EQUIVARIANT POLYNOMIAL FUNCTIONAL NETWORKS

Anonymous authors
Paper under double-blind review

ABSTRACT

Neural Functional Networks (NFNs) have gained increasing interest due to their wide range of applications, including extracting information from implicit representations of data, editing network weights, and evaluating policies. A key design principle of NFNs is their adherence to the permutation and scaling symmetries inherent in the connectionist structure of the input neural networks. Recent NFNs have been proposed with permutation and scaling equivariance based on either graph-based message-passing mechanisms or parameter-sharing mechanisms. Compared to graph-based models, parameter-sharing-based NFNs built upon equivariant linear layers exhibit lower memory consumption and faster running time. However, their expressivity is limited due to the large size of the symmetric group of the input neural networks. The challenge of designing a permutation and scaling equivariant NFN that maintains low memory consumption and running time while preserving expressivity remains unresolved. In this paper, we propose a novel solution with the development of MAGEP-NFN (**M**onomial m**A**trix **G**roup **E**quivariant **P**olynomial **NFN**). Our approach follows the parameter-sharing mechanism but differs from previous works by constructing a nonlinear equivariant layer represented as a polynomial in the input weights. This polynomial formulation enables us to incorporate additional relationships between weights from different input hidden layers, enhancing the model's expressivity while keeping memory consumption and running time low, thereby addressing the aforementioned challenge. We provide empirical evidence demonstrating that MAGEP-NFN achieves competitive performance and efficiency compared to existing baselines.

1 INTRODUCTION

Deep neural networks (DNNs) have become versatile tools, finding applications across various domains such as natural language processing (Rumelhart et al., 1986; Hochreiter & Schmidhuber, 1997; Vaswani et al., 2017; Devlin et al., 2019), computer vision (He et al., 2015; Szegedy et al., 2015; Krizhevsky et al., 2012), and the natural sciences (Raissi et al., 2019; Jumper et al., 2021). Neural functional networks (NFNs) (Zhou et al., 2024b) have recently gained prominence as specialized frameworks designed to process key aspects of DNNs, such as their weights, gradients, or sparsity masks, treating these as input data. NFNs serve a broad array of purposes, including optimizing training processes through learnable optimizers (Bengio et al., 2013; Runarsson & Jonsson, 2000; Andrychowicz et al., 2016; Metz et al., 2022), extracting features from implicit data representations (Stanley, 2007; Mildenhall et al., 2021; Runarsson & Jonsson, 2000), editing network parameters for corrective purposes (Sinitsin et al., 2020; De Cao et al., 2021; Mitchell et al., 2021), policy evaluation (Harb et al., 2020), and enabling Bayesian inference by using networks as evidence (Sokota et al., 2021).

Designing NFNs is inherently complex due to the high-dimensional nature of the structures they model. Early approaches tackled this challenge by restricting the training process to smaller, constrained weight spaces (Dupont et al., 2021; Bauer et al., 2023; De Luigi et al., 2023). More recent advancements, however, have focused on creating permutation equivariant NFNs, capable of accommodating neural network weights without imposing such constraints (Navon et al., 2023; Zhou et al., 2024b; Kofinas et al., 2024; Zhou et al., 2024c). These methods leverage permutation equivariance to respect symmetries arising from neuron reordering within hidden layers.

Despite these improvements, many existing techniques overlook additional key symmetries inherent in weight spaces. Examples include weight scaling invariance in ReLU networks (Bui Thi Mai & Lampert, 2020; Neyshabur et al., 2015; Badrinarayanan et al., 2015) and sign-flipping transformations in sin and tanh networks (Chen et al., 1993; Fefferman & Markel, 1993; Kurkova & Kainen, 1994). Addressing these symmetries remains an open challenge (Godfrey et al., 2022; Bui Thi Mai & Lampert, 2020).

NFNs that are equivariant to both permutations and scaling or sign-flipping have been introduced in (Kalogeropoulos et al., 2024) using a graph-based message-passing mechanism and in (Tran et al., 2024) with a parameter sharing mechanism. However, similar to other graph-based neural functional networks, treating the entire input neural network as a graph and utilizing graph neural networks causes the graph-based equivariant NFNs in (Kalogeropoulos et al., 2024) to have very high memory consumption and running time. In contrast, the NFNs built upon equivariant linear layers using the parameter sharing mechanism in (Tran et al., 2024) exhibit much lower memory consumption and running time. Nevertheless, the equivariant linear layers introduced in (Tran et al., 2024) possess weak expressive properties, as the weights of the input hidden layers are updated solely by the corresponding weights of the same input hidden layers. The challenge of designing an equivariant layer based on the parameter-sharing mechanism that maintains both lower memory consumption and running time while preserving expressivity remains unresolved.

Contribution. This paper aims to develop a novel NFN that is equivariant to both permutations and scaling/sign-flipping symmetries, called MAGEP-NFN (**M**onomial m**A**trix **G**roup **E**quivariant **P**olynomial **NFN**). We follow the parameter-sharing mechanism as described in (Tran et al., 2024); however, unlike (Tran et al., 2024), we construct a nonlinear equivariant layer, which is represented as a polynomial in the input weights. This polynomial formulation enables us to incorporate additional relationships between weights from different input hidden layers, thereby addressing the challenges posed in (Tran et al., 2024) and enhancing the expressivity of MAGEP-NFN. In particular, our contribution is as follows:

1. We introduce new polynomials in the input weights, called *stable polynomial terms*, such that these polynomials are stable under the group action of the weight space.
2. We conduct a comprehensive study of the linear independence of stable polynomial terms, which are essential for the parameter-sharing computations of equivariant polynomial layers defined as their linear combinations.
3. We design MAGEP-NFN, a family of monomial matrix equivariant NFNs based on the parameter-sharing mechanism that maintains both lower memory consumption and running time while preserving expressivity. The main building blocks of MAGEP-NFN are the equivariant and invariant polynomial layers for processing weight spaces.

We evaluate MAGEP-NFNs on three tasks: predicting CNN generalization from weights using Small CNN Zoo (Unterthiner et al., 2020), weight space style editing, and classifying INRs using INRs data (Zhou et al., 2024b). Experimental results show that our model achieves competitive performance and efficiency compared to existing baselines.

Organization. After recalling some related works in Section 2, we reformulate the definitions of weight spaces for MLPs and CNNs, as well as the action of monomial matrix groups on these weight spaces. In Section 4, we construct polynomial equivariant and invariant layers, which serve as the main building blocks for our MAGEP-NFNs. Several experiments are conducted in Section 5 to verify the applicability and efficiency of our models in comparison with previous ones in the literature. The paper concludes with a summary in Section 6.

2 Related Work

Symmetries in Neural Network Weight Spaces. The study of symmetries in the weight spaces of neural networks, which involves analyzing the functional equivalence of these networks, has a long-standing history. This topic was first introduced by Hecht-Nielsen (Hecht-Nielsen, 1990), providing a foundational perspective on the relationship between weight symmetries and network functionality. Over time, numerous works have expanded on this framework, presenting results tailored to different network architectures (Chen et al., 1993; Fefferman & Markel, 1993; Kurkova & Kainen, 1994; Albertini & Sontag, 1993b;a; Bui Thi Mai & Lampert, 2020). These studies build on earlier insights into convergence, gradient dynamics, and structural properties of neural networks, as explored in works such as (Allen-Zhu et al., 2019; Du et al., 2019; Frankle & Carbin, 2018; Belkin et al., 2019; Novak et al., 2018).

Neural Functional Networks. Recent advances have focused on creating effective representations of trained classifiers to evaluate their generalization capabilities and uncover insights into neural network dynamics (Baker et al., 2017; Eilertsen et al., 2020; Unterthiner et al., 2020; Schürholt et al., 2021; 2022a;b). In particular, low-dimensional encodings for Implicit Neural Representations (INRs) have been developed to support a wide range of downstream applications (Dupont et al., 2022; De Luigi et al., 2023). These approaches typically involve either flattening the network parameters or deriving parameter statistics for further processing using standard multi-layer perceptrons (MLPs) (Unterthiner et al., 2020). To involve the symmetric structure of the input neural networks, Schurholt et al. introduced neuron permutation augmentations to better align model representations with their functional equivalence (Schürholt et al., 2021). Other studies have expanded on these ideas by focusing on encoding and decoding neural network parameters, primarily for reconstruction and generative modeling (Peebles et al., 2022; Ashkenazi et al., 2022; Knyazev et al., 2021; Erkoç et al., 2023).

Equivariant Neural Functional Networks (NFNs). Significant strides have been made in addressing the limitations of permutation equivariant neural networks by incorporating specialized layers designed to enforce equivariance. These permutation equivariant layers rely on sophisticated weight-sharing mechanisms (Navon et al., 2023; Zhou et al., 2024b; Kofinas et al., 2024; Zhou et al., 2024c), as well as set-based (Andreis et al., 2023) or graph-based structures (Lim et al., 2023; Kofinas et al., 2024; Zhou et al., 2024a) to achieve the desired symmetry properties. These advancements enable the creation of both permutation invariant and equivariant networks, preserving the inherent symmetry associated with the reordering of neurons within each layer. Despite these developments, current approaches often overlook additional symmetries present in neural networks. For instance, weight scaling symmetries in ReLU networks and weight sign-flipping symmetries in $\sin$ and $\tanh$ networks remain underexplored. Recent efforts have addressed these gaps by introducing NFNs that are equivariant to both permutations and scaling, referred to as monomial equivariant NFNs (Kalogeropoulos et al., 2024; Tran et al., 2024).

However, the graph-based equivariant NFNs proposed in (Kalogeropoulos et al., 2024) suffer from high memory consumption and significant runtime overhead. While, the Monomial-NFNs constructed using equivariant linear layers and a parameter-sharing mechanism in (Tran et al., 2024) exhibit limited expressive power. In contrast, our MAGEP-NFNs are built upon equivariant polynomial layers, leveraging a parameter-sharing mechanism that achieves both lower memory consumption and reduced runtime while preserving strong expressivity.

3 Weight Space and Group Action on Weight Space

Let $\mathcal{U}, \mathcal{V}$ be two sets and assume that a group G acts on them. A function $f : \mathcal{U} \to \mathcal{V}$ is called G-equivariant if $f(g \cdot x) = g \cdot f(x)$ for all $x \in \mathcal{U}$ and $g \in G$. In case G acts trivially on $\mathcal{Y}$, the function f is called G-invariant. In the context of this paper, $\mathcal{U}$ and $\mathcal{V}$ are weight spaces of a fixed neural network architecture, while $\mathcal{G}$ is a direct product of the groups of monomial matrices.

3.1 Monomial Matrix Group

Let us start with the definition of monomial and permutation matrices.

Definition 3.1 (See (Rotman, 2012, page 46))**.** Let n be a positive integer.

- A matrix $M \in \mathrm{GL}_n(\mathbb{R})$ is called a *monomial matrix* (or *generalized permutation matrix*) if it has exactly one non-zero entry in each row and each column, and zeros elsewhere. We denote by $\mathcal{M}_n$ the set of such all monomial matrices.
- A matrix $P \in \mathrm{GL}_n(\mathbb{R})$ is called a *permutation matrix* if it is a monomial matrix and all nonzero entries are equal to 1. We denote by $\mathcal{P}_n$ the set of such all permutation matrices.

Let $\mathcal{D}_n$ the set of diagonal matrices in $\mathrm{GL}_n(\mathbb{R})$. Then, $\mathcal{M}_n$, $\mathcal{P}_n$ and $\mathcal{D}_n$ are subgroups of of the general linear group $\mathrm{GL}_n(\mathbb{R})$. Moreover, every monomial matrix can be written as a product of an invertible diagonal matrix and a permutation matrix.

Remark (Permutation matrix vs permutation). For every permutation matrix $P \in \mathcal{P}_n$, there exists a unique permutation $\pi \in \mathcal{S}_n$ such that P is obtained by permuting the n columns of the identity matrix I_n according to π. In this case, we write $P := P_\pi$ and call it the *permutation matrix* corresponding to π. Here, $\mathcal{S}_n$ is the group of all permutations of the set $\{1, 2, \ldots, n\}$.

3.2 Weight space of MLPs and CNNs

Following (Tran et al., 2024), we write the weight space of an MLP or CNN with L layers and n_i channels at i-th layer in the general form $\mathcal{U} = \mathcal{W} \times \mathcal{B}$, where:

$$\begin{aligned} \mathcal{W} &= \mathbb{R}^{w_L \times n_L \times n_{L-1}} \times \ldots \times \mathbb{R}^{w_2 \times n_2 \times n_1} \times \mathbb{R}^{w_1 \times n_1 \times n_0}, \\ \mathcal{B} &= \mathbb{R}^{b_L \times n_L \times 1} \times \ldots \times \mathbb{R}^{b_2 \times n_2 \times 1} \times \mathbb{R}^{b_1 \times n_1 \times 1}. \end{aligned} \tag{1}$$

Here, n_i is the number of channels at the i-th layer, in particular, n_0 and n_L are the number of channels of input and output; w_i is the dimension of weights and b_i is the dimension of the biases in each channel at the i-th layer. The dimension of the weight space $\mathcal{U}$ is:

$$\dim \mathcal{U} = \sum_{i=1}^{L} \left(w_i \times n_i \times n_{i-1} + b_i \times n_i \times 1 \right). \tag{2}$$

To emphasize the weights and biases of an element U of $\mathcal{U}$, we will write $U = ([W], [b])$, with the weights

$$[W] = \left([W]^{(L)}, \ldots, [W]^{(1)} \right) \in \mathcal{W}, \tag{3}$$

and biases

$$[b] = \left([b]^{(L)}, \ldots, [b]^{(1)} \right) \in \mathcal{B}. \tag{4}$$

The square brackets will be convenient in the next section when we determine polynomials in the entries of U.

Remark. The space $\mathbb{R}^{w_i \times n_i \times n_{i-1}} = (\mathbb{R}^{w_i})^{n_i \times n_{i-1}}$ at the i-th layer is interpreted as the space of $n_i \times n_{i-1}$ matrices, where each entry consists of real vectors in $\mathbb{R}^{w_i}$. Specifically, the symbol $[W]^{(i)}$ represents a matrix in $\mathbb{R}^{w_i \times n_i \times n_{i-1}} = (\mathbb{R}^{w_i})^{n_i \times n_{i-1}}$, with $[W]^{(i)}_{jk} \in \mathbb{R}^{w_i}$ indicating the entry located at row j and column k of $[W]^{(i)}$. Similarly, the term $[b]^{(i)}$ denotes a bias column vector in $\mathbb{R}^{b_i \times n_i \times 1} = (\mathbb{R}^{b_i})^{n_i \times 1}$, while $[b]^{(i)}_j \in \mathbb{R}^{b_i}$ denotes the entry at row j of the vector $[b]^{(i)}$. The dimensions of weights w_i and biases b_i at each channel do not affect the definition of group action in the subsequent sections.

3.3 Action of monomial matrix group on a weight space

The monomial matrix group $\mathcal{M}_n$ acts on the left and the right of $\mathbb{R}^n$ and $\mathbb{R}^{n \times m}$ in a canonical way (by matrix-vector or matrix-matrix multiplications). This action induces a canonical action of the larger monomial matrix group

$$\mathcal{G}_{\mathcal{U}} := \mathcal{M}_{n_L} \times \mathcal{M}_{n_{L-1}} \times \ldots \times \mathcal{M}_{n_0}, \tag{5}$$

to on the weight space $\mathcal{U}$ of an MLP or CNN defined Equation (1). Each element $g \in \mathcal{G}_{\mathcal{U}}$ has the form

$$g = \left(g^{(L)}, \ldots, g^{(0)} \right),$$

where $g^{(i)} = D^{(i)} \cdot P_{\pi_i}$ for some diagonal matrix $D^{(i)} = \mathrm{diag}(d^{(i)}_1, \ldots, d^{(i)}_{n_i})$ in $\mathcal{D}_{n_i}$ and permutation $\pi_i \in \mathcal{S}_{n_i}$. The action of $\mathcal{G}_{\mathcal{U}}$ on $\mathcal{U}$ is defined formally as follows.

Definition 3.2 (Group action on weight spaces)**.** With the notation as above, the *group action* of $\mathcal{G}_{\mathcal{U}}$ on $\mathcal{U}$ is defined to be the map $\mathcal{G}_{\mathcal{U}} \times \mathcal{U} \to \mathcal{U}$ with $(g, U) \mapsto gU = ([gW], [gb])$, where:

$$[gW]^{(i)} := \left(g^{(i)}\right) \cdot [W]^{(i)} \cdot \left(g^{(i-1)}\right)^{-1} \quad \text{and} \quad [gb]^{(i)} := \left(g^{(i)}\right) \cdot [b]^{(i)}. \tag{6}$$

Or equivalently,

$$[gW]^{(i)}_{jk} := \frac{d^{(i)}_j}{d^{(i-1)}_k} \cdot [W]^{(i)}_{\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)} \quad \text{and} \quad [gb]^{(i)}_j := d^{(i)}_j \cdot [b]^{(i)}_{\pi_i^{-1}(j)}. \tag{7}$$

In this section, we formally describe a subgroup G of $\mathcal{G}_{\mathcal{U}}$ such that its action on $\mathcal{U}$ satisfies the condition: elements of $\mathcal{U}$ in the same orbit under the action of G define the same function.

From now on, we will fix the activation σ to be the rectified linear unit $\sigma = \mathrm{ReLU}$. The case when σ is another typical activation, such as semilinear (e.g., $\mathrm{LeakyReLU}$) or odd (e.g., $\sin, \tanh$), can be derived similarly.

To determine the group G, we observe that:

- Since the neurons in a hidden layer $i = 1, \ldots, L-1$ have no inherent ordering, the network is invariant to the symmetric group $\mathcal{S}_{n_i}$ of permutations of the neurons in the i-th layer. This results in the symmetries of permutation type of $\mathcal{U}$.
- Since $\sigma(\lambda x) = \lambda\sigma(x)$ for all positive numbers λ and real numbers x, multiplying the bias and all incoming weights at a neuron of the MLP by the same positive number λ leads to scaling its output by λ. This results in the symmetries of scaling type of $\mathcal{U}$.

Based on the above observation, we define the group G of the form

$$G := \{I_{n_L}\} \times \mathcal{M}^{>0}_{n_{L-1}} \times \ldots \times \mathcal{M}^{>0}_{n_1} \times \{I_{n_0}\}, \tag{8}$$

where I_n is the identity matrix of size $n \times n$, $\mathcal{M}^{>0}_n = \mathcal{D}^{>0}_n \rtimes \mathcal{P}_n$ which is the semidirect product of $\mathcal{D}^{>0}_n$ and $\mathcal{P}_n$, and $\mathcal{D}^{>0}_n$ is the group of invertible diagonal matrices whose the nonzero entries are all positive.

With notation as above, it is well-known that the function $f = f(\cdot\,;\,U, \sigma)$ be an MLP or CNN given in Equation (1) with the weight space $U \in \mathcal{U}$ and an activation $\sigma = \mathrm{ReLU}$ will be G-invariant under the action of G, i.e.

$$f(\mathbf{x}\,;\,U, \sigma) = f(\mathbf{x}\,;\,gU, \sigma) \tag{9}$$

for all $g \in G$, $U \in \mathcal{U}$ and $\mathbf{x} \in \mathbb{R}^{n_0}$.

4 Equivariant and Invariant Polynomial Functional Networks

In this section, we introduce a new family of NFNs, called monomial matrix group equivariant polynomial neural functionals (MAGEP-NFNs). These NFNs are equivariant to the monomial matrix group described in the previous section. The equivariant and invariant layers, which are the main building blocks of MAGEP-NFNs, will be presented in Subsections 4.2 after we introduce stable polynomial terms in Subsection 4.1.

4.1 Stable polynomial terms

We follow the parameter-sharing mechanism used in (Tran et al., 2024) for constructing equivariant and invariant layers from $\mathcal{U}$ to ensure low memory consumption and computational efficiency in our model. However, unlike the linear layers utilized in (Tran et al., 2024), we employ polynomial layers. This choice allows us to capture additional relationships between weights from different hidden layers of the input network, thereby enhancing the expressivity of our model. To achieve this, we identify specific polynomials in the input weights that remain "stable" under the action of the group G. The formal definition and determination of these "stable polynomials" are provided in this subsection.

Intuitively, a *stable polynomial term* is a polynomial in the entries of $U \in \mathcal{U}$ such that it is "stable" under the action of G (see Definition 4.1 below). These terms are the main components of our equivariant and invariant polynomial layers in the next subsections.

Recall that $\mathcal{U} = (\mathcal{W}, \mathcal{B})$ is the weight space with L layers, n_i channels at i^{th} layer, and the dimensions of weight and bias are w_i and b_i, respectively (see Equation (1)). While the symmetries of the weight space is given by the group

$$G = \{I_{n_L}\} \times \mathcal{M}_{n_{L-1}}^{>0} \times \ldots \times \mathcal{M}_{n_1}^{>0} \times \{I_{n_0}\}.$$

Consider the case where the weight spaces have the same number of dimensions across all channels, which means $w_i = b_i = d$ for all i.

Definition 4.1 (Stable polynomial terms)**.** Let $U = ([W], [b])$ be an element of $\mathcal{U}$ with weights $[W] = \left([W]^{(L)}, \ldots, [W]^{(1)}\right)$ and biases $[b] = \left([b]^{(L)}, \ldots, [b]^{(1)}\right)$. For each $L \geqslant s > t \geqslant 0$, we define:

$$\begin{aligned}
[W]^{(s,t)} &:= [W]^{(s)} \cdot [W]^{(s-1)} \cdot \ldots \cdot [W]^{(t+1)} && \in \mathbb{R}^{d \times n_s \times n_t}, \\
[Wb]^{(s,t)(t)} &:= [W]^{(s,t)} \cdot [b]^{(t)} && \in \mathbb{R}^{d \times n_s \times 1}.
\end{aligned} \tag{10}$$

In addition, for each $L \geqslant s, t \geqslant 0$, and matrices $\Psi^{(s)(L,t)} \in \mathbb{R}^{1 \times n_L}$ and $\Psi^{(s,0)(L,t)} \in \mathbb{R}^{n_0 \times n_L}$, we also define

$$\begin{aligned}
[bW]^{(s)(L,t)} &:= [b]^{(s)} \cdot \Psi^{(s)(L,t)} \cdot [W]^{(L,t)} && \in \mathbb{R}^{d \times n_s \times n_t}, \\
[WW]^{(s,0)(L,t)} &:= [W]^{(s,0)} \cdot \Psi^{(s,0)(L,t)} \cdot [W]^{(L,t)} && \in \mathbb{R}^{d \times n_s \times n_t}.
\end{aligned} \tag{11}$$

The entries of the matrices $[W]^{(s,t)}$, $[Wb]^{(s,t)(t)}$, $[bW]^{(s)(L,t)}$ and $[WW]^{(s,0)(L,t)}$ defined above are called *stable polynomial terms* of U under the action of G.

In the above definition, we use the notation $[W]$ and $[WW]$ to denote products of the weight matrices $[W]^{(i)}$ with the appropriate index i. The notation $[Wb]$ indicates that this is a product of several weight matrices $[W]^{(i)}$ and a bias vector $[b]^{(j)}$, with appropriate indices i and j. For the indices, we use the notation (s, t) to signify that the considered product contains weight matrices with indices ranging from s down to $t+1$. When the index has two components, for example $[Wb]^{(s,t)(t)}$, the first component (s, t) specifies the range of indices for $[W]$, while the second component (t) indicates the index of the bias vector $[b]$. Specifically, the last two terms $[bW]^{(s)(L,t)}$ and $[WW]^{(s,0)(L,t)}$ contain the matrices Ψ^{-} to multiply two matrices of different sizes from the left and the right.

Proposition 4.2 (Stable polynomial terms as generalization of weights and biases)**.** *With notation as above, then for all $L \geqslant s > t > r \geqslant 0$, we have*

$$\begin{aligned}
[W]^{(s,s-1)} &= [W]^{(s)} && \in \mathbb{R}^{d \times n_s \times n_{s-1}}, \\
[W]^{(s,t)} \cdot [W]^{(t,r)} &= [W]^{(s,r)} && \in \mathbb{R}^{d \times n_s \times n_r}. \\
[bW]^{(s)(s,t)} \cdot [W]^{(t,r)} &= [bW]^{(s,r)} && \in \mathbb{R}^{d \times n_s \times n_r}, \\
[W]^{(s,t)} \cdot [Wb]^{(t,r)} &= [Wb]^{(s,r)} && \in \mathbb{R}^{d \times n_s \times n_r}.
\end{aligned}$$

The above proposition shows that the stable polynomial terms can be viewed as a generalization of the entries of the weight matrices $[W]^{(i)}$ and bias vectors $[b]^{(i)}$. The stable polynomial terms defined above are actually "stable" under the action of G in the sense presented in the following theorem.

Theorem 4.3 (Stable polynomial terms are "stable")**.** *With notation as above, let $gU = ([gW], [gb])$ be the element of $\mathcal{U}$ obtained by acting $g = (g^{(L)}, \ldots, g^{(0)}) \in G$ on $U = ([W], [b])$. Then we have*

$$\begin{aligned}
[gW]^{(s,t)} &= \left(g^{(s)}\right) \cdot [W]^{(s,t)} \cdot \left(g^{(t)}\right)^{-1}, \\
[gb]^{(s)} &= \left(g^{(s)}\right) \cdot [b]^{(s)}, \\
[gWgb]^{(s,t)(t)} &= \left(g^{(s)}\right) \cdot [Wb]^{(s,t)(t)}, \\
[gbgW]^{(s)(L,t)} &= \left(g^{(s)}\right) \cdot [bW]^{(s)(L,t)} \cdot \left(g^{(t)}\right)^{-1}, \\
[gWgW]^{(s,0)(L,t)} &= \left(g^{(s)}\right) \cdot [WW]^{(s,0)(L,t)} \cdot \left(g^{(t)}\right)^{-1}.
\end{aligned}$$

Intuitively, the above theorem states that the stable polynomials are compatible with the action of the group G. This property facilitates the efficient computation of equivariant and invariant layers when employing the weight-sharing mechanism.

