An Exposition of the Group Invariant Scattering Transform

Mathematics 858W: Selected Topics in Analysis; Wavelets, Time-Frequency Analysis, and Frames

David Nahmias
dnahmias@umd.edu

Abstract

This paper explores the Group Invariant Scattering Transform proposed by Stéphane Mallat in 2012.[4] A scattering transform and propagator is defined as a path-ordered product of non-linear non-commuting operators, which are each the modulus of a wavelet transform. The scattering transform is shown to be a translation invariant operator on $\mathbf{L}^2(\mathbb{R}^d)$*, which is also Lipschitz continuous to the action diffeomorphisms. Translation invariance and Lipschitz continuity to the action diffeormorphisms are desired properties of a transform since they allow for objects similar to each other also be represented as similar objects. These results present improvements to existing transformations. As such, the scattering transform has found applications in object representation and recognition. Unless otherwise cited, the results presented in this paper are presented by Stéphane Mallat in [4].*

December, 2016
Department of Mathematics, University of Maryland-College Park

1. Background and Motivation

The Group Invariant Scattering is introduced and defined by Stéphen Mallat in 2012 as a novel representation method.[4] The scattering transform is defined as translation-invariant representations of $\mathbf{L}^2(\mathbb{R}^d)$ functions, which are Lipschitz-continuous to the action of diffeomorphisms. These representations preserve high-frequency information to discriminate different types of signals. This is an improvement to the Fourier Transform which has instablities to deformations at high-frequencies.

1.1 The Fourier Transform and its limitation

Definition 1.1. The *Fourier Transform*, $\widehat{f}(\omega)$, of $f(x) \in \mathbf{L}^2(\mathbb{R}^d)$ is

$$\widehat{f}(\omega) = \int f(x)e^{-i\omega x}\mathrm{d}x, \tag{1}$$

which is also denoted as $\mathcal{F}[f(x)] = \widehat{f}(\omega)$.

Further,

Definition 1.2. Let $T_c f(x) = f(x-c)$ denote the translation of $f \in \mathbf{L}^2(\mathbb{R}^d)$ by $c \in \mathbb{R}^d$. Further, $\Phi : \mathbf{L}^2(\mathbb{R}^d) \to \mathcal{H}$, a Hilbert Space, is *translation-invariant* if

$$\Phi(T_c f) = \Phi(f)\ \forall f \in \mathbf{L}^2(\mathbb{R}^d) \text{ and } c \in \mathbb{R}^d. \tag{2}$$

Which leads to,

Lemma 1.1. *The Fourier Transform modulus, $\Phi = |\mathcal{F}[f(x)]|$, is translation invariant.*

Proof. For some translation, $c \in \mathbb{R}^d$, $\mathcal{F}[T_c f(x)] = e^{-ic\omega}\widehat{f}(\omega) = \widehat{T_c f}(\omega)$. Thus,
$|e^{-ic\omega}\widehat{f}(\omega)| = |\widehat{T_c f}(\omega)| = |\widehat{f}(\omega)|$. □

However, invariance to time-shifts is often not enough. Consider $f(x)$ to be not just translated but time-warped, $f(x-\tau(x))$. It is now needed to define the following.

Definition 1.3. A differentiable function $f : X \to \Omega$, where X and Ω are manifolds, is a *Diffeomorphism* if f is a bijection, both one-to-one, injective, and onto, surjective, and its inverse, $f^{-1} : \Omega \to X$, is also differentiable.

Further,

Definition 1.4. The weak topology on $\mathbf{C}^2$ diffeomorphisms defines a *distance* between $\mathbb{1} - \tau$ and $\mathbb{1}$ over any compact subset $\Omega \subset \mathbb{R}^d$ by

$$d_\Omega(\mathbb{1}, \mathbb{1} - \tau) = \sup_{x\in\Omega} |\tau(x)| + \sup_{x\in\Omega} |\nabla\tau(x)| + \sup_{x\in\Omega} |H\tau(x)|, \tag{3}$$

where $|\tau(x)|$ is the euclidean norm in $\mathbb{R}^d$, $|\nabla\tau(x)|$ the sup norm of the matrix $\nabla\tau(x)$ and $|H\tau(x)|$ the sup norm of the Hessian tensor. Further, let $\|\tau\|_\infty \triangleq \sup_{x\in\mathbb{R}^d} |\tau|$.

A representation $\Phi(f)$ is stable to deformations if its Euclidean norm, $\left\|\Phi(f) - \Phi(T_{\tau(x)}f(x)\right\|$ is small when the deformation is small, where the deformation is measured by $d_\Omega(\mathbb{1}, \mathbb{1} - \tau)$. So stability is achieved when the following is met.

Definition 1.5. A Translation-invariant operator Φ is *Lipschitz continuous* to the action of $\mathbf{C}^2$ diffeomorphisms if for any compact $\Omega \subset \mathbb{R}^d$ there exists $C \in \mathbb{R}^d$ such that for all $f \in \mathbf{L}^2(\mathbb{R}^d)$ supported in Ω and $\tau \in \mathbf{C}^2(\mathbb{R}^d)$

$$\left\|\Phi(f) - \Phi(T_{\tau(x)}f)\right\|_{\mathcal{H}} \leq C\,\|f\|\,(\|\nabla\tau\|_\infty + \|H\tau\|_\infty)\,, \tag{4}$$

with $(\|\nabla\tau\|_\infty + \|H\tau\|_\infty) < 1$, to ensure the deformation is invertable.

Note that the Lipschitz upper bound does not depend on the maximum translation amplitude $\|\tau(X)\|_\infty$ from (3) since Φ here is translation invariant. This Lipschitz continuity property implies that time-warping deformations are locally linearized by Φ. Further, Lipschitz continuous operators are differentiable almost everywhere. Thus, $\Phi(f) - \Phi(T_{\tau(x)}f)$ can be approximated by a linear operator if $(\|\nabla\tau\|_\infty + \|H\tau\|_\infty)$ is small. This leads to the following limitation of the Fourier Transform.

