Towards Time-Frequency Deformation Stability Bounds for Deep Convolutional Neural Networks

Albert Chua
Department of Mathematics
Michigan State University
East Lansing, MI
chuaalbe@msu.edu

***Abstract*—In this paper, we examine deformations from the lens of an operator similar to Fourier Integral Operators (FIOs). First, we provide a generalization of the deformation operator seen first in Mallat's seminal paper Group Invariant Scattering, which we call a time-frequency deformation. Next, we provide a generalized deformation stability bound for our time-frequency deformation in the specific case when the time and frequency portions of the deformation are separable for the deep feature extractor considered by Wiatowski and Bölcskei as well as Mallat's Windowed Scattering Transform.**

***Index Terms*—Machine Learning, Deep Convolutional Neural Network, Wavelet Scattering Transform, Time-Frequency Deformations**

I. Introduction

A. Notation

The inner product between $x, y \in \mathbb{C}^n$ is given by $\langle x, y\rangle := \sum_{i=1}^n x_i, \overline{y}_i$ where $\overline{y}_i$ denotes the complex conjugate of y_i. For a function f, we have Df and $D^2 f$ to denote the Jacobian and Hessian of f, respectively. The space $\mathbf{L}^p(\mathbb{R}^n)$, with $1 \leq p < \infty$, is the set of functions such that $\|f\|_p^p := \int_{\mathbb{R}^n} |f(x)|^p \, dx < \infty$, where dx denotes integration with respect to Lebesgue measure. For $p = \infty$, we let $\|f\|_\infty :=$ ess sup$|f|$, where ess sup denotes the supreumum of f, excluding sets of Lebesgue measure zero. For $m \in \mathbb{N}$, define the function space $C^m(\mathbb{R}^n)$ as the set of functions that are m-times continuously differentiable; when $m = 0$, $C^m(\mathbb{R}^n) = C(\mathbb{R}^n)$ refers to the space of continuous functions. The convolution of two functions is defined as $(f * g)(x) = \int_{\mathbb{R}^n} f(y)g(x-y)\, dy$. The Fourier transform of a function $f \in \mathbf{L}^2(\mathbb{R}^n) \cup \mathbf{L}^1(\mathbb{R}^n)$ is given by $\hat{f}(\xi) = \int_{\mathbb{R}^n} e^{-i\langle x, \xi\rangle} f(x)\, dx$. We will also use $\mathcal{F}$ and $\mathcal{F}^{-1}$ for the Fourier Transform, and inverse Fourier Transform, respectively. Lastly, define the translation operator $(T_t f)(x) := f(x-t), t \in \mathbb{R}^n$, flip operator $(Rf)(x) := f(-x)$, and involution operator $(If)(x) := \overline{f(-x)}$.

B. Preliminaries

In recent years, deep convolutional neural networks (deep CNNs) have been utilized to produce state of the art results for a variety of object recognition tasks [1]. At a high level, deep CNNs perform feature extraction via a cascade of convolutions with a filter, applying a pointwise nonlinearity, and downsample via a pooling operation. However, due to the complexity of deep CNNs, it has been a challenge for researchers to explore the intrinsic properties that allow them to achieve state of the art results.

In his seminal paper [2], Mallat proposed the Scattering Transform, a simple form of a convolutional neural network using a cascade of convolutions via handcrafted wavelet filters instead of learned filters and using the complex modulus as a pointwise nonlinearity. The Scattering Transform, and its variants [3]–[11], have intrinsic properties that are desirable to deep CNNs for various signal processing tasks, such as translation invariance and stability to small deformations with respect to an appropriate norm while also being applicable to practical machine learning tasks.

Wiatowski and Bölcskei [12] extended Mallat's framework to include general semi-discrete frames generated via functions such as curvelets, shearlets, and ridgelets while also allowing for general Lipschitz continuous pointwise nonlinearties, such as ReLU. In a smilar fashion to Mallat, [12] were able to alternative establish deformation sensitivity and vertical translation invariance bounds for $f \in \mathbf{L}^2(\mathbb{R}^n)$, which depend only on the network structure and not the regularity of the filter bank. Additionally, further study has been conducted on extending deformation stability results for the scattering transform to $C^\alpha(\mathbb{R}^n)$ deformations with $1 < \alpha < 2$ in [13], providing stronger stability results in the architecture in [12] for the space of Cartoon-like functions [14], and results for Sobolev spaces in [15].