Our equivariant and invariant polynomial maps will be defined to be a linear combinations of input weights and the considered stable polynomials. In particular, we define a polynomial map $I : \mathcal{U} \to \mathbb{R}^{d'}$ with maps each element $U \in \mathcal{U}$ to the vector $I(U) \in \mathbb{R}^{d'}$ of the form

$$\begin{aligned} I(U) := & \sum_{L \geqslant s > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,t):pq} \cdot [W]^{(s,t)}_{pq} + \sum_{L \geqslant s > 0} \sum_{p=1}^{n_s} \Phi_{(s):p} \cdot [b]^{(s)}_{p} \\ & + \sum_{L \geqslant s > t > 0} \sum_{p=1}^{n_s} \Phi_{(s,t)(t):p} \cdot [Wb]^{(s,t)(t)}_{p} \\ & + \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s)(L,t):pq} \cdot [bW]^{(s)(L,t)}_{pq} \\ & + \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,0)(L,t):pq} \cdot [WW]^{(s,0)(L,t)}_{pq} + \Phi_1. \end{aligned} \tag{12}$$

Intuitively speaking, $I(U)$ is a linear combination of the entries from the input weights $[W]^{(s)}$ and biases $[b]^{(s)}$, as well as all entries from the stable polynomial terms $[W]^{(s,t)}$, $[Wb]^{(s,t)(t)}$, $[bW]^{(s)(s,t)}$, and $[WW]^{(s,L)(0,t)}$ for all appropriate indices s and t. Here, the coefficients Φ_{-} and the connection matrix Ψ^{-} (inside $[bW]^{(s)(s,t)}$ and $[WW]^{(s,L)(0,t)}$) are learnable parameters. The stable nature of the stable polynomial terms will be helpful in the computation of equivariant and invariant layers via weight-sharing mechanism.

The following theorem describes a linear dependence of the terms in $I(U)$. A formal statement of this theorem can be found in Theorem B.6 in the appendix.

Theorem 4.4 (Linear dependence of stable polynomials)**.** *For a given pair of coefficients matrix Φ_{-} and Ψ^{-}, if $I(U)$ given in Equation 12 is equal to zero for all input weights $U \in \mathcal{U}$, then we have*

$$\sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} + \sum_{L > s > 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq} = 0, \tag{13}$$

and

$$\sum_{p=1}^{n_L} \Phi_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_{p} + \sum_{p=1}^{n_t} \sum_{q=1}^{n_t} \Phi_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq} = 0, \quad L > t > 0, \tag{14}$$

and all entries of Φ_{-} and Ψ^{-}, except those appear in the above two equations, are equal to zero.

Intuitively speaking, almost stable polynomial terms are linearly independent over the reals $\mathbb{R}$, except those in Equation 13 and Equation 14. This linear dependence property of the stable polynomials is essential in the computation of equivariant and invariant polynomial layers using weight-sharing mechanism. The proofs of Proposition 4.2, Theorem 4.3, and Theorem 4.4 can be found in Appendix B.

4.2 Polynomial Invariant and Equivariant Layers

We now proceed to construct G-invariant polynomial layers. The construction of G-equivariant polynomial layers is similar and will be derived in detail in Appendix C. These polynomial layers serve as the fundamental building blocks for our MAGEP-NFNs.

We define a polynomial map $I : \mathcal{U} \to \mathbb{R}^{d'}$ with maps each element $U \in \mathcal{U}$ to the vector $I(U) \in \mathbb{R}^{d'}$ of the form given in Equation 12. To make I to be G-invariant, the learnable parameters Φ_{-} and Ψ^{-} must satisfy a system of constraints (usually called *parameter sharing*), which are induced from the condition $I(gU) = I(U)$ for all $g \in G$ and $U \in \mathcal{U}$. We show in details what are these constraints and how to derive the concrete formula of I in Appendix C. The formula of I is then determined by

$$\begin{aligned} I(U) =& \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L,0)(L,0):pq} \cdot [WW]_{pq}^{(L,0)(L,0)} + \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L,0):pq} \cdot [W]_{pq}^{(L,0)} \\ &+ \sum_{L>s>0}\sum_{p=1}^{n_s} \Phi_{(s,0)(L,s):\bullet\bullet} \cdot [WW]_{pp}^{(s,0)(L,s)} + \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L)(L,0):pq} \cdot [bW]_{pq}^{(L)(L,0)} \\ &+ \sum_{L>t>0}\sum_{p=1}^{n_L} \Phi_{(L,t)(t):p} \cdot [Wb]_{p}^{(L,t)(t)} + \sum_{L>t>0}\sum_{p=1}^{n_t} \Phi_{(t)(L,t):\bullet\bullet} \cdot [bW]_{pp}^{(t)(L,t)} \\ &+ \sum_{p=1}^{n_L} \Phi_{(L):p} \cdot [b]_p^{(L)} + \Phi_1. \end{aligned} \tag{15}$$

In the above formula, the bullet $\bullet$ indicates that the value of the corresponding coefficient is independent of the index at the bullet.

To conclude, we have:

Theorem 4.5. *With notation as above, the polynomial map $I \colon \mathcal{U} \to \mathbb{R}^{d'}$ defined by Equation (15) is G-invariant. Moreover, if a map given in Equation (12) is G-invariant, then it has the form given in Equation (15).*

Remark (Comparison to the invariant/equivariant linear layers in (Tran et al., 2024)). Equation equation 15 describes the invariant polynomial layer derived from the parameter-sharing mechanism of our MAGEP-NFNs. In contrast, the invariant equivariant layer proposed in (Tran et al., 2024) is an ad hoc formulation and does not result from a parameter-sharing mechanism. Consequently, there is no direct relationship between our invariant layer and the invariant layer in (Tran et al., 2024).

However, the equivariant polynomial layer in our MAGEP-NFNs and the equivariant linear layer from (Tran et al., 2024) are related. Specifically, the equivariant layer in (Tran et al., 2024) is exactly the linear component of our equivariant polynomial layer. Due to the lengthy formulation and construction process, we have provided the details of the equivariant polynomial layers in Appendix D.4.

5 Experimental Results

In this session, we empirically evaluate the performance of our Monomial Matrix Group Polynomial Equivariant NFNs (MAGEP-NFNs) across various equivariant and invariant tasks. For invariant tasks, we apply our model to classifying Implicit Neural Representations of images and predicting CNN generalization from weights. The equivariant task involves weight space style editing. Our experiments aim to demonstrate that MAGEP-NFNs achieves superior or competitive performance compared to other baselines with equivalent parameter counts. We conduct 5 runs for each experiment and report the averaged results. For comprehensive details on hyperparameter settings and training details, please refer to Appendix E.

5.1 Classifying Implicit Neural Representations of Images

Experiment setup. In this experiment, we aim to determine which class each pretrained Implicit Neural Representation (INR) weight was trained on. Following (Tran et al., 2024), we employ three distinct INR weight datasets (Zhou et al., 2024b), each was trained on a different image dataset: CIFAR-10 (Krizhevsky & Hinton, 2009), FashionMNIST (Xiao et al., 2017), and MNIST (LeCun & Cortes, 2005). Each INR weight is trained to encode a single image from its respective class, capturing the image structure by mapping pixel coordinates (x, y) to the corresponding pixel color values—represented as 3-channel RGB values for CIFAR-10 and 1-channel grayscale values for MNIST and FashionMNIST. The varying complexity and diversity of the datasets provide a robust test for evaluating MAGEP-NFN's performance, demonstrating its effectiveness and benchmarking it against existing models.

Table 1: Classification train and test accuracies (%) for implicit neural representations of MNIST, FashionMNIST, and CIFAR-10. Uncertainties indicate standard error over 5 runs, baseline results are from (Tran et al., 2024).

	MNIST	CIFAR-10	FashionMNIST
MLP	10.62 ± 0.54	10.48 ± 0.74	9.95 ± 0.36
NP (Zhou et al., 2024b)	$\underline{69.82 \pm 0.42}$	33.74 ± 0.26	58.21 ± 0.31
HNP (Zhou et al., 2024b)	66.02 ± 0.51	31.61 ± 0.22	57.43 ± 0.46
Monomial-NFN (Tran et al., 2024)	68.43 ± 0.51	$\underline{34.23 \pm 0.33}$	$\underline{61.15 \pm 0.55}$
MAGEP-NFNs (ours)	$\mathbf{77.55 \pm 0.68}$	$\mathbf{37.18 \pm 0.30}$	$\mathbf{62.83 \pm 0.57}$

Results. We present the performance of our model alongside several baseline models, including MLP, NP (Zhou et al., 2024b), HNP (Zhou et al., 2024b), and Monomial-NFN (Tran et al., 2024). As shown in Table 1, our model achieves the highest test accuracies across all INR datasets. Notably, it outperforms the second-best model on the MNIST dataset, by a significant margin of 7.73%. For the CIFAR-10 and FashionMNIST datasets, our model also demonstrates substantial improvements, with accuracy gains of 2.95% and 1.68%, respectively, over the existing baselines. These results indicate that our model leverages the embedded information from the pretrained INRs more effectively than any of the compared baselines. This consistent superior performance across various INR datasets highlights the effectiveness of MAGEP-NFN. It also suggests that our model generalizes well to INR weights embedded with different image structures and complexities.

5.2 Predicting CNN generalization from weights

Experiment setup. For this experiment, we focus on predicting the generalization performance of pretrained CNNs based solely on their weights, without evaluating them on test data. We utilize the Small CNN Zoo dataset (Unterthiner et al., 2020), which contains various pretrained CNN models trained with different combinations of hyperparameters and activation functions. For our study, we split the Small CNN Zoo into two subsets: one comprising networks using ReLU activations and the other using Tanh activations. These two types of CNNs follow different group actions: $\mathcal{M}_n^{>0}$ for Relu networks (see Equation (8)) and $\mathcal{M}_{n_i}^{\pm 1}$ for Tanh networks (see Remark **??**).

To evaluate the robustness of our model to input transformations under group actions, we augment the ReLU dataset by applying randomly sampled group actions $\mathcal{M}_n^{>0}$. Specifically, we randomly sampling the diagonal elements $\mathcal{D}_{n,ii}^{>0}$ of the matrix $\mathcal{D}_n^{>0}$, with each element drawn from uniform distributions over different ranges, defined as $\mathcal{U}[1, 10^i]$ for $i = 1, 2, 3, 4$. To further diversify the transformations, we also randomly sample the permutation matrix $\mathcal{P}_n$.

Results. Table 2 illustrates the performance of all models trained on the ReLU subset, where our MAGEP-NFNs model clearly outperforms all other baselines. Notably, it demonstrates robustness to scale and permutation symmetry, similar to Monomial-NFN, while consistently surpassing its performance across both the original and all augmented dataset settings. This suggests that incorporating polynomial layers allows our model to capture more information from the weights across different hidden layers, compared to Monomial-NFN, thereby enhancing expressivity. On the original dataset, our model achieves a Kendall's τ performance gap of 0.007 over other baselines, and maintaining at least a 0.012 advantage in all other augmented settings. Similarly, Table 3 reveals that MAGEP-NFNs achieves the highest Kendall's τ with Tanh activation, further reinforcing its superior accuracy across different network configurations.

5.3 Weight space style editing

Experiment setup. In this experiment, we focus on modifying the weights of SIREN (Sitzmann et al., 2020) to modify the image encoded within each model. We utilize the pretrained models from paper (Zhou et al., 2024b), which encode images from the CIFAR-10 and MNIST datasets. Specifically, we address two tasks aimed at modifying the embedded information: enhancing the contrast of CIFAR-10 images and dilating MNIST images encoded in the SIREN models. We report

Table 2: Performance prediction of CNNs on the ReLU subset of Small CNN Zoo with varying scale augmentations. We use Kendall's τ as the evaluation metric. The uncertainty bars indicate the standard deviation across 5 runs.

	Augment settings				
	No augment	$\mathcal{U}[1, 10^1]$	$\mathcal{U}[1, 10^2]$	$\mathcal{U}[1, 10^3]$	$\mathcal{U}[1, 10^4]$
STATNet (Unterthiner et al., 2020)	0.915 ± 0.002	0.894 ± 0.0001	0.853 ± 0.007	0.523 ± 0.02	0.516 ± 0.001
NP (Zhou et al., 2024b)	0.920 ± 0.003	0.900 ± 0.002	0.898 ± 0.003	0.884 ± 0.002	0.884 ± 0.002
HNP (Zhou et al., 2024b)	0.926 ± 0.003	0.913 ± 0.001	0.903 ± 0.003	0.891 ± 0.003	0.601 ± 0.02
Monomial-NFN (Tran et al., 2024)	0.922 ± 0.001	0.920 ± 0.001	0.919 ± 0.001	0.920 ± 0.002	0.920 ± 0.001
MAGEP-NFNs (ours)	$\mathbf{0.933 \pm 0.001}$	$\mathbf{0.933 \pm 0.001}$	$\mathbf{0.933} \pm 0.001$	$\mathbf{0.932} \pm 0.001$	$\mathbf{0.932} \pm 0.001$

Table 3: Performance prediction of CNNs on the Tanh subset of Small CNN Zoo. We use Kendall's τ as the evaluation metric. The uncertainty bars indicate the standard deviation across 5 runs.

Model	Kendall's τ
STATNet (Unterthiner et al., 2020)	0.913 ± 0.0012
NP (Zhou et al., 2024b)	0.925 ± 0.0013
HNP (Zhou et al., 2024b)	0.933 ± 0.0019
Monomial-NFN (Tran et al., 2024)	0.939 ± 0.0004
MAGEP-NFNs (ours)	$\mathbf{0.940 \pm 0.001}$

the MSE loss between the images encoded in the modified SIREN network and the ground truth contrast-enhanced CIFAR-10 images or dilated MNIST images.

Results. Table 4 demonstrates that our model achieves performance comparable to other baselines. Specifically, MAGEP-NFNs matches the performance of NP and Monomial-NFN in the contrast-enhancing task on the CIFAR-10 dataset. Additionally, our model outperforms Monomial-NFN in the dilation task on the MNIST dataset, while achieving similar results to NP. Interestingly, NP remains a strong candidate in the weight editing tasks, and our model consistently performs on par with NP across both experiments.

Table 4: Test mean squared error (lower is better) between weight-space editing methods and ground-truth image-space transformations. Uncertainties indicate standard error over 5 runs

	Contrast (CIFAR-10)	Dilate (MNIST)
MLP	0.031 ± 0.001	0.306 ± 0.001
NP (Zhou et al., 2024b)	$\mathbf{0.020 \pm 0.002}$	$\mathbf{0.068 \pm 0.002}$
HNP (Zhou et al., 2024b)	0.021 ± 0.002	0.071 ± 0.001
Monomial-NFN (Tran et al., 2024)	$\mathbf{0.020 \pm 0.001}$	0.069 ± 0.002
MAGEP-NFNs (ours)	$\mathbf{0.020 \pm 0.001}$	$\mathbf{0.068 \pm 0.002}$

6 Conclusion

We have developed MAGEP-NFN, a novel NFN that is equivariant to both permutations and scaling symmetries. Our approach follows a parameter-sharing mechanism; however, unlike previous works, we construct an equivariant polynomial layer that incorporates stable polynomial terms. This polynomial formulation enables us to capture relationships between weights from different input hidden layers, thereby enhancing the expressivity of MAGEP-NFN while maintaining low memory consumption and efficient running time. Experimental results demonstrate that our model achieves competitive performance and efficiency compared to existing baselines.

One limitation of the equivariant polynomial layers proposed in this paper is that they are applied to a specific architecture design. However, since our method is based on a parameter-sharing mechanism, it is applicable to other architectures with additional operators (such as layer normalization, softmax, pooling) and different activation functions, provided that the symmetric group of the weight network is known.

Ethics Statement. Given the nature of the work, we do not foresee any negative societal and ethical impacts of our work.

Reproducibility Statement. Source codes for our experiments are provided in the supplementary materials of the paper. The details of our experimental settings are given in Section 5 and the Appendix E. All datasets that we used in the paper are published, and they are easy to access in the Internet.

REFERENCES

Francesca Albertini and Eduardo D Sontag. Identifiability of discrete-time neural networks. In *Proc. European Control Conference*, pp. 460–465. Springer Berlin, 1993a.

Francesca Albertini and Eduardo D Sontag. For neural networks, function determines form. *Neural networks*, 6(7):975–990, 1993b.

Zeyuan Allen-Zhu, Yuanzhi Li, and Zhao Song. A convergence theory for deep learning via over-parameterization. In *International conference on machine learning*, pp. 242–252. PMLR, 2019.

Bruno Andreis, Soro Bedionita, and Sung Ju Hwang. Set-based neural network encoding. *arXiv preprint arXiv:2305.16625*, 2023.

Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando De Freitas. Learning to learn by gradient descent by gradient descent. *Advances in neural information processing systems*, 29, 2016.

Maor Ashkenazi, Zohar Rimon, Ron Vainshtein, Shir Levi, Elad Richardson, Pinchas Mintz, and Eran Treister. Nern–learning neural representations for neural networks. *arXiv preprint arXiv:2212.13554*, 2022.

Vijay Badrinarayanan, Bamdev Mishra, and Roberto Cipolla. Understanding symmetries in deep networks. *arXiv preprint arXiv:1511.01029*, 2015.

Bowen Baker, Otkrist Gupta, Ramesh Raskar, and Nikhil Naik. Accelerating neural architecture search using performance prediction. *arXiv preprint arXiv:1705.10823*, 2017.

Matthias Bauer, Emilien Dupont, Andy Brock, Dan Rosenbaum, Jonathan Richard Schwarz, and Hyunjik Kim. Spatial functa: Scaling functa to imagenet classification and generation. *arXiv preprint arXiv:2302.03130*, 2023.

Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019.

Samy Bengio, Yoshua Bengio, Jocelyn Cloutier, and Jan Gecsei. On the optimization of a synaptic learning rule. In *Optimality in Biological and Artificial Networks?*, pp. 265–287. Routledge, 2013.

Phuong Bui Thi Mai and Christoph Lampert. Functional vs. parametric equivalence of relu networks. In *8th International Conference on Learning Representations*, 2020.

An Mei Chen, Haw-minn Lu, and Robert Hecht-Nielsen. On the geometry of feedforward neural network error surfaces. *Neural computation*, 5(6):910–927, 1993.

Nicola De Cao, Wilker Aziz, and Ivan Titov. Editing factual knowledge in language models. *arXiv preprint arXiv:2104.08164*, 2021.

Luca De Luigi, Adriano Cardace, Riccardo Spezialetti, Pierluigi Zama Ramirez, Samuele Salti, and Luigi Di Stefano. Deep learning on implicit neural representations of shapes. *arXiv preprint arXiv:2302.05438*, 2023.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: pre-training of deep bidirectional transformers for language understanding. In Jill Burstein, Christy Doran, and Thamar Solorio (eds.), *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT 2019, Minneapolis, MN, USA, June 2-7, 2019, Volume 1 (Long and Short Papers)*, pp. 4171–4186. Association for Computational Linguistics, 2019. doi: 10.18653/V1/N19-1423. URL https://doi.org/10.18653/v1/n19-1423.

Simon Du, Jason Lee, Haochuan Li, Liwei Wang, and Xiyu Zhai. Gradient descent finds global minima of deep neural networks. In *International conference on machine learning*, pp. 1675–1685. PMLR, 2019.

Emilien Dupont, Yee Whye Teh, and Arnaud Doucet. Generative models as distributions of functions. *arXiv preprint arXiv:2102.04776*, 2021.

Emilien Dupont, Hyunjik Kim, SM Eslami, Danilo Rezende, and Dan Rosenbaum. From data to functa: Your data point is a function and you can treat it like one. *arXiv preprint arXiv:2201.12204*, 2022.

Gabriel Eilertsen, Daniel Jönsson, Timo Ropinski, Jonas Unger, and Anders Ynnerman. Classifying the classifier: dissecting the weight space of neural networks, 2020.

Ziya Erkoç, Fangchang Ma, Qi Shan, Matthias Nießner, and Angela Dai. Hyperdiffusion: Generating implicit neural fields with weight-space diffusion. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 14300–14310, 2023.

Charles Fefferman and Scott Markel. Recovering a feed-forward net from its output. *Advances in neural information processing systems*, 6, 1993.

Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. *arXiv preprint arXiv:1803.03635*, 2018.

Charles Godfrey, Davis Brown, Tegan Emerson, and Henry Kvinge. On the symmetries of deep learning models and their internal representations. *Advances in Neural Information Processing Systems*, 35:11893–11905, 2022.

Jean Harb, Tom Schaul, Doina Precup, and Pierre-Luc Bacon. Policy evaluation networks. *arXiv preprint arXiv:2002.11833*, 2020.

Kaiming He, X. Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2015. URL https://api.semanticscholar.org/CorpusID:206594692.

Robert Hecht-Nielsen. On the algebraic structure of feedforward network weight spaces. In *Advanced Neural Computers*, pp. 129–135. Elsevier, 1990.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Comput.*, 9(8): 1735–1780, 1997. doi: 10.1162/NECO.1997.9.8.1735. URL https://doi.org/10.1162/neco.1997.9.8.1735.

John M. Jumper, Richard O. Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russell Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, R. D. Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David L. Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, July 2021. doi: 10.1038/s41586-021-03819-2. URL https://lens.org/103-212-983-826-945. https://europepmc.org/article/MED/34265844 ; https://www.nature.com/articles/s41586-021-03819-2 ; https://www.mendeley.com/catalogue/bde88f33-525c-3af0-823a-3bb305a93020/ ; https://facultyopinions.com/prime/740477161 ; https://pubmed.ncbi.nlm.nih.gov/34265844/

; https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8371605/ ; https://www.nature.com/articles/s41586-021-03819-2.pdf ; https://europepmc.org/abstract/MED/34265844 ; https://econpapers.repec.org/RePEc:nat:nature:v:596:y:2021:i:7873:d:10.1038_s41586-021-03819-2 ; https://www.scienceopen.com/document?vid=b1f93136-f0f0-4c78-8e45-27bd037e9bb9 ; http://rucweb.tsg211.com/http/77726476706e69737468656265737421e7e056d229317c456c0dc7af9758/articles/s41586-021-03819-2 ; http://pubmed02.keyan123.cn/34265844/.

Ioannis Kalogeropoulos, Giorgos Bouritsas, and Yannis Panagakis. Scale equivariant graph metanetworks. *arXiv preprint arXiv:2406.10685*, 2024.

Boris Knyazev, Michal Drozdzal, Graham W Taylor, and Adriana Romero Soriano. Parameter prediction for unseen deep architectures. *Advances in Neural Information Processing Systems*, 34:29433–29448, 2021.

Miltiadis Kofinas, Boris Knyazev, Yan Zhang, Yunlu Chen, Gertjan J Burghouts, Efstratios Gavves, Cees GM Snoek, and David W Zhang. Graph neural networks for learning equivariant representations of neural networks. *arXiv preprint arXiv:2403.12143*, 2024.

Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical Report 0, University of Toronto, Toronto, Ontario, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. In Peter L. Bartlett, Fernando C. N. Pereira, Christopher J. C. Burges, Léon Bottou, and Kilian Q. Weinberger (eds.), *Advances in Neural Information Processing Systems 25: 26th Annual Conference on Neural Information Processing Systems 2012. Proceedings of a meeting held December 3-6, 2012, Lake Tahoe, Nevada, United States*, pp. 1106–1114, 2012. URL https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html.

Vera Kurkova and Paul C Kainen. Functionally equivalent feedforward neural networks. *Neural Computation*, 6(3):543–558, 1994.

Yann LeCun and Corinna Cortes. The mnist database of handwritten digits. 2005. URL https://api.semanticscholar.org/CorpusID:60282629.

Derek Lim, Haggai Maron, Marc T Law, Jonathan Lorraine, and James Lucas. Graph metanetworks for processing diverse neural architectures. *arXiv preprint arXiv:2312.04501*, 2023.

Luke Metz, James Harrison, C Daniel Freeman, Amil Merchant, Lucas Beyer, James Bradbury, Naman Agrawal, Ben Poole, Igor Mordatch, Adam Roberts, et al. Velo: Training versatile learned optimizers by scaling up. *arXiv preprint arXiv:2211.09760*, 2022.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

Eric Mitchell, Charles Lin, Antoine Bosselut, Chelsea Finn, and Christopher D Manning. Fast model editing at scale. *arXiv preprint arXiv:2110.11309*, 2021.

Aviv Navon, Aviv Shamsian, Idan Achituve, Ethan Fetaya, Gal Chechik, and Haggai Maron. Equivariant architectures for learning in deep weight spaces. In *International Conference on Machine Learning*, pp. 25790–25816. PMLR, 2023.

Behnam Neyshabur, Russ R Salakhutdinov, and Nati Srebro. Path-sgd: Path-normalized optimization in deep neural networks. *Advances in neural information processing systems*, 28, 2015.

Roman Novak, Yasaman Bahri, Daniel A Abolafia, Jeffrey Pennington, and Jascha Sohl-Dickstein. Sensitivity and generalization in neural networks: an empirical study. *arXiv preprint arXiv:1802.08760*, 2018.

William Peebles, Ilija Radosavovic, Tim Brooks, Alexei A Efros, and Jitendra Malik. Learning to learn with generative models of neural network checkpoints. *arXiv preprint arXiv:2209.12892*, 2022.

M. Raissi, P. Perdikaris, and G.E. Karniadakis. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, 378:686–707, 2019. ISSN 0021-9991. doi: https://doi.org/10.1016/j.jcp.2018.10.045. URL https://www.sciencedirect.com/science/article/pii/S0021999118307125.

Joseph J Rotman. *An introduction to the theory of groups*, volume 148. Springer Science & Business Media, 2012.

David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning internal representations by error propagation. 1986. URL https://api.semanticscholar.org/CorpusID:62245742.

Thomas Philip Runarsson and Magnus Thor Jonsson. Evolution and design of distributed learning rules. In *2000 IEEE Symposium on Combinations of Evolutionary Computation and Neural Networks. Proceedings of the First IEEE Symposium on Combinations of Evolutionary Computation and Neural Networks (Cat. No. 00*, pp. 59–63. IEEE, 2000.

Konstantin Schürholt, Dimche Kostadinov, and Damian Borth. Self-supervised representation learning on neural network weights for model characteristic prediction. *Advances in Neural Information Processing Systems*, 34:16481–16493, 2021.

Konstantin Schürholt, Boris Knyazev, Xavier Giró-i Nieto, and Damian Borth. Hyper-representations as generative models: Sampling unseen neural network weights. *Advances in Neural Information Processing Systems*, 35:27906–27920, 2022a.

Konstantin Schürholt, Diyar Taskiran, Boris Knyazev, Xavier Giró-i Nieto, and Damian Borth. Model zoos: A dataset of diverse populations of neural network models. *Advances in Neural Information Processing Systems*, 35:38134–38148, 2022b.

Anton Sinitsin, Vsevolod Plokhotnyuk, Dmitriy Pyrkin, Sergei Popov, and Artem Babenko. Editable neural networks. *arXiv preprint arXiv:2004.00345*, 2020.

Vincent Sitzmann, Julien N. P. Martel, Alexander W. Bergman, David B. Lindell, and Gordon Wetzstein. Implicit neural representations with periodic activation functions. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/53c04118df112c13a8c34b38343b9c10-Abstract.html.

Samuel Sokota, Hengyuan Hu, David J Wu, J Zico Kolter, Jakob Nicolaus Foerster, and Noam Brown. A fine-tuning approach to belief state modeling. In *International Conference on Learning Representations*, 2021.

Kenneth O Stanley. Compositional pattern producing networks: A novel abstraction of development. *Genetic programming and evolvable machines*, 8:131–162, 2007.

Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott E. Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2015, Boston, MA, USA, June 7-12, 2015*, pp. 1–9. IEEE Computer Society, 2015. doi: 10.1109/CVPR.2015.7298594. URL https://doi.org/10.1109/CVPR.2015.7298594.