Lemma 1.2. *The Fourier Transform modulus, $\Phi = |\mathcal{F}[f(x)]|$, is not stable to deformations and thus not Lipschitz continuous.*

Proof. By example, consider a small dialation $\tau(x) = \epsilon x$, with $0 < \epsilon \ll 1$. So, $(\|\nabla\tau\|_\infty + \|H\tau\|_\infty)$ here is equivalent to $\sup_x |\nabla\tau(x)| = \epsilon$. Thus, the Lipschitz continuity condition is

$$\left\||\widehat{f}| - |\widehat{T_{\tau(x)}f}|\right\| \leq C\,\|f\|\,\epsilon, \tag{5}$$

where $\widehat{T_{\tau(x)}f} = f(\widehat{x - \epsilon x}) = f(\widehat{(1-\epsilon)}x) = \frac{\widehat{f}(\omega/(1-\epsilon))}{(1-\epsilon)}$.
This dialation shifts the frequency component at ω_0 by $\epsilon|\omega_0|$. Now considering a harmonic signal, $f(x) = g(t)\sum_n a_n \cos(n\zeta x) \xrightarrow{\mathcal{F}} \widehat{f}(\omega) = \sum_n \frac{a_n}{2}\left(\widehat{g}(\omega + n\zeta) + \widehat{g}(\omega - n\zeta)\right)$. So, after time-warping by $\tau(x)$, each partial $\widehat{g}(\omega \pm n\zeta)$ is translated by $\epsilon n\zeta$. Even with ϵ small, at higher frequencies $\epsilon n\zeta$ will be larger than the bandwidth of $\widehat{g}$. Therefore, the Euclidean distance of $|\widehat{f}|$ and $|\widehat{T_{\tau(x)}f}|$ does not decrease proportionally to ϵ with large amplitude a_n at high frequencies. Thus the condition cannot be satisfied for any $C > 0$ in (5).[1] □

The frequency displacement from $n\zeta$ to $(1-\epsilon)n\zeta$, from Lemma 1.2, would have small impact if the sinusoidal waves are replaced by localized functions whose Fourier Transforms have support that are wider at higher frequencies. This can be achieved by a wavelet transform. So, with this limitation on the Fourier Transform modulus, another modulus transform, utilizing wavelet transforms, is presented.

2. The Scattering Transform

As discussed, high frequency instabilities to deformations can be avoided by grouping frequencies into packets in $\mathbb{R}^d$ with a wavelet transform. However, a wavelet transform is traditionally not translation invariant. So, a translation-invariant operator is constructed with a scattering procedure along multiple paths, which preserves the Lipschitz stability of wavelets to the action of diffeomorphisms. A scattering propagator is first defined as a path-ordered product of non-linear and non-commuting operators, each of which computes the modulus of wavelet transform. Further expanded, a Windowed Scattering Transform is a non-expansive operator that locally integrates the scattering propagator output that also preserves translation invariance and Lipschitz continuity.

2.1 Wavelets and the Scattering Transform

Beginning with wavelets, a wavelet transform is constructed by dialating a wavelet $\psi \in \mathbf{L}^2(\mathbb{R}^d)$ with a sequence $\{a^j\}_{j\in\mathbb{Z}}$ for $a \geq 1$. Dilated wavelets are also rotated with elements, r, of finite rotation group, G. So,

Definition 2.1. A *mother wavelet ψ that is dilated by a^j and rotated by $r \in G$* is written as

$$\psi_{a^j,r}(x) = a^{-dj}\psi(a^{-j}r^{-1}x). \tag{6}$$

Where with normalized wavelets in $\mathbf{L}^1(\mathbb{R}^d)$, such that $\|\psi_{a^jr}\|_1 = \|\psi\|_1$,

$$\mathcal{F}[\psi_{a^j,r}] = \widehat{\psi}_{a^j,r}(\omega) = \widehat{\psi}(a^j r\omega). \tag{7}$$

Without loss of generality, dyadic wavelets with $a = 2$ are considered. Further, for notation, let $\lambda = (2^j r) \in 2^{\mathbb{Z}} \times G$, with $|\lambda| = 2^j$.

Definition 2.2. A *Scattering Transform* is computed with wavelets that can be written as

$$\psi(x) = e^{i\eta x}\theta(x), \tag{8}$$

where $\theta(x)$ is a low frequency window and $\mathcal{F}[\theta(x)] = \widehat{\theta}(\omega)$ is a real function centered at $\omega = 0$ with bandwidth of order π.

So, by the Fourier Transform, $\widehat{\psi}(\omega) = \widehat{\theta}(\omega - \eta)$. Thus, $\widehat{\psi}(\omega)$ is real and concentrated in a frequency window of the same bandwidth of order π but centered at $\omega = \eta$. After dilation and rotation, $\widehat{\psi}_\lambda(\omega) = \widehat{\theta}(\lambda\omega - \eta)$ covers a window centered at $\lambda^{-1}\eta$ with radius proportional to 2^{-j}, thus specifying the frequency localization and spread of $\widehat{\psi}_\lambda$.

2.2 Littlewood-Paley: Wavelets and Condition

Now, as opposed to a standard wavelet bases, the following is considered.

