

## Article

# Approximation of Time-Frequency Shift Equivariant Maps by Neural Networks

Dae Gwan Lee 

Department of Mathematics and Big Data Science, Kumoh National Institute of Technology,  
Gumi 39177, Gyeongsangbuk-do, Republic of Korea; daegwan@kumoh.ac.kr or daegwans@gmail.com

**Abstract:** Based on finite-dimensional time-frequency analysis, we study the properties of time-frequency shift equivariant maps that are generally nonlinear. We first establish a one-to-one correspondence between  $\Lambda$ -equivariant maps and certain phase-homogeneous functions and also provide a reconstruction formula that expresses  $\Lambda$ -equivariant maps in terms of these phase-homogeneous functions, leading to a deeper understanding of the class of  $\Lambda$ -equivariant maps. Next, we consider the approximation of  $\Lambda$ -equivariant maps by neural networks. In the case where  $\Lambda$  is a cyclic subgroup of order  $N$  in  $\mathbb{Z}_N \times \mathbb{Z}_N$ , we prove that every  $\Lambda$ -equivariant map can be approximated by a shallow neural network whose affine linear maps are simply linear combinations of time-frequency shifts by  $\Lambda$ . This aligns well with the proven suitability of convolutional neural networks (CNNs) in tasks requiring translation equivariance, particularly in image and signal processing applications.

**Keywords:** neural networks; equivariance; time-frequency shifts; time-frequency analysis

**MSC:** 68T07; 42C15; 94A12



**Citation:** Lee, D.G. Approximation of Time-Frequency Shift Equivariant Maps by Neural Networks.

*Mathematics* **2024**, *12*, 3704. <https://doi.org/10.3390/math12233704>

Academic Editor: Quanxin Zhu

Received: 27 October 2024

Revised: 18 November 2024

Accepted: 24 November 2024

Published: 26 November 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Over the past decade, machine learning techniques based on deep neural networks, commonly referred to as *deep learning* [1], have achieved significant breakthroughs across a wide range of fields, including image recognition [2,3], speech recognition [4], language translation [5,6], and game playing [7], among others. These advancements are largely driven by the availability of increasingly large training datasets and greater computational resources. Another important factor is the development of specialized neural network architectures, including convolutional neural networks [2], residual networks [3], recurrent networks (notably LSTMs [5]), and transformer networks [6].

A common theme in the design of neural network architectures is the necessity to respect the symmetries inherent in the task at hand. For instance, in image classification, the classification result should remain invariant under small translations of the input image, making convolutional neural networks a suitable choice. Likewise, in audio classification [8], the classification result should be invariant to shifts in time or changes in pitch. In principle, a fully connected neural network can learn to respect such symmetries provided that training data are sufficiently given. Nevertheless, architectures that are inherently aligned with these symmetries tend to exhibit improved generalization and thus show better performance.

In mathematical terms, symmetries can be expressed as follows. Let  $V$  be a vector space and let  $GL(V)$  be the general linear group of  $V$ . For a group  $G$  and a map  $\rho : G \rightarrow GL(V)$ , we say that a map  $F : V \rightarrow V$  is *equivariant under group actions of  $G$*  (or simply  *$G$ -equivariant*) if  $F\rho(\lambda) = \rho(\lambda)F$  for all  $\lambda \in G$ , and *invariant under group actions of  $G$*  (or simply  *$G$ -invariant*) if  $F\rho(\lambda) = F$  for all  $\lambda \in G$ . We will be focusing on the case where  $V$  is a Hilbert space and  $\rho(\lambda)$  is a unitary operator for all  $\lambda \in G$ . (A Hilbert space is a vector space equipped with an inner product that induces a distance function, making it a complete metric space.

Examples of Hilbert spaces include  $\mathbb{R}^d$  and  $\mathbb{C}^d$ , and Hilbert spaces are often regarded as natural generalizations of signal spaces.)

A particularly important and well-studied example of equivariance involves translations. It is well known that translation-equivariant linear operators are exactly the convolution operators (see, e.g., Section 2.3 of [9], Theorem 4.12 of [10], and Theorem 2.17 of [11]), and that convolutional neural networks (CNNs) are well-suited for approximating these operators. As a natural generalization of CNNs, Cohen and Welling [12] introduced the so-called group equivariant convolutional neural networks (GCNNs), which can handle more general symmetry groups than just translations. Later, Cohen et al. [13] developed a general framework for GCNNs on homogeneous spaces such as  $\mathbb{R}^d$  and  $S^2$ , and Yarotsky [14] investigated the approximation of equivariant operators using equivariant neural networks. More recently, Cahill et al. [15] introduced the so-called group-invariant max filters, which are particularly useful for classification tasks involving symmetries, and Balan and Tsoukanis [16,17] constructed stable embeddings on quotient space modulo group action, yielding group-invariant representations via coorbits. Further advances include the work of Huang et al. [18], who designed approximately group-equivariant graph neural networks by focusing on active symmetries, and Blum-Smith and Villar [19], who introduced a method for parameterizing invariant and equivariant functions based on invariant theory. In addition, Wang et al. [20] provided a theoretical analysis of data augmentation and equivariant neural networks applied to non-stationary dynamics forecasting.

In this paper, we are particularly interested in the setting of finite-dimensional time-frequency analysis, which provides a versatile framework for a wide range of signal processing applications, see, e.g., [21,22]. It is known that every linear map from  $\mathbb{C}^N$  to  $\mathbb{C}^N$  can be expressed as a linear combination of compositions of translations and modulations (see (3) below). We consider maps  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  that are generally *nonlinear* and are  $\Lambda$ -equivariant for a given subgroup  $\Lambda$  of  $\mathbb{Z}_N \times \mathbb{Z}_N$ , that is,  $F \circ \pi(k, \ell) = \pi(k, \ell) \circ F$  for all  $(k, \ell) \in \Lambda$ . Here,  $\pi(k, \ell) := M^\ell T^k$  represents the time-frequency shift by  $(k, \ell)$ , where  $T, M : \mathbb{C}^N \rightarrow \mathbb{C}^N$  are the translation and modulation operators defined as  $Tx = (x_{N-1}, x_0, x_1, \dots, x_{N-2})$  and  $Mx = (\omega^0 x_0, \omega^1 x_1, \dots, \omega^{N-1} x_{N-1})$ ,  $\omega := e^{2\pi i/N}$ , for  $x = (x_0, x_1, \dots, x_{N-1}) \in \mathbb{C}^N$ , respectively (see Section 2.1 for further details). For any  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  and any nonzero  $v \in \mathbb{C}^N$ , we define  $F_v : \mathbb{C}^N \rightarrow \mathbb{C}$  by  $F_v(x) = \langle F(x), v \rangle$ ,  $x \in \mathbb{C}^N$ . For any  $\Omega \subset \mathbb{Z}_N$ , we say that a function  $H : \mathbb{C}^N \rightarrow \mathbb{C}$  is  $\Omega$ -phase homogeneous if  $H(e^{2\pi i s/N} x) = e^{2\pi i s/N} H(x)$  for all  $s \in \Omega$  and  $x \in \mathbb{C}^N$ .

We first address the properties of the mapping  $F \mapsto F_v$  from the space of  $\Lambda$ -equivariant functions  $\mathbb{C}^N \rightarrow \mathbb{C}^N$  to the space of certain phase homogeneous functions.

**Theorem 1** (see Theorem 3 below). *Assume that  $\text{span}\{\pi(k, \ell)v : (k, \ell) \in \Lambda\} = \mathbb{C}^N$  for some subgroup  $\Lambda$  of  $\mathbb{Z}_N \times \mathbb{Z}_N$  and some vector  $v \in \mathbb{C}^N$ . Then, the mapping  $F \mapsto F_v$  is an injective map from the space of  $\Lambda$ -equivariant functions  $\mathbb{C}^N \rightarrow \mathbb{C}^N$  to the space of  $\Omega_\Lambda$ -phase homogeneous functions  $\mathbb{C}^N \rightarrow \mathbb{C}$ , where  $\Omega_\Lambda := \{k\ell' \bmod N : (k, \ell), (k', \ell') \in \Lambda\}$ . Moreover, if  $\{\pi(k, \ell)u\}_{(k, \ell) \in \Lambda}$  is a dual frame of  $\{\pi(k, \ell)v\}_{(k, \ell) \in \Lambda}$  in  $\mathbb{C}^N$ , then a  $\Lambda$ -equivariant function  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  can be expressed as*

$$F(x) = \sum_{(k, \ell) \in \Lambda} e^{-2\pi i k \ell / N} F_v(\pi(-k, -\ell)x) \pi(k, \ell)u.$$

*If  $|\Lambda| = N$ , then the mapping  $F \mapsto F_v$  is a bijective map from the space of  $\Lambda$ -equivariant functions  $\mathbb{C}^N \rightarrow \mathbb{C}^N$  to the space of  $\Omega_\Lambda$ -phase homogeneous functions  $\mathbb{C}^N \rightarrow \mathbb{C}$ .*

We then consider the approximation of  $\Lambda$ -equivariant maps. In particular, we show that if  $\Lambda$  is a cyclic subgroup of order  $N$  in  $\mathbb{Z}_N \times \mathbb{Z}_N$ , then every  $\Lambda$ -equivariant map can be easily approximated by a shallow neural network whose affine linear maps consist of linear combinations of time-frequency shifts by  $\Lambda$ .

**Theorem 2** (see Theorem 5 below). Assume that  $\sigma : \mathbb{C} \rightarrow \mathbb{C}$  is shallow universal and satisfies  $\sigma(e^{\pi i/N} z) = e^{\pi i/N} \sigma(z)$  for all  $z \in \mathbb{C}$ . Let  $\Lambda = \{(0,0), (1,s), \dots, (N-1, (N-1)s)\}$  for some  $s \in \{0, 1, \dots, N-1\}$ . Then, any continuous  $\Lambda$ -equivariant map  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  can be approximated (uniformly on compact sets) by a shallow neural network

$$x \mapsto \sum_{j=1}^J c_j \sigma(A_j x + b_j v),$$

where  $A_j \in \text{span}\{\pi(k, \ell) : (k, \ell) \in \Lambda\}$ ,  $b_j \in \mathbb{C}$  for  $j = 1, \dots, J$ , and  $v \in \mathbb{C}^N$  satisfies  $\pi(k, \ell)v = e^{k\ell\pi i/N} v$  for all  $(k, \ell) \in \Lambda$ . Moreover, every map of this form is  $\Lambda$ -equivariant.

In the case  $s = 0$ , i.e.,  $\Lambda = \{(0,0), (1,0), \dots, (N-1,0)\}$ , the  $\Lambda$ -equivariant maps  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  are precisely those that are translation equivariant, meaning that  $FT = TF$ . Furthermore, if  $F$  is linear, then  $F$  is just a convolutional map, which can be expressed as a linear combination of  $T^k$ ,  $k = 0, \dots, N-1$ , or simply as an  $N \times N$  circulant matrix. If  $F$  is nonlinear, then Theorem 2 shows that  $F$  can be approximated by a shallow neural network whose affine linear maps are convolutional maps, i.e., by a shallow convolutional neural network. This agrees with the well-established fact that convolutional neural networks (CNNs) are particularly well-suited for applications involving translation equivariance, especially in image and signal processing.

