d\mu(t) \\ &= 1+\sum_{k=1}^{\infty}2\left(\int_0^{2\pi}e^{-ik\theta}\,d\mu(\theta)\right)z^k\end{aligned}$$

in $z\in\mathbb{D}$, hence the desired result follows. □

Herglotz' theorem assigns a probability measure μ to a Carathéodory function f by a weak* limit

$$\lim_{r\uparrow 1}\frac{1}{2\pi}\operatorname{Re} f(re^{i\theta})\,d\theta = d\mu.$$

This method allows us to construct measures using complex analytic functions. We now introduce several examples.

Example 2.4 (Dirac measures)**.** Identify $\mathbb{T}=\mathbb{R}/2\pi\mathbb{Z}$ with the interval $[0,2\pi)$. For each $\psi\in[0,2\pi)$, the Möbius transform $f_\psi(z)=(e^{i\psi}+z)/(e^{i\psi}-z)$ corresponds to the Dirac measure δ_ψ, defined as

$$\delta_\psi(E):=\begin{cases}1 & \text{, if } \psi\in E,\\ 0 & \text{, if } \psi\notin E\end{cases}$$

for Borel measurable $E\subset[0,2\pi)$. This is not only a direct consequence of the Herglotz representation theorem, but can also be viewed as a property of the Poisson kernel. Recall that the measure μ in the Herglotz theorem is constructed as the weak* limit of $(2\pi)^{-1}\operatorname{Re} f(re^{i\theta})\,d\theta$ with $r\uparrow 1$. The Poisson kernel is given as the real part of the Möbius transform

$$P_r(\psi-\theta)=\frac{1-r^2}{1-2r\cos(\theta-\psi)+r^2}=\operatorname{Re}\left(\frac{1+re^{i(\theta-\psi)}}{1-re^{i(\theta-\psi)}}\right)=\operatorname{Re} f_\psi(re^{i\theta}).$$

Since

$$\lim_{r\uparrow 1}\frac{1}{2\pi}\int g(\theta)P_r(\psi-\theta)\,d\theta = g(\psi)=\int g(\theta)\,d\delta_\psi(\theta)$$

for all $g\in C(\mathbb{T})$, we have $(2\pi)^{-1}\operatorname{Re} f(re^{i\theta})\,d\theta\to\delta_\psi$ in weak* topology of $C(\mathbb{T})^*$.

Example 2.5 (Continuous restrictions)**.** Let f be a Carathéodory function and $\tau:\mathbb{D}\to\mathbb{D}$ be an analytic function on the open unit disk $\mathbb{D}$. Then, the composition $f\circ\tau$ is Carathéodory.

Suppose we have an additional condition that τ can be continuously extended to $\tau:\overline{\mathbb{D}}\to\mathbb{D}$. The probability measure on $\mathbb{T}$ corresponded to the composition $f\circ\tau$ via the Herglotz theorem can be constructed as the weak* limit of $(2\pi)^{-1}\operatorname{Re} f(\tau(re^{i\theta}))\,d\theta$ as $r\uparrow 1$. Since $f\circ\tau$ is a continuous function on the closed disk $\overline{\mathbb{D}}$, the limit is described as the continuous density function $\mathbb{T}=\mathbb{R}/2\pi\mathbb{Z}\to\mathbb{R}:\theta\mapsto \operatorname{Re} f(\tau(e^{i\theta}))$.

Example 2.6 (The nth power map)**.** For a Carathéodory function f, we have a new family of functions in the Carathéodory class, the composition with the nth power map $z\mapsto f(z^n)$. If μ is a probability measure on $\mathbb{T}=\mathbb{R}/2\pi\mathbb{Z}$ that satisfies

$$f(z)=\int \frac{e^{i\theta}+z}{e^{i\theta}-z}\,d\mu(\theta),$$

then we have

$$f(z^n)=\int \frac{e^{i\theta}+z}{e^{i\theta}-z}\,d\mu_n(\theta)$$

for each positive integer n, where

$$\begin{aligned}\mu_n(E)&=\lim_{r\uparrow 1}\frac{1}{2\pi}\int_E \operatorname{Re} f(r^n e^{in\theta})\,d\theta\\ &=\lim_{r\uparrow 1}\sum_{j=0}^{n-1}\int_{(E-2\pi j/n)\cap[0,2\pi/n)}\operatorname{Re} f(r^n e^{in\theta})\,d\theta\\ &=\lim_{r\uparrow 1}\frac{1}{n}\sum_{j=0}^{n-1}\int_{(nE-2\pi j)\cap[0,2\pi)}\operatorname{Re} f(r^n e^{i\theta})\,d\theta\\ &=\frac{1}{n}\sum_{j=0}^{n-1}\mu((nE-2\pi j)\cap[0,2\pi)).\end{aligned}$$

If μ is absolutely continuous with respect to the Lebesgue measure, then the density of μ_n is the pull back by the kneading transformation

$$T_n(\theta):=n\theta-2\pi\left\lfloor\frac{n\theta}{2\pi}\right\rfloor.$$

The corresponding positive definite sequence is transformed from $(c_k)_{k\in\mathbb{Z}}$ to

$$(\cdots,0,c_{-2},0,\cdots,0,c_{-1},0,\cdots,0,c_0,0,\cdots,0,c_1,0,\cdots,0,c_2,0,\cdots),$$

where $n-1$ zeros are between c_k and c_{k+1}.

The rest of the proof of Proposition 2.3. Recall that K denotes the closed convex hull of the curve $\theta\mapsto(e^{-i\theta},e^{-i2\theta},\cdots)\in\mathbb{C}^{\mathbb{N}}$. We first claim that a point on this curve is an extreme point of K. Fix $\theta\in[0,2\pi)$ and suppose two complex sequences $(c_1,c_2,\cdots)$ and $(d_1,d_2,\cdots)$ in $\mathbb{C}^{\mathbb{N}}$ are contained in K and satisfy

$$\frac{c_k+d_k}{2}=e^{-ik\theta}$$

for all $k\in\mathbb{N}$. For each k, since all components of K are bounded by one so that $|c_k|\le 1$ and $|d_k|\le 1$, and since $e^{-ik\theta}$ is an extreme point of the closed unit disk $\overline{\mathbb{D}}\subset\mathbb{C}$, we have $c_k=d_k=e^{-ik\theta}$, we deduce the desired claim.

For the converse, take a point $(c_1,c_2,\cdots)$ in K such that no θ satisfies $c_k=e^{-ik\theta}$ for all $k\in\mathbb{N}$. As we have seen, there is a probability Borel measure μ on $\mathbb{T}$ that corresponds to $(c_1,c_2,\cdots)$. Since μ is not a Dirac measure, the support of μ contains at least two points. Partition the support of μ into two non-trivial subsets A and B. Then, for two measures μ_A and μ_B given by $\mu_A(E):=\mu(E\cap A)/\mu(A)$ and $\mu_B(E):=\mu(E\cap B)/\mu(B)$ for Borel sets $E\subset\mathbb{T}$, the measure μ is a non-trivial convex combination of μ_A and μ_B. By paraphrasing this fact in terms of the positive definite sequences, we can see that $(c_1,c_2,\cdots)$ is not an extreme point. $\square$

3 Bochner's theorem on $\mathbb{R}$: probability theory

In this chapter, we prove the Bochner theorem on the additive group $\mathbb{R}$ using Lévy's continuity theorem. The Lévy continuity theorem relates the *weak convergence* of probability measures and the pointwise convergence of positive definite functions. Before we provide the statements for the theorems, we first introduce several fundamental theorems about the topology of weak convergence, such as the Portmanteau theorem, theorems on the Lévy Prokhorov metric, and the Prokhorov theorem on the compactness in the space of probability measures.

It is known that the systematic study of positive definite functions to investigate the convergence of measures began in probability theory. A celebrated research was done in the book [6] by Paul Lévy. Recall that a probability distribution is defined as a measure of norm one on a "state space", which is $\mathbb{R}$ for usual random variables. Some classical problems including central limit theorems and laws of large numbers that arise in probability theory describe limit behaviors of a sequence of probability distributions. The Lévy continuity theorem tells us that it is easier to see the limits via the *Fourier transforms* of probability measures, than to see the measures directly.

3.1 Topologies on the space of probability measures

First, we will investigate topologies on the space of probability measures. In probability theory, the topology of weak convergence is the most often given when considering convergence of probability measures.

Definition 3.1 (Weak convergence). Let $(\mu_\alpha)_\alpha$ be a net of probability Borel measures on a topological space S. We say μ_α *converges weakly* to another probability Borel measure μ if

$$\int g\, d\mu_\alpha \to \int g\, d\mu$$

for any $g \in C_b(S)$, where $C_b(S)$ denotes the space of continuous and bounded functions. We often write $\mu_\alpha \Rightarrow \mu$ when μ_α converges weakly to μ.

In fact, for its own interests in probability theory, the state space S is usually taken to be $\mathbb{R}$, or more generally a metrizable space. However, we temporarily define the weak convergence in the meaningless general setting, the topological spaces, to further compare with another topology on the space of measures. Some reasons why we require the metrizability of S will be addressed later.

Vague convergence is another convergence that reveals a more functional analytic nature of measures. Recall that the Riesz-Markov-Kakutani representation theorem states that on a locally compact Hausdorff space the space of regular Borel finite (complex) measures has a natural identification to the continuous dual of the space of continuous functions vanishing at infinity.

Definition 3.2 (Vague convergence). Let $(\mu_\alpha)_\alpha$ and μ be probability regular Borel measures on a locally compact Hausdorff space Ω. We say μ_α *converges vaguely* to another probability regular Borel measures μ if

$$\int g\, d\mu_\alpha \to \int g\, d\mu$$

for any $g \in C_0(\Omega)$, where $C_0(\Omega)$ denotes the space of continuous functions vanishing at infinity. By the Riesz-Markov-Kakutani representation theorem, the topology of vague convergence coincides with the weak* topology of the dual space $C_0(\Omega)^*$.

We warn that the *regular* Borel measures in the Riesz-Markov-Kakutani representation theorem for locally compact Hausdorff spaces are different from what we usually define as *regular* Borel measures in probability theory. For the convenience of further discussions, here we clarify the concept of regular measures.

Definition 3.3 (Regular Borel measures)**.** If Ω is locally compact and Hausdorff, then we say a Borel measure μ on Ω is *regular* if

$$\mu(E) = \sup\{\,\mu(K) : K \text{ is compact in } E\,\} = \inf\{\,\mu(U) : U \text{ is open containing } E\,\}$$

for all Borel measurable E. If S is metrizable, then we say a Borel measure μ on S is *regular* if

$$\mu(E) = \sup\{\,\mu(F) : F \text{ is closed in } E\,\} = \inf\{\,\mu(U) : U \text{ is open containing } E\,\}$$

for all Borel measurable E. Note that even if a topological space is both locally compact Hausdorff and metrizable, the two notions are not equivalent — we avert possible confusion by mentioning which type the underlying space is. We denote the space of all probability regular Borel measures on Ω and S by $\mathrm{Prob}(\Omega)$ and $\mathrm{Prob}(S)$, respectively.

Lemma 3.1 (Probability measure is regular on metrizable spaces)**.** *Let S be a metrizable space. Then, every finite Borel measure μ on S is regular.*

We have omitted the regularity condition on probability measures on $\mathbb{T}$ in Chapter 2 because every finite Borel measure on a compact metric space is regular in both senses.

Vague convergence is less important in probability theory because there are situations that we have to deal with probability measures on nowhere locally compact spaces, for example, the separable Hilbert space or the space of continuous functions $C([0,1])$. This viewpoint frequently occurs and is useful when we try to analyze a stochastic process as a single random variable.

Nevertheless, vague convergence is what we will mainly consider throughout Chapter 4. Recall that we also have used weak* topology as well in Chapter 2. In this regard, we need to connect vauge convergence to weak convergence to describe our subjects in probabilistic languages: the following theorem is one such result.

Theorem 3.2. *Let Ω be a locally compact Hausdorff space. The topology of weak convergence and the topology of vague convergence are identical in* $\mathrm{Prob}(\Omega)$*, the space of probability regular Borel measures on Ω.*

Note that the topology of weak and vague convergence is the topology generated by the family of subsets

$$U_{\mu,\varepsilon,g} := \{\,\nu : |\textstyle\int g\,d\mu - \int g\,d\nu| < \varepsilon\,\},$$

where $\mu \in \mathrm{Prob}(\Omega)$, $\varepsilon > 0$, and g is contained in $C_b(\Omega)$ and $C_0(\Omega)$ respectively. The topologies are not sequential in general, we must prove it using nets.

Proof. One direction is clear, since the topology of vague convergence is coarser than the topology of weak convergence. For the opposite, let $(\mu_\alpha)_\alpha$ be a net in $\mathrm{Prob}(\Omega)$ that converges vaguely to $\mu \in \mathrm{Prob}(\Omega)$, and take $g \in C_b(\Omega)$. Since $\mu(\Omega) = \|\mu\| = 1$, there is $\varphi \in C_0(\Omega)$ such that $\|\varphi\| = 1$ and $\int \varphi\,d\mu > 1-\varepsilon$. We may assume $\varphi \geq 0$ without loss of generality by taking $\max\{\varphi, 0\}$. Then, since $g\varphi$ vanishes at infinity and $\int \varphi\,d\mu_\alpha$ converges to $\int \varphi\,d\mu$, we have

$$|\int g\,d\mu_\alpha - \int g\,d\mu| \leq |\int g\varphi\,d\mu_\alpha - \int g\varphi\,d\mu| + \|g\| \int (1-\varphi)\,d(\mu_\alpha + \mu)$$

so that

$$\limsup_\alpha |\int g\,d\mu_\alpha - \int g\,d\mu| \leq 2\|g\|\varepsilon$$

for arbitrary $\varepsilon > 0$. Therefore, we have the weak convergence of μ_α to μ. $\square$

Example 3.1 (Escaping to the infinity)**.** The two topologies are different if we consider the space of finite measures or measures bounded by one, instead of the space of probability measures. A terse example is the shifting sequence of dirac measures δ_n, which converges to the zero measure in the topology generated by C_0, but diverges in the topology generated by C_b.

According to this result, under the assumption that the base space is locally compact and Hausdorff, we do not have to distinguish the topology of weak and vague convergence. Now we return to probability theory. Two classical theorems of the space of probability measures on a metric space, the metrizability and a compactness criteria for the space of probability measures will be introduced. They will be applied to see weak* convergences of probability measures on $\mathbb{R}$, and it is doable because $\mathbb{R}$ is both a metric space and a locally compact space.

We are now going to see some useful theorems on weak convergence of probability measures. For a deeper discussion on the topology of weak convergence, see the textbook of Parthasarathy [9].

Lemma 3.3 (The Portmanteau theorem)**.** *Let S be a metric space, and μ_α be a net of probability Borel measures on S. The following statements are all equivalent:*

(a) $\int g\,d\mu_\alpha \to \int g\,d\mu$ *for every* $g \in C_b(S)$*, i.e. weakly convergent.*

(b) $\int g\,d\mu_\alpha \to \int g\,d\mu$ *for every uniformly continuous* $g \in C_b(S)$*.*

(c) $\limsup_\alpha \mu_\alpha(F) \le \mu(F)$ *for every closed* $F \subset S$*.*

(d) $\liminf_\alpha \mu_\alpha(U) \ge \mu(U)$ *for every open* $U \subset S$*.*

(e) $\lim_\alpha \mu_\alpha(A) = \mu(A)$ *for every Borel set* $A \subset S$ *such that* $\mu(\partial A) = 0$*.*

Proof. (a)⇒(b) Clear.

