Applied and Computational Harmonic Analysis 70 (2024) 101635

Contents lists available at ScienceDirect

Applied and Computational Harmonic Analysis

journal homepage: www.elsevier.com/locate/acha

Full Length Article

Geometric scattering on measure spaces

Joyce Chew, Matthew Hirn, Smita Krishnaswamy *, Deanna Needell, Michael Perlmutter *, Holly Steach, Siddharth Viswanath, Hau-Tieng Wu

ARTICLE INFO

Communicated by Thomas Strohmer

Keywords:
Geometric deep learning
Manifold learning
Scattering transforms
Stability and invariance

ABSTRACT

The scattering transform is a multilayered, wavelet-based transform initially introduced as a mathematical model of convolutional neural networks (CNNs) that has played a foundational role in our understanding of these networks' stability and invariance properties. In subsequent years, there has been widespread interest in extending the success of CNNs to data sets with non-Euclidean structure, such as graphs and manifolds, leading to the emerging field of geometric deep learning. In order to improve our understanding of the architectures used in this new field, several papers have proposed generalizations of the scattering transform for non-Euclidean data structures such as undirected graphs and compact Riemannian manifolds without boundary. Analogous to the original scattering transform, these works prove that these variants of the scattering transform have desirable stability and invariance properties and aim to improve our understanding of the neural networks used in geometric deep learning.

In this paper, we introduce a general, unified model for geometric scattering on measure spaces. Our proposed framework includes previous work on compact Riemannian manifolds without boundary and undirected graphs as special cases but also applies to more general settings such as directed graphs, signed graphs, and manifolds with boundary. We propose a new criterion that identifies to which groups a useful representation should be invariant and show that this criterion is sufficient to guarantee that the scattering transform has desirable stability and invariance properties. Additionally, we consider finite measure spaces that are obtained from randomly sampling an unknown manifold. We propose two methods for constructing a data-driven graph on which the associated graph scattering transform approximates the scattering transform on the underlying manifold. Moreover, we use a diffusion-maps based approach to prove quantitative estimates on the rate of convergence of one of these approximations as the number of sample points tends to infinity. Lastly, we showcase the utility of our method on spherical images, a directed graph stochastic block model, and on high-dimensional single-cell data.

1. Introduction

Many popular machine learning algorithms and architectures either explicitly or implicitly rely on producing a hidden, or transformed, representation of the input data. For example, popular algorithms such as word2vec [18], node2vec [37], and graph2vec [62] explicitly associate each input in a text corpus, network, or collection of networks to a point in a high-dimensional vector space. This transformed representation can then be used for a variety of tasks such as clustering or classification. Deep neural networks, on the

* Corresponding author.
E-mail addresses: smita.krishnaswamy@yale.edu (S. Krishnaswamy), mperlmutter@boisestate.edu (M. Perlmutter).

https://doi.org/10.1016/j.acha.2024.101635
Received 17 August 2022; Received in revised form 1 August 2023; Accepted 26 January 2024
Available online 6 February 2024

other hand, use multilayered architectures to classify an input signal. In this case, the early layers of the network may be viewed as producing a transformed representation of the input and the final layer may be viewed as a classifier acting on the transformed data. In either case, there is a fundamental question. What properties should these hidden representations satisfy in order to be useful for downstream tasks?

In order to help answer this question, Mallat introduced the scattering transform [57], a wavelet-based architecture which models the hidden representation produced by the early layers of a convolutional neural network (CNN). Given a function $f \in \mathbf{L}^2(\mathbb{R}^N)$ and a scale parameter J, the windowed scattering transform of [57] is a countable collection of functions

$$S_J f := \{S_J[p]f : p = (j_1, \dots, j_m), j_i \leq J, m \geq 0\}, \tag{1}$$

where the scattering coefficients $S_J[p]f$ are defined through an alternating sequence of m wavelet convolutions (at scales j_i) and nonlinear activations followed by a final convolution against a low-pass averaging filter at scale 2^J. If one is interested in classifying many signals $\{f_i\}_{i=1}^{N_{\text{signals}}}$, they may first transform the input data by computing $S_J f_i$ for each i and then use these transformed representations as input to a classification model such as a support vector machine.

One of the key insights of [57] is that convolutional architectures naturally have desirable invariance and equivariance properties with respect to the action of the translation group. Specifically, if τ_c is the translation operator $\tau_c f(x) := f(x - c)$, we have the equivariance relationship

$$S_J[p](\tau_c f) = \tau_c S_J[p]f, \tag{2}$$

where on the right-hand side τ_c is applied term by term. Moreover, when the scale parameter J tends to infinity, we have the approximate invariance relationship

$$S_J[p](\tau_c f) \approx S_J[p]f. \tag{3}$$

Furthermore, Mallat also shows that the scattering transform is stable to the perturbations of the form $f(x - c(x))$ where is $c(x)$ is a function with bounded gradient and Hessian.

In addition to being a theoretical model, the scattering transform has also proven to be a practical object. A notable difference between the scattering transform and other CNN-like architectures is that it uses predesigned wavelet filters, rather than filters learned from training data. In settings where labeled data is abundant, this may be viewed as a limitation on the expressive power of the scattering transform. However, in the context of unsupervised learning, or limited data environments, it may be difficult or impossible to train a traditional neural network. In these settings, the lack of trainable filters *increases* the practical utility of the scattering transform [49]. For instance, [68] applied the scattering transform to Sonar data to detect unexploded bombs on the ocean floor despite there only being 14 objects in the data set. Additionally, the scattering transform can also be used for a variety of other tasks in addition to classification. For example, [9] applied it to the texture synthesis problem and [75] combined the scattering transform with nonnegative matrix factorization in order to achieve audio source separation.

While CNNs have had tremendous success for tasks related to images, audio signals, and other data with a Euclidean grid-like structure, many modern data sets have an irregular structure and are naturally modeled as more complex structures such as graphs and manifolds. This has led to the new field of *geometric deep learning* [8] which aims to extend the success of CNNs to these irregular domains. In these more general settings, the concept of translation is not well defined. However, invariance and equivariance still play a critical role. For example, nearly all popular graph neural networks are designed so that they are naturally invariant or equivariant to the action of the permutation group, i.e., reordering of the vertices. More generally, one of the principal goals of geometric deep learning is to design architectures that respect the intrinsic symmetries and invariances of the data, which are typically modeled by group actions [7,11].

There are many possible ways to accomplish this goal, but here we will focus on spectral methods based on the eigendecomposition of a suitable Laplace type operator such as the graph Laplacian or Laplace-Beltrami operator on a manifold. These methods, which have been popularized through work such as [71], view the eigenvectors/eigenfunctions of the Laplace operator as generalized Fourier modes and define convolution as multiplication in the Fourier basis analogous to well-known results in the Euclidean case. This notion of convolution is used in popular graph neural networks such as [10], [26], [50], and [45] and has also been used in manifold neural networks such as [81] and [82].

Following the rise of these spectral networks, several works have introduced versions of the scattering transform for (undirected, unsigned) graphs [89,34,35,64] and smooth compact manifolds without boundary [63]. In these works, the authors assume that one is given a signal f defined on the graph or manifold and use generalizations of the wavelet transform [39,21] to construct scattering coefficients similar to (1) through an alternating sequence of generalized convolutions and nonlinearities. They then provide detailed stability and invariance analysis of their respective versions of the scattering transform proving results analogous to (2) and (3). Thereby, they help improve our understanding of spectral networks used in geometric deep learning, analogous to how [57] helps us understand Euclidean CNNs. Moreover, there has also been work applying the graph scattering transform to combinatorial optimization [60] problems and to graph synthesis [88,3], an important problem with potential applications to drug discovery.

1.1. Contributions and main results

In this work, we extend the scattering transform to a general class of measure spaces $\mathcal{X} := (X, \mathcal{F}, \mu)$. Our framework applies both to the settings considered in previous work on geometric scattering, i.e., compact Riemannian manifolds without boundary and unweighted, unsigned graphs, and also to other interesting examples such as signed or directed graphs, which have not been previously considered in the literature about the scattering transform. The generality of our construction is motivated in part by the recent book [7] which laments "there is a veritable zoo of neural network architectures for various kinds of data, but few unifying principles." In the same spirit, we look for the general themes that unite spectral networks on different domains and formulate a general theory of scattering networks on measure spaces.

Analogous to, e.g., [64] and [63], we will construct two versions of the scattering transform on $\mathcal{X}$. For both of these transforms, we will assume that we are given a signal f defined on $\mathcal{X}$ and analogous to (1) will represent f via a sequence of scattering coefficients obtained through alternating sequences of convolutions and pointwise nonlinearities followed by a final aggregation. In the first version, which we refer to as the *windowed scattering transform*, the aggregation step is given by convolution against a low-pass filter that can be viewed as a local-averaging operator. We also define a *non-windowed scattering transform* where the final aggregation is computed via a global integration. Importantly, we note that the windowed scattering transform outputs a sequence of vectors (i.e., functions) whereas the non-windowed scattering transform outputs a sequence of scalars.

We will examine the invariance and equivariance properties of these representations and establish results analogous to (2) and (3). Towards this end, we let $\mathcal{G}$ be a group of bijections from X to X with proper structure. We let $\mathcal{G}$ act on $\mathbf{L}^2(\mathcal{X})$ by composition and on a Laplacian-type operator $\mathcal{L}$ by conjugation. Specifically, for $\zeta \in \mathcal{G}$ we define

$$V_\zeta f(x) := f(\zeta^{-1}(x)) \quad \text{and} \quad \mathcal{L}_\zeta := V_\zeta \circ \mathcal{L} \circ V_\zeta^{-1}.$$

In the case where $\mathcal{X}$ is a graph or a manifold, it is natural to take $\mathcal{G}$ to be the permutation group or the isometry group respectively. However, on an arbitrary measure space, it is not obvious what groups our representation should be invariant to.

Perhaps the most natural idea would be the group of all bijections that preserves measures in the sense that $\mu(\zeta^{-1}(B)) = \mu(B)$ for all $\zeta \in \mathcal{G}$ and $B \in \mathcal{F}$. Indeed, for the windowed scattering transform, our analysis will show that this condition is needed prove a result analogous to (2) establishing invariance in the limit as the scale parameter tends to infinity. However, it will not be needed in order to establish our other primary invariance and equivariance results. This is fortunate because conservation of measure is actually a stronger condition than it appears at first glance. For example, it does not hold when $\mathcal{X}$ is a graph, $\mathcal{G}$ is the permutation group, and different vertices are assigned different weights. Instead, our other invariance and equivariance results will only require the weaker assumption that V_ζ is an isometry on $\mathbf{L}^2(\mathcal{X})$, i.e.,

$$\langle V_\zeta f_1, V_\zeta f_2 \rangle_{\mathbf{L}^2(\mathcal{X})} = \langle f_1, f_2 \rangle_{\mathbf{L}^2(\mathcal{X})}.$$

One may verify that this condition holds for the permutation group on graphs for arbitrary choices of the measure.

In addition to significantly generalizing previous constructions of the geometric scattering transform, we also use the methods based on diffusion maps [20] and Laplacian Eigenmaps [1,2] to show that the scattering transform on manifolds can be interpreted as the limit of the scattering transform on data-driven graphs. In short, if we have a collection of points $\{x_i\}_{i=0}^{N-1} \subseteq \mathbb{R}^D$ that lie on a d-dimensional manifold for some $d \ll D$, we will use a kernel to construct an affinity matrix W which can be interpreted as the adjacency matrix of a weighted graph. We use this affinity matrix to construct a data-driven approximation of the Laplace-Beltrami operator which we then use to construct an approximation of the windowed and non-windowed manifold scattering transforms. We will then prove theorems guaranteeing the rates of convergence of these methods as the number of sample points tends to infinity. To the best of our knowledge, this is the first attempt to prove such convergence guarantees for any neural-network-like architecture constructed from the Laplace-Beltrami operator.

In summary, we provide a theoretically justified model for understanding spectral neural networks on measure spaces paralleling the role of the original scattering transform [57] in understanding CNNs. Towards this end, we note that equivariance results similar to ours can likely be obtained for other networks such as the ones considered in [81] or [87] constructed through the spectrum of the appropriate Laplace operator. Similarly, our methods can likely be adapted to study the convergence of other spectral manifold neural networks.

1.2. Notation and organization

Throughout, we will let $\mathcal{X} = (X, \mathcal{F}, \mu)$ be a measure space with set X, σ-algebra $\mathcal{F}$, and measure μ. We let $\mathcal{H} = \mathbf{L}^2(\mathcal{X})$ denote the Hilbert space of functions which are square integrable on $\mathcal{X}$ and for $f \in \mathcal{H}$ we will denote its norm by either $\|f\|_{\mathcal{H}}$ or $\|f\|_{\mathbf{L}^2(\mathcal{X})}$. Similarly, for $f, g \in \mathcal{H}$, we shall denote their inner product by $\langle f, g \rangle_{\mathcal{H}}$ or $\langle f, g \rangle_{\mathbf{L}^2(\mathcal{X})}$. If T is an operator on $\mathcal{H}$, we will let $\|T\|_{\mathcal{H}}$ denote its operator norm. If $\mathbf{x}$ and $\mathbf{y} \in \mathbb{R}^N$ are vectors, we shall use $\|\mathbf{x}\|_2$ and $\langle \mathbf{x}, \mathbf{y} \rangle_2$ to denote their ℓ^2-norm and inner product. Similarly, if A is a matrix, we will let $\|A\|_2$ denote its operator norm on ℓ^2. We shall let $\mathcal{L}$ be a positive semidefinite, self-adjoint operator on $\mathcal{H}$ and denote its eigenfunctions and eigenvalues by φ_k and λ_k for k in some at most countable indexing set $\mathcal{I}$. If $\{f_j\}_{j \in \mathcal{J}}$ is an at most countable collection of elements in $\mathcal{H}$, we shall define $\|\{f_j\}_{j \in \mathcal{J}}\|_{\ell^2(\mathcal{H})}$ by

$$\|\{f_j\}_{j \in \mathcal{J}}\|_{\ell^2(\mathcal{H})}^2 := \sum_{j \in \mathcal{J}} \|f\|_{\mathcal{H}}^2.$$

We shall let $\mathcal{G}$ denote a group of bijections $X \to X$, and for $\zeta \in \mathcal{G}$, we shall let V_ζ denote the operator defined by $V_\zeta f(x) = f(\zeta^{-1}(x))$.

Our construction of the scattering transform will be based on a collection of wavelets $\mathcal{W} := \{W_j\}_{j\in\mathcal{J}} \cup \{A\}$ where $\mathcal{J}$ is an at most countable indexing set. We will let $p := (j_1, \dots, j_m)$ denote a scattering path of length m, and for $f \in \mathcal{H}$ let $S[p]f$ and $\overline{S}[p]f$ denote corresponding windowed and non-windowed scattering coefficients. We shall let $\{H_t\}_{t\geq 0}$ denote a semigroup of operators defined on $\mathcal{H}$ defined in terms of a spectral function $g : [0,\infty) \to [0,\infty)$. When notationally convenient, we will write H^t instead of H_t. In Section 6, we will consider finite subsets $X_N \subseteq X$ of cardinality N and let $\mathcal{X}_N$ be a corresponding measure space. In this setting, we will denote objects corresponding to $\mathcal{X}_N$ with either a subscript or superscript N.

The rest of this paper is organized as follows. In Section 2, we will define convolution, the wavelet transform, and the scattering transform on a measure space $\mathcal{X}$. In Section 3, we will discuss examples of measure spaces included in our framework, some of which have been considered in previous work on the scattering transform and some which have not. In Section 4, we will establish fundamental continuity and invariance properties of the measure space scattering transform and in Section 5 we will consider stability to perturbations. In Section 6, we will introduce numerical methods for implementing the scattering transform in the case where $\mathcal{X}$ is a manifold, but one only has access to $\mathcal{X}$ through a finite number of samples. We will also prove the convergence of these methods as the number of sample points tends to infinity. In Section 7, we will present numerical experiments on both synthetic data and on real-world biomedical data before providing a brief conclusion in Section 8.

2. Definitions

In this section, we first define convolution and wavelets on a measure space $\mathcal{X}$ and then use these wavelets to define the measure space scattering transform.

Let $\mathcal{X} = (X, \mathcal{F}, \mu)$ be a measure space with set X, σ-algebra $\mathcal{F}$, and measure μ. Let $\mathrm{vol}(\mathcal{X}) := \mu(X)$, and let $\mathcal{H} = \mathbf{L}^2(\mathcal{X})$ denote the Hilbert space of measurable functions such that

$$\|f\|_{\mathcal{H}}^2 := \|f\|_{\mathbf{L}^2(\mathcal{X})}^2 := \int_X |f|^2 d\mu < \infty.$$

Let $\mathcal{L}$ be a self-adjoint and positive semidefinite operator on $\mathcal{H}$, and let $\mathcal{I}$ be an at most countable set of nonnegative integers. Without loss of generality, we assume either $\mathcal{I}$ is the natural numbers $\mathbb{N} \cup \{0\}$ or that $\mathcal{I} = \{0, \dots, N-1\}$ for some $N \in \mathbb{N}$. We assume that there is a collection of functions $\{\varphi_k\}_{k\in\mathcal{I}} \subset \mathcal{H}$ such that $\mathcal{L}\varphi_k = \lambda_k \varphi_k$, with $\lambda_0 = 0 < \lambda_1$ and $\lambda_k \leq \lambda_{k+1}$ for $k \geq 1$. We also assume that $\{\varphi_k\}_{k\in\mathcal{I}}$ forms an orthonormal basis for $\mathcal{H}$.

2.1. Convolution and wavelet transforms

For $f \in \mathcal{H}$, we define its generalized Fourier coefficients $\widehat{f}(k), k \in \mathcal{I}$, by

$$\widehat{f}(k) := \langle f, \varphi_k \rangle_{\mathcal{H}}.$$

Since $\{\varphi_k\}_{k\in\mathcal{I}}$ is an orthonormal basis, we obtain the generalized Fourier series

$$f = \sum_{k\in\mathcal{I}} \widehat{f}(k)\varphi_k,$$

where the convergence is in the $\mathbf{L}^2(\mathcal{X})$ sense if $\mathcal{I}$ is infinite. In the case when $\mathcal{X}$ is the unit circle, it is well known that convolution corresponds to multiplication in the Fourier domain. Therefore, for any $h \in \mathcal{H}$, we define a convolution operator T_h by

$$T_h f := h \star f := \sum_{k\in\mathcal{I}} \widehat{h}(k)\widehat{f}(k)\varphi_k. \tag{4}$$

One may verify that for any $n \geq 0$ we have

$$(T_h)^n f = \sum_{k\in\mathcal{I}} \widehat{h}(k)^n \widehat{f}(k)\varphi_k. \tag{5}$$

Therefore, if $\widehat{h}(k)$ is nonnegative for all k we may define, for $t \in \mathbb{R}$, T_h^t by

$$T_h^t f := \sum_{k\in\mathcal{I}} \widehat{h}(k)^t \widehat{f}(k)\varphi_k. \tag{6}$$

In the case where $t = 1/2$, we note that we have $T_h^{1/2} T_h^{1/2} = T_h$. Therefore, we will refer to $T_h^{1/2}$ as the square root of T_h.

We will use this notion of spectral convolution to construct a diffusion operator H. To do this, we let $g : [0,\infty) \to [0,\infty)$ be a nonnegative and nonincreasing function with

$$g(0) = 1 \quad \text{and } g(t) < 1 \quad \text{for all } t > 0. \tag{7}$$

For $t \geq 0$, we define H^t to be the operator corresponding to convolution against $\sum_{k\in\mathcal{I}} g(\lambda_k)^t \varphi_k$, i.e.,

$$H^t f := \sum_{k\in\mathcal{I}} g(\lambda_k)^t \widehat{f}(k)\varphi_k. \tag{8}$$

We note that by construction, $\{H^t\}_{t\geq 0}$ forms a semigroup since $H^t H^s = H^{t+s}$ and $H^0 = \mathrm{Id}$ is the identity operator. We also note that $H^t = (H^1)^t$, where the exponents are interpreted as in (5) and (6). Motivated by the interpretation of t as an exponent, we will occasionally write H in place of H^1 when convenient.

As our primary example, we will take $g(\lambda) = e^{-\lambda}$, in which case, one may verify that, for sufficiently well-behaved functions, $u_f(x,t) := H^t f(x)$ satisfies the heat equation

$$\partial_t u_f = -\mathcal{L}_x u_f, \quad \lim_{t\to 0} u(t,x) = f(x),$$

since we may compute

$$\begin{aligned}\partial_t H^t f(x) &= \partial_t \sum_{k\in\mathcal{I}} e^{-\lambda_k t}\widehat{f}(k)\varphi_k(x)\\ &= \sum_{k\in\mathcal{I}} -\lambda_k e^{-\lambda_k t}\widehat{f}(k)\varphi_k(x)\\ &= -\mathcal{L}_x H^t f(x).\end{aligned} \tag{9}$$

Therefore, in this case, $\{H^t\}_{t\geq 0}$ is referred to as the *heat semigroup* and t is referred to as the *diffusion time*.

Remark 1. The definition of H does not depend on the choice of eigenbasis, even when some eigenvalues have multiplicity greater than one. To see this, let Λ be the set of distinct eigenvalues of $\mathcal{L}$ and note that

$$Hf = \sum_{k\in\mathcal{I}} g(\lambda_k)\widehat{f}(k)\varphi_k = \sum_{\lambda\in\Lambda} g(\lambda) \sum_{k:\lambda_k=\lambda} \widehat{f}(k)\varphi_k = \sum_{\lambda\in\Lambda} g(\lambda)\pi_\lambda(f),$$

where, for $\lambda\in\Lambda$, π_λ is the operator which projects a function onto the eigenspace corresponding to λ.

Given this diffusion operator we define the wavelet transform

$$\mathcal{W}_J f := \{W_j f\}_{j=0}^J \cup \{A_J f\}, \tag{10}$$

where $W_0 := \mathrm{Id} - H^1$, $A_J := H^{2^J}$, and for $1\leq j\leq J$

$$W_j := H^{2^{j-1}} - H^{2^j}.$$

The wavelets aim to capture the geometry of $\mathcal{X}$ at different scales. In particular, the W_j track changes between the geometry at different diffusion times. The operator A_J performs a localized averaging operation and may be interpreted as a low-pass filter. Our construction uses a minimal time scale of 1 for simplicity and notational convenience. However, if one wishes to obtain wavelets which are sensitive to smaller time scales, they may simply change the spectral function g. For example, if $g_1(\lambda) = e^{-\lambda}$ and $g_2(\lambda) = e^{-\lambda/2}$ then the associated diffusion operators would satisfy $H_2^1 = H_1^{1/2}$.

The following result shows that $\mathcal{W}_J$ is a nonexpansive frame on $\mathcal{H}$. Its proof is identical to the proof of Proposition 2.2 of [64]. For completeness, we give full details in Appendix A.

Proposition 1. *There exists a universal constant $c>0$ such that for all $f\in\mathcal{H}$*

$$c\|f\|_{\mathcal{H}}^2 \leq \|\mathcal{W}_J f\|_{\ell^2(\mathcal{H})}^2 := \sum_{j=0}^J \|W_j f\|_{\mathcal{H}}^2 + \|A_J f\|_{\mathcal{H}}^2 \leq \|f\|_{\mathcal{H}}^2.$$

Remark 2. If we instead define our wavelets by $W_0' = \sqrt{Id - H}$, $W_j' = \sqrt{H^{2^{j-1}} - H^{2^j}}$ for $1\leq j\leq J$, and $A_J' = \sqrt{H^{2^J}}$, we can obtain a similar result for $\mathcal{W}_J' f = \{W_j'\}_{j=0}^J \cup \{A_J' f\}$ but with $c=1$, so that the wavelet transform is norm-preserving, i.e.,

$$\sum_{j=0}^J \|W_j' f\|_{\mathcal{H}}^2 + \|A_J' f\|_{\mathcal{H}}^2 = \|f\|_{\mathcal{H}}.$$

The proof is identical to the proof of Proposition 2.1 of [64].

2.2. The scattering transform

In this section, we will construct the scattering transform as a multilayered architecture built off of a filter bank $\mathcal{W}$. For the sake of generality, we will not require our $\mathcal{W}$ to be the diffusion wavelets constructed in the previous subsection. Instead, we let $\mathcal{J}$ be an arbitrary countable indexing set and assume

$$\mathcal{W} = \{W_j\}_{j\in\mathcal{J}} \cup \{A\}$$

is any collection of operators such that

$$c\|f\|_{\mathcal{H}}^2 \leq \|\mathcal{W}f\|_{\ell^2(\mathcal{H})}^2 = \sum_{j\in\mathcal{J}} \|W_j f\|_{\mathcal{H}}^2 + \|Af\|_{\mathcal{H}}^2 \leq \|f\|_{\mathcal{H}}^2 \tag{11}$$

for some $c > 0$. This generality is motivated both by the fact that several different versions of the graph scattering transform [89,34,35] have used different wavelet constructions and also by various works which have constructed versions of the Euclidean scattering transform using generalized, non-wavelet filter banks [24,36,84,85]. Here, we note that the letter of A is chosen because we typically interpret A as an averaging operator analogous to the low-pass operator A_J considered in (10). However, we emphasize that this is merely suggestive notation.

The scattering transform consists of an alternating sequence of linear filterings (typically wavelet transforms) and nonlinear activations. Towards this end, we let σ be an nonlinear function defined on either $\mathbb{R}$ or $\mathbb{C}$ such that the real part of $\sigma(x)$ is nonnegative and σ is non-expansive in the sense that $|\sigma(x) - \sigma(y)| \leq |x - y|$. In a slight abuse of notation let $\sigma : \mathcal{H} \to \mathcal{H}$ also denote the operator defined by $(\sigma f)(x) := \sigma(f(x))$. We note that in the case where admissible choices of σ include the absolute value function which is commonly used in papers concerning the scattering transform, the rectified linear unit (ReLU) which is commonly used in other neural network architectures, and the complex version of ReLU considered in [87].

Given $\mathcal{W}_J$ and σ, we define the *windowed* scattering transform $S : \mathcal{H} \to \ell^2(\mathcal{H})$ by

$$Sf := \{S[p]f : m \geq 0, p := (j_1, \ldots, j_m) \in \mathcal{J}^m\},$$

where $\mathcal{J}^m$ is the m-fold Cartesian product of $\mathcal{J}$, and the windowed scattering coefficients $S[p]$ corresponding to the path $p = (j_1, \ldots, j_m) \in \mathcal{J}^m$ are defined by

$$S[p]f := AU[p]f, \quad U[p]f := \sigma W_{j_m} \ldots \sigma W_{j_1} f$$

for $m \geq 1$, and when $m = 0$ and p_e is the "empty path", we declare that

$$S[p_e]f := Af. \tag{12}$$

We also define an operator U by

$$Uf := \{U[p]f : m \geq 0, p = (j_1, \ldots, j_m) \in \mathcal{J}^m\} \tag{13}$$

and a *non-windowed* scattering transform by

$$\overline{S}f := \{\overline{S}[p]f : m \geq 0, p = (j_1, \ldots, j_m) \in \mathcal{J}^m\},$$

where the non-windowed scattering coefficients are given by

$$\overline{S}[p]f := \left| \int_X (U[p]f)\bar{\varphi}_0 d\mu \right| = \left| \langle U[p]f, \varphi_0 \rangle_{\mathcal{H}} \right|.$$

In the case where $\mathcal{J} = \{0, \ldots, J\}$ and S is constructed from the diffusion wavelets $\mathcal{W}_J$ defined in (10), we will occasionally write $S_J[p]f$ in place of $S[p]f$ if we want to emphasize the dependency of the parameter J. We note that the primary difference between the windowed and non-windowed scattering transform is the use of the localized averaging operator A rather than a global integration against φ_0. Indeed, the term "windowed" refers to the idea that an average is computed within a neighborhood of each point. In particular, the windowed scattering transform should not be confused with constructions, such as those appearing in [24], which construct scattering transforms (on $\mathbb{R}^N$) using Gabor filters.