### Organization of the Paper

In Section 2, we begin by reviewing some basic properties of time-frequency shift operators, followed by a discussion on time-frequency group equivariant maps, and then prove our first main result, Theorem 1, which establishes a 1:1 correspondence between  $\Lambda$ -equivariant maps and certain phase-homogeneous functions. Section 3 is devoted to the approximation of  $\Lambda$ -equivariant maps. We first discuss the embedding of  $\Lambda$  into the Weyl–Heisenberg group, which allows for the use of tools from group representation theory. (The finite Weyl–Heisenberg group  $\mathbb{H}_N$  is the set  $\mathbb{Z}_N \times \mathbb{Z}_N \times \mathbb{Z}_N$  equipped with group operation  $(k, \ell, s) + (k', \ell', s') := (k + k', \ell + \ell', s + s' - k\ell')$ . The noncommutativity of  $\mathbb{H}_N$  plays an important role in finite-dimensional time-frequency analysis; see, e.g., [21,23].) After reviewing key concepts from group representation theory, we consider the case of cyclic subgroups of  $\mathbb{Z}_N \times \mathbb{Z}_N$ , where group representations can be defined directly without embedding into the Weyl–Heisenberg group. Section 3 concludes with the proof of our second main result, Theorem 2, which establishes the approximation of  $\Lambda$ -equivariant maps by a shallow neural network whose affine linear maps consist of linear combinations of time-frequency shifts by  $\Lambda$ .

## 2. Time-Frequency Shift Equivariant Maps

### 2.1. Time-Frequency Shift Operators

We define the *translation (time shift) operator*  $T : \mathbb{C}^N \rightarrow \mathbb{C}^N$  by

$$Tx = (x_{N-1}, x_0, x_1, \dots, x_{N-2}), \quad x = (x_0, x_1, \dots, x_{N-1}) \in \mathbb{C}^N,$$

and the *modulation (frequency shift) operator*  $M : \mathbb{C}^N \rightarrow \mathbb{C}^N$  by

$$Mx = (\omega^0 x_0, \omega^1 x_1, \dots, \omega^{N-1} x_{N-1}) \quad \text{with } \omega := e^{2\pi i/N}.$$

These operators are linear unitary operators, which can be represented by  $N \times N$  unitary matrices:

$$T = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \quad M = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & \omega & 0 & \cdots & 0 & 0 \\ 0 & 0 & \omega^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \omega^{N-2} & 0 \\ 0 & 0 & 0 & \cdots & 0 & \omega^{N-1} \end{bmatrix}.$$

Note that since  $T^N = M^N = I_N$ , we have  $T^{N+k} = T^k$  and  $M^{N+\ell} = M^\ell$  for all integers  $k$  and  $\ell$ . The time-frequency shift by  $(k, \ell) \in \mathbb{Z}_N \times \mathbb{Z}_N$  is defined by  $\pi(k, \ell) := M^\ell T^k$ . Since  $T, M : \mathbb{C}^N \rightarrow \mathbb{C}^N$  are linear unitary operators, the operator  $\pi(k, \ell)$  is also linear and unitary.

For a Hilbert space  $\mathcal{H}$ , we will denote the class of all linear operators on  $\mathcal{H}$  by  $\mathcal{L}(\mathcal{H})$ , and the class of all linear unitary operators on  $\mathcal{H}$  by  $\mathcal{U}(\mathcal{H})$ .

**Proposition 1.** For any  $k, \ell = 0, \dots, N-1$ , we have

$$M^\ell T^k = \omega^{k\ell} T^k M^\ell. \quad (1)$$

This implies  $(M^\ell T^k)(M^q T^p) = \omega^{-kq} M^{\ell+q} T^{k+p} = \omega^{\ell p - kq} (M^q T^p)(M^\ell T^k)$  for  $k, \ell, p, q = 0, \dots, N-1$ , and consequently, the operators  $M^\ell T^k$  and  $M^q T^p$  commute if and only if  $[(k, \ell), (p, q)] := \ell p - kq$  is a multiple of  $N$ . Moreover, for any  $k, \ell = 0, \dots, N-1$ , we have

$$(M^\ell T^k)^{-1} = \omega^{-k\ell} M^{-\ell} T^{-k}, \quad (2)$$

that is,  $\pi(k, \ell)^{-1} = \omega^{-k\ell} \pi(-k, -\ell)$ .

**Proof.** The relation (1) is easily seen by computation. Using (1), we obtain

$$(M^\ell T^k)^{-1} = T^{-k} M^{-\ell} = \omega^{-k\ell} M^{-\ell} T^{-k},$$

which is exactly (2).  $\square$

**Remark 1.** The definition of  $[\cdot, \cdot]$  remains unchanged for time-frequency shift operators of the form  $T^k M^\ell$  with  $(k, \ell) \in \mathbb{Z}_N \times \mathbb{Z}_N$ . Indeed, (1) implies  $(T^k M^\ell)(T^p M^q) = \omega^{\ell p - kq} (T^p M^q)(T^k M^\ell)$  for  $k, \ell, p, q = 0, \dots, N-1$ , and consequently, the operators  $T^k M^\ell$  and  $T^p M^q$  commute if and only if  $[(k, \ell), (p, q)] = \ell p - kq$  is a multiple of  $N$ .

For a subgroup  $\Lambda$  of  $\mathbb{Z}_N \times \mathbb{Z}_N$ , its adjoint group is defined by

$$\begin{aligned} \Lambda^\circ &:= \{(p, q) \in \mathbb{Z}_N \times \mathbb{Z}_N : (M^\ell T^k)(M^q T^p) = (M^q T^p)(M^\ell T^k) \text{ for all } (k, \ell) \in \Lambda\} \\ &= \{(p, q) \in \mathbb{Z}_N \times \mathbb{Z}_N : \ell p - kq \in N\mathbb{Z} \text{ for all } (k, \ell) \in \Lambda\}. \end{aligned}$$

Since  $\{M^\ell T^k : k, \ell = 0, \dots, L-1\}$  forms a basis for  $\mathcal{L}(\mathbb{C}^N)$  (see, e.g., Lemma 1 of [24]), every linear operator  $F \in \mathcal{L}(\mathbb{C}^N)$  can be expressed as

$$F = \sum_{k, \ell=0}^{L-1} a_{k, \ell} M^\ell T^k \quad \text{for some } a_{k, \ell} \in \mathbb{C}, \quad k, \ell = 0, \dots, L-1. \quad (3)$$

If  $F$  commutes with  $M^\ell T^k$  for  $(k, \ell) \in \Lambda$ , then we must have  $a_{k, \ell} = 0$  for  $(k, \ell) \notin \Lambda^\circ$ , so that

$$F = \sum_{(k, \ell) \in \Lambda^\circ} a_{k, \ell} M^\ell T^k.$$

Therefore, the commutant (or centralizer) of  $(\pi, \Lambda)$  (see, e.g., Proposition 4.14 of [25]) is given by

$$\begin{aligned} C(\pi, \Lambda) &:= \{F \in \mathcal{L}(\mathbb{C}^N) : F\pi(k, \ell) = \pi(k, \ell)F \text{ for all } (k, \ell) \in \Lambda\} \\ &= \text{span}\{\pi(k, \ell) : (k, \ell) \in \Lambda^\circ\}. \end{aligned} \quad (4)$$

**Remark 2.** For a subgroup  $\Lambda$  of  $\mathbb{Z}_L \times \mathbb{Z}_L$ , its adjoint group  $\Lambda^\circ$  has cardinality  $L^2/|\Lambda|$ . While this fact is somewhat considered folklore, we could not find a suitable reference in the literature, so we provide a short proof of this fact in Appendix A.

## 2.2. $\Lambda$ -Equivariant Maps

**Definition 1.** For any  $\Lambda \subset \mathbb{Z}_N \times \mathbb{Z}_N$ , we say that a map  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  is  $\Lambda$ -equivariant if

$$F \circ \pi(k, \ell) = \pi(k, \ell) \circ F \quad \text{for all } (k, \ell) \in \Lambda. \quad (5)$$

Clearly, the set of  $\Lambda$ -equivariant linear maps is precisely  $C(\pi, \Lambda)$ , the commutant of  $(\pi, \Lambda)$ . According to (4), every  $(\pi, \Lambda)$ -equivariant linear map is of the form

$$\sum_{(k, \ell) \in \Lambda^\circ} a_{k, \ell} M^\ell T^k \quad \text{for some } \{a_{k, \ell}\}_{(k, \ell) \in \Lambda^\circ} \in \mathbb{C}^{\Lambda^\circ}.$$

Since the case of linear maps is obvious, our consideration of  $\Lambda$ -equivariant maps will be focused on nonlinear maps.

We first observe some necessary conditions for  $\Lambda$ -equivariance.

**Proposition 2.** Let  $\Lambda$  be a subgroup of  $\mathbb{Z}_N \times \mathbb{Z}_N$ , and assume that  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  is  $\Lambda$ -equivariant. If  $\Lambda$  contains  $(k, \ell)$  and  $(k', \ell')$  with  $s = \gcd(k\ell', N)$ , then

$$F(e^{2\pi i s/N} x) = e^{2\pi i s/N} F(x) \quad \text{for all } x \in \mathbb{C}^N. \quad (6)$$

**Proof.** Note that since  $\Lambda$  is a subgroup, we have  $(k + k', \ell + \ell') \in \Lambda$ . Using Proposition 1 and (5), we have

$$\begin{aligned} \omega^{-k\ell'} \pi(k + k', \ell + \ell') F(x) &= \pi(k, \ell) \pi(k', \ell') F(x) = F(\pi(k, \ell) \pi(k', \ell') x) \\ &= F(\omega^{-k\ell'} \pi(k + k', \ell + \ell') x) \\ &= \pi(k + k', \ell + \ell') F(\omega^{-k'\ell} x), \end{aligned}$$

so that  $\omega^{-k\ell'} F(x) = F(\omega^{-k'\ell} x)$ . Since  $\gcd(k\ell', N) = s$ , there exist some  $p, q \in \mathbb{Z}$  with  $p(-k'\ell) + qN = s$ . In fact, we can choose  $p \in \{0, \dots, N-1\}$  such that  $p(-k'\ell) \equiv s \pmod{N}$ . Then, for any  $x \in \mathbb{C}^N$ , we have

$$\omega^s F(x) = (\omega^{-k\ell'})^p F(x) = (\omega^{-k\ell'})^{p-1} F(\omega^{-k'\ell} x) = \dots = F((\omega^{-k'\ell})^p x) = F(\omega^s x),$$

which is equivalent to (6).  $\square$

It is easily seen that for a subgroup  $\Lambda$  of  $\mathbb{Z}_N \times \mathbb{Z}_N$ , the set

$$\Omega_\Lambda := \{k\ell' \pmod{N} : (k, \ell), (k', \ell') \in \Lambda\}$$

forms a subgroup of  $\mathbb{Z}_N$ ; in fact,  $\Omega_\Lambda = s_0 \mathbb{Z} / N\mathbb{Z}$ , where  $s_0 := \min\{\gcd(k\ell', N) : (k, \ell), (k', \ell') \in \Lambda\}$ . This leads to the following definition.