II. A CNN-like Feature Extractor

In this section, we provide a brief summary of the CNN-like feature extractor described in [12].

Definition 1. *Let $\{g_\lambda\}_{\lambda\in\Lambda}$ with $g_\lambda \in \mathbf{L}^1(\mathbb{R}^n) \cap \mathbf{L}^2(\mathbb{R}^n)$, where Λ is a countable index set. The set $\Psi_\Lambda = \{T_b I g_\lambda\}_{(b,\lambda)\in\mathbb{R}^n\times\Lambda}$ is a semi-discrete frame for $\mathbf{L}^2(\mathbb{R}^n)$ if and only if there exist $A, B > 0$ such that*

$$A\|f\|_2^2 \leq \sum_{\lambda\in\Lambda} \int_{\mathbb{R}^n} |\langle f, T_b I g_\lambda\rangle|^2 \, db \leq B\|f\|_2^2. \tag{1}$$

Next, we define our pooling operator, which is a continuous version of the average pooling operator.

Definition 2. *Define the pooling operator as*

$$f \mapsto S^{n/2} P(f)(S\cdot) \tag{2}$$

for $f \in L^2(\mathbb{R}^n)$, where P is a Lipschitz continuous function satisfying $Pf = 0$ if $f = 0$ and S is a subsampling factor.

Next, we define the building blocks for the CNN-like feature extractor.

Definition 3. *For $k \in \mathbb{N}$, let Ψ_k be a semi-discrete frame for $\mathbf{L}^2(\mathbb{R}^n)$, M_k and P_k be Lipschitz operators such that $M_k f = 0$ and $P_k f = 0$ when $f = 0$. Then*

$$\Omega := ((\Psi_k, M_k, P_k))_{k\in\mathbb{N}}$$

is known as a module-sequence. Furthermore, suppose that a module sequence has frame bounds $B_k > 0$ for Ψ_k, Lipschitz constant L_k for each M_k, and Lipschitz constant R_k for each P_k. We refer to

$$\max\{B_k, B_k L_k^2 R_k^2\} \le 1, \qquad \forall k \in \mathbb{N} \tag{3}$$

as the module-sequence admissibility condition.

Definition 4. *Let $\Omega = ((\Psi_k, M_k, P_k))_{k\in\mathbb{N}}$ be a module sequence and $\{\gamma_{\lambda_k}\}_{\lambda_k\in\Lambda_k}$ be the atoms of Ψ_k with subsampling factors $S_k \ge 1$ for all $k \in \mathbb{N}$. The operator $U_k : \Lambda_k \times \mathbf{L}^2(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^n)$ is given by*

$$U_k[\lambda_k]f := S_k^{n/2} P_k(M_k(f * g_{\lambda_k}))(S_k\cdot). \tag{4}$$

Additionally, define $\Lambda^k = \Pi_{i=1}^k \Lambda_k$ and let a path be $q = (\lambda_1, \dots, \lambda_k) \in \Lambda^k$ with $\Lambda^0 = \emptyset$ and $U_0[\emptyset] = f$.

Definition 5. *Suppose we use the feature extractor above and let $k \in \mathbb{N}$. Define*

$$\Phi_\Omega^k(f) := \{U[q]f * \chi_k\}_{q\in\Lambda^k}, \tag{5}$$

where $\chi_{k-1} = g_{\lambda_k^*}$ with $\lambda_k^* \in \Lambda_k$. The feature extractor maps f to feature vector defined by:

$$\Phi_\Omega(f) := \bigcup_{k=0}^{\infty} \Phi_\Omega^k(f). \tag{6}$$

Note that we will use the same abuse of notation in [12] here.