The following result relates the non-windowed scattering transform $\overline{S}$ to the limit of the windowed scattering transform S_J as $J \to \infty$. In particular, if $\mathcal{L}$ is either the Laplace-Beltrami operator on a manifold or the unnormalized Laplacian on a graph, then φ_0 is constant. Therefore, the following result shows that the windowed scattering coefficients $S_J[p]f(x)$ converge to a constant multiple of $\overline{S}[p]f$. Please see Appendix B for a proof.

Proposition 2. *Let S_J be the windowed scattering transform build on top of the diffusion wavelet frame $\mathcal{W}_J$ defined in (10) and assume $\lambda_1 > 0$. Then for all $f \in \mathcal{H}$, and every path p we have*

$$\lim_{J\to\infty} \| |S_J[p]f| - (\overline{S}[p]f)|\varphi_0| \|_{\mathcal{H}} = 0. \tag{14}$$

Remark 3. In the case where $\mathcal{L} = -\nabla \cdot \nabla$ is the Laplace-Beltrami operator on a manifold or the unnormalized graph Laplacian, one may take φ_0 to be the constant function $\varphi_0(x) = \mathrm{vol}(\mathcal{X})^{1/2}$ and it is known that the associated heat semigroup $\{e^{-t\mathcal{L}}\}_{t\geq 0}$ is positivity preserving (see, e.g., [25,44]). Therefore $S_J[p]f$ will be nonnegative, and (14) implies

$$\frac{1}{\mathrm{vol}(\mathcal{X})^{1/2}} \lim_{J\to\infty} S_J[p]f = (\overline{S}[p]f).$$

3. Examples and relationship to prior work

Several versions of the scattering transform for graphs [35,34,89] and smooth Riemannian manifolds without boundary [63,59] have been introduced in previous work. In this section, we will discuss how these constructions relate to our framework. We also discuss several other examples of measure spaces included in our theory that have not been previously considered in the scattering transform literature. Most of the techniques used to prove our theoretical results in Sections 2, 4, and 5 are natural generalization of the techniques used in these previous works on geometric (and even Euclidean) scattering. Indeed, our definitions were developed by carefully examining these papers and designing our framework in such a way that these techniques could be extended to more general settings. Additionally, we note that our convergence results (Theorems 10, 11, 12, 13, and 14 stated in Section 6), do not, to the best of our knowledge, have direct analogs in any previous work on the scattering transform.

3.1. Undirected, unsigned graphs

Several works have introduced different definitions of the *graph scattering transform*. These works differ primarily in two respects, i) the definition of the wavelets and ii) whether they use a windowed or unwindowed version of the graph scattering transform. Below, we explain how these constructions are related to our framework. Throughout this subsection, we let $G = (V, E, W)$ be a weighted graph with weighted adjacency matrix A and weighted degree matrix $D = \text{diag}(\mathbf{d})$, where $\mathbf{d}$ denotes the degree vector. Notably, all of the work discussed in this subsection focuses on undirected, unsigned graphs, i.e., graphs for which A is symmetric and has nonnegative entries.

In [34], the authors define wavelets of the form $T^{2^{j-1}} - T^{2^j}$, where $T := \frac{1}{2}(I + D^{-1/2}AD^{-1/2})$, for $1 \leq j \leq J$ for some maximal scale J.[1] In order to obtain these wavelets from our framework, we may choose μ to be the uniform measure which gives weight 1 to each vertex, let $\mathcal{L}$ be the symmetric normalized Laplacian $L_{\text{sym}} = I - D^{-1/2}AD^{-1/2}$ and choose

$$g(\lambda) = \begin{cases} 1 - \lambda/2 & \text{if } 0 \leq \lambda \leq 2 \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

in (8). The authors of [34] are primarily concerned with graph level tasks and therefore use a non-windowed version of the scattering transform.

In [35], the authors use wavelets of the form $P^{2^{j-1}} - P^{2^j}$ where

$$P := \frac{1}{2}(I + AD^{-1}) = D^{1/2}TD^{-1/2}$$

is the lazy random walk matrix, i.e., the matrix whose entries describe the transition probabilities of a lazy random walk on the graph. In order to incorporate these wavelets into our framework, we define μ by the rule $\mu(\{v_i\}) = \frac{1}{\mathbf{d}_i}$, where $v_i \in V$ and $\mathbf{d}_i$ is the i-th entry of $\mathbf{d}$, and choose $\mathcal{L}$ to be the random walk normalized Laplacian $L_{\text{RW}} = I - AD^{-1} = D^{1/2}L_{\text{sym}}D^{-1/2}$. Using the fact that L_{RW} is similar to L_{sym}, one may imitate the proof of Lemma 1.1 of [64] to verify that L_{RW} is self adjoint for this choice of μ. Therefore, one can recover the wavelets from [35] by again choosing g as in (15). Similar to [34], the authors of [35] are primarily concerned with graph level tasks and therefore also use a non-windowed version of the scattering transform.

The wavelets used in [34] and [35] are based on [21]. By contrast, [89] uses a different wavelet construction based on [39]. Rather than using a single spectral function g, a family of wavelets $\{\psi_j\}_{j\in\mathbb{Z}}$ is constructed on the real line and used to define wavelet convolution with respect to the spectral decomposition of the unnormalized Laplacian $L_{\text{un}} := D - A$. In our framework, this corresponds to defining

$$W_j f = \sum_{k=0}^{N-1} \psi_j(\lambda_k)\widehat{f}(k)\varphi_k,$$

where N is the number of vertices and $\{(\lambda_k, \varphi_k)\}_{k=0}^{N-1}$ are eigenpairs of L_{un}. We note that in Theorem 1 and in Section 5.2 we do not assume that our wavelets are constructed as in (10) and therefore some of our results may be applied to the scattering transform constructed from these wavelets as well. We also note that analogs of many of our other results were previously established in [89] in this case.

3.2. Signed graphs, directed graphs, hypergraphs, and simplicial complexes

A directed graph is a graph where the adjacency matrix is not symmetric. This makes it non-obvious how to apply spectral methods since naive extensions of the (unnormalized or normalized) graph Laplacian are in general not diagonalizable on the standard unweighted inner product space. Nevertheless, directed graphs are a natural model for many phenomena such as email networks or traffic networks, and so there have been several attempts to define directed graph Laplacians which are either real symmetric or complex Hermitian.

[1] [34] also uses a wavelet $I - T$ for when $j = 0$.

In [17], the author defines a directed Laplacian via a non-reversible Markov chain and provides an extensive analysis of these matrices' spectral properties. This matrix was later used as the basis for spectral directed graph neural networks in [56] and [78]. An alternative approach, dating back to at least [51], is to construct a complex Hermitian adjacency matrix known as the magnetic Laplacian, which may be viewed as a special case of the graph connection Laplacian (see, e.g., [72,73]). This matrix represents the undirected geometry of the graph in the magnitude of its entries and incorporates directional information in the phases. It has been studied by the graph signal processing community [33] and also applied to numerous data science applications such as clustering and community detection [30,19,28,29]. Recently, [87] showed that the Magnetic Laplacian could be effectively incorporated into a graph neural network. Analogously, there has also been work [23] using spectral clustering methods on signed graphs, i.e., graphs with both positive "friend" edges and negative "enemy" edges using methods based on signed Laplacians. Very recently, [32,40,74,47] proposes various signed magnetic Laplacian and uses these matrices to construct a signed and directed graph neural network. Similarly, [31] has proposed a neural network on hypergraphs (graphs where generalized edges may consist of more than two nodes) based on a generalized Laplacian.

In this paper, we are agnostic to the question of which Laplacian is the best for signed and/or directed graphs. We merely note that our theory applies to all of the Laplacians discussed here and any of these Laplacians can be used to define scattering transforms on signed and/or directed graphs. Additionally, we note that there has been work developing spectral clustering methods on hypergraphs using matrices which do not fit within our framework because they are not self-adjoint (see [16] and the references within). Developing variants of our theory which utilize these operators would be an interesting direction of future work.

We also note the recent work [66], which uses the Hodge Laplacian to construct wavelets on simplicial complices. These wavelets were then used as a basis for simplicial complex scattering transforms in the follow up work [67]. Furthermore, we also note several papers which have applied Hodge Laplacians to directed graphs [52,69]. We remark that in some of these cases, the condition that $0 = \lambda_0 < \lambda_1$ may not hold. In these settings, both U and the windowed scattering transform are still well-defined and most of our analysis carries through unchanged. The definition of the windowed scattering transform, however, should be modified to either be projection onto the 0-eigenspace, in the case where 0 has multiplicity, or to be defined via a global summation, i.e., $\overline{S}[p]f := \sum_{v \in V} U[p]f(v)$, in the case where $0 < \lambda_1$. However, it is important to note that in the latter case it is no longer true, in general, that the non-windowed scattering transform is the limit of the windowed scattering transform. (It follows from the proof of Proposition 2 that the windowed scattering transform converges to zero in this case.)

3.3. Manifolds

In [63], the authors constructed a scattering transform for smooth and compact manifolds without boundary via the spectral decomposition of the Laplace-Beltrami operator. If we choose $g(\lambda) = e^{-\lambda}$, the wavelets proposed in Section 2.1 are a minor variation of those considered there. Indeed, if we add an additional square root term as discussed in Remark 2, then the wavelets from Section 2.1 will exactly coincide with those considered in [63]. We also note [59] which replaced with wavelets used in [63] with wavelets optimized for fast computation on the sphere. As with the wavelets considered in [89], these wavelets are not a special case of the wavelets constructed in Section 2.1. However, it is likely that one could derive analogs of most of our results for this version of the scattering transform as well.

Our framework can also be applied to other interesting setups not considered in previous work on the scattering transform. For example, when the Laplace-Beltrami operator is equipped with suitable boundary conditions, our methods may also be applied to manifolds with boundary. Moreover, it may also be applied to weighted Laplacians such as those considered in [41] or [42] or anisotropic Laplacians such as those applied to two-dimensional surfaces in [6].

4. Continuity and invariance

In this section, we establish the fundamental continuity and invariance properties of the windowed and non-windowed scattering transform. In Section 4.1, we show that both the windowed and non-windowed scattering transforms are Lipschitz continuous with respect to additive noise, and then, in Section 4.2, we establish invariance and equivariance properties for the scattering transforms under certain group actions.

4.1. Lipschitz continuity with respect to additive noise

The following two theorems show that the windowed and non-windowed scattering transforms are Lipschitz continuous on $\mathcal{H}$. Our first result, Theorem 1, shows that the windowed scattering is nonexpansive. Its proof is based on analogous theorems in works such as [57], [63], [64], and [89] which consider specific measure spaces.

Theorem 1. *Let S be the scattering transform built on top of the wavelet frame $\mathcal{W}$. Then the scattering transform is a nonexpansive operator from $\mathcal{H} \to \ell^2(\mathcal{H})$, i.e., for all $f_1, f_2 \in \mathcal{H}$,*

$$\|Sf_1 - Sf_2\|_{\ell^2(\mathcal{H})} \leq \|f_1 - f_2\|_{\mathcal{H}}.$$

Theorem 2 shows that the non-windowed scattering transform is Lipschitz continuous on $\mathcal{H}$. Its proof is a generalization of Theorem 3.2 of [64]. Notably, unlike Theorem 1, Theorem 2 requires that we use the wavelets defined in (10).

Theorem 2. *Let S_J be the scattering transform built on top of diffusion wavelets $\mathcal{W}_J$ defined in (10). Assume that $\inf_x |\varphi_0(x)| > 0$ and $\lambda_1 > 0$. Then*

$$\|\overline{S}f_1 - \overline{S}f_2\|_2^2 \leq \frac{1}{\min_x |\varphi_0(x)|^2 vol(\mathcal{X})} \|f_1 - f_2\|_{\mathcal{H}}.$$

For proofs of Theorems 1 and 2, please see Appendix C.

Remark 4. In many cases of interest such as when either i) $\mathcal{X}$ is a compact Remannian manifold without boundary and $\mathcal{L}$ is the Laplace-Beltrami operator or ii) $\mathcal{X}$ is an unweighted and undirected graph and $\mathcal{L}$ is the unnormalized graph Laplacian, we have that $\varphi_0(x)$ is constant and therefore $\frac{1}{\min_x |\varphi_0(x)|^2 \text{vol}(\mathcal{X})} = 1$.

Remark 5. Inspecting the proof, we see that results analogous to Theorem 2 can be derived for the non-windowed scattering transform built upon other frames as long as one is able to establish a result similar to Proposition 2.

4.2. Invariance and equivariance

Let $\mathcal{G}$ be a collection of bijections $\zeta : X \to X$ which form a group under composition. For $\zeta \in \mathcal{G}$, let

$$\mathcal{X}_\zeta := (X, \mathcal{F}_\zeta, \mu_\zeta) \tag{16}$$

be the measure space with σ-algebra $\mathcal{F}_\zeta$ and measure μ_ζ given by

$$\mathcal{F}_\zeta := \{\zeta^{-1}(B) : B \in \mathcal{F}\}, \quad \mu_\zeta(B) := \mu(\zeta^{-1}(B)).$$

We let $\mathcal{G}$ act on $\mathcal{H}$ by function composition and we let it act on the set of linear operators by conjugation. Let $\mathcal{H}^{(\zeta)}$ be the Hilbert space of functions on $\mathcal{X}_\zeta$ which are square integrable with respect to μ_ζ. Let $V_\zeta : \mathcal{H} \to \mathcal{H}^{(\zeta)}$ denote the operator $V_\zeta f := f \circ \zeta^{-1}$ and let $\mathcal{L}_\zeta$ denote the operator on $\mathcal{H}^{(\zeta)}$ defined by

$$\mathcal{L}_\zeta := V_\zeta \circ \mathcal{L} \circ V_\zeta^{-1}.$$

The following lemma relates the eigenfunctions of $\mathcal{L}$ and $\mathcal{L}_\zeta$.

Lemma 1. *If φ is an eigenfunction of $\mathcal{L}$ with $\mathcal{L}\varphi = \lambda\varphi$, then $V_\zeta\varphi$ is an eigenfunction of $\mathcal{L}_\zeta$ and $\mathcal{L}_\zeta V_\zeta \varphi = \lambda V_\zeta \varphi$.*

Proof. The proof is immediate from the definition:

$$\mathcal{L}_\zeta V_\zeta \varphi = V_\zeta \mathcal{L} V_\zeta^{-1} V_\zeta \varphi = V_\zeta \mathcal{L}\varphi = V_\zeta \lambda \varphi = \lambda V_\zeta \varphi. \quad \square$$

In the case where $\mathcal{X}$ is a graph or a manifold, the standard choice of $\mathcal{G}$ is the permutation group or the isometry group. The key desired property of this group is that the associated group action is an isometry from $\mathcal{H}$ to $\mathcal{H}^{(\zeta)}$. This motivates the following definition.

Definition 1. We say that $\mathcal{G}$ preserves inner products on $\mathcal{H}$ if for all $\zeta \in \mathcal{G}$ and all $f_1, f_2 \in \mathcal{H}$ we have

$$\langle V_\zeta f_1, V_\zeta f_2 \rangle_{\mathcal{H}^{(\zeta)}} = \langle f_1, f_2 \rangle_{\mathcal{H}}.$$

Importantly, we note that Definition 1 is satisfied both when $\mathcal{X}$ is a compact Riemannian manifold, $\mathcal{G}$ is the isometry group, and μ is the Riemannian volume and also when $\mathcal{X}$ is a graph, $\mathcal{G}$ is the permutation group, and μ is *any* measure, including both the uniform measure and measures which assign different weights to vertices depending upon their degrees.

Lemma 2. *Suppose $\mathcal{G}$ preserves inner products on $\mathcal{H}$. Then, for all $\zeta \in \mathcal{G}$, $\mathcal{L}_\zeta$ is self-adjoint on $\mathcal{H}^{(\zeta)}$.*

Proof. Using the definition of $\mathcal{L}_\zeta$, the fact that $\mathcal{L}$ is self-adjoint on $\mathcal{H}$, and the assumption that $\mathcal{G}$ preserves inner products implies

$$\begin{aligned}
\langle \mathcal{L}_\zeta f_1, f_2 \rangle_{\mathcal{H}^{(\zeta)}} &= \langle V_\zeta \mathcal{L} V_\zeta^{-1} f_1, V_\zeta V_\zeta^{-1} f_2 \rangle_{\mathcal{H}^{(\zeta)}} = \langle \mathcal{L} V_\zeta^{-1} f_1, V_\zeta^{-1} f_2 \rangle_{\mathcal{H}} \\
&= \langle V_\zeta^{-1} f_1, \mathcal{L} V_\zeta^{-1} f_2 \rangle_{\mathcal{H}} = \langle V_\zeta^{-1} f_1, V_\zeta^{-1} V_\zeta \mathcal{L} V_\zeta^{-1} f_2 \rangle_{\mathcal{H}} \\
&= \langle V_\zeta^{-1} f_1, V_\zeta^{-1} \mathcal{L}_\zeta f_2 \rangle_{\mathcal{H}} = \langle f_1, \mathcal{L}_\zeta f_2 \rangle_{\mathcal{H}^{(\zeta)}}. \quad \square
\end{aligned}$$

Recall the operators $H^t : \mathcal{H} \to \mathcal{H}$

$$H^t f = \sum_{k \in \mathcal{I}} g(\lambda_k)^t \widehat{f}(k) \varphi_k = \sum_{k \in \mathcal{I}} g(\lambda_k)^t \langle f, \varphi_k \rangle_{\mathcal{H}} \varphi_k,$$

and define

$$H^t_\zeta f := \sum_{k \in \mathcal{I}} g(\lambda_k)^t \langle f, \varphi_k^{(\zeta)} \rangle_{\mathcal{H}^{(\zeta)}} \varphi_k^{(\zeta)}$$

to be the corresponding operator on $\mathcal{H}^{(\zeta)}$, where $\{\varphi_k^{(\zeta)}\}_{k \in \mathcal{I}}$ is an orthonormal basis of eigenfunctions for $\mathcal{L}_\zeta$. Let $W_j^{(\zeta)}$ and $A^{(\zeta)}$ denote wavelets and averaging operators on $\mathcal{H}^{(\zeta)}$, and let $U^{(\zeta)}$, $S^{(\zeta)}$, and $\overline{S^{(\zeta)}}$ be the analogs of U, S, and $\overline{S}$ on $\mathcal{H}^{(\zeta)}$. We observe that by Remark 1, the definition of H^t_ζ does not depend on the choice of eigenbasis. Therefore, by Lemma 1, we may assume without loss of generality that $\varphi_k^{(\zeta)} = V_\zeta \varphi_k$ and therefore that

$$H^t_\zeta f = \sum_{k \in \mathcal{I}} g(\lambda_k)^t \langle f, V_\zeta \varphi_k \rangle_{\mathcal{H}^{(\zeta)}} V_\zeta \varphi_k. \tag{17}$$

In light of (17), if $\mathcal{G}$ preserves inner products, we see that H^t commutes with V_ζ in the sense that

$$H^t_\zeta V_\zeta f = V_\zeta H^t f \quad \text{for all } t \geq 0 \tag{18}$$

since we may compute

$$\begin{aligned} H^t_\zeta V_\zeta f &= \sum_{k \in \mathcal{I}} g(\lambda_k)^t \langle V_\zeta f, V_\zeta \varphi_k \rangle_{\mathcal{H}^{(\zeta)}} V_\zeta \varphi_k = \sum_{k \in \mathcal{I}} g(\lambda_k)^t \langle f, \varphi_k \rangle_{\mathcal{H}} V_\zeta \varphi_k \\ &= V_\zeta \sum_{k \in \mathcal{I}} g(\lambda_k)^t \langle f, \varphi_k \rangle_{\mathcal{H}} \varphi_k = V_\zeta H^t f. \end{aligned}$$

This readily leads to the following theorem which shows that the condition that $\mathcal{G}$ preserves inner products on $\mathcal{H}$ is sufficient to produce equivariance results for the wavelet transform and the windowed scattering transform analogous to (2) mentioned in the introduction. For a proof, please see Appendix D.

Theorem 3. *Let $\mathcal{J} = \{0, \dots, J\}$, and let $\mathcal{W} = \mathcal{W}_J$ be the diffusion wavelets constructed in (10) and assume $\lambda_1 > 0$. Then, if $\mathcal{G}$ preserves inner products, then $\mathcal{G}$ commutes with both the wavelet transform, the operator U defined in (13) and the scattering transform. That is, for all $\zeta \in \mathcal{G}$, $f \in \mathcal{H}$ and $0 \leq j \leq J$, we have*

$$W_j^{(\zeta)} V_\zeta f = V_\zeta W_j f, \quad A^{(\zeta)} V_\zeta f = V_\zeta A f, \quad U^{(\zeta)} V_\zeta f = V_\zeta U f \quad \text{and} \quad S^{(\zeta)} V_\zeta f = V_\zeta S f.$$

Remark 6. Equations analogous to (18) hold for any spectral filter of the form (4), as long as $\widehat{h}(k)$ is a function of λ_k, i.e., $\widehat{h}(k) = \widetilde{h}(\lambda_k)$ for some function $\widetilde{h}$. Therefore, results similar to Theorem 3 can be derived for any network constructed from such filters and pointwise nonlinearities σ. Additionally, analogous results can also be derived for the scattering transform built upon other geometric wavelet constructions. For example, the conclusions of Proposition 4.1 of [89] are similar to those of Theorem 3 above.

Our next result shows that the non-windowed scattering transform $\overline{S}$ is fully invariant under the assumption that $\mathcal{G}$ preserves inner products on $\mathcal{H}$. Importantly, we note that the windowed scattering transform is not in general invariant. Intuitively, this distinction arises from the fact that $\overline{S}$ is the composition of an equivariant operator U together with a final global aggregation operator whereas S uses a localized averaging operator A.

Theorem 4. *Let $\mathcal{J} = \{0, \dots, J\}$, and let $\mathcal{W} = \mathcal{W}_J$ be the diffusion wavelets constructed in (10). Assume $\mathcal{L}$ has a spectral gap, i.e., $\lambda_1 > 0$. Then, if $\mathcal{G}$ preserves inner products, the non-windowed scattering transform is invariant to the action of $\mathcal{G}$, i.e.,*

$$\overline{S^{(\zeta)}} V_\zeta f = \overline{S} f \text{ for all } \zeta \in \mathcal{G} \text{ and all } f \in \mathcal{H}.$$

Proof. Since $\lambda_1 > 0$, the eigenspace corresponding to $\lambda = 0$ has dimension one. Therefore, $\varphi_0^{(\zeta)} = c V_\zeta \varphi_0$, for some constant c with $|c| = 1$, and so

$$\begin{aligned} \overline{S^{(\zeta)}}[p] V_\zeta f &= |\langle U^{(\zeta)}[p] V_\zeta f, c V_\zeta \varphi_0 \rangle_{\mathcal{H}^{(\zeta)}}| \\ &= |\langle V_\zeta U[p] f, V_\zeta \varphi_0 \rangle_{\mathcal{H}^{(\zeta)}}| = |\langle U[p] f, \varphi_0 \rangle_{\mathcal{H}}| = \overline{S}[p] f. \quad \square \end{aligned}$$

Unlike the non-windowed scattering transform, $\overline{S}$, the windowed scattering transform S_J is not in general permutation invariant, even in the limit as $J \to \infty$. If one wishes the windowed-scattering transform to be invariant to the action of $\mathcal{G}$, then one must also require that $\mathcal{G}$ preserves the measure μ as defined below.

Definition 2. We say that $\mathcal{G}$ preserves the measure μ if $\mathcal{F}_\zeta = \mathcal{F}$ and $\mu_\zeta(B) = \mu(B)$ for all $\zeta \in \mathcal{G}$ and all $B \in \mathcal{F}$.

To better understand this definition, we note that if $\mathcal{X}$ is a Riemannian manifold, then the isometry group preserves μ when μ is the Riemannian measure, but not if a general μ is chosen. Similarly, if $\mathcal{X}$ is a graph, the permutation group will preserve μ if it gives equal weight to each vertex, but not if, for example, μ gives different weights to vertices depending on their degrees.

Under the assumption that $\mathcal{G}$ preserves the measure μ, we are able to show that the windowed scattering transform is invariant to the action of $\mathcal{G}$ in the limit as $J \to \infty$ at an exponential rate.

Theorem 5. *Let $\mathcal{J} = \{0, \dots, J\}$, and let $\mathcal{W} = \mathcal{W}_J$ be the diffusion wavelets constructed in (10). Suppose that $\varphi_0(x)$ is constant and assume $\mathcal{G}$ preserves both measures and inner products. Then for all $\zeta \in \mathcal{G}$, we have*

$$\|S_J f - S_J^{(\zeta)} V_\zeta f\|_{\ell^2(\mathcal{H})} \leq 2|g(\lambda_1)|^{2^J} \|Uf\|_{\ell^2(\mathcal{H})}.$$

The proof of Theorem 5 is based on Lemma 3 as well as the observation that $\lim_{J\to\infty} \|A_J V_\zeta - A_J\|_{\mathcal{H}} = 0$. We note that while Theorem 5 assumes that the $\mathcal{W} = \mathcal{W}_J$ are the diffusion wavelets constructed in (10), Lemma 3 does not. (Note that other geometric wavelet constructions such as the one utilized in [89] also lead to versions of the scattering transform where the (19) condition is satisfied.) For a proof of both Theorem 5 and Lemma 3, please see Appendix E.

Lemma 3. *Assume $\mathcal{G}$ preserves both measures and inner products and that S is equivariant with respect to the action of $\mathcal{G}$ in the sense that*

$$S^{(\zeta)} V_\zeta f = V_\zeta S f. \tag{19}$$

Then for all $\zeta \in \mathcal{G}$, we have

$$\|Sf - S^{(\zeta)} V_\zeta f\|_{\ell^2(\mathcal{H})} \leq \|V_\zeta A - A\|_{\mathcal{H}} \|Uf\|_{\ell^2(\mathcal{H})}.$$

Remark 7. One limitation of Theorem 5 is that the right-hand side is given in terms of $\|Uf\|_{\ell^2(\mathcal{H})}$ instead of $\|f\|_{\mathcal{H}}$. This is a common issue with many asymptotic invariance results for the windowed scattering transform. However, as first noted in [57], one may use (11) and the fact that σ is nonexpansive to see

$$\sum_{p\in\mathcal{J}^m} \|U[p]f\|_{\mathcal{H}}^2 \leq \sum_{p\in\mathcal{J}^{m-1}} \|U[p]f\|_{\mathcal{H}}^2$$

for any $m \geq 1$. Therefore,

$$\sum_{p\in\mathcal{J}^m} \|U[p]f\|_{\mathcal{H}}^2 \leq \sum_{p\in\mathcal{J}^{m-1}} \|U[p]f\|_{\mathcal{H}}^2 \leq \dots \leq \sum_{p\in\mathcal{J}} \|U[p]f\|_{\mathcal{H}}^2 \leq \|f\|_{\mathcal{H}}^2 \tag{20}$$

and so, if one only uses M scattering layers, the total energy of Uf may be bounded by

$$\sum_{m\leq M} \left(\sum_{p\in\mathcal{J}^m} \|U[p]f\|_{\mathcal{H}}^2 \right) \leq (M+1)\|f\|_{\mathcal{H}}^2.$$

Therefore, if one only uses finitely many scattering layers, the right-hand side of Theorem 5 may be controlled in terms of $\|f\|_{\mathcal{H}}$. Moreover, in the case where $\mathcal{X}$ is a graph, for certain classes of wavelets we have

$$\sum_{p\in\mathcal{J}^m} \|U[p]f\|_{\mathcal{H}}^2 \leq r \sum_{p\in\mathcal{J}^{m-1}} \|U[p]f\|_{\mathcal{H}}^2 \tag{21}$$

for some $r < 1$ (see, e.g., Proposition 3.3 of [89] or Theorem 3.4 of [64]). Therefore, in this case, one has

$$\|Uf\|_{\ell^2(\mathcal{H})}^2 = \sum_{m=0}^{\infty} \left(\sum_{p\in\mathcal{J}^m} \|U[p]f\|_{\mathcal{H}}^2 \right) \leq \sum_{m=0}^{\infty} r^m \|f\|_{\mathcal{H}}^2 = \frac{1}{1-r}\|f\|_{\mathcal{H}}^2$$

independent of the number of layers used.

