Geometric Scattering on Manifolds

Michael Perlmutter[*]
Dept. of CMSE[†]
Michigan State University
East Lansing, MI, USA
perlmut6@msu.edu

Guy Wolf[‡]
Dept. of Math. and Stat.
Université de Montréal
Montreal, QC, Canada
guy.wolf@umontreal.edu

Matthew Hirn[§]
Dept. of CMSE[*]
Dept. of Mathematics
Michigan State University
East Lansing, MI, USA
mhirn@msu.edu

Abstract

The Euclidean scattering transform was introduced nearly a decade ago to improve the mathematical understanding of the success of convolutional neural networks (ConvNets) in image data analysis and other tasks. Inspired by recent interest in geometric deep learning, which aims to generalize ConvNets to manifold and graph-structured domains, we generalize the scattering transform to compact manifolds. Similar to the Euclidean scattering transform, our geometric scattering transform is based on a cascade of designed filters and pointwise nonlinearities, which enables rigorous analysis of the feature extraction provided by scattering layers. Our main focus here is on theoretical understanding of this geometric scattering network, while setting aside implementation aspects, although we remark that application of similar transforms to graph data analysis has been studied recently in related work. Our results establish conditions under which geometric scattering provides localized isometry invariant descriptions of manifold signals, and conditions in which the filters used in our network are stable to families of diffeomorphisms formulated in intrinsic manifolds terms. These results not only generalize the deformation stability and local roto-translation invariance of Euclidean scattering, but also demonstrate the importance of linking the used filter structures (e.g., in geometric deep learning) to the underlying manifold geometry, or the data geometry it represents.

1 Introduction

Characterizing variability in data is one of the most fundamental aspects of modern data analysis, which appears at the core of both supervised and unsupervised learning tasks. In particular, a crucial part of typical machine learning methods is to separate informative sources of variability from disruptive sources, which are considered via deformation groups or noise models. For example, linear classification methods aim to find a separating hyperplane that captures undesired (i.e., irrelevant for discriminating between classes) variance, and then eliminate it by projecting the data on normal directions of the hyperplane. Principal component analysis (PCA), on the other hand, aims to find a hyperplane that maximizes the captured variance, while treating directions of minimal variance as noise directions. Nonlinear methods extend such notions to consider richer models (beyond linear hyperplanes and projections) for capturing or eliminating variability - either task dependent or following an assumed model for the data.

In the past decade, deep learning methods have shown impressive potential in capturing complex nonlinear variability by using a cascade of linear transformations and simple

---

[*]Corresponding author

[†]Computational Mathematics, Science & Engineering

[‡]http://guywolf.org/

[§]https://matthewhirn.com/

nonlinear activations that together form approximations of target functions. While there is a multitude of applications where such methods are effective, some of their most popular achievements are in fields traditionally associated with signal processing in general, and image processing (i.e., computer vision) in particular. In such settings, the data has an inherent spatial (or temporal) structure, and collected observations form signals over it. Convolutional neural networks (ConvNets) utilize this structure to treat their linear transformations as convolutions of input signals with filters that are learned in the training process. In classification applications, the resulting deep cascade of convolutional filters (and nonlinearities) eliminates intra-class heterogeneity and results in high accuracy classifiers. Furthermore, hidden ConvNet layers yield coefficients that isolate semantic features and can be utilized for unsupervised feature extraction, e.g., as part of dimensionality reduction or generative models.

In an effort to improve mathematical understanding of deep convolutional networks and their learned features, Mallat [2010, 2012] presented the *scattering transform*. This transform has an architecture similar to ConvNets, based on a cascade of convolutional filters and simple pointwise nonlinearities in the form of complex modulus or absolute value. However, unlike deep learning methods, this transform does not learn its filters from data, but rather has them designed to provide guaranteed stability to a given family of deformations. As shown in Mallat [2012], under some admissibility conditions, one can use appropriate wavelet filter banks in the scattering transform to provide invariance to the actions of Lie groups. Moreover, the resulting scattering features also provide Lipschitz stability to small diffeomorphisms, where the size of a diffeomorphism is quantified by its deviation from a translation. These notions were made concrete in Bruna and Mallat [2011, 2013], Sifre and Mallat [2012, 2013, 2014] and Oyallon and Mallat [2015] using groups of translations, rotations, and scaling operations, with applications in image and texture classification. Further applications of the scattering transform and its deep filter bank approach were shown effective in several fields, such as audio processing [Andén and Mallat, 2011, 2014, Wolf et al., 2014, 2015, Andén et al., 2018], medical signal processing [Chudáček et al., 2014], and quantum chemistry [Hirn et al., 2017, Eickenberg et al., 2017, 2018, Brumwell et al., 2018].

Another structure often found in modern data is that of a graph between collected observations or features. Such structure naturally arises, for example, in transportation and social networks, where weighted edges represent connections between locations (e.g., roads or other routes) or user profiles (e.g., friendships or exchanged communications). They are also common when representing molecular structures in chemical data, or biological interactions in biomedical data. Relatedly, signals supported on manifolds or manifold valued data is also becoming increasing prevalent, in particular in shape matching and computer graphics. Manifold models for high dimensional data also arise in the field of manifold learning [e.g., Tenenbaum et al., 2000, Coifman and Lafon, 2006a, van der Maaten and Hinton, 2008], in which unsupervised algorithms infer data-driven geometries and use them to capture intrinsic structure and patterns in data. As such, a large body of work has emerged to explore the generalization of spectral and signal processing notions to manifolds [e.g., Coifman and Lafon, 2006b] and graphs [Shuman et al., 2013, and references therein]. In these settings, functions are supported on the manifold or the vertices of the graph, and the eigenfunctions of the Laplace-Beltrami operator or the eigenvectors of the graph Laplacian serve as the Fourier harmonics.

This increasing interest in non-Euclidean data, particularly graphs and manifolds, has led to a new research direction known as *geometric deep learning*, which aims to generalize convolutional networks to graph and manifold structured data [Bronstein et al., 2017, and references therein]. Unlike classical ConvNets, in which filters are learned on collected data features (i.e., in spatial or temporal domain), many geometric deep learning approaches for manifolds learn spectral coefficients of their filters [Bruna et al., 2014, Defferrard et al., 2016, Levie et al., 2017, Yi et al., 2017]. The frequency spectrum is defined by the eigenvalues of the Laplace-Beltrami operator on the manifold, which links geometric deep learning with manifold and graph signal processing.

Inspired by geometric deep learning, recent works have also proposed an extension of the scattering transform to graph domains. These mostly focused on finding features that

represent a graph structure (given a fixed set of signals on it) while being stable to graph perturbations. In Gama et al. [2018], a cascade of diffusion wavelets from Coifman and Maggioni [2006] was proposed, and its Lipschitz stability was shown with respect to a global diffusion-inspired distance between graphs. A similar construction discussed in Zou and Lerman [2018] was shown to be stable to permutations of vertex indices, and to small perturbations of edge weights. Finally, Gao et al. [2018] established the viability of scattering coefficients as universal graph features for data analysis tasks (e.g., in social networks and biochemistry data).

In this paper, we take the less charted path and consider the manifold aspect of geometric deep learning. In this setting, one needs to process signals over a manifold, and in particular, to represent them with features that are stable to orientations, noise, or deformations over the manifold geometry. In order to work towards these aims, we define a scattering transform on compact smooth Riemannian manifolds without boundary, which we call *geometric scattering*. Our construction is based on convolutional filters defined spectrally via the eigendecomposition of the Laplace-Beltrami operator over the manifold, as discussed in Section 3. We show that these convolutional operators can be used to construct a frame, which, with appropriately chosen low-pass and high-pass filters, forms a wavelet frame similar to the diffusion wavelets constructed in Coifman and Maggioni [2006]. Then, in Section 4, a cascade of these generalized convolutions and pointwise absolute value operations is used to map signals on the manifold to scattering coefficients that encode approximate local invariance to isometries, which correspond to translations, rotations, and reflections in Euclidean space. In Section 5, we analyze the commutators of the filters used in our network with the action of diffeomorphisms. We believe that the results presented there will allow us to study the stability of our scattering network to these diffeomorphisms in future work using a notion of stability that analogous to the Lipschitz stability considered in Mallat [2012] on Euclidean space. As we discuss in Section 5, this requires a quantitative notion of deformation size. We consider three formulations of such a notion, which measure how far a given diffeomorphism is from being an isometry, and explore the type of stability that can be achieved for the geometric scattering transform with respect to each of them. Our results provide a path forward for utilizing the scattering mathematical framework to analyze and understand geometric deep learning, while also shedding light on the challenges involved in such generalization to non-Euclidean domains.