The norm for the operator $\Phi_\Omega : (\mathbf{L}^2(\mathbb{R}^n))^{\bigcup_{k=0}^{\infty}\Lambda^k} \to \mathbf{L}^2(\mathbb{R}^n)$ is given by

$$|||\Phi_\Omega(f)|||^2 := \sum_{k=0}^{\infty} \sum_{q\in\Lambda^k} \|U[q]f\|_2^2. \tag{7}$$

An important result for our purposes is the following non-expansiveness result:

Theorem 1 (Theorem 2, [12]). *Suppose that $\Omega = ((\Psi_k, M_k, P_k))_{k\in\mathbb{N}}$ is an admissible module-sequence and $S_k \ge 1$ for $k \in \mathbb{N}$. Then for $f, g \in \mathbf{L}^2(\mathbb{R}^n)$, we have*

$$|||\Phi_\Omega(f) - \Phi_\Omega(g)||| \le \|f - g\|_2. \tag{8}$$

III. Diffeomorphism Stability and Time Frequency Deformations

Let Φ be a translation invariant representation which acts from $\mathbf{L}^2(\mathbb{R}^n)$ to a Hilbert Space, say $\mathcal{H}$. Consider the operator that deforms f by a small deformation τ: $L_\tau f(x) := f(x - \tau(x))$, where $\tau \in C^2(\mathbb{R}^n) \cap \mathbf{L}^\infty(\mathbb{R}^n)$ with $\|D\tau\|_\infty < \frac{1}{2n}$. We say that the representation Φ is Lipschitz continuous to the action of diffeomorphisms when

$$\|\Phi f - \Phi L_\tau f\|_{\mathcal{H}} \le C(\|D\tau\|_\infty + \|D^2\tau\|_\infty)\|f\|_2 \tag{9}$$

for some constant $C > 0$. Mallat was able to prove a result for the Windowed Scattering Transform in [2]. For the feature extractor in the previous section, a similar result holds for bandlimited functions, which we will discuss later in this paper.

Regarding the deformations considered above, when $f \in \mathbf{L}^2(\mathbb{R}^n)$, via Fourier inversion, one has

$$(L_\tau f)(x) = \int_{\mathbb{R}^n} e^{i\langle \xi, x-\tau(x)\rangle} \hat{f}(\xi)\, d\xi \tag{10}$$

Notably, the form of (10) resembles a Fourier Integral Operator [16], which takes the form

$$\int_{\mathbb{R}^n} e^{i\Psi(x,\xi)} a(x,\xi) f(\xi)\, d\xi, \tag{11}$$

where $\Psi(x,\xi)$ is known as a phase function and $a(x,\xi)$ is known as an amplitude function.

With this interpretation in mind, a natural extension is to also add small frequency deformation to (10) via an amplitude function, which results in the following time-frequency deformation operator

$$K_{\tau_1,b} f(x) := \int_{\mathbb{R}^n} e^{i\langle \xi, x-\tau_1(x)\rangle} (1 + b(x,\xi)) \hat{f}(\xi)\, d\xi \tag{12}$$

with $\tau_1 \in C^2(\mathbb{R}^n) \cap \mathbf{L}^\infty(\mathbb{R}^n)$ and $\|\tau_1\|_\infty < \frac{1}{2n}$. Here, the small time-frequency deformation is the term

$$K_{\tau_1,b} f(x) := \int_{\mathbb{R}^n} e^{i\langle \xi, x-\tau_1(x)\rangle} b(x,\xi) \hat{f}(\xi)\, d\xi. \tag{13}$$

However, numerous technical challenges come with examining (12), so we will consider the case where $b(x,\xi)$ is separable and can be written as $b(x,\xi) := \tau_2(\xi)\tau_3(x)$. The corresponding operator we consider is

$$K_{\tau_1,\tau_2,\tau_3} f(x) := \int_{\mathbb{R}^n} e^{i\langle \xi, x-\tau_1(x)\rangle} (1 + \tau_2(\xi)\tau_3(x)) \hat{f}(\xi)\, d\xi \tag{14}$$

with $\tau_1 \in C^2(\mathbb{R}^n) \cap \mathbf{L}^\infty(\mathbb{R}^n)$ and $\|\tau_1\|_\infty < \frac{1}{2n}$.