The main results of this section, Theorems 3, 4, and 5, can be summarized as follows: If $\mathcal{G}$ preserves inner products, and the scattering transform is constructed using the diffusion wavelets defined in Section 2.1, then the windowed scattering transform is equivariant and the non-windowed scattering transform is invariant to the action of $\mathcal{G}$. If we further assume that $\mathcal{G}$ preserves measure and that φ_0 is constant, then we also have that the windowed scattering transform is invariant in the limit as $J \to \infty$.

As alluded to in the introduction, these invariance and equivariance results show that the scattering transform respects the intrinsic structure of the data and therefore is well-suited for a variety of machine learning tasks. In particular, the equivariance result, Theorem 3, shows that it is well-suited for point-level tasks such as the node classification task which we will consider in

Section 7.3. Similarly, the invariance results Theorems 4 and 5, show that it is well-equipped to handle shape-level tasks such as the manifold classification tasks considered in Sections 7.1 and 7.2.

We also note that the assumption that $\mathcal{G}$ preserves inner products is quite natural. It is satisfied both when μ is the Riemannian measure on a manifold and $\mathcal{G}$ is the isometry group and when $\mathcal{X}$ is a (possibly signed, possibly directed) graph, $\mathcal{G}$ is the permutation group, and μ is any measure. The conditions that $\mathcal{G}$ preserves volumes and φ_0 is constant are a bit stronger. For example, when $\mathcal{X}$ is a graph, permutations do not preserve measure if the μ gives different vertices different weights. Moreover, if we take $\mathcal{L}$ to be the symmetric normalized graph Laplacian (on an undirected, unsigned graph), then φ_0 is given by $\varphi_0(x) \sim \text{degree}(x)^{1/2}$ and therefore is not constant unless the graph is regular.

5. Stability

In this section, we show that the measure space scattering transform is robust to small perturbations to the measure μ and the diffusion operator H. In particular, we consider a measure space $\mathcal{X} = (X, \mathcal{F}, \mu)$ and another measure space $\mathcal{X}' = (X', \mathcal{F}', \mu')$ which we interpret as a perturbed version of $\mathcal{X}$. We assume that these two spaces have the same underlying sets and σ-algebras and that measures are mutually absolutely continuous with bounded Radon-Nikodyn derivatives, i.e., we have $X = X', \mathcal{F} = \mathcal{F}'$ and that there exist Radon-Nikodyn derivatives such that

$$d\mu = \frac{d\mu}{d\mu'} d\mu' \quad \text{and} \quad d\mu' = \frac{d\mu'}{d\mu} d\mu.$$

To quantify the distortion between measures, we let $\mathcal{H} = \mathbf{L}^2(\mathcal{X})$ and $\mathcal{H}' = \mathbf{L}^2(\mathcal{X}')$, and we introduce two quantities, $R = R(\mathcal{H}, \mathcal{H}')$ and $\kappa = \kappa(\mathcal{H}, \mathcal{H}')$, defined by

$$R := R(\mathcal{H}, \mathcal{H}') := \max \left\{ \left\| \frac{d\mu'}{d\mu} \right\|_\infty, \left\| \frac{d\mu}{d\mu'} \right\|_\infty \right\} \tag{22}$$

and

$$\kappa(\mathcal{H}, \mathcal{H}') = \max \left\{ \left\| 1 - \frac{d\mu}{d\mu'} \right\|_\infty, \left\| 1 - \frac{d\mu'}{d\mu} \right\|_\infty \right\}.$$

We note that these two quantities are closely related to their analogs in [64] which focused on the special case where $\mathcal{X}$ was an undirected, unsigned graph. In the case where $\mu = \mu'$, we have $R(\mathcal{H}, \mathcal{H}') = 1$ and $\kappa(\mathcal{H}, \mathcal{H}') = 0$. Therefore, we will consider μ and μ' to be close to one another if $R \approx 1$ and $\kappa \approx 0$.

To further understand these definitions, consider the case where $\mathcal{X}$ and $\mathcal{X}'$ are two (possibly signed, possibly directed) graphs with N vertices and identify both vertex sets with $\{0, 1, \dots, N-1\}$. If μ and μ' are both the uniform measure, then we automatically have $R(\mathcal{H}, \mathcal{H}') = 1$ and $\kappa(\mathcal{H}, \mathcal{H}') = 0$. In this case, bounds produced in Theorem 6 will simplify considerably as discussed below. Another natural choice of measure in the graph setting is to let $\mu(i) = \mathbf{d}_i^{-1} = \text{degree}(i)^{-1}$ since this is the measure needed in order to make the random-walk Laplacian $I - AD^{-1}$ self-adjoint. In this case, we have $R(\mathcal{H}, \mathcal{H}') = \max_{0 \le i \le N-1} \max \left\{ \frac{\mathbf{d}_i}{\mathbf{d}_i'}, \frac{\mathbf{d}_i'}{\mathbf{d}_i} \right\}$. In particular, if both $\mathbf{d}$ and $\mathbf{d}'$ satisfy the entrywise bound $0 < m \le \mathbf{d}_i, \mathbf{d}_i' \le M < \infty$, we have $R(\mathcal{H}, \mathcal{H}') \le \frac{M}{m}$.

Observe that the assumption $R(\mathcal{H}, \mathcal{H}') < \infty$ implies that the sets with measure zero with respect to μ are the same as those with measure zero with respect to μ'. Therefore, each function $f \in \mathcal{H}$ can be uniquely identified with an element of $\mathcal{H}' = \mathbf{L}^2(\mathcal{X}')$ (and vice-versa) and so we may regard the Hilbert spaces $\mathcal{H}$ and $\mathcal{H}'$ as having the same elements. Therefore, if $f \in \mathcal{H}$ and $\widetilde{f} \in \mathcal{H}'$, the subtraction $f - \widetilde{f}$ is well defined. We also note that

$$\|f\|_{\mathcal{H}}^2 = \int_X |f|^2 d\mu = \int_X |f|^2 \frac{d\mu}{d\mu'} d\mu' \le R(\mathcal{H}, \mathcal{H}') \|f\|_{\mathcal{H}'}^2, \tag{23}$$

and similarly,

$$\|f\|_{\mathcal{H}'}^2 \le R(\mathcal{H}, \mathcal{H}') \|f\|_{\mathcal{H}}^2. \tag{24}$$

We also observe that

$$|\langle f, g \rangle_{\mathcal{H}} - \langle f, g \rangle_{\mathcal{H}'}| = \left| \int_X f \bar{g} \left(1 - \frac{d\mu'}{d\mu} \right) d\mu \right| \le \kappa(\mathcal{H}, \mathcal{H}') \|f\|_{\mathcal{H}} \|g\|_{\mathcal{H}}. \tag{25}$$

Let $\mathcal{L}$ and $\mathcal{L}'$ be self-adjoint positive semidefinite operators on $\mathcal{H}$ and $\mathcal{H}'$ respectively, and let $\{\varphi_k\}_{k=0}^\infty$, $\{\varphi_k'\}_{k=0}^\infty$ be the associated eigenbases. Let g be a spectral function satisfying the same assumptions as described in Section 2.1 and let H and H' be the associated operators defined as in (8). Importantly, we note that we use the same function g when constructing both H and H', so we may interpret H and H' as being analogous operators on different spaces. For example, in the case where $\mathcal{X}$ and $\mathcal{X}'$ are manifolds and $g(\lambda) = e^{-\lambda}$, one may check that $\{H^t\}_{t \ge 0}$ is the heat semigroup on $\mathcal{X}$ and $\{(H^t)'\}_{t \ge 0}$ is the heat semigroup on $\mathcal{X}'$.

Below, we prove a stability result for the wavelet transform. Our result will give bounds in terms of $R(\mathcal{H}, \mathcal{H}')$ and $\kappa(\mathcal{H}, \mathcal{H}')$, which measure how much μ differs from μ'. However, these terms are not by themselves necessarily sufficient to characterize how

different $\mathcal{X}$ is from $\mathcal{X}'$. For example, consider the case where $\mathcal{X}$ is a complete graph with N vertices, $\mathcal{X}'$ is a cycle graph of N vertices, and $\mathcal{L}$ and $\mathcal{L}'$ are the unnormalized graph Laplacians on $\mathcal{X}$ and $\mathcal{X}'$. In both of these cases, the natural choice of measure is to assign equal mass to each vertex, and so we will have $\mu(\{x\}) = \mu'(\{x\})$ for every vertex $x \in X = X'$. It follows, that $\frac{d\mu'}{d\mu} = 1$ uniformly, and therefore, we have $R(\mathcal{H}, \mathcal{H}') = 1$ and $\kappa(\mathcal{H}, \mathcal{H}') = 0$. However, a complete graph and a cycle graph are clearly very far from being isomorphic as graphs in any reasonable sense. In particular, one way in which these graphs differ is that heat will diffuse much more rapidly through a fully connected graph than through a directed cycle. This motivates us to follow the lead of [34] (see also [20] and [61]) and consider the term

$$\|H - H'\|_{\mathcal{H}}. \tag{26}$$

We note that since we assume that $R(\mathcal{H}, \mathcal{H}')$ is finite, the operator H' is well-defined on $\mathcal{H}$. We also note that unlike [34], (26) does not take the infimum over the orbits of $\mathcal{G}$. This is because the wavelet transform is not invariant to the action of $\mathcal{G}$, but is merely equivariant. Therefore, no infimum will appear in Theorem 6 stated below which establishes the stability of the wavelet transform. The scattering transform, by contrast, is invariant to the action of $\mathcal{G}$ and therefore such infimums will emerge in Theorems 7 and 8 which establish stability for the windowed and non-windowed scattering transforms.

5.1. Stability of the wavelet transform

We will decompose H and H' by

$$H = \widetilde{H} + \overline{H}, \quad H' = \widetilde{H}' + \overline{H}' \tag{27}$$

where

$$\widetilde{H} f = \widehat{f}(0)\varphi_0, \quad \text{and} \quad \overline{H} f = \sum_{k \geq 1} g(\lambda_k)\widehat{f}(k)\varphi_k,$$

and $\widetilde{H}'$ and $\overline{H}'$ are defined similarly.

$$\|\overline{H} f\|_{\mathcal{H}}^2 = \|\sum_{k \geq 1} g(\lambda_k)\widehat{f}(k)\varphi_k\|_{\mathcal{H}}^2 \leq g(\lambda_1)^2 \|f\|_{\mathcal{H}}^2 \tag{28}$$

and similarly,

$$\|\widetilde{H} f\|_{\mathcal{H}}'^2 \leq g(\lambda_1')^2 \|f\|_{\mathcal{H}'}^2. \tag{29}$$

Moreover, combining (29) with (23) and (24) implies

$$\|\widetilde{H}' f\|_{\mathcal{H}}^2 \leq R(\mathcal{H}, \mathcal{H}') g(\lambda_1')^2 \|f\|_{\mathcal{H}'}^2 \leq R(\mathcal{H}, \mathcal{H}')^2 g(\lambda_1')^2 \|f\|_{\mathcal{H}}^2.$$

Therefore,

$$\beta \leq \max\{g(\lambda_1), g(\lambda_1') R(\mathcal{H}, \mathcal{H}')\}. \tag{30}$$

In light of (30), in order for the requirement that $\beta < 1$ to hold it suffices for μ and μ' to be well-aligned enough so that $R(\mathcal{H}, \mathcal{H}') < g(\lambda_1')^{-1}$. Therefore, Theorem 6 stated below can be interpreted as a local stability result where the radius of convergence depends on the spectral gap λ_1'.

Theorem 6. *Let $\mathcal{W}_J$ be the diffusion wavelets on $\mathcal{X}$ defined as in (10), and let $\mathcal{W}_J'$ be the analogous wavelets on $\mathcal{X}'$. Let $\beta = \max\{\|\overline{H}\|_{\mathcal{H}}, \|\overline{H}'\|_{\mathcal{H}}\}$ and assume that $\beta < 1$. Then,*

$$\begin{aligned} &\|\mathcal{W}_J - \mathcal{W}_J'\|_{\ell^2(\mathcal{H})}^2 \\ \leq &C(\beta)\left[\|\varphi_0 - \varphi_0'\|_{\mathcal{H}}^2 R(\mathcal{H}, \mathcal{H}') + R(\mathcal{H}, \mathcal{H}')^2 \kappa(\mathcal{H}, \mathcal{H}')^2 + \|H - H'\|_{\mathcal{H}}^2\right] \end{aligned}$$

where $C(\beta) = C\frac{\beta^2+1}{(1-\beta^2)^3}$ for some absolute constant $C > 0$.

For a Proof of Theorem 6, please see Appendix F. As noted above, in the case where $\mathcal{X}$ is a graph and μ is the uniform measure, we have $R(\mathcal{H}, \mathcal{H}') = 1$ and $\kappa(\mathcal{H}, \mathcal{H}') = 0$. Therefore, the result of Theorem 6 simplifies to

$$\|\mathcal{W}_J - \mathcal{W}_J'\|_{\ell^2(\mathcal{H})}^2 \leq C(\beta)\left[\|\varphi_0 - \varphi_0'\|_{\mathcal{H}}^2 + \|H - H'\|_{\mathcal{H}}^2\right].$$

Furthermore, if $\mathcal{L}$ is the unnormalized graph Laplacian, we have $\varphi_0 = \varphi_0'$, and the result further simplifies to $\|\mathcal{W}_J - \mathcal{W}_J'\|_{\ell^2(\mathcal{H})}^2 \leq C(\beta)\|H - H'\|_{\mathcal{H}}^2$.

5.2. Stability of the scattering transform

In this section, we prove the stability of the windowed and non-windowed scattering transforms. As in Section 4.1, and following the lead of [64], in this section, we will not assume that the scattering transform is constructed using the diffusion wavelets constructed in Section 2.1. Instead, as in Section 2.2, we will let $\mathcal{J}$ be an arbitrary countable indexing set and assume that

$$\mathcal{W} = \{W_j, A\}_{j\in\mathcal{J}} \quad \text{and} \quad \mathcal{W}' = \{W_j', A'\}_{j\in\mathcal{J}}$$

are any frames on $\mathcal{H}$ and $\mathcal{H}'$ such that (11) holds. We do this because, for any given measure space, there may be many possible ways to construct wavelets, or more generally frames satisfying (11) and in the Euclidean setting there have been various works defining the scattering transform using more general non-wavelet frames [24,36,84,85]. Therefore, we will show that the stability of the underlying frame directly implies the stability of the resulting scattering transforms. Throughout this section, we will let S^ℓ denote the set of all ℓ-th order scattering coefficients, on $\mathcal{X}$, i.e.,

$$S^\ell f := \{S[p]f \,:\, p = (j_1, \dots, j_\ell)\},$$

and let $(S^\ell)'$ denote the corresponding set of scattering coefficients on $\mathcal{X}'$. We will also continue to assume that the sets X and X' and the σ-algebras $\mathcal{F}$ and $\mathcal{F}'$ are the same and also that $R(\mathcal{H},\mathcal{H}')$ and $\kappa(\mathcal{H},\mathcal{H}')$ are finite. We recall that, as noted prior to (23), this means that $\mathcal{H}$ and $\mathcal{H}'$ can be regarded as having the same elements and so the subtraction of elements $\mathcal{H}'$ from elements of $\mathcal{H}$ is well defined.

Theorem 7 *(Stability for the windowed scattering transform). Let $\mathcal{X} = (X,\mathcal{F},\mu)$ and $\mathcal{X}' = (X',\mathcal{F}',\mu')$ be measure spaces with $X = X'$ and $\mathcal{F} = \mathcal{F}'$. Let $\mathcal{H} = \mathbf{L}^2(\mathcal{X}), \mathcal{H}' = \mathbf{L}^2(\mathcal{X}')$ and let $\mathcal{J}$ be a countable indexing set. Let $\mathcal{W} = \{W_j, A\}_{j\in\mathcal{J}}$ and $\mathcal{W}' = \{W_j', A'\}_{j\in\mathcal{J}}$ be frames on $\mathcal{H}$ and $\mathcal{H}'$ such that (11) holds. Let S^ℓ and $(S^\ell)'$ be the ℓ-th layers of the windowed scattering transforms on $\mathcal{X}$ and $\mathcal{X}'$ constructed from $\mathcal{W}$ and $\mathcal{W}'$. Further assume that S^ℓ is equivariant to the action of $\mathcal{G}$ and also invariant up to a factor of $\mathcal{B}$ in the sense that*

$$V_\zeta S^\ell f = S^{\ell,(\zeta)} V_\zeta f, \quad \text{and} \quad \left\| V_\zeta S^\ell f - S^\ell f \right\|_{\ell^2(\mathcal{H})} \leq \mathcal{B}\|f\|_{\mathcal{H}} \tag{31}$$

for all $f \in \mathcal{H}$ and $\zeta \in \mathcal{G}$. Then for all $f \in \mathcal{H}$ and $\widetilde{f} \in \mathcal{H}'$, we have

$$\begin{aligned} &\left\| S^\ell f - (S^\ell)' \widetilde{f} \right\|_{\ell^2(\mathcal{H})} \\ &\leq \inf_{\zeta\in\mathcal{G}} \left[\mathcal{B}\|f\|_{\mathcal{H}} + R\left(\mathcal{H},\mathcal{H}^{(\zeta)}\right) \|V_\zeta f - \widetilde{f}\|_{\mathcal{H}} \right. \\ &\qquad \left. + \left(\sqrt{2} R\left(\mathcal{H},\mathcal{H}^{(\zeta)}\right) \|\mathcal{W}^{(\zeta)} - \mathcal{W}'\|_{\mathcal{H}^{(\zeta)}} \left(\sum_{k=0}^{\ell} \|\mathcal{W}'\|_{\mathcal{H}^{(\zeta)}}^k \right) \right) \cdot \|\widetilde{f}\|_{\mathcal{H}} \right]. \end{aligned} \tag{32}$$

For a proof of Theorem 7, please see Appendix G. We note that if $\mathcal{W}$ are the diffusion wavelets constructed in Section 2.1, $\mathcal{G}$ preserves measures, and φ_0 is constant, then Theorem 5 and Remark 7 imply condition (31) holds with $\mathcal{B} = \sqrt{(\ell+1)|g(\lambda_1)|^{2^J}}$ (which converges to zero as $J \to \infty$). In particular, these conditions are satisfied both when $\mathcal{X}$ is a Riemannian manifold, $\mathcal{L}$ is the Laplace-Beltrami operator, and μ is the Riemannian volume form and when $\mathcal{X}$ is a graph, μ is the uniform measure, and $\mathcal{L}$ is the unnormalized graph Laplacian.

We also note that we can interpret each of the terms on the right-hand side of (32). We are looking for a bijection $\zeta \in \mathcal{G}$ which will simultaneously align both the wavelets $\mathcal{W}$ (which are typically constructed from the operators $\mathcal{L}$ of $\mathcal{X}$), the Hilbert spaces $\mathcal{H}$, and the signal f. Therefore, the term $R(\mathcal{H},\mathcal{H}^{(\zeta)})$ measures how well aligned the Hilbert spaces are, the term $\|\mathcal{W}^{(\zeta)} - \mathcal{W}'\|_{\mathcal{H}^{(\zeta)}}$ measures how well aligned the wavelets are, and the term $\|V_\zeta f - \widetilde{f}\|_{\mathcal{H}}$ measures how well aligned the signals are. We also note that in the case where $\mathcal{W}$ are the diffusion wavelets constructed in Section 2.1, we can control the term $\|\mathcal{W}^{(\zeta)} - \mathcal{W}'\|_{\mathcal{H}^{(\zeta)}}$ by applying Theorem 6.

The next result is the analogue of Theorem 7 for the non-windowed scattering transform. We note that the terms on the right-hand side of (34) have similar interpretations as those in Theorem 7. Additionally, by Theorems 2 and 4, we note that the condition (33) is satisfied whenever $\inf_x |\varphi_0(x)| > 0$, $\lambda_1 > 0$, $\mathcal{G}$ preserves inner products and $\mathcal{W} = \mathcal{W}_J$ are the diffusion wavelets constructed in (10).

Theorem 8 *(Stability for the non-windowed scattering transform). Let $\mathcal{X} = (X,\mathcal{F},\mu)$ and $\mathcal{X}' = (X',\mathcal{F}',\mu')$ be measure spaces with $X = X'$ and $\mathcal{F} = \mathcal{F}'$. Let $\mathcal{H} = \mathbf{L}^2(\mathcal{X}), \mathcal{H}' = \mathbf{L}^2(\mathcal{X}')$ and let $\mathcal{J}$ be a countable indexing set. Let $\mathcal{W} = \{W_j, A\}_{j\in\mathcal{J}}$ and $\mathcal{W}' = \{W_j', A'\}_{j\in\mathcal{J}}$ be frames on $\mathcal{H}$ and $\mathcal{H}'$ such that (11) holds. Let $\overline{S^\ell}$ and $(\overline{S^\ell})'$ be the ℓ-th layers of the non-windowed scattering transforms on $\mathcal{X}$ and $\mathcal{X}'$ constructed from $\mathcal{W}$ and $\mathcal{W}'$. Assume that $\overline{S}$ is fully invariant to the action of $\mathcal{G}$ and also Lipschitz continuous on $\mathcal{H}$ with constant C_L in the sense that*

$$\|\overline{S} f_1 - \overline{S} f_2\|_2^2 \leq C_L \|f_1 - f_2\|_{\mathcal{H}} \quad \text{and} \quad \overline{S^{(\zeta)}} V_\zeta f_1 = \overline{S} f_1 \tag{33}$$

Then for all $f \in \mathcal{H}$ and $\widetilde{f} \in \mathcal{H}'$, we have

$$\left\| \overline{S^\ell} f - (\overline{S^\ell})' \widetilde{f} \right\|_2^2 \tag{34}$$

$$\leq 3 \inf_{\zeta \in \mathcal{G}} \Bigg[2C_L \|V_\zeta f - \tilde{f}\|^2_{\mathcal{H}^{(\zeta)}} + R(\mathcal{H}^{(\zeta)}, \mathcal{H}')^2 \|\varphi_0^{(\zeta)} - \varphi_0'\|^2_{\mathcal{H}'} \|\widetilde{f}\|^2_{\mathcal{H}'}$$

$$+ 2\|\mathcal{W}^{(\zeta)} - \mathcal{W}'\|^2_{\mathcal{H}^{(\zeta)}} \left(\sum_{k=0}^{\ell-1} \|\mathcal{W}'\|^k_{\mathcal{H}^{(\zeta)}} \right)^2 \|\tilde{f}\|^2_{\mathcal{H}^{(\zeta)}} + \kappa(\mathcal{H}', \mathcal{H}^{(\zeta)}) \|\widetilde{f}\|_{\mathcal{H}'} \cdot \Bigg]$$

For a proof of Theorem 8, please see Appendix H.

6. Implementing the manifold scattering transform from point-cloud data

In [63], the authors showed that the manifold scattering transform was effective for classification tasks on known two-dimensional surfaces with predefined meshes. However, in many applications of interest, one is not given a predefined manifold. Instead, one is given a collection of points $\{x_i\}_{0=1}^{N-1}$ embedded in some high-dimensional Euclidean space $\mathbb{R}^D$ and one makes a modeling assumption that these points lie on (or near) a comparatively low-dimensional manifold. Thus, in this section, we will assume that $\mathcal{X}$ is a smooth d-dimensional Riemannian manifold without boundary which is embedded in $\mathbb{R}^D$ for some $D \gg d$ and that $\{x_i\}_{i=0}^{N-1}$ is a discrete subset randomly and independently sampled from $\mathcal{X}$. We will use the x_i to construct a weighted graph $\mathcal{X}_N$ and present two methods which use $\mathcal{X}_N$ to implement an approximation of the manifold scattering transform when one only has access to these sample points.

Both of these methods rely on an affinity kernel $K_\epsilon(\cdot,\cdot)$ to construct a data-driven graph $\mathcal{X}_N$ with weighted adjacency matrix $W^{(N)}$. In our first method, we simply define an approximate heat semigroup at time $t = 1$ by $H^1_{N,\epsilon} := (D^{(N)})^{-1} W^{(N)}$, where $D^{(N)} = W^{(N)}\mathbf{1}$ is the degree matrix associated to $W^{(N)}$. One may then approximate H^{2^j} by, e.g., matrix multiplication. We note that while in principle, $H^1_{N,\epsilon}$ is a dense matrix, most of its entries will typically be small and therefore one may apply a threshold operator and use sparse matrix multiplications to implement an approximation of the wavelet transform. (Notably, if one imitates the method used in [77], there is no need to ever form a dense matrix after the initial thresholding.) In our second method, we use $W^{(N)}$ to construct a data-driven graph Laplacian $L_{N,\epsilon}$. We then define a discrete approximation of the heat semigroup using the eigenvectors and eigenvalues of $L_{N,\epsilon}$.

In either case, once we have our approximations of H^t, it is then straightforward to implement the wavelet transform and therefore the scattering transform. The advantage of the second, eigenvector-based method is that we will be able to use results from [27,15] to prove a quantitative rate of convergence for the scattering transform. The first method, on the other hand, is more computationally efficient for large N (if one uses a thresholding operator to promote sparsity as discussed above) since it does not require one to compute an eigendecomposition. We are not able to prove a convergence rate for the scattering transform computed using this method, but we note that the approximation $H^1 \approx H^1_{N,\epsilon} = (D^{(N)})^{-1} W^{(N)}$ was shown to converge pointwise [20], albeit without a rate.

In order to avoid confusion, we will typically denote objects corresponding to $\mathcal{X}_N$ with a subscript or superscript N and objects corresponding to $\mathcal{X}$ without such subscript or subscript. For example, we will let W_j denote a wavelet on $\mathcal{X}$ at scale 2^j and $W_{j,N}$ denote the corresponding wavelet on $\mathcal{X}_N$. Throughout the section, we will choose $\mathcal{L} = -\nabla \cdot \nabla$ to be the Laplace-Beltrami operator on $\mathcal{X}$, where ∇ is the intrinsic gradient. We will also choose $g(\lambda) = e^{-\lambda}$, in which case $\{H^t\}_{t \geq 0}$ is the heat semigroup (see Equation (9)). We will let $h_t(x,y)$ denote the heat kernel so that $H_t f(x) = \int_{\mathcal{X}} h_t(x,y) f(y) d\mu(y)$, where $d\mu$ is the Riemannian measure, normalized so that

$$\mu(\mathcal{X}) = 1. \tag{35}$$

It is well known that

$$\int_{\mathcal{X}} h_t(x,y) d\mu(y) = 1 \tag{36}$$

for all $x \in \mathcal{X}$ and all $t > 0$, and

$$h_t(x,y) = \sum_{k=0}^{\infty} e^{-t\mu_k} \varphi_k(x) \varphi_k(y), \tag{37}$$

where in (37), and throughout this section, we will use μ_k to denote eigenvalues of the Laplace-Beltrami operator $\mathcal{L} = -\nabla \cdot \nabla$ and will reserve λ_k (sometimes with additional superscripts) for eigenvalues of the data-driven graph Laplacian which we will define below.

We now construct a weighted graph. We let $K(\cdot,\cdot)$ be an affinity kernel such as

$$K(x,x') := K_\epsilon(x,x') := \epsilon^{-d/2} \exp\left(-\frac{\|x - x'\|_2^2}{\epsilon} \right), \quad \epsilon > 0 \tag{38}$$

where in the above equation $\|x - x'\|_2$ refers to the Euclidean distance between two points in $\mathbb{R}^D$ and ϵ is a bandwidth parameter.[2] Given this kernel, we define an affinity matrix $W^{(N)}$ and a diagonal degree matrix $D^{(N)}$ by

$$W^{(N)}_{i,j} := K(x_i, x_j) \text{ and } D^{(N)}_{i,i} := \sum_{j=0}^{N-1} W^{(N)}_{i,j}.$$

Given $W^{(N)}$ and $D^{(N)}$, one may then approximate H^1 by

$$H^1_{N,\epsilon} := (D^{(N)})^{-1} W^{(N)}. \tag{39}$$

While the primary motivation of this method is to avoid computing eigenvectors and eigenvalues, we do note that (39) may also be equivalently obtained from (8) by choosing $\mathcal{L}$ to be the Markov normalized Graph Laplacian $I^{(N)} - (D^{(N)})^{-1} W^{(N)}$ on $\mathcal{X}_N$ and choosing $g(\lambda) = 1 - \lambda$.

