

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

Ikhan Choi

May 30, 2022

## Abstract

The Bochner theorem states that the image of finite Borel measures on an abelian topological group under the Fourier-Stieltjes transform is the set of continuous positive definite functions. This thesis will describe, prove, and investigate several examples of the Bochner theorem in its historical contexts within three different fields of mathematics, complex analysis, probability theory, and representation theory.

## Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Positive definite functions . . . . .	2
1.2	Brief history of Bochner's theorem . . . . .	3
<b>2</b>	<b>On the group <math>\mathbb{Z}</math>: complex analysis</b>	<b>5</b>
2.1	The Carathéodory coefficient problem . . . . .	5
2.2	Toeplitz's algebraic condition . . . . .	9
2.3	The Herglotz-Riesz representation theorem . . . . .	11
<b>3</b>	<b>On the group <math>\mathbb{R}</math>: probability theory</b>	<b>15</b>
3.1	Topologies on the space of probability measures . . . . .	15
3.2	The Lévy continuity theorem . . . . .	20
3.3	Further developments for non-locally compact groups . . . . .	23
<b>4</b>	<b>On locally compact abelian groups: representation theory</b>	<b>24</b>
4.1	The Fourier transform . . . . .	24
4.2	The Pontryagin duality . . . . .	26
4.3	Cyclic representations and Bochner's theorem . . . . .	28
4.4	Further developments for non-abelian groups . . . . .	29

## Acknowledgement

# 1 Introduction

## 1.1 Positive definite functions

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

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

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

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 even more, positive *semi*-definite, regardless of any choices of  $(x_1, \dots, x_n) \in G^n$ . We give some several remarkable properties and examples of positive definite functions as follows:

**Proposition 1.1.** 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)  $\bar{f}_1$  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  $f = \lim_{m \rightarrow \infty} f_m$  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  $0 \leq \bar{\xi} f(e) \xi$  implies  $f(e) \in \mathbb{R}$ . Since

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

we have

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

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

(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

$$0 \leq \begin{pmatrix} 1 & \bar{\xi} \end{pmatrix} \begin{pmatrix} f(e) & f(x^{-1}) \\ f(x) & f(e) \end{pmatrix} \begin{pmatrix} 1 \\ \xi \end{pmatrix} = f(e)(1 + |\xi|^2) + 2 \text{Re}(f(x)\bar{\xi}).$$

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

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

(g) Let  $f = f_1$  and write

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

If  $\eta = -\xi$ , then

$$0 \leq f(e) + 2(f(e) - \operatorname{Re} f(h))|\xi|^2 + 2\operatorname{Re}((f(x) - f(xh))\bar{\xi}).$$

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

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

so that we have

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

Since  $\varepsilon$  can be taken arbitrarily,  $f$  is right uniformly continuous. The left uniform continuity is shown in the same manner.  $\square$

**Example 1.1** (Positive definite functions not positive). Let  $G = \mathbb{R}$ . Then,  $f(x) := \cos x$  is positive definite since

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

## 1.2 Brief history of Bochner's theorem

This thesis follows the historical flows to extract mathematical ideas behind the positive definite functions. In particular, we are concerned with the results like the following *Bochner-type theorems*:

**Theorem 1.2.** A function  $c : \mathbb{Z} \rightarrow \mathbb{C}$  is positive definite if and only if there is a unique finite regular Borel measure  $\mu$  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 1.3.** A continuous function  $\varphi : \mathbb{R} \rightarrow \mathbb{C}$  is positive definite if and only if there is a unique finite regular Borel measure  $\mu$  on  $\mathbb{R}$  such that

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

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

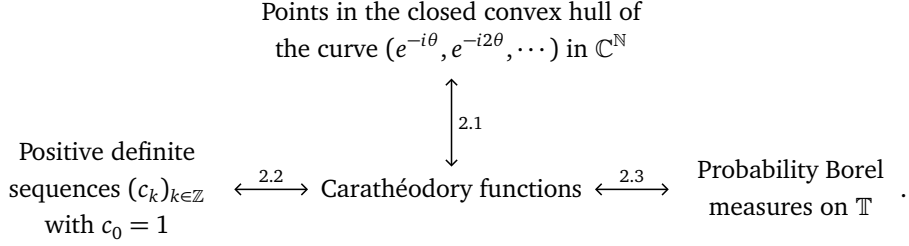
They have similar forms in that they describe the necessary and sufficient conditions for a function to have a Fourier-Stieltjes integral representation of a finite regular Borel measure. One of our primary goals is to investigate the nature of positive definite functions and their harmonic-analytic relation to Borel measures within more familiar cases of  $G = \mathbb{Z}$  or  $\mathbb{R}$ . Now then, we finally extend the Bochner-type results in the more general setting, where  $G$  is a locally compact group, and assign a new perspective of measures in terms of the representation theory of groups.

Each theorem above has its own taste in different subfields of mathematics. Theorem 1.1, which is a corollary of the celebrated Herglotz-Riesz representation theorem, is related to a classical problem in complex analysis that asks to give a characterization of a special class of analytic functions on the open unit disk  $\mathbb{D}$  called the Carathéodory class. The positive definiteness arises as a property of coefficients of functions in the Carathéodory class, and their connection to Fourier coefficients leads the complex analysis problem into harmonic analysis. In Section 2, with the methods of elementary complex variable function theory, our first Bochner-type theorem will be proved, giving a geometric description of the space of positive definite functions in addition.

In Section 3, we review the well-known results of the positive definite functions on the real line and their “weak convergence”. They have been studied by probabilists, to attack the weak convergence of probability measures. Recall that a probability distribution of a real-valued random variable is defined by a probability measure on  $\mathbb{R}$ . The extended Fourier transform, but reversing the sign convention on the phase term, with respect to not only integrable functions but also finite measures, called Fourier-Stieltjes transform, of a probability measure  $\mu$  is called a characteristic function of the distribution  $\mu$ . In terms of probability theory, it is nothing but the function defined by the expectation  $\varphi(t) := Ee^{itX}$ , where  $X$  is a random variable of law  $\mu$ . The Bochner theorem states that the necessary and sufficient condition for being a characteristic function is the positive definiteness and continuity.