(b)⇒(c) Let U be an open set such that $F \subset U$. There is uniformly continuous $g \in C_b(S)$ such that $\mathbf{1}_F \le g \le \mathbf{1}_U$. Therefore,

$$\limsup_\alpha \mu_\alpha(F) \le \limsup_\alpha \mu_\alpha(g) = \mu(g) \le \mu(U).$$

By the outer regularity of μ, we obtain $\limsup_\alpha \mu_\alpha(F) \le \mu(F)$.

(c)⇔(d) Clear by taking complements.

(c)+(d)⇒(e) It easily follows from

$$\limsup_\alpha \mu_\alpha(\overline{A}) \le \mu(\overline{A}) = \mu(A) = \mu(A^\circ) \le \liminf_\alpha \mu_\alpha(A^\circ).$$

(e)⇒(a) Let $g \in C_b(S)$ and $\varepsilon > 0$. Since the pushforward measure $g_*\mu$ has at most countably many mass points, there is a partition $(t_i)_{i=0}^n$ of an interval containing $[-\|g\|, \|g\|]$ such that $|t_{i+1} - t_i| < \varepsilon$ and $\mu(\{x : g(x) = t_i\}) = 0$ for each i. Let $(A_i)_{i=0}^{n-1}$ be a Borel decomposition of S given by $A_i := g^{-1}([t_i, t_{i+1}))$, and define $f_\varepsilon := \sum_{i=0}^{n-1} t_i \mathbf{1}_{A_i}$ so that we have $\sup_{x\in S} |g_\varepsilon(x) - g(x)| \le \varepsilon$. From

$$\begin{aligned}|\mu_\alpha(g) - \mu(g)| &\le |\mu_\alpha(g - g_\varepsilon)| + |\mu_\alpha(g_\varepsilon) - \mu(g_\varepsilon)| + |\mu(g_\varepsilon - g)| \\ &\le \varepsilon + \sum_{i=0}^{n-1} |t_i||\mu_\alpha(A_i) - \mu(A_i)| + \varepsilon,\end{aligned}$$

we get

$$\limsup_\alpha |\mu_\alpha(g) - \mu(g)| < 2\varepsilon.$$

Since ε is arbitrary, we are done. □

Theorem 3.4 (Lévy-Prokhorov metric)**.** *Let (S, d) be a metric space, and* $\mathrm{Prob}(S)$ *be the set of probability Borel measures on S. Denote by $\mathcal{B}(S)$ the σ-algebra of all Borel sets. Define a function* $\pi : \mathrm{Prob}(S) \times \mathrm{Prob}(S) \to [0, \infty)$ *such that*

$$\pi(\mu, \nu) := \inf\{\, \varepsilon > 0 : \mu(E) \le \nu(E^\varepsilon) + \varepsilon,\ \nu(E) \le \mu(E^\varepsilon) + \varepsilon,\ \forall E \in \mathcal{B}(S) \,\},$$

where E^ε denotes the ε-neighborhood of a, $E^\varepsilon := \bigcup_{x\in E} B(x,\varepsilon)$. The set in the definition of π contains $\varepsilon = 1$ so that it is always non-empty.

(a) *The function π is a metric.*

(b) *For a sequence $\mu_n \in \mathrm{Prob}(S)$, if $\mu_n \to \mu$ in π, then $\mu_n \Rightarrow \mu$.*

(c) *For a net $\mu_\alpha \in \mathrm{Prob}(S)$, if $\mu_\alpha \Rightarrow \mu$, then $\mu_\alpha \to \mu$ in π, given S is separable.*

(d) *The metric space $(\mathrm{Prob}(S), \pi)$ is separable if and only if (S, d) is separable.*

Proof. (a) We will only prove the two non-trivial parts: non-degeneracy and triangle inequality. Let $d(\mu, \nu) = 0$ so that there is a sequence $\varepsilon_n \downarrow 0$ such that for every Borel E we have

$$\mu(E) \le \nu(E^{\varepsilon_n}) + \varepsilon_n, \quad \nu(E) \le \mu(E^{\varepsilon_n}) + \varepsilon_n,$$

Taking limit $n \to 0$, we obtain

$$\mu(E) \le \nu(\overline{E}), \quad \nu(E) \le \mu(\overline{E})$$

for all Borel sets E. Thus $\mu(F) = \nu(F)$ for all closed F, and the inner regularity proves $\mu = \nu$. For the triangle inequality, take $\mu, \nu, \lambda \in \mathrm{Prob}(S)$. Take sequences $a_n \downarrow d(\mu,\lambda)$ and $b_n \downarrow d(\lambda, \nu)$ such that

$$\mu(E) \le \lambda(E^{a_n}) + a_n \le \nu((E^{a_n})^{b_n}) + a_n + b_n \le \nu(E^{a_n+b_n}) + a_n + b_n$$

and

$$\nu(E) \le \lambda(E^{b_n}) + b_n \le \mu((E^{b_n})^{a_n}) + a_n + b_n \le \mu(E^{a_n+b_n}) + a_n + b_n$$

for all Borel sets E. Taking limit $n \to \infty$ we get $d(\mu, \nu) \le \inf_n(a_n + b_n) = d(\mu, \lambda) + d(\lambda, \nu)$.

(b) Take $\varepsilon_n \downarrow 0$ such that $\mu_n(E) \le \mu(E^{\varepsilon_n}) + \varepsilon_n$ for every Borel E, which deduces $\limsup_{n\to\infty} \mu_n(F) \le \mu(F)$ for every closed F. Therefore, $\mu_n \Rightarrow \mu$ by the Portmanteau theorem.

(c) Let E be Borel and fix $\varepsilon > 0$. Note that since an open interval is uncountable, there is r in the interval such that $\mu(\partial B(x, r)) = 0$ for any point $x \in S$ because uncountable sums of positive numbers always diverge to infinity. If $\{x_i\}_{i=1}^\infty$ is dense in S, then

$$S = \bigcup_{i=1}^\infty B(x_i, \varepsilon_i)$$

for some $\varepsilon_i \in (\varepsilon/4, \varepsilon/2)$ such that $\mu(\partial B(x_i, \varepsilon_i)) = 0$. Define

$$B := \Big(\bigcup_{i=1}^n B(x_i, \varepsilon_i)\Big)^c$$

for sufficiently large n such that $\mu(B) < \varepsilon/3$. Define A to be the union of all $B(x_i, \varepsilon_i)$ such that $1 \le i \le n$ and $B(x_i, \varepsilon_i) \cap E \ne \varnothing$. Then, $E \subset A \cup B$ and $A \subset E^\varepsilon$ since $\varepsilon_i < \varepsilon/2$.

Since $\mu(\partial B(x_i, \varepsilon_i)) = 0$ for all i, we have $\mu(\partial A) = 0$ and $\mu(\partial B) = \mu(\partial(B^c)) = 0$, we can take α_0 by the Portmanteau theorem such that $\alpha \succ \alpha_0$ implies

$$\max\{\,|\mu_\alpha(A) - \mu(A)|, |\mu_\alpha(B) - \mu(B)|\,\} < \frac{\varepsilon}{3}.$$

Then, $d(\mu_\alpha, \mu) \le \varepsilon$ for all $\alpha \succ \alpha_0$ since

$$\mu(E) \le \mu(A) + \mu(B) \le \mu(A) + \frac{1}{3}\varepsilon \le \mu_\alpha(A) + \frac{2}{3}\varepsilon < \mu(E^\varepsilon) + \varepsilon$$

and

$$\mu_\alpha(E) \le \mu_\alpha(A) + \mu_\alpha(B) \le \mu_\alpha(A) + \frac{2}{3}\varepsilon \le \mu(A) + \varepsilon \le \mu(E^\varepsilon) + \varepsilon.$$

(d) Let $\{x_i\}_{i=1}^{\infty}$ be dense in S. We want to show

$$\mathcal{M} := \left\{ \begin{array}{c} \text{rational coefficient convex combination of} \\ \text{Dirac measures } \delta_{x_i} \end{array} \right\}$$

is dense in $\mathrm{Prob}(S)$. Let $\mu \in \mathrm{Prob}(S)$ and suppose $g \in C_b(S)$ is uniformly continuous so that for fixed $\varepsilon > 0$ we can take $\delta > 0$ such that $|x-y| < \delta$ implies $|g(x)-g(y)| < \varepsilon/4$. Since $S = \bigcup_{i=1}^{\infty} B(x_i, \delta)$, we can have a partition $\{A_1, \cdots, A_n, B\}$ of S such that $A := B(x_i, \delta) \setminus A_{i-1}$ and $\mu(B) < \varepsilon/8\|g\|$. Take any $y \in B$.

Define $\nu \in \mathcal{M}$ such that

$$\nu := \sum_{i=1}^{n} (\mu(A_i) + \varepsilon_i)\delta_{x_i} + (\mu(B) - \sum_{i=1}^{n} \varepsilon_i)\delta_y,$$

with perturbations ε_i such that $\mu(A_i) + \varepsilon_i \in \mathbb{Q}$ and $\sum_{i=1}^{n} |\varepsilon_i| < \varepsilon/4$. The measure $\nu \in \mathcal{M}$ depends on ε. Then,

$$\begin{aligned}
|\int g\, d\nu - \int g\, d\mu| &\le \sum_{i=1}^{n} |\int_{A_i} g\, d\nu - \int_{A_i} g\, d\mu| + |\int_B g\, d\nu - \int_B g\, d\mu| \\
&\le \sum_{i=1}^{n} \int_{A_i} |g(x_i) - g(x)| d\mu(x) + \int_B |g(y) - g(x)|\, d\mu(x) + \frac{\varepsilon}{2} \\
&\le \sum_{i=1}^{n} \int_{A_i} \frac{\varepsilon}{4}\, d\mu + \frac{\varepsilon}{8M} 2M + \frac{\varepsilon}{2} < \varepsilon.
\end{aligned}$$

Therefore, $\mathcal{M}$ is dense in $\mathrm{Prob}(S)$. □

Definition 3.4 (Polish spaces)**.** A topological space X is called *Polish* if it is homeomorphic to a complete separable metric space.

Polish spaces are measure-theoretically well-behaved topological spaces that occur as one of the most fundamental assumptions in probability theory. The above theorem about the Prokhorov metric states that if S is Polish then so is $\mathrm{Prob}(S)$. The importance of Polish spaces can be found in several theorems such as the Prokhorov theorem and the Kolmogorov extension theorem.

The Prokhorov theorem is a compactness theorem, and will be critically used to construct a limit of a sequence of measures. *Tightness* is the measure-theoretic paraphrase of the compactness in the probability measure space according to the Prokhorov theorem.

Definition 3.5 (Tight measures)**.** Let M be a set of probability Borel measures on a metric space S. We say M is *tight* if for every $\varepsilon > 0$ there is a compact $K \subset S$ such that $\mu(K) > 1 - \varepsilon$ for all $\mu \in M$

Theorem 3.5 (The Prokhorov theorem)**.** *Let M be a subset of* $\mathrm{Prob}(S)$ *for a Polish space S. The set M is relatively compact in the topology of weak convergence if and only if it is tight.*

Proof. ($\Rightarrow$) Suppose M is relatively compact. We first claim that for a given countable open cover $\{U_i\}_{i=1}^{\infty}$ of S and for each $\varepsilon > 0$ we can find n such that

$$\inf_{\mu \in M} \mu\Big(\bigcup_{i=1}^{n} U_i\Big) \ge 1 - \varepsilon.$$

Assume that it is not true so that there is a sequence $\mu_n \in M$ such that

$$\mu_n\Big(\bigcup_{i=1}^{n} U_i\Big) < 1 - \varepsilon.$$

If we take a subsequence $(\mu_{n_k})_k$ that converges weakly to $\mu \in \overline{M}$ using the compactness of $\overline{M}$, then by the Portmanteau theorem we have

$$\mu\Big(\bigcup_{i=1}^{n} U_i\Big) \leq \liminf_{k\to\infty} \mu_{n_k}\Big(\bigcup_{i=1}^{n} U_i\Big) \leq \liminf_{k\to\infty} \mu_{n_k}\Big(\bigcup_{i=1}^{n_k} U_i\Big) \leq 1-\varepsilon,$$

which leads to a contradiction $\mu(S) \leq 1-\varepsilon$.

Let $\{x_i\}_{i=1}^{\infty}$ be a dense set in S. Then, since $\{B(x_i,1/m)\}_{i=1}^{n}$ is a countable open cover of S for each integer $m>0$, there is $n_m>0$ such that

$$\inf_{\mu\in M} \mu\Big(\bigcup_{i=1}^{n_m} B(x_i,1/m)\Big) \geq 1-\frac{\varepsilon}{2^m}.$$

Define

$$K := \bigcap_{m=1}^{\infty} \bigcup_{i=1}^{n_m} \overline{B(x_i,1/m)}.$$

It is clearly closed in a complete metric space $\mathrm{Prob}(S)$, and is totally bounded since for any $\varepsilon>0$ we have $K \subset \bigcup_{i=1}^{n_m} B(x_i,\varepsilon)$ if m satisfies $1/m<\varepsilon$, so K is compact. Moreover, we can verify

$$1-\mu(K) = \mu\Big(\bigcup_{m=1}^{\infty} \bigcap_{i=1}^{n_m} \overline{B(x_i,1/m)}^c\Big) \leq \sum_{m=1}^{\infty}\left(1-\mu\Big(\bigcup_{i=1}^{n_m} B(x_i,1/m)\Big)\right) \leq \varepsilon$$

for every $\mu\in M$, so M is tight.

($\Leftarrow$) Suppose M is tight and let μ_α be any net in M. We claim that it has a convergent subnet in $\mathrm{Prob}(S)$. Let βS be the Stone-Čech compactification of S. The inclusion $\iota : S \to \beta S$ is a topological embedding because S is completely regular. Pushforward the measures μ_α to make them probability Borel measures $\nu_\alpha := \iota_*\mu_\alpha$ on βS. We want to take a convergent subnet of $\nu_\alpha \in \mathrm{Prob}(\beta S)$, and to show the limit is in fact contained in $\mathrm{Prob}(S)$.

Our first claim is that the measure ν_α is regular for each α, that is, $\nu_\alpha \in \mathrm{Prob}(\beta S)$. For any Borel $E \subset \beta S$ and any $\varepsilon>0$, there is $F \subset E\cap S$ that is closed in S such that $\mu_\alpha(E\cap S) < \mu_\alpha(F)+\varepsilon/2$ by inner regularity, and there is K that is compact in S such that $\mu_\alpha(S\setminus K)<\varepsilon/2$ by tightness. Then, the inequality

$$\nu_\alpha(E) = \mu_\alpha(E\cap S) < \mu_\alpha(F)+\frac{\varepsilon}{2} < \mu_\alpha(F\cap K)+\varepsilon = \nu_\alpha(F\cap K)+\varepsilon$$

proves the regularity of ν_α since $F\cap K$ is compact in both S and βS with $F\cap K\subset E$. The space $\mathrm{Prob}(\beta S)$ is compact by the Banach-Alaoglu theorem and the Riesz-Markov-Kakutani representation theorem. Therefore, ν_α has a subnet ν_β that converges to $\nu\in\mathrm{Prob}(\beta S)$.

Recall that μ_β is tight. For each $\varepsilon>0$, there is a compact $K\subset S$ such that $\nu_\beta(K)=\mu_\beta(K)\geq 1-\varepsilon$ for all β. Then, by the Portmanteau theorem, we have

$$\nu(S) \geq \nu(K) \geq \limsup_{\beta} \nu_\beta(K) \geq 1-\varepsilon.$$

Since ε is arbitrary, ν is concentrated on S, i.e. $\nu(S)=1$. Now we restrict ν to S in order to obtain μ, which is a probability Borel measure on S.