Our second method constructs approximations of H^t based on (37). In our implementation, we may only use finitely many eigenvalues. This motivates us to define the truncated heat semigroup by

$$H^\kappa_t f(x) := \int_{\mathcal{X}} h^\kappa_t(x, y) f(y) d\mu(y), \text{ where } h^\kappa_t(x, y) := \sum_{k=0}^{\kappa} e^{-t\mu_k} \varphi_k(x) \varphi_k(y),$$

where κ is chosen by the user. Our goal is to construct a good discrete approximation of $\mathcal{L}$. This will require controlling the two sources of error: (i) that we only use the first $\kappa + 1$ eigenvalues and (ii) that we do not know the eigenvalues or eigenfunctions of the Laplace-Beltrami operator $\mathcal{L}$ and must instead use the eigenvalues and the eigenvectors of the data-driven Laplacian defined below. The following lemma addresses (i) by bounding the error induced by only using finitely many eigenvalues. For a proof please see Appendix I.

Lemma 4. *For $\kappa \geq 0$ and $f \in \mathbf{L}^2(\mathcal{X})$, we have*

$$\|H^\kappa_t f - H_t f\|_{\mathbf{L}^2(\mathcal{X})} \leq e^{-t\mu_{\kappa+1}} \|f\|_{\mathbf{L}^2(\mathcal{X})} \tag{40}$$

and also

$$\|H^\kappa_t f - H_t f\|_\infty \leq C_{\mathcal{X}} \|f\|_\infty, \tag{41}$$

where $C_{\mathcal{X}}$ is a constant which depends on the geometry of $\mathcal{X}$ but does not depend on κ, t, or f.

Next, we construct an unnormalized data-driven graph Laplacian by

$$L_{N,\epsilon} := \frac{1}{\epsilon N}(D^{(N)} - W^{(N)}).$$

We will interpret $W^{(N)}$ and $L_{N,\epsilon}$ as the adjacency matrix and Laplacian matrix of a data-driven graph $\mathcal{X}_N$. We will denote the eigenvectors and eigenvalues of $L_{N,\epsilon}$ by $\lambda^{N,\epsilon}_k$ and $\mathbf{u}^{N,\epsilon}_k$ so that

$$L_{N,\epsilon} \mathbf{u}^{N,\epsilon}_k = \lambda^{N,\epsilon}_k \mathbf{u}^{N,\epsilon}_k. \tag{42}$$

When convenient, we make the dependence on N and ϵ implicit and simply write $\mathbf{u}_k$ and λ_k in place of $\mathbf{u}^{N,\epsilon}_k$ and $\lambda^{N,\epsilon}_k$. We define the discrete truncated heat-kernel matrix by

$$H_{N,\epsilon,\kappa,t} := \sum_{k=0}^{\kappa} e^{-t\lambda^{N,\epsilon}_k} \mathbf{u}^{N,\epsilon}_k (\mathbf{u}^{N,\epsilon}_k)^T. \tag{43}$$

To accomplish goal (ii), we will need discrete approximations of our eigenfunctions φ_k of the Laplace-Beltrami operator, which motivates us to introduce the normalized evaluation operator $\rho : C(\mathcal{X}) \to \mathbb{R}^N$ given by

$$\rho f := \frac{1}{\sqrt{N}}(f(x_0), \ldots, f(x_{N-1})).$$

We then define

$$\mathbf{v}_k := \rho \varphi_k.$$

We note these definitions differ slightly from [15]. There, the authors do not include the normalization term $\frac{1}{\sqrt{N}}$ in the definition of the evaluation operator ρ but instead include it in the definition of the vector $\mathbf{v}_k$. Importantly, we note that in either case the

[2] Notably, our construction is sensitive to the choice of this bandwidth parameter. For more on this issue, we refer the reader to [53] which discusses some remedies to this sensitivity.

definition of $\mathbf{v}_k$ is ultimately the same, i.e., $(\mathbf{v}_k)_i = \frac{1}{\sqrt{N}}\varphi_k(x_i)$ (although [15] uses the letter ϕ instead of $\mathbf{v}$)). Our convention is chosen so that $\mathbb{E}\|\rho f\|_2^2 = \|f\|_{\mathbf{L}^2(\mathcal{X})}^2$. Additionally, we may also use Hoeffding's inequality to derive the following lemma which shows that $\|\rho f\|_2^2 \approx \|f\|_{\mathbf{L}^2(\mathcal{X})}^2$ with high probability as $N \to \infty$. For a proof please see Appendix J.

Lemma 5. *Assume that the points $\{x_i\}_{i=0}^{N-1}$ are drawn i.i.d. uniformly at random, and let $f, g \in C(\mathcal{X})$. Then, with probability at least $1 - \frac{2}{N^9}$, we have*

$$|\langle \rho f, \rho g\rangle_2 - \langle f, g\rangle_{\mathbf{L}^2(\mathcal{X})}| \leq \sqrt{\frac{18\log N}{N}}\|fg\|_\infty.$$

Our goal is to show that for a large fixed κ, in the limit as $N \to \infty$ and $\epsilon \to 0$, we have $H_{N,\epsilon,\kappa,t}\rho f \approx \rho H_t f$ in the sense that κ is large enough so that $\|\rho H_t f - \rho H_t^\kappa f\|_2$ is negligible, which follows for large κ from Lemmas 4 and 5, and

$$\|H_{N,\epsilon,\kappa,t}\rho f - \rho H_t f\|_2 \to 0 \quad \text{as } N \to \infty.$$

In order to do this, we need the following result that shows that $\lambda_k^{N,\epsilon}$ and $\mathbf{u}_k^{N,\epsilon}$ are good approximations of μ_k and $\mathbf{v}_k$. It is a special case of Theorem 5.4 from [15], which follows by setting $\epsilon \sim N^{-2/(d+6)}$.

Theorem 9 *(Theorem 5.4 of [15]). Assume that the points $\{x_i\}_{i=0}^{N-1}$ are drawn i.i.d. uniformly at random and that the first $\kappa+2$ eigenvalues of $\mathcal{L}$, $\mu_0, \ldots, \mu_{\kappa+1}$, all have single multiplicity. As in (42), let $\mathbf{u}_k^{N,\epsilon}$ and $\lambda_k^{N,\epsilon}$ be the eigenvectors and eigenvalues of the data-driven Laplacian constructed via the Gaussian affinity kernel K_ϵ defined as in (38), and let $\kappa > 0$ be fixed. Assume that $\epsilon \to 0$ and $N \to \infty$ at a rate where $\epsilon \sim N^{-2/(d+6)}$. Then, with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$, there exist scalars α_k with*

$$|\alpha_k| = 1 + o(1)$$

such that for all $0 \leq k \leq \kappa$

$$|\mu_k - \lambda_k^{N,\epsilon}| = \mathcal{O}\left(N^{-\frac{2}{d+6}}\right), \quad \|\mathbf{u}_k^{N,\epsilon} - \alpha_k \mathbf{v}_k\|_2 = \mathcal{O}\left(N^{-\frac{2}{d+6}}\sqrt{\log N}\right),$$

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

Remark 8. Inspecting the proofs of Theorem 5.4 of [15] and the related results in that paper shows that when $\epsilon \sim N^{-2/(d+6)}$, we have

$$\max\left\{||\alpha_k| - 1|, \left|\frac{1}{|\alpha_k|} - 1\right|\right\} \leq \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right) + \mathcal{O}\left(\frac{\log(N)}{N^{4/(d+6)}}\right).$$

Please see Appendix K for details.

Given Theorem 9, we may use Lemma 5 to derive the following result which shows that $H_{N,\epsilon,\kappa,t}\rho f$ converges to $\rho H_t^\kappa f$ as $N \to \infty$. Moreover, the rate of the convergence for $H_{N,\epsilon,\kappa,t}\rho f$, is the same (up to logarithmic factors) as the convergence rate for the eigenvectors and eigenvalues provided in Theorem 9.

Theorem 10. *Let $f \in C(\mathcal{X})$. Then, under the assumptions of Theorem 9 we have*

$$\|H_{N,\epsilon,\kappa,t}\rho f - \rho H_t^\kappa f\|_2^2 \leq \max\{t^2, 1\}\mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right)\left(\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \sqrt{\frac{\log N}{N}}\|f\|_\infty^2\right)$$

with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$ if $d \geq 2$. In the case where $d = 1$, if the assumptions of Theorem 9 hold, then we have

$$\begin{aligned}&\|H_{N,\epsilon,\kappa,t}\rho f - \rho H_t^\kappa f\|_2^2 \\ &\leq \max\{t^2, 1\}\left(\mathcal{O}\left(\frac{\log N}{N^{4/7}}\right)\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \mathcal{O}\left(\frac{\log N}{N}\right)\|f\|_\infty^2\right).\end{aligned}$$

In both cases, the implied constants depend both on κ and on the geometry of $\mathcal{X}$.

If we combine Theorem 10 with Lemma 4, we may then obtain the following corollary.

Corollary 1. *Let $f \in C(\mathcal{X})$. Then, under the assumptions of Theorem 9, we have*

$$\|H_{N,\epsilon,\kappa,t}\rho f - \rho H_t f\|_2^2 \tag{44}$$

$$\leq \max\{t^2, 1\}\left[\left(\mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right) + 2e^{-2t\mu_{\kappa+1}}\right)\|f\|^2_{\mathbf{L}^2(\mathcal{X})} + \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right)\|f\|^2_\infty\right] \tag{45}$$

with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$*, where the constants implied by the* $\mathcal{O}$ *notation depend both on* κ *and the geometry of* $\mathcal{X}$*.*

For proofs of Theorem 10 and Corollary 1, please see Appendix L.

In our eigenvector based method, where we approximate the heat semigroup via (43), we next define a data-driven wavelet transform

$$\mathcal{W}_{J,N}\mathbf{x} := \{W_{j,N}\mathbf{x}, A_{J,N}\mathbf{x}\}_{j=0}^{J},$$

where $W_{0,N}\mathbf{x} = (I_N - H_{N,\epsilon,\kappa,1})\mathbf{x}$, $A_{J,N}\mathbf{x} = H_{N,\epsilon,\kappa,2^J}\mathbf{x}$ and for $1 \leq j \leq J$,

$$W_{j,N}\mathbf{x} = H_{N,\epsilon,\kappa,2^{j-1}}\mathbf{x} - H_{N,\epsilon,\kappa,2^j}\mathbf{x}.$$

We note that these wavelets implicitly depend on both κ and ϵ in addition to N, but we suppress these dependencies in order to avoid cumbersome notation. Analogously to Section 2.2, for a path $p = (j_1, \ldots, j_m)$ we define

$$U_N[p]\mathbf{x} := \sigma W_{j_m,N} \ldots \sigma W_{j_1,N}\mathbf{x}$$

and define data-driven scattering coefficients by

$$S_{J,N}[p]\mathbf{x} := A_{J,N}U_N[p]\mathbf{x} \quad \text{and} \quad \overline{S}_N[p]\mathbf{x} := |\langle U_N[p]\mathbf{x}, \mathbf{u}_0\rangle_2|.$$

In the case where we approximate the heat semigroup via (39) rather than (43), we define $W_{j,N}, U_N[p]$, and $S_{J,N}$ similarly, but with $H^{2^j}_{N,\epsilon}$ in place of $H_{N,\epsilon,\kappa,2^j}$ and we define the non-windowed scattering transform by $\overline{S}_N[p]\mathbf{x} = \|U_N[p]\mathbf{x}\|_1$ in order to avoid needing to compute any eigenvalues.

The following theorem uses Corollary 1 to bound the discretization error of the wavelets.

Theorem 11. *Let* $f \in C(\mathcal{X})$ *and assume that the heat semigroup is approximated as in* (43)*. Then, under the assumptions of Theorem 9, we have that*

$$\begin{aligned}&\|W_{j,N}\rho f - \rho W_j f\|_2^2\\ \leq& 2^{2j}\left[\left(\mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right) + \mathcal{O}\left(e^{-2^j\mu_{\kappa+1}}\right)\right)\|f\|^2_{\mathbf{L}^2(\mathcal{X})} + \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right)\|f\|^2_\infty\right],\end{aligned}$$

with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$*, where the constants implied by the big-*$\mathcal{O}$ *notation depend both on* κ *and on the geometry of* $\mathcal{X}$*.*

Proof. For $j \geq 1$,

$$\begin{aligned}\|W_{j,N}\rho f - \rho W_j f\|_2^2 &\leq \|(H_{N,\epsilon,\kappa,2^{j-1}}\rho f - H_{N,\epsilon,\kappa,2^j}\rho f) - (\rho H^{2^{j-1}} f - \rho H^{2^j} f)\|_2^2\\ &\leq 2\|H_{N,\epsilon,\kappa,2^{j-1}}\rho f - \rho H^{2^{j-1}} f\|_2^2 + 2\|H_{N,\epsilon,\kappa,2^j}\rho f - \rho H^{2^j} f\|_2^2.\end{aligned}$$

Therefore, the result follows from Corollary 1. For the case where $j = 0$, we note that $I_N\rho f = \rho \mathrm{Id} f$. Therefore,

$$W_{j,N}\rho f - \rho W_j f = H_{N,\epsilon,K,1}\rho f - \rho H^1 f$$

and we may again conclude by applying Corollary 1. □

Iteratively applying Theorem 11, one may obtain the following bound for the discretization error of $U_N[p]\rho f$. For a proof, please see Appendix M.

Theorem 12. *Let* $f \in C(\mathcal{X})$ *and assume that the heat semigroup is approximated as in* (43)*. Let* $p = (j_1, \ldots, j_m)$ *be a path of length* m *for some* $m \geq 1$*, and let* $j_{\max} = \max_{1\leq i\leq m} j_i$*. Then, under the assumptions of Theorem 9, we have that*

$$\begin{aligned}&\|U_N[p]\rho f - \rho U[p] f\|_2^2\\ \leq& 2^{2j_{\max}}\left[\left(\mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right) + \mathcal{O}\left(e^{-\mu_{\kappa+1}}\right)\right)\|f\|^2_{\mathbf{L}^2(\mathcal{X})} + \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right)\|f\|^2_\infty\right],\end{aligned}$$

where the constants implied by the $\mathcal{O}$ *notation depend on* m*,* κ*, and the geometry of* $\mathcal{X}$*.*

Inspecting the proof of Theorem 12, one may observe that the constants implied by the big-$\mathcal{O}$ notation increase exponentially with respect to m. However, in practice, one typically only uses two or three scattering layers, so we do not view this as a major limitation. We also note that a similar exponential dependence on the number of layers was observed for the generalization bounds for message passing networks arising from the discretization of graphons in [58]. Additionally, we note that, by inspecting the proof, it is clear that the implied constants in the term $\mathcal{O}(e^{-\mu_{\kappa+1}})$ do not depend on κ. A similar remark holds for the analogous terms in our subsequent results.

The next two results establish convergence of the windowed and non-windowed scattering coefficients as $N \to \infty$. For proofs, please see Appendix N.

Theorem 13. *Let $f \in C(\mathcal{X})$ and assume that the heat semigroup is approximated as in (43). Let $p = (j_1, \dots, j_m)$ be a path of length m for some $m \geq 1$. Then, under the assumptions of Theorem 9, we have*

$$\begin{aligned}&\|S_{J,N}[p]\rho f - \rho S_J[p]f\|_2^2 \\ \leq & 2^{2J}\left[\left(\mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right) + \mathcal{O}\left(e^{-\mu_{\kappa+1}}\right)\right)\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right)\|f\|_\infty^2\right],\end{aligned}$$

with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$, where the constants implied by the big-$\mathcal{O}$ notation depend on m, κ, and the geometry of $\mathcal{X}$.

The following is the analog of Theorem 13 for the non-windowed scattering transform.

Theorem 14. *Let $f \in C(\mathcal{X})$ and assume that the heat semigroup is approximated as in (43). Let $p = (j_1, \dots, j_m)$ be a path of length m for some $m \geq 1$. Then, under the assumptions of Theorem 9, we have*

$$\begin{aligned}&|\overline{S}_N[p]\rho f - \overline{S}[p]f| \\ \leq & 2^J\left[\left(\mathcal{O}\left(\frac{\sqrt{\log N}}{N^{\frac{2}{d+6}}}\right) + \mathcal{O}\left(e^{-\mu_{\kappa+1}/2}\right)\right)\|f\|_{\mathbf{L}^2(\mathcal{X})} + \mathcal{O}\left(\left(\frac{\log N}{N}\right)^{1/4}\right)\|f\|_\infty\right]\end{aligned}$$

with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$, where the constants implied by the big-$\mathcal{O}$ notation depend on m, κ, and the geometry of $\mathcal{X}$.

The convergence guarantees presented in this section may be summarized as follows. Theorem 9 is a result from [15] which provides convergence rates for the eigenvectors and eigenvalues of $L_{N,\epsilon}$. We then use this result to obtain convergence rates for our discretization of the heat semigroup, the wavelet transform, and the scattering transform. In all of these results, both the assumptions on the manifold and the convergence rate with respect to N are the same as in Theorem 9. Moreover, inspecting the proofs, one will observe that any future work which builds upon Theorem 9 by, e.g., relaxing the assumption that the λ_k have single multiplicity, will readily lead to improved versions of our convergence results for the scattering transform. We also note that, given a point cloud, there are many possible ways to construct a graph Laplacian which approximates the Laplace-Beltrami operator. For example, [12] proves a result analogous to Theorem 9 for nearest neighbor graphs and ϵ graphs. One could readily modify our method to define approximations of the manifold scattering transform using these graphs, and it is likely that one could imitate the methods presented here in order to obtain convergence results as $N \to \infty$. Additionally, we note that under certain assumptions on the generation of data points $\{x_i\}_{i=0}^{N-1}$, for example, when the sampling is not uniform, the users could add additional terms which account for the density of the data (see, e.g., the α-normalization approach [20,27,54]) when constructing $W^{(N)}$ or to implement methods based on other data-driven Laplacians such as the longest-leg path distance Laplacian considered in [55].

7. Numerical results

The numerical effectiveness of the graph scattering transform for tasks such as node classification, graph classification, and even graph synthesis has been demonstrated in numerous works such as [35,34,89,83,88] and [3]. However, the numerical effectiveness of the manifold scattering transform is much less well established. Indeed, the initial work [63] only provided numerical experiments on two-dimensional surfaces with predefined meshes. Here, in Sections 7.1 and 7.2, we will show that the methods proposed in Section 6 are effective for both synthetic and real-world data. As in Section 6, we assume that we may only access the manifold though a finite collection of random samples $\{x_i\}_{i=0}^{N-1}$ in both Sections 7.1 and 7.2. Additionally, in Section 7.3, we will show that our proposed method is effective for node classifications on directed graphs.

First, in Section 7.1 we will show that the manifold scattering transform is effective for learning on two-dimensional surfaces, even without a mesh. In particular, we will consider the same toy data sets that were analyzed with a mesh-based approach in [63]. These experiments aim to provide validation for our methods and show that the manifold scattering transform can still produce good results in the more challenging setting where one does not have access to the entire manifold. Having established proof of concept on toy data sets, in Section 7.2 we apply the manifold scattering transform to high-dimensional biomedical data where one models the data as lying upon some unknown manifold. In both of these settings, we will follow the lead of [35] and [3] and augment the expressive power of the scattering transform by considering higher q-th order scattering moments for $1 \leq q \leq Q$ defined by

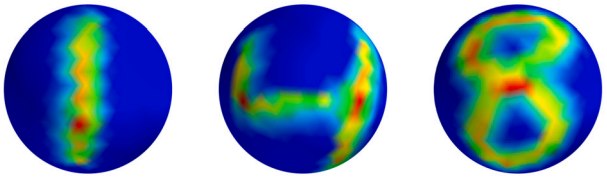


Fig. 1. The MNIST dataset projected onto the sphere.

Table 1
Classification accuracies for spherical MNIST averaged over 10 realizations.

Data type	N	κ	Q	Accuracy (%)
Point cloud	1200	200	4	79 ± 0.9
Point cloud	1200	400	4	88 ± 0.2
Point cloud	1200	642	4	84 ± 0.7
Mesh	642	642	1	91 ± 0.2

$$\overline{S}[p,q]\rho f = \frac{1}{N}\|U_N[p]\rho f\|_q^q.$$

Using these higher-order moments instead of the standard non-windowed scattering transform increases the expressive power of our representation and helps compensate for the lack of global knowledge of the manifold. Notably, these scattering moments are invariant to the ordering of the data points, since by Theorem 3 each $U_N[p]$ is equivariant to permutations (i.e., reorderings) and each $\overline{S}[p,q]$ is defined via a global summation. Additionally, we note that if the kernel $K(x_i, x_j)$ is a function of the Euclidean distance between x_i and x_j then the $\overline{S}_N[p,q]$ will be invariant to rigid motions in the embedded space. Throughout this section, we shall report all accuracies as mean $\pm$ standard deviation.

7.1. Two-dimensional surfaces without a mesh

When implementing convolutional networks on two-dimensional surfaces, it is standard, e.g., [5,6] to use triangular meshes. In this section, we show that mesh-free methods can also work well in this setting. Importantly, note that we are *not* claiming that mesh-free methods are *better* for two-dimensional surfaces. Instead, we aim to show that these methods can work relatively well thereby justifying their use in higher-dimensional settings.

We conduct experiments using both mesh-based and mesh-free methods on a spherical version of MNIST and on the FAUST dataset which were previously considered in [63]. In both methods, we use the wavelets defined in Section 2.1 with two scattering layers and $J = 8$ and use a radial basis function (RBF) kernel support vector machine (SVM) see, for example, [22,14] with cross-validated hyperparameters as our classifier. For the mesh-based methods, we use the same discretization scheme as in [63] and set $Q = 1$ which was the setting implicitly assumed there. For our mesh-free experiments, we use the eigenvector-based method discussed in Section 6 and set $Q = 4$. We show that the information captured by the higher-order moments can help compensate for the structure lost by not using a mesh. For all of our experiments on spherical MNIST and FAUST, we used an 80/20 train-test split with 10-fold cross-validation.

We first study the MNIST dataset projected onto the sphere as visualized in Fig. 1. We uniformly sampled N points from the unit two-dimensional sphere, and then applied random rotations to the MNIST dataset and projected each digit onto the spherical point cloud to generate a collection of signals $\{f_i\}$ on the sphere. Table 1 shows that for properly chosen κ, the mesh-free method can achieve similar performance to the mesh-based method. As noted in Section 6, the implied constants in our theoretical results depend on κ. By inspecting the proof of Theorem 5.4 of [15] we see that for larger values of κ, more sample points are needed to ensure the convergence of the first κ eigenvectors in Theorem 9. Thus, we want κ to be large enough to get a good approximation of H^1, but also not too large.

Next, we consider the FAUST dataset, a collection of surfaces corresponding to scans of ten people in ten different poses [4] as shown in Fig. 2. As in [63], we use 352 SHOT descriptors [76] as our signals. We use the first $\kappa = 80$ eigenvectors and eigenvalues of the approximate Laplace-Beltrami operator of each point cloud to generate scattering moments. We achieved $94 \pm 3.7\%$ classification accuracy over 10 realizations for the task of classifying different poses. This is comparable with the 95% accuracy obtained with meshes in [63].

7.2. Single-cell datasets

In this section, we present two experiments showing the utility of manifold scattering in analyzing single-cell data. We will formulate these experiments as manifold classification tasks, where each patient will correspond to a different manifold and the goal is to predict patient outcomes. In particular, each patient will correspond to a collection of cells, and each cell will correspond to a

Fig. 2. Wavelets on the FAUST dataset with $g(\lambda) = e^{-0.0005\lambda}$, $j = 1, 3, 5$ from left to right. Positive values are red, while negative values are blue. (For interpretation of the colors in the figure(s), the reader is referred to the web version of this article.)

point in high-dimensional space.[3] Therefore, each patient will be described by a high-dimensional point cloud which we model as lying upon a low-dimensional manifold. In order to classify the patients, we compute the scattering transform on each manifold with signals corresponding to protein expression and then feed this representation into a classifier. For both of the experiments described in this section, we used a 75/25 train-test split. Notably, in both of the data sets we consider, the number of patients in fairly small. Therefore, the fact that the scattering transform uses predesigned filters is particularly advantageous in this setting.

On these datasets, we deviate slightly from our theory and demonstrate that our method can be effectively utilized with different graph constructions. In our first data set, which focuses on data derived from melanoma patients, we use a k-NN graph with $k = 5$. On our second data set, which is derived from COVID-19 patients, we use a Gaussian kernel with an adaptive bandwidth which is designed to account for non-uniform density of the data points. Specifically, we set

$$K_{\text{k-nn}}(x, x') = \frac{1}{2}\left(\exp\left(-\frac{\|x - x'\|_2^2}{\sigma_k(x)^2}\right) + \exp\left(-\frac{\|x - x'\|_2^2}{\sigma_k(x')^2}\right)\right), \tag{46}$$

where $\sigma_k(x)$ is the distance from x to its k-th nearest neighbor ($k = 3$). We then approximate H^1 via (39). For the COVID data, we used three scattering layers with $J = 8$ and $Q = 4$, imitating the settings used in [35]. We then apply principal component analysis (PCA) to the scattering features and train a decision tree classifier on the top 10 principal components. For the melanoma patients, we used 2 scattering layers with $J = 4$ and $Q = 4$, followed by a multilayer perceptron with a single hidden layer. Additionally, with the melanoma data, in order increased the effective size of our training data, we subsample point clouds of 400 points each and repeat this procedure 10 times for each point cloud (so the data set consists of 540 graphs rather than 54). Importantly, we note that we do this subsampling after splitting the data into train and test in order to ensure that no patient is in both the train and test set. As a baseline comparison, we compare our scattering-based method against a method which first preprocesses the data by using a k-means clustering based approach to extract features and then applies a decision tree classifier. For details on this baseline, please see Appendix O.

We first consider data collected in [65] on patients with various stages of melanoma. All patients received checkpoint blockade immunotherapy, a treatment that licenses patient T cells to kill tumor cells. (For details on this therapy, see [43].) In this dataset, 11,862 T lymphocytes from core tissue sections were taken from each of 54 patients diagnosed with melanoma, and 30 proteins were measured per cell. Therefore, we model our data as consisting of 54 manifolds embedded in 30-dimensional space (with one dimension corresponding to each of the proteins) with 11,862 points per manifold. We achieved 71% accuracy when using scattering moments based on protein expression feature signals[4] with a decision tree classifier compared to 46% accuracy using our baseline method.

We next consider data previously studied in [48] comprised of 209 blood samples from 148 people.[5] Of the 209 samples, 61 were taken from healthy controls, 123 were taken from patients who were COVID+ but recovered, and 25 were taken from patients who were COVID+ and died. Here, our goal is to predict whether the person corresponding to each blood sample died of COVID, recovered from COVID, or was a control. This task is particularly challenging because COVID outcome depends on a wide variety of known and unknown immunoregulatory pathways, unlike response to checkpoint blockade immunotherapy which targets a specific known immunoregulatory axis (T-cell inhibition). We focus on innate immune (myeloid) cells, a population that has previously been shown to be predictive of patient mortality [48]. Fourteen proteins were measured on $1,502,334$ total cells, approximately $10,000$ cells per patient. To accommodate the size of these data sets lying in $\mathbb{R}^{14}$, we first aggregate data points for each patient into less than 500 clusters via the diffusion condensation algorithm [48]. We treat the centroids (with respect to Euclidean distance) of each cluster as single data points in the high-dimensional immune state space when implementing the manifold scattering transform. As

[3] In order to turn the cells into points, we take single-cell protein measurements and apply a logarithmic transformation followed by ℓ^1 normalization.

[4] Proteins were selected on the basis of having a known functional role in T cell regulation and included CD4, CD8, CD45RO, CD56, FOXP3, Granzyme B, Ki-67, LAG3, PD-1, and TIM-3.

[5] In [48] the data was taken from 168 patients. However, here we focus on the 148 patients for whom sufficient monocyte data was available.

Table 2
Classification accuracies for patient outcome prediction.

