Extended Abstract Track

Convergence of Manifold Filter-Combine Networks

Abstract

In order to better understand manifold neural networks (MNNs), we introduce Manifold Filter-Combine Networks (MFCNs). The filter-combine framework parallels the popular aggregate-combine paradigm for graph neural networks (GNNs) and naturally suggests many interesting families of MNNs which can be interpreted as the manifold analog of various popular GNNs. We then propose a method for implementing MFCNs on high-dimensional point clouds that relies on approximating the manifold by a sparse graph. We prove that our method is consistent in the sense that it converges to a continuum limit as the number of data points tends to infinity.

Keywords: Geometric Deep Learning, Manifold Learning, Spectral Convergence

1. Introduction

Geometric deep learning (Bronstein et al., 2017, 2021) aims to extend the success of deep learning from data such as images, with a regular grid-like structure, to more irregular domains such as graphs and manifolds. Notably, graph neural networks (GNNs), e.g., have rapidly emerged as an extremely active area of research (Zhou et al., 2020).

By contrast, the manifold side of geometric deep learning is much less explored, and much of the existing literature on manifold deep learning is limited to 2D surfaces (Masci et al., 2015a,b; Schonsheck et al., 2022) and cannot be applied to higher-dimensional manifolds. This is despite the fact that unsupervised manifold learning algorithms (Coifman and Lafon, 2006; van der Maaten and Hinton, 2008) are commonly used for representing higher-dimensional data (Van Dijk et al., 2018; Moyle et al., 2021)).

Inspired by the successes of GNNs and manifold learning, several recent works have proposed *manifold neural networks* (MNNs) (Wang et al., 2021a,b) that define convolution in terms of a manifold Laplacian $\mathcal{L}$, paralleling spectral GNNs (Defferrard et al., 2016; Levie et al., 2018) that define convolution in terms of the graph Laplacian $\mathbf{L}$. Moreover, several recent papers Chew et al. (2024, 2022a,b); Wang et al. (2023) have introduced numerical methods for implementing MNNs on point clouds satisfying the manifold hypothesis and establish the accuracy and statistical consistency of these methods by proving they converge to a continuum limit as the number of data points tends to infinity under various assumptions.

In this work, in order to better understand MNNs, we introduce *Manifold Filter-Combine Networks*. The filter-combine paradigm parallels the aggregate-combine framework introduced in Xu et al. (2019) to understand GNNs. It leads one to consider many interesting classes of MNNs which may be thought of as the manifold counterparts of various popular GNNs. We provide sufficient conditions for such networks to converge to a continuum limit as the number of sample points, n, tends to infinity. Notably, our analysis shows that if the weights of the network are properly normalized, then the rate of convergence depends linearly on the depth of the network, in contrast to previous results exhibiting exponential dependence.

2. Background

2.1. Graph Signal Processing and Spectral Graph Neural Networks

Graph signal processing (GSP) Shuman et al. (2013) extends Fourier analysis to graphs. For a function (signal) $\mathbf{x}$ defined on the vertices of a graph $G = (V, E)$, $V = \{v_1, \ldots, v_n\}$, one can define its graph Fourier transform by $\widehat{\mathbf{x}}(i) = \langle \mathbf{u}_i, \mathbf{x} \rangle_2$, where $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is an orthonormal basis of eigenvectors for the graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{A}$, $\mathbf{L}\mathbf{u}_i = \lambda_i \mathbf{u}_i$.[1] Since the $\mathbf{u}_i$ form an orthonormal basis, we obtain the Fourier inversion formula $\mathbf{x} = \sum_{i=1}^{n} \widehat{\mathbf{x}}(i)\mathbf{u}_i$. We can then define convolutional operators in the Fourier domain $w(\mathbf{L})\mathbf{x} = \sum_{i=1}^{n} w(\lambda_i)\widehat{\mathbf{x}}(i)\mathbf{u}_i$, and use these convolutions to define spectral GNNs such as ChebNet (Defferrard et al., 2016).

2.2. Manifold Learning

Manifold learning algorithms (Lin et al., 2015; Moon et al., 2018) aim to detect non-linear structure in the data, analogous to the manner in which PCA is used to detect linear structure. Given a high-dimensional point cloud $\{x_i\}_{i=1}^{n} \subseteq \mathbb{R}^D$, which is assumed to lie upon an unknown d-dimensional manifold $\mathcal{M}$ ($d \ll D$), they aim to produce a low-dimensional representation of the data points x_i that approximates the intrinsic geometry of $\mathcal{M}$.

Many popular manifold learning algorithms such as Diffusion Maps (Coifman and Lafon, 2006) and Laplacian Eigenmaps (Belkin and Niyogi, 2003) operate by constructing a graph $G_n = (V_n, E_n)$ where the vertices are the data points, i.e., $V_n = \{x_i\}_{i=1}^{n}$. For instance, Laplacian Eigenmaps map each $x_j \in \mathbb{R}^D$ to the point $(\phi_2^n(j), \ldots, \phi_{m+1}^n(j)) \in \mathbb{R}^m$, where $\phi_1^n, \ldots, \phi_n^n$ are the eigenvectors of the graph Laplacian $\mathbf{L}_n$.[2] In order to justify the intuition that Laplacian Eigenmaps and related algorithms capture the intrinsic geometry of the data, one may aim to prove that the graph Laplacian $\mathbf{L}_n$ converges to a manifold Laplacian such as the (negative) Laplace-Beltrami Operator, $-\Delta = -\text{div} \circ \nabla$ as the number of points $n \to \infty$. Results along these lines have been established in numerous works such as Dunson et al. (2021); Cheng and Wu (2022); Calder and Trillos (2022); Belkin and Niyogi (2008).

3. Methods

3.1. Signal Processing and Spectral Convolution on Manifolds

Let $\mathcal{M}$ be a compact, connected, d-dimensional Riemannian manifold embedded in $\mathbb{R}^D$, $D \gg d$. Let μ be a probability distribution on $\mathcal{M}$ with a smooth, non-vanishing density ρ, and let $L^2(\mathcal{M})$ denote the set of functions where $\|f\|_{L^2(\mathcal{M})}^2 = \langle f, f \rangle_{L^2(\mathcal{M})} = \int_{\mathcal{M}} |f|^2 d\mu < \infty$.

We let $\mathcal{L} = -\frac{1}{2\rho}\text{div}(\rho^2 \nabla f)$ denote the weighted manifold Laplacian. It is known that $\mathcal{L}$ has an orthonormal basis of eigenfunctions $\mathcal{L}\phi_i = \lambda_i \phi_i$. Thus, we define a generalized Fourier series by $\widehat{f}(i) = \langle f, \phi_i \rangle_{L^2(\mathcal{M})}$, and obtain the Fourier inversion formula $f(x) = \sum_{i=1}^{\infty} \widehat{f}(i)\phi_i(x)$. We may then define spectral convolution operators[3], $w(\mathcal{L})$, for $w \in L^\infty([0, \infty))$ by

$$w(\mathcal{L})f = \sum_{i=1}^{\infty} w(\lambda_i)\widehat{f}(i)\phi_i. \tag{1}$$

---

1. Here and throughout, we identify the function $\mathbf{x}$, with the vector $\mathbf{x} \in \mathbb{R}^n$, $x_i = \mathbf{x}(v_i)$.
2. The first eigenvector is omitted because it is constant.
3. $w(\mathcal{L})$ is independent of the choice of eigenbasis. See Remark 1 of Chew et al. (2024) for details.

Extended Abstract Track

3.2. Manifold Filter-Combine Networks

We now introduce the *filter-combine* paradigm, a novel framework, for thinking about manifold neural networks, paralleling the aggregate-combine framework introduced in Xu et al. (2019) in order to understand GNNs.[4]

We will assume that our input data is a vector-valued function $F = (f_1, \ldots, f_C)$, where each $f_i \in L^2(\mathcal{M})$. Each hidden layer of the network will consist of the following five steps: (i) We *filter* each input channel f_k by a family of spectral operators $w_{j,k}(\mathcal{L})$, $1 \leq j \leq J$. (ii) For each fixed j, we *combine* the filtered feature functions $\tilde{f}_{j,k} = (w_{j,k}(\mathcal{L})f_k)$ into new feature functions $g_{j,k}$ where each $g_{j,k}$ is a linear combination of the $\tilde{f}_{j,k}$. (iii) For each fixed k, we perform a *cross-filter convolution* that maps $\{g_{j,k}\}_{j=1}^J$ to $\{\tilde{g}_{j,k}\}_{j=1}^{J'}$ where each $\tilde{g}_{j,k}$ is a linear combination of the $g_{j,k}$. (iv) We apply some *non-linear, nonexpansive pointwise activation function* σ to each of the $\tilde{g}_{j,k}$, to obtain $h_{j,k} = \sigma \circ \tilde{g}_{j,k}$. (v) Lastly, we *reshape* the $\{h_{i,j}\}_{1 \leq i \leq \tilde{C}, 1 \leq j \leq J'}$ into $\{f'_i\}_{i=1}^{C'}$, where $C' = \tilde{C}J'$. We note that each of these steps can be effectively omitted by choosing parameters in a suitable manner (e.g., choosing certain matrices to be the identity). For further details and discussion, please see Appendix A.

