Geometric Scattering on Non-Euclidean Data

Joyce A. Chew

Department of Mathematics
University of California, Los Angeles

Acknowledgements

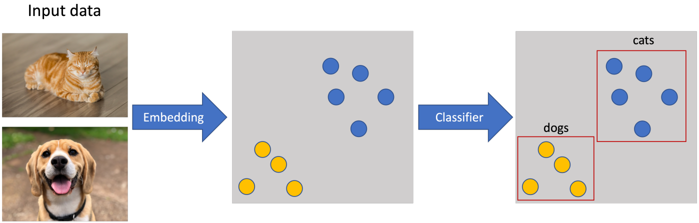
This talk is based on joint work with Matthew Hirn, Smita Krishnaswamy, Deanna Needell, Michael Perlmutter, Holly Steach, Siddharth Viswanath, and Hau-Tieng Wu, arXiv:2208.08561.

Outline

- The Euclidean Scattering Transform
- A General Scattering Framework
- Examples
- Manifold Scattering on Point Clouds
- Numerical Experiments

Deep Neural Networks

- A deep neural network can be thought of as an embedding together with a classifier.
- The embedding transforms each input into an element of a high-dimensional vector space.
- The classifier makes a final prediction.



Invariance and Equivariance

- Let τ_c be the translation operator $\tau_c f(x) = f(x - c)$.
- Equivariance (where are the eyes?): Want a transformation S such that $S\tau_c f = \tau_c S f$ (i.e. the transformation commutes with translations)
- Invariance (are the eyes open?): Want a transformation $\overline{S}$ such that $\overline{S}\tau_c f = \overline{S} f$ (i.e. the transformation is unchanged by translations)

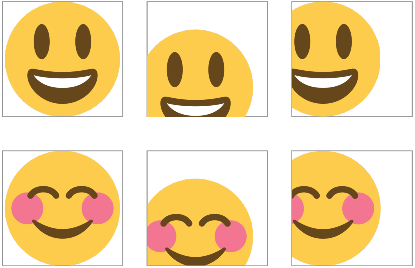


Figure: Created by Holly Steach

The (Euclidean) Scattering Transform

Group Invariant Scattering (S. Mallat 2012):

- Model of Convolutional Neural Networks.
- Predefined (wavelet) filters.
- Highlights the symmetries of such networks with respect to group actions

Advantages:

- Provable stability and invariance properties.
- Very good numerical results in certain situations.
- Needs less training data.

The Wavelet Transform

Setup:

- Mean-zero function ψ: $\int_{\mathbb{R}} \psi(x)dx = 0$
- Non-negative scaling function ϕ: $\int_{\mathbb{R}} \phi(x)dx = 1$
- Dilations: $\psi_j(x) = 2^{-j}\psi\left(\frac{x}{2^j}\right)$, $\phi_J(x) = 2^{-J}\phi\left(\frac{x}{2^J}\right)$
- Convolution Operators: $W_j f = \psi_j \star f$, $A_J f = \phi_J \star f$

The Transform:

- $\mathcal{W}_J := \{W_j\}_{j\leq J} \cup \{A_J\}$
- Captures information about the input at different scales of resolution or frequency bands
- Isometry property:

$$\|\mathcal{W}_J f(x)\|^2 := \sum_{j\leq J} \|W_j f\|^2 + \|A_J f\|^2 = \|f\|^2$$

The Scattering Transform

Windowed and Non-Windowed Transforms

- Multilayered cascade of nonlinear measurements.
- Each "layer" uses a wavelet transform $\mathcal{W}_J$ and a nonlinearity.
- $U[j]f(x) = MW_jf(x) = |W_jf(x)|, \quad j \leq J,$
- Path of scales $p = (j_1, \ldots, j_m)$
- $U[p]f(x) = U[j_m] \ldots U[j_1]f(x)$
- Windowed scattering transform:

$$S_J[p]f(x) = A_J U[p]f(x)$$

- Non-windowed scattering transform:

$$\overline{S}[p] = \lim_{J \to \infty} S_J[p]f(x) \cong \|U[p]f\|_1$$

Invariance and Equivariance

Theorem: (Mallat 2012)

Let τ_c be the translation operator $\tau_c f(x) = f(x - c)$

- The windowed scattering transform S_J is *equivariant*:

$$S_J[p]\tau_c f = \tau_c S_J[p]f$$

- The non-windowed scattering transform $\overline{S}$ is *invariant*:

$$\overline{S}[p]\tau_c f = \overline{S}[p]f.$$

Extract Invariance from Equivariance

- The invariance of $\overline{S}$ follows from the facts:
  - The operator U is translation equivariant.
  - $\overline{S}[p]f \cong \|U[p]\|_1 f.$
  - $\|\cdot\|_1$ is translation invariant.