1.1 Notation

Let $\mathcal{M}$ be a smooth, compact, and connected, d-dimensional Riemannian manifold without boundary contained in $\mathbb{R}^n$. Let $r : \mathcal{M} \times \mathcal{M} \to \mathbb{R}$ denote the geodesic distance between two points, and let Δ be the Laplace-Beltrami operator on $\mathcal{M}$. The eigenfunctions and non-unique eigenvalues of $-\Delta$ are denoted by φ_k and λ_k, respectively. Since $\mathcal{M}$ is compact, the spectrum of $-\Delta$ is countable and we may assume that $\{\varphi_k\}_{k \in \mathbb{N}}$ forms an orthonormal basis for $\mathbf{L}^2(\mathcal{M})$. We define the Fourier transform of $f \in \mathbf{L}^2(\mathcal{M})$ as the sequence $\widehat{f} \in \ell^2$ defined by $\widehat{f}(k) = \langle f, \varphi_k \rangle$. The set of unique eigenvalues of $-\Delta$ is denoted by Λ, and for $\lambda \in \Lambda$ we let $m(\lambda)$ and E_λ denote the corresponding multiplicities and eigenspaces. The diffeomorphism group of $\mathcal{M}$ is Diff($\mathcal{M}$), and the isometry group is Isom($\mathcal{M}$) $\subset$ Diff($\mathcal{M}$). For a diffeomorphism $\zeta \in$ Diff($\mathcal{M}$), we let V_ζ be the operator $V_\zeta f(x) = f(\zeta^{-1}(x))$, and let $\|\zeta\|_\infty = \sup_{x \in \mathcal{M}} r(x, \zeta(x))$. We let $C(\mathcal{M})$ denote a constant which depends only on the manifold $\mathcal{M}$, and for two operators A and B defined on $\mathbf{L}^2(\mathcal{M})$, we let $[A, B] := AB - BA$ denote their commutator.

2 Problem Setup

We consider signals $f \in \mathbf{L}^2(\mathcal{M})$ and representations $\Theta : \mathbf{L}^2(\mathcal{M}) \to \ell^2(\mathbf{L}^2(\mathcal{M}))$, which encode information about the signal, often referred to as the "features" of f. A common goal in machine learning tasks is to classify or cluster signals based upon these features. However, we often want to consider two signals, or even two manifolds, to be equivalent if they differ by the action of a global isometry. Therefore, we seek to construct a family of representations, $(\Theta_t)_{t \in (0, \infty)}$, which are invariant to isometric transformations of any

$f \in \mathbf{L}^2(\mathcal{M})$ up to the scale t. Such a representation should satisfy a condition similar to,

$$\|\Theta_t(f) - \Theta_t(V_\zeta f)\|_{2,2} \leq \alpha(\zeta)\beta(t)\|f\|_2, \quad \forall\, f \in \mathbf{L}^2(\mathcal{M}), \zeta \in \mathrm{Isom}(\mathcal{M}), \tag{1}$$

where $\alpha(\zeta)$ measures the size of the isometry with $\alpha(\mathrm{id}) = 0$, and $\beta(t)$ decreases to zero as the scale t grows to infinity.

Along similar lines, it is desirable that the action of small diffeomorphisms on f, or on the underlying manifold $\mathcal{M}$, should not have a large impact on the representation of the inputted signal for tasks such as classification or regression. However, the set of $\mathrm{Diff}(\mathcal{M})$ is a large group and invariance over $\mathrm{Diff}(\mathcal{M})$ would collapse the variability even between vastly different signals. Thus, in this case, we want a family $(\Theta_t)_{t \in (0,\infty)}$ that is stable to diffeomorphism actions on f, but not invariant. This leads to a condition such as:

$$\|\Theta_t(f) - \Theta_t(V_\zeta f)\|_{2,2} \leq A(\zeta)\|f\|_2, \quad \forall\, t \in (0,\infty),\ f \in \mathbf{L}^2(\mathcal{M}), \zeta \in \mathrm{Diff}(\mathcal{M}), \tag{2}$$

where $A(\zeta)$ measures how much ζ differs from being an isometry, with $A(\zeta) = 0$ if $\zeta \in \mathrm{Isom}(\mathcal{M})$ and $A(\zeta) > 0$ if $\zeta \notin \mathrm{Isom}(\mathcal{M})$.

Combining (1) and (2), we have the following goal:

$$\|\Theta_t(f) - \Theta_t(V_\zeta f)\|_{2,2} \leq [\alpha(\zeta)\beta(t) + A(\zeta)]\|f\|_2, \tag{3}$$
$$\forall\, t \in (0,\infty),\ f \in \mathbf{L}^2(\mathcal{M}), \zeta \in \mathrm{Diff}(\mathcal{M}).$$

At the same time, the representations $(\Theta_t)_{t \in (0,\infty)}$ should not be trivial. Indeed, distinguishing different classes or types of signals often requires leveraging subtle information in the signals. This information is often stored in the high frequencies, i.e., $\widehat{f}(k)$ for large k. Our problem is thus to find a family of representations that are stable to diffeomorphisms, discriminative between different types of signals, which allow one to control the scale of isometric invariance, and do so for data that is supported on a manifold. The wavelet scattering transform of Mallat [2012] achieves goals analogous to the ones presented here, but for Euclidean supported signals. Therefore, we seek to construct a geometric version of the scattering transform, using filters corresponding to the spectral geometry of $\mathcal{M}$, and to show that it has similar properties to its Euclidean counterpart.

3 Spectral Integral Operators

Similar to traditional ConvNets, the Euclidean scattering transform constructed in Mallat [2012] consists of an alternating cascade of convolutions and nonlinearities. In the manifold setting, it is not immediately clear how to define convolution operators, because translation is not well defined. In this section, we introduce a family of operators on $\mathbf{L}^2(\mathcal{M})$ that are analogous to convolution operators on Euclidean space due to the characterization of Euclidean convolution operators as Fourier multipliers. These operators will then be used in Sec. 4 to construct the geometric scattering transform.

For a function $\eta : \Lambda \to \mathbb{R}$, we define a *spectral kernel* K_η by

$$K_\eta(x,y) = \sum_{k \in \mathbb{N}} \eta(\lambda_k)\varphi_k(x)\overline{\varphi}_k(y) \tag{4}$$

and refer to the integral operator T_η, with kernel K_η, as a *spectral integral operator*. Grouping together the φ_k belonging to each eigenspace E_λ, we write

$$K_\eta(x,y) = \sum_{\lambda \in \Lambda} \eta(\lambda)K^{(\lambda)}(x,y), \tag{5}$$

where $K^{(\lambda)}(x,y) = \sum_{\lambda_k = \lambda} \varphi_k(x)\overline{\varphi}_k(y)$. Using the fact that $\{\varphi_k\}_{k \in \mathbb{N}}$ is an orthonormal basis for $\mathbf{L}^2(\mathcal{M})$, a simple calculation shows that

$$\|T_\eta f\|_2^2 = \sum_{k \in \mathbb{N}} |\eta(\lambda_k)|^2 |\langle f, \varphi_k \rangle|^2 \leq \|\eta\|_\infty^2 \|f\|^2, \tag{6}$$

4

and so, if $\eta \in \mathbf{L}^\infty$, then T_η is a bounded operator of $\mathbf{L}^2(\mathcal{M})$ with operator norm $\|\eta\|_\infty$. In particular, if $\|\eta\|_\infty \leq 1$, then T_η is nonexpansive. Operators of this form are analogous to convolution operators defined on $\mathbb{R}^d$ since the latter are diagonalized in the Fourier basis. To further emphasize this connection, we note the following theorem, which shows that spectral integral operators are equivariant with respect to isometries.