For convenience in the theoretical analysis of (14), we rewrite it as

$$\begin{aligned} K_{\tau_1,\tau_2,\tau_3} f(x) &= \int_{\mathbb{R}^n} e^{i\langle \xi, x-\tau_1(x)\rangle} \hat{f}(\xi)\, d\xi \\ &\quad + \int_{\mathbb{R}^n} e^{i\langle \xi, x-\tau_1(x)\rangle} \tau_2(\xi)\tau_3(x) \hat{f}(\xi)\, d\xi \\ &:= L_{\tau_1} f(x) + Q_{\tau_1,\tau_2,\tau_3} f(x) \end{aligned} \tag{15}$$

with

$$Q_{\tau_1,\tau_2,\tau_3}f(x) := \int_{\mathbb{R}^n} e^{i\langle\xi, x-\tau_1(x)\rangle}\tau_2(\xi)\tau_3(x)\hat{f}(\xi)\,d\xi. \quad (16)$$

The first lemma below provides an $\mathbf{L}^2$ boundedness result of Q_{τ_1,τ_2,τ_3} in terms of $\|\tau_2\|_\infty$ and $\|\tau_3\|_\infty$.

Lemma 1. *Suppose that $\tau_2 \in \mathbf{L}^1(\mathbb{R}^n) \cap \mathbf{L}^\infty(\mathbb{R}^n) \cap C(\mathbb{R}^n)$ and $\tau_3 \in \mathbf{L}^\infty(\mathbb{R}^n)$. We have*

$$\|Q_{\tau_1,\tau_2,\tau_3}f\|_2 \le \sqrt{2}\|\tau_3\|_\infty\|\tau_2\|_\infty\|f\|_2.$$

Proof. Since $Q_{\tau_1,\tau_2,\tau_3}f \in \mathbf{L}^2(\mathbb{R}^n)$, we notice that $Q_{\tau_1,\tau_2,\tau_3}f$ is an inverse Fourier Transform, which via Convolution Theorem, results in

$$\begin{aligned}(Q_{\tau_1,\tau_2,\tau_3}f)(x) &= \tau_3(x)(\mathcal{F}^{-1}(\tau_2) * f)(x - \tau_1(x)) \\ &= \tau_3(x)(L_{\tau_1}(\mathcal{F}^{-1}(\tau_2) * f))(x).\end{aligned}$$

For general $g \in \mathbf{L}^2(\mathbb{R}^n)$, since $\|D\tau_1\|_\infty < \frac{1}{2n}$, let $z = \gamma(x) = x - \tau_1(x)$. Since γ is a diffeomorphism, we can write $x = \gamma^{-1}(z)$. Thus, applying change of variables yields

$$\begin{aligned}\|L_{\tau_1}g\|_2^2 &= \int_{\mathbb{R}^n} |g(x-\tau_1(x))|^2\,dx \\ &= \int_{\mathbb{R}^n} |g(z)|^2 \frac{dz}{\det(I - D\tau_1(\gamma^{-1}(z)))} \\ &\le \int_{\mathbb{R}^n} |g(z)|^2 \frac{dz}{1 - n\|D\tau_1\|_\infty} \\ &\le 2\int_{\mathbb{R}^n} |g(z)|^2\,dz \\ &= 2\|g\|_2^2.\end{aligned}$$

That is, we have $\|L_{\tau_1}g\|_2 \le \sqrt{2}\|g\|_2$ and the spectral norm of L_{τ_1} satisfies $\|L_{\tau_1}\| \le \sqrt{2}$.