## 2 On the group $\mathbb{Z}$ : complex analysis

In this section, we are going to investigate the origin of positive definiteness that occurs in the context of complex analysis via establishing the following one-to-one correspondences:



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

### 2.1 The Carathéodory coefficient problem

The concept of positive definiteness of functions were 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 values of 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.** 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 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 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  $\operatorname{Re} f : \mathbb{D} \rightarrow \mathbb{R}$  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 Carathéodory class and the positive harmonic functions on the open unit disk that has the value one at zero.

Carathéodory's result intuitively tells us that every function in the Carathéodory class can be constructed by convex combinations the Möbius transforms  $f_\theta$ . As a result, they can be viewed as “extreme points” in the Carathéodory class. We discuss about 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 coefficient 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  $c_{-k} := \bar{c}_k$ .*

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

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

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

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

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

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

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 is extended well to the functions that may have harsh behavior on the boundary. Furthermore, by loosening the regularity requirement at 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  $(c_1, \dots, c_n) \in \mathbb{C}^n$  belongs to the convex hull of the curve  $(e^{-i\theta}, \dots, e^{-in\theta}) \in \mathbb{C}^n$  parametrized by  $\theta \in [0, 2\pi)$ .*

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

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

with coefficients  $\lambda_{n,j} \geq 0$  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,

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

implies

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

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

( $\Rightarrow$ ) Conversely, suppose that  $f$  is in the Carathéodory class. Let  $(\gamma_1, \dots, \gamma_n)$  be any point on the surface  $\partial K_n$  of  $K_n$  and  $S$  any supporting hyperplane of  $K_n$  tangent at  $(\gamma_1, \dots, \gamma_n)$ . Let  $(u_1, \dots, u_n)$  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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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 Carathéodory's theorem is equivalent to the following: the point  $(c_1, c_2, \dots) \in \mathbb{C}^{\mathbb{N}}$  belongs to the closed convex hull of the curve  $(e^{-i\theta}, e^{-i2\theta}, \dots) \in \mathbb{C}^{\mathbb{N}}$  parametrized by  $\theta \in [0, 2\pi)$ .*

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



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

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

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

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

is contained in and converges to the point  $(c_1, c_2, \dots)$  in the product topology as  $n \rightarrow \infty$ , so we are done with the desired result. For the opposite direction, let  $(c_1, c_2, \dots) \in K$ . By definition of  $K$  we have an expression

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

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

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

belongs to  $K_n$  because  $K_n$  is closed.

Fix  $\theta \in [0, 2\pi)$  and suppose two complex sequences  $(c_1, c_2, \dots)$  and  $(d_1, d_2, \dots)$  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  $|c_k| \leq 1$  and  $|d_k| \leq 1$ , 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 that  $(e^{-i\theta}, e^{-i2\theta}, \dots)$  is an extreme point of  $K$ . Conversely, every extreme point of  $K$  is contained in the curve  $(e^{-i\theta}, e^{-i2\theta}, \dots)$  by Milman's "converse" theorem of the Krein-Milman theorem[citation: Phelps].  $\square$

## 2.2 Toeplitz's algebraic condition

Toeplitz discovered the coefficient condition addressed in the Carathéodory's paper which regards convex bodies enveloped by a curve 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 determinant for any  $n$ . This algebraic condition is equivalent to that  $H_n$  are all positive semi-definite matrices. Since the principal minors of a positive semi-definite matrix is positive semi-definite, and since a hermitian matrix such that every leading principal minor has non-negative determinant is positive semi-definite, the bilateral sequence  $(c_k)_{k=-\infty}^\infty$  is 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

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

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

**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  $c_{-k} = \bar{c}_k$ .*

*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

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

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

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

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

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

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

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

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

### 2.3 The Herglotz-Riesz representation theorem

Herglotz proved another equivalent condition for the Carathéodory class in 1911, considered as the first Bochner-type theorem, which states the correspondence between the Carathéodory class and probability Borel measure on the unit circle. The Carathéodory theorem states that the function  $f$  in the Carathéodory class is a limit 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-Riesz representation theorem, states that in fact  $f$  is directly represented by the integral of the Möbius transforms with respect to a newly constructed probability measure, instead of limiting process of convex sums.

The essential difficulty comes from the construction of a measure, and here we resolve this in the aid of 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 manipulate 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  $\{z : |z| \leq (r+1)/2\}$  for each fixed  $r < 1$ , we have

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

**Theorem 2.6** (The Herglotz-Riesz 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  $\mu$  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  $\mu$  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 open unit

disk onto the right half plane. The function  $f$  belongs to the Carathéodory class by the open mapping theorem since

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

( $\Rightarrow$ ) Fix  $z$  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 the Helly's selection theorem. Let

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

Then, we have  $F(0) = 0$  and  $F(2\pi) = 1$ , and  $F_{n_j}$  converges to  $F$  at every continuity point  $\theta$  of  $F$ . It means  $F_{n_j}$  converges to  $F$  weakly as  $j \rightarrow \infty$ , so by the Portmanteau theorem, we get

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

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

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

as  $j \rightarrow \infty$ . Therefore, by the uniqueness of limit, we have

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

where  $\mu$  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  $l[1] = \lim_{j \rightarrow \infty} l_{n_j}[1] = 1$  because  $1 \in C(\mathbb{T})$ . (Notice that it does not valid if the domain space,  $\mathbb{T}$  here, is not compact, and we will see this problem more carefully in the next chapter.)

