ON THE WAVELET SCATTERING TRANSFORM AND ITS GENERALIZATIONS

By

Albert Chua

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Mathematics—Doctor of Philosophy

2023

ABSTRACT

In this thesis, we look into generalizations of Mallat's wavelet scattering transform. In the second chapter, we generalize finite depth wavelet scattering transforms, which we formulate as $\mathbf{L}^q(\mathbb{R}^n)$ norms of a cascade of continuous wavelet transforms (or dyadic wavelet transforms) and contractive nonlinearities. We then provide norms for these operators, prove that these operators are well-defined, and are Lipschitz continuous to the action of C^2 diffeomorphisms in specific cases; additionally, we extend our results to formulate an operator invariant to the action of rotations $R \in \mathrm{SO}(n)$ and an operator that is equivariant to the action of rotations of $R \in \mathrm{SO}(n)$. In the third and fourth chapters, we generalize our results to stochastic process and signals on compact manifolds, respectively.

ACKNOWLEDGEMENTS

First, and most importantly, I would like to thank my advisor, Yang Yang, for all the support that he has provided me during the one and a half years we have worked with each other. In spite of the fact that I always try "interesting" ideas for my research, he has never been dismissive, and has been very supportive by providing help whenever he can. I'll always appreciate the support he has provided me.

I also would like to thank my committee members, Jianliang Qian, Ekaterina Rapinchuk, and Rongrong Wang for all the support they have provided me thoughout my time at MSU. Without their thoughtful suggestions for this thesis, it would be much worse condition compared to now.

I also appreciate the help Michael Perlmutter, who was one of the first people to provide edits to what became the journal version of Chapter 2 of this thesis. I definitely would not have caught most of those errors myself. I also appreciate his advice for how/where to publish the paper, because I wasn't really sure of what to do with it at the time.

I would like to also Anna Little, who was one of the coauthors on the journal version of Chapter 2 in this thesis. She provided a lot of help with foundational parts of the work, which made my life a lot easier. I also appreciate the time she took to help me through the review process for the paper, which I was completely unprepared for on my own.

Matt Hirn, who was my advisor during the third and fourth years of my PhD, is also someone I would like to thank. He was the one who introduced me to the scattering transform, and helped me while I started my research on nonwindowed scattering transforms. Despite all the trouble I had at first, he was very patient with me and helped me with any trouble I had. I appreciate all the proofs he checked, especially all the "proofs" with mistakes that he caught.

Lastly, I would like to thank anyone who I have missed on this list. I don't think I could have finished this thesis without the help of many people in my life.

TABLE OF CONTENTS

CHAPTER 1 INTRODUCTION . . . 1
- 1.1 Notation . . . 1
- 1.2 Machine Learning and Model Fitting . . . 1
- 1.3 Background On Convolutional Neural Networks . . . 2
- 1.4 Invariance, Equivariance, Stability, Frequency Representations, and Machine Learning . . . 3
- 1.5 Wavelets . . . 5
- 1.6 Scattering Transforms . . . 7
- 1.7 Contributions . . . 11

CHAPTER 2 GENERALIZING THE NONWINDOWED SCATTERING TRANSFORM . . . 16
- 2.1 Fourier Transforms and Hardy Spaces . . . 16
- 2.2 Wavelet Scattering is a Bounded Operator . . . 17
- 2.3 Stability to Dilations . . . 28
- 2.4 Stability to Diffeomorphisms . . . 39
- 2.5 Equivariance and Invariance to Rotations . . . 54

CHAPTER 3 EXPECTED SCATTERING TRANSFORMS . . . 60
- 3.1 Background . . . 60
- 3.2 Wavelet Transforms for Stochastic Processes . . . 60
- 3.3 Scattering Moments and the Expected Scattering Transform . . . 61
- 3.4 The Expected Scattering Transform When $q = 2$ . . . 62
- 3.5 The Expected Scattering Transform When $1 < q < 2$ . . . 66

CHAPTER 4 NONWINDOWED SCATTERING ON COMPACT RIEMANNIAN MANIFOLDS . . . 69
- 4.1 Notation for Scattering on Manifolds . . . 69
- 4.2 Spectral Filters and the Geometric Wavelet Transform . . . 70
- 4.3 The Geometric Scattering Transform . . . 72
- 4.4 Generalizing Geometric Scattering Transforms . . . 74

CHAPTER 5 CONCLUSIONS . . . 78

BIBLIOGRAPHY . . . 79

CHAPTER 1

INTRODUCTION

1.1 Notation

Set $\mathbb{R}_+$ to be the positive real numbers, i.e. $\mathbb{R}_+ := (0, \infty)$. The gradient of a function $f : \mathbb{R}^n \to \mathbb{C}$ is given by ∇f, the Jacobian of a function $f : \mathbb{R}^n \to \mathbb{R}^m$ is given by Df, and the Hessian is given by $D^2 f$. For $1 \leq q < \infty$, the $\mathbf{L}^q(\mathbb{R}^n)$ norm of a function $f : \mathbb{R}^n \to \mathbb{C}$ is $\|f\|_q := \left[\int_{\mathbb{R}^n} |f(x)|^q \, dx\right]^{1/q}$. When $q = \infty$, $\|f\|_\infty := \text{ess sup}|f|$. We will also use the notation, $\|\Delta f\|_\infty = \sup_{x,y \in \mathbb{R}^d} |f(x) - f(y)|$, for the first two chapters of this thesis (which should not be mistaken for applying a Laplacian operator). Greek letters with a vector symbol, such as $\vec{\alpha} = (\alpha_1, \cdots, \alpha_n)$, will be a multi-index of nonnegative integers; additionally, we write $|\vec{\alpha}| = \alpha_1 + \cdots + \alpha_n$, and the usage will be clear from context. The operator $D^{\vec{\alpha}}$ is a multi-index of derivatives: $D^{\vec{\alpha}} f = \frac{\partial^{|\vec{\alpha}|}}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}} f$. For integer $s \geq 0$, we define the function space $\mathbf{H}^s(\mathbb{R}^n) = \{f \in \mathbf{L}^2(\mathbb{R}^n) : D^{\vec{\alpha}} f \in \mathbf{L}^2(\mathbb{R}^n) \text{ for } |\vec{\alpha}| \leq s\}$.

1.2 Machine Learning and Model Fitting

The following material is based on [1]. A functional perspective on supervised learning is the following. Suppose we have a set of data that is split into a training set $T = \{(x_i, y_i)\}_{i=1}^N$, which has known data $\{x_i\} \subset \mathcal{X}$ and labels $\{y_i\}_{i=1}^N$. Our goal is to find a model F_θ, parameterized by a set of weights $\theta \in \mathbb{R}^n$, that best fits the data with respect to some metric (i.e. mean squared loss). To check if our model F_θ actually fits the data, we are given a set of test points T_{test}, which are only accessible for evaluating the fit of F_θ.

Define the set of all possible models $\mathcal{F}$ as

$$\mathcal{F} = \{f_\theta(x) \ : \ \theta \in \mathbb{R}^n\},$$

where each f_θ is a model parameterized by weights $\theta \in \mathbb{R}^n$. One needs to narrow down the search space by choosing an appropriate model to fit the data. One such instance is when one has prior knowledge of the distribution of data. For example, consider linear regression; suppose that $\{x_i\}_{i=1}^N \subset \mathbb{R}^n$ and $y_i \subset \mathbb{R}$ with $y_i = w^T x_i + \varepsilon$, where $w \in \mathbb{R}^n$ is a set of unknown weights and ε is a small noise. Lastly, assume that we want to find a representation that minimizes the mean squared error:

$$\sum_{i=1}^{N}(f_\theta(x_i) - y_i)^2.$$

At this point, it is natural to restrict the set of functions to have the following representation:

$$f_\theta(x) = \theta^T x, \qquad \theta \in \mathbb{R}^n.$$

Note that this example is relatively simple. For more complex representations, such as images, one needs to consider more sophisticated representations. Over the past two decades, convolutional neural networks have show remarkable success for image recognition tasks. For example, [2, 3, 4, 5] have gradually redefined state-of-the art on benchmark datasets in the 2010s. However, the mechanisms behind how they work have not been fully understood until recently [6, 7, 8, 9, 10].

1.3 Background On Convolutional Neural Networks

Before we provide more discussion about invariants in machine learning, we will discuss the architecture for convolutional neural networks.

Consider two discrete functions: $a_1 : \mathbb{Z} \to \mathbb{R}$ and $b_1 : \mathbb{Z} \to \mathbb{R}$. Practitioners in deep learning generally define the convolution (which is cross-correlation) as

$$(a_1 * b_1)(i) = \sum_{j\in\mathbb{Z}} a_1(i + j)b_1(j). \tag{1.1}$$

More generally, we can assume that we have two dimensional functions $a_2 : \mathbb{Z}^2 \to \mathbb{R}$ and $b_2 : \mathbb{Z}^2 \to \mathbb{R}$. The two dimensional convolution is given by

$$(a_2 * b_2)(i_1, i_2) = \sum_{(j_1,j_2)\in\mathbb{Z}^2} a_2(i_1 + j_1, i_2 + j_2)b_2(j_1, j_2). \tag{1.2}$$

This is the first building block for convolution operations similar to the operations seen in deep learning libraries, such as "Conv2d" in PyTorch. However, in practice, these operations generally are implemented with finite filters rather than infinite filters like above.

To construct a full Conv2d layer, suppose that we have a set of N_1 functions, and the goal is to get a representation with N_2 functions via a set of convolutions. Define a set of functions $\{F_{n_1,n_2}\}$

with indexes $1 \leq n_1 \leq N_1$ and $1 \leq n_2 \leq N_2$. The Conv2d layer can be mathematically expressed as

$$C(f) = \sum_{n_1=1}^{N_1} F_{n_1,n_2} * f. \tag{1.3}$$

After applying C, a nonlinearity is applied to each entry of the result, and some form of subsampling is done to reduce the data necessary for the representation. A convolutional neural network, less formally speaking, is a cascade of applying a Conv2d layer, a nonlinearity, and a subsampling operator, in that exact order.

1.4 Invariance, Equivariance, Stability, Frequency Representations, and Machine Learning

Let $\mathcal{B}_1$,$\mathcal{B}_2$ be Banach Spaces and $\Phi : \mathcal{B}_1 \to \mathcal{B}_2$ be an operator, let $T : \mathcal{B}_1 \to \mathcal{B}_1$ be an operator. We say that Φ is invariant to T if

$$\Phi T f = \Phi f, \qquad \forall f \in \mathcal{B}_1,$$

and Φ is a T-invariant operator. Similarly, for $T : \mathcal{B}_1 \to \mathcal{B}_2$, for we say that Φ is equivariant with respect to T if

$$\Phi T f = T \Phi f, \qquad \forall f \in \mathcal{B}_1.$$

Similar to the regression example, CNNs restrict the the possible set of models we consider. With respect to images, convolution has two properties that are helpful for image recognition tasks:

- Convolution is inherently a local operation and depends on neighboring pixels. That is to say, we utilize the underlying geometry of an image.
- Convolution is equivariant with respect to translation. In other words, translating a function and translating a function after convolution yield the same output.

However, it is not necessarily useful to have translation equivariance. Suppose we have the following two tasks:

- Determine if a cat is in the picture.
- Determine where the cat is in the picture.

For the first task, the location of the cat does not matter, so translating the cat in the picture is irrelevant. Thus, we would like a representation that is invariant to translation. On the other hand, in the second task, to keep track of the location of the cat, we would like a representation that is equivariant with respect to translations. This example illustrates the following point. Using relevant information about our task is a way of restricting down the search space for possible models.

Along with some type of invariance or equivariance, stability is also an important property for our representation. Let $L_\gamma f(x) = f(\gamma^{-1}(x))$, where $\gamma(x) := x - \tau(x)$ for $\tau \in C^2(\mathbb{R}^n)$ suitably small. We would like a representation such that

$$\|\Phi f - \Phi L_\tau f\|_{\mathcal{B}_2} \leq K(\tau)\|f\|_{\mathcal{B}_1},$$

and $K(\tau)$ get smaller as τ get smaller. The intuition is that small deformations of the signal will not change the representation too much.

An important aspect of convolutional models is their ability to discern frequency information. Empirically, high frequency information is important for image recognition. In Figure 1.1, one can

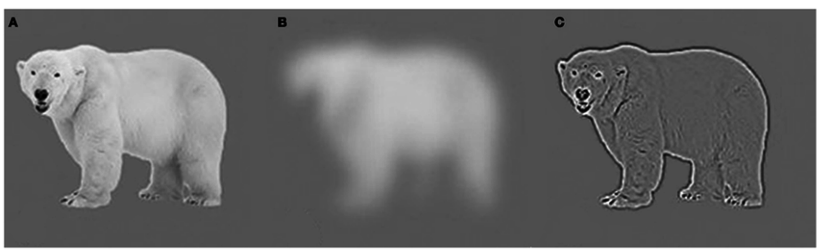

Figure 1.1 **Left**: Polar Bear. **Middle**: Low Pass filtering. **Right**: High Pass filtering.

see that the high frequency information is what allows us to determine that the image is in fact an image of a polar bear, so it is important that a representation can extract high frequency information properties. Notably, convolutions are useful for this task because of the convolution theorem.

Lastly, one also needs sufficient model complexity to retain enough meaningful information, which is a key ingredient of deep convolutional neural networks. For example, notice that using the representation $\|f\|_2^2$ yields a translation invariant operator, but any meaningful information about the function f is lost, including high frequency information.

Since convolutional neural networks learn the best model via optimizing a set of weights, it is hard to study their mathematical properties. Instead, one can consider a proxy by using unlearned filters to simplify the analysis. Ideally, the representation should have the following properties:

- Has some invariance/equivariance properties.
- Stable to small deformations.
- Keeps meaningful information and is sufficiently complex.

Regarding the third point, choosing an operator with sufficient complexity, invariance, and stability is not an easy task. For now, consider a simple dilation operator $L_c f(x) = f((1-c)x)$ for $|c| < \frac{1}{2n}$. A feasible way to extract more information is via a low pass filtering (e.g. define an operator $K_\phi f = f * \phi$, where $\hat{\phi}(\omega) = 1B_R(0)$ for some $R > 0$). One can check that for functions f such that $\hat{f}$ is supported in $B_R(0)$, we have

$$\begin{aligned} \|f - L_c f\|_2^2 &= \|f * \phi - L_c f * \phi\|_2^2 \\ &\leq c^2 \cdot C_R \|f\|_2^2 \end{aligned}$$

for some constant C_R. However, high frequency information is lost because $\hat{f}$ is only supported in some bounded ball.

To keep high frequency information, a feasible translation invariant operator to consider is the fourier modulus. However, this operator is not even stable with respect dilations with respect to the 2-norm. The following informal argument from [11]. Suppose that $f(x) = e^{i\xi x}\theta(x)$, where θ is regular with fast decay. Then one can prove that

$$\||\widehat{L_\tau f}| - |\hat{f}|\|_2 \approx |c||\xi|\|\theta\|_2.$$

Since ξ is arbitrary, we see that we can choose it so that the Fourier modulus is not stable to dilations. The main point of these examples is to show that Fourier invariants, which are a natural choice for a feature extractor, are simply not enough. Even for the most simple class of dilations, we do not have any stability result that can contain high frequency information. To create an operator with the properties mentioned above, we consider using wavelets.

1.5 Wavelets

We let $\psi \in \mathbf{L}^1(\mathbb{R}^n) \cap \mathbf{L}^2(\mathbb{R}^n)$ be a wavelet, which means it is a function that is localized in both space and frequency and has zero average, i.e.,

$$\int_{\mathbb{R}^n} \psi(x)\, du = 0\,.$$

Assume $f \in \mathbf{L}^2(\mathbb{R}^n)$. The continuous wavelet transform $\mathcal{W}f \in \mathbf{L}^2(\mathbb{R}^n \times \mathbb{R}_+)$ is defined as:

$$\forall\, (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}_+\,, \quad \mathcal{W}f(x, \lambda) := f * \psi_\lambda(x)\,.$$

Furthermore, if ψ satisfies the following admissibility condition

$$\int_0^\infty \frac{|\widehat{\psi}(\lambda\omega)|^2}{\lambda} d\lambda = C_\psi\,, \quad \forall\, \omega \in \mathbb{R}^n \setminus \{0\}\,, \tag{1.4}$$

for some $C_\psi > 0$, then we will say that ψ is a Littlewood-Paley wavelet for the continuous wavelet transform. If ψ satisfies (1.4), one can show that the norm $\mathcal{W}f$ computed with a weighted measure $(dx, d\lambda/\lambda^{n+1})$ on $\mathbb{R}^n \times \mathbb{R}_+$ is well defined:

$$\begin{aligned} \|\mathcal{W}f\|^2_{\mathbf{L}^2(\mathbb{R}^n \times \mathbb{R}_+)} &:= \int_0^\infty \int_{\mathbb{R}^n} |\mathcal{W}f(x, \lambda)|^2\, dx\, \frac{d\lambda}{\lambda^{n+1}} \\ &= \int_0^\infty \int_{\mathbb{R}^n} |f * \psi_\lambda(x)|^2\, dx\, \frac{d\lambda}{\lambda^{n+1}} \\ &= \int_0^\infty \|f * \psi_\lambda\|_2^2\, \frac{d\lambda}{\lambda^{n+1}}\,. \end{aligned}$$

We note, in fact, that one can show:

$$\|\mathcal{W}f\|^2_{\mathbf{L}^2(\mathbb{R}^n \times \mathbb{R}_+)} = \beta \cdot C_\psi \|f\|_2^2\,.$$

where

$$\beta = \begin{cases} 1/2 & \text{if } \psi \text{ is real valued} \\ 1 & \text{if } \psi \text{ is complex valued} \end{cases}\,. \tag{1.5}$$

For a function $f \in \mathbf{L}^2(\mathbb{R}^n)$ we define the dyadic wavelet transform $Wf \in \ell^2(\mathbf{L}^2(\mathbb{R}^n))$ as

$$Wf = \left(f * \psi_j\right)_{j \in \mathbb{Z}}\,.$$

If ψ satisfies

$$\sum_{j \in \mathbb{Z}} |\widehat{\psi}(2^j\omega)|^2 = \hat{C}_\psi, \quad \forall \omega \in \mathbb{R}^n \setminus \{0\}\,, \tag{1.6}$$

for some $\hat{C}_\psi > 0$, then we will say that ψ is a Littlewood-Paley wavelet for the dyadic wavelet transform. If ψ satisfies (1.6), one can show that the norm Wf given below is well defined:

$$\|Wf\|^2_{\ell^2(\mathbf{L}^2(\mathbb{R}^n))} := \sum_{j \in \mathbb{Z}} \|f * \psi_j\|_2^2\,.$$

In fact, we have the following norm equivalence:

$$\|Wf\|^2_{\ell^2(\mathbf{L}^2(\mathbb{R}^n))} = \beta \cdot \hat{C}_\psi \|f\|_2^2\,,$$

where β is defined in (1.5). Wavelets are an ideal choice because the wavelet transform provides a decomposition of a function into frequency bins.

1.6 Scattering Transforms

We now introduce the windowed scattering transform, which is a simple model for convolutional neural network with desirable mathematical properties. Let $\phi : \mathbb{R}^n \to \mathbb{R}$ be a low pass filter ($\hat{\phi}(0) \neq 0$) , $\psi : \mathbb{R}^n \to \mathbb{C}$ a suitable mother wavelet ($\hat{\psi}(0) = 0$), and let G be a rotation group and $G^+ = G/\{-\mathbf{1}, \mathbf{1}\}$, where $\mathbf{1}$ is the identity element for the group. Define a set of rotations and dilations by

$$\Lambda_J := \{\lambda = 2^j r : r \in G^+, j > -J\} \text{ if } J \neq \infty \tag{1.7}$$

and

$$\Lambda_\infty := \{2^j r : r \in G^+,\ j \in \mathbb{Z}\}. \tag{1.8}$$

Let $\lambda = 2^j r \in \Lambda_J$. We further assume that our wavelet satisfies the following unitary frame condition:

$$|\phi(2^J\omega)|^2 + \sum_{\lambda \in \Lambda_J} |\psi(\lambda^{-1}\omega)|^2 = 1$$

is ψ is a complex wavelet, and

$$|\phi(2^J\omega)|^2 + \frac{1}{2}\sum_{\lambda \in \Lambda_J} \left[|\psi(\lambda^{-1}\omega)|^2 + |\psi(-\lambda^{-1}\omega)|^2\right] = 1$$

if ψ is a real wavelet.

Consider the operator

$$U[\lambda] = \left|\int_{\mathbb{R}^n} f(u) 2^{nj} \psi(2^j r^{-1}(x-u))\, du\right| \tag{1.9}$$

For a tuple of rotations and dilations in Λ_J, define a path of length m as the tuple $p := (\lambda_1, \ldots, \lambda_m)$ and let $\mathcal{P}_J$ be the set of all finite paths. The scattering propagator for $f \in \mathbf{L}^2(\mathbb{R}^n)$ and $p \in \mathcal{P}_J$ is

$$U[p]f := U[\lambda_m] \cdots U[\lambda_1] f, \tag{1.10}$$

which gathers high frequency information via a cascade of wavelet transforms and nonlinearities. The scattering operator is

$$\overline{S}f(p) = \frac{1}{\mu_p}\int_{\mathbb{R}^n} U[p]f(x)\,dx \tag{1.11}$$

with $\mu_p := \int_{\mathbb{R}^n} U[p]\delta(x)\,dx$. Additionally, to aggregate features similar to pooling, the author of [11] define the scattering operator for $f \in \mathbf{L}^2(\mathbb{R}^n)$ and $p \in \mathcal{P}_J$ as

$$S_J[p]f(x) = \int_{\mathbb{R}^n} U[p]f(u)2^{-nJ}\phi(2^{-J}(x-u))\,du. \tag{1.12}$$

Additionally, the windowed scattering transform is the set of functions

$$S_J[\mathcal{P}_J]f = \{S_J[p]f\}_{p\in\mathcal{P}_J}. \tag{1.13}$$

This operator is similar to a convolution neural network because along each path (analogous to each layer of a convolutional neural network) a convolution, a nonlinearity is applied, and feature aggregation occurs via the low pass filter. The scattering norm for any set of paths Ω is

$$\|S_J[\Omega]f\|^2 = \sum_{p\in\Omega}\|S_J[p]f\|_2^2. \tag{1.14}$$

Notably, we see that the windowed scattering transform has a structure similar to a convolutional neural network. Since it is important for a feature extractor to extract high frequency information, we will provide an informal explanation for how the modulus nonlinearity does this.

Suppose $f \in L^2(\mathbb{R}^n)$. Then

$$\widehat{(f*\psi_j)}(0) = \hat{f}(0)\hat{\psi}_j(0) = 0,$$

and assume that ψ is C^∞ without any loss of generality. Assume $f*\psi_j \neq 0$ on a set of positive measure. Then

$$\widehat{|f*\psi_j|}(0) = \int_{\mathbb{R}^n}|f*\psi_j|(x)\,dx > 0.$$

Since $|f*\psi_j|$ is continuous, we can find a neighborhood around the origin where $|\widehat{(f*\psi_j)}(x)|$ is nonzero. In other words, high frequency information is pushed down to lower frequency bins.

Before we discuss the theoretical properties of scattering transforms, we provide empirical justification of scattering architectures for feature extraction. First, the seminal paper [12] provided justification for using the windowed scattering transform for small benchmark datasets. From then on, scattering features have shown competitive results for audio tasks [13, 14, 15] and image tasks [16, 17]. Adding learning, like in [18, 19], have been shown to help improve performance in classification tasks as well.

Moving on to theoretical properties of the windowed scattering transform, the windowed scattering transform has the following properties, which are desirable for a feature extractor. The first property is energy preservation, under strict assumptions on the wavelet.

Theorem 1 ([11]). *A scattering wavelet ψ is said to be admissible 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*

$$\hat{\Psi}(\omega) = |\rho(\omega - \eta)|^2 - \sum_{k=1}^{\infty} k \left(1 - |\hat{\rho}(2^{-k}(\omega - \eta))|^2\right) \tag{1.15}$$

satisfies

$$\alpha = \inf_{1 \leq |\omega| \leq 2} \sum_{j=-\infty}^{\infty} \sum_{r \in G} \hat{\Psi}(2^{-j} r^{-1} \omega) |\hat{\psi}(2^{-j} r^{-1} \omega)|^2 > 0. \tag{1.16}$$

If a wavelet is admissible, then $\|S_J[P_J]\| = \|f\|$.

The problem with the admissibility condition in above is that there are very few classes of wavelets that are admissible. The author of [11] mentions an analytic cubic spline Battle-Lemarié wavelet is admissible in one dimension, but provides no other examples. On a related note, [20] has shown that scattering coefficients have exponential decay for $n = 1$ under relatively mild assumptions, but her proof only applies for $n = 1$, which makes the admissibility condition still necessary for $n \geq 2$. Additionally, to our knowledge, there are no examples in the literature of wavelets that satisfy the admissibility condition when $n > 1$.

The second property is that the windowed scattering transform is nonexpansive.

Theorem 2 ([11]). *Suppose ψ is an admissible wavelet. For all $f, h \in L^2(\mathbb{R}^n)$,*

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

The third property is an "almost translation invariance" property.

Theorem 3 ([11])**.** *Define* $L_c f(u) = f(u - c)$*. For admissible wavelets,*

$$\lim_{J\to\infty} \|S_J[P_J]f - S_J[P_J]L_c f\| = 0.$$

for all $c \in \mathbb{R}^n$ *and for all* $f \in L^2(\mathbb{R}^n)$*.*

The last property is a deformation stability bound.

Theorem 4 ([11], informal)**.** *Let* $\tau \in C^2(\mathbb{R}^n)$ *and* $L_\tau f = f(u - \tau(u))$*. For* $f \in L^2(\mathbb{R}^n)$ *and* $\|D\tau\|_\infty < \frac{1}{2n}$*,*

$$\|S_J[P_J]L_\tau f - S_J[P_J]f\| \leq K(\tau)\|f\|_2$$

with $K(\tau) \to 0$ *as* $\|\tau\|_\infty + \|D\tau\|_\infty + \|D^2\tau\|_\infty \to 0$*.*

Deformation stability bounds have become a major point of importance in mathematical deep learning. Since Mallat's work, other works have tried to find feature extractors with similar mathematical properties. For example, [21, 22] consider a generalization of the scattering transform where one uses a general frame instead of a wavelet frame. Another set of related works are [23, 24], which uses a generalization of gabor frames, called uniform covering frames, as a convolution layer. Convolutional kernel networks, as seen in [25, 6], also have desirable mathematical properties. Additionally, rather than working on Euclidean space, a better intrinsic representation can be found by working on a graph or manifold (e.g. point cloud data); works such as [26, 27, 28, 29, 30] focus on feature extractors on noneuclidean data. We will provide a preliminary generalization of [28] in Chapter 4 of this thesis.

Notably, other than [11, 23, 24] all these feature extractors for Euclidean data only provide stability bounds for bandlimited functions, or the set of functions that satisfy

$$\{\hat{f} \ : \ \hat{f} \text{ has compact support}\}.$$

This assumption is reasonable for actual signals because real-world implementation of signals are implemented on a domain with compact time and frequency support.

The work in [23, 24] makes a slight generalization to $(\epsilon - R)$ bandlimited functions. Let

$$Q_R(x) = \{y \in \mathbb{R}^n \ : \ \|y - x\|_\infty < R\}.$$

A function $f \in \mathbf{L}^2(\mathbb{R}^n)$ is (ϵ, R) bandlimited for some $\varepsilon \in [0, 1)$ and $R > 0$ if

$$\|\hat{f}\|_{\mathbf{L}^2(Q_R(0))} \geq (1-\varepsilon)\|f\|_2.$$

However, their stability result is slightly weaker because there are terms that are independent of the deformation in their bound.