Definition 2.3. A *Littlewood-Paley wavelet transform* is a redundant representation which computes convolutions at all $x \in \mathbb{R}^d$, without subsampling, such that

$$W[\lambda]f(x) = f * \psi_\lambda(x) = \int f(u)\psi_\lambda(x-u)\mathrm{d}u,\ \forall x \in R^d,\ \forall \lambda \in 2^{\mathbb{Z}} \times G. \tag{9}$$

Where $\mathcal{F}[W[\lambda]f(x)] = \widehat{f}(\omega)\widehat{\psi}_\lambda(\omega) = \widehat{f}(\omega)\widehat{\psi}(\lambda\omega)$. Note then that if f is real, that is $\widehat{f}(-\omega) = \widehat{f}^*(\omega)$, and $\widehat{\psi}(\omega)$ is chosen to be real, then $W[-\lambda]f = W[\lambda]f^*$. This implies that rotations r and $-r$ are equivalent and can thus consider rotations $r \in G^+ = G \setminus \{-\mathbb{1}, \mathbb{1}\}$.
A wavelet with a finite scale, 2^J, only retains wavelets of frequencies, subbands, satisfying $2^j \leq 2^J$. The lower frequencies not captured by these wavelets are provided by a lowpass filter, ϕ, with spatial domain proportional to 2^J, $\phi_{2^J}(x) = 2^{-dJ}\phi(2^{-J}x)$. The operator, A_J is defined as $A_J f = f * \phi_{2^J}$. So,

Definition 2.4. If f is real, then $W_J f = \{A_J f, (W[\lambda]f)_{\lambda\in\Lambda_J}\}$ which is indexed by $\Lambda_J = \{\lambda = 2^j r : 2^j \leq 2^J, r \in G^+\}$. If $J = \infty$ then $(W[\lambda]f)_{\lambda\in\Lambda_\infty}\}$ with $\Lambda_\infty = 2^{\mathbb{Z}} \times G^+$. If f is complex, then all rotations, $r \in G$ are considered.

The norm $\|W_J f\|^2 = \|A_J f\|^2 + \sum_{\lambda\in\Lambda_J}\|W[\lambda]f\|^2 = \|f * \phi_{2^J}\|^2 + \sum_{\lambda\in\Lambda_J}\|f * \psi_\lambda\|^2$ and when $J = \infty$, $\|W_\infty f\|^2 = \sum_{\lambda\in\Lambda_\infty}\|W[\lambda]f\|^2$. So, W_J is a linear operator from $\mathbf{L}^2(\mathbb{R}^d)$ to a product space generated by copies of $\mathbf{L}^2(\mathbb{R}^d)$. Further, W_J defines a frame characterized by the following.

Lemma 2.1. *Littlewood-Paley Condition: If for any $J \in \mathbb{Z}$ and for almost all $\omega \in \mathbb{R}^d$, $\exists \epsilon > 0$ such that*

$$1 - \epsilon \leq |\widehat{\phi}(2^J \omega)|^2 + \frac{1}{2} \sum_{j \leq J} \sum_{r \in G} |\widehat{\psi}(2^j r \omega)|^2 \leq 1, \tag{10}$$

then W_J is a frame with bounds $1 - \epsilon$ and 1,

$$(1 - \epsilon) \|f\|^2 \leq \|W_J f\|^2 \leq \|f\|^2, \; f \in \mathbf{L}^2(\mathbb{R}^d) \tag{11}$$

and further W_J is unitary and preserves the Euclidean norm if and only if $\epsilon = 0$.

Assuming $\widehat{\psi}$ is real, $\widehat{\psi}(0) = \int \psi(x)\mathrm{d}x = 0$ and thus the wavelet has at least one vanishing moment. Further, assuming $\widehat{\phi}$ is real and symmetric, $|\widehat{\phi}(r\omega)| = |\widehat{\phi}(\omega)|$.
Through the wavelet transform integral, if $f(x)$ is scaled and rotated by $2^{\tilde{j}}\tilde{r} \in 2^{\mathbb{Z}} \times G$, $2^{\tilde{j}}\tilde{r} \circ f(x) = f(2^{\tilde{j}}\tilde{r}x)$, the wavelet transform is scaled and rotated in the following manner,

$$W[\lambda](2^{\tilde{j}}\tilde{r} \circ f) = 2^{\tilde{j}}\tilde{r} \circ W[2^{-\tilde{j}}\tilde{r}\lambda]f. \tag{12}$$

2.3 A Path-Ordered Scattering Transform

Convolutions with wavelets define operators that are Lipschitz continuous under action diffeomorphisms, since wavelets are regular and localized functions. However, a wavelet transform is not translation-invariant, where in particular $W[\lambda]f = f * \psi_\lambda$ translates when f is translated. A scattering operator computes translation-invariant representations and coefficients that remain stable under the action diffeomorphisms while also retaining high-frequency information.
Defining an operator $U[\lambda]$ on $\mathbf{L}^2(\mathbb{R}^d)$, not necessarily linear but which commutes with translations, $\int U[\lambda]f(x)\mathrm{d}x$ is translation invariant if finite. Further, $W[\lambda]f = f * \psi_\lambda$ commutes with translations but $\int W[\lambda]f(x)\mathrm{d}x = 0$ since wavelets have zero mean, that is $\int \psi(x)\mathrm{d}x = 0$.
So, to obtain such a non-zero invariant operator, $U[\lambda]f = M[\lambda]W[\lambda]f$, where $M[\lambda]$ is a non-linear demodulation that maps $W[\lambda]f$ to a lower-frequency function that has a non-zero integral, enabling informative average value to be extracted from each subband λ. Note that $M[\lambda]$ must preserve Lipschitz continuity to diffeormorphisms actions and for stability in $\mathbf{L}^2(\mathbb{R}^d)$ be non-expansive. Further, $M[\lambda]$ needs to be a pointwise operator on f and preserve norm, that is $\|M[\lambda]f\| = \|f\|$ for all $f \in \mathbf{L}^2(\mathbb{R}^d)$ and thus $|M[\lambda]f| = |f|$. So, the most regular results are obtained through defining the non-linearity $M[\lambda]$ as the complex modulus, $M[\lambda]f = |f|$.