By the Riesz-Markov-Kakutani representation theorem, there is a probability Borel measure  $\mu$  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

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

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  $c : \mathbb{Z} \rightarrow \mathbb{C}$  is positive-definite and  $c_0 = 1$  if and only if there is a probability Borel measure  $\mu$  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  $\mu$  be a probability Borel measure on  $\mathbb{T}$ . Then, the sequence defined in the statement is positive definite because

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

for any  $(\xi_1, \dots, \xi_n) \in \mathbb{C}^n$ , 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  $z \mapsto 1 + \sum_{k=1}^{\infty} 2c_k z^k$  is in the Carathéodory class. By the Herglotz-Riesz representation theorem, there is a probability Borel measure  $\mu$  on  $\mathbb{T}$  such that

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

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

**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  $\delta_\psi$ , 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-Riesz representation theorem, but also viewed as a property of the Poisson kernel. Recall that the measure  $\mu$  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  $(2\pi)^{-1} \operatorname{Re} f(re^{i\theta}) d\theta \rightarrow \delta_\psi$  in weak\* topology of  $C(\mathbb{T})^*$ .

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

Assume  $\tau$  is continuously extended to  $\tau : \bar{\mathbb{D}} \rightarrow \bar{\mathbb{D}}$ . The probability measure on  $\mathbb{T}$  corresponded to the composition  $f \circ \tau$  via the Herglotz theorem is given by 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  $\bar{\mathbb{D}}$ , the limit is described as the continuous density function  $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z} \rightarrow \mathbb{R} : \theta \mapsto \operatorname{Re} f(\tau(e^{i\theta}))$ .

**Example 2.6** (The  $n$ th power map). As another example, if a Carathéodory function  $f$  and a probability measure  $\mu$  on  $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$  satisfies

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

then we have a new family of Carathéodory functions

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

for each positive integer  $n$ , where

$$\mu_n(E) = \frac{1}{n} \sum_{j=0}^{n-1} \mu((nE - 2\pi j) \cap [0, 2\pi)).$$

Intuitively, if  $\mu$  is absolutely continuous with respect to the Lebesgue measure, then the density of  $\mu_n$  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  $c_{k+1}$ .

### 3 On the group $\mathbb{R}$ : probability theory

We have seen the relation of positive definite sequences and measures on the unit circle  $\mathbb{T}$ . On the real line  $\mathbb{R}$ , predictably, we can also prove that there exists a correspondence between measures and positive definite functions. The previous chapter used measures to characterize certain complex functions and positive definite sequences, but from this section, we will see how the positive functions are used in studying measures.

The systematic study of positive definite functions to study measures virtually starts in probability theory 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 problems and laws of large numbers arisen in probability theory want to describe limit behaviors of probability distributions. Lévy’s discovery was that it is easier to verify the convergence of probability distributions via the *Fourier transforms* of probability measures, instead of the measures themselves.

The Fourier transform (often called as *Fourier-Stieltjes* transform when we emphasize the *measures*, the objects being transformed) of a probability measure is called a *characteristic function*. One of possible statement of Bochner’s theorem is that a complex function on a real line is a characteristic function of a probability measure if and only if it is continuous and positive definite, i.e. the theorem gives the one-to-one correspondence of probability measures on  $\mathbb{R}$  and the continuous positive definite functions on  $\mathbb{R}$ , under the Fourier-Stieltjes transform.

#### 3.1 Topologies on the space of probability measures

First, we will investigate topologies on the space of probability measures. In probability theory, the following concept of convergence is the most usual one 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  $\mu_\alpha$  *converges weakly* to another probability Borel measure  $\mu$  if

$$\int g d\mu_\alpha \rightarrow \int g d\mu$$

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  $\mu_\alpha$  converges weakly to  $\mu$ .

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 comparison with the other topology on the sapce of measures. Some reasons why we require the metrizability on  $S$  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 idetification to the continuous dual of a continuous function space.

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

$$\int g d\mu_\alpha \rightarrow \int g d\mu$$

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  $C_0(\Omega)^*$ .

Be cautious that in the Riesz-Markov-Kakutani representation theorem for locally compact Hausdorff spaces we are concerned with *regular* Borel measures that differ to what we call *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  $\Omega$  is locally compact and Hausdorff, then we say a Borel measure  $\mu$  on  $\Omega$  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  $\mu$  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, two notions are not equivalent. We denote the space of all probability regular Borel measures on  $\Omega$  and  $S$  by  $\text{Prob}(\Omega)$  and  $\text{Prob}(S)$ , respectively.

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

**Example 3.1** (Dieudonné measure).

We have omitted the regularity condition on measures in Chapter 2 because every finite Borel measure on a compact metric space is regular in both senses. The vague convergence is less important in probability theory because there are situations that we have to deal with probability measures on a 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 one stochastic process as a single random variable.

Nevertheless, the vague(weak\*) convergence is what we will mainly consider throughout this thesis as a preparation for Chapter 4. Recall that we have used weak\* topology as well in Chapter 2. In this regard, we need to connect the vague convergence to the weak convergence to describe our subjects in probabilistic languages, and the following theorem is one result.

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

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

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

where  $\mu \in \text{Prob}(\Omega)$ ,  $\varepsilon > 0$ , 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  $\text{Prob}(\Omega)$  that converges vaguely to  $\mu \in \text{Prob}(\Omega)$ , 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 maximum with zero. 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  $\mu_\alpha$  to  $\mu$ . □



**Example 3.2** (Escaping to the infinity). 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  $\delta_n$ , 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 have no need to distinguish the topology of weak convergence and the weak\* topology. Now we return to the probability theory. Two classical theorems of the space of probability measures on a metric spaces, 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.

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

- (a)  $\int g d\mu_\alpha \rightarrow \int g d\mu$  for every  $g \in C_b(S)$ , i.e. weakly convergent.
- (b)  $\int g d\mu_\alpha \rightarrow \int g d\mu$  for every uniformly continuous  $g \in C_b(S)$ .
- (c)  $\limsup_\alpha \mu_\alpha(F) \leq \mu(F)$  for every closed  $F \subset S$ .
- (d)  $\liminf_\alpha \mu_\alpha(U) \geq \mu(U)$  for every open  $U \subset S$ .
- (e)  $\lim_\alpha \mu_\alpha(E) = \mu(E)$  for every Borel set  $E \subset S$  such that  $\mu(\partial E) = 0$ .

