Group Invariant Scattering

Stéphane Mallat
CMAP, Ecole Polytechnique, Palaiseau, France[*]

January 12th, 2011

Abstract

Pattern classification often requires using translation invariant representations, which are stable and hence Lipschitz continuous to deformations. A Fourier transform does not provide such Lipschitz stability. Scattering operators are obtained by iterating on wavelet transforms and modulus operators. The resulting representation is proved to be translation invariant and Lipschitz continuous to deformations, up to a log term. It is computed with a non-linear convolution network, which scatters functions along an infinite set of paths. Invariance to the action of any compact Lie subgroup of $GL(\mathbb{R}^d)$ is obtained with a combined scattering, which iterates over wavelet transforms defined on this group. Scattering representations yield new metrics on stationary processes, which are stable to random deformations.

1 Introduction

A core issue of classification is to build metrics which reflect the similarity of signals according to the classification goals. Signals of a same class should have a small distance whereas the distance with elements of other classes should be large. If a class is invariant to rigid transformations such as translations or rotations then a metric having the same invariant reduces intra-class distances. Signals and images classes often satisfy stronger local invariance to small elastic deformations. Taking advantage of these higher

[*]This work is funded by the French ANR.

dimensional invariants requires using a metric which satisfy some form of Lipschitz continuity to linear deformations.

Kernel classification algorithms rely on distances that can be written

$$d(f,g) = \|\Phi(f) - \Phi(g)\| \ ,$$

where Φ is an operator which maps f and g in a Hilbert space which is typically much larger then the original space where they live in [18]. Imposing invariance rather defines operators that reduce the dimensionality, which provides the ability to learn with limited number of training examples. Reducing dimensionality may however also reduce the distance and hence the discriminability of signals which belong to different classes.

Let us first consider translation invariance. A signal class $\mathcal{C}$ is translation invariant if for all $f \in \mathcal{C}$ and $\tau \in \mathbb{R}^d$, $T_\tau f(x) = f(x-\tau) \in \mathcal{C}$. Translation invariant metrics incorporate this invariance by imposing that

$$d(f, T_\tau f) = \|\Phi(f) - \Phi(T_\tau f)\| = 0 \ , \tag{1}$$

which implies that Φ is invariant: $\Phi(T_\tau f) = \Phi(f)$. Global translation invariance is a weak invariant whose equivalence classes $\{f(x-\tau)\}_{\tau \in \mathbb{R}^d}$ are d-dimensional manifolds in $\mathbf{L}^2(\mathbb{R}^d)$. Local translation invariance is a much stronger condition, which requires deformation stability.

Signal and image classes are often invariant to small deformations [1, 2, 22]. For example, small deformations of hand-written digits do not change their classification. An operator $D_\tau f(x) = f(x - \tau(x))$ can be approximated by a translation in the neighborhood of x if $\tau(x)$ has a small gradient tensor $\nabla\tau(x)$. This gradient yields a small linear deformation $L_\beta f(u) = f(\beta u)$, where $\beta = \mathbf{1} - \nabla\tau(x)$ is in the group $GL(\mathbb{R}^d)$ of linear invertible matrices. We write $|\beta|$ the matrix sup norm of β and $\mathbf{1}$ the identity matrix. To be nearly invariant to an elastic deformation D_τ, the representation must have small variations when $|\nabla\tau(x)| = |\mathbf{1} - \beta|$ is small. This is expressed as a Lipschitz regularity condition of Φ relatively to $\beta \in GL(\mathbb{R}^d)$. The representation Φ is said to be locally Lipschitz continuous if for all β in the neighborhood of $\mathbf{1}$ there exists $C > 0$ such that

$$\|\Phi(f) - \Phi(T_\tau L_\beta f)\| \leq C \, \|f\| \, |\mathbf{1} - \beta| \ . \tag{2}$$

For $\beta = \mathbf{1} - \nabla\tau(x)$, it becomes a Lipschitz continuity property relatively to elastic deformations:

$$\|\Phi(f) - \Phi(D_\tau f)\| \leq C \, \|f\| \, \sup_x |\nabla\tau(x)| \ . \tag{3}$$

The Fourier transform modulus $\Phi(f) = |\hat{f}|$ defines a translation invariant representation, but Section 2 proves that it is not Lipschitz continuous to deformations. Indeed, local deformations may severely affect high frequencies. Moreover, the loss of the Fourier phase strongly reduces the discriminability of signals such as images or audio recordings, which include localized transient structures. Preserving discriminability while removing the phase with a modulus operator requires using sparse representations.

We study a general class of scattering operators Φ introduced in [14], which are proved to be globally and locally translation invariant, thanks to their Lipschitz continuity relatively to deformations. They are constructed by iterating on a contractive operator which is the modulus of a complex wavelet transform. This wavelet transform modulus maps signal high frequencies into lower frequency interferences. The resulting scattering metric is contractive and preserves $\mathbf{L}^2(\mathbb{R}^d)$ norms for appropriate complex wavelets. A scattering operator transforms f into infinite number of functions $S_J(p)f$ indexed by a path variables p. This is analogous to multiple path scattering propagations in physics. It is computed with a cascade of convolutions and modulus operators, which belongs to the general class of convolution network architectures introduced by LeCun et. al. [11]. Translation invariance and Lipschitz regularity linearizes deformations. Deformed patterns such as hand-written digits [10] can thus be classified with simple affine space models on their scattering transform [5].

Translation invariance leads to an invariant scattering representation of stationary processes, which is proved to be stable to random elastic deformations. The scattering metric can discriminate stationary processes having same power spectra, and it remains Lipschitz continuous to random deformations, up to a log term. Consistent estimators discriminate stationary processes from single realizations, which applies to image texture classification [5].

Classification may require more then translation invariance. Invariance to the action of compact Lie subgroups G of $GL(\mathbb{R}^d)$, such as rotations, is obtained with combined scattering operators in $\mathbf{L}^2(\mathbb{R}^d)$ and in $\mathbf{L}^2(G)$. As a result, the scattering distance $\|\Phi(f) - \Phi(D_\tau f)\|$ does not depend on the component of $\nabla\tau$ which is in $\mathbf{L}^2(G)$. For rotation invariance, the scattering is computed in space and along rotation angle variables.

Section 2 begins by explaining the inability of Fourier modulus and wavelet representations to provide translation invariance together with Lipschitz continuity to deformations. Difficulties arise because of high frequencies. Scattering operators are introduced in Section 3. A first theorem proves that the energy of scattered signals $S(p)f$ is concentrated along very few paths p, and that it preserves the $\mathbf{L}^2(\mathbb{R}^d)$ norm. Translation invariance and Lipschitz continuity to deformations are proved in the main theorem of Section 3.4. It relies on the properties of wavelet transform commutators with D_τ. Section 4 introduces a periodic scattering for compactly supported functions, with a fast computational algorithm.

Scattering representations of stationary processes are studied in Section 5. A theorem proves that Lipschitz continuity is preserved for stationary random deformations. The consistency of scattering metric estimation is studied with an open conjecture. Section 6 extends translation invariance to compact Lie groups G, with combined scattering operators. A combined scattering is proved to be locally invariant to translations and to the action of G.

2 Fourier and Wavelet Invariance

A Fourier transform modulus yields a translation invariant representation, but Section 2.1 shows that high frequencies are not stable to elastic deformations. The resulting phase removal also produces an important loss of discriminability, which is related to sparsity in Section 2.2. Wavelets are localized functions and the resulting transform is stable to deformations. It also provides sparse representations of large classes functions, but Section 2.3 shows that high frequencies are not translation invariant.

2.1 Fourier Modulus Invariance

The Fourier transform of $f \in \mathbf{L}^2(\mathbb{R}^d)$ is written $\hat{f}(\omega)$ for $\omega \in \mathbb{R}^d$. A translation $T_\tau f(x) = f(x-\tau)$ produces a Fourier transform phase shift: $\widehat{T_\tau f}(\omega) = e^{i\tau\omega}\hat{f}(\omega)$. The Fourier modulus $\Phi(f) = |\hat{f}|$ is thus invariant to global translations: $|\widehat{T_\tau f}| = |\hat{f}|$.

Local translation invariance requires stability to elastic deformations, which is expressed as a Lipschitz regularity relatively to linear transformations. The Fourier transform of $L_\beta f(x) = f(\beta x)$ with $\beta \in GL(\mathbb{R}^d)$ is

$$\widehat{L_\beta f}(\omega) = |\det \beta|^{-1}\, \hat{f}(\beta^{-1*}\,\omega) \tag{4}$$

where β^{-1*} is the adjoint of β^{-1} and $\det \beta$ is the determinant of β. Local Lipschitz continuity means that there exists $C > 0$ such that for all $f \in \mathbf{L}^2(\mathbb{R}^d)$ and all $\beta \in GL(\mathbb{R}^d)$ in a neighborhood of $\mathbf{1}$

$$\| |\widehat{L_\beta f}| - |\hat{f}| \| \leq C \, \|f\| \, |\mathbf{1} - \beta| \, . \tag{5}$$

This property is not valid for the Fourier transform modulus because eventhough $|\mathbf{1} - \beta|$ may be small, $|\beta^{-1*}\omega - \omega|$ can be large when $|\omega|$ is sufficiently large. For example, if $\beta = (1+\epsilon)^{-1}\,\mathbf{1}$ then $|\widehat{L_\beta f}(\omega)| = |1+\epsilon| \, |\hat{f}((1+\epsilon)\,\omega)|$. Suppose that $f(x) = \mathrm{e}^{i\xi x}\,\theta(x)$, where $\theta(x) = e^{-x^2/2}$. Since $\hat{f}(\omega) = \hat{\theta}(\omega - \xi)$, we verify that if $|\xi| > c > 0$ there exists $A > 0$ such that

$$\| |\widehat{L_\beta f}| - |\hat{f}| \| \geq A \, \|f\| \, \min(|\epsilon| \, |\xi|, 1) \, , \tag{6}$$

which contradicts the existence of a constant C satisfying the Lipschitz continuity (5). Small signal deformations may produce large modifications of the Fourier transform modulus at high frequencies.

2.2 Modulus Discriminability and Sparsity

Translation invariance requires to define an equivalence class including the d-dimensional manifold $\{f(x-\tau)\}_{\tau \in \mathbb{R}^d}$ for each $f \in \mathbf{L}^2(\mathbb{R}^d)$. Removing the Fourier phase eliminates one complex angular coordinate per frequency which is a much higher dimensional reduction. Signal discriminability may thus be considerably reduced. High frequency signals as different as a Dirac and a chirp e^{ix^2} have a Fourier modulus transform which is equal, and constant in this case. For periodic signals, Fourier coefficients are decomposition coefficients in an orthonormal basis. The loss of discriminability produced by a modulus operator on decomposition coefficients in an orthonormal basis $\mathcal{B} = \{g_m\}_{m \in \mathbb{Z}}$ is related to their sparsity. Let z^* be the complex conjugate of $z \in \mathbb{C}$. As in a Fourier basis of $\mathbf{L}^2[0,\Delta]^d$, we suppose that g_0 is a real function and that $g_{-m} = g_m^*$ for $m \neq 0$.

All along the paper, f is a real valued function. The modulus of its complex coefficients in $\mathcal{B}$ is written $\overline{U} f = \{\langle f, g_0 \rangle \, , |\langle f, g_m \rangle|\}_{m \neq 0}$. Since $g_{-m} = g_m^*$, $\overline{U} h = \overline{U} f$ if and only if there exists $\{\phi_m\}_{m>0}$ such that

$$h = \langle f, g_0 \rangle g_0 + 2 \sum_{m>0} |\langle f, g_m \rangle| \, \mathrm{Real}(e^{i\phi_m} g_m) \, . \tag{7}$$

Let $\{h : \overline{U} h = \overline{U} f\}$ be the set of functions which can not be discriminated from f after applying $\overline{U}$. It is thus a manifold whose dimension is equal to the number M of non-zero coefficients $\langle f, g_m \rangle$ for $m > 0$. In a Fourier basis, $M = 1$ corresponds to translated sinusoidal waves, but M is usually much larger. Indeed, signals that include high frequency transient structures such as singularities do not have a sparse representation in a Fourier basis. Removing their Fourier phase then produces a considerable loss of discriminability.

The loss of discriminability of a signal set $\Theta \subset \mathbf{L}^2(\mathbb{R}^d)$ can be measured by comparing the size of Θ and of

$$\overline{\Theta} = \{h \in \mathbf{L}^2(\mathbb{R}^d) \,:\, \exists f \in \Theta \,,\, \overline{U}h = \overline{U}f\} \,.$$

Minimax learning rates depend upon the geometry of signal classes and their complexity measured by the Kolmogorov ϵ-entropy [23]. The ϵ-entropy $H_\epsilon(\Theta)$ is the infimum of $\log N_\epsilon$ over all ϵ-approximation nets $\{a_n\}_{1 \leq n \leq N_\epsilon}$ of size N_ϵ satisfying

$$\sup_{f \in \Theta} \inf_{n \leq N_\epsilon} \|f - a_n\| \leq \epsilon \,.$$

Since $\Theta \subset \overline{\Theta}$ it results that $H_\epsilon(\Theta) \leq H_\epsilon(\overline{\Theta})$. The loss of discriminability is then measured by the increase of $H_\epsilon(\overline{\Theta})$ relatively to $H_\epsilon(\Theta)$.

A modulus operator removes the phase but preserves the support of coefficients in $\mathcal{B}$. For sparse signals whose support variability is the main source of complexity, the loss of phase has a marginal impact on discriminability. Let us consider an ϵ-approximation net $\{a_n\}_n$ of functions obtained by quantizing uniformly the basis coefficients of all $f \in \Theta$:

$$a_n = \sum_m Q(\langle f, g_m \rangle)\, g_m$$

with $Q(x) = n\,\Delta_\epsilon$ for $x \in [(n-1/2)\Delta_\epsilon, (n+1/2)\Delta_\epsilon)$. The quantization step Δ_ϵ is adjusted so that $\|a_n - f\| \leq \epsilon$. The support of a_n in $\mathcal{B}$ is $\{m \in \mathbb{Z} : |\langle f, g_m \rangle| > \Delta_\epsilon/2\}$. Let $\tilde{N}_\epsilon$ be the number distinct support sets for all $\{a_n\}_n$ and hence for all $f \in \Theta$. It is the same in Θ and in $\overline{\Theta}$ because the modulus does not change the support. When the Kolmogorov complexity is dominated by this support variability in $\mathcal{B}$, which means that $H_\epsilon(\overline{\Theta}) \sim H_\epsilon(\Theta) \sim \log \tilde{N}_\epsilon$ as ϵ goes to 0, then the loss of discriminability produced by $\overline{U}$ is negligible. This is for example the case if Θ is a ball of functions in a Besov space having sparse wavelet coefficients in l^p with $p < 2$ [7].

2.3 Wavelets Non-Invariance

As opposed to sinusoidal waves, wavelets are localized functions which are thus more stable to local deformations and they can provide sparse representations of signals including singularities. We consider Littlewood-Paley wavelet transforms, which satisfy frame type energy conservations but are not discretized and thus do not define bases of $\mathbf{L}^2(\mathbb{R}^d)$. Wavelet construction and deformation properties are briefly reviewed.

A Littlewood-Paley wavelet transform projects $f \in \mathbf{L}^2(\mathbb{R}^d)$ over a finite family of wavelets $\{\psi_\gamma\}_{\gamma\in\Gamma}$ which are scaled and translated by a continuous variable $x \in \mathbb{R}^d$ [15]. It can be written as convolutions

$$W_{j,\gamma} f(x) = f \star \psi_{j,\gamma}(x) = \int f(u)\; \psi_{j,\gamma}(x-u)\, du$$

with

$$\psi_{j,\gamma}(u) = 2^{-dj} \psi_\gamma(2^{-j}u)\ .$$

The cardinal of Γ is written $|\Gamma|$. Since $\hat{\psi}_{j,\gamma}(\omega) = \hat{\psi}_\gamma(2^j\omega)$ the Fourier transform of $W_{j,\gamma} f(x)$ along x is

$$\widehat{W_{j,\gamma} f}(\omega) = \hat{f}(\omega)\, \hat{\psi}_\gamma(2^j\omega)\ .$$

Each wavelet has a Fourier transform $\hat{\psi}_\gamma(\omega)$ which vanishes at $\omega = 0$ and can thus be interpreted as a "band-pass filter". Wavelet coefficients are computed at scales $2^j > 2^J$ and low frequencies are kept by a convolution with a scaling function, which averages f at a scale 2^J:

$$A_J f(x) = f \star \phi_J(x) \ \text{ with } \ \phi_J(u) = 2^{-dJ}\, \phi(2^{-J}u)\ .$$

The energy of $\hat{\phi}(\omega)$ is concentrated at low frequencies so ϕ_J acts as a low-pass filter. The wavelets ψ_γ may be complex functions but ϕ is real.

The norm of $\overline{W}_J f = \{A_J f, W_{j,\gamma} f\}_{j<J, 1\leq\gamma\leq|\Gamma|}$ is

$$\|\overline{W}_J f\|^2 = \|A_J f\|^2 + \sum_{j=-\infty}^{J-1} \sum_{\gamma=1}^{\Gamma} \|W_{j,\gamma} f\|^2 \tag{8}$$

with

$$\|A_J f\|^2 = \int |A_J f(x)|^2\, dx \ \text{ and } \ \|W_{j,\gamma} f\|^2 = \int |W_{j,\gamma} f(x)|^2\, dx\ .$$

The following proposition gives a Littlewood-Paley condition so that the wavelet transform is contracting and potentially unitary over real valued functions.

Proposition 1 *For any $f \in \mathbf{L}^2(\mathbb{R}^d)$ with $f(x) \in \mathbb{R}$ and any $J \in \mathbb{Z}$ the wavelet transform satisfies*

$$\|f\|^2(1-\delta) \leq \|\overline{W}_J f\|^2 \leq \|f\|^2 \quad \text{with} \;\; \delta < 1 \tag{9}$$

if and only if for almost all $\omega \in \mathbb{R}^d$

$$(1-\delta) \leq |\hat{\phi}(\omega)|^2 + \frac{1}{2} \sum_{j=-\infty}^{-1} \sum_{\gamma \in \Gamma} \Big(|\hat{\psi}_\gamma(2^j \omega)|^2 + |\hat{\psi}_\gamma(-2^j \omega)|^2 \Big) \leq 1 \; . \tag{10}$$

Proof: The Fourier transform of $W_{j,\gamma} f = f \star \psi_{j,\gamma}$ is $\widehat{W_{j,\gamma} f}(\omega) = \hat{f}(\omega)\, \hat{\psi}_\gamma(2^j \omega)$. Since f is real, $\hat{f}(-\omega) = \hat{f}^*(\omega)$. If (10) is satisfied, multiplying (10) by $|\hat{f}(\omega)|^2$, integrating in ω and applying the Plancherel formula proves (9).