Data set	N_{patients}	Baseline	Scattering
Melanoma	54	$46.0 \pm 7.1\%$	$71.0 \pm 9.0\%$
COVID	148	$40.1 \pm 2.2\%$	$47.7 \pm 0.5\%$

with our melanoma experiments, we used signals related to protein expression, averaged across cells in each cluster. For the baseline method, k-means clustering, we set $k = 3$ based on expected monocyte subtypes (classical, non-classical, intermediate). We achieved 48% accuracy with scattering and a decision tree classifier compared to 40% via the baseline method. See Table 2 for a summary of the results for both of the data sets discussed in this subsection.

7.3. Directed graphs

Next, we apply our framework to weighted and directed graphs $G = (V, E, W)$ with vertices V, edges E, and edge weights W. We turn G into a measure space $\mathcal{X} = (X, \mathcal{F}, \mu)$ by setting $X = V$, letting $\mathcal{F}$ be the set of all subsets of V, and letting μ be the uniform measure such that $\mu(\{v\}) = 1$ for all $v \in V$. In our experiments in this section, we will take $\mathcal{L}$ to be the normalized magnetic Laplacian described in detail below.

We let A denote the asymmetric, weighted adjacency matrix of G, and let $A^{(s)} = \frac{1}{2}(A + A^T)$ be its symmetric counterpart. Next, we define the symmetric, diagonal degree matrix $D^{(s)}$ by $D^{(s)}_{i,i} = \sum_{j=0}^{N-1} A^{(s)}_{i,j}$, where $N = |V|$, and $D^{(s)}_{i,j} = 0$ if $i \neq j$. We then let $\Theta = A - A^T$ and define the Hermitian adjacency matrix by

$$H^{(q)} = A^{(s)} \odot \exp(2\pi \mathbb{i} q \Theta),$$

where $\mathbb{i} = \sqrt{-1}$, $\odot$ denotes Hadamard product (componentwise multiplication), q is a "charge" parameter,[6] and exponentiation is defined componentwisely, i.e.,

$$\exp(2\pi \mathbb{i} q \Theta)_{i,j} = \exp(2\pi \mathbb{i} q \Theta_{i,j}).$$

Notably, $H^{(q)}$ encodes the undirected geometry of the graph in the magnitude of its entries and directional information via its phases. The charge parameter q allows one to balance the relationship between directed and undirected information as desired.

Given $H^{(q)}$, we define the unnormalized and normalized magnetic Laplacians by

$$L^{(q)}_U = D^{(s)} - H^{(q)}$$

and

$$\begin{aligned} L^{(q)}_N &= (D^{(s)})^{-1/2} L^{(q)}_U (D^{(s)})^{-1/2} \\ &= I - (D^{(s)})^{-1/2} H^{(q)} (D^{(s)})^{-1/2}. \end{aligned}$$

By construction, both $L^{(q)}_U$ and $L^{(q)}_N$ are Hermitian and one may check (see, e.g., Theorem 1 of [87]) that they are positive semidefinite. Therefore, both of these matrices fit within our framework as admissible choices of $\mathcal{L}$ and can be used to define scattering transforms on directed graphs.

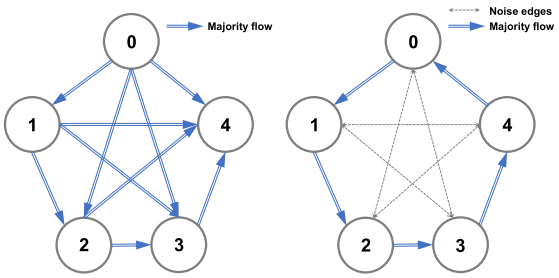
In our experiments, we will choose $\mathcal{L} = L^{(q)}_N$ and consider the task of node classification on the following directed stochastic block model considered in [87]. We first divide the N vertices into n_c equally-sized clusters $C_1, \ldots, C_{n_c}$ for some n_c which divides N. We let $\{\alpha_{i,j}\}_{1 \leq i,j \leq n_c}$ to be a collection of probabilities, with $\alpha_{i,j} = \alpha_{j,i}$ and $0 < \alpha_{i,j} \leq 1$. For an unordered pair of vertices, $u, v \in V$, $u \neq v$ we create an undirected edge between u and v with probability $\alpha_{i,j}$ if $u \in C_i, v \in C_j$. We then define $\{\beta_{i,j}\}_{1 \leq i,j \leq n_c}$ to be a collection of probabilities such that $\beta_{i,j} + \beta_{j,i} = 1$ and $0 \leq \beta_{i,j} \leq 1$. We then replace each undirected edge $\{u, v\}$, with a directed edge which points from u to v with probability $\beta_{i,j}$ if $u \in C_i$ and $v \in C_j$, and otherwise points from v to u. Notably, if $\alpha_{i,j}$ is constant, then the only way to determine the clusters will be from the directional information.

For our experiments, we set $n_c = 5$ and consider three meta-graphs: ordered, cyclic, and noisy cyclic. For all meta-graphs, we set $\beta_{i,i} = 0.5$. For the ordered meta-graph (Fig. 3a), we set $\alpha_{i,j} = 0.1$ for all i, j and set $\beta_{i,j} = 0.95$ for $i < j$. For the cyclic meta-graph (Fig. 3b, but without the dashed gray edges), we set

$$\alpha_{i,j} = \begin{cases} 0.1 & i = j \\ 0.1 & i = (j \pm 1) \bmod 5 \\ 0 & \text{otherwise} \end{cases} \text{ and } \beta_{i,j} = \begin{cases} 0.5 & i = j \\ 0.95 & i = (j-1) \bmod 5 \\ 0.05 & j = (i-1) \bmod 5 \\ 0 & \text{otherwise} \end{cases}.$$

Finally, for the noisy cyclic meta-graph (Fig. 3b), we set $\alpha_{i,j} = 0.1$ for all i, j and set the edge direction probabilities as

[6] This term, as well as the name Magnetic Laplacian originates from the Magnetic Laplaican serving as the quantum mechanical Hamiltonian of a particle under magnetic flux [51].



(a) Ordered (b) Noisy cyclic

Fig. 3. Meta-graphs for synthetic datasets (reproduced from [87]).

Table 3
Node classification accuracy on our directed stochastic block model with different meta-graph structures (ordered, cyclic, or noisy cyclic) and different graph methods.

Method/Meta-graph	Ordered	Cyclic	Noisy cyclic
MagNet [87]	99.6 ± 0.2	100.0 ± 0.0	80.5 ± 1.0
ChebNet [26]	19.9 ± 0.7	74.7 ± 16.5	18.3 ± 3.1
GCN [45]	68.6 ± 2.2	78.87 ± 30.0	24.2 ± 6.8
APPNP [46]	97.4 ± 1.8	19.6 ± 0.5	17.4 ± 1.8
SAGE [38]	20.2 ± 1.2	88.6 ± 8.3	26.4 ± 7.7
GIN [86]	57.9 ± 6.3	75.3 ± 21.5	24.7 ± 6.4
GAT [80]	42.0 ± 4.8	98.3 ± 2.2	27.4 ± 6.9
DGCN [79]	81.4 ± 1.1	83.7 ± 23.1	37.3 ± 6.1
DiGraph [78]	82.5 ± 1.4	39.1 ± 33.6	18.0 ± 1.8
DiGraphIB [78]	99.2 ± 0.4	84.8 ± 17.0	43.4 ± 10.1
Scattering	97.8 ± 1.2	99.8 ± 0.2	88.5 ± 4.0
Parameters	$J = 9, q = 0.25$	$J = 9, q = 0$	$J = 10, q = 0.2$

$$\beta_{i,j} = \begin{cases} 0.95 & i = (j-1) \bmod 5 \\ 0.05 & j = (i-1) \bmod 5 \\ 0.5 & \text{otherwise} \end{cases}.$$

Motivated by the so-called residual convolution operators used in [83], for improved numerical performance, we use a modified version of the windowed scattering transform given by $S_J^{\text{res}}[p] = H^1 U[p]$ in our experiments. We chose our input signals to be i.i.d. standard Gaussian random vectors and used paths of length $m \in \{0, 1, 2\}$. Following the settings used in [87], we set $N = 2500$ and $n_c = 500$ for the ordered and cyclic meta-graphs and $N = 500$ and $n_c = 100$ for the noisy meta-graph, and we used 2%, 10%, and 60% of the nodes in each cluster for training for the ordered, cyclic, and noisy cyclic meta-graphs, respectively. On all three data sets, we used 20% of the nodes for validation and the remaining nodes were used for testing. Details on our validation procedure are provided in Appendix P. After computing the scattering transform, we used an SVM with an RBF kernel for classification. In Table 3, we report our results for each meta-graph along with the maximum scale J used to compute the scattering coefficients and parameter q used to compute the magnetic Laplacian.[7] As we can see, scattering performs well on all three versions of the stochastic block model and is the top-performing method on the noisy cyclic stochastic block model.[8]

8. Conclusion

In this work, we have extended the geometric scattering transform to a broad class of measure spaces. In particular, our construction extends several previous works defining the scattering transform on undirected, unsigned graphs and smooth compact Riemannian manifolds without boundary as special cases and also includes many other examples as discussed extensively in Section 3. Our invariance and equivariance results help clarify the relationship between the invariance / equivariance of the scattering transform and the group of bijections to which it is invariant or equivariant. Namely, they show that the critical property for $\mathcal{G}$ to possess is that for every $\zeta \in \mathcal{G}$, the operator V_ζ, defined by $V_\zeta f(x) = f(\zeta^{-1}(x))$, is an isometry on $\mathbf{L}^2(\mathcal{X})$. Additionally, we provide two numerical schemes for implementing the manifold scattering transform when one only has access to finite point clouds and provide quantitative convergence rates for one of these schemes as the number of sample points grows to infinity. The proof of this convergence result utilizes previous work showing the convergence of the eigenvectors and eigenvalues of the Laplace-Beltrami operator.

[7] All baseline results taken from [87].

[8] For methods not designed for di-graphs, the reported accuracies are the maximum of those obtained by a) symmetrizing the adjacency matrix as a preprocessing step and b) running the algorithm as is.

While we do not know whether or not our convergence rate is optimal, we do note that both the assumptions of our convergence theorems and our convergence rates are the same as for the previous work on the convergence of the eigenvectors and eigenvalues. Therefore, our convergence results should be interpreted as showing that the number of sample points needed to apply scattering to high-dimensional point cloud data is the same as other manifold learning based methods.

We believe our work opens up several new exciting avenues for future research. The framework presented here provides a theoretical foundation for defining neural networks on manifolds from point-cloud data, a relatively unexplored topic except in the setting of two-dimensional surfaces. Additionally, it would be interesting to extend our methods to higher-order operators such as the connection Laplacian or to implement versions of our method that utilize anisotropic diffusions. Lastly, it would be interesting to improve on our convergence results by relaxing the assumptions on the data generation or developing quantitative convergence guarantees that do not require the explicit computation of eigenvectors or eigenvalues.

Data availability

Data will be made available on request.

Appendix A. The proof of Proposition 1

Proof. To prove the upper bound, we note,

$$\begin{aligned}
&\sum_{j=0}^{J} \|W_j f\|_{\mathcal{H}}^2 + \|A_J f\|_{\mathcal{H}}^2 \\
&= \sum_{k\in\mathcal{I}} \sum_{j=0}^{J} \left(|\widehat{W_j f}(k)|^2 + |\widehat{A_J f}(k)|^2 \right) \\
&= \sum_{k\in\mathcal{I}} \left(|1-g(\lambda_k)|^2 + \sum_{j=1}^{J} \left| g(\lambda_k)^{2^{j-1}} - g(\lambda_k)^{2^j} \right|^2 + \left| g(\lambda_k)^{2^J} \right|^2 \right) |\widehat{f}(k)|^2 \\
&\leq \sum_{k\in\mathcal{I}} \left(1 - g(\lambda_k) + \sum_{j=1}^{J} \left(g(\lambda_k)^{2^{j-1}} - g(\lambda_k)^{2^j} \right) + g(\lambda_k)^{2^J} \right)^2 |\widehat{f}(k)|^2 \qquad (47) \\
&= \sum_{k\in\mathcal{I}} |\widehat{f}(k)|^2 \\
&= \|f\|_{\mathcal{H}}^2,
\end{aligned}$$

where in (47), we used the fact that g is nonnegative and decreasing.

In order to prove the lower bound, we define $p_0(t) := (1-t)$, $p_j(t) := (t^{2^{j-1}} - t^{2^j})$ if $1 \leq j \leq J$, and $p_{J+1}(t) := t^{2^J}$ and observe that since g is positive and decreasing we have

$$\begin{aligned}
&\sum_{j=0}^{J} \|W_j f\|_{\mathcal{H}}^2 + \|A_J f\|_{\mathcal{H}}^2 \\
&= \sum_{k\in\mathcal{I}} \left(|1-g(\lambda_k)|^2 + \sum_{j=1}^{J} \left| g(\lambda_k)^{2^{j-1}} - g(\lambda_k)^{2^j} \right|^2 + \left| g(\lambda_k)^{2^J} \right|^2 \right) |\widehat{f}(k)|^2 \\
&= \sum_{k\in\mathcal{I}} \left(p_0(g(\lambda_k))^2 + \sum_{j=1}^{J} p_j(g(\lambda_k))^2 + p_{J+1}(g(\lambda_k))^2 \right) |\widehat{f}(k)|^2 \\
&\geq \min_{0\leq t\leq 1} \sum_{j=0}^{J+1} p_j(t)^2 \|f\|_{\mathcal{H}}^2,
\end{aligned}$$

where in the final inequality we use Plancherel's identity and the fact that $0 \leq g(\lambda_k) \leq g(0) = 1$.

Therefore it suffices to show that $\min_{0\leq t\leq 1} \sum_{j=0}^{J+1} p_j(t)^2 \geq c > 0$. To do so, we let $0 \leq t \leq 1$ and consider three cases. First, if $0 \leq t \leq 1/2$, then

$$\sum_{j=0}^{J+1} p_j(t)^2 \geq p_0(t)^2 = (1-t)^2 \geq \left(1 - \frac{1}{2}\right)^2 = 1/4.$$

Secondly, if $t^{2^J} \geq 1/2$,

$$\sum_{j=0}^{J+1} p_j(t)^2 \geq p_{J+1}(t)^2 = \left(t^{2^J}\right)^2 \geq \left(\frac{1}{2}\right)^2 = 1/4.$$

In the final case where $t^{2^J} < \frac{1}{2} < t$, there exists a unique j_0, $1 \leq j_0 \leq J$, such that $t^{2^{j_0}} < 1/2 \leq t^{2^{j_0-1}}$. Since $t^{2^{j_0-1}} \geq 1/2$ and $t^{2^{j_0-1}} t^{2^{j_0-1}} = t^{2^{j_0}}$ it follows that $1/4 \leq t^{2^{j_0}} < 1/2$ and thus $1/2 \leq t^{2^{j_0-1}} < 1/\sqrt{2}$. Therefore, in this case we have

$$\sum_{j=0}^{J+1} p_j(t)^2 \geq p_{j_0}(t)^2 = (t^{2^{j_0-1}} - t^{2^{j_0}})^2 \geq \inf_{x \in [\frac{1}{2}, \frac{1}{\sqrt{2}}]} (x - x^2)^2 =: c > 0. \quad \square$$

Appendix B. The proof of Proposition 2

Proof. Using the definition on the scattering transform as well as the fact that $g(0) = 1$, one may compute

$$\begin{aligned}\left| |S_J[p]f(x)| - \overline{S}[p]f|\varphi_0(x)| \right| &= \left| \left| \sum_{k \geq 0} g(\lambda_k)^{2^J} \langle U[p]f, \varphi_k \rangle_{\mathcal{H}} \varphi_k(x) \right| - |\langle U[p]f, \varphi_0 \rangle_{\mathcal{H}}| |\varphi_0(x)| \right| \\ &\leq \left| \sum_{k \geq 1} g(\lambda_k)^{2^J} \langle U[p]f, \varphi_k \rangle_{\mathcal{H}} \varphi_k(x) \right|.\end{aligned}$$

Therefore, Parseval's identity implies that

$$\begin{aligned}\left\| |S_J[p]f(x)| - \overline{S}[p]f|\varphi_0(x)| \right\|_{\mathcal{H}}^2 &\leq \sum_{k \geq 1} \left| g(\lambda_k)^{2^J} \right|^2 |\langle U[p]f, \varphi_k \rangle_{\mathcal{H}}|^2 \\ &\leq g(\lambda_1)^{2^{J+1}} \sum_{k \geq 1} |\langle U[p]f, \varphi_k \rangle_{\mathcal{H}}|^2 \\ &\leq g(\lambda_1)^{2^{J+1}} \|U[p]f\|_{\mathcal{H}}^2.\end{aligned}$$

Since $\lambda_1 > 0$, (7) implies $g(\lambda_1) < 1$, and so the right-hand side converges to zero as $J \to \infty$. $\square$

Appendix C. The proof of Theorems 1 and 2

The proof of Theorem 1 is based on the following lemma.

Lemma 6. *For all $f_1, f_2 \in \mathcal{H}$, we have*

$$\sum_{p \in \mathcal{J}^m} \|U[p]f_1 - U[p]f_2\|_{\mathcal{H}}^2 \geq \sum_{p \in \mathcal{J}^{m+1}} \|U[p]f_1 - U[p]f_2\|_{\mathcal{H}}^2 + \sum_{p \in \mathcal{J}^m} \|S[p]f_1 - S[p]f_2\|_{\mathcal{H}}^2. \tag{48}$$

Moreover, for all $f \in \mathcal{H}$

$$\sum_{p \in \mathcal{J}^m} \|U[p]f\|_{\mathcal{H}}^2 \geq \sum_{p \in \mathcal{J}^{m+1}} \|U[p]f\|_{\mathcal{H}}^2 + \sum_{p \in \mathcal{J}^m} \|S[p]f\|_{\mathcal{H}}^2. \tag{49}$$

The Proof of Lemma 6. The assumption (11) implies that for all $p \in \mathcal{J}^m$ we have that

$$\|U[p]f_1 - U[p]f_2\|_{\mathcal{H}}^2 \geq \sum_{j_{m+1} \in \mathcal{J}} \|W_{j_{m+1}}(U[p]f_1 - U[p]f_2)\|_{\mathcal{H}}^2 + \|A(U[p]f_1 - U[p]f_2)\|_{\mathcal{H}}^2.$$

Therefore,

$$\begin{aligned}&\sum_{p \in \mathcal{J}^m} \|U[p]f_1 - U[p]f_2\|_{\mathcal{H}}^2 \\ &\geq \sum_{p \in \mathcal{J}^m} \left(\sum_{j_{m+1} \in \mathcal{J}} \|W_{j_{m+1}}(U[p]f_1 - U[p]f_2)\|_{\mathcal{H}}^2 + \|A(U[p]f_1 - U[p]f_2)\|_{\mathcal{H}}^2 \right) \qquad (50) \\ &= \sum_{p \in \mathcal{J}^m} \left(\sum_{j_{m+1} \in \mathcal{J}} \|W_{j_{m+1}}U[p]f_1 - W_{j_{m+1}}U[p]f_2\|_{\mathcal{H}}^2 + \|A(U[p]f_1 - U[p]f_2)\|_{\mathcal{H}}^2 \right) \\ &\geq \sum_{p \in \mathcal{J}^m} \left(\sum_{j_{m+1} \in \mathcal{J}} \|\sigma W_{j_{m+1}}U[p]f_1 - \sigma W_{j_{m+1}}U[p]f_2\|_{\mathcal{H}}^2 + \|A(U[p]f_1 - U[p]f_2)\|_{\mathcal{H}}^2 \right) \qquad (51)\end{aligned}$$

$$= \sum_{p\in\mathcal{J}^m}\left(\sum_{j_{m+1}\in\mathcal{J}}\|U[j_{m+1}]U[p]f_1 - U[j_{m+1}]U[p]f_2\|_{\mathcal{H}}^2 + \|A(U[p]f_1 - U[p]f_2)\|_{\mathcal{H}}^2\right)$$
$$= \sum_{p\in\mathcal{J}^{m+1}}\|U[p]f_1 - U[p]f_2\|_{\mathcal{H}}^2 + \sum_{p\in\mathcal{J}^m}\|S[p]f_1 - S[p]f_2\|_{\mathcal{H}}^2.$$

This completes the proof of (48). (49) follows from setting $f_2 = 0$ and noting that in this case equality holds in (50) and (51). □

Proof of Theorem 1. Applying Lemma 6, and recalling that $U[p_e]f = f$, we see

$$\begin{aligned}\|Sf_1 - Sf_2\|_{\ell^2(\mathcal{H})}^2 &= \lim_{N\to\infty}\sum_{m=0}^{N}\sum_{p\in\mathcal{J}^m}\|S[p]f_1 - S[p]f_2\|_{\mathcal{H}}^2\\ &\le \lim_{N\to\infty}\sum_{m=0}^{N}\left(\sum_{p\in\mathcal{J}^m}\|U[p]f_1 - U[p]f_2\|_{\mathcal{H}}^2 - \sum_{p\in\mathcal{J}^{m+1}}\|U[p]f_1 - U[p]f_2\|_{\mathcal{H}}^2\right)\\ &\le \|f_1 - f_2\|_{\mathcal{H}}^2 - \limsup_{N\to\infty}\sum_{p\in\mathcal{J}^{N+1}}\|U[p]f_1 - U[p]f_2\|_{\mathcal{H}}^2\\ &\le \|f_1 - f_2\|_{\mathcal{H}}^2. \quad \square\end{aligned}$$

The Proof of Theorem 2. Note that

$$\left\|\frac{|S_J[p]f_i|}{|\varphi_0|} - \overline{S}[p]f_i\right\|_{\mathcal{H}} \le \frac{1}{\min_x|\varphi_0(x)|}\||S_J[p]f_i| - \overline{S}[p]f_i|\varphi_0|\|_{\mathcal{H}}.$$

Therefore, by Proposition 2,

$$\lim_{J\to\infty}\left\|\frac{|S_J[p]f_i|}{|\varphi_0|} - \overline{S}[p]f_i\right\|_{\mathcal{H}} = 0$$

which in turn implies that

$$\lim_{J\to\infty}\left\|\frac{|S_J[p]f_i|}{|\varphi_0|}\right\|_{\mathcal{H}} = \overline{S}[p]f_i\mathrm{vol}(\mathcal{X})^{1/2}.$$

Thus, using Fatou's lemma, we have

$$\begin{aligned}\|\overline{S}f_1 - \overline{S}f_2\|_2^2 &= \sum_p|\overline{S}[p]f_1 - \overline{S}[p]f_2|^2\\ &= \frac{1}{\mathrm{vol}(\mathcal{X})}\sum_p\lim_{J\to\infty}\left|\left\|\frac{|S_J[p]f_1|}{|\varphi_0|}\right\|_{\mathcal{H}} - \left\|\frac{|S_J[p]f_2|}{|\varphi_0|}\right\|_{\mathcal{H}}\right|^2\\ &\le \frac{1}{\mathrm{vol}(\mathcal{X})}\liminf_{J\to\infty}\sum_p\left|\left\|\frac{|S_J[p]f_1|}{|\varphi_0|}\right\|_{\mathcal{H}} - \left\|\frac{|S_J[p]f_2|}{|\varphi_0|}\right\|_{\mathcal{H}}\right|^2\\ &\le \frac{1}{\mathrm{vol}(\mathcal{X})}\liminf_{J\to\infty}\sum_p\left\|\frac{S_J[p]f_1 - S_J[p]f_2}{\varphi_0}\right\|_{\mathcal{H}}^2\\ &\le \frac{1}{\min_x|\varphi_0(x)|^2\mathrm{vol}(\mathcal{X})}\liminf_{J\to\infty}\sum_p\|S_J[p]f_1 - S_J[p]f_2\|_{\mathcal{H}}^2\\ &\le \frac{1}{\min_x|\varphi_0(x)|^2\mathrm{vol}(\mathcal{X})}\|f_1 - f_2\|_{\mathcal{H}},\end{aligned}$$

where in the last line we applied Theorem 1. □

Appendix D. The proof of Theorem 3

Proof. Since $A_j = H^{2^J}$ and $W_j = H^{2^{j-1}} - H^{2^j}$ it follows from (18) that

$$A^{(\zeta)}V_\zeta f = V_\zeta Af, \quad \text{and} \quad W_j^{(\zeta)}V_\zeta f = V_\zeta W_j f. \tag{52}$$

By definition, for all $\zeta\in\mathcal{G}$, we have that V_ζ commutes with σ since

$$(V_\zeta\sigma f)(x) = (\sigma f)(\zeta^{-1}(x)) = \sigma(f(\zeta^{-1}(x)) = \sigma(V_\zeta f(x)) = (\sigma V_\zeta f)(x).$$

Therefore, since by definition, $U[p]f = \sigma W_{j_m}f\ldots\sigma W_{j_1}f$ it follows that $U^{(\zeta)}V_\zeta f = V_\zeta Uf$. Lastly since $S = AU$, we have

$$S^{(\zeta)}V_\zeta f = A^{(\zeta)}U^{(\zeta)}V_\zeta f = A^{(\zeta)}V_\zeta U f = V_\zeta A U f = V_\zeta S f. \quad \square$$

Appendix E. The proof of Theorem 5

In this Section, we prove both Theorem 5 and Lemma 3.