To our knowledge, a result similar to Mallat's stability bound, which does not rely on the function being bandlimited, does not exist for other feeature extractors in the current literature. An interesting line of work appears in [31], where one relaxes the assumption on τ in Theorem 4 from $\tau \in C^2(\mathbb{R}^n)$ to $\tau \in \mathbb{C}^{1+\alpha}(\mathbb{R}^n)$ for $\alpha \in (0, 1)$. Similar results also apply to our stability bound in Chapter 2 as well.

1.7 Contributions

Windowed Scattering Transforms are useful when the representation does not need to be rigid. For example, object detection does not require translation invariance, so a Windowed Scattering Transform would be appropriate since a smaller choice of J would not have coeffiicents that would be nearly translation invariant. For a task like classification that needs rigid translation invariance, windowed scattering coefficients are not necessarily the best option. Since the set of functions $\{\phi_J\}$ forms an approximate identity,

$$\lim_{J\to\infty} S[p]f = \lim_{J\to\infty} 2^{nJ} \int_{\mathbb{R}^n} U[p](f * \phi_J)(x)\, dx = \phi(0)\|U[p]f\|_1.$$

Here, the norm acts as the global pooling layer instead of a local pooling layer with the low pass filter. Mallat considered the set of all nonwindowed scattering coefficients, given by $\overline{S}[P_\infty]f$, which provides a rigid representation. However, he was not able to provide stability results for the norm he considered.

We consider a slightly different problem than Mallat did for the nonwindowed scattering transform. As mentioned before, the nonwindowed scattering transform introduced in [11] was a collection of $\mathbf{L}^1(\mathbb{R}^n)$ norms of various cascades of dyadic wavelet convolutions and modulus nonlinearities applied to a signal. Here, we extend the definition of the scattering transform to the continuous wavelet transform and for $\mathbf{L}^q(\mathbb{R}^n)$ norms with $q \in [1,2]$. For a continuous dilation parameter $\lambda \in \mathbb{R}_+$ we define the dilations of ψ as:

$$\forall\, \lambda \in \mathbb{R}_+\,, \quad \psi_\lambda(x) := \lambda^{-n/2}\psi(\lambda^{-1}x)\,,$$

which preserves the $\mathbf{L}^2(\mathbb{R}^n)$ norm of ψ:

$$\|\psi_\lambda\|_2 = \|\psi\|_2\,, \quad \forall\, \lambda \in \mathbb{R}_+\,.$$

For the continuous wavelet transform, the one layer wavelet scattering transform with $\mathbf{L}^q(\mathbb{R}^n)$ norm is the function $S_{\text{cont},q} : \mathbb{R}_+ \longrightarrow \mathbb{R}$ defined as:

$$\forall\, \lambda \in \mathbb{R}_+\,, \quad S_{\text{cont},q} f(\lambda) := \|f * \psi_\lambda\|_q\,. \tag{1.17}$$

For a dyadic dilation parameter $j \in \mathbb{Z}$ we define dilations of ψ as:

$$\forall\, j \in \mathbb{Z}\,, \quad \psi_j(x) = 2^{-nj}\psi(2^{-j}x)\,,$$

which preserves the $\mathbf{L}^1(\mathbb{R}^n)$ norm of ψ:

$$\|\psi_j\|_1 = \|\psi\|_1\,, \quad \forall\, j \in \mathbb{Z}\,.$$

The one layer wavelet scattering transform for the dyadic wavelet transform is the function $S_{\text{dyad},q} f :$ $\mathbb{Z} \longrightarrow \mathbb{R}$ defined as:

$$\forall\, j \in \mathbb{Z}\,, \quad S_{\text{dyad},q} f(j) := \|f * \psi_j\|_q\,. \tag{1.18}$$

More generally, the m-layer wavelet scattering transforms $S^m_{\text{cont},q} f : \mathbb{R}^m_+ \longrightarrow \mathbb{R}$ and $S^m_{\text{dyad},q} f :$ $\mathbb{Z}^m \longrightarrow \mathbb{R}$ are defined as

$$S^m_{\text{cont},q} f(\lambda_1, \ldots, \lambda_m) := \||||f * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \cdots | * \psi_{\lambda_m}\|_q\,, \tag{1.19}$$

$$S^m_{\text{dyad},q} f(j_1, \ldots, j_m) := \||||f * \psi_{j_1}| * \psi_{j_2}| * \cdots | * \psi_{j_m}\|_q\,. \tag{1.20}$$

This is similar to working with a windowed scattering transform with a finite number of layers. However, our operator is different from the operator S_J in [11] because it does not contain the filter A_J to aggregate low frequency information, so the scale parameter in our formulation is not bounded above or below. Additionally, because the averaging filter is replaced $\mathbf{L}^q(\mathbb{R}^n)$ norms, our representation is fully translation invariant rather than translation invariant as $J \to \infty$.

As for the significance of using $\mathbf{L}^q(\mathbb{R}^n)$ norms to replace the averaging filter, there is one area with direct application: quantum energy regression tasks [32], where a representation that is similar to the rotation invariant representation in Section 6.2 has already been used for quantum energy regression.

Given a configuration of atoms, we would like to estimate the ground state energy of the configuration. Suppose we have a molecule with K atoms with nuclear charges z_k and nuclear positions p_k with $k = 1, \ldots, K$. The state x of a molecule is given by

$$x = \{(p_k, z_k) \in \mathbb{R}^3 \times \mathbb{R} \,:\, k = 1 \ldots, K\}, \tag{1.21}$$

Due to how we have defined our state, we would like our representation to have the following properties:

- **Permutation Invariance**: the energy should not depend on the index of the molecules.
- **Deformation Stability**: small deformations of the molecule should only lead to small changes in energy of the system.
- **Isometry Invariance**: the energy should be invariant to group actions such as translations, rotations, and other general isometries.
- **Multiscale Interactions**: molecules have many interactions terms, and these interaction terms depend on the pairwise distance between atoms (i.e. short range covalent bonds and longer range Van Der Waals interactions).

The rotation invariant version of our scattering transform in Chapter 6 satisfies permutation invariance, deformation stability, and has multiscale interactions based on the proofs we've provided. We do not prove isometry invariance, but the operator is rotation and translation invariant.

Motivated by DFT theory, the paper [32] uses a dictionary of one and two layer scattering norms with $q = 1$ and $q = 2$ to get (at the time) state-of-the-art results for energy regression tasks for planar molecules. In particular, scattering operators with $q = 1$ scaled with the number of atoms in the system and $q = 2$ encoded pairwise interactions. The motivation for using $1 < q < 2$ comes from [33, 34], which based on the Thomas–Fermi–Dirac–von Weizsäcker model [35], also use scattering norms with $q = 4/3, 5/3$. Later papers, like [33, 34], use a similar representation, involving spherical harmonics, for 3D quantum energy regression.

Remark 1. We can replace all the modulus operators with any contraction mapping (or use different contraction mappings in each layer) in the definition above, and all the proofs in the rest of this paper will still work. In particular, the modulus can be replaced with a complex version of the rectified linear unit (ReLU) nonlinearity, $\max(0, \mathrm{Re}(a_i))_{i=1,\dots,n}$ for $a \in \mathbb{C}^n$, which is a popular choice for complex neural networks. Nonetheless, we will use the modulus operator throughout this paper without any loss of generality.

We provide a general roadmap for this chapter. First, we will cover notation, basic properties about wavelets and the wavelet scattering operator, and harmonic analysis that will be necessary for the paper. We then provide norms for an m-layer wavelet scattering transforms and prove that the operators are well defined mappings into specific spaces when $1 \leq q \leq 2$. Next, we explore conditions under which the m-layer scattering transform is stable to dilations, and we generalize our results to diffeomorphisms. Lastly, in the last section of this chapter, we formulate two new translation invariant operators that are stable to diffeomorphisms. The first is rotation equivariant, and the second is rotation invariant. Our contributions include, but are not limited to, the following:

- We formulate an extension of the dyadic wavelet scattering operator for a finite, arbitrary number of layers with parameter $q \in [1, 2]$ by applying $\mathbf{L}^q(\mathbb{R}^n)$ norms instead of $\mathbf{L}^1(\mathbb{R}^n)$ norms. Additionally, we formulate a wavelet scattering operator with $q \in [1, 2]$ that uses a continuous scale parameter, like the continuous wavelet transform.
- We create a new finite depth scattering norm using dyadic and continuous scales in the case when $q \in [1, 2]$, and prove that the mappings are well defined and provide theoretical justification for a broader class of wavelets that make the scattering transform Lipchitz continuous to the action of C^2 diffeomorphisms. However, the trade-off is that our stability bound depends on the number of layers.

- We provide a condition for norm equivalence in the case of $q = 2$ that is less stringent.
- In the case of $q \in (1, 2]$, we prove that our norm is stable to diffeomorphisms $\tau \in C^2(\mathbb{R}^n)$ provided that $\|\tau\|_\infty < \frac{1}{2n}$ and the wavelet and its first and second partial derivatives have sufficient decay. In the case of $q = 1$, we show stability to dilations.
- We extend our formulation to include invariance or equivariance to the action of rotations $R \in \mathrm{SO}(n)$.

CHAPTER 2

GENERALIZING THE NONWINDOWED SCATTERING TRANSFORM

The contents of this chapter were a joint work with Matthew Hirn and Anna Little. A journal version of this chapter is published in [36]. We start by providing basic prerequisite knowledge that will be necessary for the results in this chapter.

2.1 Fourier Transforms and Hardy Spaces

The Fourier transform of a function $f \in \mathbf{L}^1(\mathbb{R}^n)$ is the function $\widehat{f} \in \mathbf{L}^\infty(\mathbb{R}^n)$ defined as:

$$\forall\, \omega \in \mathbb{R}^n\,, \quad \widehat{f}(\omega) := \int_{\mathbb{R}^n} f(x) e^{-ix\cdot\omega}\, dx\,.$$

The Hilbert transform of a function $f \in \mathbf{L}^1(\mathbb{R})$ is denoted by Hf and is defined as:

$$Hf(x) := \lim_{\epsilon\to 0} \int_{|x-y|>\epsilon} \frac{f(y)}{x-y}\, dy\,.$$

The map H is a convolution operator in which f is convolved against the function $1/x$. We note that

$$H : \mathbf{L}^q(\mathbb{R}) \to \mathbf{L}^q(\mathbb{R})\,, \quad \forall\, 1 < q < \infty\,,$$

however the result is not true for $q = 1$, i.e., if $f \in \mathbf{L}^1(\mathbb{R})$ it is not necessarily true that $Hf \in \mathbf{L}^1(\mathbb{R})$. We thus introduce the Hardy space. We denote the Hardy space as $\mathbf{H}^1(\mathbb{R})$ and it consists of those functions $f \in \mathbf{L}^1(\mathbb{R})$ such that $Hf \in \mathbf{L}^1(\mathbb{R})$ as well. For $f \in \mathbf{H}^1(\mathbb{R})$ the Hardy space norm is $\|f\|_{\mathbf{H}^1(\mathbb{R})}$, which we define as (see Corollary 2.4.7 of [37])

$$\|f\|_{\mathbf{H}^1(\mathbb{R})} := \|f\|_1 + \|Hf\|_1\,. \tag{2.1}$$

One can show that if $f \in \mathbf{H}^1(\mathbb{R})$, then f must necessarily have zero average. An important property of the Hilbert transform and convolution is the following:

$$H(f * g) = Hf * g = f * Hg\,, \quad f \in \mathbf{L}^p(\mathbb{R})\,,\, g \in \mathbf{L}^q(\mathbb{R})\,, \quad 1 < \frac{1}{p} + \frac{1}{q}\,.$$

We have a similar definition for Hardy spaces when $n \geq 2$. For $1 \leq j \leq n$, define the j^{th} Riesz transform as

$$R_j f(x) = \lim_{\varepsilon\to 0} \int_{|x-y|>\varepsilon} \frac{x_j - y_j}{|x-y|^{n+1}} f(y)\, dy\,, \tag{2.2}$$

where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. The Hardy space $f \in \mathbf{H}^1(\mathbb{R}^n)$ consists of functions f such that $f \in \mathbf{L}^1(\mathbb{R}^n)$ and $R_j f \in \mathbf{L}^1(\mathbb{R}^n)$ for $1 \le j \le n$ as well. For $f \in \mathbf{H}^1(\mathbb{R}^n)$ the Hardy space norm is $\|f\|_{\mathbf{H}^1(\mathbb{R}^n)}$, which we define as (see Corollary 2.4.7 of [37])

$$\|f\|_{\mathbf{H}^1(\mathbb{R}^n)} := \|f\|_1 + \sum_{j=1}^{n} \|R_j f\|_1 \,. \tag{2.3}$$

2.1.1 Operator Valued Spaces

Consider a Banach space $\mathcal{B}$. Suppose $f : \mathbb{R}^n \to \mathcal{B}$ and $x \to \|f(x)\|_{\mathcal{B}}$ is measurable in the Lebesgue sense. Define $\mathbf{L}^p_{\mathcal{B}}(\mathbb{R}^n)$ for $1 \le p < \infty$ to be

$$\|f\|^p_{\mathbf{L}^p_{\mathcal{B}}(\mathbb{R}^n)} = \int_{\mathbb{R}^n} \|f(x)\|^p_{\mathcal{B}} \, dx \,.$$

Also, for $1 \le p < \infty$, define

$$\|f\|_{\mathbf{L}^{p,\infty}_{\mathcal{B}}(\mathbb{R}^n)} = \sup_{\delta > 0} \delta \cdot m(\{x \in \mathbb{R}^n : \|f(x)\|_{\mathcal{B}} > \delta\})^{1/p} \,.$$

We also have the following relation:

$$\|f\|_{\mathbf{L}^{p,\infty}_{\mathcal{B}}(\mathbb{R}^n)} \le \|f\|_{\mathbf{L}^p_{\mathcal{B}}(\mathbb{R}^n)} \,.$$

Note that for $f : \mathbb{R}^n \to \mathbb{R}^n$,

$$\|f\|^p_{\mathbf{L}^p_{\mathbb{R}^n}(\mathbb{R}^n)} = \int_{\mathbb{R}^n} \|f(x)\|^p_{\mathbb{R}^n} \, dx = \int_{\mathbb{R}^n} |f(x)|^p \, dx = \|f\|^p_p \,.$$

2.2 Wavelet Scattering is a Bounded Operator

In this chapter we explore for which $q > 0$ and $m \ge 1$ the wavelet scattering transforms $S^m_{\mathrm{cont},q} f$ and $S^m_{\mathrm{dyad},q} f$ are well-defined as functions in some Banach space (i.e., have finite norm), and under what circumstances.

Let ψ be a wavelet. We assume that ψ has the following properties:

$$|\psi(x)| \le A(1 + |x|)^{-n-\varepsilon} \tag{2.4}$$

$$\int_{\mathbb{R}^n} |\psi(x - y) - \psi(x)| \, dx \le A|y|^{\varepsilon'} \,, \tag{2.5}$$

for some constants $A, \varepsilon', \varepsilon > 0$ and for all $h \neq 0$.

Consider the Littlewood-Paley G-function

$$G_{\psi}(f)(x) = \left(\int_{(0,\infty)} |f * t^{-n}\psi(x/t)|^2 \frac{dt}{t} \right)^{1/2} . \tag{2.6}$$

Let $\mathcal{B} = \mathbf{L}^2 \left((0,\infty), \frac{dt}{t} \right)$. We can rewrite this as a Bochner integral by considering the function $K(x) = (t^{-n/2}\psi_t(x))_{t>0}$. This is a mapping $K : \mathbb{R}^n \to \mathcal{B}$ and the function $x \to \|K(x)\|_{\mathcal{B}}$ is measurable. Also, if we let

$$\mathcal{T}(f)(x) = \left(\int_{\mathbb{R}^n} t^{-n/2}\psi_t(x-y) f(y)\, dy \right)_{t>0} = \left((t^{-n/2}\psi_t * f)(x) \right)_{t>0} ,$$

we observe that

$$G_{\psi}(f)(x) = \|\mathcal{T}(f)(x)\|_{\mathcal{B}}$$

and

$$\|G_{\psi}(f)\|_p^p = \|\mathcal{T}(f)\|^p_{L^p_{\mathcal{B}}(\mathbb{R}^n)} .$$

From Problem 6.1.4 of [38], the two properties above for the wavelet ψ imply that

$$\|K(x)\|_{\mathcal{B}} \leq \frac{c_n A}{|x|^n} , \tag{2.7}$$

and

$$\sup_{y \in \mathbb{R}^n \setminus \{0\}} \int_{|x| \geq 2|y|} \|K(x-y) - K(x)\|_{\mathcal{B}} dx \leq c'_n A , \tag{2.8}$$

where c_n and c'_n depend only on n, ε, and ε'. We will omit the dependence on ε and ε' throughout the rest of this paper, and this will have no effect on any of our proofs.

Remark 2. For the rest of this paper, we will write G in place of G_{ψ} when referring to the G-function because the dependence on the mother wavelet is clear.

Remark 3. Note that (2.5) holds under the alternative condition

$$|\nabla\psi(x)| \leq A(1+|x|)^{-n-1-\epsilon'} . \tag{2.9}$$

This is a consequence of Mean Value Theorem.

We have the following result taken from Problem 6.1.4 of [38] and from Chapter V of [39].

Lemma 5 ([38, 39])**.** *Assume that ψ is defined as above and satisfies (2.7) and (2.8). Then the operator G is bounded from $\mathbf{L}^2(\mathbb{R}^n)$ to $\mathbf{L}^2(\mathbb{R}^n)$. Also, for $p \in (1, \infty)$ and $\mathcal{B} = \mathbf{L}^2(\mathbb{R}_+, dt/t)$, we have*

$$\|\mathcal{T} f\|_{\mathbf{L}^p_{\mathcal{B}}(\mathbb{R}^n)} \leq C_n A \max(p, (p-1)^{-1}) \|f\|_{\mathbf{L}^p(\mathbb{R}^n)}\,, \tag{2.10}$$

for some C_n. For all $f \in \mathbf{L}^1(\mathbb{R}^n)$, we also have

$$\|\mathcal{T} f\|_{\mathbf{L}^{1,\infty}_{\mathcal{B}}(\mathbb{R}^n)} \leq C'_n A \|f\|_{\mathbf{L}^1(\mathbb{R}^n)} \tag{2.11}$$

and

$$\|\mathcal{T} f\|_{\mathbf{L}^1_{\mathcal{B}}(\mathbb{R}^n)} \leq C'_n A \|f\|_{\mathbf{H}^1(\mathbb{R}^n)}\,, \tag{2.12}$$

for some C'_n.

Remark 4. We can also formulate similar bounds for the Littlewood-Paley $\mathfrak{g}$ operator

$$\mathfrak{g}(f)(x) := \left[\sum_{j\in\mathbb{Z}} |\psi_j * f(x)|^2\right]^{1/2} \tag{2.13}$$

using similar arguments.

Remark 5. Let ψ be a wavelet that has properties (2.4) and (2.5). Then with the $\mathbf{L}^2$ normalized dilations, the Littlewood-Paley G-function can be written as:

$$G(f)(x) = \left[\int_0^\infty |f * \psi_\lambda(x)|^2 \, \frac{d\lambda}{\lambda^{n+1}}\right]^{1/2} . \tag{2.14}$$

Note that the λ measure for $G(f)$ matches the measure in defining the norm of $\mathcal{W} f$.

2.2.1 The $\mathbf{L}^2(\mathbb{R}^n)$ Wavelet Scattering Transform

In this section we prove the $\mathbf{L}^2(\mathbb{R}^n)$ scattering transforms are bounded operators. More specifically, we prove that $S^m_{\text{cont},2} : \mathbf{L}^2(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^m_+)$, where $\mathbf{L}^2(\mathbb{R}^m_+)$ has the weighted measure defined by

$$\|S^m_{\text{cont},2} f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)} := \int_0^\infty \cdots \int_0^\infty |S^m_{\text{cont},2} f(\lambda_1, \ldots, \lambda_m)|^2 \, \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}$$

and we show that $\|S^m_{\text{cont},2}f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \le C\|f\|_{\mathbf{L}^2(\mathbb{R}^n)}$. We also show that $S^m_{\text{dyad},2} : \mathbf{L}^2(\mathbb{R}^n) \to \ell^2(\mathbb{Z}^m)$, where

$$\|S^m_{\text{dyad},2}f\|^2_{\ell^2(\mathbb{Z}^m)} := \sum_{j_m\in\mathbb{Z}} \cdots \sum_{j_1\in\mathbb{Z}} |S^m_{\text{dyad},2}f(j_1,\ldots,j_m)|^2.$$

Proposition 6. *For any wavelet satisfying* (2.4) *and* (2.5), *we have* $S^m_{cont,2} : \mathbf{L}^2(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^m_+)$ *and* $S^m_{dyad,2} : \mathbf{L}^2(\mathbb{R}^n) \to \ell^2(\mathbb{Z}^m)$.

Proof. The proof of the dyadic case is essentially identical to the proof given below and is thus omitted. The case of $m = 1$ follows by an application of Fubini's Theorem:

$$\begin{aligned}
\|S_{\text{cont},2}f\|^2_{\mathbf{L}^2(\mathbb{R}_+)} &= \int_0^\infty \|f * \psi_\lambda\|_2^2 \frac{d\lambda}{\lambda^{n+1}} \\
&= \int_0^\infty \int_{\mathbb{R}^n} |(f * \psi_\lambda)(x)|^2\, dx\, \frac{d\lambda}{\lambda^{n+1}} \\
&= \int_{\mathbb{R}^n} |G(f)(x)|^2\, dx \\
&\le C\|f\|_2^2
\end{aligned}$$

by boundedness of the G-function. Now we proceed by using induction. Assume that we have $\|S^m_{\text{cont},2}f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)} \le C_m\|f\|_2^2$. Let $\mathcal{W}_t f = f * \psi_t$, define $Mf = |f|$, and $U_\lambda = MW_\lambda$ for notational brevity. Then notice that

$$\||||f * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \cdots * \psi_{\lambda_m}| * \psi_{\lambda_{m+1}}\|_2^2 = \|\mathcal{W}_{\lambda_{m+1}} U_{\lambda_m} \cdots U_{\lambda_1} f\|_2^2.$$

Substituting yields

$$\begin{aligned}
\|S^{m+1}_{\text{cont},2}f\|_{\mathbf{L}^2(\mathbb{R}^{m+1}_+)} &= \int_0^\infty \cdots \int_0^\infty \|\mathcal{W}_{\lambda_{m+1}} U_{\lambda_m} \cdots U_{\lambda_1} f\|_2^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_{m+1}}{\lambda_{m+1}^{n+1}} \\
&= \int_0^\infty \cdots \int_0^\infty \int_0^\infty \|(U_{\lambda_m} \cdots U_{\lambda_1} f) * \psi_{\lambda_{m+1}}\|_2^2 \frac{d\lambda_{m+1}}{\lambda_{m+1}^{n+1}} \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}} \\
&= \int_0^\infty \cdots \int_0^\infty \|U_{\lambda_m} \cdots U_{\lambda_1} f\|^2_{\mathbf{L}^2(\mathbb{R}_+)} \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}} \\
&\le C \int_0^\infty \cdots \int_0^\infty \|U_{\lambda_m} \cdots U_{\lambda_1} f\|_2^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}} \\
&= C \int_0^\infty \cdots \int_0^\infty |S^m_{\text{cont},2}(\lambda_1,\ldots,\lambda_m)|^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}} \\
&\le C^{m+1}\|f\|_2^2,
\end{aligned}$$

where we used the induction hypothesis in the last line. This completes the proof. □

Proposition 7. *Suppose ψ is a Littlewood-Paley wavelet satisfying* (2.4) *and* (2.5). *Then $S^m_{cont,2}f$: $\mathbf{L}^2(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^m_+)$ and specifically $\|S^m_{cont,2}f\|_1 = C^m_\psi \|f\|_2^2$. Also, $S^m_{dyad,2} : \mathbf{L}^2(\mathbb{R}^n) \to \ell^2(\mathbb{Z}^m)$ and $\|S^m_{dyad,2}f\|_1 = \hat{C}^m_\psi \|f\|_2^2$.*

Proof. We only provide the proof of the continuous case again. First consider the case $m = 1$. We have:

$$\begin{aligned}
\|S_{\mathrm{cont},2}f\|^2_{\mathbf{L}^2(\mathbb{R}_+)} &= \int_0^\infty \|f * \psi_\lambda\|_2^2 \,\frac{d\lambda}{\lambda^{n+1}} \\
&= \frac{1}{(2\pi)^n}\int_0^\infty \|\hat{f}\cdot\hat{\psi}_\lambda\|_2^2 \,\frac{d\lambda}{\lambda^{n+1}} \\
&= \frac{1}{(2\pi)^n}\int_0^\infty \left(\int_{\mathbb{R}^n} |\hat{f}(\omega)|^2 |\hat{\psi}_\lambda(\omega)|^2 \, d\omega\right) \frac{d\lambda}{\lambda^{n+1}} \\
&= \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \left(\int_0^\infty |\hat{\psi}(\lambda\omega)|^2 \frac{d\lambda}{\lambda}\right) |\hat{f}(\omega)|^2 \, d\omega \\
&= \frac{1}{(2\pi)^n}\int_{\mathbb{R}^n} \left(C_\psi |\hat{f}(\omega)|^2\right) \, d\omega \\
&= \frac{1}{(2\pi)^n} C_\psi \|\hat{f}\|_2^2 \\
&= C_\psi \|f\|_2^2.
\end{aligned}$$

Thus the claim holds for $m = 1$. Now assume that it holds through m. Then by the inductive hypothesis,

$$\|S^m_{\mathrm{cont},2}f\|^2_{\mathbf{L}^2(\mathbb{R}_+)} = \int_0^\infty \cdots \int_0^\infty \| | | |f * \psi_{\lambda_1}| * \psi_{\lambda_2}| * \cdots * \psi_{\lambda_m}\|_2^2 \,\frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}} = C^m_\psi \|f\|_2^2.$$

Now consider the case of $m + 1$. Similar to the previous proposition, we have

$$\begin{aligned}
\|S^{m+1}_{\mathrm{cont},2}f\|^2_{\mathbf{L}^2(\mathbb{R}_+)} &= \int_0^\infty \cdots \int_0^\infty \left(\int_0^\infty \|(U_{\lambda_m}\cdots U_{\lambda_1} f) * \psi_{\lambda_{m+1}}\|_2^2 \frac{d\lambda_{m+1}}{\lambda_{m+1}^{n+1}}\right) \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}} \\
&= C_\psi \int_0^\infty \cdots \int_0^\infty |S^m_{\mathrm{cont},2}f(\lambda_1, \ldots, \lambda_m)|^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}} \\
&= C_\psi \|S^m_{\mathrm{cont},2}f\|^2_{\mathbf{L}^2(\mathbb{R}_+)} \\
&= C^{m+1}_\psi \|f\|_2^2.
\end{aligned}$$

Thus, the claim is proven by induction. □

2.2.2 The $\mathbf{L}^1(\mathbb{R}^n)$ Wavelet Scattering Transform

Define the notation $\mathcal{W}_t f = f * \psi_t$, $Mf = |f|$, and $U_t = M\mathcal{W}_t$. We now try to prove that for $m \in \mathbb{N}$, $S^m_{\text{cont},1} : \mathbf{H}^1(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^m_+)$. The norm for $S^m_{\text{cont},1} f$ is:

$$\|S^m_{\text{cont},1} f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} := \left(\int_0^\infty \int_0^\infty \cdots \int_0^\infty |S^m_{\text{cont},1} f(\lambda_1, \lambda_2, \ldots, \lambda_m)|^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \frac{d\lambda_2}{\lambda_2^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2}$$
$$= \left(\int_0^\infty \int_0^\infty \cdots \int_0^\infty \left\|(U_{\lambda_{m-1}} \cdots U_{\lambda_1} f) * \psi_{\lambda_m}\right\|_1^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \frac{d\lambda_2}{\lambda_2^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} .$$

An analogous result will also hold for the operator $\mathbf{H}^1(\mathbb{R}^n) \to \ell^2(\mathbb{Z}^m_+)$ with norm

$$\|S^m_{\text{dyad},1} f\|_{\ell^2(\mathbb{Z}^m)} := \left(\sum_{j_m \in \mathbb{Z}} \cdots \sum_{j_1 \in \mathbb{Z}} |S^m_{\text{dyad},1} f(j_1, \ldots, j_m)|^2\right)^{1/2} .$$

Before we begin, we will need an important multiplier property of the individual Riesz Transforms:

$$\widehat{R_j f}(\omega) = -i \frac{\omega_j}{|\omega|} \hat{f}(\omega) . \tag{2.15}$$

Let $\vec{\alpha} = (\alpha_1, \ldots, \alpha_n)$ be a multi-index with n-elements, and let $t = (t_1, \ldots, t_n) \in \mathbb{R}^n$. We say that ψ has k vanishing moments if for all $|\vec{\alpha}| < k$, we have

$$\int_{\mathbb{R}^n} \left(\Pi_{i=1}^n t_i^{\alpha_i}\right) \psi(t) dt = 0. \tag{2.16}$$

The following lemmas will be necessary.

