Group Invariant Scattering

STÉPHANE MALLAT
Centre de Mathématiques Appliquées de l'École Polytechnique (CMAP)
Institut des Hautes Études Scientifiques (IHÉS)

Abstract

This paper constructs translation-invariant operators on $\mathbf{L}^2(\mathbb{R}^d)$, which are Lipschitz-continuous to the action of diffeomorphisms. A scattering propagator is a path-ordered product of nonlinear and noncommuting operators, each of which computes the modulus of a wavelet transform. A local integration defines a windowed scattering transform, which is proved to be Lipschitz-continuous to the action of $\mathbf{C}^2$ diffeomorphisms. As the window size increases, it converges to a wavelet scattering transform that is translation invariant. Scattering coefficients also provide representations of stationary processes. Expected values depend upon high-order moments and can discriminate processes having the same power spectrum. Scattering operators are extended on $\mathbf{L}^2(G)$, where G is a compact Lie group, and are invariant under the action of G. Combining a scattering on $\mathbf{L}^2(\mathbb{R}^d)$ and on $\mathbf{L}^2(SO(d))$ defines a translation- and rotation-invariant scattering on $\mathbf{L}^2(\mathbb{R}^d)$. © 2012 Wiley Periodicals, Inc.

1 Introduction

Symmetry and invariants, which play a major role in physics [6], are making their way into signal information processing. The information content of sounds or images is typically not affected under the action of finite groups such as translations or rotations, and it is stable to the action of small diffeomorphisms that deform signals [21]. This motivates the study of translation-invariant representations of $\mathbf{L}^2(\mathbb{R}^d)$ functions, which are Lipschitz-continuous to the action of diffeomorphisms and which keep high-frequency information to discriminate different types of signals. We then study invariance to the action of compact Lie groups and rotations.

We first concentrate on translation invariance. Let $L_c f(x) = f(x-c)$ denote the translation of $f \in \mathbf{L}^2(\mathbb{R}^d)$ by $c \in \mathbb{R}^d$. An operator Φ from $\mathbf{L}^2(\mathbb{R}^d)$ to a Hilbert space $\mathcal{H}$ is *translation-invariant* if $\Phi(L_c f) = \Phi(f)$ for all $f \in \mathbf{L}^2(\mathbb{R}^d)$ and $c \in \mathbb{R}^d$. Canonical translation-invariant operators satisfy $\Phi(f) = L_a f$ for some $a \in \mathbb{R}^d$ that depends upon f [15]. The modulus of the Fourier transform of f is an example of a noncanonical translation-invariant operator. However, these translation-invariant operators are not Lipschitz-continuous to the action of diffeomorphisms. Instabilities to deformations are well-known to appear at high

Communications on Pure and Applied Mathematics, Vol. LXV, 1331–1398 (2012)

frequencies [10]. The major difficulty is to maintain the Lipschitz continuity over high frequencies.

To preserve stability in $\mathbf{L}^2(\mathbb{R}^d)$ we want Φ to be nonexpansive:

$$\forall (f,h) \in \mathbf{L}^2(\mathbb{R}^d)^2 \quad \|\Phi(f) - \Phi(h)\|_{\mathcal{H}} \le \|f - h\|.$$

It is then sufficient to verify its Lipschitz continuity relative to the action of small diffeomorphisms close to translations. Such a diffeomorphism transforms $x \in \mathbb{R}^d$ into $x - \tau(x)$, where $\tau(x) \in \mathbb{R}^d$ is the displacement field. Let $L_\tau f(x) = f(x - \tau(x))$ denote the action of the diffeomorphism $\mathbb{1} - \tau$ on f. Lipschitz stability means that $\|\Phi(f) - \Phi(L_\tau f)\|$ is bounded by the "size" of the diffeomorphism and hence by the distance between the $\mathbb{1} - \tau$ and $\mathbb{1}$, up to a multiplicative constant multiplied by $\|f\|$. Let $|\tau(x)|$ denote the euclidean norm in $\mathbb{R}^d$, $|\nabla\tau(x)|$ the sup norm of the matrix $\nabla\tau(x)$, and $|H\tau(x)|$ the sup norm of the Hessian tensor. The weak topology on $\mathbf{C}^2$ diffeomorphisms defines a distance between $\mathbb{1}-\tau$ and $\mathbb{1}$ over any compact subset Ω of $\mathbb{R}^d$ by

$$d_\Omega(\mathbb{1}, \mathbb{1} - \tau) = \sup_{x\in\Omega} |\tau(x)| + \sup_{x\in\Omega} |\nabla\tau(x)| + \sup_{x\in\Omega} |H\tau(x)|. \tag{1.1}$$

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

$$\|\Phi(f) - \Phi(L_\tau f)\|_{\mathcal{H}} \le C\|f\|\Big(\sup_{x\in\mathbb{R}^d} |\nabla\tau(x)| + \sup_{x\in\mathbb{R}^d} |H\tau(x)|\Big). \tag{1.2}$$

Since Φ is translation invariant, the Lipschitz upper bound does not depend upon the maximum translation amplitude $\sup_x |\tau(x)|$ of the diffeomorphism metric (1.1). The Lipschitz continuity (1.2) implies that Φ is invariant to global translations, but it is much stronger. Φ is almost invariant to "local translations" by $\tau(x)$, up to the first- and second-order deformation terms.

High-frequency instabilities to deformations can be avoided by grouping frequencies into dyadic packets in $\mathbb{R}^d$ with a wavelet transform. However, a wavelet transform is not translation invariant. A translation-invariant operator is constructed with a scattering procedure along multiple paths, which preserves the Lipschitz stability of wavelets to the action of diffeomorphisms. A scattering propagator is first defined as a path-ordered product of nonlinear and noncommuting operators, each of which computes the modulus of a wavelet transform [13]. This cascade of convolutions and modulus can also be interpreted as a convolutional neural network [11]. A windowed scattering transform is a nonexpansive operator that locally integrates the scattering propagator output. For appropriate wavelets, the main theorem in Section 2 proves that a windowed scattering preserves the norm: $\|\Phi(f)\|_{\mathcal{H}} = \|f\|$ for all $f \in \mathbf{L}^2(\mathbb{R}^d)$, and it is Lipschitz-continuous to $\mathbf{C}^2$ diffeomorphisms.

When the window size increases, windowed scattering transforms converge to a translation-invariant scattering transform defined on a path set $\overline{\mathcal{P}}_\infty$ that is not countable. Section 3 introduces a measure μ and a metric on $\overline{\mathcal{P}}_\infty$ and proves that scattering transforms of functions in $\mathbf{L}^2(\mathbb{R}^d)$ belong to $\mathbf{L}^2(\overline{\mathcal{P}}_\infty, d\mu)$. A scattering transform has striking similarities with a Fourier transform modulus but a different behavior relative to the action of diffeomorphisms. Numerical examples are shown. An open conjecture remains on conditions for strong convergence in $\mathbf{L}^2(\mathbb{R}^d)$.

The representation of stationary processes with the Fourier power spectrum results from the translation invariance of the Fourier modulus. Similarly, Section 4 defines an expected scattering transform that maps stationary processes to an $\mathbf{l}^2$ space. Scattering coefficients depend upon high-order moments of stationary processes and thus can discriminate processes having the same second-order moments. As opposed to the Fourier spectrum, a scattering representation is Lipschitz-continuous to random deformations up to a log term. For large classes of ergodic processes, it is numerically observed that the scattering transform of a single realization provides a mean-square consistent estimator of the expected scattering transform.

Section 5 extends scattering operators to build invariants to actions of compact Lie groups G. The left action of $g \in G$ on $f \in \mathbf{L}^2(G)$ is denoted $L_g f(r) = f(g^{-1}r)$. An operator Φ on $\mathbf{L}^2(G)$ is invariant to the action of G if $\Phi(L_g f) = \Phi(f)$ for all $f \in \mathbf{L}^2(G)$ and all $g \in G$. Invariant scattering operators are constructed on $\mathbf{L}^2(G)$ with a scattering propagator that iterates on a wavelet transform defined on $\mathbf{L}^2(G)$, and a modulus operator that removes complex phases. A translation- and rotation-invariant scattering on $\mathbf{L}^2(\mathbb{R}^d)$ is obtained by combining a scattering on $\mathbf{L}^2(\mathbb{R}^d)$ and a scattering on $\mathbf{L}^2(SO(d))$.

Available at `www.cmap.polytechnique.fr/scattering` is a package of software to reproduce numerical experiments. Applications to audio and image classification can be found in [1, 3, 4, 18].

NOTATION. $\|\tau\|_\infty := \sup_{x\in\mathbb{R}^d} |\tau(x)|$, $\|\Delta\tau\|_\infty := \sup_{(x,u)\in\mathbb{R}^{2d}} |\tau(x)-\tau(u)|$, $\|\nabla\tau\|_\infty := \sup_{x\in\mathbb{R}^d} |\nabla\tau(x)|$, and $\|H\tau\|_\infty := \sup_{x\in\mathbb{R}^d} |H\tau(x)|$ where $|H\tau(x)|$ is the norm of the Hessian tensor. The inner product of $(x, y) \in \mathbb{R}^{2d}$ is $x \cdot y$. The norm of f in a Hilbert space is $\|f\|$ and in $\mathbf{L}^2(\mathbb{R}^d)$ $\|f\|^2 = \int |f(x)|^2\,dx$. The norm in $\mathbf{L}^1(\mathbb{R}^d)$ is $\|f\|_1 = \int |f(x)|dx$. We denote the Fourier transform of f by $\hat{f}(\omega) := \int f(x)e^{-ix\cdot\omega}\,d\omega$. We denote by $g \circ f(x) = f(gx)$ the action of a group element $g \in G$. An operator R parametrized by p is denoted by $R[p]$ and $R[\Omega] = \{R[p]\}_{p\in\Omega}$. The sup norm of a linear operator A in $\mathbf{L}^2(\mathbb{R}^d)$ is denoted by $\|A\|$, and the commutator of two operators is $[A, B] = AB - BA$.

2 Finite Path Scattering

To avoid high-frequency instabilities under the action of diffeomorphisms, Section 2.2 introduces scattering operators that iteratively apply wavelet transforms and remove complex phases with a modulus. Section 2.3 proves that a scattering is nonexpansive and preserves $\mathbf{L}^2(\mathbb{R}^d)$ norms. Translation invariance and Lipschitz continuity to deformations are proved in Sections 2.4 and 2.5.

2.1 From Fourier to Littlewood-Paley Wavelets

The Fourier transform modulus $\Phi(f) = |\widehat{f}|$ is translation invariant. Indeed, for $c \in \mathbb{R}^d$, the translation $L_c f(x) = f(x-c)$ satisfies $\widehat{L_c f}(\omega) = e^{-ic\cdot\omega}\,\widehat{f}(\omega)$ and hence $|\widehat{L_c f}| = |\widehat{f}|$. However, deformations lead to well-known instabilities at high frequencies [10]. This is illustrated with a small scaling operator, $L_\tau f(x) = f(x-\tau(x)) = f((1-s)x)$ for $\tau(x) = sx$ and $\|\nabla\tau\|_\infty = |s| < 1$. If $f(x) = e^{i\xi\cdot x}\,\theta(x)$, then scaling by $1-s$ translates the central frequency ξ to $(1-s)\xi$. If θ is regular with a fast decay, then

$$\| |\widehat{L_\tau f}| - |\widehat{f}| \| \sim |s|\,|\xi|\,\|\theta\| = \|\nabla\tau\|_\infty\,|\xi|\,\|f\|. \tag{2.1}$$

Since $|\xi|$ can be arbitrarily large, $\Phi(f) = |\widehat{f}|$ does not satisfy the Lipschitz continuity condition (1.2) when scaling high frequencies. The frequency displacement from ξ to $(1-s)\xi$ has a small impact if sinusoidal waves are replaced by localized functions having a Fourier support that is wider at high frequencies. This is achieved by a wavelet transform [7, 14] whose properties are briefly reviewed in this section.

A wavelet transform is constructed by dilating a wavelet $\psi \in \mathbf{L}^2(\mathbb{R}^d)$ with a scale sequence $\{a^j\}_{j\in\mathbb{Z}}$ for $a > 1$. For image processing, usually $a = 2$ [3, 4]. Audio processing requires a better frequency resolution with typically $a \le 2^{1/8}$ [1]. To simplify notation, we normalize $a = 2$, with no loss of generality. Dilated wavelets are also rotated with elements of a finite rotation group G, which also includes the reflection $-\mathbb{1}$ with respect to 0: $-\mathbb{1}x = -x$. If d is even, then G is a subgroup of $SO(d)$; if d is odd, then G is a finite subgroup of $O(d)$. A mother wavelet ψ is dilated by 2^{-j} and rotated by $r \in G$,

$$\psi_{2^j r}(x) = 2^{dj}\,\psi(2^j r^{-1} x). \tag{2.2}$$

Its Fourier transform is $\widehat{\psi}_{2^j r}(\omega) = \widehat{\psi}(2^{-j} r^{-1}\omega)$. A scattering transform is computed with wavelets that can be written

$$\psi(x) = e^{i\eta\cdot x}\theta(x) \quad \text{and hence} \quad \widehat{\psi}(\omega) = \widehat{\theta}(\omega - \eta), \tag{2.3}$$

where $\widehat{\theta}(\omega)$ is a real function concentrated in a low-frequency ball centered at $\omega = 0$ whose radius is of the order of π. As a result, $\widehat{\psi}(\omega)$ is real and concentrated in a frequency ball of the same radius but centered at $\omega = \eta$. To simplify notation, we denote $\lambda = 2^j r \in 2^{\mathbb{Z}} \times G$, with $|\lambda| = 2^j$. After dilation and rotation, $\widehat{\psi}_\lambda(\omega) = \widehat{\theta}(\lambda^{-1}\omega - \eta)$ covers a ball centered at $\lambda\eta$ with a radius proportional to $|\lambda| = 2^j$. The index λ thus specifies the frequency localization and spread of $\widehat{\psi}_\lambda$.

As opposed to wavelet bases, a Littlewood-Paley wavelet transform [7, 14] is a redundant representation that computes convolution values at all $x \in \mathbb{R}^d$ without subsampling:

$$\forall x \in \mathbb{R}^d \quad W[\lambda]f(x) = f \star \psi_\lambda(x) = \int f(u)\psi_\lambda(x-u)\,du. \tag{2.4}$$

Its Fourier transform is

$$\widehat{W[\lambda]f}(\omega) = \hat{f}(\omega)\hat{\psi}_\lambda(\omega) = \hat{f}(\omega)\hat{\psi}(\lambda^{-1}\omega).$$

If f is real then $\hat{f}(-\omega) = \hat{f}^*(\omega)$, and if $\hat{\psi}(\omega)$ is real then $W[-\lambda]f = W[\lambda]f^*$. Let G^+ denote the quotient of G with $\{-\mathbb{1}, \mathbb{1}\}$, where two rotations r and $-r$ are equivalent. It is sufficient to compute $W[2^j r]f$ for "positive" rotations $r \in G^+$. If f is complex, then $W[2^j r]f$ must be computed for all $r \in G = G^+ \times \{-\mathbb{1}, \mathbb{1}\}$.

A wavelet transform at a scale 2^J only keeps wavelets of frequencies $2^j > 2^{-J}$. The low frequencies that are not covered by these wavelets are provided by an averaging over a spatial domain proportional to 2^J:

$$A_J f = f \star \phi_{2^J} \quad \text{with } \phi_{2^J}(x) = 2^{-dJ}\phi(2^{-J}x). \tag{2.5}$$

If f is real, then the wavelet transform $W_J f = \{A_J f, (W[\lambda]f)_{\lambda \in \Lambda_J}\}$ is indexed by $\Lambda_J = \{\lambda = 2^j r : r \in G^+, 2^j > 2^{-J}\}$. Its norm is

$$\|W_J f\|^2 = \|A_J f\|^2 + \sum_{\lambda \in \Lambda_J} \|W[\lambda]f\|^2. \tag{2.6}$$

If $J = \infty$ then $W_\infty f = \{W[\lambda]f\}_{\lambda \in \Lambda_\infty}$ with $\Lambda_\infty = 2^{\mathbb{Z}} \times G^+$. Its norm is $\|W_\infty f\|^2 = \sum_{\lambda \in \Lambda_\infty} \|W[\lambda]f\|^2$. For complex-valued functions f, all rotations in G are included by defining $W_J f = \{A_J f, (W[\lambda]f)_{\lambda, -\lambda \in \Lambda_J}\}$ and $W_\infty f = \{W[\lambda]f\}_{\lambda, -\lambda \in \Lambda_\infty}$.The following proposition gives a standard Littlewood-Paley condition [7] so that W_J is unitary.

PROPOSITION 2.1. *For any $J \in \mathbb{Z}$ or $J = \infty$, W_J is unitary in the spaces of real-valued or complex-valued functions in $\mathbf{L}^2(\mathbb{R}^d)$ if and only if for almost all $\omega \in \mathbb{R}^d$*

$$\beta \sum_{j=-\infty}^{\infty} \sum_{r \in G} |\hat{\psi}(2^{-j} r^{-1}\omega)|^2 = 1 \quad \textit{and} \quad |\hat{\phi}(\omega)|^2 = \beta \sum_{j=-\infty}^{0} \sum_{r \in G} |\hat{\psi}(2^{-j} r^{-1}\omega)|^2, \tag{2.7}$$

where $\beta = 1$ for complex functions and $\beta = \frac{1}{2}$ for real functions.

PROOF. If f is complex, $\beta = 1$, and one can verify that (2.7) is equivalent to

$$\forall J \in \mathbb{Z} \quad |\hat{\phi}(2^J\omega)|^2 + \sum_{j > -J, r \in G} |\hat{\psi}(2^{-j} r^{-1}\omega)|^2 = 1. \tag{2.8}$$

Since $\widehat{W[2^j r]f}(\omega) = \hat{f}(\omega)\hat{\psi}_{2^j r}(\omega)$, multiplying (2.8) by $|\hat{f}(\omega)|^2$ and applying the Plancherel formula proves that $\|W_J f\|^2 = \|f\|^2$. For $J = \infty$ the same result is obtained by letting J go to ∞.

Conversely, if $\|W_J f\|^2 = \|f\|^2$, then (2.8) is satisfied for almost all ω. Otherwise, one can construct a function $f \neq 0$ where $\widehat{f}$ has a support in the domain of ω where (2.8) is not valid. With the Plancherel formula we verify that $\|W_J f\|^2 \neq \|f\|^2$, which contradicts the hypothesis.

If f is real then $|\widehat{f}(\omega)| = |\widehat{f}(-\omega)|$ so $\|W[2^j r]f\| = \|W[-2^j r]f\|$. Hence $\|W_J f\|$ remains the same if r is restricted to G^+ and ψ is multiplied by $\sqrt{2}$, which yields condition (2.7) with $\beta = \frac{1}{2}$. □

In all the following, $\widehat{\psi}$ is a real function that satisfies the unitary condition (2.7). It implies that $\widehat{\psi}(0) = \int \psi(x)\,dx = 0$ and $|\widehat{\phi}(r\omega)| = |\widehat{\phi}(\omega)|$ for all $r \in G$. We choose $\widehat{\phi}(\omega)$ to be real and symmetric so that ϕ is also real and symmetric and $\phi(rx) = \phi(x)$ for all $r \in G$. We also suppose that ϕ and ψ are twice differentiable and that their decay as well as the decay of their partial derivatives of order 1 and 2 is $O((1+|x|)^{-d-2})$.

A change of variable in the wavelet transform integral shows that if f is scaled and rotated, $2^l g \circ f = f(2^l g x)$ with $2^l g \in 2^{\mathbb{Z}} \times G$, then the wavelet transform is scaled and rotated according to

$$W[\lambda](2^l g \circ f) = 2^l g \circ W[2^{-l} g\lambda] f. \tag{2.9}$$

Since ϕ is invariant to rotations in G, we verify that A_J commutes with rotations in G: $A_J(g \circ f) = g \circ A_J f$ for all $g \in G$.

In dimension $d = 1$, $G = \{-\mathbb{1}, \mathbb{1}\}$. To build a complex wavelet ψ concentrated on a single frequency band according to (2.3), we set $\widehat{\psi}(\omega) = 0$ for $\omega < 0$. Following (2.7), W_J is unitary if and only if

$$\beta \sum_{j\in\mathbb{Z}} |\widehat{\psi}(2^{-j}|\omega|)|^2 = 1 \quad \text{and} \quad |\widehat{\phi}(\omega)|^2 = \beta \sum_{j=-\infty}^{0} |\widehat{\psi}(2^{-j}|\omega|)|^2. \tag{2.10}$$

If $\widetilde{\psi}$ is a real wavelet that generates a dyadic orthonormal basis of $\mathbf{L}^2(\mathbb{R})$ [14], then $\widehat{\psi} = 2\widehat{\widetilde{\psi}}\, \mathbb{1}_{\omega>0}$ satisfies (2.7). Numerical examples in the paper are computed with a complex wavelet ψ calculated from a cubic-spline orthogonal Battle-Lemarié wavelet $\widetilde{\psi}$ [14].

In any dimension $d \geq 2$, $\widehat{\psi} \in \mathbf{L}^2(\mathbb{R}^d)$ can be defined as a separable product in frequency polar coordinates $\omega = |\omega|\eta$, with η in the unit sphere $\mathbb{S}^d$ of $\mathbb{R}^d$:

$$\forall (|\omega|, \eta) \in \mathbb{R}^+ \times \mathbb{S}^d \quad \widehat{\psi}(|\omega|\eta) = \widehat{\psi}(|\omega|)\gamma(\eta).$$

The one-dimensional function $\widehat{\psi}(|\omega|)$ is chosen to satisfy (2.10). The Littlewood-Paley condition (2.7) is then equivalent to

$$\forall \eta \in \mathbb{S}^d \quad \sum_{r\in G} |\gamma(r^{-1}\eta)|^2 = 1.$$

2.2 Path-Ordered Scattering

Convolutions with wavelets define operators that are Lipschitz-continuous under the action of diffeomorphisms, because wavelets are regular and localized functions. However, a wavelet transform is not invariant to translations, and $W[\lambda]f = f \star \psi_\lambda$ translates when f is translated. The main difficulty is to compute translation-invariant coefficients, which remain stable under the action of diffeomorphisms and retain high-frequency information provided by wavelets. A scattering operator computes such a translation-invariant representation. We first explain how to build translation-invariant coefficients from a wavelet transform, while maintaining stability under the action of diffeomorphisms. Scattering operators are then defined, and their main properties are summarized.

If $U[\lambda]$ is an operator defined on $\mathbf{L}^2(\mathbb{R}^d)$, not necessarily linear but which commutes with translations, then $\int U[\lambda]f(x)dx$ is translation invariant if finite. $W[\lambda]f = f \star \psi_\lambda$ commutes with translations but $\int W[\lambda]f(x)dx = 0$ because $\int \psi(x)dx = 0$. More generally, one can verify that any linear transformation of $W[\lambda]f$, which is translation invariant, is necessarily 0. To get a nonzero invariant, we set $U[\lambda]f = M[\lambda]W[\lambda]f$ where $M[\lambda]$ is a nonlinear "demodulation" that maps $W[\lambda]f$ to a lower-frequency function having a nonzero integral. The choice of $M[\lambda]$ must also preserve the Lipschitz continuity to diffeomorphism actions.

If $\psi(x) = e^{i\eta\cdot x}\theta(x)$, then $\psi_\lambda(x) = e^{i\lambda\eta\cdot x}\,\theta_\lambda(x)$, and hence

$$W[\lambda]f(x) = \mathrm{e}^{i\lambda\eta\cdot x}(f^\lambda \star \theta_\lambda(x)) \quad \text{with } f^\lambda(x) = \mathrm{e}^{-i\lambda\eta\cdot x}\, f(x). \tag{2.11}$$

The convolution $f^\lambda \star \theta_\lambda$ is a low-frequency filtering because $\widehat{\theta}_\lambda(\omega) = \widehat{\theta}(\lambda^{-1}\omega)$ covers a frequency ball centered at $\omega = 0$, of radius proportional to $|\lambda|$. A nonzero invariant can thus be obtained by canceling the modulation term $\mathrm{e}^{i\lambda\eta\cdot x}$ with $M[\lambda]$. A simple example is

$$M[\lambda]h(x) = e^{-i\lambda\eta\cdot x}e^{-i\Phi(\widehat{h}(\lambda\eta))}h(x) \tag{2.12}$$

where $\Phi(\widehat{h}(\lambda\eta))$ is the complex phase of $\widehat{h}(\lambda\eta)$. This nonlinear phase registration guarantees that $M[\lambda]$ commutes with translations. From (2.11) we have that $\int M[\lambda]W[\lambda]f(x)dx = |\widehat{f}(\lambda\eta)|\,|\widehat{\theta}(0)|$. It recovers the Fourier modulus representation, which is translation invariant but not Lipschitz-continuous to diffeomorphisms as shown in (2.1). Indeed, the demodulation operator $M[\lambda]$ in (2.12) commutes with translations but does not commute with the action of diffeomorphisms and in particular with dilations. The commutator norm of $M[\lambda]$ with a dilation is equal to 2, even for arbitrarily small dilations, which explains the resulting instabilities.

Lipschitz continuity under the action of diffeomorphisms is preserved if $M[\lambda]$ commutes with the action of diffeomorphisms. For $\mathbf{L}^2(\mathbb{R}^d)$ stability, we also impose that $M[\lambda]$ is nonexpansive. One can prove [4] that $M[\lambda]$ is then necessarily a pointwise operator, which means that $M[\lambda]h(x)$ depends only on the value of h at x. We further impose that $\|M[\lambda]h\| = \|h\|$ for all $h \in \mathbf{L}^2(\mathbb{R}^d)$, which then implies that $|M[\lambda]h| = |h|$. The most regular functions are obtained with $M[\lambda]h = |h|$, which eliminates all phase variations. We derive from (2.11) that this modulus maps $W[\lambda]f$ into a lower-frequency envelope:

$$M[\lambda]W[\lambda]f = |W[\lambda]f| = |f^\lambda \star \theta_\lambda|.$$

Lower frequencies created by a modulus result from interferences. For example, if $f(x) = \cos(\xi_1 \cdot x) + a\cos(\xi_2 \cdot x)$ where ξ_1 and ξ_2 are in the frequency band covered by $\widehat{\psi}_\lambda$, then $|f \star \psi_\lambda(x)| = 2^{-1}|\widehat{\psi}_\lambda(\xi_1) + a\widehat{\psi}_\lambda(\xi_2)e^{i(\xi_2-\xi_1)\cdot x}|$ oscillates at the interference frequency $|\xi_2 - \xi_1|$, which is smaller than $|\xi_1|$ and $|\xi_2|$.

The integration $\int U[\lambda]f(x)dx = \int |f \star \psi_\lambda(x)|dx$ is translation invariant but it removes all the high frequencies of $|f \star \psi_\lambda(x)|$. To recover these high frequencies, a scattering also computes the wavelet coefficients of each $U[\lambda]f$: $\{U[\lambda]f \star \psi_{\lambda'}\}_{\lambda'}$. Translation-invariant coefficients are again obtained with a modulus $U[\lambda']U[\lambda]f = |U[\lambda]f \star \psi_{\lambda'}|$ and an integration $\int U[\lambda']U[\lambda]f(x)dx$. If $f(x) = \cos(\xi_1 \cdot x) + a\cos(\xi_2 \cdot x)$ with $a < 1$, and if $|\xi_2 - \xi_1| \ll |\lambda|$ with $|\xi_2 - \xi_1|$ in the support of $\widehat{\psi}_{\lambda'}$, then $U[\lambda']U[\lambda]f$ is proportional to $a|\psi_\lambda(\xi_1)|\,|\psi_{\lambda'}(|\xi_2 - \xi_1|)|$. The second wavelet $\widehat{\psi}_{\lambda'}$ captures the interferences created by the modulus, between the frequency components of f in the support of $\widehat{\psi}_\lambda$.

We now introduce the scattering propagator, which extends these decompositions.

DEFINITION 2.2. An ordered sequence $p = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ with $\lambda_k \in \Lambda_\infty = 2^{\mathbb{Z}} \times G^+$ is called a path. The empty path is denoted $p = \varnothing$. Let $U[\lambda]f = |f \star \psi_\lambda|$ for $f \in \mathbf{L}^2(\mathbb{R}^d)$. A scattering propagator is a path-ordered product of noncommutative operators defined by

$$U[p] = U[\lambda_m] \cdots U[\lambda_2]U[\lambda_1], \tag{2.13}$$

with $U[\varnothing] = \mathrm{Id}$.

The operator $U[p]$ is well-defined on $\mathbf{L}^2(\mathbb{R}^d)$ because $\|U[\lambda]f\| \le \|\psi_\lambda\|_1\|f\|$ for all $\lambda \in \Lambda_\infty$. The scattering propagator is a cascade of convolutions and modulus:

$$U[p]f = \big||f \star \psi_{\lambda_1}| \star \psi_{\lambda_2}\big| \cdots | \star \psi_{\lambda_m}|. \tag{2.14}$$

Each $U[\lambda]$ filters the frequency component in the band covered by $\widehat{\psi}_\lambda$ and maps it to lower frequencies with the modulus. The index sequence $p = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ is thus a frequency path variable. The scaling and rotation by $2^l g \in 2^{\mathbb{Z}} \times G$ of a path p is written $2^l gp = (2^l g\lambda_1, 2^l g\lambda_2, \ldots, 2^l g\lambda_m)$. The concatenation of two paths is denoted $p + p' = (\lambda_1, \lambda_2, \ldots, \lambda_m, \lambda'_1, \lambda'_2, \ldots, \lambda'_{m'})$; in particular, $p + \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m, \lambda)$. It results from (2.13) that

$$U[p + p'] = U[p']U[p]. \tag{2.15}$$

Section 2.1 explains that if f is complex valued then its wavelet transform is $W_\infty f = \{W[\lambda]f\}_{\lambda, -\lambda \in \Lambda_\infty}$, whereas if f is real then $W_\infty f = \{W[\lambda]f\}_{\lambda \in \Lambda_\infty}$. If f is complex then at the next iteration $U[\lambda_1]f = |W[\lambda_1]f|$ is real, so next-stage wavelet transforms are computed only for $\lambda_k \in \Lambda_\infty$. The scattering propagator of a complex function is thus defined over "positive" paths $p = (\lambda_1, \lambda_2, \dots, \lambda_m) \in \Lambda_\infty^m$ and "negative" paths denoted $-p = (-\lambda_1, \lambda_2, \dots, \lambda_m)$. This is analogous to the positive and negative frequencies of a Fourier transform. If f is real, then $W[-\lambda_1]f = W[\lambda_1]f^*$ so $U[-\lambda_1]f = U[\lambda_1]f$ and hence $U[-p]f = U[p]f$. To simplify explanations, all results are proved on real functions with scattering propagators restricted to positive paths. These results apply to complex functions by including negative paths.