Conversely, if (10) is satisfied then (9) is satisfied for almost all ω. Otherwise, one can construct a real function $f \neq 0$ where $\hat{f}$ has a support in the domain of ω where (9) is not valid. With the Plancherel formula we verify that f does not satisfy (10). $\square$

If $\delta = 0$ then the wavelet transform is a unitary operator. In the following, we shall always suppose that the wavelet transform satisfies condition (10) and hence defines a contractive operator. It results that $\hat{\psi}_\gamma(0) = \int \psi_\gamma(x)\, dx = 0$. We also suppose that ϕ and all ψ_γ are twice differentiable and that their decay as well as the decay of their partial derivatives of order 1 and 2 is $O((1+|x|)^{-d-2})$.

With multiple wavelets, $\hat{\psi}_\gamma$ may be chosen to have an arbitrarily small support while satisfying the frequency covering condition (10). In $d = 1$ dimension, wavelets having a small bandwidth are needed to precisely discriminate close frequencies in sounds. In $d = 2$ dimensions, wavelets having a narrow angular frequency support provide directional information on image structures such as edges or textures. The parameter γ then specifies an angle [19]. In general, γ brings out explicit dependence of the signal relatively to groups of deformations, which is used by Section 6 to builds invariant representations relatively to these groups.

Wavelets may be real in which case $\hat{\psi}_\gamma(-\omega) = \hat{\psi}^*_\gamma(\omega)$. To build invariant representations, we shall consider complex wavelets whose Fourier transform have a support mostly concentrated on one half of $\mathbb{R}^d$. Such wavelets can be written $\psi_\gamma(x) = \theta_\gamma(x)\, e^{i\xi_\gamma x}$ where $\hat{\theta}_\gamma(\omega)$ is concentrated over low frequencies.

In $d = 1$ dimension, the Littlewood-Paley condition may be satisfied with a single complex wavelet ψ_γ which is an analytic function. If ψ is a real wavelet which generates an orthonormal basis of $\mathbf{L}^2(\mathbb{R}^d)$ [15] then it satisfies (10) for $\delta = 0$. The analytic part ψ_γ of ψ is defined by $\hat{\psi}_\gamma(\omega) = 2\hat{\psi}(\omega)$ for $\omega \geq 0$ and $\hat{\psi}_\gamma(\omega) = 0$ for $\omega < 0$. It then also satisfies the frequency covering (10) for $\delta = 0$ and thus defines a unitary wavelet transform. Numerical examples in the paper are computed with the analytic part of a cubic-spline Battle-Lemarié wavelet [13].

Wavelet transforms in several dimensions may be defined with separable products of a one-dimensional wavelet ψ and its scaling function ϕ. The separable scaling function is $\phi(x_1) \dots \phi(x_d)$. Writing $h_0 = \phi$ and $h_1 = \psi$, we define $|\Gamma| = 2^d - 1$ separable wavelets for binary $\gamma = (\epsilon_1, ..., \epsilon_d) \neq 0$:

$$\psi_\gamma(x_1, ..., x_d) = h_{\epsilon_1}(x_1) \dots h_{\epsilon_d}(x_d) \,.$$

If $\hat{\psi}$ and $\hat{\phi}$ satisfy (10) for $\delta = 0$ then the $2^d - 1$ separable ψ_γ also satisfy (10) for $\delta = 0$.

A wavelet transform is computed with convolutions and thus commutes with translation operators $T_\tau f(x) = f(x - \tau)$:

$$W_{j,\gamma} T_\tau f(x) = W_{j,\gamma} f(x - \tau) = T_\tau W_{j,\gamma} f(x) \ .$$

It is therefore not translation invariant but if $2^{-j}\,|\tau| \ll 1$ then $W_{j,\gamma} T_\tau f$ and $W_{j,\gamma} f$ are nearly equal because both functions have derivatives of the order of 2^{-j}.

Let us now consider the case where f is deformed into $D_\tau f(x) = f(x - \tau(x))$. Let $|\nabla\tau(x)|$ be the sup matrix norm of the deformation gradient and $|\nabla\tau|_\infty = \sup_x |\nabla\tau(x)|$. We write $|\tau|_\infty = \sup_x |\tau(x)|$ the maximum displacement amplitude. The following theorem computes an upper bound of $\|W_{j,\gamma} D_\tau f - W_{j,\gamma} f\|$. We write $\|A\|$ the sup norm of an operator A in $\mathbf{L}^2(\mathbb{R}^d)$. The commutator of two operators is written $[A, B] = AB - BA$. The $\mathbf{L^1}(\mathbb{R}^d)$ norm is $\|h\|_1 = \int |h(x)|\, dx$.

Proposition 2 *If $|\nabla\tau|_\infty \leq 1 - \epsilon$ with $\epsilon > 0$ then there exists C such that*

$$\|W_{j,\gamma} D_\tau f - W_{j,\gamma} f\| \leq C\, \|f\| \Big(2^{-j} |\tau|_\infty + |\nabla\tau|_\infty \Big) \ . \tag{11}$$

Proof: Observe that

$$W_{j,\gamma} D_\tau - W_{j,\gamma} = D_\tau W_{j,\gamma} - W_{j,\gamma} + [W_{j,\gamma}, D_\tau]$$

and hence

$$\|W_{j,\gamma} D_\tau f - W_{j,\gamma} f\| \leq \|D_\tau W_{j,\gamma} f - W_{j,\gamma} f\| + \| \, [W_{j,\gamma}, D_\tau] f \, \| \ . \tag{12}$$

Appendix C provides an upper bound of the commutator term, as a particular case of the more general Lemma 5. It shows in (132) that there exists $C > 0$ such that

$$\| \, [W_{j,\gamma}, D_\tau] \, f \| \leq C \, \|f\| \, |\nabla\tau|_\infty \ . \tag{13}$$

The following lemma applied to $h = \psi_\gamma$ shows that

$$\|D_\tau W_{j,\gamma} f - W_{j,\gamma} f\| \leq C \, \|f\| \, 2^{-j} \ |\tau|_\infty,$$

which proves (11). The lemma proof is in Appendix A.

Lemma 1 *Let $h \in \mathbf{L}^2(\mathbb{R}^d)$ with $\|h\|_1 < +\infty$, $\|\nabla h\|_1 < +\infty$ and $\||x| \, \nabla h(x)\|_1 < +\infty$. If $Z_j f = f \star h_j$ with $h_j(x) = 2^{-dj} h(2^{-j}x)$ and $|\nabla\tau|_\infty < 1 - \epsilon$ with $\epsilon > 0$ then there exists C such that*

$$\|D_\tau Z_j f - Z_j f\| \leq C \, \|f\| \, 2^{-j} \ |\tau|_\infty \ . \tag{14}$$

□

The condition $|\nabla\tau|_\infty \leq 1 - \epsilon$ guaranties that $x - \tau(x)$ and D_τ are invertible. Proposition 2 proves in (11) that small elastic deformations produce small modifications of the wavelet transform, bounded by $|\nabla\tau|_\infty$. As expected, the translation component of D_τ yields an error term proportional to $2^{-j} \, |\tau|_\infty$. It is large at higher frequencies corresponding to small scales $2^j \leq |\tau|_\infty$.

3 Scattering Representation and Metric

Similarly to the Fourier case, a translation invariant representation is constructed by removing the complex phase of wavelet coefficients with a modulus operator. Section 3.1 shows that this wavelet modulus operator represents the signal through interference within each frequency octave. Iterating on this operator defines a scattering operator which becomes translation invariant. It scatters signals across multiple wavelet paths along a convolution network. Section 3.3 proves that this scattering transform is contractive and preserves the $\mathbf{L}^2(\mathbb{R}^d)$ norm for appropriate complex wavelets. Translation invariance and Lipschitz continuity are proved in Section 3.4.

3.1 Wavelet Modulus and Interferences

The translation non-invariance of wavelet coefficients is due to their high frequency variations at fine scales. By removing the complex wavelet phase, we show that a modulus operator maps wavelet high frequency coefficients to lower frequencies which carry local interference information [14].

Suppose that

$$\psi_\gamma(x) = \theta_\gamma(x)\,\mathrm{e}^{i\xi_\gamma x} \;\;\text{with}\;\; \theta_\gamma(x) \in \mathbb{R}\ ,$$

where $|\hat\theta_\gamma(\omega)|$ has an energy concentrated over a low frequency domain. To simplify notations, we write $\lambda = (j,\gamma)$, $\Lambda = \{\lambda = (j,\gamma)\ :\ j \in \mathbb{Z}\,,\ \gamma \in \Gamma\}$, $W_\lambda f = f \star \psi_\lambda(x)$ and $\psi_\lambda(x) = \mathrm{e}^{i\xi_\lambda x}\,\theta_\lambda(x)$ with $\xi_\lambda = 2^{-j}\xi_\gamma$ and $\theta_\lambda = 2^{-j}\theta(2^{-j}x)$. The wavelet transform can be written:

$$W_\lambda f(x) = \mathrm{e}^{i\xi_\lambda x}\left(f_\lambda \star \theta_\lambda(x)\right) \;\;\text{with}\;\; f_\lambda(x) = \mathrm{e}^{-i\xi_\lambda x}\,f(x)\ .$$

The high frequency modulation $\mathrm{e}^{i\xi_\lambda x}$ is removed by a modulus operator:

$$U_\lambda f = |W_\lambda f| = |f_\lambda \star \theta_\lambda|\ .$$

The Fourier transform of $|W_\lambda f|$ has therefore an energy mostly concentrated over the low frequency domain covered by $\hat\theta_\lambda(\omega)$. However, $|f_\lambda \star \theta_\lambda(x)|$ is singular at vanishing locations, because $|z|$ is singular at $z = 0$, which also brings energy at higher frequencies.

The modulus operator is not applied to $A_J f = f \star \phi_J$ which is already a low frequency signal. The resulting wavelet modulus operator $\overline{U}_J$ is

$$\overline{U}_J f = \Big\{A_J f = f \star \phi_J\,,\ U_\lambda f = |f \star \psi_\lambda|\Big\}_{\lambda\in\Lambda}\ . \tag{15}$$

As explained in Section 2.2, the loss of discriminability produced by such a modulus operator depends on the representation sparsity. Defining sparse signal representations is thus an important criteria for choosing the wavelet family $\{\psi_\gamma\}_{\gamma\in\Gamma}$.

Let us show that the modulus non-linearity can be interpreted as an interference calculation in each frequency octave. Suppose that $f(x) = \sum_n c_n\,\cos(\omega_n x)$. If the sign of ω_n is adjusted so that $\hat\psi_\lambda(-\omega_n) = 0$ then

$$W_\lambda f(x) = \sum_n a_n\,\mathrm{e}^{i\omega_n x} \;\;\text{with}\;\; a_n = c_n\,\hat\psi_\lambda(\omega_n)\,/2\,.$$

It results that

$$U_\lambda f(x) = (e_\lambda^2 + \epsilon_\lambda(x))^{1/2} = e_\lambda + \frac{\epsilon_\lambda(x)}{2e_\lambda} + O\Big(\frac{\epsilon_\lambda^2(x)}{e_\lambda^3}\Big)~, \tag{16}$$

where $e_\lambda^2 = \sum_n |a_n|^2$ and $\epsilon_\lambda(x)$ carries the interferences of the filtered sinusoidal components in each octave:

$$\epsilon_\lambda(x) = \sum_{n \neq n'} a_n\, a_{n'} \cos[(\omega_n - \omega_{n'})x + \alpha_{\lambda,n,n'}]~.$$

Sinusoids of frequency ω_n and $\omega_{n'}$ are transformed into sinusoidal interferences of smaller frequencies $\omega_n - \omega_{n'}$, with a phase shift. The square root in (16) also creates singularities when $e_\lambda^2 + \epsilon_\lambda(x)$ vanishes, which produces higher frequency harmonics of $\omega_n - \omega_{n'}$. The operator U_λ computes interferences which lose information on frequency locations within an octave, but it keeps frequency interval values. This interference calculation is local since wavelets have a local support. A similar phenomena appears in audio perception. We easily recognize frequency intervals between musical notes (minor third, major third intervals) but it is much more difficult to identify the frequency of an isolated musical note.

3.2 Scattering Operator

The wavelet modulus operator is iteratively applied to progressively map high signal frequencies to lower interference signals. The resulting scattering operator is defined from $A_J f = f \star \phi_J$ and $U_\lambda f = |f \star \psi_\lambda|$.

Definition 1 *A wavelet path is an index sequence $p = \{\lambda_n\}_{1 \leq n \leq |p|}$. A scattering operator at the scale 2^J is defined over a set P_J of paths $p = \{\lambda_n = (j_n, \gamma_n)\}_{1 \leq n \leq |p|}$ for which $\max_n j_n < J$:*

$$S_J(p) f = A_J\, S(p) f \quad \textit{with} \quad S(p) f = \prod_{n=1}^{|p|} U_{\lambda_n} f~. \tag{17}$$

For $p = 0$, $S(0)f = f$ and $S_J(0) f = A_J f$.

The scattering operator S_J implements a sequence of wavelet convolutions and modulus, followed by a convolution with ϕ_J:

$$S_J(p) f = |\cdots|f \star \psi_{\lambda_1}| \star \psi_{\lambda_2}| \cdots| \star \psi_{\lambda_{|p|}}| \star \phi_J.$$

It is a non-linear operator but preserves amplitude factors:

$$\forall \mu \in \mathbb{R} \ , \ S_J(0)(\mu f) = \mu \, S_J(0) f \ \text{ and } \ S_J(p)(\mu f) = |\mu| \, S_J(p) f \text{ if } p \neq 0.$$

It also results that

$$S_J(p) f \leq |f| \star \theta_p \ \text{ with } \ \theta_p = |\psi_{\lambda_1}| \star \cdots \star |\psi_{\lambda_{|p|-1}}| \star |\phi_J| \ .$$

If the mother wavelets ψ_γ and ϕ have a compact support of radius bounded by Δ then the dilated wavelets ψ_λ for $\lambda = (j, \gamma)$ and ϕ_J have a support of radius bounded respectively by $2^j \Delta$ and $2^J \Delta$. The envelop θ_p then has a compact support of radius bounded by $l(p)\,\Delta$ with $l(p) = \sum_{n=0}^{|p|} 2^{j_n} + 2^J$.

An analogy can be made with path integral formulations of quantum field theory [9], if one considers f as the wave function of a particle. This wavefunction evolves in time through the propagators U_λ, which model interactions with other particles having a state parametrized by $\lambda = (j, \gamma)$. The scale 2^j plays the role of a propagation time interval between two interactions. Different paths correspond to different interaction sequences. The path length is the number of interactions along a path and $l(p)$ can be interpreted as the total duration of a path. Particles are collected on a detector after a propagation in free space, resulting in a final diffusion given by the convolution with ϕ_J. The operator S_J computes the wavefunction modulus of particles reaching this detector through the path p. If $\|f\|^2 = 1$ then $\|S_J(p) f\|^2$ can be interpreted as the probability of reaching the detector through path p. This analogy gives an intuition of the underlying mathematics but is not a formal physical model.

The scattering $S_J(p) f$ averages positive coefficients $S(p) f(x)$ and when J goes to ∞, the following proposition proves that $2^{-Jd}\, S_J f(x)$ converges to a constant which does not depend upon x.

Proposition 3 *If $f \in \mathbf{L}^1(\mathbb{R}^d)$ then for all $p \in P_J$*

$$\forall x \in \mathbb{R}^d \quad , \quad \lim_{J \to \infty} 2^{Jd}\, S_J(p) f(x) = \phi(0) \int S(p) f(u)\, du \ . \tag{18}$$

Proof: Since $f \in \mathbf{L}^1(\mathbb{R}^d)$ and $S(p) f$ is obtained by convolutions with wavelets $\psi_\lambda \in \mathbf{L}^1(\mathbb{R}^d)$, it results that $S(p) f \in \mathbf{L}^1(\mathbb{R}^d)$. Moreover, ϕ is continuous at 0 and $\phi_J(x) = 2^{-dJ} \phi(2^{-J} x)$ so we verify that

$$\lim_{J \to \infty} 2^{Jd}\, S(p) f \star \phi_J(x) = \phi(0) \int S(p) f(u)\, du$$

which proves (18). $\square$

This proposition proves that scattering coefficients converge to $\mathbf{L}^1(\mathbb{R}^d)$ norms computed on iterated wavelet coefficients. If $p = \{\lambda_1\}$ then $S(p)f = |f \star \psi_{\lambda_1}|$ and $2^{Jd}\, S_J(p)f(x)$ converges to $\phi(0)\, \|f \star \psi_{\lambda_1}\|_1$. If $p = \{\lambda_1, \lambda_2\}$ then $2^{Jd}\, S_J(p)f(x)$ converges to $\phi(0)\, \||f \star \psi_{\lambda_1}| \star \psi_{\lambda_2}\|_1$.

The following proposition shows that for J fixed, the maximum amplitude of $S_J(p)f(x)$ decreases exponentially as the path length $|p|$ increases.

Proposition 4 *Let*

$$\eta = \max_{\gamma \in \Gamma} \|\psi_\gamma\|_1 / 2 \ . \tag{19}$$

For any scattering path $p = \{\lambda_n\}_{1 \leq n \leq |p|}$

$$\|S(p)f\|_\infty \leq \|W_{\lambda_1} f\|_\infty\, \eta^{|p|-1} \tag{20}$$

and

$$\|S_J(p)f\|_\infty \leq 2\, \|f\|_\infty\, \|\phi\|_1\, \eta^{|p|}. \tag{21}$$

Proof: Let us first prove (20) by induction on $m = |p|$. If $m = 1$ then $S(p)f = |W_{\lambda_1} f|$ so the property is evident. Suppose that the induction hypothesis is valid for m. For any path p of length $|p| = m+1$, the truncated path $p' = \{\lambda_n\}_{n < |p|}$ is of length m, and

$$S(p)f = |S(p')f \star \psi_{\lambda_{|p|}}| \ .$$

Since $\int \psi_\lambda(u)\, du = 0$, if $f(u) = c$ is constant then $f \star \psi_\lambda = 0$. So we can subtract the constant $\|S(p')f\|_\infty/2$ to $S(p')f$:

$$S(p)f = \left| \Big(S(p')f - \|S_J(p')f\|_\infty/2 \Big) \star \psi_{\lambda_{|p|}} \right| \ .$$

Since $S(p')f \geq 0$, $\|S(p')f - \|S(p')f|_\infty/2\|_\infty = \|S(p')f|_\infty/2$ so

$$\|S(p)f\|_\infty \leq \|S(p')f\|_\infty/2\, \|\psi_{\lambda_{|p|}}\|_1 = \|S_J(p')f\|_\infty\, \|\psi_{\lambda_{|p|}}\|_1/2 \ .$$

The induction hypothesis (19) applied on p' proves that

$$\|S(p)f\|_\infty \leq \|W_{\lambda_1} f\|_\infty\, \eta^{|p|-1}$$

This finishes the induction proof of (20).

