A Applications of Sign Equivariance

A.1 Improving Invariant Eigenvector Networks

Neural networks that are invariant to eigenvector symmetries have been shown to empirically improve graph learning models and achieve theoretically high expressive power. SignNet [Lim et al., 2023], a sign invariant neural network, takes the form

$$f(v_1, \ldots, v_k) = \rho(\phi(v_1) + \phi(-v_1), \ldots, \phi(v_k) + \phi(-v_k)) \tag{8}$$

for neural networks ρ and ϕ. This directly enforces invariant representations, without any intermediate equivariant representations. However, many successful invariant models first have many equivariant layers before a final invariant operation as equivariant layers are more expressive: this includes convolutional neural networks [LeCun et al., 1989], message passing graph neural networks [Gilmer et al., 2017], invariant graph networks [Maron et al., 2018], and group convolutional neural networks [Cohen and Welling, 2016]. Thus, sign equivariant layers may lead to better sign invariant networks. Moreover, sign equivariant layers may improve on other aspects of SignNet, such as expressiveness of node features (Proposition 1) and efficiency (Appendix A.2)

A.2 Efficiency Gains from Sign Equivariant Networks

Here, we show that our sign equivariant models can reduce the complexity of equivariant or invariant networks for two different types of applications. Throughout, we consider functions $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$, and we consider our permutation equivariant and sign equivariant DSS-based architecture from Section 3.4.

The time cost (in floating point operations) per layer of our DSS-based model is $\mathcal{O}(n(kd + d^2))$, where d is the maximum hidden dimension of the MLP and we assume constant depth MLPs. To see this, note that we can precompute $\sum_{j=1}^n V_{j,:}$, so that each $\sum_{j\neq i} V_{j,:}$ can be computed in constant time by subtracting $V_{i,:}$ from the total sum. Then for each of the n rows, the MLPs require $\mathcal{O}(kd+d^2)$ to evaluate matrix multiplications. In this process, we only form tensors of size $\mathcal{O}(n(k + d))$, as the inputs and outputs are of size $\mathcal{O}(nk)$, and the hidden layers of the MLPs form tensors of size $\mathcal{O}(nd)$.

A.2.1 Efficient Orthogonally Equivariant Networks

Consider the case of $O(k)$ equivariant models $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$ such that $f(XQ) = f(X)Q$ for all orthogonal matrices $Q \in O(k)$. There are many orthogonally equivariant neural architectures that are specialized to the special case of $k = 3$, which is very useful for applications in the physical sciences [Thomas et al., 2018, Fuchs et al., 2020]. Here we consider models that directly work for general dimension k.

Frame averaging approaches [Puny et al., 2022, Atzmon et al., 2022] require 2^k forward passes of a base network f_θ, one for each sign flip of the principal components. Letting their base network be a permutation equivariant DeepSets [Zaheer et al., 2017], this means that they require $\mathcal{O}(n(kd + d^2)2^k)$ time to evaluate their model, where d is the hidden dimension of the base model. Note that this has an extra exponential 2^k factor compared to our $\mathcal{O}(n(kd + d^2))$ cost.

Another general approach with universality guarantees comes from Villar et al. [2021], who analyze invariant polynomials to develop equivariant architectures. However, their method for $O(k)$ invariance or equivariance requires forming $XX^\top$, an $n \times n$ matrix. Thus, the complexity is at least $\mathcal{O}(n^2)$, which is a problem in applications, since oftentimes n is much larger than k. Variants of their method do not need to compute all $\mathcal{O}(n^2)$ inner products, but it is unclear how to maintain permutation equivariance when doing this.

A.2.2 Efficient Sign Invariant Networks

Consider again the form of SignNet [Lim et al., 2023], $f(V) = \rho([\phi(v_i) + \phi(-v_i)]_{i=1,\ldots,k})$. In the permutation equivariant version, e.g. when ϕ is a DeepSets [Zaheer et al., 2017] or a message passing neural network [Gilmer et al., 2017], ϕ maps from $\mathbb{R}^n \to \mathbb{R}^{n\times d}$, where d is the hidden dimension. Thus, computing $\phi(v_i) + \phi(-v_i)$ for all k vectors v_i require an $\mathcal{O}(nkd)$ sized tensor to be formed (even if the output space of ϕ is $\mathbb{R}^n$, a vectorized implementation computes all $\phi(v_i) + \phi(-v_i)$ in two batched inference calls to ϕ, which would require $\mathcal{O}(nkd)$ sized intermediate tensors). This is a multiplicative factor larger than the sign equivariant requirement of $\mathcal{O}(n(k+d))$ sized tensors. Moreover, it would take $\mathcal{O}(nkd^2)$ time to compute $\phi(v_i)+\phi(-v_i)$ for each i, which is a multiplicative factor larger than the $\mathcal{O}(n(kd+d^2))$ time for the sign equivariant architecture.

A.3 Edge Representations and Link Prediction

A.3.1 Sign Invariant Link Prediction Decoders

Here, we present an ansatz for universal permutation invariant and sign invariant functions for $n=2$, that is $f:\mathbb{R}^{2\times k}\to\mathbb{R}^{d_{\text{out}}}$. Note that SignNet is only known to be universal for such functions for $n=1$, where there are no permutation symmetries [Lim et al., 2023].

We will parameterize such functions as

$$f(v_1,\dots v_k)=\varphi\left(v_1\odot v_1, v_1\odot \operatorname{rev}(v_1),\dots,v_k\odot v_k, v_k\odot\operatorname{rev}(v_k)\right). \tag{9}$$

Here, rev : $\mathbb{R}^2\to\mathbb{R}^2$ reverses the vector, so $\operatorname{rev}(a)_1=a_2$ and $\operatorname{rev}(a)_2=a_1$. Moreover, $\varphi:\mathbb{R}^{2\times 2k}\to\mathbb{R}^{d_{\text{out}}}$ is a permutation invariant neural network, so $\varphi(PX)=\varphi(X)$ for all 2×2 permutation matrices P. Note that it is easy to parameterize permutation invariant functions φ in a maximally expressive way, e.g. via DeepSets [Zaheer et al., 2017]. Now, we show that this parameterization is universal:

Proposition 4. *Functions $f:\mathbb{R}^{2\times k}\to\mathbb{R}^{d_{\text{out}}}$ of the above form are permutation invariant and sign invariant, and they universally approximate permutation invariant and sign invariant functions.*

Proof. Invariance of f is easy to see; let P be a 2×2 permutation matrix and $s_i\in\{-1,1\}$ for each i. Then

$$\begin{aligned}
f(Pv_1s_1,\dots,Pv_ks_k) &= \varphi\left((Pv_1s_1)\odot(Pv_1s_1),(Pv_1s_1)\odot\operatorname{rev}(Pv_1s_1),\dots\right) && (10)\\
&= \varphi\left(P(v_1s_1\odot v_1s_1),P(v_1s_1\odot\operatorname{rev}(v_1s_1)),\dots\right) && (11)\\
&= \varphi\left(P(v_1\odot v_1),P(v_1\odot\operatorname{rev}(v_1)),\dots\right) && (12)\\
&= \varphi\left(v_1\odot v_1,v_1\odot\operatorname{rev}(v_1),\dots\right) && (13)\\
&= f(v_1,\dots,v_k), && (14)
\end{aligned}$$

where the second to last inequality is by permutation invariance of φ. Next, we show universal approximation.