We shall refer to networks constructed using the layers above as Manifold Filter-Combine Networks (MFCNs). As illustrated in the examples below, the MFCN framework naturally allows one to consider many different subfamilies of networks. Indeed, for nearly any (spectral) GNN, there is a corresponding MFCN. Moreover, if desired, one could adapt the MFCN framework to consider the counterparts of non-spectral GNNs by allowing the filters in step (i) to be generic linear operators on $L^2(\mathcal{M})$.

Example 1 (MCNs) *In order to obtain a network analogous to GCN Kipf and Welling (2016), we let $\mathcal{A} = w(\mathcal{L})$ for a decreasing function w, e.g., $w(\lambda) = e^{-\lambda}$, so that $\mathcal{A}$ may be thought of as a low-pass filter. We will omit cross-filter convolutions (step (iii)) and use a learnable weight matrix Θ to combine the filtered features in step (ii). We thus obtain $f_k^{\ell+1} = \sum_{i=1}^{C_\ell} \theta_{i,k} \widetilde{A} f_k$ which may be written compactly as $F^{(\ell+1)} = \sigma\left(\widetilde{A} F^{(\ell)} \Theta^{(\ell)}\right)$.*

Example 2 (Manifold ChebNets) *In order to obtain a network analogous to ChebNet (Defferrard et al., 2016), one can construct a network where the filters take the form $p_{j,k}(\mathcal{P})$, where each $p_{j,k}$ is a polynomial and $\mathcal{P} = e^{-\mathcal{L}}$ is the heat-kernel. (Note that $\mathcal{P}$ has the same eigenfunctions as $\mathcal{L}$ and that its eigenvalues are given by $0 \leq \omega_k = e^{-\lambda_k} \leq 1$. Therefore, polynomials of $\mathcal{P}$ are still spectral filters of the form (1).)*

Example 3 (The Manifold Scattering Transform) *The manifold scattering transform (Perlmutter et al., 2020) is a hand-crafted, wavelet-based method for deep learning on manifolds inspired by analogous constructions for Euclidean data (Mallat, 2012) and graphs (Gama et al., 2019a,b; Zou and Lerman, 2019; Gao et al., 2019). Here, we omit steps (ii) and (iii) and consider a family of wavelets $\{w_j(\lambda)\}_{j=1}^J$, chosen to be a band-pass filters such as or $w_j(\lambda) = e^{-2^{j-1}\lambda} - e^{-2^j\lambda}$. One then defines ℓ-th order scattering coefficients by $U[j_1, \ldots, j_\ell]f_k = \sigma(w_{j_\ell}(\mathcal{L})U[j_1, \ldots, j_{\ell-1}]f_k)$, $U[j_1]f_k = \sigma(w_{j_\ell}(\mathcal{L})f_k)$. We could also consider variations which include steps (ii) and (iii), paralleling analogous graph constructions (Tong et al., 2022; Wenkel et al., 2022).*

---

4. We use the term "filter" rather than "aggregate" because our filters are not required to be localized averaging operations such as those used in common message-passing GNNs.

3.3. Implementation on MFCNs on Point Clouds

We now consider the setting where we do not know the manifold $\mathcal{M}$, but are merely given finitely many sample points $\{x_i\}_{i=1}^n \in \mathbb{R}^D$ (i.e., a point cloud) assumed to lie upon an unknown manifold $\mathcal{M}$ sampled from density ρ. We construct a graph $G_n = (V_n, E_n)$, whose vertices are the sample points, and edges are defined by $\{x_i, x_j\} \in E_n$ if $\|x_i - x_j\|_2 < \epsilon$, where we set $\epsilon \sim (\log(n)/n)^{\frac{1}{d+4}}$ following the lead of Calder and Trillos (2022). We then approximate $\mathcal{L}$ by the graph Laplacian $\mathbf{L}_n = \frac{d+2}{v_d n \epsilon^{d+2}} (\mathbf{D}_n - \mathbf{A}_n)$ where $\mathbf{D}_n$ and $\mathbf{A}_n$ are the degree and adjacency matrices of G_n and v_d is the volume of the unit ball in $\mathbb{R}^d$.

For an input signal f, we let $\mathbf{x} = P_n f$, where $P_n f \in \mathbb{R}^n$ is the vector $P_n f(i) = f(x_i)/\sqrt{n}$. We then approximate $w(\mathcal{L})f$ by $w(\mathbf{L}_n)\mathbf{x} = \sum_{i=1}^\infty w(\lambda_i^n)\widehat{\mathbf{x}}(i)\phi_i^n$, where $\{\phi_i^n\}_{i=1}^n$ is an an orthonormal basis of eigenvectors, $\mathbf{L}_n\phi_i^n = \lambda_i^n\phi_i^n$, and implement the MFCN as in Section 3.2, but with this approximation used in step (i). (See Appendix A for more details.)

Below, in Theorem 2, we prove that $\lim_{n\to\infty} \|\mathbf{x}_k^\ell - P_n f_k^\ell\|_2 = 0$, where $\mathbf{x}_k^\ell$ is the k-th signal in the ℓ-th layer of the discrete implementation of the network. In other words, with sufficiently many data points, the result of our discrete implementation will be nearly the same as if one implemented the entire network on the manifold $\mathcal{M}$, using global knowledge of both $\mathcal{M}$ and the function F and then discretized the corresponding output. We first prove an intermediate result on the convergence of spectral filters.

Theorem 1 *Let $w : [0, \infty) \to \mathbb{R}$, $\|w\|_{L^\infty([0,\infty))} \leq 1$, and assume that w is normalized Lipschitz, i.e., $|w(\lambda_1) - w(\lambda_2)| \leq |\lambda_1 - \lambda_2|$. Assume that f has $\kappa < \infty$ nonzero Fourier coefficients. Then, with probability $1 - \mathcal{O}(n^{-9})$, we have*

$$\|w(\mathbf{L}_n)P_n f - P_n w(\mathcal{L})f\|_2 \leq \mathcal{O}\left(\left(\frac{\log(n)}{n}\right)^{\frac{1}{d+4}}\right)\|f\|_{L^2(\mathcal{M})} + \mathcal{O}\left(\left(\frac{\log(n)}{n}\right)^{\frac{1}{4}}\right)\|f\|_{L^4(\mathcal{M})},$$

where the constants implied by the big-$\mathcal{O}$ notation depend on κ and the geometry of $\mathcal{M}$.

An numerical illustration of Theorem 1 is provided in Figures 2 and 4 in Appendix E.[5] We plot the discretization error $\|w(\mathbf{L}_n)P_n f - P_n w(\mathcal{L})f\|_2$ for the spectral filter applied to a simple function f constructed as the sum of two spherical harmonics (i.e., eigenfunctions).[6] For further details see Appendix E. Next, we use Theorem 1 to prove the following result showing $\lim_{n\to\infty} \|\mathbf{x}_k^\ell - P_n f_k^\ell\|_2 = 0$.

Theorem 2 *Let w and each of the f_k satisfy the assumptions of Theorem 2. Then,*

$$\|\mathbf{x}_k^\ell - P_n f_k^\ell\|_2 \leq \ell \left(\mathcal{O}\left(\left(\frac{\log(n)}{n}\right)^{\frac{1}{d+4}}\right) \max_k \|f_k\|_{L^2(\mathcal{M})} + \mathcal{O}\left(\left(\frac{\log(n)}{n}\right)^{1/4}\right) \max_k \|f_k\|_{L^4(\mathcal{M})}\right)$$

with probability $1 - \mathcal{O}(n^{-9})$, where the constants implied by the big-$\mathcal{O}$ notation depend on the geometry of $\mathcal{M}$ and the weights used in steps (ii) and (iii) of the MFCN.

---

5. Code available at https://anonymous.4open.science/r/mfcn-31CD

6. While our method is designed for generic manifolds, here we consider the sphere since there are known formulas for the eigenfunctions which allows us to compare against ground truth.