**Definition 2.** Let  $m, n \in \mathbb{N}$  and  $N \in \mathbb{N}$ . For any  $\Omega \subset \mathbb{Z}_N$ , we say that a map  $F : \mathbb{C}^n \rightarrow \mathbb{C}^m$  is  $\Omega$ -phase homogeneous if

$$F(e^{2\pi i s/N} x) = e^{2\pi i s/N} F(x) \quad \text{for all } s \in \Omega, x \in \mathbb{C}^n.$$

**Definition 3.** For any  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  and any nonzero  $v \in \mathbb{C}^N$ , we define  $F_v : \mathbb{C}^N \rightarrow \mathbb{C}$  by

$$F_v(x) = \langle F(x), v \rangle, \quad x \in \mathbb{C}^N.$$

We now present our first main theorem, which addresses the properties of the mapping  $F \mapsto F_v$ . Note that if  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  is  $\Lambda$ -equivariant for a subgroup  $\Lambda$  of  $\mathbb{Z}_N \times \mathbb{Z}_N$ , then it is  $\Omega_\Lambda$ -phase homogeneous by Proposition 2, and so is  $F_v$ .

Before stating the theorem, we note that  $\{\pi(k, \ell)v : (k, \ell) \in \mathbb{Z}_N \times \mathbb{Z}_N\}$  is a tight frame for  $\mathbb{C}^N$  whenever  $v \neq 0$  (see, e.g., Proposition 2 of [24]). Moreover, there exists a nonzero

vector  $v \in \mathbb{C}^N$  such that every  $N$  elements of  $\{\pi(k, \ell)v : (k, \ell) \in \mathbb{Z}_N \times \mathbb{Z}_N\}$  are linearly independent in  $\mathbb{C}^N$ . In fact, such vectors form a dense open set  $W_N$  of full measure in  $\mathbb{C}^N$  (see Theorem 1 of [24]). If  $\Lambda \subset \mathbb{Z}_N \times \mathbb{Z}_N$  is a set of cardinality at least  $N$ , then for any  $v \in W_N$  we have  $\text{span}\{\pi(k, \ell)v : (k, \ell) \in \Lambda\} = \mathbb{C}^N$ , in which case  $\{\pi(k, \ell)v\}_{(k, \ell) \in \Lambda}$  forms a frame of  $\mathbb{C}^N$ .

**Theorem 3.** Assume that  $\text{span}\{\pi(k, \ell)v : (k, \ell) \in \Lambda\} = \mathbb{C}^N$  for some subgroup  $\Lambda$  of  $\mathbb{Z}_N \times \mathbb{Z}_N$  and some vector  $v \in \mathbb{C}^N$ . Then, the mapping  $F \mapsto F_v$  is an injective map from the space of  $\Lambda$ -equivariant functions  $\mathbb{C}^N \rightarrow \mathbb{C}^N$  to the space of  $\Omega_\Lambda$ -phase homogeneous functions  $\mathbb{C}^N \rightarrow \mathbb{C}$ . Moreover, if  $\{\pi(k, \ell)u\}_{(k, \ell) \in \Lambda}$  is a dual frame of  $\{\pi(k, \ell)v\}_{(k, \ell) \in \Lambda}$  in  $\mathbb{C}^N$ , then a  $\Lambda$ -equivariant function  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  can be expressed as

$$F(x) = \sum_{(k, \ell) \in \Lambda} \omega^{-k\ell} F_v(\pi(-k, -\ell)x) \pi(k, \ell)u. \quad (7)$$

If  $|\Lambda| = N$ , then the mapping  $F \mapsto F_v$  is a bijective map from the space of  $\Lambda$ -equivariant functions  $\mathbb{C}^N \rightarrow \mathbb{C}^N$  to the space of  $\Omega_\Lambda$ -phase homogeneous functions  $\mathbb{C}^N \rightarrow \mathbb{C}$ .

**Proof.** To prove the injectivity of  $F \mapsto F_v$ , suppose that  $F_v = H_v$  for some  $\Lambda$ -equivariant functions  $F, H : \mathbb{C}^N \rightarrow \mathbb{C}^N$ . Then, for any  $(k, \ell) \in \Lambda$  and  $x \in \mathbb{C}^N$ , we have

$$\begin{aligned} \langle F(x), \pi(k, \ell)v \rangle &= \langle \pi(k, \ell)^{-1}F(x), v \rangle = \langle \omega^{-k\ell} \pi(-k, -\ell)F(x), v \rangle \\ &= \langle \omega^{-k\ell} F(\pi(-k, -\ell)x), v \rangle = \omega^{-k\ell} F_v(\pi(-k, -\ell)x) \\ &= \omega^{-k\ell} H_v(\pi(-k, -\ell)x) = \langle \omega^{-k\ell} H(\pi(-k, -\ell)x), v \rangle \\ &= \langle \omega^{-k\ell} \pi(-k, -\ell)H(x), v \rangle = \langle \pi(k, \ell)^{-1}H(x), v \rangle \\ &= \langle H(x), \pi(k, \ell)v \rangle. \end{aligned}$$

Since  $\{\pi(k, \ell)v : (k, \ell) \in \Lambda\}$  is complete in  $\mathbb{C}^N$ , we obtain that  $F(x) = H(x)$  for all  $x \in \mathbb{C}^N$ .

Now, let  $\{\pi(k, \ell)u\}_{(k, \ell) \in \Lambda}$  be a dual frame of  $\{\pi(k, \ell)v\}_{(k, \ell) \in \Lambda}$  in  $\mathbb{C}^N$ , which means that

$$z = \sum_{(k, \ell) \in \Lambda} \langle z, \pi(k, \ell)v \rangle \pi(k, \ell)u, \quad z \in \mathbb{C}^N.$$

Then, for any  $x \in \mathbb{C}^N$ , we have

$$\begin{aligned} \sum_{(k, \ell) \in \Lambda} \omega^{-k\ell} F_v(\pi(-k, -\ell)x) \pi(k, \ell)u &= \sum_{(k, \ell) \in \Lambda} \omega^{-k\ell} \langle F(\pi(-k, -\ell)x), v \rangle \pi(k, \ell)u \\ &= \sum_{(k, \ell) \in \Lambda} \langle \omega^{-k\ell} \pi(-k, -\ell)F(x), v \rangle \pi(k, \ell)u \\ &= \sum_{(k, \ell) \in \Lambda} \langle \pi(k, \ell)^{-1}F(x), v \rangle \pi(k, \ell)u \\ &= \sum_{(k, \ell) \in \Lambda} \langle F(x), \pi(k, \ell)v \rangle \pi(k, \ell)u \\ &= F(x), \end{aligned}$$

which establishes (7).

Finally, assume that  $|\Lambda| = N$ . Then,  $\{\pi(k, \ell)v\}_{(k, \ell) \in \Lambda}$  forms a Riesz basis for  $\mathbb{C}^N$ , so there exists a unique dual Riesz basis  $\{\pi(k, \ell)u\}_{(k, \ell) \in \Lambda}$  of  $\{\pi(k, \ell)v\}_{(k, \ell) \in \Lambda}$  in  $\mathbb{C}^N$ , which is necessarily biorthogonal to  $\{\pi(k, \ell)v\}_{(k, \ell) \in \Lambda}$  (see, e.g., [26]). To prove the surjectivity of  $F \mapsto F_v$ , we pick any  $\Omega_\Lambda$ -phase homogeneous function  $g : \mathbb{C}^N \rightarrow \mathbb{C}$  and set  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  by

$$F(x) := \sum_{(k, \ell) \in \Lambda} \omega^{-k\ell} g(\pi(-k, -\ell)x) \pi(k, \ell)u, \quad x \in \mathbb{C}^N.$$



Then, for any  $(p, q) \in \Lambda$  and  $x \in \mathbb{C}^N$ , we have

$$\begin{aligned}
 F(\pi(p, q)x) &= \sum_{(k, \ell) \in \Lambda} \omega^{-k\ell} g(\pi(-k, -\ell)\pi(p, q)x) \pi(k, \ell)u \\
 &= \sum_{(k, \ell) \in \Lambda} \omega^{-(k+p)(\ell+q)} g(\pi(-k-p, -\ell-q)\pi(p, q)x) \pi(k+p, \ell+q)u \\
 &\stackrel{(8)}{=} \sum_{(k, \ell) \in \Lambda} \omega^{-(k+p)(\ell+q)} g(\omega^{-(k+p)q} \pi(-k, -\ell)x) \omega^{p\ell} \pi(p, q)\pi(k, \ell)u \\
 &= \sum_{(k, \ell) \in \Lambda} \omega^{-(k+p)(\ell+q)} \omega^{-(k+p)q} g(\pi(-k, -\ell)x) \omega^{p\ell} \pi(p, q)\pi(k, \ell)u \\
 &= \sum_{(k, \ell) \in \Lambda} \omega^{-k\ell} g(\pi(-k, -\ell)x) \pi(p, q)\pi(k, \ell)u \\
 &= \pi(p, q) \sum_{(k, \ell) \in \Lambda} \omega^{-k\ell} g(\pi(-k, -\ell)x) \pi(k, \ell)u \\
 &= \pi(p, q)F(x),
 \end{aligned}$$

which shows that  $F$  is  $\Lambda$ -equivariant. Since  $\{\pi(k, \ell)u\}_{(k, \ell) \in \Lambda}$  and  $\{\pi(k, \ell)v\}_{(k, \ell) \in \Lambda}$  are biorthogonal, it holds for any  $x \in \mathbb{C}^N$  that

$$\begin{aligned}
 F_v(x) &= \langle F(x), v \rangle = \left\langle \sum_{(k, \ell) \in \Lambda} \omega^{-k\ell} g(\pi(-k, -\ell)x) \pi(k, \ell)u, v \right\rangle \\
 &= \left\langle \sum_{(k, \ell) \in \Lambda} g(\omega^{-k\ell} \pi(-k, -\ell)x) \pi(k, \ell)u, v \right\rangle \\
 &= \left\langle \sum_{(k, \ell) \in \Lambda} g(\pi(k, \ell)^{-1}x) \pi(k, \ell)u, v \right\rangle \\
 &= g(x).
 \end{aligned}$$

Hence, we conclude that the mapping  $F \mapsto F_v$  is also surjective.  $\square$

**Remark 3.** As one would expect, the mapping  $F \mapsto F_v$  is not surjective if  $|\Lambda| > N$ . Indeed, if  $|\Lambda| > N$  and  $\text{span}\{\pi(k, \ell)v : (k, \ell) \in \Lambda\} = \mathbb{C}^N$ , then there are many dual frames of  $\{\pi(k, \ell)v : (k, \ell) \in \Lambda\}$  in  $\mathbb{C}^N$ . If  $g = F_v$  for some  $F$  and  $v$ , then for any dual frames  $\{\pi(k, \ell)w\}_{(k, \ell) \in \Lambda}$  and  $\{\pi(k, \ell)\tilde{w}\}_{(k, \ell) \in \Lambda}$  of  $\{\pi(k, \ell)v\}_{(k, \ell) \in \Lambda}$  we have

$$\sum_{(k, \ell) \in \Lambda} \omega^{-k\ell} g(\pi(-k, -\ell)x) \pi(k, \ell)w = F(x) = \sum_{(k, \ell) \in \Lambda} \omega^{-k\ell} g(\pi(-k, -\ell)x) \pi(k, \ell)\tilde{w}$$

for all  $x \in \mathbb{C}^N$  by (7). Certainly, not every  $\Omega_\Lambda$ -phase homogeneous function  $g : \mathbb{C}^N \rightarrow \mathbb{C}$  satisfies this property.

