Time-Frequency Scattering Transforms: Theory and Applications

Weilin Li

Norbert Wiener Center
Department of Mathematics
University of Maryland, College Park
http://www.norbertwiener.umd.edu

7-th International Conference on Computational Harmonic Analysis
Vanderbilt University
May 15, 2018

Acknowledgements
NSF DMS-1440140, DTRA 1-13-1-0015

Norbert Wiener Center for Harmonic Analysis and Applications

Collaborators

(a) Wojciech Czaja

(b) Ilya Kavalerov

Outline

Scattering transforms

Let P be a countable index set and fix a sequence of $L^2(\mathbb{R}^d)$ functions,

$$G = \{g,\ g_p\}_{p \in P}.$$

Associate $p \in P^k$ with the *scattering propagator* $U[p]$, formally defined as

$$U[p](f) = \begin{cases} |f * g_p| & \text{if } p \in P, \\ U[p_k] \cdots U[p_2]U[p_1]f & \text{if } p = (p_1, p_2, \ldots, p_k) \in P^k. \end{cases}$$

The *scattering transform* S_G associated with G is formally defined as

$$S_G(f) = \{f * g\} \cup \{U[p](f) * g\}_{p \in P^k, k \geq 1}.$$

- Zero order coefficient: $\{f * g\}$.
- First order coefficients: $\{|f * g_p| * g\}_{p \in P}$.
- Second order coefficients: $\{||f * g_{p_1}| * g_{p_2}| * g\}_{(p_1, p_2) \in P^2}$.
- Etc.

Scattering network

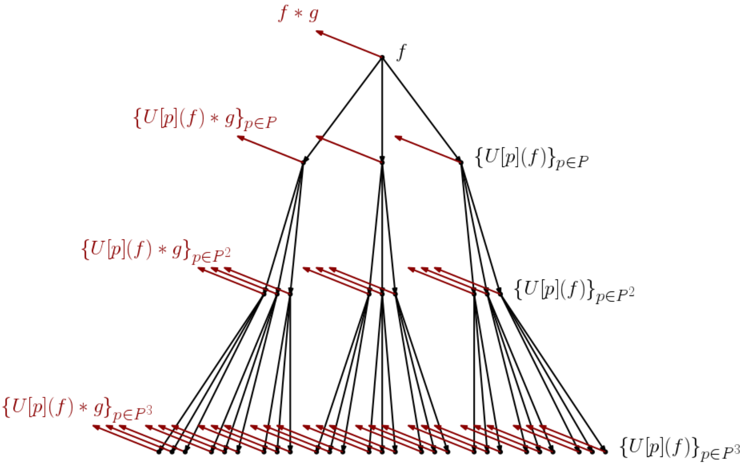


Figure : Network structure of the scattering transform.

Mallat scattering transforms

Theorem [Mallat '12]

Consider the Parseval semi-discrete wavelet frame $W = \{\varphi, \psi_\lambda\}_{\lambda \in \Lambda}$, where ψ a certain technical condition. If S_W is the wavelet scattering transform associated with W, then

1. (Energy preservation) For all $f \in X$, $\|S_W(f)\|_{L^2\ell^2} = \|f\|_{L^2}$.
2. (Non-expansiveness) For all $f, h \in L^2$, $\|S_W(f) - S_W(h)\|_{L^2\ell^2} \leq \|f - g\|_{L^2}$.
3. (Translation stability) There exists a constant $C(J) > 0$ such that for all $y \in \mathbb{R}^d$ and f in a certain logarithmic Sobolev space,
$$\|S_W(f) - S_W(f(\cdot - y))\|_{L^2\ell^2} \leq C(J)|y|(\|f\|_{L^2} + \|f\|_X).$$
4. (Diffeomorphism stability) For any diffeomorphism $\tau \in C^2(\mathbb{R}^d; \mathbb{R}^d)$ such that $Id - \tau$ is sufficiently small, there exists $C(J, \tau) > 0$ such that for all f in a certain logarithmic Sobolev space,
$$\|S_W(f) - S_W(f(\tau(\cdot)))\|_{L^2\ell^2} \leq C(J, \tau)\Big(\|f\|_{L^2} + \sum_{k=1}^{\infty} \sum_{\lambda \in \Lambda^k} \|U[\lambda](f)\|_{L^2}\Big).$$