Lemma 8 ([40])**.** *Suppose that ψ has N vanishing moments, let $M > 1$ be an integer, let $\vec{\alpha}$ be defined as before, and let $\vec{\beta} = (\beta_1, \ldots, \beta_n)$ be a multi-index. Assume that ψ satisfies the following properties:*

- $\psi \in \mathbf{H}^s(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ *for some* $s > M + \frac{n}{2}$.
- *There exists* $A > 0$ *and* $\epsilon \in [0, 1)$ *such that* ψ *satisfies*

$$|D^{\vec{\alpha}} \psi| \leq A(1 + |x|)^{-n-N-|\vec{\alpha}|+\varepsilon} \textit{ for } 0 \leq |\vec{\alpha}| \leq M.$$

- *For* $0 \leq |\vec{\alpha}| \leq M - 1$ *and* $|\vec{\beta}| < N + |\vec{\alpha}|$,

$$\int_{\mathbb{R}^n} \Pi_{i=1}^n t_i^{\beta_i} D^{\vec{\alpha}} \psi(t)\, dt = 0.$$

Then

$$|D^{\vec{\alpha}} R_i \psi(x)| = |R_i D^{\vec{\alpha}} \psi(x)| \le A(1+|x|)^{-n-N-|\vec{\alpha}|+\varepsilon+\delta}$$

for some $0 < \delta < 1 - \varepsilon$ *and* $D^{\vec{\alpha}} R_i \psi$ *has vanishing moments up to degree* $N - 1 + |\vec{\alpha}|$.

An immediate consequence is the following Lemma, which we will provide without proof.

Lemma 9. *Suppose that* ψ *satisfies the following conditions:*

- $\psi \in \mathbf{H}^s(\mathbb{R}^d) \cap C(\mathbb{R}^d)$ *for some* $s > 2 + \frac{n}{2}$.
- *There exists* $A > 0$ *and* $\epsilon \in [0, 1)$ *such that* ψ *satisfies*

$$|D^{\vec{\alpha}} \psi| \le A(1+|x|)^{-n-2-|\vec{\alpha}|+\varepsilon} \text{ for } 0 \le |\vec{\alpha}| \le 3.$$

- *For* $0 \le |\vec{\alpha}| \le 2$ *and* $|\vec{\beta}| < 2 + |\vec{\alpha}|$,

$$\int_{\mathbb{R}^n} \Pi_{i=1}^n t_i^{\beta_i} D^{\vec{\alpha}} \psi(t)\, dt = 0.$$

Then $R_j \psi$ *and all of its first and second partial derivatives have* $O((1+|x|)^{-n-1+\eta})$ *decay for some* $\eta \in (0, 1)$.

The first implication to take note of is that $R_j \psi$ is a wavelet with "good" decay of itself and all its first and second partial derivatives. Note that the strict decay on the partial derivatives is necessary for technical reasons in later proofs, but decay on all second partial derivatives can be relaxed for the following theorem.

Theorem 10. *Let* ψ *be a wavelet satisfying Lemma 9 and let* $S^m_{cont,1}$ *be defined as above. Then for* $f \in \mathbf{H}^1(\mathbb{R}^n)$, *there exists a constant* C_m *such that*

$$\|S^m_{cont,1} f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \le C_m \|f\|_{\mathbf{H}^1(\mathbb{R}^n)}\,.$$

Additionally,

$$\|S^m_{dyad,1} f\|_{\ell^2(\mathbb{Z}^m)} \le C_m \|f\|_{\mathbf{H}^1(\mathbb{R}^n)}.$$

Proof. We proceed by induction and only provide a proof for the continuous case because the dyadic case follows by almost identical reasoning. Let $f \in \mathbf{H}^1(\mathbb{R}^n)$ throughout the proof. By Minkowski's integral inequality ([41], Theorem 202), we have

$$
\begin{aligned}
\|S_{\text{cont},1} f\|_{\mathbf{L}^2(\mathbb{R}_+)} &= \left(\int_0^\infty \|f * \psi_\lambda\|_1^2 \frac{d\lambda}{\lambda^{n+1}}\right)^{1/2} \\
&= \left(\int_0^\infty \left(\int_{\mathbb{R}^n} |f * \psi_\lambda(x)|\, dx\right)^2 \frac{d\lambda}{\lambda^{n+1}}\right)^{1/2} \\
&\le \left(\int_{\mathbb{R}^n} \left(\int_0^\infty |f * \psi_\lambda(x)|^2 \frac{d\lambda}{\lambda^{n+1}}\right)^{1/2} dx\right) \\
&= \int_{\mathbb{R}^n} G(f)(x)\, dx \\
&= \|G(f)\|_1 \\
&\le C\|f\|_{\mathbf{H}^1(\mathbb{R}^n)}\,,
\end{aligned}
$$

where in the last inequality we used Lemma 5.

Now we assume that there exists some $m \ge 1$ such that

$$\|S^m_{\text{cont},1} f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \le C_m \|f\|_{\mathbf{H}^1(\mathbb{R}^n)}.$$

We have

$$
\begin{aligned}
&\|S^{m+1}_{\mathrm{cont},1} f\|_{\mathbf{L}^2(\mathbb{R}^{m+1}_+)} \\
&= \left(\int_0^\infty \cdots \int_0^\infty \left\|(U_{\lambda_m}\cdots U_{\lambda_1} f) * \psi_{\lambda_{m+1}}\right\|_1^2 \; \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_{m+1}}{\lambda_{m+1}^{n+1}}\right)^{1/2} \\
&= \left(\int_0^\infty \cdots \int_0^\infty \left(\int_{\mathbb{R}^n} \left|(U_{\lambda_m}\cdots U_{\lambda_1} f) * \psi_{\lambda_{m+1}}\right| \, dx\right)^2 \; \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_{m+1}}{\lambda_{m+1}^{n+1}}\right)^{1/2} \\
&\le \left(\int_0^\infty \cdots \int_0^\infty \left(\int_{\mathbb{R}^n} \left[\int_0^\infty \left|(U_{\lambda_m}\cdots U_{\lambda_1} f) * \psi_{\lambda_{m+1}}\right|^2 \; \frac{d\lambda_{m+1}}{\lambda_{m+1}^{n+1}}\right]^{1/2} dx\right)^2 \; \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} \\
&= \left(\int_0^\infty \cdots \int_0^\infty \left[\int_{\mathbb{R}^n} G(U_{\lambda_m}\cdots U_{\lambda_1} f)(x)\, dx\right]^2 \; \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} \\
&= \left(\int_0^\infty \cdots \int_0^\infty \|G(U_{\lambda_m}\cdots U_{\lambda_1} f)\|_1^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} \\
&= \left(\int_0^\infty \cdots \int_0^\infty \|G(\mathcal{W}_{\lambda_m} U_{\lambda_{m-1}}\cdots U_{\lambda_1} f)\|_1^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2}
\end{aligned}
$$

since the G function has a modulus already.

It follows that

$$
\|S^m_{\mathrm{cont},1} f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \le C \left(\int_0^\infty \cdots \int_0^\infty \|\mathcal{W}_{\lambda_m} U_{\lambda_{m-1}}\cdots U_{\lambda_1} f\|^2_{\mathbf{H}^1(\mathbb{R}^n)} \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2}.
$$

Now use the definition of the $\mathbf{H}^1(\mathbb{R}^n)$ norm to write

$$
\begin{aligned}
\|\mathcal{W}_{\lambda_m} U_{\lambda_{m-1}}\cdots U_{\lambda_1} f\|_{\mathbf{H}^1(\mathbb{R}^n)} &= \|\mathcal{W}_{\lambda_m} U_{\lambda_{m-1}}\cdots U_{\lambda_1} f\|_{\mathbf{L}^1(\mathbb{R}^n)} \\
&\quad + \sum_{j=1}^n \left\|\left(R_j \mathcal{W}_{\lambda_m}\right)\left(U_{\lambda_{m-1}}\cdots U_{\lambda_1} f\right)\right\|_{\mathbf{L}^1(\mathbb{R}^n)}.
\end{aligned}
$$

Thus, since $R_j \mathcal{W}_{\lambda_m} h = h * \left(R_j \psi_{\lambda_m}\right)$ and $R_j \psi$ wavelet, we can use our induction hypothesis and the previous lemma to get

$$
\begin{aligned}
& C\left(\int_0^\infty \cdots \int_0^\infty \|\mathcal{W}_{\lambda_m}(U_{\lambda_{m-1}}\cdots U_{\lambda_1} f)\|^2_{\mathbf{H}^1(\mathbb{R}^n)} \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} \\
& \leq C\left(\int_0^\infty \cdots \int_0^\infty \|\mathcal{W}_{\lambda_m}(U_{\lambda_{m-1}}\cdots U_{\lambda_1} f)\|^2_{\mathbf{L}^1(\mathbb{R}^n)} \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} \\
& + C\sum_{j=1}^n \left(\int_0^\infty \cdots \int_0^\infty \left\|\left(R_j \mathcal{W}_{\lambda_m}\right)(U_{\lambda_{m-1}}\cdots U_{\lambda_1} f)\right\|^2_{\mathbf{L}^1(\mathbb{R}^n)} \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} \\
& \leq C_{m+1}\|f\|_{\mathbb{H}^1(\mathbb{R}^n)}.
\end{aligned}
$$

Thus, the theorem is proved by induction. □

The case of $n = 1$ is a little trickier. We have the following multiplier property for the Hilbert Transform:

$$
\widehat{Hf}(\omega) = \begin{cases} +i\widehat{f}(\omega) & \omega < 0 \\ -i\widehat{f}(\omega) & \omega > 0 \end{cases} \tag{2.17}
$$

Unfortunately, this yields less regularity for $\widehat{Hf}$ at the origin without additional assumptions. However, notice that the Hilbert transform commutes with dilations, so in particular:

$$
H(\psi_\lambda) = H(\psi)_\lambda \quad \text{and} \quad H(\psi_j) = H(\psi)_j\,.
$$

Using the calculation of $\widehat{Hf}$ in (2.17) we see that

$$
H\psi = -i\psi\,, \quad \text{if } \psi \text{ is complex analytic.}
$$

Thus, we have the following corollary.

Corollary 11. *Let ψ be a complex analytic wavelet such that* (2.4) *and* (2.5) *hold. Then for $f \in \mathbf{H}^1(\mathbb{R})$, there exists a constant C_m such that*

$$
\|S^m_{cont,1} f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \leq C_m \|f\|_{\mathbf{H}^1(\mathbb{R})}\,.
$$

Additionally,

$$
\|S^m_{dyad,1} f\|_{\ell^2(\mathbb{Z}^m)} \leq C_m \|f\|_{\mathbf{H}^1(\mathbb{R})}.
$$

2.2.3 $\mathbf{L}^q(\mathbb{R}^n)$ Wavelet Scattering Transform

In this section, assume $1 < q < 2$. We prove that for $m \in \mathbb{N}$, $S^m_{\text{cont},q} : \mathbf{L}^q(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^m_+)$. The norm for $S^m_{\text{cont},q} f$ is:

$$\begin{aligned}
\|S^m_{\text{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} &:= \left(\int_0^\infty \int_0^\infty \cdots \int_0^\infty |S^m_{\text{cont},q} f(\lambda_1, \lambda_2, \ldots, \lambda_m)|^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \frac{d\lambda_2}{\lambda_2^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{q/2} \\
&= \left(\int_0^\infty \int_0^\infty \cdots \int_0^\infty \left(\left\|(U_{\lambda_{m-1}} \cdots U_{\lambda_1} f) * \psi_{\lambda_m}\right\|_q\right)^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \frac{d\lambda_2}{\lambda_2^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{q/2} .
\end{aligned}$$

There is also an analagous result for

$$\|S^m_{\text{dyad},q} f\|^q_{\ell^2(\mathbb{Z}^m)} := \left(\sum_{j_m \in \mathbb{Z}} \cdots \sum_{j_m \in \mathbb{Z}} |S^m_{\text{dyad},q} f(\lambda_1, \lambda_2, \ldots, \lambda_m)|^2\right)^{q/2} .$$

Theorem 12. *Let $1 < q < 2$. Also, let ψ be a wavelet that satisfies properties* (2.4) *and* (2.5) *and let $S^m_{cont,q}$ and $S^m_{dyad,q}$ be defined as above. Then there exists a universal constant $C_m > 0$ such that $\|S^m_{cont,q} f\|^q_{\mathbf{L}^2(\mathbb{R}_+)} \leq C_m \|f\|^q_q$ for all $f \in \mathbf{L}^q(\mathbb{R}^n)$, and furthermore $\|S^m_{dyad,q} f\|^q_{\ell^2(\mathbb{Z})} \leq C_m \|f\|^q_q$.*

Proof. We proceed by induction and consider the case of $m = 1$ first. Let $f \in \mathbf{L}^q(\mathbb{R}^n)$. For the continuous wavelet transform, we apply Minkowski's integral inequality:

$$\begin{aligned}
\|S_{\text{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}_+)} &= \left[\int_0^\infty \left(\|f * \psi_\lambda\|_q\right)^q \frac{d\lambda}{\lambda^{n+1}}\right]^{1/2} \\
&= \left[\int_0^\infty \left(\int_{\mathbb{R}^n} |f * \psi_\lambda(x)|^q \, dx\right)^{2/q} \frac{d\lambda}{\lambda^{n+1}}\right]^{q/2} \\
&\leq \int_{\mathbb{R}^n} \left(\int_0^\infty |f * \psi_\lambda(x)|^2 \frac{d\lambda}{\lambda^{n+1}}\right)^{q/2} dx \\
&= \|G(f)\|^q_q \\
&\leq C\|f\|^q_q .
\end{aligned}$$

where in the last inequality we used Theorem 5.

Now, let us assume that

$$\|S^m_{\text{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} \leq C^{m \cdot q} \|f\|^q_{\mathbf{L}^q(\mathbb{R}^n)} .$$

We apply Minkowski's Integral inequality [41] to swap and then bound:

$$
\begin{aligned}
&\|S^{m+1}_{\mathrm{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^{m+1}_+)} \\
&= \left[\int_0^\infty \cdots \int_0^\infty \left(\left\|(U_{\lambda_1}\cdots U_{\lambda_1} f) * \psi_{\lambda_{m+1}}\right\|_q\right)^{2/q} \ \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_{m+1}}{\lambda_{m+1}^{n+1}}\right]^{q/2} \\
&= \left[\int_0^\infty \cdots \int_0^\infty \left(\int_{\mathbb{R}^n} |(U_{\lambda_1}\cdots U_{\lambda_1} f) * \psi_{\lambda_{m+1}}(x)|^q \, dx\right)^{2/q} \ \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_{m+1}}{\lambda_{m+1}^{n+1}}\right]^{q/2} \\
&= \left[\int_0^\infty \cdots \int_0^\infty \left[\int_0^\infty \left(\int_{\mathbb{R}^n} |(U_{\lambda_1}\cdots U_{\lambda_1} f) * \psi_{\lambda_{m+1}}(x)|^q \, dx\right)^{2/q} \ \frac{d\lambda_{m+1}}{\lambda_{m+1}^{n+1}}\right]^{\frac{q}{2}\cdot\frac{2}{q}} \ \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right]^{q/2} \\
&\leq \left[\int_0^\infty \cdots \int_0^\infty \left[\int_{\mathbb{R}^n} \left(\int_0^\infty |(U_{\lambda_1}\cdots U_{\lambda_1} f) * \psi_{\lambda_{m+1}}(x)|^2 \, \frac{d\lambda_{m+1}}{\lambda_{m+1}^{n+1}}\right)^{q/2} \ dx\right]^{\frac{2}{q}} \ \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right]^{q/2} \\
&= \left[\int_0^\infty \cdots \int_0^\infty \|G(U_{\lambda_1}\cdots U_{\lambda_1} f)\|_q^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right]^{q/2} \\
&\leq C^q \left[\int_0^\infty \cdots \int_0^\infty \|(U_{\lambda_1}\cdots U_{\lambda_1}) f\|_q^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right]^{q/2} \\
&= C^q \|S^m_{\mathrm{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&\leq C^{(m+1)q} \|f\|_q^q.
\end{aligned}
$$

This proves the desired claim. □

2.3 Stability to Dilations

We now consider dilations defined by $\tau(x) = cx$ for some constant c, so that $L_\tau f(x) = f((1-c)x)$. We will start by proving a lemma that will be useful for our work.

Lemma 13. *Assume L_τ is defined as above. Then*

$$L_\tau f * \psi_\lambda(x) = (1-c)^{-n/2} \left(f * \psi_{(1-c)\lambda}\right) ((1-c)x).$$

Proof. Notice that

$$L_\tau f * \psi_\lambda(x) = \int_{\mathbb{R}^n} f((1-c)y)\psi_\lambda(x-y) \, dy.$$

We make the substitution $z = (1-c)y$. Then it follows that

$$\begin{aligned}
L_\tau f * \psi_\lambda(x) &= (1-c)^{-n} \int_{\mathbb{R}^n} f(z)\psi_\lambda(x - (1-c)^{-1}z)\, dz \\
&= (1-c)^{-n} \int_{\mathbb{R}^n} f(z)\lambda^{-n/2}\psi\left(\lambda^{-1}(x - (1-c)^{-1}z)\right)\, dz \\
&= (1-c)^{-n/2} \int_{\mathbb{R}^n} f(z)[(1-c)\lambda]^{-n/2}\psi\left([(1-c)\lambda]^{-1}\left((1-c)x - z\right)\right)\, dz \\
&= (1-c)^{-n/2} \int_{\mathbb{R}^n} f(z)\psi_{(1-c)\lambda}\left((1-c)x - z\right)\, dz \\
&= (1-c)^{-n/2} f * \psi_{(1-c)\lambda}\left((1-c)x\right) \\
&= (1-c)^{-n/2} L_\tau \left(f * \psi_{(1-c)\lambda}\right)(x).
\end{aligned}$$

□

Remark 6. We also have

$$L_\tau \mathcal{W}_\lambda f(x) = (f * \psi_\lambda)(x(1-c)).$$

Before we begin the next Lemma, we explain the general idea behind our approach to explain the necessity of Lemma 14. Define

$$\Psi(x) = (1-c)^{-n/2}\psi_{(1-c)}(x) - \psi(x). \tag{2.18}$$

We want to prove that Ψ satisfies (2.4) and (2.5) with a linear dependence on c for future stability lemmas.

Lemma 14. *Suppose that ψ is a wavelet that satisfies the following three conditions:*

$$|\psi(x)| \leq \frac{A}{(1+|x|)^{n+1+\alpha}} \quad x \in \mathbb{R}^n, \tag{2.19}$$

$$|\nabla\psi(x)| \leq \frac{A}{(1+|x|)^{n+1+\beta}} \quad x \in \mathbb{R}^n, \tag{2.20}$$

$$\|D^2\psi(x)\|_\infty \leq \frac{A}{(1+|x|)^{n+1+\kappa}} \quad x \in \mathbb{R}^n, \tag{2.21}$$

for $\alpha, \beta, \kappa > 0$. Consider

$$\Psi(x) = (1-c)^{-n/2}\psi_{(1-c)}(x) - \psi(x).$$

for $c < \frac{1}{2n}$. Then Ψ is a wavelet satisfying (2.4) and (2.5).

Proof. Without loss of generality, assume $\alpha < \beta < \kappa < 1$. First, it's clear that $\int_{\mathbb{R}^n} \Psi = 0$. We now just need to verify properties (2.4) and (2.5). Assume $c > 0$. We can modify the proof accordingly if $c < 0$. Then

$$\begin{aligned}|\Psi(x)| &= \left|(1-c)^{-n/2}\psi_{(1-c)}(x) - \psi(x)\right| \\ &= (1-c)^{-n}\left|\psi\left(\frac{x}{(1-c)}\right) - (1-c)^n\psi(x)\right| \\ &\leq (1-c)^{-n}\left|\psi\left(\frac{x}{1-c}\right) - \psi\left(\frac{1-c}{1-c}x\right)\right| + (1-c)^{-n}\sum_{j=1}^{n}\binom{n}{j}c^j\,|\psi(x)|\,.\end{aligned}$$

Now use mean value theorem on the first term to choose a point z on the segment connecting $\frac{x}{1-c}$ and x such that

$$\frac{c}{1-c}\left|[\nabla\psi(z)]^T x\right| = \left|\psi\left(\frac{x}{1-c}\right) - \psi\left(\frac{1-c}{1-c}x\right)\right|.$$

We now use Cauchy-Schwarz to bound the left side:

$$\frac{c}{1-c}\left|[\nabla\psi(z)]^T x\right| \leq \frac{c}{1-c}\frac{A|x|}{(1+|z|)^{n+1+\beta}}.$$

Since z lies on the segment connecting $\frac{x}{1-c}$ and x, we see that for some $t \in [0,1]$, we have

$$\begin{aligned}z &= (1-t)\frac{x}{1-c} + tx \\ &= \frac{1-t}{1-c}x + \frac{t-tc}{1-c}x \\ &= \frac{1-t+t-tc}{1-c}x \\ &= \frac{1-tc}{1-c}x.\end{aligned}$$

Thus, $|z| \geq |x|$. It now follows that

$$\frac{c}{1-c}\frac{A|x|}{(1+|z|)^{n+1+\beta}} \leq \frac{c}{1-c}\frac{A}{(1+|x|)^{n+\beta}}.$$

Finally, we get

$$\begin{aligned}|\Psi_\lambda(x)| &\leq \frac{c}{(1-c)^{n+1}}\frac{A}{(1+|x|)^{n+\beta}} + \frac{\sum_{j=1}^{n}\binom{n}{j}c^j}{(1-c)^{n+1}}\frac{A}{(1+|x|)^{n+\alpha}} \\ &\leq 2A\left(\frac{2n}{2n-1}\right)^{-n-1}\frac{\sum_{j=1}^{n}\binom{n}{j}c^j}{(1+|x|)^{n+\alpha}} \\ &\leq \frac{A_n c}{(1+|x|)^{n+\alpha}}\end{aligned}$$

for some constant A_n since we assume $\alpha < \beta$ and $c < \frac{1}{2n}$. Thus, (2.4) is satisfied.

We use a similar idea for proving (2.5) holds. Assume $c > 0$ without loss of generality and further assume that $|x| \geq 2|y|$. By Mean Value Theorem, there exists z on the line segment connecting x and $x - y$ such that

$$|\Psi(x-y) - \Psi(x)| = |\nabla\Psi(z)||y|.$$

Like before, we notice that

$$\begin{aligned}
|\nabla\Psi(z)| &= \left|(1-c)^{-n/2}\nabla\psi_{(1-c)}(z) - \nabla\psi(z)\right| \\
&= \left|(1-c)^{-n-1}\nabla\psi\left(\frac{z}{1-c}\right) - \nabla\psi(z)\right| \\
&= (1-c)^{-n-1}\left|\nabla\psi\left(\frac{z}{1-c}\right) - (1-c)^{n+1}\nabla\psi(z)\right| \\
&\leq (1-c)^{-n-1}\left|\nabla\psi\left(\frac{z}{1-c}\right) - \nabla\psi\left(\frac{1-c}{1-c}z\right)\right| + (1-c)^{-n-1}\sum_{j=1}^{n+1}\binom{n+1}{j}c^j\left|\nabla\psi\left(z\right)\right|.
\end{aligned}$$

Let S be the set of points on the segment connecting $\frac{z}{1-c}$ and z. By Mean Value Inequality, since S is closed and bounded, we have

$$\left|\nabla\psi\left(\frac{z}{1-c}\right) - \nabla\psi\left(\frac{1-c}{1-c}z\right)\right| \leq \frac{c}{1-c}\max_{w\in S}\left\|D^2\psi(w)\right\|_\infty |z|.$$

The maximum for the quantity above is attained in S, so let us say the maximizer is $w_1 = (1-t)\frac{z}{1-c}+tz$ for some $t \in [0,1]$. Now use decay of the Hessian to bound the right side:

$$\frac{c}{1-c}\max_{w\in S}\left\|D^2\psi(w)\right\|_\infty |z| \leq \frac{c}{1-c}\frac{A|z|}{(1+|w_1|)^{n+1+\kappa}}.$$

It follows that

$$\begin{aligned}
w_1 &= (1-t)\frac{z}{1-c} + tz \\
&= \frac{1-t}{1-c}z + \frac{t-tc}{1-c}z \\
&= \frac{1-t+t-tc}{1-c}z \\
&= \frac{1-tc}{1-c}z.
\end{aligned}$$

Thus, $|w_1| \geq |z|$. We conclude

$$\frac{c}{1-c}\frac{A|z|}{(1+|w_1|)^{n+1+\kappa}} \leq \frac{c}{1-c}\frac{A}{(1+|z|)^{n+\kappa}}.$$

For bounding $|\nabla\Psi(z)|$, we see

$$\begin{aligned}
|\nabla\Psi(z)| &\leq \frac{c}{(1-c)^{n+2}}\frac{A}{(1+|z|)^{n+\kappa}} + \frac{\sum_{j=1}^{n+1}\binom{n+1}{j}c^j}{(1-c)^{n+1}}\frac{A}{(1+|z|)^{n+1+\beta}} \\
&\leq A(1-c)^{-n-2}\frac{2\sum_{j=1}^{n+1}\binom{n+1}{j}c^j}{(1+|z|)^{n+\kappa}} \\
&\leq \left(\frac{2n}{2n-1}\right)^{n+2}\frac{2A\sum_{j=1}^{n+1}\binom{n+1}{j}c^j}{(1+|z|)^{n+\kappa}}.
\end{aligned}$$

Going back to proving (2.5) holds for Ψ,

$$|\Psi(x-y)-\Psi(x)| = |\nabla\Psi(z)||y| \leq \left(\frac{2n}{2n-1}\right)^{n+2}\frac{2A\sum_{j=1}^{n+1}\binom{n+1}{j}c^j|y|}{(1+|z|)^{n+\kappa}}.$$

since the point z lies on the lines on a line segment connecting $x-y$ and x with $|x| \geq 2|y|$, we can use an argument similar to above to conclude

$$|\Psi(x-y)-\Psi(x)| \leq 2^{n+1+\kappa}\left(\frac{2n}{2n-1}\right)^{n+2}\frac{A\sum_{j=1}^{n+1}\binom{n+1}{j}c^j}{(1+|x|)^{n+\kappa}}|y|.$$

Now integrate to get

$$\begin{aligned}
\int_{|x|\geq 2|y|}|\Psi(x-y)-\Psi(x)|\,dx &\leq 2^{n+1+\kappa}\left(\frac{2n}{2n-1}\right)^{n+2} A\sum_{j=1}^{n+1}\binom{n+1}{j}c^j|y|\int_{|x|\geq 2|y|}\frac{dx}{|x|^{n+\kappa}} \\
&= 2^{n+1+\kappa}\left(\frac{2n}{2n-1}\right)^{n+2} AI_n\sum_{j=1}^{n+1}\binom{n+1}{j}c^j|y|^{1-\kappa},
\end{aligned}$$

where I_n is some constant associated with the integration. Finally, we have a bound of

$$\int_{|x|\geq 2|y|}|\Psi(x-y)-\Psi(x)|\,dx \leq \tilde{A}_n c|y|^{1-\kappa}.$$

for some constant $\tilde{A}_n$ only dependent on the dimension n. Thus, (2.5) holds with exponent $1-\kappa \in (0,1)$. Let $\hat{A}_n = \max\{A_n, \tilde{A}_n\}$. It follows that

$$\begin{aligned}
|\Psi_\lambda(x)| &\leq \frac{\hat{A}_n c}{(1+|x|)^{n+\alpha}} \\
\int_{|x|\geq 2|y|}|\Psi(x-y)-\Psi(x)|\,dx &\leq \hat{A}_n c|y|^{1-\kappa}.
\end{aligned}$$

$\square$

It follows from Problem 6.1.2 in [38] that the bound in the G-function depends linearly on the constant A from Theorem 5 when proving $\mathbf{L}^2(\mathbb{R}^n)$ boundedness. Thus, the following corollaries hold.