Proof of Lemma 3. We first note that, under the assumption that $\mathcal{G}$ preserves the measure μ, the Hilbert spaces $\mathcal{H}$ and $\mathcal{H}^{(\zeta)}$ have the same elements and so the subtraction $Sf - SV_\zeta f$ is well defined. Similarly, we may identify V_ζ with an operator mapping $\mathcal{H}$ into itself. Therefore, by the assumption (19), we have

$$\begin{aligned}\|Sf - S^{(\zeta)}V_\zeta f\|_{\ell^2(\mathcal{H})} &= \|AUf - V_\zeta Sf\|_{\ell^2(\mathcal{H})} \\ &= \|AUf - V_\zeta AUf\|_{\ell^2(\mathcal{H})} \\ &\leq \|V_\zeta A - A\|_{\mathcal{H}}\|Uf\|_{\ell^2(\mathcal{H})}. \quad \square\end{aligned}$$

Proof of Theorem 5. Let $\zeta \in \mathcal{G}$. The assumption that φ_0 is constant implies that $V_\zeta\varphi_0 - \varphi_0 = 0$. Therefore,

$$\begin{aligned}\|V_\zeta A_J f - A_J f\|_{\mathcal{H}} &= \left\|\sum_{k\in\mathcal{I}} g(\lambda_k)^{2^J}\langle f,\varphi_k\rangle_{\mathcal{H}}(V_\zeta\varphi_k - \varphi_k)\right\|_{\mathcal{H}} \\ &= \left\|\sum_{k\geq 1} g(\lambda_k)^{2^J}\langle f,\varphi_k\rangle_{\mathcal{H}}(V_\zeta\varphi_k - \varphi_k)\right\|_{\mathcal{H}} \\ &\leq \left\|\sum_{k\geq 1} g(\lambda_k)^{2^J}\langle f,\varphi_k\rangle_{\mathcal{H}}V_\zeta\varphi_k\right\|_{\mathcal{H}} + \left\|\sum_{k\geq 1} g(\lambda_k)^{2^J}\langle f,\varphi_k\rangle_{\mathcal{H}}\varphi_k\right\|_{\mathcal{H}}.\end{aligned}$$

The assumption that $\mathcal{G}$ preserves inner products together with the assumption that it preserves the measure implies that for $f, g \in \mathcal{H}$

$$\langle V_\zeta f, V_\zeta g\rangle_{\mathcal{H}} = \int_X V_\zeta f\overline{V_\zeta g}\,d\mu = \int_X V_\zeta f\overline{V_\zeta g}\,d\mu^{(\zeta)} = \langle V_\zeta f, V_\zeta g\rangle_{\mathcal{H}^{(\zeta)}} = \langle f, g\rangle_{\mathcal{H}}.$$

Thus, $\{V_\zeta\varphi_k\}_{k\in\mathcal{I}}$ forms an orthonormal basis for $\mathcal{H}$, and so applying Parseval's identity together with the assumption that g is decreasing implies

$$\|V_\zeta Af - A_J f\|_{\mathcal{H}} \leq 2|g(\lambda_1)|^{2^J}\|f\|_{\mathcal{H}}.$$

Therefore, the result now follows from Lemma 3. $\square$

Appendix F. The proof of Theorem 6

Proof. Recall, from (27) the decomposition

$$H = \widetilde{H} + \overline{H}, \quad H' = \widetilde{H}' + \overline{H}'.$$

The operator $\widetilde{H}$ projects a function onto the zero eigenspace $\mathrm{span}(\varphi_0)$ and the operator $\overline{H}$ maps a function into its orthogonal complement $\mathrm{span}(\varphi_0)^\perp$. Therefore, we have $\widetilde{H}\overline{H} = \overline{H}\widetilde{H} = 0$, and we also have $\widetilde{H}^{2^j} = \widetilde{H}$ for all $j \geq 0$. Therefore,

$$H^{2^j} = \widetilde{H} + \overline{H}^{2^j},$$

which implies

$$H^{2^{j+1}} - H^{2^j} = \overline{H}^{2^{j+1}} - \overline{H}^{2^j}$$

(with similar equations holding for H' and $\overline{H}'$). Therefore,

$$\begin{aligned}&\|\mathcal{W}_J - \mathcal{W}'_J\|^2_{\ell^2(\mathcal{H})} \\ \leq&\|H^{2^J} - (H')^{2^J}\|^2_{\mathcal{H}} + \sum_{j=0}^{J-1}\|H^{2^j} - H^{2^{j+1}} - \left((H')^{2^j} - (H')^{2^{j+1}}\right)\|^2_{\mathcal{H}} + \|H - H'\|^2_{\mathcal{H}} \\ =&\|\widetilde{H} - \widetilde{H}' + (\overline{H})^{2^J} - (\overline{H})'^{2^J}\|^2_{\mathcal{H}} + \sum_{j=0}^{J-1}\|(\overline{H})^{2^j} - (\overline{H})^{2^{j+1}} - \left((\overline{H}')^{2^j} - (\overline{H}')^{2^{j+1}}\right)\|^2_{\mathcal{H}} \\ &+ \|\widetilde{H} - \widetilde{H}' + \overline{H} - \overline{H}'\|^2_{\mathcal{H}}\end{aligned}$$

$$\begin{aligned}
\leq & 4\|\widetilde{H}-\widetilde{H}'\|_{\mathcal{H}}^2+2\|(\overline{H})^{2^J}-(\overline{H}')^{2^J}\|_{\mathcal{H}}^2 \\
& +2\sum_{j=0}^{J-1}\|(\overline{H})^{2^j}-(\overline{H}')^{2^j}\|_{\mathcal{H}}^2+2\sum_{j=0}^{J-1}\|(\overline{H}')^{2^{j+1}}-(\overline{H})^{2^{j+1}}\|_{\mathcal{H}}^2+2\|\overline{H}-\overline{H}'\|_{\mathcal{H}}^2 \\
\leq & 4\left(\|\widetilde{H}-\widetilde{H}'\|_{\mathcal{H}}^2+\sum_{j=0}^{J}\|(\overline{H})^{2^j}-(\overline{H}')^{2^j}\|_{\mathcal{H}}^2\right). \qquad (53)
\end{aligned}$$

The following Lemma is a variant of Eq. (23) in [34] (see also Lemma L.1 of [64]).

Lemma 7. *Let* $\beta=\max\left\{\|\overline{H}\|_{\mathcal{H}},\|\overline{H}'\|_{\mathcal{H}}\right\}$ *and assume that* $\beta<1$. *Then*

$$\sum_{j=0}^{J}\left\|\overline{H}^{2^j}-(\overline{H}')^{2^j}\right\|_{\mathcal{H}}^2\leq C_0(\beta)\left\|\overline{H}-\overline{H}'\right\|_{\mathcal{H}}^2,$$

where $C_0(\beta):=\frac{\beta^2+1}{(1-\beta^2)^3}$.

Proof. Letting $A_j(t)=(t\overline{H}+(1-t)\overline{H}')^{2^j}$, we may check that

$$\left\|\overline{H}^{2^j}-(\overline{H}')^{2^j}\right\|_{\mathcal{H}}=\|A_j(1)-A_j(0)\|_{\mathcal{H}}\leq\int_0^1\|A_j'(t)\|_{\mathcal{H}}dt\leq\sup_{0\leq t\leq 1}\|A_j'(t)\|_{\mathcal{H}}dt.$$

Since,

$$A_j'(t)=\sum_{\ell=0}^{2^j-1}\left(t\overline{H}+(1-t)\overline{H}'\right)^{\ell}\left(\overline{H}-\overline{H}'\right)\left(t\overline{H}+(1-t)\overline{H}'\right)^{2^j-\ell-1},$$

and $\|\overline{H}\|_{\mathcal{H}},\|\overline{H}'\|_{\mathcal{H}}\leq\beta$, this implies

$$\|A_j'(t)\|_2\leq 2^j\beta^{2^j-1}\left\|\overline{H}-\overline{H}'\right\|_{\mathcal{H}}.$$

Therefore,

$$\sum_{j=0}^{J+1}\left\|\overline{H}^{2^j}-(\overline{H}')^{2^j}\right\|_{\mathcal{H}}^2\leq\sum_{j=0}^{\infty}(2^j\beta^{2^j-1})^2\left\|\overline{H}-\overline{H}'\right\|_{\mathcal{H}}^2=:C_0(\beta)\left\|\overline{H}-\overline{H}'\right\|_{\mathcal{H}}^2.$$

Lastly, one may compute

$$C_0(\beta)=\sum_{j=0}^{\infty}(2^j\beta^{2^j-1})^2=\beta^{-2}\sum_{j=0}^{\infty}(2^j\beta^{2^j})^2\leq\beta^{-2}\sum_{n=0}^{\infty}n^2\beta^{2n}=\beta^{-2}\frac{\beta^2(\beta^2+1)}{(1-\beta^2)^3}=\frac{\beta^2+1}{(1-\beta^2)^3},$$

where we used the Taylor expansion $\frac{x^2(x^2+1)}{(1-x^2)^3}=\sum_{n=0}^{\infty}n^2x^{2n}$. □

Returning to the proof of the theorem, we note that by the triangle inequality we have

$$\|\overline{H}-\overline{H}'\|_{\mathcal{H}}\leq\|H-H'\|_{\mathcal{H}}+\|\widetilde{H}-\widetilde{H}'\|_{\mathcal{H}}.$$

Therefore, combining (53) with Lemma 7, and using the fact that $(a+b)^2\leq 2(a^2+b^2)$ for all $a,b\in\mathbb{R}$, we have

$$\|\mathcal{W}_J-\mathcal{W}_J'\|_{\ell^2(\mathcal{H})}^2\leq C(\beta)\left(\|\widetilde{H}-\widetilde{H}'\|_{\mathcal{H}}^2+\|H-H'\|_{\mathcal{H}}^2\right), \qquad (54)$$

where $C(\beta)=CC_0(\beta)$ for some absolute constant C.

To estimate $\|\widetilde{H}-\widetilde{H}'\|_{\mathcal{H}}^2$, we note that for all $f\in\mathcal{H}$ we have

$$\begin{aligned}
\|\widetilde{H}f-\widetilde{H}'f\|_{\mathcal{H}} &= \|\langle f,\varphi_0\rangle_{\mathcal{H}}\varphi_0-\langle f,\varphi_0'\rangle_{\mathcal{H}'}\varphi_0'\|_{\mathcal{H}} \\
&\leq \|\langle f,\varphi_0-\varphi_0'\rangle_{\mathcal{H}}\varphi_0\|_{\mathcal{H}}+\|\langle f,\varphi_0'\rangle_{\mathcal{H}}(\varphi_0-\varphi_0')\|_{\mathcal{H}}+\|\,|\langle f,\varphi_0'\rangle_{\mathcal{H}}-\langle f,\varphi_0'\rangle_{\mathcal{H}'}|\varphi_0'\|_{\mathcal{H}} \\
&\leq \|\varphi_0-\varphi_0'\|_{\mathcal{H}}\|f\|_{\mathcal{H}}+\|\varphi_0-\varphi_0'\|_{\mathcal{H}}\|\varphi_0'\|_{\mathcal{H}}\|f\|_{\mathcal{H}}+|\langle f,\varphi_0'\rangle_{\mathcal{H}}-\langle f,\varphi_0'\rangle_{\mathcal{H}'}|\,\|\varphi_0'\|_{\mathcal{H}}.
\end{aligned}$$

By (23) and by (25)

$$\|\varphi_0'\|_{\mathcal{H}}\leq R(\mathcal{H},\mathcal{H}')^{1/2}\quad\text{and}\quad|\langle f,\varphi_0'\rangle_{\mathcal{H}}-\langle f,\varphi_0'\rangle_{\mathcal{H}'}|\leq\kappa(\mathcal{H},\mathcal{H}')R(\mathcal{H},\mathcal{H}')^{1/2}\|f\|_{\mathcal{H}}.$$

Therefore, we have

$$\|\widetilde{H} - \widetilde{H}'\|_{\mathcal{H}} \leq \|\varphi_0 - \varphi_0'\|_{\mathcal{H}}(1 + R(\mathcal{H}, \mathcal{H}')^{1/2}) + R(\mathcal{H}, \mathcal{H}')\kappa(\mathcal{H}, \mathcal{H}').$$

Thus,

$$\|\mathcal{W}_J - \mathcal{W}_J'\|^2_{\ell^2(\mathcal{H})} \leq C(\beta)\left[\|\varphi_0 - \varphi_0'\|^2_{\mathcal{H}} R(\mathcal{H}, \mathcal{H}') + R(\mathcal{H}, \mathcal{H}')^2\kappa(\mathcal{H}, \mathcal{H}')^2 + \|H - H'\|^2_{\mathcal{H}}\right]$$

as desired. □

Appendix G. The proof of Theorem 7

In order to prove Theorem 7, we will need the following Lemma.

Lemma 8. *Under the assumptions of Theorem 7, we have*

$$\left\|S^\ell f - (S^\ell)' f\right\|_{\ell^2(\mathcal{H})} \leq \sqrt{2}\|\mathcal{W} - \mathcal{W}'\|_{\mathcal{H}}\left(\sum_{k=0}^{\ell}\|\mathcal{W}'\|^k_{\mathcal{H}}\right)\|f\|_{\mathcal{H}} \quad \text{for all } f \in \mathcal{H}. \tag{55}$$

Before proving Lemma 8, we will show how it is used to prove Theorem 7.

Proof of Theorem 7. Let $\zeta \in \mathcal{G}$, and let $\mathcal{X}_\zeta$ be defined as in (16). By (31), we have $V_\zeta S^\ell f = S^{\ell,(\zeta)} V_\zeta f$. Therefore, the triangle inequality implies

$$\left\|S^\ell f - (S^\ell)'\widetilde{f}\right\|_{\ell^2(\mathcal{H})} \leq \left\|S^\ell f - V_\zeta S^\ell f\right\|_{\ell^2(\mathcal{H})} + \|S^{\ell,(\zeta)} V_\zeta f - S^{\ell,(\zeta)}\widetilde{f}\|_{\ell^2(\mathcal{H})} + \|S^{\ell,(\zeta)}\widetilde{f} - (S^\ell)'\widetilde{f}\|_{\ell^2(\mathcal{H})}. \tag{56}$$

The assumption (31) also implies that

$$\|S^\ell f - V_\zeta S^\ell f\|_{\ell^2(\mathcal{H})} \leq \mathcal{B}\|f\|_{\mathcal{H}}. \tag{57}$$

Similarly, by Theorem 1 and (23), we have that

$$\begin{aligned}\|S^{\ell,(\zeta)} V_\zeta f - S^{\ell,(\zeta)}\widetilde{f}\|_{\ell^2(\mathcal{H})} &\leq R\left(\mathcal{H}, \mathcal{H}^{(\zeta)}\right)^{1/2}\|S^{\ell,(\zeta)} V_\zeta f - S^{\ell,(\zeta)}\widetilde{f}\|_{\ell^2(\mathcal{H}^{(\zeta)})} \\ &\leq R\left(\mathcal{H}, \mathcal{H}^{(\zeta)}\right)^{1/2}\|V_\zeta f - \widetilde{f}\|_{\mathcal{H}^{(\zeta)}} \\ &\leq R\left(\mathcal{H}, \mathcal{H}^{(\zeta)}\right)\|V_\zeta f - \widetilde{f}\|_{\mathcal{H}}.\end{aligned} \tag{58}$$

Applying Lemma 8 and (23) yields

$$\begin{aligned}\|S^{\ell,(\zeta)}\widetilde{f} - (S^\ell)'\widetilde{f}\|_{\ell^2(\mathcal{H})} &\leq R\left(\mathcal{H}, \mathcal{H}^{(\zeta)}\right)^{1/2}\|S^{\ell,(\zeta)}\widetilde{f} - (S^\ell)'\widetilde{f}\|_{\ell^2(\mathcal{H}^{(\zeta)})} \\ &\leq \sqrt{2}R\left(\mathcal{H}, \mathcal{H}^{(\zeta)}\right)^{1/2}\|\mathcal{W}^{(\zeta)} - \mathcal{W}'\|_{\mathcal{H}^{(\zeta)}}\left(\sum_{k=0}^{\ell}\|\mathcal{W}'\|^k_{\mathcal{H}^{(\zeta)}}\right)\|\widetilde{f}\|_{\mathcal{H}^{(\zeta)}} \\ &\leq \sqrt{2}R\left(\mathcal{H}, \mathcal{H}^{(\zeta)}\right)\|\mathcal{W}^{(\zeta)} - \mathcal{W}'\|_{\mathcal{H}^{(\zeta)}}\left(\sum_{k=0}^{\ell}\|\mathcal{W}'\|^k_{\mathcal{H}^{(\zeta)}}\right)\|\widetilde{f}\|_{\mathcal{H}}.\end{aligned}$$

Thus, infimizing over ζ completes the proof. □

The Proof of Lemma 8. Let $\mathcal{A} := \|\mathcal{W} - \mathcal{W}'\|_{\mathcal{H}}$ and $C := \|\mathcal{W}'\|_{\mathcal{H}}$.

To prove (55), we need to show

$$\sum_{p\in\mathcal{J}^\ell}\|S[p]f - S'[p]f\|^2_{\mathcal{H}} \leq 2\mathcal{A}^2 \cdot \left(\sum_{k=0}^{\ell} C^k\right)^2\|f\|^2_{\mathcal{H}}. \tag{59}$$

For $\ell = 0$, we recall from (12) that the zeroth-order windowed scattering coefficient of f is given by $S[p_e]f = Af$, where p_e is the empty-index. Therefore, by the definition of $\mathcal{A}$ we have

$$\sum_{p\in\mathcal{J}^0}\|S[p]f - S'[p]f\|^2_{\mathcal{H}} = \|Af - A'f\|^2_{\mathcal{H}} \leq \|\mathcal{W}f - \mathcal{W}'f\|^2_{\mathcal{H}} \leq \mathcal{A}^2\|f\|^2_{\mathcal{H}},$$

and so (59) holds when $\ell = 0$. For the case where $\ell \geq 1$, we note that for all $p \in \mathcal{J}^\ell$, we have

$$\begin{aligned}\|S[p]f - S'[p]f\|_{\mathcal{H}} &= \|AU[p]f - A'U'[p]f\|_{\mathcal{H}} \\ &\leq \|(A-A')U[p]f\|_{\mathcal{H}} + \|A'U[p]f - A'U'[p]f\|_{\mathcal{H}} \\ &\leq \|A-A'\|_{\mathcal{H}}\|U[p]f\|_{\mathcal{H}} + \|A'\|_{\mathcal{H}}\|U[p]f - U'[p]f\|_{\mathcal{H}},\end{aligned}$$

and so using the fact that $(a+b)^2 \leq 2a^2 + 2b^2$ and summing over p implies

$$\sum_{p\in\mathcal{J}^\ell} \|S[p]f - S'[p]f\|_{\mathcal{H}}^2 \leq 2\|A-A'\|_{\mathcal{H}}^2 \sum_{p\in\mathcal{J}^\ell} \|U[p]f\|_{\mathcal{H}}^2 + 2\|A'\|_{\mathcal{H}}^2 \sum_{p\in\mathcal{J}^\ell} \|U[p]f - U'[p]f\|_{\mathcal{H}}^2.$$

Therefore, (59) and thus (55), follow from applying Lemma 9 stated below, noting that $\|A-A'\|_{\mathcal{H}}^2 \leq \mathcal{A}^2$ and $\|A'\|_{\mathcal{H}}^2 \leq C^2$, and using the fact that $a^2 + b^2 \leq (a+b)^2$ when $a, b \geq 0$. □

Lemma 9. *Let $\mathcal{A} := \|\mathcal{W} - \mathcal{W}'\|_{\mathcal{H}}$ and $C := \|\mathcal{W}'\|_{\mathcal{H}}$. Then, for all $\ell \geq 1$,*

$$\sum_{p\in\mathcal{J}^\ell} \|U[p]f\|_{\mathcal{H}}^2 \leq \|f\|_{\mathcal{H}}^2, \quad \text{and} \quad \sum_{p\in\mathcal{J}^\ell} \|U[p]f - U'[p]f\|_{\mathcal{H}}^2 \leq \mathcal{A}^2 \left(\sum_{k=0}^{\ell-1} C^k\right)^2 \|f\|_{\mathcal{H}}^2.$$

Proof. When $\ell = 1$, the first inequality follows immediately from (11) and the fact that σ is nonexpansive. Now, suppose by induction that the first inequality holds for ℓ. Let $f \in \mathcal{H}$. Then

$$\begin{aligned}\sum_{p\in\mathcal{J}^{\ell+1}} \|U[p]f\|_{\mathcal{H}}^2 &= \sum_{p\in\mathcal{J}^{\ell+1}} \|\sigma W_{j_{\ell+1}} \cdots \sigma W_{j_1} f\|_{\mathcal{H}}^2 \\ &= \sum_{p\in\mathcal{J}^{\ell}} \left(\sum_{j_{\ell+1}\in\mathcal{J}} \|\sigma W_{j_{\ell+1}} (\sigma W_{j_\ell} \cdots \sigma W_{j_1} f)\|_{\mathcal{H}}^2 \right) \\ &\leq \sum_{p\in\mathcal{J}^{\ell}} \|\sigma W_{j_\ell} \cdots \sigma W_{j_1} f\|_{\mathcal{H}}^2 \\ &\leq \|f\|_{\mathcal{H}}^2, \end{aligned} \tag{60}$$

with the last inequality following from the inductive assumption.

To prove the second inequality, let $t_\ell := \left(\sum_{p\in\mathcal{J}^\ell} \|U[p]f - U'[p]f\|_{\mathcal{H}}^2\right)^{1/2}$. Since σ is nonexpansive, the definition of $\mathcal{A}$ implies $t_1 \leq \mathcal{A}\|f\|_{\mathcal{H}}$. Now, by induction, suppose the result holds for ℓ. Then, recalling that $U[p] = \sigma W_{j_\ell} \cdots \sigma W_{j_1}$, we have

$$\begin{aligned} t_{\ell+1} &= \left(\sum_{p\in\mathcal{J}^{\ell+1}} \|\sigma W_{j_{\ell+1}} \cdots \sigma W_{j_1} f - \sigma W'_{j_{\ell+1}} \cdots \sigma W'_{j_1} f\|_{\mathcal{H}}^2 \right)^{1/2} \\ &\leq \left(\sum_{p\in\mathcal{J}^{\ell+1}} \|(W_{j_{\ell+1}} - W'_{j_{\ell+1}})\sigma W_{j_\ell} \cdots \sigma W_{j_1} f\|_{\mathcal{H}}^2 \right)^{1/2} \\ &\quad + \left(\sum_{p\in\mathcal{J}^{\ell+1}} \|W'_{j_{\ell+1}} (\sigma W_{j_i} \cdots \sigma W_{j_1} f - \sigma W'_{j_\ell} \cdots \sigma W'_{j_1} f)\|_{\mathcal{H}}^2 \right)^{1/2} \\ &\leq \mathcal{A} \left(\sum_{p\in\mathcal{J}^{\ell+1}} \|\sigma W_{j_\ell} \cdots \sigma W_{j_1} f\|_{\mathcal{H}}^2 \right)^{1/2} \\ &\quad + C \left(\sum_{p\in\mathcal{J}^{\ell}} \|\sigma W_{j_\ell} \cdots \sigma W_{j_1} f - \sigma W'_{j_\ell} \cdots \sigma W'_{j_1} f\|_{\mathcal{H}}^2 \right)^{1/2} \\ &\leq \mathcal{A}\|f\|_{\mathcal{H}} + t_\ell C \|f\|_{\mathcal{H}} \end{aligned}$$

by the definitions of $\mathcal{A}$ and C and by (60). By the inductive hypothesis, we have that

$$t_\ell \leq \mathcal{A} \sum_{k=0}^{\ell-1} C^k \|f\|_{\mathcal{H}}.$$

Therefore,

$$t_{\ell+1} \leq \mathcal{A}\|f\|_{\mathcal{H}} + \mathcal{A}\sum_{k=0}^{\ell-1} C^{k+1}\|f\|_{\mathcal{H}} = \mathcal{A}\sum_{k=0}^{\ell} C^k \|f\|_{\mathcal{H}}.$$

Squaring both sides completes the proof of the second inequality. □

Appendix H. The proof of Theorem 8

Proof. Let $\zeta \in \mathcal{G}$, and let $\mathcal{X}_\zeta$ be as in (16). By (33), we have $\overline{S^{(\zeta)}}V_\zeta f = \overline{S}f$. Therefore, we may use the definitions of the non-windowed scattering transform to see that for each path p we have

$$\begin{aligned}
&|\overline{S}[p]f - \overline{S'}[p]\widetilde{f}| \\
=&|\overline{S^{(\zeta)}}[p]V_\zeta f - \overline{S'}[p]\widetilde{f}| \\
\leq&|\langle U^{(\zeta)}[p]V_\zeta f, \varphi_0^{(\zeta)}\rangle_{\mathcal{H}^{(\zeta)}} - \langle U'[p]\widetilde{f}, \varphi_0'\rangle_{\mathcal{H}'}| \\
\leq&|\langle U^{(\zeta)}[p]V_\zeta f - U'[p]\widetilde{f}, \varphi_0^{(\zeta)}\rangle_{\mathcal{H}^{(\zeta)}}| + |\langle U'[p]\widetilde{f}, \varphi_0^{(\zeta)} - \varphi_0'\rangle_{\mathcal{H}^{(\zeta)}}| + |\langle U'[p]\widetilde{f}, \varphi_0'\rangle_{\mathcal{H}^{(\zeta)}} - \langle U'[p]\widetilde{f}, \varphi_0'\rangle_{\mathcal{H}'}| \\
=:& I[p] + II[p] + III[p].
\end{aligned}$$

To bound $I[p]$, we use the Cauchy Schwarz inequality to observe

$$\begin{aligned}
|\langle U^{(\zeta)}[p]V_\zeta f - U'[p]\widetilde{f}, \varphi_0^{(\zeta)}\rangle_{\mathcal{H}^{(\zeta)}}| &\leq |\langle U^{(\zeta)}[p]V_\zeta f - U^{(\zeta)}[p]\widetilde{f}, \varphi_0^{(\zeta)}\rangle_{\mathcal{H}^{(\zeta)}}| + |\langle U^{(\zeta)}[p]\widetilde{f} - U'[p]\widetilde{f}, \varphi_0^{(\zeta)}\rangle_{\mathcal{H}^{(\zeta)}}| \\
&\leq |\overline{S^{(\zeta)}}[p]V_\zeta f - \overline{S^{(\zeta)}}[p]\widetilde{f}| + \|U^{(\zeta)}[p]\widetilde{f} - U'[p]\widetilde{f}\|_{\mathcal{H}^{(\zeta)}}.
\end{aligned}$$

Therefore, applying (33) and Lemma 9 yields

$$\sum_{p\in\mathcal{J}^\ell} I[p]^2 \leq 2C_L\|V_\zeta f - \tilde{f}\|_{\mathcal{H}^{(\zeta)}}^2 + 2\|\mathcal{W}^{(\zeta)} - \mathcal{W}'\|_{\mathcal{H}^{(\zeta)}}^2 \left(\sum_{k=0}^{\ell-1}\|\mathcal{W}'\|_{\mathcal{H}^{(\zeta)}}^k\right)^2 \|\tilde{f}\|_{\mathcal{H}^{(\zeta)}}^2. \tag{61}$$

For $II[p]$, we again use the Cauchy Schwarz inequality and (23) to see

$$|\langle U'[p]\widetilde{f}, \varphi_0^{(\zeta)} - \varphi_0'\rangle_{\mathcal{H}^{(\zeta)}}| \leq R(\mathcal{H}^{(\zeta)}, \mathcal{H}')\|\varphi_0^{(\zeta)} - \varphi_0'\|_{\mathcal{H}'}\|U'[p]\widetilde{f}\|_{\mathcal{H}'}$$

Therefore, again applying Lemma 9 implies

$$\sum_{p\in\mathcal{J}^\ell} II[p]^2 \leq R(\mathcal{H}^{(\zeta)}, \mathcal{H}')^2\|\varphi_0^{(\zeta)} - \varphi_0'\|_{\mathcal{H}'}^2\|\widetilde{f}\|_{\mathcal{H}'}^2. \tag{62}$$

Lastly, to bound $III[p]$, we note that by (25) and the Cauchy Schwarz inequality, we have

$$\begin{aligned}
|\langle U'[p]\widetilde{f}, \varphi_0'\rangle_{\mathcal{H}^{(\zeta)}} - \langle U'[p]\widetilde{f}, \varphi_0'\rangle_{\mathcal{H}'}| &\leq \kappa(\mathcal{H}', \mathcal{H}^{(\zeta)})\|U'[p]\widetilde{f}\|_{\mathcal{H}'}\|\varphi_0'\|_{\mathcal{H}'} \\
&\leq \kappa(\mathcal{H}', \mathcal{H}^{(\zeta)})\|U'[p]\widetilde{f}\|_{\mathcal{H}'},
\end{aligned}$$

and so summing over p and once more applying Lemma 9 gives

$$\sum_{p\in\mathcal{J}^\ell} III[p]^2 \leq \kappa(\mathcal{H}', \mathcal{H}^{(\zeta)})\|\widetilde{f}\|_{\mathcal{H}'}.$$

Therefore, combining this with (61) and (62) yields

$$\begin{aligned}
&\sum_{p\in\mathcal{J}^\ell} |\overline{S}[p]f - \overline{S}'[p]\widetilde{f}|^2 \\
\leq&3\Bigg(2C_L\|V_\zeta f - \tilde{f}\|_{\mathcal{H}^{(\zeta)}}^2 + R(\mathcal{H}^{(\zeta)}, \mathcal{H}')^2\|\varphi_0^{(\zeta)} - \varphi_0'\|_{\mathcal{H}'}^2\|\widetilde{f}\|_{\mathcal{H}'}^2 \\
&\quad + 2\|\mathcal{W}^{(\zeta)} - \mathcal{W}'\|_{\mathcal{H}^{(\zeta)}}^2\left(\sum_{k=0}^{\ell-1}\|\mathcal{W}'\|_{\mathcal{H}^{(\zeta)}}^k\right)^2\|\tilde{f}\|_{\mathcal{H}^{(\zeta)}}^2 + \kappa(\mathcal{H}', \mathcal{H}^{(\zeta)})\|\widetilde{f}\|_{\mathcal{H}'}\Bigg).
\end{aligned}$$

The result follows by taking the infimum over $\zeta \in \mathcal{G}$. □

Appendix I. The proof of Lemma 4

Proof. By definition, we have

$$H_t f(x) - H_t^\kappa f(x)$$

$$
\begin{aligned}
&= \int_X h_t(x,y) f(y) d\mu(y) - \int_X h_t^\kappa(x,y) f(y) d\mu(y) \\
&= \int_X \sum_{k=0}^{\infty} e^{-t\mu_k} \varphi_k(x)\varphi_k(y) f(y) d\mu(y) - \int_X \sum_{k=0}^{\kappa} e^{-t\mu_k} \varphi_k(x)\varphi_k(y) f(y) d\mu(y) \\
&= \int_X \sum_{k=\kappa+1}^{\infty} e^{-t\mu_k} \varphi_k(x)\varphi_k(y) f(y) d\mu(y) \\
&= \sum_{k=\kappa+1}^{\infty} e^{-t\mu_k} \langle \varphi_k, f \rangle_{\mathbf{L}^2(\mathcal{X})} \varphi_k(x).
\end{aligned} \tag{63}
$$