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

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

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

- (a) The function  $\pi$  is a metric.
- (b) For a sequence  $\mu_n \in \text{Prob}(S)$ , if  $\mu_n \rightarrow \mu$  in  $\pi$ , then  $\mu_n \Rightarrow \mu$ .
- (c) For a net  $\mu_\alpha \in \text{Prob}(S)$ , if  $\mu_\alpha \Rightarrow \mu$ , then  $\mu_\alpha \rightarrow \mu$  in  $\pi$ , given  $S$  is separable.
- (d) The metric space  $(\text{Prob}(S), \pi)$  is separable if and only if  $(S, d)$  is separable.
- (e) The metric space  $(\text{Prob}(S), \pi)$  is complete if and only if  $(S, d)$  is complete.

*Proof.* (a) We will only prove two non-triviality: 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

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

Taking limit  $n \rightarrow \infty$ , we obtain

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

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  $\mu, \nu, \lambda \in \text{Prob}(S)$ . Take sequences  $a_n \downarrow d(\mu, \lambda)$  and  $b_n \downarrow d(\lambda, \nu)$  such that

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

and

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

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

(b) Take  $\varepsilon_n \downarrow 0$  such that  $\mu_n(E) \leq \mu(E^{\varepsilon_n}) + \varepsilon_n$  for every Borel  $E$ , which deduces  $\limsup_{n \rightarrow \infty} \mu_n(F) \leq \mu(F)$  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 sum of positive numbers always diverges 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

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

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  $1 \leq i \leq n$  and  $B(x_i, \varepsilon_i) \cap E \neq \emptyset$ . 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  $\alpha_0$  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,  $d(\mu_\alpha, \mu) \leq \varepsilon$  for all  $\alpha \succ \alpha_0$  since

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

and

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

(d)

(e)

□

**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 are admitted as the most fundamental assumption in probability theory. The above theorem about the Prokhorov metric states that if  $S$  is Polish then so is  $\text{Prob}(S)$ . 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 set of probability Borel measures on 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

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

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

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

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

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

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

Let  $\{x_i\}_{i=1}^\infty$  be a dense set in  $S$ . Then, for each integer  $m > 0$  there is  $n_m > 0$  such that

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

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  $\text{Prob}(S)$ , 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

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

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

( $\Leftarrow$ ) Suppose  $M$  is tight and let  $\mu_\alpha$  be any net in  $M$ . We claim that it has a convergent subnet. Let  $\beta S$  be the Stone-C ech compactification of  $S$ . The inclusion  $\iota : S \rightarrow \beta S$  is a topological embedding because  $S$  is completely regular. Pushforward the measures  $\mu_\alpha$  to make them probability Borel measures  $\nu_\alpha := \iota_* \mu_\alpha$  on  $\beta S$ . Our first claim is that the measure  $\nu_\alpha$  is regular for each  $\alpha$ , that is,  $\nu_\alpha \in \text{Prob}(\beta S)$ . For any Borel  $E \subset \beta S$  and any  $\varepsilon > 0$  we have  $F \subset E \cap S$  that is closed in  $S$  and  $K$  compact in  $S$  such that  $\mu_\alpha(E \cap S) < \mu_\alpha(F) + \varepsilon/2$  and  $\mu_\alpha(S \setminus K) < \varepsilon/2$ . 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  $\nu_\alpha$  since  $F \cap K$  is compact in both  $S$  and  $\beta S$  with  $F \cap K \subset E$ . Note that  $\beta S$  is quite far from metrizable spaces, so the regularity of measures is required for  $\text{Prob}(\beta S)$ . The space  $\text{Prob}(\beta S)$  is compact by the Banach-Alaoglu theorem and the Riesz-Markov-Kakutani representation theorem. Therefore,  $\nu_\alpha$  has a subnet  $\nu_\beta$  that converges to  $\nu \in \text{Prob}(\beta S)$ .

Recall that  $\mu_\beta$  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  $\beta$ . Then, by the Portmanteau theorem, we have

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

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

From the definition of weak convergence we have

$$\int_{\beta S} f d\nu_\beta \rightarrow \int_{\beta S} f d\nu$$

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

$$\int_S f d\mu_\beta \rightarrow \int_S f d\mu$$

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

In this proof of the theorem, we can add a new interpretation of tightness; a restriction condition of a limit point from compactification. The tightness keeps measures from escaping the image of  $S$  in the compactification, and lets the limit point concentrated on it. We can also see the naturality of the topology of weak convergence by recognizing it as the induced topology from the Stone-Ćech compactification.

### 3.2 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. One of the direct connection 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 point-wise 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, but conventionally reversed the sign on the phase term, of a probability measure. Characteristic functions take an advantage that we can learn the information about probability measures by studying the continuous functions instead of measures themselves.

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

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

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

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

*Proof.* It follows clearly that

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

and

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

**Example 3.3.** 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	$f(x) = \frac{1}{2} \cdot \mathbf{1}_{[-1,1]}(x)$	$\varphi(t) = \text{sinc } t$
Exponential	$f(x) = e^{-x} \cdot \mathbf{1}_{[0,\infty)}(x)$	$\varphi(t) = (1 - it)^{-1}$
Cauchy	$f(x) = 1/\pi(1 + x^2)$	$\varphi(t) = e^{- t }$
Polya	$f(x) = (1 - \cos x)/\pi x^2$	$\varphi(t) = \max\{1 -  t , 0\}$

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 to be 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 with a reasonable estimate of the tail probability.

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

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

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

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

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

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

so we are done.  $\square$

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

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

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

as  $n \rightarrow \infty$ .

