EXPRESSIVE SIGN EQUIVARIANT NETWORKS FOR SPECTRAL GEOMETRIC LEARNING

Derek Lim
MIT CSAIL
dereklim@mit.edu

Joshua Robinson
MIT CSAIL

Stefanie Jegelka
MIT CSAIL

Yaron Lipman
Weizmann Institute of Science and Meta AI (FAIR)

Haggai Maron
NVIDIA Research

ABSTRACT

Recent work has shown the utility of developing machine learning models that respect the symmetries of eigenvectors. These works promote sign invariance, since for any eigenvector v the negation $-v$ is also an eigenvector. In this work, we demonstrate that sign *equivariance* is useful for applications such as building orthogonally equivariant models and link prediction. To obtain these benefits, we develop novel sign equivariant neural network architectures. These models are based on our analytic characterization of the sign equivariant polynomials and thus inherit provable expressiveness properties.

1 INTRODUCTION

The need to process eigenvectors is ubiquitous in machine learning and the computational sciences. For instance, there is often a need to process eigenvectors of operators associated with manifolds or graphs, principal components (PCA) of arbitrary datasets, and eigenvectors arising from implicit or explicit matrix factorization methods. However, eigenvectors are not merely unstructured data—they have important structure in the form of symmetries (Rustamov et al., 2007; Lim et al., 2023).

Specifically, eigenvectors have sign and basis symmetries. An eigenvector v is sign symmetric in the sense that the sign-flipped vector $-v$ is also an eigenvector of the same eigenvalue. Basis symmetries occur when there is a repeated eigenvalue, as then there are infinitely many choices of eigenvector basis for the same eigenspace. As such, prior work has developed neural networks that are invariant to these symmetries, improving empirical performance in several settings (Lim et al., 2023).

The first contribution of this work is to show that sign *equivariant* architectures are more natural than sign *invariant* architectures for several applications. First, we show that sign and basis invariant networks are theoretically limited in expressive power for learning edge representations (and more generally multi-node representations) because they learn structural node embeddings that are known to be limited for link prediction and multi-node tasks (Srinivasan & Ribeiro, 2019; Zhang et al., 2021). In contrast, we show that sign equivariant models can bypass this limitation. Furthermore, we show that sign equivariance combined with PCA can be used to parameterize orthogonally equivariant point cloud models, thus giving an efficient alternative to PCA-based frame averaging (Puny et al., 2022; Atzmon et al., 2022).

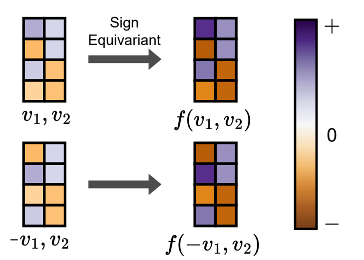


Figure 1: Illustration of a sign equivariant function f. When column 1 of the input is negated, column 1 of the output is also negated.

The second contribution of this work is to develop sign equivariant neural network architectures, with provable expressiveness guarantees. We achieve this by deriving a complete characterization of sign equivariant polynomials, whose form directly leads to our equivariant network architectures and are crucial to our expressivity analysis. Initial numerical experiments support our theory and demonstrate the utility of sign equivariant models.

Preliminaries. Let $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$ be a function that takes eigenvectors $v_1, \ldots, v_k \in \mathbb{R}^n$ of an underlying matrix as input, and outputs representations $f(v_1, \ldots, v_k)$. We often concatenate the eigenvectors into a matrix $V = [v_1, \ldots, v_k] \in \mathbb{R}^{n\times k}$, and then write $f(V)$ as the application of f.

Sign equivariance means that if we flip the sign of an eigenvector, then the corresponding column of the output has its sign flipped. In other words, for all choices of signs $s_1, \ldots, s_k \in \{-1, 1\}^k$,

$$f(s_1 v_1, \ldots, s_k v_k)_{:,j} = s_j f(v_1, \ldots, v_k)_{:,j}, \tag{1}$$

where $A_{:,j}$ is the j-th column of an $n \times k$ matrix A. See Figure 1 for an illustration. In matrix form, letting $\mathrm{diag}(\{-1, 1\}^k)$ represent all $k \times k$ diagonal matrices with -1 or 1 on the diagonal, f is sign equivariant if $f(VS) = f(V)S$ for all $S \in \mathrm{diag}(\{-1, 1\}^k)$.

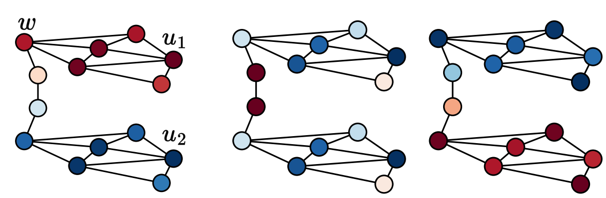
Permutation equivariance is often also a desirable property of our functions f. We say that f is *permutation equivariant* if $f(PV) = Pf(V)$ for all $n \times n$ permutation matrices P. For instance, eigenvectors of matrices associated to simple graphs of size n have such permutation symmetries, as the ordering of nodes is arbitrary (Lim et al., 2023).

2 Applications of Sign Equivariance

In this section, we summarize several applications for which modeling networks with sign equivariant architectures is beneficial. Appendix B.1 also details how sign equivariant architectures could improve sign invariant architectures.

2.1 Multi-Node Representations and Link Prediction

In Figure 2, we plot the first nontrivial Laplacian eigenvector of an example graph, viewed as a node feature. Laplacian eigenvectors are an example of positional encodings (Srinivasan & Ribeiro, 2019) that capture useful information of graphs (Chung, 1997); for instance, u_1 and u_2 are far in the graph, and the eigenvector has very different values on these two nodes. However, u_1 and u_2 are also automorphic, so certain node embeddings — structural encodings — assign them the same representation (e.g. most GNNs). This is known to be problematic for link prediction and other multi-node tasks (Srinivasan & Ribeiro, 2019; Zhang et al., 2021); for instance, a function on structural node encodings assigns both u_1 and u_2 the same probability of connecting to w, whereas a function on a positional encoding (e.g. eigenvectors) can assign a higher probability for u_1 to connect to w.