DEFINITION 2.3. Let $\mathcal{P}_\infty$ be the set of all finite paths. The scattering transform of $f \in \mathbf{L}^1(\mathbb{R}^d)$ is defined for any $p \in \mathcal{P}_\infty$ by

$$\overline{S}f(p) = \frac{1}{\mu_p} \int U[p]f(x)\,dx \quad \text{with } \mu_p = \int U[p]\delta(x)\,dx. \tag{2.16}$$

A scattering is a translation-invariant operator that transforms $f \in \mathbf{L}^1(\mathbb{R}^d)$ into a function of the frequency path variable p. The normalization factor μ_p results from a path measure introduced in Section 3. Conditions are given so that μ_p does not vanish. This transform is then well-defined for any $f \in \mathbf{L}^1(\mathbb{R}^d)$ and any p of finite length m. Indeed, $\|\psi_\lambda\|_1 = \|\psi\|_1$ so (2.14) implies that $\|U[p]f\|_1 \leq \|f\|_1 \, \|\psi\|_1^m$. We shall see that a scattering transform shares similarities with the Fourier transform modulus, where the path p plays the role of a frequency variable. However, as opposed to a Fourier modulus, a scattering transform is stable under the action of diffeomorphisms, because it is computed by iterating on wavelet transforms and modulus operators, which are stable. For complex-valued functions, $\overline{S}f$ is also defined on negative paths, and $\overline{S}f(-p) = \overline{S}f(p)$ if f is real.

If $p \neq \varnothing$ then $\overline{S}f(p)$ is nonlinear but it preserves amplitude factors:

$$\forall \mu \in \mathbb{R} \quad \overline{S}(\mu f)(p) = |\mu| \overline{S}f(p). \tag{2.17}$$

A scattering has similar scaling and rotation covariance properties as a Fourier transform. If f is scaled and rotated, $2^l g \circ f(x) = f(2^l g x)$, then (2.9) implies that $U[\lambda](2^l g \circ f) = 2^l g \circ U[2^{-l} g \lambda]f$, and cascading this result shows that

$$\forall p \in \mathcal{P}_\infty \quad U[p](2^l g \circ f) = 2^l g \circ U[2^{-l} g p]f. \tag{2.18}$$

Inserting this result in the definition (2.16) proves that

$$\overline{S}(2^l g \circ f)(p) = 2^{-dl} \overline{S}f(2^{-l} g p). \tag{2.19}$$

Rotating f thus rotates identically its scattering, whereas if f is scaled by 2^l, then the frequency paths p are scaled by 2^{-l}. The extension of the scattering transform in $\mathbf{L}^2(\mathbb{R}^d)$ is done as a limit of windowed scattering transforms, which we now introduce.

DEFINITION 2.4. Let $J \in \mathbb{Z}$ and $\mathcal{P}_J$ be a set of finite paths $p = (\lambda_1, \lambda_2, \dots, \lambda_m)$ with $\lambda_k \in \Lambda_J$ and hence $|\lambda_k| = 2^{j_k} > 2^{-J}$. A windowed scattering transform is defined for all $p \in \mathcal{P}_J$ by

$$S_J[p]f(x) = U[p]f \star \phi_{2^J}(x) = \int U[p]f(u)\phi_{2^J}(x-u)\,du. \tag{2.20}$$

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

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

It defines an infinite family of functions indexed by $\mathcal{P}_J$, denoted by

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

For complex-valued functions, negative paths are also included in $\mathcal{P}_J$, and if f is real, $S_J[-p]f = S_J[p]f$.

Section 2.3 proves that for appropriate wavelets, $\|f\|^2 = \sum_{p \in \mathcal{P}_J} \|S_J[p]f\|^2$. However, the signal energy is mostly concentrated on a much smaller set of frequency-decreasing paths $p = (\lambda_k)_{k \le m}$ for which $|\lambda_{k+1}| \le |\lambda_k|$. Indeed, the propagator $U[\lambda]$ progressively pushes the energy towards lower frequencies. The main theorem of Section 2.5 proves that a windowed scattering is Lipschitz-continuous under the action of diffeomorphisms.

Since $\phi(x)$ is continuous at 0, if $f \in \mathbf{L}^1(\mathbb{R}^d)$ then its windowed scattering transform converges pointwise to its scattering transform when the scale 2^J goes to ∞:

$$\forall x \in \mathbb{R}^d \quad \lim_{J \to \infty} 2^{dJ} S_J[p]f(x) = \phi(0) \int U[p]f(u)\,du = \phi(0)\mu_p \overline{S}(p). \tag{2.21}$$

However, when J increases, the path set $\mathcal{P}_J$ also increases. Section 3 shows that $\{\mathcal{P}_J\}_{J \in \mathbb{Z}}$ defines a multiresolution path approximation of a much larger set $\overline{\mathcal{P}}_\infty$ including paths of infinite length. This path set is not countable as opposed to each $\mathcal{P}_J$, and Section 3 introduces a measure μ and a metric on $\overline{\mathcal{P}}_\infty$.

Section 3.2 extends the scattering transform $\overline{S}f(p)$ to all $f \in \mathbf{L}^2(\mathbb{R}^d)$ and to all $p \in \overline{\mathcal{P}}_\infty$, and proves that $\overline{S}f \in \mathbf{L}^2(\overline{\mathcal{P}}_\infty, d\mu)$. A sufficient condition is given to guarantee a strong convergence of $S_J f$ to $\overline{S}f$, and it is conjectured that it is valid on $\mathbf{L}^2(\mathbb{R}^d)$. Numerical examples illustrate this convergence and show that a scattering transform has strong similarities to a Fourier transforms modulus when mapping the path p to a frequency variable $\omega \in \mathbb{R}^d$.

2.3 Scattering Propagation and Norm Preservation

We prove that a windowed scattering S_J is nonexpansive and preserves the $\mathbf{L}^2(\mathbb{R}^d)$ norm. We denote by $S_J[\Omega] := \{S_J[p]\}_{p \in \Omega}$ and $U[\Omega] := \{U[p]\}_{p \in \Omega}$ a family of operators indexed by a path set Ω.

A windowed scattering can be computed by iterating on the *one-step propagator* defined by

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

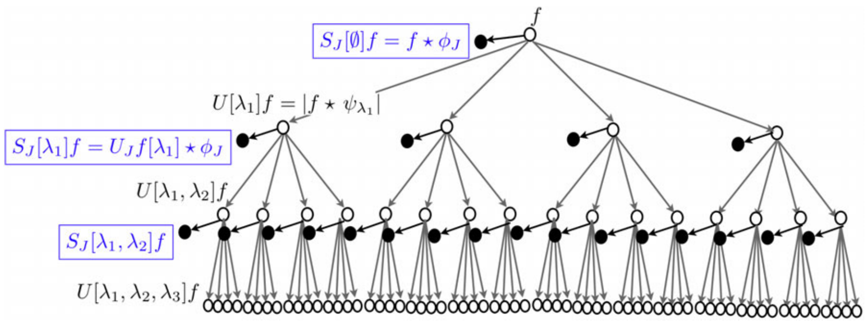

FIGURE 2.1. A scattering propagator U_J applied to f computes each $U[\lambda_1]f = |f \star \psi_{\lambda_1}|$ and outputs $S_J[\varnothing]f = f \star \phi_{2^J}$. Applying U_J to each $U[\lambda_1]f$ computes all $U[\lambda_1, \lambda_2]f$ and outputs $S_J[\lambda_1] = U[\lambda_1] \star \phi_{2^J}$. Applying iteratively U_J to each $U[p]f$ outputs $S_J[p]f = U[p]f \star \phi_{2^J}$ and computes the next path layer.

with $A_J f = f \star \phi_{2^J}$ and $U[\lambda]f = |f \star \psi_\lambda|$. After calculating $U_J f$, applying again U_J to each $U[\lambda]f$ yields a larger infinite family of functions. The decomposition is further iterated by recursively applying U_J to each $U[p]f$. Since $U[\lambda]U[p] = U[p + \lambda]$ and $A_J U[p] = S_J[p]$, it holds that

$$U_J U[p]f = \{S_J[p]f, (U[p + \lambda]f)_{\lambda \in \Lambda_J}\}. \tag{2.22}$$

Let Λ_J^m be the set of paths of length m with $\Lambda_J^0 = \{\varnothing\}$. It is propagated into

$$U_J U[\Lambda_J^m]f = \{S_J[\Lambda_J^m]f, U[\Lambda_J^{m+1}]f\}. \tag{2.23}$$

Since $\mathcal{P}_J = \bigcup_{m \in \mathbb{N}} \Lambda_J^m$, one can compute $S_J[\mathcal{P}_J]f$ from $f = U[\varnothing]f$ by iteratively computing $U_J U[\Lambda_J^m]f$ for m going from 0 to ∞, as illustrated in Figure 2.1.

Scattering calculations follow the general architecture of convolution neural networks introduced by LeCun [11]. Convolution networks cascade convolutions and a "pooling" nonlinearity, which is here the modulus of a complex number. Convolution networks typically use kernels that are not predefined functions such as wavelets but which are learned with back-propagation algorithms. Convolution network architectures have been successfully applied to a number of recognition tasks [11] and are studied as models for visual perception [2, 17]. Relations between scattering operators and path formulations of quantum field physics are also studied in [9].

The propagator $U_J f = \{A_J f, (|W[\lambda]f|)_{\lambda \in \Lambda_J}\}$ is nonexpansive because the wavelet transform W_J is unitary and a modulus is nonexpansive in the sense that $||a| - |b|| \le |a - b|$ for any $(a, b) \in \mathbb{C}^2$. This is valid whether f is real or complex.

As a consequence,

$$\|U_J f - U_J h\|^2 = \|A_J f - A_J h\|^2 + \sum_{\lambda \in \Lambda_J} \| |W[\lambda] f| - |W[\lambda] h| \|^2$$
$$\le \|W_J f - W_J h\|^2 \le \|f - h\|^2. \tag{2.24}$$

Since W_J is unitary, setting $h = 0$ also proves that $\|U_J f\| = \|f\|$, so U_J preserves the norm.

For any path set Ω the norms of $S_J[\Omega] f$ and $U[\Omega] f$ are

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

Since $S_J[\mathcal{P}_J]$ iterates on U_J, which is nonexpansive, the following proposition proves that $S_J[\mathcal{P}_J]$ is also nonexpansive [12].

PROPOSITION 2.5. *The scattering transform is nonexpansive*:

$$\forall (f, h) \in \mathbf{L}^2(\mathbb{R}^d)^2 \quad \|S_J[\mathcal{P}_J] f - S_J[\mathcal{P}_J] h\| \le \|f - h\|. \tag{2.25}$$

PROOF. Since U_J is nonexpansive, it results from (2.23) that

$$\begin{aligned} &\|U[\Lambda_J^m] f - U[\Lambda_J^m] h\|^2 \\ &\quad \ge \|U_J U[\Lambda_J^m] f - U_J U[\Lambda_J^m] h\|^2 \\ &\quad = \|S_J[\Lambda_J^m] f - S_J[\Lambda_J^m] h\|^2 + \|U[\Lambda_J^{m+1}] f - U[\Lambda_J^{m+1}] h\|^2. \end{aligned}$$

Summing these equations for m going from 0 to ∞ proves that

$$\begin{aligned} \|S_J[\mathcal{P}_J] f - S_J[\mathcal{P}_J] h\|^2 &= \sum_{m=0}^{\infty} \|S_J[\Lambda_J^m] f - S_J[\Lambda_J^m] h\|^2 \\ &\le \|f - h\|^2. \end{aligned} \tag{2.26}$$

□

Section 2.2 explains that each $U[\lambda] f = |f \star \psi_\lambda|$ captures the frequency energy of f over a frequency band covered by $\widehat{\psi}_\lambda$ and propagates this energy towards lower frequencies. The following theorem proves this result by showing that the whole scattering energy ultimately reaches the minimum frequency 2^{-J} and is trapped by the low-pass filter ϕ_{2^J}. The propagated scattering energy thus goes to 0 as the path length increases, and the theorem implies that $\|S_J[\mathcal{P}_J] f\| = \|f\|$. This result also applies to complex-valued functions by incorporating negative paths $(-\lambda_1, \lambda_2, \ldots, \lambda_m)$ in $\mathcal{P}_J$.

THEOREM 2.6. *A scattering wavelet ψ is said to be admissible if there exists $\eta \in \mathbb{R}^d$ and $\rho \ge 0$, with $|\widehat{\rho}(\omega)| \le |\widehat{\phi}(2\omega)|$ and $\widehat{\rho}(0) = 1$, such that the function*

$$\widehat{\Psi}(\omega) = |\widehat{\rho}(\omega - \eta)|^2 - \sum_{k=1}^{+\infty} k (1 - |\widehat{\rho}(2^{-k}(\omega - \eta))|^2) \tag{2.27}$$

satisfies

$$\alpha = \inf_{1\le|\omega|\le 2} \sum_{j=-\infty}^{+\infty} \sum_{r\in G} \widehat{\Psi}(2^{-j} r^{-1}\omega) |\widehat{\psi}(2^{-j} r^{-1}\omega)|^2 > 0. \tag{2.28}$$

If the wavelet is admissible, then for all $f \in \mathbf{L}^2(\mathbb{R}^d)$

$$\lim_{m\to\infty} \|U[\Lambda_J^m] f\|^2 = \lim_{m\to\infty} \sum_{n=m}^{\infty} \|S_J[\Lambda_J^n] f\|^2 = 0 \tag{2.29}$$

and

$$\|S_J[\mathcal{P}_J] f\| = \|f\|. \tag{2.30}$$

PROOF. We first prove that $\lim_{m\to\infty} \|U[\Lambda_J^m] f\| = 0$ is equivalent to having $\lim_{m\to\infty} \sum_{n=m}^{\infty} \|S_J[\Lambda_J^n] f\|^2 = 0$ and to $\|S_J[\mathcal{P}_J] f\| = \|f\|$. Since $\|U_J h\| = \|h\|$ for any $h \in \mathbf{L}^2(\mathbb{R}^d)$ and $U_J U[\Lambda_J^n] f = \{S_J[\Lambda_J^n] f, U[\Lambda_J^{n+1}]\}$,

$$\|U[\Lambda_J^n] f\|^2 = \|U_J U[\Lambda_J^n] f\|^2 = \|S_J[\Lambda_J^n] f\|^2 + \|U[\Lambda_J^{n+1}] f\|^2. \tag{2.31}$$

Summing for $m \le n < \infty$ proves that $\lim_{m\to\infty} \|U[\Lambda_J^m] f\| = 0$ is equivalent to $\lim_{m\to\infty} \sum_{n=m}^{\infty} \|S_J[\Lambda_J^n] f\|^2 = 0$. Since $f = U[\Lambda_J^0] f$, summing (2.31) for $0 \le n < m$ also proves that

$$\|f\|^2 = \sum_{n=0}^{m-1} \|S_J[\Lambda_J^n] f\|^2 + \|U[\Lambda_J^m] f\|^2, \tag{2.32}$$

so

$$\|S_J[\mathcal{P}_J] f\|^2 = \sum_{n=0}^{\infty} \|S_J[\Lambda_J^n] f\|^2 = \|f\|^2$$

if and only if $\lim_{m\to\infty} \|U[\Lambda_J^m]\| = 0$. □

We now prove that condition (2.27) implies that $\lim_{m\to\infty} \|U[\Lambda_J^m] f\|^2 = 0$. It relies on the following lemma, which gives a lower bound of $|f \star \psi_\lambda|$ convolved with a positive function.

LEMMA 2.7. *If* $h \ge 0$ *then for any* $f \in \mathbf{L}^2(\mathbb{R}^d)$

$$|f \star \psi_\lambda| \star h \ge \sup_{\eta\in\mathbb{R}^d} |f \star \psi_\lambda \star h_\eta| \quad \text{with } h_\eta(x) = h(x) e^{i\eta x}. \tag{2.33}$$

The lemma is proved by computing

$$\begin{aligned} |f \star \psi_\lambda| \star h(x) &= \int \left| \int f(v) \psi_\lambda(u - v) dv \right| h(x - u) du \\ &= \int \left| \int f(v) \psi_\lambda(u - v) e^{i\eta(x-u)} h(x - u) dv \right| du \\ &\ge \left| \iint f(v) \psi_\lambda(u - v) h(x - u) e^{i\eta(x-u)}\, dv\, du \right| = \end{aligned}$$

$$= \left| \int f(v) \int \psi_\lambda(x - v - u') h(u') e^{i\eta u'}\, du'\, dv \right|$$

$$= \left| \int f(v) \psi_\lambda \star h_\eta(x - v) dv \right| = |f \star \psi_\lambda \star h_\eta|,$$

which finishes the lemma's proof.

Appendix A uses this lemma to show that the scattering energy propagates progressively towards lower frequencies and proves the following lemma:

LEMMA 2.8. *If* (2.28) *is satisfied and*

$$\|f\|_w^2 = \sum_{j=0}^{\infty} \sum_{r \in G^+} j\, \|W[2^j r] f\|^2 < \infty \tag{2.34}$$

then

$$\frac{\alpha}{2} \|U[\mathcal{P}_J] f\|^2 \le \max(J + 1, 1) \|f\|^2 + \|f\|_w^2. \tag{2.35}$$

PROOF. The class of functions for which $\|f\|_w < \infty$ is a logarithmic Sobolev class corresponding to functions that have an average modulus of continuity in $\mathbf{L}^2(\mathbb{R}^d)$. Since

$$\|U[\mathcal{P}_J] f\|^2 = \sum_{m=0}^{+\infty} \|U[\Lambda_J^m] f\|^2,$$

if $\|f\|_w < \infty$ then (2.35) implies that $\lim_{m\to\infty} \|U[\Lambda_J^m] f\| = 0$. This result is extended in $\mathbf{L}^2(\mathbb{R}^d)$ by density. Since $\phi \in \mathbf{L}^1(\mathbb{R}^d)$ and $\widehat{\phi}(0) = 1$, any $f \in \mathbf{L}^2(\mathbb{R}^d)$ satisfies $\lim_{n\to-\infty} \|f - f_n\| = 0$, where $f_n = f \star \phi_{2^n}$ and $\phi_{2^n}(x) = 2^{-nd}\phi(2^{-n}x)$. We prove that $\lim_{m\to\infty} \|U[\Lambda_J^m] f_n\|^2 = 0$ by showing that $\|f_n\|_w < \infty$. Indeed,

$$\|W[2^j r] f_n\|^2 = \int |\widehat{f}(\omega)|^2\, |\widehat{\phi}(2^n \omega)|^2\, |\widehat{\psi}(2^{-j} r^{-1} \omega)|^2\, d\omega$$

$$\le C\, 2^{-2n-2j} \int |\widehat{f}(\omega)|^2\, d\omega,$$

because ψ has a vanishing moment so $|\widehat{\psi}(\omega)| = O(|\omega|)$, and the derivatives of ϕ are in $\mathbf{L}^1(\mathbb{R}^d)$ so $|\omega|\, |\widehat{\phi}(\omega)|$ is bounded. Thus we have that $\|f_n\|_w < \infty$.

Since $U[\Lambda^m]$ is nonexpansive, $\|U[\Lambda_J^m] f - U[\Lambda_J^m] f_n\| \le \|f - f_n\|$, so

$$\|U[\Lambda_J^m] f\| \le \|f - f_n\| + \|U[\Lambda_J^m] f_n\|.$$

Since $\lim_{n\to-\infty} \|f - f_n\| = 0$ and $\lim_{m\to\infty} \|U[\Lambda_J^m] f_n\| = 0$, we have

$$\lim_{m\to\infty} \|U[\Lambda_J^m] f\|^2 = 0$$

for any $f \in \mathbf{L}^2(\mathbb{R}^d)$. □

The proof shows that the scattering energy propagates progressively towards lower frequencies. The energy of $U[p]f$ is mostly concentrated along frequency-decreasing paths $p = (\lambda_k)_{k \le m}$ for which $|\lambda_{k+1}| < |\lambda_k|$. For example, if $f = \delta$, then paths of length 1 have an energy $\|U[2^j r]\delta\|^2 = \|\psi_{2^j r}\|^2 = 2^{-dj}\|\psi\|^2$. This energy is then propagated among all paths $p \in \mathcal{P}_J$. For a cubic spline wavelet in dimension $d = 1$, over 99.5% of this energy is concentrated along frequency-decreasing paths. Numerical implementations of scattering transforms thus limit computations to these frequency-decreasing paths. The scattering transform of a signal of size N is computed along all frequency-decreasing paths, with $O(N \log N)$ operations, by using a filter bank implementation [13].

The decay of $\sum_{n=m}^{\infty} \|S_J[\Lambda_J^n]f\|^2$ implies that we can neglect all paths of length larger than some $m > 0$. The numerical decay of $\|S_J[\Lambda_J^n]f\|^2$ appears to be exponential in image and audio processing applications. The path length is limited to $m = 3$ in classification applications [1, 3].

Theorem 2.6 requires a unitary wavelet transform and hence an admissible wavelet that satisfies the Littlewood-Paley condition $\beta \sum_{(j,r)\in\mathbb{Z}\times G} |\widehat{\psi}(2^j r\omega)|^2 = 1$. There must also exist $\rho \ge 0$ and $\eta \in \mathbb{R}^d$ with $|\widehat{\rho}(\omega)| \le |\widehat{\phi}(2\omega)|$ such that

$$\sum_{(j,r)\in\mathbb{Z}\times G} |\widehat{\psi}(2^j r\omega)|^2 \, |\widehat{\rho}(2^j r\omega - \eta)|^2$$

is sufficiently large so that $\alpha > 0$. This can be obtained if according to (2.3), $\psi(x) = e^{i\eta x}\theta(x)$ and hence $\widehat{\psi}(\omega) = \widehat{\theta}(\omega - \eta)$, where $\widehat{\theta}$ and $\widehat{\rho}$ have their energy concentrated over nearly the same low-frequency domains. For example, an analytic cubic spline Battle-Lemarié wavelet is admissible in one dimension with $\eta = 3\pi/2$. This is verified by choosing ρ to be a positive cubic box spline, in which case a numerical evaluation of (2.28) gives $\alpha = 0.2766 > 0$.

2.4 Translation Invariance

We show that the scattering distance $\|S_J[\overline{\mathcal{P}}_J]f - S_J[\overline{\mathcal{P}}_J]h\|$ is nonincreasing when J increases, and thus converges when J goes to ∞. It defines a limit distance that is proved to be translation invariant. Section 3 studies the convergence of $S_J[\mathcal{P}_J]f$, when J goes to ∞, to the translation-invariant scattering transform $\overline{S}f$.

PROPOSITION 2.9. *For all* $(f,h) \in \mathbf{L}^2(\mathbb{R}^d)^2$ *and* $J \in \mathbb{Z}$,

$$\|S_{J+1}[\mathcal{P}_{J+1}]f - S_{J+1}[\mathcal{P}_{J+1}]h\| \le \|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\|. \tag{2.36}$$

PROOF. Any $p' \in \mathcal{P}_{J+1}$ can be uniquely written as an extension of a path $p \in \mathcal{P}_J$ where p is the longest prefix of p' that belongs to $\mathcal{P}_J$, and $p' = p + q$ for some $q \in \mathcal{P}_{J+1}$. The set of all extensions of $p \in \mathcal{P}_J$ in $\mathcal{P}_{J+1}$ is

$$\mathcal{P}_{J+1}^p = \{p\} \cup \{p + 2^{-J}r + p''\}_{r\in G^+, p''\in\mathcal{P}_{J+1}}. \tag{2.37}$$

It defines a nonintersecting partition of $\mathcal{P}_{J+1} = \bigcup_{p\in\mathcal{P}_J} \mathcal{P}^p_{J+1}$. We shall prove that such extensions are nonexpansive,

$$\sum_{p'\in\mathcal{P}^p_{J+1}} \|S_{J+1}[p']f - S_{J+1}[p']h\|^2 \leq \|S_J[p]f - S_J[p]h\|^2. \tag{2.38}$$

To later prove Proposition 3.3, we also verify that it preserves energy,

$$\sum_{p'\in\mathcal{P}^p_{J+1}} \|S_{J+1}[p']f\|^2 = \|S_J[p]f\|^2. \tag{2.39}$$

Summing (2.38) on all $p \in \mathcal{P}_J$ proves (2.36).

Appendix A proves in (A.6) that for all $g \in \mathbf{L}^2(\mathbb{R}^d)$

$$\|g \star \phi_{2^{J+1}}\|^2 + \sum_{r\in G^+} \|g \star \psi_{2^{-J}r}\|^2 = \|g \star \phi_{2^J}\|^2.$$

Applying it to $g = U[p]f - U[p]h$ together with $U[p]f \star \phi_{2^J} = S_J[p]f$ and $|U[p]f \star \psi_{2^{-J}r}| = U[p + 2^{-J}r]f$ gives

$$\begin{aligned}\|S_J[p]f - S_J[p]h\|^2 \geq{}& \|S_{J+1}[p]f - S_{J+1}[p]h\|^2 \\ &+ \sum_{r\in G^+} \|U[p + 2^{-J}r]f - U[p + 2^{-J}r]h\|^2.\end{aligned} \tag{2.40}$$

Since

$$S_{J+1}[\mathcal{P}_{J+1}]U[p + 2^{-J}r]f = \{S_{J+1}[p + 2^{-J}r + p'']\}_{p''\in\mathcal{P}_{J+1}}$$

and $S_{J+1}[\mathcal{P}_{J+1}]f$ is nonexpansive, it implies

$$\begin{aligned}&\|S_J[p]f - S_J[p]h\|^2 \\ &\quad\geq \|S_{J+1}[p]f - S_{J+1}[p]h\|^2 \\ &\qquad + \sum_{p''\in\mathcal{P}_{J+1}} \sum_{r\in G^+} \big\|S_{J+1}[p + 2^{-J}r + p'']f - S_{J+1}[p + 2^{-J}r + p'']h\big\|^2,\end{aligned}$$

which proves (2.38). Since $S_J[\mathcal{P}_{J+1}]f$ preserves the norm, setting $h = 0$ in (2.40) gives an equality

$$\|S_J[p]f\|^2 = \|S_{J+1}[p]f\|^2 + \sum_{p''\in\mathcal{P}_{J+1}} \sum_{G^+} \|S_{J+1}[p + 2^{-J}r + p'']f\|^2,$$

which proves (2.39). □

This proposition proves that $\|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\|$ is positive and nonincreasing when J increases, and thus converges. Since $S_J[\mathcal{P}_J]$ is nonexpansive, the limit metric is also nonexpansive:

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

For admissible scattering wavelets that satisfy (2.28), Theorem 2.6 proves that $\|S_J[\mathcal{P}_J]f\| = \|f\|$ so $\lim_{J\to\infty}\|S_J[\mathcal{P}_J]f\| = \|f\|$. The following theorem proves that the limit metric is translation invariant:

THEOREM 2.10. *For admissible scattering wavelets*

$$\forall f \in \mathbf{L}^2(\mathbb{R}^d),\ \forall c \in \mathbb{R}^d \qquad \lim_{J\to\infty}\|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]L_c f\| = 0.$$

PROOF. Since $S_J[\mathcal{P}_J]\,L_c = L_c\,S_J[\mathcal{P}_J]$ and $S_J[\mathcal{P}_J]f = A_J\,U[\mathcal{P}_J]f$,

$$\begin{aligned}\|S_J[\mathcal{P}_J]L_c f - S_J[\mathcal{P}_J]f\| &= \|L_c A_J U[\mathcal{P}_J]f - A_J U[\mathcal{P}_J]f\| \\ &\le \|L_c A_J - A_J\|\,\|U[\mathcal{P}_J]f\|.\end{aligned} \tag{2.41}$$

LEMMA 2.11. *There exists C such that for all $\tau \in \mathbf{C}^2(\mathbb{R}^d)$ with $\|\nabla\tau\|_\infty \le \frac{1}{2}$ we have*

$$\|L_\tau A_J f - A_J f\| \le C\,\|f\|2^{-J}\|\tau\|_\infty. \tag{2.42}$$

This lemma is proved in Appendix B. Applying it to $\tau = c$ and hence $\|\tau\|_\infty = |c|$ proves that

$$\|L_c A_J - A_J\| \le C\,2^{-J}\,|c|. \tag{2.43}$$

Inserting this into (2.41) gives

$$\|L_c S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f\| \le C\,2^{-J}|c|\,\|U[\mathcal{P}_J]f\|. \tag{2.44}$$

Since the admissibility condition (2.28) is satisfied, Lemma 2.8 proves in (2.35) that for $J > 1$

$$\frac{\alpha}{2}\|U[\mathcal{P}_J]f\|^2 \le (J+1)\|f\|^2 + \|f\|_w^2. \tag{2.45}$$

If $\|f\|_w < \infty$ then from (2.44) we have

$$\|L_c S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f\|^2 \le ((J+1)\|f\|^2 + \|f\|_w^2)C^2\,2\,\alpha^{-1}\,2^{-2J}|c|^2$$

so $\lim_{J\to\infty}\|L_c S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f\| = 0$.

We then prove that $\lim_{J\to\infty}\|L_c S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f\| = 0$ for all $f \in \mathbf{L}^2(\mathbb{R}^d)$, with a similar density argument as in the proof of Theorem 2.6. Any $f \in \mathbf{L}^2(\mathbb{R}^d)$ can be written as a limit of $\{f_n\}_{n\in\mathbb{N}}$ with $\|f_n\|_w < \infty$, and since $S_J[\mathcal{P}_J]$ is nonexpansive and L_c unitary, one can verify that

$$\|L_c S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f\| \le \|L_c S_J[\mathcal{P}_J]f_n - S_J[\mathcal{P}_J]f_n\| + 2\|f - f_n\|.$$

Letting n go to ∞ proves that $\lim_{J\to\infty}\|L_c S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f\| = 0$, which finishes the proof. □

2.5 Lipschitz Continuity to Actions of Diffeomorphisms

This section proves that a windowed scattering is Lipschitz-continuous under the action of diffeomorphisms. A diffeomorphism of $\mathbb{R}^d$ sufficiently close to a translation maps x to $x - \tau(x)$ where $\tau(x)$ is a displacement field such that $\|\nabla\tau\|_\infty < 1$. The diffeomorphism action on $f \in \mathbf{L}^2(\mathbb{R}^d)$ is $L_\tau f(x) = f(x - \tau(x))$. The maximum increment of τ is denoted by $\|\Delta\tau\|_\infty := \sup_{(x,u)\in\mathbb{R}^{2d}} |\tau(x) - \tau(u)|$. Let S_J be a windowed scattering operator computed with an admissible scattering wavelet that satisfies (2.28). The following theorem computes an upper bound of $\|S_J[\mathcal{P}_J]L_\tau f - S_J[\mathcal{P}_J]f\|$ as a function of a mixed $(\mathbf{l}^1, \mathbf{L}^2(\mathbb{R}^d))$ scattering norm:

$$\|U[\mathcal{P}_J]f\|_1 = \sum_{m=0}^{+\infty} \|U[\Lambda_J^m]f\|. \tag{2.46}$$

We denote by $\mathcal{P}_{J,m}$ the subset of $\mathcal{P}_J$ of paths of length strictly smaller than m, and $(a \vee b) := \max(a, b)$.