Definition 2.5. An ordered sequence $p = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ with $\lambda_k \in 2^{\mathbb{Z}} \times G^+$ is defined as a path. With an empty path, $p = \emptyset$. Define

$$U[\lambda]f = M[\lambda]W[\lambda]f = |f * \psi_\lambda| \text{ for } f \in \mathbf{L}^2(\mathbb{R}^d). \tag{13}$$

The Path-Ordered Scattering Transform is a path-ordered product of non-commutative operators, $U[\lambda]f$, such that

$$U[p] = U[\lambda_m] \cdots U[\lambda_2]U[\lambda_1], \tag{14}$$

with $U[\emptyset] = Id$ thus, $U[\emptyset]f = f$.

The operator $U[p]$ is well defined on $\mathbf{L}^2(\mathbb{R}^d)$ since $\|U[\lambda]f\| \leq \|\psi_\lambda\|_1 \|f\| \, \forall \lambda \in 2^{\mathbb{Z}} \times G^+$.
From this, the Path-Ordered Scattering Transform is a cascade of convolutions and modulus,

$$U[p] = ||f * \psi_{\lambda_1}| * \psi_{\lambda_2}| \cdots | * \psi_{\lambda_m}|. \tag{15}$$

Each $U[\lambda]$ filters the frequency component in the subband covered by $\widehat{\psi}_\lambda$ and maps it to lower frequencies with the modulus. Scaling and rotating path p by $2^{\tilde{j}}\tilde{r} \in 2^{\mathbb{Z}} \times G$ is written as

$2^{\tilde{j}}\tilde{r}p = (2^{\tilde{j}}\tilde{r}\lambda_1, 2^{\tilde{j}}\tilde{r}\lambda_2, \ldots, 2^{\tilde{j}}\tilde{r}\lambda_m)$. Further, combining path p with path $\tilde{p}$ is written as $p + \tilde{p} = (\lambda_1, \lambda_2, \ldots, \lambda_m, \tilde{\lambda}_1, \tilde{\lambda}_2, \ldots, \tilde{\lambda}_m)$. Thus, by (14),

$$U[p + \tilde{p}] = U[\tilde{p}]U[p]. \tag{16}$$

Definition 2.6. Let $\mathcal{P}_\infty$ be the set of all finite paths. *The Integral Scattering Transform* of $f \in \mathbf{L}^1(\mathbb{R}^d)$ is defined for any $p \in \mathcal{P}_\infty$ by

$$\overline{S}f(p) = \frac{1}{\mu_p}\int U[p]f(x)\mathrm{d}x, \tag{17}$$

where $\mu_p = \int U[p]\delta(x)\mathrm{d}x$, a non-vanishing normalization factor resulting from further development of a path measure.

So, a scattering is a translation-invariant operator that transforms $f \in \mathbf{L}^1(\mathbb{R}^d)$ into a function with frequency path variable p. If $p \neq \emptyset$, then $\overline{S}f(p)$ is non-linear but preserves amplitude factor, $\overline{S}(\mu f)(p) = |\mu|\overline{S}f(p)$, $\forall \mu \in \mathbb{R}$.

Lemma 2.2. *For some scaling and rotation $2^{\tilde{j}}\tilde{r} \in 2^{\mathbb{Z}} \times G$, the Integral Scattering Transform of a scaled and rotated f is,*

$$\overline{S}(2^{\tilde{j}}\tilde{r} \circ f)(p) = 2^{-d\tilde{j}}\overline{S}f(2^{-\tilde{j}}\tilde{r}p). \tag{18}$$

Proof. If f is scaled and rotated by $2^{\tilde{j}}\tilde{r} \in 2^{\mathbb{Z}} \times G$, that is $2^{\tilde{j}}\tilde{r} \circ f(x) = f(2^{\tilde{j}}\tilde{r}x)$, then by (12), $U[\lambda](2^{\tilde{j}}\tilde{r} \circ f) = 2^{\tilde{j}}\tilde{r} \circ U[2^{-\tilde{j}}\tilde{r}\lambda]f$. Through cascading, this yields $U[p](2^{\tilde{j}}\tilde{r} \circ f) = 2^{\tilde{j}}\tilde{r} \circ U[2^{-\tilde{j}}\tilde{r}p]f$, $\forall p \in \mathcal{P}_\infty$. Applying this to (17) shows that $\overline{S}(2^{\tilde{j}}\tilde{r} \circ f)(p) = 2^{-d\tilde{j}}\overline{S}f(2^{-\tilde{j}}\tilde{r}p)$. □

A direct extension of the Integral Scattering Transform in $\mathbf{L}^2(\mathbb{R}^d)$ is a limit of windowed Scattering Transforms.

Definition 2.7. Let $J \in \mathbb{Z}$ and $\mathcal{P}_J$ be a set of finite paths $p = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ with $\lambda_k \in \Lambda_J$ and thus $2^{j_k} \leq 2^J$. A *Windowed Scattering Transform* is defined for all $p \in \mathcal{P}_J$ by

$$S_J[p]f(x) = U[p]f * \phi_{2^J}(x) = \int U[p]f(u)\phi_{2^J}(x-u)\mathrm{d}u. \tag{19}$$

The convolution with $\phi_{2^J}(x) = 2^{-dJ}\phi(2^{-J}x)$ localizes the scattering transform over spatial domains of size proportional to 2^J:

$$S_J[p]f(x) = ||f * \psi_{\lambda_1}| * \psi_{\lambda_2}| \cdots | * \psi_{\lambda_m}| * \phi_{2^J}(x). \tag{20}$$

This defines an infinite family of functions indexed by $\mathcal{P}_J$, denoted by $S_J[\mathcal{P}_J] \triangleq \{S_J[p]f\}_{p \in \mathcal{P}_J}$. Since $\phi(x)$ is continuous at 0, if $f \in \mathbf{L}^1(\mathbb{R}^d)$, then its Windowed Scattering Transform converges poitwise when the scale 2^J goes to ∞.

$$\lim_{J\to\infty} 2^{dJ}S_J[p]f(x) = \phi(0)\int U[p]f(u)\mathrm{d}u = \phi(0)\mu_p\overline{S}(p),\ \forall x \in \mathbb{R}^d \tag{21}$$

where recall $\mu_p = \int U[p]\delta(x)\mathrm{d}x$ is a non-vanishing normalization factor. The limit and convergence of the Windowed Scattering Transform is formalized in [4] by defining a measure over the path.