Corollary 15. *Assume $|c| < \frac{1}{2n}$. For ψ satisfying the conditions of Lemma 14, when $1 < p < \infty$, there exist constants $C_{n,p}$ and $\hat{C}_{n,p}$ such that*

$$\left\| \left(\int_0^\infty |f * \Psi_\lambda(x)|^2 \frac{d\lambda}{\lambda^{n+1}} \right)^{1/2} \right\|_{\mathbf{L}^p(\mathbb{R}^n)} \leq c \cdot C_{n,p} \max\{p, (p-1)^{-1}\} \|f\|_{\mathbf{L}^p(\mathbb{R}^n)}$$

and

$$\left\| \left(\sum_{j \in \mathbb{Z}} |f * \Psi_j(x)|^2 \right)^{1/2} \right\|_{\mathbf{L}^p(\mathbb{R}^n)} \leq c \cdot \hat{C}_n \max\{p, (p-1)^{-1}\} \|f\|_{\mathbf{L}^p(\mathbb{R}^n)}.$$

Alternatively, if one of the following holds:

- *$n = 1$, ψ is complex analytic and satisfies the conditions of Lemma 14,*
- *$n \geq 2$ and ψ satisfies the conditions of Lemma 9,*

there exist constants H_n and $\hat{H}_n$ such that

$$\left\| \left(\int_0^\infty |f * \Psi_\lambda(x)|^2 \frac{d\lambda}{\lambda^{n+1}} \right)^{1/2} \right\|_{\mathbf{L}^1(\mathbb{R}^n)} \leq c \cdot H_n \|f\|_{\mathbf{H}^1(\mathbb{R}^n)}$$

and

$$\left\| \left(\sum_{j \in \mathbb{Z}} |f * \Psi_j(x)|^2 \right)^{1/2} \right\|_{\mathbf{L}^1(\mathbb{R}^n)} \leq c \cdot \hat{H}_n \|f\|_{\mathbf{H}^1(\mathbb{R}^n)}.$$

Now we can use the results above for our main dilation stability results.

Theorem 16. *Suppose that ψ is a wavelet that satisfies the conditions of Lemma 14. Then there exists a constants $K_{n,m}$ and $\hat{K}_{n,m}$ only dependent on n and m such that*

$$\|S^m_{cont,2} f - S^m_{cont,2} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \leq |c| \cdot K_{n,m} \|f\|_2$$

and

$$\|S^m_{dyad,2} f - S^m_{dyad,2} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \leq |c| \cdot \hat{K}_{n,m} \|f\|_2$$

for any $|c| < \frac{1}{2n}$.

Proof. Without loss of generality, assume $c > 0$. Let

$$\begin{aligned}
\mathcal{W}_t f &= f * \psi_t \\
M f &= |f| \\
U_t &= M \mathcal{W}_t \\
A_q f &= \left(\int_{\mathbb{R}^n} f^q(x)\, dx \right)^{1/q}.
\end{aligned}$$

It follows that $S^m_{\text{cont},2} = A_2 M W_{\lambda_m} U_{\lambda_{m-1}} \cdots U_{\lambda_1}$. We will also let $V_{m-1} = U_{\lambda_{m-1}} \cdots U_{\lambda_1}$, with V_0 being the identity operator, and make a slight abuse of notation by denoting $\mathcal{W}_{\lambda_m}$ as $\mathcal{W}$. First, we will add and subtract $A_2 M L_\tau \mathcal{W} V_{m-1} f$ and apply triangle inequality:

$$\begin{aligned}
\|S^m_{\text{cont},2} f - S^m_{\text{cont},2} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} &= \|A_2 M \mathcal{W} V_{m-1} f - A_2 M \mathcal{W} V_{m-1} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&\leq \|A_2 M \mathcal{W} V_{m-1} f - A_2 M L_\tau \mathcal{W} V_{m-1} f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&+ \|A_2 M L_\tau \mathcal{W} V_{m-1} f - A_2 M \mathcal{W} V_{m-1}, L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)}.
\end{aligned}$$

We'll start by bounding the first term. We see that $g = \mathcal{W} V_{m-1} f \in \mathbf{L}^2(\mathbb{R}^n)$. Thus

$$|A_2 M \mathcal{W} V_{m-1} f - A_2 M L_\tau \mathcal{W} V_{m-1} f| = |\|g\|_2 - \|L_\tau g\|_2|.$$

Now use a change of variables:

$$\|L_\tau g\|_2^2 = \int_{\mathbb{R}^n} |g((1-c)x)|^2\, dx = (1-c)^{-n} \|g\|_2^2.$$

It then follows that

$$|\|L_\tau g\|_2 - \|g\|_2| = \|g\|_2 \left(\frac{1}{(1-c)^{n/2}} - 1 \right) \leq \|g\|_2 \left(\frac{1}{(1-c)^n} - 1 \right).$$

Taking the scattering norm yields

$$\begin{aligned}
\|A_2 M \mathcal{W} V_{m-1} f - A_2 M L_\tau \mathcal{W} V_{m-1} f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)} &\le \left(\frac{1}{(1-c)^n} - 1\right)^2 \|S^m_{\mathrm{cont},2} f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&= \left(\frac{1-(1-c)^n}{(1-c)^n}\right)^2 \|S^m_{\mathrm{cont},2} f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&= \left(\frac{1}{(1-c)^n} \sum_{j=1}^n \binom{n}{j} c^j\right)^2 \|S^m_{\mathrm{cont},2} f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&\le \left[\left(\frac{2n}{2n-1}\right)^n \sum_{j=1}^n \binom{n}{j} c^j\right]^2 \|S^m_{\mathrm{cont},2} f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&\le c^2 \cdot C_{m,n} \|f\|_2^2.
\end{aligned}$$

For the second term, apply Minkwoski's inequality for 2 norms:

$$\begin{aligned}
&\|A_2 M L_\tau \mathcal{W} V_{m-1} f - A_2 M \mathcal{W} V_{m-1} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&= \left(\int_0^\infty \cdots \int_0^\infty \left|\|L_\tau \mathcal{W} V_{m-1} f\|_2 - \|\mathcal{W} L_\tau V_{m-1} f\|_2\right|^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} \\
&\le \left(\int_0^\infty \cdots \int_0^\infty \|L_\tau \mathcal{W} V_{m-1} f - \mathcal{W} L_\tau V_{m-1} f\|_2^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} \\
&= \|A_2 M [\mathcal{W} V_{m-1}, L_\tau] f\|_{\mathbf{L}^2(\mathbb{R}^m_+)}.
\end{aligned}$$

Now this is a commutator term, and we can now bound:

$$\begin{aligned}
\|A_2 M [\mathcal{W} V_{m-1}, L_\tau] f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)} &= \int_0^\infty \cdots \int_0^\infty \|[\mathcal{W} V_{m-1}, L_\tau] f\|_2^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}} \\
&= \||[\mathcal{W} V_{m-1}, L_\tau] f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+ \times \mathbb{R}^n)} \\
&\le \|[\mathcal{W} V_{m-1}, L_\tau]\|^2_{\mathbf{L}^2(\mathbb{R}^m_+ \times \mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^n)} \|f\|_2^2.
\end{aligned}$$

We examine the commutator term more closely. Without a loss of generality, assume $m \ge 2$. By expanding it, we see that each term contains $[\mathcal{W}, L_\tau]$. It follows that

$$\begin{aligned}
&\|[\mathcal{W} V_{m-1}, L_\tau]\|_{\mathbf{L}^2(\mathbb{R}^m_+ \times \mathbb{R}^n)} \\
&\le m \|\mathcal{W}\|^{m-1}_{\mathbf{L}^2(\mathbb{R}_+ \times \mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^n)} \|M\|^{m-1}_{\mathbf{L}^2(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^n)} \|[\mathcal{W}, L_\tau]\|_{\mathbf{L}^2(\mathbb{R}_+ \times \mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^n)} \\
&\le C_m \|[\mathcal{W}, L_\tau]\|_{\mathbf{L}^2(\mathbb{R}_+ \times \mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^n)}.
\end{aligned}$$

Thus, once we bound this quantity appropriately, we will finish the proof. We start by writing

$$\|[\mathcal{W}, L_\tau]f\|^2_{\mathbf{L}^2(\mathbb{R}_+\times\mathbb{R}^n)} = \int_0^\infty \|(L_\tau f) * \psi_\lambda - L_\tau (f * \psi_\lambda)\|_2^2 \, \frac{d\lambda}{\lambda^{n+1}} .$$

By substitution with $z = (1-c)x$ and Lemma 13,

$$\begin{aligned}
&\|(L_\tau f) * \psi_\lambda - L_\tau (f * \psi_\lambda)\|_2^2 \\
&= \int_{\mathbb{R}^n} |(L_\tau f * \psi_\lambda)(x) - L_\tau (f * \psi_\lambda)(x)|^2 \, dx \\
&= \int_{\mathbb{R}^n} \left|(1-c)^{-n/2} \left(f * \psi_{(1-c)\lambda}\right)((1-c)x) - (f * \psi_\lambda)((1-c)x)\right|^2 \, dx \\
&= (1-c)^{-n} \int_{\mathbb{R}^n} \left|(1-c)^{-n/2} \left(f * \psi_{(1-c)\lambda}\right)(z) - (f * \psi_\lambda)(z)\right|^2 \, dz \\
&= (1-c)^{-n} \int_{\mathbb{R}^n} \left|f * \left((1-c)^{-n/2}\psi_{(1-c)\lambda} - \psi_\lambda\right)\right|^2 \, dz \\
&= (1-c)^{-n} \int_{\mathbb{R}^n} |(f * \Psi_\lambda)(z)|^2 \, dz, \\
&= (1-c)^{-n} \|f * \Psi_\lambda\|_2^2 .
\end{aligned}$$

Thus, we obtain

$$\begin{aligned}
\int_0^\infty \|(L_\tau f) * \psi_\lambda - L_\tau (f * \psi_\lambda)\|_2^2 \, \frac{d\lambda}{\lambda^{n+1}} &= (1-c)^{-n} \int_0^\infty \|f * \Psi_\lambda\|_2^2 \frac{d\lambda}{\lambda^{n+1}} \\
&= (1-c)^{-n} \int_{\mathbb{R}^n} \int_0^\infty |f * \Psi_\lambda(x)|^2 \, \frac{d\lambda}{\lambda^{n+1}} \, dx \\
&= (1-c)^{-n} \left\| \left(\int_0^\infty |f * \Psi_\lambda(x)|^2 \, \frac{d\lambda}{\lambda^{n+1}} \right)^{1/2} \right\|_2^2 \\
&\leq c^2 \cdot \left(\frac{2n}{2n-1} \right)^n C_{n,p} \|f\|_2^2 .
\end{aligned}$$

It follows that

$$\|S^m_{\text{cont},2} f - S^m_{\text{cont},2} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \leq |c| \cdot K_{n,m} \|f\|_2$$

for any $c < \frac{1}{2n}$. □

As is customary at this point, we have the following corollaries. We start with the case where $1 < q < 2$.

Corollary 17. *Assume $|c| < \frac{1}{2n}$. For $q \in (1,2)$, there exists constants $K_{n,m,q}$ and $\hat{K}_{n,m,q}$ such that*

$$\|S^m_{cont,q} f - S^m_{cont,q} L_\tau f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} \leq |c|^q \cdot K_{n,m,q} \|f\|^q_q$$

and

$$\|S^m_{dyad,q} f - S^m_{dyad,q} L_\tau f\|^q_{\ell^2(\mathbb{Z}^m)} \leq |c|^q \cdot \hat{K}_{n,m,q} \|f\|^q_q.$$

Proof. Without loss of generality again, assume $c > 0$. First, we will add and subtract $A_q M L_\tau \mathcal{W} V_{m-1} f$ and apply triangle inequality:

$$\begin{aligned} \|S^m_{\mathrm{cont},q} f - S^m_{\mathrm{cont},q} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} &= \|A_q M \mathcal{W} V_{m-1} f - A_q M \mathcal{W} V_{m-1} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \\ &\leq \|A_q M \mathcal{W} V_{m-1} f - A_q M L_\tau \mathcal{W} V_{m-1} f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \\ &+ \|A_q M L_\tau \mathcal{W} V_{m-1} f - A_q M \mathcal{W} V_{m-1}, L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)}. \end{aligned}$$

We'll start by bounding the first term again. Define $g = \mathcal{W} V_{m-1} f \in \mathbf{L}^q(\mathbb{R}^n)$. and we have

$$|A_q M \mathcal{W} V_{m-1} f - A_q M L_\tau \mathcal{W} V_{m-1} f| = \left| \|g\|_q - \|L_\tau g\|_q \right|.$$

By change of variables,

$$\left| \|g\|_q - \|L_\tau g\|_q \right| = \|g\|_q \left(\frac{1}{(1-c)^{n/q}} - 1 \right) \leq \|g\|_q \left(\frac{1}{(1-c)^n} - 1 \right).$$

Again, we have

$$\begin{aligned} \|A_q M \mathcal{W} V_{m-1} f - A_q M L_\tau \mathcal{W} V_{m-1} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} &\leq \left(\frac{1}{(1-c)^{n/q}} - 1 \right)^q \|S^m_{\mathrm{cont},2} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} \\ &\leq \left(\frac{1}{(1-c)^n} - 1 \right)^q \|S^m_{\mathrm{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} \\ &= \left[\frac{1-(1-c)^n}{(1-c)^n} \right]^q \|S^m_{\mathrm{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} \\ &= \left[\frac{1}{(1-c)^n} \sum_{j=1}^n \binom{n}{j} c^j \right]^q \|S^m_{\mathrm{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} \\ &\leq \left[\left(\frac{2n}{2n-1} \right)^n \sum_{j=1}^n \binom{n}{j} c^j \right]^q \|S^m_{\mathrm{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} \\ &\leq |c|^q \cdot C_{m,n} \|f\|^q_q. \end{aligned}$$

For the second term, apply Minkoski's inequality for q norms:

$$\begin{aligned}
&\|A_q M L_\tau \mathcal{W} V_{m-1} f - A_q M \mathcal{W} V_{m-1}, L_\tau f\|_{\mathbf{L}^2(\mathbb{R}_+^m)} \\
&= \left(\int_0^\infty \cdots \int_0^\infty \left|\|L_\tau \mathcal{W} V_{m-1} f\|_q - \|\mathcal{W} L_\tau V_{m-1} f\|_q\right|^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} \\
&\leq \left(\int_0^\infty \cdots \int_0^\infty \|L_\tau \mathcal{W} V_{m-1} f - \mathcal{W} L_\tau V_{m-1} f\|_q^2 \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{1/2} \\
&= \|A_q M [\mathcal{W} V_{m-1}, L_\tau] f\|_{\mathbf{L}^2(\mathbb{R}_+^m)}.
\end{aligned}$$

Via a similar reduction technique for Theorem 16, we can reduce to a commutator bound above. Additionally, we have

$$\|(L_\tau f) * \psi_\lambda - L_\tau (f * \psi_\lambda)\|_q^q = (1-c)^{-n} \|f * \Psi_\lambda\|_q^q \,.$$

Thus,

$$\begin{aligned}
\|A_q M [\mathcal{W}, L_\tau] f\|_{\mathbf{L}^2(\mathbb{R}_+^m)}^q &= \left(\int_0^\infty \|(L_\tau f) * \psi_\lambda - L_\tau (f * \psi_\lambda)\|_q^2 \frac{d\lambda}{\lambda^{n+1}}\right)^{q/2} \\
&= (1-c)^{-n} \left(\int_0^\infty \|f * \Psi_\lambda\|_q^2 \frac{d\lambda}{\lambda^{n+1}}\right)^{q/2} \\
&\leq (1-c)^{-n} \left\| \left(\int_0^\infty |f * \Psi_\lambda(x)|^2 \frac{d\lambda}{\lambda^{n+1}}\right)^{1/2} \right\|_q^q \\
&\leq |c|^q \cdot \tilde{C}_n \|f\|_q^q.
\end{aligned}$$

It follows that

$$\|S_{\mathrm{cont},q}^m f - S_{\mathrm{cont},q}^m L_\tau f\|_{\mathbf{L}^2(\mathbb{R}_+^m)}^q \leq |c|^q \cdot K_{n,m} \|f\|_q^q$$

for any $|c| < \frac{1}{2n}$. □

Additionally, for the case of $q = 1$, we have the following corollary that we will state, but not prove, since the idea is the same as the previous corollary.

Corollary 18. *Suppose one of the following holds:*

- $n = 1$, ψ *is complex analytic and satisfies the conditions of Lemma 14,*

- *$n \geq 2$ and ψ satisfies the conditions of Lemma 9,*

then there exist constants $K_{H,m}$ and $\hat{K}_{H,m}$ such that

$$\|S^m_{cont,1} f - S^m_{cont,1} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \leq c \cdot K_{H,m} \|f\|_{\mathbf{H}^1(\mathbb{R}^n)}$$

and

$$\|S^m_{dyad,1} f - S^m_{dyad,1} L_\tau f\|_{\ell^2(\mathbb{Z}^m)} \leq c \cdot \hat{K}_{H,m} \|f\|_{\mathbf{H}^1(\mathbb{R}^n)}.$$

2.4 Stability to Diffeomorphisms

We now focus on the stability of $S^m_{\text{cont},q} f$ for general diffeomorphisms with $\|D\tau\|_\infty < \frac{1}{2n}$. The corresponding operator for diffeomorphisms is defined as $L_\tau f(x) = f(x - \tau(x))$.

2.4.1 Stability to Diffeomorphisms When $q = 2$

Proposition 19. *Assume ψ and its first and second order derivatives have decay* in $O((1+|x|)^{-n-3})$, and $\int_{\mathbb{R}^n} \psi(x)\, dx = 0$. Then for every $\tau \in C^2(\mathbb{R}^n)$ with $\|D\tau\|_\infty \leq \frac{1}{2n}$, there exists $\tilde{C}_n > 0$ such that:*

$$\|[\mathcal{W}, L_\tau]\|_{\mathbf{L}^2(\mathbb{R}_+ \times \mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^n)} \leq \tilde{C}_n \left(\|D\tau\|_\infty \left(\log \frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty} \vee 1 \right) + \|D^2\tau\|_\infty \right).$$

Proof. The argument is a continuous version of Lemma 2.14 in [11]. We will first show how to transform our commutator term into an analogous commutator term from [11]. To shorten notation, we will denote $\|[\mathcal{W}, L_\tau]\|_{\mathbf{L}^2(\mathbb{R}_+ \times \mathbb{R}^n)}$ as $\|[\mathcal{W}, L_\tau]\|$. We have

$$\begin{aligned}
\|[\mathcal{W}, L_\tau] f\|^2_{\mathbf{L}^2(\mathbb{R}_+ \times \mathbb{R}^n)} &= \int_0^\infty \|[\mathcal{W}_t, L_\tau] f\|_2^2 \, \frac{dt}{t^{n+1}} \\
&= \int_0^\infty \|\psi_t * (L_\tau f) - L_\tau(\psi_t * f)\|_2^2 \frac{dt}{t^{n+1}} \\
&= \int_0^\infty \int_{\mathbb{R}^n} |\psi_t * (L_\tau f) - L_\tau(\psi_t * f)|^2 \, dx \, \frac{dt}{t^{n+1}}.
\end{aligned}$$

Notice that $\psi_{\frac{1}{t}}(x) = t^{n/2} \psi(tx)$. Use the change of variables $t = \frac{1}{\lambda}$ to get

$$\begin{aligned}
\|[\mathcal{W}, L_\tau] f\|^2_{\mathbf{L}^2(\mathbb{R}_+ \times \mathbb{R}^n)} &= \int_0^\infty \left\| \psi_{\frac{1}{\lambda}} * (L_\tau f) - L_\tau(\psi_{\frac{1}{\lambda}} * f) \right\|_2^2 \lambda^{n-1} \, d\lambda \\
&= \int_0^\infty \left\| \lambda^{n/2} \psi_{\frac{1}{\lambda}} * (L_\tau f) - L_\tau(\lambda^{n/2} \psi_{\frac{1}{\lambda}} * f) \right\|_2^2 \frac{d\lambda}{\lambda}.
\end{aligned}$$

*Similar to [31], we have found that there needs to be $O((1 + |x|)^{-n-2+\alpha})$ decay for some $\alpha > 0$ to bound (E.26) in [11].

Define $\mathscr{W}_\lambda f = f * \lambda^{n/2}\psi_{\frac{1}{\lambda}}$ with $\lambda^{n/2}\psi_{\frac{1}{\lambda}}(x) = \lambda^n\psi(\lambda x)$. In other words, $\mathscr{W}_t$ is a convolution with an $\mathbf{L}^1$ normalized wavelet, which matches with the normalization in [11]. Now we have

$$\|[\mathcal{W}, L_\tau]f\|^2_{\mathbf{L}^2(\mathbb{R}_+\times\mathbb{R}^n)} = \int_0^\infty \|[\mathscr{W}_\lambda, L_\tau]f\|_2^2 \,\frac{d\lambda}{\lambda}.$$

This implies

$$[\mathcal{W}, L_\tau]^*[\mathcal{W}, L_\tau] = \int_0^\infty [\mathscr{W}_\lambda, L_\tau]^*[\mathscr{W}_\lambda, L_\tau] \,\frac{d\lambda}{\lambda}$$

Defining $K_\lambda = \mathscr{W}_\lambda - L_\tau \mathscr{W}_\lambda L_\tau^{-1}$ so that $[\mathscr{W}_\lambda, L_\tau] = K_\lambda L_\tau$, we have:

$$\begin{aligned}
\|[\mathcal{W}, L_\tau]\| &= \|[\mathcal{W}, L_\tau]^*[\mathcal{W}, L_\tau]\|^{1/2} \\
&= \left\| \int_0^\infty [\mathscr{W}_\lambda, L_\tau]^*[\mathscr{W}_\lambda, L_\tau] \,\frac{d\lambda}{\lambda} \right\|^{1/2} \\
&= \left\| \int_0^\infty L_\tau^* K_\lambda^* K_\lambda L_\tau \,\frac{d\lambda}{\lambda} \right\|^{1/2} \\
&\leq \|L_\tau\| \cdot \left\| \int_0^\infty K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda} \right\|^{1/2},
\end{aligned}$$

Since $\|L_\tau f\|_2^2 \leq \left(\frac{1}{1-n\|D\tau\|_\infty}\right)\|f\|_2^2$,

$$\|L_\tau\| \leq \frac{1}{1 - n\|D\tau\|_\infty} \leq 2$$

and the problem is reduced to bounding $\left\|\int_0^\infty K_\lambda^* K_\lambda \,\lambda^{-1}\, d\lambda\right\|^{1/2}$. Let $\gamma \geq 1$. The integral is divided into three pieces:

$$\begin{aligned}
\left\| \int_0^\infty K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda} \right\|^{1/2} &\leq \left(\left\| \int_0^{2^{-\gamma}} K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda} \right\| + \left\| \int_{2^{-\gamma}}^1 K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda} \right\| + \left\| \int_1^\infty K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda} \right\| \right)^{1/2} \\
&\leq \left\| \int_0^{2^{-\gamma}} K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda} \right\|^{1/2} + \left\| \int_{2^{-\gamma}}^1 K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda} \right\|^{1/2} + \left\| \int_1^\infty K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda} \right\|^{1/2} \\
&= P_1 + P_2 + P_3.
\end{aligned}$$

To bound P_1, we decompose $K_\lambda = \tilde{K}_{\lambda,1} + \tilde{K}_{\lambda,2}$, where the kernels defining $\tilde{K}_{\lambda,1}, \tilde{K}_{\lambda,2}$ are

$$\begin{aligned}
\tilde{k}_{\lambda,1}(x,u) &:= (1 - \det(I - D\tau(u)))\lambda^n\psi(\lambda(x-u)) \\
&:= a(u)\lambda^n\psi(\lambda(x-u)), \\
\tilde{k}_{\lambda,2}(x,u) &:= \det(I - D\tau(u))(\lambda^n\psi(\lambda(x-u)) - \lambda^n\psi(\lambda(x - \tau(x) - u + \tau(u))),
\end{aligned}$$

respectively. Since our normalization matches with [11], E.13 implies that there exists a constant C_n such that

$$\|\tilde{K}_{\lambda,2}\| \leq C_n \lambda \|\Delta\tau\|_\infty.$$

We want to prove that

$$\left\| \int_0^1 \tilde{K}^*_{\lambda,1}\tilde{K}_{\lambda,1} \, \frac{d\lambda}{\lambda} \right\|^{1/2} \leq C_n \|D\tau\|_\infty.$$

Let $f \in \mathbf{L}^2(\mathbb{R}^n)$ be arbitrary and define $\tilde{\psi}(t) = \psi^*(-t)$. Based on [11], the kernel of $\tilde{K}^*_{\lambda,1}\tilde{K}_{\lambda,1}$ is given by

$$\tilde{k}_\lambda(y,z) := a(y)a(z)\lambda^{n/2}\tilde{\psi}_{\frac{1}{\lambda}} * \lambda^{n/2}\tilde{\psi}_{\frac{1}{\lambda}}(z-y).$$

Thus, it is sufficient to bound the quantity

$$\int_0^1 \|\tilde{K}^*_{\lambda,1}\tilde{K}_{\lambda,1} f\|_2^2 \frac{d\lambda}{\lambda}.$$

We see that $\|a\|_\infty \leq n\|D\tau\|_\infty$. Substituting in the kernel and bounding yields

$$\begin{aligned}
\int_0^1 \|\tilde{K}^*_{\lambda,1}\tilde{K}_{\lambda,1} f\|_2^2 \frac{d\lambda}{\lambda} &= \int_0^1 \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} a(y)a(z) \left(\lambda^{n/2}\tilde{\psi}_{\frac{1}{\lambda}} * \lambda^{n/2}\psi_{\frac{1}{\lambda}}\right)(z-y) f(y)\, dy \right|^2 dz \, \frac{d\lambda}{\lambda} \\
&= \int_0^1 \int_{\mathbb{R}^n} |a(z)|^2 \left| \int_{\mathbb{R}^n} a(y) \left(\lambda^{n/2}\tilde{\psi}_{\frac{1}{\lambda}} * \lambda^{n/2}\psi_{\frac{1}{\lambda}}\right)(z-y) f(y)\, dy \right|^2 dz \, \frac{d\lambda}{\lambda} \\
&\leq n^2 \|D\tau\|_\infty^2 \int_0^1 \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} a(y) \left(\lambda^{n/2}\tilde{\psi}_{\frac{1}{\lambda}} * \lambda^{n/2}\psi_{\frac{1}{\lambda}}\right)(z-y) f(y)\, dy \right|^2 dz \, \frac{d\lambda}{\lambda}.
\end{aligned}$$

Let $F(y) = a(y)f(y) \in \mathbf{L}^2(\mathbb{R}^n)$ and let $\mathcal{F}$ represent taking the Fourier Transform. Then we substitute $F(y)$ for $a(y)f(y)$ in the last line of the inequality above to get

$$\begin{aligned}
& n^2 \|D\tau\|_\infty^2 \int_0^1 \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} a(y) \left(\lambda^{n/2}\tilde{\psi}_{\frac{1}{\lambda}} * \lambda^{n/2}\psi_{\frac{1}{\lambda}}\right)(z-y) f(y)\, dy \right|^2 dz \, \frac{d\lambda}{\lambda} \\
&= n^2 \|D\tau\|_\infty^2 \int_0^1 \int_{\mathbb{R}^n} \left| \int_{\mathbb{R}^n} \left(\lambda^{n/2}\tilde{\psi}_{\frac{1}{\lambda}} * \lambda^{n/2}\psi_{\frac{1}{\lambda}}\right)(z-y) F(y)\, dy \right|^2 dz \, \frac{d\lambda}{\lambda} \\
&= n^2 \|D\tau\|_\infty^2 \int_0^1 \int_{\mathbb{R}^n} \left| \mathcal{F}\left(\lambda^{n/2}\tilde{\psi}_{\frac{1}{\lambda}} * \lambda^{n/2}\psi_{\frac{1}{\lambda}}\right)(\omega)\hat{F}(\omega) \right|^2 dz \, \frac{d\lambda}{\lambda} \\
&= n^2 \|D\tau\|_\infty^2 \int_{\mathbb{R}^n} |\hat{F}(\omega)|^2 \left(\int_0^1 |\hat{\psi}(\tfrac{\omega}{\lambda})|^4 \frac{d\lambda}{\lambda} \right) d\omega.
\end{aligned}$$

To finish up the argument, we make a substitution to rewrite

$$\int_0^1 |\hat{\psi}(\tfrac{\omega}{\lambda})|^4 \frac{d\lambda}{\lambda} = \int_1^\infty |\hat{\psi}(\lambda\omega)|^4 \frac{d\lambda}{\lambda}.$$

Using our decay assumptions on ψ and its partial derivatives, from Problem 6.1.3 in [38], we know that

$$|\hat{\psi}(\omega)| \leq M_\psi \min\{|\omega|, |\omega|^{-2}\}$$

for some constant M_ψ. Now, consider the quantity $\int_0^\infty |\hat{\psi}(\lambda\omega)|^4 \frac{d\lambda}{\lambda}$. Without loss of generality, assume that $|\omega| = 1$ since dilations of ω do not change the integral. It follows that

$$\int_0^\infty |\hat{\psi}(\lambda\omega)|^4 \frac{d\lambda}{\lambda} \leq M_\psi \int_0^1 \lambda^3 d\lambda + M_\psi \int_1^\infty \lambda^{-9} d\lambda < \infty,$$

and we conclude that

$$\int_1^\infty |\hat{\psi}(\lambda\omega)|^4 \frac{d\lambda}{\lambda} \leq A_\psi$$

for some constant A_ψ. To finish up,

$$\begin{aligned} n^2\|D\tau\|_\infty^2 \int_{\mathbb{R}^n} |\hat{F}(\omega)|^2 \left(\int_0^1 |\hat{\psi}(\tfrac{\omega}{\lambda})|^4 \frac{d\lambda}{\lambda} \right) d\omega &\leq n^2\|D\tau\|_\infty^2 A_\psi \int_{\mathbb{R}^n} |\hat{F}(\omega)|^2 \, d\omega \\ &\leq n^2\|D\tau\|_\infty^2 A_\psi \int_{\mathbb{R}^n} |a(z) f(z)|^2 \, dz \\ &\leq n^2\|D\tau\|_\infty^2 A_\psi \|f\|_2^2. \end{aligned}$$

Thus, we have the desired bound on $\left\| \int_0^1 \tilde{K}_{\lambda,1}^* \tilde{K}_{\lambda,1} \, \frac{d\lambda}{\lambda} \right\|^{1/2}$.