Now we can apply the bound on $\|L_\tau\|$ above to get

$$\begin{aligned}\|\tau_3 L_{\tau_1}(\mathcal{F}^{-1}\tau_2 * f)\|_2 &\le \|\tau_3\|_\infty\|L_{\tau_1}(\mathcal{F}^{-1}\tau_2 * f)\|_2 \\ &\le \sqrt{2}\|\tau_3\|_\infty\|\mathcal{F}^{-1}\tau_2 * f\|_2.\end{aligned}$$

Since τ_2 is continuous and absolutely integrable, we know that $\mathcal{F}^{-1}\tau_2 = R\mathcal{F}\tau_2$. Now, via Young's convolution inequality, we see that

$$\|\mathcal{F}^{-1}\tau_2 * f\|_2 \le \|\mathcal{F}^{-1}\tau_2\|_1\|f\|_2 = \|R\hat{\tau}_2\|_1\|f\|_2 = \|\hat{\tau}_2\|_1\|f\|_2.$$

We know that $\hat{\tau}_2 \in \mathbf{L}^1(\mathbb{R}^n)$ since τ_2 is continuous and bounded, which implies $\mathcal{F}^{-1}\tau_2 * f \in \mathbf{L}^2(\mathbb{R}^n)$. Additionally, we also have $\mathcal{F}\mathcal{F}^{-1}\tau_2(x) = \tau_2(x)$ for all $x \in \mathbb{R}^n$. Thus, we can apply Plancherel Theorem and Convolution Theorem to get

$$\begin{aligned}\|\mathcal{F}^{-1}\tau_2 * f\|_2^2 &= \|\mathcal{F}(\mathcal{F}^{-1}\tau_2 * f)\|_2^2 \\ &= \int_{\mathbb{R}^n} |\tau_2(\xi)|^2|\hat{f}(\xi)|^2\,d\xi \\ &\le \|\tau_2\|_\infty^2\|f\|_2^2.\end{aligned}$$

Finally, we have derived the desired bound. □

For the remainder of this paper, we will assume that the deformations τ_2 and τ_3 satisfy the conditions of Lemma 1.

IV. Time Frequency Deformation Stability for a CNN-like Feature Extractor

The deformation sensitivity result in [12] can be summarized in the following theorem:

Theorem 2 (Theorem 2, [12])**.** *Define the deformation operator*

$$F_{\tau,\omega}(x) := M_\omega L_\tau f(x) := e^{2\pi i\omega(x)} f(x - \tau(x))$$

with $M_\omega f(x) := e^{2\pi i\omega(x)}f(x)$. Suppose that $\Omega = ((\Psi_k, M_k, P_k))_{k\in\mathbb{N}}$ is an admissible module sequence, $S_k \ge 1$ for $k \in \mathbb{N}$, and $\hat{f}$ is supported in $B(0,R)$ for some $R > 0$. Then

$$|||\Phi_\Omega(F_{\tau,\omega}f) - \Phi_\Omega(f)||| \le C_1(R\|\tau\|_\infty + \|\omega\|_\infty)\|f\|_2$$

for some constant $C_1 > 0$.

Regarding the operator $F_{\tau,\omega}(x)$, we can incorporate the term $e^{2\pi i\omega(x)}$ into our framework by noticing that

$$\begin{aligned}& e^{2\pi i\omega(x)}L_{\tau_1,b}f(x) \\ &= e^{2\pi i\omega(x)}\tau_3(x)Q_{\tau_1,\tau_2,\tau_3}f(x) \\ &= e^{2\pi i\omega(x)}\tau_3(x)\int_{\mathbb{R}^n} e^{i\langle\xi,x-\tau_1(x)\rangle}\tau_2(\xi)\hat{f}(\xi)\,d\xi.\end{aligned}$$

We now state and prove one of the main theorems of this paper:

Theorem 3. *Suppose that $\Omega = ((\Psi_k, M_k, P_k))_{k\in\mathbb{N}}$ is an admissible module sequence, $S_k \ge 1$ for $k \in \mathbb{N}$, and $\hat{f}$ is supported in $B(0,R)$ for some $R > 0$. Then*

$$|||\Phi_\Omega(M_\omega K_{\tau_1,\tau_2,\tau_3}f) - \Phi_\Omega(f)||| \le S(\tau_1,\tau_2,\tau_3,\omega)\|f\|_2$$

with

$$S(\tau_1,\tau_2,\tau_3,\omega) = C_1(R\|\tau\|_\infty + \|\omega\|_\infty) + \sqrt{2}\|\tau_3\|_\infty\|\tau_2\|_\infty.$$