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

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

for all  $n > N$ . For each  $n \leq N$ , since every single measure is tight, there is 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  $\mu_n$  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  $\lim_{j \rightarrow \infty} \varphi_{n_j} = \varphi$ , by the first half of this theorem. Let  $\mu$  be this unique limit. Then,  $\mu_n$  converges weakly to  $\mu$  since the tightness guarantees that every subsequence of  $\mu_n$  has a further subsequence, which converges to  $\mu$  weakly.  $\square$

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 theorem 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  $\varphi : \mathbb{R} \rightarrow \mathbb{C}$  is continuous and positive-definite such that  $\varphi(0) = 1$  if and only if there is a probability Borel measure  $\mu$  on  $\mathbb{R}$  such that*

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

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

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

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

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

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

Conversely, suppose  $\varphi$  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-Riesz representation theorem, there is a finite regular Borel measure  $\nu_\delta$  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  $\mu_\delta$  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  $\varphi_\delta$  is the characteristic function of  $\mu_\delta$ .

Note that  $\nu_\delta$  converges to the Dirac measure  $\delta_0$  as  $\delta \rightarrow 0$  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  $\nu_\delta$  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

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

and

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

as  $\delta \rightarrow 0$ .

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

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

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

$$|\varphi_\delta(t) - \varphi(t)| \leq |\varphi_\delta(t) - \varphi_\delta(t_\delta)| + 0 + |\varphi(t_\delta) - \varphi(t)| \rightarrow 0$$

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

**Corollary 3.10** (Polya's criterion). *Let  $\varphi : \mathbb{R} \rightarrow [0, \infty)$  be an even function such that  $\varphi(0) = 1$ . If  $\varphi$  is decreasing and convex on  $(0, \infty)$ , then  $\varphi$  is a characteristic function.*

*Proof.*  $\square$

**Example 3.4** (Lévy's  $\alpha$ -stable distribution).

### 3.3 Further developments for non-locally compact groups

## 4 On locally compact abelian groups: representation theory

In this chapter, we extend the domain of Fourier transforms on locally compact abelian groups. Recall that the Fourier transform on  $\mathbb{R}$  is given by the integral

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

with an exponential term.

This exponential term can be recognized as continuous group homomorphism from  $G$  to circle group  $\mathbb{T}$  in general, so we will introduce the notion of space of these homomorphisms and define Fourier transform on  $G$  by using this.

Remark that the homomorphisms  $G \rightarrow \mathbb{T}$  are called character in algebra context. At the same time, when  $G$  is abelian, they form another abelian group, and they are just another form of irreducible representations over  $\mathbb{C}$ . In representation theory of groups, the group algebra  $\mathbb{C}[G]$  has a fundamental position for investigation of the structure of representations of  $G$ .

In locally compact abelian groups, the convolution algebra  $L^1(G)$  replaces the algebra  $\mathbb{C}[G]$ .

### 4.1 The Fourier transform

We will always mean locally compact *Hausdorff* abelian groups by locally compact abelian groups. For a locally compact abelian group  $G$ , we also always denote the identity of  $G$  by  $e$  and a fixed Haar measure of  $G$  by  $dx$ . For clarity, we do not use the hat notation to indicate Fourier transform, for which the curly alphabet  $\mathcal{F}$  will be used.

**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*  $\hat{G}$  of  $G$  is the group of all continuous group homomorphisms  $\chi : G \rightarrow \mathbb{T}$ , endowed with the topology of compact convergence.

Consider the a Banach space  $L^1(G)$ . Then,  $L^1(G)$  is an abelian Banach algebra with multiplication

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

which is called convolution. Here we shortly introduce spectral theory of abelian Banach algebras. The spectrum of an abelian Banach algebra  $\mathcal{A}$  is the set of all non-zero algebra homomorphisms  $\mathcal{A} \rightarrow \mathbb{C}$ , and denoted by  $\hat{\mathcal{A}}$  or  $\sigma(\mathcal{A})$ . If we endow on it 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.

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

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

on  $L^1(G)$ , which is a non-zero algebra homomorphism, so induces a map  $\hat{G} \rightarrow (L^1(G))^\wedge$ . In fact this map is a homeomorphism and considered as a canonical identification of the spectra. 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. This correspondence reduces to a starting point to construct a connection between the groups and algebras.

**Proposition 4.1.** *The map  $\hat{G} \rightarrow (L^1(G))^\wedge$  is a homeomorphism.*

*Proof.*

□



**Corollary 4.2.** *If  $G$  is a locally compact abelian group, then  $\widehat{G}$  is also a locally compact abelian group.*

**Example 4.1** (Real line and circle).

**Example 4.2** (Finite abelian groups).

**Example 4.3** (Infinite product).

We now define Fourier transform.

**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 *Fourier transform* is a linear operator  $\mathcal{F} : L^1(G) \rightarrow \mathbb{C}^{\widehat{G}}$  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  $\mathcal{F} : M(G) \rightarrow \mathbb{C}^{\widehat{G}}$  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  $\text{Prob}(G)$ . We will also often use the adjoint Fourier transform  $\mathcal{F}^* : M(\widehat{G}) \rightarrow \mathbb{C}^{\widehat{\widehat{G}}}$  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.

First, the Fourier transform  $\mathcal{F} : L^1(G) \rightarrow C_0(\widehat{G}) \subset B(L^2(\widehat{G}))$  can be considered as an extension of a special unitary representation of  $G$  itself; Gelfand representation. It acts on the Hilbert space  $L^2(\widehat{G})$  by multiplication. This representation is Note that as in vector spaces, there is a natural homomorphism from a group to its double dual group.

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

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

$$\begin{array}{ccc} G & \xrightarrow{\Phi} & \widehat{\widehat{G}} \\ \downarrow & & \downarrow \\ M(G) & \xrightarrow{\mathcal{F}^*} & C_b(\widehat{G}). \end{array}$$

In particular,  $\Phi$  is injective.

**Proposition 4.3.** *Property*

- (a)  $\mathcal{F} : L^1(G) \rightarrow C_0(\widehat{G})$  is an embedding with dense image.
- (b)  $\mathcal{F} : M(G) \rightarrow C_b(\widehat{G})$  is an embedding.