Since $S_J(p)f = S(p)f \star \phi_J$,

$$\|S_J(p)f\|_\infty \leq \|S(p)f\|_\infty \, \|\phi_J\|_1 = \|S(p)f\|_\infty \, \|\phi\|_1 \ .$$

Applying (20) proves (21) by observing that

$$\|W_{\lambda_1} f\|_1 = \|f \star \psi_{\lambda_1}\|_1 \leq \|f\|_\infty \, \|\psi_{\lambda_1}\| \leq 2 \, \|f\|_\infty \, \eta \ .$$

$\square$

If $\eta < 1$ then (20) proves that the maximum scattering amplitude $\|S_J(p)f\|_\infty$ has an exponential decay as a function of the path length $|p|$. For example, in $d = 1$ dimension, if ψ_γ is a complex Battle-Lemarié cubic spline wavelet then $\eta = 0.8872$. Next section gives a finer result which proves that the total scattering energy over all paths of length larger then m decays to zero as m increases.

3.3 Scattering Propagation and Metric Contraction

The key properties of scattering operators result from their factorization as a product of wavelet modulus propagators U_J, which implement a multilayer convolution network. The main theorem proves that the scattering energy is concentrated on a small subset of paths, and that it decreases to zero when the path length increases. It results that scattering operators can preserve $\mathbf{L}^2(\mathbb{R}^d)$ norms.

The following proposition proves that the wavelet transform modulus operator $\overline{U}_J f = \Big\{ A_J f \, , \, U_\lambda f \Big\}_{\lambda \in \Lambda}$ acts as a propagator.

Proposition 5 *For any* $m \geq 0$

$$\overline{U}_J \{S(p)f\}_{p \in P_J, |p| = m} = \{S_J(p)f\}_{p \in P_J, |p| = m} \cup \{S(p)f\}_{p \in P_J, |p| = m+1} \ . \qquad (22)$$

Proof: The continuation of a path $p = \{\lambda_n\}_{n \leq |p|}$ with λ is a path written $p + \lambda$, whose last element is λ. Since $\overline{U}_J f = \Big\{ A_J f \, , \, U_\lambda f \Big\}_{\lambda \in \Lambda}$

$$\overline{U}_J S(p) f = \{ A_J S(p) f = S_J(p) f \, , \, U_\lambda S(p) f = S(p + \lambda) f \}_{\lambda \in \Lambda} \ . \qquad (23)$$

Since all paths of length $m + 1$ are obtained as a continuation of a path of length m, we derive (22) from (23). $\square$

This proposition shows that all $S_J(p)f$ can be computed by iterating on the propagator $\overline{U}_J$, layer by layer. It begins with $S(0)f = f$ at the layer 0.

Applying $\overline{U}_J$ computes the layer 1 of all $S(p)f$ with $|p| = 1$ and it updates $S(0)f$ into $S_J(0)f$. Applying again $\overline{U}_J$ on all $S(p)f$ of layer 1 computes the layer 2 of all $S(p)$ with $|p| = 2$, and it transforms all $S(p)f$ in layer 1 into $S_J(p)f$.

Let us verify by induction that applying m times $\overline{U}_J$ computes the first $m-1$ layers of $S_J(p)f$ for all $1 \leq |p| < m$ plus the layer m of $S(p)f$ for $|p| = m$. This is written:

$$\overline{U}_J^m f = \{S_J(p)f\}_{p\in P_J,|p|<m} \cup \{S(p)f\}_{p\in P_J,|p|=m} \ . \tag{24}$$

Indeed, applying (22) to the induction hypothesis (24) gives

$$\begin{aligned} \overline{U}_J^{m+1} f &= \{S_J(p)f\}_{p\in P_J,|p|<m} \cup \{\overline{U}_J S(p)f\}_{p\in P_J,|p|=m} \\ &= \{S_J(p)f\}_{p\in P_J,|p|<m+1} \cup \{S(p)f\}_{p\in P_J,|p|=m+1} \ , \end{aligned} \tag{25}$$

which verifies the induction hypothesis for $m+1$. Section 3.1 explains that a wavelet transform modulus $\overline{U}_J$ computes interferences between frequency components within different octaves. The layer m of $\{S_J(p)f\}_{|p|=m}$ thus carries m^{th} order interferences.

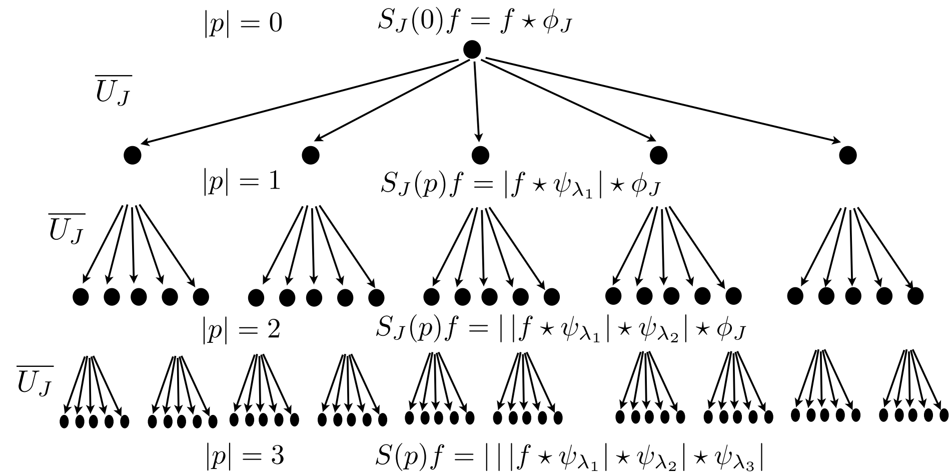

Figure 1: A scattering network iterates m times on the propagator $\overline{U}_J$ to compute $S_J(p)f$ at all layers $|p| < m$ and $S(p)f$ at the last layer $|p| = m$.

Figure 1 illustrates this scattering decomposition implemented along a convolution network. Convolution networks are general computational architectures introduced by LeCun [11], that involve convolutions and non-linear operators. They have been successfully applied to number of recognition tasks [11] with back-propagation learning algorithms. Convolution networks have also been studied as models for visual perception [4, 17], maxima detections. Scattering operators are implemented with modulus non-linearities and with filters which are dilated wavelets.

Since S_J iterates on $\overline{U}_J$, it inherits many of its properties. If wavelets satisfy the Littlewood-Paley condition (10) then Proposition 1 proves that $\overline{W}_J f = \{A_J f\,,\, W_\lambda f\}_{\lambda\in\Lambda}$ is a linear contrative operator $\|\overline{W}_J f\| \leq \|f\|$. Since $||a| - |b|| \leq |a-b|$ for any $(a,b) \in \mathbb{C}^2$, it results that $\overline{U}_J$ is a non-linear contracting operator:

$$\|\overline{U}_J f - \overline{U}_J g\|^2 = \sum_{\lambda\in\Lambda)} \|U_\lambda f - U_\lambda g\|^2 + \|A_J f - A_J g\|^2 \leq \|f-g\|^2 .$$

The scattering metric for $(f,g) \in \mathbf{L}^2(\mathbb{R}^d)$ is defined by

$$\|S_J f - S_J g\|^2 = \sum_{p\in P_J} \|S_J(p) f - S_J(p) g\|^2 \;.$$

The following proposition derives from (25) that S_J it also contractive [12].

Proposition 6 *The operator S_J is contracting*

$$\forall (f,g) \in \mathbf{L}^2(\mathbb{R}^d)^2 \quad,\quad \|S_J f - S_J g\| \leq \|f-g\| \;. \tag{26}$$

.

Proof: Since $\overline{U}_J$ is contractive, iterations on $\overline{U}_J$ are also contracting. It results from (25) that

$$\begin{aligned} \|\overline{U}_J^m f - \overline{U}_J^m g\|^2 &= \sum_{p\in P_J, |p|<m} \|S_J(p) f - S_J(p) g\|^2 + \sum_{p\in P_J, |p|=m} \|S(p) f - S(p) g\|^2 \\ &\leq \|f-g\|^2 \;. \end{aligned} \tag{27}$$

Taking the limit over m proves that

$$\forall (f,g) \in \mathbf{L}^2(\mathbb{R}^d)^2 \quad,\quad \|S_J f - S_J g\|^2 = \sum_{p\in P_J} \|S_J(p) f - S_J(p) g\|^2 \leq \|f-g\|^2 \;.$$

□

Proposition 1 proves that a wavelet transform is unitary if and only if

$$|\hat\phi(\omega)|^2 + \frac{1}{2} \sum_{j=-\infty}^{-1} \sum_{\gamma\in\Gamma} \Big(|\hat\psi_\gamma(2^j\omega)|^2 + |\hat\psi_\gamma(-2^j\omega)|^2 \Big) = 1 \;. \tag{28}$$

If $\|\overline{W}_J f\| = \|f\|$ then the propagator $\overline{U}_J$ also preserves norms

$$\forall f \in \mathbf{L}^2(\mathbb{R}^d) \quad , \quad \|\overline{U}_J f\|^2 = \|f\|^2 \; .$$

Iterating on $\overline{U}_J$ thus also preserves norms

$$\|\overline{U}_J^m f\|^2 = \sum_{p \in P_J, |p| < m} \|S_J(p) f\|^2 + \sum_{p \in P_J, |p| = m} \|S(p) f\|^2 = \|f\|^2 \; . \tag{29}$$

It results that $\|S_J f\| = \|f\|$ if and only if

$$\lim_{m \to \infty} \sum_{p \in P_J, |p| = m} \|S(p) f\|^2 = 0 \; .$$

This decay condition means that the scattering energy is attenuated as it propagates along the convolution network, and that it tends to zero as the path length increases. Theorem 1 gives a sufficient condition on wavelets to obtain this decay.

Theorem 1 *Suppose that the wavelet transform is unitary. If for each $\gamma \in \Gamma$ there exist $\xi_\gamma \in \mathbb{R}^d$ and $\rho_\gamma \geq 0$ with $|\hat\rho_\gamma(\omega)| \leq |\hat\phi(2\omega)|$, $\hat\rho_\gamma(0) = 1$ such that the functions*

$$\hat\Psi_\gamma(\omega) = |\hat\rho_\gamma(\omega - \xi_\gamma)|^2 - \sum_{k=1}^{+\infty} k \, (1 - |\hat\rho_\gamma(2^{-k}(\omega - \xi_\gamma))|^2) \tag{30}$$

satisfy

$$\alpha = \inf_{\pi \leq |\omega| \leq 2\pi} \sum_{j=-\infty}^{+\infty} \sum_{\gamma=1}^{|\Gamma|} \Big(\hat\Psi_\gamma(2^j \omega) \, |\hat\psi_\gamma(2^j \omega)|^2 + \hat\Psi_\gamma(-2^j \omega) \, |\hat\psi_\gamma(-2^j \omega)|^2 \Big) > 0 \tag{31}$$

then for all real functions $f \in \mathbf{L}^2(\mathbb{R}^d)$

$$\lim_{m \to \infty} \sum_{p \in P_J, |p| = m} \|S(p) f\|^2 = 0 \quad , \quad \lim_{m \to \infty} \sum_{p \in P_J, |p| > m} \|S_J(p) f\|^2 = 0 \tag{32}$$

and

$$\|S_J f\| = \|f\| \; . \tag{33}$$

Proof: If $\lim_{m\to\infty} \sum_{p\in P_J,|p|=m} \|S(p)f\|^2 = 0$ then we saw that $\|S_J f\| = \|f\|$. It also implies the asymptotic decay of S_J in (32) because

$$\sum_{p\in P_J,|p|>m} \|S_J(p)f\|^2 \leq \sum_{p\in P_J,|p|=m} \|S(p)f\|^2 . \tag{34}$$

Indeed, Proposition 5 proves that $\{S_J(p)f\}_{p\in P_J,|p|>m}$ is computed from $\{S(p)f\}_{p\in P_J,|p|=m}$ by iteratively applying $\overline{U}_J$. Since $\overline{U}_J$ is contracting, it implies (34).

To prove that $\lim_{m\to\infty} \sum_{p\in P_J,|p|=m} \|S(p)f\|^2 = 0$ we show that the signal energy mostly propagates along paths along which the scale increases, and hence towards lower frequencies. The evaluation of this property is based on the following lemma, which gives a lower bound of the energy of $|f\star\psi_\lambda|$ filtered by a positive low frequency function. This lemma is a key technical element of the proof, which both takes into account and suppresses the non-linearity of the modulus operator.

Lemma 2 *If $h(x) \geq 0$ then for any $f \in \mathbf{L}^2(\mathbb{R}^d)$*

$$\||f\star\psi_\lambda|\star h\| \geq \sup_{\eta\in\mathbb{R}^d} \|f\star\psi_\lambda\star h_\eta\| \quad \textit{with} \quad h_\eta(x) = h(x)\, e^{i\eta x} . \tag{35}$$

The lemma is proved by computing

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

which finishes the lemma proof.

Appendix D uses this lemma to show that the scattering energy propagates progressively towards larger scales. It proves the following lemma which gives an upper bound on the scattering energy along all paths $p \in P_J$.

Lemma 3 *If (31) is satisfied then*

$$\frac{\alpha}{2} \sum_{p\in P_J} \|S(p)f\|^2 \leq \|f\|^2 + \sum_{j<J} \sum_{\gamma=1}^{|\Gamma|} (J-j) \|W_{j,\gamma} f\|^2 . \tag{36}$$

Since $\|f\|^2 = \|f \star \phi_J\|^2 + \sum_{j<J} \sum_{\gamma=1}^{|\Gamma|} \|W_{j,\gamma} f\|^2$ the left hand side of (36) is finite if and only if

$$\|f\|_w^2 = \sum_{j=-\infty}^{0} \sum_{\gamma \in \Gamma} |j| \, \|W_{j,\gamma} f\|^2 < \infty \ . \tag{37}$$

Since

$$\sum_{p \in P_J} \|S(p) f\|^2 = \sum_{m=0}^{+\infty} \sum_{p \in P_J, |p|=m} \|S(p) f\|^2,$$

if $\|f\|_w < \infty$ then (36) implies that $\lim_{m \to \infty} \sum_{p \in P_J, |p|=m} \|S(p) f\|^2 = 0$.

This result is extended in $\mathbf{L}^2(\mathbb{R}^d)$ with a limit argument. If $f \in \mathbf{L}^2(\mathbb{R}^d)$ and $f_n = f \star \phi_n$, since $\phi \in \mathbf{L}^1(\mathbb{R}^d)$, $\hat{\phi}(0) = 1$ and $\hat{\phi}_n(\omega) = \hat{\phi}(2^n \omega)$, one can verify that $\lim_{n \to -\infty} \|f - f_n\| = 0$. We prove that $\lim_{m \to \infty} \sum_{p \in P_J, |p|=m} \|S(p) f_n\|^2 = 0$ by showing that $\|f_n\|_w < \infty$. Indeed

$$\|W_{j,\gamma} f_n\|^2 = \int |\hat{f}(\omega)|^2 \, |\hat{\phi}(2^n \omega)| \, |\hat{\psi}_\gamma(2^j \omega)|^2 \, d\omega \leq C \, 2^{j-n} \int |\hat{f}(\omega)|^2 \, d\omega,$$

because ψ_γ has a vanishing moment so $|\hat{\psi}_\gamma(\omega)| = O(|\omega|)$, and the derivatives of ϕ are in $\mathbf{L}^1(\mathbb{R}^d)$ so $|\omega| \, |\hat{\phi}(\omega)|$ is bounded. It results from (37) that $\|f_n\|_w < \infty$.

Since $\overline{U}_J$ is contracting, (27) implies that

$$\sum_{p \in P_J, |p|=m} \|S(p) f - S(p) f_n\|^2 \leq \|f - f_n\|^2$$

so

$$\Big(\sum_{p \in P_J, |p|=m} \|S(p) f\|^2 \Big)^{1/2} \leq \|f - f_n\| + \Big(\sum_{p \in P_J, |p|=m} \|S(p) f_n\|^2 \Big)^{1/2} .$$

Since $\lim_{n \to -\infty} \|f - f_n\| = 0$ and $\lim_{m \to \infty} \sum_{p \in P_J, |p|=m} \|S(p) f_n\|^2 = 0$ it results that $\lim_{m \to \infty} \sum_{p \in P_J, |p|=m} \|S(p) f\|^2 = 0$ for any $f \in \mathbf{L}^2(\mathbb{R}^d)$. $\square$

Since $\lim_{m \to \infty} \sum_{p \in P_J, |p|>m} \|S_J(p) f\|^2 = 0$, for m sufficiently large we can neglect and therefore not compute $S_J(p) f$ for $|p| \leq m$. This maximum path length is the depth of the convolution network illustrated in Figure 1. The proof also shows that the energy propagates towards larger scales. The energy of $S_J(p) f$ is thus mostly concentrated along scale increasing paths $p = \{\lambda_n = (j_n, \gamma_n)\}_{n \leq |p|}$ with $j_{n+1} > j_n$. In numerical applications, computations are restricted to such scale increasing paths.

To satisfy condition (31), for each $\gamma \in \Gamma$ we look for $\rho_\gamma \geq 0$ with $|\hat{\rho}_\gamma(\omega)| \leq |\hat{\phi}(2\omega)|$ and a frequency ξ_γ such that $|\hat{\rho}_\gamma(\omega - \xi_\gamma)|$ is as large as possible over frequencies ω where the energy of $\hat{\psi}_\gamma$ is concentrated. To obtain $\alpha > 0$ we choose complex wavelets so that $\hat{\psi}_\gamma(\omega)$ is concentrated over frequencies close to some ξ_γ. For an analytic cubic spline Battle-Lemarié wavelet in $d = 1$ dimension, $\rho_\gamma(x)$ is chosen to be a positive cubic box spline with $\xi_\gamma = 3\pi/2$. The numerical evaluation of (31) gives $\alpha = 0.2766 > 0$. A scattering operator computed with analytic cubic spline wavelets thus preserves the $\mathbf{L}^2(\mathbb{R}^d)$ norm.

To build translation invariant operators, the scattering scale 2^J must be increased up to size of the function support. The following proposition proves that the scattering metric is progressively more contracting as J increases.