From the definition of weak convergence we have

$$\int_{\beta S} f\,d\nu_\beta \to \int_{\beta S} f\,d\nu$$

for all $f\in C(\beta S)$. Since $\nu_\beta(\beta S\setminus S)=\nu(\beta S\setminus S)=0$ and the restriction $C(\beta S)\to C_b(S)$ is an isomorphism due to the universal property of βS,

$$\int_S f\,d\mu_\beta \to \int_S f\,d\mu$$

for all $f\in C_b(S)$, so μ_β converges weakly to $\mu\in\mathrm{Prob}(S)$. □

In this proof of the theorem, we can add a new interpretation of tightness; any limit measures defined on βS must be concentrated on the original state space S. The tightness keeps measures from escaping the image of S in the compactification, and lets the limit point be concentrated on it. We can also recognize the topology of weak convergence as the induced topology from the Stone-Čech compactification.

3.2 Proof by the Lévy continuity theorem

In this section, we will only focus on probability distributions on the real line $\mathbb{R}$ and concrete examples on it, rather than other abstract spaces S. One of the direct connections in probability theory between convergences in two different realms, measures and positive definite functions, is encoded in the Lévy continuity theorem. This theorem connects the weak convergence of probability measures and pointwise convergence of *characteristic functions*. In this section, we will prove Bochner's theorem on $\mathbb{R}$ with the aid of the Lévy continuity theorem.

A characteristic function is defined as the Fourier transform of a probability measure, with reversed sign on the phase term. Characteristic functions have an advantage that we can learn the information about probability measures by studying continuous functions instead of the measures themselves.

Definition 3.6 (Characteristic functions)**.** Let μ be a probability Borel measure on $\mathbb{R}$. The *characteristic function* of μ is a function $\varphi : \mathbb{R} \to \mathbb{C}$ defined by

$$\varphi(t) := \int e^{itx}\, d\mu(x).$$

Equivalently, if μ is the distribution of a random variable X, then $\varphi(t) = Ee^{itX}$.

Proposition 3.6. *Let φ be a characteristic function of a probability Borel measure μ on $\mathbb{R}$. Then, φ is positive definite and uniformly continuous.*

Proof. It follows clearly that

$$\sum_{k,l=1}^{n} \varphi(t_k - t_l)\xi_k\overline{\xi}_l = \int \left| \sum_{k=1}^{n} e^{it_k x}\xi_k \right|^2 d\mu(x) \geq 0,$$

and

$$|\varphi(t) - \varphi(s)| \leq \int |e^{itx} - e^{its}|\, d\mu(x) \leq |t - s|. \qquad \square$$

Example 3.2. Many continuous positive definite functions are computed from probability distributions:

Name	mass or density functions	characteristic functions
Constant	$p(x) = \mathbf{1}_{\{c\}}(x)$	$\varphi(t) = e^{ict}$
Bernoulli	$p(x) = \frac{1}{2} \cdot \mathbf{1}_{\{\pm 1\}}(x)$	$\varphi(t) = \cos t$
Normal	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$	$\varphi(t) = e^{-t^2/2}$
Uniform	$f(x) = \frac{1}{2} \cdot \mathbf{1}_{[-1,1]}(x)$	$\varphi(t) = \operatorname{sinc} t$
Exponential	$f(x) = e^{-x} \cdot \mathbf{1}_{[0,\infty)}(x)$	$\varphi(t) = (1 - it)^{-1}$
Cauchy	$f(x) = 1/\pi(1 + x^2)$	$\varphi(t) = e^{-\lvert t \rvert}$
Polya	$f(x) = (1 - \cos x)/\pi x^2$	$\varphi(t) = \max\{1 - \lvert t \rvert, 0\}$

In the proof of the continuity theorem, by characteristic functions, we will show the tightness of associated probability measures to see their weak convergence. To verify that a family of probability measures is tight, their tail probabilities ought to be uniformly controlled. The following lemma is useful in bounding tail probabilities in terms of characteristic functions; the averaging of $1-\varphi$ near zero provides a reasonable estimate of the tail probability.

Lemma 3.7. *Let μ be a probability Borel measure on $\mathbb{R}$ and φ be its characteristic function. Then,*

$$\mu([-\tfrac{2}{\delta},\tfrac{2}{\delta}]^c) \le 2\cdot\frac{1}{2\delta}\int_{-\delta}^{\delta}(1-\varphi(t))\,dt$$

for any $\delta > 0$. In particular, a single measure is tight.

Proof. Write the average with the sinc function as

$$\begin{aligned}\frac{1}{2\delta}\int_{-\delta}^{\delta}\varphi(t)\,dt &= \int \frac{1}{2\delta}\int_{-\delta}^{\delta} e^{itx}\,dt\,d\mu(x)\\ &= \int \frac{1}{2\delta}\cdot\frac{e^{i\delta x}-e^{-i\delta x}}{ix}\,d\mu(x)\\ &= \int \frac{\sin\delta x}{\delta x}\,d\mu(x).\end{aligned}$$

Then, for appropriate constant $R > 0$ we have the following estimate of the sinc function term

$$\begin{aligned}\int \frac{\sin\delta x}{\delta x}\,d\mu(x) &\le \int_{|x|\le R} 1\,d\mu(x) + \int_{|x|>R}\frac{1}{|\delta x|}\,d\mu(x)\\ &= 1-\int_{|x|>R}\left(1-\frac{1}{|\delta x|}\right)d\mu(x).\end{aligned}$$

If we take $R = \frac{2}{\delta}$, then the Chebyshev inequality has

$$\frac{1}{2}\mu([-\tfrac{2}{\delta},\tfrac{2}{\delta}]^c) \le \int_{|x|>\frac{2}{\delta}}\left(1-\frac{1}{|\delta x|}\right)d\mu(x) \le 1-\frac{1}{2\delta}\int_{-\delta}^{\delta}\varphi(t)\,dt,$$

so we are done. □

Theorem 3.8 (The Lévy continuity theorem). *Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of probability Borel measures on $\mathbb{R}$ and φ_n their characteristic functions. Then, μ_n converges weakly to a probability Borel measure μ if and only if φ_n converges pointwise to a function φ that is continuous at zero.*

Proof. ($\Rightarrow$) Suppose μ_n converges weakly to a probability Borel measure μ on $\mathbb{R}$. Let φ be the characteristic function of μ. Then, φ is continuous at zero. Since e^{itx} is continuous and bounded for each $t\in\mathbb{R}$, we have

$$\varphi_n(t) = \int e^{itx}\,d\mu_n(x) \to \int e^{itx}\,d\mu(x) = \varphi(t)$$

as $n\to\infty$.

($\Leftarrow$) Let φ_n be the characteristic functions of μ_n, and suppose φ_n converges pointwise to a function φ. Suppose further that φ is continuous at zero. For $\varepsilon > 0$, take $\delta > 0$ using the continuity of φ such that

$$\frac{1}{2\delta}\int_{-\delta}^{\delta}(1-\varphi(t))\,dt < \frac{\varepsilon}{4}.$$

By the bounded convergence theorem, there is $N > 0$ such that

$$\frac{1}{2\delta}\int_{-\delta}^{\delta}|\varphi_n(t)-\varphi(t)|\,dt < \frac{\varepsilon}{4}$$

so that we have

$$\mu_n([-\tfrac{2}{\delta},\tfrac{2}{\delta}]^c) \le 2\cdot\frac{1}{2\delta}\int_{-\delta}^{\delta}(1-\varphi_n(t))\,dt < \varepsilon$$

for all $n > N$. For each $n \le N$, since every single measure is tight, there is a compact $K_n \subset \mathbb{R}$ such that $\mu(K_n^c) < \varepsilon$. If we define a compact set $K := [-\frac{2}{\delta},\frac{2}{\delta}] \cup \bigcup_{n=1}^N K_n$, then $\mu_n(K^c) < \varepsilon$ for all n, so the sequence μ_n is tight.

Let $(\mu_{n_j})_j$ be any subsequence that converges weakly to a probability measure. The limit of this subsequence is independent on the choice of the subsequence since its characteristic function is given by the pointwise limit $\lim_{j\to\infty}\varphi_{n_j} = \varphi$, by the first half of this theorem. Let μ be this unique limit. Then, μ_n converges weakly to μ since the tightness guarantees that every subsequence of μ_n has a further subsequence by the Prokhorov theorem, which converges to μ weakly. □

There are various ways to prove Bochner's theorem on $\mathbb{R}$. For example, we can prove it using either Helly's selection theorem or the Riesz-Markov-Kakutani representation theorem in the same manner as we did in the previous chapter. We introduce a new proof that follows from the Herglotz representation theorem, in order to see the relation of two Bochner's theorems on $\mathbb{Z}$ and $\mathbb{R}$. In this proof, the Lévy continuity theorem is used as a key lemma.

Corollary 3.9 (Bochner's theorem on $\mathbb{R}$). *A function $\varphi : \mathbb{R} \to \mathbb{C}$ is continuous and positive-definite such that $\varphi(0) = 1$ if and only if there is a probability Borel measure μ on $\mathbb{R}$ such that*

$$\varphi(t) = \int e^{itx}\,d\mu(x).$$

Proof. Let μ be a probability Borel measure on $\mathbb{R}$. Then, the function φ defined in the statement is positive definite because

$$\begin{aligned}\sum_{k,l=1}^{n}\varphi(t_k-t_l)\xi_k\overline{\xi_l} &= \sum_{k,l=1}^{n}\int e^{i(t_k-t_l)x}\,d\mu(x)\xi_k\overline{\xi_l}\\ &= \int\left|\sum_{k=1}e^{it_kx}\xi_k\right|^2 d\mu(x) \ge 0.\end{aligned}$$

It is continuous because a single probability measure μ is tight so that for every $\varepsilon > 0$ there is $M > 0$ such that

$$\begin{aligned}|\varphi(t)-\varphi(s)| &\le \int |e^{itx}-e^{isx}|\,d\mu(x) = \int |2\sin(\frac{t-s}{2}x)|\,d\mu(x)\\ &\le \int_{|x|\le M}|(t-s)x|\,d\mu(x) + \int_{|x|>M}d\mu(x)\\ &\le M|t-s| + \frac{\varepsilon}{2} < \varepsilon\end{aligned}$$

whenever $|t-s| < \varepsilon/2M$. The normalization condition $f(0) = 1$ is clear.

Conversely, suppose that φ is continuous and positive definite. For each small $\delta > 0$, since the sequence $(\varphi(\delta k))_{k\in\mathbb{Z}}$ is positive definite, by the Herglotz representation theorem, there is a finite regular Borel measure ν_δ on $[-\pi,\pi)$ such that

$$\varphi(\delta k) = \int_{-\pi}^{\pi} e^{-ik\theta}\,d\nu_\delta(\theta)$$

for every $k \in \mathbb{Z}$. If we define a measure μ_δ on $\mathbb{R}$ such that the support is contained in $[-\pi/\delta,\pi/\delta]$ and $\mu_\delta(E) := \nu_\delta(-\delta E)$ for Borel sets $E \subset [-\pi/\delta,\pi/\delta)$, then

$$\varphi(\delta k) = \int_{-\pi/\delta}^{\pi/\delta} e^{i\delta kx}\,d\mu_\delta(x) = \varphi_\delta(\delta k)$$

for every $k \in \mathbb{Z}$, where φ_δ is the characteristic function of μ_δ.

Note that ν_δ converges to the Dirac measure δ_0 as $\delta \to 0$ in weak* topology of $C(\mathbb{T})^*$ where $\mathbb{T}$ is identified with the interval $[-\pi, \pi)$. This is because trigonometric polynomials are uniformly dense in $C(\mathbb{T})$ and ν_δ are uniformly bounded in norm; for any $\varepsilon > 0$ and $g \in C(\mathbb{T})$, there is a trigonometric polynomial $h = \sum_k c_k e^{-ik\theta}$ such that $\|g - h\|_{C(\mathbb{T})} < \varepsilon/2$, which implies

$$\begin{aligned}|\langle g, \nu_\delta\rangle - g(0)| &\le |\langle g - h, \nu_\delta\rangle| + |\langle h, \nu_\delta\rangle - h(0)| + |h(0) - g(0)| \\ &\le \frac{\varepsilon}{2} + |\sum_k c_k \varphi(\delta k) - h(0)| + \frac{\varepsilon}{2}\end{aligned}$$

and

$$\sum_k c_k \varphi(\delta k) \to \sum_k c_k = h(0)$$

as $\delta \to 0$.

For each $t \in \mathbb{R}$ and $\delta > 0$, take t_δ such that $|t - t_\delta| \le \delta/2$ and $t_\delta \in \delta\mathbb{Z}$. Then, we get

$$\begin{aligned}|\varphi_\delta(t) - \varphi_u(t_\delta)| &= |\int (e^{itx} - e^{it_\delta x})\, d\mu_\delta(x)| \\ &= |\int_{-\pi}^{\pi} (e^{i\frac{t}{\delta}\theta} - e^{i\frac{t_\delta}{\delta}\theta})\, d\nu_\delta(\theta)| \\ &\le \int_{-\pi}^{\pi} \left|\left(\frac{t}{\delta} - \frac{t_\delta}{\delta}\right)\theta\right|\, d\nu_\delta(\theta) \\ &\le \frac{1}{2}\int_{-\pi}^{\pi} |\theta|\, d\nu_\delta(\theta) \to 0\end{aligned}$$

as $\delta \to 0$ since the function $\theta \mapsto |\theta|$ is a continuous function on $\mathbb{T}$ if we view it as $[-\pi, \pi)$. Therefore, the pointwise convergence is verified as

$$|\varphi_\delta(t) - \varphi(t)| \le |\varphi_\delta(t) - \varphi_\delta(t_\delta)| + 0 + |\varphi(t_\delta) - \varphi(t)| \to 0$$

as $\delta \to 0$, and since φ is continuous at zero, we can conclude that φ is a characteristic function by the Lévy continuity theorem. $\square$

4 Bochner's theorem on locally compact abelian groups: representation theory

In this chapter, we extend Bochner's theorem to a locally compact Hausdorff abelian group G. We will prove the theorem by two different approaches in Section 4.1 and 4.2, respectively; one is by generalized Fourier transforms on G, and the other uses representation theory of G. In Section 4.3, we will prove the Pontryagin duality, one of the most famous applications of the Bochner theorem.

We always mean locally compact *Hausdorff* abelian groups by locally compact abelian groups. For a locally compact abelian group G, we will denote the identity of G by e and a fixed Haar measure of G by dx. Note that the Haar measure is neither in general σ-finite nor inner regular, unless the group G is σ-compact. However, Fubini's theorem and the Riesz representation $L^1(G)^* \cong L^\infty(G)$ can be applied up to minor changes (For example, we need to modify the definition of $L^\infty(G)$ since it is smaller than the dual space $L^1(G)^*$, but the proofs in this thesis only consider the inclusion $C_b(G) \subset L^\infty(G)$ to endow the weak* topology on $C_b(G)$ or $C_0(G)$, so it does not cause any problems). We will not discuss the modifications carefully, but for details, we can refer to Section 2.3 of Folland's book [4]. We also note that substantial parts of the expositions in this chapter are inspired by the same book [4].