Mallat scattering transforms

Applications:

- Handwritten digit analysis [Bruna, Mallat]
- Texture classification [Bruna, Mallat], [Sifre, Mallat]
- Music classification [Andèn, Mallat]
- Classification of molecules [Eickenberg, Exarchakis, Hirn, Mallat]

Related theory:

- Scattering on graphs [Cheng, Chen, Mallat], [Lerman, Zou]
- Wavelet phase retrieval [Waldspurger, Mallat]
- Alternative admissibility conditions [Waldspurger]
- More general scattering networks [Wiatowski, Bölcskei]
- The Lipschitz problem [Balan, Zou]

Time-frequency scattering?

We would like to use a different set of analyzing functions in the scattering transform instead of a wavelet frame. Our reasons:

- The convolution kernels that are learned in a convolutional neural network typically are not related by some algebraic structure such as scaling.
- Learned filters in the first few layers of neural networks almost always are localized, oriented, band-pass filters, which *resemble* Gabor functions.
- Biological evidence suggests that simple cells in the mammalian visual cortex are modeled by modulations and rotations of a fixed 2-dimensional Gaussian.
- The short-time (or windowed) Fourier transform

$$V_g f(x,\xi) = \int_{\mathbb{R}^d} f(y)\overline{g(y-x)}e^{-2\pi i \xi\cdot y}\,dy,$$

has been used as a feature extractor for various audio and image classification problems. Most notably, for $d=1$, the function $|V_g f|^2$ is the spectrogram of an audio signal f.

Outline

Uniform covering frame

From here onwards, we assume that G is a *uniform covering frame* (UCF):

- *Mild regularity and integrability*. $g \in L^1 \cap L^2 \cap C^1$ and $g_p \in L^1 \cap L^2$ for each $p \in P$.
- *Frequency support conditions*. $|\widehat{g}(0)| = 1$ and supp$(\widehat{g_p})$ is a compact and connected set for each $p \in P$.
- *Uniform covering property*. For any $R > 0$, there exists an integer N, such that for any $p \in P$, the supp$(\widehat{g_p})$ can be covered by N cubes of side length $2R$.

  The uniform covering property and the connectedness assumption implies the family of sets $\{$supp$(\widehat{g_p})\}$ have uniformly bounded diameters.

  Any wavelet frame violates the uniform covering property, so it is not a UCF.
- *Frame condition*. We have

$$|\widehat{g}|^2 + \sum_{p \in P} |\widehat{g_p}|^2 = 1.$$

  This is equivalent to: For all $f \in L^2$,

$$\|f * g\|_{L^2}^2 + \sum_{p \in P} \|f * g_p\|_{L^2}^2 = \|f\|_{L^2}^2.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Example 1: Gabor frame

Let g be smooth and assume that $\widehat{g}$ is compactly supported and for all $\xi \in \mathbb{R}^d$,

$$\sum_{m \in \mathbb{Z}^d} |\widehat{g}(\xi - m)|^2 = 1$$

Let $P = \mathbb{Z}^d \setminus \{0\}$.
For each $p \in P$, define the function g_p by the formula,

$$g_p(x) = e^{2\pi i p \cdot x} g(x).$$

The sequence of functions,

$$G = \{g,\ g_p\}_{p \in P}$$

is a *semi-discrete Gabor frame* as well as a UCF.
This example readily generalizes to other lattices.

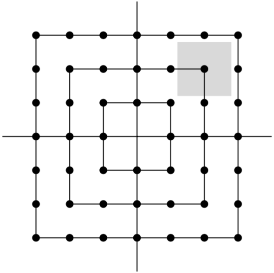


Figure : Tiling of the Fourier domain induced by a Gabor system.