THEOREM 2.12. *There exists C such that all $f \in \mathbf{L}^2(\mathbb{R}^d)$ with $\|U[\mathcal{P}_J]f\|_1 < \infty$ and all $\tau \in \mathbf{C}^2(\mathbb{R}^d)$ with $\|\nabla\tau\|_\infty \leq \frac{1}{2}$ satisfy*

$$\|S_J[\mathcal{P}_J]L_\tau f - S_J[\mathcal{P}_J]f\| \leq C\,\|U[\mathcal{P}_J]f\|_1\, K(\tau) \tag{2.47}$$

with

$$K(\tau) = 2^{-J}\|\tau\|_\infty + \|\nabla\tau\|_\infty\left(\log\frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty} \vee 1\right) + \|H\tau\|_\infty, \tag{2.48}$$

and for all $m \geq 0$

$$\|S_J[\mathcal{P}_{J,m}]L_\tau f - S_J[\mathcal{P}_{J,m}]f\| \leq C m \|f\| K(\tau). \tag{2.49}$$

PROOF. Let $[S_J[\mathcal{P}_J], L_\tau] = S_J[\mathcal{P}_J]L_\tau - L_\tau S_J[\mathcal{P}_J]$. We have

$$\begin{aligned}\|S_J[\mathcal{P}_J]L_\tau f - S_J[\mathcal{P}_J]f\| &\leq \|L_\tau S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f\| \\ &\quad + \|[S_J[\mathcal{P}_J], L_\tau]f\|.\end{aligned} \tag{2.50}$$

Similarly to (2.41), the first term on the right satisfies

$$\|L_\tau S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f\| \leq \|L_\tau A_J - A_J\|\,\|U[\mathcal{P}_J]f\|. \tag{2.51}$$

Since

$$\|U[\mathcal{P}_J]f\| = \left(\sum_{m=0}^{+\infty}\|U[\Lambda_J^m]f\|^2\right)^{1/2} \leq \sum_{m=0}^{+\infty}\|U[\Lambda_J^m]f\|,$$

we have that

$$\|L_\tau S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]f\| \leq \|L_\tau A_J - A_J\|\,\|U[\mathcal{P}_J]f\|_1. \tag{2.52}$$

Since $S_J[\mathcal{P}_J]$ iterates on U_J, which is nonexpansive, Appendix D proves the following upper bound on scattering commutators:

LEMMA 2.13. *For any operator L on $\mathbf{L}^2(\mathbb{R}^d)$*

$$\|[S_J[\mathcal{P}_J], L]f\| \leq \|U[\mathcal{P}_J]f\|_1\,\|[U_J, L]\|. \tag{2.53}$$

The operator $L = L_\tau$ also satisfies

$$\|[U_J, L_\tau]\| \le \|[W_J, L_\tau]\|. \tag{2.54}$$

Indeed, $U_J = M\,W_J$, where $M\{h_J, (h_\lambda)_{\lambda\in\Lambda_J}\} = \{h_J, (|h_\lambda|)_{\lambda\in\Lambda_J}\}$ is a nonexpansive modulus operator. Since $ML_\tau = L_\tau M$

$$\|[U_J, L_\tau]\| = \|M_J\,[W_J, L_\tau]\| \le \|[W_J, L_\tau]\|. \tag{2.55}$$

Inserting (2.53) with (2.54) and (2.52) in (2.50) gives

$$\|S_J[\mathcal{P}_J]L_\tau f - S_J[\mathcal{P}_J]f\| \le \|U[\mathcal{P}_J]f\|_1(\|L_\tau A_J - A_J\| + \|[W_J, L_\tau]\|). \tag{2.56}$$

Lemma 2.11 proves that $\|L_\tau A_J - A_J\| \le C\,2^{-J}\,\|\tau\|_\infty$. This inequality and (2.56) imply that

$$\|S_J[\mathcal{P}_J]L_\tau f - S_J[\mathcal{P}_J]f\| \le C\,\|U[\mathcal{P}_J]f\|_1(2^{-J}\|\tau\|_\infty + \|[W_J, L_\tau]\|). \tag{2.57}$$

To prove (2.47), the main difficulty is to compute an upper bound of $\|[W_J, L_\tau]\|$, and hence of $\|[W_J, L_\tau]\|^2 = \|[W_J, L_\tau]^*\,[W_J, L_\tau]\|$, where A^* is the adjoint of an operator A. The wavelet commutator applied to f is

$$[W_J, L_\tau]f = \{[A_J, L_\tau]f,\ ([W[\lambda], L_\tau]f)_{\lambda\in\Lambda_J}\},$$

whose norm is

$$\|[W_J, L_\tau]f\|^2 = \|[A_J, L_\tau]f\|^2 + \sum_{\lambda\in\Lambda_J} \|[W[\lambda], L_\tau]f\|^2. \tag{2.58}$$

From this we get that

$$[W_J, L_\tau]^*\,[W_J, L_\tau] = [A_J, L_\tau]^*\,[A_J, L_\tau] + \sum_{\lambda\in\Lambda_J} [W[\lambda], L_\tau]^*\,[W[\lambda], L_\tau].$$

The operator $[W_J, L_\tau]^*\,[W_J, L_\tau]$ has a singular kernel along the diagonal, but Appendix E proves that its norm is bounded.

LEMMA 2.14. *There exists* $C > 0$ *such that all* $J \in \mathbb{Z}$ *and all* $\tau \in \mathbf{C}^2(\mathbb{R}^d)$ *with* $\|\nabla\tau\|_\infty \le \frac{1}{2}$ *satisfy*

$$\|[W_J, L_\tau]\| \le C\left(\|\nabla\tau\|_\infty\left(\log\frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty} \vee 1\right) + \|H\tau\|_\infty\right). \tag{2.59}$$

Inserting the wavelet commutator bound (2.59) in (2.57) proves the theorem inequality (2.47). One can verify that (2.47) remains valid when replacing $\mathcal{P}_J$ by the subset of paths of length smaller than m: $\mathcal{P}_{J,m} = \bigcup_{n<m} \Lambda_J^n$ if we replace $\|U[\mathcal{P}_J]f\|_1$ by $\|U[\mathcal{P}_{J,m}]f\|_1$. The inequality (2.49) results from

$$\|U[\mathcal{P}_{J,m}]f\|_1 = \sum_{n=0}^{m-1} \|U[\Lambda_J^n]f\| \le m\|f\|. \tag{2.60}$$

This is obtained by observing that

$$\|U[\Lambda_J^n]f\| \le \|U[\Lambda_J^{n-1}]f\| \le \|f\|, \tag{2.61}$$

because $U[\Lambda_J^n]f$ is computed in (2.22) by applying the norm-preserving operator U_J on $U[\Lambda_J^{n-1}]f$. □

The condition $\|\nabla\tau\|_\infty \le \frac{1}{2}$ can be replaced by $\|\nabla\tau\|_\infty < 1$ if C is replaced by $C\,(1-\|\nabla\tau\|_\infty)^{-d}$. Indeed, $\|S_J[\mathcal{P}_J]f\| = \|f\|$ and

$$\|S_J[\mathcal{P}_J]L_\tau f\| \le \|f\|(1-\|\nabla\tau\|_\infty)^{-d}.$$

This remark applies to all subsequent theorems where the condition $\|\nabla\tau\|_\infty \le \frac{1}{2}$ appears. The theorem proves that the distance $\|S_J[\mathcal{P}_J]L_\tau f - S_J[\mathcal{P}_J]f\|$ produced by the diffeomorphism action L_τ is bounded by a translation term proportional to $2^{-J}\|\tau\|_\infty$ and a deformation error proportional to $\|\nabla\tau\|_\infty$. This deformation term results from the wavelet transform commutator $[W_J, L_\tau]$. The term $\log(\|\Delta\tau\|_\infty/\|\nabla\tau\|_\infty)$ can also be replaced by $\max(J,1)$ in the proof of Theorem 2.12. For compactly supported functions f, Corollary 2.15 replaces this term by the log of the support radius.

If $f \in \mathbf{L}^2(\mathbb{R}^d)$ has a weak form of regularity such as an average modulus of continuity in $\mathbf{L}^2(\mathbb{R}^d)$, then Lemma 2.8 proves that

$$\|U[\mathcal{P}_J]f\|^2 = \sum_{n=0}^{\infty} \|U[\Lambda_J^n]f\|^2$$

is finite. Numerical experiments indicate that $\|U[\Lambda_J^n]f\|$ has exponential decay for a large class of functions, but we do not characterize here the class of functions for which $\|U[\mathcal{P}_J]f\|_1 = \sum_{n=0}^{\infty}\|U[\Lambda_J^n]f\|$ is finite. In audio and image processing applications [1, 3], the percentage of scattering energy becomes negligible over paths of length larger than 3 so (2.49) is applied with $m = 4$.

The following corollary derives from Theorem 2.12 that a windowed scattering is Lipschitz-continuous under the action of diffeomorphisms over compactly supported functions.

COROLLARY 2.15. *For any compact $\Omega \subset \mathbb{R}^d$ there exists C such that for all $f \in \mathbf{L}^2(\mathbb{R}^d)$ supported in Ω with $\|U[\mathcal{P}_J]f\|_1 < \infty$ and for all $\tau \in \mathbf{C}^2(\mathbb{R}^d)$ with $\|\nabla\tau\|_\infty \le \frac{1}{2}$, if $2^J \ge \|\tau\|_\infty/\|\nabla\tau\|_\infty$, then*

$$\|S_J[\mathcal{P}_J]L_\tau f - S_J[\mathcal{P}_J]f\| \le C\,\|U[\mathcal{P}_J]f\|_1(\|\nabla\tau\|_\infty + \|H\tau\|_\infty). \tag{2.62}$$

PROOF. The inequality (2.62) is proved by applying (2.47) to a $\tilde{\tau}$ with $L_{\tilde{\tau}}f = L_\tau f$ and showing that there exists C' that depends only on Ω such that

$$2^{-J}\|\tilde{\tau}\|_\infty + \|\nabla\tilde{\tau}\|_\infty\left(\log\frac{\|\Delta\tilde{\tau}\|_\infty}{\|\nabla\tilde{\tau}\|_\infty} \vee 1\right) + \|H\tilde{\tau}\|_\infty \le C'(\|\nabla\tau\|_\infty + \|H\tau\|_\infty). \tag{2.63}$$

Since f has a support in Ω, $L_{\widetilde{\tau}} f = L_\tau f$ is equivalent to $\widetilde{\tau}(x) = \tau(x)$ for all $x \in \Omega_\tau = \{x : x - \tau(x) \in \Omega\}$ and $\widetilde{\tau}^{-1}(\Omega) = \Omega_\tau$. If Ω has a radius R, then the radius of Ω_τ is smaller than $2R$, because $\|\nabla\tau\|_\infty \leq \frac{1}{2}$. We define $\widetilde{\tau}$ as a regular extension of τ equal to $\tau(x)$ for $x \in \Omega_\tau$ and to the constant $\min_{x \in \Omega_\tau} \tau(x)$ outside a compact $\widetilde{\Omega}_\tau$ of radius $(4R+2)$ including Ω_τ. From this we have

$$\|\Delta\widetilde{\tau}\|_\infty = \sup_{(x,u)\in\widetilde{\Omega}_\tau^2} |\widetilde{\tau}(x) - \widetilde{\tau}(u)| \leq (4R+2)\|\nabla\widetilde{\tau}\|_\infty. \tag{2.64}$$

The extension in $\widetilde{\Omega}_\tau - \Omega_\tau$ can be made regular in the sense that

$$\|\nabla\widetilde{\tau}\|_\infty + \|H\widetilde{\tau}\|_\infty \leq \alpha\,(\|\nabla\tau\|_\infty\| + \|H\tau\|_\infty)$$

for some $\alpha > 0$ that depends on Ω. This property together with (2.64) proves (2.63). □

Similarly to Theorem 2.12, if $\mathcal{P}_J$ is replaced by the subset $\mathcal{P}_{J,m}$ of paths of length smaller than m, then $\|U[\mathcal{P}_J]f\|_1$ is replaced by $m\|f\|$ in (2.62). The upper bound (2.62) is proportional to $m|s|\,\|f\|$ if $L_\tau f(x) = f((1-s)x)$ with $|\nabla\tau(x)| = |s| < 1$. In this case, a lower bound is simply obtained by observing that since $\|S_J[\mathcal{P}_J]f\| = \|f\|$ and $\|S_J[\mathcal{P}_J]L_\tau f\| = \|L_\tau f\| = (1-s)^{-1}\,\|f\|$,

$$\|S_J[\mathcal{P}_J]L_\tau f - S_J[\mathcal{P}_J]f\| \geq |\|L_\tau f\| - \|f\|| > 2^{-1}\,s\,\|f\|.$$

Together with the upper bound (2.62), it proves that if $\tau(x) = sx$, then the scattering distance of f and $L_\tau f$ is of the order of $\|\nabla\tau\|_\infty\,\|f\|$.

The next theorem reduces the error term $2^{-J}\,\|\tau\|_\infty$ in Theorem 2.12 to a second-order term $2^{-2J}\,\|\tau\|_\infty^2$, with a first-order Taylor expansion of each $S_J[p]f$. We denote

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

THEOREM 2.16. *There exists C such that all $f \in \mathbf{L}^2(\mathbb{R}^d)$ with $\|U[\mathcal{P}_J]f\|_1 < \infty$ and all $\tau \in \mathbf{C}^2(\mathbb{R}^d)$ with $\|\nabla\tau\|_\infty \leq \frac{1}{2}$ satisfy*

$$\|S_J[\mathcal{P}_J]L_\tau f - S_J[\mathcal{P}_J]f + \tau\cdot\nabla S_J[\mathcal{P}_J]f\| \leq C\,\|U[\mathcal{P}_J]f\|_1\,K(\tau) \tag{2.65}$$

with

$$K(\tau) = 2^{-2J}\,\|\tau\|_\infty^2 + \|\nabla\tau\|_\infty\left(\log\frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty} \vee 1\right) + \|H\tau\|_\infty. \tag{2.66}$$

PROOF. The proof proceeds similarly to the proof of Theorem 2.12. Replacing $S_J[\mathcal{P}_J]L_\tau - S_J[\mathcal{P}_J]$ by $S_J[\mathcal{P}_J]L_\tau - S_J[\mathcal{P}_J] + \tau\cdot\nabla S_J[\mathcal{P}_J]$ in the derivation steps of the proof of Theorem 2.12 amounts to replacing $L_\tau A_J - A_J$ by $L_\tau A_J - A_J + \nabla A_J$. Equation (2.56) then becomes

$$\|S_J[\mathcal{P}_J]L_\tau f - S_J[\mathcal{P}_J]f + \tau\cdot\nabla S_J[\mathcal{P}_J]\| \leq$$
$$\|U[\mathcal{P}_J]f\|_1(\|L_\tau A_J - A_J + \nabla A_J\| + \|[W_J, L_\tau]\|).$$

Appendix C proves that there exists $C > 0$ such that

$$\|L_\tau A_J f - A_J + \nabla A_J\| \leq C\, 2^{-2J} \|\tau\|_\infty^2. \tag{2.67}$$

Inserting the upper bound (2.59) of $\|[W_J, L_\tau]\|$ proves (2.65). □

If $2^J \gg \|\tau\|_\infty$ and $\|\nabla\tau\|_\infty + \|H\tau\|_\infty \ll 1$, then $K(\tau)$ becomes negligible and $\tau(x)$ can be estimated at each x by solving the system of linear equations resulting from (2.65):

$$\forall p \in \mathcal{P}_J \quad S_J[p]L_\tau f(x) - S_J[p]f(x) + \tau(x) \cdot \nabla S_J[p]f(x) \approx 0. \tag{2.68}$$

In dimension d, the displacement $\tau(x)$ has d coordinates that can be computed if the system (2.68) has rank d. Estimating $\tau(x)$ has many applications. In image processing, the displacement field $\tau(x)$ between two consecutive images of a video sequence is proportional to the optical flow velocity of image points.

3 Normalized Scattering Transform

To define the convergence of $S_J[\mathcal{P}_J]$, all countable sets $\mathcal{P}_J$ are embedded in a noncountable set $\overline{\mathcal{P}}_\infty$. Section 3.1 constructs a measure μ and a metric in $\overline{\mathcal{P}}_\infty$. Section 3.2 redefines the scattering transform $\overline{S} f$ as the limit of windowed scattering transforms over $\overline{\mathcal{P}}_\infty$, with $\overline{S} f \in \mathbf{L}^2(\overline{\mathcal{P}}_\infty, d\mu)$ for $f \in \mathbf{L}^2(\mathbb{R}^d)$. Numerical comparisons between $\overline{S} f$ and $|\hat{f}|$ are given in Section 3.3.

3.1 Dirac Scattering Measure and Metric

A path $p \in \mathcal{P}_J$ can be extended into an infinite set of paths in $\mathcal{P}_{J+1}$ that refine p. In that sense, $\mathcal{P}_{J+1}$ is a set of higher-resolution paths. When J increases to ∞, these progressive extensions converge to paths of infinite length that belong to an uncountable path set $\overline{\mathcal{P}}_\infty$. A measure and a metric are defined on $\overline{\mathcal{P}}_\infty$.

A path $p = (\lambda_1, \lambda_2, \dots, \lambda_m)$ of length m belongs to the finite product set Λ_∞^m with $\Lambda_\infty = 2^{\mathbb{Z}} \times G^+$. An infinite path p is an infinite ordered string that belongs to the infinite product set Λ_∞^∞. For complex-valued functions, adding negative paths $(-\lambda_1, \lambda_2, \dots, \lambda_m)$ doubles the size of Λ_∞^m and Λ_∞^∞. We concentrate on positive paths $(\lambda_1, \lambda_2, \dots, \lambda_m)$ and the same construction applies to negative paths. Since $\Lambda_\infty = 2^{\mathbb{Z}} \times G^+$ is a discrete group, its natural topology is the discrete topology where basic open sets are individual elements. Open elements of the product topology of Λ_∞^∞ are cylinders defined for any $\lambda \in \Lambda_\infty$ and $n \geq 0$ by $C_n(\lambda) = \{q = \{q_k\}_{k>0} \in \Lambda_\infty^\infty : q_{n+1} = \lambda\}$ [22]. Cylinder sets are intersections of a finite number of open cylinders:

$$C_n(\lambda_1, \lambda_2, \dots, \lambda_m) = \{q \in \Lambda_\infty^\infty : q_{n+1} = \lambda_1, \lambda_2, \dots, q_{n+m} = \lambda_m\}$$
$$= \bigcap_{i=1}^{m} C_{n+i}(\lambda_i).$$

As elements of the topology, cylinder sets are open sets but are also closed. Indeed, the complement of a cylinder set is a union of cylinders and is thus closed. As a result, the topology is a sigma algebra on which a measure μ can be defined. The measure of a cylinder set C is written $\mu(C)$.

Let $\mathcal{P}_\infty$ be the set of all finite paths including the $\varnothing$ path: $\mathcal{P}_\infty = \bigcup_{m\in\mathbb{N}} \Lambda_\infty^m$. To any $p = (\lambda_1, \lambda_2, \ldots, \lambda_m) \in \mathcal{P}_\infty$, we associate a cylinder set

$$C(p) = C_0(p) = \{q \in \Lambda_\infty^\infty : q_1 = \lambda_1, \lambda_2, \ldots, q_m = \lambda_m\}.$$

This family of cylinder sets generates the same sigma algebra as open cylinders, because open cylinders can be written $C_n(\lambda) = \bigcup_{(\lambda_1,\lambda_2,\ldots,\lambda_n)\in\Lambda_\infty^n} C(\lambda_1, \lambda_2, \ldots, \lambda_n, \lambda)$. The following proposition defines a measure on Λ_∞^∞ from the scattering of a Dirac:

$$U[p]\delta = |\,||\psi_{\lambda_1}| \star \psi_{\lambda_2}| \star \cdots | \star \psi_{\lambda_m}|.$$

PROPOSITION 3.1. *There exists a unique σ-finite Borel measure μ, called the Dirac scattering measure, such that $\mu(C(p)) = \|U[p]\delta\|^2$ for all $p \in \mathcal{P}_\infty$. For all $2^l g \in \Lambda_\infty$ and $p \in \mathcal{P}_\infty$, $\mu(C(2^l gp)) = 2^{dl}\mu(C(p))$. If $|\widehat{\psi}(\omega)| + |\widehat{\psi}(-\omega)| \neq 0$ almost everywhere, then $\|U[p]\delta\| \neq 0$ for $p \in \mathcal{P}_\infty$.*

PROOF. The Dirac scattering measure is defined as a subdivision measure over the tree that generates all paths. Each finite path p corresponds to a node of the subdivision tree. Its sons are the $\{p+\lambda\}_{\lambda\in\Lambda_\infty}$, and $C(p) = \bigcup_{\lambda\in\Lambda_\infty} C(p+\lambda)$ is a nonintersecting partition. Since

$$\|U[p]\delta\|^2 = \|U_J U[p]\delta\|^2 = \sum_{\lambda\in\Lambda_\infty} \|U[p+\lambda]\delta\|^2,$$

we have that $\mu(C(p)) = \sum_{\lambda\in\Lambda_\infty} \mu(C(p+\lambda))$. The sigma additivity of the Dirac measure over all cylinder sets results from the tree structure, and the decomposition of the measure of a node $\mu(C(p))$ as a sum of the measures $\mu(C(p+\lambda))$ of all its sons. This subdivision measure is uniquely extended to the Borel sigma algebra through the sigma additivity. Since $\Lambda_\infty^\infty = \bigcup_{\lambda\in\Lambda_\infty} C(\lambda)$ and $\mu(C(\lambda)) = \|U[\lambda]\delta\|^2 = \|\psi_\lambda\|^2$, this measure is σ-finite.

We showed in (2.18) that $U[p](2^l g \circ f) = 2^l g \circ U[2^{-l}gp]f$. Since $2^l g \circ \delta = 2^{-dl}\delta$, we have $\|U[2^{-l}gp]\delta\|^2 = 2^{-dl}\|U[p]\delta\|^2$ and hence $\mu(C(2^l gp)) = 2^{dl}\mu(C(p))$.

For the set of $\omega \in \mathbb{R}^d$ where $\widehat{\psi}(\omega) = 0$ and $\widehat{\psi}(-\omega) = 0$ is of measure 0, let us prove by induction on the path length that $U[p]f \neq 0$ if $f \in \mathbf{L}^2(\mathbb{R}^d) \bigcup \mathbf{L}^1(\mathbb{R}^d)$ or if $f = \delta$. We suppose that $U[p]f \neq 0$ and verify that $U[p+\lambda]f \neq 0$ for any $\lambda \in \Lambda_\infty$. Since $U[p]f$ is real, $|\widehat{U[p]f}(\omega)| = |\widehat{U[p]f}(-\omega)|$. But $\widehat{\psi}_\lambda(\omega) = \widehat{\psi}(\lambda^{-1}\omega)$, so $\widehat{\psi}_\lambda(\omega)$ and $\widehat{\psi}_\lambda(-\omega)$ vanish simultaneously on a set of measure 0. We have that $\widehat{U[p+\lambda]f} = \widehat{U[p]f}\,\widehat{\psi}_\lambda \neq 0$ if $\widehat{U[p]f} \neq 0$ so $U[p+\lambda]f$ is a nonzero function. $\square$

A topology and a metric can now be constructed on the path set Λ_∞^∞. Neighborhoods are defined with cylinder sets of frequency resolution 2^J:

$$C_J(p) = \bigcup_{\substack{\lambda\in\Lambda_\infty \\ |\lambda|\le 2^{-J}}} C(p+\lambda) \subset C(p). \tag{3.1}$$

Clearly $C_{J+1}(p) \subset C_J(p)$. The following proposition proves that $\mu(C_J(p))$ decreases at least like 2^{-dJ} when 2^J increases, and it defines a distance from these measures. The set Λ_∞^∞ of infinite paths is not complete with this metric. It is completed by embedding the set $\mathcal{P}_\infty$ of finite paths, and we denote by $\overline{\mathcal{P}}_\infty := \mathcal{P}_\infty \cup \Lambda_\infty^\infty$ the completed set. This embedding is defined by adding each finite path $p \in \mathcal{P}_\infty$ to $C(p)$ and to each $C_J(p)$ for all $J \in \mathbb{Z}$ without modifying their measure. We still denote $C_J(p)$ the resulting subsets of $\overline{\mathcal{P}}_\infty$. For complex-valued functions, the size of $\overline{\mathcal{P}}_\infty$ is doubled by adding finite and infinite negative paths $(-\lambda_1, \lambda_2, \ldots, \lambda_m, \ldots)$.

PROPOSITION 3.2. *If $p \in \mathcal{P}_\infty$ is a path of length m, then*

$$\mu(C_J(p)) = \|S_J\delta[p]\|^2 \le 2^{-dJ}\,\|\phi\|^2\,\|\psi\|_1^{2m}. \tag{3.2}$$

Suppose that $|\widehat{\psi}(\omega)| + |\widehat{\psi}(-\omega)| \neq 0$ almost everywhere. For any $q \neq q' \in \overline{\mathcal{P}}_\infty$

$$\bar{d}(q,q') = \inf_{(q,q')\in C_J(p)^2} \mu(C_J(p)) \quad \textit{and} \quad \bar{d}(q,q) = 0 \tag{3.3}$$

defines a distance on $\overline{\mathcal{P}}_\infty$, and $\overline{\mathcal{P}}_\infty$ is complete for this metric.

PROOF. According to (3.1)

$$\mu(C_J(p)) = \sum_{\substack{\lambda\in\Lambda_\infty \\ |\lambda|\le 2^{-J}}} \mu(C(p+\lambda)) = \sum_{\substack{\lambda\in\Lambda_\infty \\ |\lambda|\le 2^{-J}}} \|U[p+\lambda]\delta\|^2.$$

Since $U[p+\lambda]\delta = |U[p]\delta \star \psi_\lambda|$ and

$$|\widehat{\phi}_{2^J}(\omega)|^2 = \sum_{\substack{\lambda\in\Lambda_\infty \\ |\lambda|\le 2^{-J}}} |\widehat{\psi}_\lambda(\omega)|^2,$$

the Plancherel formula implies

$$\mu(C_J(p)) = \sum_{\substack{\lambda\in\Lambda_\infty \\ |\lambda|\le 2^{-J}}} \|U[p]\delta \star \psi_\lambda\|^2 = \|U[p]\delta \star \phi_{2^J}\|^2 = \|S_J[p]\delta\|^2.$$

Since $S_J[p]\delta = U[p]\delta \star \phi_{2^J}$, Young's inequality implies

$$\|S_J[p]\delta\| \le \|U[p]\delta\|_1\,\|\phi_{2^J}\|.$$

Moreover, $\|U[\lambda]f\|_1 \le \|\psi_\lambda\|_1\,\|f\|_1$ with $\|\psi_\lambda\|_1 = \|\psi\|_1$, so we verify by induction that $\|U[p]\delta\|_1 \le \|\psi\|^m$. Inserting $\|\phi_{2^J}\|^2 = 2^{-dJ}\|\phi\|^2$ proves (3.2).

Let us now prove that $\bar{d}$ defines a distance. If $q \neq q'$, we denote by $\bar{p} \in \mathcal{P}_\infty$ their common prefix of longest size m, which may be 0, and we show that

$\overline{d}(q,q') \neq 0$. Let $|q_{m+1}| = 2^{j_{m+1}}$ and $|q'_{m+1}| = 2^{j'_{m+1}}$ be the frequencies of their first different coordinate. If $2^{-J} = \max(|q_{m+1}|, |q'_{m+1}|)$ then $(q,q') \in C_J(\overline{p})^2$ and it is the smallest set including both paths so $\overline{d}(q,q') = \mu(C_J(\overline{p}))$. We have that $\overline{d}(q,q') \neq 0$ because $\mu(C_J(\overline{p})) \geq \mu(C(\overline{p} + 2^{-J}r))$ for $r \in G^+$ and Proposition 3.1 proves that $\mu(C(p)) \neq 0$ for all $p \in \mathcal{P}_\infty$, so $\overline{d}(q,q') \neq 0$.

The triangle inequality is proved by showing that

$$\forall (q,q',q'') \in \overline{\mathcal{P}}_\infty^3 \quad \overline{d}(q',q'') \leq \max(\overline{d}(q,q'), \overline{d}(q,q'')). \tag{3.4}$$

This is verified by writing $\overline{d}(q,q') = \mu(C_J(\overline{p}))$, $\overline{d}(q',q'') = \mu(C_{J'}(\overline{p}'))$, and $\overline{d}(q',q'') = \mu(C_{J''}(\overline{p}''))$. Necessarily $\overline{p}$ is a substring of $\overline{p}'$ or vice versa, and $\overline{p}''$ is larger than the smallest of the two. If $\overline{p}''$ is strictly larger then the smallest, say $\overline{p}$, then $\mu(C_{J''}(\overline{p}'')) \leq \mu(C(\overline{p}'')) \leq C_J(\overline{p})$, so (3.4) is satisfied. If $\overline{p}'' = \overline{p} = \overline{p}'$, then $2^{-J''} \leq \max(2^{-J}, 2^{-J'})$ and (3.4) is satisfied. Otherwise $\overline{p}'' = \overline{p}$ is strictly smaller than $\overline{p}'$ and necessarily $2^{J''} = 2^J$ so (3.4) is also satisfied.

To prove that $\overline{\mathcal{P}}_\infty$ is complete, consider a Cauchy sequence $\{q_j\}_{j\in\mathbb{N}}$ in $\overline{\mathcal{P}}_\infty$. Let p_k be the common prefix of maximum length m_k among all q_j for $j \geq k$. It is a growing string that either converges to a finite string $p \in \mathcal{P}_\infty$ if m_k is bounded or to an infinite string $p \in \Lambda_\infty^\infty$. Among all paths $\{q_j\}_{j\geq k}$ whose maximum common prefix with p has a length m_k, let q_{j_k} be a path whose next element λ_{m_k+1} has a maximum frequency amplitude $|\lambda_{m_k+1}|$. One can verify that

$$\sup_{j,j'\geq k} \overline{d}(q_j, q_{j'}) = \overline{d}(q_{j_k}, p) = \sup_{j\geq k} \overline{d}(p, q_j).$$

The convergence of $\sup_{j,j'\geq k} \overline{d}(q_j, q_{j'})$ to 0 as k increases also implies the convergence of $\sup_{j\geq k} \overline{d}(p, q_j)$ to 0 and hence the convergence of $\{q_j\}_{j\in\mathbb{N}}$ to p. □

3.2 Scattering Convergence

For $h \in \mathbf{L}^2(\overline{\mathcal{P}}_\infty, d\mu)$, we denote $\|h\|^2_{\overline{\mathcal{P}}_\infty} = \int_{\overline{\mathcal{P}}_\infty} |h(q)|^2 \, d\mu(q)$, where μ is the Dirac scattering measure. This section redefines the scattering transform $\overline{S} f$ as a limit of windowed scattering transforms and proves that $\overline{S} f \in \mathbf{L}^2(\overline{\mathcal{P}}_\infty, d\mu)$ for all $f \in \mathbf{L}^2(\mathbb{R}^d)$. We suppose that ψ is an admissible scattering wavelet and that $|\widehat{\psi}(\omega)| + |\widehat{\psi}(-\omega)| \neq 0$ almost everywhere.