(a) Eigenvector 1 (b) Sign Invariant (c) Sign Equivariant

Figure 2: (a) First nontrivial normalized Laplacian eigenvector of a graph, which is positional. Nodes u_1 and u_2 are far apart in the graph, but automorphic. (b) Sign invariant node features, which are structural. Nodes u_1 and u_2 have the same feature. (c) Sign equivariant node features, which are positional. Nodes u_1 and u_2 have opposite signs. A link prediction model with sign invariant node features assigns u_1 and u_2 the same probability of connecting to w, while sign equivariant node features could give higher probability to u_1.

When processing eigenvectors of matrices associated to graphs, invariance to the sign and basis symmetries of the eigenvectors has been found useful (Dwivedi et al., 2022a; Lim et al., 2023), especially for graph classification tasks. However, we show that exact invariance to these symmetries removes positional information and thus the outputs of sign invariant or basis invariant networks (Lim et al., 2023) are in fact *structural encodings* (see Appendix B.4).[1] Hence, eigenvector-symmetry-invariant networks cannot learn node representations that distinguish automorphic nodes, and thus face the aforementioned difficulties when used for link prediction or multi-node tasks:

Proposition 1. *Let $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times d_{\text{out}}}$ 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. Denote $Z = f(V)$. Let nodes i and j be automorphic, and let z_i and $z_j \in \mathbb{R}^{d_{\text{out}}}$ be their embeddings.*

[1]When there are repeated eigenvalues, sign invariant embeddings still have some positional information.

If f is sign invariant and the eigenvalues associated to 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 the first k eigenspaces, then $z_i = z_j$.

The problem $z_i = z_j$ arises from the sign/basis invariances. We instead propose using sign *equivariant* networks to learn node representations $z_i = f(V)_{i,:} \in \mathbb{R}^k$. These representations z_i maintain positional information for each node (see Figure 2 (c)). Then we use a sign invariant decoder $f_{\text{decode}}(z_i, z_j) = f_{\text{decode}}(Sz_i, Sz_j)$ for $S \in \text{diag}(\{-1, 1\}^k)$ to obtain edge representations. For instance, the commonly used $f_{\text{decode}} = \text{MLP}(z_i \odot z_j)$ is sign invariant. When the eigenvalues are distinct, this approach has the desired invariances (giving structural edge representations) and also maintains positional information in the node embeddings (see Appendix B.4). More details and the proof of Proposition 1 are in Appendix B.3.

2.2 Orthogonal Equivariance

For various applications in modelling physical systems, we desire equivariance to rigid transformations; thus, orthogonally equivariant models have been a fruitful research direction in recent years (Thomas et al., 2018; Deng et al., 2021; Satorras et al., 2021). We say that a function $f : \mathbb{R}^{n \times k} \to \mathbb{R}^{n \times k}$ is orthogonally equivariant if $f(XQ) = f(X)Q$ for any orthogonal $Q \in O(k)$. Several works have approached this problem using so-called Principal Component Analysis (PCA) based frames (Puny et al., 2022; Atzmon et al., 2022; Xiao et al., 2020).

PCA-frame methods take an input $X \in \mathbb{R}^{n \times k}$, compute orthonormal eigenvectors R_X of the covariance matrix $\text{cov}(X) = (X - \frac{1}{n}\mathbf{1}\mathbf{1}^\top X)^\top (X - \frac{1}{n}\mathbf{1}\mathbf{1}^\top X)$ (assumed to have distinct eigenvalues), then average outputs of a base model h for each of the 2^k sign-flipped inputs XR_XS, where $S \in \text{diag}(\{-1, 1\}^k)$. We instead suggest using a sign equivariant network to parameterize an efficient $O(k)$ equivariant model. For a sign equivariant network h, we define our model f to be

$$f(X) = h(XR_X)R_X^\top. \tag{2}$$

Intuitively, this first transforms X by R_X into a nearly canonical orientation that is unique up to sign flips, processes XR_X using the model h that respects the sign symmetries, then incorporates orientation information back into the output by post-multiplying by $R_X^\top$. Our approach only requires one forward pass through h, whereas frame averaging requires 2^k forward passes. The following proposition shows that f is indeed $O(k)$ equivariant, and inherits universality properties of h.[2]

Proposition 2. *Consider a domain $\mathcal{X} \subseteq \mathbb{R}^{n \times k}$ such that each $X \in \mathcal{X}$ has distinct covariance eigenvalues, and let R_X be a choice of orthonormal eigenvectors of $\text{cov}(X)$ for each $X \in \mathcal{X}$. If $h : \mathcal{X} \subseteq \mathbb{R}^{n \times k} \to \mathbb{R}^{n \times k}$ is sign equivariant, and if $f(X) = h(XR_X)R_X^\top$, then f is well defined and orthogonally equivariant. Moreover, if h is from a universal class of sign equivariant functions, then the f of the above form universally approximate $O(k)$ equivariant functions on $\mathcal{X}$.*

3 Sign Equivariant Polynomials and Networks

In this section, we analytically characterize the sign equivariant polynomials, and use this characterization to develop sign equivariant architectures. As equivariant polynomials universally approximate continuous equivariant functions (Yarotsky, 2022), our architectures inherit universality guarantees.

3.1 Sign Equivariant Polynomials

Consider polynomials $p : \mathbb{R}^{n \times k} \to \mathbb{R}^{n' \times k}$ that are sign equivariant, meaning $p(VS) = p(V)S$ for $S \in \text{diag}(\{-1, 1\}^k)$. We can show that a polynomial p is sign equivariant if and only if it can be written as the elementwise product of a simple (linear) sign equivariant polynomial and a general sign invariant polynomial, followed by another linear sign equivariant map.