2.4 Scattering Propagation

Let $S_J[\Omega] \triangleq \{S_J[p]\}_{p\in\Omega}$ and $U[\Omega] \triangleq \{U[p]\}_{p\in\Omega}$ be the set of operators determined by path set Ω. The Windowed Scattering Transform can be computed through the following procedure.

Definition 2.8. A *one-step propagator* is

$$U_J f = \{A_J f, (U[\lambda]f)_{\lambda \in \Lambda_J}\}, \tag{22}$$

where recall $A_J f = f * \phi_{2^J}$ and $U[\lambda]f = |f * \psi_\lambda|$.

So after computing $U_J f$, U_J is again applied to each $U[\lambda]f$ and recursively to each $U[p]f$. Since $U[\lambda]U[p] = U[\lambda + p]$ from (16) and $A_J U[p] = S_J[p]$ from (19),

$$U_J U[p]f = \{S_J[p]f, (U[p+\lambda]f)_{\lambda \in \Lambda_J}\}. \tag{23}$$

Now let Λ_J^m be the set of paths of length m with $\Lambda_J^0 = \{\emptyset\}$. Then,

$$U_J U[\Lambda_J^m]f = \{S_J[\Lambda_J^m]f, (U[\Lambda_J^{m+1}]f)_{\lambda \in \Lambda_J}\}. \tag{24}$$

So, since $\mathcal{P}_J = \bigcup_{m \in \mathbb{N}} \Lambda_J^m$, the Windowed Scattering Transform of f, $S_J[\mathcal{P}_J]f$, is obtained from $f = U[\emptyset]f$ by iteratively computing $U_J U[\Lambda_J^m]f_{m=0 \to \infty}$. This procedure can be seen in Figure 1.

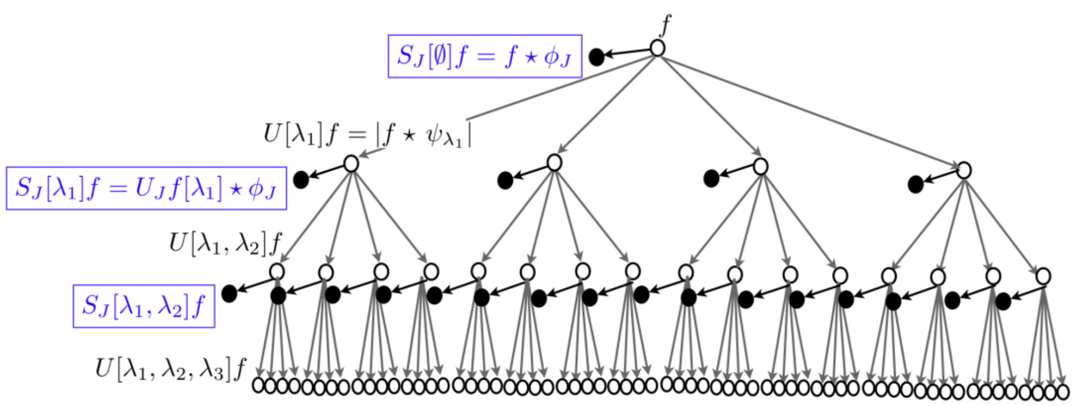

Figure 1: *A scattering propagator U_J applied to f computes each $U[\lambda_1]f = |f * \psi_{\lambda_1}|$ and outputs $S[\emptyset]f = f * \phi_{2^J}$. Applying U_J to each $U[\lambda_1]f$ computes all $U[\lambda_1, \lambda_2]f$ and outputs $S_J[\lambda_1] = U[\lambda_1] * \phi_{2^J}$. Applying iteratively U_J to each $U[p]f$ outputs $S_J[p]f = U[p]f * \phi_{2^J}$ and computes the next path layer.[4]*

Note that all results and definitions of the Path-Ordered Scattering Transform and Windowed Scattering Transform are extendable to complex functions by also considering negative paths, denoted as $-p = (-\lambda_1, \lambda_2, \ldots, \lambda_m)$. Where if f is real, $W[-\lambda_1]f = W[\lambda_1]f^* \Rightarrow U[-\lambda_1]f = U[\lambda_1]f \Rightarrow U[-p]f = U[p]f \Rightarrow S_J[-p]f = S_J[p]f$.

3. Properties of The Windowed Scattering Transform

3.1 Norm Preservation of The Windowed Scattering Transform

To preserve stability in $\mathbf{L}^2(\mathbb{R}^d)$ operators, Φ, need to be non-expansive.

Lemma 3.1. *The propagator $U_J f = \{A_J f, (W[\lambda]f)_{\lambda \in \Lambda_J}\}$ is non-expansive and preserves norm.*

Proof. The wavelet transform W_J is unitary, by Lemma 2.1, and a modulus is non-expansive since $||a| - |b|| \leq |a - b|, \forall (a,b) \in \mathbb{C}^2$. So, for $f, h \in \mathbb{R}$ or $f, h \in \mathbb{C}$,

$$\begin{aligned} \|U_J f - U_J h\|^2 = & \quad \|A_J f - A_J h\|^2 + \textstyle\sum_{\lambda \in \Lambda_J} \||W[\lambda]f| - |W[\lambda]h|\| \\ \leq & \quad \|W_J f - W_J h\|^2 \\ \leq & \quad \|f - h\|^2 . \end{aligned} \tag{25}$$

Further, since W_J is unitary, with $h = 0$, U_J preserves the norm, $\|U_j f\| = \|f\|$.