Let $\mathbb{1}_{C_J(p)}(q)$ be the indicator function of $C_J(p)$ in $\overline{\mathcal{P}}_\infty$. A windowed scattering $S_J[\mathcal{P}_J]f(x) = \{S_J[p]f(x)\}_{p\in\mathcal{P}_J}$ is first extended into a normalized function of $(q,x) \in \overline{\mathcal{P}}_\infty \times \mathbb{R}^d$

$$S_J f(q,x) = \sum_{p\in\mathcal{P}_J} \frac{S_J[p]f(x)}{\|S_J[p]\delta\|} \mathbb{1}_{C_J(p)}(q). \tag{3.5}$$

It satisfies $S_J f(p,x) = S_J[p]f(x)/\|S_J[p]\delta\|$ for $p \in \mathcal{P}_\infty$. Since $\mu(C_J(p)) = \|S_J[p]\delta\|^2$, for all $(f,h) \in \mathbf{L}^2(\mathbb{R}^d)^2$

$$\int_{\overline{\mathcal{P}}_\infty} \int_{\mathbb{R}^d} |S_J f(q,x) - S_J h(q,x)|^2 \, d\mu(q) dx = \|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\|^2 \leq \|f-h\|^2,$$

and

$$\int_{\overline{\mathcal{P}}_\infty} \int_{\mathbb{R}^d} |S_J f(q,x)|^2 \, d\mu(q) dx = \|S_J[\mathcal{P}_J]f\|^2 = \|f\|^2,$$

so $S_J f(q,x)$ can be interpreted as a scattering energy density in $\overline{\mathcal{P}}_\infty \times \mathbb{R}^d$.

The windowed scattering $S_J f(q,x)$ has a spatial resolution 2^{-J} along x and a resolution 2^J along the frequency path q. When J goes to ∞, $S_J f(q,x)$ loses its spatial localization, and Theorem 2.10 proves that the asymptotic metric on $S_J[\mathcal{P}_J]f$ and hence on $S_J f(q,x)$ is translation invariant. The convergence of $S_J f(q,x)$ to a function that depends only on $q \in \overline{\mathcal{P}}_\infty$ is studied by introducing the marginal $\mathbf{L}^2(\mathbb{R}^d)$ norm of $S_J f(q,x)$ along x for q fixed:

$$\forall q \in \overline{\mathcal{P}}_\infty \quad \overline{S}_J f(q) = \int |S_J f(q,x)|^2 \, dx = \sum_{p \in \mathcal{P}_J} \frac{\|S_J[p]f\|}{\|S_J[p]\delta\|} \mathbb{1}_{C_J(p)}(q). \tag{3.6}$$

It is a piecewise constant function of the path variable q whose resolution increases with J. Since $\mu(C_J(p)) = \|S_J[p]\delta\|^2$,

$$\|\overline{S}_J f - \overline{S}_J h\|^2_{\overline{\mathcal{P}}_\infty} = \int_{\overline{\mathcal{P}}_\infty} |\overline{S}_J f(q) - \overline{S}_J h(q)|^2 \, d\mu(q) = \sum_{p \in \mathcal{P}_J} \big| \|S_J[p]f\| - \|S_J[p]h\| \big|^2. \tag{3.7}$$

The following proposition proves that $\overline{S}_J$ is a nonexpansive operator that preserves the norm.

PROPOSITION 3.3. *For all* $(f,h) \in \mathbf{L}^2(\mathbb{R}^d)^2$ *and* $J \in \mathbb{Z}$

$$\|\overline{S}_J f - \overline{S}_J h\|_{\overline{\mathcal{P}}_\infty} \leq \|\overline{S}_{J+1} f - \overline{S}_{J+1} h\|_{\overline{\mathcal{P}}_\infty}, \tag{3.8}$$

$$\|\overline{S}_J f - \overline{S}_J h\|_{\overline{\mathcal{P}}_\infty} \leq \|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\| \leq \|f-h\|, \tag{3.9}$$

$$\|\overline{S}_J f\|_{\overline{\mathcal{P}}_\infty} = \|f\|. \tag{3.10}$$

PROOF. We proved in (2.39) that

$$\|S_J[p]f\|^2 = \sum_{p' \in \mathcal{P}^p_{J+1}} \|S_{J+1}[p']f\|^2, \tag{3.11}$$

where $\mathcal{P}_{J+1} = \bigcup_{p \in \mathcal{P}_J} \mathcal{P}_{J+1}^p$ is a disjoint partition. Applying this to f and h implies

$$\big| \|S_J[p]f\| - \|S_J[p]h\| \big|^2 \le \sum_{p' \in \mathcal{P}_{J+1}^p} \big| \|S_{J+1}[p']f\| - \|S_{J+1}[p']h\| \big|^2.$$

Summing over $p \in \mathcal{P}_J$ and inserting (3.7) proves (3.8).

Since $|\|S_J[p]f\| - \|S_J[p]h\|| \le \|S_J[p]f - S_J[p]h\|$, summing this inequality over $p \in \mathcal{P}_J$ and inserting (3.7) proves the first inequality of (3.9). The second inequality is obtained because $S_J[\mathcal{P}_J]$ is nonexpansive. Setting $h = 0$ proves that $\|\overline{S}_J f\|_{\overline{\mathcal{P}}_\infty} = \|S_J[\mathcal{P}_J]f\|$, and Theorem 2.6 proves $\|S_J[\mathcal{P}_J]f\| = \|f\|$, which gives (3.10). □

Since $\|\overline{S}_J f - \overline{S}_J h\|_{\overline{\mathcal{P}}_\infty}$ is nondecreasing and bounded when J increases, it converges to a limit that is smaller than the limit of the nonincreasing sequence $\|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\|$. The following proposition proves that $\overline{S}_J f$ converges pointwise to the scattering transform on $\mathcal{P}_\infty$ introduced in Definition 2.3:

PROPOSITION 3.4. *If $f \in \mathbf{L}^1(\mathbb{R}^d)$, then*

$$\forall p \in \mathcal{P}_\infty \quad \lim_{J\to\infty} \overline{S}_J f(p) = \overline{S} f(p) = \frac{1}{\mu_p} \int U[p]f(x)\,dx \tag{3.12}$$

with $\mu_p = \int U[p]\delta(x)\,dx$.

PROOF. If $p \in \mathcal{P}_\infty$, then $\overline{S}_J f(p) = \|S_J[p]f\| / \|S_J[p]\delta\|$ for J sufficiently large. Let us prove that

$$\lim_{J\to\infty} 2^{dJ/2} \|S_J[p]f\| = \|\phi\| \int U[p]f(x)\,dx \tag{3.13}$$

and that this equality also holds for $f = \delta$. Since $S_J[p]f = U[p]f \star \phi_{2^J}$, the Plancherel formula implies

$$2^{dJ} \|S_J[p]f\|^2 = 2^{dJ} (2\pi)^{-d} \int |\widehat{U[p]f}(\omega)|^2 \, |\widehat{\phi}(2^J\omega)|^2 \, d\omega. \tag{3.14}$$

Since derivatives of ϕ are in $\mathbf{L}^1(\mathbb{R}^d)$, we have $\widehat{\phi}(\omega) = O((1+|\omega|)^{-1})$, and hence $(2\pi)^{-d} 2^{dJ} |\widehat{\phi}(2^J\omega)|^2$ converges to $\|\phi\|^2 \, \delta(\omega)$. Moreover, if $f \in \mathbf{L}^1(\mathbb{R}^d)$ then $U[p]f \in \mathbf{L}^1(\mathbb{R}^d)$ so $\widehat{U[p]f}(\omega)$ is continuous at $\omega = 0$. We have from (3.14) that $\lim_{J\to\infty} 2^{dJ} \|S_J[p]f\|^2 = |\widehat{U[p]f}(0)|^2 \, \|\phi\|^2$, which proves (3.13). The same derivations hold to prove this result for $f = \delta$.

Since $|\widehat{\psi}(\omega)| + |\widehat{\psi}(-\omega)| \neq 0$ almost everywhere, Proposition 3.1 proves that $U[p]\delta \neq 0$. Since it is positive, it has a nonzero integral. We have from (3.13) that $\lim_{J\to\infty} \|S_J[p]f\| / \|S_J[p]\delta\| = \int U[p]f(x)\,dx / \int U[p]\delta(x)\,dx$, which proves (3.12). □

The scattering transform $\overline{S} f$ can now be extended to $\overline{\mathcal{P}}_\infty$ as a windowed scattering limit:

$$\forall q \in \overline{\mathcal{P}}_\infty \quad \overline{S} f(q) = \liminf_{J\to\infty} \overline{S}_J f(q).$$

Proposition 3.3 proves that $\|\overline{S}_J f\|_{\overline{\mathcal{P}}_\infty} = \|f\|$, so Fatou's lemma implies that $\overline{S} f \in \mathbf{L}^2(\overline{\mathcal{P}}_\infty, d\mu)$. The following theorem gives a sufficient condition so that $\overline{S}_J f$ converges strongly to $\overline{S} f$, which then preserves the $\mathbf{L}^2(\mathbb{R}^d)$ norm of f.

THEOREM 3.5. *If for $f \in \mathbf{L}^2(\mathbb{R}^d)$ there exists $\Omega_J^f \subset \mathcal{P}_J$ with*

$$\lim_{J\to\infty} \|S_J[\Omega_J^f] f\|^2 = 0 \quad \textit{and}$$
$$\lim_{J\to\infty} \sup_{p\in\mathcal{P}_J-\Omega_J^f} \left\| \frac{S_J[p]f}{\|S_J[p]f\|} - \frac{S_J[p]\delta}{\|S_J[p]\delta\|} \right\| = 0, \tag{3.15}$$

then $\lim_{J\to\infty} \|\overline{S}_J f - \overline{S} f\|_{\overline{\mathcal{P}}_\infty} = 0$ *with* $\|\overline{S} f\|_{\overline{\mathcal{P}}_\infty} = \|f\|$ *and*

$$\forall p \in \mathcal{P}_\infty \quad \int_{C(p)} |\overline{S} f(q)|^2 \, d\mu(q) = \|U[p]f\|^2. \tag{3.16}$$

If $(f, h) \in \mathbf{L}^2(\mathbb{R}^d)^2$ satisfy (3.15)*, then*

$$\lim_{J\to\infty} \|S_J[\mathcal{P}_J] f - S_J[\mathcal{P}_J]h\| = \|\overline{S} f - \overline{S} h\|_{\overline{\mathcal{P}}_\infty}. \tag{3.17}$$

If (3.15) *is satisfied in a dense subset of $\mathbf{L}^2(\mathbb{R}^d)$, then $S_J f$ converges strongly to $\overline{S} f$ for all $f \in \mathbf{L}^2(\mathbb{R}^d)$, and both* (3.16) *and* (3.17) *are satisfied in $\mathbf{L}^2(\mathbb{R}^d)$.*

PROOF. The following lemma proves that $\{\overline{S}_J f\}_{J\in\mathbb{N}}$ is Cauchy and hence converges in norm to $\overline{S} f \in \mathbf{L}^2(\overline{\mathcal{P}}_\infty, d\mu)$. The proof is in Appendix F.

LEMMA 3.6. *If $f \in \mathbf{L}^2(\mathbb{R}^d)$ satisfies* (3.15)*, then $\{\overline{S}_J f\}_{J\in\mathbb{N}}$ is a Cauchy sequence in $\mathbf{L}^2(\overline{\mathcal{P}}_\infty, d\mu)$.*

Since $\mathbf{L}^2(\overline{\mathcal{P}}_\infty, d\mu)$ is complete, $\overline{S}_J f(q)$ converges in norm to its limit inf $\overline{S} f$. Since $\|\overline{S}_J f\| = \|f\|$, it also implies that $\|\overline{S} f\|_{\overline{\mathcal{P}}_\infty} = \|f\|$. Also, $U[p+q] = U[q]U[p]$, so

$$\|\overline{S}_J U[p] f\|^2_{\overline{\mathcal{P}}_\infty} = \int_{C(p)} |\overline{S}_J f(q)|^2 \, d\mu(q).$$

Since $\|\overline{S}_J U[p] f\|^2_{\overline{\mathcal{P}}_\infty} = \|U[p]f\|^2$, taking the limit when J goes to ∞ proves (3.16).

The windowed scattering convergence (3.17) relies on the following lemma:

LEMMA 3.7. *If $(f, h) \in \mathbf{L}^2(\mathbb{R}^d)^2$ satisfy* (3.15)*, then*

$$\lim_{J\to\infty} \|S_J[\mathcal{P}_J] f - S_J[\mathcal{P}_J]h\| = \lim_{J\to\infty} \|\overline{S}_J f - \overline{S}_J h\|_{\overline{\mathcal{P}}_\infty}. \tag{3.18}$$

PROOF. Since (3.15) implies that $\overline{S}_J f$ and $\overline{S}_J h$, respectively, converge in norm to $\overline{S} f$ and $\overline{S} h$, the convergence (3.17) results from (3.18). Proving (3.18) is equivalent to proving that $\lim_{J\to\infty} \sum_{p\in\mathcal{P}_J} I_J(f,h)[p] = 0$ for

$$I_J(f,h)[p] = \|S_J[p]f - S_J[p]h\|^2 - \big|\|S_J[p]f\| - \|S_J[p]h\|\big|^2.$$

Observe that

$$\begin{aligned} I_J(f,h)[p] &= \|S_J[p]f\|\,\|S_J[p]h\| \left\| \frac{S_J[p]f}{\|S_J[p]f\|} - \frac{S_J[p]h}{\|S_J[p]h\|} \right\|^2 \\ &\leq 2\|S_J[p]f\|\,\|S_J[p]h\| \left(\left\| \frac{S_J[p]f}{\|S_J[p]f\|} - \frac{S_J[p]\delta}{\|S_J[p]\delta\|} \right\|^2 \right. \\ &\qquad \left. + \left\| \frac{S_J[p]h}{\|S_J[p]h\|} - \frac{S_J[p]\delta}{\|S_J[p]\delta\|} \right\|^2 \right). \end{aligned} \tag{3.19}$$

When summing over $p \in \mathcal{P}_J$, we separate $\Omega_J^f \cup \Omega_J^h$ from its complement in $\mathcal{P}_J$. Since we have

$$\lim_{J\to\infty} \|S_J[\Omega_J^f]f\|^2 = 0, \quad \|S_J[\mathcal{P}_J]f\|^2 = \|f\|^2,$$
$$\lim_{J\to\infty} \|S_J[\Omega_J^h]h\|^2 = 0, \quad \|S_J[\mathcal{P}_J]h\|^2 = \|h\|^2,$$

dividing the sum over Ω_J^f and Ω_J^h and applying Cauchy-Schwarz proves that

$$\lim_{J\to\infty} \sum_{p\in\Omega_J^f\cup\Omega_J^h} \|S_J[p]f\|\,\|S_J[p]h\| = 0$$

and $\sum_{p\in\mathcal{P}_J} \|S_J[p]f\|\,\|S_J[p]h\| \leq \|f\|\,\|h\|$. The hypothesis (3.15) applied to f and h gives

$$\lim_{J\to\infty} \sup_{p\in\mathcal{P}_J-\Omega_J^f\cup\Omega_J^h} \left(\left\| \frac{S_J[p]f}{\|S_J[p]f\|} - \frac{S_J[p]\delta}{\|S_J[p]\delta\|} \right\|^2 + \left\| \frac{S_J[p]h}{\|S_J[p]h\|} - \frac{S_J[p]\delta}{\|S_J[p]\delta\|} \right\|^2 \right) = 0$$

so (3.19) implies that $\lim_{J\to\infty} \sum_{p\in\mathcal{P}_J} I_J(f,h)[p] = 0$, which finishes the proof of the lemma. □

Suppose that (3.15) is satisfied in a dense subset of $\mathbf{L}^2(\mathbb{R}^d)$. Any $f \in \mathbf{L}^2(\mathbb{R}^d)$ is the limit of $\{f_n\}_{n>0}$ in this dense set. Since $\overline{S}$ and $\overline{S}_J$ are nonexpansive,

$$\|\overline{S} f - \overline{S}_J f\|_{\overline{\mathcal{P}}_\infty} \leq 2\|f - f_n\| + \|\overline{S} f_n - \overline{S}_J f_n\|_{\overline{\mathcal{P}}_\infty}.$$

Since f_n satisfies (3.15), we proved that $\overline{S}_J f_n$ converges in norm to $\overline{S} f_n$. Letting n go to ∞ implies that $\overline{S}_J f$ converges in norm to $\overline{S}$. The previous derivations then imply that both (3.16) and (3.17) are satisfied in $\mathbf{L}^2(\mathbb{R}^d)$. □

If $f \in \mathbf{L}^1(\mathbb{R}^d)$ and $p \in \mathcal{P}_\infty$, applying the Plancherel formula proves that

$$\lim_{J\to\infty} \left\| \frac{S_J[p]f}{\|S_J[p]f\|} - \frac{S_J[p]\delta}{\|S_J[p]\delta\|} \right\|^2 = 0 \tag{3.20}$$

since $S_J[p]f(x) = U[p]f \star \phi_{2^J}$ and $\|U[p]f\|_1 < \infty$. This is, however, not sufficient to prove (3.15) because the sup is taken over all $p \in \mathcal{P}_J - \Omega_J^f$, which grows when J increases. For $f \in \mathbf{L}^1(\mathbb{R}^d)$, one can find paths $p_J \in \mathcal{P}_J$, which are not frequency decreasing, where $S_J[p_J]f/\|S_J[p_J]f\|$ does not converge to $S_J[p_J]\delta/\|S_J[p_J]\delta\|$. The main difficulty is to prove that over the set Ω_J^f of all such paths, a windowed scattering transform has a norm $\|S_J[\Omega_J^f]f\|$ that converges to 0. Numerical experiments indicate that this property could be valid for all $f \in \mathbf{L}^1(\mathbb{R}^d)$. It also seems that if $f \in \mathbf{L}^1(\mathbb{R}^d)$, then $\overline{S}f(q)$ is a continuous function of the path q relative to the Dirac scattering metric. This is analogous to Fourier transform continuity when $f \in \mathbf{L}^1(\mathbb{R}^d)$.

CONJECTURE 3.8. *Condition* (3.15) *holds for all* $f \in \mathbf{L}^1(\mathbb{R}^d)$. *Moreover, if* $f \in \mathbf{L}^1(\mathbb{R}^d)$, *then* $\overline{S}f(q)$ *is continuous in* $\overline{\mathcal{P}}_\infty$ *relative to the Dirac scattering metric.*

If this conjecture is valid, since $\mathbf{L}^1(\mathbb{R}^d)$ is dense in $\mathbf{L}^2(\mathbb{R}^d)$, then Theorem 3.5 proves that $\overline{S}_J$ converges strongly to $\overline{S}f$ for all $f \in \mathbf{L}^2(\mathbb{R}^d)$ and $\|\overline{S}f\|_{\overline{\mathcal{P}}_\infty} = \|f\|$. In addition, property (3.17) proves that $\|S_J[\mathcal{P}_J]f - S_J[\mathcal{P}_J]h\|$ converges to $\|\overline{S}f - \overline{S}h\|_{\overline{\mathcal{P}}_\infty}$ as J goes to ∞. Through this limit, the Lipschitz continuity of S_J under the action of diffeomorphisms can then be extended to the scattering transform $\overline{S}$.

3.3 Numerical Comparisons with Fourier

Let $\mathbb{R}^{d+}$ be the half frequency space of all $\omega = (\omega_1, \omega_2, \dots, \omega_d) \in \mathbb{R}^d$ with $\omega_1 \geq 0$ and $\omega_k \in \mathbb{R}$ for $k > 1$. To display numerical examples for real functions, the following proposition constructs a function from $\mathbb{R}^{d+}$ to $\overline{\mathcal{P}}_\infty$ that maps the Lebesgue measure of $\mathbb{R}^{d+}$ into the Dirac scattering measure. It provides a representation of $\overline{S}f$ over $\mathbb{R}^{d+}$. We assume that ψ is an admissible scattering wavelet and that $|\widehat{\psi}(\omega)| + |\widehat{\psi}(-\omega)| \neq 0$ almost everywhere.

PROPOSITION 3.9. *There exists a surjective function* $q(\omega)$ *from* $\mathbb{R}^{d+}$ *onto* $\overline{\mathcal{P}}_\infty$ *such that for all measurable sets* $\Omega \subset \overline{\mathcal{P}}_\infty$

$$\mu(\Omega) = \int_{q^{-1}(\Omega)} d\omega. \tag{3.21}$$

PROOF. The proof first constructs the inverse q^{-1} by mapping each cylinder $C(p)$ for $p \in \mathcal{P}_\infty$ into a set $q^{-1}(C(p)) \subset \mathbb{R}^{d+}$ satisfying the following properties: $\mu(C(p)) = \int_{q^{-1}(C(p))} d\omega$ and $q^{-1}(C(p)) \cap q^{-1}(C(p')) = \varnothing$ if $C(p) \cap C(p') = \varnothing$, and $q^{-1}(C(p)) \subset q^{-1}(C(p'))$ if $C(p) \subset C(p')$. Let $\overline{q^{-1}(C(p))}$ be the closure of $q^{-1}(C(p))$ in $\mathbb{R}^{d+}$. For all $p \neq \varnothing$, we also impose that the frontier of $q^{-1}(C(p))$ is a set of measure 0 in $\mathbb{R}^{d+}$ and that $\overline{q^{-1}(C(p+\lambda))} \subset q^{-1}(C(p))$ for all $\lambda \in \Lambda_\infty$. The cylinders $C(p)$ generate the sigma algebra on which the measure μ is defined. A measurable set Ω can be approximated by sets Ω_k that are the union of disjoint cylinder sets $C(p)$ with $\lim_{k\to\infty}\mu(\Omega-\Omega_k)=0$. The properties of q^{-1} on the cylinders $C(p)$ imply that $\int_{q^{-1}(\Omega_k)} d\omega = \mu(\Omega_k)$, and when k goes to ∞ we get (3.21).

Once all $q^{-1}(C(p))$ are constructed, the inverse $q(\omega)$ is uniquely defined for all $\omega \in \mathbb{R}^{d+}$ as follows: Let p_m be the prefix of $\overline{q} \in \overline{\mathcal{P}}_\infty$ of length m. We define $q^{-1}(\overline{q}) = \bigcap_{m\in\mathbb{N}} q^{-1}(C(p_m))$. Since $\overline{q^{-1}(C(p+\lambda))} \subset q^{-1}(C(p))$ for all $\lambda \in \Lambda_\infty$, we have that $\bigcap_{m\in\mathbb{N}} q^{-1}(C(p_m)) = \bigcap_{m\in\mathbb{N}} \overline{q^{-1}(C(p_m))}$. It is a closed nonempty set because $\overline{q^{-1}(C(p_m))} \subset \overline{q^{-1}(C(p_{m-1}))}$ is a nonempty set of measure $\|U[p_m]\delta\| \neq 0$. We verify that $q(\omega) = \overline{q}$ for all $\omega \in q^{-1}(\overline{q})$ defines a surjective function on $\mathbb{R}^{d+}$ by showing that $\bigcup_{\overline{q}\in\overline{\mathcal{P}}_\infty} q^{-1}(\overline{q})$ is a partition of $\mathbb{R}^{d+}$. If $\mathcal{P}_m$ is the set of all paths of length m, then $\bigcup_{p\in\mathcal{P}_m} C(p)$ is a partition of $\overline{\mathcal{P}}_\infty$, so the recursive construction of q^{-1} implies that $\bigcup_{p\in\mathcal{P}_m} q^{-1}(C(p))$ is a partition of $\mathbb{R}^{d+}$. Letting m go to infinity proves that $\bigcup_{\overline{q}\in\overline{\mathcal{P}}_\infty} q^{-1}(\overline{q})$ is a partition of $\mathbb{R}^{d+}$.

The sets $q^{-1}(C(p))$ satisfying the previously mentioned properties are defined recursively on the path length, with a subdivision procedure. In dimension $d=1$, each $q^{-1}(C(p))$ is recursively defined as an interval of $\mathbb{R}^+$. We begin with paths $p = 2^j$ of length 1 by defining $q^{-1}(C(2^j)) = [2^j\|\psi\|^2, 2^{j+1}\|\psi\|^2)$, whose width is $2^j\|\psi\|^2 = \mu(C(2^j))$. Suppose now that $q^{-1}(C(p))$ is an interval of width equal to $\mu(C(p))$. All $q^{-1}(C(p+2^j))$ for $j \in \mathbb{Z}$ are defined as consecutive intervals $[a_j, a_{j-1})$ that define a partition of $q^{-1}(C(p)) = \bigcup_{j\in\mathbb{Z}}[a_j, a_{j-1})$ with $a_{j-1} - a_j = \|U[p+2^j]\delta\|^2 = \mu(C(p+2^j))$. One can verify that this recursive construction defines intervals $q^{-1}(C(p))$ that satisfy all mentioned properties. Moreover, in this case the resulting function $q(\omega)$ is bijective from $\mathbb{R}^+$ to $\overline{\mathcal{P}}_\infty$.

In higher dimensions $d \geq 1$, this construction is extended as follows: All cylinders $C(\lambda)$ for all paths $p = \lambda = 2^j r$ of length 1 are mapped to nonintersecting hyperrectangles $q^{-1}(C(\lambda))$ of measure

$$\int_{q^{-1}(C(2^j r))} d\omega = \mu(C(2^j r)) = \|U[2^j r]\delta\|^2 = 2^{dj}\|\psi\|^2.$$

These hyperrectangles are chosen to define a partition of $\mathbb{R}^{d+}$, and hence $\mathbb{R}^{d+} = \bigcup_{\lambda\in\Lambda_\infty} q^{-1}(C(\lambda))$ with $q^{-1}(C(\lambda)) \cap q^{-1}(C(\lambda')) = \varnothing$ for $\lambda \neq \lambda'$. Suppose now that $q^{-1}(C(p))$, with $\int_{q^{-1}(C(p))} dx = \|U[p]\delta\|^2$, is defined for all paths p of length m. The operator U preserves the norm $\sum_{\lambda\in\Lambda_\infty}\|U[p+\lambda]\delta\|^2 = \|U[p]\delta\|^2$. We can thus partition $q^{-1}(C(p))$ into subsets $\{q^{-1}(C(p+\lambda))\}_{\lambda\in\Lambda_\infty}$ with

$$\int_{q^{-1}(C(p+\lambda)} d\omega = \|U[p+\lambda]\delta\|^2,$$

and with frontiers that are zero-measure piecewise hyperplanes of dimension $d-1$.

The property $\overline{q^{-1}(C(p+\lambda))} \subset q^{-1}(C(p))$ for all $\lambda \in \Lambda_\infty$ is obtained with a progressive packing strategy. We first construct $q^{-1}(C(p+\lambda))$ for all $\lambda = 2^j r$

with $j \geq 0$ by defining a partition of a closed subset of $q^{-1}(C(p))$ of measure $\sum_{\lambda\in\Lambda_\infty,|\lambda|\geq 1}\|U[p+\lambda]\delta\|^2$. The remaining $q^{-1}(C(p+\lambda))$ are then progressively constructed for $\lambda = 2^j r$ and j going from -1 to $-\infty$, within the remaining closed subset of $q^{-1}(C(p))$ not already allocated. This is possible since we guarantee that the frontier of each $q^{-1}(C(p))$ has a zero measure. □

The function $q(\omega)$ maps the Lebesgue measure into the Dirac scattering measure, but it is discontinuous at all $\omega \in \mathbb{R}^{d+}$ such that $q(\omega) \in \mathcal{P}_\infty$. Indeed, these ω are then at a boundary of the subdivision procedure used to construct $q(\omega)$. As a result, if ω and ω' are on opposite sides of a subdivision boundary, then they are mapped to paths $q(\omega)$ and $q(\omega')$ whose distance $\bar{d}(q(\omega), q(\omega'))$ does not converge to 0 as $|\omega - \omega'|$ goes to 0.

Measure preservation (3.21) implies that $q(\omega)$ defines a function $\overline{S}f(q(\omega))$, which belongs to $\mathbf{L}^2(\mathbb{R}^{d+})$ and

$$\|\overline{S}f(q(\omega))\|^2_{\mathbb{R}^{d+}} = \int_{\mathbb{R}^{d+}} |\overline{S}f(q(\omega))|^2\,d\omega = \int_{\bar{\mathcal{P}}_\infty} |\overline{S}f(q)|^2\,d\mu(q) = \|\overline{S}f\|^2_{\bar{\mathcal{P}}_\infty}.$$

If f is a complex-valued function, then $\bar{\mathcal{P}}_\infty$ is a union of positive paths $q = (\lambda_1, \lambda_2, \dots)$ and negative paths $-q = (-\lambda_1, \lambda_2, \dots)$. Setting $q(-\omega) = -q(\omega)$ defines a surjective function from $\mathbb{R}^d$ to $\bar{\mathcal{P}}_\infty$ that satisfies (3.21). We have that $\overline{S}f(q(-\omega)) = \overline{S}f(-q(\omega))$ for all $\omega \in \mathbb{R}^d$, and $\overline{S}f(q(\omega)) \in \mathbf{L}^2(\mathbb{R}^d)$ with $\|\overline{S}f(q(\omega))\| = \|\overline{S}f\|_{\bar{\mathcal{P}}_\infty}$.

If f satisfies (3.15), then $\overline{S}f(q(\omega))$ and $|\hat{f}(\omega)|$ have an equivalent decay over dyadic frequency bands, because their norm is equal over these frequency bands. Indeed, for a frequency band $\lambda = 2^j r$ of radius proportional to $|\lambda| = 2^j$, the measure preservation (3.21) together with (3.16) proves that $\|U[\lambda]f\| = \|f \star \psi_\lambda\|$ satisfies

$$\int_{q^{-1}(C(\lambda))} |\overline{S}f(q(\omega))|^2\,d\omega = \|U[\lambda]f\|^2 = \frac{1}{2\pi}\int |\hat{f}(\omega)|^2\,|\hat{\psi}(\lambda^{-1}\omega)|^2\,d\omega. \tag{3.22}$$

If Conjecture 3.8 is valid, then this is true for all $f \in \mathbf{L}^2(\mathbb{R}^d)$. In dimension $d = 1$, $q^{-1}(C(2^j)) = [\|\psi\|^2 2^j, \|\psi\|^2 2^{j+1})$ and $|\hat{\psi}(2^j\omega)|$ is nonnegligible on a similar dyadic frequency interval. Hence $\overline{S}f(q(\omega))$ and $|\hat{f}(\omega)|$ have equivalent energy over dyadic frequency intervals.