Proof. We first use the fact our feature extractor is nonexpansive to get

$$\begin{aligned}& |||\Phi_\Omega(M_\omega K_{\tau_1,\tau_2,\tau_3}f) - \Phi_\Omega(f)||| \\ &\le \|M_\omega K_{\tau_1,\tau_2,\tau_3}f - f\|_2 \\ &= \|M_\omega L_{\tau_1}f + M_\omega Q_{\tau_1,\tau_2,\tau_3}f - f\|_2 \\ &\le \|M_\omega L_{\tau_1}f - f\|_2 + \|M_\omega Q_{\tau_1,\tau_2,\tau_3}f\|_2.\end{aligned}$$

For the first term above, Theorem 2 yields

$$\|M_\omega L_{\tau_1}f - f\|_2 \le C_1(R\|\tau\|_\infty + \|\omega\|_\infty)\|f\|_2.$$

For the second term, we see that

$$\begin{aligned}\|M_\omega Q_{\tau_1,\tau_2,\tau_3}f\|_2 &\le \|Q_{\tau_1,\tau_2,\tau_3}f\|_2 \\ &\le \sqrt{2}\|\tau_3\|_\infty\|\tau_2\|_\infty\|f\|_2.\end{aligned}$$

Thus, we have

$$|||\Phi_\Omega(M_\omega K_{\tau_1,\tau_2,\tau_3}f) - \Phi_\Omega(f)||| \le S(\tau_1,\tau_2,\tau_3,\omega)\|f\|_2,$$

which completes the proof. □

The above result is in the same spirit as the results in [2], [12]. We see that as $\|\tau_1\|_\infty + \|\tau_2\|_\infty + \|\tau_3\|_\infty + \|\omega\|_\infty \to 0$, we have $S(\tau_1, \tau_2, \tau_3, \omega) \to 0$, which means that a smaller deformation in our signal results in a smaller deformation of our representation with Φ_Ω. Additionally, akin to [12], our bound depends on the architecture chosen rather than the specific filter choice.

Example 1. *A set of deformations where Lemma 1 is satisfied consist of the frequency deformation $\tau_2(\xi) = c_2 e^{-|\xi|^2}$ and time deformation $\tau_3(x) = c_3 e^{-|x|^2}$, where c_2 and c_3 are deformation amplitude parameters with $|c_2| < 1$ and $|c_3| < 1$.*

V. Time Frequency Deformation Stability for the Windowed Scattering Transform

As an additional application, we consider the specific case of Mallat's Windowed Scattering Transform [2]. We provide a brief overview below and refer the reader to [2] for a comprehensive overview.

Let G be a set of rotations in $\mathrm{SO}(n)$ with determinant 1 and G^+ be $G \setminus \{I, -I\}$, where I is the identity element of G. Suppose that we have a wavelet, a function $\psi \in \mathbf{L}^1(\mathbb{R}^n) \cap \mathbf{L}^2(\mathbb{R}^n)$ with $\hat{\psi}(0) = 0$; dilations and rotations of the wavelet are defined as $\psi_\lambda(x) := 2^{nj}\psi(2^j r^{-1} x)$ with $j \in \mathbb{Z}$ and $r \in G^+$. In addition to our wavelet, fix $J \in \mathbb{Z}$ and define a low pass filter $\phi_J(x) = 2^{-nJ}\phi(2^{-J}x)$. We say that our wavelet frame is a satisfies the Littlewood Paley condition if

$$\beta = \sum_{j \in \mathbb{Z}} |\hat{\psi}(2^{-j}|\omega|)|^2 = 1$$

and

$$|\hat{\phi}(\omega)|^2 = \beta \sum_{j \in \mathbb{Z}} |\hat{\psi}(2^{-j}|\omega|)|^2 = 1,$$

where $\beta = 1/2$ if ψ is real and $\beta = 1$ if ψ is complex.