Therefore, since the φ_k form an orthonormal basis, twice applying Plancherel's theorem implies that

$$
\begin{aligned}
\| H_t^\kappa f(x) - H_t f(x) \|_{\mathbf{L}^2(\mathcal{X})}^2 &= \sum_{k=\kappa+1}^{\infty} e^{-2t\mu_k} |\langle \varphi_k, f \rangle_{\mathbf{L}^2(\mathcal{X})}|^2 \\
&\leq e^{-2t\mu_{\kappa+1}} \sum_{k=\kappa+1}^{\infty} |\langle \varphi_k, f \rangle_{\mathbf{L}^2(\mathcal{X})}|^2 \\
&\leq e^{-2t\mu_{\kappa+1}} \| f \|_{\mathbf{L}^2(\mathcal{X})}^2.
\end{aligned}
$$

This completes the proof of (40). To prove (41), we note that (63) implies

$$
\| H_t^\kappa f - H_t f \|_\infty \leq \sup_{x,y \in \mathcal{X}} \left| \sum_{k=\kappa+1}^{\infty} e^{-t\mu_k} \varphi_k(x)\varphi_k(y) \right| \| f \|_\infty.
$$

In [27], the proof of Theorem 3, it is shown that

$$
\sup_{x,y \in \mathcal{X}} \left| \sum_{k=\kappa+1}^{\infty} e^{-t\mu_k} \varphi_k(x)\varphi_k(y) \right| \leq C_{\mathcal{X}} e^{-C'_{\mathcal{X}} t} \leq C_{\mathcal{X}}
$$

and so the result follows. □

Appendix J. The proof of Lemma 5

Proof. Let $f, g \in C(\mathcal{X})$, and define random variables $X_i = f(x_i) g(x_i)$. Since the x_i are sampled i.i.d. uniformly at random, we have

$$
\langle \rho f, \rho g \rangle_2 = \frac{1}{N} \sum_{i=0}^{N-1} X_i
$$

and (35) implies

$$
\mathbb{E}\left(\frac{1}{N} \sum_{i=0}^{N-1} X_i \right) = \langle f, g \rangle_{\mathbf{L}^2(\mathcal{X})}.
$$

Therefore, by Hoeffding's inequality, we have

$$
\begin{aligned}
\mathbb{P}\left(|\langle \rho f, \rho g \rangle_2 - \langle f, g \rangle_{\mathbf{L}^2(\mathcal{X})}| > \eta \right) &= \mathbb{P}\left(\left| \frac{1}{N} \left(\sum_{i=0}^{N-1} X_i - \mathbb{E} \sum_{i=0}^{N-1} X_i \right) \right| > \eta \right) \\
&= \mathbb{P}\left(\left| \left(\sum_{i=0}^{N-1} X_i - \mathbb{E} \sum_{i=0}^{N-1} X_i \right) \right| > N\eta \right) \\
&\leq 2 \exp\left(\frac{-2N^2\eta^2}{4N \| fg \|_\infty^2} \right) \\
&= 2 \exp\left(\frac{-N\eta^2}{2 \| fg \|_\infty^2} \right).
\end{aligned}
$$

The result now follows by setting $\eta = \sqrt{\frac{18 \log N}{N}} \| fg \|_\infty$. □

Appendix K. The proof of Remark 8

To see this we note that the term α_k in Theorem 5.4 of [15] is first introduced in Proposition 5.2. We observe, by Equation ((42)), that

$$\left|\left|(\mathbf{u}_k^{N,\epsilon})^T \mathbf{v}_k\right| - 1\right| = \left|\frac{1}{|\alpha_k|} - 1\right| = \mathcal{O}(|\mathrm{Err}_{\mathrm{norm}}| + \mathrm{Err}_{\mathrm{pt}}^2).$$

(Please see [15] for the definitions of $\mathrm{Err}_{\mathrm{norm}}$ and $\mathrm{Err}_{\mathrm{pt}}^2$.) Since $|\alpha_k|$ converges to 1, for sufficiently large N, we have $\frac{1}{2}||\alpha_k| - 1| \leq \left|\frac{1}{|\alpha_k|} - 1\right| \leq 2||\alpha_k| - 1|$ and therefore, we also have that

$$||\alpha_k| - 1| = \mathcal{O}(|\mathrm{Err}_{\mathrm{norm}}| + \mathrm{Err}_{\mathrm{pt}}^2).$$

Immediately prior to Equation (42), the authors note

$$\mathrm{Err}_{\mathrm{norm}} = \mathcal{O}\left(\sqrt{\frac{\log(N)}{N}}\right),$$

and Equation (40) shows that

$$\mathrm{Err}_{\mathrm{pt}} = \mathcal{O}(\epsilon) + \mathcal{O}\left(\sqrt{\frac{\log(N)}{N\epsilon^{d/2+1}}}\right)$$

In particular, if we set $\epsilon \sim N^{-2/(d+6)}$ we have

$$\mathrm{Err}_{\mathrm{pt}} = \mathcal{O}(N^{-2/(d+6)}) + \mathcal{O}\left(\sqrt{\frac{\log(N)}{N^{4/(d+6)}}}\right) = \mathcal{O}\left(\sqrt{\frac{\log(N)}{N^{4/(d+6)}}}\right)$$

Appendix L. The proof of Theorem 10 and Corollary 1

Proof of Theorem 10. To avoid cumbersome notation, within this proof we will drop explicit dependence on N and ϵ and simply write λ_k in place of $\lambda_k^{N,\epsilon}$.

Let $\tilde{\mathbf{u}}_k = \mathrm{sgn}(\alpha_k)\mathbf{u}_k$ where sgn is the standard signum function. Then,

$$\begin{aligned}
&H_{N,\epsilon,\kappa,t}\rho f - \rho H_t^\kappa f \\
&= \sum_{k=0}^{\kappa} e^{-\lambda_k t}\mathbf{u}_k\mathbf{u}_k^T\rho f - \rho\sum_{k=0}^{\kappa} e^{-\mu_k t}\langle f,\varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}\varphi_k \\
&= \sum_{k=0}^{\kappa} e^{-\lambda_k t}\tilde{\mathbf{u}}_k\tilde{\mathbf{u}}_k^T\rho f - \rho\sum_{k=0}^{\kappa} e^{-\mu_k t}\langle f,\varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}\varphi_k \\
&= \sum_{k=0}^{\kappa} e^{-\lambda_k t}\langle\tilde{\mathbf{u}}_k,\rho f\rangle_2\tilde{\mathbf{u}}_k - \sum_{k=0}^{\kappa} e^{-\mu_k t}\langle f,\varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}\mathbf{v}_k \\
&= \sum_{k=0}^{\kappa}(e^{-\lambda_k t} - e^{-\mu_k t})\langle\tilde{\mathbf{u}}_k,\rho f\rangle_2\tilde{\mathbf{u}}_k \\
&\qquad + \sum_{k=0}^{\kappa} e^{-\mu_k t}\left(\langle\tilde{\mathbf{u}}_k,\rho f\rangle_2 - \langle f,\varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}\right)\tilde{\mathbf{u}}_k \\
&\qquad\quad + \sum_{k=0}^{\kappa} e^{-\mu_k t}\langle f,\varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}(\tilde{\mathbf{u}}_k - \mathbf{v}_k).
\end{aligned} \tag{64}$$

Since $|\mathrm{sgn}(\alpha_k)| = 1$, $\{\widetilde{\mathbf{u}}_k\}_{k=0}^{\kappa}$ is an orthonormal basis for the span of $\{\mathbf{u}_k\}_{k=0}^{\kappa}$. Therefore, to bound the first of the above terms, we may apply Parseval's theorem to see

$$\begin{aligned}
&\|\sum_{k=0}^{\kappa}(e^{-\lambda_k t} - e^{-\mu_k t})\langle\tilde{\mathbf{u}}_k,\rho f\rangle_2\tilde{\mathbf{u}}_k\|_2^2 \\
&= \sum_{k=0}^{\kappa}|e^{-\lambda_k t} - e^{-\mu_k t}|^2|\langle\tilde{\mathbf{u}}_k,\rho f\rangle_2|^2 \\
&\leq \max_{0\leq k\leq\kappa}|e^{-\lambda_k t} - e^{-\mu_k t}|^2\sum_{k=0}^{\kappa}|\langle\tilde{\mathbf{u}}_k,\rho f\rangle_2|^2
\end{aligned}$$

$$\leq \max_{0\leq k\leq \kappa} |e^{-\lambda_k t} - e^{-\mu_k t}|^2 \|\rho f\|_2^2. \tag{65}$$

By Theorem 9, we have

$$\begin{aligned}\max_{0\leq k\leq \kappa} |e^{-\lambda_k t} - e^{-\mu_k t}| &\leq t \max_{0\leq k\leq \kappa} |\lambda_k - \mu_k| \\ &= t\mathcal{O}(N^{-2/(d+6)})\end{aligned} \tag{66}$$

with probability at least $1 - \mathcal{O}(N^{-9})$.

By Lemma 5 we have

$$\|\rho f\|_2^2 \leq \|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \sqrt{\frac{18\log N}{N}}\|f\|_\infty^2$$

with probability at least $1 - 2/N^9$. Therefore, combining (65) and (66), yields

$$\begin{aligned}&\left\|\sum_{k=0}^{\kappa}(e^{-\lambda_k t} - e^{-\mu_k t})\langle \tilde{\mathbf{u}}_k, \rho f\rangle_2 \tilde{\mathbf{u}}_k\right\|_2^2 \\ &\leq \max_{0\leq k\leq \kappa} |e^{-\lambda_k t} - e^{-\mu_k t}|^2 \|\rho f\|_2^2 \\ &\leq t^2\left(\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \sqrt{\frac{\log N}{N}}\|f\|_\infty^2\right)\mathcal{O}(N^{-4/(d+6)})\end{aligned} \tag{67}$$

with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$.

To bound the second term from (64), we use Parseval's Identity to see

$$\begin{aligned}&\left\|\sum_{k=0}^{\kappa} e^{-\mu_k t}\left(\langle \tilde{\mathbf{u}}_k, \rho f\rangle_2 - \langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}\right)\tilde{\mathbf{u}}_k\right\|_2^2 \\ &\leq \sum_{k=0}^{\kappa} |\langle \tilde{\mathbf{u}}_k, \rho f\rangle_2 - \langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}|^2 \\ &\leq 2\sum_{k=0}^{\kappa}(|\langle \tilde{\mathbf{u}}_k, \rho f\rangle_2 - \langle \mathbf{v}_k, \rho f\rangle_2|^2 + |\langle \mathbf{v}_k, \rho f\rangle_2 - \langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}|^2) \\ &\leq 2\sum_{k=0}^{\kappa}(\|\tilde{\mathbf{u}}_k - \mathbf{v}_k\|_2^2\|\rho f\|_2^2 + |\langle \rho\varphi_k, \rho f\rangle_2 - \langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}|^2).\end{aligned} \tag{68}$$

By Remark 8,

$$\max\left\{||\alpha_k| - 1|, \left|\frac{1}{|\alpha_k|} - 1\right|\right\} \leq \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right) + \mathcal{O}\left(\frac{\log(N)}{N^{4/(d+6)}}\right).$$

Therefore,

$$\left(\frac{|\alpha_k| - 1}{\alpha_k}\right)^2 \leq \mathcal{O}\left(\frac{\log N}{N}\right) + \mathcal{O}\left(\frac{\log(N)^2}{N^{8/(d+6)}}\right),$$

and so we may recall the definition of $\tilde{\mathbf{u}}_k$, and use Theorem 9 to see

$$\begin{aligned}\|\tilde{\mathbf{u}}_k - \mathbf{v}_k\|_2^2 &= \|\mathrm{sgn}(\alpha_k)\mathbf{u}_k - \mathbf{v}_k\|_2^2 \\ &= \frac{1}{\alpha_k^2}\|\,|\alpha_k|\mathbf{u}_k - \alpha_k\mathbf{v}_k\|_2^2 \\ &\leq \frac{2}{\alpha_k^2}(\|(|\alpha_k| - 1)\mathbf{u}_k\|^2 + \|\mathbf{u}_k - \alpha_k\mathbf{v}_k\|_2^2) \\ &\leq 2\left(\frac{|\alpha_k| - 1}{\alpha_k}\right)^2 + \frac{2}{|\alpha_k|^2}\|\mathbf{u}_k - \alpha_k\mathbf{v}_k\|_2^2 \\ &\leq \mathcal{O}\left(\frac{\log N}{N}\right) + \mathcal{O}\left(\frac{\log(N)^2}{N^{8/(d+6)}}\right) + \mathcal{O}(N^{-\frac{4}{d+6}}\log(N)) \\ &= \mathcal{O}(N^{-\frac{4}{d+6}}\log N).\end{aligned} \tag{69}$$

As noted earlier, by Lemma 5, we have

$$\|\rho f\|_2^2 \leq \|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \sqrt{\frac{18\log N}{N}}\|f\|_\infty^2$$

with probability at least $1 - \frac{2}{N^9}$ and again applying Lemma 5 we have

$$|\langle \rho\varphi_k, \rho f\rangle_2 - \langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}| \leq \sqrt{\frac{18\log N}{N}}\|f\varphi_k\|_\infty$$

with probability at least $1 - \frac{2}{N^9}$.

It is known (see, e.g., [70]) that $\|\varphi_k\|_\infty \leq C_{\mathcal{X}}\lambda_k^{(d-1)/4}$. Weyl's asymptotic formula (see, e.g., [13] Theorem 72) implies that $\lambda_k \leq C_{\mathcal{X}}k^{2/d}$. Therefore,

$$\|\varphi_k\|_\infty \leq C_{\mathcal{X}}k^{\frac{2}{d}\frac{d-1}{4}} = C_{\mathcal{X}}k^{(d-1)/2d} = \mathcal{O}(1),$$

where the final equality uses the fact that the implied constants depend on κ and the geometry of $\mathcal{X}$. Therefore, by (68),

$$\begin{aligned}
&\|\sum_{k=0}^{\kappa} e^{-\mu_k t}\left(\langle\tilde{\mathbf{u}}_k, \rho f\rangle_2 - \langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}\right)\tilde{\mathbf{u}}_k\|_2^2 \\
&\leq 2\sum_{k=0}^{\kappa}(\|\tilde{\mathbf{u}}_k - \mathbf{v}_k\|_2^2\|\rho f\|_2^2 + |\langle\rho\varphi_k, \rho f\rangle_2 - \langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}|^2) \\
&\leq \kappa\left(\mathcal{O}(N^{-\frac{4}{d+6}}\log N)\left(\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \sqrt{\frac{18\log N}{N}}\|f\|_\infty^2\right) + \mathcal{O}\left(\frac{\log N}{N}\right)\max_{0\leq k\leq\kappa}\|f\varphi_k\|_\infty^2\right) \\
&\leq \kappa\left(\mathcal{O}(N^{-\frac{4}{d+6}}\log N)\left(\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \sqrt{\frac{\log N}{N}}\|f\|_\infty^2\right) + \mathcal{O}\left(\frac{\log N}{N}\right)\kappa^{(d-1)/d}\|f\|_\infty^2\right) \\
&\leq \mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right)\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \left(\mathcal{O}\left(\frac{(\log N)^{3/2}}{N^{\frac{4}{d+6}+\frac{1}{2}}}\right) + \mathcal{O}\left(\frac{\log N}{N}\right)\right)\|f\|_\infty^2.
\end{aligned} \tag{70}$$

Finally, to bound the third term in (64), we use (69) to see

$$\begin{aligned}
&\|\sum_{k=0}^{\kappa} e^{-\mu_k t}\langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}(\tilde{\mathbf{u}}_k - \mathbf{v}_k)\|_2^2 \\
&\leq \kappa\sum_{k=0}^{\kappa}|\langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}|^2\|\tilde{\mathbf{u}}_k - \mathbf{v}_k\|_2^2 \\
&\leq \kappa\max_{0\leq k\leq\kappa}\|\tilde{\mathbf{u}}_k - \mathbf{v}_k\|_2^2\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 \\
&\leq \mathcal{O}(N^{-\frac{4}{d+6}}\log(N))\|f\|_{\mathbf{L}^2(\mathcal{X})}^2.
\end{aligned} \tag{71}$$

Combining (67), (70), and (71) with (64) implies that in the case $d \geq 2$ we have

$$\begin{aligned}
&\|H_{N,\epsilon,\kappa,t}\rho f - \rho H_t^\kappa f\|_2^2 \\
&\leq 3\|\sum_{k=0}^{\kappa}(e^{-\lambda_k^{N,\epsilon}t} - e^{-\mu_k t})\langle\tilde{\mathbf{u}}_k, \rho f\rangle_2\tilde{\mathbf{u}}_k\|_2^2 \\
&\quad + 3\|\sum_{k=0}^{\kappa} e^{-\mu_k t}\left(\langle\tilde{\mathbf{u}}_k, \rho f\rangle_2 - \langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}\right)\tilde{\mathbf{u}}_k\|_2^2 \\
&\quad + 3\|\sum_{k=0}^{\kappa} e^{-\mu_k t}\langle f, \varphi_k\rangle_{\mathbf{L}^2(\mathcal{X})}(\tilde{\mathbf{u}}_k - \mathbf{v}_k)\|_2^2 \\
&\leq t^2\left(\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \sqrt{\frac{\log N}{N}}\|f\|_\infty^2\right)\mathcal{O}(N^{-4/(d+6)}) \\
&+ \mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right)\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \left(\mathcal{O}\left(\frac{(\log N)^{3/2}}{N^{\frac{4}{d+6}+\frac{1}{2}}}\right) + \mathcal{O}\left(\frac{\log N}{N}\right)\right)\|f\|_\infty^2 \\
&+ \mathcal{O}(N^{-\frac{4}{d+6}}\log(N))\|f\|_{\mathbf{L}^2(\mathcal{X})}^2
\end{aligned}$$

$$\leq \max\{t^2, 1\}\left(\mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right)\|f\|^2_{\mathbf{L}^2(\mathcal{X})} + \mathcal{O}\left(\frac{(\log N)^{3/2}}{N^{\frac{4}{d+6}+\frac{1}{2}}}\right)\|f\|^2_\infty\right) \tag{72}$$
$$= \max\{t^2, 1\}\mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right)\left(\|f\|^2_{\mathbf{L}^2(\mathcal{X})} + \sqrt{\frac{\log N}{N}}\|f\|^2_\infty\right)$$

where in (72) we used the fact that $d \geq 2$. Repeating the final string of inequalities in the case where $d = 1$, we instead obtain

$$\|H_{N,\epsilon,\kappa,t}\rho f - \rho H^\kappa_t f\|^2_2 \leq \max\{t^2, 1\}\left(\mathcal{O}\left(\frac{\log N}{N^{4/7}}\right)\|f\|^2_{\mathbf{L}^2(\mathcal{X})} + \mathcal{O}\left(\frac{\log N}{N}\right)\|f\|^2_\infty\right)$$

as desired. □

Proof of Corollary 1. We first note that

$$\|H_{N,\epsilon,\kappa,t}\rho f - \rho H_t f\|^2_2 \leq 2\|H_{N,\epsilon,\kappa,t}\rho f - \rho H^\kappa_t f\|^2_2 + 2\|\rho H_t f - \rho H^\kappa_t f\|^2_2.$$

Lemma 5 implies that with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$

$$\|\rho(H_t f - H^\kappa_t f)\|^2_2 \leq \|H_t f - H^\kappa_t f\|^2_{\mathbf{L}^2(\mathcal{X})} + \|H_t f - H^\kappa_t f\|^2_\infty\sqrt{\frac{18\log N}{N}}.$$

Therefore, applying Lemma 4 implies

$$\|\rho(H_t f - H^\kappa_t f)\|^2_2 \leq e^{-2t\mu_{\kappa+1}}\|f\|^2_{\mathbf{L}^2(\mathcal{X})} + \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right)\|f\|^2_\infty.$$

Applying Theorem 10 thus completes the proof. □

Appendix M. The proof of Theorem 12

In order to prove Theorem 12, we will need two lemmas.

Lemma 10. *Let $f \in \mathbf{L}^2(\mathcal{X})$, and let $p = (j_1, \ldots, j_m)$ be a path of length m, then*

$$\|U[p]f\|_\infty \leq 2^m\|f\|_\infty.$$

Proof of Lemma 10. Young's inequality and (36) implies that for all $t > 0$ we have $\|H_t f\|_\infty \leq \|f\|_\infty$. Therefore, the case where $m = 1$ follows from the triangle inequality and the fact that σ is non-expansive. The general case follows from the fact that $\|U[j_1, \ldots, j_m] = U[j_m]\ldots U[j_1]$. □

Lemma 11. *For all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$ and all $0 \leq j \leq J$ we have*

$$\|A_{J,N}\mathbf{x} - A_{J,N}\mathbf{y}\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_2$$

and

$$\|U_N[j]\mathbf{x} - U_N[j]\mathbf{y}\|_2 \leq \|W_{j,N}\mathbf{x} - W_{j,N}\mathbf{y}\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_2.$$

Proof. By construction we have, for $1 \leq j \leq J$

$$W_{j,N}\mathbf{x} - W_{j,N}\mathbf{y} = (H_{N,\epsilon,\kappa,2^{j-1}} - H_{N,\epsilon,\kappa,2^j})(\mathbf{x} - \mathbf{y})$$
$$= \sum_{k=0}^{\kappa}(e^{-\lambda^{N,\epsilon}_k 2^{j-1}} - e^{-\lambda^{N,\epsilon}_k 2^j})\mathbf{u}_k\mathbf{u}^T_k(\mathbf{x} - \mathbf{y}).$$

Therefore, the fact that $\|W_{j,N}\mathbf{x} - W_{j,N}\mathbf{y}\|_2 \leq \|\mathbf{x} - \mathbf{y}\|_2$ follows from the fact that the vectors $\{\mathbf{u}_k\}_{k=0}^\kappa$ are an orthonormal basis for their span and the fact that

$$|e^{-\lambda^{N,\epsilon}_k 2^{j-1}} - e^{-\lambda^{N,\epsilon}_k 2^j}| \leq 1.$$

The bounds for $W_{0,N}$ and $A_{J,N}$ follow similarly and the bound for $U_N[j]$ follows from the fact that σ is nonexpansive. □

Proof of Theorem 12. We argue by induction on m. To establish the base case, we let $p = (j_1)$ and observe that σ commutes with ρ. Therefore, we have

$$\begin{aligned}\|U_N[j_1]\rho f - \rho U[j_1]f\|_2^2 &= \|\sigma W_{j_1,N}\rho f - \rho\sigma W_j f\|_2^2\\ &= \|\sigma W_{j_1,N}\rho f - \sigma\rho W_j f\|_2^2\\ &\leq \|W_{j_1,N}\rho f - \rho W_j f\|_2^2,\end{aligned}$$

where the final inequality follows from the fact that σ is non-expansive. Therefore, the case where $m = 1$ now follows from Theorem 11.

Now suppose the theorem is true for $m-1$. Let $p = (j_1, \dots, j_m)$ be a path of length m. Let $p_{m-1} = (j_1, \dots, j_{m-1})$ so that $U[p] = U[j_m]U[p_{m-1}]$ and $U_N[p] = U_N[j_m]U_N[p_{m-1}]$. Then,

$$\begin{aligned}&\|U_N[p]\rho f - \rho U[p]f\|_2^2\\ =&\|U_N[j_m]U_N[p_{m-1}]\rho f - \rho U[j_m]U[p_{m-1}]f\|_2^2\\ =&\|U_N[j_m]U_N[p_{m-1}]\rho f - U_N[j_m]\rho U[p_{m-1}]f + U_N[j_m]\rho U[p_{m-1}]f - \rho U[j_m]U[p_{m-1}]f\|_2^2\\ \leq&2\|U_N[j_m]U_N[p_{m-1}]\rho f - U_N[j_m]\rho U[p_{m-1}]f\|_2^2 + 2\|U_N[j_m]\rho U[p_{m-1}]f - \rho U[j_m]U[p_{m-1}]f\|_2^2\\ \leq&2\|U_N[p_{m-1}]\rho f - \rho U[p_{m-1}]f\|_2^2 + 2\|U_N[j_m]\rho U[p_{m-1}]f - \rho U[j_m]U[p_{m-1}]f\|_2^2,\end{aligned}$$

where in the final inequality we used Lemma 11. The term $\|U_N[p_{m-1}]\rho f - \rho U[p_{m-1}]f\|_2^2$ may be immediately bounded by the inductive hypothesis. Moreover, we may also apply the inductive hypothesis with $U[p_{m-1}]f$ in place of f to see

$$\begin{aligned}&\|U_N[j_m]\rho U[p_{m-1}]f - \rho U[j_m]U[p_{m-1}]f\|_2^2\\ &\leq 2^{2j_{\max}}\left(\left(\mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right) + \mathcal{O}(e^{-\mu_{\kappa+1}})\right)\|U[p_{m-1}]f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right)\|U[p_{m-1}]f\|_\infty^2\right)\end{aligned}$$

Iteratively applying Proposition 1 implies that $\|U[p_{m-1}]f\|_{\mathbf{L}^2(\mathcal{X})} \leq \|f\|_{\mathbf{L}^2(\mathcal{X})}$ and Lemma 10 implies$\|U[p_{m-1}]f\|_\infty \leq 2^{m-1}\|f\|_\infty$. Therefore, the result follows. □

Appendix N. The proofs of Theorems 13 and 14

The Proof of Theorem 13.

$$\begin{aligned}&\|S_{J,N}[p]\rho f - \rho S_J[p]f\|_2^2\\ =&\|A_{J,N}U_{J,N}[p]\rho f - \rho A_J U[p]f\|_2^2\\ \leq&2\|A_{J,N}U_{J,N}[p]\rho f - A_{J,N}\rho U[p]f\|_2^2 + 2\|A_{J,N}\rho U[p]f - \rho A_J U[p]f\|_2^2\\ \leq&2\|A_{J,N}\|_2\|U_{J,N}[p]\rho f - \rho U[p]f\|_2^2 + 2\|A_{J,N}\rho U[p]f - \rho A_J U[p]f\|_2^2\\ \leq&2\|U_{J,N}[p]\rho f - \rho U[p]f\|_2^2 + 2\|A_{J,N}\rho U[p]f - \rho A_J U[p]f\|_2^2,\end{aligned}$$

where the last inequality uses Lemma 11. To bound $\|U_{J,N}[j]\rho f - \rho U[j]f\|_2^2$, we may apply Theorem 12. To bound the second term, we apply Corollary 1 with $t = 2^J$ to obtain

$$\begin{aligned}&\|A_{J,N}\rho U[p]f - \rho A_J U[p]\|_2^2\\ \leq&2^{2J}\left(\left(\mathcal{O}\left(\frac{\log N}{N^{\frac{4}{d+6}}}\right) + \mathcal{O}(e^{-2^{J+1}\mu_{\kappa+1}})\right)\|f\|_{\mathbf{L}^2(\mathcal{X})}^2 + \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right)\|f\|_\infty^2\right).\end{aligned}$$

Iteratively applying Proposition 1 implies that $\|U[p_{m-1}]f\|_{\mathbf{L}^2(\mathcal{X})} \leq \|f\|_{\mathbf{L}^2(\mathcal{X})}$ and Lemma 10 implies$\|U[p_{m-1}]f\|_\infty \leq 2^{m-1}\|f\|_\infty$. Therefore, the result follows. □