Figure 3.1(b,c,d) shows the convergence of the piecewise constant $\overline{S}_J f(q(\omega))$ when J increases, for a Gaussian second derivative f. $\overline{S}_J f(q(\omega))$ is constant if $q(\omega) = p$ is constant and hence if $\omega \in q^{-1}(C_J(p))$. The frequency interval

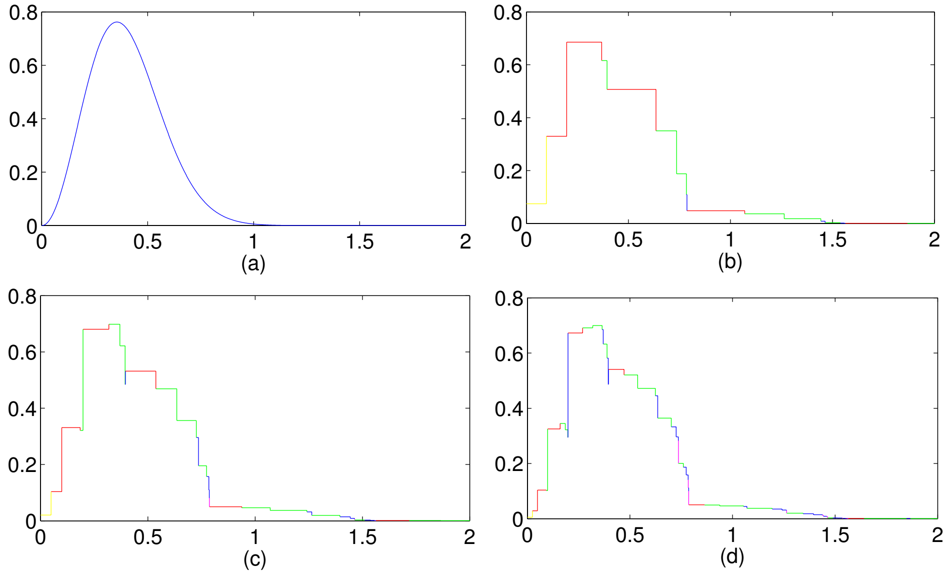

FIGURE 3.1. (a) Fourier modulus $|\hat{f}(\omega)|$ of a Gaussian second derivative as a function of $\omega \in [0,2]$. (b,c,d) Piecewise constant graphs of $\overline{S}_J f(q(\omega))$ as a function of $\omega \in [0,2]$. The color specifies the length of each path $q(\omega)$: 0 is yellow, 1 red, 2 green, 3 blue, 4 violet. The frequency resolution 2^J increases from (b) to (c) to (d), and $\overline{S}_J f(q(\omega))$ converges to a limit function $\overline{S} f(q(\omega))$.

$q^{-1}(C_J(p))$ has a width $\mu(C_J(p)) = \|S_J \delta[p]\|^2$, which goes to 0 as J goes to ∞ as shown by (3.2). When J increases, each $q^{-1}(C_J(p))$ is subdivided into smaller intervals $q^{-1}(C_{J+1}(p'))$ corresponding to paths p that are prolongations of p. For each ω, the graph color specifies the length of the path $p = q(\omega)$. At low frequencies, $q(\omega) = \varnothing$ is shown as a yellow interval. Paths $q(\omega)$ of length 1 to 4 are, respectively, coded in red, green, blue, and violet.

In these numerical examples, the total energy of $\overline{S}_J f(q(\omega))$ on frequency-decreasing paths $q(\omega)$ is about 10^5 times larger than the energy of scattering coefficients on all other paths. We thus only compute $\overline{S}_J f(q(\omega))$ for frequency-decreasing paths, with an $O(N \log N)$ filter bank algorithm described in [13]. It is implemented with the complex cubic spline Battle-Lemarié wavelet ψ. As expected from (3.22), $\overline{S}_J f(q(\omega)) f$ has an amplitude and a frequency localization that is similar to the Fourier modulus $|\hat{f}(\omega)|$ shown in Figure 3.1(a). The discontinuities of $\overline{S}(q(\omega)) f$ along ω are produced by the discontinuities of the mapping $q(\omega)$, as opposed to discontinuities of $\overline{S}(q) f$ relative to the scattering metric in $\overline{\mathcal{P}}_\infty$.

Figure 3.2 compares $\overline{S}(q(\omega)) f_i$ and $|\hat{f}_i(\omega)|$ for four functions f_i, $1 \leq i \leq 4$. For $f_1 = 1_{[0,1]}$, the first row of Figure 3.2 shows that $|\hat{f}_1(\omega)| = O((1 + |\omega|)^{-1})$

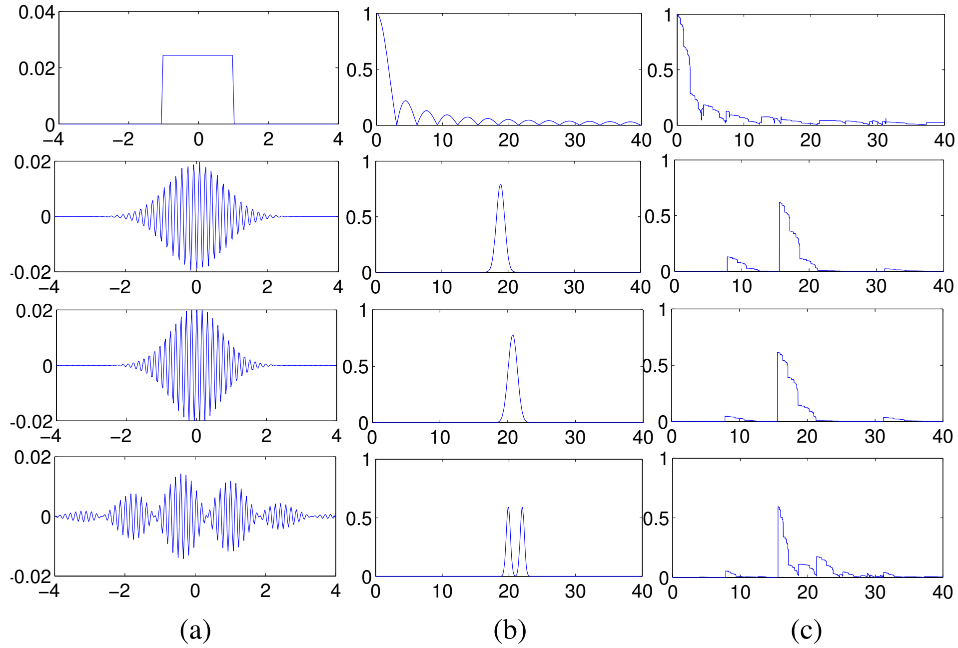

FIGURE 3.2. (a) Each row $1 \le i \le 4$ gives an example of function $f_i(x)$. (b) Graphs of the Fourier modulus $|\widehat{f}_i(\omega)|$ as a function of ω. (c) Graphs of the scattering $\overline{S} f_i(q(\omega))$ as a function of ω.

has the same decay in ω as $\overline{S} f_1(q(\omega))$. The second row corresponds to a Gabor function $f_2(x) = e^{i\xi x} e^{-x^2/2}$, and the third row shows a small scaling $f_3(x) = f_2((1-s)x)$ with $s = -0.1$. The support of $\widehat{f}_3(\omega) = (1-s)^{-1} \widehat{f}_2((1-s)^{-1}\omega)$ is shifted towards higher frequencies relative to the support of $\widehat{f}_2$. A numerical computation gives $\||\widehat{f}_2| - |\widehat{f}_3|\| = C|s| \|f_2\|$ with $C = 13.5$. As shown by (2.1), the constant C grows proportionally to the center frequency ξ of $\widehat{f}_2$. It illustrates the instability of the Fourier modulus under the action of diffeomorphisms.

On the contrary, the scattering distance remains stable. We numerically obtain $\|\overline{S} f_2 - \overline{S} f_3\| = C|s| \|f_2\|$ with $C = 1.5$, and this constant does not grow with ξ. It illustrates the Lipschitz continuity of a scattering relative to deformations. In the fourth row, f_4 is a sum of two high-frequency Gabor functions, and $|\widehat{f}_4(\omega)|$ includes two narrow peaks localized within the support of $\widehat{f}_3$. The wavelet transform has a bad frequency localization at such high frequencies and cannot discriminate the two frequency peaks of $\widehat{f}_4$ from $\widehat{f}_3$. However, these two frequency peaks create low-frequency interferences, which appear in the graph of f_4, and which are captured by second-order scattering coefficients. As a result, $\overline{S} f_4$ is very different from $\overline{S} f_3$, which illustrates the high-frequency resolution of a scattering transform obtained through interferences.

4 Scattering Stationary Processes

A scattering defines a representation of stationary processes in $\mathbf{l}^2(\mathcal{P}_\infty)$ having different properties from a Fourier power spectrum. The Fourier power spectrum depends only on second-order moments. A scattering transform incorporates higher-order moments that can discriminate processes having the same second-order moments. Section 4.2 shows that it is Lipschitz-continuous to random deformations up to a log term.

4.1 Expected Scattering

The properties of a scattering transform in $\mathbf{L}^2(\mathbb{R}^d)$ are extended to stationary processes $X(x)$ with finite second-order moments. The role of the $\mathbf{L}^2(\mathbb{R}^d)$ norm on functions is replaced by the mean-square norm $E(|X(x)|^2)^{1/2}$ on stationary stochastic processes, which does not depend upon x and is thus denoted $E(|X|^2)^{1/2}$. Convolutions as well as a modulus preserve stationarity. If $X(x)$ is stationary, we have that $U[p]X(x)$ is also stationary, and its expected value thus does not depend upon x.

DEFINITION 4.1. The expected scattering transform of a stationary process X is defined for all $p = (\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathcal{P}_\infty$ by

$$\overline{S}X(p) = E(U[p]X) = E(|\,|X \star \psi_{\lambda_1}| \star \cdots | \star \psi_{\lambda_m}|).$$

This definition replaces the normalized integral of the scattering transform (2.16) by an expected value. The expected scattering distance between two stationary processes X and Y is

$$\|\overline{S}X - \overline{S}Y\|^2 = \sum_{p \in \mathcal{P}_\infty} |\overline{S}X(p) - \overline{S}Y(p)|^2.$$

Scattering coefficients depend upon normalized high-order moments of X. This is shown by decomposing

$$|U[p]X(x)|^2 = E(|U[p]X|^2)(1 + \epsilon(x)).$$

A first-order approximation assumes that $|\epsilon| \ll 1$. Since $\int \psi_\lambda(x)dx = 0$ and $U[p]X(x) = \sqrt{|U[p]X(x)|^2}$, computing $U[p+\lambda]X = |U[p]X \star \psi_\lambda|$ with $\sqrt{1+\epsilon} \approx 1 + \epsilon/2$ gives

$$U[p+\lambda]X \approx \frac{||U[p]X|^2 \star \psi_\lambda|}{2\,E(|U[p]X|^2)^{1/2}}. \tag{4.1}$$

Iterating on (4.1) proves that $\overline{S}X(p) = E(U[p]X)$ for $p = (\lambda_1, \lambda_2, \dots, \lambda_m)$ depends on normalized moments of X of order 2^m, successively filtered by the wavelets ψ_{λ_k} for $1 \leq k \leq m$.

The expected scattering transform is estimated by computing a windowed scattering transform of a realization $X(x)$:

$$S_J[\mathcal{P}_J]X = \{S_J[p]X\}_{p \in \mathcal{P}_J} \quad \text{with } S_J[p]X = U[p]X \star \phi_{2^J}.$$

Since $\int \phi_{2^J}(x)dx = 1$, we have that $E(S_J[p]X) = E(U[p]X) = \overline{S}X(p)$. So $S_J[\mathcal{P}_J]X$ is an unbiased estimator of $\{\overline{S}X(p)\}_{p\in\mathcal{P}_J}$.

The autocovariance of a real stationary process X is denoted

$$RX(\tau) = E\big((X(x) - E(X))(X(x-\tau) - E(X))\big).$$

Its Fourier transform $\widehat{R}X(\omega)$ is the power spectrum of X. The mean-square norm of $S_J[\mathcal{P}_J]X = \{S_J[p]X\}_{p\in\mathcal{P}_J}$ is written

$$E(\|S_J[\mathcal{P}_J]\,X\|^2) = \sum_{p\in\mathcal{P}_J} E(|S_J[p]X|^2).$$

The following proposition proves that $S_J[\mathcal{P}_J]X$ and $\overline{S}X$ are nonexpansive and that $\overline{S}X \in \mathbf{l}^2(\mathcal{P}_\infty)$. The wavelet ψ is assumed to satisfy the Littlewood-Paley condition (2.7).

PROPOSITION 4.2. *If X and Y are finite second-order stationary processes, then*

$$E(\|S_J[\mathcal{P}_J]X - S_J[\mathcal{P}_J]Y\|^2) \le E(|X-Y|^2), \tag{4.2}$$

$$\|\overline{S}X - \overline{S}Y\|^2 \le E(|X-Y|^2), \tag{4.3}$$

$$\|\overline{S}X\|^2 \le E(|X|^2). \tag{4.4}$$

PROOF. We first show that $W_J X = \{A_J X, (W[\lambda]X)_{\lambda\in\Lambda_J}\}$ is unitary over stationary processes. Let us denote

$$E(\|W_J X\|^2) = E(|A_J X|^2) + \sum_{\lambda\in\Lambda_J} E(|W[\lambda]X|^2).$$

Both $A_J X = X \star \phi_{2^J}$ and $W[\lambda]X = X \star \psi_\lambda$ are stationary. Since $\int \phi_{2^J}(x)dx = 1$ and $\int \psi_\lambda(x)dx = 0$, we have that $E(A_J X) = E(X)$ and $E(W[\lambda]X) = 0$. Since the power spectrum of $A_J X$ and $W[\lambda]X$ is, respectively, $\widehat{R}X(\omega)|\hat{\phi}(2^J\omega)|^2$ and $\widehat{R}X(\omega)|\hat{\psi}_\lambda(\omega)|^2$, we get

$$E(|A_J X|^2) = \int \widehat{R}X(\omega)|\hat{\phi}(2^J\omega)|^2\, d\omega + E(X)^2$$

and

$$E(|W[\lambda]X|^2) = \int \widehat{R}X(\omega)|\hat{\psi}_\lambda(\omega)|^2\, d\omega.$$

Since $E(|X|^2) = \int \widehat{R}X(\omega)d\omega + E(X)^2$, the same proof as in Proposition 2.1 shows that the wavelet condition (2.7) implies that $E(\|W_J X\|^2) = E(|X|^2)$.

The propagator $U_J X = \{A_J X, (|W[\lambda]X|)_{\lambda\in\Lambda_J}\}$ satisfies

$$E(\|U_J X - U_J Y\|^2) \le E(\|W_J X - W_J Y\|^2) = E(|X-Y|^2)$$

and is thus nonexpansive on stationary processes. We verify as in (2.23) that

$$U_J\, U[\Lambda_J^m]X = \{S_J[\Lambda_J^m]X, U[\Lambda_J^{m+1}]X\}.$$

Since $\mathcal{P}_J = \bigcup_{m=0}^{+\infty} \Lambda_J^m$, one can compute $S_J[\mathcal{P}_J]X$ by iteratively applying the nonexpansive operator U_J. The nonexpansive property (4.2) is derived from the fact that U_J is nonexpansive, as in Proposition 2.5.

Let us prove (4.3). Since $\overline{S}X(p) = E(S_J[p]X)$ and $\overline{S}Y(p) = E(S_J[p]Y)$,

$$\sum_{p \in \mathcal{P}_J} |\overline{S}X(p) - \overline{S}Y(p)|^2 \le E(\|S_J[\mathcal{P}_J]X - S_J[\mathcal{P}_J]Y\|^2) \le E(|X - Y|^2).$$

Letting J go to ∞ proves (4.3). The last inequality (4.4) is obtained by setting $Y = 0$. □

Paralleling the scattering norm preservation in $\mathbf{L}^2(\mathbb{R}^d)$, the following theorem proves that $S_J[\mathcal{P}_J]$ preserves the mean-square norm of stationary processes:

THEOREM 4.3. *If the wavelet satisfies the admissibility condition* (2.28) *and if X is stationary with $E(|X|^2) < \infty$, then*

$$E(\|S_J[\mathcal{P}_J]X\|^2) = E(|X|^2). \tag{4.5}$$

PROOF. The proof of (4.5) is almost identical to the proof of (2.29) in Theorem 2.6 if we replace f by X, $|\widehat{f}(\omega)|^2$ by the power spectrum $\widehat{R}X(\omega)$, and $\|f\|^2$ by $E(|X|^2)$. We proved that $E(\|W_J X\|^2) = E(|X|^2)$, so we also have $E(\|U_J X\|^2) = E(|X|^2)$. In the derivations of Lemma 2.8, replacing $f_p = U[p]f$ by $X_p = U[p]X$ and $|\widehat{f}_p(\omega)|^2$ by $\widehat{R}X_p(\omega)$ proves that

$$\frac{\alpha}{2} E(\|U[\mathcal{P}_J]X\|^2) \le \max(J+1, 1) E(|X|^2) + \sum_{j>0} \sum_{r \in G+} j E(|X \star \psi_{2^j r}|^2).$$

Since $\mathcal{P}_J = \bigcup_{m \in \mathbb{N}} \Lambda_J^m$, if the right-hand side term is finite, then

$$\lim_{m \to \infty} E(\|U[\Lambda_J^m]X\|^2) = 0. \tag{4.6}$$

The same density argument as in the proof of Theorem 2.6 proves that (4.6) also holds if $E(|X|^2) < \infty$ because $\widehat{R}X(\omega)$ is integrable.

Since $E(\|U_J X\|^2) = E(|X|^2)$ and $U_J U[\Lambda_J^m]X = \{S_J[\Lambda_J^m]X, U[\Lambda_J^{m+1}]X\}$, iterating m times on U_J proves as in (2.32) that

$$E(|X|^2) = \sum_{n=0}^{m-1} E(\|S_J[\Lambda_J^n]X\|^2) + E(\|U[\Lambda_J^m]X\|^2).$$

When m goes to ∞, (4.6) implies (4.5). □

A windowed scattering $S_J[p] = U[p]X \star \phi_{2^J}$ averages $U[p]X$ over a domain whose size is proportional to 2^J. If $U[p]X$ is ergodic, it thus converges to $\overline{S}X(p) = E(U[p]X)$ when J goes to ∞. The windowed transformed scattering $S_J[\mathcal{P}_J]X$ is said to be a *mean-square consistent* estimator of $\overline{S}X$ if its total variance over all paths converges to 0:

$$\lim_{J \to \infty} E(\|S_J[\mathcal{P}_J]X - \overline{S}_J X\|^2) = \lim_{J \to \infty} \sum_{p \in \mathcal{P}_J} E(|S_J[p]X - \overline{S}X(p)|^2) = 0.$$

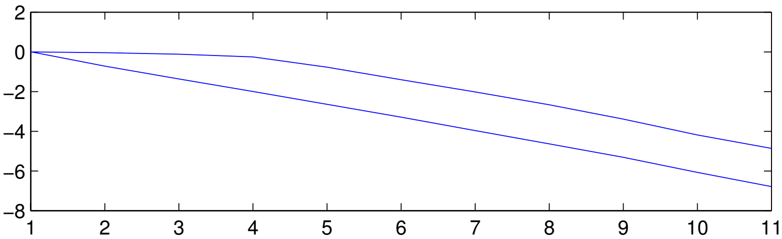

FIGURE 4.1. Decay of $\log_2 E(\|S_J[\mathcal{P}_J]X - \overline{S}X\|^2)$ as a function of J for a Gaussian white noise X (bottom line) and a moving average Gaussian process (top line) along frequency-decreasing paths.

Mean-square convergence implies convergence in probability and therefore that $S_J[\mathcal{P}_J]X$ converges to $\overline{S}X$ with probability 1.

For a large class of ergodic processes X, including Gaussian processes, mean-square convergence is observed numerically, with $E(|S_J[\mathcal{P}_J]X - \overline{S}_J X)|^2) \leq C\, 2^{-\alpha J}$ for $C > 0$ and $\alpha > 0$. When J increases, the global variance of $S_J[\mathcal{P}_J]X$ decreases despite the path subdivision into new paths because each modulus reduces the variance by removing random phase variations. The variance of $S_J[p]X$ thus decreases when the path length increases, and it is concentrated over a small number of frequency-decreasing paths. For a Gaussian white noise and a moving average Gaussian process of unit variance, Figure 4.1 shows that, computed over all frequency-decreasing paths, $\log E(\|S_J[\mathcal{P}_J]X - \overline{S}X\|^2)$ decays linearly as a function J. For the correlated Gaussian process, the decay begins for $2^J \geq 2^4$, which is the correlation length of this process. Indeed, the averaging by ϕ_{2^J} effectively reduces the estimator variance when 2^J is bigger than the correlation length.

CONJECTURE 4.4. *If X is a Gaussian stationary process with $\|RX\|_1 < \infty$, then $S_J[\mathcal{P}_J X]$ is a mean-square consistent estimator of $\overline{S}X$.*

The following corollary of Theorem 4.3 proves that mean-square consistency implies an expected scattering energy conservation.

COROLLARY 4.5. *For an admissible scattering wavelet that satisfies condition* (2.28)*, $S_J[\mathcal{P}_J]X$ is mean-square consistent if and only if*

$$\|\overline{S}X\|^2 = E(|X|^2), \tag{4.7}$$

and mean-square consistency implies that for all $\lambda \in \Lambda_\infty$

$$\sum_{p \in \mathcal{P}_\infty} |\overline{S}X(\lambda + p)|^2 = E(|X \star \psi_\lambda|^2). \tag{4.8}$$

PROOF. We have from Theorem 4.3 that $E(\|S_J[\mathcal{P}_J]X\|^2) = E(|X|^2)$. Since

$$E(\|S_J[\mathcal{P}_J]X\|^2) = \sum_{p \in \mathcal{P}_J} E(S_J[p]X)^2 + E(|S_J[\mathcal{P}_J]X - E(S_J[\mathcal{P}_J]X)|^2)$$

and $E(S_J[p]X) = \overline{S}X(p)$, we derive that

$$\lim_{J\to\infty} E(\|S_J[\mathcal{P}_J]X - E(S_J[\mathcal{P}_J]X)\|^2) = 0$$

if and only if $\|\overline{S}X\|^2 = E(|X|^2)$. Moreover, since $U[p]U[\lambda]X = U[\lambda + p]X$ for all $\lambda \in \Lambda_\infty$, applying (4.7) to $U[\lambda]X$ instead of X proves (4.8). □

The expected scattering can be represented by a singular scattering spectrum in $\overline{\mathcal{P}}_\infty$. Similarly to Section 3.2, we associate to $\overline{S}X(p) = E(U[p]X)$ a function that is piecewise constant in $\overline{\mathcal{P}}_\infty$,

$$\forall q \in \overline{\mathcal{P}}_\infty \quad P_J X(q) = \sum_{p\in\mathcal{P}_J} \overline{S}X(p)^2 \frac{\mathbb{1}_{C_J(p)}(q)}{\|S_J[p]\delta\|^2}. \tag{4.9}$$

The following proposition proves that P_J converges to a singular measure, called a *scattering power spectrum.*

PROPOSITION 4.6. *$P_J X(q)$ converges in the sense of distributions to a Radon measure in $\overline{\mathcal{P}}_\infty$, supported in $\mathcal{P}_\infty$:*

$$PX(q) = \lim_{J\to\infty} P_J X(q) = \sum_{p\in\mathcal{P}_\infty} \overline{S}X(p)^2 \delta(q - p). \tag{4.10}$$

PROOF. For any $p \in \mathcal{P}_\infty$, the Dirac $\delta(p - q)$ is defined as a linear form satisfying $\int_{\overline{\mathcal{P}}_\infty} f(q)\delta(p - q)\,d\mu(q) = f(p)$ for all continuous functions $f(q)$ of $\overline{\mathcal{P}}_\infty$ relative to the scattering metric. For all $J \in \mathbb{Z}$, $\mu(C_J(p)) = \|S_J[p]\delta\|^2$, $p \in C_J(p)$, and $\lim_{J\to\infty} \mu(C_J(p)) = 0$. We thus obtain the following convergence in the sense of distributions:

$$\lim_{J\to\infty} \frac{\mathbb{1}_{C_J(p)}(q)}{\|S_J[p]\delta\|^2} = \delta(q - p).$$

Letting J go to ∞ in (4.9) proves (4.10). □

If $S_J[\mathcal{P}_J]X$ is mean-square consistent, then (4.8) implies that the scattering spectrum $PX(q)$ is related to the Fourier power spectrum $\widehat{R}X(\omega)$ by

$$\int_{C(\lambda)} PX(q)\,d\mu(q) = E(|X \star \psi_\lambda|^2) = \frac{1}{2\pi} \int \widehat{R}X(\omega)|\widehat{\psi}(\lambda^{-1}\omega)|^2\, d\omega. \tag{4.11}$$

Let $q(\omega)$ be the function of Proposition 3.9, which maps the Lebesgue measure of $\mathbb{R}^{d+}$ into the Dirac scattering measure of $\overline{\mathcal{P}}_\infty$. It defines a scattering power spectrum $PX(q(\omega))$ over the half-frequency space $\omega \in \mathbb{R}^{d+}$. In dimension $d = 1$, $q^{-1}(C(2^j)) = [\|\psi\|^2 2^j, \|\psi\|^2 2^{j+1})$, so (4.11) implies

$$\int_{\|\psi\|^2 2^j}^{\|\psi\|^2 2^{j+1}} PX(q(\omega))\,d\omega = \frac{1}{2\pi} \int \widehat{R}X(\omega)|\widehat{\psi}(2^j\omega)|^2\, d\omega.$$

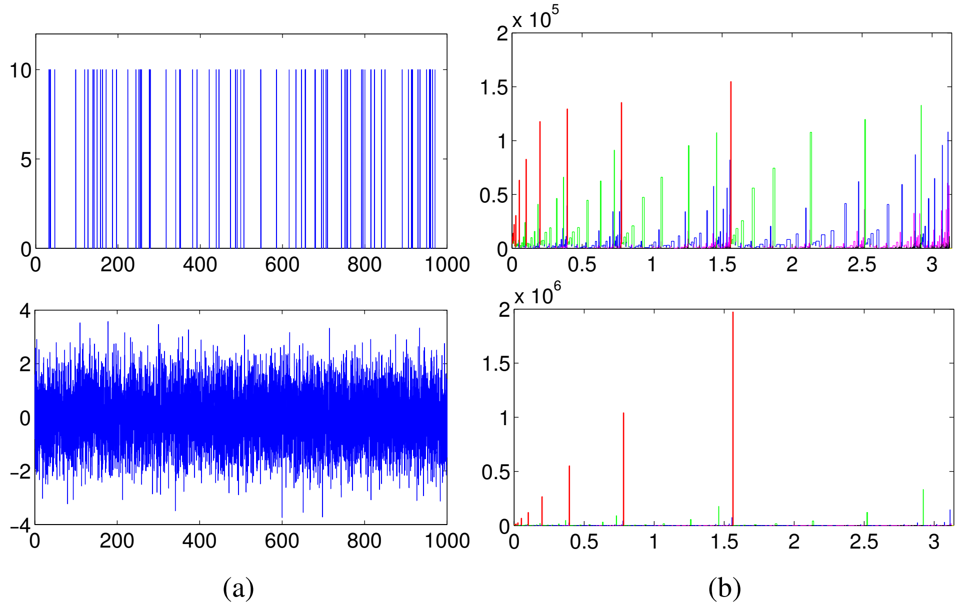

FIGURE 4.2. (a) Realization of a Bernoulli process $X_1(x)$ at the top and a Gaussian white noise $X_2(x)$ at the bottom, both having a unit variance. (b) Scattering power spectrum $PX_i(q(\omega))$ of each process as a function of $\omega \in [0, \pi]$. The values of $PX_i(q(\omega))$ are displayed in red, green, blue, and violet for paths $q(\omega)$ of lengths 1, 2, 3, and 4, respectively.

Although $PX(q(\omega))$ and $\widehat{R}X(\omega)$ have the same integral over dyadic frequency intervals, they have very different distributions within each of these intervals. Indeed, (4.1) shows that if p is of length m, then $E(U[p]X)$ depends upon normalized moments of X of order 2^m. We have that $PX(q(\omega))$ depends upon arbitrarily high-order moments of X, whereas $\widehat{R}X(\omega)$ only depends upon moments of order 2. Hence, $PX(q)$ can discriminate different stationary processes having the same Fourier power spectrum and thus the same second-order moments.

Figure 4.2 gives the scattering power spectrum of a Gaussian white noise X_2 and of a Bernoulli process X_1 in dimension $d = 1$, estimated from a realization sampled over $N = 10^4$ integer points. Both processes have a constant Fourier power spectrum $\widehat{R}X_i(\omega) = 1$ but very different scattering spectra. Their scattering spectrum $PX_i(q(\omega))$ is estimated by $P_J X_i(q(\omega))$ in (4.9) at the maximum scale $2^J = N$. It is a sum of spikes in Figure 4.2(b), which converges to a Radon measure supported in $\mathcal{P}_\infty$ when increasing $2^J = N$. A Gaussian white noise X_2 has a scattering spectrum mostly concentrated on paths $q(\omega) = (2^j)$ of length 1. These scattering coefficients appear as large-amplitude red spikes at dyadic positions in the bottom graph of Figure 4.2(b). Their amplitude is proportional to $\overline{S}X_2(2^j)^2 \sim 2^j$. Other spikes in green correspond to paths $q(\omega) = (\lambda_1, \lambda_2)$ of length 2. They have a much smaller amplitude. Scattering coefficients for paths of length 3 and 4, in blue and violet, are so small that they are not visible. The top of Figure 4.2(b) shows the scattering spectrum $P_J X_1(q(\omega))$ of a Bernouilli process X_1. It has a maximum amplitude for paths $q(\omega)$ of length 1 (in red), but longer paths shown in green, blue, and violet also produce large scattering coefficients as opposed to a Gaussian white noise scattering. Scattering coefficients for paths p of length m depend upon the moments of X up to the order 2^m. For $m > 1$, large scattering coefficients indicate a strongly non-Gaussian behavior of high-order moments.

4.2 Random Deformations

We now show that the scattering transform is nearly Lipschitz-continuous under the action of random deformations. If τ is a random process with $\|\nabla\tau\|_\infty = |\nabla\tau(x)| < 1$, then $x - \tau(x)$ is a random diffeomorphism. If $X(x)$ and $\tau(x)$ are independent stationary processes, then the action of this random diffeomorphism on $X(x)$ defines a randomly deformed process $L_\tau X(x) = X(x-\tau(x))$ that remains stationary.

The following theorem adapts the result of Theorem 2.12 by proving that the scattering distance produced by a random deformation is dominated by a first-order term proportional to $E(\|\nabla\tau\|_\infty^2)$. Let us denote

$$E(\|U[\mathcal{P}_J]X\|_1) := \sum_{m=0}^{+\infty}\Big(\sum_{p\in\Lambda_J^m} E(|U[p]X|^2)\Big)^{1/2}$$

where Λ_J^m is the set of paths $p = (\lambda_k)_{k\le m}$ of length m with $|\lambda_k| < 2^J$.