Proposition 3. *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*

$$p(V) = W^{(2)} \left((W^{(1)} V) \odot p_{\text{inv}}(V) \right) \tag{3}$$

[2]A class of functions $\mathcal{F}_{\text{model}}$ from $\mathcal{X} \to \mathcal{Y}$ is universal with respect to a target class $\mathcal{F}_{\text{target}}$ if for all compact $\mathcal{D} \subseteq \mathcal{X}$, $f_{\text{target}} \in \mathcal{F}_{\text{target}}$, and $\epsilon > 0$, there is an $f_{\text{model}} \in \mathcal{F}_{\text{model}}$ such that $\|f_{\text{model}}(x) - f_{\text{target}}(x)\| < \epsilon$ for all $x \in \mathcal{D}$.

for sign equivariant linear $W^{(2)}$ and $W^{(1)}$, and a sign invariant polynomial $p_{\text{inv}} : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$.

The proof of this statement is in Appendix C. Sign equivariant linear maps $W : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ take a simple form, as they act independently on each column; $W[v_1, \ldots, v_k] = [W_1 v_1, \ldots, W_k v_k]$ for some arbitrary linear maps $W_1, \ldots, W_k : \mathbb{R}^n \to \mathbb{R}^{n'}$.

3.2 Sign Equivariance without Permutation Symmetries

From the form of the sign equivariant polynomials, we can now develop our sign equivariant architectures. For now, we do not enforce permutation equivariance. We parameterize sign equivariant $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ as a composition of layers f_l, each of the form

$$f_l(V) = [W_1^{(l)} v_1, \ldots, W_k^{(l)} v_k] \odot \text{SignNet}_l(V), \tag{4}$$

in which the $W_i^{(l)} : \mathbb{R}^n \to \mathbb{R}^{n'}$ are arbitrary linear maps, and $\text{SignNet}_l : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$ is sign invariant (Lim et al., 2023). In the case of $n = n' = 1$, there is a simple universal form; we can write sign equivariant $f : \mathbb{R}^k \to \mathbb{R}^k$ as $f(v) = v \odot \text{MLP}(|v|)$, where $|v|$ is the elementwise absolute value. These two architectures are universal because they can approximate sign equivariant polynomials.

Proposition 4. *Functions of the form $v \mapsto v \odot \text{MLP}(|v|)$ 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 4 universally approximate continuous sign equivariant functions $f : \mathbb{R}^{n\times k} \to \mathbb{R}^{n'\times k}$.

3.3 Sign Equivariance and Permutation Equivariance

We can build on the above architectures that *do not* satisfy permutation equivariance to develop sign equivariant architectures that *are* permutation equivariant as well. A natural way to do this is by using the DeepSets for Symmetric Elements (DSS) framework (Maron et al., 2020). Each layer $f_l : \mathbb{R}^{n\times k} \to \mathbb{R}^{n\times k}$ of our DSS-based sign equivariant network takes the following form on row i:

$$f_l(V)_{i,:} = f_l^{(1)}(V_{i,:}) + f_l^{(2)}\Big(\sum_{j\neq i} V_{j,:}\Big) \tag{5}$$

Where $f_l^{(1)}$ and $f_l^{(2)}$ are sign equivariant functions as in Section 3.2. DSS has universal approximation guarantees (Maron et al., 2020), but they only apply for groups that act as permutation matrices, whereas our sign group $\{-1, 1\}^k$ does not. The universal approximation properties of our proposed DSS-based architecture are still an open question.

4 Experiments

In this section, we experimentally test our sign equivariant model on a link prediction task; more experiments on other tasks are in Appendix E.

Table 1: Link prediction AUC and runtime per epoch for structural edge models.

Model	Test AUC	Runtime (s)
GCN (constant input)	.497±.01	.120±.00
SignNet	.504±.01	1.24±.26
$V_{i,:}^\top V_{j,:}$	.562±.02	.005±.01
$\text{MLP}(V_{i,:} \odot V_{j,:})$	.629±.03	.052±.01
Sign Equivariant	**.737**±.01	.117±.02

Link Prediction in Nearly Symmetric Graphs. We present a synthetic link prediction task that numerically demonstrates the benefits of sign equivariance in link prediction, as theoretically explained in Section 2.1. First, we generate a random graph H of 500 nodes. Then we form a larger graph G that contains two disjoint copies of H, along with 1000 uniformly-randomly added edges Without the random edges, each node in one copy of H is automorphic to the corresponding node in the other copy. In Table 1, we show the link prediction performance of several models that learn structural edge representations — the methods that use eigenvectors have a sign invariant final prediction for each edge. GCN (Kipf & Welling, 2017) with all ones input and SignNet (Lim et al., 2023) both completely fail at this task (these two models map automorphic nodes to the same embedding), while our sign equivariant model outperforms all methods. Further, the equivariant model takes comparable runtime to GCN, and is significantly faster than SignNet. See Appendix F.1 for more details.

ACKNOWLEDGMENTS

We would like to thank Maks Ovsjanikov for noting that basis invariant models give automorphic nodes the same representation, and also Johannes Lutzeyer and Michael Murphy for helpful comments. DL is supported by an NSF Graduate Fellowship. SJ acknowledges support from NSF AI Institute NSF CCF-2112665, and NSF Award 2134108 and NSF CAREER Award 1553284.

REFERENCES

Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

Matan Atzmon, Koki Nagano, Sanja Fidler, Sameh Khamis, and Yaron Lipman. Frame averaging for equivariant shape space learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 631–641, 2022.

Fan Chung. *Spectral graph theory*. American Mathematical Soc., 1997.

Taco Cohen and Max Welling. Group equivariant convolutional networks. In *International conference on machine learning*, pp. 2990–2999. PMLR, 2016.

Leonardo Cotta, Beatrice Bevilacqua, Nesreen Ahmed, and Bruno Ribeiro. Causal lifting and link prediction. *arXiv preprint arXiv:2302.01198*, 2023.

Congyue Deng, Or Litany, Yueqi Duan, Adrien Poulenard, Andrea Tagliasacchi, and Leonidas J Guibas. Vector neurons: A general framework for so (3)-equivariant networks. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 12200–12209, 2021.

Vijay Prakash Dwivedi, Chaitanya K Joshi, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Benchmarking graph neural networks. In *JMLR*, 2022a.

Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson. Graph neural networks with learnable structural and positional representations. In *Int. Conference on Learning Representations (ICLR)*, volume 10, 2022b.

Nadav Dym and Steven J Gortler. Low dimensional invariant embeddings for universal geometric learning. *arXiv preprint arXiv:2205.02956*, 2022.