Let $h:\mathbb{R}^{2\times k}\to\mathbb{R}^{d_{\text{out}}}$ be a continuous permutation invariant and sign invariant function. Then by the decomposition theorem in Lim et al. [2023], we can write

$$h(v_1,\dots,v_k)=\rho(\phi(v_1v_1^\top),\dots,\phi(v_kv_k^\top)), \tag{15}$$

for continuous functions ρ and ϕ. As a composition of continuous functions, the function $\psi:B\subseteq\mathbb{R}^{2\times 2k}\to\mathbb{R}^{d_{\text{out}}}$ given by $\psi(A_1,\dots,A_k)=\rho(\phi(A_1),\dots,\phi(A_k))$ is continuous, where B is the subset of $\mathbb{R}^{2\times 2k}$ consisting of $(v_1v_1^\top,\dots,v_kv_k^\top)$ such that each $v_i\in\mathbb{R}^2$. Note that ψ is permutation invariant on B, in the sense that for any 2×2 permutation matrix P, we have

$$\psi(PA_1P^\top,\dots,PA_kP^\top)=\psi(A_1,\dots,A_k), \tag{16}$$

because if $v_iv_i^\top=A_i$, then

$$\psi(PA_1P^\top,\dots,PA_kP^\top)=h(Pv_1,\dots,Pv_k)=h(v_1,\dots,v_k)=\psi(A_1,\dots,A_k), \tag{17}$$

by permutation invariance of h.

Now, we define our permutation invariant function $\varphi:C\subseteq\mathbb{R}^{2\times 2k}\to\mathbb{R}^{d_{\text{out}}}$, on the domain

$$C=\{[v_1\odot v_1,v_1\odot\operatorname{rev}(v_1),\dots,v_k\odot v_k,v_k\odot\operatorname{rev}(v_k)]:v_i\in\mathbb{R}^2\}. \tag{18}$$

We define φ by

$$\varphi(A)=\psi\left(\begin{bmatrix}A_{1,1} & A_{2,2}\\ A_{2,2} & A_{2,1}\end{bmatrix},\begin{bmatrix}A_{1,3} & A_{2,4}\\ A_{2,4} & A_{2,3}\end{bmatrix},\dots,\begin{bmatrix}A_{1,2k-1} & A_{2,2k}\\ A_{2,2k} & A_{2,2k-1}\end{bmatrix}\right). \tag{19}$$

To see that φ is permutation invariant, we need only consider the case where $P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, in which case

$$\begin{aligned}
\varphi(PA) &= \psi\left(\begin{bmatrix} A_{2,1} & A_{1,2} \\ A_{1,2} & A_{1,1} \end{bmatrix}, \begin{bmatrix} A_{2,3} & A_{1,4} \\ A_{1,4} & A_{1,3} \end{bmatrix}, \ldots, \begin{bmatrix} A_{2,2k-1} & A_{1,2k} \\ A_{1,2k} & A_{1,2k-1} \end{bmatrix}\right) && (20)\\
&= \psi\left(P\begin{bmatrix} A_{1,1} & A_{2,2} \\ A_{2,2} & A_{2,1} \end{bmatrix}P^\top, P\begin{bmatrix} A_{1,3} & A_{2,4} \\ A_{2,4} & A_{2,3} \end{bmatrix}P^\top, \ldots, P\begin{bmatrix} A_{1,2k-1} & A_{2,2k} \\ A_{2,2k} & A_{2,2k-1} \end{bmatrix}P^\top\right) && (21)\\
&= \psi\left(\begin{bmatrix} A_{1,1} & A_{2,2} \\ A_{2,2} & A_{2,1} \end{bmatrix}, \begin{bmatrix} A_{1,3} & A_{2,4} \\ A_{2,4} & A_{2,3} \end{bmatrix}, \ldots, \begin{bmatrix} A_{1,2k-1} & A_{2,2k} \\ A_{2,2k} & A_{2,2k-1} \end{bmatrix}\right) \quad (\psi \text{ perm. inv.}) && (22)\\
&= \varphi(A), && (23)
\end{aligned}$$

where in the second equality, we use the fact that $A_{2,2j} = A_{1,2j}$, $j = 1, \ldots, k$ for $A \in C$, because $A_{2,2j} = (v_j \odot \mathrm{rev}(v_j))_2 = (v_j \odot \mathrm{rev}(v_j))_1 = A_{1,2j}$ for some $v_j \in \mathbb{R}^2$. Moreover, φ is clearly continuous and sign invariant. Defining $f : \mathbb{R}^{2\times k} \to \mathbb{R}^{d_{\mathrm{out}}}$ using this φ, we compute that

$$\begin{aligned}
f(v_1, \ldots v_k) &= \varphi\left(v_1 \odot v_1, v_1 \odot \mathrm{rev}(v_1), \ldots, v_k \odot v_k, v_k \odot \mathrm{rev}(v_k)\right) && (24)\\
&= \psi\left(\begin{bmatrix} v_{1,1}^2 & v_{1,1}v_{1,2} \\ v_{1,1}v_{1,2} & v_{1,2}^2 \end{bmatrix}, \ldots, \begin{bmatrix} v_{k,1}^2 & v_{k,1}v_{k,2} \\ v_{k,1}v_{k,2} & v_{k,2}^2 \end{bmatrix}\right) && (25)\\
&= \psi\left(v_1v_1^\top, \ldots, v_kv_k^\top\right) && (26)\\
&= h(v_1, \ldots, v_k), && (27)
\end{aligned}$$

so we are done.

If φ instead comes from a universally approximating class of permutation invariant neural networks (rather than being an arbitrary continuous permutation invariant function), then on a compact domain we can get ϵ approximation of f to h by letting φ approximate ψ to ϵ accuracy. □

A.3.2 Proof of Proposition 1

Proposition 1. *Let $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times d_{\mathrm{out}}}$ be a permutation equivariant function, and let $V = [v_1, \ldots, v_k] \in \mathbb{R}^{n\times k}$ be k orthonormal eigenvectors of an adjacency matrix A. Let nodes i and j be automorphic, and let z_i and $z_j \in \mathbb{R}^{d_{\mathrm{out}}}$ be their embeddings, i.e, the ith and jth row of $Z = f(V)$.*

- *If f is sign invariant and the eigenvalues associated with the v_l are distinct, then $z_i = z_j$.*
- *If f is basis invariant and $v_1, \ldots, v_k$ are a basis for some number of eigenspaces of A then $z_i = z_j$.*

Proof. We only prove the basis invariance claim, as the sign invariance claim is a special case; basis invariance is sign invariance when eigenvalues are distinct.

Let $P \in \mathbb{R}^{n\times n}$ be a permutation matrix associated to an automorphism that maps node i to node j, so $PAP^\top = A$ and $Pe_i = e_j$, where e_l is the lth standard basis vector. Let $V_t = [v_{r_1}, \ldots, v_{r_{d_t}}]$ be the matrix whose columns are the eigenvectors v_{r_l} that are associated to eigenvalue λ_i. The columns of V_t are thus an orthonormal basis for the eigenspace associated to λ_t. Note that for any of these eigenvectors, we have

$$A(Pv_{r_l}) = PAP^\top(Pv_{r_l}) = PAv_{r_l} = P\lambda_i v_{r_l} = \lambda_t(Pv_{r_l}), \tag{28}$$

so Pv_{r_l} is also an eigenvector of A with eigenvalue λ_t. As P is orthogonal, note that $Pv_{r_1}, \ldots, Pv_{r_{d_t}}$ is still an orthonormal basis of the eigenspace. Thus, there exists an orthogonal matrix $Q_t \in \mathbb{R}^{d_t\times d_t}$ such that $PV_t = V_tQ_t$—see Lim et al. [2023].