THEOREM 4.7. *There exists a constant C such that for all independent stationary processes τ and X satisfying $\|\nabla\tau\|_\infty \le \frac{1}{2}$ with probability* 1, *if $E(\|U[\mathcal{P}_J]X\|_1) < \infty$, then*

$$E(\|S_J[\mathcal{P}_J]L_\tau X - S_J[\mathcal{P}_J]X\|^2) \le CE(\|U[\mathcal{P}_J]X\|_1)^2\, K(\tau) \tag{4.12}$$

with

$$K(\tau) = E\bigg(\bigg(2^{-J}\|\tau\|_\infty + \|\nabla\tau\|_\infty\bigg(\log\frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty}\vee 1\bigg) + \|H\tau\|_\infty\bigg)^2\bigg). \tag{4.13}$$

Over the subset $\mathcal{P}_{J,m}$ of paths in $\mathcal{P}_J$ of length strictly smaller than m,

$$E(\|S_J[\mathcal{P}_{J,m}]L_\tau X - S_J[\mathcal{P}_{J,m}]X\|^2) \le CmE(|X|^2)K(\tau). \tag{4.14}$$

PROOF. Similarly to the proof of Theorem 2.12, we decompose

$$\begin{aligned} E(\|S_J[\mathcal{P}_J]L_\tau X - S_J[\mathcal{P}_J]X\|^2) &\le 2E(\|L_\tau S_J[\mathcal{P}_J]X - S_J[\mathcal{P}_J]X\|^2) \\ &\quad + 2E(\|[S_J[\mathcal{P}_J], L_\tau]X\|^2). \end{aligned}$$

Appendix H proves that $E(\|[S_J[\mathcal{P}_J], L_\tau]X\|^2) \leq E(\|U[\mathcal{P}_J]X\|_1)^2 \, B(\tau)$ with

$$B(\tau) = C^2 \, E\biggl(\biggl(\|\nabla\tau\|_\infty\biggl(\log\frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty} \vee 1\biggr) + \|H\tau\|_\infty\biggr)^2\biggr), \tag{4.15}$$

and since

$$E(\|L_\tau S_J[\mathcal{P}_J]X - S_J[\mathcal{P}_J]X\|^2) \leq C^2 \, E(\|U[\mathcal{P}_J]X\|_1)^2 \, E(2^{-J}\|\tau\|_\infty^2), \tag{4.16}$$

we get (4.12). The commutator $[S_J[\mathcal{P}_J], L_\tau]$ and $L_\tau S_J[\mathcal{P}_J] - S_J[\mathcal{P}_J]$ are random operators since τ is a random process. The key argument of the proof is provided by the following lemma, which relates the expected $\mathbf{L}^2(\mathbb{R}^d)$ sup norm of a random operator to its norm on stationary processes. This lemma is proved in Appendix G.

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

$$E(k_\tau(x,u)k_\tau^*(x,u')) = \overline{k}_\tau(x-u, x-u') \quad \textit{and}$$
$$\iint |\overline{k}_\tau(v,v')| \, |v-v'| dv \, dv' < \infty,$$

then for any stationary process Y independent of τ, $E(|K_\tau Y(x)|^2)$ does not depend upon x and

$$E(|K_\tau Y|^2) \leq E(\|K_\tau\|^2) E(|Y|^2), \tag{4.17}$$

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

This result remains valid when replacing $S_J[\mathcal{P}_J]$ by $S_J[\mathcal{P}_{J,m}]$ and $U[\mathcal{P}_J]$ by $U[\mathcal{P}_{J,m}]$. With the same argument as in the proof of (2.60), we verify that

$$E(\|U[\mathcal{P}_{J,m}]X\|_1) \leq m E(|X|^2)^{1/2},$$

which proves (4.14). □

Small stationary deformations of stationary processes result in small modifications of the scattering distance, which is important to characterize deformed stationary processes as in image textures [3]. The following corollary proves that the expected scattering transform is almost Lipschitz-continuous in the size of the stochastic deformation gradient $\nabla\tau$, up to a log term.

COROLLARY 4.9. *There exists C such that for all independent stationary processes τ and X satisfying $\|\nabla\tau\|_\infty \leq \frac{1}{2}$ with probability 1, if $E(\|U[\mathcal{P}_\infty]X\|_1) < \infty$, then*

$$\|\overline{S}L_\tau X - \overline{S}X\|^2 \leq CE(\|U[\mathcal{P}_\infty]\|_1)E(|X|^2)K(\tau) \tag{4.18}$$

with

$$K(\tau) = E\biggl\{\biggl(\|\nabla\tau\|_\infty\biggl(\log\frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty} \vee 1\biggr) + \|H\tau\|_\infty\biggr)^2\biggr\}. \tag{4.19}$$

PROOF. Because

$$E(\|S_J[\mathcal{P}_J]L_\tau X - S_J[\mathcal{P}_J]X\|^2) \le \|E(S_J[\mathcal{P}_J]L_\tau X) - E(S_J[\mathcal{P}_J]X)\|^2,$$

letting J go to ∞ in (4.12) proves (4.18). □

5 Invariance to Actions of Compact Lie Groups

Invariant scattering is extended to the action of compact Lie groups G. Section 5.1 builds scattering operators in $\mathbf{L}^2(G)$ that are invariant to the action of G. Section 5.2 defines a translation- and rotation-invariant operator on $\mathbf{L}^2(\mathbb{R}^d)$ by combining a scattering operator on $\mathbf{L}^2(\mathbb{R}^d)$ and a scattering operator on $\mathbf{L}^2(SO(d))$.

5.1 Compact Lie Group Scattering

Let G be a compact Lie group and $\mathbf{L}^2(G)$ be the space of measurable functions $f(r)$ such that $\|f\|^2 = \int_G |f(r)|^2\,dr < \infty$, where dr is the Haar measure of G. The left action of $g \in G$ on $f \in \mathbf{L}^2(G)$ is defined by $L_g f(r) = f(g^{-1}r)$. This section introduces a scattering transform on $\mathbf{L}^2(G)$ that is invariant under the action of G. It is obtained with a scattering propagator that cascades the modulus of wavelet transforms defined on $\mathbf{L}^2(G)$.

The construction of Littlewood-Paley decompositions on compact manifolds and in particular on compact Lie groups was developed by Stein [19]. Different wavelet constructions have been proposed over manifolds [16]. Geller and Pesenson [8] have built unitary wavelet transforms on compact Lie groups, which can be viewed as analogues of unitary wavelet transforms on the circle in $\mathbb{R}^2$. In place of sinusoids, they use the eigenvectors of the Laplace-Beltrami operator of an invariant metric defined on the group. Similarly to Meyer wavelets [14], these basis elements are regrouped into dyadic subbands with appropriate windowing. For any $2^L \le 1$, it defines a scaling function $\widetilde{\phi}_{2^L}(r)$ and a family of wavelets $\{\widetilde{\psi}_{2^j}(r)\}_{-L<j<\infty}$ that are in $\mathbf{L}^2(G)$ [8]. The wavelet coefficients of $f \in \mathbf{L}^2(G)$ are computed with left convolutions on the group G for each $\widetilde{\lambda} = 2^j$,

$$\widetilde{W}[\widetilde{\lambda}]f(r) = f \star \widetilde{\psi}_{\widetilde{\lambda}}(r) = \int_G f(g)\widetilde{\psi}_{\widetilde{\lambda}}(g^{-1}r)dg, \tag{5.1}$$

and the scaling function performs an averaging on G,

$$\widetilde{A}_L f(r) = f \star \widetilde{\phi}_{2^L}(r) = \int_G f(g)\widetilde{\phi}_{2^L}(g^{-1}r)dg. \tag{5.2}$$

The resulting wavelet transform of $f \in \mathbf{L}^2(G)$ is

$$\widetilde{W}_L f = \{\widetilde{A}_L f, (\widetilde{W}[\widetilde{\lambda}]f)_{\widetilde{\lambda}\in\widetilde{\Lambda}_L}\} \quad \text{with } \widetilde{\Lambda}_L = \{\widetilde{\lambda} = 2^j : j > -L\}.$$

At the maximum scale $2^L = 1$, since $\widetilde{\phi}_1(r) = (\int_G dg)^{-1} = |G|^{-1}$ is constant, the operator $\widetilde{A}_0$ performs an integration on the group,

$$\widetilde{A}_0 f(r) = |G|^{-1} \int_G f(g)dg = \text{const.} \tag{5.3}$$

Wavelets are constructed to obtain a unitary operator [8]

$$\|\widetilde{W}_L f\| = \|f\| \quad \text{with } \widetilde{W}_L f\|^2 = \|\widetilde{A}_L f\|^2 + \sum_{\widetilde{\lambda}\in\widetilde{\Lambda}_L} \|\widetilde{W}[\widetilde{\lambda}]f\|^2. \tag{5.4}$$

The Abelian group $G = SO(2)$ of rotations in $\mathbb{R}^2$ is a simple example parametrized by an angle in $[0, 2\pi]$. The space $\mathbf{L}^2(G)$ is thus equivalent to $\mathbf{L}^2[0, 2\pi]$. Wavelets in $\mathbf{L}^2(G)$ are the well-known periodic wavelets in $\mathbf{L}^2[0, 2\pi]$ [14]. They are obtained by periodizing a scaling function $\phi_{2^L}(x) = 2^{-L}\phi(2^{-L}x)$ and wavelets $\psi_{2^j}(x) = 2^j\psi(2^j x)$ with $(\phi, \psi) \in \mathbf{L}^2(\mathbb{R})^2$:

$$\widetilde{\phi}_{2^L}(x) = \sum_{m\in\mathbb{Z}} \phi_{2^L}(x - 2\pi m) \quad \text{and} \quad \widetilde{\psi}_{2^j}(x) = \sum_{m\in\mathbb{Z}} \psi_{2^j}(x - 2\pi m). \tag{5.5}$$

We suppose that $\hat{\phi}(0) = 1$ and $\hat{\phi}(2k\pi) = 0$ for $k \in \mathbb{Z} - \{0\}$. The Poisson formula implies that $\widetilde{\phi}(x) = \sum_{n\in\mathbb{Z}} \phi(x - n) = 2\pi$. Convolutions (5.1) and (5.2) on the rotation group are circular convolutions of periodic functions in $\mathbf{L}^2[0, 2\pi]$. With the Poisson formula, one can prove that the periodic wavelet transform $\widetilde{W}_L$ is unitary if and only if $(\phi, \psi) \in \mathbf{L}^2(\mathbb{R})^2$ satisfy the Littlewood-Paley equalities (2.7).

For a general compact Lie group G, we define a wavelet modulus operator by $\widetilde{U}[\widetilde{\lambda}]f = |\widetilde{W}_L[\widetilde{\lambda}]f|$, and the resulting one-step propagator is

$$\widetilde{U}_L f = \{\widetilde{A}_L f, (\widetilde{U}[\widetilde{\lambda}]f|)_{\widetilde{\lambda}\in\widetilde{\Lambda}_L}\}.$$

Since $\widetilde{W}_L$ is unitary, we verify as in (2.24) that $\widetilde{U}_L$ is nonexpansive and preserves the norm in $\mathbf{L}^2(G)$.

A scattering operator on $\mathbf{L}^2(G)$ applies $\widetilde{U}_L$ iteratively. Let $\widetilde{\mathcal{P}}_L$ denote the set of all finite paths $\widetilde{p} = \{\widetilde{\lambda}_1, \widetilde{\lambda}_2, \dots, \widetilde{\lambda}_m\}$ of length m, where $\widetilde{\lambda}_k = 2^{j_k} \in \widetilde{\Lambda}_L$. Following Definition 2.2, a scattering propagator on $\mathbf{L}^2(G)$ is a path ordered product of noncommutative wavelet modulus operators

$$\widetilde{U}[\widetilde{p}] = \widetilde{U}[\widetilde{\lambda}_m]\cdots\widetilde{U}[\widetilde{\lambda}_2]\widetilde{U}[\widetilde{\lambda}_1],$$

with $\widetilde{U}[\varnothing] = \mathrm{Id}$.

Following Definition 2.4, a windowed scattering is defined by averaging $\widetilde{U}[\widetilde{p}]f$ through a group convolution with $\widetilde{\phi}_{2^L}$

$$\widetilde{S}_L[\widetilde{p}]f(r) = \widetilde{A}_L\widetilde{U}[\widetilde{p}]f(r) = \int_G U[\widetilde{p}]f(g)\widetilde{\phi}_{2^L}(g^{-1}r)dg. \tag{5.6}$$

It yields an infinite family of functions $\widetilde{S}_L[\widetilde{\mathcal{P}}_L]f = \{\widetilde{S}_L[\widetilde{p}]f\}_{\widetilde{p}\in\widetilde{\mathcal{P}}_L}$ whose norm is

$$\|\widetilde{S}_L[\widetilde{\mathcal{P}}_L]f\|^2 = \sum_{\widetilde{p}\in\widetilde{\mathcal{P}}_L} \|\widetilde{S}_L[\widetilde{p}]f\|^2 \quad \text{with } \|\widetilde{S}_L[\widetilde{p}]f\|^2 = \int_G |\widetilde{S}_L[\widetilde{p}]f(r)|^2\,dr.$$

Since $\widetilde{U}[\widetilde{\mathcal{P}}_L]$ is obtained by cascading the nonexpansive operator $\widetilde{U}_L$, the same proof as in Proposition 2.5 shows that it is nonexpansive:

$$\forall (f,h) \in \mathbf{L}^2(G)^2 \quad \|\widetilde{S}_L[\widetilde{\mathcal{P}}_L]f - \widetilde{S}_L[\widetilde{\mathcal{P}}_L]h\| \le \|f-h\|.$$

Since $\widetilde{U}[\widetilde{\mathcal{P}}_L]$ preserves the norm in $\mathbf{L}^2(G)$, to also prove as in Theorem 2.6 that $\|\widetilde{S}_L[\widetilde{\mathcal{P}}_L]f\| = \|f\|$, it is necessary to verify that $\lim_{m\to\infty}\|\widetilde{U}[\widetilde{\Lambda}_L^m]\|^2 = 0$. For the translation group where $G = \mathbb{R}^d$, Theorem 2.6 proves this result by imposing a condition on the Fourier transform of the wavelet. The extension of this result is not straightforward on $\mathbf{L}^2(G)$ for general compact Lie groups G, but it remains valid for the rotation group $G = SO(2)$ in $\mathbb{R}^2$. Indeed, periodic wavelets $\widetilde{\psi}_{\widetilde{\lambda}} \in \mathbf{L}^2(SO(2)) = \mathbf{L}^2[0,2\pi]$ are obtained by periodizing wavelets $\psi_{\widetilde{\lambda}} \in \mathbf{L}^2(\mathbb{R})$ in (5.5), which is equivalent to subsampling uniformly their Fourier transform. If ψ satisfies the admissibility condition of Theorem 2.6, then by replacing convolutions with circular convolutions in the proof, we verify that the periodized wavelets $\widetilde{\psi}_{\widetilde{\lambda}}$ define a scattering transform of $\mathbf{L}^2[0,2\pi]$ that preserves the norm $\|\widetilde{S}_L[\widetilde{\mathcal{P}}_L]f\| = \|f\|$.

When $2^L = 1$, $\widetilde{A}_0$ is the integration operator (5.3) on the group, so $\widetilde{S}_0[\widetilde{p}]f(r)$ does not depend on r. Following Definition 2.3, it defines a scattering transform that maps any $f \in \mathbf{L}^2(G)$ into a function of the path variable $\widetilde{p}$:

$$\forall \widetilde{p} \in \widetilde{\mathcal{P}}_0 \quad \widetilde{S}_0[\widetilde{p}]f = |G|^{-1}\int_G U[\widetilde{p}]f(g)dg. \tag{5.7}$$

Over a compact Lie group, the scattering transform $\widetilde{S}_0[\widetilde{\mathcal{P}}_0]f = \{\widetilde{S}_0[\widetilde{p}]f\}_{\widetilde{p}\in\widetilde{\mathcal{P}}_0}$ is a discrete sequence in $\mathbf{l}^2(\widetilde{\mathcal{P}}_0)$. The following proposition proves that it is invariant under the action $L_g f(r) = f(g^{-1}r)$ of $g \in G$ on $f \in \mathbf{L}^2(G)$:

PROPOSITION 5.1. *For any $f \in \mathbf{L}^2(G)$ and $g \in G$*

$$\widetilde{S}_0[\widetilde{\mathcal{P}}_0]L_g f = \widetilde{S}_0[\widetilde{\mathcal{P}}_0]f. \tag{5.8}$$

PROOF. Since $\widetilde{A}_0$ and $\widetilde{W}[\widetilde{\lambda}]f$ are computed with left convolutions on G, they commute with L_g. We have that $\widetilde{U}[\widetilde{\lambda}]$ and hence $\widetilde{S}_0[\widetilde{\mathcal{P}}_L]$ also commutes with L_g. If $\widetilde{p} \in \widetilde{\mathcal{P}}_0$, since $\widetilde{S}_0[\widetilde{p}]f(r)$ is constant in r, $\widetilde{S}_0[\widetilde{p}]L_g f = L_g\widetilde{S}_0[\widetilde{p}]f = \widetilde{S}_0[\widetilde{p}]f$, which proves (5.8). □

As in the translation case, the Lipchitz continuity of $\widetilde{S}_L$ under the action of diffeomorphisms relies on the Lipschitz continuity of the wavelet transform $\widetilde{W}_L$. The action of a small diffeomorphism on $f \in \mathbf{L}^2(G)$ can be written $L_\tau f(r) = f(\tau(r)^{-1}r)$ with $\tau(r) \in G$. The proof of Theorem 2.12 on Lipschitz continuity applies to any compact Lie groups G. The main difficulty is to prove Lemma 2.14, which proves the Lipschitz continuity of $\widetilde{W}_L$ by computing an upper bound of the commutator norm $\|[\widetilde{W}_L, L_\tau]\|$. The proof of this lemma can still be carried by applying Cotlar's lemma, but integration by parts and the resulting bounds require appropriate hypotheses on the regularity and decay of $\widetilde{\psi}_{\widetilde{\lambda}}$. If $G = SO(2)$, then the proof can be directly adapted from the proof on the translation group by replacing convolutions with circular convolutions. It proves that $\widetilde{S}_L$ is Lipschitz-continuous under the action of diffeomorphisms on $\mathbf{L}^2(SO(2))$.

5.2 Combined Translation and Rotation Scattering

We construct a scattering operator on $\mathbf{L}^2(\mathbb{R}^d)$ that is invariant to translations and rotations by combining a translation-invariant scattering operator on $\mathbf{L}^2(\mathbb{R}^d)$ and a rotation-invariant scattering operator on $\mathbf{L}^2(SO(d))$.

Let G be a rotation subgroup of the general linear group of $\mathbb{R}^d$, which also includes the reflection $-\mathbb{1}$ defined by $-\mathbb{1}x = -x$. According to (2.2), the wavelet transform in $\mathbf{L}^2(\mathbb{R}^d)$ is defined for any $\lambda = 2^j r \in 2^{\mathbb{Z}} \times G$ by $W[\lambda]f = f \star \psi_\lambda$, where $\psi_\lambda(x) = 2^{dj}\psi(2^j r^{-1}x)$. Section 2.1 considers the case of a finite group G that is a subgroup of $SO(d)$ if d is even or a subgroup of $O(d)$ if d is odd, while including $-\mathbb{1}$. The extension to a compact subgroup potentially equal to $SO(d)$ or $O(d)$ is straightforward. We still denote by G^+ the quotient of G by $\{-\mathbb{1}, \mathbb{1}\}$. The wavelet transform of a complex-valued function is defined over all $\lambda \in 2^{\mathbb{Z}} \times G$ but is restricted to $2^{\mathbb{Z}} \times G^+$ if f is real. All discrete sums on G and G^+ are replaced by integrals with the Haar measure dr. The group is compact and thus has a finite measure $|G| = \int_G dr$. We have that these integrals behave as finite sums in all derivations of this paper. The theorems proved for a finite group G remains valid for a compact group G. In the following we concentrate on real-valued functions.

Let $\mathcal{P}_J$ be the countable set of all finite paths $p = (\lambda_1, \lambda_2, \ldots, \lambda_m)$ with $\lambda_k \in \Lambda_J = \{\lambda = 2^j r : j > -J, r \in G^+\}$. The windowed scattering $S_J[\mathcal{P}_J]f = \{S_J[p]f\}_{p \in \mathcal{P}_J}$ is defined in Definition 2.4, but Λ_J and $\mathcal{P}_J$ are not discrete sets anymore. The scattering norm is defined by summing the $\mathbf{L}^2(\mathbb{R}^d)$ norms of all $S_J[p]f$ for all $p = (2^{j_1}r_1, \ldots, 2^{j_m}r_m) \in \overline{\mathcal{P}}_J$, with the Haar measure,

$$\|S_J[\mathcal{P}_J]f\|^2 = \sum_{m=0}^{\infty} \sum_{j_1 > -J, \ldots, j_m > -J} \int_{G^{+m}} \|S_J[2^{j_1}r_1, \ldots, 2^{j_m}r_m]f\|^2 \, dr_1 \cdots dr_m,$$

which is written

$$\|S_J[\mathcal{P}_J]f\|^2 = \int_{\mathcal{P}_J} \|S_J[p]f\|^2 \, dp.$$

One can verify that $S_J[\mathcal{P}_J]$ is nonexpansive as in the case where G is a finite group. For an admissible scattering wavelet satisfying (2.28), Theorem 2.6 remains valid and $\|S_J[\mathcal{P}_J]f\| = \|f\|$.

The scattering S_J is covariant to rotations. Invariance to rotations in G is obtained by applying the scattering operator $\widetilde{S}_L$ defined on $\mathbf{L}^2(G)$ by (5.6). Any $p = (\lambda_1, \lambda_2, \dots, \lambda_m) \in \mathcal{P}_J$ with $\lambda_1 = r2^{j_1}$ can be written as a rotation $p = r\overline{p}$ of a normalized path $\overline{p} = (\overline{\lambda}_1, \overline{\lambda}_2, \dots, \overline{\lambda}_m)$, where $\overline{\lambda}_k = r^{-1}\lambda_k$ and hence where $\overline{\lambda}_1 = 2^{j_1}$ is a scaling without rotation. We have that

$$S_J[p]f(x) = S_J[r\overline{p}]f(x).$$

For each x and $\overline{p}$ fixed, $S_J[r\overline{p}]f(x)$ is a function of r that belongs to $\mathbf{L}^2(G)$. We can thus apply the scattering operator $\widetilde{S}_L[\widetilde{p}]$ to this function of r. The result is denoted $\widetilde{S}_L[\widetilde{p}]S_J[r\overline{p}]f(x)$ for all $\widetilde{p} \in \widetilde{\mathcal{P}}_L$. This output can be indexed by the original path variable $p = r\overline{p}$, and we denote the combined scattering by

$$\widetilde{S}_L[\widetilde{p}]S_J[p]f(x) := \widetilde{S}_L[\widetilde{p}]S_J[r\overline{p}]f(x). \tag{5.9}$$

This combined scattering cascades wavelet transforms and hence convolutions along the spatial variable x and along the rotation r, which is factorized from each path. In $d = 2$ dimensions then $\mathbf{L}^2(SO(2)) = \mathbf{L}^2[0, 2\pi]$. The wavelet transform along rotations is implemented by circular convolutions along the rotation angle variable in $[0, 2\pi]$, with the periodic wavelets (5.5).

A combined scattering transform computes

$$\widetilde{S}_L[\widetilde{\mathcal{P}}_L]S_J[\mathcal{P}_J]f = \{\widetilde{S}_L[\widetilde{p}]S_J[p]f\}_{p\in\mathcal{P}_J, \widetilde{p}\in\widetilde{\mathcal{P}}_L}.$$

Its norm is computed by summing the $\mathbf{L}^2(\mathbb{R}^d)$ norms $\|\widetilde{S}_L[\widetilde{p}]S_J[p]f\|^2$:

$$\|\widetilde{S}_L[\widetilde{\mathcal{P}}_L]S_J[\mathcal{P}_J]f\|^2 = \sum_{\widetilde{p}\in\widetilde{\mathcal{P}}_L} \int_{\mathcal{P}_J} \|\widetilde{S}_L[\widetilde{p}]S_J[p]f\|^2 \, dp. \tag{5.10}$$

Since $\widetilde{S}_L[\widetilde{\mathcal{P}}_L]$ and $S_J[\mathcal{P}_J]$ are nonexpansive, their cascade is also nonexpansive:

$$\forall (f, h) \in \mathbf{L}^2(\mathbb{R}^d)^2 \quad \|\widetilde{S}_L[\widetilde{\mathcal{P}}_L]S_J[\mathcal{P}_J]f - \widetilde{S}_L[\widetilde{\mathcal{P}}_L]S_J[\mathcal{P}_J]h\| \le \|f - h\|.$$

When J goes to ∞, $S_J[\mathcal{P}_J]$ converges to the scattering transform $\overline{S}f$, which is translation invariant and covariant to rotations: $\overline{S}(r \circ f)(p) = \overline{S}f(rp)$ for all $p \in \overline{\mathcal{P}}_\infty$. By setting $2^L = 1$, Proposition 5.1 proves that $\widetilde{S}_0[\widetilde{\mathcal{P}}_0]$ is invariant to rotations in G. If G is the full rotation group $SO(d)$, then the combined scattering $\widetilde{S}_0[\widetilde{p}]\overline{S}f(p)$ for $(\widetilde{p}, p) \in \widetilde{\mathcal{P}}_0 \times \overline{\mathcal{P}}_\infty$ defines a translation- and rotation-invariant representation. Such translation- and rotation-invariant scattering representations are used for the rotation-invariant classification of image textures [18]. For any $(c, g) \in \mathbb{R}^d \times SO(d)$, we denote $L_{c,g}f(x) := f(g^{-1}(x - c))$.

PROPOSITION 5.2. *For all* $(c, g) \in \mathbb{R}^d \times SO(d)$ *and all* $f \in \mathbf{L}^2(\mathbb{R}^d)$

$$\forall (\widetilde{p}, p) \in \widetilde{\mathcal{P}}_0 \times \overline{\mathcal{P}}_\infty \quad \widetilde{S}_0[\widetilde{p}]\overline{S}(p)L_{c,g}f = \widetilde{S}_0[\widetilde{p}]\overline{S}(p)f. \tag{5.11}$$

PROOF. The scattering transform in $\mathbf{L}^2(\mathbb{R}^d)$ is translation invariant and covariant to rotations: $\overline{S} L_{c,g} f(p) = \overline{S} f(g^{-1} p)$ for all $p \in \overline{\mathcal{P}}_\infty$. Since g^{-1} acts as a rotation on the path p, Proposition 5.1 proves that $\widetilde{S}_0[\widetilde{p}]\overline{S} f(g^{-1} p) = \widetilde{S}_0[\widetilde{p}]\overline{S} f(p)$, which gives (5.11). □

Appendix A Proof of Lemma 2.8

The proof of (2.35) shows that the scattering energy propagates towards lower frequencies. It computes the average arrival log frequency of the scattering energy $\|U[p]f\|$ for paths of length m and shows that it increases when m increases. The arrival log frequency of $p = \{\lambda_k = 2^{j_k} r_k\}_{k \le m}$ is the log frequency index $\log_2 |\lambda_m| = j_m$ of the last path element.

Let us denote $e_m = \|U[\Lambda_J^m]f\|^2$ and $\overline{e}_m = \|S_J[\Lambda_J^m]f\|^2$. The average arrival log frequency among paths of length m is

$$a_m = e_m^{-1} \sum_{p \in \Lambda_J^m} j_m \, \|U[p]f\|^2 \ge -J. \tag{A.1}$$

The following lemma shows that when m increases by 1, then a_m decreases by nearly $\alpha/2$, where α is defined in (2.28).

LEMMA A.1. *If* (2.28) *is satisfied, then*

$$\forall m > 0 \quad \frac{\alpha}{2} e_{m-1} \le (a_m + J)e_m - (a_{m+1} + J)e_{m+1} + e_{m-1} - e_m. \tag{A.2}$$

We first show that (A.2) implies (2.35) and then prove this lemma. Summing over (A.2) gives

$$\frac{\alpha}{2} \sum_{k=0}^{m-1} e_k \le (a_1 + J)e_1 - (a_{m+1} + J)e_{m+1} + e_0 - e_m \le e_0 + (a_1 + J)e_1. \tag{A.3}$$

For $m = 1$, $p = 2^j r$ so $a_1 \, e_1 = \sum_{j > -J} \sum_{r \in G^+} j \, \|W[2^j r]f\|^2$. Moreover, $e_0 = \|f\|^2$, so

$$e_0 + (a_1 + J)e_1 = \|f\|^2 + \sum_{j > -J} \sum_{r \in G^+} (j + J)\|W[2^j r]f\|^2.$$

Inserting this into (A.3) for $m = \infty$ proves (2.35).