## 4.2 The Pontryagin duality

In the previous section, we defined Fourier transform on a locally compact abelian group  $G$  that maps a function on  $G$  to another function on  $\hat{G}$ . Then, repeated Fourier transform will map a function on  $G$  to a function on the double dual  $\hat{\hat{G}}$ . 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 transformed function  $\mathcal{F}f$  on  $\hat{G} = \mathbb{R}$  or  $\mathbb{Z}$  we can reconstruct the original function  $f$  by doing the almost same operation, the adjoint Fourier transform. In other words, although the domain of  $\mathcal{F}^*\mathcal{F}f$  is in principle  $\hat{\hat{G}}$ , but it can be considered as the original function  $f$  on the original group  $G$ . Furthermore, in a suitable setting of function spaces, the adjoint Fourier transform  $\mathcal{F}^*$  plays a role of inverse Fourier transform  $\mathcal{F}^{-1}$ .

We are interested in the generalization for the conditions of the group  $G$  such that the recovery of the original group from the dual group  $\hat{G}$  is successful. This kind of question of recovery is called *duality*, and the most famous result is the Pontryagin duality. Pontryagin proved in 1934 the duality for compact second countable abelian groups, and van Kampen generalized in the next year for the case of locally compact abelian groups. In present the Pontryagin duality refers to the duality result for locally compact abelian groups.

For a locally compact abelian group  $G$ , we can discover in Bochner's theorem an actual mechanism to pullback the doubly transformed function on  $\hat{\hat{G}}$  to the original group  $G$  without any loss of information. To see this, we will reformulate Bochner's theorem as a corollary after the proof of the theorem.

**Theorem 4.4** (Bochner's theorem). *Let  $G$  be a locally compact abelian group, A function  $f : G \rightarrow \mathbb{C}$  is continuous and positive definite if and only if there is a unique non-negative  $\mu \in M(\hat{G})$  such that*

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

for all  $x \in G$ .

*Proof.*

□

**Definition 4.4** (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 4.5.** *Let  $G$  be a locally compact abelian group, and  $\Phi : G \rightarrow \hat{\hat{G}}$  be the canonical homomorphism. Then,  $\Phi^* \circ \mathcal{F}^* : M(\hat{G}) \rightarrow B(G)$  is a well-defined algebra isomorphism, where  $\Phi^*f = f \circ \Phi$ .*

Classical Fourier inversion theorems go further than Bochner's theorem; not only is the Fourier transform bijective, but the inverse is given by its adjoint. We rigorously state them as follows.

**Theorem 4.6** (Fourier inversion). *Let  $G$  be a locally compact abelian group, and  $\hat{G}$  be its dual group. By adjusting the constant of Haar measure on  $\hat{G}$ , we have the following statements:*

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

*Proof.* (a) Without loss of generality, assume  $f \in P(G) \cap L^1(G)$ , where  $P(G)$  denotes the space of all continuous positive definite functions on  $G$ . By the Bochner theorem, there is a non-negative measure  $\mu \in M(\hat{G})$  such that  $\Phi^* \circ \mathcal{F}^*\mu = f$ . Our claim is that there is a Haar measure  $d\chi$  on  $\hat{G}$  such that  $d\mu(\chi) = \mathcal{F}f d\chi$ .

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

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

where  $f \in P(G) \cap L^1(G)$  satisfies  $\mathcal{F}f > 0$  on  $\text{supp } \varphi$ . The existence of such  $f$  is shown as follows:

The linear functional  $I$  is well-defined since

(b) Note that we can slightly modify the Bochner theorem to state that  $\Phi^* \circ \mathcal{F} : M(\hat{\hat{G}}) \rightarrow B(\hat{G})$  is also an algebra isomorphism. From the part (a), we have  $\mathcal{F}\varphi \in L^1(\hat{\hat{G}})$  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  $\mathcal{F}^* \circ \Phi^* \circ \mathcal{F}\varphi = \varphi$  by the injectivity of  $\Phi^* \circ \mathcal{F}$ .  $\square$

**Theorem 4.7** (The Plancherel theorem). *Let  $G$  be a locally compact abelian group, and  $\hat{G}$  be its dual group. Then,*

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

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

*Proof.*  $\square$

**Theorem 4.8** (Pontryagin duality). *Let  $G$  be a locally compact abelian group, and  $\hat{G}$  be its dual group. Then, the canonical homomorphism  $\Phi : G \rightarrow \hat{\hat{G}}$  is a topological isomorphism.*

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

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

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

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

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

*Proof of the Pontryagin duality.* Since we have shown the injectivity of  $\Phi$ , we claim that the image  $\Phi(G)$  is closed and dense in  $\hat{\hat{G}}$  to show the surjectivity of  $\Phi$ .

(Closedness) We first show  $\Phi$  is a topological embedding. Suppose a net  $x_\alpha$  does not converge to  $x$  in  $G$ . We may assume there is an open neighborhood  $U$  of  $x$  in  $G$  such that  $x_\alpha \notin U$  for all  $\alpha$ . By the previous lemma, there is  $\varphi \in L^1(\hat{G})$  such that  $\text{supp } \mathcal{F}^*\varphi \subset U$ , and we have  $\mathcal{F}^*\varphi(x_\alpha) = 0$  for all  $\alpha$  but  $\mathcal{F}^*\varphi(x) \neq 0$ . Then, since  $\mathcal{F}^*\varphi(y) = \int \Phi^*(y)(\chi) \varphi(\chi) d\chi$ , so  $\Phi^*(x_\alpha)$  does not converges to  $\Phi^*(x)$  in the weak\* topology of  $L^\infty(\hat{G})$ , which is same as the topology on  $\hat{\hat{G}} = (L^1(\hat{G}))^\wedge$ .

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

(Density) If  $\alpha(G)$  is not dense in  $\widehat{\widehat{G}}$ , then a non-zero function  $f \in \mathcal{F}^*(M(\widehat{G}))$  vanishes on  $\alpha(G)$  by the previous lemma. For  $\mu \in M(\widehat{G})$  such that  $f = \mathcal{F}^*\mu$ , we have  $\alpha^* \circ \mathcal{F}^*\mu = f|_{\alpha(G)} = 0$ , so  $\mu = 0$  by the Bochner theorem, and it leads a contradiction to  $f \neq 0$ .  $\square$

### 4.3 Cyclic representations and Bochner's theorem

We give in this section a representation-theoretic interpretation of Bochner's theorem: each element of  $M(\widehat{G})$  can be considered as a cyclic representation of  $G$  with a canonical cyclic vector. The following theorem will be defined and proved.