### 3. Approximation of $\Lambda$ -Equivariant Maps

In this section, we consider an approximation of continuous  $\Lambda$ -equivariant maps  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  that are generally nonlinear, where  $\Lambda$  is a subgroup of  $\mathbb{Z}_N \times \mathbb{Z}_N$  and the  $\Lambda$ -equivariance is defined by (5). For instance, the map  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  given by  $F(x) = \|x\|^p x$  with  $p > 0$ , is a nonlinear continuous  $\Lambda$ -equivariant map.

As seen in Section 2.2 (particularly in Theorem 3 and its proof), working with the time-frequency shift operators  $\pi(k, \ell)$ ,  $(k, \ell) \in \Lambda$ , usually requires careful bookkeeping of extra multiplicative phase factors due to the non-commutativity of  $T$  and  $M$ . (The non-commutativity of  $T$  and  $M$  can often be frustrating. However, it is precisely this

non-commutativity that has given rise to the deep and rich theory of time-frequency analysis [23].) In fact, the map

$$\pi|_{\Lambda} : \Lambda \rightarrow \mathcal{U}(\mathbb{C}^N), \quad (k, \ell) \mapsto M^{\ell} T^k,$$

is generally not a group homomorphism; indeed,

$$\pi(k, \ell)\pi(k', \ell') = e^{-2\pi i k \ell' / N} \pi(k + k', \ell + \ell') \quad (8)$$

is equal to  $\pi(k + k', \ell + \ell')$  only if  $k\ell'$  is a multiple of  $N$  (see Proposition 1). (Although  $\pi|_{\Lambda}$  is not a group homomorphism and thus not a group representation, it is often referred to as a *projective group representation* of  $G$  on  $\mathbb{C}^N$ . In general, a map  $\rho : G \rightarrow \mathcal{U}(\mathcal{H})$  is called a *projective group representation* of  $G$  on  $\mathcal{H}$  if for each pair of  $g_1, g_2 \in G$ , there exists a unimodular  $c(g_1, g_2) \in \mathbb{C}$  such that  $\rho(g_1 g_2) = c(g_1, g_2) \rho(g_1) \rho(g_2)$ ; see, e.g., [25].) Obviously, the computations involved would be simplified significantly if  $\pi|_{\Lambda}$  were a group homomorphism in general. Note that, as mentioned in Section 1, a group homomorphism  $\rho : G \rightarrow \mathcal{U}(\mathcal{H})$  whose images are unitary operators on  $\mathcal{H}$  is called a (*unitary*) *group representation* of  $G$  on  $\mathcal{H}$ , where  $G$  is a group and  $\mathcal{H}$  is a separable Hilbert space. Therefore, the map  $\pi|_{\Lambda}$  would be a unitary representation if it were a group homomorphism.

In the following, we first discuss a systematic method of avoiding such extra multiplicative phase factors by embedding  $\Lambda \subset \mathbb{Z}_N \times \mathbb{Z}_N$  into the Weyl–Heisenberg group. After briefly reviewing essential concepts on group representations and neural networks, we consider cyclic subgroups of  $\mathbb{Z}_N \times \mathbb{Z}_N$ , in which case the map  $\pi|_{\Lambda}$  can be replaced by a unitary group representation. We show that if  $\Lambda$  is a cyclic subgroup of  $\mathbb{Z}_N \times \mathbb{Z}_N$ , then any  $\Lambda$ -equivariant map  $\mathbb{C}^N \rightarrow \mathbb{C}^N$  can be approximated with shallow neural networks involving the adjoint group  $\Lambda^{\circ}$ , which has significantly fewer degrees of freedom compared with standard shallow neural networks.

### 3.1. Embedding of $\Lambda$ into the Weyl–Heisenberg Group

To avoid the bookkeeping of extra multiplicative phase factors, we can simply embed the subgroups of  $\mathbb{Z}_N \times \mathbb{Z}_N$  into the finite Weyl–Heisenberg group  $\mathbb{H}_N = \mathbb{Z}_N \times \mathbb{Z}_N \times \mathbb{Z}_N$ , on which group representations can be defined. There exists a group representation  $\tau : \mathbb{H}_N \rightarrow \mathcal{U}(\mathbb{C}^N)$ , known as the *Schrödinger representation*, which satisfies  $\tau(k, \ell, 0) = \pi(k, \ell)$  for all  $(k, \ell) \in \mathbb{Z}_N \times \mathbb{Z}_N$ . In fact, for any subgroup  $\Lambda$  of  $\mathbb{Z}_N \times \mathbb{Z}_N$  and any subgroup  $\Omega$  of  $\mathbb{Z}_N$  containing  $\Omega_{\Lambda} := \{k\ell' \bmod N : (k, \ell), (k', \ell') \in \Lambda\}$ , the map

$$\tau : \Lambda \times \Omega \rightarrow \mathcal{U}(\mathcal{H}), \quad \tau(k, \ell, s) = e^{2\pi i s / N} M^{\ell} T^k, \quad (9)$$

is a group representation of  $G = \Lambda \times \Omega$  on  $\mathbb{C}^N$ , with the group operation on  $G$  given by

$$(k, \ell, s) + (k', \ell', s') := (k + k', \ell + \ell', s + s' - k\ell').$$

Clearly, we have  $\tau(k, \ell, 0) = M^{\ell} T^k = \pi(k, \ell)$  for all  $(k, \ell) \in \Lambda$ .

It is clear that a map  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  is  $\Lambda$ -equivariance in the sense of (5) if and only if it is  $(\tau, \Lambda \times \{0\})$ -equivariant in the sense of Definition 4. Moreover, in this case, Proposition 2 implies that  $F$  is  $\Omega_{\Lambda}$ -phase homogeneous, which is equivalent to  $F \circ \tau(0, 0, s) = \tau(0, 0, s) \circ F$  for all  $s \in \Omega_{\Lambda}$ . Consequently, we have the following proposition.

**Proposition 3.** *For any subgroup  $\Lambda$  of  $\mathbb{Z}_N \times \mathbb{Z}_N$  and any  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$ , the following are equivalent.*

- (i)  $F$  is  $\Lambda$ -equivariant;
- (ii)  $F$  is  $(\tau, \Lambda \times \{0\})$ -equivariant;
- (iii)  $F$  is  $(\tau, \Lambda \times \Omega_{\Lambda})$ -equivariant.



Using the true group representation  $\tau$  instead of  $\pi|_{\Lambda}$  allows us to avoid the tedious bookkeeping of extra multiplicative phase factors. Note, however, that  $\tau$  requires three input parameters, while  $\pi|_{\Lambda}$  involves only two. In fact, the description of the extra phase factors is simply transferred to the third parameter of  $\tau$ . Nevertheless, an important advantage of using  $\tau$  instead of  $\pi|_{\Lambda}$  is that it allows for the use of tools from group representation theory.

### 3.2. Group Representations and Neural Networks

In this section, we review some concepts and tools from group representation theory and introduce the so-called  $\mathfrak{h}$ -transform and its inverse transform for later use. We also review the basic structure of neural networks and the universal approximation theorem.

We assume that  $G$  is a finite group, and consider maps of the form  $F : \mathcal{H} \rightarrow \mathcal{H}$ , where  $\mathcal{H}$  is a finite-dimensional Hilbert spaces on which a unitary representation  $\rho$  of  $G$  is defined. This means that for each  $\lambda \in G$ , the map  $\rho(\lambda) : \mathcal{H} \rightarrow \mathcal{H}$  is a linear unitary operator, and that  $\rho : G \rightarrow \mathcal{U}(\mathcal{H})$  is a group homomorphism, i.e.,  $\rho(\lambda_1 \lambda_2) = \rho(\lambda_1) \rho(\lambda_2)$  for all  $\lambda_1, \lambda_2 \in G$ . Let us formally state the definition of equivariance and invariance in this setting.

**Definition 4** (Equivariance and Invariance). *For a group  $G$  and a unitary representation  $\rho$  of  $G$  on a Hilbert space  $\mathcal{H}$ , we say that a map  $F : \mathcal{H} \rightarrow \mathcal{H}$  is*

- $(\rho, G)$ -equivariant if  $F\rho(\lambda) = \rho(\lambda)F$  for all  $\lambda \in G$ ;
- $(\rho, G)$ -invariant if  $F\rho(\lambda) = F$  for all  $\lambda \in G$ .

Note that a  $(\rho, G)$ -equivariant/invariant map  $F : \mathcal{H} \rightarrow \mathcal{H}$  is not necessarily linear or bounded.

**Definition 5.** *For a group  $G$ , the left translation of a vector  $x \in \mathbb{C}^G$  by  $\lambda \in G$  is given by*

$$L_{\lambda}x(v) := x(\lambda^{-1}v) \quad \text{for } v \in G.$$

In fact, the map  $\lambda \mapsto L_{\lambda}$  is a group homomorphism from  $G$  to  $\mathcal{U}(\mathbb{C}^G)$ , that is,  $L_{\lambda_1 \lambda_2} = L_{\lambda_1} L_{\lambda_2}$  for all  $\lambda_1, \lambda_2 \in G$ , and therefore, it induces a group representation of  $G$  on  $\mathbb{C}^G$ . We say that a map  $\Phi : \mathbb{C}^G \rightarrow \mathbb{C}^G$  is left  $G$ -translation equivariant if  $\Phi L_{\lambda} = L_{\lambda} \Phi$  for all  $\lambda \in G$ .