4.1 Proof by Fourier transforms

Recall that the Fourier transform on $\mathbb{R}$ is given by the integral operator

$$\mathcal{F}f(\xi) = \int_{\mathbb{R}} e^{-ix\xi} f(x)\, dx$$

with an exponential term $e^{-ix\xi}$. The exponential terms parametrized by $\xi \in \mathbb{R}$ can be recognized as continuous group homomorphisms from $G = \mathbb{R}$ to the circle group $\mathbb{T}$, so we will introduce the space of these group homomorphisms to define generalized Fourier transforms on G.

Definition 4.1 (Dual group)**.** Let G be a locally compact abelian group, and let $\mathbb{T} = \{z \in \mathbb{C} : |z| = 1\}$ be the circle group. The *dual group* $\widehat{G}$ of G is the group of all continuous group homomorphisms $\chi : G \to \mathbb{T}$, endowed with the topology of compact convergence. The elements of the dual group are said to be *characters*.

First, we want to show $\widehat{G}$ is again a locally compact abelian group. To see this, consider the Banach space $L^1(G)$. The function space $L^1(G)$ is an commutative Banach algebra with multiplication structure

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

which is called the *convolution*. Here we briefly introduce spectral theory of commutative Banach algebras. The *spectrum* of an commutative Banach algebra $\mathcal{A}$ is the set of all non-zero algebra homomorphisms $\mathcal{A} \to \mathbb{C}$, which we denote by $\widehat{\mathcal{A}}$ or $\sigma(\mathcal{A})$. If we endow the weak*-topology induced from the dual space $\mathcal{A}^*$ as a Banach space, then the spectrum becomes locally compact and Hausdorff in light of the Banach-Alaoglu theorem. Proof can be found in [8] or [3]. The convolution algebra $L^1(G)$ has some additional properties:

Lemma 4.1. *Let G be a locally compact abelian group, and $L^1(G)$ be the convolution algebra.*

(a) *The algebra $L^1(G)$ admits an approximate identity $(e_\alpha)_\alpha$ such that $e_\alpha(x) = e_\alpha(x^{-1}) = \overline{e_\alpha(x)}$.*

(b) *For $g \in L^1(G)$, we have the limit $L_x g \to g$ in $L^1(G)$ as $x \to e$.*

Proof. (a) Let $\mathcal{N}$ be a local base of symmetric open neighborhoods at the identity $e \in G$, and assign $\psi_U \in C_c(G)$ to each $U \in \mathcal{N}$ that satisfies $\operatorname{supp}\psi_U \subset U$ and $\int_G \psi_U(x)\,dx = 1$. Then,

$$\|\psi_U * g - g\|_{L^1(G)} \le \iint_{G^2} |\psi_U(y)(g(y^{-1}x) - g(x))|\,dx\,dy \le \sup_{y\in U} \|L_y g - g\|_{L^1(G)} \to 0$$

as $U \to \{e\}$, so the net $(\psi_U)_{U\in\mathcal{N}}$ is an approximate identity for $L^1(G)$. The additional properties are trivially satisfied.

(b) We approximate g by a function $h \in C_c(G)$. Since each h is uniformly continuous, if we let K be a compact neighborhood of $\operatorname{supp} h$, then

$$\|L_x h - h\|_{L^1(G)} = |K|\|L_x h - h\|_{L^\infty(G)} \to 0$$

as $x \to e \in G$, because $\operatorname{supp}(L_x h - h) \subset K$ if x is sufficiently near to $e \in G$. If we take $h \in C_c(G)$ such that $\|g - h\|_{L^1(G)} < \varepsilon$ for a fixed $\varepsilon > 0$, then

$$\|L_x g - g\|_{L^1(G)} \le \|L_x(g-h)\|_{L^1(G)} + \|L_x h - h\|_{L^1(G)} + \|h - g\|_{L^1(G)} < 2\varepsilon + \|L_x h - h\|_{L^1(G)}$$

proves the desired result by taking $x \to e$ and $\varepsilon \to 0$. □

Let $\chi \in \widehat{G}$. Then, it defines a linear functional

$$L^1(G) \to \mathbb{C} : f \mapsto \int_G \chi(x) f(x)\,dx$$

on $L^1(G)$, which is a non-zero algebra homomorphism, so induces a map $\widehat{G} \to (L^1(G))^\frown$. In fact this map is a homeomorphism and considered as a canonical identification of the two spectra(the dual group $\widehat{G}$ is sometimes called the spectrum of G). It has an analogy with a locally compact version of the theorem that complex representations of a finite group G has a one-to-one correspondence to $\mathbb{C}[G]$-modules, because $\widehat{G}$ and $L^1(G)^\frown$ can be recognized as the space of irreducible representations of G and $L^1(G)$, respectively. This correspondence provides a starting point to construct a bridge between the groups and algebras.

Proposition 4.2. *The map $\widehat{G} \to (L^1(G))^\frown$ is a homeomorphism. In particular, on $\widehat{G}$, the topology of compact convergence coincides with the weak* topology in $L^1(G)^*$.*

Proof. (Injectivity) If $\chi_1, \chi_2 \in \widehat{G}$ satisfy $\int_G \chi_1(x) f(x)\,dx = \int \chi_2(x) f(x)\,dx$ for all $f \in L^1(G)$, then by the Riesz representation $L^1(G)^* \cong L^\infty(G)$, we have $\chi_1 = \chi_2$.

(Surjectivity) Let $\varphi \in (L^1(G))^\frown$. Define $\chi \in L^\infty(G)$ such that $\varphi(g) = \int_G \chi(x) g(x)\,dx$ for all $g \in L^1(G)$, using the Riesz representation theorem. Then,

$$\begin{aligned}\int_G \varphi(f)\chi(x)g(x)\,dx &= \varphi(f)\varphi(g) = \varphi(f * g)\\ &= \iint_{G^2} \chi(y) f(yx^{-1}) g(x)\,dx\,dy = \int \varphi(L_x f) g(x)\,dx\end{aligned}$$

for $f, g \in L^1(G)$, so we have $\varphi(f)\chi(x) = \varphi(L_x f)$ almost everywhere, where $L_x f(y) = f(x^{-1}y)$. Then, we can take $f \in L^1(G)$ with $\varphi(f) \ne 0$ so that χ has a new representation $\varphi(L_x f)/\varphi(f)$.

We can check that it gives a continuous version of χ by approximation of f by uniformly continuous functions. It is also a group homomorphism since

$$\frac{\varphi(L_{xy} f)}{\varphi(f)} = \frac{\varphi(L_x(L_y f))}{\varphi(L_y f)} \frac{\varphi(L_y f)}{\varphi(f)}.$$

Finally, the boundedness of χ implies $|\chi(x)| = |\chi(x^n)|^{1/n} \le \|\chi\|_{L^\infty(G)}^{1/n} \to 1$ for any $x \in G$ as $n \to \infty$, and by applying x^{-1} once more, we have $\chi : G \to \mathbb{T}$.

(Continuity) Suppose $\chi_\alpha \to \chi$ in the topology of compact convergence in $\widehat{G} \subset C_b(G)$. Let $g \in L^1(G)$ and $\varepsilon > 0$. Take a compact set $K \subset G$ such that

$$\int_{K^c} |g(x)|\, dx < \varepsilon.$$

Then, by taking the limit of α on

$$\left|\int_G (\chi_\alpha - \chi)(x) g(x)\right| dx \le \sup_{x\in K} |\chi_\alpha(x) - \chi(x)| \int_K |g| + \varepsilon,$$

we have

$$\limsup_\alpha \left|\int_G (\chi_\alpha - \chi)(x) g(x)\right| dx \le \varepsilon.$$

Since ε was chosed to be arbitrary, we are done.

(Continuity of inverse) Suppose $\chi_\alpha \to \chi$ in the weak* topology of $L^1(G)^*$. Let K be a compact subset of G and take $\varepsilon > 0$. We will bound $|\chi_\alpha(x) - \chi(x)|$ by averaging. Using the continuity of χ, fix a small compact neighborhood U of the identity e in G such that

$$\frac{1}{|U|} \int_U |1 - \chi(y)|\, dy < \varepsilon.$$

Then, for all $x \in G$,

$$\begin{aligned} \left|\frac{\mathbf{1}_U}{|U|} * \chi(x) - \chi(x)\right| \le (2\varepsilon)^{1/2} \le \frac{1}{|U|} \int_U |\chi(y^{-1}x) - \chi(x)|\, dy \\ = \frac{1}{|U|} \int_U |1 - \chi(y)|\, dy < \varepsilon. \end{aligned}$$

Similarly, we also have for any $x \in G$ that

$$\begin{aligned} \left|\chi_\alpha(x) - \frac{\mathbf{1}_U}{|U|} * \chi_\alpha(x)\right| &\le \frac{1}{|U|} \int_U |\chi_\alpha(x) - \chi_\alpha(y^{-1}x)|\, dy \\ &= \frac{1}{|U|} \int_U \sqrt{2 - 2\operatorname{Re} \chi_\alpha(y)}\, dy \\ &\le \Big(\frac{1}{|U|} \int_U (2 - 2\operatorname{Re} \chi_\alpha(y))\, dy\Big)^{1/2} \\ &\le \Big(2\varepsilon + \frac{2}{|U|} \Big|\int_U (\chi(y) - \chi_\alpha(y))\, dy\Big|\Big)^{1/2}, \end{aligned}$$

so we have

$$\limsup_\alpha \left|\chi_\alpha(x) - \frac{\mathbf{1}_U}{|U|} * \chi_\alpha(x)\right| \le (2\varepsilon)^{1/2}.$$

Since the map $K \to L^1(G) : x \mapsto \mathbf{1}_{x^{-1}U}$ is continuous so that $\{\mathbf{1}_{x^{-1}U} : x \in K\}$ is compact in L^1, there is a finite sequence $(x_j)_{j=1}^n \subset K$ such that for every $x \in K$there is j satisfying $\int |\mathbf{1}_{x^{-1}U} - \mathbf{1}_{x_j^{-1}U}| = \|\mathbf{1}_{x^{-1}U} - \mathbf{1}_{x_j^{-1}U}\|_{L^1(G)} < \varepsilon$. Then,

$$\begin{aligned} |\mathbf{1}_U * (\chi_\alpha - \chi)(x)| &= \Big|\int \mathbf{1}_{x^{-1}U}(\chi_\alpha - \chi)\Big| \\ &\le \Big|\int (\mathbf{1}_{x^{-1}U} - \mathbf{1}_{x_j^{-1}U})\chi_\alpha\Big| + \Big|\int \mathbf{1}_{x_j^{-1}U}(\chi_\alpha - \chi)\Big| + \Big|\int (\mathbf{1}_{x_j^{-1}U} - \mathbf{1}_{x^{-1}U})\chi\Big| \\ &\le \varepsilon + \max_{1\le j\le n} \Big|\int \mathbf{1}_{x_j^{-1}U}(\chi_\alpha - \chi)\Big| + \varepsilon \end{aligned}$$

implies

$$\limsup_{\alpha} |\mathbf{1}_U * (\chi_\alpha - \chi)(x)| \le 2\varepsilon.$$

By summing up the above three terms, the inequality

$$|\chi_\alpha(x) - \chi(x)| \le |\chi_\alpha(x) - \frac{\mathbf{1}_U}{|U|} * \chi_\alpha(x)| + |\frac{\mathbf{1}_U}{|U|} * (\chi_\alpha - \chi)(x)| + |\frac{\mathbf{1}_U}{|U|} * \chi(x) - \chi(x)|$$

implies

$$\limsup_{\alpha} \sup_{x \in K} |\chi_\alpha(x) - \chi(x)| \le (2\varepsilon)^{1/2} + \frac{2\varepsilon}{|U|} + \varepsilon,$$

and it completes the proof by limiting $\varepsilon \to 0$. □

Corollary 4.3. *If G is a locally compact abelian group, then $\widehat{G}$ is also a locally compact abelian group.*

Example 4.1 (Real line)**.** We have an isomorphism of topological groups $\mathbb{R} \cong \widehat{\mathbb{R}} : t \mapsto e^{itx}$. The group $\mathbb{R}$ is additive, and the group $\widehat{\mathbb{R}}$ is multiplicative. The only non-trivial argument is the surjectivity. If $\chi \in \widehat{\mathbb{R}}$, then there is $\varepsilon > 0$ such that $c := \int_0^\varepsilon \chi(x)\,dx \neq 0$, and

$$\int_0^\varepsilon \chi(x)\,dx = \int_y^{y+\varepsilon} \chi(x-y)\,dx = \chi(y)^{-1} \int_y^{y+\varepsilon} \chi(x)\,dx.$$

By differentiating with respect to y, we get a differential equation

$$\chi'(y) = c^{-1}(\chi(y+\varepsilon) - \chi(y)) = c^{-1}(\chi(\varepsilon) - 1)\chi(y),$$

therefore, $\chi(0) = 1$ and $|\chi(x)| = 1$ implies $\chi(x) = e^{itx}$ for some $t \in \mathbb{R}$.

Example 4.2 (Circle and integer)**.** Using the above result, we can also show $\widehat{\mathbb{T}} \cong \mathbb{Z}$. From the identification $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$, a character χ of $\mathbb{T}$ can be characterized as a character e^{itx} of $\mathbb{R}$ that factors through $\mathbb{T}$, which means $e^{itx} = 1$, and it is equivalent to $t \in 2\pi\mathbb{Z}$. The characters of $\mathbb{Z}$ is parametrized by the value at one, so $\widehat{\mathbb{Z}} \cong \mathbb{T}$.

We now define the Fourier transform. For clarity, we use the curly alphabet $\mathcal{F}$ instead of the hat notation $\widehat{f}$ for Fourier transforms.

Definition 4.2 (Fourier transform)**.** Let G be a locally compact abelian group, and $\widehat{G}$ be its dual group. Let $f \in L^1(G)$. The *Foureir transform* is a linear operator $\mathcal{F} : L^1(G) \to \mathbb{C}^{\widehat{G}}$ defined by

$$\mathcal{F}f(\chi) := \int_G \overline{\chi(x)} f(x)\,dx$$

for $\chi \in \widehat{G}$. The extended Fourier transform for measures $\mathcal{F} : M(G) \to \mathbb{C}^{\widehat{G}}$ is called the *Fourier-Stieltjes transform* and given by

$$\mathcal{F}\mu(\chi) := \int_G \overline{\chi(x)}\,d\mu(x)$$

for $\chi \in \widehat{G}$, where $M(G)$ denotes the space of all finite complex regular Borel measures; it is the complex linear span of $\operatorname{Prob}(G)$. We will also often use the adjoint Fourier transform $\mathcal{F}^* : M(\widehat{G}) \to \mathbb{C}^{\widehat{\widehat{G}}}$ defined by

$$\mathcal{F}^*\mu(x) := \int_{\widehat{G}} x(\chi)\,d\mu(\chi)$$

for $x \in \widehat{\widehat{G}}$. Note that the Fourier transform of functions in $L^1(\widehat{G})$ depends on the choice of Haar measure $d\chi$ on $\widehat{G}$, up to constant, and the reasonable constant will be determined by the Fourier inversion theorem in Section 4.3.

The notion of the following *canonical homomorphism* will be useful in the analysis of Fourier transforms on a locally compact abelian group.

Definition 4.3 (Canonical homomorphism)**.** Let G be a locally compact abelian group, and $\widehat{\widehat{G}}$ be its double dual group. We define the *canonical homomorphism* of G to be the map $\Phi : G \to \widehat{\widehat{G}}$ such that $\Phi(x)(\chi) := \chi(x)$.

Note that we do not have any additional information about the homomorphism Φ for now. None of injectivity, surjectivity, and continuity can be deduced by simple arguments at this stage. The goal of this thesis paper is to verify that Φ is a topological isomorphism.