Repeat the above argument to get such a Q_t for each of the eigenbases $V_1, \ldots, V_l$. We can then see that

$$\begin{aligned}
z_j &= f(V_1, \ldots, V_l)_{j,:} \\
&= f(V_1Q_1, \ldots, V_lQ_l)_{j,:} && \text{basis invariance}\\
&= f(PV_1, \ldots, PV_l)_{j,:} && \text{choice of } Q_t\\
&= (Pf(V_1, \ldots, V_l))_{j,:} && \text{permutation equivariance}\\
&= f(V_1, \ldots, V_l)_{i,:} && \text{choice of } P\\
&= z_i.
\end{aligned}$$

So we are done. $\square$

A.4 Sign Invariance and Structural Node or Node-Pair Encodings

In this section, we show that when the eigenvalues $\lambda_1, \ldots, \lambda_k$ are distinct, then sign invariant functions of the orthonormal eigenvectors $v_1, \ldots, v_k$ give structural node or node-pair representations. This can also be generalized in a straightforward way to larger tuples of nodes beyond pairs, though we only consider nodes and node-pairs for ease of exposition. First, we give formal definitions.

Definition 1 (Structural Representations [Srinivasan and Ribeiro, 2019])**.** *Let $A \in \mathbb{R}^{n \times n}$ be the adjacency matrix of a graph on node set $\{1, \ldots, n\}$.*

A function $f : \mathbb{R}^{n \times n} \to \mathbb{R}^n$ is a node structural representation if $f(PAP^\top) = Pf(A)$ for all $n \times n$ permutation matrices P.

A function $f : \mathbb{R}^{n \times n} \to \mathbb{R}^{n \times n}$ is a node-pair structural representation if $f(PAP^\top) = Pf(A)P^\top$ for all $n \times n$ permutation matrices P.

Importantly, these structural representations are permutation equivariant functions of adjacency matrices, not arbitrary matrices. For each adjacency matrix A, let $V(A) = [v_1(A), \ldots, v_k(A)]$ be a choice of orthonormal eigenvectors for the first k eigenvalues $\lambda_1(A), \ldots, \lambda_k(A)$. We assume in this section that these first k eigenvalues are distinct for all A under consideration, so $V(A)$ is defined up to sign flips. Let $h : \mathbb{R}^{n \times k} \to \mathbb{R}^n$ be a permutation equivariant function of sets, so $h(PX) = Ph(X)$ for all permutations matrices P. Then of course $h(PV(A)) = Ph(V(A))$, but this does not make h a node structural encoding. This is because $A \mapsto h(V(A))$ is in general not a well-defined function of the adjacency, since the choice of $V(A)$ is not well-defined (the choices of sign are arbitrary). If we constrain h to not depend on the signs (sign invariance), or to depend on the signs in a predictable way (sign equivariance), then we can compute structural node or node-pair encodings from eigenvectors.

We capture these observations in the below proposition. First, we define three types of functions:

- Let $f_{\text{node}} : \mathbb{R}^{n \times k} \to \mathbb{R}^n$ be sign invariant and permutation equivariant; that is, $f_{\text{node}}(Pv_1 s_1, \ldots, Pv_k s_k) = Pf_{\text{node}}(v_1, \ldots, v_k)$ for $s_i \in \{-1, 1\}$ and P a permutation matrix.
- Let $f_{\text{decode}} : \mathbb{R}^{2 \times k} \to \mathbb{R}$ be sign invariant; that is, $f_{\text{decode}}(Sz_i, Sz_j) = f_{\text{decode}}(z_i, z_j)$ for $S \in \text{diag}(\{-1, 1\}^k)$.
- Let $f_{\text{equiv}} : \mathbb{R}^{n \times k} \to \mathbb{R}^{n \times k}$ be a permutation equivariant and sign equivariant function; that is, $f_{\text{equiv}}(PV(A)S) = Pf_{\text{equiv}}(V(A))S$ for $S \in \text{diag}(\{-1, 1\}^k)$ and P a permutation matrix.

Proposition 5. *Let $\mathcal{A} \subseteq \mathbb{R}^{n \times n}$ denote the matrices with distinct first-k eigenvalues. For $A \in \mathcal{A}$, let $V(A) = [v_1(A), \ldots, v_k(A)]$ be a choice of orthonormal eigenvectors of A, associated to the first-k (distinct) eigenvalues $\lambda_1(A), \ldots, \lambda_k(A)$. Then*

(a) The map $q_{\text{node}} : \mathcal{A} \to \mathbb{R}^n$ given by $q_{\text{node}}(A)_i = f_{\text{node}}\left(f_{\text{equiv}}(V(A))\right)_i$ is well-defined and gives a structural node representation.

(b) The map $q_{\text{pair}} : \mathcal{A} \to \mathbb{R}^{n \times n}$ defined by $q_{\text{pair}}(A)_{i,j} = f_{\text{decode}}\left(f_{\text{equiv}}(V(A))_{i,:}, f_{\text{equiv}}(V(A))_{j,:}\right)$ is well-defined and gives a structural node-pair representation.

Note that the identity mapping $V(A) \mapsto V(A)$ is permutation equivariant and sign equivariant, so using f_{node} or f_{decode} directly on eigenvectors also gives structural representations. The statement (b) means that our link prediction pipeline with sign equivariant node features and sign invariant decoding produces structural node-pair representations.

Proof. **Part (a)** We first show that $q_{\text{node}} : \mathcal{A} \to \mathbb{R}^n$ is well-defined. Suppose we had another choice of eigenvectors, so the eigenvectors we input are $V(A)S$ for some $S \in \text{diag}(\{-1, 1\}^k)$. Then

$$f_{\text{node}}\left(f_{\text{equiv}}(V(A)S)\right) = f_{\text{node}}\left(f_{\text{equiv}}(V(A))S\right) = f_{\text{node}}\left(f_{\text{equiv}}(V(A)\right), \tag{29}$$

where the first equality is by sign equivariance, and the second equality by sign invariance. Thus, the value of $q_{\text{node}}(A)$ is unchanged.

Now, let P be any permutation matrix. Then for each eigenvector $v_i(A)$, $i \in [k]$, we have $(PAP^\top)Pv_i(A) = PAv_i(A) = \lambda_i(A)Pv_i(A)$, so $Pv_i(A)$ is an eigenvector of $PAP^\top$ associated to $\lambda_i(A) = \lambda_i(PAP^\top)$. Hence, we denote $v_i(PAP^\top) = Pv_i(A)$ (the choice of sign does not matter as q does not depend on the sign. Now, we have that

$$\begin{aligned} q_{\text{node}}(PAP^\top) &= f_{\text{node}}\left(f_{\text{equiv}}(V(PAP^\top))\right) && (30)\\ &= f_{\text{node}}\left(f_{\text{equiv}}(PV(A))\right) && (31)\\ &= Pf_{\text{node}}\left(f_{\text{equiv}}(V(A))\right) && (32)\\ &= Pq_{\text{node}}(A) && (33) \end{aligned}$$

where the second to last equality is by permutation equivariance of f_{node} and f_{equiv}.