The Proof of Theorem 14. Let α_0 be the scalar from Theorem 9 with $k = 0$. By Remark 1, and the definition of the non-windowed scattering coefficients, we may assume without loss of generality that α_0 is non-negative (since $-\varphi_0$ is also an eigenfunction). Thus, recalling that $\mathbf{v}_0 = \rho\varphi_0$, we see that by the definition of the non-windowed scattering coefficients, the triangle inequality, and the Cauchy-Schwarz inequality we have

$$\begin{aligned}&|\overline{S}_N[p]\rho f - \overline{S}[p]f| \qquad (73)\\ \leq&|\langle U_N[p]\rho f, \mathbf{u}_0\rangle_2 - \langle U[p]f, \varphi_0\rangle_{\mathbf{L}^2(\mathcal{X})}|\\ \leq&|\langle U_N[p]\rho f, \mathbf{u}_0\rangle_2 - \langle \rho U[p]f, \mathbf{v}_0\rangle_2| + |\langle \rho U[p]f, \mathbf{v}_0\rangle_2 - \langle U[p]f, \varphi_0\rangle_{\mathbf{L}^2(\mathcal{X})}|\\ =&|\langle U_N[p]\rho f, \mathbf{u}_0\rangle_2 - \langle \frac{1}{\alpha_0}\rho U[p]f, \alpha_0\mathbf{v}_0\rangle_2| + |\langle \rho U[p]f, \mathbf{v}_0\rangle_2 - \langle U[p]f, \varphi_0\rangle_{\mathbf{L}^2(\mathcal{X})}|\end{aligned}$$

$$\leq |\langle U_N[p]\rho f, \mathbf{u}_0 - \alpha_0 \mathbf{v}_0 \rangle_2| + |\langle U_N[p]\rho f - \frac{1}{\alpha_0}\rho U[p]f, \alpha_0 \mathbf{v}_0 \rangle_2| + |\langle \rho U[p]f, \mathbf{v}_0 \rangle_2 - \langle U[p]f, \varphi_0 \rangle_{\mathbf{L}^2(\mathcal{X})}|$$

$$\leq \|U_N \rho f\|_2 \|\mathbf{u}_0 - \alpha_0 \mathbf{v}_0\|_2 + \|U_N[p]\rho f - \frac{1}{\alpha_0}\rho U[p]f\|_2 \|\alpha_0 \mathbf{v}_0\|_2 + |\langle \rho U[p]f, \rho\varphi_0 \rangle_2 - \langle U[p]f, \varphi_0 \rangle_{\mathbf{L}^2(\mathcal{X})}|. \tag{74}$$

Lemmas 5 and 11 together with the inequality $\sqrt{a^2+b^2} \leq |a| + |b|$ imply

$$\|U_N \rho f\|_2 \leq \|\rho f\|_2 \leq \|f\|_{\mathbf{L}^2(\mathcal{X})} + \left(\frac{18 \log N}{N}\right)^{1/4} \|f\|_\infty$$

with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$ and Theorem 9 implies that

$$\|\mathbf{u}_0 - \alpha_0 \mathbf{v}_0\|_2 = \mathcal{O}\left(N^{-\frac{2}{d+6}} \sqrt{\log N}\right),$$

again with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$. Therefore,

$$\|U_N \rho f\|_2 \|\mathbf{u}_0 - \alpha_0 \mathbf{v}_0\|_2 \leq \mathcal{O}\left(N^{-\frac{2}{d+6}} \sqrt{\log N}\right) \|f\|_{\mathbf{L}^2(\mathcal{X})} + \mathcal{O}\left(N^{-\frac{2}{d+6}-\frac{1}{4}} (\log N)^{3/4}\right) \|f\|_\infty. \tag{75}$$

Theorem 9 shows that $|\alpha_0| = 1 + o(1)$, and (35) implies that $\|\varphi_0\|_{\mathbf{L}^2(\mathcal{X})} = \|\varphi_0\|_\infty = 1$. Therefore, Lemma 5 implies

$$\|\alpha_0 \mathbf{v}_0\|_2 \leq (1 + o(1)) \|\rho\varphi_0\|_2 \leq (1 + o(1)) \left(\|\varphi_0\|^2_{\mathbf{L}^2(\mathcal{X})} + \sqrt{\frac{\log N}{N}} \|\varphi_0\|^2_\infty \right) = \mathcal{O}(1). \tag{76}$$

Proposition 1 and a simple induction argument implies $\|U[p]f\|_{\mathbf{L}^2(\mathcal{X})} \leq \|f\|_{\mathbf{L}^2(\mathcal{X})}$, and Remark 8 implies

$$\left| \frac{1}{\alpha_k} - 1 \right| \leq \mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right) + \mathcal{O}\left(\frac{\log(N)}{N^{4/(d+6)}}\right).$$

Therefore, by Theorem 12, Lemma 5, and Lemma 10, we have

$$\begin{aligned}
\|U_N[p]\rho f - \frac{1}{\alpha_0}\rho U[p]f\|_2 &\leq \|U_N[p]\rho f - \rho U[p]f\|_2 + \left|\frac{1}{\alpha_0} - 1\right| \|\rho U[p]f\|_2 \\
&\leq 2^J \left[\left(\mathcal{O}\left(\frac{\sqrt{\log N}}{N^{\frac{2}{d+6}}} \right) + \mathcal{O}(e^{-\mu_{\kappa+1}/2}) \right) \|f\|_{\mathbf{L}^2(\mathcal{X})} + \mathcal{O}\left(\left(\frac{\log N}{N} \right)^{1/4} \right) \|f\|_\infty \right] \\
&\quad + \left(\mathcal{O}\left(\sqrt{\frac{\log N}{N}} \right) + \mathcal{O}\left(\frac{\log(N)}{N^{4/(d+6)}} \right) \right) \|f\|_{\mathbf{L}^2(\mathcal{X})} \\
&\quad + \left(\mathcal{O}\left(\left(\frac{\log N}{N} \right)^{3/4} \right) + \mathcal{O}\left(\frac{\log^{5/4}(N)}{N^{4/(d+6)+1/4}} \right) \right) \|f\|_\infty \\
&= \left(\mathcal{O}\left(\frac{\sqrt{\log N}}{N^{\frac{2}{d+6}}} \right) 2^J + \mathcal{O}\left(\sqrt{\frac{\log N}{N}} \right) + \mathcal{O}(e^{-\mu_{\kappa+1}/2}) 2^J \right) \|f\|_{\mathbf{L}^2(\mathcal{X})} \\
&\quad + \mathcal{O}\left(\left(\frac{\log N}{N} \right)^{1/4} \right) 2^J \|f\|_\infty.
\end{aligned} \tag{77}$$

Lastly, we again apply Lemma 5 and Lemma 10 to see that

$$\begin{aligned}
|\langle \rho U[p]f, \rho\varphi_0 \rangle_2 - \langle U[p]f, \varphi_0 \rangle_{\mathbf{L}^2(\mathcal{X})}| &= \mathcal{O}\left(\sqrt{\frac{\log N}{N}} \right) \|U[p]f\,\varphi_0\|_\infty \\
&= \mathcal{O}\left(\sqrt{\frac{\log N}{N}} \right) \|f\|_\infty
\end{aligned} \tag{78}$$

with probability at least $1 - \mathcal{O}\left(\frac{1}{N^9}\right)$. Combining (74) with (75), (76), (77), and (78) yields

$$\begin{aligned}
&|\overline{S}_N[p]\rho f - \overline{S}[p]f| \\
&\leq \|U_N \rho f\|_2 \|\mathbf{u}_0 - \alpha_0 \mathbf{v}_0\|_2 + \|U_N[p]\rho f - \frac{1}{\alpha_0}\rho U[p]f\|_2 \|\alpha_0 \mathbf{v}_0\|_2 + |\langle \rho U[p]f, \rho\varphi_0 \rangle_2 - \langle U[p]f, \varphi_0 \rangle_{\mathbf{L}^2(\mathcal{X})}| \\
&\leq \mathcal{O}\left(N^{-\frac{2}{d+6}} \sqrt{\log N} \right) \|f\|_{\mathbf{L}^2(\mathcal{X})} + \mathcal{O}\left(N^{-\frac{2}{d+6}-\frac{1}{4}} (\log N)^{3/4} \right) \|f\|_\infty
\end{aligned}$$

$$+\left(\mathcal{O}\left(\frac{\sqrt{\log N}}{N^{\frac{2}{d+6}}}\right)2^J+\mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right)+\mathcal{O}(e^{-\mu_{\kappa+1}/2})2^J\right)\|f\|_{\mathbf{L}^2(\mathcal{X})}$$
$$+\mathcal{O}\left(\left(\frac{\log N}{N}\right)^{1/4}\right)2^J\|f\|_\infty+\mathcal{O}\left(\sqrt{\frac{\log N}{N}}\right)\|f\|_\infty$$
$$\leq 2^J\left[\left(\mathcal{O}\left(\frac{\sqrt{\log N}}{N^{\frac{2}{d+6}}}\right)+\mathcal{O}\left(e^{-\mu_{\kappa+1}/2}\right)\right)\|f\|_{\mathbf{L}^2(\mathcal{X})}+\mathcal{O}\left(\left(\frac{\log N}{N}\right)^{1/4}\right)\|f\|_\infty\right]. \quad \square$$

Appendix O. Details on the baseline method

For both biomedical datasets, in our baseline classification method, we first performed k-means clustering on all cells from all patients (modeled as points in either $\mathbb{R}^{30}$ or $\mathbb{R}^{14}$). The value of k was based on expected subsets of immune cells: for the melanoma data we set $k=3$ based on expected subsets of CD4+ T helper cells, CD8+ killer T cells, and FOXP3+ T regulatory cells, and in COVID data we again set $k=3$ based on expected subsets of CD14+CD16++ non-classical monocytes, CD14++CD16 intermediate monocytes, and CD14++CD16- classical monocytes. Then, for each patient, we identified the proportion of cells corresponding to that patient lying within each cluster. We then used these features as input to a decision tree classifier.

Appendix P. Training details for Section 7.3

The results for baseline methods presented in Table 3 are taken directly from [87]. Therefore, for a fair comparison, we use the same validation procedure when training our method as was used in [87]. For each of the three meta-graphs, we independently, randomly generated 5 realizations of the DSBM. For each of these realizations, we randomly generated 10 training/test/validation splits. To tune our hyperparameters, J, q, c and γ (the latter two of which are hyperparameters of the SVM), we picked a single realization of each model and performed a grid search, choosing the parameters with the best average validation accuracy over the 10 splits. We then used these hyperparameters for all five realizations of each model (following the standard procedure of training on the training set and testing on the test set, holding out the validation set). The results reported in Table 3 are the test accuracies averaged over both the 5 realizations of each model and the 10 training/test/validation splits (i.e., over all 50 of the test sets). In our search, we selected J from a pool of $\{2,3,\dots,12\}$, magnetic Laplacian charge parameter q from a pool of $\{0,.05,.10,.15,.20,.25\}$, and SVM parameters from pools of $c\in\{25,100,250,500,1000\}$ and $\gamma\in\{10^{-5},10^{-4},10^{-3},10^{-2},10^{-1}\}$.

References

[1] Mikhail Belkin, Partha Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396.
[2] Mikhail Belkin, Partha Niyogi, Convergence of Laplacian eigenmaps, in: Advances in Neural Information Processing Systems, 2007, pp. 129–136.
[3] Dhananjay Bhaskar, Jackson D. Grady, Michael A. Perlmutter, Smita Krishnaswamy, Molecular graph generation via geometric scattering, in: 2022 IEEE 32nd International Workshop on Machine Learning for Signal Processing (MLSP), 2022.
[4] Federica Bogo, Javier Romero, Matthew Loper, Michael J. Black, FAUST: dataset and evaluation for 3D mesh registration, in: Proceedings IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2014.
[5] Davide Boscaini, Jonathan Masci, Simone Melzi, Michael M. Bronstein, Umberto Castellani, Pierre Vandergheynst, Learning class-specific descriptors for deformable shapes using localized spectral convolutional networks, in: Computer Graphics Forum, vol. 34, Wiley Online Library, 2015, pp. 13–23.
[6] Davide Boscaini, Jonathan Masci, Emanuele Rodolà, Michael Bronstein, Learning shape correspondence with anisotropic convolutional neural networks, in: Advances in Neural Information Processing Systems, vol. 29, 2016, pp. 3189–3197.
[7] Michael M. Bronstein, Joan Bruna, Taco Cohen, Petar Veličković, Geometric deep learning: grids, groups, graphs, geodesics, and gauges, arXiv preprint arXiv:2104.13478, 2021.
[8] Michael M. Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, Pierre Vandergheynst, Geometric deep learning: going beyond Euclidean data, IEEE Signal Process. Mag. 34 (4) (2017) 18–42.
[9] Joan Bruna, Stéphane Mallat, Multiscale sparse microcanonical models, Math. Stat. Learn. 1 (3/4) (2018) 257–315.
[10] Joan Bruna, Wojciech Zaremba, Arthur Szlam, Yann LeCun, Spectral networks and locally connected networks on graphs, in: Yoshua Bengio, Yann LeCun (Eds.), 2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings, 2014.
[11] Jameson Cahill, Joseph W. Iverson, Dustin G. Mixon, Daniel Packer, Group-invariant max filtering, arXiv preprint arXiv:2205.14039, 2022.
[12] Jeff Calder, Nicolas Garcia Trillos, Improved spectral convergence rates for graph Laplacians on ε-graphs and k-nn graphs, Appl. Comput. Harmon. Anal. 60 (2022) 123–175.
[13] Yaiza Canzani, Analysis on manifolds via the Laplacian, in: Course Notes for Math vol. 253, Fall 2013, Harvard University, 2013.
[14] Chih-Chung Chang, Chih-Jen Lin, LIBSVM: a library for support vector machines, ACM Trans. Intell. Syst. Technol. 2 (3) (2011) 1–27.
[15] Xiuyuan Cheng, Nan Wu, Eigen-convergence of Gaussian kernelized graph Laplacian by manifold heat interpolation, Appl. Comput. Harmon. Anal. 61 (2022) 132–190.
[16] Philip Chodrow, Nicole Eikmeier, Jamie Haddock, Nonbacktracking spectral clustering of nonuniform hypergraphs, SIAM J. Math. Data Sci. 5 (2) (2023) 251–279.
[17] Fan Chung, Laplacians and the Cheeger inequality for directed graphs, Ann. Comb. 9 (1) (2005) 1–19.
[18] Kenneth Ward Church, Word2vec, Nat. Lang. Eng. 23 (1) (2017) 155–162.
[19] Alexander Cloninger, A note on Markov normalized magnetic eigenmaps, Appl. Comput. Harmon. Anal. 43 (2) (2017) 370–380.
[20] Ronald R. Coifman, Stéphane Lafon, Diffusion maps, Appl. Comput. Harmon. Anal. 21 (2006) 5–30.
[21] Ronald R. Coifman, Mauro Maggioni, Diffusion wavelets, Appl. Comput. Harmon. Anal. 21 (1) (2006) 53–94.
[22] Nello Cristianini, John Shawe-Taylor, et al., An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods, Cambridge University Press, 2000.

[23] Mihai Cucuringu, Apoorv Vikram Singh, Déborah Sulem, Hemant Tyagi, Regularized spectral methods for clustering signed networks, J. Mach. Learn. Res. 22 (264) (2021) 1–79.
[24] Wojciech Czaja, Weilin Li, Analysis of time-frequency scattering transforms, Appl. Comput. Harmon. Anal. 47 (1) (2019) 149–171.
[25] E.B. Davies, B. Simon, Ultracontractivity and the heat kernel for Schrödinger operators and Dirichlet Laplacians, J. Funct. Anal. 59 (2) (1984) 335–395.
[26] Michaël Defferrard, Xavier Bresson, Pierre Vandergheynst, Convolutional neural networks on graphs with fast localized spectral filtering, in: Advances in Neural Information Processing Systems, vol. 29, 2016, pp. 3844–3852.
[27] David B. Dunson, Hau-Tieng Wu, Nan Wu, Spectral convergence of graph Laplacian and heat kernel reconstruction in l-infinity from random samples, Appl. Comput. Harmon. Anal. 55 (2021) 282–336.
[28] Bruno Messias F. de Resende, Luciano da F. Costa, Characterization and comparison of large directed networks through the spectra of the magnetic Laplacian, Chaos 30 (7) (2020) 073141.
[29] Michaël Fanuel, Carlos M. Alaíz, Ángela Fernández, Johan A.K. Suykens, Magnetic eigenmaps for the visualization of directed networks, Appl. Comput. Harmon. Anal. 44 (1) (2018) 189–199.
[30] Michaël Fanuel, Carlos M. Alaiz, Johan AK Suykens, Magnetic eigenmaps for community detection in directed networks, Phys. Rev. E 95 (2) (2017) 022302.
[31] Yifan Feng, Haoxuan You, Zizhao Zhang, Rongrong Ji, Yue Gao, Hypergraph neural networks, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 33, 2019, pp. 3558–3565.
[32] Stefano Fiorini, Stefano Coniglio, Michele Ciavotta, Enza Messina Sigmanet, One Laplacian to rule them all, in: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 37, 2023, pp. 7568–7576.
[33] Satoshi Furutani, Toshiki Shibahara, Mitsuaki Akiyama, Kunio Hato, Aida Masaki, Graph signal processing for directed graphs based on the Hermitian Laplacian, in: Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Springer, 2019, pp. 447–463.
[34] Fernando Gama, Alejandro Ribeiro, Joan Bruna, Diffusion scattering transforms on graphs, in: International Conference on Learning Representations, 2019.
[35] Feng Gao, Guy Wolf, Matthew Hirn, Geometric scattering for graph data analysis, in: Proceedings of the 36th International Conference on Machine Learning, in: PMLR, vol. 97, 2019, pp. 2122–2131.
[36] Philipp Grohs, Thomas Wiatowski, Helmut Bölcskei, Deep convolutional neural networks on cartoon functions, in: IEEE International Symposium on Information Theory, 2016, pp. 1163–1167.
[37] Aditya Grover, Jure Leskovec, node2vec: scalable feature learning for networks, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016, pp. 855–864.
[38] William L. Hamilton, Rex Ying, Jure Leskovec, Inductive representation learning on large graphs, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, Curran Associates Inc., Red Hook, NY, USA, 2017, pp. 1025–1035.
[39] David K. Hammond, Pierre Vandergheynst, Rémi Gribonval, Wavelets on graphs via spectral graph theory, Appl. Comput. Harmon. Anal. 30 (2011) 129–150.
[40] Yixuan He, Michael Perlmutter, Gesine Reinert, Mihai Cucuringu, Msgnn: a spectral graph neural network based on a novel magnetic signed Laplacian, in: Learning on Graphs Conference, PMLR, 2022, pp. 40:1–40:39.
[41] Matthias Hein, Jean-Yves Audibert, Ulrike von Luxburg, Graph Laplacians and their convergence on random neighborhood graphs, J. Mach. Learn. Res. 8 (6) (2007).
[42] Franca Hoffmann, Bamdad Hosseini, Assad A. Oberai, Andrew M. Stuart, Spectral analysis of weighted Laplacians arising in data clustering, Appl. Comput. Harmon. Anal. 56 (2022) 189–249.
[43] Alexander C. Huang, Roberta Zappasodi, A decade of checkpoint blockade immunotherapy in melanoma: understanding the molecular basis for immune sensitivity and resistance, Nat. Immunol. 23 (5) (2022) 660–670.
[44] Matthias Keller, Daniel Lenz, Radosław K. Wojciechowski, Large time behavior of the heat kernel, Springer International Publishing, Cham, 2021, pp. 241–254.
[45] T. Kipf, M. Welling, Semi-supervised classification with graph convolutional networks, in: Proc. of ICLR, 2016.
[46] Johannes Klicpera, Aleksandar Bojchevski, Stephan Günnemann, Predict then propagate: graph neural networks meet personalized pagerank, in: ICLR, 2019.
[47] Taewook Ko, Yoonhyuk Choi, Chong-Kwon Kim, A spectral graph convolution for signed directed graphs via magnetic Laplacian, Neural Netw. 164 (2023) 562–574.
[48] Manik Kuchroo, et al., Multiscale PHATE identifies multimodal signatures of COVID-19, in: Nature Biotechnology, 2022.
[49] Roberto Leonarduzzi, Haixia Liu, Yang Wang, Scattering transform and sparse linear classifiers for art authentication, Signal Process. 150 (2018) 11–19.
[50] Ron Levie, Federico Monti, Xavier Bresson, Michael M. Bronstein, Cayleynets: graph convolutional neural networks with complex rational spectral filters, IEEE Trans. Signal Process. 67 (1) (2018) 97–109.
[51] Elliott H. Lieb, Michael Loss, Fluxes, Laplacians, and Kasteleyn's theorem, in: Statistical Mechanics, Springer, 1993, pp. 457–483.
[52] Lek-Heng Lim, Hodge Laplacians on graphs, SIAM Rev. 62 (3) (2020) 685–715.
[53] Ofir Lindenbaum, Moshe Salhov, Arie Yeredor, Amir Averbuch, Gaussian bandwidth selection for manifold learning and classification, Data Min. Knowl. Discov. 34 (2020) 1676–1712.
[54] Anna Little, Daniel McKenzie, James M. Murphy, Balancing geometry and density: path distances on high-dimensional data, SIAM J. Math. Data Sci. 4 (1) (2022) 72–99.
[55] Anna V. Little, Mauro Maggioni, James M. Murphy, Path-based spectral clustering: guarantees, robustness to outliers, and fast algorithms, J. Mach. Learn. Res. 21 (2020).
[56] Yi Ma, Jianye Hao, Yaodong Yang, Han Li, Junqi Jin, Guangyong Chen, Spectral-based graph convolutional network for directed graphs, arXiv preprint arXiv: 1907.08990, 2019.
[57] Stéphane Mallat, Group invariant scattering, Commun. Pure Appl. Math. 65 (10) (October 2012) 1331–1398.
[58] Sohir Maskey, Ron Levie, Yunseok Lee, Gitta Kutyniok, Generalization analysis of message passing neural networks on large random graphs, Adv. Neural Inf. Process. Syst. 35 (2022) 4805–4817.
[59] Jason McEwen, Christopher Wallis, Augustine N. Mavor-Parker, Scattering networks on the sphere for scalable and rotationally equivariant spherical CNNs, in: International Conference on Learning Representations, 2022.
[60] Yimeng Min, Frederik Wenkel, Michael Perlmutter, Guy Wolf, Can hybrid geometric scattering networks help solve the maximum clique problem?, Adv. Neural Inf. Process. Syst. 35 (2022) 22713–22724.
[61] Boaz Nadler, Stéphane Lafon, Ronald R. Coifman, Ioannis G. Kevrekidis, Diffusion maps, spectral clustering and reaction coordinates of dynamical systems, Appl. Comput. Harmon. Anal. 21 (1) (2006) 113–127.
[62] Annamalai Narayanan, Mahinthan Chandramohan, Rajasekar Venkatesan, Lihui Chen, Yang Liu, Shantanu Jaiswal, graph2vec: learning distributed representations of graphs, CoRR, arXiv:1707.05005 [abs], 2017.
[63] Michael Perlmutter, Feng Gao, Guy Wolf, Matthew Hirn, Geometric scattering networks on compact Riemannian manifolds, in: Mathematical and Scientific Machine Learning Conference, 2020.
[64] Michael Perlmutter, Alexander Tong, Feng Gao, Guy Wolf, Matthew Hirn, Understanding graph neural networks with generalized geometric scattering transforms, arXiv preprint arXiv:1911.06253, 2019.
[65] Jason Ptacek, Matthew Vesely, David Rimm, Monirath Hav, Murat Aksoy, Ailey Crow, Jessica Finn, 52 characterization of the tumor microenvironment in melanoma using multiplexed ion beam imaging (MIBI), J. ImmunoTher. Cancer 9 (Suppl 2) (2021) A59.
[66] Naoki Saito, Stefan C. Schonsheck, Eugene Shvarts, Multiscale transforms for signals on simplicial complexes, arXiv preprint arXiv:2301.02136, 2022.

[67] Naoki Saito, Stefan C. Schonsheck, Eugene Shvarts, Multiscale Hodge scattering networks for data analysis, arXiv preprint arXiv:2311.10270, 2023.
[68] Naoki Saito, David S. Weber, Underwater object classification using scattering transform of sonar signals, in: Wavelets and Sparsity XVII, vol. 10394, SPIE, 2017, pp. 103–115.
[69] Michael T. Schaub, Yu Zhu, Jean-Baptiste Seby, T. Mitchell Roddenberry, Santiago Segarra, Signal processing on higher-order networks: livin'on the edge... and beyond, Signal Process. 187 (2021) 108149.
[70] Yiqian Shi, Bin Xu, Gradient estimate of an eigenfunction on a compact Riemannian manifold without boundary, Ann. Glob. Anal. Geom. 38 (2010) 21–26.
[71] David I. Shuman, Sunil K. Narang, Pascal Frossard, Antonio Ortega, Pierre Vandergheynst, The emerging field of signal processing on graphs: extending high-dimensional data analysis to networks and other irregular domains, IEEE Signal Process. Mag. 30 (3) (2013) 83–98.
[72] Amit Singer, Hau-Tieng Wu, Orientability and diffusion maps, Appl. Comput. Harmon. Anal. 31 (1) (2011) 44–58.
[73] Amit Singer, Hau-Tieng Wu, Vector diffusion maps and the connection Laplacian, Commun. Pure Appl. Math. 65 (8) (2012) 1067–1144.
[74] Rahul Singh, Yongxin Chen, Signed graph neural networks: a frequency perspective, arXiv preprint arXiv:2208.07323, 2022.
[75] Pablo Sprechmann, Joan Bruna, Yann LeCun, Audio source separation with discriminative scattering networks, in: International Conference on Latent Variable Analysis and Signal Separation, Springer, 2015, pp. 259–267.
[76] Federico Tombari, Samuele Salti, Luigi Di Stefano, Unique signatures of histograms for local surface description, in: European Conference on Computer Vision, 2010, pp. 356–369.
[77] Alexander Tong, Frederik Wenkel, Dhananjay Bhaskar, Kincaid Macdonald, Jackson Grady, Michael Perlmutter, Smita Krishnaswamy, Guy Wolf, Learnable filters for geometric scattering modules, arXiv preprint arXiv:2208.07458, 2022.
[78] Zekun Tong, Yuxuan Liang, Changsheng Sun, Xinke Li, David Rosenblum, Andrew Lim, Digraph inception convolutional networks, Adv. Neural Inf. Process. Syst. 33 (2020) 17907–17918.
[79] Zekun Tong, Yuxuan Liang, Changsheng Sun, David S. Rosenblum, Andrew Lim, Directed graph convolutional network, arXiv:2004.13970, 2020.
[80] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Liò, Yoshua Bengio, Graph attention networks, in: International Conference on Learning Representations, 2018.
[81] Zhiyang Wang, Luana Ruiz, Alejandro Ribeiro, Stability of neural networks on Riemannian manifolds, in: 2021 29th European Signal Processing Conference (EUSIPCO), 2021, pp. 1845–1849.
[82] Zhiyang Wang, Luana Ruiz, Alejandro Ribeiro, Stability of neural networks on manifolds to relative perturbations, in: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2022, pp. 5473–5477.
[83] Frederik Wenkel, Yimeng Min, Matthew Hirn, Michael Perlmutter, Guy Wolf, Overcoming oversmoothness in graph convolutional networks via hybrid scattering networks, arXiv preprint arXiv:2201.08932, 2022.
[84] Thomas Wiatowski, Helmut Bölcskei, Deep convolutional neural networks based on semi-discrete frames, in: Proceedings of IEEE International Symposium on Information Theory, 2015, pp. 1212–1216.
[85] Thomas Wiatowski, Helmut Bölcskei, A mathematical theory of deep convolutional neural networks for feature extraction, IEEE Trans. Inf. Theory 64 (3) (2018) 1845–1866.
[86] Keyulu Xu, Weihua Hu, Jure Leskovec, Stefanie Jegelka, How powerful are graph neural networks?, in: International Conference on Learning Representations, 2019.
[87] Xitong Zhang, Yixuan He, Nathan Brugnone, Michael Perlmutter, Matthew Hirn MagNet, A neural network for directed graphs, Adv. Neural Inf. Process. Syst. 34 (2021).
[88] Dongmian Zou, Gilad Lerman, Encoding robust representation for graph generation, in: International Joint Conference on Neural Networks, 2019.
[89] Dongmian Zou, Gilad Lerman, Graph convolutional neural networks via scattering, Appl. Comput. Harmon. Anal. 49(3) (3) (2019) 1046–1074.