Example 2: Rotational UCF

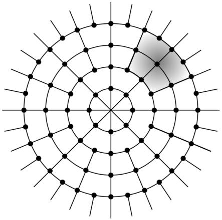
For a fixed integer $A \geq 1$ and each integer $m \geq 1$, let U_m be the finite rotation group on $\mathbb{R}^2$ generated by rotations by $2\pi/(m^* A)$, where $m^* = 2^{\lfloor \log_2(m) \rfloor}$. Define the index set,

$$P = \{(m, r) \colon m \geq 1, r \in U_m\}.$$

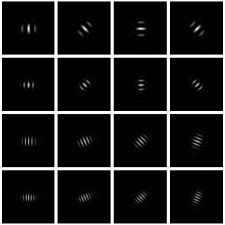
It is possible to select g and $\{g_m\}_{m \geq 1}$ appropriately such that if we let $g_{m,r}(x) = g_m(rx)$ for each $r \in U_m$, then

$$G = \{g,\ g_p\}_{p \in P}$$

is a UCF that is partially generated by rotations. This readily generalizes to d-dimensions, but is more complicated to write down.



(a) Tiling of the Fourier domain induced by the Rotational UCF for $A = 8$.



(b) Intensity plot of the Rotational UCF in the spatial domain.

Fourier scattering transform

The *Fourier scattering transform* (FST) S_F associated with a uniform covering frame $F = \{g, g_p\}_{p \in P}$ is formally defined as

$$S_F(f) = \{f * g\} \cup \{U[p](f) * g\}_{p \in P^k, k \geq 1}.$$

Theoretical questions:

- What does this transform do?
- Does it provide a useful representation of data?

Computational concerns:

- Transform needs to be truncated to be used in practice.
- Appears to be computationally expensive.

Outline

Norbert Wiener Center for Harmonic Analysis and Applications

Proposition: Exponential decay of energy

There exists a constant $0 < C_{decay} < 1$ depending only on the UCF such that for all $f \in L^2$ and $n \geq 1$,

$$\sum_{p \in P^n} \|U[p](f) * g\|_{L^2}^2 \leq \sum_{p \in P^n} \|U[p](f)\|_{L^2}^2 \leq C_{decay}^{n-1} \|f\|_{L^2}^2.$$

This property is not (known to be) true for the wavelet scattering transform. To prove this, the main challenge is to obtain the lower bound,

$$\||f * g_p| * g\|_{L^2}^2 \geq (1 - C_{decay}) \|f * g_p\|_{L^2}^2.$$

Higher order terms can be controlled by iterating this inequality and using the partition of unity property to obtain some cancellation in the form of a telescoping series.

Let $f_p = |f * g_p|$. From a Fourier perspective, the inequality of interest is equivalent to

$$\int_{\mathbb{R}^d} |\widehat{f_p}(\xi)|^2 |\widehat{g}(\xi)|^2 \, d\xi \geq (1 - C_{decay}) \int_{\mathbb{R}^d} |\widehat{f_p}(\xi)|^2 \, d\xi$$

Proof. Let $\phi \geq 0$ such that $|\widehat{\phi}| \leq |\widehat{g}| \leq 1$. Choose $R > 0$ sufficiently small such that $|\widehat{\phi}|^2 \geq C_\phi$ on $Q_R(0)$. By the uniform covering property, there exists an integer $N \geq 1$ such that $S_p = supp(\widehat{f_p})$ can be covered by N cubes of side length $2R$, for any $p \in P$. Let $\{\xi_{p,n}\}_{n=1}^N \subseteq \mathbb{R}^d$ be the center of these cubes, and so

$$\begin{aligned}
\|f * g_p\|_{L^2}^2 &\leq \frac{1}{C_\phi} \sum_{n=1}^{N} \int_{\mathbb{R}^d} |\widehat{f}(\xi)|^2 |\widehat{g_p}(\xi)|^2 |\widehat{\phi}(\xi - \xi_{p,n})|^2 \, d\xi \\
&= \frac{1}{C_\phi} \sum_{n=1}^{N} \|f * g_p * M_{p,n}\phi\|_{L^2}^2 \\
&\leq \frac{1}{C_\phi} \sum_{n=1}^{N} \||f * g_p| * \phi\|_{L^2}^2 \\
&\leq \frac{N}{C_\phi} \||f * g_p| * g\|_{L^2}^2
\end{aligned}$$