Nadav Dym and Haggai Maron. On the universality of rotation equivariant point cloud networks. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=6NFBvWlRXaG.

Fabian Fuchs, Daniel Worrall, Volker Fischer, and Max Welling. Se (3)-transformers: 3d roto-translation equivariant attention networks. *Advances in Neural Information Processing Systems*, 33:1970–1981, 2020.

Simon Geisler, Yujia Li, Daniel Mankowitz, Ali Taylan Cemgil, Stephan Günnemann, and Cosmin Paduraru. Transformers meet directed graphs. *arXiv preprint arXiv:2302.00049*, 2023.

Justin Gilmer, Samuel S Schoenholz, Patrick F Riley, Oriol Vinyals, and George E Dahl. Neural message passing for quantum chemistry. In *International conference on machine learning*, pp. 1263–1272. PMLR, 2017.

Xiaoxin He, Bryan Hooi, Thomas Laurent, Adam Perold, Yann LeCun, and Xavier Bresson. A generalization of vit/mlp-mixer to graphs. *arXiv preprint arXiv:2212.13350*, 2022.

Jinwoo Kim, Tien Dat Nguyen, Seonwoo Min, Sungjun Cho, Moontae Lee, Honglak Lee, and Seunghoon Hong. Pure transformers are powerful graph learners. *arXiv preprint arXiv:2207.02505*, 2022.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *Int. Conference on Learning Representations (ICLR)*, volume 5, 2017.

Devin Kreuzer, Dominique Beaini, Will Hamilton, Vincent Létourneau, and Prudencio Tossou. Rethinking graph transformers with spectral attention. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, 2021.

Yann LeCun, Bernhard Boser, John Denker, Donnie Henderson, Richard Howard, Wayne Hubbard, and Lawrence Jackel. Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, 2, 1989.

Derek Lim, Joshua David Robinson, Lingxiao Zhao, Tess Smidt, Suvrit Sra, Haggai Maron, and Stefanie Jegelka. Sign and basis invariant networks for spectral graph representation learning. In *International Conference on Learning Representations*, 2023.

Haggai Maron, Heli Ben-Hamu, Nadav Shamir, and Yaron Lipman. Invariant and equivariant graph networks. In *Int. Conference on Learning Representations (ICLR)*, volume 6, 2018.

Haggai Maron, Ethan Fetaya, Nimrod Segol, and Yaron Lipman. On the universality of invariant networks. In *Int. Conference on Machine Learning (ICML)*, pp. 4363–4371. PMLR, 2019.

Haggai Maron, Or Litany, Gal Chechik, and Ethan Fetaya. On learning sets of symmetric elements. In *International Conference on Machine Learning*, pp. 6734–6744. PMLR, 2020.

Grégoire Mialon, Dexiong Chen, Margot Selosse, and Julien Mairal. GraphiT: Encoding graph structure in transformers. In *preprint arXiv:2106.05667*, 2021.

Michael Murphy, Stefanie Jegelka, Ernest Fraenkel, Tobias Kind, David Healey, and Thomas Butler. Efficiently predicting high resolution mass spectra with graph neural networks. *arXiv preprint arXiv:2301.11419*, 2023.

Omri Puny, Matan Atzmon, Heli Ben-Hamu, Edward J Smith, Ishan Misra, Aditya Grover, and Yaron Lipman. Frame averaging for invariant and equivariant network design. In *Int. Conference on Learning Representations (ICLR)*, volume 10, 2022.

Ladislav Rampasek, Mikhail Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini. Recipe for a general, powerful, scalable graph transformer. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=lMMaNf6oxKM.

Raif M Rustamov et al. Laplace-beltrami eigenfunctions for deformation invariant shape representation. In *Symposium on geometry processing*, volume 257, pp. 225–233, 2007.

Vıctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E (n) equivariant graph neural networks. In *International conference on machine learning*, pp. 9323–9332. PMLR, 2021.

Nimrod Segol and Yaron Lipman. On universal equivariant set networks. In *Int. Conference on Learning Representations (ICLR)*, volume 7, 2019.

Balasubramaniam Srinivasan and Bruno Ribeiro. On the equivalence between positional node embeddings and structural graph representations. In *Int. Conference on Learning Representations (ICLR)*, 2019.

Nathaniel Thomas, Tess Smidt, Steven Kearnes, Lusann Yang, Li Li, Kai Kohlhoff, and Patrick Riley. Tensor field networks: Rotation-and translation-equivariant neural networks for 3d point clouds. *arXiv preprint arXiv:1802.08219*, 2018.

Soledad Villar, David Hogg, Kate Storey-Fisher, Weichi Yao, and Ben Blum-Smith. Scalars are universal: Equivariant machine learning, structured like classical physics. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, 2021.

Haorui Wang, Haoteng Yin, Muhan Zhang, and Pan Li. Equivariant and stable positional encoding for more powerful graph neural networks. In *Int. Conference on Learning Representations (ICLR)*, volume 10, 2022.

Zelin Xiao, Hongxin Lin, Renjie Li, Lishuai Geng, Hongyang Chao, and Shengyong Ding. Endowing deep 3d models with rotation invariance based on principal component analysis. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. IEEE, 2020.

Dmitry Yarotsky. Universal approximations of invariant maps by neural networks. *Constructive Approximation*, 55(1):407–474, 2022.

Jiaxuan You, Jonathan M Gomes-Selman, Rex Ying, and Jure Leskovec. Identity-aware graph neural networks. In *Association for the Advancement of Artificial Intelligence (AAAI)*, volume 35, pp. 10737–10745, 2021.

Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 30, pp. 3391–3401, 2017.

Muhan Zhang and Yixin Chen. Link prediction based on graph neural networks. *Advances in neural information processing systems*, 31, 2018.

Muhan Zhang, Pan Li, Yinglong Xia, Kai Wang, and Long Jin. Labeling trick: A theory of using graph neural networks for multi-node representation learning. *Advances in Neural Information Processing Systems*, 34:9061–9073, 2021.