Part (b) That $q_{\text{pair}} : \mathcal{A} \to \mathbb{R}^{n \times n}$ is well-defined follows from a similar argument to the q_{node} case. Let P be a permutation matrix, and $\sigma : [n] \to [n]$ its underlying permutation. We compute that

$$\begin{aligned} q_{\text{pair}}(PAP^\top)_{i,j} &= f_{\text{decode}}\left(f_{\text{equiv}}(V(PAP^\top))_{i,:}, f_{\text{equiv}}(V(PAP^\top))_{j,:}\right) && (34)\\ &= f_{\text{decode}}\left(f_{\text{equiv}}(PV(A))_{i,:}, f_{\text{equiv}}(PV(A))_{j,:}\right) && (35)\\ &= f_{\text{decode}}\left([Pf_{\text{equiv}}(V(A))]_{i,:}, [Pf_{\text{equiv}}(V(A))]_{j,:}\right) && (36)\\ &= f_{\text{decode}}\left(f_{\text{equiv}}(V(A))_{\sigma^{-1}(i),:}, f_{\text{equiv}}(V(A))_{\sigma^{-1}(j),:}\right) && (37)\\ &= q_{\text{pair}}(A)_{\sigma^{-1}(i),\sigma^{-1}(j)} && (38)\\ &= (Pq_{\text{pair}}(A)P^\top)_{i,j} && (39) \end{aligned}$$

□

A.4.1 Sign Equivariance is Provably More Expressive for Link Prediction

Our arguments in Section 2.1 and Figure 2 explain why we can expect sign equivariant models to be more powerful than sign invariant models in link prediction. To give a theoretically rigorous explanation, here we provide an example where sign equivariant models can provably compute more expressive link representations than sign invariant models.

Consider a cycle graph C_{2k} for some even length $2k$, where $k \geq 3$. All nodes are automorphic in this graph, so any model based on structural node representations must assign the same representation to each node-pair. For instance, consider the eigenvalue -2 of the adjacency matrix, which is a simple eigenvalue with eigenvector $[1, -1, 1, -1, \ldots, 1, -1]$ [Lee et al., 1992]. Then a sign invariant model will lose the sign information and map each node to the same encoding, which means that each node-pair will also have the same encoding. However, a sign equivariant model can preserve the sign of each node (for instance by learning the identity function). Then for any pair of nodes that are one hop away, it can take a dot product to compute the pair representation -1, whereas it can take a dot product between any nodes that are two hops away to compute the pair representation 1. Of course, using more eigenvectors would allow for more complex representations to be computed.

A.4.2 More on Sign Equivariance and Link Prediction

Key to our method is the ability to update a positional node embedding in an equivariant way, which respects the graph symmetries. To elaborate, consider the aforementioned definition of node positional encodings as samples from a permutation equivariant probability distribution over node features [Srinivasan and Ribeiro, 2019]. Laplacian eigenvector positional embeddings are samples from the distribution of orthonormal bases of the eigenspaces of the Laplacian. Our sign equivariance based approach is possible because the randomness in Laplacian eigenvector positional encodings is exceptionally structured (consisting only of sign flips when eigenvalues are distinct). In contrast, a general way to obtain structural pair representations from node positional embeddings is to average some function over the randomness of the positional encoding (i.e., over many samples of the positional encoding) [Srinivasan and Ribeiro, 2019], but this is highly expensive, often intractable, and introduces substantial variance into the learning procedure. For instance, one may have to average samples of the $n!$ assignments of unique node identifiers [Murphy et al., 2019] or approximate an integral over Gaussian random features [Abboud et al., 2021].

A.5 Proof of Proposition 2, Orthogonal Equivariance via Sign Equivariance

Proposition 2. *Consider a domain $\mathcal{X} \subseteq \mathbb{R}^{n\times d}$ such that each $X \in \mathcal{X}$ has distinct covariance eigenvalues, and let R_X be a choice of orthonormal eigenvectors of* $\mathrm{cov}(X)$ *for each $X \in \mathcal{X}$. If $h : \mathcal{X} \subseteq \mathbb{R}^{n\times d} \to \mathbb{R}^{n\times d}$ is sign equivariant, and if $f(X) = h(XR_X)R_X^\top$, then f is well defined and orthogonally equivariant.*

Moreover, is h is from a universal class of sign equivariant functions, then the f of the above form universally approximate $O(k)$ equivariant functions on $\mathcal{X}$.

Proof. First, we show that f is well defined. R_X is only unique up to sign flips, as $R_X S$ is an orthonormal set of eigenvectors of $\mathrm{cov}(X)$ for $S \in \mathrm{diag}(\{-1,1\}^k)$. However, no matter the choice of signs, $f(X)$ takes the same value, since

$$
\begin{aligned}
h(XR_XS)(R_XS)^\top &= h(XR_XS)S^\top R_X^\top && (40)\\
&= h(XR_X)SS^\top R_X^\top && \text{sign equivariance} \quad (41)\\
&= h(XR_X)R_X^\top. && (42)
\end{aligned}
$$

Next, we show that f is $O(k)$ equivariant. Let $Q \in O(k)$ be any orthogonal matrix. Note that

$$
\mathrm{cov}(XQ) = \left(XQ - \frac{1}{n}\mathbf{1}\mathbf{1}^\top XQ\right)^\top \left(XQ - \frac{1}{n}\mathbf{1}\mathbf{1}^\top XQ\right) = Q^\top \mathrm{cov}(X)Q. \tag{43}
$$

Thus, $Q^\top R_X$ is an orthonormal set of eigenvectors of $\mathrm{cov}(XQ)$. This means that there is a choice of signs $S \in \mathrm{diag}(\{-1,1\}^k)$ such that $Q^\top R_X S = R_{XQ}$. Hence, we have that

$$
\begin{aligned}
f(XQ) &= h(XQR_{XQ})R_{XQ}^\top && (44)\\
&= h(XQQ^\top R_XS)(Q^\top R_XS)^\top && (45)\\
&= h(XR_X)SS^\top R_X^\top Q && \text{sign equivariance} \quad (46)\\
&= h(XR_X)R_X^\top Q && (47)\\
&= f(X)Q^\top, && (48)
\end{aligned}
$$

so f is $O(k)$ equivariant.

Universal Approximation. Our proof of the universality of this class of functions builds on the proof of the universality of frame averaging [Puny et al., 2022]. Let f_{target} be a continuous $O(k)$ equivariant function and let $\epsilon > 0$ be a desired approximation accuracy. Then f_{target} is also sign equivariant (as the sign matrices $S \in \mathrm{diag}(\{-1,1\}^k)$ are orthogonal).

Hence, by sign equivariant universality, we can choose a sign equivariant h such that $\|h(X) - f_{\text{target}}(X)\| < \epsilon$ for all $X \in \mathcal{X}$ (where $\|\cdot\|$ is the Frobenius norm). Define the $O(k)$ equivariant $f(X) = h(XR_X)R_X^\top$. Then for all $X \in \mathcal{X}$ we have that

$$
\begin{aligned}
\|f_{\text{target}}(X) - f(X)\| &= \left\|f_{\text{target}}(X) - h(XR_X)R_X^\top\right\| && (49)\\
&= \left\|f_{\text{target}}(X)R_XR_X^\top - h(XR_X)R_X^\top\right\| && R_X \text{ orthogonal} \quad (50)\\
&= \left\|f_{\text{target}}(XR_X)R_X^\top - h(XR_X)R_X^\top\right\| && \text{orthogonal equivariance} \quad (51)\\
&= \|f_{\text{target}}(XR_X) - h(XR_X)\| && R_X \text{ orthogonal} \quad (52)\\
&< \epsilon. && (53)
\end{aligned}
$$

So f approximates f_{target} within ϵ accuracy on $\mathcal{X}$, and we are done. □

B Characterization of Sign Equivariant Polynomials

In this Appendix, we characterize the form of the sign equivariant polynomials. This is useful, because for a finite group, equivariant polynomials universally approximate equivariant continuous functions [Yarotsky, 2022]; thus, if a model universally approximates equivariant polynomials, then it universally approximates equivariant continuous functions. Using equivariant polynomials to analyze or develop equivariant machine learning models has been done successfully in many contexts [Zaheer et al., 2017, Yarotsky, 2022, Segol and Lipman, 2019, Dym and Maron, 2021, Maron et al., 2019, 2020, Villar et al., 2021, Dym and Gortler, 2022, Puny et al., 2023].