Extended Abstract Track

References

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15(6):1373–1396, 2003.

Mikhail Belkin and Partha Niyogi. Towards a theoretical foundation for laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308, 2008.

Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42, 2017.

Michael M Bronstein, Joan Bruna, Taco Cohen, and Petar Veličković. Geometric deep learning: Grids, groups, graphs, geodesics, and gauges. *arXiv preprint arXiv:2104.13478*, 2021.

Jeff Calder and Nicolas Garcia Trillos. Improved spectral convergence rates for graph laplacians on ε-graphs and k-nn graphs. *Applied and Computational Harmonic Analysis*, 60:123–175, 2022.

Xiuyuan Cheng and Nan Wu. Eigen-convergence of gaussian kernelized graph laplacian by manifold heat interpolation. *Applied and Computational Harmonic Analysis*, 61:132–190, 2022.

Joyce Chew, Deanna Needell, and Michael Perlmutter. A convergence rate for manifold neural networks. *arXiv preprint arXiv:2212.12606*, 2022a.

Joyce Chew, Holly Steach, Siddharth Viswanath, Hau-Tieng Wu, Matthew Hirn, Deanna Needell, Matthew D Vesely, Smita Krishnaswamy, and Michael Perlmutter. The manifold scattering transform for high-dimensional point cloud data. In *Topological, Algebraic and Geometric Learning Workshops 2022*, pages 67–78. PMLR, 2022b.

Joyce Chew, Matthew Hirn, Smita Krishnaswamy, Deanna Needell, Michael Perlmutter, Holly Steach, Siddharth Viswanath, and Hau-Tieng Wu. Geometric scattering on measure spaces. *Applied and Computational Harmonic Analysis*, 70:101635, 2024. ISSN 1063-5203. doi: https://doi.org/10.1016/j.acha.2024.101635. URL https://www.sciencedirect.com/science/article/pii/S1063520324000125.

Ronald R. Coifman and Stéphane Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.

Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. In *Advances in Neural Information Processing Systems 29*, pages 3844–3852, 2016.

David B Dunson, Hau-Tieng Wu, and Nan Wu. Spectral convergence of graph laplacian and heat kernel reconstruction in l-infinity from random samples. *Applied and Computational Harmonic Analysis*, 55:282–336, 2021.

Fernando Gama, Joan Bruna, and Alejandro Ribeiro. Stability of graph scattering transforms. In *Advances in Neural Information Processing Systems 33*, 2019a.

Fernando Gama, Alejandro Ribeiro, and Joan Bruna. Diffusion scattering transforms on graphs. In *International Conference on Learning Representations*, 2019b.

Feng Gao, Guy Wolf, and Matthew Hirn. Geometric scattering for graph data analysis. In *Proceedings of the 36th International Conference on Machine Learning, PMLR*, volume 97, pages 2122–2131, 2019.

T. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. *Proc. of ICLR*, 2016.

Ron Levie, Federico Monti, Xavier Bresson, and Michael M Bronstein. Cayleynets: Graph convolutional neural networks with complex rational spectral filters. *IEEE Transactions on Signal Processing*, 67(1):97–109, 2018.

Binbin Lin, Xiaofei He, and Jieping Ye. A geometric viewpoint of manifold learning. *Applied Informatics*, 2(1):3, 2015.

Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, October 2012.

Jonathan Masci, Davide Boscaini, Michael Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 37–45, 2015a.

Jonathan Masci, Davide Boscaini, Michael M. Bronstein, and Pierre Vandergheynst. Geodesic convolutional neural networks on riemannian manifolds. In *IEEE International Conference on Computer Vision Workshop (ICCVW)*, 2015b.

Kevin R. Moon, Jay S. Stanley, Daniel Burkhardt, David van Dijk, Guy Wolf, and Smita Krishnaswamy. Manifold learning-based methods for analyzing single-cell RNA-sequencing data. *Current Opinion in Systems Biology*, 7:36–46, 2 2018. ISSN 24523100. doi: 10.1016/j.coisb.2017.12.008.

Mark W Moyle, Kristopher M Barnes, Manik Kuchroo, Alex Gonopolskiy, Leighton H Duncan, Titas Sengupta, Lin Shao, Min Guo, Anthony Santella, Ryan Christensen, et al. Structural and developmental principles of neuropil assembly in c. elegans. *Nature*, 591 (7848):99–104, 2021.

Michael Perlmutter, Feng Gao, Guy Wolf, and Matthew Hirn. Geometric scattering networks on compact Riemannian manifolds. In *Mathematical and Scientific Machine Learning Conference*, 2020.

Stefan C Schonsheck, Bin Dong, and Rongjie Lai. Parallel transport convolution: Deformable convolutional networks on manifold-structured data. *SIAM Journal on Imaging Sciences*, 15(1):367–386, 2022.

Extended Abstract Track

David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

Alexander Tong, Frederik Wenkel, Dhananjay Bhaskar, Kincaid Macdonald, Jackson Grady, Michael Perlmutter, Smita Krishnaswamy, and Guy Wolf. Learnable filters for geometric scattering modules, 2022.

Laurens van der Maaten and Geoffrey Hinton. Visualizing high-dimensional data using t-SNE. *Journal of Machine Learning Research*, 9:2579–2605, 2008.

David Van Dijk, Roshan Sharma, Juozas Nainys, Kristina Yim, Pooja Kathail, Ambrose J Carr, Cassandra Burdziak, Kevin R Moon, Christine L Chaffer, Diwakar Pattabiraman, et al. Recovering gene interactions from single-cell data using data diffusion. *Cell*, 174 (3):716–729, 2018.

Zhiyang Wang, Luana Ruiz, and Alejandro Ribeiro. Stability of manifold neural networks to deformations. *arXiv preprint arXiv:2106.03725*, 2021a.

Zhiyang Wang, Luana Ruiz, and Alejandro Ribeiro. Stability of neural networks on manifolds to relative perturbations. *arXiv preprint arXiv:2110.04702*, 2021b.

Zhiyang Wang, Luana Ruiz, and Alejandro Ribeiro. Convergence of graph neural networks on relatively sparse graphs. 2023.

Frederik Wenkel, Yimeng Min, Matthew Hirn, Michael Perlmutter, and Guy Wolf. Overcoming oversmoothness in graph convolutional networks via hybrid scattering networks. *arXiv preprint arXiv:2201.08932*, 2022.

Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. How powerful are graph neural networks? In *International Conference on Learning Representations*, 2019.

Jie Zhou, Ganqu Cui, Shengding Hu, Zhengyan Zhang, Cheng Yang, Zhiyuan Liu, Lifeng Wang, Changcheng Li, and Maosong Sun. Graph neural networks: A review of methods and applications. *AI Open*, 1:57–81, 2020. ISSN 2666-6510. doi: https://doi.org/10. 1016/j.aiopen.2021.01.001. URL https://www.sciencedirect.com/science/article/pii/S2666651021000012.

Dongmian Zou and Gilad Lerman. Graph convolutional neural networks via scattering. *Applied and Computational Harmonic Analysis*, 49(3)(3):1046–1074, 2019.

Appendix A. Further Details on MFCNs

Here, we provide explicit layerwise update rules for MFCNs and also add some further discussion. In the ℓ-th layer, given input $F^{(\ell)} = (f_1^{(\ell)}, \ldots, f_{C_\ell}^{(\ell)})$, we define $F^{(\ell+1)} = (f_1^{(\ell+1)}, \ldots, f_{C_{\ell+1}}^{(\ell+1)})$ by:

$$\text{Filtering:} \quad \tilde{f}_{j,k}^{(\ell)} = w_{j,k}^{(\ell)}(\mathcal{L})f_k^{(\ell)}, \quad 1 \leq j \leq J_\ell, 1 \leq k \leq C_\ell$$

$$\text{Combine:} \quad g_{j,k}^{(\ell)} = \sum_{i=1}^{C_\ell} \tilde{f}_{j,i}^{(\ell)} \theta_{i,k}^{(\ell,j)}, \quad 1 \leq j \leq J_\ell, 1 \leq k \leq C_\ell'$$

$$\text{Cross-Filter Convolution:} \quad \tilde{g}_{j,k}^{(\ell)} = \sum_{i=1}^{J_\ell} \alpha_{j,i}^{(\ell,k)} g_{i,k}, 1 \leq j \leq J_\ell', 1 \leq k \leq C_\ell'$$

$$\text{Activation:} \quad h_{j,k}^{(\ell)} = \sigma^{(\ell)} \circ \tilde{g}_{j,k}^{(\ell)}, \quad 1 \leq j \leq J_\ell', 1 \leq k \leq C_\ell'$$