Proposition 7 *If the wavelet transform is unitary then*

$$\forall (f, g) \in \mathbf{L}^2(\mathbb{R}^d)^2 \quad , \quad \|S_{J+1} f - S_{J+1} g\| \leq \|S_J f - S_J g\| \ . \tag{38}$$

Proof: Let $p = \{\lambda_n\}_{n \leq |p|}$ with $\lambda_n = (j_n, \gamma_n)$. The truncation of p at the scale 2^J is defined by stopping at the first $\lambda_m = (j_m, \gamma_m)$ with $j_m \geq J$:

$$a_J(p) = \{\lambda_n\}_{n<m} \ \text{ with } m = \min\{n \,:\, j_n \geq J\} \ .$$

We are going show that if $p \in P_J$ then summing the scattering distance over paths $q \in P_{J+1}$ such that $a_J(q) = p$ gives

$$\|S_J(p) f - S_J(p) g\|^2 \geq \sum_{\substack{q \in P_{J+1} \\ a_J(q) = p}} \|S_{J+1}(q) f - S_{J+1}(q) g\|^2 \ . \tag{39}$$

Summing on both sides of this inequality for all $p \in P_J$ proves the proposition result (38).

To prove (39), let us consider $p \in P_J$. Since wavelets satisfy the unitary property (10) for $\delta = 0$, Appendix D proves in (178) that

$$\|h \star \phi_J\|^2 = \|h \star \phi_{J+1}\|^2 + \sum_{\gamma=1}^{|\Gamma|} \|h \star \psi_{J,\gamma}\|^2 \ . \tag{40}$$

For $h = S(p) f - S(p) g$ it yields

$$\|S_J(p) f - S_J(p) g\|^2 = \|S_{J+1}(p) f - S_{J+1}(p) g\|^2 + \sum_{\gamma=1}^{|\Gamma|} \|(S(p) f - S(p) g) \star \psi_{J,\gamma}\|^2 \tag{41}$$

Let $q \in P_{J+1}$ with $a_J(q) = p$. Either $q = p$ in which case $S_{J+1}f(q) = S_{J+1}(p)f$. Otherwise $|q| > |p|$ in which case q is a continuation of $p + (J,\gamma)$ for some $\gamma \in \Gamma$. The set of all $S_{J+1}f(q)$ where $q \in P_{J+1}$ is a continuation of $p + (J,\gamma)$ is computed from $S(p + (J,\gamma))f$ by iteratively applying the contractive operator $\overline{U}_{J+1}$. It results that

$$\sum_{\gamma\in\Gamma} \|S(p+(J,\gamma))f - S(p+(J,\gamma))g\|^2 \geq \sum_{\substack{q\in P_{J+1}\\ a_J(q)=p,|q|>|p|}} \|S_{J+1}(q)f - S_{J+1}(q)g\|^2 .$$

Since $S(p+(J,\gamma))f = |S(p)f \star \psi_{J,\gamma}|$, inserting this in (41) proves (39). $\square$

When J increases, Proposition 7 proves that $\|S_J f - S_J g\|$ is a positive decreasing sequence. It thus converges to a limit metric when J goes to ∞. The limit metric satisfies the same contracting property

$$\forall (f,g) \in \mathbf{L}^2(\mathbb{R}^d)^2 \quad , \quad \lim_{J\to+\infty} \|S_J f - S_J g\| \leq \|f - g\| .$$

If (31) is satisfied then it also preserves the $\mathbf{L}^2(\mathbb{R}^d)$ norm

$$\lim_{J\to+\infty} \|S_J f\| = \|f\| .$$

3.4 Translation and Deformation Invariance

Translation invariance is the result of maintaining the commutativity of scattering operators with translations, and the fact $S_J(p)f$ that converges to a constant as J increases. The main theorem proves that a scattering operator is also Lipschitz continuous relatively to deformations, up to a log term.

Since $S_J(p)f$ is computed with convolutions and modulus operators which commute with translation, it also commutes with translations T_τ:

$$S_J(p)T_\tau f(x) = S_J(p)f(x-\tau) = T_\tau S_J(p)f(x) . \tag{42}$$

The following theorem proves that the scattering metric converges to a metric which is invariant to translation operators.

Theorem 2 *If the wavelet transform is unitary and the wavelet condition (31) is satisfied then*

$$\forall f \in \mathbf{L}^2(\mathbb{R}^d) \quad , \quad \lim_{J\to\infty} \|S_J f - S_J T_\tau f\| = 0 .$$

Proof: We saw in (42) that a translation T_τ commutes with S_J and hence that $\|S_J T_\tau - S_J\| = \|T_\tau S_J - S_J\|$. Since $S_J(p) f = A_J\, S(p) f$

$$\begin{aligned} \|T_\tau S_J f - S_J f\|^2 &= \sum_{p \in P_J} \|T_\tau A_J S(p) f - A_J S(p) f\|^2 \\ &\leq \|T_\tau A_J - A_J\|^2 \sum_{p \in P_J} \|S(p) f\|^2\,. \end{aligned} \quad (43)$$

Lemma 1 applied to $D_\tau = T_\tau$ proves in (14), for $h_j = \phi_J$, that there exists $C > 0$ such that

$$\|T_\tau A_J - A_J\| \leq C\ 2^{-J}\ |\tau|_\infty\ . \quad (44)$$

Inserting this in (43) gives

$$\|T_\tau S_J f - S_J f\|^2 \leq C\ 2^{-J}\, |\tau|_\infty \sum_{p \in P} \|S(p) f\|^2\,. \quad (45)$$

If condition (31) is satisfied then Lemma 3 proves in (36) that

$$\frac{\alpha}{2} \sum_{p \in P_J} \|S(p) f\|^2 \leq \|f \star \phi_J\|^2 + \sum_{j<J} \sum_{1 \leq \gamma \leq |\Gamma|} (J-j)\, \|W_{j,\gamma} f\|^2 \leq \max(J,1)\, \|f\|^2 + \|f\|_w^2, \quad (46)$$

with $\|f\|_w^2 = \sum_{j<0} |j| \sum_{1 \leq \gamma \leq |\Gamma|} \|W_{j,\gamma} f\|^2$. If $\|f\|_w < \infty$ then it results from (45) that

$$\|T_\tau S_J f - S_J f\|^2 \leq (J \|f\|^2 + \|f\|_w^2)\, C\, 2\, \alpha^{-1}\, 2^{-J}\ |\tau|_\infty$$

so $\lim_{J \to \infty} \|T_\tau S_J f - S_J f\| = 0$.

Let us now prove that $\lim_{J \to \infty} \|T_\tau S_J f - S_J f\| = 0$ for all $f \in \mathbf{L}^2(\mathbb{R}^d)$, with a similar density argument as in the proof of Theorem 1. If $f_n = f \star \phi_n$ then $\lim_{n \to -\infty} \|f - f_n\| = 0$ and we showed that $\|f_n\|_w < \infty$. We decompose

$$\|T_\tau S_J f - S_J f\| \leq \|T_\tau S_J f_n - T_\tau S_J f\| + \|T_\tau S_J f_n - S_J f_n\| + \|S_J f_n - S_J f\|\ .$$

Since $\lim_{J \to \infty} \|T_\tau S_J f_n - S_J f_n\| = 0$ and $\|T_\tau S_J f_n - T_\tau S_J f\| = \|S_J f_n - S_J f\| \leq \|f - f_n\|$, by letting n go to $-\infty$ we prove that $\lim_{J \to \infty} \|T_\tau S_J f - S_J f\| = 0$, which finishes the proof.$\square$

Lipschitz continuity is studied by considering deformations defined by a non-constant displacement field $\tau(x)$: $D_\tau f(x) = f(x - \tau(x))$. We write $H\tau(x)$ the Hessian tensor of τ in x and $|H\tau(x)|$ its tensor norm. For any $(y,z) \in \mathbb{R}^d$ it satisfies $|y\, H\tau(x)\, z| \leq |H\tau(x)|\, |y|\, |z|$. The sup norm of this tensor over all

x is written $|H\tau|_\infty = \sup_{x\in\mathbb{R}^d} |H\tau(x)|$. The following theorem is a central result of this paper, which computes an upper bound of $\|S_J - S_J D_\tau\|$. We denote by $\|Sf\|_{1,P_J}$ the mixed $\mathbf{l}^1$ and $\mathbf{l}^2$ norm defined by

$$\|Sf\|_{1,P_J} = \sum_{m=0}^{+\infty} \Big(\sum_{p\in P_J, |p|=m} \|S(p)f\|^2 \Big)^{1/2} . \tag{47}$$

The theorem also computes the deformation error over the restriction $S_{J,m}$ of S_J to paths $|p| < m$. We write $(a \vee b) = \max(a,b)$.

Theorem 3 *If $|\nabla\tau|_\infty \leq 1-\epsilon$ with $\epsilon > 0$ then there exists C such that for all $f \in \mathbf{L}^2(\mathbb{R}^d)$ with $\|Sf\|_{1,P_J} < \infty$*

$$\|S_J D_\tau f - S_J f\| \leq C\, \|Sf\|_{1,P_J} \Big(2^{-J}|\tau|_\infty + |\nabla\tau|_\infty (\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty \Big) , \tag{48}$$

and for all $m \geq 0$

$$\|S_{J,m} D_\tau f - S_{J,m} f\| \leq C\, m\, \|f\|^2 \Big(2^{-J}|\tau|_\infty + |\nabla\tau|_\infty (\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty \Big) . \tag{49}$$

Proof: Let $[S_J, D_\tau] = S_J\, D_\tau - D_\tau\, S_J$,

$$\|S_J D_\tau f - S_J f\| \leq \|D_\tau S_J f - S_J f\| + \|[S_J\, ,\, D_\tau] f\| . \tag{50}$$

Similarly to (43) the first term on the right satisfies

$$\|D_\tau S_J f - S_J f\| \leq \|D_\tau A_J - A_J\| \Big(\sum_{p\in P_J} \|S(p)f\|^2 \Big)^{1/2} . \tag{51}$$

Since

$$\Big(\sum_{p\in P_J} \|S(p)f\|^2 \Big)^{1/2} \leq \sum_{m=0}^{+\infty} \Big(\sum_{p\in P_J, |p|=m} \|S(p)f\|^2 \Big)^{1/2}$$

it results that

$$\|D_\tau S_J f - S_J f\| \leq \|D_\tau A_J - A_J\|\, \|Sf\|_{1,P_J} . \tag{52}$$

Since S_J iterates on the propagator $\overline{U}_J$ which is contractive, Appendix E proves the following upper bound on scattering commutators.

Lemma 4 *For any operator O over $\mathbf{L}^2(\mathbb{R}^d)$*

$$\|[S_J\,,\,O]f\| \leq \|Sf\|_{1,P_J}\,\|[\overline{U}_J\,,\,O]\|\ . \tag{53}$$

If $O = D_\tau$ then we also have

$$\|[\overline{U}_J, D_\tau]\| \leq \|[\overline{W}_J, D_\tau]\|\ . \tag{54}$$

Indeed, $\overline{U}_J = M_J\,\overline{W}_J$, where $M_J\{h_J\,,\,h_j\}_{-\infty<j<J} = \{h_J\,,\,|h_j|\}_{-\infty<j<J}$ is a contracting modulus operator. Since $M_J D_\tau = D_\tau M_J$

$$\|[\overline{U}_J, D_\tau]\| = \|M_J\,[\overline{W}_J, D_\tau]\| \leq \|[\overline{W}_J, D_\tau]\|\ . \tag{55}$$

Inserting (53) with (54) and (52) in (50) gives

$$\|S_J D_\tau f - S_J f\| \leq \|Sf\|_{1,P_J}\Big(\|D_\tau A_J - A_J\| + \|[\overline{W}_J, D_\tau]\|\Big)\ . \tag{56}$$

Lemma 1 proves that

$$\|D_\tau A_J - A_J\| \leq C\,2^{-J}\,|\tau|_\infty\ . \tag{57}$$

It result from (56) and (57) that

$$\|S_J D_\tau f - S_J f\| \leq C\,\|Sf\|_{1,P_J}\Big(2^{-J}\,|\tau|_\infty + \|[\overline{W}_J, D_\tau]\|\Big)\ . \tag{58}$$

To prove (48), the main difficulty is to compute an upper bound of $\|[\overline{W}_J, D_\tau]\|$, and hence of $\|[\overline{W}_J, D_\tau]\|^2 = \|[\overline{W}_J, D_\tau]^*\,[\overline{W}_J, D_\tau]\|$, where O^* is the adjoint of an operator O.

The wavelet commutator applied to f is

$$[\overline{W}_J, D_\tau]f = \{[W_\lambda, D_\tau]f\ ,\ [A_J, D_\tau]f\}_{\lambda\in\Lambda}\ .$$

whose norm is

$$\|[\overline{W}_J, D_\tau]f\|^2 = \sum_{\lambda\in\Lambda}\|[W_\lambda, D_\tau]f\|^2 + \|[A_J, D_\tau]f\|^2\ . \tag{59}$$

It results that

$$[\overline{W}_J, D_\tau]^*\,[\overline{W}_J, D_\tau] = \sum_{\lambda\in\Lambda}[W_\lambda, D_\tau]^*\,[W_\lambda, D_\tau] + [A_J, D_\tau]^*\,[A_J, D_\tau]\ ,$$

The operator $[\overline{W}_J, D_\tau]^*\,[\overline{W}_J, D_\tau]$ is singular but Appendix C proves that its norm satisfies the following bound.

Lemma 5 *If* $|\nabla\tau|_\infty \leq 1-\epsilon$ *with* $\epsilon > 0$ *then there exists* $C > 0$ *such that for all* $J \in \mathbb{Z}$

$$\|[\overline{W}_J, D_\tau]\| \leq C\left(|\nabla\tau|_\infty(\log\frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty\right) . \tag{60}$$

Inserting the wavelet commutator bound (60) in (58) proves the first theorem inequality (48). The second theorem inequality (49) applies to the restriction of S_J to paths of length $|p| < m$. The same proof shows that

$$\begin{aligned}\|S_{J,m}D_\tau f - S_{J,m}f\| \leq \ & C \sum_{n=0}^{m-1}\left(\sum_{p\in P_J,|p|=n}\|S(p)f\|^2\right)^{1/2} \\ & \left(2^{-J}|\tau|_\infty + |\nabla\tau|_\infty(\log\frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty\right) .\end{aligned} \tag{61}$$

Proposition 5 shows in (22) that paths of length n are obtained by applying the contractive operator $\overline{U}_J$ on paths of length $n-1$, so

$$\sum_{p\in P_J,|p|=n}\|S(p)f\|^2 \leq \sum_{p\in P_J,|p|=n-1}\|S(p)f\|^2 \leq \|f\|^2 , \tag{62}$$

hence

$$\sum_{n=0}^{m-1}\left(\sum_{p\in P_J,|p|=n}\|S(p)f\|^2\right)^{1/2} \leq m\,\|f\| . \tag{63}$$

Inserting this in (61) proves the second theorem inequality (49). $\square$

In this theorem formulation, the constant C depends upon $\epsilon > 0$. This dependency is removed by multiplying the upper bounds in (48) and (49) by $(1-|\nabla\tau|_\infty)^{-3d}$. The proof in Appendix C shows that the constant C then does not depends upon τ, and it is sufficient to impose that $|\nabla\tau| < 1$. The same remark applies to all subsequent theorems where the condition $|\nabla\tau|_\infty \leq 1-\epsilon$ appears. Applications [5] are usually implemented with a maximum path length m equal to 3 or 4, because the total energy of $S_J(p)f$ is most often negligible for $|p| > 3$.

The theorem includes two error terms. The elastic deformation term is dominated by $|\nabla\tau|_\infty$ and does not depend upon J. The translation term is proportional to $2^{-J}|\tau|_\infty$ and requires to choose a sufficiently large scale $2^J \gg |\tau|_\infty^{-1}$. A scattering transform can increase arbitrarily the scale 2^J while preserving high frequency information through multiple interferences.

Similar error terms appear in the wavelet transform Proposition 2, but the requirement to use large scales then impose to remove all high frequencies.

The following corollary derives that the scattering metric is almost Lipschitz continuous relatively to small deformations, up to a log term.

Corollary 1 *If* $|\nabla\tau|_\infty \leq 1-\epsilon$ *with* $\epsilon > 0$ *and* $J \geq \log \frac{|\tau|_\infty}{|\nabla\tau|_\infty}$ *then*

$$\|S_{J,m} D_\tau f - S_{J,m} f\| \leq C\, m\, \|f\| \left(|\nabla\tau|_\infty (\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty \right) . \quad (64)$$

This corollary proves that scattering operators are not only translation invariant but also stable to elastic deformations. The translation error term $2^{-J}\|\tau\|_\infty$ in Theorem 3 can be reduced with a first order Taylor expansion $S_J f - \tau\,.\,\nabla S_J f$ of $S_J D_\tau f$. The following theorem proves that it yields a second order term proportional to $2^{-2J}|\tau|^2_\infty$.

Theorem 4 *If* $|\nabla\tau|_\infty \leq 1-\epsilon$ *with* $\epsilon > 0$ *then there exists* C *such that for all* $f \in \mathbf{L}^2(\mathbb{R}^d)$

$$\|S_J D_\tau f - S_J f + \tau\,.\,\nabla S_J f\| \leq C\, \|Sf\|_{1,P_J} \Big(2^{-2J} |\tau|^2_\infty + |\nabla\tau|_\infty (\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty) \Big) . \quad (65)$$

Proof: The proof proceeds as the proof of Theorem 3. Replacing $S_J D_\tau - S_J$ by $S_J D_\tau - S_J + \tau\,.\,\nabla S_J$ in the derivation steps of the proof of Theorem 3 amounts to replace $D_\tau A_J - A_J$ by $D_\tau A_J - A_J + \nabla A_J$. Equation (56) then becomes

$$\|S_J D_\tau f - S_J f + \tau\,.\,\nabla S_J\|_{\tilde{P}_J} \leq \|Sf\|_{1,P_J} \Big(\|D_\tau A_J - A_J + \nabla A_J\| + \|[\overline{W}_J, D_\tau]\| \Big) . \quad (66)$$

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

$$\|D_\tau A_J f - A_J + \nabla A_J\| \leq C\, 2^{-2J}\, |\tau|^2_\infty . \quad (67)$$

Inserting the upper bound (60) of $\|[\overline{W}_J, D_\tau]\|$ proves (65). $\square$

If the scattering norm is reduced to path of order at most m then $\|Sf\|_{1,P_J}$ in (66) can be replaced by the upper bound $m\,\|f\|$. This theorem gives an approximation of $S_J D_\tau f$ which is linear in τ. If we neglect the residual error, then $\tau(x)$ can be estimated at each x by solving the system of linear equations

$$\forall p \in P_J \quad , \quad S_J D_\tau f(x,p) - S_J f(x,p) + \tau(x)\,.\,\nabla S_J f(x,p) \approx 0. \quad (68)$$

In dimension d, $\tau(x)$ has d coordinates which can be computed if this resulting system has rank d. Estimating $\tau(x)$ has many applications. For example, the displacement field $\tau(x)$ between two consecutive images of a video sequence is proportional to the optical flow velocity of the image points [3]. Similarly, the relative displacement field of points between two images of a stereo pair is inversely proportional to the distance between the camera and the points in the three dimensional scene. The residual error of the system (68) is small only if $2^J \gg |\tau|_\infty$. However, as opposed to algorithms which rely on lower frequencies to compute large displacements, the system (68) of scattering coefficients also takes into account high frequency signal information to compute $\tau(x)$ at large scales 2^J.