We can embed G into the algebra $M(G)$ by Dirac measures. One way to understand the Fourier transform is a lifting of the canonical homomorphism $\Phi : G \to \widehat{\widehat{G}}$, as described in the following commutative diagram:

$$\begin{tikzcd} G \arrow[r, "\Phi"] \arrow[d, hook] & \widehat{\widehat{G}} \arrow[d, hook] \\ M(G) \arrow[r, "\mathcal{F}^*"] & \mathbb{C}^{\widehat{G}}. \end{tikzcd}$$

The idea of considering the Fourier transform as the extension of the canonical homomorphism from G to its double dual group $\widehat{\widehat{G}}$ provides a fundamental framework for the general theory of commutative Banach algebras; the Gelfand transform. For a commutative Banach algebra $\mathcal{A}$, an element $a \in \mathcal{A}$ defines a function $\widehat{a} : \widehat{\mathcal{A}} \to \mathbb{C} : \varphi \mapsto \varphi(a)$ by evaluation at a. This function is continuous by definition of the weak* topology and and vanishes at infinity since the set $\{\, \varphi \in \widehat{\mathcal{A}} : |\varphi(a)| \geq \varepsilon \,\}$ is weak* compact for any $a \in \mathcal{A}$ and $\varepsilon > 0$. This defines an algebra homomorphism

$$\Gamma : \mathcal{A} \to C_0(\widehat{\mathcal{A}}) : a \mapsto \widehat{a},$$

and this homomorphism is called the *Gelfand transform*. It is not hard to see that when $\mathcal{A} = L^1(G)$ the Gelfand transform $\Gamma : L^1(G) \to C_0(L^1(G)^\wedge)$ corresponds to the adjoint Fourier transform $\mathcal{F}^* : L^1(G) \to C_0(\widehat{G})$ under the identification $C_0(L^1(G)^\wedge) \xrightarrow{\sim} C_0(\widehat{G})$.

We state some basic properties of Fourier transform in the following propositions:

Proposition 4.4. *Let G be a locally compact abelian group.*

(a) *For $\mu, \nu \in M(G)$, $\mathcal{F}(\mu * \nu) = \mathcal{F}\mu\mathcal{F}\nu$.*

(b) *For $\mu \in M(\widehat{G})$ and $\nu \in M(G)$, $\int_{\widehat{G}} \mathcal{F}\nu(\chi)\, d\mu(\chi) = \int_G \mathcal{F}\mu(\Phi(x))\, d\nu(x)$.*

(c) *If $f^*(x) := \overline{f(x^{-1})}$ for $f \in L^1(G)$, then $\mathcal{F}f^*(\chi) = \overline{\mathcal{F}f(\chi)}$.*

Proof. (a) We have

$$\mathcal{F}(\mu * \nu)(\chi) = \iint_{G^2} \chi(xy)\, d\mu(x)\, d\nu(y) = \mathcal{F}\mu(\chi)\mathcal{F}\nu(\chi).$$

(b) We have

$$\begin{aligned} \int_{\widehat{G}} \mathcal{F}\nu(\chi)\, d\mu(\chi) &= \int_{\widehat{G}} \int_G \overline{\chi(x)}\, d\nu(x)\, d\mu(\chi) \\ &= \int_G \int_{\widehat{G}} \overline{\chi(x)}\, d\mu(\chi)\, d\nu(x) \\ &= \int_G \mathcal{F}\mu(\Phi(x))\, d\nu(x). \end{aligned}$$

(c) We have

$$\mathcal{F}f^*(\chi) = \int_G \overline{\chi(x) f(x^{-1})}\, dx = \Big(\int_G \chi(x^{-1}) f(x)\, dx \Big)^- = \overline{\mathcal{F}f(\chi)}. \qquad \square$$

Proposition 4.5. *Let G be a locally compact abelian group. Let $\Phi : G \to \widehat{\widehat{G}}$ be the canonical homomorphism, and denote $(\Phi^* f)(x) := f(\Phi(x))$.*

(a) $\mathcal{F} : L^1(G) \to C_0(\widehat{G})$ *is well-defined.*

(b) $\mathcal{F} : L^1(G) \to C_0(\widehat{G})$ *has dense image.*

(c) $\Phi^* \circ \mathcal{F}^* : M(\widehat{G}) \to C_b(G)$ *is well-defined.*

(d) $\Phi^* \circ \mathcal{F}^* : M(\widehat{G}) \to C_b(G)$ *is injective.*

Proof. (a) The vanishing at infinity and the continuity of $\mathcal{F}f$ for $f \in L^1(G)$ are due to the fact that the (adjoint) Fourier transform coincides with the Gelfand representation of commutative Banach algebras.

(b) Since $L^1(G)$ is closed under convolution and the involution defined as $f^*(x) := \overline{f(x^{-1})}$, the image $\mathcal{F}(L^1(G))$ is a $*$-subalgebra of $C_0(\widehat{G})$. It separates points and vanishes nowhere since for $\chi_1 \neq \chi_2 \in \widehat{G}$ we have $f \in L^1(G)$ such that

$$\int \overline{(\chi_1 - \chi_2)(x)} f(x)\, dx \neq 0,$$

so $\mathcal{F}(L^1(G))$ is dense in $C_0(\widehat{G})$ by the Stone-Weierstrass theorem.

(c) Because the boundedness is clear from the inequality $|\int_{\widehat{G}} \chi(x)\, d\mu(\chi)| \leq \|\mu\|_{M(\widehat{G})}$, it suffices to show $\Phi^* \circ \mathcal{F}^*\mu$ is continuous for $\mu \in M(G)$. We may suppose $\mu \geq 0$ and $\mu(\widehat{G}) = 1$. By the inner regularity, for any $\varepsilon > 0$ there is a compact set $K \subset \widehat{G}$ such that $\mu(K) > 1 - \frac{\varepsilon}{4}$.

We claim

$$W := \{x \in G : \sup_{\chi \in K} |\chi(x) - 1| < \frac{\varepsilon}{2}\}$$

is an open neighborhood of the identity $e \in G$. (In fact, we can show that W is open) Note that $G \times \widehat{G} \to \mathbb{T} : (x, \chi) \mapsto \chi(x)$ is continuous since for each $\chi \in \widehat{G}$ there exists $f \in L^1(G)$ with $\mathcal{F}f(\chi) \neq 0$ and we have $\chi(x) = \mathcal{F}(L_x f)(\chi)/\mathcal{F}f(\chi)$. Thus, for each $\eta \in K$ we can find an open set $U_\eta \times V_\eta \in G \times \widehat{G}$ such that $(e, \eta) \in U_\eta \times V_\eta$ and $|\chi(x) - 1| < \frac{\varepsilon}{2}$ for all $(x, \chi) \in U_\eta \times V_\eta$. By the compactness of K, we can choose a finite sequence $(\eta_j)_{j=1}^n$ such that the union of V_{η_j} covers K. If $x \in \bigcap_{j=1}^n U_{\eta_j}$, then for any χ, since there is j such that $\chi \in V_{\eta_j}$, so $(x, \chi) \in U_{\eta_j} \times V_{\eta_j}$ implies $|\chi(x) - 1| < \frac{\varepsilon}{2}$. It means an open set $\bigcap_{j=1}^n U_{\eta_j}$ is a subset of W containing e, so W is a neighborhood of $e \in G$.

If $x_\alpha \to e$ in G so that x_α eventually in W, then since $x_\alpha \in W$ implies

$$|\Phi^* \circ \mathcal{F}^*\mu(x_\alpha) - 1| \leq \int_K |\chi(x_\alpha) - 1|\, d\mu(\chi) + \int_{\widehat{G} \setminus K} |\chi(x_\alpha) - 1|\, d\mu(\chi) < \frac{\varepsilon}{2} + 2 \cdot \frac{\varepsilon}{4} = \varepsilon,$$

we are done by limiting $\varepsilon \to 0$ for

$$\limsup_\alpha |\Phi^* \circ \mathcal{F}^*\mu(x_\alpha) - 1| \leq \varepsilon.$$

(d) If $\mu \in M(\widehat{G})$ satisfies $\Phi^* \circ \mathcal{F}^*\mu(x) = \int_{\widehat{G}} \chi(x)\, d\mu(\chi) = 0$ for all $x \in G$, then we have

$$\int_{\widehat{G}} \mathcal{F}f(\chi)\, d\mu(\chi) = \int_G \mathcal{F}\mu(\Phi(x)) f(x)\, dx = 0 = \int_G (\int_{\widehat{G}} \chi(x)\, d\mu(\chi))^- f(x)\, dx = 0$$

for all $f \in L^1(G)$. Since $\mathcal{F}(L^1(G))$ is dense in $C_0(\widehat{G})$, we have $\mu = 0$. $\square$

Remark. The injectivity of $\mathcal{F}$ on $L^1(G)$ is a difficult problem. We can show it as a corollary of the part (d) of the above theorem if Φ is injective. However, the injectivity of the canonical homomorphism is equivalent to either the Gelfand-Raikov theorem for abelian groups or the locally compact version of the Peter-Weyl theorem, which states the dual group $\widehat{G}$ separates points G. It is remarkable to compare with standard texts in which we prove that the Fourier transforms $\mathcal{F} : L^1(\mathbb{R}) \to C_0(\mathbb{R})$ and $\mathcal{F} : L^1(\mathbb{T}) \to c_0(\mathbb{Z})$ are injective by using summability methods. This is one direction of the Pontryagin duality theorem, and we will prove it in Section 4.3 via the Fourier inversion theorem.

We finally prove Bochner's theorem. In Chapter 2 and Chapter 3, we constructed a measure indirectly as a weak* limit, but here we will directly define a positive linear functional on a continuous function space to show the existence of a measure.

Theorem 4.6 (Bochner's theorem)**.** *Let G be a locally compact abelian group. A function $f : G \to \mathbb{C}$ is continuous and positive definite if and only if there is a unique non-negative $\mu \in M(\widehat{G})$ such that*

$$f(x) = \int_{\widehat{G}} \chi(x)\, d\mu(\chi)$$

for all $x \in G$.

Proof. ($\Leftarrow$) The continuity is trivially satisfied. The positive definiteness is also clear from the fact that

$$\sum_{k,l=1}^{n} f(x_l^{-1}x_k)\xi_k\overline{\xi_l} = \int_{\widehat{G}} \Big|\sum_{k=1}^{n} \chi(x_k)\xi_k\Big|^2 d\mu(\chi) \geq 0.$$

($\Rightarrow$) The uniqueness directly follows from the parts (d) of Proposition 4.5. We will show the existence of μ. We claim the inequality

$$|\int_G g(x)\overline{f(x)}\,dx| \leq \|\mathcal{F}g\|_{C_0(\widehat{G})}$$

for $g \in L^1(G)$. Then, although we do not know whether $\mathcal{F} : L^1(G) \to C_0(\widehat{G})$ is injective, we can define

$$\mathcal{F}g \mapsto \int_G g(x)\overline{f(x)}\,dx$$

the linear functional on $\mathcal{F}(L^1(G))$ that is bounded with respect to the uniform norm induced from $C_0(\widehat{G})$, and its norm is less than or equal to $\|f\|_{C_b(G)} = f(e)$.

If the claim is true, then since $\mathcal{F}(L^1(G))$ is dense in $C_0(\widehat{G})$, there is a unique bounded linear functional on $C_0(\widehat{G})$ that extends the above linear functional, so we have a complex measure $\mu \in M(\widehat{G})$ such that

$$\int_G g(x)\overline{f(x)}\,dx = \int_{\widehat{G}} \mathcal{F}g(\chi)\,d\mu(\chi) = \int_G g(x)\int_{\widehat{G}} \overline{\chi(x)}\,d\mu(\chi)\,dx$$

for all $g \in L^1(G)$, which implies the equation in the Bochner theorem. Finally,

$$f(e) = \mu(\widehat{G}) \leq \|\mu\|_{M(\widehat{G})} \leq \|f\|_{C_b(G)} = f(e)$$

concludes the non-negativity of μ, so we are done.

Now we prove the claim. Since the positive definiteness of $\overline{f}$ implies that

$$\langle g, h\rangle_{\overline{f}} := \int_G h^* * g(x)\overline{f(x)}\,dx = \iint_{G^2} \overline{h(y)}g(x)\overline{f(y^{-1}x)}\,dx\,dy$$

is a positive semi-definite Hermitian form, where we denote $h^*(x) := \overline{h(x^{-1})}$, we have by the Schwarz inequality and by using the approximate identity that

$$|\int_G g(x)\overline{f(x)}\,dx|^2 \leq \|f\|_{C_b(G)} \int_G g^* * g(x)\overline{f(x)}\,dx.$$

Applying this inequality inductively, we get

$$\begin{aligned}
|\int_G g(x)\overline{f(x)}\,dx| &\leq \|f\|_{C_b(G)}^{1-1/2^n}\Big(\int_G (g^* * g)^{*2^{n-1}}(x)\overline{f(x)}\,dx\Big)^{1/2^n} \\
&\leq \|f\|_{C_b(G)}\|(g^* * g)^{*2^{n-1}}\|_{L^1(G)}^{1/2^n} \\
&\to \|f\|_{C_b(G)}\|\mathcal{F}(g^* * g)\|_{C_0(\widehat{G})}^{1/2} = \|f\|_{C_b(G)}\|\mathcal{F}g\|_{C_0(\widehat{G})}
\end{aligned}$$

as $n \to \infty$ by Gelfand's formula of the spectral radius. Consequently, the claim is true. □

4.2 Proof by the Gelfand-Naimark-Segal construction

We give in this section a representation-theoretic proof of Bochner's theorem. Before we give a precise formulation, we define several notions of representations of locally compact abelian groups.

Definition 4.4 (Strongly continuous unitary representation)**.** Let G be a locally compact abelian group. A *strongly continuous unitary representation* or just a *representation* of G is a continuous group homomorphism $\rho : G \to U(H)$, where $U(H)$ denotes the group of unitary operators on a Hilbert space H with the strong operator topology.

Definition 4.5 (Cyclic representation)**.** Let G be a locally compact abelian group. A *cyclic representation* of G is a representation $\rho : G \to U(H_\rho)$ for which there exists a vector $\psi_\rho \in H_\rho$ called a *cyclic vector* that satisfies the closed linear span of $\rho(G)\psi_\rho$ is equal to H_ρ. A *pointed cyclic representation* of G is a pair (ρ, ψ_ρ) of a cyclic representation ρ of G and a unit cyclic vector $\psi_\rho \in H_\rho$.

Definition 4.6 (Unitary equivalence)**.** Let (ρ_1, ψ_1) and (ρ_2, ψ_2) be pointed cyclic representations of a locally compact abelian group G. We say that they are *unitarily equivalent* if there is a unitary operator $u : H_{\rho_1} \to H_{\rho_2}$ such that $\rho_2(x) = u\rho_1(x)u^*$ for all $x \in G$ and $\psi_2 = u\psi_1$.

Now, we need to figure out the maps that connect measures, positive definite functions, and representations. The idea is based on a famous result of C*-algebra theory called the *Gelfand-Naimark-Segal representation*, ot the GNS representation. The GNS representation is a construction method of cyclic representations of a C*-algebra from a normalized positive linear functional, which is called a *state* in the C*-algebra theory. In commutative C*-algebras, the positive linear functional is nothing but the finite regular Borel measure on a locally compact Hausdorff space, so the GNS construction can be paraphrased into a mapping that sends a probability regular Borel measure to a cyclic representation. For details on the general non-commutative GNS representation, see Chapter 3 and 5 in [8]. We are not going to use the general theory of C*-algebras, but follow and apply the key idea of the GNS construction directly on the commutative C*-algebra $C_0(\widehat{G})$.