□

So, for any path set Ω,

$$\|S_J[\Omega]f\|^2 = \sum_{p\in\Omega} \|S_J[p]f\|^2 \text{ and } \|U[\Omega]f\|^2 = \sum_{p\in\Omega} \|U[p]f\|^2 . \tag{26}$$

Therefore,

Theorem 3.2. *The Windowed Scattering Transform is non-expansive,*

$$\|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\| \leq \|f - h\|, \ \forall (f,h) \in \mathbf{L}^2(\mathbb{R}^d). \tag{27}$$

Proof. U_J is non-expansive by Lemma 3.1. So,

$$\begin{aligned} \left\|U[\Lambda_J^m]f - U[\Lambda_J^m]h\right\|^2 \geq & \left\|U_J U[\Lambda_J^m]f - U_J U[\Lambda_J^m]h\right\|^2 \\ = & \left\|S_J[\Lambda_J^m]f - S_J[\Lambda_J^m]h\right\|^2 + \left\|U[\Lambda_J^{m+1}]f - U[\Lambda_J^{m+1}]h\right\|^2 \\ = & \|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\|^2 , \text{ by (24) and scattering propagation.} \end{aligned} \tag{28}$$

So summing over $m = 0 \to \infty$,

$$\begin{aligned} \|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\|^2 = & \textstyle\sum_{m=0}^{\infty} \left\|S_J[\Lambda_J^m]f - S_J[\Lambda_J^m]h\right\|^2 \\ \leq & \|f - h\|^2 \end{aligned} \tag{29}$$

yielding the desired result. □

The conditions for when the Windowed Scattering Transform preserves norm follow.

Theorem 3.3. *A scattering wavelet, ψ, is admissible if $\exists \eta \in \mathbb{R}^d$ and $0 \leq \rho \in \mathbf{L}^2(\mathbb{R}^d)$, with $|\widehat{\rho}(\omega)| \leq |\widehat{\phi}(2\omega)|$ and $\widehat{\rho}(0) = 1$, such that*

$$\widehat{\Psi}(\omega) = |\widehat{\rho}(\omega - \eta)|^2 - \sum_{k=1}^{\infty} k(1 - |\widehat{\rho}(2^{-k}(\omega - \eta))|^2) \tag{30}$$

satisfies

$$\inf_{1\leq|\omega|\leq 2} \sum_{j=-\infty}^{\infty} \sum_{r\in G} \widehat{\Psi}(2^{-j}r^{-1}\omega)|\widehat{\psi}(2^{-j}r^{-1}\omega)|^2 > 0. \tag{31}$$

Further, if the wavelet, ϕ, satisfies the Littlewood-Paley Condition in Lemma 2.1, with $\epsilon = 0$, and is admissible, then $\forall f \in \mathbf{L}^2(\mathbb{R}^d)$,

$$\lim_{m\to\infty} \left\|U[\Lambda_J^m]f\right\|^2 = \lim_{m\to\infty} \sum_{n\geq m}^{\infty} \left\|S_J[\Lambda_J^n]f\right\|^2 = 0 \tag{32}$$

and

$$\|S_J[\mathcal{P}_J]f\| = \|f\| . \tag{33}$$

The proof of this theorem shows that scattering transform propagates progressively energy towards lower frequencies because of the demodulation effect of the modulus. Further, this fact leads to a Lemma that is used in the proof of Theorem 3.3.

Lemma 3.4. *If the condition in Theorem 3.3 is satisfied and*

$$\|f\|_w^2 = \sum_{j=0}^{\infty} \sum_{r \in G^+} j \left\| W[2^j r] f \right\|^2 < \infty \tag{34}$$

then,

$$\frac{\alpha}{2} \left\| U[\mathcal{P}_J] f \right\|^2 \leq \max(J+1, 1) \|f\|^2 + \|f\|_w^2, \tag{35}$$

where $\alpha = \inf_{1 \leq |\omega| \leq 2} \sum_{j=-\infty}^{\infty} \sum_{r \in G} \widehat{\Psi}(2^{-j} r^{-1} \omega) |\widehat{\psi}(2^{-j} r^{-1} \omega)|^2$, *(31) from Theorem 3.3.*

As a note, the class of functions where $\|f\|_w < \infty$ is called a logarithmic Sobolev class which corresponds here to functions that have an average modulus of continuity in $\mathbf{L}^2(\mathbb{R}^d)$.

3.2 Translation Invariance of The Windowed Scattering Transform

The scattering distance is defined as $\left\| S_J[\mathcal{P}_J] f - S_J[\mathcal{P}_J] h \right\|$. It is non-increasing as J increases and thus defines a limit distance that converges as $J \to \infty$. So,

Lemma 3.5. *For all* $(f, h) \in \mathbf{L}^2(\mathbb{R}^d)^2$ *and* $J \in \mathbb{Z}$,

$$\left\| S_{J+1}[\mathcal{P}_{J+1}] f - S_{J+1}[\mathcal{P}_{J+1}] h \right\| \leq \left\| S_J[\mathcal{P}_J] f - S_J[\mathcal{P}_J] h \right\|. \tag{36}$$

Further, since $S_J[\mathcal{P}_J]$ is non-expansive from Theorem 3.2, the limit metric is also non-expansive and thus

$$\lim_{J \to \infty} \left\| S_J[\mathcal{P}_J] f - S_J[\mathcal{P}_J] h \right\| \leq \|f - h\|, \ \forall (f, h) \in \mathbf{L}^2(\mathbb{R}^d)^2. \tag{37}$$

In addition, for admissible scattering wavelets, from Theorem 3.3, $\left\| S_J[\mathcal{P}_J] f \right\| = \|f\|$ and so $\lim_{J \to \infty} \left\| S_J[\mathcal{P}_J] f \right\| = \|f\|$. To show the limit metric is translation invariant, the following preliminary results are needed.