Zhaocheng Zhu, Zuobai Zhang, Louis-Pascal Xhonneux, and Jian Tang. Neural bellman-ford networks: A general graph neural network framework for link prediction. *Advances in Neural Information Processing Systems*, 34:29476–29490, 2021.

A Related Work

Especially for link prediction, the need for structural node-pair representations that are not obtained from structural node representations has been discussed in several works (Srinivasan & Ribeiro, 2019; Zhang et al., 2021; Cotta et al., 2023). As such, several works have developed methods for learning structural node-pair representations that are not limited by structural node representations. SEAL and other labeling-trick based methods (Zhang & Chen, 2018; Zhang et al., 2021) use added node features depending on the node-pair that we want a representation of. This is empirically successful in many tasks, but typically requires a separate forward pass through a GNN for each node-pair under consideration. PEG (Wang et al., 2022) maintains positional information by using eigenvector distances in each layer of a GNN, but do not update eigenvector representations. Identity-aware GNNs (You et al., 2021) and Neural Bellman-Ford Networks (Zhu et al., 2021) learn pair representations by conditioning on a source node from the pair.

When using eigenvectors of graphs as node positional encodings for graph models like GNNs and Graph Transformers, many works have noted the need to deal with the sign ambiguity of the eigenvectors. This is often done by encouraging sign invariance through data augmentation — the signs of the eigenvectors are chosen randomly in each iteration of training (Dwivedi et al., 2022a;b; Kreuzer et al., 2021; Mialon et al., 2021; Kim et al., 2022; He et al., 2022). In contrast, SignNet (Lim et al., 2023) enforces exact sign invariance, by processing eigenvectors with a sign invariant neural architecture; this approach has been taken by some recent works (Rampasek et al., 2022; Geisler et al., 2023; Murphy et al., 2023).

B Applications of Sign Equivariance

B.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 (Lim et al., 2023). SignNet (Lim et al., 2023), a sign invariant neural network, takes the form

$$f(v_1, \ldots, v_k) = \rho(\phi(v_1) + \phi(-v_1), \ldots, \phi(v_k) + \phi(-v_k)) \tag{6}$$

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 & 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 B.2)

B.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 3.3.

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 $\mathcal{O}(nd)$.

B.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.

B.2.2 Efficient Sign Invariant Networks

Consider again the form of SignNet (Lim et al., 2023), $f(V) = \rho([\phi(v_i) + \phi(-v_i)]_{i=1,\dots,k})$. 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.

B.3 Edge Representations and Link Prediction

B.3.1 Sign Invariant Link Prediction Decoders

We have an ansatz for universal permutation invariant and sign invariant functions at $n = 2$, 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 at $n = 1$, where there are no permutation symmetries (Lim et al., 2023).

We will parameterize such functions as

$$f(v_1, \dots v_k) = \varphi\left(v_1 \odot v_1, v_1 \odot \text{rev}(v_1), \dots, v_k \odot v_k, v_k \odot \text{rev}(v_k)\right). \tag{7}$$

Here, $\text{rev} : \mathbb{R}^2 \to \mathbb{R}^2$ reverses the vector, so $\text{rev}(a)_1 = a_2$ and $\text{rev}(a)_2 = a_1$. 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. by DeepSets. Now, we show that this parameterization is universal:

Proposition 5. *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

$$\begin{aligned}
f(Pv_1s_1, \dots, Pv_ks_k) &= \varphi\left((Pv_1s_1) \odot (Pv_1s_1), (Pv_1s_1) \odot \text{rev}(Pv_1s_1), \dots\right) && (8)\\
&= \varphi\left(P(v_1s_1 \odot v_1s_1), P(v_1s_1 \odot \text{rev}(v_1s_1)), \dots\right) && (9)\\
&= \varphi\left(P(v_1 \odot v_1), P(v_1 \odot \text{rev}(v_1)), \dots\right) && (10)\\
&= \varphi\left(v_1 \odot v_1, v_1 \odot \text{rev}(v_1), \dots\right) && (11)\\
&= f(v_1, \dots, v_k), && (12)
\end{aligned}$$

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

$$h(v_1, \dots, v_k) = \rho(\phi(v_1v_1^\top), \dots, \phi(v_kv_k^\top)), \tag{13}$$

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

$$\psi(PA_1P^\top, \dots, PA_kP^\top) = \psi(A_1, \dots, A_k), \tag{14}$$

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

$$\psi(PA_1P^\top, \dots, PA_kP^\top) = h(Pv_1, \dots, Pv_k) = h(v_1, \dots, v_k) = \psi(A_1, \dots, A_k), \tag{15}$$

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

$$C = \{[v_1 \odot v_1, v_1 \odot \text{rev}(v_1), \dots, v_k \odot v_k, v_k \odot \text{rev}(v_k)] : v_i \in \mathbb{R}^2\}. \tag{16}$$

We define φ by

$$\varphi(A) = \psi\left(\begin{bmatrix} A_{1,1} & A_{2,2} \\ A_{2,2} & A_{2,1} \end{bmatrix}, \begin{bmatrix} A_{1,3} & A_{2,4} \\ A_{2,4} & A_{2,3} \end{bmatrix}, \dots, \begin{bmatrix} A_{1,2k-1} & A_{2,2k} \\ A_{2,2k} & A_{2,2k-1} \end{bmatrix}\right). \tag{17}$$

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

$$\varphi(PA) = \psi\left(\begin{bmatrix} A_{2,1} & A_{1,2} \\ A_{1,2} & A_{1,1} \end{bmatrix}, \begin{bmatrix} A_{2,3} & A_{1,4} \\ A_{1,4} & A_{1,3} \end{bmatrix}, \dots, \begin{bmatrix} A_{2,2k-1} & A_{1,2k} \\ A_{1,2k} & A_{1,2k-1} \end{bmatrix}\right) \tag{18}$$