$$\text{Reshaping:} \quad f_{(j-1)C_\ell+k}^{(\ell+1)} = h_{j,k}^{(\ell)}, 1 \leq j \leq J_\ell', 1 \leq k \leq C_\ell',$$

where $C_{\ell+1} = J_\ell' C_\ell'$ (and we set $F_0 = F, C_0 = C$). A graphical depiction of these operations performed is given in Figure 1. We note that the reshaping operator is merely included so that each layer both inputs and outputs a vector of functions, allowing for multiple layers to be stacked upon each other.
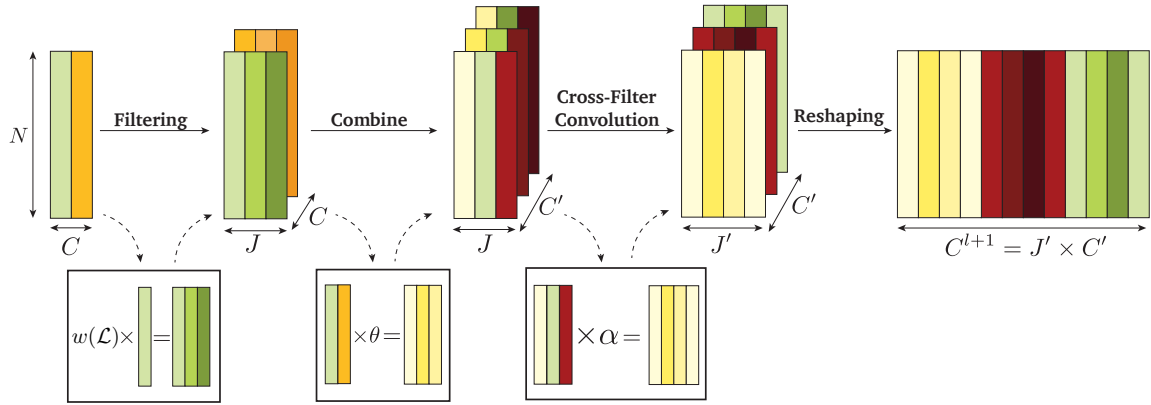

Figure 1: **Manifold Filter-Combine Network** architecture. Starting from a C-dimensional row vector-valued function, each layer in turn filters, combines, convolves channel-wise, applies a point-wise nonlinearity, and reshapes. (For conciseness, we do not visualize the activation step).

We also note one may effectively omit the combine step by setting each matrix $\Theta^{(\ell,j)} := (\theta_{i,k}^{(\ell,j)})_{1 \leq i,k \leq C_\ell}$ equal to the identity. Similarly, one may omit the cross-filter convolutions by setting the matrices $(\alpha_{j,i}^{(\ell,k)})_{1 \leq i,j \leq J_\ell}$ to the identity. Additionally, we observe that because of the generality of our framework, it is possible to write the same network as an MFCN in more than one way. For instance, if one omits the cross channel convolutions, uses a

8

Extended Abstract Track

shared filter bank $\{w^{(\ell)}(\mathcal{L})_j\}_{1 \leq j \leq J}$ (independent of k) and chooses the combine step to be independent of j (i.e., $\theta^{(\ell,j)}_{i,k} = \theta^{(\ell)}_{i,k}$) then we have $f^{(\ell+1)}_{(j-1)C_\ell+k} = \sigma^{(\ell)}\left(\sum_{i=1}^{C_\ell} w^{(\ell)}(\mathcal{L})_j \theta^{(\ell)}_{i,k} f_i\right)$, which may also be obtained by using filters of the form $\widetilde{w}^{(\ell)}_{(j-1)C_\ell+k,i}(\mathcal{L}) = w_j(\mathcal{L})\theta^{(\ell)}_{i,k}$ and using a combine step with $\tilde{\theta}^{(\ell,j)}_{i,k} = 1$.

In our discrete implementation, we assume we have an $n \times C$ data matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_C)$, where $\mathbf{x}_k = P_n f_k$, where as before, $P_n f \in \mathbb{R}^n$ is the vector defined by $P_n f(i) = \frac{1}{\sqrt{n}} f(x_i)$. We may then apply the following discrete update rules paralleling those theoretically conducted in the continuum.

$$\text{Filtering:} \quad \tilde{\mathbf{x}}^{(\ell)}_{j,k} = w^{(\ell)}_{j,k}(\mathbf{L}_n)\mathbf{x}^{(\ell)}_k, \quad 1 \leq j \leq J_\ell, 1 \leq k \leq C_\ell$$

$$\text{Combine:} \quad \mathbf{y}^{(\ell)}_{j,k} = \sum_{i=1}^{C_\ell} \tilde{\mathbf{x}}^{(\ell)}_{j,i}\theta^{(\ell,j)}_{i,k}, \quad 1 \leq j \leq J_\ell, 1 \leq k \leq C'_\ell$$

$$\text{Cross-Filter Convolution:} \quad \tilde{\mathbf{y}}^{(\ell)}_{j,k} = \sum_{i=1}^{J_\ell} \alpha^{(\ell,k)}_{j,i}\mathbf{y}_{i,k}, 1 \leq j \leq J'_\ell, 1 \leq k \leq C'_\ell$$

$$\text{Activation:} \quad \mathbf{z}^{(\ell)}_{j,k} = \sigma \circ \tilde{\mathbf{y}}^{(\ell)}_{j,k}, \quad 1 \leq j \leq J_\ell, 1 \leq k \leq C'_\ell$$

$$\text{Reshaping:} \quad \mathbf{x}^{(\ell+1)}_{(j-1)C_\ell+k} = \mathbf{z}^{(\ell)}_{j,k}, 1 \leq j \leq J'_\ell, 1 \leq k \leq C'_\ell.$$

We note that the filtering step, in principle, requires the eigendecomposition of the $\mathcal{L}_n$ which bears an $\mathcal{O}(n^3)$ computational cost (if all n eigenvalues are utilized). However, this cost can be reduced by approximating $w_{j,k}(\lambda)$ by a polynomial which will allow one to implement the filters without directly computing an eigendecomposition.

Appendix B. Auxilliary Results

In order to prove Theorem 1, we must account for three sources of discretization error: (i) The graph eigenvalue λ^n_i does not exactly equal the manifold eigenvalue λ_i. (ii) The graph eigenvector ϕ^n_i does not exactly equal $P_n\phi_i$, the discretization of the true continuum eigenfunction. (iii) The discrete Fourier coefficients $\widehat{\mathbf{x}}(i)$ are not exactly equal to the continuum Fourier coefficients $\widehat{f}(i)$.

In order to control the first two sources of error, we recall the following result from Calder and Trillos (2022).

Theorem 3 (Theorems 2.4,2.7 of Calder and Trillos (2022)) *Assume that G_n is constructed as in Section 3.3. Fix $\kappa > 0$ and let $\alpha_n = \max_{1 \leq i \leq \kappa} |\lambda_i - \lambda^n_i|$ and $\beta_n = \max_{1 \leq i \leq \kappa} \|\phi^n_i - P_n\phi_i\|_2$. Then, with probability $1 - \mathcal{O}(n^{-9})$,*

$$\alpha_n = \mathcal{O}\left(\frac{\log(n)}{n}\right)^{\frac{1}{d+4}}, \quad and \quad \beta_n = \mathcal{O}\left(\frac{\log(n)}{n}\right)^{\frac{1}{d+4}}, \tag{2}$$

where the implied constants depend on the geometry of the manifold $\mathcal{M}$.

Additionally, we need the following theorem, which is a consequence of Bernstein's inequality, which will help us prove that $\lim_{n \to \infty} \widehat{P_n f}(i) = \widehat{f}(i)$ (since $\widehat{f}(i) = \langle f, \phi_i \rangle_{L^2(\mathcal{M})}$ and $\widehat{\mathbf{x}}(i) = \langle \mathbf{x}, \phi^n_i \rangle_2$).

Proposition 4 *Let f and g be continuous functions defined on $\mathcal{M}$. Then for sufficiently large n, with probability at least $1 - \frac{2}{n^9}$, we have*

$$|\langle P_n f, P_n g\rangle_2 - \langle f, g\rangle_{L^2(\mathcal{M})}| \leq 6\sqrt{\frac{\log(n)}{n}}\|f\|_{L^4(\mathcal{M})}\|g\|_{L^4(\mathcal{M})},$$

where $\|f\|_{L^4(\mathcal{M})} = \left(\int_\mathcal{M} |f|^4 d\mu\right)^{1/4}$.