Hoang V Tran, Thieu N Vo, Tho H Tran, An T Nguyen, and Tan Minh Nguyen. Monomial matrix group equivariant neural functional networks. *arXiv preprint arXiv:2409.11697*, 2024.

Thomas Unterthiner, Daniel Keysers, Sylvain Gelly, Olivier Bousquet, and Ilya Tolstikhin. Predicting neural network accuracy from weights. *arXiv preprint arXiv:2002.11448*, 2020.

Thomas Unterthiner, Daniel Keysers, Sylvain Gelly, Olivier Bousquet, and Ilya Tolstikhin. Predicting neural network accuracy from weights, 2021.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 5998–6008, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *CoRR*, abs/1708.07747, 2017. URL http://arxiv.org/abs/1708.07747.

Allan Zhou, Chelsea Finn, and James Harrison. Universal neural functionals. *arXiv preprint arXiv:2402.05232*, 2024a.

Allan Zhou, Kaien Yang, Kaylee Burns, Adriano Cardace, Yiding Jiang, Samuel Sokota, J Zico Kolter, and Chelsea Finn. Permutation equivariant neural functionals. *Advances in Neural Information Processing Systems*, 36, 2024b.

Allan Zhou, Kaien Yang, Yiding Jiang, Kaylee Burns, Winnie Xu, Samuel Sokota, J Zico Kolter, and Chelsea Finn. Neural functional transformers. *Advances in Neural Information Processing Systems*, 36, 2024c.

Supplement to "Equivariant Polynomial Functional Networks"

Table of Contents

A PRELIMINARIES

This section contains notations and basic results on matrices and polynomials that will be used throughout the paper. We will mainly focus on matrices with real entries (real matrices) and polynomials with real coefficients (real polynomials). We will omit almost all of the proofs in this section as they are well-known. These results will be use in proofs in the rest of the paper.

A.1 ENTRIES OF MATRICES

A real matrix A with m rows and n columns is an element of $\mathbb{R}^{m\times n}$:

$$A = (a_{ij})_{1\leqslant i\leqslant m, 1\leqslant j\leqslant n} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \in \mathbb{R}^{m\times n}.$$

The entry in the i^{th}-row and the j^{th}-column of A, or the (i,j) entry of A, is denoted by $A_{ij} = a_{ij}$. The i^{th}-row of A and j^{th}-column of A are, respectively, denoted by:

$$\begin{aligned} A_{i*} &= (a_{i1}\ ,\ a_{i2}\ ,\ \dots\ ,\ a_{in}) && \in \mathbb{R}^{1\times n}, \\ A_{*j} &= (a_{1j}\ ,\ a_{2j}\ ,\ \dots\ ,\ a_{mj})^\top && \in \mathbb{R}^{m\times 1}. \end{aligned}$$

Remark. Sometimes, a comma is added between two subscript indices to make sure there will be no confusion, i.e. $A_{i,j}, a_{i,j}, A_{i,*}, A_{*,j}$.

Let $A^{(L)}, \dots, A^{(2)}, A^{(1)}$ be L matrices such that the matrix product:

$$A^{(L)} \cdot \ldots \cdot A^{(2)} \cdot A^{(1)},$$

is well-defined.

Proposition A.1. *The (i,j) entry of $A^{(L)} \cdot \ldots \cdot A^{(2)} \cdot A^{(1)}$ is equal to:*

$$\begin{aligned} \left(A^{(L)} \cdot \ldots \cdot A^{(2)} \cdot A^{(1)}\right)_{ij} &= A^{(L)}_{i,*} \cdot A^{(L-1)} \cdot \ldots \cdot A^{(2)} \cdot A^{(1)}_{*,j} \\ &= \sum_{k_{L-1},\dots,k_2,k_1} a^{(L)}_{i,k_{L-1}} \cdot a^{(L)}_{k_{L-1},k_{L-2}} \cdot \ldots \cdot a^{(2)}_{k_2,k_1} \cdot a^{(1)}_{k_1,j}. \end{aligned}$$

In the case where $L = 1$, the above equation is simply $A^{(1)}_{ij} = a^{(1)}_{ij}$.

We set a denotation for matrices that have only one nonzero entry with value 1. The matrix with the 1 in the i^{th}-row and the j^{th}-column, and the rest are 0, is denoted by E_{ij}. Matrix E_{ij} can have any shape, but its shape are usually defined by context, and will be omitted without confusion. The product of matrices of this type is presented as below.

Proposition A.2. *Let $E_{i_1,j_1}, E_{i_2,j_2}, \dots, E_{i_L,j_L}$ be L matrix units such that the product:*

$$E_{i_1,j_1} \cdot E_{i_2,j_2} \cdot \ldots \cdot E_{i_L,j_L},$$

is well-defined. Then:

$$E_{i_1,j_1} \cdot E_{i_2,j_2} \cdot \ldots \cdot E_{i_L,j_L} = \left(\delta_{j_1,i_2} \cdot \delta_{j_2,i_3} \cdot \ldots \cdot \delta_{j_{L-1},i_L}\right) \cdot E_{i_1,j_L},$$

where δ_{ij} is the Kronecker delta:

$$\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j, \\ 1 & \text{if } i = j. \end{cases}$$

We have a direct corollary for $E_{1,1}$'s.

Corollary A.3. *We have:*

$$E_{1,1} \cdot E_{1,1} \cdot \ldots \cdot E_{1,1} = E_{1,1}.$$

A.2 Evaluation of polynomials

Denote $\mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ be the ring of all polynomials with real coefficients in n indeterminates $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

Definition A.4. A *monomial* of $\mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ is a polynomial of $\mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ that has one term.

Remark. In some contexts, a monomial is defined as a polynomial that has one term with coefficient 1. We will use *both* of these definitions simultaneously.

Proposition A.5. *$\mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ is naturally a vector space over $\mathbb{R}$. It is an infinite-dimensional vector space; moreover, the set of all monomials with coefficient 1 in $\mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$ is a basis for the vector space.*

Remark. For $f \in \mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$, by saying monomials in f, we refer to all monomials that appeared in the expression of f.

Polynomial evaluation is computing of the value of a polynomial when the indeterminates are substituted for some values. We have the well-known result.

Proposition A.6. *Let f, g be two polynomials of $\mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$. If f, g are equal at every evaluations, i.e.*

$$f(x_1, \ldots, x_n) = g(x_1, \ldots, x_n) \ , \ \forall (x_1, \ldots, x_n) \in \mathbb{R}^n, \tag{16}$$

then $f = g$. In other words, the only polynomial of $\mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$, that has $\mathbb{R}^n$ as its zero set, is the polynomial $0 \in \mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$.

Remark. The result still holds if $\mathbb{R}$ is replaced by an arbitrary infinite field, but does not hold if $\mathbb{R}$ is replaced by a finite field.

We have a direct corollary.

Corollary A.7. *Let f be a nonzero polynomial of $\mathbb{R}[\mathbf{x}_1, \ldots, \mathbf{x}_n]$. Then there exists $(x_1, \ldots, x_n) \in \mathbb{R}^n$ such that $f(x_1, \ldots, x_n) \neq 0$.*

A.3 Entries of tensors

Proposition A.8. *Let $a = (a_i)_{1 \leqslant i \leqslant n}$ and $b = (b_i)_{1 \leqslant i \leqslant n}$ be two vectors in $\mathbb{R}^n$. If:*

$$a_i \cdot b_j + a_j \cdot b_i = 0, \tag{17}$$

for all $1 \leqslant i, j \leqslant n$, then $a = 0$ or $b = 0$.

Proof. Assume that both of a and b are not equal to 0, then there exists i, j such that a_i and b_j are non-zero. From Equation (17), we have:

$$a_i \cdot b_i + a_i \cdot b_i = 0, \tag{18}$$

so $a_i \cdot b_i = 0$. Since a_i is non-zero, then $b_i = 0$. It implies that:

$$a_i \cdot b_j + a_j \cdot b_i = a_i \cdot b_j + 0 = a_i \cdot b_j \neq 0, \tag{19}$$

which contradicts to Equation (17). So at least one of a and b is equal to 0. ☐

Proposition A.9. *Let $A = (a_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$ and $B = (b_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant n}$ be two matrices in $\mathbb{R}^{m \times n}$. If:*

$$a_{ij} \cdot b_{kl} + a_{kj} \cdot b_{il} + a_{il} \cdot b_{kj} + a_{kl} \cdot b_{ij} = 0, \tag{20}$$

for all $1 \leqslant i, k \leqslant m$ and $1 \leqslant j, l \leqslant n$, then $A = 0$ or $B = 0$.

Proof. Consider Equation (20) when $1 \leqslant j = l \leqslant n$, we have:

$$\begin{aligned} 0 &= a_{ij} \cdot b_{kj} + a_{kj} \cdot b_{ij} + a_{ij} \cdot b_{kj} + a_{kj} \cdot b_{ij} \quad (21) \\ &= 2 \cdot (a_{ij} \cdot b_{kj} + a_{kj} \cdot b_{ij}) \, , \quad (22) \end{aligned}$$

which means:

$$a_{ij} \cdot b_{kj} + a_{kj} \cdot b_{ij} = 0. \tag{23}$$

This holds for all $1 \leqslant i, k \leqslant m$. Apply Proposition A.8, we have $a_{ij} = 0$ for all $1 \leqslant i \leqslant m$, or $b_{ij} = 0$ for all $1 \leqslant i \leqslant m$, which means $A_{*,j} = 0$ or $B_{*,j} = 0$. This holds for all $1 \leqslant j \leqslant n$. Similarly, we have $A_{i,*} = 0$ or $B_{i,*} = 0$ for $1 \leqslant i \leqslant m$. Now, assume that, both of A and B are not equal to 0, then there exists i, j and k, l such that a_{ij} and b_{kl} are non-zero. By previous observation, we have $B_{i,*} = B_{*,j} = A_{k,*} = A_{*,l} = 0$. It implies that:

$$a_{ij} \cdot b_{kl} + a_{kj} \cdot b_{il} + a_{il} \cdot b_{kj} + a_{kl} \cdot b_{ij} = a_{ij} \cdot b_{kl} + 0 + 0 + 0 = a_{ij} \cdot b_{kl} \neq 0, \tag{24}$$

which contradicts to Equation (20). So at least one of A and B is equal to 0. □

Proposition A.8 and Proposition A.9 are, respectively, one-dimensional and two-dimensional cases. By using the same arguments, we will obtain the d-dimensional version belows.

Proposition A.10. *Let d be a positive integer and $n_1, n_2, \ldots, n_d$ be d positive integers. Let:*

$$\begin{aligned} A &= (A_{i_1,i_2,\ldots,i_d})_{1\leqslant i_1\leqslant n_1, 1\leqslant i_2\leqslant n_2,\ldots,1\leqslant i_d\leqslant n_d} &\in \mathbb{R}^{n_1\times n_2\times\ldots\times n_d}, \\ B &= (B_{i_1,i_2,\ldots,i_d})_{1\leqslant i_1\leqslant n_1, 1\leqslant i_2\leqslant n_2,\ldots,1\leqslant i_d\leqslant n_d} &\in \mathbb{R}^{n_1\times n_2\times\ldots\times n_d}. \end{aligned}$$

If for all $1 \leqslant i_1^0, i_1^1 \leqslant n_1, 1 \leqslant i_2^0, i_2^1 \leqslant n_2, \ldots, 1 \leqslant i_d^0, i_d^1 \leqslant n_d$, we have:

$$\sum_{(\alpha_1,\ldots,\alpha_d)\in\{0,1\}^d} \left(A_{i_1^{\alpha_1}, i_2^{\alpha_2}, \ldots, i_d^{\alpha_d}}\right) \cdot \left(B_{i_1^{1-\alpha_1}, i_2^{1-\alpha_2}, \ldots, i_d^{1-\alpha_d}}\right) = 0, \tag{25}$$

then $A = 0$ or $B = 0$.

B Stable Polynomial Terms

Intuitively, a *stable polynomial term* is a polynomial in the entries of $U \in \mathcal{U}$ that is "stable" under the action of G (see Definition B.1 below). The equivariant polynomial layers we aim to construct are linear combinations of these stable polynomial terms. In Subsection B.1, we provide a formal definition for stable polynomial terms as well as their properties. We will study the linear dependence of stable polynomial terms in the language of polynomial rings with real coefficients in Subsections B.2 and B.3. These properties play a central role in determining the parameter-sharing computation of equivariant polynomial layers in the next section.

B.1 Definitions and Basic Properties

Recall the weight space $\mathcal{U}$ given by:

$$\begin{aligned} \mathcal{U} &= \mathcal{W} \times \mathcal{B}, && \text{where:} \\ \mathcal{W} &= \mathbb{R}^{w_L\times n_L\times n_{L-1}} \times \ldots \times \mathbb{R}^{w_2\times n_2\times n_1} \times \mathbb{R}^{w_1\times n_1\times n_0}, \\ \mathcal{B} &= \mathbb{R}^{b_L\times n_L\times 1} \times \ldots \times \mathbb{R}^{b_2\times n_2\times 1} \times \mathbb{R}^{b_1\times n_1\times 1}. \end{aligned}$$

Let us consider the case where the weight spaces have the same number of dimensions across all channels, which means $w_i = b_i = d$ for all i.

Definition B.1 (Stable polynomial terms)**.** Let $U = ([W], [b])$ be an element of $\mathcal{U}$ with weights $[W] = \left([W]^{(L)}, \ldots, [W]^{(1)}\right)$ and biases $[b] = \left([b]^{(L)}, \ldots, [b]^{(1)}\right)$. For each $L \geqslant s > t \geqslant 0$, we define:

$$\begin{aligned} [W]^{(s,t)} &:= [W]^{(s)} \cdot [W]^{(s-1)} \cdot \ldots \cdot [W]^{(t+1)} &&\in \mathbb{R}^{d\times n_s\times n_t}, \\ [Wb]^{(s,t)(t)} &:= [W]^{(s,t)} \cdot [b]^{(t)} &&\in \mathbb{R}^{d\times n_s\times 1}. \end{aligned} \tag{26}$$

In addition, for each $L \geqslant s, t \geqslant 0$, and matrices $\Psi^{(s)(L,t)} \in \mathbb{R}^{1\times n_L}$ and $\Psi^{(s,0)(L,t)} \in \mathbb{R}^{n_0\times n_L}$, we also define

$$\begin{aligned} [bW]^{(s)(L,t)} &:= [b]^{(s)} \cdot \Psi^{(s)(L,t)} \cdot [W]^{(L,t)} &&\in \mathbb{R}^{d\times n_s\times n_t}, \\ [WW]^{(s,0)(L,t)} &:= [W]^{(s,0)} \cdot \Psi^{(s,0)(L,t)} \cdot [W]^{(L,t)} &&\in \mathbb{R}^{d\times n_s\times n_t}. \end{aligned} \tag{27}$$

The entries of the matrices $[W]^{(s,t)}$, $[Wb]^{(s,t)(t)}$, $[bW]^{(s)(L,t)}$ and $[WW]^{(s,0)(L,t)}$ defined above are called *stable polynomial terms* of U under the action of G.

The following observations are direct implications from the definition.

- For all $L \geqslant s > t > r \geqslant 0$:
$$[W]^{(s,s-1)} = [W]^{(s)} \in \mathbb{R}^{d \times n_s \times n_{s-1}}, \tag{28}$$
and
$$[W]^{(s,t)} \cdot [W]^{(t,r)} = [W]^{(s,r)} \in \mathbb{R}^{d \times n_s \times n_r}, \tag{29}$$
by definition. For $g = \left(g^{(L)}, \ldots, g^{(0)}\right) \in \mathcal{G}_{\mathcal{U}}$:
$$[gW]^{(s,t)} = \left(g^{(s)}\right) \cdot [W]^{(s,t)} \cdot \left(g^{(t)}\right)^{-1} \in \mathbb{R}^{d \times n_s \times n_t}. \tag{30}$$
- If $g \in G$, then:
$$[gW]^{(L,t)} = [W]^{(L,t)} \cdot \left(g^{(t)}\right)^{-1} \in \mathbb{R}^{d \times n_L \times n_t} \tag{31}$$
$$[gW]^{(s,0)} = \left(g^{(s)}\right) \cdot [W]^{(s,0)} \in \mathbb{R}^{d \times n_s \times n_0}. \tag{32}$$
- For all $L \geqslant s > t > 0$, we have
$$[gW]^{(s,t)} \cdot [gb]^{(t)} = \left(g^{(s)}\right) \cdot [W]^{(s,t)} \cdot [b]^{(t)} \in \mathbb{R}^{d \times n_s \times 1} \tag{33}$$
- For all $L \geqslant s > 0$, $L > t \geqslant 0$ and $\Psi^{(s,0)(L,t)} \in \mathbb{R}^{d \times n_0 \times n_L}$, we have:
$$\begin{aligned} &[gW]^{(s,0)} \cdot \Psi^{(s,0)(L,t)} \cdot [gW]^{(L,t)} \\ &= \left(g^{(s)}\right) \cdot [W]^{(s,0)} \cdot \Psi^{(s,0)(L,t)} \cdot [W]^{(L,t)} \cdot \left(g^{(t)}\right)^{-1} \in \mathbb{R}^{d \times n_s \times n_t}. \end{aligned} \tag{34}$$
In particular, if $t = s - 1$, we have:
$$\begin{aligned} &[gW]^{(s,0)} \cdot \Psi^{(s,0)(L,s-1)} \cdot [gW]^{(L,s-1)} \\ &= \left(g^{(s)}\right) \cdot [W]^{(s,0)} \cdot \Psi^{(s,0)(L,s-1)} \cdot [W]^{(L,s-1)} \cdot \left(g^{(s-1)}\right)^{-1} \in \mathbb{R}^{d \times n_s \times n_{s-1}}. \end{aligned} \tag{35}$$
- For all $L \geqslant s > 0$ and $\Psi^{(s)(L,t)} \in \mathbb{R}^{d \times 1 \times n_L}$, we have:
$$\begin{aligned} &[gb]^{(s)} \cdot \Psi^{(s)(L,t)} \cdot [gW]^{(L,t)} \\ &= \left(g^{(s)}\right) \cdot [b]^{(s)} \cdot \Psi^{(s)(L,t)} \cdot [W]^{(L,t)} \cdot \left(g^{(t)}\right)^{-1} \in \mathbb{R}^{d \times n_s \times n_t}. \end{aligned} \tag{36}$$

Based on the above observations, we can determine the image of the stable polynomial terms under the action of an element $g \in \mathcal{G}_{\mathcal{U}}$ as follows:
$$\begin{aligned} [gW]^{(s,t)} &= \left(g^{(s)}\right) \cdot [W]^{(s,t)} \cdot \left(g^{(t)}\right)^{-1}, \\ [gb]^{(s)} &= \left(g^{(s)}\right) \cdot [b]^{(s)}, \\ [gWgb]^{(s,t)(t)} &= \left(g^{(s)}\right) \cdot [Wb]^{(s,t)(t)}, \\ [gbgW]^{(s)(L,t)} &= \left(g^{(s)}\right) \cdot [bW]^{(s)(L,t)} \cdot \left(g^{(t)}\right)^{-1}, \\ [gWgW]^{(s,0)(L,t)} &= \left(g^{(s)}\right) \cdot [WW]^{(s,0)(L,t)} \cdot \left(g^{(t)}\right)^{-1}. \end{aligned}$$

In concrete, we have:
$$\begin{aligned} [gW]^{(s,t)}_{pq} &= \frac{d^{(s)}_p}{d^{(t)}_q} \cdot [W]^{(s,t)}_{\pi_s^{-1}(p), \pi_t^{-1}(q)}, \\ [gb]^{(s)}_p &= d^{(s)}_p \cdot [b]^{(s)}_{\pi_s^{-1}(p)}, \\ [gWgb]^{(s,t)(t)}_p &= d^{(s)}_p \cdot [Wb]^{(s,t)(t)}_{\pi_s^{-1}(p)}, \\ [gbgW]^{(s)(L,t)}_{pq} &= \frac{d^{(s)}_p}{d^{(t)}_q} \cdot [bW]^{(s)(L,t)}_{\pi_s^{-1}(p), \pi_t^{-1}(q)}, \\ [gWgW]^{(s,0)(L,t)} &= \frac{d^{(s)}_p}{d^{(t)}_q} \cdot [WW]^{(s,0)(L,t)}_{\pi_s^{-1}(p), \pi_t^{-1}(q)}. \end{aligned}$$

B.2 INPUT WEIGHTS AS INDETERMINATES

To simplify the technical difficulties, we consider the weight space $\mathcal{U}$ in the case where $d = 1$, i.e.,

$$\begin{aligned} \mathcal{U} &= \mathcal{W} \times \mathcal{B}, \quad \text{where:} \\ \mathcal{W} &= \mathbb{R}^{n_L \times n_{L-1}} \times \ldots \times \mathbb{R}^{n_2 \times n_1} \times \mathbb{R}^{n_1 \times n_0}, \\ \mathcal{B} &= \mathbb{R}^{n_L \times 1} \times \ldots \times \mathbb{R}^{n_2 \times 1} \times \mathbb{R}^{n_1 \times 1}. \end{aligned}$$

We introduce the set I consists of indeterminates defined by:

$$I \coloneqq \{\mathbf{x}_{jk}^{(i)} \,:\, 1 \leqslant i \leqslant L, 1 \leqslant j \leqslant n_i, 1 \leqslant k \leqslant n_{i-1}\} \cup \{\mathbf{y}_j^{(i)} \,:\, 1 \leqslant i \leqslant L, 1 \leqslant j \leqslant n_i\}.$$

We have $|I| = \dim \mathcal{U}$. Denote $R = \mathbb{R}[I]$, which is the ring of all polynomials with indeterminates are all elements of I. For $1 \leqslant i \leqslant L$, we define:

$$\begin{aligned} [\mathbf{W}]^{(i)} &\coloneqq \left(\mathbf{x}_{jk}^{(i)}\right)_{1 \leqslant j \leqslant n_i, 1 \leqslant k \leqslant n_{i-1}} && \in R^{n_i \times n_{i-1}}, \\ [\mathbf{b}]^{(i)} &\coloneqq \left(\mathbf{y}_j^{(i)}\right)_{1 \leqslant j \leqslant n_i} && \in R^{n_i \times 1}, \end{aligned}$$

and

$$\begin{aligned} [\mathbf{W}]^{(s,t)} &\coloneqq [\mathbf{W}]^{(s)} \cdot [\mathbf{W}]^{(s-1)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)} && \in R^{n_s \times n_t}, \\ [\mathbf{Wb}]^{(s,t)(t)} &\coloneqq [\mathbf{W}]^{(s,t)} \cdot [\mathbf{b}]^{(t)} && \in R^{n_s \times 1}, \\ [\mathbf{bW}]^{(s)(L,t)} &\coloneqq [\mathbf{b}]^{(s)} \cdot \Psi^{(s)(L,t)} \cdot [\mathbf{W}]^{(L,t)} && \in R^{n_s \times n_t}, \\ [\mathbf{WW}]^{(s,0)(L,t)} &\coloneqq [\mathbf{W}]^{(s,0)} \cdot \Psi^{(s,0)(L,t)} \cdot [\mathbf{W}]^{(L,t)} && \in R^{n_s \times n_t}. \end{aligned}$$

with feasible indices (s,t). The coefficients $\Psi^{(-)}$'s are fixed real matrices and they are omitted from the notations.

Note that the entries of these matrices are stable polynomial terms in which the entries of U are now viewed as indeterminates of the polynomial ring R.

B.3 LINEAR DEPENDENCE OF STABLE POLYNOMIAL TERMS

In this subsection, we derive a necessary condition for the coefficients Φ_{-} and Ψ^{-} such that the following linear combination of stable polynomial terms are identically zero:

$$\begin{aligned} \alpha(\Phi, \Psi) \coloneqq & \sum_{L \geqslant s > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,t):pq} \cdot [\mathbf{W}]_{pq}^{(s,t)} + \sum_{L \geqslant s > 0} \sum_{p=1}^{n_s} \Phi_{(s):p} \cdot [\mathbf{b}]_p^{(s)} \\ & + \sum_{L \geqslant s > t > 0} \sum_{p=1}^{n_s} \Phi_{(s,t)(t):p} \cdot [\mathbf{Wb}]_p^{(s,t)(t)} \\ & + \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s)(L,t):pq} \cdot [\mathbf{bW}]_{pq}^{(s)(L,t)} \\ & + \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,0)(L,t):pq} \cdot [\mathbf{WW}]_{pq}^{(s,0)(L,t)} + \Phi_1 \cdot 1. \end{aligned} \tag{37}$$

Here, α is parameterized by Φ and Ψ, where Φ is a collection of real scalars Φ_{-}'s appeared in the linear combination and Ψ is a collection of real matrices Ψ^{-}'s that be used to define $[\mathbf{bW}]^{(-)}$'s and $[\mathbf{WW}]^{(-)}$'s. The index of each scalar Φ_{-} naturally presents its corresponding polynomial in $\alpha(\Phi, \Psi)$. This necessary and sufficient condition enables us to determine the equivariant polynomial map via parameter sharing later.

We first take a look at entries of $[\mathbf{W}]^{(-)}$'s, $[\mathbf{b}]^{(-)}$'s, $[\mathbf{Wb}]^{(-)}$'s, $[\mathbf{bW}]^{(-)}$'s, $[\mathbf{WW}]^{(-)}$'s. It is clear that for one of these matrices, its entries are *homogeneous polynomials with the same degree*. For example:

- $[\mathbf{W}]^{(-)}$: The polynomial $[\mathbf{W}]^{(s,t)}_{pq}$ has degree $s-t$. All of its monomial terms consist of one $\mathbf{x}^{(i)}_{-}$ for each $s \geqslant i > t$.
- $[\mathbf{b}]^{(-)}$: The polynomial $[\mathbf{b}]^{(s)}_{p}$ has degree 1. All of its monomial terms consist of one $\mathbf{y}^{s}_{-}$.
- $[\mathbf{Wb}]^{(-)}$: The polynomial $[\mathbf{Wb}]^{(s,t)(t)}_{p}$ has degree $s-t+1$. All of its monomial terms consist of one $\mathbf{x}^{(i)}_{-}$ for each $s \geqslant i > t$ and one $\mathbf{y}^{(t)}_{-}$.
- $[\mathbf{bW}]^{(-)}$: The polynomial $[\mathbf{bW}]^{(s)(L,t)}_{p}$ has degree $L-t+1$. All of its monomial terms consist of one $\mathbf{y}^{(s)}_{-}$ and one $\mathbf{x}^{(i)}_{-}$ for each $L \geqslant i > t$.
- $[\mathbf{WW}]^{(-)}$: The polynomial $[\mathbf{WW}]^{(s,0)(L,t)}_{pq}$ has degree $L+s-t$. All of its monomial terms consist of one $\mathbf{x}^{(i)}_{-}$ for each $s \geqslant i > 0$ and one $\mathbf{x}^{(i)}_{-}$ for each $L \geqslant i > t$.
- 1: The polynomial $1 \in R$.