Theorem 3.1. *For every spectral integral operator T_η, and for every $f \in \mathbf{L}^2(\mathcal{M})$,*

$$T_\eta V_\zeta f = V_\zeta T_\eta f, \quad \forall \zeta \in \mathrm{Isom}(\mathcal{M}).$$

Proof. Let ζ be an isometry and let $\psi_k(x) = \varphi_k(\zeta(x))$. For $\lambda \in \Lambda$ and $x \in \mathcal{M}$, we define $\vec{\varphi}_\lambda(x) \in \mathbb{C}^{m(\lambda)}$ and $\vec{\psi}_\lambda(x) \in \mathbb{C}^{m(\lambda)}$ by

$$\vec{\varphi}_\lambda(x) = (\varphi_k(x))_{k:\lambda_k=\lambda} \quad \text{and} \quad \vec{\psi}_\lambda(x) = (\psi_k(x))_{k:\lambda_k=\lambda}.$$

Since ζ is an isometry, $\{\varphi_k\}_{k:\lambda_k=\lambda}$ and $\{\psi_k\}_{k:\lambda_k=\lambda}$ are both orthonormal bases for E_λ. Therefore, there exists an $m(\lambda) \times m(\lambda)$ unitary matrix A_λ (that does not depend upon x) such that $\vec{\psi}_\lambda(x) = A_\lambda \vec{\varphi}_\lambda(x)$. Using this fact, we see

$$K^{(\lambda)}(\zeta(x), \zeta(y)) = \sum_{k:\lambda_k=\lambda} \varphi_k(\zeta(x))\overline{\varphi}_k(\zeta(y)) = \sum_{k:\lambda_k=\lambda} \psi_k(x)\overline{\psi}_k(y) = \langle \vec{\psi}_\lambda(x), \vec{\psi}_\lambda(y) \rangle$$

$$= \langle A_\lambda \vec{\varphi}_\lambda(x), A_\lambda \vec{\varphi}_\lambda(y) \rangle = \langle \vec{\varphi}_\lambda(x), \vec{\varphi}_\lambda(y) \rangle = K^{(\lambda)}(x, y).$$

Therefore, by (5), we see that $K_\eta(\zeta(x), \zeta(y)) = K_\eta(x, y)$ for all $(x, y) \in \mathcal{M} \times \mathcal{M}$. Now letting $x = \zeta(\tilde{x})$ and changing variables $y = \zeta(z)$, we have

$$T_\eta V_\zeta f(x) = \int_\mathcal{M} K_\eta(x, y) f\left(\zeta^{-1}(y)\right) dV(y) = \int_\mathcal{M} K_\eta(x, \zeta(z)) f(z) dV(z)$$

$$= \int_\mathcal{M} K_\eta(\zeta(\tilde{x}), \zeta(z)) f(z) dV(z) = \int_\mathcal{M} K_\eta(\tilde{x}, z) f(z) dV(z)$$

$$= T_\eta f(\tilde{x}) = T_\eta f\left(\zeta^{-1}(x)\right) = V_\zeta T_\eta f(x)$$

as desired. $\qquad \square$

We will consider frame analysis operators that are constructed using a countable family of spectral integral operators. We call a spectral function $g : [0, \infty) \to \mathbb{R}$ a *low-pass filter* if

$$|g(\lambda)| \leq g(0) = 1, \forall \lambda \geq 0, \quad \text{and} \quad \lim_{\lambda \to \infty} g(\lambda) = 0. \tag{7}$$

Similarly, we call $h : [0, \infty) \to \mathbb{R}$ a *high-pass filter* if

$$h(0) = 0 \quad \text{and} \quad \|h\|_\infty \leq 1.$$

Figure 1(a) illustrates a low-pass operator T_g, via its impulse responses, on the popular Stanford bunny [Turk and Levoy, 2005] manifold.

We will assume that we have a low-pass filter g, and a family of high-pass filters $\{h_\gamma\}_{\gamma \in \Gamma}$, which satisfy a Littlewood-Paley type condition

$$A \leq m(\lambda) \left[|g(\lambda)|^2 + \sum_{\gamma \in \Gamma} |h_\gamma(\lambda)|^2 \right] \leq B, \quad \forall \lambda \in \Lambda \tag{8}$$

for some $0 < A \leq B$. A frame analysis operator $\Phi : \mathbf{L}^2(\mathcal{M}) \to \ell^2(\mathbf{L}^2(\mathcal{M}))$ is then defined by

$$\Phi f = \left\{ T_g f, T_{h_\gamma} f : \gamma \in \Gamma \right\},$$

where T_g and T_{h_γ} are the spectral integral operators corresponding to g and h_γ, respectively. The Littlewood-Paley type condition (8) implies that the filters used to define Φ evenly cover the frequencies of $\mathcal{M}$. The following proposition shows that if g and $\{h_\gamma\}_{\gamma \in \Gamma}$ satisfy (8), then Φ has corresponding upper and lower frame bounds.

5
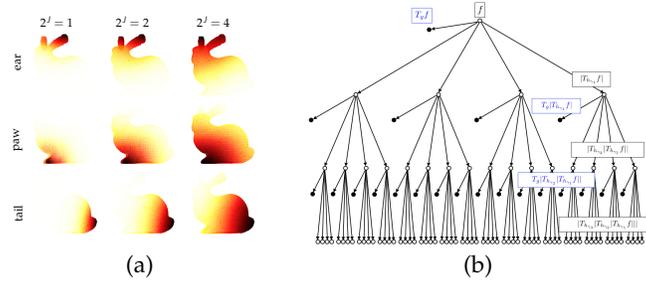
Figure 1: *(a) Illustration of $T_g\delta_x$ over the Stanford bunny manifold [Turk and Levoy, 2005], with g a Gaussian restricted to $[0,\infty)$ at scale 2^j for three $\log_2$-scales J, and three Diracs δ_x centered at different regions of the manifold. (b) The geometric scattering transform. Black: Equivariant intermediate layers. Blue: Invariant output coefficients at each layer.*

Proposition 3.2. *Under the Littlewood-Paley condition (8), Φ is a bounded operator from $\mathbf{L}^2(\mathcal{M})$ to $\ell^2(\mathbf{L}^2(\mathcal{M}))$ and*

$$A\|f\|_2^2 \leq \|\Phi f\|_{2,2}^2 := \|T_g f\|_2^2 + \sum_{\gamma \in \Gamma} \|T_{h_\gamma} f\|_2^2 \leq B\|f\|_2^2, \ \ \forall f \in \mathbf{L}^2(\mathcal{M}).$$

In particular, if $A = B = 1$, then Φ is an isometry.

Proposition 3.2 is proved by using (6) to write

$$\|\Phi f\|_{2,2}^2 = \|T_g f\|_2^2 + \sum_{\gamma \in \Gamma} \|T_{h_\gamma} f\|_2^2 = \sum_{k \in \mathbb{N}} |g(\lambda_k)|^2 |\langle f, \varphi_k \rangle|^2 + \sum_{\gamma \in \Gamma} \sum_{k \in \mathbb{N}} |h_\gamma(\lambda_k)|^2 |\langle f, \varphi_k \rangle|^2.$$

We then group together the terms corresponding to each eigenspace, E_λ, apply (8), and then use Parseval's Identity. Note that for each $\lambda \in \Lambda$, $|g(\lambda)|^2$ and $|h(\lambda)|^2$ each appear $m(\lambda)$ times in the above sum. It is for this reason that the term $m(\lambda)$ is needed in (8). For full details of the proof, see Appendix A.

3.1 Geometric Wavelet Transforms

As an important example of the frame analysis operator Φ, we define a geometric wavelet transform in terms of a single low-pass filter. Given a spectral function, $\eta : [0,\infty) \to \mathbb{R}$, we define the dilation at scale 2^j by

$$\eta_j(\lambda) = \eta(2^j\lambda),$$

and a normalized spectral function $\widetilde{\eta}_j : \Lambda \to \mathbb{R}$ by

$$\widetilde{\eta}_j(\lambda) = \frac{\eta_j(\lambda)}{\sqrt{m(\lambda)}}. \tag{9}$$

The normalization factor of $\frac{1}{\sqrt{m(\lambda)}}$ will ensure that the wavelet frame constructed below satisfies the Littlewood-Paley condition (8).