4 Compact and Discrete Scattering

A periodic scattering is introduced for compactly supported functions. Section 4.2 derives a fast discrete scattering algorithm, computed along scale increasing paths.

4.1 Periodic Scattering

The scattering operator S_J is redefined over functions having a support included in a compact set $\Omega \subset \mathbb{R}^d$, by replacing the wavelet transform operator $\overline{W}_J$ defined in $\mathbf{L}^2(\mathbb{R}^d)$ by a wavelet transform $\overline{W}_J$ defined over $\mathbf{L}^2(\Omega)$. At the boundary of Ω, wavelets must be adjusted so that their support remains within Ω. Many approaches have been developed for this purpose, depending upon the boundary properties of Ω. When Ω is a hyperrectangle, the simplest approach is a periodic extension [13, 15]. Similarly to Fourier series, a periodic scattering operator maps periodic functions into $\mathbf{l}^2$ sequences.

Let us consider functions having a support in $[0,\Delta]^d$. Each $f \in \mathbf{L}^2[0,\Delta]^d$ is extended into a Δ periodic function $\tilde{f}(x) = \sum_{m\in\mathbb{Z}^d} f(x - \Delta m)$. The periodic wavelet transform of f is calculated as a wavelet transform of $\tilde{f}$. Let $\psi_\lambda(x) = 2^{-dj}\psi_\gamma(2^{-j}x)$ for $\gamma \in \Gamma$ and $\lambda = (j,\gamma)$,

$$\widetilde{W}_\lambda f(x) = \int \tilde{f}(u)\, \psi_\lambda(x-u)\, du = \tilde{f} \star \psi_\lambda(x) \ . \tag{69}$$

Since $\tilde{f}(x) = \sum_{m\in\mathbb{Z}^d} f(x-\Delta m)$,

$$\widetilde{W}_\lambda f(x) \ \ = \ \ \int f(u)\, \tilde{\psi}_\lambda(x-u)\, du = f \star \tilde{\psi}_\lambda(x)$$

$$= \int_{[0,\Delta]^d} \tilde f(u)\, \tilde\psi_\lambda(x-u)\, du = \tilde f \star \tilde\psi_\lambda(x) \tag{70}$$

which is a circular convolution with periodized wavelets

$$\tilde\psi_\lambda(x) = \sum_{m\in\mathbb{Z}^d} \psi_\lambda(x - \Delta m)\ . \tag{71}$$

Similarly, the scaling function ϕ is periodized $\tilde\phi_J(x) = \sum_{m\in\mathbb{Z}^d} \phi_J(x-\Delta m)$

$$\widetilde{A}_J f(x) = \int f(u)\, \tilde\phi_J(x-u)\, du = \widetilde{A}_J \tilde f(x) = \tilde f \star \tilde\phi_J(x)\ . \tag{72}$$

We suppose that

$\forall n \in \mathbb{Z}^d - \{0\}$, $\hat\phi(2n\pi/\Delta) = 0$, $\hat\phi(0) = 1$ and $\hat\phi(\omega) \neq 0$ for $|\omega| < 2\pi/\Delta$.

The Poisson formula implies that

$$\tilde\phi(x) = \sum_{n\in\mathbb{Z}^d} \phi(x-\Delta n) = \Delta^{-d}.$$

Since all integrals are circular convolutions with periodized wavelets and scaling functions, we do not distinguish anymore f and $\tilde f$.

The periodic wavelet transform $\overline{\widetilde{W}}_J f = \{\widetilde{A}_J f\ ,\ \widetilde{W}_\lambda f\}_{\lambda\in\Lambda}$ is composed of periodic functions whose restriction to $[0,\Delta]^d$ are in $\mathbf{L}^2[0,\Delta]^d$. It inherits all properties of the original wavelet transform in $\mathbf{L}^2(\mathbb{R}^d)$. If the Littlewood-Paley property (10) is satisfied then one can verify that the periodic wavelet transform is contractive for the $\mathbf{L}^2[0,\Delta]^d$ norm $\|f\|^2 = \int_{[0,\Delta]^d} |f(x)|^2\, dx$:

$$(1-\delta)\|f\|^2 \leq \|\overline{\widetilde{W}}_J f\|^2 = \|\widetilde{A}_J f\|^2 + \sum_{\lambda\in\Lambda} \|\widetilde{W}_\lambda f\|^2 \leq \|f\|^2\ . \tag{73}$$

Indeed, the Fourier series coefficients of $\tilde\psi_\lambda$ are $\Delta^{-d}\hat\psi_\lambda(2n\pi/\Delta)$. The Littlewood-Paley property (10) on the wavelets ψ_λ thus implies an equivalent property on the periodic wavelets $\tilde\psi_\lambda$, which implies (73).

A periodic wavelet transform modulus propagator is defined by

$$\overline{\widetilde{U}}_J f = \{\widetilde{A}_J f\ ,\ \widetilde{U}_\lambda f = |\widetilde{W}_\lambda f| = |f \star \tilde\psi_\lambda|\}_{\lambda\in\Lambda}\ . \tag{74}$$

Since the wavelet transform is contractive, $\overline{\widetilde{U}}_J$ is also contractive and it preserves the norm if the wavelet transform is unitary. Iterating on this propagator defines a scattering operator over any path $p = \{\lambda_n\}_{n\leq|p|} \in P_J$

$$\widetilde{S}(p) f = \prod_{n=1}^{|p|} \widetilde{U}_{\lambda_n} f = |\cdots|f \star \tilde\psi_{\lambda_1}| \star \tilde\psi_{\lambda_2}| \cdots| \star \tilde\psi_{\lambda_{|p|}}|\ ,$$

and

$$\widetilde{S}_J(p)f = A_J S(p)f = S(p)f \star \tilde{\phi}_J \ . \tag{75}$$

Since $\tilde{\phi}(x) = \Delta^{-d}$, at the maximum scale $2^J = 1$, $\widetilde{S}_0(p)f$ is a constant for each $p = \{\lambda_n\}_{n \leq |p|}$:

$$\widetilde{S}_0(p)f = \Delta^{-d} \int_{[0,\Delta]^d} |\cdots|f \star \tilde{\psi}_{\lambda_1}| \star \tilde{\psi}_{\lambda_2}| \cdots | \star \tilde{\psi}_{\lambda_{|p|}}(x)|\, dx \ . \tag{76}$$

Similarly to a Fourier transform, a periodic scattering operator associates to $f \in \mathbf{L}^2[0,\Delta]^d$ a discrete sequence of values $\{\widetilde{S}_0(p)f\}_{p \in \widetilde{P}_0}$. For $p = 0$, $\widetilde{S}_0(0)f$ is the average of f over $[0,\Delta]^d$. For $|p| = 1$, $\widetilde{S}_0(\lambda_1)f$ is proportional to the $\mathbf{L}^1[0,\Delta]^d$ norm of wavelet coefficients $W_{\lambda_1}f$. For $|p| \geq 2$ these coefficients gives the $\mathbf{L}^1[0,\Delta]^d$ norm of higher order interferences.

For any $J \leq 0$, the scattering metric is defined by

$$\|\widetilde{S}_J f - \widetilde{S}_J g\|^2 = \sum_{p \in P_J} \|\widetilde{S}_J(p)f - \widetilde{S}_J(p)g\|^2$$

with $\|\widetilde{S}_J(p)f\|^2 = \int_{[0,\Delta]^d} |\widetilde{S}_J(p)f(x)|^2\, dx$. When $J = 0$, since $\widetilde{S}_0(p)f$ is constant, $\|\widetilde{S}_0(p)f\|^2 = \Delta^d |\widetilde{S}_0(p)f|^2$ and

$$\|\widetilde{S}_0 f - \widetilde{S}_0 g\|^2 = \Delta^d \sum_{p \in \widetilde{P}_0} |\widetilde{S}_0(p)f - \widetilde{S}_0(p)g|^2 \ .$$

The following proposition proves that $\widetilde{S}_J$ is contractive and preserves the norm for appropriate wavelets.

Proposition 8 *For Littlewood-Paley wavelets, a periodic scattering operator is contractive:*

$$\forall (f,g) \in \mathbf{L}^2[0,\Delta]^d \quad , \quad \|\widetilde{S}_J f - \widetilde{S}_J g\| \leq \|f - g\| \ . \tag{77}$$

If the wavelet transform is unitary and satisfies (31) then $\|\widetilde{S}_J f\| = \|f\|$.

Proof: The proof of (77) proceeds as the proof of (26), by writing $\widetilde{S}_J$ through an iteration (24) of a periodic one-step propagators $\overline{\widetilde{U}}_J$ in (74), which is contracting.

The proof of the norm preservation is identical to the proof in Theorem 1. This proof relies on the wavelet Fourier transform properties. Since periodic wavelet have a Fourier transform which is a uniform sampling of non-periodic wavelets, the Fourier transform condition (31) over all frequencies $\omega \in \mathbb{R}^d$ is a sufficient condition for the periodic case. $\square$

Let us now show that a periodic scattering operator $\widetilde{S}_0$ provides invariant representations for circular translations. It is used in Section 6.2 to construct image representations that are invariant to rotations in $\mathbb{R}^2$. A circular translation is defined by

$$\widetilde{T}_\tau f(x) = f((x-\tau)\mathrm{mod}\Delta)$$

where $x \bmod \Delta = x - n\Delta \in [0,\Delta)^d$ with $n \in \mathbb{Z}$. Circular convolutions commute with circular translations. Since $\widetilde{S}_J f$ is computed with a cascade of circular convolutions and modulus operators, it also commutes with circular translations. For any path p and any $J \geq 0$:

$$\widetilde{S}_J(p)(\widetilde{T}_\tau f)(x) = \widetilde{S}_J(p) f((x-\tau)\mathrm{mod}\Delta) = \widetilde{T}_\tau \widetilde{S}_J(p) f(x)\ . \qquad (78)$$

Since $\widetilde{S}_0(p)f$ is a constant, it results that for all p and f, $\widetilde{S}_0(p)(\widetilde{T}_\tau f) = \widetilde{S}_0(p)f$ and hence that $\widetilde{S}_0\widetilde{T}_\tau = \widetilde{S}_0$.

4.2 Fast Discrete Scattering

A discrete periodic scattering transform is computed over signals $f[n]$ of size N^d, which are considered as discretizations of functions having a support in $[0,\Delta]^d$. The finer scale is $2^L = N^{-1}$. The periodic wavelet transform is replaced by a discrete periodic wavelet transform at scales $2^L \leq 2^j < 2^J$ computed with discrete circular convolutions.

The scattering transform is calculated along scale increasing paths $p = \{\lambda_n = (j_n, \gamma_n)\}_{n \leq |p|}$ for which $j_n < j_{n+1} < J$ because they carry most of the scattering energy. The last scale $j_{|p|}$ is thus the largest one. The Fourier transform of $\widetilde{S}(p)f$ is concentrated over a frequency domain of width proportional to $2^{-j_{|p|}}$ and is thus subsampled at intervals $a\,N\,2^{j_{|p|}}$, where a is an oversampling factor adjusted to reduce aliasing. For one-dimensional cubic splines wavelets, we set $a = 1/2$. The scattering transform is obtained with a succession of subsampled convolutions described below.

The algorithm is initialized with $\widetilde{S}(0)f[n] = f[n]$. The scattering transform is calculated layer per layer, for paths of increasing length m. The maximum length of scale increasing paths is $J - L \leq \log_2 N$. Calculations are usually restricted to paths of maximum length m_0 typically equal to 4, because the scattering energy is negligible beyond this length.

For m going from 0 to $m_0 - 1$, paths of length $m+1$ are computed from paths p of length m by calculating for all $\lambda = (j, \gamma) \in \Lambda$ with $j > j_{|p|}$:

$$\widetilde{S}(p+\lambda)f[n] = |\,\widetilde{S}(p)f \star \tilde{\psi}_\lambda[n\alpha]\,| \quad \text{with} \quad \alpha = a\,N\,2^{j-j_{|p|}}\,. \tag{79}$$

Scattering along paths of length m are then averaged by ϕ_J:

$$\widetilde{S}_J(p)f[n] = \widetilde{S}(p)f \star \tilde{\phi}_J[n\alpha] \quad \text{with} \quad \alpha = a\,N\,2^{J-j_{|p|}}\,. \tag{80}$$

For $m = m_0$, we only compute the averaging (80).

If there are $|\Gamma|$ different mother wavelets then the number of scale increasing paths of length $|p| \leq m_0$ is $O(|\Gamma|^{m_0}\,(\log_2 N + J)^{m_0})$. Since all signals $S_J(p)f$ include $a^{-1}2^{-dJ}$ samples, the scattering transform along all scale increasing paths of length smaller then m_0 has a total of $O(|\Gamma|^{m_0}(\log_2 N + J)^{m_0} 2^{-dJ})$ samples. At the maximum scale $J = 0$, the scattering size is $O(|\Gamma|^{m_0}\,(\log_2 N)^{m_0})$. Convolutions in (79) and (80) are circular convolutions computed with an FFT. The scattering transform at any scale 2^J then requires $O(|\Gamma|\,N^d\,\log_2 N)$ operations.
Software is available at *www.cmap.polytechnique.fr/scattering*.

5 Scattering of Stationary Processes

A translation invariant scattering leads to a new metric on stationary processes, which is Lipschitz continuous to random deformations. Section 5.2 studies the metric estimation and its consistency.

5.1 Scattering Distance over Processes

The Fourier power spectrum provides a representation of stationary processes F based on their second order moments. We write $E\{F\} = E\{F(x)\}$ an expected value which does not depend upon x. The power spectrum $\hat{R}_F(\omega)$ is the Fourier transform of the auto-covariance $R_F(x-y) = E\{(F(x) - E\{F\})(F(y) - E\{F\})\}$. It characterizes stationary Gaussian processes but is not sufficient to discriminate non-Gaussian processes whose properties depend upon high order moments. Moreover, the power spectrum is not Lipschitz continuous to deformations because the Fourier transform modulus is not Lipschitz continuous as explained in Section 2.1. A small deformation of a stationary process may produce a large modification of its power spectrum at high frequencies. We introduce a scattering metric over stationary processes and prove that it is almost Lipschitz continuous to deformations, up to a log term. The $\mathbf{L}^2(\mathbb{R}^d)$ norm is replaced by the mean square norm $E\{|F|^2\}$ over stationary processes. Scattering operator properties in $\mathbf{L}^2(\mathbb{R}^d)$ are transposed over stationary processes by exchanging these norms.

A scattering operator is computed with a sequence of convolutions and modulus operator. Each $S_J(p)F(x)$ is therefore a stationary process if F is stationary. Since $\int \phi_J(x)\,dx = 1$, the expected value of $S_J(p)F(x) = S(p)F \star \phi_J(x)$ does not depend upon J and x and is written:

$$\overline{S_J(p)F} = E\{S_J(p)F(x)\} = E\{S(p)F(x)\} = \overline{S(p)F}\ . \tag{81}$$

If $p = 0$ then $\overline{S(0)F} = E\{F\}$. Paths of length one, $p = \{\lambda_1\}$, give the first order moments of wavelet coefficients:

$$\overline{S(\lambda_1)F} = E\{|F \star \psi_{\lambda_1}|\}\ .$$

For $|p| \geq 2$, $\overline{S(p)F}$ depends upon higher order statistics of F.

The scattering distance between two stationary processes F and G at a scale 2^J is defined by:

$$\|\overline{S_J\,F} - \overline{S_J\,G}\|^2 = \sum_{p \in P_J} |\overline{S(p)F} - \overline{S(p)G}|^2\ . \tag{82}$$

The following proposition proves that this scattering norm is contractive over stationary processes. We assume that wavelets satisfy the Littlewood-Paley condition (10) for some $\delta < 1$.

Proposition 9 *If F is stationary then*

$$\forall J \in \mathbb{Z}\quad,\quad \|\overline{S_J\,F}\|^2 \leq E\{|F|^2\}\ . \tag{83}$$

Proof: We first show that $\overline{W_J}$ is a contracting operator over stationary processes. Let us write

$$E\{|\overline{W_J}F|^2\} = E\{|A_JF|^2\} + \sum_{\lambda \in \Lambda} E\{|W_\lambda F|^2\}\ .$$

Both $A_J F = F \star \phi_J$ and $W_\lambda F = F \star \psi_\lambda$ are stationary. Since $\int \phi_J(x)\,dx = 1$ and $\int \psi_\lambda(x)\,dx = 0$ it results that $E\{A_J F\} = E\{F\}$ and $E\{W_\lambda F\} = 0$. Hence

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

and

$$E\{|W_\lambda F|^2\} = \int \hat{R}_F(\omega)\,|\hat{\psi}_\lambda(\omega)|^2\,d\omega.$$

Since $E\{|F|^2\} = \int \hat{R}_F(\omega)\,d\omega + E\{F\}^2$, the Littlewood-Paley condition (10) implies that

$$(1-\delta) E\{|F|^2\} \leq E\{|\overline{W}_J F|^2\} \leq E\{|F|^2\}\ . \tag{84}$$

The propagator $\overline{U_J} F = \{A_J F\,,\,|W_\lambda F|\}_{\lambda\in\Lambda}$ satisfies

$$E\{|\overline{U_J} F|^2\} = E\{|\overline{W_J} F|^2\} \leq E\{|F|^2\}$$

and is thus also contractive over stationary processes. By decomposing S_J as iterations on $\overline{U}_J$, we prove as in (26) that

$$E\{|S_J F|^2\} = \sum_{p\in P_J} E\{|S_J(p) F|^2\} \leq E\{|F|^2\}\ . \tag{85}$$

But

$$|\overline{S(p)F}|^2 = |E\{S_J(p)F\}|^2 \leq E\{|S_J(p)F|^2\}$$

so (85) implies (83). $\Box$

It results from (83) that

$$\|\overline{S_J\,F} - \overline{S_J\,G}\|^2 = \sum_{p\in P_J} |\overline{S(p)F} - \overline{S(p)G}|^2 \leq 2(E\{|F|^2\} + E\{|G|^2\}). \tag{86}$$

Since $P_J \subset P_{J+1}$, $\|\overline{S_J\,F} - \overline{S_J\,G}\|$ is monotonically increasing and since it is bounded it converges when J goes to ∞. Let $P_\infty = \cup_{J\in N} P_J$ be the set of all paths $p = \{\lambda_n = (j_n, \gamma_n)\}_{n\leq|p|}$ with no restriction on $\max_{n\leq|p|} j_n$. The limit scattering metric is

$$\lim_{J\to\infty} \|\overline{S_J\,F} - \overline{S_J\,G}\|^2 = \sum_{p\in P_\infty} |\overline{S(p)F} - \overline{S(p)G}|^2\ . \tag{87}$$

Similarly to Theorem 1, the following theorem proves that for appropriate wavelets the scattering energy goes to zero as the path length increases, so the scattering distance is dominated by shorter path lengths. It results a scattering energy conservation.