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

$$= \psi\left(\begin{bmatrix} A_{1,1} & A_{2,2} \\ A_{2,2} & A_{2,1} \end{bmatrix}, \begin{bmatrix} A_{1,3} & A_{2,4} \\ A_{2,4} & A_{2,3} \end{bmatrix}, \dots, \begin{bmatrix} A_{1,2k-1} & A_{2,2k} \\ A_{2,2k} & A_{2,2k-1} \end{bmatrix}\right) \quad (\psi \text{ perm. invariant}) \tag{20}$$

$$= \varphi(A), \tag{21}$$

where in the second equality, we use the fact that $A_{2,2j} = A_{1,2j}$, $j = 1, \dots, k$ 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 $f : \mathbb{R}^{2\times k} \to \mathbb{R}^{d_{\text{out}}}$ using this φ, we compute that

$$f(v_1, \dots v_k) = \varphi\left(v_1 \odot v_1, v_1 \odot \mathrm{rev}(v_1), \dots, v_k \odot v_k, v_k \odot \mathrm{rev}(v_k)\right) \tag{22}$$

$$= \psi\left(\begin{bmatrix} v_{1,1}^2 & v_{1,1}v_{1,2} \\ v_{1,1}v_{1,2} & v_{1,2}^2 \end{bmatrix}, \dots, \begin{bmatrix} v_{k,1}^2 & v_{k,1}v_{k,2} \\ v_{k,1}v_{k,2} & v_{k,2}^2 \end{bmatrix}\right) \tag{23}$$

$$= \psi\left(v_1 v_1^\top, \dots, v_k v_k^\top\right) \tag{24}$$

$$= h(v_1, \dots, v_k), \tag{25}$$

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. □

B.3.2 PROOF OF PROPOSITION 1

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

If f is sign invariant and the eigenvalues associated to the v_l are distinct, then $z_i = z_j$.

If f is basis invariant and $V_{\lambda_i} \cap \{v_1, \dots, v_k\} = \dim(V_{\lambda_i})\ \forall i \in [k]$, 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. The condition $V_{\lambda_i} \cap \{v_1, \dots, v_k\} = \dim(V_{\lambda_i})\ \forall i \in [k]$ means that if there is an eigenvector from some eigenspace in the input $\{v_1, \dots, v_k\}$, then there is a basis for that eigenspace in $\{v_1, \dots, v_k\}$ (e.g. we don't allow taking as input only one eigenvector of a two dimensional eigenspace).

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 $V_t = [v_{r_1}, \dots, v_{r_{d_t}}]$ 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

$$A(Pv_{r_l}) = PAP^\top(Pv_{r_l}) = PAv_{r_l} = P\lambda_i v_{r_l} = \lambda_t(Pv_{r_l}), \tag{26}$$

so Pv_{r_l} is also an eigenvector of A with eigenvalue λ_t. As P is orthogonal, note that $Pv_{r_1}, \dots, Pv_{r_{d_t}}$ 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 $PV_t = V_t Q_t$ — see Lim et al. (2023).

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

$$\begin{aligned} z_j &= f(V_1, \dots, V_l)_{j,:} \\ &= f(V_1 Q_1, \dots, V_l Q_l)_{j,:} && \text{basis invariance} \\ &= f(PV_1, \dots, PV_l)_{j,:} && \text{choice of } Q_t \\ &= (Pf(V_1, \dots, V_l))_{j,:} && \text{permutation equivariance} \\ &= f(V_1, \dots, V_l)_{i,:} && \text{choice of } P \\ &= z_i. \end{aligned}$$

So we are done. □

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

In this section, we show that when the eigenvalues $\lambda_1, \dots, \lambda_k$ are distinct, then sign invariant functions of the orthonormal eigenvectors $v_1, \dots, v_k$ 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 a formal definitions.

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

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.

For the below proposition, 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, $f_{\text{node}}(Pv_1 s_1, \dots, Pv_k s_k) = Pf_{\text{node}}(v_1, \dots, v_k)$ 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 6. *Let $\mathcal{A} \subseteq \mathbb{R}^{n\times n}$ denote the matrices with distinct first-k eigenvalues. For $A \in \mathcal{A}$, let $V(A) = [v_1(A), \dots, v_k(A)]$ be orthonormal eigenvectors of A, associated to the first-k (distinct) eigenvalues $\lambda_1(A), \dots, \lambda_k(A)$. (These eigenvectors are thus defined up to a sign flip). 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. We prove that these functions are well-defined, because a general function applied to the eigenvectors may not be due to sign ambiguity; in our case, the choice of signs does not affect the output of the function.

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

$$f_{\text{node}}\left(f_{\text{equiv}}(V(A)S)\right) = f_{\text{node}}\left(f_{\text{equiv}}(V(A))S\right) = f_{\text{node}}\left(f_{\text{equiv}}(V(A)\right), \tag{27}$$

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

$$
\begin{aligned}
q_{\text{node}}(PAP^\top) &= f_{\text{node}}\left(f_{\text{equiv}}(V(PAP^\top))\right) && (28)\\
&= f_{\text{node}}\left(f_{\text{equiv}}(PV(A))\right) && (29)\\
&= Pf_{\text{node}}\left(f_{\text{equiv}}(V(A))\right) && (30)\\
&= Pq_{\text{node}}(A) && (31)
\end{aligned}
$$

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

$$
\begin{aligned}
q_{\text{pair}}(PAP^\top)_{i,j} &= f_{\text{decode}}\left(f_{\text{equiv}}(V(PAP^\top))_{i,:}, f_{\text{equiv}}(V(PAP^\top))_{j,:}\right) && (32)\\
&= f_{\text{decode}}\left(f_{\text{equiv}}(PV(A))_{i,:}, f_{\text{equiv}}(PV(A))_{j,:}\right) && (33)\\
&= f_{\text{decode}}\left([Pf_{\text{equiv}}(V(A))]_{i,:}, [Pf_{\text{equiv}}(V(A))]_{j,:}\right) && (34)\\
&= f_{\text{decode}}\left(f_{\text{equiv}}(V(A))_{\sigma^{-1}(i),:}, f_{\text{equiv}}(V(A))_{\sigma^{-1}(j),:}\right) && (35)\\
&= q_{\text{pair}}(A)_{\sigma^{-1}(i),\sigma^{-1}(j)} && (36)\\
&= (Pq_{\text{pair}}(A)P^\top)_{i,j} && (37)
\end{aligned}
$$