**Definition 6.** *Let  $G$  be a group and let  $\rho$  be a unitary representation of  $G$  on a Hilbert space  $\mathcal{H}$ . Given a window  $g \in \mathcal{H}$ , the set  $\{\rho(\lambda)g : \lambda \in G\}$  is called the orbit of  $g$  under  $\rho(\lambda)$  for  $\lambda \in G$ . The map  $U_g : \mathcal{H} \rightarrow \mathbb{C}^G$  defined by*

$$U_g(f) = \{\langle f, \rho(\lambda)g \rangle\}_{\lambda \in G}$$

*is called the analysis operator of  $\{\rho(\lambda)g : \lambda \in G\}$ , and its adjoint operator  $U_g^* : \mathbb{C}^G \rightarrow \mathcal{H}$  given by*

$$U_g^*(x) = \sum_{\lambda \in G} x_{\lambda} \rho(\lambda)g$$

*is called the synthesis operator of  $\{\rho(\lambda)g : \lambda \in G\}$ .*

It is easy to check that

$$U_g \rho(\lambda) = L_{\lambda} U_g \quad \text{and} \quad U_g^* L_{\lambda} = \rho(\lambda) U_g^*, \quad \lambda \in G. \quad (10)$$

We are particularly interested in the case where the orbit of  $g$  spans  $\mathcal{H}$ , that is,  $\text{span}\{\rho(\lambda)g : \lambda \in G\} = \mathcal{H}$ . Since  $\mathcal{H}$  is finite-dimensional, this implies that  $\{\rho(\lambda)g : \lambda \in G\}$  is a frame for  $\mathcal{H}$  and the associated frame operator  $S_g := U_g^* U_g$  is a positive, self-adjoint bounded operator on  $\mathcal{H}$ . It follows from (10) that  $S_g \rho(\lambda) = \rho(\lambda) S_g$  and thus  $S_g^{-1} \rho(\lambda) = \rho(\lambda) S_g^{-1}$  for all  $\lambda \in G$ . For any  $f \in \mathcal{H}$ , we have

$$\begin{aligned} f &= S_g^{-1} S_g f = S_g^{-1} \left( \sum_{\lambda \in G} U_g f(\lambda) \rho(\lambda) g \right) = \sum_{\lambda \in G} U_g f(\lambda) S_g^{-1}(\rho(\lambda) g) \\ &= \sum_{\lambda \in G} U_g f(\lambda) \rho(\lambda) S_g^{-1}(g) = \sum_{\lambda \in G} U_g f(\lambda) \rho(\lambda) g^* = U_g^* U_g f, \end{aligned}$$

where  $g^* := S_g^{-1}(g) \in \mathcal{H}$ . This shows that  $U_g^* U_g$  is the identity operator on  $\mathcal{H}$ , i.e.,

$$U_g^* U_g = \text{Id}_{\mathcal{H}}, \quad (11)$$

and correspondingly,  $\{\rho(\lambda)g^* : \lambda \in G\}$  is the canonical dual frame of  $\{\rho(\lambda)g : \lambda \in G\}$ .

In light of (11), we newly introduce a transform which lifts a map  $\mathcal{H} \rightarrow \mathcal{H}$  to a map  $\mathbb{C}^G \rightarrow \mathbb{C}^G$ , and also its inverse transform.

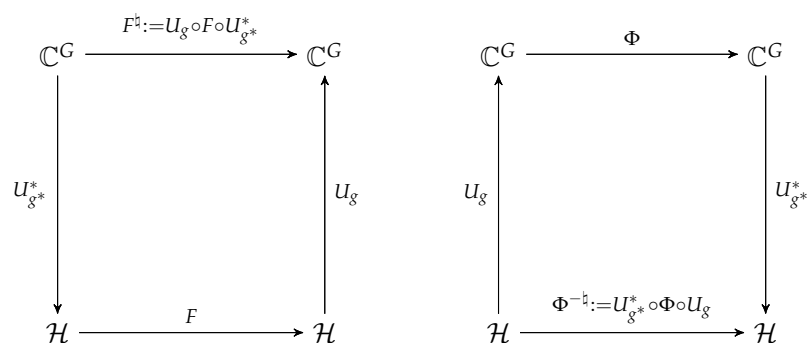
**Definition 7.** Let  $G$  be a finite group and let  $\rho$  be a unitary representation of  $G$  on a finite-dimensional Hilbert space  $\mathcal{H}$ . Assume that  $\text{span}\{\rho(\lambda)g : \lambda \in G\} = \mathcal{H}$ , and let  $S_g := U_g^* U_g$  and  $g^* := S_g^{-1}(g)$ . For any map  $F : \mathcal{H} \rightarrow \mathcal{H}$ , the  $\natural$ -transform of  $F$  is defined by

$$F^\natural := U_g \circ F \circ U_g^* : \mathbb{C}^G \rightarrow \mathbb{C}^G.$$

For any map  $\Phi : \mathbb{C}^G \rightarrow \mathbb{C}^G$ , the inverse  $\natural$ -transform of  $\Phi$  is defined by

$$\Phi^{-\natural} := U_g^* \circ \Phi \circ U_g : \mathcal{H} \rightarrow \mathcal{H}.$$

As shown in Figure 1, the  $\natural$ -transform converts a map  $\mathcal{H} \rightarrow \mathcal{H}$  into a map  $\mathbb{C}^G \rightarrow \mathbb{C}^G$ , and the inverse  $\natural$ -transform converts a map  $\mathbb{C}^G \rightarrow \mathbb{C}^G$  into a map  $\mathcal{H} \rightarrow \mathcal{H}$ .



**Figure 1.** The  $\natural$ -transform and its inverse transform

**Proposition 4.** Let  $G$  be a finite group, and let  $\rho$  be a unitary representation of  $G$  on a finite-dimensional Hilbert space  $\mathcal{H}$ . Assume that  $\text{span}\{\rho(\lambda)g : \lambda \in G\} = \mathcal{H}$ , and let  $S_g := U_g^* U_g$  and  $g^* := S_g^{-1}(g)$ . Then, the following hold.

- (i)  $(F^\natural)^{-\natural} = F$  for any map  $F : \mathcal{H} \rightarrow \mathcal{H}$ .
- (ii) A map  $F : \mathcal{H} \rightarrow \mathcal{H}$  is continuous if and only if  $F^\natural$  is continuous.
- (iii) A map  $F : \mathcal{H} \rightarrow \mathcal{H}$  is  $(\rho, G)$ -equivariant if and only if  $F^\natural$  is left  $G$ -translation equivariant.

**Proof.** (i) It follows from (11) that  $(F^\natural)^{-\natural} = U_g^* (U_g F U_g^*) U_g = F$  for any  $F : \mathcal{H} \rightarrow \mathcal{H}$ .  
(ii) Since the maps  $U_g : \mathcal{H} \rightarrow \mathbb{C}^G$  and  $U_g^* : \mathbb{C}^G \rightarrow \mathcal{H}$  are bounded linear operators, the continuity of  $F$  implies the continuity of  $F^\natural = U_g F U_g^*$ . Similarly, the continuity of  $F^\natural$  implies the continuity of  $F = (F^\natural)^{-\natural} = U_g^* F^\natural U_g$ .  
(iii) It follows from (10) that the  $G$ -equivariance of  $F$  implies the left  $G$ -translation equivariance of  $F^\natural = U_g F U_g^*$ . Similarly, the left  $G$ -translation equivariance of  $F^\natural$  implies the  $G$ -equivariance of  $F = (F^\natural)^{-\natural} = U_g^* F^\natural U_g$ .  $\square$

We now provide a brief review of neural networks and the universal approximation theorem.

Let  $\mathbb{K}$  be either  $\mathbb{R}$  or  $\mathbb{C}$ . An *activation function* is a function  $\sigma : \mathbb{K} \rightarrow \mathbb{K}$  that acts componentwise on vectors; that is,  $\sigma(x_1, \dots, x_n) = (\sigma(x_1), \dots, \sigma(x_n))$  for any  $(x_1, \dots, x_n) \in \mathbb{K}^n$ .

A fully connected feedforward neural network with  $P$  hidden layers is given by

$$\Psi : \mathbb{K}^d \rightarrow \mathbb{K}^n, \quad \Psi(x) = R^{(P)} \circ (\sigma \circ R^{(P-1)}) \circ \dots \circ (\sigma \circ R^{(0)}), \quad (12)$$

where  $R^{(p)} : \mathbb{K}^{N_p} \rightarrow \mathbb{K}^{N_{p+1}}$ ,  $x \mapsto A^{(p)}x + b^{(p)}$  is affine-linear with  $N_0 = d$  and  $N_{P+1} = n$ . Such a function  $\Psi$  is often called a *neural network*, but we will call it a  $\sigma$ -*neural network* to specify the activation function employed.

A *shallow neural network* is a neural network with a single ( $P = 1$ ) hidden layer. In particular, a shallow neural network with output dimension  $n = 1$  is given by

$$\Psi : \mathbb{K}^d \rightarrow \mathbb{K}, \quad \Psi(x) = \sum_{j=1}^J c_j \sigma(w_j^T x + b_j) \quad \text{with some } J \in \mathbb{N}, \quad c_j, b_j \in \mathbb{K}, \quad w_j \in \mathbb{K}^d. \quad (13)$$

**Definition 8.** A function  $\sigma : \mathbb{K} \rightarrow \mathbb{K}$  is called *shallow universal* if the set of  $\mathbb{K}$ -valued shallow  $\sigma$ -networks is dense in the set of all continuous functions  $f : \mathbb{K}^d \rightarrow \mathbb{K}$ , with respect to locally uniform convergence.

The following theorem, known as the *universal approximation theorem*, is a fundamental result in the theory of neural networks.

**Theorem 4** (The universal approximation theorem; see [27–31] for  $\mathbb{K} = \mathbb{R}$ , and [32] for  $\mathbb{K} = \mathbb{C}$ ). Let  $d \in \mathbb{N}$ .

- A function  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  is shallow universal if and only if  $\sigma$  is not a polynomial.
- A function  $\sigma : \mathbb{C} \rightarrow \mathbb{C}$  is shallow universal if and only if  $\sigma$  is not a polyharmonic. Here, a function  $\tau : \mathbb{C} \rightarrow \mathbb{C}$  is called *polyharmonic* if there exists  $m \in \mathbb{N}$  such that  $\tau \in C^{2m}$  in the sense of real variables and  $\Delta^m \sigma \equiv 0$ , where  $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$  is the usual Laplace operator on  $\mathbb{C} \cong \mathbb{R}^2$ .

In 1996, Mhaskar [33] obtained a quantitative result for approximation of  $C^n$  functions using shallow networks with smooth activation functions. More recently, Yarotsky [34] derived a quantitative approximation result for deep ReLU networks, where ReLU networks are given by (12) with  $\mathbb{K} = \mathbb{R}$  and the ReLU activation function  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$ ,  $\sigma(x) = \max\{x, 0\}$ , and “deep” refers to having a large  $P \in \mathbb{N}$  in (12). For the case of complex-valued deep neural networks, we refer to [35].

### 3.3. Cyclic Subgroups $\Lambda$ of $\mathbb{Z}_N \times \mathbb{Z}_N$

We now consider the case of cyclic subgroups of  $\mathbb{Z}_N \times \mathbb{Z}_N$ , where group representations can be defined directly without embedding into the Weyl–Heisenberg group. The cyclic subgroups of order  $N$  in  $\mathbb{Z}_N \times \mathbb{Z}_N$  are given by

$$\begin{aligned} \Lambda_s &= \{(0, 0), (1, s), \dots, (N-1, (N-1)s)\} = \langle (1, s) \rangle, \quad s = 0, \dots, N-1, \\ \Lambda_\infty &= \{(0, 0), (0, 1), \dots, (0, N-1)\} = \langle (0, 1) \rangle. \end{aligned}$$

If  $N$  is prime, these are the only nontrivial proper subgroups of  $\mathbb{Z}_N \times \mathbb{Z}_N$ , but if  $N$  is composite, there exist noncyclic subgroups of order  $N$  in  $\mathbb{Z}_N \times \mathbb{Z}_N$ ; for instance,  $\{0, 2, 4\} \times \{0, 3\}$  is a noncyclic subgroup of order 6 in  $\mathbb{Z}_6 \times \mathbb{Z}_6$ . It is easily seen that the adjoint group of  $\Lambda_s$  in  $\mathbb{Z}_N \times \mathbb{Z}_N$  is  $\Lambda_s$  itself; that is,  $(\Lambda_s)^\circ = \Lambda_s$  (see Section 2.1).