Definition 4.7 (Representations of C*-algerbas)**.** Let $\mathcal{A}$ be a C*-algerba. A *strongly continuous representation* or just a *representation* of $\mathcal{A}$ is a continuous $*$-homomorphism $\pi : \mathcal{A} \to B(H_\pi)$, where $B(H_\pi)$ is the algebra of bounded linear operators on a Hilbert space H_π with the strong operator topology.

We say a representation $\pi : \mathcal{A} \to$ is *cyclic* if there exists a vector $\psi_\pi \in H_\pi$ such that the closure of $\pi(\mathcal{A})\psi_\pi$ is equal to H_π. A *pointed cyclic representation* of $\mathcal{A}$ is a pair (π, ψ_π) of a cyclic representation of $\mathcal{A}$ and a unit cyclic vector $\psi_\pi \in H_\pi$. For pointed cyclic representations (π_1, ψ_1) and (π_2, ψ_2), we say they are *unitarily equivalent* if there is a unitary operator $u : H_{\pi_1} \to H_{\pi_2}$ such that $\pi_2(a) = u\pi_1(a)u^*$ for all $a \in \mathcal{A}$ and $\psi_2 = u\psi_1$.

Remark. The continuity condition for a representation $\mathcal{A} \to B(H)$ of a C*-algebra $\mathcal{A}$ is redundant. It is because every $*$-homomorphism between C*-algebras is norm-decreasing. Since the strong operator topology is weaker than the norm topology, boundedness implies the continuity.

The following commutative diagram might be helpful to understand our picture.

$$
\begin{array}{ccc}
\left\{\begin{array}{c}\text{regular Borel}\\ \text{probability measures on } \widehat{G}\end{array}\right\} & \xrightarrow{\text{GNS}} & \left\{\begin{array}{c}\text{unitary equivalence classes of}\\ \text{pointed cyclic representations of } C_0(\widehat{G})\end{array}\right\} \\
{\scriptstyle \Phi^*\circ\mathcal{F}^*}\Big\downarrow & & \Big\downarrow{\scriptstyle \text{surjective ?}} \\
\left\{\begin{array}{c}\text{normalized continuous}\\ \text{positive definite functions on } G\end{array}\right\} & \xrightarrow{\text{“GNS”}} & \left\{\begin{array}{c}\text{unitary equivalence classes of}\\ \text{pointed cyclic representations of } G\end{array}\right\}
\end{array}
$$

One of our goals in the rest of this section is to verify that the two horizontal arrows in the above diagram are bijective. Then, we set the vertical arrow on the left side to be the adjoint Fourier transform. After that, if we were to show the vertical arrow on the right side which makes the diagram commute is a bijection, then the Bochner theorem would follow. However, the surjectivity of the vertical arrow on the right side is tough to be proved directly. If we try to construct a representation of $C_0(\widehat{G})$ from a representation of G, we encounter an approximation problem: we have to find a suitable notion of convergence in $C_b(\widehat{G})$ to approximate $\varphi \in C_0(\widehat{G})$ by a net $\varphi_\alpha \in \operatorname{span}\Phi(G)$. These kinds of technical issues with regard to the Fourier transformation of non-integrable functions on $\widehat{G}$ such as $\Phi(x)$, which are traced back to the main difficulties in the proof of Bochner's theorem in Section 4.1.

Instead, we will prove the Bochner theorem by showing that the extreme points of the set of normalized continuous positive definite functions on G is in fact the union $\widehat{G} \cup \{0\}$ of the dual group and a singleton. Here is the place where we use the representation theory. Then, by applying the Krein-Milman theorem, we finish the proof of surjectivity.

The most important step is to show that an extreme point is indeed a character, for which we use the horizontal arrow named the "GNS construction" in the second row. It is not the authentic GNS construction because it does not provide a representation of a C*-algebra but of a group G, or its convolution algebra $L^1(G)$, which is not a C*-algebra. Nevertheless, we can mimic the idea to construct a cyclic representation of a locally compact abelian group G, which will be addressed in Theorem 4.8. We first establish the one-to-one correspondence between regular Borel measures on $\widehat{G}$ and the unitary equivalence classes of pointed cyclic representations of $C_0(\widehat{G})$. We will not use the general theory of C*-algebra in order to see every step in the ideas of GNS construction.

Theorem 4.7 (GNS representation for regular Borel measures). *Let G be a locally compact abelian group. Then, there is a one-to-one correspondence*

$$\left\{ \begin{array}{c} \textit{regular Borel} \\ \textit{probability measures on } \widehat{G} \end{array} \right\} \xrightarrow{\sim} \left\{ \begin{array}{c} \textit{unitary equivalence classes of} \\ \textit{pointed cyclic representations of } C_0(\widehat{G}) \end{array} \right\}.$$

Proof. (Well-definedness) We will define the map in the statement of the theorem, which turns out to be identical to the the GNS representation of the C*-algebra $C_0(\widehat{G})$. Let μ be a regular Borel probability measure on $\widehat{G}$ (it is a *state* of $C_0(\widehat{G})$, by the Riesz-Markov-Kakutani representation theorem). Then, μ defines a positive semi-definite Hermitian form on $C_0(\widehat{G})$ by

$$\langle \gamma, \eta \rangle_\mu := \int_{\widehat{G}} \overline{\eta(\chi)} \gamma(\chi)\, d\mu(\chi).$$

The *left kernel* of μ is defined as the set L_μ of elements of $C_0(\widehat{G})$ that have zero as the value of the Hermitian form defined by μ, and it is equal to the kernel of the restriction operator onto the support of μ;

$$L_\mu := \{ f \in C_0(\widehat{G}) : \int |f|^2\, d\mu = 0 \} = \{ f \in C_0(\widehat{G}) : f|_{\operatorname{supp}\mu} = 0 \}.$$

Recall that one way to describe the support of a non-negative measure μ is the complement of the union of all open null sets. Therefore, since the restriction $C_0(\widehat{G}) \to C_0(\operatorname{supp}\mu)$ is surjective by the Urysohn lemma, we obtain the isomorphism $C_0(\widehat{G})/L_\mu \cong C_0(\operatorname{supp}\mu)$. If we induce the Hermitian for $\langle -, - \rangle_\mu$ on $C_0(\operatorname{supp}\mu)$, then it becomes positive definite; an inner product. We can complete the inner product space $C_0(\operatorname{supp}\mu)$ to obtain the Hilbert space $H_\mu = L^2(\operatorname{supp}\mu, \mu)$.

The Gelfand-Naimark-Segal representation of $C_0(\widehat{G})$ with respect to μ is now the $*$-algebra homomorphism

$$\pi_\mu : C_0(\widehat{G}) \to B(H_\mu) : \varphi \mapsto M_\varphi,$$

where M_φ denotes the multiplication operator such that $M_\varphi(\gamma) = \varphi\gamma$. This $*$-homomorphism is a representation, that is, strongly continuous because if $\varphi_n \to \varphi \in C_0(\widehat{G})$, then

$$\|M_{\varphi_n}\gamma - M_\varphi\gamma\|_{H_\mu}^2 = \int_{\operatorname{supp}\mu} |(\varphi_n - \varphi)(\chi)\gamma(\chi)|^2\, d\mu(\chi) \le \|\varphi_n - \varphi\|_{C(\operatorname{supp}\mu)}^2 \cdot \|\gamma\|_{H_\mu}^2 \to 0.$$

If we let $\psi_\mu := \mathbf{1}_{\operatorname{supp}\mu} \in H_\mu$, then it is a unit cyclic vector because $\pi_\mu(C_0(\widehat{G}))\psi_\mu = C(\operatorname{supp}\mu)$ is dense in H_μ, so the pair (π_μ, ψ_μ) is a pointed cyclic representation of $C_0(\widehat{G})$. We call this cyclic vector ψ_μ the canonical cyclic vector, and we assign the unitary equivalence class of (π_μ, ψ_μ) to the measure μ.

(Injectivity) Suppose we have two regular Borel probability measures μ_1 and μ_2 on $\widehat{G}$ such that the pointed cyclic representations $(\pi_{\mu_1}, \psi_{\mu_1})$ and $(\pi_{\mu_2}, \psi_{\mu_2})$ of $C_0(\widehat{G})$ defined as above are unitarily equivalent. Let $u : H_{\mu_1} \to H_{\mu_2}$ be a unitary operator such that $\pi_{\mu_2}(\varphi) = u\pi_{\mu_1}(\varphi)u^*$ for all $\varphi \in C_0(\widehat{G})$ and $\psi_{\mu_2} = u\psi_{\mu_1}$. Then,

$$u\pi_{f_1}(\varphi)\psi_{f_1} = \pi_{f_2}(\varphi)u\psi_{f_1} = \pi_{f_2}(\varphi)\psi_{f_2}$$

implies

$$\begin{aligned}\int_{\widehat{G}} \varphi(\chi)\, d\mu_1(\chi) &= \langle \pi_{\mu_1}(\varphi)\psi_{\mu_1}, \psi_{\mu_1}\rangle_{H_{\mu_1}} = \langle u\pi_{\mu_1}(\varphi)\psi_{\mu_1}, u\psi_{\mu_1}\rangle_{H_{\mu_1}} \\ &= \langle \pi_{\mu_2}(\varphi)\psi_{\mu_2}, \psi_{\mu_2}\rangle_{H_{\mu_2}} = \int_{\widehat{G}} \varphi(\chi)\, d\mu_2(\chi)\end{aligned}$$

for $\varphi \in C_0(\widehat{G})$, and it proves $\mu_1 = \mu_2$ as bounded linear functionals on $C_0(\widehat{G})$.

(Surjectivity) Let (π, ψ) be a pointed cyclic representation of $C_0(\widehat{G})$ with the underlying Hilbert space H. Then, since $C_0(\widehat{G}) \to \mathbb{C} : \varphi \mapsto \langle \pi(\varphi)\psi, \psi\rangle_H$ is a linear functional and has norm one since it satisfies that $|\langle \pi(\varphi)\psi, \psi\rangle_H| \le \|\pi\|\|\varphi\|_{C_0(\widehat{G})}\|\psi\|_H^2 \le \|\varphi\|_{C_0(\widehat{G})}$ and $\lim_\alpha \langle \pi(e_\alpha)\psi, \psi\rangle_H = \langle \psi, \psi\rangle_H = 1$ where e_α denotes an approximate identity of $C_0(\widehat{G})$. The bound $\|\pi\| \le 1$ is due to the fact that every $*$-homomorphism between C*-homomorphism has at most norm one. Therefore, by the Riesz-Markov-Kakutani representation theorem, there is a regular Borel probability measure μ on $\widehat{G}$ such that

$$\langle \pi(\varphi)\psi, \psi\rangle_H = \int_{\widehat{G}} \varphi(\chi)\, d\mu(\chi)$$

for all $\varphi \in C_0(\widehat{G})$.

With this measure μ, construct a pointed cyclic representation (π_μ, ψ_μ) of $C_0(\widehat{G})$ as we did above. Define a bounded linear operator

$$u : H \to H_\mu : \pi(\varphi)\psi \mapsto \pi_\mu(\varphi)\psi_\mu$$

using the cyclicity of π. Then, u is a unitary operator since it is an isometry by

$$\|\pi_\mu(\varphi)\psi_\mu\|_{H_\mu}^2 = \int_{\widehat{G}} |\varphi(\chi)|^2\, d\mu(\chi) = \langle \pi(|\varphi|^2)\psi, \psi\rangle_H = \|\pi(\varphi)\psi\|_H^2,$$

and since it is surjective by the cyclicity of π_μ. Because $\psi_\mu = u\psi$ and

$$[u^*\pi_\mu(\varphi)u](\pi(\gamma)\psi) = u^*\pi_\mu(\varphi)\pi_\mu(\gamma)\psi_\mu = u^*\pi_\mu(\varphi\gamma)\psi_\mu = \pi(\varphi\gamma)\psi = \pi(\varphi)(\pi(\gamma)\psi)$$

for $\varphi, \gamma \in C_0(\widehat{G})$, u is a unitary equivalence between the pointed cyclic representations π and π_μ. □

The next step is to apply the same idea to positive definite functions. The statement and the proof of the "GNS representation theorem" for positive definite functions is as follows:

Theorem 4.8 ("GNS representation" for positive definite functions)**.** *Let G be a locally compact abelian group. Then, there is a one-to-one correspondence*

$$\left\{ \begin{array}{c} \textit{normalized continuous} \\ \textit{positive definite functions on } G \end{array} \right\} \xrightarrow{\sim} \left\{ \begin{array}{c} \textit{unitary equivalence class of} \\ \textit{pointed cyclic representations of } G \end{array} \right\}.$$

Proof. (Well-definedness) We first define the map that sends a normalized continuous positive definite function on G to a pointed cyclic representation of G. Let f be a continuous positive definite function on G such that $\|f\|_{C_b(G)} = f(e) = 1$. The function f defines a positive semi-definite Hermitian form on $L^1(G)$ given by

$$\langle g, h\rangle_f := \int_G h^* * g(y)f(y)\,dy = \iint_{G^2} \overline{h(z^{-1})}g(y)f(zy)\,dz\,dy.$$

Define the left kernel

$$L_f := \{\, g \in L^1(G) : \langle g, g\rangle_f = 0\,\}.$$

Then, by the Cauchy-Schwarz inequality, the Hermitian form induces another Hermitian form on $L^1(G)/L_f$ that is positive definite, in other words, an inner product. Complete the inner product space $L^1(G)/L_f$ to define a Hilbert space H_f, and denote the inner product by $\langle -, -\rangle_{H_f}$.

For each $x \in G$, we can uniquely define a bounded linear operator $\rho_f(x) \in B(H_f)$ such that

$$\rho_f(x)(g + L_f) = L_x g + L_f$$

for $g + L_f \in L^1(G)/L_f$, where $L_x g(y) = g(x^{-1}y)$, because the identity

$$\begin{aligned}
\|\rho_f(x)(g + L_f)\|_{H_f}^2 &= \|L_x g + L_f\|_{H_f}^2 = \|L_x g\|_f^2 \\
&= \iint_{G^2} \overline{L_x g(z^{-1})}L_x g(y)f(zy)\,dz\,dy \\
&= \iint_{G^2} \overline{g(x^{-1}z^{-1})}g(x^{-1}y)f(zy)\,dz\,dy \\
&= \iint_{G^2} \overline{g(z^{-1})}g(y)f((zx^{-1})(xy))\,dz\,dy \\
&= \|g\|_f^2 = \|g + L_f\|_{H_f}^2
\end{aligned}$$

proves the boundedness of $\rho_f(x)$. We claim that $\rho_f : G \to B(H_f)$ is a cyclic representation.