By these above observations, we have:

- $[\mathbf{W}]^{(-)}$, $[\mathbf{WW}]^{(-)}$: Each of the polynomials $[\mathbf{W}]^{(-)}_{-}$'s and $[\mathbf{WW}]^{(-)}_{-}$'s is 0 or a non-constant element in R, and it is a real polynomial of at least one indeterminate from $\mathbf{x}^{(-)}_{-}$'s.
- $[\mathbf{b}]^{(-)}$: Each of the polynomials $[\mathbf{b}]^{(-)}_{-}$'s is a non-constant element in R, and it is a real polynomial of one indeterminate from $\mathbf{y}^{(-)}_{-}$'s.
- $[\mathbf{Wb}]^{(-)}$, $[\mathbf{bW}]^{(-)}$: Each of the polynomials $[\mathbf{Wb}]^{(-)}_{-}$'s and $[\mathbf{bW}]^{(-)}_{-}$'s is 0 or a non-constant element in R, and it is a real polynomial of at least one indeterminate from $\mathbf{x}^{(-)}_{-}$'s and one indeterminate from $\mathbf{y}^{(-)}_{-}$'s.

Therefore, if $\alpha(\Phi, \Psi) = 0$, we must have

$$0 = \sum_{L\geqslant s>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,t):pq} \cdot [\mathbf{W}]^{(s,t)}_{pq} + \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,0)(L,t):pq} \cdot [\mathbf{WW}]^{(s,0)(L,t)}_{pq}, \tag{38}$$

$$0 = \sum_{L\geqslant s>0} \sum_{p=1}^{n_s} \Phi_{(s):p} \cdot [\mathbf{b}]^{(s)}_{p}, \tag{39}$$

$$0 = \sum_{L\geqslant s>t>0} \sum_{p=1}^{n_s} \Phi_{(s,t)(t):p} \cdot [\mathbf{Wb}]^{(s,t)(t)}_{p} + \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s)(L,t):pq} \cdot [\mathbf{bW}]^{(s)(L,t)}_{pq}, \tag{40}$$

$$0 = \Phi_1 \cdot 1. \tag{41}$$

We induce the constraints on Φ and Ψ in these above equations by using the fact that a set of distinct monomials of R is a linear independent set (see Proposition A.5).

- Equation (38): Observe that
  - If $L \geqslant s > t \geqslant 0$ and $(s,t) \neq (L,0)$, then the monomials $\mathbf{x}^{(s)}_{-} \cdot \ldots \cdot \mathbf{x}^{(t+1)}_{-}$'s only appear in the polynomials $[\mathbf{W}]^{(s,t)}_{-}$'s. They do not appear in the polynomials $[\mathbf{W}]^{(s',t')}_{-}$'s for all pairs $(s',t') \neq (s,t)$, and do not appear in the polynomials $[\mathbf{WW}]^{(s',0)(L,t')}_{-}$'s for all pairs (s',t').
  - If $L \geqslant s > 0$, $L > t \geqslant 0$, and $s \neq t$, then the monomials $\left(\mathbf{x}^{(s)}_{-} \cdot \ldots \cdot \mathbf{x}^{(1)}_{-}\right) \cdot \left(\mathbf{x}^{(L)}_{-} \cdot \ldots \cdot \mathbf{x}^{(t+1)}_{-}\right)$'s only appear in the polynomials $[\mathbf{WW}]^{(s,0)(L,t)}_{-}$'s. They do not appear in the polynomials $[\mathbf{W}]^{(s',t')}_{-}$'s for all pairs (s',t'), and do not appear in the polynomials $[\mathbf{WW}]^{(s',0)(L,t')}_{-}$'s for all pairs $(s',t') \neq (s,t)$.

So from Equation (38), it implies that

$$0 = \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi_{(s,t):pq} \cdot [\mathbf{W}]_{pq}^{(s,t)}, \tag{42}$$

for all $(s,t) \neq (L,0)$, and

$$0 = \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi_{(s,0)(L,t):pq} \cdot [\mathbf{WW}]_{pq}^{(s,0)(L,t)}, \tag{43}$$

for all (s,t) that $s \neq t$. The rest of the terms in Equation (38) is

$$0 = \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L,0):pq} \cdot [\mathbf{W}]_{pq}^{(L,0)} + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):pq} \cdot [\mathbf{WW}]_{pq}^{(s,0)(L,s)}. \tag{44}$$

- Equation (39): Since the set of all $[\mathbf{b}]_{-}^{(-)}$'s, which is the set of all monomials $\mathbf{y}_{-}^{(-)}$'s, is a linear independent set in R, it implies that

$$0 = \Phi_{(s):p}, \tag{45}$$

for all $L \geqslant s > 0$ and $1 \leqslant p \leqslant n_s$.

- Equation (40): Observe that
  - If $L > s > t > 0$, then the monomials $\left(\mathbf{x}_{-}^{(s)} \cdot \ldots \cdot \mathbf{x}_{-}^{(t+1)}\right) \cdot \mathbf{y}_{-}^{(t)}$'s only appear in the polynomials $[\mathbf{Wb}]_{-}^{(s,t)(t)}$'s. They do not appear in the polynomials $[\mathbf{Wb}]_{-}^{(s',t')(t')}$'s for all $L \geqslant s' > t' > 0$ that $(s',t') \neq (s,t)$, and do not appear in polynomials $[\mathbf{bW}]_{-}^{(s')(L,t')}$'s for all $L \geqslant s' > 0$ and $L > t' \geqslant 0$.
  - If $L \geqslant s > 0$, $L > t \geqslant 0$, and $s \neq t$, then the monomials $\mathbf{y}_{-}^{(s)} \cdot \left(\mathbf{x}_{-}^{(L)} \cdot \ldots \cdot \mathbf{x}_{-}^{(t+1)}\right)$'s only appear in the polynomials $[\mathbf{bW}]_{-}^{(s)(L,t)}$'s. They do not appear in the polynomials $[\mathbf{Wb}]_{-}^{(L,t')(t')}$'s for all $L > t' > 0$, and do not appear in the polynomials $[\mathbf{bW}]_{-}^{(s')(L,t')}$'s for all pairs $(s',t') \neq (s,t)$.
  - If $L > t > 0$, then the monomials $\mathbf{y}_{-}^{(t)} \cdot \left(\mathbf{x}_{-}^{(L)} \cdot \ldots \cdot \mathbf{x}_{-}^{(t+1)}\right)$'s only appear in the polynomials $[\mathbf{Wb}]_{-}^{(L,t)(t)}$'s and appear in the polynomials $[\mathbf{bW}]_{-}^{(t)(L,t)}$'s. They do not appear in the polynomials $[\mathbf{Wb}]_{-}^{(L,t')(t')}$'s for all $L > t' > 0$ that $t' \neq t$, and do not appear in polynomials $[\mathbf{bW}]_{-}^{(s')(L,t')}$'s for all pair (s',t') that $t' \neq t$.

  So from Equation (40), it implies that

$$0 = \sum_{p=1}^{n_s} \Phi_{(s,t)(t):p} \cdot [\mathbf{Wb}]_{p}^{(s,t)(t)}, \tag{46}$$

  for all $L > s > t > 0$, and

$$0 = \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi_{(s)(L,t):pq} \cdot [\mathbf{bW}]_{pq}^{(s)(L,t)} \tag{47}$$

  for all (s,t) that $s \neq t$, and

$$0 = \sum_{p=1}^{n_L} \Phi_{(L,t)(t):p} \cdot [\mathbf{Wb}]_{p}^{(L,t)(t)} + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi_{(t)(L,t):pq} \cdot [\mathbf{bW}]_{pq}^{(t)(L,t)} \tag{48}$$

  for all $L > t > 0$.

- Equation (41): Clearly, it implies that

$$0 = \Phi_1. \tag{49}$$

There are 8 equations, Equations (42)-(49), that are derived. In Equations (45) and (49), the corresponding Φ_{-}'s are directly characterized, and in Equations (42), (43), (44), (46), (47), (48), the corresponding Φ_{-}'s are not. We will characterize the Φ_{-}'s and Ψ_{-}'s in Equations (42), (43), (46) and (47) below by Lemma B.2, Lemma B.3 and Lemma B.5, respectively. These lemma are stated as below (their proofs will be postponed to Section B.4).

Lemma B.2. *For a pair (s,t) such that $L \geqslant s > t \geqslant 0$ and $(s,t) \neq (L,0)$, if*

$$0 = \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,t):pq} \cdot [\mathbf{W}]_{pq}^{(s,t)}, \tag{50}$$

then $\Phi_{(s,t):pq} = 0$ for all p, q.

Lemma B.3. *For a pair (s,t) such that $L \geqslant s > 0$, $L > t \geqslant 0$, if*

$$0 = \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,0)(L,t):pq} \cdot [\mathbf{W}\mathbf{W}]_{pq}^{(s,0)(L,t)}, \tag{51}$$

then $\Phi_{(s,0)(L,t):pq} = 0$ for all p, q or $\Psi^{(s,0)(L,t)} = 0$.

Lemma B.4. *For a pair (s,t) such that $L \geqslant s > t > 0$, if*

$$0 = \sum_{p=1}^{n_s} \Phi_{(s,t)(t):p} \cdot [\mathbf{W}\mathbf{b}]_{p}^{(s,t)(t)}, \tag{52}$$

then $\Phi_{(s,t)(t):p} = 0$ for all p.

Lemma B.5. *For a pair (s,t) such that $L \geqslant s > 0$, $L > t \geqslant 0$, if*

$$0 = \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s)(L,t):pq} \cdot [\mathbf{b}\mathbf{W}]_{pq}^{(s)(L,t)}, \tag{53}$$

then $\Phi_{(s)(L,t):pq} = 0$ for all p, q or $\Psi^{(s)(L,t)} = 0$.

Remark. The reason that we skip the characterizations of Φ_{-}'s and Ψ_{-}'s in Equations (44) and (48) is they can be concretely characterized. For instance, consider the case when $n_L = \ldots = n_2 = n_1 = n_0 = 1$. From Equation (44), we have

$$\begin{aligned}
0 &= \sum_{p=1}^{1} \sum_{q=1}^{1} \Phi_{(L,0):pq} \cdot [\mathbf{W}]_{pq}^{(L,0)} + \sum_{L>s>0} \sum_{p=1}^{1} \sum_{q=1}^{1} \Phi_{(s,0)(L,s):pq} \cdot [\mathbf{W}\mathbf{W}]_{pq}^{(s,0)(L,s)} \\
&= \Phi_{(L,0):1,1} \cdot [\mathbf{W}]_{1,1}^{(L,0)} + \sum_{L>s>0} \Phi_{(s,0)(L,s):1,1} \cdot [\mathbf{W}\mathbf{W}]_{1,1}^{(s,0)(L,s)} \\
&= \Phi_{(L,0):1,1} \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(2)} \cdot \mathbf{x}_{1,1}^{(1)} \right) \\
&\quad + \sum_{L>s>0} \Phi_{(s,0)(L,s):1,1} \cdot \left(\mathbf{x}_{1,1}^{(s)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(2)} \cdot \mathbf{x}_{1,1}^{(1)} \right) \cdot \Psi^{(s,0)(L,s)} \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(s+2)} \cdot \mathbf{x}_{1,1}^{(s+1)} \right) \\
&= \Phi_{(L,0):1,1} \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(2)} \cdot \mathbf{x}_{1,1}^{(1)} \right) \\
&\quad + \sum_{L>s>0} \Phi_{(s,0)(L,s):1,1} \cdot \Psi^{(s,0)(L,s)} \cdot \left(\mathbf{x}_{1,1}^{(s)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(2)} \cdot \mathbf{x}_{1,1}^{(1)} \right) \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(s+2)} \cdot \mathbf{x}_{1,1}^{(s+1)} \right) \\
&= \Phi_{(L,0):1,1} \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(2)} \cdot \mathbf{x}_{1,1}^{(1)} \right) \\
&\quad + \sum_{L>s>0} \Phi_{(s,0)(L,s):1,1} \cdot \Psi^{(s,0)(L,s)} \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(2)} \cdot \mathbf{x}_{1,1}^{(1)} \right) \\
&= \left(\Phi_{(L,0):1,1} + \sum_{L>s>0} \Phi_{(s,0)(L,s):1,1} \cdot \Psi^{(s,0)(L,s)} \right) \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(2)} \cdot \mathbf{x}_{1,1}^{(1)} \right).
\end{aligned}$$

It implies that

$$0 = \Phi_{(L,0):1,1} + \sum_{L>s>0} \Phi_{(s,0)(L,s):1,1} \cdot \Psi^{(s,0)(L,s)}. \tag{54}$$

From Equation (54), we can not derive a more concrete relation on the Φ_-'s and Ψ_-'s. Similarly, from Equation (48), we have

$$\begin{aligned}
0 &= \sum_{p=1}^{1} \Phi_{(L,t)(t):p} \cdot [\mathbf{Wb}]_p^{(L,t)(t)} + \sum_{p=1}^{1}\sum_{q=1}^{1} \Phi_{(t)(L,t):pq} \cdot [\mathbf{bW}]_{pq}^{(t)(L,t)} \\
&= \Phi_{(L,t)(t):1} \cdot [\mathbf{Wb}]_1^{(L,t)(t)} + \Phi_{(t)(L,t):1,1} \cdot [\mathbf{bW}]_{1,1}^{(t)(L,t)} \\
&= \Phi_{(L,t)(t):1} \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(t+2)} \cdot \mathbf{x}_{1,1}^{(t+1)}\right) \cdot \left(\mathbf{y}_{1,1}^{(t+1)}\right) \\
&\qquad + \Phi_{(t)(L,t):1,1} \cdot \left(\mathbf{y}_{1,1}^{(t+1)}\right) \cdot \Psi^{(t)(L,t)} \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(t+2)} \cdot \mathbf{x}_{1,1}^{(t+1)}\right) \\
&= \Phi_{(L,t)(t):1} \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(t+2)} \cdot \mathbf{x}_{1,1}^{(t+1)}\right) \cdot \left(\mathbf{y}_{1,1}^{(t+1)}\right) \\
&\qquad + \Phi_{(t)(L,t):1,1} \cdot \Psi^{(t)(L,t)} \cdot \left(\mathbf{y}_{1,1}^{(t+1)}\right) \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(t+2)} \cdot \mathbf{x}_{1,1}^{(t+1)}\right) \\
&= \Phi_{(L,t)(t):1} \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(t+2)} \cdot \mathbf{x}_{1,1}^{(t+1)}\right) \cdot \left(\mathbf{y}_{1,1}^{(t+1)}\right) \\
&\qquad + \Phi_{(t)(L,t):1,1} \cdot \Psi^{(t)(L,t)} \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(t+2)} \cdot \mathbf{x}_{1,1}^{(t+1)}\right) \cdot \left(\mathbf{y}_{1,1}^{(t+1)}\right) \\
&= \left(\Phi_{(L,t)(t):1} + \Phi_{(t)(L,t):1,1} \cdot \Psi^{(t)(L,t)}\right) \cdot \left(\mathbf{x}_{1,1}^{(L)} \cdot \ldots \cdot \mathbf{x}_{1,1}^{(t+2)} \cdot \mathbf{x}_{1,1}^{(t+1)}\right) \cdot \left(\mathbf{y}_{1,1}^{(t+1)}\right)
\end{aligned}$$

It implies that

$$0 = \Phi_{(L,t)(t):1} + \Phi_{(t)(L,t):1,1} \cdot \Psi^{(t)(L,t)}. \tag{55}$$

From Equation (55), we can not derive a more concrete relation on the Φ_-'s and Ψ_-'s.

Combining the discussions above, we obtain the following necessary for the coefficients Φ and Ψ such that $\alpha(\Phi, \Psi) = 0$.

Theorem B.6. *Let $\alpha(\Phi, \Psi)$ be a polynomial given in equation 37. If $\alpha(\Phi, \Psi) = 0$, then the following condition holds:*

1. *For all $L \geqslant s > t \geqslant 0$ with $(s,t) \neq (L,0)$, and for all p, q, we have*
$$\Phi_{(s,t):pq} = 0. \tag{56}$$

2. *For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have*
$$\Phi_{(s,0)(L,t):pq} = 0, \tag{57}$$
for all p, q, or
$$\Psi^{(s,0)(L,t)} = 0. \tag{58}$$

3. *We have*
$$0 = \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L,0):pq} \cdot [\mathbf{W}]_{pq}^{(L,0)} + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):pq} \cdot [\mathbf{WW}]_{pq}^{(s,0)(L,s)}. \tag{59}$$

4. *For all $L > s > t > 0$, and for all p, we have*
$$\Phi_{(s,t)(t):p} = 0. \tag{60}$$

5. *For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have*
$$\Phi_{(s)(L,t):pq} = 0, \tag{61}$$
for all p, q, or
$$\Psi^{(s)(L,t)} = 0. \tag{62}$$

6. *For all $L > t > 0$, we have*

$$0 = \sum_{p=1}^{n_L} \Phi_{(L,t)(t):p} \cdot [\mathbf{Wb}]_p^{(L,t)(t)} + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi_{(t)(L,t):pq} \cdot [\mathbf{bW}]_{pq}^{(t)(L,t)}. \tag{63}$$

7. *For all $L \geqslant s > 0$ and for all p, we have*

$$\Phi_{(s):p} = 0. \tag{64}$$

8. *We have*

$$\Phi_1 = 0. \tag{65}$$

Proof. By the previous observations, $\alpha(\Phi, \Psi) = 0$ implies four Equations (38)-(41). These equations imply Equations (42)-(49). The proof of all parts in Theorem B.6 are as follows.

1. It comes from Equation (42) and Lemma B.2.
2. It comes from Equation (43) and Lemma B.3.
3. It comes from Equation (44).
4. It comes from Equation (46) and Lemma B.4
5. It comes from Equation (47) and Lemma B.5.
6. It comes from Equation (48).
7. It comes from Equation (45).
8. It comes from Equation (49).

The proof is finsihed. □

The following corollary is a direct consequence of Theorem B.6.

Corollary B.7. *If for some Φ, Φ', and Ψ, we have $\alpha(\Phi, \Psi) = \alpha(\Phi', \Psi)$, then:*

1. *For all $L \geqslant s > t \geqslant 0$ with $(s,t) \neq (L,0)$, and for all p, q, we have*

$$\Phi_{(s,t):pq} = \Phi'_{(s,t):pq}. \tag{66}$$

2. *For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have*

$$\Phi_{(s,0)(L,t):pq} = \Phi'_{(s,0)(L,t):pq}, \tag{67}$$

   for all p, q, or

$$\Psi^{(s,0)(L,t)} = 0. \tag{68}$$

3. *We have*

$$\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L,0):pq} \cdot [\mathbf{W}]_{pq}^{(L,0)} + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):pq} \cdot [\mathbf{WW}]_{pq}^{(s,0)(L,s)} \tag{69}$$

$$= \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi'_{(L,0):pq} \cdot [\mathbf{W}]_{pq}^{(L,0)} + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi'_{(s,0)(L,s):pq} \cdot [\mathbf{WW}]_{pq}^{(s,0)(L,s)}. \tag{70}$$

4. *For all $L > s > t > 0$, and for all p, we have*

$$\Phi_{(s,t)(t):p} = \Phi'_{(s,t)(t):p}. \tag{71}$$

5. *For all* $L \geqslant s > 0$*,* $L > t \geqslant 0$ *with* $s \neq t$*, we have*

$$\Phi_{(s)(L,t):pq} = \Phi'_{(s)(L,t):pq}, \tag{72}$$

for all p, q*, or*

$$\Psi^{(s)(L,t)} = 0. \tag{73}$$

6. *For all* $L > t > 0$*, we have*

$$\sum_{p=1}^{n_L} \Phi_{(L,t)(t):p} \cdot [\mathbf{Wb}]_p^{(L,t)(t)} + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi_{(t)(L,t):pq} \cdot [\mathbf{bW}]_{pq}^{(t)(L,t)} \tag{74}$$

$$= \sum_{p=1}^{n_L} \Phi'_{(L,t)(t):p} \cdot [\mathbf{Wb}]_p^{(L,t)(t)} + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi'_{(t)(L,t):pq} \cdot [\mathbf{bW}]_{pq}^{(t)(L,t)}. \tag{75}$$

7. *For all* $L \geqslant s > 0$ *and for all* p*, we have*

$$\Phi_{(s):p} = \Phi'_{(s):p}. \tag{76}$$

8. *We have*

$$\Phi_1 = \Phi'_1. \tag{77}$$

Proof. Since

$$0 = \alpha(\Phi, \Psi) - \alpha(\Phi', \Psi) = \alpha(\Phi - \Phi', \Psi),$$

so the results come directly from Theorem B.6. □

B.4 Proofs of Lemmas B.2-B.5

The proofs of Lemmas B.2 through B.5 are directly from the coefficient comparison of two equal polynomials and the following lemma. We omit the proofs of Lemmas B.2 through B.5 and show only the proof for Lemma B.8.

Lemma B.8. *For a feasible tuple* (s, t, p, q)*, if*

$$[\mathbf{WW}]_{pq}^{(s,0)(L,t)} = \left([\mathbf{W}]^{(s,0)} \cdot \Psi^{(s,0)(L,t)} \cdot [\mathbf{W}]^{(L,t)}\right)_{pq} = 0 \in R. \tag{78}$$

Then $\Psi^{(s,0)(L,t)} = 0 \in \mathbb{R}^{n_0 \times n_L}$*.*

Proof. We first consider the case where $s = L$ and $t = 0$, then the case $s \leqslant t$, and finally the case $s > t$. Note that, the proof for the last case will be by combining the arguments of the first two cases.

Case $s = L$ and $t = 0$. From Equation (78), we have

$$[\mathbf{WW}]_{pq}^{(L,0)(L,0)} = \left([\mathbf{W}]^{(L,0)} \cdot \Psi^{(L,0)(L,0)} \cdot [\mathbf{W}]^{(L,0)}\right)_{pq} = 0, \tag{79}$$

which means

$$\left([\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(1)} \cdot \Psi^{(L,0)(L,0)} \cdot [\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(1)}\right)_{pq} = 0. \tag{80}$$

Here, we consider three cases, where $L = 1$, $L = 2$ and $L \geqslant 3$.

- Case $L = 1$. Equation (80) becomes

$$\left([\mathbf{W}]^{(1)} \cdot \Psi^{(1,0)(1,0)} \cdot [\mathbf{W}]^{(1)}\right)_{pq} = 0. \tag{81}$$

By Proposition A.1, this is equivalent to

$$[\mathbf{W}]^{(1)}_{p,*} \cdot \Psi^{(1,0)(1,0)} \cdot [\mathbf{W}]^{(1)}_{*,q} = 0, \tag{82}$$

which is:

$$\sum_{i=1}^{n_0} \sum_{j=0}^{n_1} \Psi^{(1,0)(1,0)}_{ij} \cdot \mathbf{x}^{(1)}_{p,i} \cdot \mathbf{x}^{(1)}_{j,q} = 0. \tag{83}$$

Since the LHS of Equation (83) is a linear combination between distinct monomials

$$\mathbf{x}^{(1)}_{p,i} \cdot \mathbf{x}^{(1)}_{j,q}, \tag{84}$$

for $1 \leqslant i \leqslant n_0$ and $1 \leqslant j \leqslant n_1$, so it implies that

$$\Psi^{(1,0)(1,0)}_{ij} = 0, \tag{85}$$

for all $1 \leqslant i \leqslant n_0$ and $1 \leqslant j \leqslant n_1$, which means

$$\Psi^{(1,0)(1,0)} = 0. \tag{86}$$

- Case $L = 2$. Equation (80) becomes

$$\left([\mathbf{W}]^{(2)} \cdot [\mathbf{W}]^{(1)} \cdot \Psi^{(2,0)(2,0)} \cdot [\mathbf{W}]^{(2)} \cdot [\mathbf{W}]^{(1)}\right)_{pq} = 0. \tag{87}$$

By Proposition A.1, this is equivalent to

$$[\mathbf{W}]^{(2)}_{p,*} \cdot [\mathbf{W}]^{(1)} \cdot \Psi^{(2,0)(2,0)} \cdot [\mathbf{W}]^{(2)} \cdot [\mathbf{W}]^{(1)}_{*,q} = 0, \tag{88}$$

which is

$$\left([\mathbf{W}]^{(2)}_{p,*} \cdot [\mathbf{W}]^{(1)}_{*,1}\,,\ [\mathbf{W}]^{(2)}_{p,*} \cdot [\mathbf{W}]^{(1)}_{*,2}\,,\ \ldots\,,\ [\mathbf{W}]^{(2)}_{p,*} \cdot [\mathbf{W}]^{(1)}_{*,n_0}\right) \cdot \Psi^{(2,0)(2,0)} \tag{89}$$

$$\cdot \left([\mathbf{W}]^{(2)}_{1,*} \cdot [\mathbf{W}]^{(1)}_{*,q}\,,\ [\mathbf{W}]^{(2)}_{2,*} \cdot [\mathbf{W}]^{(1)}_{*,q}\,,\ \ldots\,,\ [\mathbf{W}]^{(2)}_{n_2,*} \cdot [\mathbf{W}]^{(1)}_{*,q}\right)^\top = 0, \tag{90}$$

which is

$$\sum_{i=1}^{n_0} \sum_{j=0}^{n_2} \Psi^{(2,0)(2,0)}_{ij} \cdot \left([\mathbf{W}]^{(2)}_{p,*} \cdot [\mathbf{W}]^{(1)}_{*,i}\right) \cdot \left([\mathbf{W}]^{(2)}_{j,*} \cdot [\mathbf{W}]^{(1)}_{*,q}\right) = 0. \tag{91}$$

Since the LHS of Equation (91) is a linear combination between elements of a linear independent collection of R, which are:

$$\left([\mathbf{W}]^{(2)}_{p,*} \cdot [\mathbf{W}]^{(1)}_{*,i}\right) \cdot \left([\mathbf{W}]^{(2)}_{j,*} \cdot [\mathbf{W}]^{(1)}_{*,q}\right), \tag{92}$$

for $1 \leqslant i \leqslant n_0$ and $1 \leqslant j \leqslant n_2$, so it implies that

$$\Psi^{(2,0)(2,0)}_{ij} = 0, \tag{93}$$

for all $1 \leqslant i \leqslant n_0$ and $1 \leqslant j \leqslant n_2$, which means

$$\Psi^{(2,0)(2,0)} = 0. \tag{94}$$

- Case $L \geqslant 3$. Equation (80) becomes

$$\left([\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(1)} \cdot \Psi^{(L,0)(L,0)} \cdot [\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(1)}\right)_{pq} = 0. \tag{95}$$

It is noted that the matrices $[\mathbf{W}]^{(L-1)}, \ldots, [\mathbf{W}]^{(2)}$ can be substituted for some matrix units at 1^{st}-row and 1^{st}-column such that the product

$$[\mathbf{W}]^{(L-1)} \cdot \ldots \cdot [\mathbf{W}]^{(2)} \tag{96}$$

becomes a matrix unit at 1st-row and 1st-column. Substitute this in Equation (95), we have

$$\left([\mathbf{W}]^{(L)} \cdot E_{11} \cdot [\mathbf{W}]^{(1)} \cdot \Psi^{(L,0)(L,0)} \cdot [\mathbf{W}]^{(L)} \cdot E_{11} \cdot [\mathbf{W}]^{(1)}\right)_{pq} = 0. \tag{97}$$

By Proposition A.1, this is equivalent to

$$[\mathbf{W}]^{(L)}_{p,*} \cdot E_{11} \cdot [\mathbf{W}]^{(1)} \cdot \Psi^{(L,0)(L,0)} \cdot [\mathbf{W}]^{(L)} \cdot E_{11} \cdot [\mathbf{W}]^{(1)}_{*,q} = 0. \tag{98}$$

which can be rewritten as

$$[\mathbf{W}]^{(L)}_{p,1} \cdot [\mathbf{W}]^{(1)}_{1,*} \cdot \Psi^{(L,0)(L,0)} \cdot [\mathbf{W}]^{(L)}_{*,1} \cdot [\mathbf{W}]^{(1)}_{1,q} = 0, \tag{99}$$

or

$$\mathbf{x}^{(L)}_{p,1} \cdot [\mathbf{W}]^{(1)}_{1,*} \cdot \Psi^{(L,0)(L,0)} \cdot [\mathbf{W}]^{(L)}_{*,1} \cdot \mathbf{x}^{(1)}_{1,q} = 0. \tag{100}$$

Since the LHS of Equation (100) is a polynomial in R, we have

$$[\mathbf{W}]^{(1)}_{1,*} \cdot \Psi^{(L,0)(L,0)} \cdot [\mathbf{W}]^{(L)}_{*,1} = 0, \tag{101}$$

which is

$$\sum_{i=1}^{n_0} \sum_{j=0}^{n_L} \Psi^{(L,0)(L,0)}_{ij} \cdot \mathbf{x}^{(1)}_{1,i} \cdot \mathbf{x}^{(L)}_{j,1} = 0. \tag{102}$$

Since the LHS of Equation (102) is a linear combination between distinct monomials

$$\mathbf{x}^{(1)}_{1,i} \cdot \mathbf{x}^{(L)}_{j,1}, \tag{103}$$

for $1 \leqslant i \leqslant n_0$ and $1 \leqslant j \leqslant n_L$, so it implies that

$$\Psi^{(L,0)(L,0)}_{ij} = 0, \tag{104}$$

for all $1 \leqslant i \leqslant n_0$ and $1 \leqslant j \leqslant n_L$, which means

$$\Psi^{(L,0)(L,0)} = 0. \tag{105}$$

In conclusion, for all cases of L, we have

$$\Psi^{(L,0)(L,0)} = 0. \tag{106}$$

We finish the proof of the case where $s = L$ and $t = 0$.