We define the map  $\rho : \Lambda_s \rightarrow \mathcal{U}(\mathbb{C}^N)$  by

$$(\rho(k, \ell)x)(n) = e^{-k\ell\pi i/N} e^{2\pi i\ell n/N} x(n-k), \quad (k, \ell) \in \Lambda_s, \quad x \in \mathbb{C}^N.$$

Setting  $\omega_0 := e^{\pi i/N}$ , we may simply write

$$\rho(k, \ell) = \omega_0^{-k\ell} M^\ell T^k = \omega_0^{k\ell} T^k M^\ell, \quad (k, \ell) \in \Lambda_s. \quad (14)$$

For any  $(k, \ell), (k', \ell') \in \Lambda_s$ , we have

$$\begin{aligned} \rho(k+k', \ell+\ell') &= \omega_0^{-(k+\ell)(k'+\ell')} M^{\ell+\ell'} T^{k+k'} = \omega_0^{-k\ell-k'\ell'-2k\ell'} M^{\ell+\ell'} T^{k+k'} \\ &\stackrel{(8)}{=} \omega_0^{-k\ell-k'\ell'} M^\ell T^k M^{\ell'} T^{k'} = \rho(k, \ell) \rho(k', \ell'), \end{aligned}$$

where we used the fact that  $k'\ell = k\ell'$  for all  $(k, \ell), (k', \ell') \in \Lambda_s$ . This shows that  $\rho$  is a group homomorphism and thus a unitary group representation of  $\Lambda_s$  on  $\mathbb{C}^N$ . Due the symmetry in (14),  $\rho$  is called the *symmetric representation* of  $\Lambda_s$  on  $\mathbb{C}^N$ .

Note that for any  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  and  $(k, \ell) \in \mathbb{Z}_N \times \mathbb{Z}_N$ , we have  $F\pi(k, \ell) = \pi(k, \ell)F$  if and only if  $F\rho(k, \ell) = \rho(k, \ell)F$ , where we used the relation  $\rho(k, \ell) = \omega_0^{-k\ell}\pi(k, \ell)$  from (14). This implies that a map  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  is  $\Lambda_s$ -equivariant in the sense of Definition 1 if and only if it is  $(\rho, \Lambda_s)$ -equivariant in the sense of Definition 4. Importantly, employing  $(\rho, \Lambda_s)$ -equivariance in place of  $\Lambda_s$ -equivariance will allow us to apply the tools from group representation theory described in Section 3.2.

We are interested in approximating  $\Lambda_s$ -equivariant (or  $(\rho, \Lambda_s)$ -equivariant) maps  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  by neural networks. For this, we need to choose a complex-valued activation function  $\sigma : \mathbb{C} \rightarrow \mathbb{C}$  (see Section 3.2) for the neural networks. Since  $\sigma$  acts componentwise on its input, i.e.,  $(x_1, \dots, x_N) \mapsto (\sigma(x_1), \dots, \sigma(x_N))$ , it clearly commutes with all translations, i.e.,  $\sigma T = T\sigma$ ; however,  $\sigma$  does not commute with modulations in general. As shown in (14), the representation  $\rho$  includes the multiplicative phase factor  $\omega_0 = e^{\pi i/N}$ , so we will assume that  $\sigma : \mathbb{C} \rightarrow \mathbb{C}$  is  $e^{\pi i/N}$ -phase homogeneous (see Definition 2):

$$\sigma(e^{\pi i/N} z) = e^{\pi i/N} \sigma(z), \quad z \in \mathbb{C},$$

which ensures that  $\sigma$  commutes with all  $\rho(k, \ell)$  and all modulations.

We first need the following lemma. Below, we denote by  $\mathbf{1}_N := (1, 1, \dots, 1) \in \mathbb{C}^N$  the vector whose entries are all equal to 1.

**Lemma 1.** Assume that  $\sigma : \mathbb{C} \rightarrow \mathbb{C}$  is shallow-universal. If a map  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  satisfies  $FT = TF$ , then there exists a shallow convolutional neural network

$$\Psi : \mathbb{C}^N \rightarrow \mathbb{C}^N, \quad \Psi(x) = \sum_{j=1}^J c_j \sigma(B_j x + b_j \mathbf{1}_N), \quad x \in \mathbb{C}^N,$$

where  $B_j \in \text{span}\{T^k : k = 0, \dots, N-1\}$  and  $b_j \in \mathbb{C}$  for  $j = 1, \dots, J$ , which approximates  $F$  uniformly on compact sets in  $\mathbb{C}^N$ .

**Proof.** Using the universal approximation theorem (see Theorem 4), the first output component map  $F_0 : \mathbb{C}^N \rightarrow \mathbb{C}, x \mapsto (Fx)(0)$ , can be approximated by a shallow network

$$\psi : \mathbb{C}^N \rightarrow \mathbb{C}, \quad x \mapsto \sum_{j=1}^J c_j \sigma(w_j^T x + b_j)$$

with some  $J \in \mathbb{N}, b_j, c_j \in \mathbb{C}, w_j \in \mathbb{C}^N$ . Note that since  $FT = TF$  and since  $T^N$  is the identity map on  $\mathbb{C}^N$ , we have  $FT^n = T^n F$  for all  $n \in \mathbb{Z}$ . This condition provides approximations for other component maps  $F_n : \mathbb{C}^N \rightarrow \mathbb{C}, x \mapsto (Fx)(n)$ , with  $n = 1, \dots, N-1$ , in terms of  $\Psi$ . In fact, we have

$$(Fx)(n) = (T^{-n} Fx)(0) = (FT^{-n} x)(0) \approx \psi(T^{-n} x), \quad x \in \mathbb{C}^N, \quad n = 1, \dots, N-1.$$

Consequently, the map  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$ ,  $x \mapsto \{(Fx)(n)\}_{n=0}^{N-1}$ , is approximated by the map  $\Psi : \mathbb{C}^N \rightarrow \mathbb{C}^N$  defined by  $(\Psi x)(n) = \psi(T^{-n}x)$  for  $n = 0, \dots, N-1$ . For  $x, y \in \mathbb{C}^N$ , let  $x * y$  be the circular convolution of  $a$  and  $b$  defined by  $(x * y)(n) = \sum_{k=0}^{N-1} x_k y_{n-k}$ , where  $x$  and  $y$  are understood as  $N$ -periodic sequences on the integers. Then, for any  $x \in \mathbb{C}^N$  and  $n = 0, \dots, N-1$ , we have

$$\Psi(T^{-n}x) = \sum_{j=1}^J c_j \sigma((w_j * x)(n) + b_j),$$

and therefore, we may write

$$\Psi : \mathbb{C}^N \rightarrow \mathbb{C}^N, \quad \Psi(x) = \{\psi(T^{-n}x)\}_{n=0}^{N-1} = \left\{ \sum_{j=1}^J c_j \sigma((w_j * x)(n) + b_j) \right\}_{n=0}^{N-1}.$$

It is easily seen that every convolutional map  $\mathbb{C}^N \rightarrow \mathbb{C}^N$ ,  $x \mapsto w * x$ , is a linear map, and in fact, a linear combination of  $T^k$ ,  $k = 0, \dots, N-1$ . Hence, the map  $\Psi : \mathbb{C}^N \rightarrow \mathbb{C}^N$  can be rewritten as

$$\Psi(x) = \sum_{j=1}^J c_j \sigma(B_j x + b_j \mathbf{1}_N), \quad x \in \mathbb{C}^N,$$

where  $B_j \in \text{span}\{T^k : k = 0, \dots, N-1\}$  for  $j = 1, \dots, J$ . The fact that  $\Psi$  approximates  $F$  uniformly on compact sets in  $\mathbb{C}^N$  follows from the uniform approximation of  $F_0$  by  $\psi$  on compact sets in  $\mathbb{C}$ . Finally, we note that  $\Psi$  expressed above is a shallow convolutional neural network described in Section 3.2. This completes the proof.  $\square$

**Theorem 5.** Assume that  $\sigma : \mathbb{C} \rightarrow \mathbb{C}$  is shallow universal and satisfies  $\sigma(e^{\pi i/N} z) = e^{\pi i/N} \sigma(z)$  for all  $z \in \mathbb{C}$ . Let  $\Lambda = \Lambda_s$  for some  $s \in \{0, 1, \dots, N-1\}$ . Then, any continuous  $(\rho, \Lambda)$ -equivariant (or  $\Lambda$ -equivariant) map  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  can be approximated (uniformly on compact sets) by a shallow neural network

$$x \mapsto \sum_{j=1}^J c_j \sigma(A_j x + b_j v),$$

where  $A_j \in \text{span}\{\rho(k, \ell) : (k, \ell) \in \Lambda\}$  and  $b_j \in \mathbb{C}$  for  $j = 1, \dots, J$ , and  $v \in \mathbb{C}^N$  satisfies  $\rho(k, \ell)v = v$  for all  $(k, \ell) \in \Lambda$ . Moreover, every map of this form is  $(\rho, \Lambda)$ -equivariant (or  $\Lambda$ -equivariant).

**Remark 4.** Since  $\rho(k, \ell) = \omega_0^{-k\ell} \pi(k, \ell)$  by (14), we have  $\text{span}\{\rho(k, \ell) : (k, \ell) \in \Lambda\} = \text{span}\{\pi(k, \ell) : (k, \ell) \in \Lambda\}$  for any  $\Lambda \subset \mathbb{Z}_N \times \mathbb{Z}_N$ . On the other hand, the vectors satisfying  $\rho(k, \ell)b = b$  can be significantly different from those satisfying  $\pi(k, \ell)b = b$ .

**Proof.** Since  $\Lambda = \Lambda_s$  is cyclic, we order its elements as  $(0, 0), (1, s), \dots, (N-1, (N-1)s)$ , and treat  $\mathbb{C}^\Lambda$  as  $\mathbb{C}^N$ , since  $\mathbb{C}^\Lambda \simeq \mathbb{C}^N$ . Then, the operators  $U_g : \mathbb{C}^N \rightarrow \mathbb{C}^\Lambda$  and  $U_g^* : \mathbb{C}^\Lambda \rightarrow \mathbb{C}^N$ , given in Definition 6, can be represented as the  $N \times N$  matrices

$$U_g = \begin{bmatrix} (\rho(0, 0)g)^* \\ \vdots \\ (\rho(N-1, (N-1)s)g)^* \end{bmatrix}, \quad U_g^* = [\rho(0, 0)g, \dots, \rho(N-1, (N-1)s)g],$$

respectively, where  $(\cdot)^*$  denotes the conjugate transpose. Setting  $g = (1, 0, \dots, 0) \in \mathbb{C}^N$ , we have

$$U_g^* = \text{diag}(e^{k^2 s \pi i / N})_{k=0}^{N-1} = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & e^{1^2 s \pi i / N} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{(N-1)^2 s \pi i / N} \end{bmatrix}, \quad (15)$$

so that  $S_g = U_g^* U_g = \text{Id}_N$  and  $g^* := S_g^{-1} g = g$ . As a result, the set  $\{\rho(k, \ell)g\}_{(k, \ell) \in \Lambda}$  forms an orthonormal basis for  $\mathbb{C}^N$ .