Theorem 5 *If the wavelet transform is unitary and wavelets satisfy condition (31) then*

$$\lim_{m\to\infty} \sum_{p\in P_J, |p|=m} E\{|S(p)F|^2\} = 0 \quad , \quad \lim_{m\to\infty} \sum_{p\in P_J, |p|>m} E\{|S_J(p)F|^2\} = 0 \quad (88)$$

and

$$\sum_{p\in P_J} E\{|S_J(p)F|^2\} = E\{|F|^2\} \ . \quad (89)$$

Proof: The wavelet transform is unitary if the Littlewood-Paley condition (10) is satisfied for $\delta = 0$ so (84) proves that $E\{|\overline{U}_J F|^2\} = E\{|\overline{W}_J F|^2\} = E\{|F|^2\}$.

The proof of (88) is almost identical to the proof of (32) in Theorem 1, if we replace f by F, $|\hat{f}(\omega)|^2$ by the power spectrum $\hat{R}_F(\omega)$ and $\|f\|^2$ by $E\{|F|^2\}$. Similarly to Lemma 2, we show that if $h(x) \geq 0$ then

$$E\{|F \star \psi_\lambda| \star h\|^2\} \geq \sup_{\eta\in\mathbb{R}^d} E\{|F \star \psi_\lambda \star h_\eta|^2\} \ \text{ with } \ h_\eta(x) = h(x)\, e^{i\eta x} \ . \quad (90)$$

In the derivations of Lemma 3, replacing $f_p = S(p)f$ and $|\hat{f}_p(\omega)|^2$ by $F_p = S(p)F$ and $\hat{R}_{F_p}(\omega)$ proves that

$$\frac{\alpha}{2} \sum_{p\in P_J} E\{|S(p)F|^2\} \leq E\{|F \star \phi_J|^2\} + \sum_{j<J} \sum_{1\leq\gamma\leq|\Gamma|} (J-j) E\{|F \star \psi_{j,\gamma}|^2\} \ .$$

The same final density argument as in the proof of Theorem 1 then proves (88).

Since $E\{|\overline{U}_J F|^2\} = E\{|F|^2\}$, iterating m times on $\overline{U_J} F$ proves as in (29) that

$$\sum_{p\in P_J, |p|<m} E\{|S_J(p)F|^2\} + \sum_{p\in P_J, |p|=m} E\{|S(p)F|^2\} = E\{|F|^2\} \ .$$

When m goes to ∞, (88) implies (89). $\square$

Let us now study the impact of random stationary deformations on the scattering distance. If $F(x)$ is stationary and $\tau(x)$ is an independent stationary process then the randomly deformed process

$$D_\tau F(x) = F(x - \tau(x))$$

remains stationary. Its scattering transform $S_J(p) D_\tau F(x)$ is therefore also stationary. The following theorem adapts the result of Theorem 3 by proving that the scattering distance produced by a random deformation is dominated by a first order term proportional to $E\{|\nabla\tau|_\infty^2\}$. A simplified expression is obtained for a scattering $S_{J,m}$ restricted to paths of order $|p| < m$: $\|\overline{S_{J,m}\,F}\|^2 = \sum_{p\in P_J,|p|<m} \|\overline{S_J(p)F}\|^2$.

Theorem 6 *If $|\nabla\tau|_\infty \le 1-\epsilon$ with probability 1 for $\epsilon > 0$ then there exists C such that*

$$\|\overline{S_J\,D_\tau F} - \overline{S_J\,F}\| \le C\,\|S\,F\|_{1,P_J} E\Big\{\Big(|\nabla\tau|_\infty (\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty\Big)^2\Big\}^{1/2}, \tag{91}$$

with $\|S\,F\|_{1,P_J} = \sum_{n=0}^{+\infty} \Big(\sum_{p\in P_J,|p|=n} E\{|S(p)F|^2\}\Big)^{1/2}$, and

$$\|\overline{S_{J,m}\,D_\tau F} - \overline{S_{J,m}\,F}\| \le C\,m\,E\{|F|^2\}^{1/2}\,E\Big\{\Big(|\nabla\tau|_\infty (\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty\Big)^2\Big\}^{1/2}. \tag{92}$$

Proof: Since $E\{S_J(p)F(x)\} = \overline{S_J(p)F}$ does not depend on x, $E\{D_\tau S_J(p)F(x)\} = \overline{S_J(p)F}$, so

$$\begin{aligned}\|\overline{S_J\,D_\tau F} - \overline{S_J\,F}\|^2 &= \sum_{p\in P_J} |E\{S_J(p)D_\tau F - D_\tau S_J(p)F\}|^2 \\ &\le \sum_{p\in P_J} E\{|S_J(p)D_\tau F - D_\tau S_J(p) F|^2\} = E\{|[S_J, D_\tau]F|^2\}.\end{aligned}$$

Appendix G proves that

$$E\{|[S_J\,,\,D_\tau]F|^2\} \le C^2\,\|SF\|_{1,P_J}^2\,E\Big\{\Big(|\nabla\tau|_\infty (\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty\Big)^2\Big\}, \tag{93}$$

which implies (91). The commutator $[S_J, D_\tau]$ is a random operator 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 over stationary processes.

Lemma 6 *Let K_τ be an integral operator with a kernel $k_\tau(x,u)$ which depends upon a random process τ. If the following two conditions are satisfied*

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

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

$$E\{|K_\tau F|^2\} \leq E\{\|K_\tau\|^2\}\, E\{|F|^2\} \ , \tag{94}$$

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

This lemma is proved in Appendix F. The inequality (92) is derived from (91) by verifying that

$$\sum_{n=0}^{m-1} \Big(\sum_{p \in P_J, |p|=n} E\{|S(p)F|^2\} \Big)^{1/2} \leq m\, E\{|F|^2\}^{1/2} \ .$$

Since $E\{|\overline{U}_J F|^2\} \leq E\{|F|^2\}$, the proof of this result is identical to the proof of (63). $\square$

This theorem proves that a scattering metric satisfies a form of Lipschitz continuity relatively to the stochastic deformation gradient $\nabla\tau$, up to a log term. Small stationary deformations of stationary processes result in small modifications of the scattering distance, which is important for the classification of deformed stationary textures [5].

5.2 Consistent Scattering Distance Estimation

For classification problems such as texture discrimination, a consistent estimation of scattering distances must be computed from single realizations of stationary processes. We study the mean-square convergence of $S_J F$ to $\overline{S_J\, F}$ and the resulting properties.

We say that $S_J F$ converges in mean-square to $\overline{S_J\, F}$ if the scattering variance tends to zero:

$$\lim_{J\to\infty} \sum_{p \in P_J} E\{|S_J(p)F - \overline{S(p)F}|^2\} = 0 \ . \tag{95}$$

The following proposition proves that this mean-square convergence is equivalent to a scattering energy conservation.

Proposition 10 *Suppose that the wavelet transform is unitary and wavelets satisfy condition (31). If F is stationary then $S_J F$ converges in mean-square to $\overline{S_J\, F}$ if and only if*

$$\lim_{J\to\infty} \|\overline{S_J\, F}\|^2 = \sum_{p \in P_\infty} |\overline{S(p)F}|^2 = E\{|F|^2\} \ . \tag{96}$$

Proof: Theorem 5 proves that

$$\sum_{p \in P_J} E\{|S_J(p)F|^2\} = E\{|F|^2\} \ .$$

Since

$$E\{|S_J(p)F|^2\} = |\overline{S_J(p)F}|^2 + E\{|S_J(p)F - \overline{S(p)F}|^2\}$$

it results that $\lim_{J \to \infty} \sum_{p \in P_J} E\{|S_J(p)F - \overline{S(p)F}|^2\} = 0$ is equivalent to (96). $\square$

The following proposition shows that mean-square convergence implies a consistent estimation of the scattering metric.

Proposition 11 *If F and G are stationary with finite second order moments and if $S_J F$ and $S_J G$ converge in mean-square to $\overline{S_J\,F}$ and $\overline{S_J\,G}$ then*

$$\forall x \in \mathbb{R}^d \quad , \quad \lim_{J \to \infty} \sum_{p \in P_J} |S_J(p)F(x) - S_J(p)G(x)|^2 = \lim_{J \to \infty} \|\overline{S_J\,F} - \overline{S_J\,G}\|^2 \quad (97)$$

with probability 1.

Proof: The estimation error is

$$X_J = \sum_{p \in P_J} |S_J(p)F(x) - S_J(p)G(x)|^2 - \sum_{p \in P_J} |\overline{S_J(p)F} - \overline{S_J(p)G}|^2 \ .$$

Let $A_J^2 = \sum_{p \in P_J} |S_J(p)F(x) - \overline{S(p)F}|^2$, $B_J^2 = \sum_{p \in P_J} |S_J(p)G(x) - \overline{S(p)G}|^2$, and $C_J^2 = \sum_{p \in P_J} |\overline{S_J(p)F} - \overline{S_J(p)G}|^2$. The mean-square convergence implies that $\lim_{J \to \infty} E\{A_J^2\} = 0$ and $\lim_{J \to \infty} E\{B_J^2\} = 0$ and hence that $\lim_{J \to \infty} A_J = 0$ and $\lim_{J \to \infty} B_J = 0$ with probability 1. A direct computation shows that

$$|X_J| \leq 4\, C_J\, (A_J + B_J) + 2\, A_J^2 + 2\, B_J^2 \ .$$

We saw in (86) that $C_J \leq 2(E\{|F|^2\} + E\{|G|^2\})$ so $\lim_{J \to \infty} X_J = 0$ with probability 1, which proves (97). $\square$

This proposition shows that mean-square convergence is sufficient to estimate the scattering metric from single realizations. Image texture classification algorithms are based on this approach [5]. For Gaussian processes and large classes of stationary textures F having a bounded spectrum, we observe numerically an exponential mean square convergence of

$\sum_{p \in P_J} E\{|S_J(p)F - \overline{S(p)F}|^2\}$ to zero when J increases. This result is not proved but explained through a conjectured property of the variance of $S_J(p)F$.

A scattering S_J reduces the variance of F by removing random phase fluctuations and by averaging with ϕ_J. If X is a complex Gaussian random variable whose real and imaginary parts are independent with same variance then one can verify that the modulus reduces the variance by $1 - \pi/4$:

$$E\{(|X| - E\{|X|\})^2\} = \Big(1 - \frac{\pi}{4}\Big)\, E\{|X - E\{X\}|^2\} \ .$$

Along a path $p = \{j_n, \gamma_n\}_{n \leq |p|}$, the scattering operator iterates $|p| - 1$ times over modulus operators

$$S(p)F = |\cdots|F \star \psi_{j_1,\gamma_1}| \star \psi_{j_2,\gamma_2}| \cdots | \star \psi_{j_{|p|},\gamma_{|p|}}| \ .$$

For Gaussian processes, it is numerically observed that this iterative random phase removal reduces the variance of $F \star \psi_{j_1,\gamma_1}$ by a factor $\alpha^{|p|-1}$ with $\alpha < 1$. The averaging $S_J(p)F = S(p)F \star \phi_J$ further reduces the variance by a factor of the order of $2^{d(J-j_1)}$. For all path p beginning with (j_1, γ_1), numerical experiments indicate that there exists $\alpha < 1$ such that

$$E\{|S_J(p)F - \overline{S_J(p)F}|^2\} \leq E\{|F \star \psi_{j_1,\gamma_1}|^2\}\, \alpha^{|p|-1}\, 2^{d(j_1-J)} \ . \qquad (98)$$

For Gaussian processes, the constant α is numerically close to the variance reduction factor $1 - \pi/4$ produced by a modulus on a complex Gaussian random variable X.

As explained in Section 3.3, the scattering energy is concentrated over a subset $\widetilde{P}_J \subset P_J$ of scale increasing paths $p = \{j_n, \gamma_n\}_{n \leq |p|}$ defined by $j_n < j_{n+1} < J$ for $n < |p|$. Numerical computations in Section 4.2 are thus restricted over these scale increasing paths. The following proposition derives from (98) that the scattering variance converges exponentially to zero over scale increasing paths. The number of wavelets $\{\psi_\gamma\}_{\gamma \in \Gamma}$ is written $|\Gamma|$.

Proposition 12 *Suppose that $|\Gamma| \leq 2^d - 1$. If F satisfies (98) and $\|\hat{R}_F\|_\infty < \infty$ then there exists $C > 0$ such that*

$$\sum_{p \in \widetilde{P}_J} E\{|S_J(p)F - \overline{S_J(p)F}|^2\} \leq C\,(\|\hat{R}_F\|_\infty^2 + R_F(0))\left(\frac{1 + |\Gamma|\,\alpha}{2^d}\right)^J \ . \qquad (99)$$

The proof is in Appendix H. Since $|\Gamma| \leq 2^d - 1$, (99) gives an exponential decay with a rate $2^{-d}(1 + |\Gamma|\,\alpha) < 1$. If $|\Gamma| \geq 2^d$ a similar result can be proved for appropriate wavelets. Numerical experiments over images textures indicate that such exponential mean-square convergence applies to a large class of stationary processes, but there is no mathematical characterization of this class.

6 Group Invariance

Many classification problems require invariance to more than the translation group. A simple example is rotation invariance, which is needed in many image object recognition problems. Local invariance to translations and to the action of a compact Lie group G is obtained by cascading a scattering in $\mathbf{L}^2(\mathbb{R}^d)$ and a scattering in $\mathbf{L}^2(G)$. Section 6.1 begins by constructing scattering operators in $\mathbf{L}^2(G)$, which are invariant to displacements in G. Combined scattering operators in $\mathbf{L}^2(\mathbb{R}^d)$ and $\mathbf{L}^2(G)$ are introduced and studied in Section 6.2.

6.1 Compact Lie Group Scattering

Let G be a compact Lie group of linear invertible operators in $GL(\mathbb{R}^d)$. Let $\mathbf{L}^2(G)$ be the space of measurable functions $h(\gamma)$ such that $\|h\|^2 = \int_G |h(\gamma)|^2\, d\gamma < \infty$, where $d\gamma$ is the Haar measure of G.

An invariant representation to displacements in G is constructed with a scattering operator obtained as a cascade of wavelet transforms in $\mathbf{L}^2(G)$ and modulus operators. The construction of Littlewood-Paley decompositions on compact manifolds and in particular on compact Lie groups was developed by Stein [20]. Different wavelet constructions have been proposed over manifolds [16]. Geller and Pesenson [8] have built unitary wavelet transforms on compact Lie groups, by replacing the Fourier transform by a decomposition over the eigenvectors of the Laplace-Beltrami operator of an invariant metric defined on the group [8]. For any $L \leq 0$, it defines a scaling function and a family of wavelets

$$\{\tilde{\phi}_L \ , \ \tilde{\psi}_l\}_{-\infty < l < L}$$

in $\mathbf{L}^2(G)$ where 2^l is a scale factor. The wavelet transform of $h \in \mathbf{L}^2(G)$ is

$\overline{\overline{\widetilde{W}}}_L h = \{\widetilde{A}_J h \, , \, \widetilde{W}_{l,\nu} h\}_{l<L}$, where

$$\widetilde{W}_l h(\beta) = \int_G h(\gamma) \, \tilde{\psi}_l(\beta \, \gamma^{-1}) \, d\gamma \tag{100}$$

and

$$\widetilde{A}_L h(\beta) = \int_G h(\gamma) \, \tilde{\phi}_L(\beta \, \gamma^{-1}) \, d\gamma \ . \tag{101}$$

If $L = 0$ then $\phi_0(\gamma) = (\int_G \ d\gamma)^{-1} = |G|^{-1}$ is constant so

$$\widetilde{A}_0 h(\beta) = |G|^{-1} \int_G h(\gamma) \, d\gamma = cst. \tag{102}$$

Wavelets are constructed so that the wavelet transform satisfies frame bounds inequalities [8]

$$(1-\delta) \, \|f\|^2 \leq \|\overline{\overline{\widetilde{W}}}_L f\|^2 \leq \|f\|^2$$

with

$$\|\overline{\overline{\widetilde{W}}}_L h\|^2 = \|\widetilde{A}_L h\|^2 + \sum_{l<L} \|\widetilde{W}_l h\|^2 \ .$$

The wavelet transform may also be unitary [8] and hence $\delta = 0$.

The periodic wavelet transform introduced in Section 4.1 is a simple example of Littlewood-Paley decompositions on the Abelian group of rotations in $\mathbb{R}^2$. The rotation matrix γ is identified to an angle in $[0, 2\pi)$, and one can write $\beta \, \gamma^{-1} = \beta - \gamma$ because the group operation commutes. The scaling function and wavelets $\{\tilde{\phi}_L \ , \ \tilde{\psi}_l\}_{-\infty<l<L}$ are then 2π periodic functions of the angle γ, and are obtained in (71) as periodization of scaling functions and wavelets in $\mathbf{L}^2(\mathbb{R})$. The resulting periodic wavelet transform is written as circular convolutions in (70).

The wavelet transform modulus operator is defined by

$$\overline{\overline{\widetilde{U}}}_L = \{\widetilde{A}_L f \, , \, \widetilde{U}_l f = |\widetilde{W}_l f|\}_{l<L} \ .$$

Since $\overline{\overline{\widetilde{W}}}_L$ is contractive, $\overline{\overline{\widetilde{U}}}_L$ is also contractive and if $\overline{\overline{\widetilde{W}}}_L$ is unitary then $\overline{\overline{\widetilde{U}}}_L$ preserves the norm in $\mathbf{L}^2(G)$. The wavelet transform is a convolution over the group which commutes with left displacement operators: $\widetilde{T}_\beta h(\gamma) = h(\beta\gamma)$. Indeed, a change of variable in (100) and (101) proves that that $\widetilde{W}_l \widetilde{T}_\beta h = \widetilde{T}_\beta \widetilde{W}_l h$ and $\widetilde{A}_L \widetilde{T}_\beta h = \widetilde{T}_\beta \widetilde{A}_L h$, and hence $\overline{\overline{\widetilde{W}}}_L \widetilde{T}_\beta = \widetilde{T}_\beta \overline{\overline{\widetilde{W}}}_L$. It results that $\overline{\overline{\widetilde{U}}}_L$ also commutes with $\widetilde{T}_\beta$.