Case $s \leqslant t$**.** From Equation (78), we have

$$[\mathbf{WW}]^{(s,0)(L,t)}_{pq} = \left([\mathbf{W}]^{(s,0)} \cdot \Psi^{(s,0)(L,t)} \cdot [\mathbf{W}]^{(L,t)}\right)_{pq} = 0, \tag{107}$$

which means

$$\left([\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(1)} \cdot \Psi^{(s,0)(L,t)} \cdot [\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)}\right)_{pq} = 0. \tag{108}$$

Since $s \leqslant t$, the $[\mathbf{W}]^{-}$'s that are in the product $[\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(1)}$, and the $[\mathbf{W}]^{-}$'s that are in the product $[\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)}$, are distinct. By directly multiplying $[\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(1)}$ and $[\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)}$, we can write these two products in the forms

$$\begin{aligned} [\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(1)} &= \left(\mathbf{f}^{(s)}_{ij}\right)_{1 \leqslant i \leqslant n_s, 1 \leqslant j \leqslant n_0} \in R^{n_s \times n_0}, \\ [\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)} &= \left(\mathbf{g}^{(t)}_{ij}\right)_{1 \leqslant i \leqslant n_L, 1 \leqslant j \leqslant n_t} \in R^{n_L \times n_t}, \end{aligned} \tag{109}$$

where all $\mathbf{f}^{(s)}_{-}$'s are nonzero and all $\mathbf{g}^{(t)}_{-}$'s are nonzero. Moreover, $\mathbf{f}^{(s)}_{-}$'s are real polynomials of indeterminates $\mathbf{x}^{(1)}_{-}$'s, $\mathbf{x}^{(2)}_{-}$'s, …, $\mathbf{x}^{(s)}_{-}$'s. Similarly, $\mathbf{g}^{(t)}_{-}$'s are real polynomials of indeterminates $\mathbf{x}^{(L)}_{-}$'s, $\mathbf{x}^{(L-1)}_{-}$'s, …, $\mathbf{x}^{(t+1)}_{-}$'s. Now, Equation (108) is equal to

$$\left(\left(\mathbf{f}^{(s)}_{ij}\right)_{1 \leqslant i \leqslant n_s, 1 \leqslant j \leqslant n_0} \cdot \Psi^{(s,0)(L,t)} \cdot \left(\mathbf{g}^{(t)}_{ij}\right)_{1 \leqslant i \leqslant n_L, 1 \leqslant j \leqslant n_t}\right)_{pq} = 0. \tag{110}$$

By Proposition A.1, this is equivalent to

$$\left(\mathbf{f}_{p,1}^{(s)}\,,\,\mathbf{f}_{p,2}^{(s)}\,,\,\ldots\,,\,\mathbf{f}_{p,n_0}^{(s)}\right)\cdot\Psi^{(s,0)(L,t)}\cdot\left(\mathbf{g}_{1,q}^{(t)}\,,\,\mathbf{g}_{2,q}^{(t)}\,,\,\ldots\,,\,\mathbf{f}_{n_L,q}^{(t)}\right)^{\top}=0, \tag{111}$$

which is

$$\sum_{i=1}^{n_0}\sum_{j=0}^{n_L}\Psi_{ij}^{(s,0)(L,t)}\cdot\mathbf{f}_{p,i}^{(s)}\cdot\mathbf{g}_{j,q}^{(t)}=0. \tag{112}$$

Since the LHS of Equation (112) is a linear combination between elements of a linear independent collection of R, which are

$$\mathbf{f}_{p,i}^{(s)}\cdot\mathbf{g}_{j,q}^{(t)}, \tag{113}$$

for $1\leqslant i\leqslant n_0$ and $1\leqslant j\leqslant n_L$. The linear dependency comes from the distinction between indeterminates of $\mathbf{f}_{-}^{(-)}$ and $\mathbf{g}_{-}^{(-)}$. It implies that

$$\Psi_{ij}^{(s,0)(L,t)}=0, \tag{114}$$

for all $1\leqslant i\leqslant n_0$ and $1\leqslant j\leqslant n_L$, which means

$$\Psi^{(s,0)(L,t)}=0. \tag{115}$$

We finish the proof of the case where $s\leqslant t$.

Case $s>t$. From Equation (78), we have:

$$[\mathbf{WW}]_{pq}^{(s,0)(L,t)}=\left([\mathbf{W}]^{(s,0)}\cdot\Psi^{(s,0)(L,t)}\cdot[\mathbf{W}]^{(L,t)}\right)_{pq}=0, \tag{116}$$

which means

$$\left([\mathbf{W}]^{(s)}\cdot\ldots\cdot[\mathbf{W}]^{(1)}\cdot\Psi^{(s,0)(L,t)}\cdot[\mathbf{W}]^{(L)}\cdot\ldots\cdot[\mathbf{W}]^{(t+1)}\right)_{pq}=0. \tag{117}$$

Assume that $\Psi^{(s,0)(L,t)}\neq 0\in\mathbb{R}^{n_0\times n_L}$. Observe that

$$[\mathbf{W}]^{(s)}\cdot\ldots\cdot[\mathbf{W}]^{(1)}\cdot\Psi^{(s,0)(L,t)}\cdot[\mathbf{W}]^{(L)}\cdot\ldots\cdot[\mathbf{W}]^{(t+1)} \tag{118}$$

$$=\left([\mathbf{W}]^{(s)}\cdot\ldots\cdot[\mathbf{W}]^{(t+1)}\cdot[\mathbf{W}]^{(t)}\cdot\ldots\cdot[\mathbf{W}]^{(1)}\right)\cdot\Psi^{(s,0)(L,t)}\cdot\left([\mathbf{W}]^{(L)}\cdot\ldots\cdot[\mathbf{W}]^{(s+1)}\cdot[\mathbf{W}]^{(s)}\cdot\ldots\cdot[\mathbf{W}]^{(t+1)}\right) \tag{119}$$

$$=\left([\mathbf{W}]^{(s)}\cdot\ldots\cdot[\mathbf{W}]^{(t+1)}\right)\cdot\left([\mathbf{W}]^{(t)}\cdot\ldots\cdot[\mathbf{W}]^{(1)}\cdot\Psi^{(s,0)(L,t)}\cdot[\mathbf{W}]^{(L)}\cdot\ldots\cdot[\mathbf{W}]^{(s+1)}\right)\cdot\left([\mathbf{W}]^{(s)}\cdot\ldots\cdot[\mathbf{W}]^{(t+1)}\right). \tag{120}$$

In the second term of Equation (120),

$$[\mathbf{W}]^{(t)}\cdot\ldots\cdot[\mathbf{W}]^{(1)}\cdot\Psi^{(s,0)(L,t)}\cdot[\mathbf{W}]^{(L)}\cdot\ldots\cdot[\mathbf{W}]^{(s+1)}. \tag{121}$$

Since $s>t$, all $[\mathbf{W}]^{(-)}$'s in Equation (121) are distinct. And since $\Psi^{(s,0)(L,t)}\neq 0\in\mathbb{R}^{n_0\times n_L}$, with the same argument as in **Case** $s\leqslant t$, we have

$$[\mathbf{W}]^{(t)}\cdot\ldots\cdot[\mathbf{W}]^{(1)}\cdot\Psi^{(s,0)(L,t)}\cdot[\mathbf{W}]^{(L)}\cdot\ldots\cdot[\mathbf{W}]^{(s+1)}\neq 0\in R^{n_t\times n_s}. \tag{122}$$

Moreover, it can be written in the form

$$[\mathbf{W}]^{(t)}\cdot\ldots\cdot[\mathbf{W}]^{(1)}\cdot\Psi^{(s,0)(L,t)}\cdot[\mathbf{W}]^{(L)}\cdot\ldots\cdot[\mathbf{W}]^{(s+1)} \tag{123}$$

$$=\left(\mathbf{c}_{ij}^{(st)}\right)_{1\leqslant i\leqslant n_t,1\leqslant j\leqslant n_s}\in R^{n_t\times n_s}, \tag{124}$$

where all $\mathbf{c}_{\_}^{(st)}$'s are real polynomials of indeterminates $\mathbf{x}_{\_}^{(1)}$'s, $\mathbf{x}_{\_}^{(2)}$'s, ..., $\mathbf{x}_{\_}^{(t)}$'s and $\mathbf{x}_{\_}^{(L)}$'s, $\mathbf{x}_{\_}^{(L-1)}$'s, ..., $\mathbf{x}_{\_}^{(s+1)}$'s, and at least one of them is a nonzero element in R. By Corollary A.7, indeterminates $\mathbf{x}_{\_}^{(1)}$'s, $\mathbf{x}_{\_}^{(2)}$'s, ..., $\mathbf{x}_{\_}^{(t)}$'s and $\mathbf{x}_{\_}^{(L)}$'s, $\mathbf{x}_{\_}^{(L-1)}$'s, ..., $\mathbf{x}_{\_}^{(s+1)}$'s can be substituted for some real values to make

$$[\mathbf{W}]^{(t)} \cdot \ldots \cdot [\mathbf{W}]^{(1)} \cdot \Psi^{(s,0)(L,t)} \cdot [\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(s+1)}, \tag{125}$$

become a nonzero matrix of $\mathbb{R}^{n_t \times n_s}$. We denote this nonzero matrix by $\overline{\Psi}^{(s,0)(L,t)} \in \mathbb{R}^{n_t \times n_s}$. Note that, since $s > t$, the substitution only applies for indeterminates in $[\mathbf{W}]^{(1)}, [\mathbf{W}]^{(2)}, \ldots, [\mathbf{W}]^{(t)}$ and $[\mathbf{W}]^{(L)}, [\mathbf{W}]^{(L-1)}, \ldots, [\mathbf{W}]^{(s+1)}$. In other words, it does not apply for $[\mathbf{W}]^{(s)}, \ldots, [\mathbf{W}]^{(t+1)}$. So, with the above substitution, Equation (118) becomes

$$\begin{aligned}
&[\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(1)} \cdot \Psi^{(s,0)(L,t)} \cdot [\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)} \\
&= [\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)} \cdot \overline{\Psi}^{(s,0)(L,t)} \cdot [\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)}.
\end{aligned} \tag{126}$$

Note that, since $s > t$, there is at least one $[\mathbf{W}]^{(-)}$ in the product $[\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)}$. Combine with $\overline{\Psi}^{(s,0)(L,t)} \neq 0$, by applying the argument of **Case** $s = L$ **and** $t = 0$, we have

$$\begin{aligned}
&\left([\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(1)} \cdot \Psi^{(s,0)(L,t)} \cdot [\mathbf{W}]^{(L)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)}\right)_{pq} \\
&= \left([\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)} \cdot \overline{\Psi}^{(s,0)(L,t)} \cdot [\mathbf{W}]^{(s)} \cdot \ldots \cdot [\mathbf{W}]^{(t+1)}\right)_{pq} \neq 0,
\end{aligned} \tag{127}$$

which contradicts to Equation (117). In conclusion

$$\Psi^{(s,0)(L,t)} = 0. \tag{128}$$

We finish the proof of the case where $s > t$.

In summary, we did consider all possible cases. The proof is finished. □

C Equivariant Polynomial Layers

We now proceed to construct a G-equivariant polynomial layer, denoted as E. These layers serve as the fundamental building blocks for our MAGEP-NFNs. Our strategy is as follows: we first express E as a polynomial layer that is a linear combination of stable polynomial terms (Subsection C.1). We then find the equivariant maps among these polynomial layers using the parameter sharing mechanism.

C.1 Equivariant layer as a linear combination of stable polynomial terms

For two weight spaces $\mathcal{U}$ and $\mathcal{U}'$ with the same number of layers L as well as the same number of channels at i^{th} layer n_i,

$$\begin{aligned}
\mathcal{U} &= \mathcal{W} \times \mathcal{B}, && \text{where:} \\
\mathcal{W} &= \mathbb{R}^{w_L \times n_L \times n_{L-1}} \times \ldots \times \mathbb{R}^{w_2 \times n_2 \times n_1} \times \mathbb{R}^{w_1 \times n_1 \times n_0}, \\
\mathcal{B} &= \mathbb{R}^{b_L \times n_L \times 1} \times \ldots \times \mathbb{R}^{b_2 \times n_2 \times 1} \times \mathbb{R}^{b_1 \times n_1 \times 1},
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{U}' &= \mathcal{W}' \times \mathcal{B}', && \text{where:} \\
\mathcal{W}' &= \mathbb{R}^{w'_L \times n_L \times n_{L-1}} \times \ldots \times \mathbb{R}^{w'_2 \times n_2 \times n_1} \times \mathbb{R}^{w'_1 \times n_1 \times n_0}, \\
\mathcal{B}' &= \mathbb{R}^{b'_L \times n_L \times 1} \times \ldots \times \mathbb{R}^{b'_2 \times n_2 \times 1} \times \mathbb{R}^{b'_1 \times n_1 \times 1}.
\end{aligned}$$

We want to build a map $E \colon \mathcal{U} \to \mathcal{U}'$ such that E is G-equivariant, where:

$$G = \{\mathrm{id}_{\mathcal{G}_{n_L}}\} \times \mathcal{G}_{n_{L-1}}^{>0} \times \ldots \times \mathcal{G}_{n_1}^{>0} \times \{\mathrm{id}_{\mathcal{G}_{n_0}}\}. \tag{129}$$

Let us consider a polynomial map $E \colon \mathcal{U} \to \mathcal{U}'$ such that, for input $U = ([W], [b])$, each entry of the output $E(U) = ([E(W)], [E(b)])$ is a linear combinations of stable polynomial terms, i.e the entries of $[W]^{(s,t)}$, $[b]^{(s)}$, $[Wb]^{(s,t)(t)}$, $[bW]^{(s)(L,t)}$, $[WW]^{(s,0)(L,t)}$, together with a bias. In concrete:

$$\begin{aligned}
[E(W)]_{jk}^{(i)} \coloneqq & \sum_{L \geqslant s > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,t):pq}^{(i):jk} \cdot [W]_{pq}^{(s,t)} + \sum_{L \geqslant s > 0} \sum_{p=1}^{n_s} \Phi_{(s):p}^{(i):jk} \cdot [b]_p^{(s)} \\
& + \sum_{L \geqslant s > t > 0} \sum_{p=1}^{n_s} \Phi_{(s,t)(t):p}^{(i):jk} \cdot [Wb]_p^{(s,t)(t)} \\
& + \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s)(L,t):pq}^{(i):jk} \cdot [bW]_{pq}^{(s)(L,t)} \\
& + \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,0)(L,t):pq}^{(i):jk} \cdot [WW]_{pq}^{(s,0)(L,t)} + \Phi_1^{(i):jk},
\end{aligned}$$

$$\begin{aligned}
[E(b)]_{j}^{(i)} \coloneqq & \sum_{L \geqslant s > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,t):pq}^{(i):j} \cdot [W]_{pq}^{(s,t)} + \sum_{L \geqslant s > 0} \sum_{p=1}^{n_s} \Phi_{(s):p}^{(i):j} \cdot [b]_p^{(s)} \\
& + \sum_{L \geqslant s > t > 0} \sum_{p=1}^{n_s} \Phi_{(s,t)(t):p}^{(i):j} \cdot [Wb]_p^{(s,t)(t)} \\
& + \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s)(L,t):pq}^{(i):j} \cdot [bW]_{pq}^{(s)(L,t)} \\
& + \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,0)(L,t):pq}^{(i):j} \cdot [WW]_{pq}^{(s,0)(L,t)} + \Phi_1^{(i):j}.
\end{aligned}$$

All Φ's are in $\mathbb{R}^{d' \times d}$, except the biases Φ_1^{-}'s are in $\mathbb{R}^{d' \times 1}$. In summary, E is parameterized by Φ_{-}'s and Ψ_{-}'s.

In order to be G-equivariant, the polynomial map E must satisfy the condition $E(gU) = gE(U)$ for all $g \in \mathcal{G}_{\mathcal{U}}$ and $U \in \mathcal{U}$. In the following subsections, we derive the computations of $E(gU)$ and $gE(U)$, and compare them in order to obtain all possible G-equivariant polynomial maps among those considered.

C.2 Compute $E(gU)$

We have

$$\begin{aligned}
[E(gW)]_{jk}^{(i)} = & \sum_{L \geqslant s > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,t):pq}^{(i):jk} \cdot [gW]_{pq}^{(s,t)} \\
& + \sum_{L \geqslant s > 0} \sum_{p=1}^{n_s} \Phi_{(s):p}^{(i):jk} \cdot [gb]_p^{(s)} \\
& + \sum_{L \geqslant s > t > 0} \sum_{p=1}^{n_s} \Phi_{(s,t)(t):p}^{(i):jk} \cdot [gWgb]_p^{(s,t)(t)} \\
& + \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s)(L,t):pq}^{(i):jk} \cdot [gbgW]_{pq}^{(s)(L,t)} \\
& + \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,0)(L,t):pq}^{(i):jk} \cdot [gWgW]_{pq}^{(s,0)(L,t)} \\
& + \Phi_1^{(i):jk}
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{L\geqslant s>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi^{(i):jk}_{(s,t):pq} \cdot \frac{d^{(s)}_p}{d^{(t)}_q} \cdot [W]^{(s,t)}_{\pi_s^{-1}(p),\pi_t^{-1}(q)} \\
&\quad + \sum_{L\geqslant s>0} \sum_{p=1}^{n_s} \Phi^{(i):jk}_{(s):p} \cdot d^{(s)}_p \cdot [b]^{(s)}_{\pi_s^{-1}(p)} \\
&\quad + \sum_{L\geqslant s>t>0} \sum_{p=1}^{n_s} \Phi^{(i):jk}_{(s,t)(t):p} \cdot d^{(s)}_p \cdot [Wb]^{(s,t)(t)}_{\pi_s^{-1}(p)} \\
&\quad + \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi^{(i):jk}_{(s)(L,t):pq} \cdot \frac{d^{(s)}_p}{d^{(t)}_q} \cdot [bW]^{(s)(L,t)}_{\pi_s^{-1}(p),\pi_t^{-1}(q)} \\
&\quad + \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi^{(i):jk}_{(s,0)(L,t):pq} \cdot \frac{d^{(s)}_p}{d^{(t)}_q} \cdot [WW]^{(s,0)(L,t)}_{\pi_s^{-1}(p),\pi_t^{-1}(q)} \\
&\quad + \Phi^{(i):jk}_1 \\
&= \sum_{L\geqslant s>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi^{(i):jk}_{(s,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [W]^{(s,t)}_{pq} \\
&\quad + \sum_{L\geqslant s>0} \sum_{p=1}^{n_s} \Phi^{(i):jk}_{(s):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} \cdot [b]^{(s)}_{p} \\
&\quad + \sum_{L\geqslant s>t>0} \sum_{p=1}^{n_s} \Phi^{(i):jk}_{(s,t)(t):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} \cdot [Wb]^{(s,t)(t)}_{p} \\
&\quad + \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi^{(i):jk}_{(s)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(s)(L,t)}_{pq} \\
&\quad + \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi^{(i):jk}_{(s,0)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [WW]^{(s,0)(L,t)}_{pq} \\
&\quad + \Phi^{(i):jk}_1
\end{aligned}
$$

$$
\begin{aligned}
[E(gb)]^{(i)}_j &= \sum_{L\geqslant s>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi^{(i):j}_{(s,t):pq} \cdot [gW]^{(s,t)}_{pq} \\
&\quad + \sum_{L\geqslant s>0} \sum_{p=1}^{n_s} \Phi^{(i):j}_{(s):p} \cdot [gb]^{(s)}_{p} \\
&\quad + \sum_{L\geqslant >t>0} \sum_{p=1}^{n_s} \Phi^{(i):j}_{(s,t)(t):p} \cdot [gWgb]^{(s,t)(t)}_{p} \\
&\quad + \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi^{(i):j}_{(s)(L,t):pq} \cdot [gbgW]^{(s)(L,t)}_{pq} \\
&\quad + \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi^{(i):j}_{(s,0)(L,t):pq} \cdot [gWgW]^{(s,0)(L,t)}_{pq} \\
&\quad + \Phi^{(i):j}_1 \\
&= \sum_{L\geqslant s>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} \Phi^{(i):j}_{(s,t):pq} \cdot \frac{d^{(s)}_p}{d^{(t)}_q} \cdot [W]^{(s,t)}_{\pi_s^{-1}(p),\pi_t^{-1}(q)} \\
&\quad + \sum_{L\geqslant s>0} \sum_{p=1}^{n_s} \Phi^{(i):j}_{(s):p} \cdot d^{(s)}_p \cdot [b]^{(s)}_{\pi_s^{-1}(p)}
\end{aligned}
$$

$$
\begin{aligned}
&+ \sum_{L\geqslant s>t>0} \sum_{p=1}^{n_s} \Phi^{(i):j}_{(s,t)(t):p} \cdot d^{(s)}_p \cdot [Wb]^{(s,t)(t)}_{\pi_s^{-1}(p)} \\
&+ \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi^{(i):j}_{(s)(L,t):pq} \cdot \frac{d^{(s)}_p}{d^{(t)}_q} \cdot [bW]^{(s)(L,t)}_{\pi_s^{-1}(p),\pi_t^{-1}(q)} \\
&+ \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi^{(i):j}_{(s,0)(L,t):pq} \cdot \frac{d^{(s)}_p}{d^{(t)}_q} \cdot [WW]^{(s,0)(L,t)}_{\pi_s^{-1}(p),\pi_t^{-1}(q)} \\
&+ \Phi^{(i):j}_1 \\
=& \sum_{L\geqslant s>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi^{(i):j}_{(s,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [W]^{(s,t)}_{pq} \\
&+ \sum_{L\geqslant s>0} \sum_{p=1}^{n_s} \Phi^{(i):j}_{(s):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} \cdot [b]^{(s)}_p \\
&+ \sum_{L\geqslant s>t>0} \sum_{p=1}^{n_s} \Phi^{(i):j}_{(s,t)(t):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} \cdot [Wb]^{(s,t)(t)}_p \\
&+ \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi^{(i):j}_{(s)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(s)(L,t)}_{pq} \\
&+ \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi^{(i):j}_{(s,0)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [WW]^{(s,0)(L,t)}_{pq} \\
&+ \Phi^{(i):j}_1 .
\end{aligned}
$$

Note that, we can move around the π's in above equations since the group G satisfy that: $G \cap \mathcal{P}_i$ is trivial (for $i = 0$ or $i = L$) or the whole $\mathcal{P}_i$ (for $0 < i < L$).