Substituting everything in yields

$$
\begin{aligned}
&\left\|\int_0^{2^{-\gamma}} K_\lambda^* K_\lambda \, \frac{d\lambda}{\lambda}\right\|^{1/2} \\
&= \left\|\int_0^{2^{-\gamma}} (\tilde{K}_{\lambda,1} + \tilde{K}_{\lambda,2})^* (\tilde{K}_{\lambda,1} + \tilde{K}_{\lambda,2}) \, \frac{d\lambda}{\lambda}\right\|^{1/2} \\
&= \left\|\int_0^{2^{-\gamma}} (\tilde{K}_{\lambda,1}^* \tilde{K}_{\lambda,1} + \tilde{K}_{\lambda,1}^* \tilde{K}_{\lambda,2} + \tilde{K}_{\lambda,2}^* \tilde{K}_{\lambda,1} + \tilde{K}_{\lambda,2}^* \tilde{K}_{\lambda,2}) \, \frac{d\lambda}{\lambda}\right\|^{1/2} \\
&\leq \left(\left\|\int_0^{2^{-\gamma}} \tilde{K}_{\lambda,1}^* \tilde{K}_{\lambda,1} \frac{d\lambda}{\lambda}\right\| + \left\|\int_0^{2^{-\gamma}} \tilde{K}_{\lambda,1}^* \tilde{K}_{\lambda,2} + \tilde{K}_{\lambda,2}^* \tilde{K}_{\lambda,1} + \tilde{K}_{\lambda,2}^* \tilde{K}_{\lambda,2} \, \frac{d\lambda}{\lambda}\right\|\right)^{1/2} \\
&\leq \left(\left\|\int_0^{2^{-\gamma}} \tilde{K}_{\lambda,1}^* \tilde{K}_{\lambda,1} \frac{d\lambda}{\lambda}\right\| + \int_0^{2^{-\gamma}} \|\tilde{K}_{\lambda,2}\|^2 \, \frac{d\lambda}{\lambda} + \int_0^{2^{-\gamma}} 2\|\tilde{K}_{\lambda,1}\|\|\tilde{K}_{\lambda,2}\| \, \frac{d\lambda}{\lambda}\right)^{1/2} \\
&\leq \left\|\int_0^{2^{-\gamma}} \tilde{K}_{\lambda,1}^* \tilde{K}_{\lambda,1} \frac{d\lambda}{\lambda}\right\|^{1/2} + \left(\int_0^{2^{-\gamma}} \|\tilde{K}_{\lambda,2}\|^2 \, \frac{d\lambda}{\lambda}\right)^{1/2} + \left(\int_0^{2^{-\gamma}} 2\|\tilde{K}_{\lambda,1}\|\|\tilde{K}_{\lambda,2}\| \, \frac{d\lambda}{\lambda}\right)^{1/2} \\
&\leq 2C_n \left(\|D\tau\|_\infty + \|\Delta\tau\|_\infty \left(\int_0^{2^{-\gamma}} \lambda^2 \, \frac{d\lambda}{\lambda}\right)^{1/2} + \|D\tau\|_\infty^{1/2} \|\Delta\tau\|_\infty^{1/2} \left(\int_0^{2^{-\gamma}} 2\lambda \, \frac{d\lambda}{\lambda}\right)^{1/2}\right) \\
&\leq 2C_n \left(\|D\tau\|_\infty + 2^{-\gamma}\|\Delta\tau\|_\infty + 2^{-\gamma/2}\|D\tau\|_\infty^{1/2}\|\Delta\tau\|_\infty^{1/2}\right) \\
&\leq 4C_n \left(\|D\tau\|_\infty + 2^{-\gamma}\|\Delta\tau\|_\infty\right).
\end{aligned}
$$

To bound P_3, we decompose $K_\lambda = K_{\lambda,1} + K_{\lambda,2}$, where the kernels defining $K_{\lambda,1}, K_{\lambda,2}$ are

$$
\begin{aligned}
k_{\lambda,1}(x,u) &= \lambda^n \psi(\lambda(x-u)) - \lambda^n \psi(\lambda(I - D\tau(u))(x-u)) \det(I - D\tau(u)) \\
k_{\lambda,2}(x,u) &= \det(I - D\tau(u)) \lambda^n \psi(\lambda(I - D\tau(u))(x-u)) - \lambda^n \psi(\lambda(x - \tau(x) - u + \tau(u))).
\end{aligned}
$$

A similar computation to the one for P_1 shows that:

$$
\begin{aligned}
\left\|\int_1^\infty K_\lambda^* K_\lambda \, \frac{d\lambda}{\lambda}\right\|^{1/2} \leq{}& \left\|\int_1^\infty K_{\lambda,1}^* K_{\lambda,1} \frac{d\lambda}{\lambda}\right\|^{1/2} + \left(\int_1^\infty \|K_{\lambda,2}\|^2 \, \frac{d\lambda}{\lambda}\right)^{1/2} \\
&+ \left(\int_1^\infty 2\|K_{\lambda,1}\|\|K_{\lambda,2}\| \, \frac{d\lambda}{\lambda}\right)^{1/2}.
\end{aligned}
$$

Letting $Q_j = K_{2^j,1}^* K_{2^j,1}$, it is shown in [11] that:

$$
\begin{aligned}
\|K_{\lambda,1}\| &\leq C_n \|D\tau\|_\infty \\
\|K_{\lambda,2}\| &\leq \min\{\lambda^{-n}\|D^2\tau\|_\infty, \|D\tau\|_\infty\} \\
\|Q_j Q_\ell\| &\leq C_n^2 2^{-|j-\ell|}(\|D\tau\|_\infty + \|D^2\tau\|_\infty)^4
\end{aligned}
$$

so that

$$\left\| \int_1^\infty K_{\lambda,1}^* K_{\lambda,1} \frac{d\lambda}{\lambda} \right\|^{1/2} = \left\| \int_0^\infty K_{2^j,1}^* K_{2^j,1} \log(2)\, dj \right\|^{1/2}$$
$$= \sqrt{\log(2)} \left\| \int_0^\infty Q_j\, dj \right\|^{1/2}.$$

We now apply a continuous version of Cotlar's Lemma (see Ch. 7 of [42], Sec. 5.5 for the continuous extension). We define:

$$\beta(j,\ell) = \begin{cases} C_n 2^{-|j-\ell|/2} (\|D\tau\|_\infty + \|D^2\tau\|_\infty)^2 & j \geq 0 \text{ and } \ell \geq 0 \\ 0 & \text{otherwise} \end{cases}.$$

Defining $Q_j = 0$ for $j < 0$, we have $\|Q_j^* Q_\ell\| \leq \beta(j,\ell)^2$ and $\|Q_j Q_\ell^*\| \leq \beta(j,\ell)^2$ for all j, ℓ. Thus by Cotlar's Lemma:

$$\left\| \int_{\mathbb{R}} Q_j\, dj \right\| \leq \sup_{j\in\mathbb{R}} \int_{\mathbb{R}} \beta(j,\ell)\, d\ell,$$
$$\left\| \int_0^\infty Q_j\, dj \right\| \leq \sup_{j\geq 0} \int_0^\infty \beta(j,\ell)\, d\ell$$
$$\leq C_n (\|D\tau\|_\infty + \|H\tau\|_\infty)^2 \left(\sup_{j\geq 0} \int_0^\infty 2^{-|j-\ell|/2}\, d\ell \right).$$

Now observing that with the change of variable $\lambda_1 = 2^j, \lambda_2 = 2^\ell$, we have $2^{-|j-\ell|/2} = \frac{\lambda_1}{\lambda_2} \wedge \frac{\lambda_2}{\lambda_1}$, we obtain:

$$\sup_{j\geq 0} \int_0^\infty 2^{-|j-\ell|/2}\, d\ell = \sup_{\lambda_1 \geq 1} \int_1^\infty \frac{(\lambda_1 \wedge \lambda_2)}{\sqrt{\lambda_1 \lambda_2}} \frac{d\lambda_2}{\ln(2)\lambda_2}$$
$$= \frac{1}{\ln(2)} \sup_{\lambda_1\geq 1} \left(\int_1^{\lambda_1} \frac{1}{\sqrt{\lambda_1\lambda_2}}\, d\lambda_2 + \int_{\lambda_1}^\infty \frac{\sqrt{\lambda_1}}{\lambda_2^{3/2}}\, d\lambda_2 \right)$$
$$= \frac{1}{\ln(2)} \sup_{\lambda_1\geq 1} \left(\frac{1}{\sqrt{\lambda_1}} (2\sqrt{\lambda_1} - 2) + \sqrt{\lambda_1} \left(\frac{2}{\sqrt{\lambda_1}} \right) \right)$$
$$= \frac{1}{\ln(2)} \sup_{\lambda_1\geq 1} \left(4 - \frac{2}{\sqrt{\lambda_1}} \right)$$
$$= \frac{4}{\ln(2)}$$

and conclude that

$$\left\| \int_1^\infty K_{\lambda,1}^* K_{\lambda,1} \frac{d\lambda}{\lambda} \right\|^{1/2} \leq 3C_n (\|D\tau\|_\infty + \|H\tau\|_\infty).$$

Thus we have:

$$\left\|\int_1^\infty K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda}\right\|^{1/2} \leq \left\|\int_1^\infty K_{\lambda,1}^* K_{\lambda,1} \frac{d\lambda}{\lambda}\right\|^{1/2} + \left(\int_1^\infty \|K_{\lambda,2}\|^2 \,\frac{d\lambda}{\lambda}\right)^{1/2}$$
$$+ \left(\int_1^\infty 2\|K_{\lambda,1}\|\|K_{\lambda,2}\| \,\frac{d\lambda}{\lambda}\right)^{1/2}.$$

Now we see that there exists a constant C_n such that

$$\begin{aligned}
\left\|\int_1^\infty K_{\lambda,1}^* K_{\lambda,1} \frac{d\lambda}{\lambda}\right\|^{1/2} &\leq C_n(\|D\tau\|_\infty + \|D^2\tau\|_\infty) \\
\left(\int_1^\infty \|K_{\lambda,2}\|^2 \,\frac{d\lambda}{\lambda}\right)^{1/2} &\leq C_n\|D^2\tau\|_\infty \left(\int_1^\infty \lambda^{-2n} \,\frac{d\lambda}{\lambda}\right)^{1/2} \\
\left(\int_1^\infty 2\|K_{\lambda,1}\|\|K_{\lambda,2}\| \,\frac{d\lambda}{\lambda}\right)^{1/2} &\leq C_n\|D\tau\|_\infty^{1/2}\|D^2\tau\|_\infty^{1/2} \left(\int_1^\infty 2\lambda^{-n} \,\frac{d\lambda}{\lambda}\right)^{1/2}.
\end{aligned}$$

and

$$\begin{aligned}
\left\|\int_1^\infty K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda}\right\|^{1/2} &\leq C_n\left(\|D\tau\|_\infty + \frac{1}{2n}\|D^2\tau\|_\infty + \frac{2}{n}\|D\tau\|_\infty^{1/2}\|D^2\tau\|_\infty^{1/2}\right) \\
&\leq C_n\left(\|D\tau\|_\infty + \frac{1}{2n}\|D^2\tau\|_\infty + \frac{1}{n}\|D\tau\|_\infty + \frac{1}{n}\|D^2\tau\|_\infty\right) \\
&\leq 2C_n(\|D\tau\|_\infty + \|D^2\tau\|_\infty).
\end{aligned}$$

Finally, we bound P_2. Note that in the previous section it was observed (shown in [11]) that

$$\begin{aligned}
\|K_{\lambda,1}\| &\leq C_n\|D\tau\|_\infty \\
\|K_{\lambda,2}\| &\leq \min\{\lambda^{-n}\|D^2\tau\|_\infty, \|D\tau\|_\infty\}.
\end{aligned}$$

The above two inequalities imply

$$\|K_\lambda\| = \|K_{\lambda,1} + K_{\lambda,2}\| \leq \|K_{\lambda,1}\| + \|K_{\lambda,2}\| \leq 2C_n\|D\tau\|_\infty$$

so that

$$\begin{aligned}
\left\|\int_{2^{-\gamma}}^1 K_\lambda^* K_\lambda \,\frac{d\lambda}{\lambda}\right\|^{1/2} &\leq \left(\int_{2^{-\gamma}}^1 \|K_\lambda\|^2 \,\frac{d\lambda}{\lambda}\right)^{1/2} \\
&\leq 2C_n\|D\tau\|_\infty \left(\int_{2^{-\gamma}}^1 \frac{d\lambda}{\lambda}\right)^{1/2} \\
&\leq 2C_n\|D\tau\|_\infty \,(-\ln(2^{-\gamma}))^{1/2} \\
&\leq 2C_n\gamma^{1/2}\|D\tau\|_\infty.
\end{aligned}$$

Putting everything together and since $\gamma \geq 1$, we obtain:

$$
\begin{aligned}
\|[\mathcal{W}, L_\tau]\| &\leq 2(P_1 + P_2 + P_3) \\
&\leq 4C_n\left(\|D\tau\|_\infty + 2^{-\gamma}\|\Delta\tau\|_\infty\right) + 2C_n\gamma^{1/2}\|D\tau\|_\infty + 3C_n(\|D\tau\|_\infty + \|D^2\tau\|_\infty) \\
&\leq \tilde{C}_n\left(\gamma\|D\tau\|_\infty + 2^{-\gamma}\|\Delta\tau\|_\infty + \|D^2\tau\|_\infty\right).
\end{aligned}
$$

Choosing $\gamma = \left(\log \frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty}\right) \vee 1$ gives

$$
\|[\mathcal{W}, L_\tau]\| \leq \tilde{C}_n\left(\left(\log \frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty} \vee 1\right)\|D\tau\|_\infty + \|D^2\tau\|_\infty\right),
$$

and the lemma is proved. □

Theorem 20. *Assume ψ and its first and second order derivatives have decay in $O((1+|x|)^{-n-3})$ and $\int_{\mathbb{R}^n}\psi(x)\,dx = 0$. Then for every $\tau \in C^2(\mathbb{R}^n)$ with $\|D\tau\|_\infty \leq \frac{1}{2n}$, there exists $C_{m,n} > 0$ and $\hat{C}_{m,n} > 0$ such that*

$$
\|S^m_{cont,2}f - S^m_{cont,2}L_\tau f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)} \leq C_{m,n}K_2(\tau)\|f\|_2^2.
$$

and

$$
\|S^m_{dyad,2}f - S^m_{dyad,2}L_\tau f\|^2_{\ell^2(\mathbb{Z}^m)} \leq \hat{C}_{m,n}K_2(\tau)\|f\|_2^2,
$$

with

$$
K_2(\tau) = \|D\tau\|^2_\infty + \left(\|D\tau\|_\infty\left(\log \frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty} \vee 1\right) + \|D^2\tau\|_\infty\right)^2.
$$

Proof. The proof is only provided for the continuous case. We have the following bound for some C_m:

$$
\begin{aligned}
\|S^m_{\text{cont},2}f - S^m_{\text{cont},2}L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} &\leq \|A_2M\mathcal{W}V_{m-1}f - A_2ML_\tau\mathcal{W}V_{m-1}f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&\quad + \|A_2M[\mathcal{W}V_{m-1}, L_\tau]f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&\leq \|A_2M\mathcal{W}V_{m-1}f - A_2ML_\tau\mathcal{W}V_{m-1}f\|_{\mathbf{L}^2(\mathbb{R}^m_+)} \\
&\quad + C_m^2\|[\mathcal{W}, L_\tau]\|^2_{\mathbf{L}^2(\mathbb{R}^m_+\times\mathbb{R}^n)\to\mathbf{L}^2(\mathbb{R}^n)}\|f\|_2^2.
\end{aligned}
$$

For the first term, we can mimic the dilation argument to get

$$|A_2 M\mathcal{W}V_{m-1}f - A_2 ML_\tau \mathcal{W}V_{m-1}f| = |\|g\|_2 - \|L_\tau g\|_2|\,.$$

The difference is the term with the diffeomorphism. Let $y = \gamma(x) = x - \tau(x)$. Then it follows that $\gamma^{-1}(y) = x$ and change of variables implies that

$$\|L_\tau f\|_2^2 = \int_{\mathbb{R}^n} |f(x-\tau(x))|^2\, dx = \int_{\mathbb{R}^n} |f(y)|^2 \frac{dy}{|\det(I - D\tau(\gamma^{-1}(y)))|}.$$

We also have

$$1 - n\|D\tau\|_\infty \leq |\det(I - D\tau(\gamma^{-1}(y)))| \leq 1 + n\|D\tau\|_\infty.$$

Thus, we obtain

$$\frac{1}{1+n\|D\tau\|_\infty}\int_{\mathbb{R}^n} |f(y)|^2\, dy \leq \|L_\tau f\|_2^2 \leq \frac{1}{1-n\|D\tau\|_\infty}\int_{\mathbb{R}^n} |f(y)|^2\, dy,$$
$$\frac{1}{1+n\|D\tau\|_\infty}\|f\|_2^2 \leq \|L_\tau f\|_2^2 \leq \frac{1}{1-n\|D\tau\|_\infty}\|f\|_2^2.$$

Since we have a bound on $\|D\tau\|_\infty$, we see that

$$\frac{1}{1+n\|D\tau\|_\infty} = \frac{1-n\|D\tau\|_\infty}{1-n^2\|D\tau\|_\infty^2} \geq 1 - n\|D\tau\|_\infty$$

since $1 > 1 - n^2\|D\tau\|_\infty^2 > 0$. Similarly,

$$\frac{1}{1-n\|D\tau\|_\infty} = \frac{1+2n\|D\tau\|_\infty}{1+n\|D\tau\|_\infty - 2n^2\|D\tau\|_\infty^2}$$

and

$$1 + n\|D\tau\|_\infty - 2n^2\|D\tau\|_\infty^2 \geq 1 + n\|D\tau\|_\infty - \frac{2n^2}{2n}\|D\tau\|_\infty = 1$$

since $\|D\tau\|_\infty \leq \frac{1}{2n}$. It follows that $\frac{1}{1-n\|D\tau\|_\infty} \leq 1 + 2n\|D\tau\|_\infty$ and

$$(1 - n\|D\tau\|_\infty)^{1/2}\|f\|_2 \leq \|L_\tau f\|_2 \leq (1 + 2n\|D\tau\|_\infty)^{1/2}\|f\|_2.$$

Since $1 - n\|D\tau\|_\infty < 1$ and $1 + 2n\|D\tau\|_\infty > 1$, Use the lower bound on $\|L_\tau f\|_2$ to get

$$\begin{aligned}\|f\|_2 - \|L_\tau f\|_2 &= \|f\|_2\left(1 - (1 - n\|D\tau\|_\infty)^{1/2}\right)\\ &\leq \|f\|_2\,(1 - (1 - n\|D\tau\|_\infty))\\ &= n\|D\tau\|_\infty\|f\|_2.\end{aligned}$$

and the upper bound to get

$$\begin{aligned}\|L_\tau f\|_2 - \|f\|_2 &= \|f\|_2 \left((1+2n\|D\tau\|_\infty)^{1/2} - 1\right)\\ &\leq \|f\|_2\left((1+2n\|D\tau\|_\infty) - 1\right)\\ &= 2n\|D\tau\|_\infty \|f\|_2.\end{aligned}$$

Finally, we have

$$|\|f\|_2 - \|L_\tau f\|_2| \leq 2n\|D\tau\|_\infty \|f\|_2$$

for any $f \in \mathbf{L}^2(\mathbb{R}^n)$. Now we mimic the argument given for dilation stability to get

$$\|A_2 M\mathcal{W}V_{m-1}f - A_2 M L_\tau \mathcal{W}V_{m-1}f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)} \leq C\|D\tau\|^2_\infty \|f\|^2_2$$

for some constant C. For the second term, we have

$$C_m^2 \|[\mathcal{W}, L_\tau]\|^2_{\mathbf{L}^2(\mathbb{R}^m_+ \times \mathbb{R}^n)\to\mathbf{L}^2(\mathbb{R}^n)} \|f\|_2^2 \leq C' \left(\|D\tau\|_\infty \left(\log\frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty} \vee 1\right) + \|D^2\tau\|_\infty\right)^2 \|f\|_2^2$$

for some constant C'. We now choose $C_{n,m} = \max\{C', C\}$ to get the desired bound. □

2.4.2 Stability to Diffeomorphisms When $1 < q < 2$

Lemma 21. *Let $\gamma(z) = z - \tau(z)$, $g(z) = f(\gamma(z))$, and*

$$K_\lambda(x,z) = \det(D\gamma(z))\psi_\lambda(\gamma(x) - \gamma(z)) - \psi_\lambda(x - z).$$

Additionally, define

$$T_\lambda g(x) = \int_{\mathbb{R}^n} g(z)K_\lambda(x,z)\,dz$$

and consider $Tg : \mathbb{R}^n \to \mathbf{L}^2(\mathbb{R}_+, \frac{d\lambda}{\lambda^{n+1}})$ defined by $Tg(x) = (T_\lambda g(x))_{\lambda\in\mathbb{R}_+}$. Then for the Banach space $\mathcal{X} = \mathbf{L}^2(\mathbb{R}_+, \frac{d\lambda}{\lambda^{n+1}})$,

$$\|Tg\|^2_{\mathbf{L}^2_{\mathcal{X}}(\mathbb{R}^n)} \leq C_{n,m}\left(\|D\tau\|_\infty \left(\log\frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty} \vee 1\right) + \|D^2\tau\|_\infty\right)^2 \|f\|_2$$

for some constant $C_{n,m} > 0$.

Proof. Notice that

$$\begin{aligned}
&\|Tg\|^2_{\mathbf{L}^2_X(\mathbb{R}^n)} \\
&= \int_{\mathbb{R}^n}\int_0^\infty |T_\lambda g(x)|^2 \frac{d\lambda}{\lambda^{n+1}} dx \\
&= \int_{\mathbb{R}^n}\int_0^\infty \left|\int_{\mathbb{R}^n} K_\lambda(x,z) g(z)\, dz\right|^2 \frac{d\lambda}{\lambda^{n+1}} dx \\
&= \int_{\mathbb{R}^n}\int_0^\infty \left|\int_{\mathbb{R}^n} f(\gamma(z))[\det(D\gamma(z))\psi_\lambda(\gamma(x)-\gamma(z)) - \psi_\lambda(x-z)]\, dz\right|^2 \frac{d\lambda}{\lambda^{n+1}} dx \\
&= \int_{\mathbb{R}^n}\int_0^\infty \left|\int_{\mathbb{R}^n} \det(D\gamma(z)) f(\gamma(z))\psi_\lambda(\gamma(x)-\gamma(z))\, dz - \int_{\mathbb{R}^n} f(\gamma(z))\psi_\lambda(x-z)\, dz\right|^2 \frac{d\lambda}{\lambda^{n+1}} dx.
\end{aligned}$$

Using the change of variables $u = \gamma(z)$, we get

$$\begin{aligned}
\|Tg\|^2_{L^2_X(\mathbb{R}^n)} &= \int_{\mathbb{R}^n}\int_0^\infty |L_\tau(f*\psi_\lambda)(x) - (L_\tau f*\psi_\lambda)(x)|^2 \frac{d\lambda}{\lambda^{n+1}}\, dx \\
&= \int_{\mathbb{R}^n}\int_0^\infty |[\mathcal{W}_\lambda, L_\tau]f(x)|^2 \frac{d\lambda}{\lambda^{n+1}}\, dx \\
&= \int_0^\infty \int_{\mathbb{R}^n} |[\mathcal{W}_\lambda, L_\tau]f(x)|^2 \, dx \frac{d\lambda}{\lambda^{n+1}} \\
&= \int_0^\infty \|[\mathcal{W}_\lambda, L_\tau]f\|_2^2 \frac{d\lambda}{\lambda^{n+1}} \\
&= \|[\mathcal{W}, L_\tau]f\|^2_{\mathbf{L}^2(\mathbb{R}_+\times\mathbb{R}^n)} \\
&\le C_{n,m}\left(\|D\tau\|_\infty\left(\log\frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty}\vee 1\right) + \|D^2\tau\|_\infty\right)\|f\|_2^2,
\end{aligned}$$

where the last inequality follows from the $q = 2$ case. □

Lemma 22 ([39], Marcinkiewicz Interpolation)**.** *Let $\mathcal{A}$ and $\mathcal{B}$ be Banach spaces and let $T : \mathcal{A} \to \mathcal{B}$ be a quasilinear operator defined on $\mathbf{L}^{p_0}_{\mathcal{A}}(\mathbb{R}^n)$ and $\mathbf{L}^{p_1}_{\mathcal{A}}(\mathbb{R}^n)$ with $0 < p_0 < p_1$. Furthermore, if T satisfies*

$$\|Tf\|_{\mathbf{L}^{p_i,\infty}_{\mathcal{B}}(\mathbb{R}^n)} \le M_i \|f\|_{\mathbf{L}^{p_i}_{\mathcal{A}}(\mathbb{R}^n)}$$

for $i = 0, 1$, then for all $p \in (p_0, p_1)$,

$$\|Tf\|_{\mathbf{L}^{p}_{\mathcal{B}}(\mathbb{R}^n)} \le N_p \|f\|_{\mathbf{L}^{p}_{\mathcal{A}}(\mathbb{R}^n)},$$

where N_p only depends on M_0, M_1, and p.