Rearranging, we have

$$\||f * g_p| * g\|_{L^2}^2 \geq \frac{C_\phi}{N} \|f * g_p\|_{L^2}^2.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Theorem [Czaja, L. '17]

Let S_F be the Fourier scattering transform associated with a UCF F, and let $PW(R, \epsilon)$ be the set of $f \in L^2$ such that $\|\widehat{f}\|_{L^2(Q_R(0))} \geq (1-\epsilon)\|f\|_{L^2}$.

1. (Energy conservation) For all $f \in L^2$, $\|S_F(f)\|_{L^2\ell^2} = \|f\|_{L^2}$.
2. (Non-expansiveness) For all $f, h \in L^2$, $\|S_F(f) - S_F(h)\|_{L^2\ell^2} \leq \|f - h\|_{L^2}$.
3. (Translation stability) There exists a constant $C > 0$ depending only on the UCF such that for all $f \in L^2$ and $y \in \mathbb{R}^d$,
$$\|S_F(f) - S_F(f(\cdot - y))\|_{L^2\ell^2} \leq C|y|\|\nabla g\|_{L^1}\|f\|_{L^2}.$$
4. (Diffeomorphism stability) There exists a universal constant $C > 0$ such that for any $\tau \in C^1(\mathbb{R}^d; \mathbb{R}^d)$ with $\|Id - \tau\|_{L^\infty}$ sufficiently small, and any $f \in PW(R, \epsilon)$,
$$\|S_F(f) - S_F(f(\tau(\cdot)))\|_{L^2\ell^2} \leq C(R\|Id - \tau\|_{L^\infty} + \epsilon)\|f\|_{L^2}.$$

Norbert Wiener Center for Harmonic Analysis and Applications

Outline

Norbert Wiener Center for Harmonic Analysis and Applications

Truncation

A natural way to truncate the transform is to generate the FST using only UCF frame elements whose Fourier transforms tile a large cube centered at the origin.

Proposition: Canonical ordering of the frame elements

For any UCF F, there exist a constant $C = C_{tiling} > 0$ and finite subsets $\{P_m\}_{m \geq 1}$ such that $P_m \nearrow P$ and for each $m \geq 1$,

$$|\widehat{g}(\xi)|^2 + \sum_{p \in P_m} |\widehat{g_p}(\xi)|^2 = \begin{cases} 1 & \text{if } \xi \in \overline{Q_{Cm}(0)}, \\ 0 & \text{if } \xi \notin \overline{Q_{C(m+1)}(0)}. \end{cases}$$

Using more covering techniques:

Proposition: Energy propagation along frequency decreasing paths

For each $m \geq 1$, there exists $C_m > 0$ such that $C_m \nearrow 1$ as $m \to \infty$ and for all $n \geq 1$, $p \in P^n$, and $f \in L^2$,

$$\|U[p](f) * g\|_{L^2}^2 + \sum_{q \in P_m} \|U[p](f) * g_q\|_{L^2}^2 \geq C_m \|U[p]f\|_{L^2}^2.$$

The *truncated Fourier scattering transform* $S_F[M,N]$ is formally defined as

$$S_F[M,N](f) = \{f * g\} \cup \{U[p](f) * g\}_{p \in P_M^n, 1 \le n \le N}.$$

It has N layers and $(\#P_M)^n$ coefficients in the n-th layer.

Theorem [Czaja, L. '17]

For any UCF F, let S_F be its associated Fourier scattering transform.