It is a group homomorphism since the identity $\rho_f(xy) = \rho_f(x)\rho_f(y)$ for bounded linear operators on $L^1(G)/L_f$ is extended to H_f. It is unitary because it is an isometry by the above identity and $\rho_f(x)$ has its inverse $\rho_f(x^{-1})$. It is strongly continuous because if a net $x_\alpha \in G$ converges to the identity e, then the inequality

$$|\langle g, h\rangle_f| \le \iint_{G^2} |\overline{h(z^{-1})}g(y)f(zy)|\,dz\,dy \le \int_G |\overline{h(z^{-1})}|\,dz \int_G |g(y)|\,dy = \|h\|_{L^1(G)}\|g\|_{L^1(G)}$$

implies

$$\|(\rho_f(x_\alpha) - \mathrm{id}_{H_f})(g + L_f)\|_{H_f} = \|L_{x_\alpha}g - g\|_f \le \|L_{x_\alpha}g - g\|_{L^1(G)} \to 0.$$

Finally, it is cyclic with a cyclic vector $\psi_f \in H_f$ defined by the weak* limit of a net $e_\alpha + L_f$, where e_α is an approximate identity of $L^1(G)$. The limit uniquely exists since we have

$$\langle e_\alpha + L_f, g + L_f\rangle_{H_f} = \langle e_\alpha, g\rangle_f = \int_G g^* * e_\alpha(y)f(y)\,dy \to \int_G g^*(y)f(y)\,dy$$

for each $g + L_f \in L^1(G)/L_f$ and $\|e_\alpha + L_f\|_{H_f} = \|e_\alpha\|_f \le \|e_\alpha\|_{L^1(G)} = 1$ is uniformly bounded. The vector ψ_f is cyclic because if $g + L_f \in L^1(G)/L_f$ satisfies $\langle \rho_f(x)\psi_f, g + L_f\rangle_{H_f} = 0$ for all $x \in G$, then

$$0 = \langle \psi_f, L_{x^{-1}}g + L_f\rangle_{H_f} = \lim_\alpha \langle e_\alpha, L_{x^{-1}}g\rangle_f = \int_G g(xy)f(y)\,dy = \int_G g(y)f(x^{-1}y)\,dy$$

implies

$$0=\int_G \overline{g(x)}\int_G g(y)f(x^{-1}y)\,dy\,dx=\langle g,g\rangle_f,$$

and it means the set $\{\rho_f(x)\psi_f : x\in G\}$ is dense in H_f. Furthermore, since

$$\begin{aligned}\langle \rho_f(x)\psi_f,\psi_f\rangle_{H_f} &= \lim_{\alpha,\beta}\langle L_x e_\alpha, e_\beta\rangle\\ &= \lim_{\alpha,\beta}\iint_{G^2}\overline{e_\beta(z^{-1})}e_\alpha(x^{-1}y)f(zy)\,dz\,dy\\ &= \lim_{\alpha,\beta} e_\alpha * e_\beta * f(x)=f(x),\end{aligned}$$

we have $\|\psi_f\|_{H_f}=\sqrt{f(e)}=1$. Therefore, (ρ_f,ψ_f) is a pointed cyclic representation of G.

(Injectivity) Suppose we have two normalized continuous positive functions f_1 and f_2 on G such that the pointed cyclic representations (ρ_{f_1},ψ_{f_1}) and (ρ_{f_2},ψ_{f_2}) defined as above are unitarily equivalent. Let $u : H_{f_1}\to H_{f_2}$ be a unitary operator such that $\rho_{f_2}(x)=u\rho_{f_1}(x)u^*$ for all $x\in G$ and $\psi_{f_2}=u\psi_{f_1}$. Then,

$$u\rho_{f_1}(x)\psi_{f_1}=\rho_{f_2}(x)u\psi_{f_1}=\rho_{f_2}(x)\psi_{f_2}$$

implies

$$f_1(x)=\langle \rho_{f_1}(x)\psi_{f_1},\psi_{f_1}\rangle_{H_{f_1}}=\langle u\rho_{f_1}(x)\psi_{f_1},u\psi_{f_1}\rangle_{H_{f_1}}=\langle \rho_{f_2}(x)\psi_{f_2},\psi_{f_2}\rangle_{H_{f_2}}=f_2(x).$$

(Surjectivity) Let (ρ,ψ) be a pointed cyclic representation of G with the underlying Hilbert space H. Then, because ρ is continuous with respect to the strong operator topology of $B(H)$ and

$$\sum_{k,l=1}^n \langle \rho(x_l^{-1}x_k)\psi,\psi\rangle_H \xi_k\overline{\xi}_l=\sum_{k=1}^n \|\xi_k\rho(x_k)\psi\|_H^2\geq 0$$

for all $(x_1,\cdots,x_n)\in G^n$ and $(\xi_1,\cdots,\xi_n)\in\mathbb{C}^n$, the function

$$f : G\to\mathbb{C} : x\mapsto \langle \rho(x)\psi,\psi\rangle$$

is continuous and positive definite.

Let (ρ_f,ψ_f) be the pointed cyclic representation of G defined as above. Define a bounded linear operator

$$u : H\to H_f : \rho(x)\psi\mapsto \rho_f(x)\psi_f$$

using the cyclicity of ρ. Then, u is an isometry since the identity

$$\langle \rho_f(x)\psi_f,\rho_f(y)\psi_f\rangle_{H_f}=\langle \rho_f(y^{-1}x)\psi_f,\psi_f\rangle_{H_f}=f(y^{-1}x)=\langle \rho(y^{-1}x)\psi,\psi\rangle_H=\langle \rho(x)\psi,\rho(y)\psi\rangle_H$$

for $x,y\in G$ implies

$$\|\sum_{k=1}^n a_k\rho_f(x_k)\psi_f\|_{H_f}^2=\|\sum_{k=1}^n a_k\rho(x_k)\psi\|_H^2,$$

and surjective since the range of u contains the linear span of $u\rho(x)\psi=\rho_f(x)\psi_f$ for all $x\in G$, which is dense in H_f. Thus the operator u is a unitary operator. Because $\psi_f=u\psi$ and

$$[u^*\rho_f(x)u](\rho(y)\psi)=u^*\rho_f(x)\rho_f(y)\psi_f=u^*\rho_f(xy)\psi_f=\rho(xy)\psi=\rho(x)(\rho(y)\psi)$$

for $x,y\in G$, u is a unitary equivalence between the pointed cyclic representations ρ and ρ_f. □

Definition 4.8 (Irreducible representations)**.** Let G be a locally compact abelian group. We say that a representation $\rho : G\to B(H)$ of G is *irreducible* if there is no non-trivial proper invariant closed subspace K of H, that is, or equivalently, if there exists a representation $\rho_K : G\to B(K)$ satisfying $\rho(x)\xi=\rho_K(x)\xi$ for every $\xi\in K$.

Lemma 4.9. *Let G be a locally compact abelian group. A representation of G is irreducible if and only if it is one-dimensional.*

Proof. It is trivially true that a one-dimensional representation is irreducible. The proof of the converse is based on Schur's lemma, which needs the Borel functional calculus. Here we only sketch the idea of the proof. We can find detailed formulations for the Borel functional calculus in [8] or [3].

If we assume a representation is not one-dimensional, then we can construct a self-adjoint operator that is not a multiple of the identity in the range of the representation. By the Borel functional calculus of this self-adjoint operator, we can find a non-trivial proper projection that commutes with all elements of the range of the representation because the set of projections generate the whole von Neumann algebra generated by the range of the representation. It means that the range of the projection is an invariant subspace, and the irreducibility of the representation fails to hold. □

Proof of Bochner's theorem. Denote the set of continuous positive definite functions f on G such that $f(e) \leq 1$ and $f(e) = 1$ by $P(G)_0$ and $P(G)_1$, and the set of regular Borel measures $\mu \geq 0$ on $\widehat{G}$ such that $\mu(\widehat{G}) \leq 1$ and $\mu(\widehat{G}) = 1$ by $M(\widehat{G})_0^+$ and $M(\widehat{G})_1^+$, respectively. Then, $P(G)_0$ and $M(\widehat{G})_0^+$ are compact convex sets in the weak* topologies of $L^1(G)^*$ and $C_0(\widehat{G})^*$ by the Banach-Alaoglu theorem.

We will only prove the surjectivity of the adjoint Fourier transform $\Phi^* \circ \mathcal{F}^* : M(\widehat{G})_1^+ \to P(G)_1$. Since the identity

$$\int_G \Phi^* \circ \mathcal{F}^* \mu(x) g(x)\, dx = \int_{\widehat{G}} \mathcal{F}^* g(\chi)\, d\mu(\chi)$$

for $g \in L^1(G)$ implies $\Phi^* \circ \mathcal{F}^*$ is continuous so that the image of $M(\widehat{G})_0^+$ is again a compact convex set. Let f be a non-zero extreme point of $P(G)_0$ so that $f(e) = 1$. Let (ρ_f, ψ_f) be the pointed cyclic representation defined by the "GNS construction" from f.

Suppose ρ_f is reducible so that the underlying Hilbert space H_f is decomposed into non-trivial invariant subspaces as $H_f = K \oplus K^\perp$. We have a decomposition $\psi_f = a\xi + b\xi^\perp$ for $\xi \in K$ and $\xi^\perp \in K^\perp$ and it satisfies $a \neq 0 \neq b$ because ψ_f is a cyclic vector that cannot belong to either K or $K^\perp$. We may assume that $a, b > 0$ and $\|\xi\|_{H_f} = \|\xi^\perp\|_{H_f} = 1$ so that $a^2 + b^2 = 1$. Define $g(x) := \langle \rho_f(x)\xi, \xi\rangle_{H_f}$ and $g^\perp := \langle \rho_f(x)\xi^\perp, \xi^\perp\rangle_{H_f}$, which are continuous and positive definite. Then, since ψ_f is a cyclic vector, we have

$$g(x) - g^\perp(x) = \langle \rho_f(x) a\xi, a^{-1}\xi\rangle_{H_f} - \langle \rho_f(x) b\xi^\perp, b^{-1}\xi^\perp\rangle_{H_f} = \langle \rho_f(x)\psi_f, a^{-1}\xi - b^{-1}\xi^\perp\rangle_{H_f} \neq 0$$

for some $x \in G$, and

$$f(x) = \langle \rho_f(x)\psi_f, \psi_f\rangle_{H_f} = a^2 g(x) + b^2 g^\perp(x)$$

implies that f is not extreme. Therefore, ρ_f is irreducible.

Since the representation ρ_f is one-dimensional, there is a character $\chi \in \widehat{G}$ such that $\rho_f(x) = \chi(x)$, which is equal to the adjoint Fourier transform $\chi(x) = \Phi^* \circ \mathcal{F}^* \delta_\chi(x)$. It means that $\Phi^* \circ \mathcal{F}^*(M(\widehat{G})_0^+)$ contains the extreme points of $P(G)_0$, and by the Krein-Milman theorem, we conclude there is $\mu \in M(\widehat{G})_0^+$ such that $\Phi^* \circ \mathcal{F}^* \mu(x) = f(x)$. Putting $x = e$, we get $1 = f(e) = \Phi^* \circ \mathcal{F}^* \mu(e) = \mu(\widehat{G})$, hence we get the surjectivity of $\Phi^* \circ \mathcal{F}^* : M(\widehat{G})_1^+ \to P(G)_1$. □

4.3 The Pontryagin duality

One of the most well-known applications of the Bochner theorem is the Pontryagin duality, which states that the canonical homomorphism $\Phi : G \to \widehat{\widehat{G}}$ for a locally compact abelian group G is always in fact an isomorphism.

The Pontryagin duality is deeply related to the Fourier inversion theorem. In Section 4.1, we defined the Fourier transform on G as an operator that maps a function on G to another function on $\widehat{G}$. Then, the composition of the Fourier transform and the adjoint Fourier transform maps a function on G to a function on the double dual $\widehat{\widehat{G}}$. However, Bochner's theorem tells us that if a function f on G is continuous and positive definite, then f can be realized as the Fourier transform of a function on $\widehat{G}$(a measure is a function in a generalized sense), instead of another hypothetical group H such that $\widehat{H} = G$.

Consider the case of $G = \mathbb{R}$ or $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$. The Fourier inversion theorem and the theorems on the convergence of Fourier series state that from a Fourier transformed function $\mathcal{F}f$ on $\widehat{G} \cong \mathbb{R}$ or $\mathbb{Z}$, we can reconstruct the original function f by the adjoint Fourier transform. In other words, although the domain of $\mathcal{F}^*\mathcal{F}f$ is in principle $\widehat{\widehat{G}}$, but it can be identified with the original function f on the original group G. Furthermore, in a suitable setting of function spaces such as the L^2 space or the Schwartz space, the adjoint Fourier transform $\mathcal{F}^*$ plays a role of the inverse Fourier transform $\mathcal{F}^{-1}$.

We are interested in the generalization of the recovery of the original group from the dual group $\widehat{G}$. This kind of question of recovery is called *duality*, and one of the most classical results of this kind is the Pontryagin duality. The duality for compact second countable abelian groups was proved by Pontryagin [11] in 1934, and van Kampen [14] generalized the result in the following year for the case of locally compact abelian groups. Nowadays, the Pontryagin duality refers to the duality result for locally compact abelian groups.

For a locally compact abelian group G, we can infer from Bochner's theorem a process to pullback the doubly-Fourier-transformed function on $\widehat{\widehat{G}}$ to the original group G does not lose informtation of the original function. To see this, we reformulate Bochner's theorem in terms of a newly defined algebra of functions as follows:

Definition 4.9 (Fourier-Stieltjes algebra). Let G be a locally compact abelian group. The *Fourier-Stieltjes algebra* $B(G)$ is the linear span of the continuous positive definite functions on G. Note that $B(G) \cap M(G) = B(G) \cap L^1(G)$.

Corollary 4.10 (A reformulation of Bochner's theorem). *Let G be a locally compact abelian group, and $\Phi : G \to \widehat{\widehat{G}}$ be the canonical homomorphism. Then, $\Phi^* \circ \mathcal{F}^* : M(\widehat{G}) \to B(G)$ is a well-defined algebra isomorphism.*

The space $B(G) \cap L^1(G)$ replaces the Schwartz space in the classical theory of Fourier transforms on the Euclidean spaces. Note that we do not have differential structure and the notion of decay growths on G. Intuitively, the inverse of $M(\widehat{G}) \to B(G)$ is used to control the $L^1(\widehat{G})$-norm of the transformed function $\mathcal{F}f$ for $f \in B(G) \cap L^1(G)$.

Classical Fourier inversion theorems on $\mathbb{R}$ and $\mathbb{Z}$ go further than Bochner's theorem; not only is the Fourier transform bijective, but the inverse is given by its adjoint. Standard proofs of the Fourier inversion theorem on $\mathbb{R}$ use the scaling of $\mathbb{R}$ by scalar multiplication, and the differentiable structure. Standard results on the convergence theorem of Fourier series also use several approximate identities such as the Dirichlet kernel and the Fejér kernel. They cannot be generalized to the case of locally compact abelian groups G, so we should find a method for our proof. The inversion theorem is rigorously stated and proved as follows:

Theorem 4.11 (Fourier inversion). *Let G be a locally compact abelian group, and $\widehat{G}$ be its dual group. By adjusting the constant of a Haar measure on $\widehat{G}$, called the* dual measure *of the Haar measure dx of G, the following statements hold:*

(a) *For $f \in B(G) \cap L^1(G)$, we have $\mathcal{F}f \in B(\widehat{G}) \cap L^1(\widehat{G})$ and $\Phi^* \circ \mathcal{F}^* \circ \mathcal{F}f = f$.*

(b) *For $\varphi \in B(\widehat{G}) \cap L^1(\widehat{G})$, we have $\mathcal{F}^* \circ \Phi^* \circ \mathcal{F}\varphi = \varphi$*

Proof. (a) Without loss of generality, assume $f \in B(G)^+ \cap L^1(G)$, where $B(G)^+$ denotes the space of all continuous positive definite functions on G. By the Bochner theorem, there is a non-negative measure $\mu \in M(\widehat{G})$ such that

$$f(x) = \Phi^* \circ \mathcal{F}^*\mu(x) = \int_{\widehat{G}} \chi(x)\, d\mu(\chi).$$

Our claim is that there is a Haar measure $d\chi$ on $\widehat{G}$ such that $d\mu(\chi) = \mathcal{F}f(\chi)\,d\chi$. If we show this, then both conclusions follow immediately.