Proof

Define random variables $\{X_i\}_{i=1}^n$ by $X_i := f(x_i)g(x_i)$ and note that by definition we have

$$\langle P_n f, P_n g\rangle_2 = \frac{1}{n}\sum_{i=1}^n f(x_i)g(x_i) = \frac{1}{n}\sum_{i=1}^n X_i.$$

Since the x_i are sampled i.i.d. with density ρ, we have

$$\mathbb{E}[X_i] = \int_\mathcal{M} f(x)g(x)\rho(x)dx = \langle f, g\rangle_{L^2(\mathcal{M})}.$$

Therefore, letting $\sigma^2 := \mathbb{E}[X_i^2] - \mathbb{E}[X_i]^2$ and $M := \|fg - \mathbb{E}[X_i]\|_{L^\infty(\mathcal{M})}$, we see that by Bernstein's inequality, we have

$$\mathbb{P}(|\langle P_n f, P_n g\rangle_2 - \langle f, g\rangle_{L^2(\mathcal{M})}| > \eta) = \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^n X_i - \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i]\right| > \eta\right)$$

$$= \mathbb{P}\left(\left|\sum_{i=1}^n X_i - \sum_{i=1}^n \mathbb{E}[X_i]\right| > n\eta\right)$$

$$\leq 2\exp\left(-\frac{\frac{1}{2}n\eta^2}{\sigma^2 + \frac{1}{3}M\eta}\right).$$

Setting $\eta = 6\sqrt{\frac{\sigma^2 \log(n)}{n}}$, we see that for n large enough so that $1 + \frac{M\eta}{3\sigma^2} < 2$, we have

$$\mathbb{P}(|\langle P_n f, P_n g\rangle_2 - \langle f, g\rangle_{L^2(\mathcal{M})}| > \eta) \leq 2\exp\left(-\frac{18\sigma^2 \log(n)}{\sigma^2 + \frac{1}{3}M\eta}\right)$$

$$= 2\exp\left(-\frac{18\log(n)}{1 + \frac{M\eta}{3\sigma^2}}\right)$$

$$< 2\exp\left(-9\log(n)\right)$$

$$= \frac{2}{n^9}.$$

We note that $\sigma^2 \leq \mathbb{E}[X_i^2] = \langle f^2, g^2 \rangle_{L^2(\mathcal{M})}$. Therefore, by the Cauchy-Schwarz inequality, with probability at least $1 - \frac{2}{n^9}$, we have

$$
\begin{aligned}
|\langle P_n f, P_n g \rangle_2 - \langle f, g \rangle_{L^2(\mathcal{M})}| &\leq 6\sqrt{\frac{\sigma^2 \log(n)}{n}} \\
&\leq 6\sqrt{\frac{\log(n)}{n}} \sqrt{\langle f^2, g^2 \rangle_{L^2(\mathcal{M})}} \\
&\leq 6\sqrt{\frac{\log(n)}{n}} \|f\|_{L^4(\mathcal{M})} \|g\|_{L^4(\mathcal{M})}.
\end{aligned}
$$

$\blacksquare$

Appendix C. The proof of Theorem 1

Let $\alpha_n = \max_{1 \leq i \leq \kappa} |\lambda_i - \lambda_i^n|$ and $\beta_n = \max_{1 \leq i \leq \kappa} \|\phi_i^n - P_n \phi_i\|_2$. Recall from Theorem 3, that with probability $1 - \mathcal{O}(n^{-9})$ we have

$$
\alpha_n = \mathcal{O}\left(\frac{\log(n)}{n}\right)^{\frac{1}{d+4}}, \quad \text{and} \quad \beta_n = \mathcal{O}\left(\frac{\log(n)}{n}\right)^{\frac{1}{d+4}}. \tag{3}
$$

Let $\gamma_n = 6\sqrt{\frac{\log(n)}{n}}$ be the constant appearing in Proposition 4. We now compute

$$
\begin{aligned}
&\|w(\mathbf{L}_n) P_n f - P_n w(\mathcal{L}) f\|_2 \\
&= \left\| \sum_{i=1}^{\kappa} w(\lambda_i^n) \langle P_n f, \phi_i^n \rangle_2 \phi_i^n - \sum_{i=1}^{\kappa} w(\lambda_i) \langle f, \phi_i \rangle_{\mathcal{M}} P_n \phi_i \right\|_2 \\
&\leq \left\| \sum_{i=1}^{\kappa} (w(\lambda_i^n) - w(\lambda_i)) \langle P_n f, \phi_i^n \rangle_2 \phi_i^n \right\|_2 + \left\| \sum_{i=1}^{\kappa} w(\lambda_i)(\langle P_n f, \phi_i^n \rangle_2 \phi_i^n - \langle f, \phi_i \rangle_{\mathcal{M}} P_n \phi_i) \right\|_2. \tag{4}
\end{aligned}
$$

To bound the first term from (4). We use the assumption that w is normalized Lipschitz, the triangle inequality, and the Cauchy-Schwarz inequality to see that if n is large enough so that $\gamma_n \leq 1$, we have

$$
\begin{aligned}
\left\| \sum_{i=1}^{\kappa} (w(\lambda_i^n) - w(\lambda_i)) \langle P_n f, \phi_i^n \rangle_2 \phi_i^n \right\|_2 &\leq \max_{1 \leq i \leq \kappa} |w(\lambda_i^n) - w(\lambda_i)| \sum_{i=1}^{\kappa} \|P_n f\|_2 \|\phi_i^n\|_2^2 \\
&\leq \alpha_n \sum_{i=1}^{\kappa} \|P_n f\|_2 \|\phi_i^n\|_2^2 \\
&\leq \kappa \alpha_n \|P_n f\|_2 \\
&\leq \kappa \left(\alpha_n \|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})} \right),
\end{aligned}
$$

where we use the fact that $\|\phi_i^n\|_2^2 = 1$ and that

$$
\|P_n f\|_2 \leq \left(\|f\|_{L^2(\mathcal{M})}^2 + \gamma_n^2 \|f\|_{L^4(\mathcal{M})}^2 \right)^{1/2} \leq \|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})}. \tag{5}
$$

11

Now, turning our attention to the second term from (4), we have

$$\left\| \sum_{i=1}^{\kappa} w(\lambda_i)(\langle P_n f, \phi_i^n \rangle_2 \phi_i^n - \langle f, \phi_i \rangle_{L^2(\mathcal{M})} P_n \phi_i) \right\|_2$$

$$\leq \left\| \sum_{i=1}^{\kappa} w(\lambda_i) \langle P_n f, \phi_i^n \rangle_2 (\phi_i^n - P_n \phi_i) \right\|_2 + \left\| \sum_{i=1}^{\kappa} w(\lambda_i)(\langle P_n f, \phi_i^n \rangle_2 - \langle f, \phi_i \rangle_{L^2(\mathcal{M})} P_n \phi_i \right\|_2. \quad (6)$$

By definition, we have $\|\phi_i^n - P_n \phi_i\|_2 \leq \beta_n$. Therefore, since $\|w\|_{L^\infty([0,\infty))]} \leq 1$, we see that if n is large enough so $\beta_n \leq 1$ then

$$\left\| \sum_{i=1}^{\kappa} w(\lambda_i) \langle P_n f, \phi_i^n \rangle_2 (\phi_i^n - P_n \phi_i) \right\|_2 \leq \kappa \max_{1 \leq i \leq \kappa} |\langle P_n f, \phi_i^n \rangle_2| \|\phi_i^n - P_n \phi_i\|_2$$

$$\leq \kappa \beta_n \|P_n f\|_2$$

$$\leq \kappa \left(\beta_n \|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})} \right), \quad (7)$$

where the final inequality follows from (5). Meanwhile, the second term from (6) can be bounded by

$$\left\| \sum_{i=1}^{\kappa} w(\lambda_i)(\langle P_n f, \phi_i^n \rangle_2 - \langle f, \phi_i \rangle_{\mathcal{M}}) P_n \phi_i \right\|_2$$

$$\leq \sum_{i=1}^{\kappa} |w(\lambda_i)| |\langle P_n f, \phi_i^n \rangle_2 - \langle f, \phi_i \rangle_{\mathcal{M}}| \|P_n \phi_i\|_2$$

$$\leq \sum_{i=1}^{\kappa} |\langle P_n f, \phi_i^n \rangle_2 - \langle f, \phi_i \rangle_{\mathcal{M}}| \|P_n \phi_i\|_2$$

$$\leq \sum_{i=1}^{\kappa} |\langle P_n f, \phi_i^n \rangle_2 - \langle P_n f, P_n \phi_i \rangle_2| \|P_n \phi_i\|_2 + \sum_{i=1}^{\kappa} |\langle P_n f, P_n \phi_i \rangle_2 - \langle f, \phi_i \rangle_{\mathcal{M}}| \|P_n \phi_i\|_2.$$