B.1 Sign Equivariant Linear Map Characterization

Here, we prove our result characterizing the form of the equivariant linear maps.

Lemma 1. *A linear map $W : \mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$ is sign equivariant if and only if it can be written as*

$$W(X) = [W_1 X_1 \ \dots \ W_k X_k] \tag{54}$$

for some linear maps $W_1, \dots, W_k : \mathbb{R}^n \to \mathbb{R}^{n'}$, where $X_i \in \mathbb{R}^n$ is the ith column of $X \in \mathbb{R}^{n \times k}$.

Proof. For one direction, suppose W can be written as in equation 4. To see that W is sign equivariant, note that for any $S \in \operatorname{diag}(\{-1, 1\}^k)$, we have

$$W(XS) = [s_1 W_1 X_1 \ \dots \ s_k W_k X_k] = [W_1 X_1 \ \dots \ W_k X_k]\, S = W(X)S. \tag{55}$$

For the other direction, let W be a sign equivariant linear map. For any $i' \in [n']$ and $j' \in [k]$, we can write the action of W as

$$W(X)_{i',j'} = \sum_{i=1}^{n} \sum_{j=1}^{k} W_{i',j'}^{i,j} X_{i,j}, \tag{56}$$

where $W_{i',j'}^{i,j} \in \mathbb{R}$ are coefficients representing the linear map. Let $c \neq j'$ be a column that is not j'. Further, for any row $l \in [n]$, let $\tilde{X} \in \mathbb{R}^{n \times k}$ be such that $\tilde{X}_{l,c} = 1$, and $\tilde{X}$ is zero elsewhere. Then we have that

$$W(\tilde{X})_{i',j'} = W_{i',j'}^{l,c}. \tag{57}$$

Now, let $S \in \operatorname{diag}(\{-1, 1\}^k)$ have a -1 in the j'th column and a 1 elsewhere. Then $\tilde{X}S = \tilde{X}$. This implies that

$$
\begin{aligned}
W_{i',j'}^{l,c} &= W(\tilde{X})_{i',j'} && (58)\\
&= W(\tilde{X}S)_{i',j'} && (59)\\
&= -W(\tilde{X})_{i',j'} && (60)\\
&= -W_{i',j'}^{l,c}, && (61)
\end{aligned}
$$

where in the second to last equality we used sign equivariance. This implies that $W_{i',j'}^{l,c} = 0$.

Hence, for any $i' \in [n']$, $j' \in [k']$, we have that $W(X)_{i',j'}$ only depends on $X_{j'}$, so we are done. □

B.2 Sign Invariant Polynomials $\mathbb{R}^k \to \mathbb{R}$

For simplicity, we start with the case of sign invariant polynomials $p : \mathbb{R}^k \to \mathbb{R}$. The sign equivariant polynomials take a very similar form. We can write any polynomial from $\mathbb{R}^k$ to $\mathbb{R}$ in the form

$$p(v) = \sum_{d_1, \dots, d_k = 0}^{D} \mathbf{W}_{d_1, \dots, d_k} v_1^{d_1} \cdots v_k^{d_k} \tag{62}$$

for some coefficients $\mathbf{W}_{d_1, \dots, d_k} \in \mathbb{R}$ and some $D \in \mathbb{N}$. Sign invariance tells us that for any $S = \operatorname{diag}(s_1, \dots, s_k) \in \operatorname{diag}(\{-1, 1\}^k)$, we must have

$$\sum_{d_1, \dots, d_k = 0}^{D} \mathbf{W}_{d_1, \dots, d_k} v_1^{d_1} \cdots v_k^{d_k} = p(v) = p(Sv) = \sum_{d_1, \dots, d_k = 0}^{D} \mathbf{W}_{d_1, \dots, d_k} s_1^{d_1} \cdots s_k^{d_k} v_1^{d_1} \cdots v_k^{d_k}. \tag{63}$$

This holds for any $v \in \mathbb{R}^k$, so for all choices of $d_1, \dots, d_k$ we must have

$$\mathbf{W}_{d_1, \dots, d_k} = s_1^{d_1} \cdots s_k^{d_k} \mathbf{W}_{d_1, \dots, d_k}, \quad \text{for all } (s_1, \dots, s_k) \in \{-1, 1\}^k. \tag{64}$$

Note that $s_i^{d_i} = 1$ if d_i is an even number. Hence, there are no constraints on $\mathbf{W}_{d_1, \dots, d_k}$ if all d_i are even. On the other hand, suppose d_j is odd for some j. Let $s_i = 1$ for $i \neq j$ and $s_j = -1$. Then the constraint says that $\mathbf{W}_{d_1,\ldots,d_k} = -\mathbf{W}_{d_1,\ldots,d_k}$, so we must have $\mathbf{W}_{d_1,\ldots,d_k} = 0$. To summarize, we have

$$\mathbf{W}_{d_1,\ldots,d_k} = \begin{cases} \text{free} & d_i \text{ even for each } i \\ 0 & \text{else} \end{cases} \tag{65}$$

Where being free means that the coefficient may take any value in $\mathbb{R}$. Thus, any sign invariant p only has terms where each variable v_i is raised to an even power. It is also easy to see that any polynomial p where each variable v_i is raised to only even powers is sign invariant, so we have the following proposition:

Proposition 6. *A polynomial $p : \mathbb{R}^k \to \mathbb{R}$ is sign invariant if and only if it can be written*

$$p(v) = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}, \tag{66}$$

for some coefficients $\mathbf{W}_{d_1,\ldots,d_k} \in \mathbb{R}$ and $D \in \mathbb{N}$.

In other words, p is sign invariant if and only if there exists a polynomial $q : \mathbb{R}^k \to \mathbb{R}$ such that $p(v) = q(v_1^2, \ldots, v_k^2)$.

B.3 Sign Equivariant Polynomials $\mathbb{R}^k \to \mathbb{R}^k$

The case of sign equivariant polynomials $p : \mathbb{R}^k \to \mathbb{R}^k$ is very similar. For $l \in [k]$, the lth output dimension of any polynomial $p : \mathbb{R}^k \to \mathbb{R}^k$ can be written

$$p(v)_l = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} v_1^{d_1} \cdots v_k^{d_k}, \tag{67}$$

where $\mathbf{W}^{(l)}_{d_1,\ldots,d_k} \in \mathbb{R}$ are coefficients (note the extra l index, so there are k times more coefficients than in the invariant case). By sign equivariance, we have

$$s_l \cdot p(v)_l = p(Sv)_l \tag{68}$$

$$s_l \cdot \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} v_1^{d_1} \cdots v_k^{d_k} = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} s_1^{d_1} \cdots s_k^{d_k} v_1^{d_1} \cdots v_k^{d_k}. \tag{69}$$

As this holds for all inputs $v \in \mathbb{R}^k$, we have the following constraints on the coefficients:

$$s_l \mathbf{W}^{(l)}_{d_1,\ldots,d_k} = s_1^{d_1} \cdots s_k^{d_k} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} \tag{70}$$

$$\mathbf{W}^{(l)}_{d_1,\ldots,d_k} = s_l \cdot s_1^{d_1} \cdots s_k^{d_k} \mathbf{W}^{(l)}_{d_1,\ldots,d_k}, \tag{71}$$

where we use the fact that $s_l = 1/s_l$ since $s_l \in \{-1, 1\}$. If d_j is odd for $j \neq l$, then similarly to the invariant case, we can take $s_i = 1$ for $i \neq j$ and $s_j = -1$ in the above equation to see that $\mathbf{W}^{(l)}_{d_1,\ldots,d_k} = 0$. If d_l is even, then $d_l + 1$ is odd, so we have that $\mathbf{W}^{(l)}_{d_1,\ldots,d_k} = 0$ by the same argument. Thus, we must have

$$\mathbf{W}^{(l)}_{d_1,\ldots,d_k} = \begin{cases} \text{free} & d_l \text{ odd, and } d_i \text{ even for each } i \neq l \\ 0 & \text{else} \end{cases}. \tag{72}$$

Thus, the lth entry $p(v)_l$ only contains monomials of the term $v_1^{2d_1} \cdots v_l^{2d_l+1} \cdots v_k^{2d_k}$, where each term besides v_l is raised to an even power. We can factor out a v_l and write such terms as $v_l \cdot v_1^{2d_1} \cdots v_k^{2d_k}$. It is also easy to see that any polynomial with monomials only of this form is sign equivariant. Thus, we have proven Proposition 7.