C.3 COMPUTE $gE(U)$

We have:

$$
\begin{aligned}
[g(E(W))]^{(i)}_{jk} =& \frac{d^{(i)}_j}{d^{(i-1)}_k} \cdot [W']^{(i)}_{\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)} \\
=& \sum_{L\geqslant s>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \frac{d^{(i)}_j}{d^{(i-1)}_k} \cdot \Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_{(s,t):pq} \cdot [W]^{(s,t)}_{pq} \\
&+ \sum_{L\geqslant s>0} \sum_{p=1}^{n_s} \frac{d^{(i)}_j}{d^{(i-1)}_k} \cdot \Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_{(s):p} \cdot [b]^{(s)}_p \\
&+ \sum_{L\geqslant s>t>0} \sum_{p=1}^{n_s} \frac{d^{(i)}_j}{d^{(i-1)}_k} \cdot \Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_{(s,t)(t):p} \cdot [Wb]^{(s,t)(t)}_p \\
&+ \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \frac{d^{(i)}_j}{d^{(i-1)}_k} \cdot \Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_{(s)(L,t):pq} \cdot [bW]^{(s)(L,t)}_{pq} \\
&+ \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \frac{d^{(i)}_j}{d^{(i-1)}_k} \cdot \Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_{(s,0)(L,t):pq} \cdot [WW]^{(s,0)(L,t)}_{pq}
\end{aligned}
$$

$$
\begin{aligned}
&+ \frac{d_j^{(i)}}{d_k^{(i-1)}} \cdot \Phi_1^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)} \\
[g(E(b))]_j^{(i)} &= d_j^{(i)} \cdot [b']^{(i)}_{\pi_i^{-1}(j)} \\
&= \sum_{L\geqslant s>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} d_j^{(i)} \cdot \Phi_{(s,t):pq}^{(i):\pi_i^{-1}(j)} \cdot [W]_{pq}^{(s,t)} \\
&\quad + \sum_{L\geqslant s>0} \sum_{p=1}^{n_s} d_j^{(i)} \cdot \Phi_{(s):p}^{(i):\pi_i^{-1}(j)} \cdot [b]_p^{(s)} \\
&\quad + \sum_{L\geqslant s>t>0} \sum_{p=1}^{n_s} d_j^{(i)} \cdot \Phi_{(s,t)(t):p}^{(i):\pi_i^{-1}(j)} \cdot [Wb]_p^{(s,t)(t)} \\
&\quad + \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} d_j^{(i)} \cdot \Phi_{(s)(L,t):pq}^{(i):\pi_i^{-1}(j)} \cdot [bW]_{pq}^{(s)(L,t)} \\
&\quad + \sum_{L\geqslant s>0} \sum_{L>t\geqslant 0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_t} d_j^{(i)} \cdot \Phi_{(s,0)(L,t):pq}^{(i):\pi_i^{-1}(j)} \cdot [WW]_{pq}^{(s,0)(L,t)} \\
&\quad + d_j^{(i)} \cdot \Phi_1^{(i):\pi_i^{-1}(j)}
\end{aligned}
$$

C.4 COMPARE $E(gU)$ AND $gE(U)$

Since $E(gU) = gE(U)$, from Corollary B.7, the parameters Φ_-^-'s have to satisfy these following conditions:

1. For all $L \geqslant s > t \geqslant 0$ with $(s,t) \neq (L,0)$, and for all p, q, we have

$$
\Phi_{(s,t):\pi_s(p)\pi_t(q)}^{(i):jk} \cdot \frac{d_{\pi_s(p)}^{(s)}}{d_{\pi_t(q)}^{(t)}} = \frac{d_j^{(i)}}{d_k^{(i-1)}} \cdot \Phi_{(s,t):pq}^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}
$$

$$
\Phi_{(s,t):\pi_s(p)\pi_t(q)}^{(i):j} \cdot \frac{d_{\pi_s(p)}^{(s)}}{d_{\pi_t(q)}^{(t)}} = d_j^{(i)} \cdot \Phi_{(s,t):pq}^{(i):\pi_i^{-1}(j)}
$$

2. For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have

$$
\Phi_{(s,0)(L,t):\pi_s(p)\pi_t(q)}^{(i):jk} \cdot \frac{d_{\pi_s(p)}^{(s)}}{d_{\pi_t(q)}^{(t)}} = \frac{d_j^{(i)}}{d_k^{(i-1)}} \cdot \Phi_{(s,0)(L,t):pq}^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}
$$

$$
\Phi_{(s,0)(L,t):\pi_s(p)\pi_t(q)}^{(i):j} \cdot \frac{d_{\pi_s(p)}^{(s)}}{d_{\pi_t(q)}^{(t)}} = d_j^{(i)} \cdot \Phi_{(s,0)(L,t):pq}^{(i):\pi_i^{-1}(j)}
$$

3. We have

$$
\begin{aligned}
&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L,0):\pi_L(p)\pi_0(q)}^{(i):jk} \cdot \frac{d_{\pi_L(p)}^{(L)}}{d_{\pi_0(q)}^{(0)}} \cdot [W]_{pq}^{(L,0)} \\
&\qquad + \sum_{L>s>0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):\pi_s(p)\pi_s(q)}^{(i):jk} \cdot \frac{d_{\pi_s(p)}^{(s)}}{d_{\pi_s(q)}^{(s)}} \cdot [WW]_{pq}^{(s,0)(L,s)} \\
&= \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \frac{d_j^{(i)}}{d_k^{(i-1)}} \cdot \Phi_{(L,0):pq}^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)} \cdot [W]_{pq}^{(L,0)} \\
&\qquad + \sum_{L>s>0} \sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \frac{d_j^{(i)}}{d_k^{(i-1)}} \cdot \Phi_{(s,0)(L,s):pq}^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)} \cdot [WW]_{pq}^{(s,0)(L,s)}
\end{aligned}
$$

$$\begin{aligned}
&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0}\Phi^{(i):j}_{(L,0):\pi_L(p)\pi_0(q)}\cdot\frac{d^{(L)}_{\pi_L(p)}}{d^{(0)}_{\pi_0(q)}}\cdot[W]^{(L,0)}_{pq}\\
&\qquad\qquad+\sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s}\Phi^{(i):j}_{(s,0)(L,s):\pi_s(p)\pi_s(q)}\cdot\frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}}\cdot[WW]^{(s,0)(L,s)}_{pq}\\
=&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0}d^{(i)}_j\cdot\Phi^{(i):\pi_i^{-1}(j)}_{(L,0):pq}\cdot[W]^{(L,0)}_{pq}\\
&\qquad\qquad+\sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s}d^{(i)}_j\cdot\Phi^{(i):\pi_i^{-1}(j)}_{(s,0)(L,s):pq}\cdot[WW]^{(s,0)(L,s)}_{pq}
\end{aligned}$$

4. For all $L > s > t > 0$, and for all p, we have

$$\Phi^{(i):jk}_{(s,t)(t):\pi_s(p)}\cdot d^{(s)}_{\pi_s(p)}=\frac{d^{(i)}_j}{d^{(i-1)}_k}\cdot\Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_{(s,t)(t):p}$$

$$\Phi^{(i):j}_{(s,t)(t):\pi_s(p)}\cdot d^{(s)}_{\pi_s(p)}=d^{(i)}_j\cdot\Phi^{(i):\pi_i^{-1}(j)}_{(s,t)(t):p}.$$

5. For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have

$$\Phi^{(i):jk}_{(s)(L,t):\pi_s(p)\pi_t(q)}\cdot\frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}}=\frac{d^{(i)}_j}{d^{(i-1)}_k}\cdot\Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_{(s)(L,t):pq}$$

$$\Phi^{(i):j}_{(s)(L,t):\pi_s(p)\pi_t(q)}\cdot\frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}}=d^{(i)}_j\cdot\Phi^{(i):\pi_i^{-1}(j)}_{(s)(L,t):pq}$$

6. For all $L > t > 0$, we have

$$\begin{aligned}
&\sum_{p=1}^{n_L}\Phi^{(i):jk}_{(L,t)(t):\pi_L(p)}\cdot d^{(L)}_{\pi_L(p)}\cdot[Wb]^{(L,t)(t)}_{p}\\
&\qquad\qquad+\sum_{p=1}^{n_t}\sum_{q=1}^{n_t}\Phi^{(i):jk}_{(t)(L,t):\pi_t(p)\pi_t(q)}\cdot\frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}}\cdot[bW]^{(t)(L,t)}_{pq}\\
=&\sum_{p=1}^{n_L}\frac{d^{(i)}_j}{d^{(i-1)}_k}\cdot\Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_{(L,t)(t):p}\cdot[Wb]^{(L,t)(t)}_{p}\\
&\qquad\qquad+\sum_{p=1}^{n_t}\sum_{q=1}^{n_t}\frac{d^{(i)}_j}{d^{(i-1)}_k}\cdot\Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_{(t)(L,t):pq}\cdot[bW]^{(t)(L,t)}_{pq}
\end{aligned}$$

$$\begin{aligned}
&\sum_{p=1}^{n_L}\Phi^{(i):j}_{(L,t)(t):\pi_L(p)}\cdot d^{(L)}_{\pi_L(p)}\cdot[Wb]^{(L,t)(t)}_{p}\\
&\qquad\qquad+\sum_{p=1}^{n_t}\sum_{q=1}^{n_t}\Phi^{(i):j}_{(t)(L,t):\pi_t(p)\pi_t(q)}\cdot\frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}}\cdot[bW]^{(t)(L,t)}_{pq}\\
=&\sum_{p=1}^{n_L}d^{(i)}_j\cdot\Phi^{(i):\pi_i^{-1}(j)}_{(L,t)(t):p}\cdot[Wb]^{(L,t)(t)}_{p}\\
&\qquad\qquad+\sum_{p=1}^{n_t}\sum_{q=1}^{n_t}d^{(i)}_j\cdot\Phi^{(i):\pi_i^{-1}(j)}_{(t)(L,t):pq}\cdot[bW]^{(t)(L,t)}_{pq}
\end{aligned}$$

7. For all $L \geqslant s > 0$ and for all p, we have

$$\Phi^{(i):jk}_{(s):\pi_s(p)}\cdot d^{(s)}_{\pi_s(p)}=\frac{d^{(i)}_j}{d^{(i-1)}_k}\cdot\Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_{(s):p}$$

$$\Phi^{(i):j}_{(s):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} = d^{(i)}_j \cdot \Phi^{(i):\pi_i^{-1}(j)}_{(s):p}.$$

8. We have

$$\Phi^{(i):jk}_1 = \frac{d^{(i)}_j}{d^{(i-1)}_k} \cdot \Phi^{(i):\pi_i^{-1}(j)\pi_{i-1}^{-1}(k)}_1$$
$$\Phi^{(i):j}_1 = d^{(i)}_j \cdot \Phi^{(i):\pi_i^{-1}(j)}_1$$

Also, since the group G satisfy that: $G \cap \mathcal{P}_i$ is trivial (for $i = 0$ or $i = L$) or the whole $\mathcal{P}_i$ (for $0 < i < L$), so we can simplify the above conditions by moving some of the permutation π's to the left hand sides. We have

1. For all $L \geqslant s > t \geqslant 0$ with $(s,t) \neq (L,0)$, and for all p, q, we have

$$\Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(s,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(s,t):pq}$$
$$\Phi^{(i):\pi_i(j)}_{(s,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s,t):pq}$$

2. For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have

$$\Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(s,0)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(s,0)(L,t):pq}$$
$$\Phi^{(i):\pi_i(j)}_{(s,0)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s,0)(L,t):pq}$$

3. We have

$$\begin{aligned}
&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(L,0):\pi_L(p)\pi_0(q)} \cdot \frac{d^{(L)}_{\pi_L(p)}}{d^{(0)}_{\pi_0(q)}} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(s,0)(L,s):\pi_s(p)\pi_s(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}} \cdot [WW]^{(s,0)(L,s)}_{pq} \\
=&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq} \\
&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi^{(i):\pi_i(j)}_{(L,0):\pi_L(p)\pi_0(q)} \cdot \frac{d^{(L)}_{\pi_L(p)}}{d^{(0)}_{\pi_0(q)}} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi^{(i):\pi_i(j)}_{(s,0)(L,s):\pi_s(p)\pi_s(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}} \cdot [WW]^{(s,0)(L,s)}_{pq} \\
=&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq}
\end{aligned}$$

4. For all $L > s > t > 0$, and for all p, we have

$$\Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(s,t)(t):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} = \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(s,t)(t):p}$$
$$\Phi^{(i):\pi_i(j)}_{(s,t)(t):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} = d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s,t)(t):p}.$$

5. For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have

$$\Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(s)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(s)(L,t):pq}$$
$$\Phi^{(i):\pi_i(j)}_{(s)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s)(L,t):pq}$$

6. For all $L > t > 0$, we have

$$\sum_{p=1}^{n_L} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(L,t)(t):\pi_L(p)} \cdot d^{(L)}_{\pi_L(p)} \cdot [Wb]^{(L,t)(t)}_p + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(t)(L,t)}_{pq}$$
$$= \sum_{p=1}^{n_L} \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq}$$
$$\sum_{p=1}^{n_L} \Phi^{(i):\pi_i(j)}_{(L,t)(t):\pi_L(p)} \cdot d^{(L)}_{\pi_L(p)} \cdot [Wb]^{(L,t)(t)}_p + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi^{(i):\pi_i(j)}_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(t)(L,t)}_{pq}$$
$$= \sum_{p=1}^{n_L} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq}$$

7. For all $L \geqslant s > 0$ and for all p, we have

$$\Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(s):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} = \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(s):p}$$
$$\Phi^{(i):\pi_i(j)}_{(s):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} = d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s):p}.$$

8. We have

$$\Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_1 = \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_1$$
$$\Phi^{(i):\pi_i(j)}_1 = d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_1$$

From the above equalities, we can determine all constraints for the coefficients according to the entries of $[E(W)]$ as follows:

1. • If $(s,t) = (i,i-1) \notin \{(L,L-1),(1,0)\}$, $p = j$, $q = k$,

$$\Phi^{(i):\pi(j)\pi'(k)}_{(i,i-1):\pi(j)\pi'(k)} = \Phi^{(i):jk}_{(i,i-1):jk}$$

• If $(s,t) = (i,i-1) = (L,L-1)$, $q = k$,

$$\Phi^{(L):j\pi_{L-1}(k)}_{(L,L-1):p\pi_{L-1}(k)} = \Phi^{(L):jk}_{(L,L-1):pk}$$

• If $(s,t) = (i,i-1) = (1,0)$, $p = j$,

$$\Phi^{(1):\pi_1(j)k}_{(1,0):\pi_1(j)q} = \Phi^{(1):jk}_{(1,0):jq}$$

2. • If $(s,t) = (i,i-1) \notin \{(L,L-1),(1,0)\}$, $p = j$, $q = k$,

$$\Phi^{(i):\pi(j)\pi'(k)}_{(i,0)(L,i-1):\pi(j)\pi'(k)} = \Phi^{(i):jk}_{(i,0)(L,i-1):jk}$$

• If $(s,t) = (i,i-1) = (L,L-1)$, $q = k$,

$$\Phi^{(L):j\pi_{L-1}(k)}_{(L,0)(L,L-1):p\pi_{L-1}(k)} = \Phi^{(L):jk}_{(L,0)(L,L-1):pk}$$

• If $(s,t) = (i,i-1) = (1,0)$, $p = j$,

$$\Phi^{(1):\pi_1(j)k}_{(1,0)(L,0):\pi_1(j)q} = \Phi^{(1):jk}_{(1,0)(L,0):jq}$$

3.

$$\begin{aligned}
&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(s,0)(L,s):\pi_s(p)\pi_s(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}} \cdot [WW]^{(s,0)(L,s)}_{pq} \\
=&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq}
\end{aligned}$$

For each $L > l > 0$, by scaling all the $d^{(l)}_{-}$'s by the same scalar, we have

$$\begin{aligned}
0 =& \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(s,0)(L,s):\pi_s(p)\pi_s(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}} \cdot [WW]^{(s,0)(L,s)}_{pq} \\
=& \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi^{(i):jk}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi^{(i):jk}_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq}
\end{aligned}$$

4. For all $L > s > t > 0$, and for all p, we have

$$\Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(s,t)(t):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} = \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(s,t)(t):p}.$$

We have

$$\Phi^{(i):jk}_{(s,t)(t):p} = 0.$$

5. For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have

$$\Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(s)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(s)(L,t):pq}.$$

- If $(s,t) = (i, i-1) \notin \{(L, L-1), (1,0)\}$, $p = j$, $q = k$,

$$\Phi^{(i):\pi(j)\pi'(k)}_{(i)(L,i-1):\pi(j)\pi'(k)} = \Phi^{(i):jk}_{(i)(L,i-1):jk}.$$

- If $i = L$, $(s,t) = (L, L-1)$, $q = k$,

$$\Phi^{(L):j\pi(k)}_{(L)(L,L-1):p\pi(k)} = \Phi^{(L):jk}_{(L)(L,L-1):pk}$$

- If $i = 1$, $(s,t) = (1,0)$, $p = j$,

$$\Phi^{(1):\pi(j)k}_{(1)(L,0):\pi(j)q} = \Phi^{(1):jk}_{(1)(L,0):jq}.$$

6. For all $L > t > 0$, we have

$$\begin{aligned}
&\sum_{p=1}^{n_L} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p \\
&\qquad + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(t)(L,t)}_{pq} \\
=&\sum_{p=1}^{n_L} \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p \\
&\qquad + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq}
\end{aligned}$$

For each $L > l > 0$, by scaling all the $d^{(l)}_{\_}$'s by the same scalar, we have

$$\begin{aligned}
0 =&\sum_{p=1}^{n_L} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p \\
&\qquad + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi^{(i):\pi_i(j)\pi_{i-1}(k)}_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(t)(L,t)}_{pq} \\
=&\sum_{p=1}^{n_L} \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p \\
&\qquad + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \frac{d^{(i)}_{\pi_i(j)}}{d^{(i-1)}_{\pi_{i-1}(k)}} \cdot \Phi^{(i):jk}_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq}
\end{aligned}$$

7. If $i = s = 1$, $p = j$,

$$\Phi^{(1):\pi(j)k}_{(1):\pi(j)} = \Phi^{(1):jk}_{(1):j}.$$

8. We have

$$\Phi_1^{(i):jk} = 0.$$

Similarly, we can also determine all constraints for the coefficients according to the entries of $[E(b)]$ as follows:

1. For all $L \geqslant s > t \geqslant 0$ with $(s,t) \neq (L,0)$, and for all p, q, we have

$$\Phi^{(i):\pi_i(j)}_{(s,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s,t):pq}$$

$(s,t) = (i,0), p = j,$

$$\Phi^{(i):\pi(j)}_{(i,0):\pi(j)q} = \Phi^{(i):j}_{(i,0):jq}$$

2. For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have

$$\Phi^{(i):\pi_i(j)}_{(s,0)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s,0)(L,t):pq}$$

- $t = 0, s = i < L, p = j,$

$$\Phi^{(i):\pi(j)}_{(i,0)(L,0):\pi(j)q} \cdot = \Phi^{(i):j}_{(i,0)(L,0):jq}$$

- $t = 0, s = i = L,$

$$\Phi^{(L):j}_{(L,0)(L,0):pq} = \Phi^{(L):j}_{(L,0)(L,0):pq}$$

3. We have

$$\begin{aligned}
&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi^{(i):\pi_i(j)}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi^{(i):\pi_i(j)}_{(s,0)(L,s):\pi_s(p)\pi_s(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}} \cdot [WW]^{(s,0)(L,s)}_{pq} \\
=&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq}
\end{aligned}$$

- If $L > i > 0$. For each $L > l > 0$, by scaling all the $d^{(l)}_{-}$'s by the same scalar, we have

$$\begin{aligned}
0 =&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi^{(i):\pi_i(j)}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi^{(i):\pi_i(j)}_{(s,0)(L,s):\pi_s(p)\pi_s(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}} \cdot [WW]^{(s,0)(L,s)}_{pq} \\
=&\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq}
\end{aligned}$$

- If $i = L$. We have

$$\sum_{p=1}^{n_L}\sum_{q=1}^{n_0}\Phi^{(L):j}_{(L,0):pq}\cdot[W]^{(L,0)}_{pq} + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s}\Phi^{(L):j}_{(s,0)(L,s):\pi_s(p)\pi_s(q)}\cdot\frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}}\cdot[WW]^{(s,0)(L,s)}_{pq}$$
$$=\sum_{p=1}^{n_L}\sum_{q=1}^{n_0}\Phi^{(L):j}_{(L,0):pq}\cdot[W]^{(L,0)}_{pq} + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s}\Phi^{(L):j}_{(s,0)(L,s):pq}\cdot[WW]^{(s,0)(L,s)}_{pq}$$

which means $\Phi^{(L):j}_{(L,0):pq}$ can be arbitrary. The rest is

$$\sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s}\Phi^{(L):j}_{(s,0)(L,s):\pi_s(p)\pi_s(q)}\cdot\frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}}\cdot[WW]^{(s,0)(L,s)}_{pq}$$
$$=\sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s}\Phi^{(L):j}_{(s,0)(L,s):pq}\cdot[WW]^{(s,0)(L,s)}_{pq}$$

For an $L > r > 0$, by letting π_r be the identity, and $d^{(r)}_p$ be 1 for all p, we have

$$\sum_{L>s>0,s\neq r}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s}\Phi^{(L):j}_{(s,0)(L,s):\pi_s(p)\pi_s(q)}\cdot\frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}}\cdot[WW]^{(s,0)(L,s)}_{pq}$$
$$=\sum_{L>s>0,s\neq r}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s}\Phi^{(L):j}_{(s,0)(L,s):pq}\cdot[WW]^{(s,0)(L,s)}_{pq},$$

so

$$\sum_{p=1}^{n_r}\sum_{q=1}^{n_r}\Phi^{(L):j}_{(r,0)(L,r):\pi_r(p)\pi_r(q)}\cdot\frac{d^{(r)}_{\pi_r(p)}}{d^{(r)}_{\pi_r(q)}}\cdot[WW]^{(r,0)(L,r)}_{pq}$$
$$=\sum_{p=1}^{n_r}\sum_{q=1}^{n_r}\Phi^{(L):j}_{(r,0)(L,r):pq}\cdot[WW]^{(r,0)(L,r)}_{pq}$$

By Lemma B.3 and Corollary A.7, we have

$$\Phi^{(L):j}_{(r,0)(L,r):\pi_r(p)\pi_r(q)}\cdot\frac{d^{(r)}_{\pi_r(p)}}{d^{(r)}_{\pi_r(q)}} = \Phi^{(L):j}_{(r,0)(L,r):pq}.$$

So

$$\Phi^{(L):j}_{(r,0)(L,r):pq} = 0$$

for $p \neq q$, and

$$\Phi^{(L):j}_{(r,0)(L,r):\pi_r(p)\pi_r(p)} = \Phi^{(L):j}_{(r,0)(L,r):pp}.$$

In conclusion, we have $\Phi^{(L):j}_{(L,0):pq}$ is arbitrary, and for $L > s > 0$,

$$\Phi^{(L):j}_{(s,0)(L,s):\pi(p)\pi(p)} = \Phi^{(L):j}_{(s,0)(L,s):pp}.$$

4. For all $L > s > t > 0$, and for all p, we have

$$\Phi^{(i):\pi_i(j)}_{(s,t)(t):\pi_s(p)}\cdot d^{(s)}_{\pi_s(p)} = d^{(i)}_{\pi_i(j)}\cdot\Phi^{(i):j}_{(s,t)(t):p}.$$

If $i = s, p = j$,

$$\Phi^{(i):\pi(j)}_{(i,t)(t):\pi(j)} = \Phi^{(i):j}_{(i,t)(t):j}.$$

5. For all $L \geqslant s > 0, L > t \geqslant 0$ with $s \neq t$, we have

$$\Phi^{(i):\pi_i(j)}_{(s)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(s)(L,t):pq}$$

- $s = i < L, t = 0, p = j,$

$$\Phi^{(i):\pi(j)}_{(i)(L,0):\pi(j)q} = \Phi^{(i):j}_{(i)(L,0):jq}$$

- $s = i = L, t = 0$

$$\Phi^{(L):j}_{(L)(L,0):pq} = \Phi^{(L):j}_{(L)(L,0):pq}$$

6. For all $L > t > 0$, we have

$$\begin{aligned}
&\sum_{p=1}^{n_L} \Phi^{(i):\pi_i(j)}_{(L,t)(t):p)} \cdot [Wb]^{(L,t)(t)}_p \\
&\qquad + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi^{(i):\pi_i(j)}_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(t)(L,t)}_{pq} \\
=& \sum_{p=1}^{n_L} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p \\
&\qquad + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq}
\end{aligned}$$

- If $L > i > 0$. For each $L > l > 0$, by scaling all the $d^{(l)}_{\_}$'s by the same scalar, we have

$$\begin{aligned}
0 =& \sum_{p=1}^{n_L} \Phi^{(i):\pi_i(j)}_{(L,t)(t):\pi_L(p)} \cdot d^{(L)}_{\pi_L(p)} \cdot [Wb]^{(L,t)(t)}_p \\
&\qquad + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi^{(i):\pi_i(j)}_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(t)(L,t)}_{pq} \\
=& \sum_{p=1}^{n_L} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p \\
&\qquad + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} d^{(i)}_{\pi_i(j)} \cdot \Phi^{(i):j}_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq}
\end{aligned}$$

- If $i = L$. We have

$$\begin{aligned}
&\sum_{p=1}^{n_L} \Phi^{(L):j}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p \\
&\qquad + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi^{(L):j}_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(t)(L,t)}_{pq} \\
=& \sum_{p=1}^{n_L} \Phi^{(L):j}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p \\
&\qquad + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \cdot \Phi^{(L):j}_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq}
\end{aligned}$$

which means $\Phi^{(L):j}_{(L,t)(t):p)}$ can be arbitrary. The rest is

$$\sum_{p=1}^{n_t}\sum_{q=1}^{n_t}\Phi^{(L):j}_{(t)(L,t):\pi_t(p)\pi_t(q)}\cdot\frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}}\cdot[bW]^{(t)(L,t)}_{pq}$$
$$=\sum_{p=1}^{n_t}\sum_{q=1}^{n_t}\Phi^{(L):j}_{(t)(L,t):pq}\cdot[bW]^{(t)(L,t)}_{pq}$$

By Lemma B.5 and Corollary A.7, we have

$$\Phi^{(L):j}_{(t)(L,t):\pi_t(p)\pi_t(q)}\cdot\frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}}=\Phi^{(L):j}_{(t)(L,t):pq}$$

So

$$\Phi^{(L):j}_{(t)(L,t):pq}=0$$

for $p\neq q$, and

$$\Phi^{(L):j}_{(t)(L,t):\pi_t(p)\pi_t(p)}=\Phi^{(L):j}_{(t)(L,t):pp}$$

In conclusion, for all $L>t>0$, we have $\Phi^{(L):j}_{(L,t)(t):p)}$ is arbitrary, and

$$\Phi^{(L):j}_{(t)(L,t):\pi(p)\pi(p)}=\Phi^{(L):j}_{(t)(L,t):pp}.$$

7. For all $L\geqslant s>0$ and for all p, we have

$$\Phi^{(i):\pi_i(j)}_{(s):\pi_s(p)}\cdot d^{(s)}_{\pi_s(p)}=d^{(i)}_{\pi_i(j)}\cdot\Phi^{(i):j}_{(s):p}.$$

- If $i=s<L$, $p=j$,

$$\Phi^{(i):\pi(j)}_{(i):\pi(j)}=\Phi^{(i):j}_{(i):j}.$$

- If $i=s=L$,

$$\Phi^{(L):j}_{(L):p}=\Phi^{(L):j}_{(L):p}.$$

8. We have

$$\Phi^{(i):\pi_i(j)}_1=d^{(i)}_{\pi_i(j)}\cdot\Phi^{(i):j}_1$$

- If $i=L$, we have

$$\Phi^{(L):j}_1=\Phi^{(L):j}_1.$$

C.5 G-Equivariant polynomial layers

Based on the above discussions, we conclude that every G-equivariant polynomial layer , which is defined as a linear combination of stable polynomial terms, is given as $E(U)=([E(W)],[E(b)])$, where the entries of $[E(W)]$ and $[E(b)]$ are given case by case as folows:

- For $i=L$, we have

$$[E(W)]^{(L)}_{jk}=\sum_{p=1}^{n_L}\Phi^{(L):j\bullet}_{(L,L-1):p\bullet}\cdot[W]^{(L,L-1)}_{pk}+\sum_{p=1}^{n_L}\Phi^{(L):j\bullet}_{(L,0)(L,L-1):p\bullet}\cdot[WW]^{(L,0)(L,L-1)}_{pk}$$
$$+\sum_{p=1}^{n_L}\Phi^{(L):j\bullet}_{(L)(L,L-1):p\bullet}\cdot[bW]^{(L)(L,L-1)}_{pk}$$

$$[E(b)]^{(L)}_{j}=\sum_{p=1}^{n_L}\sum_{q=1}^{n_0}\Phi^{(L):j}_{(L,0)(L,0):pq}\cdot[WW]^{(L,0)(L,0)}_{pq}+\sum_{p=1}^{n_L}\sum_{q=1}^{n_0}\Phi^{(L):j}_{(L,0):pq}\cdot[W]^{(L,0)}_{pq}$$

$$
\begin{aligned}
&+ \sum_{L>s>0} \sum_{p=1}^{n_s} \Phi^{(L):j}_{(s,0)(L,s):\bullet\bullet} \cdot [WW]^{(s,0)(L,s)}_{pp} + \sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi^{(L):j}_{(L)(L,0):pq} \cdot [bW]^{(L)(L,0)}_{pq} \\
&+ \sum_{L>t>0} \sum_{p=1}^{n_L} \Phi^{(L):j}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_{p} + \sum_{L>t>0} \sum_{p=1}^{n_t} \Phi^{(L):j}_{(t)(L,t):\bullet\bullet} \cdot [bW]^{(t)(L,t)}_{pp} \\
&+ \sum_{p=1}^{n_L} \Phi^{(L):j}_{(L):p} \cdot [b]^{(L)}_{p} + \Phi^{(L):j}_{1}
\end{aligned}
$$

$$
\begin{aligned}
[E(b)]^{(L)}_{j} =& \sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi^{(L):j}_{(L,0)(L,0):pq} \cdot [WW]^{(L,0)(L,0)}_{pq} + \sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi^{(L):j}_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&+ \sum_{L>s>0} \Phi^{(L):j}_{(s,0)(L,s):\bullet\bullet} \cdot \sum_{p=1}^{n_s} [WW]^{(s,0)(L,s)}_{pp} + \sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi^{(L):j}_{(L)(L,0):pq} \cdot [bW]^{(L)(L,0)}_{pq} \\
&+ \sum_{L>t>0} \sum_{p=1}^{n_L} \Phi^{(L):j}_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_{p} + \sum_{L>t>0} \Phi^{(L):j}_{(t)(L,t):\bullet\bullet} \cdot \sum_{p=1}^{n_t} [bW]^{(t)(L,t)}_{pp} \\
&+ \sum_{p=1}^{n_L} \Phi^{(L):j}_{(L):p} \cdot [b]^{(L)}_{p} + \Phi^{(L):j}_{1}
\end{aligned}
$$

- For $i = 1$, we have

$$
\begin{aligned}
[E(W)]^{(1)}_{jk} =& \sum_{q=1}^{n_0} \Phi^{(1):\bullet k}_{(1,0):\bullet q} \cdot [W]^{(1,0)}_{jq} + \sum_{q=1}^{n_0} \Phi^{(1):\bullet k}_{(1,0)(L,0):\bullet q} \cdot [WW]^{(1,0)(L,0)}_{jq} \\
&+ \sum_{q=1}^{n_0} \Phi^{(1):\bullet k}_{(1)(L,0):\bullet q} \cdot [bW]^{(1)(L,0)}_{jq} + \Phi^{(1):\bullet k}_{(1):\bullet} \cdot [b]^{(1)}_{j} \\
[E(b)]^{(1)}_{j} =& \sum_{q=1}^{n_0} \Phi^{(1):\bullet}_{(1,0):\bullet q} \cdot [W]^{(1,0)}_{jq} + \sum_{q=1}^{n_0} \Phi^{(1):\bullet}_{(1,0)(L,0):\bullet q} \cdot [WW]^{(1,0)(L,0)}_{jq} \\
&+ \sum_{q=1}^{n_0} \Phi^{(1):\bullet}_{(1)(L,0):\bullet q} \cdot [bW]^{(1)(L,0)}_{jq} + \Phi^{(1):\bullet}_{(1):\bullet} \cdot [b]^{(1)}_{j}
\end{aligned}
$$

- For $L > i > 1$, we have

$$
\begin{aligned}
[E(W)]^{(i)}_{jk} =& \Phi^{(i):\bullet\bullet}_{(i,i-1):\bullet\bullet} \cdot [W]^{(i,i-1)}_{jk} + \Phi^{(i):\bullet\bullet}_{(i,0)(L,i-1):\bullet\bullet} \cdot [WW]^{(i,0)(L,i-1)}_{jk} \\
&+ \Phi^{(i):\bullet\bullet}_{(i)(L,i-1):\bullet\bullet} \cdot [bW]^{(i)(L,i-1)}_{jk} \\
[E(b)]^{(i)}_{j} =& \sum_{q=1}^{n_0} \Phi^{(i):\bullet}_{(i,0):\bullet q} \cdot [W]^{(i,0)}_{jq} + \sum_{q=1}^{n_0} \Phi^{(i):\bullet}_{(i,0)(L,0):\bullet q} \cdot [WW]^{(t,0)(L,0)}_{jq} \\
&+ \sum_{i>t>0} \sum_{p=1}^{n_i} \Phi^{(i):\bullet}_{(i,t)(t):\bullet} \cdot [Wb]^{(i,t)(t)}_{j} + \sum_{q=1}^{n_t} \Phi^{(i):\bullet}_{(i)(L,0):\bullet q} \cdot [bW]^{(i)(L,0)}_{jq} \\
&+ \Phi^{(i):\bullet}_{(i):\bullet} \cdot [b]^{(i)}_{j}
\end{aligned}
$$

In the above formulas, the bullet • indicates that the corresponding coefficient is independent of the indices at the bullet.

D INVARIANT POLYNOMIAL LAYERS

In this section, we construct a polynomial map $I \colon \mathcal{U} \to \mathbb{R}^{d'}$ that is G-invariant, i.e., $I(gU) = I(U)$ for all $g \in \mathcal{G}_{\mathcal{U}}$ and $U \in \mathcal{U}$.

D.1 INVARIANT LAYER AS A LINEAR COMBINATION OF STABLE POLYNOMIAL TERMS

Similar to the equivariant maps, we seek the invariant map among polynomial maps that is a linear combination of stable polynomial terms, specifically the entries of $[W]^{(s,t)}$, $[b]^{(s)}$, $[Wb]^{(s,t)(t)}$, $[bW]^{(s)(L,t)}$, and $[WW]^{(s,0)(L,t)}$, along with a bias. In concrete:

$$\begin{aligned}I(U) \coloneqq &\sum_{L\geqslant s>t\geqslant 0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_t}\Phi_{(s,t):pq}\cdot[W]^{(s,t)}_{pq} + \sum_{L\geqslant s>0}\sum_{p=1}^{n_s}\Phi_{(s):p}\cdot[b]^{(s)}_{p}\\ &+\sum_{L\geqslant s>t>0}\sum_{p=1}^{n_s}\Phi_{(s,t)(t):p}\cdot[Wb]^{(s,t)(t)}_{p}\\ &+\sum_{L\geqslant s>0}\sum_{L>t\geqslant 0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_t}\Phi_{(s)(L,t):pq}\cdot[bW]^{(s)(L,t)}_{pq}\\ &+\sum_{L\geqslant s>0}\sum_{L>t\geqslant 0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_t}\Phi_{(s,0)(L,t):pq}\cdot[WW]^{(s,0)(L,t)}_{pq}+\Phi_1.\end{aligned}$$

All Φ_{-}'s are in $\mathbb{R}^{d'\times d}$, except the bias Φ_1 is in $\mathbb{R}^{d'\times 1}$. In summary, I is parameterized by Φ_{-}'s and Ψ_{-}'s. We need I to be G-invariant, which means $I(gU) = I(U)$.

D.2 COMPUTE $I(gU)$

From the definition of $I(U)$, we have:

$$\begin{aligned}I(gU) = &\sum_{L\geqslant s>t\geqslant 0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_t}\Phi_{(s,t):pq}\cdot[gW]^{(s,t)}_{pq}\\ &+\sum_{L\geqslant s>0}\sum_{p=1}^{n_s}\Phi_{(s):p}\cdot[gb]^{(s)}_{p}\\ &+\sum_{L\geqslant s>t>0}\sum_{p=1}^{n_s}\Phi_{(s,t)(t):p}\cdot[gWgb]^{(s,t)(t)}_{p}\\ &+\sum_{L\geqslant s>0}\sum_{L>t\geqslant 0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_t}\Phi_{(s)(L,t):pq}\cdot[gbgW]^{(s)(L,t)}_{pq}\\ &+\sum_{L\geqslant s>0}\sum_{L>t\geqslant 0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_t}\Phi_{(s,0)(L,t):pq}\cdot[gWgW]^{(s,0)(L,t)}_{pq}\\ &+\Phi_1\\ = &\sum_{L\geqslant s>t\geqslant 0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_t}\Phi_{(s,t):pq}\cdot\frac{d^{(s)}_p}{d^{(t)}_q}\cdot[W]^{(s,t)}_{\pi_s^{-1}(p),\pi_t^{-1}(q)}\\ &+\sum_{L\geqslant s>0}\sum_{p=1}^{n_s}\Phi_{(s):p}\cdot d^{(s)}_p\cdot[b]^{(s)}_{\pi_s^{-1}(p)}\\ &+\sum_{L\geqslant s>t>0}\sum_{p=1}^{n_s}\Phi_{(s,t)(t):p}\cdot d^{(s)}_p\cdot[Wb]^{(s,t)(t)}_{\pi_s^{-1}(p)}\\ &+\sum_{L\geqslant s>0}\sum_{L>t\geqslant 0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_t}\Phi_{(s)(L,t):pq}\cdot\frac{d^{(s)}_p}{d^{(t)}_q}\cdot[bW]^{(s)(L,t)}_{\pi_s^{-1}(p),\pi_t^{-1}(q)}\\ &+\sum_{L\geqslant s>0}\sum_{L>t\geqslant 0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_t}\Phi_{(s,0)(L,t):pq}\cdot\frac{d^{(s)}_p}{d^{(t)}_q}\cdot[WW]^{(s,0)(L,t)}_{\pi_s^{-1}(p),\pi_t^{-1}(q)}\end{aligned}$$

$$
\begin{aligned}
&+ \Phi_1 \\
&= \sum_{L \geqslant s > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [W]^{(s,t)}_{pq} \\
&+ \sum_{L \geqslant s > 0} \sum_{p=1}^{n_s} \Phi_{(s):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} \cdot [b]^{(s)}_p \\
&+ \sum_{L \geqslant s > t > 0} \sum_{p=1}^{n_s} \Phi_{(s,t)(t):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} \cdot [Wb]^{(s,t)(t)}_p \\
&+ \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(s)(L,t)}_{pq} \\
&+ \sum_{L \geqslant s > 0} \sum_{L > t \geqslant 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_t} \Phi_{(s,0)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [WW]^{(s,0)(L,t)}_{pq} \\
&+ \Phi_1.
\end{aligned}
$$

D.3 Compare $I(gU)$ and $I(U)$

Since $I(gU) = I(U)$, from Corollary B.7, the parameters Φ_-'s have to satisfy these following conditions:

1. For all $L \geqslant s > t \geqslant 0$ with $(s,t) \neq (L,0)$, and for all p, q, we have

$$\Phi_{(s,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = \Phi_{(s,t):pq}$$

2. For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have

$$\Phi_{(s,0)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = \Phi_{(s,0)(L,t):pq}$$

3. We have

$$
\begin{aligned}
&\sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi_{(L,0):\pi_L(p)\pi_0(q)} \cdot \frac{d^{(L)}_{\pi_L(p)}}{d^{(0)}_{\pi_0(q)}} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L > s > 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):\pi_s(p)\pi_s(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}} \cdot [WW]^{(s,0)(L,s)}_{pq} \\
= &\sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&\qquad + \sum_{L > s > 0} \sum_{p=1}^{n_s} \sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq}
\end{aligned}
$$

4. For all $L > s > t > 0$, and for all p, we have

$$\Phi_{(s,t)(t):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} = \Phi_{(s,t)(t):p}.$$

5. For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have

$$\Phi_{(s)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = \Phi_{(s)(L,t):pq}$$

6. For all $L > t > 0$, we have

$$\sum_{p=1}^{n_L} \Phi_{(L,t)(t):\pi_L(p)} \cdot d^{(L)}_{\pi_L(p)} \cdot [Wb]^{(L,t)(t)}_p + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(t)(L,t)}_{pq}$$
$$= \sum_{p=1}^{n_L} \Phi_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_p + \sum_{p=1}^{n_t}\sum_{q=1}^{n_t} \Phi_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq}$$

7. For all $L \geqslant s > 0$ and for all p, we have

$$\Phi_{(s):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} = \Phi_{(s):p}.$$

8. We have

$$\Phi_1 = \Phi_1.$$

Solve these equations, we have

1. For all $L \geqslant s > t \geqslant 0$ with $(s,t) \neq (L,0)$, and for all p, q, we have

$$\Phi_{(s,t):pq} = 0$$

2. For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have

$$\Phi_{(s,0)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = \Phi_{(s,0)(L,t):pq}$$

If $(s,t) = (L, 0$, we have

$$\Phi_{(L,0)(L,0):pq} = \Phi_{(L,0)(L,0):pq}.$$

3. We have

$$\sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):\pi_s(p)\pi_s(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}} \cdot [WW]^{(s,0)(L,s)}_{pq}$$
$$= \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} + \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq}$$

which means $\Phi_{(L,0):pq}$ can be arbitrary. The rest is

$$\sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):\pi_s(p)\pi_s(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}} \cdot [WW]^{(s,0)(L,s)}_{pq}$$
$$= \sum_{L>s>0}\sum_{p=1}^{n_s}\sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq}$$

For an $L > r > 0$, by letting π_r be the identity, and $d_p^{(r)}$ be 1 for all p, we have

$$\sum_{L>s>0, s\neq r} \sum_{p=1}^{n_s} \sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):\pi_s(p)\pi_s(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(s)}_{\pi_s(q)}} \cdot [WW]^{(s,0)(L,s)}_{pq}$$
$$= \sum_{L>s>0, s\neq r} \sum_{p=1}^{n_s} \sum_{q=1}^{n_s} \Phi_{(s,0)(L,s):pq} \cdot [WW]^{(s,0)(L,s)}_{pq},$$

so

$$\sum_{p=1}^{n_r} \sum_{q=1}^{n_r} \Phi_{(r,0)(L,r):\pi_r(p)\pi_r(q)} \cdot \frac{d^{(r)}_{\pi_r(p)}}{d^{(r)}_{\pi_r(q)}} \cdot [WW]^{(r,0)(L,r)}_{pq}$$
$$= \sum_{p=1}^{n_r} \sum_{q=1}^{n_r} \Phi_{(r,0)(L,r):pq} \cdot [WW]^{(r,0)(L,r)}_{pq}$$

By Lemma B.3 and Corollary A.7, we have

$$\Phi_{(r,0)(L,r):\pi_r(p)\pi_r(q)} \cdot \frac{d^{(r)}_{\pi_r(p)}}{d^{(r)}_{\pi_r(q)}} = \Phi_{(r,0)(L,r):pq}.$$

So

$$\Phi_{(r,0)(L,r):pq} = 0,$$

for $p \neq q$, and

$$\Phi_{(r,0)(L,r):\pi_r(p)\pi_r(p)} = \Phi_{(r,0)(L,r):pp}.$$

In conclusion, we have $\Phi_{(L,0):pq}$ is arbitrary, and for $L > s > 0$,

$$\Phi_{(s,0)(L,s):\pi(p)\pi(p)} = \Phi_{(s,0)(L,s):pp}.$$

4. For all $L > s > t > 0$, and for all p, we have

$$\Phi_{(s,t)(t):p} = 0.$$

5. For all $L \geqslant s > 0$, $L > t \geqslant 0$ with $s \neq t$, we have

$$\Phi_{(s)(L,t):\pi_s(p)\pi_t(q)} \cdot \frac{d^{(s)}_{\pi_s(p)}}{d^{(t)}_{\pi_t(q)}} = \Phi_{(s)(L,t):pq}$$

If $(s,t) = (L,0)$, we have

$$\Phi_{(L)(L,0):pq} = \Phi_{(L)(L,0):pq}.$$

6. For all $L > t > 0$, we have

$$\sum_{p=1}^{n_L} \Phi_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_{p} + \sum_{p=1}^{n_t} \sum_{q=1}^{n_t} \Phi_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(t)(L,t)}_{pq}$$
$$= \sum_{p=1}^{n_L} \Phi_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_{p} + \sum_{p=1}^{n_t} \sum_{q=1}^{n_t} \Phi_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq}$$

which means $\Phi_{(L,t)(t):p}$ can be arbitrary. The rest is

$$\sum_{p=1}^{n_t} \sum_{q=1}^{n_t} \Phi_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} \cdot [bW]^{(t)(L,t)}_{pq} = \sum_{p=1}^{n_t} \sum_{q=1}^{n_t} \Phi_{(t)(L,t):pq} \cdot [bW]^{(t)(L,t)}_{pq}$$

By Lemma B.5 and Lemma A.7, we have

$$\Phi_{(t)(L,t):\pi_t(p)\pi_t(q)} \cdot \frac{d^{(t)}_{\pi_t(p)}}{d^{(t)}_{\pi_t(q)}} = \Phi_{(t)(L,t):pq}.$$

So

$$\Phi_{(t)(L,t):pq} = 0$$

for $p \neq q$, and

$$\Phi_{(t)(L,t):\pi_t(p)\pi_t(p)} = \Phi_{(t)(L,t):pp}.$$

In conclusion, for all $L > t > 0$, we have $\Phi_{(L,t)(t):p}$ is arbitrary, and

$$\Phi_{(t)(L,t):\pi(p)\pi(p)} = \Phi_{(t)(L,t):pp}$$

7. For all $L \geqslant s > 0$ and for all p, we have

$$\Phi_{(s):\pi_s(p)} \cdot d^{(s)}_{\pi_s(p)} = \Phi_{(s):p}.$$

If $s = L$, we have

$$\Phi_{(L):p} = \Phi_{(L):p}.$$

8. We have

$$\Phi_1 = \Phi_1.$$

D.4 G-INVARIANT POLYNOMIAL LAYERS

Based on the above discussions, we conclude that every G-invariant polynomial layer, which is defined as a linear combination of stable polynomial terms, is given as:

$$\begin{aligned}
I(U) =& \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L,0)(L,0):pq} \cdot [WW]^{(L,0)(L,0)}_{pq} + \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L,0):pq} \cdot [W]^{(L,0)}_{pq} \\
&+ \sum_{L>s>0}\sum_{p=1}^{n_s} \Phi_{(s,0)(L,s):\bullet\bullet} \cdot [WW]^{(s,0)(L,s)}_{pp} + \sum_{p=1}^{n_L}\sum_{q=1}^{n_0} \Phi_{(L)(L,0):pq} \cdot [bW]^{(L)(L,0)}_{pq} \\
&+ \sum_{L>t>0}\sum_{p=1}^{n_L} \Phi_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_{p} + \sum_{L>t>0}\sum_{p=1}^{n_t} \Phi_{(t)(L,t):\bullet\bullet} \cdot [bW]^{(t)(L,t)}_{pp} \\
&+ \sum_{p=1}^{n_L} \Phi_{(L):p} \cdot [b]^{(L)}_{p} + \Phi_1
\end{aligned}$$

In the above formula, the bullet $\bullet$ indicates that the corresponding coefficient is independent of the index at the bullet.

E ADDITIONAL EXPERIMENTAL DETAILS

E.1 PREDICTING GENERALIZATION FROM WEIGHTS

Dataset. The Tanh subset from the CNN Zoo dataset has 5,949 training instances and 1,488 testing instances, while the original ReLU subset consists of 6,050 training instances and 1,513 testing instances. We do the augmentation for ReLU subset, with an augmentation factor of 2, effectively doubling the size of the dataset by adding one augmented version of each original instance. The overall dataset sizes, including both the original and augmented data, are summarized in Table 5.

Table 5: Datasets information for predicting generalization task.

Dataset	Train size	Val size
Original ReLU	6050	1513
Augment ReLU	12100	3026
Tanh	5949	1488

Table 6: Number of parameters of all models for prediciting generalization task.

Model	ReLU dataset	Tanh dataset
STATNN	1.06M	1.06M
NP	2.03M	2.03M
HNP	2.81M	2.81M
Monomial-NFN	0.25M	1.41M
MAGEP-NFN (ours)	0.99M	0.99M

Table 7: Hyperparameters for Monomial-NFN on prediciting generalization task.

MLP hidden	Loss	Optimizer	Learning rate	Batch size	Epoch
500	Binary cross-entropy	Adam	0.001	8	50

Table 8: Dataset size for Classifying INRs task.

	Train	Validation	Test
CIFAR-10	45000	5000	10000
MNIST size	45000	5000	10000
Fashion-MNIST	45000	5000	20000

Baselines In this experiment, we compare our model with five other baselines:

- **STATNN** (Unterthiner et al., 2021): utilizes statistical features of the weights and biases
- **Graph-NN** (Kofinas et al., 2024): represents input network parameters as graphs and processes using Graph Neural Networks.
- **NP and HNP** (Zhou et al., 2024b): incoporates the permutation symmetries of neurons into neural functional networks.
- **Monomial-NFN** (Tran et al., 2024): extends the group action on weights from group of permutation matrices to the group of monomial matrices by incoporates scaling/sign-flipping symmetries.

Implementation Details. Our architecture begins with a Monomial-NFN layer featuring 20 channels, aligning the dimensions of the weights and biases. This is followed by two equivariant MAGEP-NFN layers, each with 20 channels, using either a ReLU activation for the ReLU dataset or a Tanh activation for the Tanh dataset. Next, the output is processed by a MAGEP-NFN Invariant layer. The final output from this layer is flattened and mapped to $\mathbb{R}^{500}$. This vector is further processed by a fully connected MLP with two hidden layers, both activated by ReLU. The final output undergoes a linear projection to a scalar, followed by a sigmoid function. The model is trained using Binary Cross Entropy (BCE) loss over 50 epochs, with early stopping determined by a threshold τ on the validation set. The entire training process on an A100 GPU takes 30 minutes. A summary of hyperparameters can be found in Table 7.

For the baseline models, we adhere to the original implementations as outlined in Zhou et al. (2024b), utilizing the official code (available at: https://github.com/AllanYangZhou/nfn), and Tran

Table 9: Hyperparameters of Monomial-NFN for each dataset in Classify INRs task.

	MNIST	Fashion-MNIST	CIFAR-10
MAGEP-NFN hidden dimension	128x3	64	64 x 3
Base model	MAGEP-Inv	NP	MAGEP-Inv
Base model hidden dimension	128	128x3	64
MLP hidden neurons	1000	500	1000
Dropout	0.1	0.1	0.1
Learning rate	0.0001	0.0001	0.0001
Batch size	32	32	32
Step	200000	200000	200000
Loss	Binary cross-entropy	Binary cross-entropy	Binary cross-entropy

Table 10: Number of parameters of all models for classifying INRs task.

	CIFAR-10	MNIST	Fashion-MNIST
MLP	2M	2M	2M
NP	16M	15M	15M
HNP	42M	22M	22M
Monomial-NFN	16M	22M	20M
MAGEP-NFN (ours)	3.4M	4.1M	4.9M

Table 11: Number of parameters of all models for Weight space style editing task.

Model	Number of parameters
MLP	4.5M
NP	4.1M
HNP	12.8M
Monomial-NFN	4.1M
MAGEP-NFN (ours)	4.1M

Table 12: Hyperparameters for Monomial-NFN on weight space style editing task.

Name	Value
MAGEP-NFN hidden dimension	16
NP dimension	128
Optimizer	Adam
Learning rate	0.001
Batch size	32
Steps	50000

et al. (2024). In the HNP, NP, and Monomial-NFN models, we employ three equivariant layers with channel configurations of 16, 16, and 5, respectively. The extracted features are then passed through an average pooling layer, followed by three MLP layers with hidden dimensions of 200 (Monomial-NFN ReLU case) and 1000 neurons (Monomial-NFN Tanh case and other models). The hyperparameters for our model, along with the parameter counts for all models involved in this task, are detailed in Table 6.

E.2 Classifying implicit neural representations of images

Dataset. We use the original INRs dataset, which contained three dataset: CIFAR-10, MNIST and Fashion-MNIST, obtained by applying a single SIREN model to every image. The detailed information about dataset is described in Zhou et al. (2024b). We use the datasett without any data augmentation as in the settings of Tran et al. (2024). The breakdown of training, validation, and test sample sizes for each dataset is detailed in Table 8.

Implementation Details. In these experiments, we use two different architectures. For both the MNIST and CIFAR datasets, the architecture begins with a Monomial-NFN layer to adjust the weight dimensions, followed by three MAGEP-NFN layers, each utilizing sine activation. The resulting weight features are then passed through a MAGEP Invariant layer. Finally, the output is flattened and processed by an MLP with two hidden layers, each containing 1,000 units and using ReLU activations.

For the Fashion-MNIST dataset, we begin with a Monomial-NFN layer with sine activation, followed by a MAGEP-NFN layer also utilizing sine activation, and then a Monomial-NFN layer with absolute activation. The architecture then aligns with the design of the NP model from Zhou et al. (2024b). Specifically, a Gaussian Fourier Transformation is applied to encode the input into sine and cosine components, mapping from 1 dimension to 256 dimensions. The encoded features are processed through IOSinusoidalEncoding, a positional encoding tailored for the NP layer, which uses a maximum frequency of 10 and 6 frequency bands. Following this, the features pass through three NP layers with ReLU activations. An average pooling is applied, after which the output is flattened, and the resulting vector is processed by an MLP with two hidden layers, each containing 1,000 units and using ReLU activations. Finally, the output is linearly projected to a scalar. We employ the Binary Cross Entropy (BCE) loss function and train the model for 200,000 steps, taking approximately 2 hours on an A100 GPU. The parameter counts for all models are presented in Table 10

E.3 Weight space style editing

Dataset. We utilize the same INRs dataset as employed in the classification task, with the sizes of the training, validation, and test sets provided in Table 8.

Implementation Details. In these experiments, our architecture begins with two MAGEP-NFN layers, each with 16 hidden dimensions. The rest of the design follows the NP model outlined in Zhou et al. (2024b). Specifically, we apply a Gaussian Fourier Transformation with a mapping size of 256, followed by IOSinusoidalEncoding. The features are then processed through three NP layers, each with 128 hidden dimensions and ReLU activation. The final output is passed through an NP layer for scalar projection and a LearnedScale layer as described in the Appendix of Zhou et al. (2024b). We use the Binary Cross Entropy (BCE) loss function and train the model for 50,000 steps, which takes approximately 35 minutes on an A100 GPU.

For the baseline models, we maintain the same settings as the official implementation. Specifically, the HNP or NP model consists of three layers, each with 128 hidden dimensions, followed by ReLU activations. An NFN of the same type is applied to map the output to one dimension, after which it is processed by a LearnedScale layer. The number of parameters for all models is detailed in Table 11, and the hyperparameters for our model are presented in Table 12.