By the Cauchy-Schwarz inequality, Proposition 4, (3), (5), and the assumption that n is large enough so that $\beta_n \leq 1$, we have

$$|\langle P_n f, \phi_i^n \rangle_2 - \langle P_n f, P_n \phi_i \rangle_2| \leq \|P_n f\|_2 \|\phi_i^n - P_n \phi_i\|_2$$

$$\leq \beta_n \left(\|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})} \right)$$

$$\leq \left(\beta_n \|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})} \right).$$

And also by Proposition 4 and (5) we have

$$|\langle P_n f, P_n \phi_i \rangle_2 - \langle f, \phi_i \rangle_2| \leq \gamma_n^2 \|f\|_{L^4(\mathcal{M})} \|\phi_i\|_{L^4(\mathcal{M})}, \quad \text{and} \quad \|P_n \phi_i\|_2 \leq 1 + \gamma_n \|\phi_i\|_{L^4(\mathcal{M})}.$$

Since κ is fixed and $\lim_{n \to \infty} \gamma_n = 0$, for sufficiently large n we have $\gamma_n \|\phi_i\|_{L^4(\mathcal{M})} \leq 1$ for all $i \leq \kappa$. This implies

$$|\langle P_n f, P_n \phi_i \rangle_2 - \langle f, \phi_i \rangle_2| \leq \gamma_n \|f\|_{L^4(\mathcal{M})}, \quad \text{and} \quad \|P_n \phi_i\|_2 \leq 1 + \gamma_n \|\phi_i\|_{L^4(\mathcal{M})} \leq 2.$$

Extended Abstract Track

Therefore, the second term from (6) can be bounded by

$$
\left\| \sum_{i=1}^{\kappa} w(\lambda_i)(\langle P_n f, \phi_i^n \rangle_2 - \langle f, \phi_i \rangle_{\mathcal{M}}) P_n \phi_i \right\|_2
$$
$$
\leq \sum_{i=1}^{\kappa} |\langle P_n f, \phi_i^n \rangle_2 - \langle P_n f, P_n \phi_i \rangle_2| \|P_n \phi_i\|_2 + \sum_{i=1}^{\kappa} |\langle P_n f, P_n \phi_i \rangle_2 - \langle f, \phi_i \rangle_2| \|P_n \phi_i\|_2
$$
$$
\leq \sum_{i=1}^{\kappa} \left(\beta_n \|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})} \right) \|P_n \phi_i\|_2 + \sum_{i=1}^{\kappa} \gamma_n \|f\|_{L^4(\mathcal{M})} \|P_n \phi_i\|_2
$$
$$
\leq 4\kappa \left(\beta_n \|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})} \right). \tag{8}
$$

Therefore, combining Equations (4) through (8) yields

$$
\|w(\mathbf{L}_n) P_n f - P_n w(\mathcal{L}) f\|_2
$$
$$
\leq \left\| \sum_{i=1}^{\kappa} (w(\lambda_i^n) - w(\lambda_i)) \langle P_n f, \phi_i^n \rangle_2 \phi_i^n \right\|_2 + \left\| \sum_{i=1}^{\kappa} w(\lambda_i)(\langle P_n f, \phi_i^n \rangle_2 \phi_i^n - \langle f, \phi_i \rangle_{\mathcal{M}} P_n \phi_i) \right\|_2
$$
$$
\leq \kappa(\alpha_n \|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})}) + 5\kappa \left(\beta_n \|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})} \right)
$$
$$
\leq 6\kappa \left((\alpha_n + \beta_n) \|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})} \right).
$$

Plugging in the values of $\alpha_n, \beta_n,$ and γ_n completes the proof.

Appendix D. The Proof of Theorem 2

Lemma 5 *Under the assumptions of Theorem 2, we have*

$$
\|w(\mathbf{L}_n)\mathbf{x} - P_n w(\mathcal{L}) f\|_2 \leq \|\mathbf{x} - P_n f\|_2 + 6\kappa \left((\alpha_n + \beta_n) \|f\|_{L^2(\mathcal{M})} + \gamma_n \|f\|_{L^4(\mathcal{M})} \right),
$$

where α_n, β_n and γ_n are as in the proof of Theorem 1 for all continuous functions f and all $\mathbf{x} \in \mathbb{R}^n$. (We do not assume $\mathbf{x} = P_n f$ here.)

Proof We first observe that since $\|w\|_{L^\infty([0,\infty))} \leq 1$, we have

$$
\|w(\mathbf{L}_n)\mathbf{x} - w(\mathbf{L}_n) P_n f\|_2 = \|w(\mathbf{L}_n)(\mathbf{x} - P_n f)\|_2
$$
$$
= \| \sum_{i=1}^{n} w(\lambda_i^n) \langle \mathbf{x} - P_n f, \phi_i^n \rangle_2 \phi_i^n \|_2 \tag{9}
$$
$$
= \left(\sum_{i=1}^{n} |w(\lambda_i^n)|^2 |\langle \mathbf{x} - P_n f, \phi_i^n \rangle_2|^2 \right)^{1/2}
$$
$$
\leq \left(\sum_{i=1}^{n} |\langle \mathbf{x} - P_n f, \phi_i^n \rangle_2|^2 \right)^{1/2}
$$
$$
= \|\mathbf{x} - P_n f\|_2. \tag{10}
$$

Therefore, by the triangle inequality, together with Theorem 1, we we have

$$\|w(\mathbf{L}_n)\mathbf{x} - P_n w(\mathcal{L})f\|_2 \leq \|w(\mathbf{L}_n)\mathbf{x} - w(\mathbf{L}_n)P_n f\|_2 + \|w(\mathbf{L}_n)P_n f - P_n w(\mathcal{L})f\|_2$$
$$\leq \|\mathbf{x} - P_n f\|_2 + 6\kappa \left((\alpha_n + \beta_n)\|f\|_{L^2(\mathcal{M})} + \gamma_n\|f\|_{L^4(\mathcal{M})} \right)$$

as desired. ∎

Proof [The proof of Theorem 2]

Let

$$\epsilon_n = \mathcal{O}\left(\left(\frac{\log(n)}{n} \right)^{\frac{1}{d+4}} \right) \max_k \|f_k\|_{L^2(\mathcal{M})} + \mathcal{O}\left(\left(\frac{\log(n)}{n} \right)^{1/4} \right) \max_k \|f_k\|_{L^4(\mathcal{M})},$$

so that by Lemma 5, together with the definitions of $\alpha_n, \beta_n,$ and γ_n we have

$$\|w(\mathbf{L}_n)\mathbf{x} - P_n w(\mathcal{L})f_k\|_2 \leq \epsilon_n.$$

Let $A_1^{(\ell)} = \max_{j,k}(|\sum_{i=1}^{C_\ell} \theta_{i,k}^{(\ell,j)}|)$, $A_2^{(\ell)} = \max_{j,k}(\sum_{i=1}^{J_\ell} |\alpha_{j,i}^{(\ell,k)}|)$, where $\theta_{i,k}^{(\ell,j)}$ and $\alpha_{j,i}^{(\ell,k)}$ are the weights used in the combination step (step (ii)) and the cross-filter step (step (iii)) of the MFCN. (See Appendix A.) The following lemma bounds the error induced in the non-filtering steps of the discrete MFCN implementation.

Lemma 6 *The errors induced by the non-filtering steps of our network may be bounded by*

$$\|\mathbf{y}_{j,k}^{(\ell)} - P_n g_{j,k}^{(\ell)}\|_2 \leq \max_{1 \leq i \leq C_\ell} \|\tilde{\mathbf{x}}_{j,k}^{(\ell)} - P_n \tilde{f}_{j,k}^{(\ell)}\|_2 \sum_{i=1}^{C_\ell} |\theta_{i,k}^{(\ell,j)}|, \tag{11}$$

$$\|\tilde{\mathbf{y}}_{j,k}^{(\ell)} - P_n \tilde{g}_{j,k}^{(\ell)}\|_2 \leq \max_{1 \leq i \leq J_\ell} \|\mathbf{y}_{j,k}^{(\ell)} - P_n g_{j,k}^{(\ell)}\|_2 \sum_{i=1}^{J_\ell} |\alpha_{j,i}^{(\ell,k)}|. \tag{12}$$

$$\|\mathbf{z}_{j,k}^{(\ell)} - P_n h_{j,k}^{(\ell)}\|_2 \leq \|\tilde{\mathbf{y}}_{j,k}^{(\ell)} - P_n \tilde{g}_{j,k}^{(\ell)}\|_2. \tag{13}$$

(where the notation is as in Appendix A).