Lastly, assume that the wavelet ψ satisfies the following admissibility condition if there exists $\eta \in \mathbb{R}^n$ and $\rho \geq 0$ with $|\hat{\rho}(\omega)| \leq |\hat{\phi}(2\omega)|$ and $\hat{\rho}(0) = 1$ such that the function

$$\hat{\Gamma}(\omega) = |\hat{\rho}(\omega - \eta)|^2 - \sum_{k=1}^{\infty} k \left(1 - |\hat{\rho}(2^{-k}(\omega - \eta))|^2\right) \quad (17)$$

satisfies

$$\alpha = \inf_{1 \leq |\omega| \leq 2} \sum_{j \in \mathbb{Z}} \sum_{r \in G} \hat{\Gamma}(2^{-j} r^{-1} \omega) |\hat{\psi}(2^{-j} r^{-1} \omega)|^2 > 0. \quad (18)$$

Using the same path notation as previously, define $\mathcal{L}_J = \{\lambda = 2^j r : r \in G^+, j > -J\}$ and take a path q consisting of $q = (\lambda_1, \ldots, \lambda_k) \in \mathcal{L}_J^k$, where $\mathcal{L}_J^k = \Pi_{i=1}^k \mathcal{L}_J$. We define

$$U[q]f = |||f * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \cdots | * \psi_{\lambda_k}| \quad (19)$$

and

$$S_J[q]f = U[q]f * \phi_J. \quad (20)$$

Denote the collection of all finite paths $q = (\lambda_1, \ldots, \lambda_k)$ with each element of the path in $\mathcal{L}_J$ as $\mathcal{P}_J$. We will consider the following norm:

$$\|S[\mathcal{P}_J]f\|^2 = \sum_{q \in \mathcal{P}_J} \|S[q]f\|_2^2.$$

Now, we provide two properties that will be necessary for the stability analysis of the Windowed Scattering Transform with respect to the time-frequency deformations we have defined.

Theorem 4 (Proposition 2.5, [2])**.** *Suppose that ψ is a wavelet satisfying the Littlewood Paley condition given above. Then for all $f, g \in \mathbf{L}^2(\mathbb{R}^n)$, we have*

$$\|S_J[P_J]f - S_J[P_J]g\| \leq \|f - g\|_2.$$

Theorem 5 (Theorem 2.16, [2])**.** *Suppose ψ is an admissible wavelet. Then*

$$\|U[\mathcal{P}_J]f\|_1 := \sum_{m=0}^{\infty} \|U[\mathcal{L}_J^m]f\| < \infty.$$

Additionally, suppose that $\|D\tau\|_\infty < \frac{1}{2n}$. Then there exists $C_2 > 0$ such that for all $f \in \mathbf{L}^2(\mathbb{R}^n)$

$$\begin{aligned} &\|S_J[P_J]K_{\tau_1,\tau_2,\tau_3} f - S_J[P_J]f\| \\ &\leq C_2 \|U[\mathcal{P}_J]\|_1 (\|D\tau\|_\infty + \|D^2\tau\|_\infty). \end{aligned}$$

Theorem 6. *Suppose that ψ is an admissible wavelet and $\|D\tau_1\|_\infty < \frac{1}{2n}$. Then we have*

$$\begin{aligned} &\|S_J[P_J]K_{\tau_1,\tau_2,\tau_3} f - S_J[P_J]f\| \\ &\leq C_2 \|U[P_J]f\|_1 (\|D\tau_1\|_\infty + \|D^2\tau_1\|_\infty) \\ &+ \sqrt{2}\|\tau_3\|_\infty \|\tau_2\|_\infty \|f\|_2. \end{aligned} \quad (21)$$

Proof. The proof idea is similar to before, and it suffices to bound

$$\underbrace{\|S_J[P_J]K_{\tau_1,\tau_2,\tau_3} f - S_J[P_J]L_{\tau_1} f\|}_{I_1} + \underbrace{\|S_J[P_J]L_{\tau_1} f - S_J[P_J]f\|}_{I_2}.$$

We start by bounding I_1. Since the Windowed Scattering Transform is nonexpansive, we have

$$\begin{aligned} &\|S_J[P_J]K_{\tau_1,\tau_2,\tau_3} f - S_J[P_J]L_{\tau_1} f\| \\ &\leq \|L_{\tau_1} f + Q_{\tau_1,\tau_2,\tau_3} f - L_{\tau_1} f\|_2 \\ &= \|Q_{\tau_1,\tau_2,\tau_3} f\|_2. \end{aligned}$$