Remark 7. Like with the scalar valued estimate, it can be shown that $N_p = \eta M_0^{\delta} M_1^{1-\delta}$, where

$$\delta = \begin{cases} \dfrac{p_0(p_1 - p)}{p(p_1 - p_0)} & p_1 < \infty, \\ \dfrac{p_0}{p} & p_1 = \infty \end{cases}$$

and

$$\eta = \begin{cases} 2\left(\dfrac{p(p_1 - p_0)}{(p - p_0)(p_1 - p)}\right)^{1/p} & p_1 < \infty, \\ 2\left(\dfrac{p_0}{p - p_0}\right)^{1/p} & p_1 = \infty. \end{cases}$$

Lemma 23. *Let T be the operator defined in Lemma 21. Let $q \in (1,2)$ and $r \in (1,q)$. Then T satisfies*

$$\|Tg\|_{\mathbf{L}_{\mathcal{X}}^{r,\infty}(\mathbb{R}^n)} \leq M_r \|f\|_{\mathbf{L}^r(\mathbb{R}^n)}$$

for some constant $M_r > 0$, which is independent of $\|D\tau\|_\infty$ and $\|D^2\tau\|_\infty$. Furthermore, T also satisfies

$$\|Tg\|^2_{\mathbf{L}_{\mathcal{X}}^{2,\infty}(\mathbb{R}^n)} \leq \tilde{C}_n \left(\|D\tau\|_\infty \left(\log \frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty} \vee 1\right) + \|D^2\tau\|_\infty\right)^2 \|f\|^2_{\mathbf{L}^2(\mathbb{R}^n)}$$

for some constant $\tilde{C}_n > 0$.

Proof. The second inequality obviously follows from strong boundedness of the operator, so we will omit the proof. For the first inequality, the norm satisfies

$$\begin{aligned} \|Tg(x)\|^2_{\mathcal{X}} &= \int_0^\infty \left| \int_{\mathbb{R}^n} \det(D\gamma(z)) f(\gamma(z)) \psi_\lambda(\gamma(x) - \gamma(z))\, dz - \int_{\mathbb{R}^n} f(\gamma(z)) \psi_\lambda(x - z)\, dz \right|^2 \frac{d\lambda}{\lambda^{n+1}} \\ &= \int_0^\infty \left| \int_{\mathbb{R}^n} f(z) \psi_\lambda(\gamma(x) - z)\, dz - \int_{\mathbb{R}^n} f(\gamma(z)) \psi_\lambda(x - z)\, dz \right|^2 \frac{d\lambda}{\lambda^{n+1}} \\ &\leq 4 \int_0^\infty \left| \int_{\mathbb{R}^n} f(z) \psi_\lambda(\gamma(x) - z)\, dz \right|^2 \frac{d\lambda}{\lambda^2} + 4 \int_0^\infty \left| \int_{\mathbb{R}^n} f(\gamma(z)) \psi_\lambda(x - z)\, dz \right|^2 \frac{d\lambda}{\lambda^{n+1}} \\ &= 4|(Gf)(\gamma(x))|^2 + 4|GL_\tau f(x)|^2. \end{aligned}$$

We see

$$\|Tg(x)\|_{\mathcal{X}} \leq \sqrt{4|(Gf)(\gamma(x))|^2 + 4|GL_\tau f(x)|^2} \leq 2|(Gf)(\gamma(x))| + 2|GL_\tau f(x)|.$$

For $\delta > 0$, Chebyshev's inequality implies that there exists A_r such that

$$\begin{aligned} m\{\|Tg(x)\|_{\mathcal{X}} > \delta\} &\leq m\{2|(Gf)(\gamma(x))| + 2|GL_\tau f(x)| > \delta\} \\ &\leq \frac{A_r}{\delta^r}(\|(Gf)(\gamma(\cdot))\|^r_{\mathbf{L}^r(\mathbb{R}^n)} + \|GL_\tau f\|^r_{\mathbf{L}^r(\mathbb{R}^n)}). \end{aligned}$$

We want to now ensure that $\|(Gf)(\gamma(\cdot))\|^r_{\mathbf{L}^r(\mathbb{R}^n)}$ can be bounded above by a constant multiple of $\|Gf\|^r_{\mathbf{L}^r(\mathbb{R}^n)}$. Since γ is a diffeomorphism, we can use change of variables to get

$$\begin{aligned} \|(Gf)(\gamma(\cdot))\|^r_{\mathbf{L}^r(\mathbb{R}^n)} &= \int_{\mathbb{R}^n} |Gf(\gamma(x))|^r \, dx \\ &= \int_{\mathbb{R}^n} |Gf(u)|^r \, \frac{du}{\det\left[(D\gamma)(\gamma^{-1}(u))\right]} \\ &\leq 2 \int_{\mathbb{R}^n} |Gf(x)|^r \, dx \\ &= 2\|Gf\|^r_{\mathbf{L}^r(\mathbb{R}^n)}. \end{aligned}$$

By Theorem 5, we get

$$\|GL_\tau f\|^r_{\mathbf{L}^r(\mathbb{R}^n)} \leq C_r \|L_\tau f\|^r_{\mathbf{L}^r(\mathbb{R}^n)} \leq 2C_r \|f\|^r_{\mathbf{L}^r(\mathbb{R}^n)}$$

for some constant C_r dependent on r. Thus, we have

$$m\{\|Tg(x)\|_{\mathcal{X}} > \delta\}^{1/r} \leq \frac{M_r}{\delta}\|f\|_{\mathbf{L}^r(\mathbb{R}^n)}$$

for some constant $M_r > 0$. □

Lemma 24. *Fix $r = \frac{1+q}{2}$ so that $r \in (1, q)$. For some constant $C_{n,q} > 0$, the operator T defined in Lemma 21 satisfies the estimate*

$$\|Tg\|^q_{\mathbf{L}^q_{\mathcal{X}}(\mathbb{R}^n)} \leq C_{n,q}\eta^q M_r^{q\delta} \left(\|D\tau\|_\infty \left(\log \frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty} \vee 1\right) + \|D^2\tau\|_\infty\right)^{q(1-\delta)} \|f\|^q_q,$$

where η and δ come from interpolation, and M_r comes from the constant for weak boundedness in Lemma 23.

Proof. Since T is an integral operator, it is clear that is quasilinear. Using the $\mathbf{L}^r(\mathbb{R}^n)$ and $\mathbf{L}^2(\mathbb{R}^n)$ estimates from the previous Lemma, we interpolate using Marcinkiewicz since $\|g\|_r \leq 2\|f\|_r \leq 4\|g\|_r$. □

Theorem 25. *Let $1 < q < 2$. Assume ψ and its first and second order derivatives have decay in $O((1+|x|)^{-n-3})$, and $\int_{\mathbb{R}^n} \psi(x)\,dx = 0$. Then for every $\tau \in C^2(\mathbb{R}^n)$ with $\|D\tau\|_\infty < \frac{1}{2n}$, there exists $C_{n,q} > 0$ such that*

$$\|S_{cont,q} f - S_{cont,q} L_\tau f\|^q_{\mathbf{L}^2(\mathbb{R}_+)} \leq C_{n,q} K_q(\tau) \|f\|_q^q$$

with

$$K_q(\tau) = \|D\tau\|_\infty^q + \eta^q M_r^{q\delta} \left(\|D\tau\|_\infty \left(\log \frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty} \vee 1 \right) + \|D^2\tau\|_\infty \right)^{q(1-\delta)}.$$

Proof. We use the same notation as Theorem 16. Using a nearly identical argument to Corollary 17, we get

$$\begin{aligned}
&\|S_{\mathrm{cont},q} f - S_{\mathrm{cont},q} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}_+)} \\
&= \|A_q M \mathcal{W} f - A_q M \mathcal{W} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}_+)} \\
&= \|A_q M \mathcal{W} f - A_q M L_\tau \mathcal{W} f + A_q M L_\tau W f - A_q M \mathcal{W} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}_+)} \\
&\leq \|A_q M \mathcal{W} f - A_q M L_\tau \mathcal{W} f\|_{\mathbf{L}^2(\mathbb{R}_+)} + \|A_q M L_\tau \mathcal{W} f - A_q M \mathcal{W} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}_+)} \\
&\leq \|(A_q M - A_q M L_\tau) \mathcal{W} f\|_{\mathbf{L}^2(\mathbb{R}_+)} + \|A_q M [\mathcal{W}, L_\tau] f\|_{\mathbf{L}^2(\mathbb{R}_+)}.
\end{aligned}$$

The first term, $\|(A_q M - A_q M L_\tau)\mathcal{W} f\|_{\mathbf{L}^2(\mathbb{R}_+)}$, can be bounded using an argument identical to the $q = 2$ case. In particular, we can prove that

$$(1 - n\|D\tau\|_\infty)\|f\|_q \leq (1 - n\|D\tau\|_\infty)^{1/q} \|f\|_q \leq \|L_\tau f\|_q$$

and

$$\|L_\tau f\|_q \leq (1 + 2n\|D\tau\|_\infty)^{1/q} \|f\|_q \leq (1 + 2n\|D\tau\|_\infty)\|f\|_q,$$

which means

$$\|(A_q M - A_q M L_\tau)\mathcal{W} f\|^q_{\mathbf{L}^2(\mathbb{R}_+)} \leq C \|D\tau\|_\infty^q \|f\|_q^q.$$

For the other term,

$$\|A_q M [\mathcal{W}, L_\tau] f\|^q_{\mathbf{L}^2(\mathbb{R}_+)} = \left(\int_0^\infty \left[\int_{\mathbb{R}^n} |(L_\tau f * \psi_\lambda)(x) - L_\tau (f * \psi_\lambda)(x)|^q \, dx \right]^{2/q} \frac{d\lambda}{\lambda^{n+1}} \right)^{q/2}.$$

Now, expand convolution and then use change of variables to get

$$
\begin{aligned}
&\|A_q M[\mathcal{W}, L_\tau] f\|^q_{\mathbf{L}^2(\mathbb{R}_+)} \\
&= \left(\int_0^\infty \left[\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} f(\gamma(z))(\det(D\gamma(z))\psi_\lambda(\gamma(x)-\gamma(z)) - \psi_\lambda(x-z))\, dz\right|^q \, dx\right]^{2/q} \frac{d\lambda}{\lambda^{n+1}}\right)^{q/2} \\
&= \left(\int_0^\infty \left[\int_{\mathbb{R}^n} \left|\int_{\mathbb{R}^n} g(z) K_\lambda(x,z)\, dz\right|^q \, dx\right]^{2/q} \frac{d\lambda}{\lambda^{n+1}}\right)^{q/2} \\
&= \left(\int_0^\infty \left[\int_{\mathbb{R}^n} |T_\lambda g(x)|^q \, dx\right]^{2/q} \frac{d\lambda}{\lambda^{n+1}}\right)^{q/2} \\
&\le \int_{\mathbb{R}^n} \left[\int_0^\infty |T_\lambda g(x)|^q \, \frac{d\lambda}{\lambda^{n+1}}\right]^{q/2} dx \\
&= \int_{\mathbb{R}^n} \left[\int_0^\infty |T_\lambda g(x)|^2 \, \frac{d\lambda}{\lambda^{n+1}}\right]^{q/2} dx \\
&= \int_{\mathbb{R}^n} \|Tg(x)\|^q_{\mathbf{L}^2\left(\mathbb{R}^+, \frac{d\lambda}{\lambda^{n+1}}\right)} dx \\
&= \|Tg\|^q_{\mathbf{L}^q_{\mathcal{X}}(\mathbb{R}^n)} \\
&\le C_n \eta^q M_r^{q\delta} \left(\|D\tau\|_\infty \left(\log \frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty} \vee 1\right) + \|D^2\tau\|_\infty\right)^{q(1-\delta)} \|f\|^q_q.
\end{aligned}
$$

Thus, the proof is complete. □

Corollary 26. *Let* $1 < q < 2$ *. Assume* ψ *and its first and second order derivatives have decay in* $O((1+|x|)^{-n-3})$*, and* $\int_{\mathbb{R}^n} \psi(x)\, dx = 0$*. Then for every* $\tau \in C^2(\mathbb{R}^n)$ *with* $\|D\tau\|_\infty < \frac{1}{2n}$*, there exist constants* $C_{n,m}, \hat{C}_{n,m} > 0$ *such that*

$$\|S^m_{cont,q} f - S^m_{cont,q} L_\tau f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)} \le C_{n,m} K_q(\tau) \|f\|^q_q$$

and

$$\|S^m_{dyad,q} f - S^m_{dyad,q} L_\tau f\|^q_{\ell^2(\mathbb{Z}^m)} \le \hat{C}_{n,m} K_q(\tau) \|f\|^q_q.$$

Remark 8. This bound is not exactly the same as the definition for stability to diffeomorphisms in [11], but the idea is similar. Since r is fixed, so is δ. It is easy to confirm that $\delta = \frac{1}{1+q} \in \left(\frac{1}{3}, \frac{1}{2}\right)$ when using Marcinkiewicz interpolation in Lemma 24, so

$$C_{n,q} \eta^q M_r^{q\delta} \left(\|D\tau\|_\infty \left(\log \frac{\|\Delta\tau\|_\infty}{\|D\tau\|_\infty} \vee 1\right) + \|D^2\tau\|_\infty\right)^{q(1-\delta)} \to 0$$

when $\|D\tau\|_\infty \to 0$ and $\|D^2\tau\|_\infty \to 0$.

2.5 Equivariance and Invariance to Rotations

We now consider adding group actions to our scattering transform and prove invariance to rotations. Let $\mathrm{SO}(n)$ be the group of $n \times n$ rotation matrices. Since $\mathrm{SO}(n)$ is a compact Lie group, we can define a Haar measure, say μ, with $\mu(\mathrm{SO}(n)) < \infty$. We say that $f \in \mathbf{L}^2(\mathrm{SO}(n))$ if and only if f is μ-measurable and $\int_{\mathrm{SO}(n)} |f(r)|^2 \, d\mu(r) < \infty$.

2.5.1 Rotation Equivariant Representations

Let $\psi : \mathbb{R}^n \to \mathbb{R}$ be a wavelet. Define

$$\psi_{\lambda,R}(x) = \lambda^{-n/2}\psi(\lambda^{-1}R^{-1}x),$$

where $R \in \mathrm{SO}(n)$ is a $n \times n$ rotation matrix. The continuous and dyadic wavelet transforms of f are given by

$$\mathcal{W}_{\mathrm{Rot}} f := \{f * \psi_{\lambda,R}(x) : x \in \mathbb{R}^n, \lambda \in (0,\infty), R \in \mathrm{SO}(n)\},$$

$$W_{\mathrm{Rot}} f := \{f * \psi_{j,R}(x) : x \in \mathbb{R}^n, j \in \mathbb{Z}, R \in \mathrm{SO}(n)\}.$$

We will first consider a translation invariant and rotation equivariant formulation of continuous and dyadic one-layer scattering using

$$\mathfrak{S}_{\mathrm{cont},q} f(\lambda, R) := \|f * \psi_{\lambda,R}\|_q,$$

$$\mathfrak{S}_{\mathrm{dyad},q} f(j, R) := \|f * \psi_{j,R}\|_q.$$

The translation invariance of our representation follows from translation invariance of the norm. For rotation equivariance, notice that if $f_{\tilde{R}}(x) := f(\tilde{R}^{-1}x)$, then we have

$$\mathfrak{S}_{\mathrm{cont},q} f_{\tilde{R}}(\lambda, R) = \mathfrak{S}_{\mathrm{cont},q} f(\lambda, \tilde{R}^{-1}R),$$

$$\mathfrak{S}_{\mathrm{dyad},q} f_{\tilde{R}}(j, R) = \mathfrak{S}_{\mathrm{dyad},q} f(j, \tilde{R}^{-1}R).$$

Now suppose we have m layers again. Then we define our m layer transforms by

$$\mathfrak{S}^m_{\mathrm{cont},q} f(\lambda_1, \ldots, \lambda_m, R_1, \ldots, R_m) := \||f * \psi_{\lambda_1,R_1}| * \ldots | * \psi_{\lambda_m,R_m}\|_q,$$

$$\mathfrak{S}^m_{\mathrm{dyad},q} f(j_1, \ldots, j_m, R_1, \ldots, R_m) := \||f * \psi_{j_1,R_1}| * \ldots | * \psi_{j_m,R_m}\|_q.$$

and rotation equivariance implies

$$\mathfrak{S}^m_{\text{cont},q} f_{\tilde{R}}(\lambda_1, \dots, \lambda_m, R_1, \dots, R_m) = \mathfrak{S}^m_{\text{cont},q} f(\lambda_1, \dots, \lambda_m, \tilde{R}^{-1}R_1, \dots, \tilde{R}^{-1}R_m),$$
$$\mathfrak{S}^m_{\text{dyad},q} f_{\tilde{R}}(j_1, \dots, j_m, R_1, \dots, R_m) = \mathfrak{S}^m_{\text{dyad},q} f(j_1, \dots, j_m, \tilde{R}^{-1}R_1, \dots, \tilde{R}^{-1}R_m).$$

The norm we will use is similar to our previous formulations. Denote the scattering norm for the continuous transform as $\|\mathfrak{S}^m_{\text{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)\times\text{SO}(n)^m}$, which is defined as

$$\left(\int_0^\infty \int_{\text{SO}(n)} \cdots \int_0^\infty \int_{\text{SO}(n)} \| |f * \psi_{j_1,R_1}| * \dots | * \psi_{j_m,R_m}\|^2_q d\mu_1(R_1)\, \frac{d\lambda_1}{\lambda_1^{n+1}} \dots d\mu_m(R_n)\, \frac{d\lambda_m}{\lambda_m^{n+1}}\right)^{q/2}.$$

For the dyadic transform, we denote the norm using $\|\mathfrak{S}^m_{\text{dyad},q} f\|^q_{\ell^2(\mathbb{Z}^m)\times\text{SO}(n)^m}$, which is given by

$$\left(\sum_{j_m\in\mathbb{Z}} \int_{\text{SO}(n)} \cdots \sum_{j_1\in\mathbb{Z}} \int_{\text{SO}(n)} \| |f * \psi_{j_1,R_1}| * \dots | * \psi_{j_m,R_m}\|^2_q d\mu_1(R_1)\ \dots d\mu_m(R_n)\right)^{q/2}.$$

We will start by proving that these formulations of the scattering transform are well defined, and prove properties about stability to diffeomorphisms like in previous chapters.

Lemma 27. *Let ψ be a wavelet that satisfies properties* (2.4) *and* (2.5).

- *If $1 < q \leq 2$, we have $\mathfrak{S}^m_{cont,q} : \mathbf{L}^q(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^m_+) \times SO(n)^m$ and $\mathfrak{S}^m_{dyad,q} : \mathbf{L}^q(\mathbb{R}^n) \to \ell^2(\mathbb{Z}^m) \times SO(n)^m$.*
- *If $q = 1$ and one of the following holds:*
  - *$n = 1$ and ψ is complex analytic,*
  - *$n \geq 2$ and ψ satisfies the conditions of Lemma 9,*

  then $\mathfrak{S}^m_{cont,1} : \mathbf{L}^1(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^m_+) \times SO(n)^m$ and $\mathfrak{S}^m_{dyad,1} : \mathbf{L}^1(\mathbb{R}^n) \to \ell^2(\mathbb{Z}^m) \times SO(n)^m$.
- *If ψ is also a Littlewood-Paley wavelet, we have*

$$\|\mathfrak{S}^m_{cont,2} f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)\times SO(n)^m} = \mu(SO(n))^m C^m_\psi \|f\|^2_2,$$
$$\|\mathfrak{S}^m_{dyad,q} f\|^2_{\ell^2(\mathbb{Z}^m)\times SO(n)^m} = \mu(SO(n))^m \hat{C}^m_\psi \|f\|^2_2.$$

Proof. We prove the first and third claim. The second claim is almost identical to the first claim, so the proof will be omitted for brevity. Note that we will only provide arguments for the continuous scattering transform since the proofs for the dyadic transform are very similar. By Fubini Theorem and boundedness of the m-layer scattering transform, there exists a constant $C_q > 0$, which is dependent on q, such that

$$
\begin{aligned}
&\|\mathfrak{S}^m_{\mathrm{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)\times \mathrm{SO}(n)^m} \\
&= \left[\int_0^\infty \int_{\mathrm{SO}(n)} \cdots \int_0^\infty \int_{\mathrm{SO}(n)} \| |f * \psi_{\lambda_1,R_1}| * \ldots | * \psi_{\lambda_m,R_m}\|_q^2 d\mu(R_m)\, \frac{d\lambda_1}{\lambda_1^{n+1}} \cdots d\mu(R_1)\, \frac{d\lambda_m}{\lambda_m^{n+1}}\right]^{q/2} \\
&\le \left[\int_{\mathrm{SO}(n)} \cdots \int_{\mathrm{SO}(n)} (C_q^{mq}\|f\|_q^q)^{2/q}\, d\mu(R_1)\cdots d\mu(R_m)\right]^{q/2} \\
&= C_q^{mq}\mu(\mathrm{SO}(n))^{mq/2}\|f\|_q^q
\end{aligned}
$$

because each ψ_{λ_i,R_i} is still a wavelet with sufficient decay even if the rotation is applied. For the third claim, we see that

$$
\begin{aligned}
&\|\mathfrak{S}^m_{\mathrm{cont},2} f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)\times \mathrm{SO}(n)^m} \\
&= \int_{\mathrm{SO}(n)} \cdots \int_{\mathrm{SO}(n)} C_\psi^m \|f\|_2^2\, d\mu(R_1)\cdots d\mu(R_m) \\
&= \mu(\mathrm{SO}(n))^m C_\psi^m \|f\|_2^2.
\end{aligned}
$$

□

Theorem 28. *Assume $|c| < \frac{1}{2n}$. Let $\tau(x) = cx$ and $L_\tau f(x) = f((1-c)x)$. Suppose that ψ is a wavelet that satisfies the conditions of Lemma 14. Then there exist constants $\tilde{K}_{n,m,q}$ and $\tilde{K}'_{n,m,q}$ dependent only on n, m, and q such that*

$$\|\mathfrak{S}^m_{cont,q} f - \mathfrak{S}^m_{cont,q} L_\tau f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)\times SO(n)^m} \le |c|^q \cdot \tilde{K}_{n,m,q}\|f\|_q^q$$

and

$$\|\mathfrak{S}^m_{dyad,q} f - \mathfrak{S}^m_{dyad,q} L_\tau f\|^q_{\ell^2(\mathbb{Z}^m)\times SO(n)^m} \le |c|^q \cdot \tilde{K}'_{n,m,q}\|f\|_q^q.$$

Alternatively, if one of the following holds:

- *$n = 1$, ψ is complex analytic and satisfies the conditions of Lemma 14,*
- *$n \ge 2$ and ψ satisfies the conditions of Lemma 9,*

there exist $\tilde{H}_{m,n}$ and $\tilde{H}'_{m,n}$ such that

$$\|\mathfrak{S}^m_{cont,1}f - \mathfrak{S}^m_{cont,1}L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)\times SO(n)^m} \leq |c| \cdot \tilde{H}_{m,n}\|f\|_{\mathbb{H}^1(\mathbb{R}^n)}.$$

and

$$\|\mathfrak{S}^m_{dyad,1}f - \mathfrak{S}^m_{dyad,1}L_\tau f\|_{\ell^2(\mathbb{Z}^m)\times SO(n)^m} \leq |c| \cdot \tilde{H}'_{m,n}\|f\|_{\mathbb{H}^1(\mathbb{R}^n)}$$

Theorem 29. *Let $\tau \in C^2(\mathbb{R}^n)$, and let $L_\tau f(x) = f(x - \tau(x))$. Suppose that ψ is a wavelet such that the wavelet and all its first and second partial derivatives have $O((1+|x|)^{-n-3})$ decay. When $q \in (1,2)$, there exists a constant $C_{n,m,q}$ dependent on $\mu(SO(n))$, n, m, and q such that*

$$\|\mathfrak{S}^m_{cont,q}f - \mathfrak{S}^m_{cont,q}L_\tau f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)\times SO(n)^m} \leq C_{n,m,q}K_q(\tau)\|f\|^q_q,$$
$$\|\mathfrak{S}^m_{dyad,q}f - \mathfrak{S}^m_{dyad,q}L_\tau f\|^q_{\ell^2(\mathbb{Z}^m)\times SO(n)^m} \leq \tilde{C}_{n,m,q}K_q(\tau)\|f\|^q_q,$$
$$\|\mathfrak{S}^m_{cont,2}f - \mathfrak{S}^m_{cont,2}L_\tau f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)\times SO(n)^m} \leq C_{n,m}K_2(\tau)\|f\|^2_2,$$
$$\|\mathfrak{S}^m_{cont,2}f - \mathfrak{S}^m_{cont,2}L_\tau f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)\times SO(n)^m} \leq C_{n,m}K_2(\tau)\|f\|^2_2.$$

2.5.2 Rotation Invariant Representations

The representation before was rotation equivariant, but in some tasks, we would rather have rotation invariance. In [11], the authors choose to integrate over each group action in a group of transformations. However, this will remove the information the relative angles between each action if we have multiple layers in our transform.

In the case of one layer, since there is only one angle, we use a similar formulation to [11] and define continuous and dyadic scattering transforms for rotation invariance as

$$\mathcal{S}_{\text{cont},q}f(\lambda) = \int_{\text{SO}(n)} \|f * \psi_{\lambda,R}\|^q_{\mathbf{L}^q(\mathbb{R}^n)} d\mu(R),$$
$$\mathcal{S}_{\text{dyad},q}f(j) = \int_{\text{SO}(n)} \|f * \psi_{j,R}\|^q_{\mathbf{L}^q(\mathbb{R}^n)} d\mu(R).$$

The corresponding norms are given by

$$\|\mathcal{S}_{\text{cont},q}f\|^q_{\mathbf{L}^2(\mathbb{R}_+)} := \left[\int_0^\infty \left[\int_{\text{SO}(n)} \|f * \psi_{\lambda,R}\|_q \mu(R)\right]^{2/q} \frac{d\lambda}{\lambda^{n+1}}\right]^{q/2},$$
$$\|\mathcal{S}_{\text{dyad},q}f\|^q_{\ell^2(\mathbb{Z})} := \left[\sum_{j\in\mathbb{Z}} \left[\int_{\text{SO}(n)} \|f * \psi_{j,R}\|_q \mu(R)\right]^{2/q}\right]^{q/2}.$$

Now we generalize to the case where $m \geq 2$. Let $R_1, \ldots, R_m \in \mathrm{SO}(n)$. Define

$$\mathcal{S}^m_{\mathrm{cont},q} f(\lambda_1, \ldots, \lambda_m, R_2, \ldots, R_m) := \int_{\mathrm{SO}(n)} \| |f * \psi_{\lambda_1, R_2 R_1}| * \cdots * |\psi_{\lambda_m, R_m R_1}\|_q^2 \, d\mu(R_1),$$
$$\mathcal{S}^m_{\mathrm{dyad},q} f(j_1, \ldots, j_m, R_2, \ldots, R_m) := \int_{\mathrm{SO}(n)} \| |f * \psi_{j_1, R_2 R_1}| * \ldots | * \psi_{j_m, R_m R_1}\|_q^2 \, d\mu(R_1).$$

The norm for the continuous transform, the norm $\|\mathcal{S}^m_{\mathrm{cont},q} f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+) \times \mathrm{SO}(n)^{m-1}}$, is given by

$$\left(\int_0^\infty \int_{\mathrm{SO}(n)} \cdots \int_0^\infty \int_{\mathrm{SO}(n)} \int_0^\infty \mathcal{S}^m_{\mathrm{cont},q} f \, \frac{d\lambda_1}{\lambda_1^{n+1}} \, d\mu_2(R_2) \, \frac{d\lambda_2}{\lambda_2^{n+1}} \ldots d\mu_m(R_m) \, \frac{d\lambda_m}{\lambda_m^{n+1}} \right)^{q/2},$$

where we use the shorthand notation

$$\mathcal{S}^m_{\mathrm{cont},q} f := \mathcal{S}^m_{\mathrm{cont},q} f(\lambda_1, \ldots, \lambda_m, R_2, \ldots, R_m)$$

and

$$\mathcal{S}^m_{\mathrm{dyad},q} f := \mathcal{S}^m_{\mathrm{dyad},q} f(\lambda_1, \ldots, \lambda_m, R_2, \ldots, R_m)$$

for brevity.