Proof To verify (11), we observe that

$$\|\mathbf{y}_{j,k}^{(\ell)} - P_n g_{j,k}^{(\ell)}\|_2 = \left\| \sum_{i=1}^{C_\ell} \tilde{\mathbf{x}}_{j,k}^{(\ell)} \theta_{i,k}^{(\ell,j)} - P_n \tilde{f}_{j,k}^{(\ell)} \theta_{i,k}^{(\ell,j)} \right\|_2$$
$$\leq \sum_{i=1}^{C_\ell} |\theta_{i,k}^{(\ell,j)}| \|\tilde{\mathbf{x}}_{j,k}^{(\ell)} - P_n \tilde{f}_{j,k}^{(\ell)}\|_2$$
$$\leq \max_{1 \leq i \leq C_\ell} \|\tilde{\mathbf{x}}_{j,k}^{(\ell)} - P_n \tilde{f}_{j,k}^{(\ell)}\|_2 \sum_{i=1}^{C_\ell} |\theta_{i,k}^{(\ell,j)}|.$$

Extended Abstract Track

The proof of (12) is identical to the proof of (11). For (13), we see that since σ is non-expansive we have

$$
\begin{aligned}
\|\mathbf{z}_{j,k}^{(\ell)} - P_n h_{j,k}^{(\ell)}\|_2^2 &= \sum_{i=1}^{n} |(\mathbf{z}_{j,k}^{(\ell)})(i) - (P_n h_{j,k}^{(\ell)})(i)|^2 \\
&= \sum_{i=1}^{n} |(\mathbf{z}_{j,k}^{(\ell)})(i) - h_{j,k}^{(\ell)}(x_i)|^2 \\
&= \sum_{i=1}^{n} |\sigma((\tilde{\mathbf{y}}_{j,k}^{(\ell)})(i)) - \sigma(\tilde{g}_{j,k}^{(\ell)}(x_i))|^2 \\
&\leq \sum_{i=1}^{n} |(\tilde{\mathbf{y}}_{j,k}^{(\ell)})(i) - \tilde{g}_{j,k}^{(\ell)}(x_i)|^2 \\
&= \|\tilde{\mathbf{y}}_{j,k}^{(\ell)} - P_n \tilde{g}_{j,k}^{(\ell)}\|_2^2.
\end{aligned}
$$

∎

Now, returning to the proof of Theorem 2, it follows from the definition of the reshaping operator

$$
\max_{k} \|\mathbf{x}_k^{(\ell+1)} - P_n f_k^{(\ell+1)}\|_2 = \max_{j,k} \|\mathbf{z}_{p,r}^{(\ell)} - P_n h_{p,r}^{(\ell)}\|_2.
$$

(Again, see Appendix A for notation.) Therefore, by Lemma 6 we have

$$
\begin{aligned}
\max_{k} \|\mathbf{x}_k^{(\ell+1)} - P_n f_k^{(\ell+1)}\|_2 &= \max_{j,k} \|\mathbf{z}_{p,r}^{(\ell)} - P_n h_{p,r}^{(\ell)}\|_2. \\
&\leq \|P_n \tilde{g}_{j,k}^{(\ell)} - \tilde{\mathbf{y}}_{j,k}^{(\ell)}\|_2. \\
&\leq A_2^{(\ell)} \max_{j,k} \|P_n g_{j,k}^{(\ell)} - \mathbf{y}_{j,k}^{(\ell)}\|_2 \\
&\leq A_2^{(\ell)} A_1^{(\ell)} \max_{j,k} \|P_n \tilde{f}_{j,k}^{(\ell)} - \tilde{\mathbf{x}}_{j,k}^{(\ell)}\|_2 \\
&\leq A_2^{(\ell)} A_1^{(\ell)} (\max_{k} \|\mathbf{x}_k^{(\ell)} - P_n f_k^{(\ell)}\|_2 + \epsilon_n).
\end{aligned}
$$

Since $\|\mathbf{x}_0^{(\ell)} - P_n f_k^{(0)}\|_2 = 0$ for all k, we may use induction to conclude that

$$
\|\mathbf{x}_k^{(\ell)} - P_n f_k^{(\ell)}\|_2 \leq \sum_{i=0}^{\ell-1} \prod_{j=i}^{\ell-1} A_1^{(j)} A_2^{(j)} \epsilon_n.
$$

In particular, if we assume that $A_1^{(j)} A_2^{(j)} \leq A$, we have

$$
\|\mathbf{x}_k^{(\ell)} - P_n f_k^{(\ell)}\|_2 \leq \ell A^\ell \epsilon_n.
$$

and if $A = 1$, we have

$$
\|\mathbf{x}_k^{(\ell)} - P_n f_k^{(\ell)}\|_2 \leq \ell \epsilon_n.
$$

∎

15

Remark 7 *Results similar to Theorems 1 and 2 can also be derived in the setting where the filters w are bandlimited, i.e., $w(\lambda_i) = 0 \, \forall i > \kappa$. They can also be obtained when G_n is constructed as a k-NN graph. Additionally, similar results can be derived if the filters have Lipschitz constants greater than one (where the bound will depend on the largest Lipschitz constant).*

Appendix E. Experiments and Extension to k-NN Graphs

In Figure 2, we numerically demonstrate the convergence of a spectral filter, complementing our theoretical analysis in Theorem 1. We focus on the two-dimensional unit sphere, embedded in $\mathbb{R}^3$, with uniform sampling, since in this setting there are known formulas for the eigenvalues and eigenfunctions (i.e., spherical harmonics) which allows us to compare our numerical implementation to ground truth. We consider a simple function f chosen to be the sum of two eigenfunctions i.e., $f = Y_1^0 + Y_2^0$, where Y_j^i is the i-th harmonic of degree j. We sample n points uniformly from the sphere evaluated f at these points, and built the graph G_n as described in Section 3.3. We then applied the spectral filter $w(\lambda) = e^{-\lambda}$.

When computing the eigendecomposition, we used only the first 64 eigenpairs as $w(\lambda)$ takes negligible values at the higher eigenvalues. We repeated this process 10 times for each value of n and calculated the discretization errors of the spectral filter as $\|w(\mathbf{L}_n)P_n f - P_n w(\mathcal{L})f\|_2$ for increasingly large values of n. Additionally, we also tracked the convergence of the first two distinct, non-zero eigenvalues (corresponding to the first eight eigenvalues counting their multiplicity), as shown in Figure 3. We see that the numerical eigenvalues converge to the true values of $\frac{\ell(\ell+1)}{8\pi}$, for $\ell = 1, 2$. (Please note that with uniform sampling on the two-dimensional unit sphere, $\mathcal{L}f = -\frac{1}{2\rho}\text{div}(\rho^2 \nabla f)$ reduces to $-\frac{1}{8\pi}\Delta$, where $-\Delta = -\text{div} \circ \nabla$ is the negative Laplace-Beltrami operator.)

E.1. Extension to k-NN graphs

In the main text, for the sake of concreteness, we have focused on ϵ-graph constructions. However, our results may be straightforwardly extended to (symmetric) k-NN graphs, $G_n = (V_n, E_n)$, in which $\{x_i, x_j\} \in E_n$ if x_i is one of the k-th closest points to x_j (with respect to Euclidean distance in $\mathbb{R}^D$) or if x_j is one of the k closest points to x_i.