Geometric Deep Learning

Modern Data Landscape

- Graphs (social networks, molecules)
- Manifolds (higher-dimensional structures, explicit and implicit)
- Goal: Generalize/extend the ideas and success of CNN-type architectures to these non-Euclidean settings.

Geometric Scattering Transforms

- Key challenge is defining wavelets.
- Once wavelets are defined, scattering is then an alternating cascade of wavelets and non-linearities.

Wavelets and Scattering on a Measure Space

Setup:

- Let $\mathcal{X} = (X, \mathcal{F}, \mu)$ be a measure space
- L a self-adjoint, positive semidefinite operator on $\mathbf{L}^2(\mathcal{X})$
- Orthonormal eigenbasis: $L\varphi_k = \lambda_k \varphi_k$, $k \geq 0$
- Heat-Semigroup: $P_t = e^{-Lt}$
- Wavelets: $W_j = P_{2^{j-1}} - P_{2^j}, \quad 0 \leq j \leq J,$
- Low-Pass Filter: $A_J = P_{2^J}$

Proposition:

$\mathcal{W} = \{W_j\}_{0 \leq j \leq J} \cup \{A_J\}$ is a non-expansive frame, on $\mathbf{L}^2(\mathcal{X})$, i.e.,

$$c\|f\|^2 \leq \sum_j \|W_j f\|^2 + \|A_J f\|^2 \leq \|f\|^2.$$

Geometric Scattering on Measure Spaces

Windowed Scattering transform

$$U[j_1, \ldots, j_m]f = MW_{j_m} \ldots MW_{j_1} f$$
$$S_J[j_1, \ldots, j_m]f = A_J U[j_1, \ldots, j_m]f$$

Non-Windowed Scattering transform

$$\overline{S}[j_1, \ldots, j_m]f = |\langle U[j_1, \ldots, j_m]f, \varphi_0 \rangle|$$

Difference from before:

Integrating against the bottom eigenvector is not in general equivalent to taking an $\mathbf{L}^1$ norm. (This issue is even present on graphs when we weight vertices by degree.)

Theorem:

$$\|S_J f_1 - S_J f_2\| \leq \|f_1 - f_2\|, \quad \|\overline{S} f_1 - \overline{S} f_2\| \leq C_{\mathcal{X}} \|f_1 - f_2\|.$$

What Groups Should We Be Invariant To?

Setup:

Let $\mathcal{G}$ be a group of bijections from X to X. For $\zeta \in \mathcal{G}$, let

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

First Guess (Preserves measures):

The scattering transform should be invariant to $\mathcal{G}$ if for all $\zeta \in \mathcal{G}$, $\mu(\zeta^{-1}(B)) = \mu(B)$ for all measurable sets B.

Problem:

What if $\mathcal{X}$ is a graph and μ weighs vertices by degree?

Weaker Condition (Preserves Inner Products):

$\mathcal{G}$ induces an isometry on $\mathbf{L}^2(\mathcal{X})$, i.e.,

$$\langle V_\zeta f, V_\zeta g \rangle = \langle f, g \rangle.$$

Equivariance and Invariance

Theorem:

If $\mathcal{G}$ preserves inner products, then the windowed scattering transform is equivariant and the non-windowed scattering transform in invariant to the action of $\mathcal{G}$, i.e.

$$S_J V_\zeta f = V_\zeta S_J f, \quad \text{and} \quad \overline{S} V_\zeta f = \overline{S} f$$

Theorem:

If $\mathcal{G}$ preserves inner products and preserves measures, and additionally φ_0 is constant, then the windowed scattering transform is invariant in the limit,

$$\lim_{J\to\infty} \|S_J V_\zeta f - S_J f\|_{\mathbf{L}^2(\mathcal{X})}.$$

Examples

- Traditional Graphs - Graph Laplacian: $D - A$
  - (Can also normalize and use $D^{-1}L, LD^{-1}$ or $D^{-1/2}LD^{-1/2}$ depending on choice of measure)
- Manifolds - Laplace-Beltrami operator: $-\Delta = -\nabla \cdot \nabla$
- Directed Graphs - Magnetic Laplacian
- Signed Graphs - Signed Laplacian
- Signed and Directed Graphs - Magnetic Signed Laplacian