Let $g : [0,\infty) \to \mathbb{R}$, be a nonincreasing low-pass filter, and let $J \in \mathbb{Z}$. In a manner analogous to classical wavelet constructions, [see for example, Meyer, 1993], we define a high-pass filter $h : [0,\infty) \to \mathbb{R}$ by

$$h(\lambda) = \left[|g(\lambda/2)|^2 - |g(\lambda)|^2\right]^{\frac{1}{2}} \tag{10}$$

and observe that

$$h_j(\lambda) = \left[|g_{j-1}(\lambda)|^2 - |g_j(\lambda)|^2\right]^{\frac{1}{2}}. \tag{11}$$

6

To emphasize the connection with classical wavelet transforms, we denote $T_{\tilde{g}_J} = A_J$ and $T_{\tilde{h}_j} = W_j$. The geometric wavelet transform $\mathcal{W}_J : \mathbf{L}^2(\mathcal{M}) \to \ell^2(\mathbf{L}^2(\mathcal{M}))$ is then defined as

$$\mathcal{W}_J f = \{A_J f, \, W_j f\}_{j \leq J}.$$

The following proposition can be proved by observing that

$$m(\lambda) \left[|\tilde{g}_J(\lambda)|^2 + \sum_{j \leq J} \left| \tilde{h}_j(\lambda) \right|^2 \right] = \left| g\left(2^J \lambda\right) \right|^2 + \sum_{j \leq J} \left[\left| g\left(2^{j-1}\lambda\right) \right|^2 - \left| g\left(2^j \lambda\right) \right|^2 \right]$$

forms a telescoping sum, and that therefore, $\{A_J, W_j\}_{j \leq J}$ satisfies the Littlewood-Paley condition (8) with $A = B = 1$. We give a complete proof in Appendix B.

Proposition 3.3. *For all $J \in \mathbb{Z}$, the geometric wavelet transform $\mathcal{W}_J$ is an isometry from $\mathbf{L}^2(\mathcal{M})$ to $\ell^2(\mathbf{L}^2(\mathcal{M}))$, i.e.,*

$$\|\mathcal{W}_J f\|_{2,2} = \|f\|_2, \quad \forall\, f \in \mathbf{L}^2(\mathcal{M}).$$

An important example is $g(\lambda) = e^{-\lambda}$. In this case the low-pass kernel $G_J(x, y)$ corresponding to T_{g_J} is the heat kernel on $\mathcal{M}$ at time $t = 2^J$, the kernel $\tilde{G}_J(x, y)$ corresponding to $A_J = T_{\tilde{g}_J}$ is a normalized version, and the corresponding wavelet operators W_j are similar to the diffusion wavelets introduced in Coifman and Maggioni [2006]. The key difference between the construction presented here and the original diffusion wavelet construction is the normalization of the spectral function according to the multiplicity of each eigenvalue, which is needed to obtain an isometric wavelet transform.

4   The Geometric Scattering Transform

The *geometric scattering transform* is a nonlinear operator $S : \mathbf{L}^2(\mathcal{M}) \to \ell^2(\mathbf{L}^2(\mathcal{M}))$ constructed through an alternating cascade of spectral integral operators and nonlinearities. Let $M : \mathbf{L}^2(\mathcal{M}) \to \mathbf{L}^2(\mathcal{M})$ be the modulus operator, $Mf(x) = |f(x)|$, and for each $\gamma \in \Gamma$, we let

$$U_\gamma f(x) = M T_{h_\gamma} f(x) = |T_{h_\gamma} f(x)|.$$

For $m \geq 1$, let Γ_m be the set of all paths of the form $\vec{\gamma} = (\gamma_1, \ldots, \gamma_m)$, and let $\Gamma_\infty = \cup_{m=1}^\infty \Gamma_m$ denote the set of a all finite paths. For $\vec{\gamma} \in \Gamma_m, m \geq 1$ let

$$U_{\vec{\gamma}} f(x) = U_{\gamma_m} \cdots U_{\gamma_1} f(x), \quad \vec{\gamma} = (\gamma_1, \ldots, \gamma_m).$$

We define an operator $U : \mathbf{L}^2(\mathcal{M}) \to \ell^2(\mathbf{L}^2(\mathcal{M}))$, called the scattering propagator, by

$$Uf = \{U_{\vec{\gamma}} f : \vec{\gamma} \in \Gamma_\infty\}.$$

In a slight abuse of notation, we include the "empty-path" of length zero in Γ_∞, and define U_γ to be the identity if γ is the empty path.

The scattering transform $S_{\vec{\gamma}}$ over a path $\vec{\gamma} \in \Gamma_\infty$ is defined as the integration of U_γ against the low-pass integral operator T_g, i.e., $S_{\vec{\gamma}} f(x) = T_g U_{\vec{\gamma}} f(x)$. The operator $S : \mathbf{L}^2(\mathcal{M}) \to \ell^2(\mathbf{L}^2(\mathcal{M}))$ given by

$$Sf = \{S_{\vec{\gamma}} f : \vec{\gamma} \in \Gamma_\infty\},$$

is referred to as the geometric scattering transform. It is a mathematical model for geometric ConvNets that are constructed out of spectral filters. The following proposition shows that S is nonexpansive if the Littlewood-Paley condition holds with $B = 1$, which we assume for the rest of the paper. The proof is nearly identical to Proposition 2.5 of Mallat [2012], and is thus omitted.

Proposition 4.1. *If the Littlewood-Paley condition (8) holds with $B = 1$, then*

$$\|Sf_1 - Sf_2\|_{2,2} \leq \|f_1 - f_2\|, \quad \forall\, f_1, f_2 \in \mathbf{L}^2(\mathcal{M}).$$

The scattering transform is invariant to the action of the isometry group on the inputted signal f up to a factor that depends upon the decay of the low-pass spectral function g. If $|g(\lambda)| \leq C_0 e^{-t\lambda}$ for some constant C_0 and $t > 0$ (e.g., the heat kernel), then the following theorem establishes isometric invariance up to the scale t^d.

Theorem 4.2. *Let $\zeta \in \mathrm{Isom}(\mathcal{M})$ and $|g(\lambda)| \leq C_0 e^{-t\lambda}$ for some constant C_0 and $t > 0$. Then there exists a constant $C(\mathcal{M}) < \infty$, such that*

$$\|Sf - SV_\zeta f\|_{2,2} \leq C_0 C(\mathcal{M}) t^{-d} \|\zeta\|_\infty \|Uf\|_{2,2}, \quad \forall\, f \in L^2(\mathcal{M}).$$

Proof. We rewrite $Sf - SV_\zeta f = [V_\zeta, S] f + Sf - V_\zeta Sf$. Theorem 3.1 proves that spectral integral operators commute with isometries. Since the modulus operator does as well, it follows that $[V_\zeta, S] f = 0$ and

$$\|Sf - SV_\zeta f\|_{2,2} = \|Sf - V_\zeta Sf\|_{2,2}.$$

Since $S = T_g U$, we see that

$$\|Sf - V_\zeta Sf\|_{2,2} = \|T_g Uf - V_\zeta T_g Uf\|_{2,2} \leq \|T_g - V_\zeta T_g\| \|Uf\|_{2,2}, \tag{12}$$

Since $|g(\lambda)| \leq C_0 e^{-t\lambda}$, Lemma 4.3 stated below shows that

$$\|T_g - V_\zeta T_g\| \leq C_0 C(\mathcal{M}) t^{-d} \|\zeta\|_\infty,$$

which completes the proof. $\qquad\square$

Lemma 4.3. *There exists a constant $C(\mathcal{M}) > 0$, such that for every spectral integral operator, T_η, and for every $\zeta \in \mathrm{Diff}(\mathcal{M})$,*

$$\|T_\eta f - V_\zeta T_\eta f\| \leq C(\mathcal{M}) \left(\sum_{k \in \mathbb{N}} \eta(\lambda_k) \lambda_k^{d/2} \right) \|\zeta\|_\infty \|f\|_2, \quad \forall\, f \in \mathbf{L}^2(\mathcal{M}).$$

Moreover, if $|\eta(\lambda)| \leq C_0 e^{-t\lambda}$ for some constant C_0 and some $t > 0$, then there exists a constant $C'(\mathcal{M}) > 0$ such that for any $\zeta \in \mathrm{Diff}(\mathcal{M})$,

$$\|T_\eta f - V_\zeta T_\eta f\| \leq C_0 C'(\mathcal{M}) t^{-d} \|\zeta\|_\infty \|f\|_2, \quad \forall\, f \in \mathbf{L}^2(\mathcal{M}).$$

We give the proof of Lemma 4.3 in Appendix C.