A scattering operator on $\mathbf{L}^2(G)$ iterates over $\widetilde{\overline{U}}_L$. As in Definition 1, an averaged scattering at the scale 2^L is defined over the set $\widetilde{P}_L$ of paths $\tilde{p} = \{l_n\}_{1\leq n\leq |p|}$ for which $\max_n l_n < L$. It is given by

$$\widetilde{S}_L(\tilde{p})h = \widetilde{A}_L \prod_{n=1}^{|\tilde{p}|} \widetilde{U}_{l_n} h \ , \tag{103}$$

with $\widetilde{S}_L(0)f = \widetilde{A}_L h$. The scattering norm for any $(h_1, h_2) \in \mathbf{L}^2(G)^2$ is

$$\|\widetilde{S}_L h_1 - \widetilde{S}_L h_2\|^2 = \sum_{\tilde{p}\in\widetilde{P}_L} \|\widetilde{S}_L(\tilde{p})h_1 - \widetilde{S}_L(\tilde{p})h_2\|^2$$

with $\|\widetilde{S}_L(\tilde{p})h\|^2 = \int_G |\widetilde{S}_L(\tilde{p})h(\gamma)|^2\, d\gamma$.

The operator $\widetilde{S}_L$ is obtained by iterating on $\widetilde{\overline{U}}_L$ and similarly to Proposition 6 we derive that it is contractive:

$$\forall (f,g) \in \mathbf{L}^2(G)^2 \quad , \quad \|\widetilde{S}_L f - \widetilde{S}_L g\| \leq \|f - g\| \ .$$

Since $\widetilde{\overline{U}}_L$ commutes with displacement operators $\widetilde{T}_\beta$ in $\mathbf{L}^2(G)$, $\widetilde{S}_L$ also commutes with $\widetilde{T}_\beta$.

When $L = 0$, $\widetilde{A}_0$ is the integration operator (102) so $\widetilde{S}_0(\tilde{p})h(\gamma)$ does not depend upon γ for all $\tilde{p}$:

$$\widetilde{S}_0(\tilde{p})h(\gamma) = |G|^{-1} \int_G \prod_{n=1}^{|\tilde{p}|} \widetilde{U}_{l_n} h(\beta)\, d\beta \ . \tag{104}$$

Together with the commutation property it proves that

$$\widetilde{S}_0(\tilde{p})\, \widetilde{T}_\beta = \widetilde{T}_\beta\, \widetilde{S}_0(\tilde{p}) = \widetilde{S}_0(\tilde{p}) \ . \tag{105}$$

The scattering operator $\widetilde{S}_0$ is therefore invariant to displacements in $\mathbf{L}^2(G)$: $\widetilde{S}_0\, \widetilde{T}_\beta = \widetilde{S}_0$.

6.2 Combined Invariant Scattering

A scattering operator which is locally invariant to the action of a compact Lie group G and to translations is constructed with two successive scattering in space and along G. The resulting combined scattering is proved to be locally invariant to translations and to the action of G, up to second order terms.

A compact Lie subgroup G of $GL(\mathbb{R}^d)$ is a compact differentiable manifold in $\mathbb{R}^{d^2}$. If $\gamma \in G$ then $\gamma^n \in G$ for all $n \in \mathbb{Z}$, so necessarily $|\det \gamma| = 1$ because the manifold is compact. A spatial wavelet transform in $\mathbf{L}^2(\mathbb{R}^d)$ is defined on the translation group with mother wavelets obtained from the orbit in G of a single wavelet $\psi \in \mathbf{L}^2(\mathbb{R}^d)$:

$$\psi_\gamma(x) = \psi(\gamma^{-1}x) \ .$$

For a rotation group, ψ has typically a narrow directional selectivity along an orientation which is modified by the rotation operator γ. The first dyadic wavelet transform is $\overline{W}_J f = \{A_J f, W_{j,\gamma} f\}_{j<J}$ where

$$W_{j,\gamma} f(x) = f \star \psi_{j,\gamma}(x) \ \text{ with } \ \psi_{j,\gamma} = 2^{-jd}\psi_\gamma(2^{-j}x) \ ,$$

and $A_J f = f \star \phi_J$ with $\phi_J = 2^{-Jd}\phi_J(2^{-J}x)$. It's norm is

$$\|\overline{W}_J f\|^2 = \|A_J f\|^2 + \sum_{j=-\infty}^{J-1} \int_G \|W_{j,\gamma} f\|^2 \, d\gamma \ , \tag{106}$$

where $d\gamma$ is the Haar measure of the compact group G. The Fourier transform in x of $W_{j,\gamma}f(x)$ is $\widehat{W_{j,\gamma}f}(\omega) = \hat{f}(\omega)\hat{\psi}_\gamma(2^j\omega)$ with $\hat{\psi}_\gamma(\omega) = \hat{\psi}(\gamma^*\omega)$, where γ^* is the adjoint of γ. Similarly to Proposition 1 we verify that if $\hat{\psi}$ and $\hat{\phi}$ satisfy for all $\omega \in \mathbb{R}^d$

$$(1-\delta) \ \le |\hat{\phi}(\omega)|^2 + \sum_{j=-\infty}^{-1} \int_G \frac{1}{2}\Big(|\hat{\psi}(2^j\gamma^*\omega)|^2 + |\hat{\psi}(-2^j\gamma^*\omega)|^2\Big)\, d\gamma \le 1 \tag{107}$$

then for any $f \in \mathbf{L}^2(\mathbb{R}^d)$ with $f(x) \in \mathbb{R}$

$$(1-\delta)\|f\|^2 \le \|\overline{W}_J f\|^2 \le \|f\|^2 \ . \tag{108}$$

The wavelet transform is unitary if $\delta = 0$ in (107) and hence

$$|\hat{\phi}(\omega)|^2 = 1 - \sum_{j=-\infty}^{-1} \int_G \frac{1}{2}\Big(|\hat{\psi}(2^j\gamma^*\omega)|^2 + |\hat{\psi}(-2^j\gamma^*\omega)|^2\Big)\, d\gamma \ . \tag{109}$$

It results that $|\hat{\phi}(\gamma^*\omega)| = |\hat{\phi}(\omega)|$ for all $\gamma \in G$. In the following, we choose the phase of $\hat{\phi}(\omega)$ so that $\hat{\phi}(\gamma^*\omega) = \hat{\phi}(\gamma^*\omega)$ and hence $\phi(\gamma^{-1}x) = \phi(x)$ for all

$\gamma \in G$. It results that ϕ is invariant to the action of G. We shall suppose that this property is valid even if the wavelet transform is not unitary.

The wavelet modulus propagator is $\overline{U}_J f = \{A_J f \, , \, U_{j,\gamma} f = |W_{j,\gamma} f|\}_{j \in \mathbb{Z}, \gamma \in G}$. Following Definition 1, a scattering operator along a path $p = \{j_n, \gamma_n\}_{1 \le n \le |p|}$ with $\max_n j_n < J$ and $\gamma_n \in \Gamma$ is defined by

$$S_J(p) f = A_J \, S(p) f \quad \text{with} \quad S(p) f = \prod_{n=1}^{|p|} U_{j_n, \gamma_n} f \ , \tag{110}$$

and $S_J(0) f = A_J f$. The scattering metric makes a sum over all paths. In this case, there is a continuum of path variables $p = \{j_n, \gamma_n\}_{1 \le n \le |p|}$ that must be integrated according to the Haar metric in G as in the wavelet transform norm (106). The scattering norm is

$$\begin{aligned} \|S_J f\|^2 &= \int_{P_J} \|S_J(p) f\|^2 dp \\ &= \sum_{m=0}^{\infty} \sum_{j_1, \dots, j_m} \int_{G^m} \int_{\mathbb{R}^d} |S_J(\{j_n, \gamma_n\}_{n \le m}) f(x)|^2 d\gamma_1 ... d\gamma_m \, dx \ . \end{aligned}$$

As in Proposition 6 we verify that S_J is a contracting operator because the wavelet modulus propagator $\overline{U}_J$ is contracting.

The action of $\beta \in G$ over a path $p = \{j_n, \gamma_n\}_{1 \le n \le |p|}$ is defined by

$$\beta p = \{j_n \, , \, \beta \gamma_n\}_{1 \le n \le |p|} \ .$$

The action of $\beta \in G$ over $f \in \mathbf{L}^2(\mathbb{R}^d)$ is given by $L_\beta f(x) = f(\beta x)$. The following proposition shows that commuting L_β with a scattering operator produces a path displacement.

Proposition 13 *For any $\beta \in G$ and any $p \in P_J$*

$$S_J(p) L_\beta f = L_\beta S_J(\beta p) f \ .$$

Proof: A change of variable in the wavelet transform integral proves that

$$W_{j,\gamma} L_\beta f(x) = L_\beta W_{j,\beta\gamma} f(x) \quad \text{and hence} \quad U_{j,\gamma} L_\beta f(x) = L_\beta U_{j,\beta\gamma} f(x) \ . \tag{111}$$

Cascading this result in (110) implies that

$$S(p) L_\beta f = L_\beta S(\beta p) f \ .$$

Since $\phi(\beta x) = \phi(x)$ we also verify that $A_J L_\beta f(x) = L_\beta A_J f(x)$, and hence that $S_J(p) L_\beta f = L_\beta S_J(\beta p) f$. $\Box$

To become insensitive to displacement of β in G we apply the scattering operator $\widetilde{S}_L$ defined in (103). For any $\tilde{p} \in \widetilde{P}_L$, $\widetilde{S}_L$ maps any $h(\gamma) \in \mathbf{L}^2(G)$ into $\widetilde{S}_L(\tilde{p}) h(\gamma)$. When $L = 0$, (104) proves that $\widetilde{S}_0(\tilde{p}) h(\gamma)$ is a constant which does not depend upon γ so $\widetilde{S}_0(\tilde{p}) h(\beta\gamma) = \widetilde{S}_0(\tilde{p}) h(\gamma)$.

Any $p = \{j_n\,,\,\gamma_n\}_{n\leq|p|} \in P_J$ can uniquely be written $p = \gamma_1\,\bar{p}$ where $\bar{p} = \{j_n\,,\,\gamma_1^{-1}\gamma_n\}_{n\leq|\bar{p}|}$ is a normalized path whose first index is $\gamma_1^{-1}\gamma_1 = \mathbf{1}$. The scattering can be written as a function of γ for x and $\bar{p}$ fixed:

$$S_J(p) f(x) = S_J(\gamma_1 \bar{p}) f(x) = h_{x,\bar{p}}(\gamma_1)\ .$$

For each x and $\bar{p}$, the operator $\widetilde{S}_L$ computes $\tilde{S}_L(\tilde{p}) h_{x,\bar{p}}(\gamma_1)$. The output can be indexed with the original path variable $p = \gamma_1\,\bar{p}$ that specifies both γ_1 and $\bar{p}$. The resulting combined scattering operator is thus defined by

$$\tilde{S}_L S_J(\tilde{p}, p) f(x) = \tilde{S}_L(\tilde{p}) h_{x,\bar{p}}(\gamma_1) \tag{112}$$

where $\bar{p} = \gamma_1^{-1} p = \{j_n\,,\,\gamma_1^{-1}\gamma_n\}_{n\leq|\bar{p}|}$ and $h_{x,\bar{p}}(\gamma_1) = S_J(p) f(x)$. The combined scattering operator $\tilde{S}_L S_J$ thus applies the scattering operator $\widetilde{S}_L$ within each layer of scattering signals $S_J(p) f$ of fixed length $|p|$, with a cascade of wavelet modulus transforms along γ_1 for x and $\bar{p}$ fixed.

A combined scattering metric is defined by summing the $\mathbf{L}^2(\mathbb{R}^d)$ norm along x over all paths p and $\tilde{p}$:

$$\|\tilde{S}_L S_J f - \tilde{S}_L S_J g\|^2 = \sum_{\tilde{p}\in\widetilde{P}_L} \int_{P_J} \|\tilde{S}_L S_J(\tilde{p}, p) f - \tilde{S}_L S_J(\tilde{p}, p) g\|^2\, dp\ .$$

Since a combined scattering operator is obtained by cascading two contractive operators $\tilde{S}_L$ and S_J, it is also contractive:

$$\forall (f, g) \in \mathbf{L}^2(\mathbb{R}^d)^2 \quad , \quad \|\tilde{S}_L S_J f - \tilde{S}_L S_J g\| \leq \|f - g\|\ .$$

We set $L = 0$ to obtain a displacement invariant scattering operator in $\mathbf{L}^2(G)$.

Let $\mathcal{G}$ be the Lie Algebra of G, which is a linear subspace of $GL(\mathbb{R}^d)$. For any deformation $D_\tau f(x) = f(x - \tau(x))$, the deformation tensor $\nabla\tau$ is decomposed as a sum of its projections $\overline{\nabla\tau}$ and $\overline{\nabla\tau}^\perp$ respectively on $\mathcal{G}$ and on the orthogonal complement of $\mathcal{G}$ in $GL(\mathbb{R}^d)$:

$$\nabla\tau(x) = \overline{\nabla\tau}(x) + \overline{\nabla\tau}^\perp(x)\ .$$

The following theorem proves that the combined scattering metric is locally invariant to the component of $\nabla\tau$ which belongs to G, with a first order term proportional to $|\overline{\nabla\tau}^{\perp}|_\infty = \sup_x |\overline{\nabla\tau}^{\perp}(x)|$.

Theorem 7 *If $|\nabla\tau|_\infty < 1-\epsilon$ with $\epsilon > 0$ then there exists C such that for all $f \in \mathbf{L}^2(\mathbb{R}^d)$*

$$\|\tilde{S}_0 S_J D_\tau f - \tilde{S}_0 S_J f\| \leq \tag{113}$$
$$C\,\|Sf\|_{1,P_J}\left(2^{-J}|\tau|_\infty + (|\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|^2_\infty)\,(\log\frac{|\tau|_\infty}{|\nabla\tau|_\infty}\vee 1) + |H\tau|_\infty\right)$$

with $\|Sf\|_{1,P_J} = \sum_{n=0}^{+\infty}\left(\sum_{p\in P_J,|p|=n}\|S(p)f\|^2\right)^{1/2}$.

Proof: The proof shows that first order terms $\|\tilde{S}_0 S_J D_\tau f - \tilde{S}_0 S_J f\|$ do not depend upon the component of $\nabla\tau$ in G. It uses the invariance of $\tilde{S}_0$ to displacements in G in order to locally compensate for the component of $\nabla\tau(x)$ in G. This local inverse displacement is obtained by mapping $\overline{\nabla\tau}(x) \in \mathcal{G}$ into G with the exponential map:

$$\beta(x) = \exp(\overline{\nabla\tau}(x)) \in G\ . \tag{114}$$

We decompose

$$\|\tilde{S}_0 S_J D_\tau f - \tilde{S}_0 S_J f\| \leq \|D_\tau \tilde{S}_0 S_J f - \tilde{S}_0 S_J f\| + \|[\tilde{S}_0 S_J\,,\,D_\tau]f\|\ . \tag{115}$$

Since $\tilde{S}_0$ is a contractive operator, we verify as in (52) that $\|D_\tau \tilde{S}_0 S_J f - \tilde{S}_0 S_J f\| \leq \|D_\tau A_J - A_J\|\,\|Sf\|_{1,P_J}$. Applying Lemma 1 to A_J implies that

$$\|D_\tau \tilde{S}_0 S_J f - \tilde{S}_0 S_J f\| \leq C\,2^{-J}\,|\tau|_\infty\,\|Sf\|_{1,P_J}\ . \tag{116}$$

The commutator of (115) is computed by compensating $\nabla\tau(x)$ with $\beta(x) \in G$. We write $\overline{U}_{J,\beta} = \{A_J f\,,\,|W_{j,\beta\gamma}f|\}_{j,\gamma}$ a propagator displaced by $\beta(x)$. The following lemma is an extension of Lemma 4, which incorporates this compensation.

Lemma 7 *For any operator O over $\mathbf{L}^2(\mathbb{R}^d)$ and any $\beta(x) \in G$*

$$\|[\tilde{S}_0 S_J\,,\,O]f\| \leq \|Sf\|_{1,P_J}\,\|\overline{U}_J O - O\overline{U}_{J,\beta}\|\ . \tag{117}$$

To compute $[\tilde{S}_0 S_J\,,\,O] = \tilde{S}_0 S_J O - O\tilde{S}_0 S_J$, we use the path displacement invariance

$$\tilde{S}_0 S_J(\tilde{p},p) O = \tilde{S}_0(\tilde{p}) S_J(p) O = \tilde{S}_0(\tilde{p}) S_J(\beta p) O\ ,$$

that results from the invariance of $\tilde{S}_0(\tilde{p})$ to the action of G. Moreover, $O\tilde{S}_0 S_J = \tilde{S}_0 O S_J$ because $\tilde{S}_0$ does not act along the variable x, so

$$\|[\tilde{S}_0 S_J\,,\,O]f\|^2 = \int_{P_J} \sum_{\tilde{p}\in\widetilde{P}_0} \|\tilde{S}_0(\tilde{p})(S_J(p)Of - OS_J(\beta p)f)\|^2\,dp\ .$$

Since $\tilde{S}_0$ is contractive

$$\|[\tilde{S}_0 S_J\,,\,O]f\|^2 \leq \int_{p\in P_J} \|S_J(p)Of - OS_J(\beta p)f\|^2\,dp \tag{118}$$

The scattering $\{S_J(\beta p)f\}_{p\in P_J}$ displaced by β is computed from displaced propagators $\overline{U}_{J,\beta}$ with the same cascade which computes $\{S_J(p)f\}_{p\in P_J}$ from $\overline{U}_J$. Replacing the commutators $[\overline{U}_J, O]$ by $\overline{U}_J O - O\overline{U}_{J,\beta}$ in Appendix E proves that

$$\int_{p\in P_J} \|S_J(p)Of - OS_J(\beta p)f\|^2\,dp \leq \|Sf\|^2_{1,P_J}\,\|\overline{U}_J O - O\overline{U}_{J,\beta}\|^2\,.$$

Inserting (118) proves the result (117) of Lemma 7.