Note that for any continuous  $(\rho, \Lambda)$ -equivariant  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$ , the map  $F^\natural := U_g F U_g^* : \mathbb{C}^\Lambda \rightarrow \mathbb{C}^\Lambda$  is continuous and left  $\Lambda$ -translation equivariant (see Proposition 4). If  $F$  is linear, then  $F^\natural$  is also linear and can be represented as a circulant matrix, equivalently,  $F^\natural = \sum_{k=0}^{N-1} c_k T^k : \mathbb{C}^\Lambda \rightarrow \mathbb{C}^\Lambda$  for some  $c_0, \dots, c_{N-1} \in \mathbb{C}$ , so that

$$F = U_g^* (U_g F U_g^*) U_g = U_g^* F^\natural U_g = \sum_{k=0}^{N-1} c_k (U_g^* T U_g)^k.$$

Therefore, the commutant of  $(\rho, \Lambda)$  is given by

$$\begin{aligned} C(\rho, \Lambda) &:= \{F \in \mathcal{L}(\mathbb{C}^N) : F\rho(k, \ell) = \rho(k, \ell)F \text{ for all } (k, \ell) \in \Lambda\} \\ &= \text{span}\{U_g^* T^k U_g : k = 0, \dots, N-1\}. \end{aligned}$$

On the other hand, since  $\rho(k, \ell) = \omega_0^{-k\ell} \pi(k, \ell)$  by (14), the commutant of  $(\rho, \Lambda)$  coincides with that of  $(\pi, \Lambda)$ , i.e.,

$$C(\rho, \Lambda) = C(\pi, \Lambda) \stackrel{(4)}{=} \text{span}\{\pi(k, \ell) : (k, \ell) \in \Lambda^\circ\} = \text{span}\{\rho(k, \ell) : (k, \ell) \in \Lambda^\circ\}.$$

Since the adjoint group of  $\Lambda = \Lambda_s$  is itself, i.e.,  $\Lambda^\circ = \Lambda$  (see Section 2.1), we obtain

$$\text{span}\{\rho(k, \ell) : (k, \ell) \in \Lambda\} = C(\rho, \Lambda) = \text{span}\{U_g^* T^k U_g : k = 0, \dots, N-1\}. \quad (16)$$

Now, we consider the general case where  $F : \mathbb{C}^N \rightarrow \mathbb{C}^N$  is possibly nonlinear. If  $F$  is nonlinear, then  $F^\natural = U_g F U_g^* : \mathbb{C}^\Lambda \rightarrow \mathbb{C}^\Lambda$  is a nonlinear left  $\Lambda$ -translation equivariant map. Since  $\Lambda = \Lambda_s = \{(0, 0), (1, s), \dots, (N-1, (N-1)s)\}$  is an additive group and since  $|\Lambda| = N$  and  $\mathbb{C}^\Lambda \simeq \mathbb{C}^N$ , the map  $F^\natural$  can be viewed as a map from  $\mathbb{C}^N$  to  $\mathbb{C}^N$ . For simplicity, we will abuse notation and write  $F^\natural : \mathbb{C}^N \rightarrow \mathbb{C}^N$  instead of  $F^\natural : \mathbb{C}^\Lambda \rightarrow \mathbb{C}^\Lambda$ ; thus, the first component of  $F^\natural(x) \in \mathbb{C}^\Lambda (\simeq \mathbb{C}^N)$  will be simply denoted by  $(F^\natural x)(0)$  instead of  $(F^\natural x)(0, 0)$ . Then, the left  $\Lambda$ -translation equivariance of  $F^\natural$  can be expressed as  $F^\natural T = T F^\natural$ . By applying Lemma 1 to  $F^\natural : \mathbb{C}^N \rightarrow \mathbb{C}^N$ , we obtain a shallow convolutional neural network

$$\Psi : \mathbb{C}^N \rightarrow \mathbb{C}^N, \quad \Psi(x) = \sum_{j=1}^J c_j \sigma(B_j x + b_j \mathbf{1}_N), \quad x \in \mathbb{C}^N,$$

where  $B_j \in \text{span}\{T^k : k = 0, \dots, N-1\}$ , and  $b_j \in \mathbb{C}$  for  $j = 1, \dots, J$ , which approximates  $F^\natural$  uniformly on compact sets in  $\mathbb{C}^N$ ; that is,

$$F^\natural(x) = (U_g F U_g^*)(x) \approx \Psi(x) = \sum_{j=1}^J c_j \sigma(B_j x + b_j \mathbf{1}_N) \quad x \in \mathbb{C}^N.$$

By the continuity of the operators  $U_g^*$  and  $U_g$ , we obtain

$$F(x) = (U_g^* (U_g F U_g^*) U_g)(x) \approx \sum_{j=1}^J c_j U_g^* \sigma(B_j U_g x + b_j \mathbf{1}_N), \quad x \in \mathbb{C}^N.$$



Note that since  $\sigma(e^{\pi i/N} z) = e^{\pi i/N} \sigma(z)$  for all  $z \in \mathbb{C}$ , the function  $\sigma : \mathbb{C} \rightarrow \mathbb{C}$  commutes with  $U_g^*$  given by (15), that is,  $U_g^* \sigma = \sigma U_g^*$ . Therefore, we have

$$F(x) \approx \sum_{j=1}^J c_j \sigma(U_g^* B_j U_g x + b_j U_g^* \mathbb{1}_N) = \sum_{j=1}^J c_j \sigma(A_j x + b_j v), \quad x \in \mathbb{C}^N,$$

where  $A_j := U_g^* B_j U_g \in \text{span}\{\rho(k, \ell) : (k, \ell) \in \Lambda\}$  by (16), and the vector  $v := U_g^* \mathbb{1}_N \in \mathbb{C}^N$  satisfies

$$\rho(k, \ell)v = \rho(k, \ell)U_g^* \mathbb{1}_N \stackrel{(4)}{=} U_g^* L_\lambda \mathbb{1}_N = U_g^* \mathbb{1}_N = v, \quad (k, \ell) \in \Lambda. \quad (17)$$

Finally, we note that for any  $(k, \ell) \in \Lambda$ ,

$$\begin{aligned} \rho(k, \ell) \left( \sum_{j=1}^J c_j \sigma(A_j x + b_j v) \right) &= \sum_{j=1}^J c_j \rho(k, \ell) \sigma(A_j x + b_j v) \\ &= \sum_{j=1}^J c_j \sigma(\rho(k, \ell) A_j x + b_j \rho(k, \ell) v) \\ &= \sum_{j=1}^J c_j \sigma(A_j \rho(k, \ell) x + b_j v), \end{aligned}$$

where we used that  $\rho(k, \ell)$  is a linear (unitary) operator commuting with  $\sigma$ , and that  $A_j \in C(\rho, \Lambda)$  by (16) and  $\rho(k, \ell)v = v$  by (17). Therefore, every map of the form  $x \mapsto \sum_{j=1}^J c_j \sigma(A_j x + b_j v)$  is  $(\rho, \Lambda)$ -equivariant.  $\square$

**Remark 5.** The proof relies on observing (16) and choosing  $g \in \mathbb{C}^N$  such that  $U_g^* \sigma = \sigma U_g^*$ . To obtain  $U_g^* \sigma = \sigma U_g^*$ , we have chosen  $g \in \mathbb{C}^N$  so that  $U_g^*$  is a diagonal matrix with exponential entries, and required an appropriate phase-homogeneity on  $\sigma$  so that  $\sigma$  commutes with those exponentials. This technique does not work for  $\Lambda_\infty$  because  $U_g^*$  cannot be expressed as a diagonal matrix for any  $g \in \mathbb{C}^N$  in that case.

**Example 1.** Let  $N = 4$  and  $s = 1$ , so that  $\Lambda = \Lambda_1 = \{(0, 0), (1, 1), (2, 2), (3, 3)\} \subset \mathbb{Z}_4 \times \mathbb{Z}_4$ . In this case, we have  $\omega = e^{2\pi i/4} = i$ ,  $\omega_0 = e^{\pi i/4} = \frac{1}{\sqrt{2}}(1 + i)$ , and  $\rho(k, \ell) = \omega_0^{-k\ell} M^\ell T^k$ . Then,

$$\rho(0, 0) = I_4,$$

$$\rho(1, 1) = \omega_0^{-1} M T = \omega_0^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -i \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} = \omega_0^{-1} \begin{bmatrix} 0 & 0 & 0 & 1 \\ i & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -i & 0 \end{bmatrix},$$

$$\rho(2, 2) = \omega_0^{-4} M^2 T^2 = - \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix},$$

$$\rho(3, 3) = \omega_0^{-9} M^2 T^2 = \omega_0^{-1} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & i \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} = \omega_0^{-1} \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -i & 0 \\ 0 & 0 & 0 & -1 \\ i & 0 & 0 & 0 \end{bmatrix},$$

and  $\rho(k, k)\rho(k', k') = \rho(k + k', k + k')$  for all  $k, k' = 0, 1, 2, 3$ . With  $g = (1, 0, 0, 0)^T$ , we have

$$U_g^* = [\rho(0,0)g \quad \rho(1,1)g \quad \rho(2,2)g \quad \rho(3,3)g] = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \omega_0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & \omega_0 \end{bmatrix},$$

$$v := U_g^* \mathbb{1}_4 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \omega_0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & \omega_0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \omega_0 \\ -1 \\ \omega_0 \end{bmatrix}.$$

It is easy to check that  $v$  is invariant under  $\rho(0,0), \rho(1,1), \rho(2,2), \rho(3,3)$ ; that is,  $\rho(k,k)v = v$  for all  $k = 0, 1, 2, 3$ . Theorem 5 shows that any  $\Lambda$ -equivariant map  $F : \mathbb{C}^4 \rightarrow \mathbb{C}^4$  can be approximated (uniformly on compact sets) by functions of the form

$$x \mapsto \sum_{m=1}^M c_m \sigma(A_m x + b_m v),$$

where  $A_m \in \text{span}\{\rho(k,k) : k = 0, 1, 2, 3\}$  and  $b_m \in \mathbb{C}$  for  $m = 1, \dots, M$ . It is worth noting that while  $\rho$  is a unitary group representation of  $\Lambda = \{(0,0), (1,1), (2,2), (3,3)\}$  on  $\mathbb{C}^4$ , the map  $\pi|_\Lambda$  given by  $\pi(k, \ell) = M^\ell T^k$  for  $(k, \ell) \in \Lambda$  is not a group representation of  $\Lambda$  on  $\mathbb{C}^4$ , since  $\pi(1,1)\pi(1,1) = (-i)\pi(2,2)$  by (8).

#### 4. Discussion

In this paper, we used finite-dimensional time-frequency analysis to investigate the properties of time-frequency shift equivariant maps that are generally nonlinear.