5 Stability to Diffeomorphisms

In the previous section, we showed that the scattering transform is invariant to the action of isometries up to a factor depending on the scale t. In Mallat [2012], it is shown that the Euclidean scattering transform is stable to the action of certain diffeomorphisms that are close to being translations. In an effort to prove analogous results for the geometric scattering transform, we introduce three quantities A_1, A_2, and A_3, which measure how far away a diffeomorphism is from being an isometry. If ζ is an isometry, then by definition $r(\zeta(x), \zeta(y)) = r(x, y)$ for all x and y in $\mathcal{M}$. This motivates us to define

$$A_1(\zeta) := \sup_{\substack{x,y \in \mathcal{M} \\ x \neq y}} \left| \frac{r(\zeta(x), \zeta(y)) - r(x,y)}{r(x,y)} \right|. \tag{13}$$

It is also known [see for example, Kobayashi and Nomizu, 1963] that isometries are volume preserving in the sense that $|\det[D\zeta(x)]| = 1$ for all $x \in \mathcal{M}$. Therefore, we introduce the quantity

$$A_2(\zeta) := \left(\sup_{x \in \mathcal{M}} \left| |\det[D\zeta(x)]| - 1 \right| \right) \left(\sup_{x \in \mathcal{M}} \left| \det[D\zeta^{-1}(x)] \right| \right), \tag{14}$$

which measures how much a ζ distorts volumes. Lastly,

$$A_3(\zeta) := \inf\{ \sup_x \{ r(\zeta(x), \zeta_1(x)) \} : \zeta_1 \in \mathrm{Isom}(\mathcal{M}) \}$$

measures the distance from ζ to the isomtery group in the sup norm. We note that

$$A_3(\zeta) = \inf\{\sup_x\{r(\zeta(x),\zeta_1(x))\} : \zeta_1 \in \text{Isom}(\mathcal{M})\}$$

$$= \inf\{\sup_x\{r\left((\zeta_1^{-1} \circ \zeta)(x), x\right)\} : \zeta_1 \in \text{Isom}(\mathcal{M})\}$$

$$= \inf\left\{\|\zeta_1^{-1} \circ \zeta\|_\infty : \zeta_1 \in \text{Isom}(\mathcal{M})\right\}$$

$$= \inf\{\|\zeta_2\|_\infty : \zeta = \zeta_1 \circ \zeta_2, \zeta_1 \in \text{Isom}(\mathcal{M})\}.$$

Thus, $A_3(\zeta)$ will be small if ζ can be factored into a global isometry ζ_1 and a small perturbation ζ_2.

A key step to showing that the Euclidean scattering transform S_{Euc} is stable to the action of certain diffeomorphisms is a bound on the commutator norm $\|[S_{\text{Euc}}, V_\zeta] f\|_{2,2}$. This bound is then combined with a bound on $\|S_{\text{Euc}} f - V_\zeta S_{\text{Euc}} f\|_{2,2}$, which is obtained by methods similar to those used to prove Theorem 4.2, and the triangle inequality to produce a bound on $\|S_{\text{Euc}} f - V_\zeta S_{\text{Euc}} f\|_{2,2}$. This motivates us to study the commutator of spectral integral operators with V_ζ for diffeomorphisms that are close to being isometries. To this end, in the two theorems below, we show that $\|[T_\eta, V_\zeta]\|$ can be controlled in terms of the quantities A_1, A_2, and A_3, as well η-dependent quantities.

Theorem 5.1. *There exists a constant $C(\mathcal{M}) > 0$ such that if T_η is a spectral integral operator and $\zeta \in \text{Diff}(\mathcal{M})$, then for all $f \in \mathbf{L}^2(\mathcal{M})$,*

$$\|[T_\eta, V_\zeta] f\|_2 \le C(\mathcal{M}) \left[\left(\sum_{k\in\mathbb{N}} \eta(\lambda_k)^2\right)^{\frac{1}{2}} A_2(\zeta) + \left(\sum_{k\in\mathbb{N}} \eta(\lambda_k)\lambda_k^{d/2}\right) A_3(\zeta)\right] \|f\|_2. \quad (15)$$

On certain manifolds, we can prove a modified version of Theorem 5.1 that replaces the term $A_3(\zeta)$ with $A_1(\zeta)$.

Definition 1. A manifold $\mathcal{M}$ is said to be *homogeneous* (with respect to the action of the isometry group), if for any two points, $x, y \in \mathcal{M}$, there exists an isometry $\tilde{\zeta} : \mathcal{M} \to \mathcal{M}$ such that $\tilde{\zeta}(x) = y$.

Definition 2. A manifold $\mathcal{M}$ is said to be *two-point homogeneous*, if for any two pairs of points, (x_1, x_2), (y_1, y_2) such that $r(x_1, x_2) = r(y_1, y_2)$, there exists an isometry $\tilde{\zeta} : \mathcal{M} \to \mathcal{M}$ such that $\tilde{\zeta}(x_1) = y_1$ and $\tilde{\zeta}(x_2) = y_2$.

Theorem 5.2. *If $\mathcal{M}$ is two-point homogeneous, then there exists a constant $C(\mathcal{M}) > 0$ such that if T_η is spectral integral operator and $\zeta \in \text{Diff}(\mathcal{M})$, then for all $f \in \mathbf{L}^2(\mathcal{M})$,*

$$\|[T_\eta, V_\zeta] f\|_2 \le C(\mathcal{M}) \left[A_1(\zeta) \left(\sum_{k\in\mathbb{N}} \eta(\lambda_k)\lambda_k^{(d+1)/4}\right) + A_2(\zeta) \left(\sum_{k\in\mathbb{N}} \eta(\lambda_k)^2\right)^{\frac{1}{2}}\right] \|f\|_2.$$

The proofs of Theorems 5.1 and 5.2 are in Appendices D and E. The keys to these proofs are Lemmas 5.3, 5.4, and 5.5 below, as well as the fact that spectral kernels are radial on two-point homogeneous manifolds. For proofs of these lemmas, please see Appendices F, G, and H. We note that the assumption that K is radial, made in Lemma 5.4, is equivalent to the assumption (16), made in Lemma 5.3, on two-point homogeneous manifolds, but is a stronger assumption in general.

Lemma 5.3. *There exists a constant $C(\mathcal{M}) > 0$ such that if $\zeta \in \text{Diff}(\mathcal{M})$ and T is a kernel integral operator on $\mathbf{L}^2(\mathcal{M})$ with a kernel that is invariant to the action of isometries, i.e.,*

$$K\left(\tilde{\zeta}(x), \tilde{\zeta}(y)\right) = K(x, y), \quad \forall \tilde{\zeta} \in \text{Isom}(\mathcal{M}), \quad (16)$$

then

$$\|[T, V_\zeta] f\|_2 \le C(\mathcal{M}) \left[\|K\|_{\mathbf{L}^2(\mathcal{M}\times\mathcal{M})} A_2(\zeta) + \|\nabla K\|_\infty A_3(\zeta)\right] \|f\|_2, \quad \forall f \in \mathbf{L}^2(\mathcal{M}).$$

9

Lemma 5.4. *There exists a constant $C(\mathcal{M}) > 0$ such that if T is a kernel integral operator with a radial kernel $K(x, y) = \kappa(r(x, y)), \kappa \in \mathbf{C}^1(\mathbb{R})$, and $\zeta \in \mathrm{Diff}(\mathcal{M})$, then*

$$\|[T, V_\zeta] f\|_2 \leq C(\mathcal{M}) \left[\|\nabla K\|_\infty A_1(\zeta) + \|K\|_{\mathbf{L}^2(\mathcal{M} \times \mathcal{M})} A_2(\zeta) \right] \|f\|_2, \quad \forall f \in \mathbf{L}^2(\mathcal{M}).$$

Lemma 5.5. *There exists a constant $C(\mathcal{M}) > 0$ such that $\left\| \nabla K^{(\lambda)} \right\|_\infty \leq C(\mathcal{M}) m(\lambda) \lambda^{d/2}$ for all $\lambda \in \Lambda$. As a consequence, if K_η is a spectral kernel, then*

$$\|\nabla K_\eta\|_\infty \leq C(\mathcal{M}) \sum_{\lambda \in \Lambda} \eta(\lambda) m(\lambda) \lambda^{d/2} = C(\mathcal{M}) \sum_{k \in \mathbb{N}} \eta(\lambda_k) \lambda_k^{d/2}.$$

Furthermore, if $\mathcal{M}$ is homogeneous, then $\left\| \nabla K^{(\lambda)} \right\|_\infty \leq C(\mathcal{M}) m(\lambda) \lambda^{(d+1)/4}$ and $\|\nabla K_\eta\|_\infty \leq C(\mathcal{M}) \sum_{k \in \mathbb{N}} \eta(\lambda_k) \lambda_k^{(d+1)/4}$.