1. (Upper bound) For all $M, N \ge 1$ and $f \in L^2$, $\|S_F[M,N](f)\|_{L^2\ell^2} \le \|f\|_{L^2}$.
2. (Lower bound) Let $M(\epsilon, R, F)$ be sufficiently large. For all $N \ge 1$ and $f \in (R, \epsilon)$,

$$\|S_F[M,N](f)\|^2_{L^2\ell^2} \ge (C_M^N(1-\epsilon^2) - C_{\text{decay}}^{N-1})\|f\|^2_{L^2}.$$

3. Non-expansiveness, translation stability, and diffeomorphism stability still hold, and the estimates are similar to the regular Fourier scattering transform.

Norbert Wiener Center for Harmonic Analysis and Applications

Fast Fourier Scattering Transform

Using standard frame theory techniques, for any UCF G, we can explicitly construct discrete and finite frame versions of G.

Fast Fourier scattering transform:

Input: vector f, parameters $M, N \geq 1$
Construct: finite UCF frame elements $\{g,\ g_p\}_{p \in P_M}$
for $n = 1, 2, \ldots, N$
  for each $p = (p', p_n) \in P_M^n$
    Compute $U[p](f) = U[(p', p_n)](f) = |U[p'](f) * g_{p_n}|$ and $U[p](f) * g$.
  end
end

The algorithm can be made more efficient by computing only the path decreasing coefficients and down-sampling the FST output. If f is real-valued and G is a standard Gabor UCF, then only approximately half the FST need to be computed.

Norbert Wiener Center for Harmonic Analysis and Applications

Outline

Hyperspectral images

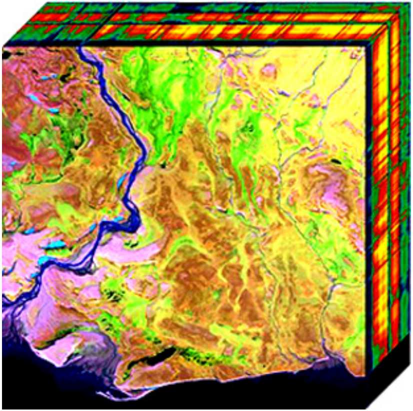


Figure : An example of a hyperspectral image, taken from Wikipedia.

Example 1: Indian Pines

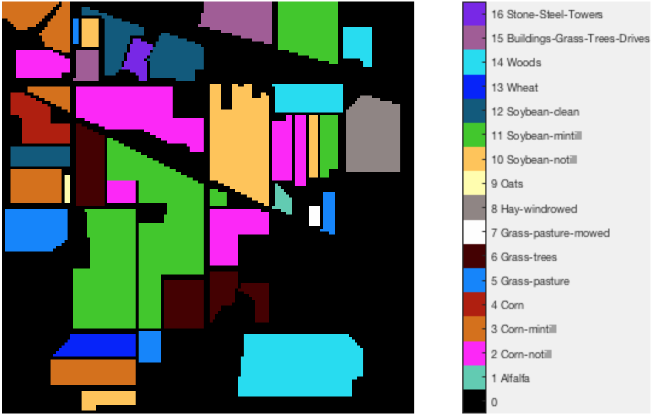


Figure : Ground-truth for the Indian Pines dataset.

Experiment setup

Supervised classification: design a classifier that uses a small portion of the data as training and classifies the remaining data points.

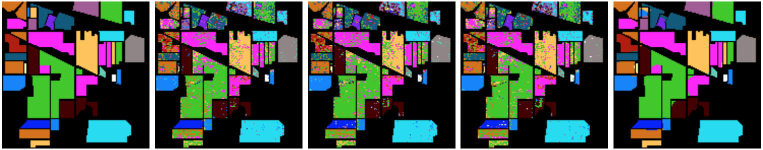
Goal: Compare the hyperspectral image classification performance of five algorithms: raw spectral, 1D WST, 1D FST, 3D WST ([Tang, Lu, Yuan]), and 3D FST.

For each transform:

1. Transform all of the labeled samples in the hyperspectral image.
2. Pick $n\%$ from each category of the labeled data as training samples.
3. Use the training samples to calculate a “1 vs. 1” support vector machine (SVM) classifier.
4. Classify the remaining $(1-n)\%$ labeled points using the SVM and compute the classification accuracies.