Proposition 7. *A polynomial $p : \mathbb{R}^k \to \mathbb{R}^k$ is sign equivariant if and only if it can be written*

$$p(v)_l = v_l \cdot \left(\sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k} \right). \tag{73}$$

In vector format, p is sign equivariant if and only if it can be written as $p(v) = v \odot p_{\mathrm{inv}}(v)$ for a sign invariant polynomial $p_{\mathrm{inv}} : \mathbb{R}^k \to \mathbb{R}^k$.

B.4 Sign Equivariant Polynomials $\mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$

Finally, we will handle the case of polynomials $p : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ equivariant to $\mathrm{diag}(\{-1,1\}^k)$. This is the case we most often deal with in practice, when we have input $V = [v_1 \quad \dots \quad v_k]$ for k eigenvectors $v_i \in \mathbb{R}^n$ of some $n \times n$ matrix. For $a \in [n']$ and $b \in [k]$, the (a,b)th output of a polynomial $\mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ is

$$p(V)_{a,b} = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} \prod_{i=1}^{n}\prod_{j=1}^{k} V_{i,j}^{d_{i,j}}, \tag{74}$$

where the sum ranges over $d_{i,j} \in \{0,\dots,D\}$ for $i \in [n]$ and $j \in [k]$, and $\mathbf{d} = (d_{1,1},\dots,d_{n,1},d_{1,2},\dots,d_{n,k})$ is a shorthand to index coefficients $\mathbf{W}_{\mathbf{d}}^{(a,b)} \in \mathbb{R}$. By sign equivariance, we have that:

$$s_b \cdot p(V)_{a,b} = p(VS)_{a,b} \tag{75}$$

$$s_b \cdot \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} \prod_{i=1}^{n}\prod_{j=1}^{k} V_{i,j}^{d_{i,j}} = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} s_1^{\tilde{d}_1}\cdots s_k^{\tilde{d}_k} \prod_{i=1}^{n}\prod_{j=1}^{k} V_{i,j}^{d_{i,j}}, \tag{76}$$

where $\tilde{d}_j = \sum_{i'=1}^n d_{i',j}$ is the number of times that an entry from column j of V appears in the product $\prod_{i=1}^n \prod_{j=1}^k V_{i,j}^{d_{i,j}}$. As this holds over all V, we thus have that

$$\mathbf{W}_{\mathbf{d}}^{(a,b)} = s_b \cdot s_1^{\tilde{d}_1}\cdots s_k^{\tilde{d}_k} \cdot \mathbf{W}_{\mathbf{d}}^{(a,b)}. \tag{77}$$

By analogous arguments to the previous subsections, if $\tilde{d}_j$ is odd for $j \neq b$, we have that the $\mathbf{W}_{\mathbf{d}}^{(a,b)} = 0$. Likewise, if $\tilde{d}_b$ is even, we have $\mathbf{W}_{\mathbf{d}}^{(a,b)} = 0$. Thus, the constraint on $\mathbf{W}$ is

$$\mathbf{W}_{\mathbf{d}}^{(a,b)} = \begin{cases} \text{free} & \sum_i d_{i,b} \text{ odd, and } \sum_i d_{i,j} \text{ even for each } j \neq b \\ 0 & \text{else} \end{cases}. \tag{78}$$

In particular, this means that the only nonzero terms in the sum that defines $p(V)_{a,b}$ have an even number of entries from column j for $j \neq b$, and an odd number of entries from column b. Thus, each term can be written as $V_{i_{\mathbf{d}},b} \cdot p_{\mathrm{inv}}(V)_{\mathbf{d}}$ for some index $i_{\mathbf{d}} \in [n]$ and sign invariant polynomial p_{inv}. Moreover, it can be seen that any polynomial that only has terms of this form is sign equivariant. Thus, we have shown the following proposition:

Proposition 8. *A polynomial $p : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ is sign equivariant if and only if it can be written as*

$$p(V)_{a,b} = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} V_{i_{\mathbf{d}},b} \cdot p_{\mathrm{inv}}(V)_{\mathbf{d}}, \tag{79}$$

where p_{inv} is a sign invariant polynomial, the sum ranges over all $\mathbf{d}$, and $i_{\mathbf{d}} \in [n]$ for each $\mathbf{d}$.

Now, we show that this implies Theorem 1. In particular, we will write p in the form

$$p(V) = W^{(2)}\left((W^{(1)}V) \odot q_{\mathrm{inv}}(V)\right), \tag{80}$$

for sign equivariant linear maps $W^{(2)}$ and $W^{(1)}$, and a sign equivariant polynomial q_{inv}. To do so, let $\tilde{D}$ denote the number of all possible $\mathbf{d}$ that the sum in equation 79 ranges over. We take $W^{(1)} : \mathbb{R}^{n\times \tilde{k}} \to \mathbb{R}^{\tilde{D}n'\times k}$ and $W^{(2)} : \mathbb{R}^{\tilde{D}n'\times k} \to \mathbb{R}^{n'\times k}$. These sign equivariant linear maps have to act independently on each column of their input, so $W^{(1)}V = [W_1^{(1)}v_1, \dots W_k^{(1)}v_k]$ for linear maps $W_i^{(1)} : \mathbb{R}^n \to \mathbb{R}^{\tilde{D}n'}$. We define $W_b^{(1)}$ to be the linear map such that $(W_b^{(1)}v_b)_{\mathbf{d},a} = W_{\mathbf{d}}^{(a,b)} V_{i_{\mathbf{d}},b}$ for $a \in [n']$. For the sign invariant polynomial q_{inv}, we take $q_{\mathrm{inv}}(V)_{\mathbf{d},a} = p_{\mathrm{inv}}(V)_{\mathbf{d}}$.

Finally, we define $W^{(2)}$ to compute the sum in equation 79. In particular, for $X = [x_1,\dots,x_k] \in \mathbb{R}^{\tilde{D}n'\times k}$ we write $W^{(2)}X = [W_1^{(2)}x_1,\dots,W_k^{(2)}x_k]$, where $(W_b^{(2)}x_b)_a = \sum_{\mathbf{d}} x_{i_{\mathbf{d}},b}$. It can be seen that with these definitions of $W^{(2)}$, $W^{(1)}$, and q_{inv}, we have written p in the desired form.

B.5 Sign Invariant Polynomials and SignNet

For completeness, here we state the form of the sign invariant polynomials $p : \mathbb{R}^{n\times k} \to \mathbb{R}$ on inputs $V = [v_1, \ldots, v_k] \in \mathbb{R}^{n\times k}$. The derivation very closely follows that of the sign equivariant polynomials from $\mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ in Appendix B.4, so we omit this derivation.