Lemma 3.6. *Schur's Lemma: For any operator* $Kf(x) = \int f(u) k(x, u) \mathrm{d}u$, *if*

$$\int |k(x,u)| \mathrm{d}x \leq C \text{ and } \int |k(x,u)| \mathrm{d}u \leq C \Rightarrow \|K\| \leq C, \tag{38}$$

where $\|K\|$ *is the* $\mathbf{L}^2(\mathbb{R}^d)$ *norm of operator* K.

Lemma 3.6 can be proved through application of a variation of the Cauchy-Schwartz Inequality and Tonelli's Theorem.[3] Further,

Lemma 3.7. *There exists* C *such that for all* $\tau \in \mathbf{C}^2(\mathbb{R}^d)$ *with* $\|\nabla \tau\|_\infty \leq 1/2$,

$$\left\| T_\tau A_J f - A_J f \right\| \leq C \|f\| \, 2^{-J} \|\tau\|_\infty. \tag{39}$$

Lemma 3.7 is proved through the application of Lemma 3.6 on of the operator norm of $k_J = T_\tau A_J - A_J$ and on first-order Taylor expansions of $|k_J|$.
With these results, it is proved that the limit metric of the Windowed Scattering Transform is translation invariant.

Theorem 3.8. *For all admissible scattering wavelets satisfying Theorem 3.3,*

$$\lim_{J \to \infty} \left\| S_J[\mathcal{P}_J] T_c f - S_J[\mathcal{P}_J] f \right\| = 0, \ \forall f \in \mathbf{L}^2(\mathbb{R}^d) \text{ and } \forall c \in \mathbb{R}^d. \tag{40}$$

Proof. With $S_J[\mathcal{P}_J]T_c = T_c S_J[\mathcal{P}_J]$ and $S_J[\mathcal{P}_J]f = A_J U[\mathcal{P}_J]f$,

$$\begin{aligned} \left\|S_J[\mathcal{P}_J]T_c f - S_J[\mathcal{P}_J]f\right\| = & \quad \left\|T_c A_J U[\mathcal{P}_J]f - A_J U[\mathcal{P}_J]f\right\| \\ \leq & \quad \left\|T_c A_J - A_J\right\| \left\|U[\mathcal{P}_J]f\right\|, \text{ by Hölder's Inequality.} \end{aligned} \tag{41}$$

Now, applying Lemma 3.7 with $\tau = c$, $\|\nabla\tau\|_\infty = |c|$ and

$$\left\|T_\tau A_J f - A_J f\right\| \leq C \|f\| 2^{-J}|c|. \tag{42}$$

Applying (42) to (41) yields,

$$\left\|S_J[\mathcal{P}_J]T_c f - S_J[\mathcal{P}_J]f\right\| \leq C \|f\| 2^{-J}|c| \left\|U[\mathcal{P}_J]f\right\|. \tag{43}$$

Since the conditions from Theorem 3.3 are assumed to be satisfied, from Lemma 3.4, for $J > 1$ and $\|f\|_w < \infty$, applying (35) to (43) obtains,

$$\left\|S_J[\mathcal{P}_J]T_c f - S_J[\mathcal{P}_J]f\right\| \leq C^2 \|f\|^2 2^{-2J}|c|^2 \left((J+1)\|f\|^2 + \|f\|_w^2\right) \frac{2}{\alpha} \tag{44}$$

Taking $J \to \infty$,

$$\lim_{J\to\infty} C^2 \|f\|^2 2^{-2J}|c|^2 \left((J+1)\|f\|^2 + \|f\|_w^2\right) \frac{2}{\alpha} = 0. \tag{45}$$

Thus, for $\|f\|_w < \infty$, $\lim_{J\to\infty} \left\|S_J[\mathcal{P}_J]T_c f - S_J[\mathcal{P}_J]f\right\| = 0$.
In general though, for all $f \in \mathbf{L}^2(\mathbb{R}^d)$, consider that any such $f \in \mathbf{L}^2(\mathbb{R}^d)$ can be written as a combination and limit of $\{f_n\}_{n\in\mathbb{N}}$, where $f_n = f * \phi_{2^n}$ with $\phi_{2^n}(x) = 2^{-nd}\phi(2^{-n}x)$. Further, $\|f_n\|_w < \infty$. So, with $S_J[\mathcal{P}_J]$ non-expansive from Theorem 3.2, implying $\left\|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f_n\right\| \leq \|f - f_n\|$, and with T_c is unitary,

$$\begin{aligned} \left\|S_J[\mathcal{P}_J]f\right\| \leq & \quad \left\|S_J[\mathcal{P}_J]f_n\right\| + \|f - f_n\| \\ \left\|S_J[\mathcal{P}_J]T_c f\right\| \leq & \quad \left\|S_J[\mathcal{P}_J]T_c f_n\right\| + \|f - f_n\| \\ \Rightarrow \left\|S_J[\mathcal{P}_J]T_c f - S_J[\mathcal{P}_J]f\right\| \leq & \quad \left\|S_J[\mathcal{P}_J]T_c f_n - S_J[\mathcal{P}_J]f_n\right\| + 2\|f - f_n\|. \end{aligned} \tag{46}$$

Since $\phi \in \mathbf{L}^1(\mathbb{R}^d)$ and $\widehat{\phi}(0) = 1$, $\lim_{n\to\infty} \|f - f_n\| = 0$, $\forall f \in \mathbf{L}^2(\mathbb{R}^d)$. Thus, as $n \to \infty$,

$$\lim_{J\to\infty} \left\|S_J[\mathcal{P}_J]T_c f - S_J[\mathcal{P}_J]f\right\| = 0, \ \forall f \in \mathbf{L}^2(\mathbb{R}^d), \tag{47}$$

thus proving the theorem. □

3.3 Lipschitz Continuity of The Windowed Scattering Transform

Finally coming to the Fourier Transform modulus' limitation shown in Lemma 1.2, the Windowed Scattering Transform is Lipschitz continuous under the action of diffeomorphisms. The diffeomorphism action on $f \in \mathbf{L}^2(\mathbb{R}^d)$ is $T_{\tau(x)}f(x) = f(x - \tau(x))$. The maximum increment of τ is written as,

$$\|\Delta\tau\|_\infty \triangleq \sup_{(x,u)\in\mathbb{R}^{2d}} |\tau(x) - \tau(u)|. \tag{48}$$

The following theorem determines the upper bound of $\left\|S_J[\mathcal{P}_J]T_\tau f - S_J[\mathcal{P}_J]f\right\|$ as a function of the scattering norm, $\left\|U[\mathcal{P}_J]f\right\|_1 = \sum_{m=0}^{\infty} \left\|U[\Lambda_J^m]f\right\|$.