For the dyadic transform, the norm $\|\mathcal{S}^m_{\mathrm{dyad},q} f\|^q_{\ell^2(\mathbb{Z}) \times \mathrm{SO}(n)^{m-1}}$ is given by

$$\left(\sum_{j_m \in \mathbb{Z}} \int_{\mathrm{SO}(n)} \cdots \sum_{j_2 \in \mathbb{Z}} \int_{\mathrm{SO}(n)} \sum_{j_1 \in \mathbb{Z}} \mathcal{S}^m_{\mathrm{dyad},q} f \, d\mu_1(R_1) \, d\mu_2(R_2) \, \ldots d\mu_m(R_m) \right)^{q/2}.$$

Like before, we will discuss the well-definedness and stability of these operators to diffeomorphisms. The proofs will be omitted since they follow directly from the previous sections with minor modifications.

Lemma 30. *Let ψ be a wavelet that satisfies properties* (2.4) *and* (2.5).

- *If $1 < q \leq 2$, we have $\mathcal{S}^m_{cont,q} : \mathbf{L}^q(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^m_+) \times SO(n)^{m-1}$ and $\mathcal{S}^m_{dyad,q} : \mathbf{L}^q(\mathbb{R}^n) \to \ell^2(\mathbb{Z}^m) \times SO(n)^{m-1}$.*
- *If $q = 1$ and one of the following holds:*
  - *$n = 1$ and ψ is complex analytic,*
  - *$n \geq 2$ and ψ satisfies the conditions of Lemma 9,*

  then $\mathcal{S}^m_{cont,1} : \mathbf{L}^1(\mathbb{R}^n) \to \mathbf{L}^2(\mathbb{R}^m_+) \times SO(n)^{m-1}$ and $\mathcal{S}^m_{dyad,1} : \mathbf{L}^1(\mathbb{R}^n) \to \ell^2(\mathbb{Z}^m) \times SO(n)^{m-1}$.

- *If $q = 2$ and ψ is also a littlewood paley wavelet, we have* $\|\mathcal{S}^m_{dyad,2} f\|_{\ell^1(\mathbb{Z}^m)\times SO(n)^{m-1}} = \mu(SO(n))^{m-1} C^m_\psi \|f\|_2^2$ *and* $\|\mathcal{S}^m_{cont,2} f\|_{\mathbf{L}^1(\mathbb{R}^m_+)\times SO(n)^{m-1}} = \mu(SO(n))^{m-1} \hat{C}^m_\psi \|f\|_2^2$.

Theorem 31. *Assume $|c| < \frac{1}{2n}$ and $1 < q < 2$. Let $\tau(x) = cx$ and let $L_\tau f(x) = f((1-c)x)$. Suppose that ψ is a wavelet that satisfies the conditions of Lemma 14. Then there exist constants $\hat{K}_{n,m,q}$ and $\hat{K}'_{n,m,q}$ dependent only on n, m, and q such that*

$$\|\mathcal{S}^m_{cont,q} f - \mathcal{S}^m_{cont,q} L_\tau f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)\times SO(n)^{m-1}} \leq |c|^q \cdot \hat{K}_{n,m,q} \|f\|_q^q$$

and

$$\|\mathcal{S}^m_{dyad,q} f - \mathcal{S}^m_{dyad,q} L_\tau f\|^q_{\ell^2(\mathbb{Z}^m)\times SO(n)^{m-1}} \leq |c|^q \cdot \hat{K}'_{n,m,q} \|f\|_q^q.$$

Additionally, if $q = 1$ and one of the following holds:

- *$n = 1$, ψ is complex analytic and satisfies the conditions of Lemma 14,*
- *$n \geq 2$ and ψ satisfies the conditions of Lemma 9,*

there exist $\hat{H}_{m,n}$ and $\hat{H}'_{m,n}$ such that

$$\|\mathcal{S}^m_{cont,1} f - \mathcal{S}^m_{cont,1} L_\tau f\|_{\mathbf{L}^2(\mathbb{R}^m_+)\times SO(n)^{m-1}} \leq |c| \cdot \hat{H}_{m,n} \|f\|_{\mathbb{H}^1(\mathbb{R}^n)}$$

and

$$\|\mathcal{S}^m_{dyad,1} f - \mathcal{S}^m_{dyad,1} L_\tau f\|_{\ell^2(\mathbb{Z}^m)\times SO(n)^{m-1}} \leq |c| \cdot \hat{H}'_{m,n} \|f\|_{\mathbb{H}^1(\mathbb{R}^n)}.$$

Theorem 32. *Let $\tau \in C^2(\mathbb{R}^n)$ and define $L_\tau f(x) = f(x - \tau(x))$ with $\|D\tau\|_\infty < \frac{1}{2n}$. Suppose that ψ is a wavelet such that the wavelet and all its first and second partial derivatives have $O((1+|x|)^{-n-3})$ decay. For $q \in (1, 2]$, there exist constants $C_{m,n}$, $\hat{C}_{m,n}$, $C_{m,n,q}$, and $\hat{C}_{m,n,q}$ such that*

$$\|\mathcal{S}^m_{cont,2} f - \mathcal{S}^m_{cont,2} L_\tau f\|^2_{\mathbf{L}^2(\mathbb{R}^m_+)\times SO(n)^{m-1}} \leq C_{m,n}\, K_2(\tau) \|f\|_2^2,$$
$$\|\mathcal{S}^m_{dyad,2} f - \mathcal{S}^m_{dyad,2} L_\tau f\|^2_{\ell^2(\mathbb{Z}^m)\times SO(n)^{m-1}} \leq \hat{C}_{m,n}\, K_2(\tau) \|f\|_2^2,$$
$$\|\mathcal{S}_{cont,q} f - \mathcal{S}^m_{cont,q} L_\tau f\|^q_{\mathbf{L}^2(\mathbb{R}^m_+)\times SO(n)^{m-1}} \leq C_{m,n,q} K_q(\tau) \|f\|_q^q,$$
$$\|\mathcal{S}^m_{dyad,q} f - \mathcal{S}^m_{dyad,q} L_\tau f\|^q_{\ell^2(\mathbb{Z}^m)\times SO(n)^{m-1}} \leq \hat{C}_{m,n,q} K_q(\tau) \|f\|_q^q$$

CHAPTER 3

EXPECTED SCATTERING TRANSFORMS

3.1 Background

Generalizing to stochastic processes, one can also consider scattering moments [11, 15], which have similar desirable properties as the nonwindowed scattering transform; other tangential works include [43, 44]. For the modeling of objects such as audio and image textures, one can think of them as realizations of highly non-Gaussian processes [15].

In the particular case of audio/image synthesis in particular, one would like generate a texture with the same statistical properties without generating a repetition of the texture. Equivariant features are more likely to lead to repetitions in textures. Thus, it is sensible to get a small number of rich descriptors that are translation invariant (e.g. using a realization of a process and calculating the nonwindowed scattering transform). In practice, instead of calculating an expectation, one takes an average of multiple realizations. Applications further applications include cosmology [45]. The main idea in all these applications is that the nonwindowed scattering transform has desirable mathematical properties and provides a small number of relevant descriptors for high dimensional, complicated data.

3.2 Wavelet Transforms for Stochastic Processes

Let X be a real valued stationary stochastic process with finite second moment. Also, let ψ be a wavelet. As a reminder, let G be a finite rotation group, and G^+ be the quotient of G with the set $\{-1, 1\}$, and let

$$\Lambda = \{(2^j, r) \ : \ j \in \mathbb{Z}, r \in G^+\}.$$

For all $\lambda \in \Lambda$, dilations of the wavelet are given by

$$\psi_\lambda(u) = 2^{-nj}\psi(2^{-j}r^{-1}u), \tag{3.1}$$

and we define the wavelet transform of X at scale 2^j as

$$X * \psi_j(t) = \int_{\mathbb{R}^n} X(u)\psi_j(t-u)\, du. \tag{3.2}$$

The dyadic wavelet transform is given by

$$WX = \{X * \psi_\lambda\}_{\lambda\in\Lambda}. \tag{3.3}$$

We say that is ψ a littlewood paley wavelet if ψ satisfies the following admissibility condition:

$$\sum_{\lambda\in\Lambda} |\hat{\psi}_\lambda(\omega)|^2 = \sum_{r\in G^+}\sum_{j\in\mathbb{Z}} |\hat{\psi}(2^j r^{-1}\omega)|^2 = C_\psi, \qquad \forall\omega \neq 0. \tag{3.4}$$

For any littlewood paley wavelet, we have the following relation between the variance $\sigma^2(X)$ and the energy of the wavelet transform:

$$\sum_{\lambda\in\Lambda} \mathbb{E}[|X * \psi_\lambda|^2] = \beta C_\psi\, \sigma^2(X), \tag{3.5}$$

where

$$\beta = \begin{cases} 1/2 & \text{if } \psi \text{ is real valued,} \\ 1 & \text{if } \psi \text{ is complex valued.} \end{cases}$$

3.3 Scattering Moments and the Expected Scattering Transform

Following [11], first order scattering moments are defined as

$$S_1X(\lambda) = \mathbb{E}\left[|X * \psi_\lambda|\right], \qquad \forall\lambda \in \Lambda. \tag{3.6}$$

Scattering moments for $m > 1$ are an iterative application of a wavelet transform followed by a modulus, which is given by:

$$S_1^m X(\lambda_1, \ldots, \lambda_m) = \mathbb{E}\left[||X * \psi_{\lambda_1}| * \cdots | * \psi_{\lambda_m}|\right], \qquad \forall(\lambda_1, \ldots, \lambda_m) \in \Lambda^m. \tag{3.7}$$

The expected scattering transform is the set of all scattering moments:

$$\overline{S}_1 X = \{S_1^m X(\lambda_1, \ldots, \lambda_m) \,:\, \forall(\lambda_1, \ldots, \lambda_m) \in \Lambda^m, \forall m \in \mathbb{N}\} \tag{3.8}$$

with norm

$$\|\overline{S}_1 X\|^2 = \sum_{m=1}^{\infty} \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} |S_1^m X(\lambda_1, \ldots, \lambda_m)|^2. \tag{3.9}$$

Additionally, suppose that Y is also a stochastic process with finite second moment. The scattering distance is given by

$$\|\overline{S}_1 X - \overline{S}_1 Y\|^2 = \sum_{m=1}^{\infty} \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} |S_1^m X(\lambda_1, \ldots, \lambda_m) - S_1^m Y(\lambda_1, \ldots, \lambda_m)|^2 \tag{3.10}$$

3.4 The Expected Scattering Transform When $q = 2$

Generalizing the norms above, we begin by defining the expected scattering transform and scattering norm when $q = 2$. The expected scattering transform is the set of all scattering moments:

$$\overline{S}_2 X = \{S_2^m X(\lambda_1, \ldots, \lambda_m) \, : \, \forall (\lambda_1, \ldots, \lambda_m) \in \Lambda^m, \, \forall m \in \mathbb{N}\} \tag{3.11}$$

with norm

$$\|\overline{S}_2 X\|_2^2 = \sum_{m=1}^{\infty} \sum_{(\lambda_1, \ldots, \lambda_m) \in \Lambda^m} |S_2^m X(\lambda_1, \ldots, \lambda_m)|^2 \tag{3.12}$$

and scattering distance

$$\|\overline{S}_2 X - \overline{S}_2 Y\|_2^2 = \sum_{m=1}^{\infty} \sum_{(\lambda_1, \ldots, \lambda_m) \in \Lambda^m} |S_2^m X(\lambda_1, \ldots, \lambda_m) - S_2^m Y(\lambda_1, \ldots, \lambda_m)|^2 \tag{3.13}$$

3.4.1 General Properties

Lemma 33. *Suppose ψ is a littlewood paley wavelet. Then we have the following bound:*

$$\sum_{(\lambda_1, \ldots, \lambda_m) \in \Lambda^m} |S_2^m X(\lambda_1, \ldots, \lambda_m)|^2 = \beta^m C_\psi^m \sigma^2(X) \leq \beta^m C_\psi^m \mathbb{E}[X^2].$$

Proof. Without a loss of generality, assume that ψ is complex and remove β from all the proofs. We proceed by induction. The base case follows directly from (3.5) since Thus, we have

$$\|S_2 X\|_{\ell^2(\mathbb{Z})}^2 = C_\psi \sigma^2(X) \leq C_\psi \mathbb{E}[X^2].$$

Now assume that for some $k \in \mathbb{N}$,

$$\sum_{(\lambda_1, \ldots, \lambda_k) \in \Lambda^k} |S_2^k X(\lambda_1, \ldots, \lambda_k)|^2 = C_\psi^k \sigma^2(X) \leq C_\psi^k \mathbb{E}[X^2].$$

Define the random variable $Y_k = ||X * \psi_{\lambda_1}| * \cdots | * \psi_{\lambda_k}|$, which is clearly stationary since the the modulus operator and wavelet transform both preserve stationarity of a stochastic process. It follows that we can write

$$\begin{aligned}
\sum_{(\lambda_1,\ldots,\lambda_{k+1})\in\Lambda^{k+1}} |S_2^{k+1}X(\lambda_1,\ldots,\lambda_{k+1})|^2 &= \sum_{(\lambda_1,\ldots,\lambda_k)\in\Lambda^k} \sum_{\lambda_{k+1}\in\Lambda} \mathbb{E}[|Y_k * \psi_{j_{k+1}}|]^2 \\
&= C_\psi \sum_{(j_1,\ldots,j_k)\in\mathbb{Z}^k} \sigma^2(Y_k) \\
&\leq C_\psi \sum_{(j_1,\ldots,j_k)\in\mathbb{Z}^k} \mathbb{E}[Y_k^2] \\
&\leq C_\psi^{k+1}\mathbb{E}[X^2].
\end{aligned}$$

□

We first begin by proving that our expected scattering transform with $q = 2$ is a nonexpansive operator.

Theorem 34 (Nonexpansive Operator)**.** *Suppose ψ is a littlewood paley wavelet with $\beta C_\psi \leq \frac{1}{2}$. Then $\|\overline{S}_2X - \overline{S}_2Y\|_2^2 \leq \mathbb{E}[|X-Y|^2]$ and $\|\overline{S}_2X\|_2^2 \leq \mathbb{E}[X^2]$.*

Proof. For notational simplicity, define

$$X_k = ||X * \psi_{\lambda_1}| * \cdots | * \psi_{\lambda_k}|,$$

$$Y_k = ||Y * \psi_{\lambda_1}| * \cdots | * \psi_{\lambda_k}|.$$

We begin by applying Minkowski's inequality and (3.5) repeatedly to get

$$
\begin{aligned}
&\sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} |S_2^m X(\lambda_1,\ldots,\lambda_m) - S_2^m Y(\lambda_1,\ldots,\lambda_m)|^2 \\
&= \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} \left| \mathbb{E}\left[|X_{m-1} * \psi_{\lambda_m}|^2\right]^{1/2} - \mathbb{E}\left[|Y_{m-1} * \psi_{\lambda_m}|^2\right]^{1/2} \right|^2 \\
&\leq \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} \mathbb{E}\left[|(X_{m-1} - Y_{m-1}) * \psi_{\lambda_m}|^2\right] \\
&\leq C_\psi \sum_{(\lambda_1,\ldots,\lambda_{m-1})\in\Lambda^{m-1}} \mathbb{E}\left[|X_{m-1} - Y_{m-1}|^2\right] \\
&\leq C_\psi \sum_{(\lambda_1,\ldots,\lambda_{m-1})\in\Lambda^{m-1}} \mathbb{E}\left[|(X_{m-2} - Y_{m-2}) * \psi_{\lambda_{m-1}}|^2\right] \\
&\leq C_\psi^2 \sum_{(\lambda_1,\ldots,\lambda_{m-2})\in\Lambda^{m-2}} \mathbb{E}\left[|X_{m-2} - Y_{m-2}|^2\right] \\
&\vdots \\
&\leq C_\psi^m \mathbb{E}[|X - Y|^2]
\end{aligned}
$$

Now sum over all m to get

$$
\|\overline{S}_2 X - \overline{S}_2 X\|_2^2 \leq \sum_{m=1}^{\infty} C_\psi^m \mathbb{E}[|X - Y|^2] = \frac{C_\psi}{1 - C_\psi} \mathbb{E}[|X - Y|^2] \leq \mathbb{E}[|X - Y|^2].
$$

Setting $Y = 0$ proves $\|\overline{S}_2 X\|_2^2 \leq \mathbb{E}[X^2]$, which completes the proof. □

3.4.2 Diffeomorphism Contraction Estimates

Let τ be a stationary random process independent of X such that $\|D\tau\|_\infty \leq \frac{1}{2n}$ with probability 1. Define the deformed process $L_\tau X(x) = X(x - \tau(x))$, which is still stationary. We will need the following lemma.

Lemma 35 ([11], Lemma 4.8)**.** *Let K_τ be an integral operator with a kernel $k_\tau(x, u)$ which depends upon a random process τ. If the following two conditions are satisfied:*

$$
\mathbb{E}\left[k_\tau(x, u) k_\tau^*(x, u')\right] = \overline{k}_\tau(x - u, x - u')
$$

and

$$
\int_{\mathbb{R}^\varkappa} \int_{\mathbb{R}^\varkappa} |\overline{k}_\tau(v, v')||v - v'| \, dv \, dv' < \infty,
$$

then for any stationary process Y independent of τ, $\mathbb{E}[|K_\tau Y(x)|^2]$ does not depend on x and

$$\mathbb{E}[|K_\tau Y|^2] \leq \mathbb{E}[\|K_\tau\|^2]\mathbb{E}[|Y|^2],$$

where $\|K_\tau\|$ is the operator norm in $\mathbf{L}^2(\mathbb{R}^n)$ for each realization of τ.

Theorem 36 (Diffeomorphism Contraction Estimate)**.** *Consider the random process $X - L_\tau X$. Assume that $\beta C_\psi < 1/2$ and $\hat{R}_{X-L_\tau X}$, the Fourier Transform of the covariance function, is bandlimited. We have the following estimate for some $C > 0$:*

$$\|\overline{S}_2 X - \overline{S}_2 L_\tau X\|_2^2 \leq (CM^2\mathbb{E}[\|\tau\|_\infty^2])\mathbb{E}[|X|^2].$$

Proof. Let ϕ be a function such that

$$\hat{\phi}(\omega) = \begin{cases} 1, & \omega \in B_1(0), \\ 0, & \omega \notin B_1(0). \end{cases}$$

Define $\phi_M(x) = M^{-n}\phi(Mx)$. Then we also know that $\int_{\mathbb{R}^n} \phi_M(x)\, dx = 1$.

Since our scattering operator is nonexpansive, we have

$$\|\overline{S}_2 X - \overline{S}_2 L_\tau X\|_2^2 \leq \mathbb{E}[|X - L_\tau X|^2],$$

where the expectation is over all possible randomness. Notice that since $\int_{\mathbb{R}^n} \phi_M(x)\, dx = 1$, we can write

$$\begin{aligned}
\mathbb{E}[|(X - L_\tau X) * \phi_M|^2] &= \int_{\mathbb{R}^n} \hat{R}_{X-L_\tau X}(\omega)|\phi_M(\omega)|^2\, d\omega + \mathbb{E}^2[(X - L_\tau X) * \phi_M] \\
&= \int_{\mathbb{R}^n} \hat{R}_{X-L_\tau X}(\omega)|\phi_M(\omega)|^2\, d\omega + \mathbb{E}^2[X - L_\tau X] \\
&= \int_{B_M(0)} \hat{R}_{X-L_\tau X}(\omega)\, d\omega + \mathbb{E}^2[X - L_\tau X] \\
&= \mathbb{E}[|X - L_\tau X|^2].
\end{aligned}$$

In other words, if we define $A_{\phi_R} f := f * \phi_R$, we can write

$$\mathbb{E}[|(X - L_\tau X) * \phi_R|^2] = \mathbb{E}[|(A_{\phi_R} - A_{\phi_R} L_\tau) f|^2].$$

From estimates given in Theorem 3.6 of [23], in the deterministic case with $f \in \mathbf{L}^2(\mathbb{R}^n)$ we have

$$\|(A_{\phi_R} - A_{\phi_R} L_\tau) f\|_2^2 \leq 4R^2 \|\nabla\phi\|_1^2 \|\tau\|_\infty^2 \|f\|_2^2,$$

where $\tau \in C^1(\mathbb{R}^n)$. It is proven in Appendix H of [11] that a operator of the form $A_{\phi_R} - A_{\phi_R} L_\tau$ has a kernel that satisfies Lemma 35. Thus, we have

$$\mathbb{E}[|X - L_\tau X|^2] \leq 4R^2 \|\nabla\phi\|_1^2 \mathbb{E}[\|\tau\|_\infty^2] \mathbb{E}[|X|^2].$$

□

3.5 The Expected Scattering Transform When $1 < q < 2$

Now we generalize to the case where $q \in (1, 2)$. The case of $q = 1$ has been addressed in [11]. The expected scattering transform is the set of all scattering moments:

$$\overline{S}_q X = \{S_q^m X(\lambda_1, \ldots, \lambda_m) \, : \, \forall (\lambda_1, \ldots, \lambda_m) \in \Lambda^m, \, \forall m \in \mathbb{N}\} \tag{3.14}$$

with norm

$$\|\overline{S}_q X\|_2^2 = \sum_{m=1}^{\infty} \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} |S_q^m X(\lambda_1, \ldots, \lambda_m)|^2 \tag{3.15}$$

and scattering distance

$$\|\overline{S}_q X - \overline{S}_q Y\|_2^2 = \sum_{m=1}^{\infty} \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} |S_q^m X(\lambda_1, \ldots, \lambda_m) - S_q^m Y(\lambda_1, \ldots, \lambda_m)|^2. \tag{3.16}$$

We start with a lemma that will help us determine when our generalized expected scattering transform is well defined.

Lemma 37. *Suppose ψ is a littlewood paley wavelet. Then we have the following bound:*

$$\sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} |S_q^m X(\lambda_1, \ldots, \lambda_m)|^2 \leq \beta^m C_\psi^m \sigma^2(X) \leq \beta^m C_\psi^m \mathbb{E}[X^2].$$

Proof. Without a loss of generality, assume that ψ is complex and remove β from all the proofs. For each $m \in \mathbb{N}$, we apply Jensen's inequality to get

$$\begin{aligned}
\sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} |S_q^m X(\lambda_1,\ldots,\lambda_m)|^2 &= \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} \mathbb{E}\left[||X * \psi_{\lambda_1}| * \cdots | * \psi_{\lambda_m}|^q\right]^{2/q} \\
&\leq \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} \mathbb{E}\left[||X * \psi_{\lambda_1}| * \cdots | * \psi_{\lambda_m}|^2\right] \\
&= \beta^m C_\psi^m \sigma^2(X) \\
&\leq \beta^m C_\psi^m \mathbb{E}[X^2].
\end{aligned}$$

□

Additionally, the expected scattering transform when $1 < q < 2$ are all nonexpansive operators because of the following lemma.

Theorem 38. *Suppose ψ is a littlewood paley wavelet with $\beta C_\psi \leq \frac{1}{2}$. Then $\|\overline{S}_q X - \overline{S}_q Y\|_2^2 \leq \mathbb{E}[|X-Y|^2]$ and $\|\overline{S}_q X\|_2^2 \leq \mathbb{E}[X^2]$.*

Proof. For notational simplicity, we use

$$\begin{aligned}
X_k &= ||X * \psi_{\lambda_1}| * \cdots | * \psi_{\lambda_k}|, \\
Y_k &= ||Y * \psi_{\lambda_1}| * \cdots | * \psi_{\lambda_k}|.
\end{aligned}$$

We have

$$\begin{aligned}
&\sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} |S_q^m X(\lambda_1,\ldots,\lambda_m) - S_q^m Y(\lambda_1,\ldots,\lambda_m)|^2 \\
&= \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} \left|\mathbb{E}\left[|X_{m-1} * \psi_{\lambda_m}|^q\right]^{1/q} - \mathbb{E}\left[|Y_{m-1} * \psi_{\lambda_m}|^q\right]^{1/q}\right|^2 \\
&\leq \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} \mathbb{E}\left[|(X_{m-1} - Y_{m-1}) * \psi_{\lambda_m}|^q\right]^{2/q} \\
&\leq \sum_{(\lambda_1,\ldots,\lambda_m)\in\Lambda^m} \mathbb{E}\left[|(X_{m-1} - Y_{m-1}) * \psi_{\lambda_m}|^2\right] \\
&\leq C_\psi^m \mathbb{E}[|X-Y|^2].
\end{aligned}$$

Now sum over all m to finish the proof. □

The following corollary also follows immediately from the proof above and the $q = 2$ case.

Corollary 39. *Suppose τ is a stochastic process independent of X and ψ is a littlewood paley wavelet with $\beta C_\psi \leq 1/2$. Consider the random process $X - L_\tau X$, and suppose that the Fourier Transform of its covariance function, $\hat{R}_{X-L_\tau X}(\omega)$, is supported on some finite ball with radius R centered at the origin: $B_R(0)$. We have the following estimate for some $C > 0$:*

$$\|\overline{S}_q X - \overline{S}_q L_\tau X\|_2^2 \leq C R^2 \mathbb{E}[\|\tau\|_\infty^2] \mathbb{E}[|X|^2].$$

CHAPTER 4

NONWINDOWED SCATTERING ON COMPACT RIEMANNIAN MANIFOLDS

In this chapter, we generalize our results with $q = 2$ to compact Riemannian manifolds. First, let us motivate why one would consider scattering transforms for non-Euclidean data. Suppose we have number written on a set spheres (i.e. spherical MNIST). We would like to classify which number is each of these spheres. A Euclidean approach would be to voxelize each of these spheres as $N \times N \times N$ discretized cubes and feed these cubes into a feature extractor (i.e. a scattering transform or a convolutional neural network). However, compared to a $N \times N$ image, this approach is N times more expensive in terms of memory because of the extra dimension. One can instead consider these as signals on the sphere, which has a lower intrinsic dimension. The point is that using Euclidean representations is not necessarily the best representation for feature extraction.

The paper [46] was the first to explore a unified framework for geometric deep learning, and [28, 27, 29] provided a mathematical framework for scattering transforms for noneuclidean data. Additionally, for spherical data, windowed scattering transforms have been generalized in [47, 48], where the convolution operation is specific to the sphere, and numerical implementations are optimized relative to [28] (with a trade-off of flexibility). As an aside, one could consider nonwindowed versions of [47, 48] for classification tasks on the sphere.

In particular, [28] defines the nonwindowed scattering transform for compact manifolds as $\mathbf{L}^1$ norms of a cascade of wavelet transforms and nonlinearities, which will be reviewed below. Similar to scattering moments and nonwindowed scattering transforms for Euclidean data, one would suspect that using $\mathbf{L}^q$ norms instead of $\mathbf{L}^1$ norms provide richer discriptors for signals on manifolds. This motivates our results for $q = 2$. Other values of q have been left to future work.