Indian Pines

Figure : Classification results on Indian Pines using 10% of the labeled points for training.



(a) Ground truth (b) Spectral (c) 1D WST (d) 1D FST (e) 3D FST

Metric	Raw Spectral	1D WST	1D FST	3D WST	3D FST
OA:	76.16 (±0.85)	74.53 (±0.70)	76.82 (±0.61)	94.46 (± 0.79)	98.12 (± 0.29)
AA:	71.64 (±1.97)	69.40 (±1.59)	71.46 (±2.30)	89.37 (± 3.35)	96.48 (± 1.26)

Norbert Wiener Center for Harmonic Analysis and Applications

Example 2: Pavia University

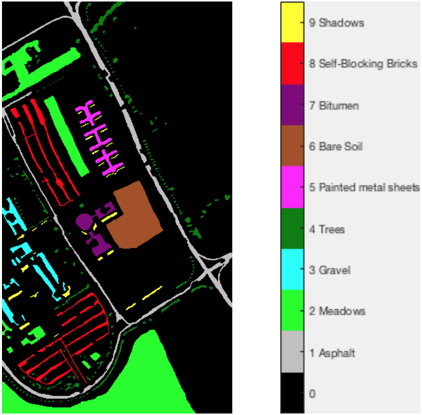
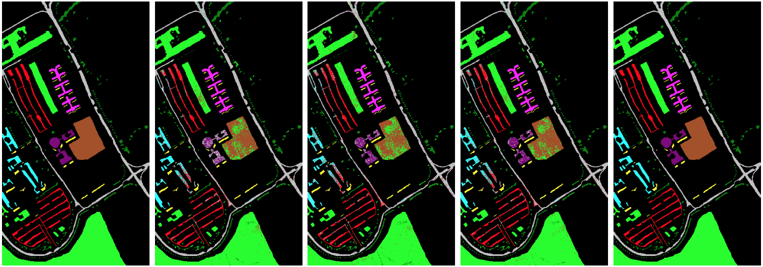


Figure : Ground-truth for the Pavia University dataset.

Pavia University

Figure : Classification results on Pavia University using 10% of the labeled points as training.



(a) Ground truth (b) Spectral (c) 1D WST (d) 1D FST (e) 3D FST

Metric	Raw Spectral	1D WST	1D FST	3D WST	3D FST
OA:	90.21 ($\pm$0.20)	87.99 ($\pm$0.17)	91.35 ($\pm$0.11)	99.30 ($\pm$ 0.12)	99.77 ($\pm$ 0.08)
AA:	84.94 ($\pm$0.58)	84.50 ($\pm$0.27)	88.60 ($\pm$0.32)	98.63 ($\pm$ 0.23)	99.56 ($\pm$0.20)

Norbert Wiener Center for Harmonic Analysis and Applications

References

Wavelet scattering:

1. Stéphane Mallat. “Group invariant scattering.” *Communications on Pure and Applied Mathematics*, 2012.
2. Joan Bruna and Stéphane Mallat. “Invariant scattering convolution networks.” *IEEE transactions on Pattern Analysis and Machine Intelligence*, 2013.

Time-frequency scattering:

1. Wojciech Czaja and Weilin Li. “Analysis of time-frequency scattering transforms.” *Applied and Computational Harmonic Analysis*, 2017.
2. Wojciech Czaja and Weilin Li. “Rotationally invariant time-frequency scattering transforms.” *Submitted*, 2017.

Hyperspectral image classification using scattering:

1. Yuan Yuan Tang, Yang Lu, and Haoliang Yuan. “Hyperspectral image classification based on three-dimensional scattering wavelet transform.” *IEEE Transactions on Geoscience and Remote sensing*, 2015.
2. Wojciech Czaja, Ilya Kavalerov, and Weilin Li. “Scattering transforms for hyperspectral data analysis .” *SPIE (invited paper)*, 2018.
3. Wojciech Czaja, Ilya Kavalerov, and Weilin Li. “3D Fourier scattering transform for hyperspectral data classification.” *In Preparation*, 2018.