6 Conclusion

The generalization of convolutional neural networks to non-Euclidean domains has been the focus of a significant body of work recently, generally referred to as geometric deep learning. In the process, various challenges have arisen, both computational and mathematical, with the fundamental question being whether the success of ConvNets on traditional signals can be replicated on more complicated geometric settings such as graphs and manifolds. We have presented geometric scattering as a mathematical framework for analyzing the theoretical potential of deep convolutional networks in such settings, in a manner analogous to the approach used by Mallat [2010, 2012] in the Euclidean setting. Our construction is also closely related to recent attempts at defining finite-graph scattering [Gama et al., 2018, Zou and Lerman, 2018, Gao et al., 2018]. Our construction is defined in a continuous setting and therefore, we view it is a more direct generalization of the original Euclidean scattering transform. Datasets defined on certain graphs can then be viewed as finite subsamples of the underlying manifold, and the graph scattering transform presented in Gao et al. [2018] can be viewed as a discretized version of the construction presented here. Finally, our stability results provided here not only establish the potential of deep cascades of *designed* filters to enable rigorous analysis in geoemtric deep learning; they also provide insights into the challenges that arise in non-Euclidean settings, which we expect will motivate further investigation in future work.

References

Joakim Andén and Stéphane Mallat. Multiscale scattering for audio classification. In *Proceedings of the 12th International Conference on Music Information Retrieval (ISMIR 2011)*, pages 657–662, 2011.

Joakim Andén and Stéphane Mallat. Deep scattering spectrum. *IEEE Transactions on Signal Processing*, 62(16):4114–4128, August 2014.

Joakim Andén, Vincent Lostanlen, and Stéphane Mallat. Classification with joint time-frequency scattering. arXiv:1807.08869, 2018.

Pierre Bérard, Gérard Besson, and Sylvestre Gallot. Embedding Riemannian manifolds by their heat kernel. *Geometric and Functional Analysis*, 4(4):373–398, 1994.

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.

Xavier Brumwell, Paul Sinz, Kwang Jin Kim, Yue Qi, and Matthew Hirn. Steerable wavelet scattering for 3D atomic systems with application to Li-Si energy prediction. In *NeurIPS Workshop on Machine Learning for Molecules and Materials*, 2018. arXiv:1812.02320.

Joan Bruna and Stéphane Mallat. Classification with scattering operators. In *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1561–1566, 2011.

Joan Bruna and Stéphane Mallat. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1872–1886, August 2013.

Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann LeCun. Spectral networks and deep locally connected networks on graphs. In *International Conference on Learning Representations (ICLR)*, 2014.

Václav Chudácek, Ronen Talmon, Joakim Andén, Stéphane Mallat, Ronald R Coifman, Patrice Abry, and Muriel Doret. Low dimensional manifold embedding for scattering coefficients of intrapartum fetal heart rate variability. In *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 6373–6376, 2014.

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

Ronald R. Coifman and Stéphane Lafon. Geometric harmonics: A novel tool for multiscale out-of-sample extension of empirical functions. *Applied and Computational Harmonic Analysis*, 21(1):31–52, July 2006b.

Ronald R. Coifman and Mauro Maggioni. Diffusion wavelets. *Applied and Computational Harmonic Analysis*, 21(1):53–94, 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.

Michael Eickenberg, Georgios Exarchakis, Matthew Hirn, and Stéphane Mallat. Solid harmonic wavelet scattering: Predicting quantum molecular energy from invariant descriptors of 3D electronic densities. In *Advances in Neural Information Processing Systems 30 (NIPS 2017)*, pages 6540–6549, 2017.

Michael Eickenberg, Georgios Exarchakis, Matthew Hirn, Stéphane Mallat, and Louis Thiry. Solid harmonic wavelet scattering for predictions of molecule properties. *Journal of Chemical Physics*, 148:241732, 2018.

Fernando Gama, Alejandro Ribeiro, and Joan Bruna. Diffusion scattering transforms on graphs. arXiv:1806.08829, 2018.

Feng Gao, Guy Wolf, and Matthew Hirn. Graph classification with geometric scattering. arXiv:1810.03068, 2018.

Evarist Giné. The addition formula for the eigenfunctions of the Laplacian. *Advances in Mathematics*, 18(1):102–107, 1975.

Matthew Hirn, Stéphane Mallat, and Nicolas Poilvert. Wavelet scattering regression of quantum chemical energies. *Multiscale Modeling and Simulation*, 15(2):827–863, 2017.

Lars Hörmander. The spectral function of an elliptic operator. *Acta Mathematica*, 121:193–218, 1968.

Shoshichi Kobayashi and Katsumi Nomizu. *Foundations of Differential Geometry*, Volume 1. Number 15 in Interscience Tracts in Pure and Applied Math. John Wiley and Sons Inc., New York, 1963.

Ron Levie, Federico Monti, Xavier Bresson, and Michael M. Bronstein. Cayleynets: Graph convolutional neural networks with complex rational spectral filters. arXiv:1705.07664, 2017.

Stéphane Mallat. Recursive interferometric representations. In *18th European Signal Processing Conference (EUSIPCO-2010)*, Aalborg, Denmark, 2010.

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

Yves Meyer. *Wavelets and Operators*, volume 1. Cambridge University Press, 1993.

Edouard Oyallon and Stéphane Mallat. Deep roto-translation scattering for object classification. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2865–2873, 2015.

Yiqian Shi and Bin Xu. Gradient estimate of an eigenfunction on a compact Riemannian manifold without boundary. *Annals of Global Analysis and Geometry*, 38:21–26, 2010.

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.

Laurent Sifre and Stéphane Mallat. Combined scattering for rotation invariant texture analysis. In *Proceedings of the 20th European Symposium on Artificial Neural Networks (ESANN 2012)*, 2012.

Laurent Sifre and Stéphane Mallat. Rotation, scaling and deformation invariant scattering for texture discrimination. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013.

Laurent Sifre and Stéphane Mallat. Rigid-motion scattering for texture classification. arXiv:1403.1687, 2014.

Joshua B. Tenenbaum, Vin de Silva, and John C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

Greg Turk and Marc Levoy. The Stanford bunny, 2005.

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

Guy Wolf, Stephane Mallat, and Shihab A. Shamma. Audio source separation with time-frequency velocities. In *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, Reims, France, 2014.

Guy Wolf, Stephane Mallat, and Shihab A. Shamma. Rigid motion model for audio source separation. *IEEE Transactions on Signal Processing*, 64(7):1822–1831, 2015.

Li Yi, Hao Su, Xingwen Guo, and Leonidas Guibas. Syncspeccnn: Synchronized spectral cnn for 3d shape segmentation. In *The Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Dongmian Zou and Gilad Lerman. Graph convolutional neural networks via scattering. arXiv:1804:00099, 2018.

A   Proof of Proposition 3.2

We note that

$$\|\pi_\lambda f\|_2^2 = \sum_{k \,:\, \lambda_k = \lambda} |\langle f, \varphi_k \rangle|^2,$$

where π_λ denotes projection onto the eigenspace E_λ. Therefore, applying (6) to $\|T_g f\|_2^2$ and $\|T_{h_\gamma} f\|_2^2$, we see that

$$
\begin{aligned}
\|\Phi f\|_{2,2}^2 &= \|T_g f\|_2^2 + \sum_{\gamma \in \Gamma} \|T_{h_\gamma} f\|_2^2 \\
&= \sum_{k \in \mathbb{N}} |g(\lambda_k)|^2 |\langle f, \varphi_k \rangle|^2 + \sum_{\gamma \in \Gamma} \sum_{k \in \mathbb{N}} |h_\gamma(\lambda_k)|^2 |\langle f, \varphi_k \rangle|^2 \\
&= \sum_{k \in \mathbb{N}} \left[|g(\lambda_k)|^2 + \sum_{\gamma \in \Gamma} |h_\gamma(\lambda_k)|^2 \right] |\langle f, \varphi_k \rangle|^2 \\
&= \sum_{\lambda \in \Lambda} m(\lambda) \left[|g(\lambda)|^2 + \sum_{\gamma \in \Gamma} |h_\gamma(\lambda)|^2 \right] \sum_{k : \lambda_k = \lambda} |\langle f, \varphi_k \rangle|^2 \\
&= \sum_{\lambda \in \Lambda} m(\lambda) \left[|g(\lambda)|^2 + \sum_{\gamma \in \Gamma} |h_\gamma(\lambda)|^2 \right] \|\pi_\lambda f\|_2^2.
\end{aligned}
$$

Therefore, the result follows from Parseval's Identity and the assumption that

$$
A \|f\|_2^2 \leq \|\Phi f\|_{2,2}^2 = \|T_g f\|_2^2 + \sum_{\gamma \in \Gamma} \|T_{h_\gamma} f\|_2^2 \leq B \|f\|_2^2. \qquad \square
$$

B   Proof of Proposition 3.3

We will show that

$$
m(\lambda) \left[|\widetilde{g}_J(\lambda)|^2 + \sum_{j \leq J} |\widetilde{h}_j(\lambda)|^2 \right] = 1, \quad \forall \lambda \in \Lambda. \tag{17}
$$

The result will then follow from Proposition 3.2.