Proposition 9. *A polynomial $p : \mathbb{R}^{n\times k} \to \mathbb{R}$ is sign invariant if and only if it can be written*

$$p(V) = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}} \prod_{i=1}^{n} \prod_{j=1}^{k} V_{i,j}^{d_{i,j}}, \tag{81}$$

where $\mathbf{W}_{\mathbf{d}} \neq 0$ for $\mathbf{d} = (d_{1,1}, \ldots, d_{n,1}, d_{1,2}, \ldots, d_{n,k})$ only if $\sum_{i=1}^{n} d_{i,j}$ is even for each column $j \in [k]$.

In particular, p is sign invariant if and only if there is a polynomial $q : \mathbb{R}^{n\times n\times k} \to \mathbb{R}$ such that $p(V) = q([V_{i_1,j} \cdot V_{i_2,j}]_{i_1\in[n], i_2\in[n], j\in[k]})$.

The polynomials $V \mapsto V_{i_1,j} \cdot V_{i_2,j}$ for $i_1, i_2 \in [n]$ and $j \in [k]$ are thus generators of the ring of sign invariant polynomials from $\mathbb{R}^{n\times k} \to \mathbb{R}$.

Notably, Lim et al. [2023] propose universal sign invariant neural architectures, but do not characterize or otherwise use the sign invariant polynomials. Instead, their proof of universality uses topological constructions and shows that all sign invariant continuous functions can be decomposed in a simple form—namely, $\rho([\phi(v_i) + \phi(-v_i)]_{i=1,\ldots,k})$ for continuous functions ρ and ϕ. Our characterization of sign invariant polynomials provides another path to developing and analyzing the expressive power of sign invariant architectures.

In particular, we can give an alternative proof for the universality of SignNet.

Proposition 10 (Universality of SignNet)**.** *Let $f : \mathcal{X} \subseteq \mathbb{R}^{n\times k} \to \mathbb{R}$ be a continuous sign invariant function on a compact domain $\mathcal{X}$, and let $\epsilon > 0$. Then there exists a continuous $\rho : \mathbb{R}^{n^2 k} \to \mathbb{R}$ and continuous $\phi : \mathbb{R}^n \to \mathbb{R}^{n^2}$ such that $|f(V) - \rho([\phi(v_i) + \phi(-v_i)]_{i=1,\ldots,k})| < \epsilon$ for all $V \in \mathcal{X}$.*

Proof. First, let p be a sign invariant polynomial that approximates f to within ϵ on $\mathcal{X}$. Then using Proposition 9, let q be a polynomial such that $p(V) = q([V_{i_1,j} \cdot V_{i_2,j}]_{i_1\in[n], i_2\in[n], j\in[k]})$.

Define $\phi : \mathbb{R}^n \to \mathbb{R}^{n^2}$ to map a $v \in \mathbb{R}^n$ to the vector of pairwise products of elements in v scaled by 1/2, that is

$$\phi(v) = \frac{1}{2}\mathrm{vec}(vv^\top) \tag{82}$$

Then $\phi(v) + \phi(-v)$ is equal to the vector of pairwise products of v. Finally, we let $\rho = q$, which gives that

$$p(V) = \rho([\phi(v_i) + \phi(-v_i)]_{i=1,\ldots,k}), \tag{83}$$

and hence

$$|f(V) - \rho([\phi(v_i) + \phi(-v_i)]_{i=1,\ldots,k})| = |f(V) - p(V)| < \epsilon \tag{84}$$

for all $V \in \mathcal{X}$. ∎

Given the form of the sign invariant polynomials, this proof is quite simple. However, it is technically weaker than the result of Lim et al. [2023], as they invoke the strong Whitney embedding theorem and only require ϕ to map to $\mathbb{R}^{2n}$ instead of $\mathbb{R}^{n^2}$. Still, further arguments could reduce the dimension required to about $2n$ in this polynomial-based proof; as the Gram matrix $vv^\top$ is rank one, it can be recovered almost always from about $2n$ of its entries [Pimentel-Alarcón et al., 2016].

C Sign Equivariant Architecture Universality

In this section, we prove Proposition 3 on the universality of our proposed sign equivariant architectures, which we restate here:

Proposition 3. *Functions of the form $v \mapsto v \odot \mathrm{MLP}(|v|)$ universally approximate continuous sign equivariant functions $f : \mathbb{R}^k \to \mathbb{R}^k$.*

Compositions $f_2 \circ f_1$ of functions f_l as in equation 6 universally approximate continuous sign equivariant functions $f : \mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$.

We prove the two statements of the proposition in the next two subsections.

C.1 Universality for functions $\mathbb{R}^k \to \mathbb{R}^k$

Proof. Let $\mathcal{X} \subseteq \mathbb{R}^k$ be a compact set, let $\epsilon > 0$, and let $f_{\mathrm{target}} : \mathcal{X} \to \mathbb{R}^k$ be a continuous sign equivariant function that we wish to approximate within ϵ. Choose a sign equivariant polynomial p that approximates f_{target} to within $\epsilon/2$ on $\mathcal{X}$. By compactness, we can choose a finite bound $B > 0$ such that $|v_i| < B$ for all $v \in \mathcal{X}$.

By Proposition 7, we can write $p(v)_l = v_l \cdot \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}$. By the universal approximation theorem for multilayer perceptrons, we can choose a $\mathrm{MLP} : \mathcal{X} \to \mathbb{R}^k$ such that approximates $q(v) = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}$ up to $\epsilon/(2B)$. Note that $q(|v|) = q(v)$, so $v \mapsto \mathrm{MLP}(|v|)$ also approximates q within $\epsilon/(2B)$ accuracy.

Thus, for all $v \in \mathcal{X}$, we have that

$$|f(v)_i - p(v)_i| = |v_i \cdot \mathrm{MLP}(|v|)_i - v_i \cdot \sum_{d=1}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}| \tag{85}$$

$$= |v_i||\mathrm{MLP}(|v|)_i - \sum_{d=1}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}| \tag{86}$$

$$\leq B \cdot |\mathrm{MLP}(|v|)_i - \sum_{d=1}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}| \tag{87}$$

$$< \epsilon/2, \tag{88}$$

so $\|f - p\|_\infty < \epsilon/2$ on $\mathcal{X}$ and we are done by the triangle inequality. □

C.2 Universality for functions $\mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$

Recall that each layer of our sign equivariant network from $\mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$ takes the form

$$f_l(V) = [W_1^{(l)} v_1, \ldots, W_k^{(l)} v_k] \odot \mathrm{SignNet}_l(V).$$

Proof. Let $\mathcal{X} \subseteq \mathbb{R}^{n \times k}$ be compact, and let $f_{\mathrm{target}} : \mathcal{X} \to \mathbb{R}^{n' \times k}$ be a continuous sign equivariant function that we wish to approximate. Since $\mathcal{X}$ is compact, we can choose a finite bound $B > 0$ such that $|V_{ij}| < B$ for all $V \in \mathcal{X}$. Let $p : \mathcal{X} \subseteq \mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$ be a sign equivariant polynomial that approximates f_{target} up to $\epsilon/2$ accuracy. Using Proposition 8, we can write

$$p(V)_{a,b} = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} V_{i_{\mathbf{d}},b} \cdot p_{\mathrm{inv}}(V)_{\mathbf{d}},$$

for some sign invariant polynomials $p_{\mathrm{inv}}(V)_{\mathbf{d}}$. We will have one network layer f_1 approximate the summands, and have the second network layer f_2 compute the sum.