4.1 Notation for Scattering on Manifolds

Let $\mathcal{M}$ will be a compact, smooth, n-dimensional Riemannian manifold without boundary contained in $\mathbb{R}^d$, where $d \geq n$ with geodesic distance between two points $x_1, x_2 \in \mathcal{M}$ given by $r(x_1, x_2)$ and Laplace-Beltrami operator denoted as Δ. The notation $\mathbf{L}^q(\mathcal{M})$ denotes the set of all functions $f : \mathcal{M} \to \mathbb{R}$ such that $\int_{\mathcal{M}} |f(x)|^q \, dx < \infty$, where dx is integration with respect to the Riemannian volume. We use the notation $\mathrm{Isom}(\mathcal{M}_1, \mathcal{M}_2)$ be the set of isometries between manifolds $\mathcal{M}_1$ and $\mathcal{M}_2$. Lastly, the set of diffeomorphisms on $\mathcal{M}$ will be denoted by $\mathrm{Diff}(\mathcal{M})$, and the maximum placement of $\gamma \in \mathrm{Diff}(\mathcal{M})$ will be given by $\|\gamma\|_\infty := \sup_{x\in\mathcal{M}} r(x, \gamma(x))$.

4.2 Spectral Filters and the Geometric Wavelet Transform

We provide a brief summary of the geometric wavelet transform, as presented in [28]. The convolution of $f, g \in L^2(\mathbb{R}^n)$ is usually defined in space as

$$(f * g)(x) = \int_{\mathbb{R}^n} f(y)g(x - y)\, dy.$$

However, for a general manifold, even under the conditions we have prescribed, a notation of translation does not necessarily exist. Instead, one can consider a spectral definition of convolution via the spectral decomposition of $-\Delta$. Denote $\mathbb{N} \cup \{0\} = \mathbb{N}_0$. Because our manifold is compact, it is well known that $-\Delta$ has a discrete spectrum, and we can order the eigenvalues in increasing order and denote them as $\{\lambda_n\}_{n\in\mathbb{N}_0}$. We will denote the corresponding eigenfunctions as $\{\phi_n(x)\}_{n\in\mathbb{N}_0}$, which form an orthonormal basis for $\mathbf{L}^2(\mathcal{M})$.

Suppose $f \in \mathbf{L}^2(\mathcal{M})$. Since the set of functions $\{\phi_n(x)\}_{n\in\mathbb{N}_0}$ forms a basis in $\mathbf{L}^2(\mathcal{M})$, we decompose

$$f(x) = \sum_{n\in\mathbb{N}_0} \langle f, \phi_n\rangle \phi_n(x) = \sum_{n\in\mathbb{N}_0} \left(\int_{\mathcal{M}} f(y)\overline{\phi}_n(y)\, dy \right) \phi_n(x), \tag{4.1}$$

which is similar to a Fourier series. Since $\phi_n(y)$, is a replacement for a Fourier node, it is natural to let

$$\hat{f}(n) = \int_{\mathcal{M}} f(y)\overline{\phi}_n(y)\, dy \tag{4.2}$$

and define convolution on $\mathcal{M}$ between functions $f, h \in \mathbf{L}^2(\mathcal{M})$ as

$$f * h(x) = \sum_{n\in\mathbb{N}_0} \hat{f}(n)\hat{h}(n)\phi_n(x). \tag{4.3}$$

Defining the operator $T_h f(x) := f * h(x)$, it is easy to verify that the kernel for T_h is given by

$$K_h(x, y) := \sum_{n\in\mathbb{N}_0} \hat{h}(n)\phi_n(x)\overline{\phi}_n(y). \tag{4.4}$$

Similar to how convolution commutes with translations on $\mathbb{R}^n$, it is important for convolution on $\mathcal{M}$ to be equivariant to a group action on $\mathcal{M}$. The authors of [28] construct an operator by convolving with functions that commute with isometries since the the geometry of $\mathcal{M}$ should be preserved by a representation.

To accomplish this goal, we use a similar definition for spectral filters. A filter $h \in \mathbf{L}^2(\mathcal{M})$ is a spectral filter if $\lambda_k = \lambda_\ell$ implies $\hat{h}(k) = \hat{h}(\ell)$. One can prove that there exists $H : [0, \infty) \to \mathbb{R}$ such that

$$H(\lambda_n) = \hat{h}(n), \qquad \forall n \in \mathbb{N}_0.$$

Let $G : [0, \infty) \to \mathbb{R}$ be be nonnegative and decreasing with $G(0) > 0$. A low-pass spectral filter ϕ is given in frequency as $\hat{\phi}(k) := G(\lambda_k)$ and its dilation at scale 2^j for $j \in \mathbb{Z}$ is $\hat{\phi}_j(k) := G(2^j \lambda_k)$.

Using the set of low pass filters, $\{\hat{\phi}_j\}_{j\in\mathbb{Z}}$, we define wavelets by

$$\hat{\psi}_j(k) := \left[|\hat{\phi}_{j-1}(k)|^2 - |\hat{\phi}_j(k)|^2\right]^{1/2}, \tag{4.5}$$

which is identical to standard constructions of Littlewood Paley wavelets in Euclidean Space.

Fix $J \in \mathbb{Z}$. Define the operators

$$A_J f := f * \phi_J,$$
$$\Psi_j f := f * \psi_j, \qquad j \leq J.$$

The windowed geometric wavelet transform is given by

$$W_J f := \{A_J f, \Psi_j f : \quad j \leq J\} \tag{4.6}$$

and the nonwindowed geometric scattering transform is given by

$$W f := \{\Psi_j f : \quad j \in \mathbb{Z}\}. \tag{4.7}$$

We have the following theorem, which provides a condition for when our wavelet frame is a nonexpansive frame.

Theorem 40. *Let $G : [0,\infty) \to \mathbb{R}$ be nonnegative and decreasing with $0 < G(0) = C$, $\lim_{x\to\infty} G(x) = 0$, and $\{\psi_j\}_{j\in\mathbb{Z}}$ is a set of wavelets generated by the low pass filter $\hat{\phi}(k) = G(\lambda_k)$. Then we have*

$$\sum_{j\in\mathbb{Z}} \|f * \psi_j\|_2^2 = C\|f\|_2^2. \tag{4.8}$$

Proof. For fixed $J > 1$, we telescope to get

$$\begin{aligned}\sum_{j=-J}^{J} |\hat{\phi}_j(k)|^2 &= \sum_{j=-J}^{J} \left[|G(2^{j-1}\lambda_k)|^2 - |G(2^j\lambda_k)|^2\right] \\ &= |G(2^{J-1}\lambda_k)|^2 - |G(2^{-J}\lambda_k)|^2.\end{aligned}$$

Since $\lim_{J\to\infty} |G(2^{J-1}\lambda_k)|^2$ and $\lim_{J\to\infty} |G(2^{-J}\lambda_k)|^2$ both exist, it follows that

$$\sum_{j\in\mathbb{Z}} |\hat{\phi}_j(k)|^2 = \lim_{J\to\infty} |G(2^{J-1}\lambda_k)|^2 - \lim_{J\to\infty} |G(2^{-J}\lambda_k)|^2 = C.$$

We can write

$$\|f * \psi_j\|_2^2 = \sum_{n\in\mathbb{N}_0} |\hat{\psi}_j(k)|^2 |\hat{f}(k)|^2.$$

Thus, it follows that

$$\begin{aligned}\sum_{j\in\mathbb{Z}} \|f * \psi_j\|_2^2 &= \sum_{j\in\mathbb{Z}} \sum_{n\in\mathbb{N}_0} |\hat{f}(k)|^2 |\hat{\psi}_j(k)|^2 \\ &= \sum_{j\in\mathbb{Z}} |\hat{f}(k)|^2 \left(\sum_{n\in\mathbb{N}_0} |\hat{\psi}_j(k)|^2\right) \\ &= C\|f\|_2^2.\end{aligned}$$

□

4.3 The Geometric Scattering Transform

In an analogous manner to the Euclidean definition of the scattering transform, one would like to find a representation that meaningfully encodes high frequency information of a signal f. Define the propagator as

$$U[j]f := |W_j f| \qquad \forall j \in \mathbb{Z}, \tag{4.9}$$

which is convolution of a wavelet and applying a nonlinearity. Similarly, we can define the windowed propogator as

$$U_J[j]f := |W_j f| \qquad \forall j \leq J. \tag{4.10}$$

Similar to Scattering Transforms on Euclidean Space, one can apply a cascade of convolutions and modulus operators repeatedly. In particular, for $m \in \mathbb{N}$, let $j_1, \ldots, j_m \in \mathbb{Z}$. The m-layer propogator is defined as

$$U[j_1, \ldots, j_m] := U[j_m] \cdots U[j_1]f = |f * \psi_{j_1}| * \psi_{j_2} \cdots * \psi_{j_m}| \tag{4.11}$$

and the m-layer windowed propogator is defined as

$$U_J[j_1, \ldots, j_m] := U_J[j_m] \cdots U_J[j_1]f := |f * \psi_{j_1}| * \psi_{j_2} \cdots * \psi_{j_m}|, \qquad j_1, \ldots, j_m \leq J \tag{4.12}$$

with $U[\emptyset]f = f$ and $U_J[\emptyset]f = f$. To aggregate low information and get local isometry invariance, one can apply a low pass filter in a manner similar to pooling to each windowed propogator to get windowed scattering coefficients:

$$S_j[j_1, \ldots, j_m] = A_J U_J[j_1, \ldots, j_m]f = U_J[j_1, \ldots, j_m]f * \phi_J,$$

where we defined $S_J[\emptyset]f = f * \phi_J$. The windowed geometric scattering transform is given by

$$S_J f = \{S_j[j_1, \ldots, j_m] : m \geq 0, \quad j_i \leq J \quad \forall 1 \leq i \leq m\}. \tag{4.13}$$

The authors of [28] were able to prove that this nonwindowed scattering operator was nonexpansive, invariant to isometries up to the scale of the low pass filter, and stable to diffeomprohisms under mild assumptions.

In addition, the authors consider a nonwindowed scattering transform, which removes the low pass filtering. For applications such as manifold classification, requires full isometry invariance instead of isometry invarance up to the scale 2^J. We see that

$$\lim_{J \to \infty} S[j_1, \ldots, j_m]f(x) = \mathrm{vol}(\mathcal{M})^{-1/2} \|U[j_1, \ldots, j_m]f\|_1. \tag{4.14}$$

As a proxy, it is more appropriate to consider

$$\overline{S}f(j_1, \ldots, j_m) = \|U[j_1, \ldots, j_m]f\|_1, \tag{4.15}$$

which motivates defining the nonwindowed geometric scattering transform as

$$\overline{S}f = \{\overline{S}[j_1, \ldots, j_m] : m \geq 0, \quad j_i \in \mathbb{Z}, \quad \forall 1 \leq i \leq m\}. \tag{4.16}$$

However, as mentioned previously, [36, 49] motivate the use of nonwindowed geometric scattering operators as 2-norms of a cascade of convolutions and modulus operators:

$$\overline{S}_q f(j_1, \ldots, j_m) = \|U[j_1, \ldots, j_m]f\|_2.$$

Additionally, one can generalize nonwindowed geometric scattering transform to

$$\overline{S}_2 f = \{\overline{S}_2[j_1, \ldots, j_m] : m \geq 0, \quad j_i \in \mathbb{Z}, \quad \forall 1 \leq i \leq m\}, \tag{4.17}$$

which we will call the 2-nonwindowed geometric scattering transform.

4.4 Generalizing Geometric Scattering Transforms

To measure stability and invariance properties of the 2-nonwindowed geometric scattering transform, we need to define appropriate norms. The original nonwindowed geometric scattering transform was a mapping $\ell^2(\mathbf{L}^1(\mathcal{M})) \to \mathbf{L}^2(\mathcal{M})$, but our interpretation is slightly different. In particular, rather than thinking of the coefficients as a sequence, we group the coefficients in each layer and define the norm

$$\|\overline{S}_2 f\|^2 = \sum_{m=1}^{\infty} \left(\sum_{(j_1,\ldots,j_m)\in\mathbb{Z}^m} |\overline{S}_2 f(j_1, \ldots, j_m)|^2 \right) \tag{4.18}$$

with scattering distance given by

$$\|\overline{S}_2 f - \overline{S}_2 g\|^2 = \sum_{m=1}^{\infty} \left(\sum_{(j_1,\ldots,j_m)\in\mathbb{Z}^m} |\overline{S}_2 f(j_1, \ldots, j_m) - \overline{S}_2 g(j_1, \ldots, j_m)|^2 \right). \tag{4.19}$$

Theorem 41. *Let $G : [0, \infty) \to \mathbb{R}$ be nonnegative and decreasing with $0 < G(0) = \frac{1}{\sqrt{2}}$, $\lim_{x\to\infty} G(x) = 0$, and $\{\psi_j\}_{j\in\mathbb{Z}}$ be a set of spectral filters generated by G. Then we have*

$$\|\overline{S}_2 f - \overline{S}_2 g\| \leq \|f - g\|_2$$

for all $f, g \in \mathbf{L}^2(\mathcal{M})$.

Proof. We begin by proving that

$$\sum_{(j_1,\ldots,j_m)\in\mathbb{Z}^m} |\overline{S}_2 f(j_1,\ldots,j_m) - \overline{S}_2 g(j_1,\ldots,j_m)|^2 \leq 2^{-m}\|f\|_2^2$$

for all $m \in \mathbb{N}$ via induction.

In the case of $m = 1$, we see that

$$\begin{aligned}
\sum_{j\in\mathbb{Z}} |\overline{S}_2 f(j) - \overline{S}_2 g(j)|^2 &= \sum_{j\in\mathbb{Z}} |\|f * \psi_j\|_2 - |g * \psi_j\|_2|^2 \\
&\leq \sum_{j\in\mathbb{Z}} \|f * \psi_j - g * \psi_j\|_2^2 \\
&= \sum_{j\in\mathbb{Z}} \|(f - g) * \psi_j\|_2^2 \\
&\leq 2^{-1}\|f - g\|_2^2.
\end{aligned}$$

We can now work recursively. It follows that we can use similar ideas to the $m = 1$ case to get

$$\begin{aligned}
&\sum_{(j_1,\ldots,j_{m+1})\in\mathbb{Z}^{m+1}} |\overline{S}_2 f(j_1,\ldots,j_{m+1}) - \overline{S}_2 g(j_1,\ldots,j_{m+1})|^2 \\
&= \sum_{(j_1,\ldots,j_{m+1})\in\mathbb{Z}^{m+1}} \left|\|U[j_1,\ldots,j_m]f * \psi_{j+1}\|_2 - \|U[j_1,\ldots,j_m]g * \psi_{j+1}\|_2\right|^2 \\
&= \sum_{(j_1,\ldots,j_{m+1})\in\mathbb{Z}^{m+1}} \left|\|(U[j_1,\ldots,j_m]f - U[j_1,\ldots,j_m]g) * \psi_{j+1}\|_2\right|^2 \\
&\leq 2^{-1} \sum_{(j_1,\ldots,j_m)\in\mathbb{Z}^m} \||U[j_1,\ldots,j_{m-1}]f * \psi_{j_m}| - |U[j_1,\ldots,j_{m-1}]g * \psi_{j_m}|\|_2^2 \\
&\leq \sum_{(j_1,\ldots,j_m)\in\mathbb{Z}^m} \|U[j_1,\ldots,j_{m-1}]f * \psi_{j_m} - U[j_1,\ldots,j_{m-1}]g * \psi_{j_m}\|_2^2 \\
&\leq 2^{-2} \sum_{(j_1,\ldots,j_{m-1})\in\mathbb{Z}^{m-1}} \|U[j_1,\ldots,j_{m-1}]f - U[j_1,\ldots,j_{m-1}]g\|_2^2 \\
&\leq 2^{-k+1}\|f - g\|_2^2.
\end{aligned}$$

Now we can sum over all m to get

$$\|\overline{S}_2 f - \overline{S}_2 g\|^2 \leq \sum_{m=1}^{\infty} 2^{-m}\|f - g\|_2^2 = \|f - g\|_2^2.$$

$\square$

Corollary 42. *Let* $G : [0, \infty) \to \mathbb{R}$ *be nonnegative and decreasing with* $0 < G(0) \leq \frac{1}{\sqrt{2}}$, $\lim_{x\to\infty} G(x) = 0$, *and* $\{\psi_j\}_{j\in\mathbb{Z}}$ *be a set of spectral filters generated by* G. *Then we have*

$$\|\overline{S}_2 f\| \leq \|f\|_2$$

for all $f \in \mathbf{L}^2(\mathcal{M})$.

Towards the point of embedding proper invariance, we provide a theorem that demonstrates that the 2-nonwindowed geometric scattering transform is invariant to isometries.

Theorem 43. *Let* $\xi \in Isom(\mathcal{M}, \mathcal{M}')$, *and let* $f \in \mathbf{L}L^2(\mathcal{M})$. *Define* $f' = V_\xi f$ *and let* $\overline{S}_2'$ *be the corresponding* 2*-nonwindowed geometric scattering transform on* $\mathcal{M}'$ *produced by a littlewood paley wavelet satisfying the conditions described in Theorem 40. We have* $\overline{S}_2' f' = \overline{S}_2 f$.

Proof. We see that $\overline{S}_2[\emptyset] f = \|f\|_2 = \|V_\xi f\|_2$ since V_ξ is an isometry. Now suppose that we consider $p = (j_1, \ldots, j_m)$. Then

$$\begin{aligned}
\overline{S}_2[j_1, \ldots, j_m] f &= \|U[p] f\|_2 \\
&= \|V_\xi U[p] f\|_2 \\
&= \|U[p] V_\xi f\|_2 \\
&= \|U[p] f'\|_2 \\
&= \overline{S}_2'[j_1, \ldots, j_m] f'.
\end{aligned}$$

Thus, we can see that $\overline{S}_2' f' = \overline{S}_2 f$. □

Additionally, we also have a diffeomorphism stability result for λ-bandlimited functions (i.e. $\hat{f})(k) = \langle f, \phi_k\rangle = 0$ whenever $\lambda_k > \lambda$).

Lemma 44 ([28])**.** *Suppose* $\xi \in Diff(\mathcal{M})$. *If* $f \in \mathbf{L}^2(\mathcal{M})$ *is* λ*-bandlimited, and* $\xi \in Diff(\mathcal{M})$ *can be decomposed as* $\xi = \xi_1 \circ \xi_2$, *where* $\xi_2 \in Diff(\mathcal{M})$ *and* $\xi_1 \in Isom(\mathcal{M})$, *then*

$$\|f - V_\xi f\|_2 \leq C(\mathcal{M}) \lambda^n \|\xi\|_\infty \|f\|_2$$

for some constant $C(\mathcal{M})$.

Theorem 45. *Let $f \in \mathbf{L}^2(\mathcal{M})$, and assume that ψ is a wavelet family satisfying the conditions of Theorem 41 with $G(\lambda) \leq e^{-\lambda}$. If $\xi \in Diff(\mathcal{M})$ can be decomposed as $\xi = \xi_1 \circ \xi_2$, where $\xi_2 \in Diff(\mathcal{M})$ and $\xi_1 \in Isom(\mathcal{M})$, then*

$$\|\overline{S}_2 f - \overline{S}_2 V_\xi f\|_2 \leq C(\mathcal{M})\lambda^n \|\xi\|_\infty \|f\|_2.$$

Proof. The transform is nonexpansive, so Lemma 44 gives the desired result. □

CHAPTER 5

CONCLUSIONS

This thesis has provided a generalization of nonwindowed scattering transforms to signals in Euclidean space, as realizations stochastic processes, and signals on compact manifolds. Future work involves the following:

- Generalize the diffeomorphism bound from chapter 2 to stochastic processes. This is possible, but this is more difficult because the techniques used in Euclidean space for Chapter 2 do not apply directly.
- Apply q-scattering moments to audio texture synthesis. Based on the results of [50], one would expect that these scattering moments yield additional, relevant signal descriptors. However, does this yield better signal synthesis?
- Generalize the results of chapter 2 to create nonwindowed scattering transforms as a cascade of wavelet transforms, nonlinearities, and $\mathbf{L}^q$ norms on a compact manifold. This is left to future work, and requires results from singular integral theory.

BIBLIOGRAPHY

[1] M. Hirn, Lecture notes for mathematics of deep learning (April 2020).

[2] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, Communications of the ACM 60 (6) (2017) 84–90.

[3] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: International Conference on Learning Representations, 2015.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1–9.

[5] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770–778.

[6] A. Bietti, J. Mairal, Group invariance, stability to deformations, and complexity of deep convolutional representations, The Journal of Machine Learning Research 20 (1) (2019) 876–924.

[7] S. Mallat, Understanding deep convolutional networks, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 374 (2065) (2016) 20150203.

[8] J. Zarka, F. Guth, S. Mallat, Separation and concentration in deep networks, arXiv preprint arXiv:2012.10424 (2020).

[9] F. Guth, J. Zarka, S. Mallat, Phase collapse in neural networks, arXiv preprint arXiv:2110.05283 (2021).

[10] F. Guth, B. Ménard, G. Rochette, S. Mallat, A rainbow in deep network black boxes, arXiv preprint arXiv:2305.18512 (2023).

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

[12] J. Bruna, S. Mallat, Invariant scattering convolution networks, IEEE transactions on pattern analysis and machine intelligence 35 (8) (2013) 1872–1886.

[13] J. Andén, S. Mallat, Deep scattering spectrum, IEEE Transactions on Signal Processing 62 (16) (2014) 4114–4128.

[14] J. Andén, V. Lostanlen, S. Mallat, Joint time–frequency scattering, IEEE Transactions on Signal Processing 67 (14) (2019) 3704–3718.

[15] J. Bruna, S. Mallat, Audio texture synthesis with scattering moments (2013). arXiv:1311.0407.

[16] T. Angles, S. Mallat, Generative networks as inverse problems with scattering transforms, arXiv preprint arXiv:1805.06621 (2018).

[17] J. Bruna, S. Mallat, Multiscale sparse microcanonical models. arxiv e-prints, arXiv preprint arXiv:1801.02013 (2018).

[18] E. Oyallon, E. Belilovsky, S. Zagoruyko, Scaling the scattering transform: Deep hybrid networks, in: Proceedings of the IEEE international conference on computer vision, 2017, pp. 5618–5627.

[19] S. Gauthier, B. Thérien, L. Alsene-Racicot, M. Chaudhary, I. Rish, E. Belilovsky, M. Eickenberg, G. Wolf, Parametric scattering networks, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5749–5758.

[20] I. Waldspurger, Exponential decay of scattering coefficients, in: 2017 international conference on sampling theory and applications (SampTA), IEEE, 2017, pp. 143–146.

[21] T. Wiatowski, H. Bölcskei, A mathematical theory of deep convolutional neural networks for feature extraction, IEEE Transactions on Information Theory 64 (3) (2017) 1845–1866.

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

[23] W. Czaja, W. Li, Analysis of time-frequency scattering transforms, Applied and Computational Harmonic Analysis 47 (1) (2019) 149–171.

[24] W. Czaja, W. Li, Rotationally invariant time–frequency scattering transforms, Journal of Fourier Analysis and Applications 26 (2020) 1–23.

[25] J. Mairal, P. Koniusz, Z. Harchaoui, C. Schmid, Convolutional kernel networks, Advances in neural information processing systems 27 (2014).

[26] X. Cheng, X. Chen, S. Mallat, Deep haar scattering networks, Information and Inference: A Journal of the IMA 5 (2) (2016) 105–133.

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

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

[29] F. Gama, A. Ribeiro, J. Bruna, Stability of graph scattering transforms, Advances in Neural Information Processing Systems 32 (2019).

[30] D. Zou, G. Lerman, Graph convolutional neural networks via scattering, Applied and Computational Harmonic Analysis 49 (3) (2020) 1046–1074.

[31] F. Nicola, S. I. Trapasso, Stability of the scattering transform for deformations with minimal regularity, arXiv preprint arXiv:2205.11142 (2022).

[32] M. Hirn, S. Mallat, N. Poilvert, Wavelet scattering regression of quantum chemical energies, Multiscale Modeling & Simulation 15 (2) (2017) 827–863.

[33] M. Eickenberg, G. Exarchakis, M. Hirn, S. Mallat, L. Thiry, Solid harmonic wavelet scattering for predictions of molecule properties, The Journal of chemical physics 148 (24) (2018) 241732.

[34] P. Sinz, M. W. Swift, X. Brumwell, J. Liu, K. J. Kim, Y. Qi, M. Hirn, Wavelet scattering networks for atomistic systems with extrapolation of material properties, The Journal of Chemical Physics 153 (8) (2020) 084109.

[35] E. Cancès, M. Defranceschi, W. Kutzelnigg, C. Le Bris, Y. Maday, Computational quantum chemistry: A primer, in: Special Volume, Computational Chemistry, Vol. 10 of Handbook of Numerical Analysis, Elsevier, 2003, pp. 3–270. doi:https://doi.org/10.1016/S1570-8659(03)10003-8.

[36] A. Chua, M. Hirn, A. Little, On generalizations of the nonwindowed scattering transform, Applied and Computational Harmonic Analysis 68 (2024) 101597.

[37] L. Grafakos, Modern fourier analysis, Vol. 250, Springer, 2009.

[38] L. Grafakos, Classical fourier analysis, Vol. 249, Springer, 2014.

[39] J. García-Cuerva, J. R. De Francia, Weighted norm inequalities and related topics, Elsevier, 1985.

[40] J. P. Ward, K. N. Chaudhury, M. Unser, Decay properties of riesz transforms and steerable wavelets, SIAM Journal on Imaging Sciences 6 (2) (2013) 984–998.

[41] G. H. Hardy, J. E. Littlewood, G. Pólya, Inequalities, Cambridge University Press, Cambridge, 1988.

[42] E. M. Stein, T. S. Murphy, Harmonic Analysis (PMS-43): Real-Variable Methods, Orthogonality, and Oscillatory Integrals, Princeton University Press, 1993.

[43] G.-R. Liu, Y.-C. Sheu, H.-T. Wu, Asymptotic analysis of higher-order scattering transform of gaussian processes, Electronic Journal of Probability 27 (2022) 1–27.

[44] G.-R. Liu, Y.-C. Sheu, H.-T. Wu, Central and noncentral limit theorems arising from the scattering transform and its neural activation generalization, SIAM Journal on Mathematical Analysis 55 (2) (2023) 1170–1213.

[45] E. Allys, F. Levrier, S. Zhang, C. Colling, B. Regaldo-Saint Blancard, F. Boulanger, P. Hennebelle, S. Mallat, The rwst, a comprehensive statistical description of the non-gaussian structures in the ism, Astronomy & Astrophysics 629 (2019) A115.

[46] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, P. Vandergheynst, Geometric deep learning: going beyond euclidean data, IEEE Signal Processing Magazine 34 (4) (2017) 18–42.

[47] J. D. McEwen, C. G. Wallis, A. N. Mavor-Parker, Scattering networks on the sphere for scalable and rotationally equivariant spherical cnns, arXiv preprint arXiv:2102.02828 (2021).

[48] Y. Xiong, W. Dai, W. Fei, S. Li, H. Xiong, Anisotropic spherical scattering networks via directional spin wavelet, IEEE Transactions on Signal Processing 71 (2023) 2981–2996. doi:10.1109/TSP.2023.3304410.

[49] J. Chew, H. Steach, S. Viswanath, H.-T. Wu, M. Hirn, D. Needell, M. D. Vesely, S. Krishnaswamy, M. Perlmutter, The manifold scattering transform for high-dimensional point cloud data, in: Topological, Algebraic and Geometric Learning Workshops 2022, PMLR, 2022, pp. 67–78.

[50] J. Bruna, S. Mallat, E. Bacry, J. F. R. Muzy, Intermittent process analysis with scattering moments, Annals of Statistics 43 (1) (2015) 323–351.