We recall from (11) that

$$
h_j(\lambda) = \left[|g_{j-1}(\lambda)|^2 - |g_j(\lambda)|^2 \right]^{\frac{1}{2}},
$$

where g is a low-pass filter assumed by (7) to satisfy

$$
|g(\lambda)| \leq g(0) = 1 \text{ for all } \lambda \geq 0 \quad \text{and} \quad \lim_{\lambda \to \infty} g(\lambda) = 0.
$$

Therefore, recalling that for any spectral function η, the corresponding normalized spectral filter is defined by $\widetilde{\eta}(\lambda) = \frac{\eta(\lambda)}{\sqrt{m(\lambda)}}$, we see $m(\lambda) \left[|\widetilde{g}_J(\lambda)|^2 + \sum_{j \leq J} |\widetilde{h}_j(\lambda)|^2 \right] = |g(2^J \lambda)|^2 + \sum_{j \leq J} \left[|g(2^{j-1}\lambda)|^2 - |g(2^j \lambda)|^2 \right] = \lim_{j \to -\infty} |g(2^j \lambda)|^2 = g(0) = 1, \quad \forall \lambda \in \Lambda.$ $\qquad \square$

C   Proof of Lemma 4.3

Let K_η be the kernel of T_η. Then by the Cauchy-Schwartz inequality and the fact that $V_\zeta f(x) = f\left(\zeta^{-1}(x)\right)$,

$$
\begin{aligned}
|T_\eta f(x) - V_\zeta T_\eta f(x)| &= \left| \int_{\mathcal{M}} \left[K_\eta(x, y) - K_\eta\left(\zeta^{-1}(x), y\right) \right] f(y) \, dV(y) \right| \\
&\leq \|f\|_2 \left(\int_{\mathcal{M}} \left| K_\eta(x, y) - K_\eta\left(\zeta^{-1}(x), y\right) \right|^2 dV(y) \right)^{1/2} \\
&\leq \|f\|_2 \|\nabla K_\eta\|_\infty \left(\int_{\mathcal{M}} \left| r\left(x, \zeta^{-1}(x)\right) \right|^2 dV(y) \right)^{1/2} \\
&\leq \|f\|_2 \sqrt{\mathrm{vol}(\mathcal{M})} \|\nabla K_\eta\|_\infty \|\zeta\|_\infty.
\end{aligned}
$$

13

It follows that

$$\|T_\eta f - V_\zeta T_\eta f\|_2 \leq \mathrm{vol}(\mathcal{M})\|\nabla K_\eta\|_\infty\|\zeta\|_\infty\|f\|_2. \tag{18}$$

Lemma 5.5 shows

$$\|\nabla K_\eta\|_\infty \leq C(\mathcal{M})\sum_{k\in\mathbb{N}}\eta(\lambda_k)\lambda_k^{d/2},$$

and therefore

$$\|T_\eta f - V_\zeta T_\eta f\|_2 \leq C(\mathcal{M})\left(\sum_{k\in\mathbb{N}}\eta(\lambda_k)\lambda_k^{d/2}\right)\|\zeta\|_\infty\|f\|_2.$$

Now suppose that $|\eta(\lambda)| \leq C_0 e^{-t\lambda}$ for some constant C_0 and $t > 0$. Theorem 2.4 of Bérard et al. [1994] proves that for any $x \in \mathcal{M}$, $\alpha \geq 0$, and $t > 0$,

$$\sum_{k\geq 1}\lambda_k^\alpha e^{-t\lambda_k}|\varphi_k(x)|^2 \leq C(\mathcal{M})(\alpha+1)t^{-(d+2\alpha)/2}.$$

Integrating both sides over $\mathcal{M}$ yields:

$$\sum_{k\geq 1}\lambda_k^\alpha e^{-t\lambda_k} \leq C(\mathcal{M})(\alpha+1)t^{-(d+2\alpha)/2}. \tag{19}$$

Using the assumption that that $|\eta(\lambda)| \leq C_0 e^{-t\lambda}$, (18), and (19) with $\alpha = d/2$, gives

$$\|Tf - V_\zeta Tf\|_2 \leq C_0 C(\mathcal{M})\left(\sum_{k\geq 1}\lambda_k^{d/2}e^{-t\lambda_k}\right)\|\zeta\|_\infty\|f\|_2 \leq C_0 C(\mathcal{M})t^{-d}\|\zeta\|_\infty\|f\|_2. \qquad \square$$

D   Proof of Theorem 5.1

Let K_η be the kernel of T_η. As shown in the proof of Theorem 4.2,

$$K_\eta\left(\tilde{\zeta}(x),\tilde{\zeta}(y)\right) = K_\eta(x,y) \tag{20}$$

for all x and y in $\mathcal{M}$ and for all isometries $\tilde{\zeta}$. Therefore, we may apply Lemma 5.3, to see that

$$\|[T_\eta, V_\zeta]f\|_2 \leq C(\mathcal{M})\left[\|\nabla K\|_\infty A_3(\zeta) + \|K\|_{\mathbf{L}^2(\mathcal{M}\times\mathcal{M})}A_2(\zeta)\right]\|f\|_2.$$

Lemma 5.5 implies that

$$\|\nabla K_\eta\|_\infty \leq C(\mathcal{M})\sum_{k\in\mathbb{N}}\eta(\lambda_k)\lambda_k^{d/2},$$

and since $\{\varphi_k\}_{k=0}^\infty$ forms an orthonormal basis for $\mathbf{L}^2(\mathcal{M})$, we may verify that

$$\|K_\eta\|_{\mathbf{L}^2(\mathcal{M}\times\mathcal{M})} = \left(\sum_{k=0}^\infty|\eta(\lambda_k)|^2\right)^{1/2}.$$

Combining the above results completes the proof. $\qquad \square$

E   Proof of Theorem 5.2

Let K_η be the kernel of T_η. For any two pairs of points, (x,y) and (x',y') such that $r(x,y) = r(x',y')$, the definition of two-point homogeneity says that there exists an isometry $\tilde{\zeta}$ mapping x to x' and y to y'. As shown in the proof of Theorem 3.1, this implies $K_\eta(x',y') = K_\eta(x,y)$. Therefore, $K_\eta(x,y)$ is radial and we may write $K_\eta(x,y) = \kappa(r(x,y))$ for some $\kappa \in \mathbf{C}^1$.

By definition, any two-point homogeneous manifold is homogeneous, and so applying Lemma 5.4, we see that

$$\|[T_\eta, V_\zeta]f\|_2 \leq C(\mathcal{M})\left[\|\nabla K_\eta\|_\infty A_1(\zeta) + \|K_\eta\|_{\mathbf{L}^2(\mathcal{M}\times\mathcal{M})}A_2(\zeta)\right]\|f\|_2.$$

14

Lemma 5.5 implies that

$$\|\nabla K_\eta\|_\infty \le C(\mathcal{M}) \sum_{k \in \mathbb{N}} \eta(\lambda_k) \lambda_k^{(d+1)/4},$$

and since $\{\varphi_k\}_{k=0}^\infty$ forms an orthonormal basis for $\mathbf{L}^2(\mathcal{M})$, it can be checked that

$$\|K_\eta\|_{\mathbf{L}^2(\mathcal{M} \times \mathcal{M})} = \left(\sum_{k=0}^\infty |\eta(\lambda_k)|^2 \right)^{1/2}.$$