First, we absorb the coefficients $\mathbf{W}_{\mathbf{d}}^{(a,b)}$ into the sign invariant part, by defining the sign invariant polynomial $q_{\mathrm{inv}}(V)_{\mathbf{d},a,b} = \mathbf{W}_{\mathbf{d}}^{(a,b)} p_{\mathrm{inv}}(V)_{\mathbf{d}}$, so we can write

$$p(V)_{a,b} = \sum_{d_{i,j}=0}^{D} V_{i_{\mathbf{d}},b} \cdot q_{\mathrm{inv}}(V)_{\mathbf{d},a,b}.$$

Now, let $d_{\mathrm{hidden}} \in \mathbb{N}$ denote the number of all possible $\mathbf{d}$ that appear in the sum, multiplied by n'. We define $f_1 : \mathcal{X} \to \mathbb{R}^{d_{\mathrm{hidden}} \times k}$ as follows. As SignNet [Lim et al., 2023] universally approximates sign invariant functions on compact sets, we can let $\text{SignNet}_1 : \mathcal{X} \to \mathbb{R}^{d_{\text{hidden}} \times k}$ be a SignNet that approximates $q_{\text{inv}}(V)$ up to $\epsilon/(2B)$ accuracy, so

$$|\text{SignNet}_1(V)_{(\mathbf{d},a),b} - q_{\text{inv}}(V)_{\mathbf{d},a,b}| < \frac{\epsilon}{2B \cdot d_{\text{hidden}}}. \tag{89}$$

For $b \in [k]$, we also define the weight matrices $W_b^{(1)} \in \mathbb{R}^{d_{\text{hidden}} \times n}$ of the layer by letting the $(\mathbf{d}, a)$th row $(W_b^{(1)})_{(\mathbf{d},a),:}$ for any $a \in [n]$ only be nonzero in the $i_{\mathbf{d}}$th index, where it is equal to 1. Thus,

$$(W_b^{(1)} v_b)_{(\mathbf{d},a)} = V_{i_{\mathbf{d}},b}. \tag{90}$$

Hence, the first layer takes the form

$$f_1(V)_{(\mathbf{d},a),:} = \begin{bmatrix} V_{i_{\mathbf{d}},1} \cdot \text{SignNet}_1(V)_{(\mathbf{d},a),1} & \dots & V_{i_{\mathbf{d}},k} \cdot \text{SignNet}_1(V)_{(\mathbf{d},a),k} \end{bmatrix} \in \mathbb{R}^k. \tag{91}$$

Now, for the second layer, we let $\text{SignNet}(V)_{i,j} = 1$ for all $i \in [n], j \in [k]$, which can be represented exactly. Then for each column $b \in [k]$ we will define weight matrices $W_b^{(2)}$ such that $(W_b^{(2)})_{a,(\mathbf{d},i)} = 1$ if $a = i$ and is 0 otherwise. Then we can see that

$$f_2 \circ f_1(V)_{a,b} = \sum_{\mathbf{d}} V_{i_{\mathbf{d}},b} \cdot \text{SignNet}_1(V)_{(d,a),b}. \tag{92}$$

To see that this approximates the polynomial p, for any $V \in \mathcal{X}$ we can bound

$$\begin{aligned}
|p(V)_{a,b} - f_2 \circ f_1(V)_{a,b}| &= \left| \sum_{\mathbf{d}} V_{i_{\mathbf{d}},b} \cdot \left(q_{\text{inv}}(V)_{\mathbf{d},a,b} - \text{SignNet}_1(V)_{(\mathbf{d},a),b} \right) \right| && (93) \\
&\leq \sum_{\mathbf{d}} |V_{i_{\mathbf{d}},b}| \left| \left(q_{\text{inv}}(V)_{\mathbf{d},a,b} - \text{SignNet}_1(V)_{(\mathbf{d},a),b} \right) \right| && (94) \\
&\leq B \sum_{\mathbf{d}} \left| \left(q_{\text{inv}}(V)_{\mathbf{d},a,b} - \text{SignNet}_1(V)_{(\mathbf{d},a),b} \right) \right| && (95) \\
&< B \sum_{\mathbf{d}} \frac{\epsilon}{2B d_{\text{hidden}}} && (96) \\
&\leq \frac{\epsilon}{2} && (97)
\end{aligned}$$

By the triangle inequality, $f_2 \circ f_1$ is ϵ-close to f_{target}, so we are done.

□

D Experimental Details

D.1 Miscellaneous Experimental Details

We ran the experiments on a HPC server with CPUs and GPUs. Each experiment was run on a single NVIDIA V100 GPU with 32GB memory. The runtimes for some of our experiments are included in the main paper. Our codes for our models and experiments will be open-sourced and permissively licensed.

D.2 Link Prediction in Nearly Synthetic Graphs

The base graphs H we generate are Erdös-Renyi or Barabási-Albert graphs with 1000 nodes. We use NetworkX [Hagberg et al., 2008] to generate and process the graphs. The Erdös-Renyi graphs have edge probability $p = .05$ and the Barabási-Albert graphs have $m = 20$ new edges per new node. Let $V = [v_1, \dots, v_k]$ be Laplacian eigenvectors of the graph. We take $k = 16$ in these experiments. The unlearned decoder baseline simply takes the predicted probability of a link between i and j to be proportional to the dot product of the eigenvectors embeddings of node i and node j; this has no learnable parameters. In other words, the node embeddings z_i and z_j are taken to be $V_{i,:}$ and $V_{j,:}$ respectively, and the edge prediction is $z_i^\top z_j$. The learned decoder baseline takes the same z_i and z_j, but takes the edge prediction to be $\mathrm{MLP}(z_i \odot z_j)$. Every other method learns node embeddings z_i and z_j, and takes the edge prediction to be $z_i^\top z_j$.

Each model is restricted to around 25,000 learnable parameters (besides the Unlearned Decoder, which has no parameters). We train each method for 100 epochs with an Adam optimizer [Kingma and Ba, 2015] at a learning rate of .01. The train/validation/test split is 80%/10%/10%, and is chosen uniformly at random.

D.3 Details on n-body Simulations

We follow the experimental setting and build on the code of Puny et al. [2022] (no license as far as we can tell) for the n-body learning task. The code for generating the data stems from Kipf et al. [2018] (MIT License) and Fuchs et al. [2020] (MIT License). There are 3000 training trajectories, 2000 validation trajectories, and 2000 test trajectories. We modify the data generation code to apply to general dimensions $d > 3$. We do not change any of the scaling factors in doing so. For each dimension d, we use the same hyperparameters for both the frame averaging model and the sign equivariant model.

D.4 Node Classification on CLUSTER

In Section 4.3, we show results for the node classification task CLUSTER [Dwivedi et al., 2022a], where the task is to cluster nodes in graphs drawn from Stochastic Block Models [Abbe, 2017]. Models are restricted to a 100k learnable parameter budget. We largely follow the experimental setting of Rampasek et al. [2022], except we report results for the eigenvector based methods on 5 runs instead of 10.

We test several eigenvector based methods within the GraphGPS framework and codebase [Rampasek et al., 2022] (MIT License), which is a state of the art Transformer / GNN hybrid. Firstly, we make use of the PEG style GraphGPS, which means that the MPNN in each GraphGPS layer takes as edge features $e_{ij} = \|V_i - V_j\|^2$, where $V_i \in \mathbb{R}^k$ is the eigenvector embedding of node i. This is fully $O(k)$ invariant (which is much stricter than sign / basis invariance), so we relax this to just be sign invariant in our model by learning a diagonal matrix D such that $e_{ij} = V_i^\top D V_j$. Also, the standard GraphGPS only updates eigenvector representations (in a non-equivariant manner) before most of the neural network modules. When we add our sign equivariant model, we update eigenvector representations within each GraphGPS layer.