Define a linear functional $I : C_c(\widehat{G}) \to \mathbb{C}$ such that for each $\varphi \in C_c(\widehat{G})$ we have

$$I(\varphi) := \int_{\widehat{G}} \varphi(\chi) \frac{d\mu(\chi)}{\mathcal{F}f(\chi)},$$

where $f \in B(G)^+ \cap L^1(G)$ such that $\mathcal{F}f > 0$ on $\operatorname{supp}\varphi$. We claim that such f always exists for any choice of $\operatorname{supp}\varphi$ and I is independent of f.

Let $h \in C_c(G)$ such that $\mathcal{F}h(e) = \int_G h(x)\,dx \neq 0$. Then, the convolution $h^* * h$ is contained in $C_c(G)$ and $\mathcal{F}(h^* * h) = |\mathcal{F}h|^2$, where $h^*(x) := \overline{h(x^{-1})}$. Using the continuity of $\mathcal{F}h$, take an open neighborhood V of $e \in \widehat{G}$ such that $\mathcal{F}h(\chi) \neq 0$ for all $\chi \in V$. For a finite sequence $\{\chi_i\}_{i=1}^n$ such that $\operatorname{supp}\varphi \subset \bigcup_i V\chi_i$, define $f_i(x) := \chi_i(x)(h^* * h)(x)$ and $f = \sum_i f_i$. Then, f can be verified to be in $C_c(G)$ and

$$\mathcal{F}f_i(\chi) = \mathcal{F}(h^* * h)(\chi_i^{-1}\chi) = |\mathcal{F}h(\chi_i^{-1}\chi)|^2 > 0$$

for $\chi \in V\chi_i$ implies $\mathcal{F}f > 0$ on $\operatorname{supp}\varphi$. The function f is also positive definite because $\mathcal{F}f$ is the sum of non-negative functions(We can show directly without Bochner's theorem).

Let $f, g \in B(G)^+ \cap L^1(G)$ such that $\mathcal{F}f, \mathcal{F}g > 0$ on $\operatorname{supp}\varphi$. Let $\mu, \nu \in M(\widehat{G})$ be such that $\Phi^* \circ \mathcal{F}^*\mu = f$ and $\Phi^* \circ \mathcal{F}^*\nu = g$, taken by the Bochner theorem. For any $h \in L^1(G)$, we have

$$\begin{aligned}
\int_{\widehat{G}} \mathcal{F}h(\chi)\mathcal{F}g(\chi)\,d\mu(\chi) &= \int_{\widehat{G}} \mathcal{F}(h * g)(\chi)\,d\mu(\chi) \\
&= \int_G h * g(x)\mathcal{F}\mu(\Phi(x))\,dx \\
&= \int_G h * g(x) f(x^{-1})\,dx \\
&= h * g * f(e),
\end{aligned}$$

and it implies by the symmetry of convolution that

$$\int_{\widehat{G}} \mathcal{F}h(\chi)\mathcal{F}g(\chi)\,d\mu(\chi) = \int_{\widehat{G}} \mathcal{F}h(\chi)\mathcal{F}f(\chi)\,d\nu(\chi).$$

Since the set of $\mathcal{F}h$ for $h \in L^1(G)$ is dense in $C_0(\widehat{G})$, we get $\mathcal{F}g(\chi)\,d\mu(\chi) = \mathcal{F}f(\chi)\,d\nu(\chi)$, which proves the well-definedness of I.

The next step is to show that I is translation-invariant: for $\varphi \in C_c(\widehat{G})$ and $\eta \in \widehat{G}$, and for $f \in B(G)^+ \cap L^1(G)$ such that $\mathcal{F}f > 0$ on $\operatorname{supp}\varphi \cup \operatorname{supp} L_\eta\varphi$, where $L_\eta\varphi(\chi) := \varphi(\eta^{-1}\chi)$, we have

$$I(L_\eta\varphi) = \int_{\widehat{G}} \varphi(\eta^{-1}\chi)\frac{d\mu(\chi)}{\mathcal{F}f(\chi)} = \int_{\widehat{G}} \varphi(\chi)\frac{d\mu(\eta\chi)}{\mathcal{F}f(\eta\chi)} = I(\varphi)$$

since the last equality is due to

$$\Phi^* \circ \mathcal{F}^*(d\mu(\eta\chi))(x) = \int_{\widehat{G}} \chi(x)\,d\mu(\eta\chi) = \int_{\widehat{G}} (\eta^{-1}\chi)(x)\,d\mu(\chi) = \eta^{-1}(x)f(x)$$

and

$$\mathcal{F}(\eta^{-1}f)(\chi) = \int_G \overline{\chi(x)}\eta^{-1}(x)f(x)\,dx = \int_G \overline{(\eta\chi)(x)}f(x)\,dx = \mathcal{F}f(\eta\chi).$$

Therefore, $d\mu/\mathcal{F}f$ is equal to a Haar measure $d\chi$ of $\widehat{G}$ on $\operatorname{supp}\mathcal{F}f$, hence $\mu(\chi) = \mathcal{F}f(\chi)\,d\chi$.

(b) Note that we can slightly modify the Bochner theorem to have an algebra isomorphism $\Phi^* \circ \mathcal{F} : M(\widehat{\widehat{G}}) \to B(\widehat{G})$. From the part (a), we have $\mathcal{F}\varphi \in L^1(\widehat{\widehat{G}})$ so that $\Phi^* \circ \mathcal{F}\varphi \in B(G) \cap L^1(G)$ and

$$\Phi^* \circ \mathcal{F} \circ (\mathcal{F}^* \circ \Phi^* \circ \mathcal{F})\varphi = (\Phi^* \circ \mathcal{F} \circ \mathcal{F}^*) \circ \Phi^* \circ \mathcal{F}\varphi = \Phi^* \circ \mathcal{F}^*\varphi,$$

hence we get $\mathcal{F}^* \circ \Phi^* \circ \mathcal{F}\varphi = \varphi$ by the injectivity of $\Phi^* \circ \mathcal{F}$. $\square$

Theorem 4.12 (Plancherel's theorem). *Let G be a locally compact abelian group, and $\widehat{G}$ be its dual group with the dual measure. Then,*

$$\|\mathcal{F}f\|_{L^2(\widehat{G})} = \|f\|_{L^2(G)}$$

for $f \in L^2(G) \cap L^1(G)$.

Proof. The convolution $f^* * f$ is in $L^1(G)$ since f and f^* are in $L^1(G)$ and satisfies $\mathcal{F}(f^* * f) = |\mathcal{F}f|^2$, where $f^*(x) := \overline{f(x^{-1})}$ for $x \in G$. It is also continuous because the translation is continuous in $L^1(G)$, and is positive definite because

$$\begin{aligned}\sum_{k,l=1}^{n} f^* * f(x_l^{-1}x_k)\xi_k\overline{\xi}_l &= \sum_{k,l=1}^{n} \int_G \overline{f(y^{-1})} f(y^{-1}x_l^{-1}x_k)\xi_k\overline{\xi}_l \, dy \\ &= \sum_{k,l=1}^{n} \int_G \overline{\xi_l f(y^{-1}x_l)} \xi_k f(y^{-1}x_k) \, dy \\ &= \int_G \Big| \sum_{k=1}^{n} \xi_k f(y^{-1}x_k) \Big|^2 \, dy \geq 0.\end{aligned}$$

By the Fourier inversion theorem, we have

$$\int_G |f(y^{-1})|^2 \, dy = f^* * f(e) = \mathcal{F}^*\mathcal{F}(f^* * f)(e) = \int_{\widehat{G}} \mathcal{F}(f^* * f)(\chi) \, d\xi = \int_{\widehat{G}} |\mathcal{F}f(\chi)|^2 \, d\xi. \qquad \square$$

Finally, we can prove the Pontryagin duality theorem.

Theorem 4.13 (Pontryagin duality). *Let G be a locally compact abelian group, and $\widehat{G}$ be its dual group. Then, the canonical homomorphism $\Phi : G \to \widehat{\widehat{G}}$ is a topological isomorphism.*

Lemma (A lemma for Pontryagin duality). *For an open subset U of $\widehat{\widehat{G}}$, there is non-zero $f \in \mathcal{F}^*(L^1(\widehat{G}))$ supported on U.*

Proof. Let V be an open set such that $VV \subset U$, and take $g \in C_c(\widehat{\widehat{G}})$ any non-negative non-zero continuous functions with $\operatorname{supp} g \subset V$ using the Urysohn lemma. If we define $f := g * g$, then $f \neq 0$ and $\operatorname{supp} f \subset (\operatorname{supp} g)(\operatorname{supp} g) \subset VV \subset U$.

By the Plancherel theorem, we have $\Phi^* \circ \mathcal{F}g \in B(\widehat{G}) \cap L^2(\widehat{G})$. Since

$$\begin{aligned}\Phi^* \circ \mathcal{F}f(\chi) &= \int_{\widehat{\widehat{G}}} x(\chi) f(x) \, dx \\ &= \int_{\widehat{\widehat{G}}} x(\chi) \int_{\widehat{\widehat{G}}} g(y) g(y^{-1}x) \, dy \, dx \\ &= \int_{\widehat{\widehat{G}}} g(y) \int_{\widehat{\widehat{G}}} x(\chi) g(y^{-1}x) \, dx \, dy \\ &= \int_{\widehat{\widehat{G}}} g(y) \int_{\widehat{\widehat{G}}} y(\chi) x(\chi) g(x) \, dx \, dy \\ &= \int_{\widehat{\widehat{G}}} y(\chi) g(y) \, dy \int_{\widehat{\widehat{G}}} x(\chi) g(x) \, dx = (\Phi^* \circ \mathcal{F}g(\chi))^2\end{aligned}$$

for all $\chi \in \widehat{G}$, we have $\Phi^* \circ \mathcal{F}f$ belongs to $B(\widehat{G}) \cap L^1(\widehat{G})$ by the Hölder inequality. Therefore, by the inversion theorem, $f = \mathcal{F}^* \circ \Phi^* \circ \mathcal{F}f$ is contained in $\mathcal{F}^*(L^1(\widehat{G}))$. $\square$

Proof of the Pontryagin duality.

(Continuity) We consider the weak* topology on $\widehat{\widehat{G}}$ as a subspace of $L^1(\widehat{G})^*$. Note that for any $\varphi \in L^1(\widehat{G})$ we have

$$\int_{\widehat{G}} \Phi(x)(\chi)\varphi(\chi) \, d\chi = \int_{\widehat{G}} \chi(x)\varphi(\chi) \, d\chi = \Phi^* \circ \mathcal{F}^*\varphi(x).$$

Since $\Phi^* \circ \mathcal{F}^*\varphi$ is a continuous function on G by the part (c) of Proposition 4.5, we obtain $\Phi(x_\alpha) \to \Phi(x)$ in $\hat{\hat{G}}$ if $x_\alpha \to x$ in G

(Embedding) We will show that Φ is a topological embedding. The injectivity of Φ clearly follows. Suppose a net x_α does not converge to e in G. We may assume by taking a subnet that there is a symmetric open neighborhood U of e in G such that $x_\alpha \notin U$ for all α. Take a non-zero function $f \in C_c(G)$ such that $f(e) \neq 0$ and $\operatorname{supp} f \subset V$, where V is a symmetric open neighborhood of $e \in G$ satisfying $VV \subset U$. Since $g = f^* * f$ is positive definite so that $f^* * f \in B(G) \cap L^1(G)$, so $\varphi := \mathcal{F}g \in L^1(\widehat{G})$ satisfies $\operatorname{supp}(\Phi^* \circ \mathcal{F}^*\varphi) = \operatorname{supp}(f^* * f) \subset U$ by the Fourier inversion.

Then, we have $\Phi^* \circ \mathcal{F}^*\varphi(x_\alpha) = 0$ for all α but $\Phi^* \circ \mathcal{F}^*\varphi(e) \neq 0$. Then, since $\Phi^* \circ \mathcal{F}^*\varphi(x) = \int \Phi(x)(\chi)\varphi(\chi)\,d\chi$, the function $\Phi(x_\alpha)$ does not converges to $\Phi(x)$ in the weak* topology of $L^1(\widehat{G})^\infty$, which is the same topology on $\hat{\hat{G}} = (L^1(\widehat{G}))\widehat{\ }$. Therefore, $\Phi : G \to \hat{\hat{G}}$ is a topological embedding.

(Surjectivity) Now let $y \in \overline{\Phi(G)}$ such that there is a net $x_\alpha \in G$ satisfying $\Phi(x_\alpha) \to y$ in $\hat{\hat{G}}$. Since $\Phi(x_\alpha)$ is Cauchy and Φ is an embedding, x_α is also Cauchy. Because every locally compact group is complete, $x_\alpha \to x$ in G. Then, $\Phi(x) = \Phi(\lim_\alpha x_\alpha) = \lim_\alpha \Phi(x_\alpha) = y$ implies $y \in \Phi(G)$, so $\Phi(G)$ is closed in $\hat{\hat{G}}$.

Now suppose that $\Phi(G)$ is not dense in $\hat{\hat{G}}$. Then a non-zero function $f \in \mathcal{F}^*(M(\widehat{G}))$ vanishes on $\Phi(G)$ by the previous lemma. For $\mu \in M(\widehat{G})$ such that $f = \mathcal{F}^*\mu$, we have $\Phi^* \circ \mathcal{F}^*\mu = f|_{\Phi(G)} = 0$, so $\mu = 0$ by the Bochner theorem, and it leads to a contradiction to $f \neq 0$. Therefore, $\Phi(G)$ is a closed dense subset of $\hat{\hat{G}}$, which proves that Φ is surjective. $\square$

References

[1] Bochner, S. *Vorlesungen über Fouriersche integrale*, vol. 12. Akademische Verlagsgesellschaft, 1932.

[2] Carathéodory, C. Über den variabilitätsbereich der koeffizienten von potenzreihen, die gegebene werte nicht annehmen. *Mathematische Annalen 64*, 1 (1907), 95–115.

[3] Conway, J. B. *A course in functional analysis*, vol. 96. Springer, 2019.

[4] Folland, G. B. *A course in abstract harmonic analysis*, vol. 29. CRC press, 2016.

[5] Herglotz, G. Uber potenzreihen mit positivem, reelen teil im einheitskreis. *Ber. Verhandl. Sachs Akad. Wiss. Leipzig, Math.-Phys. Kl. 63* (1911), 501–511.

[6] Lévy, P. *Calcul des probabilités*. Gauthier-Villars, 1925.

[7] Mathias, M. Über positive fourier-integrale. *Mathematische Zeitschrift 16* (1923), 103–125.

[8] Murphy, G. J. *C*-algebras and operator theory*. Academic press, 2014.

[9] Parthasarathy, K. R. *Probability measures on metric spaces*, vol. 352. American Mathematical Soc., 2005.

[10] Phelps, R. R. *Lectures on Choquets theorem*. Springer, 2001.

[11] Pontrjagin, L. The theory of topological commutative groups. *Annals of Mathematics* (1934), 361–388.

[12] Stewart, J. Positive definite functions and generalizations, an historical survey. *Rocky Mountain J. Math. 6*, 4 (1976), 409–434.

[13] Toeplitz, O. Über die fouriersche entwickelung positiver funktionen. *Rendiconti del Circolo Matematico di Palermo (1884-1940) 32*, 1 (1911), 191–192.

[14] van Kampen, E. R. Locally bicompact abelian groups and their character groups. *Annals of Mathematics* (1935), 448–463.