□

B.5 Proof of Proposition 2

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 $\text{cov}(X)$ 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_XS is an orthonormal set of eigenvectors of $\text{cov}(X)$ for $S \in \text{diag}(\{-1,1\}^k)$. However, no matter the choice of signs, $f(X)$ takes the same value, since

$$
\begin{aligned}
h(XR_XS)(R_XS)^\top &= h(XR_XS)S^\top R_X^\top && && (38)\\
&= h(XR_X)SS^\top R_X^\top && \text{sign equivariance} && (39)\\
&= h(XR_X)R_X^\top. && && (40)
\end{aligned}
$$

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

$$
\text{cov}(XQ) = \left(XQ - \frac{1}{n}\mathbf{1}\mathbf{1}^\top XQ\right)^\top \left(XQ - \frac{1}{n}\mathbf{1}\mathbf{1}^\top XQ\right) = Q^\top \text{cov}(X)Q. \quad (41)
$$

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

$$
\begin{aligned}
f(XQ) &= h(XQR_{XQ})R_{XQ}^\top && && (42)\\
&= h(XQQ^\top R_XS)(Q^\top R_XS)^\top && && (43)\\
&= h(XR_X)SS^\top R_X^\top Q && \text{sign equivariance} && (44)\\
&= h(XR_X)R_X^\top Q && && (45)\\
&= f(X)Q^\top, && && (46)
\end{aligned}
$$

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 $S \in \text{diag}(\{-1,1\}^k)$ 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

$$\begin{aligned}
\|f_{\text{target}}(X) - f(X)\| &= \left\|f_{\text{target}}(X) - h(XR_X)R_X^\top\right\| && (47)\\
&= \left\|f_{\text{target}}(X)R_X R_X^\top - h(XR_X)R_X^\top\right\| && R_X \text{ orthogonal} \quad (48)\\
&= \left\|f_{\text{target}}(XR_X)R_X^\top - h(XR_X)R_X^\top\right\| && \text{orthogonal equivariance} \quad (49)\\
&= \|f_{\text{target}}(XR_X) - h(XR_X)\| && R_X \text{ orthogonal} \quad (50)\\
&< \epsilon. && (51)
\end{aligned}$$

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

C 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 & Lipman, 2019; Dym & Maron, 2021; Maron et al., 2019; 2020; Villar et al., 2021; Dym & Gortler, 2022).

C.1 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

$$p(v) = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{d_1} \cdots v_k^{d_k} \tag{52}$$

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

$$\sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{d_1} \cdots v_k^{d_k} = p(v) = p(Sv) = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}_{d_1,\ldots,d_k} s_1^{d_1} \cdots s_k^{d_k} v_1^{d_1} \cdots v_k^{d_k}. \tag{53}$$

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

$$\mathbf{W}_{d_1,\ldots,d_k} = s_1^{d_1} \cdots s_k^{d_k} \mathbf{W}_{d_1,\ldots,d_k}, \quad \text{for all } (s_1,\ldots,s_k) \in \{-1,1\}^k. \tag{54}$$

Note that $s_i^{d_i} = 1$ if d_i is an even number. Hence, there are no constraints on $\mathbf{W}_{d_1,\ldots,d_k}$ 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

$$\mathbf{W}_{d_1,\ldots,d_k} = \begin{cases} \text{free} & d_i \text{ even for each } i \\ 0 & \text{else} \end{cases} \tag{55}$$

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 if sign invariant, so we have the following proposition:

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

$$p(v) = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}, \tag{56}$$

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

C.2 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

$$p(v)_l = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} v_1^{d_1} \cdots v_k^{d_k}, \tag{57}$$

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

$$s_l \cdot p(v)_l = p(Sv)_l \tag{58}$$

$$s_l \cdot \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} v_1^{d_1} \cdots v_k^{d_k} = \sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} s_1^{d_1} \cdots s_k^{d_k} v_1^{d_1} \cdots v_k^{d_k}. \tag{59}$$

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

$$s_l \mathbf{W}^{(l)}_{d_1,\ldots,d_k} = s_1^{d_1} \cdots s_k^{d_k} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} \tag{60}$$

$$\mathbf{W}^{(l)}_{d_1,\ldots,d_k} = s_l \cdot s_1^{d_1} \cdots s_k^{d_k} \mathbf{W}^{(l)}_{d_1,\ldots,d_k}, \tag{61}$$

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

$$\mathbf{W}^{(l)}_{d_1,\ldots,d_k} = \begin{cases} \text{free} & d_l \text{ odd, and } d_i \text{ even for each } i \neq l \\ 0 & \text{else} \end{cases}. \tag{62}$$

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 8.

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

$$p(v)_l = v_l \cdot \left(\sum_{d_1,\ldots,d_k=0}^{D} \mathbf{W}^{(l)}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k} \right). \tag{63}$$

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

C.3 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 $\text{diag}(\{-1, 1\}^k)$. This is the case we most often deal with in practice, when we have input $V = [v_1 \quad \ldots \quad v_k]$ 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

$$p(V)_{a,b} = \sum_{d_{i,j}=0}^{D} \mathbf{W}^{(a,b)}_{\mathbf{d}} \prod_{i=1}^{n} \prod_{j=1}^{k} V_{i,j}^{d_{i,j}}, \tag{64}$$

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:

$$s_b \cdot p(V)_{a,b} = p(VS)_{a,b} \tag{65}$$

$$s_b \cdot \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} \prod_{i=1}^{n} \prod_{j=1}^{k} V_{i,j}^{d_{i,j}} = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} s_1^{\tilde{d}_1} \cdots s_k^{\tilde{d}_k} \prod_{i=1}^{n} \prod_{j=1}^{k} V_{i,j}^{d_{i,j}}, \tag{66}$$

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

$$\mathbf{W}_{\mathbf{d}}^{(a,b)} = s_b \cdot s_1^{\tilde{d}_1} \cdots s_k^{\tilde{d}_k} \cdot \mathbf{W}_{\mathbf{d}}^{(a,b)}. \tag{67}$$

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

$$\mathbf{W}_{\mathbf{d}}^{(a,b)} = \begin{cases} \text{free} & \sum_i d_{i,b} \text{ odd, and } \sum_i d_{i,j} \text{ even for each } j \neq b \\ 0 & \text{else} \end{cases}. \tag{68}$$

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 $V_{i_{\mathbf{d}},b} \cdot p_{\text{inv}}(V)_{\mathbf{d}}$ 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 9. *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*

$$p(V)_{a,b} = \sum_{d_{i,j}=0}^{D} \mathbf{W}_{\mathbf{d}}^{(a,b)} V_{i_{\mathbf{d}},b} \cdot p_{\text{inv}}(V)_{\mathbf{d}}, \tag{69}$$

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 Proposition 3. In particular, we will write p in the form

$$p(V) = W^{(2)} \left((W^{(1)} V) \odot q_{\text{inv}}(V) \right), \tag{70}$$

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 69 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 $q_{\text{inv}}(V)_{\mathbf{d},a} = p_{\text{inv}}(V)_{\mathbf{d}}$.