F Implementation of Equivariant and Invariant Layers

We provide the multi-channel implementations of the $\mathcal{G}_{\mathcal{U}}$-equivariant map $E\colon \mathcal{U}^d \to \mathcal{U}^e$ and the $\mathcal{G}_{\mathcal{U}}$-invariant map $I\colon \mathcal{U}^d \to \mathbb{R}^{e\times d'}$. For uniformity in implementing Equivariant and Invariant layers from Appendix C.5 and Appendix D.4, we employ einops-style pseudocode as a consistent framework.

We summarize the key dimensions in Table 13 and outline the shapes of the input terms in Table 14.

Table 13: Summary of key dimensions involved in the implementation

Symbol	Description
d	Number of input channels for the equivariant and invariant layer
e	Number of output channels for the equivariant and invariant layer
b	Batch size
n_i	Number of channels at the $i_t h$ layer
d'	Embedding dimension of the invariant layer's output

Table 14: Shapes of input terms used in the implementation

Term	Shape
$[W]^{(s,t)}$	$[b, d, n_s, n_t]$
$[Wb]^{(s,t)(t)}$	$[b, d, n_s]$
$[bW]^{(s)(L,t)}$	$[b, d, n_s, n_t]$
$[WW]^{(s,0)(L,t)}$	$[b, d, n_s, n_t]$

F.1 Equivariant Layers Pseudocode

F.1.1 Pseudocode for case $i = L$

From the formula for $[E(W)]_{jk}^{(L)}$:

$$[E(W)]_{jk}^{(L)} = \sum_{p=1}^{n_L} \Phi_{(L,L-1):p}^{(L):j} \cdot [W]_{pk}^{(L,L-1)} + \sum_{p=1}^{n_L} \Phi_{(L,0)(L,L-1):p}^{(L):j} \cdot [WW]_{pk}^{(L,0)(L,L-1)} + \sum_{p=1}^{n_L} \Phi_{(L)(L,L-1):p}^{(L):j} \cdot [bW]_{pk}^{(L)(L,L-1)}$$

We define the pseudocode for each term:

For $\Phi_{(L,L-1):p}^{(L):j} \cdot [W]_{pk}^{(L,L-1)}$,

with $[W]_{pk}^{(L,L-1)}$ of shape $[b, d, n_L, n_{L-1}]$ and $\Phi_{(L,L-1):p}^{(L):j}$ of shape $[e, d, n_L, n_L]$,

Corresponding pseudocode: $\texttt{einsum}(edpj, bdpk \to bejk)$

For $\Phi_{(L,0)(L,L-1):p}^{(L):j} \cdot [WW]_{pk}^{(L,0)(L,L-1)}$,

with $[WW]_{pk}^{(L,0)(L,L-1)}$ of shape $[b, d, n_L, n_{L-1}]$ and $\Phi_{(L,0)(L,L-1):p}^{(L):j}$ of shape $[e, d, n_L, n_L]$,

Corresponding pseudocode: $\texttt{einsum}(edpj, bdpk \to bejk)$

For $\Phi_{(L)(L,L-1):p}^{(L):j} \cdot [bW]_{pk}^{(L)(L,L-1)}$,

with $[bW]_{pk}^{(L)(L,L-1)}$ of shape $[b, d, n_L, n_{L-1}]$ and $\Phi_{(L)(L,L-1):p}^{(L):j}$ of shape $[e, d, n_L, n_L]$,

Corresponding pseudocode: $\texttt{einsum}(edpj, bdpk \to bejk)$

From the formula for $[E(b)]_j^{(L)}$:

$$[E(b)]_j^{(L)} = \sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi_{(L,0)(L,0):pq}^{(L):j} \cdot [WW]_{pq}^{(L,0)(L,0)} + \sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi_{(L,0):pq}^{(L):j} \cdot [W]_{pq}^{(L,0)} + \sum_{L>s>0} \sum_{p=1}^{n_s} \Phi_{(s,0)(L,s):}^{(L):j} \cdot [WW]_{pp}^{(s,0)(L,s)} + \sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi_{(L)(L,0):pq}^{(L):j} \cdot [bW]_{pq}^{(L)(L,0)}$$

$$+\sum_{L>t>0}\sum_{p=1}^{n_L}\Phi^{(L):j}_{(L,t)(t):p}\cdot[Wb]^{(L,t)(t)}_{p}+\sum_{L>t>0}\sum_{p=1}^{n_t}\Phi^{(L):j}_{(L)(L,t):}\cdot[bW]^{(t)(L,t)}_{pp}$$
$$+\sum_{p=1}^{n_L}\Phi^{(L):j}_{(L):p}\cdot[b]^{(L)}_{p}+\Phi^{(L):j}_{1}$$

We define the pseudocode for each term:

For $\Phi^{(L):j}_{(L,0)(L,0):pq}\cdot[WW]^{(L,0)(L,0)}_{pq}$,
with $[WW]^{(L,0)(L,0)}_{pq}$ of shape $[b,d,n_L,n_0]$ and $\Phi^{(L):j}_{(L,0)(L,0):pq}$ of shape $[e,d,n_L,n_0,n_L]$,
Corresponding pseudocode: $\texttt{einsum}(edpqj,bdpq\to bej)$

For $\Phi^{(L):j}_{(L,0):pq}\cdot[W]^{(L,0)}_{pq}$, with $[W]^{(L,0)}_{pq}$ of shape $[b,d,n_L,n_0]$ and $\Phi^{(L):j}_{(L,0):pq}$ of shape $[e,d,n_L,n_0,n_L]$,
Corresponding pseudocode: $\texttt{einsum}(edpqj,bdpq\to bej)$

For $\Phi^{(L):j}_{(s,0)(L,s):}\cdot[WW]^{(s,0)(L,s)}_{pp}$,
with $[WW]^{(s,0)(L,s)}_{pp}$ of shape $[b,d,n_s,n_s]$ and $\Phi^{(L):j}_{(s,0)(L,s):}$ of shape $[e,d,n_s,n_s,n_L]$,
Corresponding pseudocode: $\texttt{einsum}(edppj,bdpp\to bej)$

For $\Phi^{(L):j}_{(L)(L,0):pq}\cdot[bW]^{(L)(L,0)}_{pq}$,
with $[bW]^{(L)(L,0)}_{pq}$ of shape $[b,d,n_L,n_0]$ and $\Phi^{(L):j}_{(L)(L,0):pq}$ of shape $[e,d,n_L,n_0,n_L]$,
Corresponding pseudocode: $\texttt{einsum}(edpqj,bdpq\to bej)$

For $\Phi^{(L):j}_{(L,t)(t):p}\cdot[Wb]^{(L,t)(t)}_{p}$, with $[Wb]^{(L,t)(t)}_{p}$ of shape $[b,d,n_L]$ and $\Phi^{(L):j}_{(L,t)(t):p}$ of shape $[e,d,n_L,n_L]$,
Corresponding pseudocode: $\texttt{einsum}(edpj,bdp\to bej)$

For $\Phi^{(L):j}_{(t)(L,t):}\cdot[bW]^{(t)(L,t)}_{pp}$, with $[bW]^{(t)(L,t)}_{pp}$ of shape $[b,d,n_t,n_t]$ and $\Phi^{(L):j}_{(t)(L,t):}$ of shape $[e,d,n_t,n_t,n_L]$,
Corresponding pseudocode: $\texttt{einsum}(edppj,bdpp\to bej)$

For $\Phi^{(L):j}_{(L):p}\cdot[b]^{(L)}_{p}$, with $[b]^{(L)}_{p}$ of shape $[b,d,n_L]$ and $\Phi^{(L):j}_{(L):p}$ of shape $[e,d,n_L,n_L]$,
Corresponding pseudocode: $\texttt{einsum}(edpj,bdp\to bej)$

For $\Phi^{(L):j}_{1}$ of shape $[e,n_L]$,
Corresponding pseudocode: $\texttt{einsum}(ej,\to ej).unsqueeze(0)$

F.1.2 Pseudocode for case $i=1$

From the formula for $[E(W)]^{(1)}_{jk}$:

$$[E(W)]^{(1)}_{jk}=\sum_{q=1}^{n_0}\Phi^{(1):\bullet k}_{(1,0):\bullet q}\cdot[W]^{(1,0)}_{jq}+\sum_{q=1}^{n_0}\Phi^{(1):\bullet k}_{(1,0)(L,0):\bullet q}\cdot[WW]^{(1,0)(L,0)}_{jq}$$
$$+\sum_{q=1}^{n_0}\Phi^{(1):\bullet k}_{(1)(L,0):\bullet q}\cdot[bW]^{(1)(L,0)}_{jq}+\Phi^{(1):\bullet k}_{(1):\bullet}\cdot[b]^{(1)}_{j}$$

We define the pseudocode for each term:

For $\Phi^{(1):\bullet k}_{(1,0):\bullet q}\cdot[W]^{(1,0)}_{jq}$, with $[W]^{(1,0)}_{jq}$ of shape $[b,d,n_1,n_0]$ and $\Phi^{(1):\bullet k}_{(1,0):\bullet q}$ of shape $[d,e,n_0,n_0]$,
Corresponding pseudocode: $\texttt{einsum}(bdjq,deqk\to bejk)$

For $\Phi^{(1):\bullet k}_{(1,0)(L,0):\bullet q}\cdot[WW]^{(1,0)(L,0)}_{jq}$,
with $[WW]^{(1,0)(L,0)}_{jq}$ of shape $[b,d,n_1,n_0]$ and $\Phi^{(1):\bullet k}_{(1,0)(L,0):\bullet q}$ of shape $[d,e,n_0,n_0]$,
Corresponding pseudocode: $\texttt{einsum}(bdjq,deqk\to bejk)$

For $\Phi^{(1):\bullet k}_{(1)(L,0):\bullet q} \cdot [bW]^{(1)(L,0)}_{jq}$,

with $[bW]^{(1)(L,0)}_{jq}$ of shape $[b, d, n_1, n_0]$ and $\Phi^{(1):\bullet k}_{(1)(L,0):\bullet q}$ of shape $[d, e, n_0, n_0]$,

Corresponding pseudocode: $\texttt{einsum}(bdjq, deqk \to bejk)$

For $\Phi^{(1):\bullet k}_{(1):\bullet} \cdot [b]^{(1)}_j$, with $[b]^{(1)}_j$ of shape $[b, d, n_1]$ and $\Phi^{(1):\bullet k}_{(1):\bullet}$ of shape $[d, e, n_0]$,

Corresponding pseudocode: $\texttt{einsum}(bdj, dek \to bejk)$

From the formula for $[E(b)]^{(1)}_j$:

$$[E(b)]^{(1)}_j = \sum_{q=1}^{n_0} \Phi^{(1):\bullet}_{(1,0):\bullet q} \cdot [W]^{(1,0)}_{jq} + \sum_{q=1}^{n_0} \Phi^{(1):\bullet}_{(1,0)(L,0):\bullet q} \cdot [WW]^{(1,0)(L,0)}_{jq} + \sum_{q=1}^{n_0} \Phi^{(1):\bullet}_{(1)(L,0):\bullet q} \cdot [bW]^{(1)(L,0)}_{jq} + \Phi^{(1):\bullet}_{(1):\bullet} \cdot [b]^{(1)}_j$$

We define the pseudocode for each term:

For $\Phi^{(1):\bullet}_{(1,0):\bullet q} \cdot [W]^{(1,0)}_{jq}$, with $[W]^{(1,0)}_{jq}$ of shape $[b, d, n_1, n_0]$ and $\Phi^{(1):\bullet}_{(1,0):\bullet q}$ of shape $[d, e, n_0]$,

Corresponding pseudocode: $\texttt{einsum}(bdjq, deq \to bej)$

For $\Phi^{(1):\bullet}_{(1,0)(L,0):\bullet q} \cdot [WW]^{(1,0)(L,0)}_{jq}$,

with $[WW]^{(1,0)(L,0)}_{jq}$ of shape $[b, d, n_1, n_0]$ and $\Phi^{(1):\bullet}_{(1,0)(L,0):\bullet q}$ of shape $[d, e, n_0]$,

Corresponding pseudocode: $\texttt{einsum}(bdjq, deq \to bej)$

For $\Phi^{(1):\bullet}_{(1)(L,0):\bullet q} \cdot [bW]^{(1)(L,0)}_{jq}$, with $[bW]^{(1)(L,0)}_{jq}$ of shape $[b, d, n_1, n_0]$ and $\Phi^{(1):\bullet}_{(1)(L,0):\bullet q}$ of shape $[d, e, n_0]$,

Corresponding pseudocode: $\texttt{einsum}(bdjq, deq \to bej)$

For $\Phi^{(1):\bullet}_{(1):\bullet} \cdot [b]^{(1)}_j$, with $[b]^{(1)}_j$ of shape $[b, d, n_1]$ and $\Phi^{(1):\bullet}_{(1):\bullet}$ of shape $[d, e]$,

Corresponding pseudocode: $\texttt{einsum}(bdj, de \to bej)$

F.1.3 PSEUDOCODE FOR CASE $1 < i < L$

From the formula for $[E(W)]^{(i)}_{jk}$:

$$[E(W)]^{(i)}_{jk} = \left(\Phi^{(i):\bullet\bullet}_{(i,i-1):\bullet\bullet}\right) \cdot [W]^{(i,i-1)}_{jk} + \left(\Phi^{(i):\bullet\bullet}_{(i,0)(L,i-1):\bullet\bullet}\right) \cdot [WW]^{(i,0)(L,i-1)}_{jk} + \left(\Phi^{(i):\bullet\bullet}_{(i)(L,i-1):\bullet\bullet}\right) \cdot [bW]^{(i)(L,i-1)}_{jk}$$

We define the pseudocode for each term:

For $\Phi^{(i):\bullet\bullet}_{(i,i-1):\bullet\bullet} \cdot [W]^{(i,i-1)}_{jk}$, with $[W]^{(i,i-1)}_{jk}$ of shape $[b, d, n_i, n_{i-1}]$ and $\Phi^{(i):\bullet\bullet}_{(i,i-1):\bullet\bullet}$ of shape $[d, e]$,

Corresponding pseudocode: $\texttt{einsum}(bdjk, de \to bejk)$

For $\Phi^{(i):\bullet\bullet}_{(i,0)(L,i-1):\bullet\bullet} \cdot [WW]^{(i,0)(L,i-1)}_{jk}$,

with $[WW]^{(i,0)(L,i-1)}_{jk}$ of shape $[b, d, n_i, n_{i-1}]$ and $\Phi^{(i):\bullet\bullet}_{(i,0)(L,i-1):\bullet\bullet}$ of shape $[d, e]$,

Corresponding pseudocode: $\texttt{einsum}(bdjk, de \to bejk)$

For $\Phi^{(i):\bullet\bullet}_{(i)(L,i-1):\bullet\bullet} \cdot [bW]^{(i)(L,i-1)}_{jk}$,

with $[bW]^{(i)(L,i-1)}_{jk}$ of shape $[b, d, n_i, n_{i-1}]$ and $\Phi^{(i):\bullet\bullet}_{(i)(L,i-1):\bullet\bullet}$ of shape $[d, e]$,

Corresponding pseudocode: $\texttt{einsum}(bdjk, de \to bejk)$

From the formula for $[E(b)]^{(i)}_j$:

$$[E(b)]^{(i)}_j = \sum_{q=1}^{n_0} \left(\Phi^{(i):\bullet}_{(i,0):\bullet q}\right) \cdot [W]^{(i,0)}_{jq} + \sum_{q=1}^{n_0} \left(\Phi^{(i):\bullet}_{(i,0)(L,0):\bullet q}\right) \cdot [WW]^{(i,0)(L,0)}_{jq}$$

$$+ \sum_{i>t>0} \left(\Phi^{(i):\bullet}_{(i,t)(t):\bullet} \right) \cdot [Wb]^{(i,t)(t)}_{j} + \sum_{q=1}^{n_t} \left(\Phi^{(i):\bullet}_{(i)(L,0):\bullet q} \right) \cdot [bW]^{(i)(L,0)}_{jq}$$
$$+ \left(\Phi^{(i):\bullet}_{(i):\bullet} \right) \cdot [b]^{(i)}_{j}$$

We define the pseudocode for each term:

For $\Phi^{(i):\bullet}_{(i,0):\bullet q} \cdot [W]^{(i,0)}_{jq}$, with $[W]^{(i,0)}_{jq}$ of shape $[b, d, n_i, n_0]$ and $\Phi^{(i):\bullet}_{(i,0):\bullet q}$ of shape $[d, e, n_0]$,
Corresponding pseudocode: $\texttt{einsum}(bdjq, deq \rightarrow bej)$

For $\Phi^{(i):\bullet}_{(i,0)(L,0):\bullet q} \cdot [WW]^{(i,0)(L,0)}_{jq}$,
with $[WW]^{(i,0)(L,0)}_{jq}$ of shape $[b, d, n_i, n_0]$ and $\Phi^{(i):\bullet}_{(i,0)(L,0):\bullet q}$ of shape $[d, e, n_0]$,
Corresponding pseudocode: $\texttt{einsum}(bdjq, deq \rightarrow bej)$

For $\Phi^{(i):\bullet}_{(i,t)(t):\bullet} \cdot [Wb]^{(i,t)(t)}_{j}$, with $[Wb]^{(i,t)(t)}_{j}$ of shape $[b, d, n_i]$ and $\Phi^{(i):\bullet}_{(i,t)(t):\bullet}$ of shape $[d, e]$,
Corresponding pseudocode: $\texttt{einsum}(bdj, de \rightarrow bej)$

For $\Phi^{(i):\bullet}_{(i)(L,0):\bullet q} \cdot [bW]^{(i)(L,0)}_{jq}$, with $[bW]^{(i)(L,0)}_{jq}$ of shape $[b, d, n_i, n_0]$ and $\Phi^{(i):\bullet}_{(i)(L,0):\bullet q}$ of shape $[d, e, n_0]$,
Corresponding pseudocode: $\texttt{einsum}(bdjq, deq \rightarrow bej)$

For $\Phi^{(i):\bullet}_{(i):\bullet} \cdot [b]^{(i)}_{j}$, with $[b]^{(i)}_{j}$ of shape $[b, d, n_i]$ and $\Phi^{(i):\bullet}_{(i):\bullet}$ of shape $[d, e]$,
Corresponding pseudocode: $\texttt{einsum}(bdj, de \rightarrow bej)$

F.2 Invariant Layers Pseudocode

From the formula for the Invariant layer $I(U)$:

$$I(U) = \sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi_{(L,0)(L,0):pq} \cdot [WW]^{(L,0)(L,0)}_{pq} + \sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi_{(L,0):pq} \cdot [W]^{(L,0)}_{pq}$$
$$+ \sum_{L>s>0} \sum_{p=1}^{n_s} \Phi_{(s,0)(L,s):\bullet\bullet} \cdot [WW]^{(s,0)(L,s)}_{pp} + \sum_{p=1}^{n_L} \sum_{q=1}^{n_0} \Phi_{(L)(L,0):pq} \cdot [bW]^{(L)(L,0)}_{pq}$$
$$+ \sum_{L>t>0} \sum_{p=1}^{n_L} \Phi_{(L,t)(t):p} \cdot [Wb]^{(L,t)(t)}_{p} + \sum_{L>t>0} \sum_{p=1}^{n_t} \Phi_{(t)(L,t):\bullet\bullet} \cdot [bW]^{(t)(L,t)}_{pp}$$
$$+ \sum_{p=1}^{n_L} \Phi_{(L):p} \cdot [b]^{(L)}_{p} + \Phi_1$$

We define the pseudocode for each term:

For $\Phi_{(L,0)(L,0):pq} \cdot [WW]^{(L,0)(L,0)}_{pq}$,
with $[WW]^{(L,0)(L,0)}_{pq}$ of shape $[b, d, n_L, n_0]$ and $\Phi_{(L,0)(L,0):pq}$ of shape $[d, e, n_L, n_0, d']$,
Corresponding pseudocode: $\texttt{einsum}(bdpq, depqk \rightarrow bek)$

For $\Phi_{(L,0):pq} \cdot [W]^{(L,0)}_{pq}$, with $[W]^{(L,0)}_{pq}$ of shape $[b, d, n_L, n_0]$ and $\Phi_{(L,0):pq}$ of shape $[d, e, n_L, n_0, d']$,
Corresponding pseudocode: $\texttt{einsum}(bdpqk, depqk \rightarrow bek)$

For $\Phi_{(s,0)(L,s):\bullet\bullet} \cdot [WW]^{(s,0)(L,s)}_{pp}$,
with $[WW]^{(s,0)(L,s)}_{pp}$ of shape $[b, d, n_s]$ and $\Phi_{(s,0)(L,s):\bullet\bullet}$ of shape $[d, e, d']$,
Corresponding pseudocode: $\texttt{einsum}(bdpk, dek \rightarrow bek)$

For $\Phi_{(L)(L,0):pq} \cdot [bW]^{(L)(L,0)}_{pq}$,
with $[bW]^{(L)(L,0)}_{pq}$ of shape $[b, d, n_L, n_0]$ and $\Phi_{(L)(L,0):pq}$ of shape $[d, e, n_L, n_0, d']$,

Corresponding pseudocode: $\texttt{einsum}(bdpqk, depqk \rightarrow bek)$

For $\Phi_{(L,t)(t):p} \cdot [Wb]_p^{(L,t)(t)}$, with $[Wb]_p^{(L,t)(t)}$ of shape $[b, d, n_L]$ and $\Phi_{(L,t)(t):p}$ of shape $[d, e, n_L, d']$,

Corresponding pseudocode: $\texttt{einsum}(bdpk, deijk \rightarrow bek)$

For $\Phi_{(t)(L,t):\bullet\bullet} \cdot [bW]_{pp}^{(t)(L,t)}$, with $[bW]_{pp}^{(t)(L,t)}$ of shape $[b, d, n_t]$ and $\Phi_{(t)(L,t):\bullet\bullet}$ of shape $[d, e, d']$,

Corresponding pseudocode: $\texttt{einsum}(bdpk, dek \rightarrow bek)$

For $\Phi_{(L):p} \cdot [b]_p^{(L)}$, with $[b]_p^{(L)}$ of shape $[b, d, n_L]$ and $\Phi_{(L):p}$ of shape $[d, e, n_L, d']$,

Corresponding pseudocode: $\texttt{einsum}(bdpk, depk \rightarrow bek)$

For $\Phi_1$of shape $[e, d']$,

Corresponding pseudocode: $\texttt{einsum}(ek \rightarrow ek).unsqueeze(0)$

G PERFORMANCE COMPARISON WITH GRAPH-BASED NFNS

Experiment Setup: Following the same experiment setup in Appendix E.1, we compare the predictive performance of our model and two graph-based baselines: GNN (Kofinas et al., 2024) and ScaleGMN (Kalogeropoulos et al., 2024), using HNP (Zhou et al., 2024b) as a reference.

Results: The results are presented in Table F.2. The GNN model exhibits a noticeable performance decline when tested on separate activation subsets. Although ScaleGMN significantly improves performance on the Tanh subset, its enhancements on the ReLU subset are comparatively modest. In contrast, our model demonstrates substantial overall improvements across both datasets, highlighting its effectiveness.

Table 15: Performance comparison with Graph-based NFNs on Small CNN Zoo task.

	ReLU subset	Tanh subset
HNP (Zhou et al., 2024b)	0.897	0.934
GNN (Kofinas et al., 2024)	0.897	0.893
ScaleGMN (Kalogeropoulos et al., 2024)	0.928	**0.942**
MAGEP-NFNs (ours)	**0.933**	0.940

H RUNTIME COMPARISON WITH GRAPH-BASED NFNS.

Table 16: Runtime of models on Small CNN Zoo task.

	ReLU subset	Tanh subset
NP (Zhou et al., 2024b)	36m40s	35m34s
HNP (Zhou et al., 2024b)	30m06s	29m37s
GNN (Kofinas et al., 2024)	4h27m29s	4h25m17s
ScaledGMN (Kalogeropoulos et al., 2024)	1h20m	1h20m
Monomial-NFN (Tran et al., 2024)	**23m47s**	**18m23s**
MAGEP-NFNs (ours)	28m43s	28m12s

Table 17: Memory consumption of models on Small CNN Zoo task.

	ReLU subset	Tanh subset
NP (Zhou et al., 2024b)	838MB	838MB
HNP (Zhou et al., 2024b)	856MB	856MB
GNN (Kofinas et al., 2024)	6390MB	6390MB
ScaledGMN (Kalogeropoulos et al., 2024)	2918MB	2918MB
Monomial-NFN (Tran et al., 2024)	**560MB**	**582MB**
MAGEP-NFNs (ours)	584MB	584MB

Table 16 and 17 provide runtime and memory consumption data for our model and other baselines in predicting CNN generalization task. For graph-based architectures, we compare with two recent works: GNN (Kofinas et al., 2024) and ScaleGMN (Kalogeropoulos et al., 2024). Our model runs significantly faster and uses much less memory than these graph-based networks and NP/HNP (Zhou et al., 2024b). Introducing additional polynomial terms slightly increases our model's runtime and memory usage compared to Monomial-NFN (Tran et al., 2024). However, this trade-off results in considerably enhanced expressivity, which is evident across many tasks like Predict CNN Generalization or INRs Classification.

I ABLATION STUDY ON THE IMPORTANCE OF COMPONENTS

We conduct an ablation study to evaluate the significance of the components introduced in our work. Specifically, we categorize the terms as follows:

- **Non Inter-Layer Terms**: These are terms that involve only the mapping of Non-inter-layer weights and biases, $(W^l, b^l)^{l=1,\cdots,L}$, to the output weight space, consistent with prior works (DWSNet(De Luigi et al., 2023), NP, HNP (Zhou et al., 2024b), Monomial-NFN(Tran et al., 2024)) on neural functional network layers
- **Inter-Layer Terms**: These are the novel terms introduced in our paper, $([W], [WW], [bW], [Wb])$, designed to capture relationships between weights and biases across multiple layers.

Experiment Setup: To assess the impact of these terms, we perform experiments on the invariant task of predicting CNN generalization on ReLU subset (follow the same setting as in Appendix E.1, using the architecture specified in Equation 15. The results of our experiments are presented in the table below.

Table 18: Ablation study assessing the contribution of each component introduced in our work, conducted on the task of predicting CNN generalization on the ReLU subset.

Components	Kendall's τ
Only Non Inter-Layer terms	0.929
Only Inter-Layer terms	<u>0.932</u>
Non Inter-Layer terms + $[W]$	0.930
Non Inter-Layer terms + $[WW]$	0.930
Non Inter-Layer terms + $[Wb]$	0.931
Non Inter-Layer terms + $[bW]$	0.931
Non Inter-Layer terms + Inter-Layer terms	**0.933**

Results: Each newly introduced Inter-Layer terms provide additional information to the network, and when combined with the Non Inter-Layer terms, the performance is boosted considerably, reaching 0.933 in Kendall's τ.