**Theorem 4.9.** *Let  $G$  be a locally compact abelian group. Then, Bochner's theorem allows to have the following one-to-one correspondence:*

$$\left\{ \begin{array}{c} \text{regular Borel} \\ \text{probability measures on } \widehat{G} \end{array} \right\} \xrightarrow{\sim} \left\{ \begin{array}{c} \text{unitary equivalence class of} \\ \text{pointed cyclic representations of } G \end{array} \right\}.$$

The importance of Bochner's theorem in this correspondence relies on its surjectivity, that is, one can construct a desired measure on  $\widehat{G}$  via a continuous positive definite function on  $G$ . In the viewpoint of Chapter 2, Theorem 4.9 can also be recognized as a “convex combination” version of the following correspondence, which is obvious by definition of  $\widehat{G}$ :

$$\left\{ \text{Dirac measures on } \widehat{G} \right\} \xrightarrow{\sim} \left\{ \begin{array}{c} \text{unitary equivalence class of} \\ \text{irreducible representations of } G \end{array} \right\}.$$

**Definition 4.5** (Unitary representation). Let  $G$  be a locally compact abelian group. A *unitary representation* of  $G$  is a group homomorphism  $\rho : G \rightarrow U(H_\rho)$ , where  $U(H_\rho)$  denotes the set of unitary operators on a Hilbert space  $H_\rho$ .

**Definition 4.6** (Cyclic representation). Let  $G$  be a locally compact abelian group. A *cyclic representation* of  $G$  is a unitary representation  $\rho : G \rightarrow U(H_\rho)$  such that there is a vector  $\psi \in H_\rho$  called a *cyclic vector* that satisfies the closed linear span of  $\rho(G)\psi$  is equal to  $H_\rho$ . A *pointed cyclic representation* of  $G$  is a pair  $(\rho, \psi)$  of a cyclic representation  $\rho$  of  $G$  and a unit cyclic vector  $\psi \in H_\rho$ .

**Definition 4.7** (Unitary equivalence). Let  $(\rho_1, \psi_1)$  and  $(\rho_2, \psi_2)$  are pointed cyclic representations of a locally compact abelian group  $G$ . We say they are *unitarily equivalent* if there is a unitary operator  $u : H_{\rho_1} \rightarrow H_{\rho_2}$  such that  $\rho_2(x) = u\rho_1(x)u^*$  for all  $x \in G$  and  $\psi_2 = u\psi_1$ .

**Definition 4.8** (State). A *state* on a  $C^*$ -algebra  $\mathcal{A}$  is a positive linear functional on  $\mathcal{A}$  of norm one.

**Lemma 4.10.** *If  $\omega$  is a state on  $\mathcal{A}$ , then  $\omega(b^*a^*ab) \leq \|a\|^2\omega(b^*b)$  for  $a, b \in \mathcal{A}$ .*

*Proof.* We omit the proof; it requires the existence of approximate identity in  $C^*$ -algebras. See [].  $\square$

**Definition 4.9** (Gelfand-Naimark-Segal representation). Let  $\mathcal{A}$  be a  $C^*$ -algebra, and  $\omega$  be a state on  $\mathcal{A}$ . Define the *left kernel*

$$L_\omega := \{a \in \mathcal{A} : \omega(a^*a) = 0\} = \{a \in \mathcal{A} : \omega(b^*a) = 0 \text{ for all } b \in \mathcal{A}\}$$

of  $\omega$ , which is a closed left ideal of  $\mathcal{A}$ . Then, we can define an inner product on a left  $\mathcal{A}$ -module  $\mathcal{A}/L_\omega$  by  $\langle a + L_\omega, b + L_\omega \rangle := \omega(b^*a)$ , and complete it with the norm induced by this inner product to obtain a Hilbert space, denoted by  $H_\omega$ . For each  $a \in \mathcal{A}$  since  $\pi(a) : \mathcal{A}/L_\omega$  defined by

$$\pi(a)(b + L) := ab + L$$

is bounded by the previous lemma, we can define a map  $\pi_\omega : \mathcal{A} \rightarrow B(H_\omega)$ . It is a cyclic representation with a *canonical cyclic vector*  $\psi$  defined such that  $\omega(a) = \langle a + L_\omega, \psi \rangle$  for all  $a \in \mathcal{A}$ , which we call *Gelfand-Naimark-Segal* representation.

**Example 4.4** (GNS representation for abelian  $C^*$ -algebras). Let  $\mathcal{A}$  be an abelian  $C^*$ -algebra, and let  $\Omega = \sigma(\mathcal{A})$  be the spectrum of  $\mathcal{A}$ . By the Gelfand-Naimark representation theorem, we can identify  $\mathcal{A} = C_0(\Omega)$  with the canonical  $*$ -isomorphism. Then, the states on  $\mathcal{A}$  are exactly the regular Borel probability measures on  $\Omega$ , by the Riesz-Kakutani representation theorem.

Consider a state on  $\mathcal{A}$  that corresponds to a regular Borel probability measure  $\mu$  on  $\Omega$ . Then, the left kernel of  $\mu$  is the kernel of the restriction operator onto the support of  $\mu$ ;

$$L_\mu = \{f \in \mathcal{A} : \int |f|^2 d\mu = 0\} = \{f \in \mathcal{A} : f|_{\text{supp } \mu} = 0\}.$$

Recall here that the complement of the support of a non-negative measure  $\mu$  can be described as the union of all open null sets. Therefore, we obtain the isomorphism  $\mathcal{A}/L_\mu \cong C(\text{supp } \mu)$  and the Hilbert space  $H_\mu = L^2(\text{supp } \mu, \mu)$  by the completion with the inner product  $\langle f, g \rangle := \int \bar{g}f d\mu$  on  $C(\text{supp } \mu)$ . The Gelfand-Naimark-Segal representation of  $\mathcal{A}$  with respect to  $\mu$  is now

$$\pi_\mu : \mathcal{A} \rightarrow B(H_\mu) : f \mapsto M_f,$$

where  $M_f$  denotes the multiplication operator such that  $M_f(g) = fg$ , and the canonical cyclic vector is the unity function on  $\text{supp } \mu$ .