In this setting, one defines the graph Laplacian by $\mathbf{L}_n = \frac{d+2}{v_d n}\left(\frac{nv_d}{k}\right)^{1+2/d}(\mathbf{D}_n - \mathbf{A}_n)$, where v_d is the volume of the d-dimensional Euclidean unit ball (and $\mathbf{D}_n$ and $\mathbf{A}_n$ are the degree and adjacency matrices associated to G_n), and the limiting manifold Laplacian is given by $\mathcal{L}f = -\frac{1}{2\rho}\text{div}(\rho^{1-2/d}\nabla f)$, which reduces to $-2\pi\Delta$ when sampling uniformly on the sphere. If one sets $k \sim \log(n)^{\frac{d}{d+4}}n^{\frac{4}{d+4}}$, then one may readily derive results analogous to Theorems 1 and 2. The only difference in the proof is that we must invoke Theorems 2.5 and 2.9 of Calder and Trillos (2022), rather than Theorems 2.4 and 2.7 (which we restate as Theorem 3). We illustrate this numerically in Figures 4 and 5, which are the analogs of Figures 2 and 3, but with the graph constructed as a k-NN graph rather than an ϵ graph.
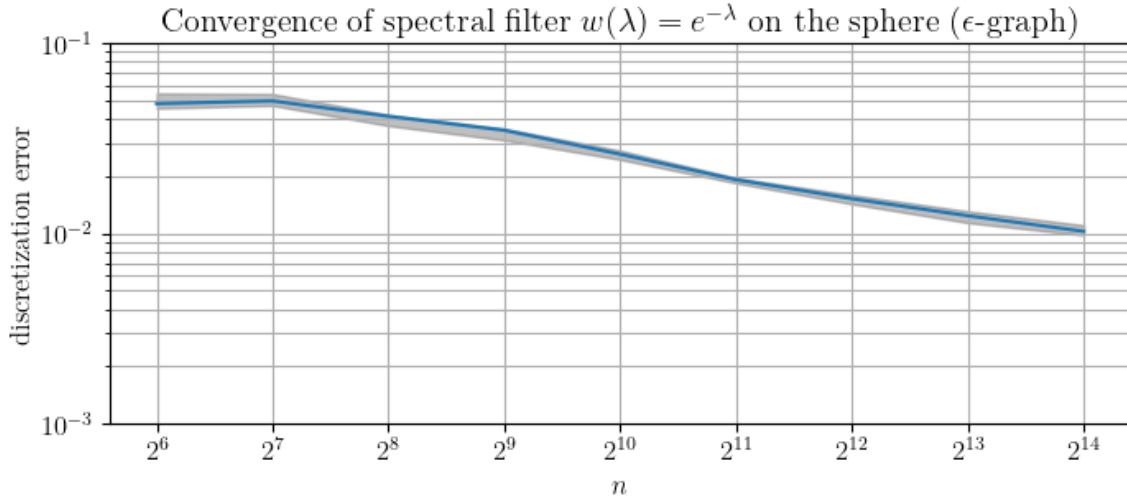
Extended Abstract Track

Figure 2: Discretization error for spectral filter $w(\lambda) = e^{-\lambda}$ applied to the sum of two spherical harmonics, for an ϵ-graph construction. The median error of 10 runs is shown in blue, against a gray band of the 25th- to 75th-percentile error range.
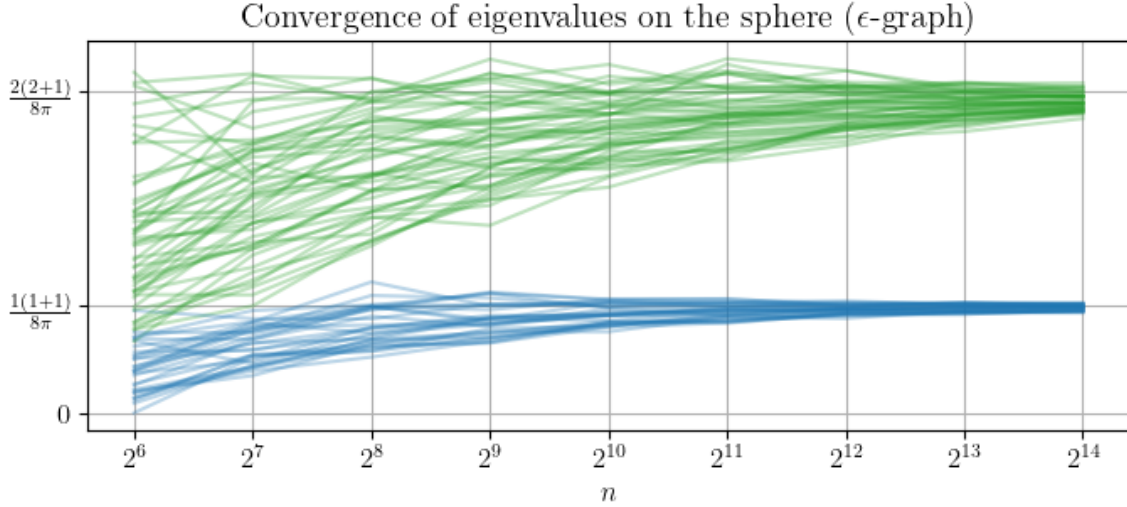

Figure 3: Convergence of first eight, non-zero eigenvalues on the sphere, for an ϵ-graph construction, all 10 runs combined. The blue lines are the first three eigenvalues that converge to the same limit (since the first non-zero eigenvalue of the spherical Laplacian has multiplicity three). Similarly, the green lines are the next five eigenvalues.
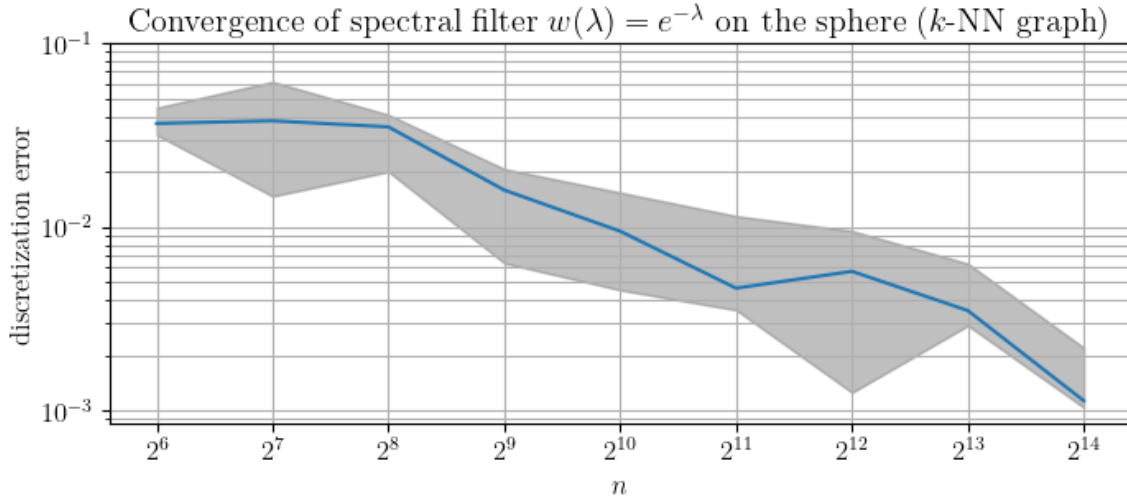
Figure 4: Discretization error for spectral filter $w(\lambda) = e^{-\lambda}$ applied to the sum of two spherical harmonics, for a k-NN graph construction. The median error of 10 runs is shown in blue, against a gray band of the 25th- to 75th-percentile error range.
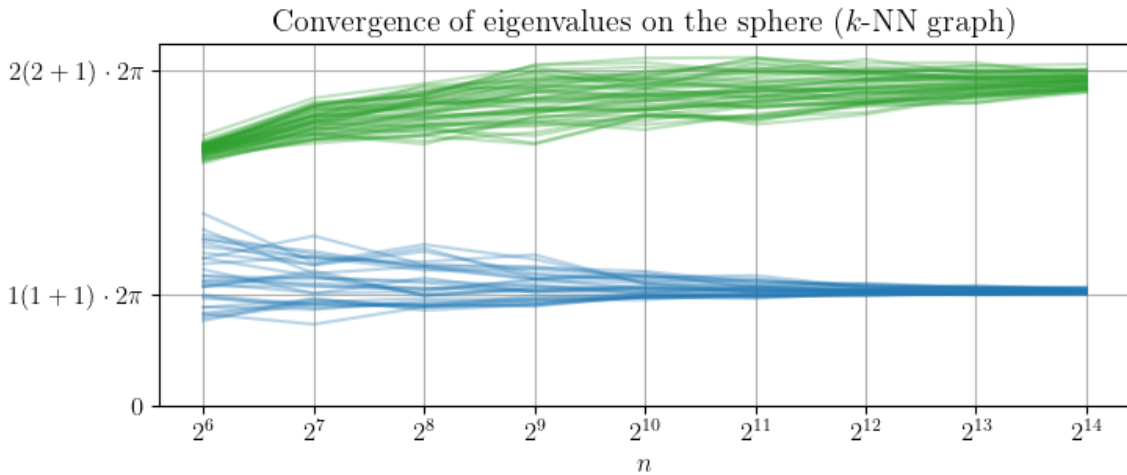

Figure 5: Convergence of first eight, non-zero eigenvalues on the sphere, for a k-NN graph construction, all 10 runs combined. The blue lines are the first three eigenvalues that converge to the same limit (since the first non-zero eigenvalue of the spherical Laplacian has multiplicity three). Similarly, the green lines are the next five eigenvalues.