Magnetic Laplacian on Directed Graphs

Hermitian Adjacency Matrix

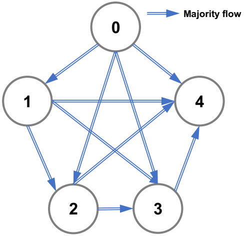
$$A_s = \frac{1}{2}(A + A^T)$$
$$\Theta = \frac{\pi}{2}(A - A^T)$$
$$H = A_s \odot \exp(i\Theta)$$

The Magnetic Laplacian

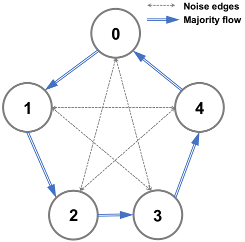
$$L = D_s - H = D_s - A_s \odot \exp(i\Theta)$$

- Undirected geometry is captured by the magnitude of entries.
- Directional information encoded by phase.

Numerical Experiments: Directed Stochastic Block Model



"Ordered" meta-graph



"Cyclic" meta-graph

- A node's cluster determines the probability of existence and direction of edges to nodes in other clusters.
- Node-level task of node classification, so windowed scattering coefficients are appropriate.
- Scattering using magnetic Laplacian achieves accuracy competitive with or exceeding that obtained from GNNs, even networks designed for directed graphs.

Point Cloud Scattering

Problem:

What if data is sampled from an underlying manifold, but we don't have knowledge of the manifold itself?

Data-Driven Graph Laplacian

- Construct an affinity matrix using a (Gaussian) kernel to determine the weights $K(x_i, x_j)$
- Approximate eigenfunctions / eigenvalues of the Laplace-Beltrami operator by the eigenvectors / eigenvalues of the graph Laplacian

Data-Driven Scattering

- Use κ eigenvectors / eigenvalues of the data-driven graph Laplacian to approximate the heat semigroup $P_t = e^{-Lt}$.
- Use this approximation to construct wavelets as before.

Convergence

Theorem:

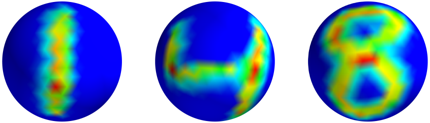
If the kernel is constructed properly, and the sample points are drawn i.i.d. uniformly at random (and several other assumptions), then with high probablity, the discretization error of the data-driven scattering transform is $\mathcal{O}(N^{-2/(d+6)})$

Remark:

This result builds on work by X. Cheng and N. Wu which guarantees the convergence of individual eigenvectors in ℓ^2 and of the eigenvalues. Our rate of convergence, with respect to N, is essentially the same as in this earlier result.

Numerical Experiments: Spherical MNIST

- Data: MNIST randomly rotated and projected onto sphere.
- Problem: signal classification on a manifold.



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

Table: Classification accuracies for spherical MNIST averaged over 10 realizations, using non-windowed scattering coefficients.

Point Cloud Scattering Cont.

Problem:

What if it is computationally infeasible to compute a sufficient number of eigenvalues / eigenvectors?

Second Method:

In this case, we use the approximation

$$P_1 \approx P_1^{(N)} := (D^{(N)})^{-1} W^{(N)}$$

where

$$W_{i,j}^{(N)} = K(x_i, x_j) \quad \text{and} \quad D_{i,i}^{(N)} = \sum_{j=0}^{N-1} W_{i,j}^{(N)},$$

and we approximate P_t by

$$P_t \approx (P_1^{(N)})^t.$$

Numerical Experiments: Single-Cell Data

Will a Melanoma Patient Respond to Immunotherapy?

- 54 Patients
- 11,862 cells per patient
- 30 proteins measured in each cell

Manifold Classification: Non-Windowed Scattering

- Each cell is a point in $\mathbb{R}^{30}$
- Each person is a point cloud of 11,862 points in $\mathbb{R}^{30}$
- We assume each person's points lie upon some d-dimensional manifold for $d < 30$.
- Scattering achieves 83% accuracy vs 48% from baseline

Conclusion

- The Euclidean scattering transform is a model of CNNs
  - Highlights the role of group invariance
  - Provable stability / invariance guarantees
- The scattering transform can be extended to graphs, manifolds, and other measure spaces with similar theoretical guarantees as the original
- The manifold scattering transform can be implemented on points sampled from unknown manifolds with provable convergence rate