Lemma A.1 is proved by calculating the evolution of a_m as m increases. We consider the advancement of a path p of length $m-1$ with two steps $p + 2^j r + 2^l r'$ and denote $f_p = U[p]f$. The average arrival log frequency a_m can be written as the average arrival log frequency of $\|U[p + 2^j r]f\|^2$ over all $2^j r$ and all p of length $m - 1$:

$$a_m \, e_m = \sum_{p \in \Lambda_J^{m-1}} \sum_{j > -J} \sum_{r \in G^+} j \, \|f_p \star \psi_{2^j r}\|^2. \tag{A.4}$$

After the second step, the average arrival log frequency of $\|U[p+2^j r+2^l r']f\|^2$ over all $p \in \Lambda_J^{m-1}$, $2^j r$, and $2^l r'$ is a_{m+1}:

$$a_{m+1}e_{m+1} = \sum_{p\in\Lambda_J^{m-1}} \sum_{j>-J} \sum_{r\in G^+} \sum_{l>-J} \sum_{r'\in G^+} l\||f_p \star \psi_{2^j r}| \star \psi_{2^l r'}\|^2.$$

The wavelet transform is unitary and hence for any $h \in \mathbf{L}^2(\mathbb{R}^d)$

$$\|h\|^2 = \sum_{l>-J} \sum_{r'\in G^+} \|h \star \psi_{2^l r'}\|^2 + \|h \star \phi_{2^J}\|^2.$$

Applied to each $h = f_p \star \psi_{2^j r}$ in (A.4) this equation, together with

$$\overline{e}_m = \sum_{p\in\Lambda_J^{m-1}} \sum_{j>-J} \sum_{r\in G^+} \||f_p \star \psi_{2^j r}| \star \phi_{2^J}\|^2\, dr,$$

shows that $I = a_m e_m - a_{m+1}e_{m+1} + J\overline{e}_m$ satisfies

$$I = \sum_{p\in\Lambda_J^{m-1}} \sum_{j>-J} \sum_{r'\in G^+} \Big(\sum_{l>-J} \sum_{r\in G^+} (j-l)\||f_p \star \psi_{2^j r}| \star \psi_{2^l r'}\|^2 + (j+J)\||f_p \star \psi_{2^j r}| \star \phi_{2^J}\|^2 \Big).$$

A lower bound of I is calculated by dividing the sum on l for $l \geq j$ and $l < j$. In the $j+J-1$ term for $l < j$, the index l is replaced by $j-1$ and the convolution with ϕ_{2^J} is incorporated in the sum:

$$I \geq \sum_{p\in\Lambda_J^{m-1}} \sum_{j>-J} \sum_{r'\in G^+} \Big[\sum_{-J<l<j} \Big(\sum_{r\in G^+} \||f_p \star \psi_{2^j r}| \star \psi_{2^l r'}\|^2 \Big) + \||f_p \star \psi_{2^j r}| \star \phi_{2^J}\|^2 - \sum_{l>j} \sum_{r\in G^+} (l-j)\||f_p \star \psi_{2^j r}| \star \psi_{2^l r'}\, dr'\|^2 \Big] dr. \tag{A.5}$$

Since wavelets satisfy the unitary property (2.7), for all real functions $f \in \mathbf{L}^2(\mathbb{R}^d)$ and all $q \in \mathbb{Z}$,

$$\sum_{-q\geq l>-J} \sum_{r\in G^+} \|f \star \psi_{2^l r}\|^2 + \|f \star \phi_{2^J}\|^2 = \|f \star \phi_{2^q}\|^2. \tag{A.6}$$

Indeed, (2.7) implies that

$$|\widehat{\phi}(2^J\omega)|^2 + \frac{1}{2} \sum_{-q\geq l>-J} \sum_{r\in G} |\widehat{\psi}(2^{-l}r^{-1}\omega)|^2 = |\widehat{\phi}(2^q\omega)|^2. \tag{A.7}$$

If f is real, then $\|f \star \psi_{2^j r}\| = \|f \star \psi_{-2^j r}\|$. Multiplying (A.7) by $|\widehat{f}(\omega)|^2$ and integrating in ω proves (A.6). Inserting (A.6) in (A.5) gives

$$I \geq \sum_{p \in \Lambda_J^{m-1}} \sum_{j>-J} \sum_{r \in G^+} \Bigg(\| |f_p \star \psi_{2^j r}| \star \phi_{2^{-j+1}} \|^2 - \sum_{l>j} (l-j) \big(\| |f_p \star \psi_{2^j r}| \star \phi_{2^{-l}} \|^2 - \| |f_p \star \psi_{2^j r}| \star \phi_{2^{-l+1}} \|^2 \big) \Bigg).$$

If $\rho \geq 0$ satisfies $|\widehat{\rho}(\omega)| \leq |\widehat{\phi}(2\omega)|$, then for any $f \in \mathbf{L}^2(\mathbb{R}^d)$ and any $l \in \mathbb{Z}$,

$$\|f \star \phi_{2^{-l+1}}\|^2 \geq \|f \star \rho_{2^l r}\|^2 \quad \text{with } \rho_{2^l r}(x) = 2^{dl} \rho(2^l r^{-1} x).$$

We have that

$$I \geq \sum_{p \in \Lambda_J^{m-1}} \sum_{j>-J} \sum_{r \in G^+} \Bigg(\| |f_p \star \psi_{2^j r}| \star \rho_{2^j r} \|^2 - \sum_{l>j} (l-j) \big(\|f_p \star \psi_{2^j r}\|^2 - \| |f_p \star \psi_{2^j r}| \star \rho_{2^l r} \|^2 \big) \Bigg).$$

Applying Lemma 2.7 for $h = \rho_{2^l r}$ and a frequency $2^j r \eta$ proves that

$$\| |f_p \star \psi_{2^j r}| \star \rho_{2^l r} \| \geq \| f_p \star \psi_{2^j r} \star \rho_{2^l r, 2^j} \| \qquad \text{with } \rho_{2^l r, 2^j}(x) = \rho_{2^l r}(x) e^{i 2^j r \eta . x}$$

and $\widehat{\rho}_{2^l r, 2^j}(\omega) = \widehat{\rho}(2^{-l} r^{-1} \omega - 2^{j-l} \eta)$. We have that

$$I \geq \sum_{p \in \Lambda_J^{m-1}} \sum_{j>-J} \sum_{r \in G^+} \Bigg(\| f_p \star \psi_{2^j r} \star \rho_{2^j r, 2^j} \|^2 - \sum_{l>j} (l-j) \big(\|f_p \star \psi_{2^j r}\|^2 - \| f_p \star \psi_{2^j r} \star \rho_{2^l r, 2^j} \|^2 \big) \Bigg).$$

We shall now rewrite this equation in the Fourier domain. Since $f_p(x) \in \mathbb{R}$, $|\widehat{f_p}(\omega)| = |\widehat{f_p}(-\omega)|$, applying Plancherel gives

$$I \geq \frac{1}{2} \sum_{p \in \Lambda_J^{m-1}} \int |\widehat{f_p}(\omega)|^2 \sum_{r \in G} \sum_{j>-J} \Bigg(|\widehat{\psi}(2^{-j} r^{-1} \omega)|^2 \, |\widehat{\rho}(2^{-j} r^{-1} \omega - \eta)|^2 - \sum_{l>j} (l-j) |\widehat{\psi}(2^{-j} r^{-1} \omega)|^2 (1 - |\widehat{\rho}(2^{-l} r^{-1} \omega - 2^{j-l} \eta)|^2) \Bigg) d\omega.$$

Inserting $\widehat{\Psi}$ defined in (2.27) by

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

with $k = l - j$ gives

$$I \geq \frac{1}{2} \sum_{p \in \Lambda_J^{m-1}} \int |\widehat{f_p}(\omega)|^2 \sum_{j > -J} b(2^{-j}\omega) d\omega$$

with $b(\omega) = \sum_{r \in G} \widehat{\Psi}(r^{-1}\omega) |\widehat{\psi}(r^{-1}\omega)|^2$. Let us add to I

$$\overline{e}_{m-1} = \sum_{p \in \Lambda_J^{m-1}} \|f_p \star \phi_{2^J}\|^2 = \sum_{p \in \Lambda_J^{m-1}} \int |\widehat{f_p}(\omega)|^2 \, |\widehat{\phi}(2^J\omega)|^2 \, d\omega.$$

Since $\rho \geq 0$, $|\widehat{\rho}(\omega)| \leq \widehat{\rho}(0) = 1$ and hence $\widehat{\Psi}(\omega) \leq 1$. The wavelet unitary property (2.7) together with $\widehat{\Psi}(\omega) \leq 1$ implies that

$$|\widehat{\phi}(2^J\omega)|^2 = \frac{1}{2} \sum_{j \leq -J} \sum_{r \in G} |\widehat{\psi}(2^{-j} r^{-1} \omega)|^2 \geq \frac{1}{2} \sum_{j \leq -J} b(2^{-j}\omega)$$

so

$$I + \overline{e}_{m-1} \geq \frac{1}{2} \sum_{p \in \Lambda_J^{m-1}} \int |\widehat{f_p}(\omega)|^2 \sum_{j=-\infty}^{+\infty} b(2^{-j}\omega) d\omega.$$

If $\alpha = \inf_{1 \leq |\omega| < 2} \sum_j b(2^{-j}\omega)$, then $\sum_j b(2^{-j}\omega) \geq \alpha$ for all $\omega \neq 0$. If hypothesis (2.28) is satisfied and hence $\alpha > 0$, then

$$\begin{aligned} I + \overline{e}_{m-1} \geq \frac{\alpha}{2} \sum_{p \in \Lambda_J^{m-1}} \int |\widehat{f_p}(\omega)|^2 \, d\omega &= \frac{\alpha}{2} \sum_{p \in \Lambda_J^{m-1}} \|f_p\|^2 \\ &= \frac{\alpha}{2} \sum_{p \in \Lambda_J^{m-1}} \|U[p]f\|^2 = \frac{\alpha}{2} e_{m-1}. \end{aligned}$$

Inserting $I = a_m \, e_m - a_{m+1} \, e_{m+1} + J \, \overline{e}_m$ proves that

$$a_m \, e_m - a_{m+1} \, e_{m+1} + J\overline{e}_m + \overline{e}_{m-1} \geq \frac{\alpha}{2} e_{m-1}. \tag{A.8}$$

Since U_J preserves the norm, $e_m = e_{m+1} + \overline{e}_m$; indeed, (2.23) proves that $U_J U[\Lambda_J^m] f = \{U[\Lambda_J^{m+1}] f, S_J[\Lambda_J^m] f\}$. Inserting $\overline{e}_m = e_m - e_{m+1}$ and $\overline{e}_{m-1} = e_{m-1} - e_m$ into (A.8) gives

$$\frac{\alpha}{2} e_{m-1} \leq (a_m + J) e_m - (a_{m+1} + J) e_{m+1} + e_{m-1} - e_m,$$

which finishes the proof of Lemma A.1.

Appendix B Proof of Lemma 2.11

Lemma 2.11 as well as all other upper bounds on operator norms are computed with Schur's lemma. For any operator $Kf(x) = \int f(u)k(x,u)du$, Schur's lemma proves that

$$\int |k(x,u)|dx \le C \quad \text{and} \quad \int |k(x,u)|du \le C \quad \Longrightarrow \quad \|K\| \le C, \tag{B.1}$$

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

The operator norm of $k_J = L_\tau A_J - A_J$ is computed by applying Schur's lemma on its kernel,

$$k_J(x,u) = \phi_{2^J}(x - \tau(x) - u) - \phi_{2^J}(x-u). \tag{B.2}$$

A first-order Taylor expansion proves that

$$\begin{aligned} |k_J(x,u)| &\le \left| \int_0^1 \nabla\phi_{2^J}(x - u - t\tau(x)) \cdot \tau(x) dt \right| \\ &\le \|\tau\|_\infty \int_0^1 |\nabla\phi_{2^J}(x-u-t\tau(x))| dt \end{aligned}$$

so

$$\int |k_J(x,u)|du \le \|\tau\|_\infty \int_0^1 \int |\nabla\phi_{2^J}(x-u-t\,\tau(x))| du\, dt. \tag{B.3}$$

Since $\nabla\phi_{2^J}(x) = 2^{-dJ-J}\nabla\phi(2^{-J}x)$, we have that

$$\begin{aligned} \int |k_J(x,u)|du &\le \|\tau\|_\infty 2^{-dJ-J} \int |\nabla\phi(2^{-J}u')|du' \\ &= 2^{-J}\|\tau\|_\infty \|\nabla\phi\|_1. \end{aligned} \tag{B.4}$$

Similarly to (B.3) we prove that

$$\int |k_J(x,u)|dx \le \|\tau\|_\infty \int_0^1 \int |\nabla\phi_{2^J}(x-u-t\,\tau(x))| dx\, dt.$$

The Jacobian of the change of variable $v = x - t\tau(x)$ is $\mathrm{Id} - t\nabla\tau(x)$, whose determinant is larger than $(1 - \|\nabla\tau\|_\infty)^d \ge 2^{-d}$, so

$$\begin{aligned} \int |k_J(x,u)|dx &\le \|\tau\|_\infty 2^d \int_0^1 \int |\nabla\phi_{2^J}(v-u)| dv\, dt \\ &= 2^{-J}\|\tau\|_\infty \|\nabla\phi\|_1 2^d. \end{aligned}$$

Schur's lemma (B.1) applied to this upper bound and (B.4) proves the lemma result:

$$\|L_\tau A_J - A_J\| \le 2^{-J+d}\|\nabla\phi\|_1 \|\tau\|_\infty.$$

Appendix C Proof of (2.67)

We prove that

$$\|L_\tau A_J f - A_J f + \tau \cdot \nabla A_J f\| \leq C \|f\| \, 2^{-2J} \|\tau\|_\infty^2 \tag{C.1}$$

by applying Schur's lemma (B.1) on the kernel of $k_J = L_\tau A_J - A_J + \tau \cdot \nabla A_J$:

$$k_J(x,u) = \phi_{2^J}(x - \tau(x) - u) - \phi_{2^J}(x-u) + \nabla\phi_{2^J}(x-u)\cdot\tau(x).$$

Let $Hf(x)$ be the Hessian matrix of a function f at x and $|Hf(x)|$ the sup matrix norm of this Hessian matrix. A Taylor expansion gives

$$|k_J(x,u)| = \left| \int_0^1 t\tau(x)\cdot H\phi_{2^J}(u - x - (1-t)\tau(u))\cdot\tau(x)\,dt \right|$$

$$\leq \|\tau\|_\infty^2 \int_0^1 |t| \, |H\phi_{2^J}(u - x - (1-t)\tau(x))|\,dt. \tag{C.2}$$

Since $\phi_{2^J}(x) = 2^{-dJ}\phi(2^{-J}x)$, $H\phi_{2^J}(x) = 2^{-Jd-2J}H\phi(2^{-J}x)$. With a change of variable, (C.2) gives

$$\begin{aligned}\int |k_J(x,u)|\,du &\leq \|\tau\|_\infty^2 2^{-dJ-2J} \int |H\phi(2^{-J}u')|\,du' \\ &= 2^{-2J}\|\tau\|_\infty^2 \, \|H\phi\|_1,\end{aligned} \tag{C.3}$$

where $\|H\phi\|_1 = \int |H\phi(u)|\,du$ is bounded. Indeed, all second-order derivatives of ϕ at u are $O((1+|u|)^{-d-1})$.

The upper bound (C.2) also implies that

$$\int |k_J(x,u)|\,dx \leq \|\tau\|_\infty^2 \int_0^1 |t| \int |H\phi_{2^J}(u - x - (1-t)\tau(x))|\,du\,dt.$$

The Jacobian of the change of variable $v = x - (1-t)\tau(x)$ is $\mathrm{Id} - (1-t)\nabla\tau(x)$, whose determinant is larger than $(1 - \|\nabla\tau\|_\infty)^d$, so

$$\begin{aligned}\int |k_J(x,u)|\,dx &\leq \|\tau\|_\infty^2 (1 - \|\nabla\tau\|_\infty)^{-d} \int_0^1\int |H\phi_{2^J}(v-u)|\,dv\,dt \\ &= 2^{-2J}\|\tau\|_\infty^2 \, \|H\phi\|_1 \, 2^d.\end{aligned} \tag{C.4}$$

The upper bounds (C.3) and (C.4) with Schur's lemma (B.1) proves (C.1).

Appendix D Proof of Lemma 2.13

This appendix proves that for any operator L and any $f \in \mathbf{L}^2(\mathbb{R}^d)$

$$\|[S_J[\mathcal{P}_J], L]f\| \leq \|[U_J, L]\| \, \|U[\mathcal{P}_J]f\|_1 = \|[U_J, L]\| \sum_{n=0}^{\infty} \|U[\Lambda_J^n]f\|. \tag{D.1}$$

For operators A and B, we denote by $\{A, B\}$ the operator defined by $\{A, B\}f = \{Af, Bf\}$. We introduce a wavelet modulus operator without averaging:

(D.2) $$V_J f = \{|W[\lambda]f| = |f \star \psi_\lambda|\}_{\lambda \in \Lambda_J} \quad \text{with } \Lambda_J = \{2^j r : j > -J, r \in G^+\},$$

and $U_J = \{A_J, V_J\}$. The propagator V_J creates all paths

$$V_J U[\Lambda_J^n] f = U[\Lambda_J^{n+1}] f \quad \text{for any } n \geq 0.$$

Since $U[\Lambda_J^0] = \mathrm{Id}$, we have that $V_J^n = U[\Lambda_J^n]$. Let $\mathcal{P}_{J,m}$ be the subset of $\mathcal{P}_J$ of paths p of length smaller than m.

To verify (D.1), we shall prove that

(D.3) $$[S_J[\mathcal{P}_{J,m}], L] = \sum_{n=0}^{m} K_{m-n} V_J^n,$$

where $K_n = \{[A_J, L], S_J[\mathcal{P}_{J,n-1}][V_J, L]\}$ satisfies

(D.4) $$\|K_n\| \leq \|[U_J, L]\|.$$

Since $V_J^n f = U[\Lambda_J^n] f$, it implies that for any $f \in \mathbf{L}^2(\mathbb{R}^d)$

$$\|[S_J[\mathcal{P}_{J,m}], L]f\| \leq \sum_{n=0}^{m} \|K_{m-n}\| \, \|V_J^n f\| \leq \|[U_J, L]\| \sum_{n=0}^{m-1} \|U[\Lambda_J^n] f\|,$$

and letting m tend to ∞ proves (D.1).

Property (D.3) is proved by first showing that

(D.5) $$S_J[\mathcal{P}_{J,m}]L = \{LA_J, S_J[\mathcal{P}_{J,m-1}]LV_J\} + K_m,$$

where $K_m = \{[A_J, L], S_J[\mathcal{P}_{J,m-1}]\,[V_J, L]\}$. Indeed, since $V_J^n = U[\Lambda_J^n]$, we have that $A_J\, V_J^n = S_J[\Lambda_J^n]$ and $\mathcal{P}_{J,m} = \bigcup_{n=0}^{m-1} \Lambda_J^n$ yields

$$S_J[\mathcal{P}_{J,m}] = \{A_J V_J^n\}_{0 \leq n < m}.$$

We have that

$$\begin{aligned} S_J[\mathcal{P}_{J,m}]L &= \{A_J\, V_J^n L\}_{0 \leq n < m} \\ &= \{LA_J + [A_J, L], A_J V_J^{n-1} L V_J + A_J V_J^{n-1}[V_J, L]\}_{1 \leq n < m} \\ &= \{LA_J, S_J[\mathcal{P}_{J,m-1}]LV_J\} + \{[A_J, L], S_J[\mathcal{P}_{J,m-1}]\,[V_J, L]\} \\ &= \{LA_J, S_J[\mathcal{P}_{J,m-1}]LV_J\} + K_m, \end{aligned}$$

which proves (D.5).

A substitution of $S_J[\mathcal{P}_{J,m-1}]L$ in (D.5) by the expression derived by this same formula gives

$$S_J[\mathcal{P}_{J,m}]L = \{LA_J, LA_J V_J, S_J[\mathcal{P}_{J,m-2}]LV_J^2\} + K_{m-1}V_J + K_m.$$

With m substitutions, we obtain

$$\begin{aligned} S_J[\mathcal{P}_{J,m}]L &= \{LA_J V_J^n\}_{0\le n<m} + \sum_{n=0}^{m} K_{m-n} V_J^n \\ &= LS_J[\mathcal{P}_{J,m}] + \sum_{n=0}^{m} K_{m-n} V_J^n, \end{aligned}$$

which proves (D.3).

Let us now prove (D.4) on $K_m = \{[A_J, L], S_J[\mathcal{P}_{J,m-1}]\,[V_J, L]\}$. Since $S_J[\mathcal{P}_J]$ is nonexpansive, its restriction $S_J[\mathcal{P}_{J,m}]$ is also nonexpansive. Given that $U_J = \{A_J, V_J\}$, we get

$$\begin{aligned} \|K_m f\|^2 &= \|[A_J, L]f\|^2 + \|S_J[\mathcal{P}_{J,m-1}][V_J, L]f\|^2 \\ &\le \|[A_J, L]f\|^2 + \|[V_J, L]f\|^2 = \|[U_J, L]f\|^2 \le \|[U_J, L]\|^2\,\|f\|^2, \end{aligned}$$

which proves (D.4).

Appendix E Proof of Lemma 2.14

This section computes an upper bound of $\|[W_J, L_\tau]\|$ by considering

$$[W_J, L_\tau]^*[W_J, L_\tau] = \sum_{r\in G^+}\sum_{j=-J+1}^{\infty}[W[2^j r], L_\tau]^*[W[2^j r], L_\tau] + [A_J, L_\tau]^*\,[A_J, L_\tau].$$

Since $\|[W_J, L_\tau]\| = \|[W_J, L_\tau]^*\,[W_J, L_\tau]\|^{1/2}$,

$$\|[W_J, L_\tau]\| \le \sum_{r\in G^+}\Big\|\sum_{j=-J+1}^{\infty}[W[2^j r], L_\tau]^*[W[2^j r], L_\tau]\Big\|^{1/2} + \|[A_J, L_\tau]^*\,[A_J, L_\tau]\|^{1/2}. \tag{E.1}$$

To prove the upper bound (2.59) of Lemma 2.14, we compute an upper bound for each term on the right under the integral and the last term, which is done by the following lemma.

LEMMA E.1. *Suppose that $h(x)$, as well as all its first- and second-order derivatives, have a decay in $O((1+|x|)^{-d-2})$. Let $Z_j f = f \star h_j$ with $h_j(x) = 2^{dj} h(2^j x)$. There exists $C > 0$ such that if $\|\nabla\tau\|_\infty \le \frac{1}{2}$, then*

$$\|[Z_j, L_\tau]\| \le C\,\|\nabla\tau\|_\infty, \tag{E.2}$$

and if $\int h(x)dx = 0$, *then*

$$\text{(E.3)}\quad \Big\| \sum_{j=-\infty}^{+\infty} [Z_j, L_\tau]^*[Z_j, L_\tau] \Big\|^{1/2} \leq C\left(\max\left(\log \frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty}, 1\right)\|\nabla\tau\|_\infty + \|H\tau\|_\infty\right).$$

Inequality (E.3) clearly remains valid if the summation is limited to $-J$ instead of $-\infty$, since $[Z_j, L_\tau]^*[Z_j, L_\tau]$ is a positive operator. Inserting in (E.1) both (E.2) with $h = \phi$ and (E.3) with $h(x) = \psi(r^{-1}x)$ for each $r \in G^+$ and replacing $-\infty$ by $-J$ proves the upper bound (2.59) of Lemma 2.14.

To prove Lemma E.1, we factorize

$$[Z_j, L_\tau] = K_j\, L_\tau \quad \text{with } K_j = Z_j - L_\tau\, Z_j\, L_\tau^{-1}.$$

Observe that

$$\text{(E.4)}\qquad \|[Z_j, L_\tau]^*[Z_j, L_\tau]\|^{1/2} = \|L_\tau^* K_j^* K_j L_\tau\|^{1/2} \leq \|L_\tau\|\, \|K_j^* K_j\|^{1/2},$$

and that

$$\text{(E.5)}\qquad \Big\| \sum_{j=-\infty}^{+\infty} [Z_j, L_\tau]^*[Z_j, L_\tau] \Big\|^{1/2} \leq \|L_\tau\| \, \Big\| \sum_{j=-\infty}^{+\infty} K_j^* K_j \Big\|^{1/2}$$

with $\|L_\tau\| \leq (1 - \|\nabla\tau\|_\infty)^{-d}$. Since $L_\tau^{-1} f(x) = f(\xi(x))$ with $\xi(x - \tau(x)) = x$, the kernel of $K_j = Z_j - L_\tau\, Z_j\, L_\tau^{-1}$ is

$$\text{(E.6)}\qquad k_j(x, u) = h_j(x - u) - h_j(x - \tau(x) - u + \tau(u)) \det(\mathrm{Id} - \nabla\tau(u)).$$

By computing upper bounds of $\|K_j\|$ and $\|\sum_{j=-\infty}^{+\infty} K_j^* K_j\|$, the lemma is proved. The sum over j is divided into three parts

$$\text{(E.7)}\quad \Big\| \sum_{j=-\infty}^{+\infty} K_j^* K_j \Big\|^{1/2} \leq \Big\| \sum_{j=-\infty}^{-\gamma-1} K_j^* K_j \Big\|^{1/2} + \Big\| \sum_{j=-\gamma}^{-1} K_j^* K_j \Big\|^{1/2} + \Big\| \sum_{j=0}^{\infty} K_j^* K_j \Big\|^{1/2}.$$

We shall first prove that

$$\text{(E.8)}\quad \Big\| \sum_{j=-\infty}^{-\gamma} K_j^* K_j \Big\|^{1/2} \leq C\big(\|\nabla\tau\|_\infty + 2^{-\gamma}\, \|\Delta\tau\|_\infty + 2^{-\gamma/2} \|\Delta\tau\|_\infty^{1/2}\, \|\nabla\tau\|_\infty^{1/2}\big).$$

Then we verify that $\|K_j\| \leq C\|\nabla\tau\|_\infty$ and hence that

$$\Big\|\sum_{j=-\gamma}^{-1} K_j^* K_j\Big\|^{1/2} \leq \gamma\|K_j\| \leq C\gamma\|\nabla\tau\|_\infty. \tag{E.9}$$

The last term carries the singular part, and we prove that

$$\Big\|\sum_{j=0}^{\infty} K_j^* K_j\Big\|^{1/2} \leq C(\|\nabla\tau\|_\infty + \|H\tau\|_\infty). \tag{E.10}$$

Choosing $\gamma = \max(\log[\|\Delta\tau\|_\infty/\|\nabla\tau\|_\infty], 1)$ yields

$$\Big\|\sum_{j=-\infty}^{+\infty} K_j^* K_j\Big\|^{1/2} \leq C\left(\max\left(\log\frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty}, 1\right)\|\nabla\tau\|_\infty + \|H\tau\|_\infty\right).$$

Inserting this result in (E.5) will prove the second lemma result (E.3). In the proof, C is a generic constant that depends only on h but that evolves along the calculations.

The proof of (E.8) is done by decomposing $K_j = \widetilde{K}_{j,1} + \widetilde{K}_{j,2}$, with a first kernel

$$\widetilde{k}_{j,1}(x,u) = a(u)h_j(x-u) \quad \text{with } a(u) = \big(1 - \det(\mathrm{Id} - \nabla\tau(u))\big), \tag{E.11}$$

and a second kernel

$$\widetilde{k}_{j,2}(x,u) = \det(\mathrm{Id} - \nabla\tau(u))\big(h_j(x-u) - h_j(x - \tau(x) - u + \tau(u))\big). \tag{E.12}$$

This kernel has a similar form to the kernel (B.2) in Appendix B with $\tau(x)$ replaced here by $\tau(x) - \tau(u)$. The same proof shows that

$$\|\widetilde{K}_{j,2}\| \leq C\, 2^j\, \|\Delta\tau\|_\infty. \tag{E.13}$$

Taking advantage of this decay, to prove (E.8) we decompose

$$\begin{aligned}\Big\|\sum_{j=-\infty}^{-\gamma} K_j^* K_j\Big\|^{1/2} \leq{}& \Big\|\sum_{j=-\infty}^{-\gamma} \widetilde{K}_{j,1}^* \widetilde{K}_{j,1}\Big\|^{1/2}\\ &+ \sum_{j=-\infty}^{-\gamma} (\|\widetilde{K}_{j,2}\| + 2^{1/2}\, \|\widetilde{K}_{j,2}\|^{1/2}\|\widetilde{K}_{j,1}\|^{1/2})\end{aligned}$$

and verify that

$$\|\widetilde{K}_{j,1}\| \leq C\|\nabla\tau\|_\infty \quad \text{and} \quad \Big\|\sum_{j=-\infty}^{0} \widetilde{K}_{j,1}^* \widetilde{K}_{j,1}\Big\|^{1/2} \leq C\|\nabla\tau\|_\infty. \tag{E.14}$$

The kernel of the self-adjoint operator $\widetilde{K}_{j,1}^* \widetilde{K}_{j,1}$ is

$$\widetilde{k}_j(y,z) = \int \widetilde{k}_{j,1}^*(x,y)\widetilde{k}_{j,1}(x,z)\,dx = a(y)a(z)\widetilde{h}_j \star h_j(z-y),$$

with $\widetilde{h}_j(u) = h_j^*(-u)$. We now have that the kernel of $\widetilde{K} = \sum_{j\le 0} \widetilde{K}_{j,1}^* \widetilde{K}_{j,1}$ is

$$\widetilde{k}(y,z) = \sum_{j\le 0} \widetilde{k}_j(y,z) = a(y)a(z)\theta(z-y) \quad \text{with } \theta(x) = \sum_{j\le 0} \widetilde{h}_j \star h_j(x).$$

Applying Young's inequality to $\|\widetilde{K} f\|$ gives

$$\|\widetilde{K}\| \le \sup_{u\in\mathbb{R}^d} |a(u)|^2 \,\|\theta\|_1.$$

Since $\widehat{\theta}(\omega) = \sum_{j\le 0} |\widehat{h}(2^{-j}\omega)|^2$ and $\widehat{h}(0) = \int h(x)\,dx = 0$ and h is regular with a polynomial decay, we verify that $\|\theta\|_1 < \infty$. Moreover, since

$$(1 - \det(\mathrm{Id} - \nabla\tau(u))) \ge (1 - \|\nabla\tau\|_\infty)^d$$

we have $\sup_u |a(u)| \le d\,\|\nabla\tau\|_\infty$, which proves that $\|\widetilde{K}\|^{1/2} \le C\,\|\nabla\tau\|_\infty$. Since $\|\widetilde{K}_{j,1}\|^2 \le \|\widetilde{K}\|$, we get the same inequality for $\|\widetilde{K}_{j,1}\|^2$, which proves the two upper bounds of (E.14).

The last sum $\sum_{j=0}^{\infty} K_j^* K_j$ carries the singular part of the operator, which is isolated and evaluated separately by decomposing $K_j = K_{j,1} + K_{j,2}$, with a first kernel

$$k_{j,1}(x,u) = h_j(x-u) - h_j((\mathrm{Id} - \nabla\tau(u))(x-u))\det(\mathrm{Id} - \nabla\tau(u)) \tag{E.15}$$

satisfying $K_{j,1}1 = \int k_{j,1}(x,u)\,du = 0$ if $\int h(x)\,dx = 0$. The second kernel is

$$\begin{aligned} k_{j,2}(x,u) = \det(\mathrm{Id} - \nabla\tau(u))\big(&h_j((\mathrm{Id} - \nabla\tau(u))(x-u)) \\ &- h_j(x - \tau(x) - u + \tau(u))\big). \end{aligned} \tag{E.16}$$

The sum $\sum_{j\ge 0} K_{j,1}^* K_{j,1}$ has a singular kernel along its diagonal, and its norm is evaluated separately with the upper bound

$$\begin{aligned} \Big\|\sum_{j=0}^{\infty} K_j^* K_j\Big\|^{1/2} \le &\Big\|\sum_{j=0}^{\infty} K_{j,1}^* K_{j,1}\Big\|^{1/2} \\ &+ \sum_{j=0}^{\infty} (\|K_{j,2}\| + 2^{1/2}\|K_{j,2}\|^{1/2}\|K_{j,1}\|^{1/2}). \end{aligned} \tag{E.17}$$

We will prove that

$$\|K_{j,1}\| \le C\,\|\nabla\tau\|_\infty \tag{E.18}$$

and

$$\|K_{j,2}\| \le C\,\min(2^{-j}\,\|H\tau\|_\infty, \|\nabla\tau\|_\infty). \tag{E.19}$$

These two inequalities imply that $\|K_j\| \le C\,\|\nabla\tau\|_\infty$. Inserting this inequality in (E.4) yields the first lemma result (E.2), and it proves (E.9). Equations (E.18) and

(E.19) also prove that

$$\sum_{j=0}^{\infty}(\|K_{j,2}\| + 2^{1/2}\|K_{j,2}\|^{1/2}\|K_{j,1}\|^{1/2}) \le C(\|\nabla\tau\|_\infty + \|H\tau\|_\infty). \tag{E.20}$$

If $\int h(x)dx = 0$, then thanks to the vanishing integrals of $k_{j,1}$ we will prove that

$$\left\|\sum_{j=0}^{\infty} K_{j,1}^* K_{j,1}\right\|^{1/2} \le C(\|\nabla\tau\|_\infty + \|H\tau\|_\infty). \tag{E.21}$$

Inserting (E.20) and (E.21) into (E.17) proves (E.10).