Theorem 3.9. *There exits C such that $\forall f \in \mathbf{L}^2(\mathbb{R}^d)$ with $\|U[\mathcal{P}_J]f\|_1 < \infty$ and all $\tau \in \mathbf{C}^2(\mathbb{R}^d)$ with $\|\nabla\tau\|_\infty \leq 1/2$ satisfy*

$$\|S_J[\mathcal{P}_J]T_\tau f - S_J[\mathcal{P}_J]f\| \leq C \|U[\mathcal{P}_J]f\|_1 K(\tau) \tag{49}$$

with

$$K(\tau) = 2^{-J}\|\tau\|_\infty + \|\nabla\tau\|_\infty \left(\max\{\log\frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty}, 1\}\right) + \|H\tau\|_\infty \tag{50}$$

and for all $m \geq 0$,

$$\|S_J[\mathcal{P}_{J,m}]T_\tau f - S_J[\mathcal{P}_{J,m}]f\| \leq Cm\|f\| K(\tau) \tag{51}$$

where $\mathcal{P}_{J,m}$ is a subset of $\mathcal{P}_J$ of paths of length strictly smaller than m.

Further, a Windowed Scattering Transform is also Lipschitz continuous under the action diffeomorphisms over compactly supported functions.

Corollary 3.10. *For any compact $\Omega \subset \mathbb{R}^d$, $\exists C$ such that $\forall f \in \mathbf{L}^2(\mathbb{R}^d)$ in the support of Ω, with $\|U[\mathcal{P}_J]f\|_1 < \infty$ and $\forall\tau \in \mathbf{C}^2(\mathbb{R}^d)$ with $\|\nabla\tau\|_\infty \leq 1/2$, if $\|\tau\|_\infty/\|\nabla\tau\|_\infty \leq 2^J$, then*

$$\|S_J[\mathcal{P}_J]T_\tau f - S_J[\mathcal{P}_J]f\| \leq C\|U[\mathcal{P}_J]f\|_1 (2^{-J}\|\tau\|_\infty + \|\nabla\tau\|_\infty + \|H\tau\|_\infty) \tag{52}$$

Finally, the translation error term, $2^{-J}\|\tau\|_\infty$, in Theorem 3.9 and Corollary 3.10, can be reduced to a second order term through a first-order Taylor expansion of each $S_J[p]f$.

Theorem 3.11. *There exits C such that $\forall f \in \mathbf{L}^2(\mathbb{R}^d)$ with $\|U[\mathcal{P}_J]f\|_1 < \infty$ and all $\tau \in \mathbf{C}^2(\mathbb{R}^d)$ with $\|\nabla\tau\|_\infty \leq 1/2$ satisfy*

$$\|S_J[\mathcal{P}_J]T_\tau f - S_J[\mathcal{P}_J]f + \tau\cdot\nabla S_J[\mathcal{P}_J]f\| \leq C\|U[\mathcal{P}_J]f\|_1 K(\tau) \tag{53}$$

with

$$K(\tau) = 2^{-2J}\|\tau\|_\infty^2 + \|\nabla\tau\|_\infty \left(\max\{\log\frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty}, 1\}\right) + \|H\tau\|_\infty, \tag{54}$$

where $\tau\cdot\nabla S_J[\mathcal{P}_J]f(x) \triangleq \{\tau(x)\cdot\nabla S_J[p]f(x)\}_{p\in\mathcal{P}_J}$.

4. Extensions and Conclusion

Several extensions are done on these fundamental definitions and results. In order to formally define a limit of the Windowed Scattering Transform, which is used to define the Integral Scattering Transform, a measure and metric is constructed over the path. This also yields the normalization found in Definition 2.6. Further, the invariant scattering transform is extended to actions of compact Lie groups, defining a scattering operator on $\mathbf{L}^2(SO(d))$. Combining this extension with results on $\mathbf{L}^2(\mathbb{R}^d)$ allows for the scattering transform to be extended to be translation and rotation invariant while maintaining Lipschitz continuity. The Windowed Scattering Transform and propagation calculations follow the general architecture of convolution neural networks. Convolution networks cascade convolutions and a pooling non-linearity, which here is the complex modulus. However, convolution networks typically use kernels that are not predefined functions such as wavelets but which are learned with back-propagation algorithms. The scattering transform has been adapted by Mallat to preform convolution network operations and has been shown to have significant applications in object representation and recognition.[2] In applications, the decay of $\sum_{n\geq m}^{\infty}\left\|S_J[\Lambda_J^n]f\right\|^2$ from Theorem 3.3 implies that all paths of length larger than some $m > 0$ can be omitted. This decay has mostly limited $m = 3$ for classification applications.

In conclusion, the Scattering Transform presents an operator that is non-expansive, norm-preserving and stable to deformations. These conditions allow for improvements over the Fourier Transform, particularly in applications of representation and recognition, where small deformations should not greatly influence the transformed object.

References

[1] J. Andén and S. Mallat. Deep scattering spectrum. *IEEE Transactions on Signal Processing*, 2014.

[2] J. Bruna and S. Mallat. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1872–1886, 2013.

[3] C. Heil. *Five Mini-Courses on Analysis*. Birkhäuser, in preparation.

[4] S. Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.