The proof follows from combining the above inequalities. $\qquad\square$

F   Proof of Lemma 5.3

We first compute

$$
\begin{aligned}
|[T, V_\zeta] f(x)| &= \left| \int_\mathcal{M} K(x,y) f\left(\zeta^{-1}(y)\right) dV(y) - \int_\mathcal{M} K\left(\zeta^{-1}(x), y\right) f(y) dV(y) \right| \\
&= \left| \int_\mathcal{M} K(x, \zeta(y)) f(y) |\det[D\zeta(y)]| dV(y) - \int_\mathcal{M} K\left(\zeta^{-1}(x), y\right) f(y) dV(y) \right| \\
&= \left| \int_\mathcal{M} f(y) \left[K(x, \zeta(y)) |\det[D\zeta(y)]| - K\left(\zeta^{-1}(x), y\right) \right] dV(y) \right| \\
&\le \left| \int_\mathcal{M} f(y) K(x, \zeta(y)) \left[|\det[D\zeta(y)]| - 1 \right] dV(y) \right| \\
&\quad + \left| \int_\mathcal{M} f(y) \left[K(x, \zeta(y)) - K\left(\zeta^{-1}(x), y\right) \right] dV(y) \right| \\
&\le \||\det[D\zeta(y)]| - 1\|_\infty \left| \int_\mathcal{M} f(y) K(x, \zeta(y)) dV(y) \right| \\
&\quad + \left| \int_\mathcal{M} f(y) \left[K(x, \zeta(y)) - K(\zeta^{-1}(x), y) \right] dV(y) \right|.
\end{aligned}
$$

Therefore, by the Cauchy-Schwartz inequality,

$$
\begin{aligned}
\|[T, V_\zeta] f\|_2 \le \|f\|_2 \Bigg[\||\det[D\zeta(y)]| - 1\|_\infty &\left(\int_\mathcal{M} \int_\mathcal{M} |K(x, \zeta(y))|^2 \, dV(y) dV(x) \right)^{\frac{1}{2}} \\
&+ \left(\int_\mathcal{M} \int_\mathcal{M} \left| K(x, \zeta(y)) - K\left(\zeta^{-1}(x), y\right) \right|^2 \, dV(y) dV(x) \right)^{\frac{1}{2}} \Bigg].
\end{aligned}
$$

We may bound the first integral by observing

$$\int_\mathcal{M} \int_\mathcal{M} |K(x, \zeta(y))|^2 \, dV(y) dV(x) \le \|\det[D\zeta^{-1}(y)]\|_\infty^2 \int_\mathcal{M} \int_\mathcal{M} |K(x,y)|^2 \, dV(y) dV(x).$$

To bound the second integral, let $\zeta = \zeta_1 \circ \zeta_2$ be a factorization of ζ into an isometry ζ_1 and a diffeomorphism ζ_2 such that $\|\zeta_2\|_\infty \le 2A_3(\zeta)$. Then

$$
\begin{aligned}
\int_\mathcal{M} \int_\mathcal{M} &\left| K(x, \zeta(y)) - K\left(\zeta^{-1}(x), y\right) \right|^2 \, dV(y) dV(x) \\
&= \int_\mathcal{M} \int_\mathcal{M} \left| K(x, \zeta_1(\zeta_2(y))) - K\left(\zeta_2^{-1}\left(\zeta_1^{-1}(x)\right), y\right) \right|^2 \, dV(y) dV(x) \\
&= \int_\mathcal{M} \int_\mathcal{M} \left| K(\zeta_1^{-1}(x), \zeta_2(y)) - K\left(\zeta_2^{-1}\left(\zeta_1^{-1}(x)\right), y\right) \right|^2 \, dV(y) dV(x)
\end{aligned}
$$

where the last equality uses the assumption (16). Now,

$$\int_{\mathcal{M}} \int_{\mathcal{M}} \left| K\left(\zeta_1^{-1}(x), \zeta_2(y)\right) - K\left(\zeta_2^{-1}(\zeta_1^{-1}(x)), y\right) \right|^2 dV(y) dV(x)$$

$$\leq 2 \int_{\mathcal{M}} \int_{\mathcal{M}} \left| K\left(\zeta_1^{-1}(x), \zeta_2(y)\right) - K\left(\zeta_1^{-1}(x), y\right) \right|^2 dV(y) dV(x)$$

$$+ 2 \int_{\mathcal{M}} \int_{\mathcal{M}} |K(\zeta_1^{-1}(x), y) - K(\zeta_2^{-1}(\zeta_1^{-1}(x)), y)|^2 dV(y) dV(x)$$

$$\leq 4 \|\zeta_2\|_\infty^2 \|\nabla K\|_\infty^2 \mathrm{vol}(\mathcal{M})^2.$$

Combining the above inequalities completes the proof. $\qquad \square$

G Proof of Lemma 5.4

We repeat the proof of Lemma 5.3, to see that it suffices to show

$$\int_{\mathcal{M}} \int_{\mathcal{M}} \left| K(x, \zeta(y)) - K\left(\zeta^{-1}(x), y\right) \right|^2 dV(y) dV(x) \leq [\|\nabla K\|_\infty A_1(\zeta) \mathrm{diam}(\mathcal{M}) \mathrm{vol}(\mathcal{M})]^2.$$

By the assumption that $K(x, y) = \kappa(r(x, y))$ is radial, we see

$$\int_{\mathcal{M}} \int_{\mathcal{M}} |K(x, \zeta(y))\ K\left(\zeta^{-1}(x), y\right)|^2 dV(y) dV(x)$$

$$= \int_{\mathcal{M}} \int_{\mathcal{M}} \left| \kappa(r(x, \zeta(y))) - \kappa\left(r\left(\zeta^{-1}(x), y\right)\right) \right|^2 dV(y) dV(x)$$

$$\leq \|\kappa'\|_\infty^2 \int_{\mathcal{M}} \int_{\mathcal{M}} \left| r(x, \zeta(y)) - r\left(\zeta^{-1}(x), y\right) \right|^2 dV(y) dV(x)$$

$$\leq [\|\kappa'\|_\infty A_1(\zeta)]^2 \int_{\mathcal{M}} \int_{\mathcal{M}} |r(x, \zeta(y))|^2 dV(y) dV(x)$$

$$\leq [\|\kappa'\|_\infty A_1(\zeta) \mathrm{diam}(\mathcal{M}) \mathrm{vol}(\mathcal{M})]^2.$$

Since $K(x, y) = \kappa(r(x, y))$, we see that $\|\nabla K\|_\infty = \|\kappa'\|_\infty$, which completes the proof. $\qquad \square$

H Proof of Lemma 5.5

For any $\lambda_k = \lambda$, it is a consequence of Hörmander's local Weyl law (Hörmander, 1968; see also Shi and Xu, 2010) that

$$\|\varphi_k\|_\infty \leq C(\mathcal{M}) \lambda^{(d-1)/4}. \tag{21}$$

Theorem 1 of Shi and Xu [2010] shows that

$$\|\nabla \varphi_k\|_\infty \leq C(\mathcal{M}) \sqrt{\lambda} \|\varphi_k\|_\infty. \tag{22}$$

Therefore,

$$\left| \nabla K^{(\lambda)}(x, y) \right|^2 = \left| \sum_{k:\lambda_k=\lambda} \nabla \varphi_k(x) \overline{\varphi}_k(y) \right|^2 \leq \left(\sum_{k:\lambda_k=\lambda} |\nabla \varphi_k(x)|^2 \right) \left(\sum_{k:\lambda_k=\lambda} |\varphi_k(y)|^2 \right)$$

$$\leq C(\mathcal{M}) m(\lambda) \lambda^{(d-1)/2} \sum_{k:\lambda_k=\lambda} |\nabla \varphi_k(x)|^2 \leq C(\mathcal{M}) m(\lambda) \lambda^{(d+1)/2} \sum_{k:\lambda_k=\lambda} \|\varphi_k\|_\infty^2$$

$$\leq C(\mathcal{M}) m(\lambda)^2 \lambda^d.$$

Furthermore, if we assume that $\mathcal{M}$ is homogeneous, then Theorem 3.2 of Giné [1975] shows that $\sum_{k:\lambda_k=\lambda} |\varphi_k(y)|^2 = C(\mathcal{M}) m(\lambda)$. Substituting this into the above string of inequalities yields

$$\left| \nabla K^{(\lambda)}(x, y) \right|^2 \leq C(\mathcal{M}) m(\lambda)^2 \lambda^{(d+1)/2}.$$

The bounds on $\|K_\eta\|_\infty$ follow from recalling that $K_\eta(x, y) = \sum_{\lambda \in \Lambda} \eta(\lambda) K^{(\lambda)}(x, y)$ and applying the triangle inequality. $\qquad \square$