Let us now first prove the upper bound (E.19) on $K_{j,2}$. The kernel of $K_{j,2}$ is

$$\begin{aligned} k_{j,2}(x,u) = \det(\mathrm{Id} - \nabla\tau(u))\big(&h_j((\mathrm{Id} - \nabla\tau(u))(x-u)) \\ &- h_j(x - \tau(x) - u + \tau(u))\big). \end{aligned}$$

A Taylor expansion of h_j together with a Taylor expansion of $\tau(x)$ gives

$$\tau(x) - \tau(u) = \nabla\tau(u)(x-u) + \alpha(u, x-u) \tag{E.22}$$

with

$$\alpha(u,z) = \int_0^1 tzH\tau(u + (1-t)z)z\,dt, \tag{E.23}$$

so

$$\begin{aligned} k_{j,2}(x,u) = -\det(\mathrm{Id} - \nabla\tau(u)) \int_0^1 \nabla h_j\big(&(\mathrm{Id} - t\nabla\tau(u))(x-u) \\ &+ (1-t)(\tau(u) - \tau(x))\big)\alpha(u, x-u)dt. \end{aligned} \tag{E.24}$$

For $j \ge 0$, we prove that $\|K_{j,2}\|$ decays like 2^{-j}. First we observe that $|\det(\mathrm{Id} - \nabla\tau(u))| \le 2^d$. Since $\nabla h_j(u) = 2^{j+dj}\nabla h(2^j u)$, the change of variable $x' = 2^j(x-u)$ in (E.24) gives

$$\begin{aligned} \int |k_{j,2}(x,u)|dx \le 2^d \int \bigg| \int_0^1 \nabla h\big(&(\mathrm{Id} - t\,\nabla\tau(u))x' \\ &+ (1-t)2^j(\tau(u) - \tau(2^{-j}x' + u))\big)2^j\alpha(u, 2^{-j}x')dt\bigg| dx'. \end{aligned}$$

For any $0 \le t \le 1$

$$\begin{aligned} \big|(\mathrm{Id} - t\nabla\tau(u))x' + (1-t)2^j(\tau(2^{-j}x' + u) - \tau(u))\big| &\ge \\ |x'|(1 - \|\nabla\tau\|_\infty) &\ge \frac{|x'|}{2}. \end{aligned}$$

Equation (E.23) also implies that

$$
\begin{aligned}
|2^j \alpha(u, 2^{-j} x')| &= 2^{-j} \left| \int_0^1 t x' H\tau(u + (1-t)2^{-j} x') x' \, dt \right| \\
&\leq 2^{-j} \|H\tau\|_\infty \frac{|x'|^2}{2}.
\end{aligned} \tag{E.25}
$$

Since $|\nabla h(u)| \leq C(1+|u|)^{-d-2}$, with the change of variable $x = x'/2$ we get

$$
\int |k_{j,2}(x,u)| dx \leq C\, 2^{-j} \|H\tau\|_\infty. \tag{E.26}
$$

For $j \leq 0$, we use a maximum error bound on the remainder α of the Taylor approximation (E.22):

$$
|2^j \, \alpha(u, 2^{-j} x')| \leq 2 \, \|\nabla \tau\|_\infty \, |x'|,
$$

which proves that $\int |k_{j,2}(x,u)| dx \leq C \|\nabla\tau\|_\infty$ and hence that

$$
\int |k_{j,2}(x,u)| dx \leq C \, \min(2^{-j} \|H\tau\|_\infty, \|\nabla \tau\|_\infty). \tag{E.27}
$$

Similarly, with the change of variable $u' = 2^j(x - u)$, we compute $\int |k_{j,2}(x,u)| du$, which leads to the same bound (E.27). Schur's lemma gives

$$
\|K_{j,2}\| \leq C \, \min(2^{-j} \|H\tau\|_\infty, \|\nabla \tau\|_\infty), \tag{E.28}
$$

which finishes the proof of (E.19).

Let us now compute the upper bound (E.18) on $K_{j,1}$. Its kernel $k_{j,1}$ in (E.15) can be written $k_{j,1}(x,u) = 2^{dj} g(u, 2^j(x-u))$ with

$$
g(u,v) = h(v) - h((\mathrm{Id} - \nabla\tau(u))v) \det(\mathrm{Id} - \nabla\tau(u)). \tag{E.29}
$$

A first-order Taylor decomposition of h gives

$$
\begin{aligned}
g(u,v) &= \big(1 - \det(\mathrm{Id} - \nabla\tau(u))\big) h(\mathrm{Id} - \nabla\tau(u)v) \\
&\quad + \int_0^1 \nabla h\big((1-t)v + t(\mathrm{Id} - \nabla\tau(u))v\big) \cdot \nabla\tau(u) v \, dt.
\end{aligned} \tag{E.30}
$$

Since $\det(\mathrm{Id} - \nabla\tau(u)) \geq (1 - \|\nabla\tau\|_\infty)^d$, we get $(1 - \det(\mathrm{Id} - \nabla\tau(u))) \leq d \, \|\nabla\tau\|_\infty$. Moreover, $\|\nabla\tau\|_\infty \leq \frac{1}{2}$, and $h(x)$ as well as its partial derivatives have a decay that is $O((1+|x|)^{-d-2})$, so

$$
|g(u,v)| \leq C \|\nabla\tau\|_\infty (1+|v|)^{-d-2}, \tag{E.31}
$$

so $k_{j,1}(x,u) = O(2^{dj} \, \|\nabla\tau\|_\infty (1 + 2^j |x-u|)^{-d-2})$. Since

$$
\int |k_{j,1}(x,u)| du = O(\|\nabla\tau\|_\infty) \quad \text{and} \quad \int |k_{j,1}(x,u)| dx = O(\|\nabla\tau\|_\infty),
$$

Schur's lemma (B.1) proves that $\|K_{j,1}\| = O(\|\nabla\tau\|_\infty)$ and hence (E.18).

Let us now prove (E.21) when $\int h(x)dx = 0$. The kernel of the self-adjoint operator $Q_j = K_{j,1}^* K_{j,1}$ is

$$\begin{aligned}\bar{k}_j(y,z) &= \int k_{j,1}^*(x,y) k_{j,1}(x,z)dx \\ &= \int 2^{2dj} g^*(y, 2^j(x-y)) g(z, 2^j(x-z))dx \\ &= \int 2^{dj} g^*(y, x' + 2^j(z-y)) g(z,x'))dx'.\end{aligned} \tag{E.32}$$

The singular kernel $\bar{k} = \sum_j \bar{k}_j$ of $\sum_j Q_j$ almost satisfies the hypotheses of the T(1) theorem of David, Journé, and Semmes [5] but not quite because it does not satisfy the decay condition $|\bar{k}(y,z) - \bar{k}(y,z')| \le C|z'-z|^\alpha |z-y|^{-d-\alpha}$ for some $\alpha > 0$. We bound this operator with Cotlar's lemma [20], which proves that if Q_j satisfies

$$\forall j,l \quad \|Q_j^* Q_l\| \le |\beta(j-l)|^2 \quad \text{and} \quad \|Q_j Q_l^*\| \le |\beta(j-l)|^2, \tag{E.33}$$

then

$$\Big\| \sum_j Q_j \Big\| \le \sum_j \beta(j). \tag{E.34}$$

Since Q_j is self-adjoint, it is sufficient to bound $\|Q_l\, Q_j\|$. The kernel of $Q_l\, Q_j$ is computed from the kernel $\bar{k}_j$ of Q_j,

$$\bar{k}_{l,j}(y,z) = \int \bar{k}_j(z,u) \bar{k}_l(y,u)du. \tag{E.35}$$

An upper bound of $\|Q_l\, Q_j\|$ is obtained with Schur's lemma (B.1) applied to $\bar{k}_{l,j}$. Inserting (E.32) in (E.35) gives

$$\begin{aligned}\int |\bar{k}_{l,j}(y,z)|dy = \int \Big| \int\!\!\int g(u,x) g(u,x') 2^{dl}\, g^*(y, x + 2^l(u-y)) \\ \times 2^{dj} g^*(z, x' + 2^j(u-z))dx\, dx'\, du \Big| dy.\end{aligned} \tag{E.36}$$

The parameters j and l have symmetrical roles and we can thus suppose that $j \ge l$.

Since $\int h(x)dx = 0$ we have from (E.29) that $\int g(u,v)dv = 0$ for all u. For $v = (v_n)_{n \le d}$, one can thus write $g(u,v) = \partial \bar{g}(u,v)/\partial v_1$, and (E.31) implies that

$$|\bar{g}(u,v)| \le C\|\nabla\tau\|_\infty (1 + |v|)^{-d-1}. \tag{E.37}$$

Let us make an integration by parts along the variable u_1 in (E.36). Since all first- and second-order derivatives of $h(x)$ have a decay that is $O((1+|x|)^{-d-2})$, we derive from (E.29) that for any $u = (u_n)_{n \le d} \in \mathbb{R}^d$ and $v = (v_n)_{n \le d} \in \mathbb{R}^d$,

$$\left| \frac{\partial g(u,v)}{\partial u_1} \right| \le C\|H\tau\|_\infty (1 + |v|(1 - \|\nabla\tau\|_\infty))^{-d-1}, \tag{E.38}$$

and from (E.30)

$$\left|\frac{\partial g(u,v)}{\partial v_1}\right| \le C\,\|\nabla\tau\|_\infty (1+|v|(1-\|\nabla\tau\|_\infty))^{-d-1}. \tag{E.39}$$

In the integration by parts, integrating $2^{dj} g(z, x' + 2^j(u-z))$ brings out a term proportional to 2^{-j}, and differentiating $g(u,x)g(u,x')2^{dl} g(y, x+2^l(u-y))$ brings out a term bounded by 2^l. An upper bound of (E.36) is obtained by inserting (E.31), (E.37), (E.38), and (E.39), which proves that there exists C such that

$$\begin{aligned}\int |\bar{k}_{l,j}(y,z)|\,dy &\le C^2 (2^{-j}\,\|\nabla\tau\|_\infty^3\,\|H\tau\|_\infty + 2^{l-j}\,\|\nabla\tau\|_\infty^4)\\ &\le C^2\,2^{l-j}(\|\nabla\tau\|_\infty + \|H\tau\|_\infty)^4.\end{aligned}$$

The same calculation proves the same bound on $\int |\bar{k}_{l,j}(y,z)|\,dz$, so Schur's lemma (B.1) implies that

$$\|Q_l\,Q_j\| \le C^2\,2^{l-j}(\|\nabla\tau\|_\infty + \|H\tau\|_\infty)^4.$$

Applying Cotlar's lemma (E.33) with $\beta(j) = C\,2^{-|j|/2}(\|\nabla\tau\|_\infty + \|H\tau\|_\infty)^2$ proves that

$$\left\|\sum_{j=-\infty}^{+\infty} K_{j,1}^* K_{j,1}\right\| = \left\|\sum_j Q_j\right\| \le C(\|\nabla\tau\|_\infty + \|H\tau\|_\infty)^2, \tag{E.40}$$

which implies (E.21).

Appendix F Proof of Lemma 3.6

We have from (3.15) that there exists ϵ_J with $\lim_{J\to\infty}\epsilon_J = 0$ such that

$$\sup_{p\in\mathcal{P}_J-\Omega_J^f}\left\|S_J[p]f - \frac{\|S_J[p]f\|}{\|S_J[p]\delta\|}S_J[p]\delta\right\|^2 \le \frac{\epsilon_J}{2}\|S_J[p]f\|^2,$$

and $\sum_{p\in\Omega_J^f}\|S_J[p]f\|^2 \le \epsilon_J\|f\|^2/8$. Since $\|S_J[\mathcal{P}_J]f\|^2 = \|f\|^2$, we get

$$\sum_{p\in\mathcal{P}_J}\left\|S_J[p]f - \frac{\|S_J[p]f\|}{\|S_J[p]\delta\|}S_J[p]\delta\right\|^2 \le \epsilon_J\|f\|^2. \tag{F.1}$$

The set of all extensions of a $p\in\mathcal{P}_J$ into $\mathcal{P}_{J+1}$ is defined in (2.37). It can be rewritten $\mathcal{P}_{J+1}^p = \mathcal{P}_{J+1}\cap C_J(p)$, and (2.38) proves that

$$\|S_J[p]f - S_J[p]h\|^2 \ge \sum_{p'\in\mathcal{P}_{J+1}\cap C_J(p)}\|S_{J+1}[p']f - S_{J+1}[p']h\|^2.$$

Iterating k times on this result yields

$$\|S_J[p]f - S_J[p]h\|^2 \ge \sum_{p'\in\mathcal{P}_{J+k}\cap C_{J+k}(p)}\|S_{J+k}[p']f - S_{J+k}[p']h\|^2.$$

Applying it to f and $h = \mu_p \delta$ with $\mu_p = \|S_J[p]f\| / \|S_J[p]\delta\|$ gives

$$\|S_J[p]f - \mu_p S_J[p]\delta\|^2 \geq \sum_{p' \in \mathcal{P}_{J+k} \cap C_{J+k}(p)} \|S_{J+k}[p']f - \mu_p S_{J+k}[p']\delta\|^2.$$

Summing over $p \in \mathcal{P}_J$ and applying (F.1) proves that

$$\sum_{p \in \mathcal{P}_J} \sum_{p' \in \mathcal{P}_{J+k} \cap C_{J+k}(p)} \left\| S_{J+k}[p']f - \frac{\|S_J[p]f\|}{\|S_J[p]\delta\|} S_{J+k}[p']\delta \right\|^2 \leq \epsilon_J \|f\|^2,$$

and hence

$$\sum_{p \in \mathcal{P}_J} \sum_{p' \in \mathcal{P}_{J+k} \cap C_{J+k}(p)} \left| \frac{\|S_{J+k}[p']f\|}{\|S_{J+k}[p']\delta\|} - \frac{\|S_J[p]f\|}{\|S_J[p]\delta\|} \right|^2 \|S_{J+k}[p']\delta\|^2 \leq \epsilon_J \|f\|^2.$$

If $q \in C_{J+k}(p')$, then $S_{J+k}(q) = \|S_{J+k}[p']f\| / \|S_{J+k}[p']\delta\|$. But $p' \in C_J(p)$ so $q \in C_J(p)$ and hence $S_J(q) = \|S_J[p]f\| / \|S_J[p]\delta\|$.

Finally, $\|S_{J+k}[p']\delta\|^2 = \mu(C_{J+k}(p'))$, so the sum can be rewritten as a path integral

$$\int_{\mathcal{P}^\infty} |S_{J+k} f(q) - S_J f(q)|^2 \, d\mu(q) \leq \epsilon_J \|f\|^2,$$

which proves that $\{S_J f\}_{J \in \mathbb{N}}$ is a Cauchy sequence in $\mathbf{L}^2(\bar{\mathcal{P}}_\infty, d\mu)$.

Appendix G Proof of Lemma 4.8

This appendix proves that

$$E(|K_\tau X|^2) \leq E(\|K_\tau\|^2) E(|X|^2), \tag{G.1}$$

as well as a generalization to the sequence of operators at the end of the appendix. The lemma result is proved by restricting X to a finite hypercube $I_T = \{(x_1, x_2, \ldots, x_d) \in \mathbb{R}^d : \forall i \leq d, |x_i| \leq T\}$ whose indicator function $\mathbb{1}_{I_T}$ defines a finite energy process $X_T(x) = X(x)\mathbb{1}_{I_T}(x)$. We shall verify that $E(|K_\tau X(x)|^2)$ does not depend upon x and that

$$E(|K_\tau X(x)|^2) = \lim_{T \to \infty} \frac{E(\|K_\tau X_T\|^2)}{(2T)^d}. \tag{G.2}$$

We first show how this result implies (G.1). The $\mathbf{L}^2(\mathbb{R}^d)$ operator norm definition implies

$$\|K_\tau X_T\|^2 = \int |K_\tau X_T(x)|^2 \, dx \leq \|K_\tau\|^2 \int |X_T(x)|^2 \, dx.$$

Since X and τ are independent processes,

$$E(\|K_\tau X_T\|^2) \leq E(\|K_\tau\|^2) E(|X|^2)(2T)^d.$$

Applying (G.2) thus proves the lemma result (G.1).

To prove (G.2), we first compute

$$E(|K_\tau X(x)|^2) = E\left(\iint k_\tau(x,u)k_\tau^*(x,u')X(u)X^*(u')du\,du'\right).$$

Since X is stationary, $E(X(u)X^*(u')) = A_X(u-u')$, and the lemma hypothesis supposes that $E(k_\tau(x,u)k_\tau^*(x,u')) = \overline{k}_\tau(x-u,x-u')$. Since X and τ are independent, the change of variable $v = x-u$ and $v' = x-u'$ gives

$$E(|K_\tau X(x)|^2) = \iint \overline{k}_\tau(x-u,x-u')A_X(u-u')du\,du'$$

$$= \iint \overline{k}_\tau(v,v')A_X(v-v')dv\,dv', \tag{G.3}$$

which proves that $E(|K_\tau X(x)|^2)$ does not depend upon x.

Similarly,

$$E(|K_\tau X_T(x)|^2) = \iint \overline{k}_\tau(v,v')A_X(v-v')\mathbb{1}_{I_T}(v-x)\mathbb{1}_{I_T}(v'-x)dv\,dv', \tag{G.4}$$

and integrating along x gives

$$(2T)^{-d}\,E(\|K_\tau X_T\|^2) = \iint \overline{k}_\tau(v,v')A_X(v-v')(1-\rho_T(v-v'))dv\,dv', \tag{G.5}$$

with

$$1-\rho_T(v-v') = (2T)^{-d}\int \mathbb{1}_{I_T}(v-x)\mathbb{1}_{I_T}(v'-x)dx$$

$$= \prod_{i=1}^d \left(1-\frac{|v_i-v_i'|}{2T}\right)\mathbb{1}_{I_T}(v-v')$$

and hence

$$0 \le \rho_T(v) \le (2T)^{-1}\sum_{i=1}^d |v_i| \le d(2T)^{-1}|v|. \tag{G.6}$$

Inserting (G.3) into (G.5) proves that

$$(2T)^{-d}E(\|K_\tau X_T\|^2) = E(|K_\tau X(x)|^2) - \iint \overline{k}_\tau(v,v')A_X(v-v')\rho_T(v-v')dv\,dv'. \tag{G.7}$$

Since $\iint |\overline{k}_\tau(v,v')|\,|v-v'|dv\,dv' < \infty$ and $A_X(v-v') \le A_X(0) = E(|X|^2)$, we have from (G.7) and (G.6) that

$$\lim_{T\to\infty}(2\,T)^{-d}E(\|K_\tau X_T\|^2) = E(|K_\tau X(x)|^2),$$

which proves (G.2).

Lemma 4.8 is extended to sequences of operators $\bar{K}_\tau = \{K_{\tau,n}\}_{n\in I}$ with kernels $\{k_{\tau,n}\}_{n\in I}$ as follows: Let us denote

$$\|\bar{K}_\tau X\|^2 = \sum_{n\in I} |K_{\tau,n} X|^2 \quad \text{and} \quad \|\bar{K}_\tau f\|^2 = \sum_{n\in I} \|K_{\tau,n} f\|^2. \tag{G.8}$$

If each average bilinear kernel is stationary,

$$E(k_{\tau,n}(x,u)k^*_{\tau,n}(x,u')) = \bar{k}_{\tau,n}(x-u, x-u') \tag{G.9}$$

and

$$\iint \Big|\sum_{n\in I} \bar{k}_{\tau,n}(v,v')\Big| |v-v'| dv\, dv' < \infty, \tag{G.10}$$

then

$$E(\|\bar{K}_\tau X\|^2) \le E(\|\bar{K}_\tau\|^2) E(|X|^2). \tag{G.11}$$

The proof of this extension follows the same derivations as the proof of (G.1) for a single operator. It just requires replacing the $\mathbf{L}^2(\mathbb{R}^d)$ norm $\|f\|^2$ by the norm $\sum_{n\in I} \|f_n\|^2$ over the space of finite energy sequences $\{f_n\}_{n\in I}$ of $\mathbf{L}^2(\mathbb{R}^d)$ functions and the sup operator norms in $\mathbf{L}^2(\mathbb{R}^d)$ by sup operator norms on a sequence of $\mathbf{L}^2(\mathbb{R}^d)$ functions.

Appendix H Proof of Theorem 4.7

This appendix proves $E(\|[S_J[\mathcal{P}_J], L_\tau]X\|^2) \le E(\|U[\mathcal{P}_J]X\|_1)^2\, B(\tau)$ with

$$B(\tau) = CE\left(\left(\|\nabla\tau\|_\infty\left(\log \frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty} \vee 1\right) + \|H\tau\|_\infty\right)^2\right) \tag{H.1}$$

and

$$E(\|U[\mathcal{P}_J]X\|_1) = \sum_{m=0}^{+\infty}\Big(\sum_{p\in\Lambda_J^m} E(|U[p]X|^2)\Big)^{1/2}.$$

For this purpose, we shall first prove that if for any stationary process X

$$E(\|[W_J, L_\tau]X\|^2) \le B(\tau)E(|X|^2) \tag{H.2}$$

where

$$E(\|[W_J, L_\tau]X\|^2) = E(|[A_J, L_\tau]X|^2) + \sum_{\lambda\in\Lambda_J} E(|[W[\lambda], L_\tau]X|^2),$$

then

$$E(\|[S_J[\mathcal{P}_J], L_\tau]X\|^2) \le B(\tau)E(\|U[\mathcal{P}_J]X\|_1)^2. \tag{H.3}$$

Since a modulus operator is nonexpansive and commutes with L_τ, with the same argument as in the proof of (2.55), we derive from (H.2) that

$$E(\|[U_J, L_\tau]X\|^2) \le B(\tau)E(|X|^2). \tag{H.4}$$

The proof of Proposition 4.2 also shows that U_J is nonexpansive for the mean-square norm on processes. Since $S_J[\mathcal{P}_J]$ is obtained by iterating on U_J, we have that

$$E(\|[S_J[\mathcal{P}_J], L_\tau]X\|^2) \le B(\tau) E(\|U[\mathcal{P}_J]X\|_1)^2.$$

The proof of this inequality follows the same derivations as in Appendix D for $L = L_\tau$ by replacing f by X, $\|f\|^2$ by $E(|X|^2)$, $\|U[p]f\|^2$ by $E(|U[p]X|^2)$, and the $\mathbf{L}^2(\mathbb{R}^d)$ sup operator norm $\|[U_J, L]\|$ by $B(\tau)$, which satisfies (H.4) for all X.

The proof of (H.1) is ended by verifying that

$$E(\|[W_J, L_\tau]X\|^2) \le E(C^2(\tau)) E(|X|^2) \tag{H.5}$$

and hence $B(\tau) = E(C^2(\tau))$ with

$$C(\tau) = C\left(\|\nabla\tau\|_\infty \left(\log \frac{\|\Delta\tau\|_\infty}{\|\nabla\tau\|_\infty} \vee 1\right) + \|H\tau\|_\infty\right).$$

The inequality (H.5) is derived from Lemma 2.14, which proves that the $\mathbf{L}^2(\mathbb{R}^d)$ operator norm of the commutator $[W_J, L_\tau]$ satisfies

$$\|[W_J, L_\tau]\| \le C(\tau). \tag{H.6}$$

Let us apply to $\overline{K}_\tau = [W_J, L_\tau] = \{[A_J, L_\tau], [W[\lambda], L_\tau]\}_{\lambda \in \Lambda_J}$ the extension (G.11) of Lemma 4.8. This extension proves that if the kernels of the wavelet commutator satisfy the conditions (G.9) and (G.10), then

$$E(\|[W_J, L_\tau]X\|^2) \le E(\|[W, L_\tau]\|^2)\, E(|X|^2).$$

Together with (H.6), it proves (H.5).

To finish the proof, we verify that the wavelet commutator kernels satisfy (G.9) and (G.10). If $Z_j f(x) = f \star h_j(x)$ with $h_j(x) = 2^{dj} h(2^j x)$, then the kernel of the integral commutator operator $[Z_j, L_\tau] = Z_j L_\tau - L_\tau Z_j$ is

$$\begin{aligned} k_{\tau,j}(x,u) &= h_j(x - u - \tau(x)) \\ &\quad - h_j\bigl(x - u - \tau\bigl(u + \tau(\beta(u))\bigr)\bigr) \bigl|\det\bigl(\mathrm{Id} - \nabla\tau\bigl(u + \tau(\beta(u))\bigr)\bigr)\bigr|^{-1} \end{aligned} \tag{H.7}$$

where β is defined by $\beta(x) = x + \tau(\beta(x))$. The kernel of $[A_J, L_\tau]$ is $k_{\tau,J}$ with $h = \phi$, and the kernel of $[W[\lambda], L_\tau]$ for $\lambda = 2^j r$ is $k_{\tau,j}$ with $h(x) = \psi(r^{-1}x)$. Since τ and $\nabla\tau$ are jointly stationary, the joint probability distribution of their values at x and $u + \tau(\beta(u))$ depends only upon $x - u$. We have that $E(k_{\tau,j}(x,u) k_{\tau,j}(x,u')) = \overline{k}_{\tau,j}(x-u, x-u')$, which proves the kernel stationarity (G.9) for wavelet commutators.

The second kernel hypothesis (G.10) is proved by showing that if $|h(x)| = O((1+|x|)^{-d-2})$, then

$$\iint \Bigl| \sum_{j \ge -J} \overline{k}_{\tau,j}(v, v') \Bigr| \, |v - v'| \, dv\, dv' < \infty.$$

Since $\overline{k}_{\tau,j}(v, v') = E(k_{\tau,j}(x, x - v)k_{\tau,j}(x, x - v'))$, it is sufficient to prove that there exists C such that for all x, with probability 1,

$$I = \sum_{j \geq -J} \iint |k_{\tau,j}(x, x - v)|\, |k_{\tau,j}(x, x - v')|\, |v - v'| dv\, dv' \leq C. \tag{H.8}$$

Since $h_j(x) = 2^{dj} h(2^j x)$ and $u + \tau(\beta(u)) = \beta(u)$, we have from (H.7) that $k_{\tau,j}(x, x - 2^{-j} w) = 2^{dj}\, \widetilde{k}_{\tau,j}(x, x - w)$ with

$$\begin{aligned} \widetilde{k}_{\tau,j}(x, x - w) &= h(w - 2^j \tau(x)) \\ &\quad - h\big(w - 2^j \tau(\beta(x - 2^{-j} w))\big)\big|\det\big(\mathrm{Id} - \nabla\tau(\beta(x - 2^{-j} w))\big)\big|^{-1}. \end{aligned} \tag{H.9}$$

The change of variable $w = 2^j v$ and $w' = 2^{-j} v'$ in (H.8) shows that $I = \sum_{j \geq -J} 2^{-j} I_j$ with

$$I_j = \iint |\widetilde{k}_{\tau,j}(x, x - w)|\, |\widetilde{k}_{\tau,j}(x, x - w')|\, |w - w'| dw\, dw'.$$

Since $|h(w)| = O((1 + |w|)^{-d-2})$ and $\|\nabla\tau\|_\infty \leq \frac{1}{2}$ with probability 1, by computing separately the integrals of each of the four terms of the product

$$|\widetilde{k}_{\tau,j}(x, x + w)|\, |k_{\tau,j}(x, x + w')|\, |w - w'|,$$

with a change of variables $y = w + 2^j \tau(x)$ and $z = w + 2^j \tau(\beta(x + 2^{-j} w))$, we verify that there exists C' such that $I_j \leq C'$ and hence that $I = \sum_{j \geq -J} 2^{-j} I_j \leq 2^{J+1} C'$ with probability 1. This proves (H.8) and hence the second kernel hypothesis (G.10).

Acknowledgment. I would like to thank Joan Bruna, Mike Glinsky, and Nir Soren for the many inspiring conversations in connection with image processing, physics, and group theory. This work was supported by grant ANR-10-BLAN-0126.

Bibliography

[1] Andén, J; Mallat, S. Multiscale scattering for audio classification. *Proceedings of the International Society for Music Information Retrieval (Miami, 2011)*, 657–662. Available at: http://ismir2011.ismir.net/papers/PS6-1.pdf

[2] Bouvrie, J.; Rosasco, L.; Poggio, T. On invariance in hierarchical models. *Advances in Neural Information Processing Systems 22*, 162–170, 2009.

[3] Bruna, J.; Mallat, S. Classification with scattering operators. *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2011), 1561–1566. doi:10.1109/CVPR.2011.5995635

[4] Bruna, J.; Mallat, S. Invariant scattering convolution networks. *IEEE Trans. PAMI*, to appear.

[5] David, G.; Journé, J.-L.; Semmes, S. Opérateurs de Calderón-Zygmund, fonctions para-accrétives et interpolation. (French) [Calderón-Zygmund operators, para-accretive functions and interpolation] *Rev. Mat. Iberoamericana* **1** (1985), no. 4, 1–56.

[6] Frankel, T. *The geometry of physics: an introduction*. 2nd ed. Cambridge University Press, Cambridge, 2004.

[7] Frazier, M.; Jawerth, B.; Weiss, G. *Littlewood-Paley theory and the study of function spaces.* CBMS Regional Conference Series in Mathematics, 79. Published for the Conference Board of the Mathematical Sciences, Washington, D.C., by the American Mathematical Society, Providence, R.I., 1991.

[8] Geller, D.; Pesenson, I. Z. Band-limited localized Parseval frames and Besov spaces on compact homogeneous manifolds. *J. Geom. Anal.* **21** (2011), no. 2, 334–371.

[9] Glinsky, M. E. A new perspective on renormalization: invariant actions, a dynamical DNA. Preprint, 2011. Available at: arXiv:1106.4369v1 [physics.gen-ph]

[10] Hörmander, L. Fourier integral operators. I. *Acta Math.* **127** (1971), no. 1, 79–183.

[11] LeCun, Y.; Kavukvuoglu, K.; Farabet, C. Convolutional networks and applications in vision. *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS) (New York, 2010)*, 253–256.

[12] Lohmiller, W.; Slotine, J. J. E. On contraction analysis for nonlinear systems. *Automatica* **34** (1998), no. 6, 683–696.

[13] Mallat, S. Recursive interferometric representation. *Proceedings of EUSICO Conference (Danemark, 2010)*.

[14] Meyer, Y. *Wavelets and operators*. Cambridge Studies in Advanced Mathematics, Vol. 37. Cambridge University Press, Cambridge, 1992.

[15] Olver, P. J. *Equivalence, invariants and symmetry*. Cambridge University Press, Cambridge, 1995.

[16] Rahman, I. U.; Drori, I.; Stodden, V. C.; Donoho, D. L.; Schröder, P. Multiscale representations for manifold-valued data. *Multiscale Model. Simul.* **4** (2005), no. 4, 1201–1232.

[17] Riesenhuber, M.; Poggio, T. Hierarchical models of object recognition in cortex. *Nat. Neurosci.* **2** (1999), 1019–1025.

[18] Sifre, L.; Mallat, S. Combined scattering for rotation invariant texture analysis. *European Symposium on Artificial Neural Networks, 2012*, submitted.

[19] Stein, E. M. *Topics in harmonic analysis related to the Littlewood-Paley theory*. Annals of Mathematics Studies, 63. Princeton University Press, Princeton, N.J.; University of Tokyo Press, Tokyo, 1970.

[20] Stein, E. M. *Harmonic analysis: real-variable methods, orthogonality, and oscillatory integrals*. (English summary) With the assistance of Timothy S. Murphy. Princeton Mathematical Series, 43. Monographs in Harmonic Analysis, III. Princeton University Press, Princeton, N.J., 1993.

[21] Trouvé, A.; Younes, L. Local geometry of deformable templates. *SIAM J. Math. Anal.* **37** (2005), no. 1, 17–59 (electronic). doi:10.1137/S0036141002404838

[22] Willard, S. *General topology*. Addison-Wesley, Reading, Mass.–London–Don Mills, Ont., 1970.

STÉPHANE MALLAT
Ecole Polytechnique, CMAP
91128 PALAISEAU CEDEX
FRANCE
E-mail: `mallat@cmap.polytechnique.fr`

Received January 2011.
Revised November 2011.