First, we established a one-to-one correspondence between  $\Lambda$ -equivariant maps and certain phase-homogeneous functions, accompanied by a reconstruction formula expressing  $\Lambda$ -equivariant maps in terms of these functions. This deepens our understanding of the structure of  $\Lambda$ -equivariant maps by connecting them to their corresponding phase-homogeneous functions.

Next, we considered the approximation of  $\Lambda$ -equivariant maps by neural networks. When  $\Lambda$  is a cyclic subgroup of order  $N$  in  $\mathbb{Z}_N \times \mathbb{Z}_N$ , we proved that every  $\Lambda$ -equivariant map can be approximated by a shallow neural network with affine linear maps formed as linear combinations of time-frequency shifts by  $\Lambda$ . For the subgroup  $\Lambda = \langle (1,0) \rangle = \{(0,0), (1,0), \dots, (N-1,0)\}$ , the  $\Lambda$ -equivariance corresponds to translation equivariance, and our result shows that every translation equivariant map can be approximated by a shallow convolutional neural network, which aligns well with the established effectiveness of convolutional neural networks (CNNs) for applications involving translation equivariance. In this context, our result extends the approximation of translation equivariant maps to general  $\Lambda$ -equivariant maps, with potential applications in signal processing.

Finally, we note that the tools used to prove the approximation result (Theorem 2) are applicable in a more general setting than the one described in Section 3.3. In particular, Definitions 6 and 7, and Proposition 4 apply to general unitary representations of arbitrary groups. Therefore, our approach can be adapted to derive similar results for general group-equivariant maps, which we leave as a direction for future research.

**Funding:** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (RS-2023-00275360).

**Data Availability Statement:** Data are contained within the article.

**Acknowledgments:** The author would like to thank Andrei Caragea, Johannes Maly, Goetz Pfander, and Felix Voigtlaender for their valuable discussions during the early stages of this paper.

**Conflicts of Interest:** The author declares no conflicts of interest.

### Appendix A. A Proof of the Fact that $|\Lambda^\circ| = L^2/|\Lambda|$ for Any Subgroup $\Lambda$ of $\mathbb{Z}_L \times \mathbb{Z}_L$ .

For finite abelian groups, it is known (see Lemma 4.2 of [25]) that the adjoint  $\Lambda^\circ$  of a subgroup  $\Lambda \subset G \times \widehat{G}$  is the symplectic analogue of the dual subgroup  $\Lambda^\perp$ , in the sense that  $\Lambda^\circ = J \Lambda^\perp$ , where

$$J = \begin{pmatrix} 0 & I_{|G|} \\ -I_{|G|} & 0 \end{pmatrix}.$$

(In fact, a similar characterization is known for locally compact Abelian groups; see, e.g., Lemma 3.5.9 and Lemma 7.7.3 of [36]. In particular, for separable subgroups  $\Lambda = \Lambda_1 \times \Lambda_2 \subset G \times \widehat{G}$ , we have  $\Lambda^\circ = \Lambda_2^\perp \times \Lambda_1^\perp$  while  $\Lambda^\perp = \Lambda_1^\perp \times \Lambda_2^\perp$ .) This implies that  $\Lambda^\circ$  has the same cardinality as  $\Lambda^\perp$ .

Here, the dual (annihilator)  $H^\perp$  of a subgroup  $H$  of  $\mathcal{G}$  is defined as

$$H^\perp = \{m \in \widehat{\mathcal{G}} : \langle m, n \rangle = 1 \text{ for all } n \in H\},$$

where  $\langle m, n \rangle = e^{2\pi i(m_1 n_1 / N_1 + \dots + m_d n_d / N_d)}$  for  $m = (m_1, \dots, m_d)$ ,  $n = (n_1, \dots, n_d)$ , if  $G = \mathbb{Z}_{N_1} \times \dots \times \mathbb{Z}_{N_d}$ . It is easily seen that  $|H^\perp| \cdot |H| = |\mathcal{G}|$ , for instance, by taking  $x = 1^\mathcal{G} = (1, 1, \dots, 1)$  in Theorem 6.3 of [21],

$$|H^\perp| \cdot \sum_{h \in H} x(h) = \sum_{m \in H^\perp} \widehat{x}(m), \quad x \in \mathbb{C}^\mathcal{G}.$$

Therefore, we have  $|\Lambda^\circ| = |\Lambda^\perp| = |\mathbb{Z}_L \times \mathbb{Z}_L|/|\Lambda| = L^2/|\Lambda|$ .

## References

1. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [CrossRef] [PubMed]
2. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. ImageNet Classification with Deep Convolutional Neural Networks. In Proceedings of the 26th International Conference on Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; Pereira, F., Burges, C.J.C., Bottou, L., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2012; Volume 25.
3. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. *arXiv* **2015**, arXiv:1512.03385.
4. Hinton, G.; Deng, L.; Yu, D.; Dahl, G.E.; Mohamed, A.; Jaitly, N.; Senior, A.; Vanhoucke, V.; Nguyen, P.; Sainath, T.N.; et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Process. Mag.* **2012**, *29*, 82–97. [CrossRef]
5. Sutskever, I.; Vinyals, O.; Le, Q.V. Sequence to Sequence Learning with Neural Networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, USA, 8–13 December 2014; Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N., Weinberger, K.Q., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2014; Volume 27.
6. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention Is All You Need. In Proceedings of the Advances in Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; Volume 30.
7. Silver, D.; Huang, A.; Maddison, C.J.; Guez, A.; Sifre, L.; Van Den Driessche, G.; Schrittwieser, J.; Antonoglou, I.; Panneershelvam, V.; Lanctot, M.; et al. Mastering the game of Go with deep neural networks and tree search. *Nature* **2016**, *529*, 484–489. [CrossRef] [PubMed]
8. Hershey, S.; Chaudhuri, S.; Ellis, D.P.W.; Gemmeke, J.F.; Jansen, A.; Moore, R.C.; Plakal, M.; Platt, D.; Saurous, R.A.; Seybold, B.; et al. CNN Architectures for Large-Scale Audio Classification. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 131–135. [CrossRef]
9. Oppenheim, A.; Schaffer, R. *Discrete-Time Signal Processing*, 3rd ed.; Pearson: Upper Saddle River, NJ, USA, 2010.
10. Walnut, D. *An Introduction to Wavelet Analysis*; Birkhäuser: Boston, MA, USA, 2002.
11. Boggess, A.; Narcowich, F.J. *A First Course in Wavelets with Fourier Analysis*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2009.
12. Cohen, T.; Welling, M. Group Equivariant Convolutional Networks. In Proceedings of the 33rd International Conference on Machine Learning, New York, NY, USA, 20–22 Jun 2016; Volume 48, pp. 2990–2999.
13. Cohen, T.; Geiger, M.; Weiler, M. A general theory of equivariant CNNs on homogeneous spaces. *arXiv* **2018**, arXiv:1811.02017.
14. Yarotsky, D. Universal approximations of invariant maps by neural networks. *Constr. Approx.* **2022**, *55*, 407–474. [CrossRef]
15. Cahill, J.; Iverson, J.W.; Mixon, D.G.; Packer, D. Group-invariant max filtering. *arXiv* **2022**, arXiv:2205.14039. [CrossRef]
16. Balan, R.; Tsoukanis, E. G-invariant representations using coorbits: Bi-lipschitz properties. *arXiv* **2023**, arXiv:2308.11784.
17. Balan, R.; Tsoukanis, E. G-invariant representations using coorbits: Injectivity properties. *arXiv* **2023**, arXiv:2310.16365.
18. Huang, N.; Levie, R.; Villar, S. Approximately equivariant graph networks. *arXiv* **2023**, arXiv:2308.10436.
19. Blum-Smith, B.; Villar, S. Machine learning and invariant theory. *arXiv* **2022**, arXiv:2209.14991. [CrossRef]

20. Wang, R.; Walters, R.; Yu, R. Data augmentation vs. equivariant networks: A theory of generalization on dynamics forecasting. *arXiv* **2022**, arXiv:2206.09450.
21. Pfander, G.E. Gabor frames in finite dimensions. In *Finite Frames*; Casazza, P.G., Kutyniok, G., Eds.; Applied and Numerical Harmonic Analysis; Birkhäuser: Boston, MA, USA, 2013; pp. 193–239.
22. Kaplan, A.; Lee, D.G.; Pfander, G.E.; Pohl, V. Sparse deterministic and stochastic channels: Identification of spreading functions and covariances. In *Compressed Sensing in Information Processing*; Kutyniok, G., Rauhut, H., Kunsch, R.J., Eds.; Springer International Publishing: New York, NY, USA, 2022; pp. 105–144.
23. Gröchenig, K. *Foundations of Time-Frequency Analysis*; Applied and Numerical Harmonic Analysis; Birkhäuser: Boston, MA, USA, 2001. [\[CrossRef\]](#)
24. Lawrence, J.; Pfander, G.E.; Walnut, D. Linear independence of Gabor systems in finite dimensional vector spaces. *J. Fourier Anal. Appl.* **2005**, *11*, 715–726. [\[CrossRef\]](#)
25. Feichtinger, H.G.; Kozek, W.; Luef, F. Gabor analysis over finite Abelian groups. *Appl. Comput. Harmon. Anal.* **2009**, *26*, 230–248. [\[CrossRef\]](#)
26. Christensen, O. *An Introduction to Frames and Riesz Bases*, 2nd ed.; Birkhäuser: New York, NY, USA, 2016.
27. Cybenko, G. Approximation by superpositions of a sigmoidal function. *Math. Control. Signals Syst.* **1989**, *2*, 303–314. [\[CrossRef\]](#)
28. Hornik, K. Approximation capabilities of multilayer feedforward networks. *Neural Netw.* **1991**, *4*, 251–257. [\[CrossRef\]](#)
29. Hornik, K.; Stinchcombe, M.; White, H. Multilayer feedforward networks are universal approximators. *Neural Netw.* **1989**, *2*, 359–366. [\[CrossRef\]](#)
30. Leshno, M.; Lin, V.; Pinkus, A.; Schocken, S. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Netw.* **1993**, *6*, 861–867. [\[CrossRef\]](#)
31. Pinkus, A. Approximation theory of the MLP model in neural networks. *Acta Numer.* **1999**, *8*, 143–195. [\[CrossRef\]](#)
32. Voigtlaender, F. The universal approximation theorem for complex-valued neural networks. *Appl. Comput. Harmon. Anal.* **2023**, *64*, 33–61. [\[CrossRef\]](#)
33. Mhaskar, H.N. Neural networks for optimal approximation of smooth and analytic functions. *Neural Comput.* **1996**, *8*, 164–177. [\[CrossRef\]](#)
34. Yarotsky, D. Error bounds for approximations with deep ReLU networks. *Neural Netw.* **2017**, *94*, 103–114. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Caragea, A.; Lee, D.G.; Maly, J.; Pfander, G.E.; Voigtlaender, F. Quantitative approximation results for complex-valued neural networks. *SIAM J. Math. Data Sci.* **2022**, *4*, 553–580. [\[CrossRef\]](#)
36. Feichtinger, H.G.; Strohmer, T. (Eds.) *Gabor Analysis and Algorithms*; Birkhäuser: Boston, MA, USA, 1998. [\[CrossRef\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.