Finally, we define $W^{(2)}$ to compute the sum in equation 69. 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.

D Sign Equivariant Architecture Universality

In this section, we prove Proposition 4 on the universality of our proposed sign equivariant architectures, which we restate here:

Proposition 4. *Functions of the form $v \mapsto v \odot \text{MLP}(|v|)$ 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 4 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.

D.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 $f_{\text{target}} : \mathcal{X} \to \mathbb{R}^k$ 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 8, 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

$$
\begin{aligned}
|f(v)_i - p(v)_i| &= |v_i \cdot \mathrm{MLP}(|v|)_i - v_i \cdot \sum_{d=1}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}| && (71)\\
&= |v_i||\mathrm{MLP}(|v|)_i - \sum_{d=1}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}| && (72)\\
&\leq B \cdot |\mathrm{MLP}(|v|)_i - \sum_{d=1}^{D} \mathbf{W}_{d_1,\ldots,d_k} v_1^{2d_1} \cdots v_k^{2d_k}| && (73)\\
&< \epsilon/2, && (74)
\end{aligned}
$$

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

D.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 $f_{\text{target}} : \mathcal{X} \to \mathbb{R}^{n'\times k}$ 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 9, we can write

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

for some sign invariant polynomials $p_{\text{inv}}(V)_{\mathbf{d}}$. 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 $q_{\text{inv}}(V)_{\mathbf{d},a,b} = \mathbf{W}_{\mathbf{d}}^{(a,b)} p_{\text{inv}}(V)_{\mathbf{d}}$, so we can write

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

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

$$|\mathrm{SignNet}_1(V)_{(\mathbf{d},a),b} - q_{\text{inv}}(V)_{\mathbf{d},a,b}| < \frac{\epsilon}{2B \cdot d_{\text{hidden}}}. \tag{75}$$

Table 2: Node clustering accuracy on CLUSTER (Dwivedi et al., 2022a), 100k parameter budget. Runtime is in seconds per epoch.

Pos. Enc.	Train Acc. (%)	Test Acc. (%)	Runtime
None	$60.9_{\pm 0.08}$	$60.0_{\pm 0.05}$	5.4
LapPE (flip)	$74.3_{\pm 0.39}$	$73.5_{\pm 0.13}$	5.4
SignNet	$76.5_{\pm 0.13}$	$74.0_{\pm 0.04}$	10.2
Sign Equivariant	$\mathbf{77.0_{\pm 0.28}}$	$\mathbf{74.3_{\pm 0.16}}$	7.0

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,

$$(W_b^{(1)} v_b)_{(\mathbf{d},a)} = V_{i_{\mathbf{d}},b}. \tag{76}$$

Hence, the first layer takes the form

$$f_1(V)_{(\mathbf{d},a),:} = \begin{bmatrix} V_{i_{\mathbf{d}},1} \cdot \text{SignNet}_1(V)_{(\mathbf{d},a),1} & \dots & V_{i_{\mathbf{d}},k} \cdot \text{SignNet}_1(V)_{(\mathbf{d},a),k} \end{bmatrix} \in \mathbb{R}^k. \tag{77}$$

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

$$f_2 \circ f_1(V)_{a,b} = \sum_{\mathbf{d}} V_{i_{\mathbf{d}},b} \cdot \text{SignNet}_1(V)_{(d,a),b}. \tag{78}$$

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

$$\begin{aligned} |p(V)_{a,b} - f_2 \circ f_1(V)_{a,b}| &= \left| \sum_{\mathbf{d}} V_{i_{\mathbf{d}},b} \cdot \left(q_{\text{inv}}(V)_{\mathbf{d},a,b} - \text{SignNet}_1(V)_{(\mathbf{d},a),b} \right) \right| & (79) \\ &\leq \sum_{\mathbf{d}} |V_{i_{\mathbf{d}},b}| \left| \left(q_{\text{inv}}(V)_{\mathbf{d},a,b} - \text{SignNet}_1(V)_{(\mathbf{d},a),b} \right) \right| & (80) \\ &\leq B \sum_{\mathbf{d}} \left| \left(q_{\text{inv}}(V)_{\mathbf{d},a,b} - \text{SignNet}_1(V)_{(\mathbf{d},a),b} \right) \right| & (81) \\ &< B \sum_{\mathbf{d}} \frac{\epsilon}{2Bd_{\text{hidden}}} & (82) \\ &\leq \frac{\epsilon}{2} & (83) \end{aligned}$$

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

□

E Further Experimental Results

E.1 Node Classification on CLUSTER

Beyond link prediction and multi-node tasks, expressive positional node embeddings can also be useful in node-level prediction tasks. For instance, consider the node-level task of community detection or semi-supervised node clustering. Here we want to give nearby nodes the same label. Far-away nodes may be automorphic or nearly automorphic, and a structural node encoding method would force these nodes to get the same lable.

In Table 2, 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 see that the sign equivariant model outperforms the other models, and takes less time per epoch than SignNet.

F Experimental Details

F.1 Link Prediction in Nearly Synthetic Graphs

The base graphs H we generate are Erdös-Renyi graphs with 500 nodes and $p = .05$ probability for each edge to be included. Let $V = [v_1, \ldots, v_k]$ 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 & Ba, 2014) at a learning rate of .01. The train/validation/test split is 80%/10%/10%, and is chosen uniformly at random.