Now we apply Lemma 1 to get

$$\|S_J[P_J]K_{\tau_1,\tau_2,\tau_3} f - S_J[P_J]L_{\tau_1} f\| \leq \sqrt{2}\|\tau_3\|_\infty \|\tau_2\|_\infty \|f\|_2.$$

For I_2, we have

$$\begin{aligned} &\|S_J[P_J]L_{\tau_1} f - S_J[P_J]f\| \\ &\leq C_2 \|U[P_J]f\|_1 (\|D\tau\|_\infty + \|D^2\tau\|_\infty) \end{aligned}$$

for some $C > 0$. Thus, we have the desired bound. ☐

VI. Conclusion

We have introduced a time-frequency generalization of lipschitz continuity to difeomorphisms for deep convolutional architectures from the lens of time-frequency transforms. Future work would include generalizing these results to convolutional kernel networks seen in [5]. Furthermore, it would be of interest to explore whether one can achieve a stability bound when the amplitude function $b(x, \xi)$ is not separable.

REFERENCES

[1] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[2] S. Mallat, "Group invariant scattering," *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, 2012.

[3] J. Bruna and S. Mallat, "Invariant scattering convolution networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 8, pp. 1872–1886, 2013.

[4] F. Gao, G. Wolf, and M. Hirn, "Geometric scattering for graph data analysis," in *International Conference on Machine Learning*, pp. 2122–2131, PMLR, 2019.

[5] A. Bietti and J. Mairal, "Group invariance, stability to deformations, and complexity of deep convolutional representations," *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 876–924, 2019.

[6] M. Eickenberg, G. Exarchakis, M. Hirn, S. Mallat, and L. Thiry, "Solid harmonic wavelet scattering for predictions of molecule properties," *The Journal of chemical physics*, vol. 148, no. 24, p. 241732, 2018.

[7] P. Sinz, M. W. Swift, X. Brumwell, J. Liu, K. J. Kim, Y. Qi, and M. Hirn, "Wavelet scattering networks for atomistic systems with extrapolation of material properties," *The Journal of Chemical Physics*, vol. 153, no. 8, p. 084109, 2020.

[8] A. Chua, M. Hirn, and A. Little, "On generalizations of the non-windowed scattering transform," *Applied and Computational Harmonic Analysis*, vol. 68, p. 101597, 2024.

[9] F. Gama, A. Ribeiro, and J. Bruna, "Stability of graph scattering transforms," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[10] M. Perlmutter, F. Gao, G. Wolf, and M. Hirn, "Geometric wavelet scattering networks on compact Riemannian manifolds," in *Mathematical and Scientific Machine Learning*, pp. 570–604, PMLR, 2020.

[11] J. Chew, M. Hirn, S. Krishnaswamy, D. Needell, M. Perlmutter, H. Steach, S. Viswanath, and H.-T. Wu, "Geometric scattering on measure spaces," *Applied and Computational Harmonic Analysis*, vol. 70, p. 101635, 2024.

[12] T. Wiatowski and H. Bölcskei, "A mathematical theory of deep convolutional neural networks for feature extraction," *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp. 1845–1866, 2017.

[13] F. Nicola and S. I. Trapasso, "Stability of the scattering transform for deformations with minimal regularity," *Journal de Mathématiques Pures et Appliquées*, vol. 180, pp. 122–150, 2023.

[14] P. Grohs, T. Wiatowski, and H. Bölcskei, "Deep convolutional neural networks on cartoon functions," in *2016 IEEE International Symposium on Information Theory (ISIT)*, pp. 1163–1167, 2016.

[15] M. Koller, J. Großmann, U. Monich, and H. Boche, "Deformation stability of deep convolutional neural networks on Sobolev spaces," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6872–6876, IEEE, 2018.

[16] L. Hörmander, *Fourier Integral Operators. I*, pp. 23–127. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994.