Following the idea of Gelfand-Naimark-Segal representations, we will establish the one-to-one correspondence between pointed cyclic representations and regular Borel probability measures even if  $L^1(G)$  is not a  $C^*$ -algebra. There is also a way to use the Gelfand-Naimark-Segal representation directly with  $C^*$ -algebras: we can complete  $L^1(G)$  with the norm of  $C_0(\hat{G})$  through the Fourier transform to obtain a  $C^*$ -algebra, which is called the *group  $C^*$ -algebra*, and it also shares the exactly same category of representations as  $G$  and  $L^1(G)$ . However, we will not introduce the group  $C^*$ -algebras.

*Proof of Theorem 4.9.* (Well-definedness) We first define the map. Let  $\mu$  be a regular Borel probability measure on  $\hat{G}$ . If we let  $H_\mu := L^2(\text{supp } \mu, \mu)$  and define a representation  $\rho_\mu : G \rightarrow U(H_\mu)$  by the multiplication operators  $\rho_\mu(x) := M_{\Phi(x)}$ , then it has the unity function  $\psi_\mu := \mathbf{1}_{\text{supp } \mu}$  as the a cyclic vector of this representation because  $\Phi(G)$  separates points in  $C(\text{supp } \mu)$  so that the linear span of the functions of the form

$$[\rho_\mu(x)\mathbf{1}_{\text{supp } \mu}](\chi) = \Phi(x)(\chi),$$

which is a unital  $*$ -subalgebra of  $C(\text{supp } \mu)$ , is dense in  $C(\text{supp } \mu)$  and  $L^2(\text{supp } \mu, \mu)$  by the Stone-Weierstrass theorem.

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

$$\mu_2(|f|^2) = \langle f, f \rangle_{H_{\mu_2}} = \int_{\hat{G}} f(\chi) d\mu_2(\chi)$$

(Surjectivity)

□

## 4.4 Further developments for non-abelian groups

Tannaka duality is formally stated in several ways, but the central idea is the same, the category of representations reconstructs the group.

Let  $\mathbf{Rep}(G)$  be the category of finite-dimensional unitary representations of  $G$ . This tensor category has three structures  $\oplus, \otimes, *$ . Consider a forgetful functor

$$F : \mathbf{Rep}(G) \rightarrow \mathbf{Vect}_{\mathbb{C}} : (\pi, H_\pi) \mapsto H_\pi.$$

Let  $\gamma$  be a natural transformation  $F \rightarrow F$  so that for each  $\pi \in \mathbf{Rep}(G)$ , we have a linear map

$$\gamma_\pi : F(\pi) \rightarrow F(\pi)$$

such that

$$\begin{array}{ccc} H_{\pi_1} & \xrightarrow{\gamma_{\pi_1}} & H_{\pi_1} \\ \downarrow \psi & & \downarrow \psi \\ H_{\pi_2} & \xrightarrow{\gamma_{\pi_2}} & H_{\pi_2} \end{array}$$

commutes for every intertwining map  $\psi : (\pi_1, H_{\pi_1}) \rightarrow (\pi_2, H_{\pi_2})$ .

Let

$$\gamma \in \prod_{\pi \in \mathbf{Rep}(G)} \text{End}(H_\pi).$$

Then,

(a)  $\gamma$  is a natural transformation  $F \rightarrow F$  if and only if

$$\gamma_{\pi_1 \oplus \pi_2} = \gamma_{\pi_1} \oplus \gamma_{\pi_2}$$

and

$$\gamma_{\pi_2} = u \gamma_{\pi_1} u^{-1}$$

whenever  $u : H_{\pi_1} \rightarrow H_{\pi_2}$  is a unitary operator such that  $\pi_2(g) = u \pi_1(g) u^{-1}$  for all  $g \in G$ . (It is because every intertwining map is a sum of direct sums of unitary equivalence.)

(b)  $\gamma$  is tensor-preserving if and only if

$$\gamma_{\pi_1 \otimes \pi_2} = \gamma_{\pi_1} \otimes \gamma_{\pi_2}.$$

(c)  $\gamma$  is self-conjugate if and only if

$$\gamma_{\pi^*} = (\gamma_\pi)^*.$$

The set of tensor-preserving and self-conjugate natural transformations is a topological group (if  $G$  is a group). It is called the *Tannakian group*  $T$ .

**Theorem 4.11** (Tannaka duality). *Let  $G$  be a compact group. Then,  $T \cong G$ .*

**Example 4.5** (Pontryagin duality for compact abelian groups). Let  $G$  be a compact abelian group. Since  $G$  is compact, every unitary representation of  $G$  is a direct sum of irreducible unitary representations, and since  $G$  is abelian, every irreducible unitary representation is one-dimensional.

Let  $M$  be the free abelian monoid generated by the dual group  $\hat{G}$ . We have a monoid homomorphism  $\mathbf{Rep}(G) \rightarrow M$  which is an isomorphism up to unitary equivalence. Since  $\text{End}(H_\pi) \cong \mathbb{C}$ , by recognizing  $T$  as

$$T \subset \prod_{\pi \in \mathbf{Rep}(G)} \text{End}(H_\pi) \cong \mathbb{C}^{\mathbf{Rep}(G)},$$

we have

$$T \cong \mathbf{Mon}(M, \mathbb{C}) \cong \mathbf{Grp}(\hat{G}, \mathbb{T}) = \hat{\hat{G}}.$$

By the Tannak duality, the double dual group is isomorphic to the original group.

## References