If $O = D_\tau$, since the modulus commutes with D_τ and is contractive we verify, as in (54) that

$$\|\overline{U}_J D_\tau - D_\tau \overline{U}_{J,\beta}\| \leq \|\overline{W}_J D_\tau - D_\tau \overline{W}_{J,\beta}\| \tag{119}$$

with $\overline{W}_{J,\beta} = \{A_J f\,,\,W_{j,\beta\gamma}f\}_{j,\gamma}$ so it results from (117) that

$$\|[\tilde{S}_0 S_J\,,\,O]f\| \leq \|Sf\|_{1,P_J}\,\|\overline{W}_J O - O\overline{W}_{J,\beta}\|\ .$$

Inserting this bound and (116) in (115) gives

$$\|S_J D_\tau f - S_J f\| \leq C\,\|Sf\|_{1,P_J}\Big(2^{-J}\,|\tau|_\infty + \|\overline{W}_J D_\tau - D_\tau \overline{W}_{J,\beta}\|\Big)\ . \tag{120}$$

The following lemma computes an upper bound of the wavelet commutator compensated by $\beta(x) = \exp(\overline{\nabla}\tau(x))$.

Lemma 8 *If $|\nabla\tau|_\infty \leq 1-\epsilon$ with $\epsilon > 0$ then there exists $C > 0$ such that for all $J \in \mathbb{Z}$*

$$\|\overline{W}_J D_\tau - D_\tau \overline{W}_{J,\beta}\| \leq C\,\Big((|\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|^2_\infty)\,(\log\frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty\Big)\ . \tag{121}$$

The lemma proof is in Appendix I. Inserting (121) in (120) proves the theorem inequality (113). $\square$

Theorem 7 proves that a combined scattering yields a metric whose first order term does not depend on the component of the deformation $\nabla\tau$ which belongs to G. It is locally invariant to the action of G up to the second order terms $|\nabla\tau|^2_\infty$ and $|H\tau|_\infty$. First order terms are proportional the projection $\overline{\nabla\tau}^\perp$ of $\nabla\tau$ on the orthogonal of the Lie Algebra $\mathcal{G}$. If G is the rotation group then $\mathcal{G}$ is the subspace of $GL(\mathbb{R}^d)$ of skew-hermitian matrices so

$$\overline{\nabla\tau}^\perp = \frac{\nabla\tau + \nabla\tau^*}{2} ~.$$

The resulting combined scattering metric has first order term which depend on the symmetric part of $\nabla\tau$.

Let $S_{J,m}$ be the restriction of S_J to paths of length $|p| < m$. At a sufficient large scale 2^J, the following corollary derives a Lipschitz type regularity property for a combined scattering restricted to $|p| < m$.

Corollary 2 *If $|\nabla\tau|_\infty \leq 1 - \epsilon$ with $\epsilon > 0$ there exists $C > 0$ such that if $J \geq \log \frac{|\tau|_\infty}{|\overline{\nabla\tau}^\perp|_\infty}$ then*

$$\|\tilde S_0 S_{J,m} D_\tau f - \tilde S_0 S_{J,m} f\| \leq \tag{122}$$
$$C\, m\, \|f\| \Big((|\overline{\nabla\tau}^\perp|_\infty + |\nabla\tau|^2_\infty)\, (\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty \Big).$$

Proof: Similarly to Theorem 3, the proof of Theorem 7 shows that if we restrict the paths length $|p|$ to $m-1$ then $\|Sf\|_{1,P_J}$ in (113) is replaced by $\sum_{n=0}^{m-1} \Big(\sum_{p \in P_J, |p|=n} \|S(p)f\|^2\Big)^{1/2}$. We saw in (63) that

$$\sum_{n=0}^{m-1} \left(\sum_{p\in P_J, |p|=n} \|S(p) f\|^2 \right)^{1/2} \leq m\, \|f\| ~,$$

and since $2^{-J}|\tau|_\infty \leq |\overline{\nabla\tau}^\perp|_\infty$, we derive (122) from (113). $\square$

This corrolary proves that a combined scattering is locally invariant to deformations in G and almost Lipschitz continuous relatively to the deformation component which is orthogonal to $\mathcal{G}$, up to second order terms.

A Proof of Lemma 1

Lemma 1 as well all other upper bounds on operator norms are computed with the Schur lemma. For any operator $Kf(x) = \int f(u)\, k(x,u)\, du$, Schur lemma proves that

$$\int |k(x,u)|\, dx \leq A \;\text{ and }\; \int |k(x,u)|\, du \leq A \implies \|K\| \leq A\ , \qquad (123)$$

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

The operator norm of $K_j = D_\tau Z_j - Z_j$ is computed by applying Schur lemma on its kernel

$$k_j(x,u) = h_j(x - \tau(x) - u) - h_j(x-u)\ . \qquad (124)$$

A first order Taylor expansion proves that

$$|k_j(x,u)| \leq |\int_0^1 \nabla h_j(x{-}u{-}t\,\tau(x))\,.\,\tau(x)\,dt| \leq |\tau|_\infty \int_0^1 |\nabla h_j(x{-}u{-}t\,\tau(x))|\,dt$$

so

$$\int |k_j(x,u)|\,du \leq |\tau|_\infty \int_0^1 \int |\nabla h_j(x-u-t\,\tau(x))|\,du\,dt\ . \qquad (125)$$

Since $\nabla h_j(x) = 2^{-dj-j} \nabla h(2^{-j}x)$, it results that

$$\int |k_j(x,u)|\,du \leq |\tau|_\infty\, 2^{-dj-j} \int |\nabla h(2^{-j}u')|\,du' = 2^{-j}\,|\tau|_\infty\,\|\nabla h\|_1\ . \qquad (126)$$

Similarly to (125) we prove that

$$\int |k_j(x,u)|\,dx \leq |\tau|_\infty \int_0^1 \int |\nabla h_j(x-u-t\,\tau(x))|\,dx\,dt\ .$$

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

$$\begin{aligned}\int |k_j(x,u)|\,dx &\leq |\tau|_\infty\,\epsilon^{-d} \int_0^1 \int |\nabla h_j(v-u)|\,dv\,dt \\ &= 2^{-j}\,|\tau|_\infty\,\|\nabla h\|_1\,\epsilon^{-d}\,.\end{aligned}$$

Schur lemma (123) applied to this upper bound and (126) proves the lemma result:

$$\|D_\tau Z_j - Z_j\| \leq \epsilon^{-d}\,\|\nabla h\|_1\; 2^{-j}\,|\tau|_\infty\ .$$

B Proof of (67)

We prove that

$$\|D_\tau A_j f - A_j f + \tau \,.\, \nabla A_j f\| \le C\, \|f\|\, 2^{-2j}\ \ |\tau|_\infty^2 \tag{127}$$

by applying Schur lemma (123) on the kernel of $K_j = D_\tau A_j - A_j + \tau \cdot \nabla A_j$:

$$k_j(x,u) = \phi_j(x - \tau(x) - u) - \phi_j(x-u) + \nabla\phi_j(x-u)\cdot\tau(x)\ .$$

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

$$\begin{aligned} |k_j(x,u)| &= \left|\int_0^1 t\tau(x)\cdot H\phi_j(u - x - (1-t)\tau(u))\cdot\tau(x)\,dt\right| \\ &\le |\tau|_\infty^2 \int_0^1 |t||H\phi_j(u - x - (1-t)\,\tau(x))|\,dt\ . \end{aligned} \tag{128}$$

Since $\phi_j(x) = 2^{-dj}\phi(2^{-j}x)$, $H\phi_j(x) = 2^{-jd-2j}H\phi(2^{-j}x)$. With a change of variable, (128) gives

$$\int |k_j(x,u)|\,du \le |\tau|_\infty^2\, 2^{-dj-2j}\int |H\phi(2^{-j}u')|\,du' = 2^{-2j}\,|\tau|_\infty^2\,\|H\phi\|_1\ , \tag{129}$$

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 (128) also implies that

$$\int |k_j(x,u)|\,dx \le |\tau|_\infty^2 \int_0^1 |t| \int |H\phi_j(u - x - (1-t)\tau(x))|\,du\,dt\ .$$

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

$$\begin{aligned} \int |k_j(x,u)|\,dx &\le |\tau|_\infty^2\,(1 - |\nabla\tau|_\infty)^{-d}\int_0^1\int |H\phi_j(v-u)|\,dv\,dt \\ &= 2^{-2j}\,|\tau|_\infty^2\,\|H\phi\|_1\,\epsilon^{-d}\,. \end{aligned} \tag{130}$$

The upper bounds (129) and (130) with Schur lemma (123) proves (127).

C Proof of Lemma 5

This section computes an upper bound of $\|[\overline{W}_J, D_\tau]\|$ by considering

$$[\overline{W}_J, D_\tau]^* \, [\overline{W}_J, D_\tau] = \sum_{\gamma\in\Gamma} \sum_{j=-\infty}^{J-1} [W_{j,\gamma}, D_\tau]^* [W_{j,\gamma}, D_\tau] + [A_J, D_\tau]^* \, [A_J, D_\tau] \ .$$

Since $\|[\overline{W}_J, D_\tau]\| = \|[\overline{W}_J, D_\tau]^* \, [\overline{W}_J, D_\tau]\|^{1/2}$,

$$\|[\overline{W}_J, D_\tau]\| \leq \sum_{\gamma\in\Gamma} \Big\| \sum_{j=-\infty}^{J-1} [W_{j,\gamma}, D_\tau]^* [W_{j,\gamma}, D_\tau] \Big\|^{1/2} + \|[A_J, D_\tau]^* \, [A_J, D_\tau]\|^{1/2} \ . \tag{131}$$

To prove the upper bound (60) of Lemma 5, we compute an upper bound for each term on the right, which is done by the following lemma.

Lemma 9 *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 $h_j(x) = 2^{-dj} h(2^{-j}x)$. If $|\nabla\tau|_\infty < 1-\epsilon$ with $\epsilon > 0$ then there exists $C > 0$ such that*

$$\|[Z_j, D_\tau]\| \leq C \, |\nabla\tau|_\infty \tag{132}$$

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

$$\Big\| \sum_{j=-\infty}^{+\infty} [Z_j, D_\tau]^* [Z_j, D_\tau] \Big\|^{1/2} \leq C \left(\max\!\Big(\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \, , \, 1\Big) \, |\nabla\tau|_\infty + |H\tau|_\infty \right) \ . \tag{133}$$

Applying (132) to $h = \psi_\gamma$ proves (13). The inequality (133) clearly remains valid if the summation is limited to a finite J instead of $+\infty$ since $[Z_j, D_\tau]^*[Z_j, D_\tau]$ is a positive operator. Inserting in (131) both (132) with $h = \phi$ and (133) with $h = \psi_\gamma$, and replacing $+\infty$ by J finite, proves the upper bound (60) of Lemma 5.

To prove Lemma 9, we factorize

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

Observe that

$$\|[Z_j, D_\tau]^* [Z_j, D_\tau]\|^{1/2} = \|D_\tau^* K_j^* K_j D_\tau\|^{1/2} \leq \|D_\tau\| \, \|K_j^* \, K_j\|^{1/2}, \tag{134}$$

and that

$$\Big\| \sum_{j=-\infty}^{+\infty} [Z_j, D_\tau]^* [Z_j, D_\tau] \Big\|^{1/2} \leq \|D_\tau\| \, \| \sum_{j=-\infty}^{+\infty} K_j^* \, K_j \|^{1/2} \ , \tag{135}$$

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

$$k_j(x,u) = h_j(x-u) - h_j(x - \tau(x) - u + \tau(u)) \, \det(\mathbf{1} - \nabla\tau(u)) \ . \tag{136}$$

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

$$\| \sum_{j=-\infty}^{+\infty} K_j^* \, K_j \|^{1/2} \leq \| \sum_{j=-\infty}^{0} K_j^* \, K_j \|^{1/2} + \| \sum_{j=1}^{\eta} K_j^* \, K_j \|^{1/2} + \| \sum_{j=\eta+1}^{+\infty} K_j^* \, K_j \|^{1/2} \ , \tag{137}$$

and we shall first prove that

$$\| \sum_{j=\eta}^{+\infty} K_j^* K_j \|^{1/2} \leq C \left(|\nabla\tau|_\infty + 2^{-\eta} \, |\tau|_\infty + 2^{-\eta/2} \, |\tau|_\infty^{1/2} \, |\nabla\tau|_\infty^{1/2} \right) \ . \tag{138}$$

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

$$\| \sum_{j=1}^{\eta} K_j^* \, K_j \|^{1/2} \leq \eta \, \|K_j\| \leq C \, \eta \, |\nabla\tau|_\infty \ . \tag{139}$$

The last term carries the singular part and we prove that

$$\| \sum_{j=-\infty}^{0} K_j^* \, K_j \|^{1/2} \leq C \, (|\nabla\tau|_\infty + |H\tau|_\infty) \ . \tag{140}$$

Choosing $\eta = \max(\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty}, 1)$ yields

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

Inserting this result in (135) will prove the second lemma result (133) . In the proof, C is a generic constant which depends only on h and ϵ but which evolves along the calculations. However, slight modifications of the proof shows that the dependancy on ϵ can be removed by multiplying the upper bounds $(1-|\nabla\tau|_\infty)^{-3d}$, and it is then sufficient to impose that $|\nabla\tau|<1$.

The proof of (138) is done by decomposing $K_j = \tilde K_{j,1} + \tilde K_{j,2}$, with a first kernel

$$\tilde k_{j,1}(x,u) = a(u)\, h_j(x-u) \quad\text{with}\quad a(u) = (1-\det(\mathbf{1}-\nabla\tau(u))) \tag{141}$$

and a second kernel

$$\tilde k_{j,2}(x,u) = \det(\mathbf{1}-\nabla\tau(u))\Big(h_j(x-u) - h_j(x-\tau(x)-u+\tau(u))\Big)\ . \tag{142}$$

This kernel has the same form as the kernel (124) in Appendix A and the same proof shows that

$$\|\tilde K_{j,2}\| \leq C\, 2^{-j}\, |\tau|_\infty\ . \tag{143}$$

Taking advantage of this decay, to prove (138), we decompose

$$\|\sum_{j=\eta}^{+\infty} K_j^* K_j\|^{1/2} \leq \|\sum_{j=\eta}^{+\infty} \tilde K_{j,1}^* \tilde K_{j,1}\|^{1/2} + \sum_{j=\eta}^{+\infty}(2^{1/2}\,\|\tilde K_{j,2}\| + \|\tilde K_{j,2}\|^{1/2}\|\tilde K_{j,1}\|^{1/2}) \tag{144}$$

and verify that

$$\|\tilde K_{j,1}\| \leq C\,|\nabla\tau|_\infty \quad\text{and}\quad \|\sum_{j=0}^{+\infty} \tilde K_{j,1}^*\tilde K_{j,1}\|^{1/2} \leq C\,|\nabla\tau|_\infty\ . \tag{145}$$

The kernel of the self-adjoint operator $\tilde K_{j,1}^*\tilde K_{j,1}$ is:

$$\tilde k_j(y,z) = \int \tilde k_{j,1}^*(x,y)\,\tilde k_{j,1}(x,z)\,dx = a(y)\,a(z)\,\tilde h_j\star h_j(z-y)$$

with $\tilde h_j(u) = h_j^*(-u)$. It results that the kernel of $\tilde K = \sum_{j\geq 0}\tilde K_{j,1}^*\tilde K_{j,1}$ is:

$$\tilde k(y,z) = \sum_{j\geq 0}\tilde k_j(y,z) = a(y)\,a(z)\,\theta(z-y) \quad\text{with}\quad \theta(x) = \sum_{j\geq 0}\tilde h_j\star h_j(x)\ .$$

Applying Cauchy-Schwartz on $\|\tilde K f\|$ shows that

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

Since $\hat\theta(\omega) = \sum_{j\geq 0}|h(2^j\omega)|^2$ and $\hat h(0) = \int h(x)\,dx = 0$ and h is both regular with a polynomial decay, we verify that $\|\theta\| < \infty$. Moreover, since $(1-$

$\det(\mathbf{1}-\nabla\tau(u))) \geq (1-|\nabla\tau|_\infty)^d$ we have $\sup_u |a(u)| \leq d\,|\nabla\tau|_\infty$ which proves that $\|\tilde{K}\|^{1/2} \leq C\,|\nabla\tau|_\infty$. Since $\|\tilde{K}_{j,1}\|^2 \leq \|\tilde{K}\|$ we get the same inequality for $\|\tilde{K}_{j,1}\|^2$, which proves the two upper bounds of (145).

The first sum $\sum_{j=-\infty}^{0} K_j^* K_j$ 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

$$k_{j,1}(x,u) = h_j(x-u) - h_j((\mathbf{1}-\nabla\tau(u))(x-u))\,\det(\mathbf{1}-\nabla\tau(u)) \tag{146}$$

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

$$k_{j,2}(x,u) = \det(\mathbf{1}-\nabla\tau(u))\Big(h_j((\mathbf{1}-\nabla\tau(u))(x-u)) - h_j(x-\tau(x)-u+\tau(u))\Big)\,. \tag{147}$$

The sum $\sum_{j\leq 0} K_{j,1}^* K_{j,1}$ is a singular operator whose norm is evaluated separately in the upper bound

$$\|\sum_{j=-\infty}^{0} K_j^* K_j\|^{1/2} \leq \|\sum_{j=-\infty}^{0} K_{j,1}^* K_{j,1}\|^{1/2} + \sum_{j=-\infty}^{0} (\|K_{j,2}\| + 2^{1/2}\|K_{j,2}\|^{1/2}\|K_{j,1}\|^{1/2})\,. \tag{148}$$

We will prove that

$$\|K_{j,1}\| \leq C\,|\nabla\tau|_\infty \tag{149}$$

and

$$\|K_{j,2}\| \leq C\,\min(2^j|H\tau|_\infty\,,\,|\nabla\tau|_\infty)\,. \tag{150}$$

It implies that $\|K_j\| \leq C\,|\nabla\tau|_\infty$. Inserting this inequality in (134) yields the first lemma result (132) and it proves (139). Equations (149) and (150) also prove that

$$\sum_{j=-\infty}^{0} (\|K_{j,2}\| + 2^{1/2}\|K_{j,2}\|^{1/2}\|K_{j,1}\|^{1/2}) \leq C(|\nabla\tau|_\infty + |H\tau|_\infty)\,. \tag{151}$$

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

$$\|\sum_{j=-\infty}^{0} K_{j,1}^* K_{j,1}\|^{1/2} \leq C\,(|\nabla\tau|_\infty + |H\tau|_\infty)\,. \tag{152}$$

Inserting (151) and (152) in (148) proves (140).

Let us now first prove the upper bound (150) on $K_{j,2}$. The kernel of $K_{j,2}$ is

$$k_{j,2}(x,u) = \det(\mathbf{1}-\nabla\tau(u))\Big(h_j((\mathbf{1}-\nabla\tau(u))(x-u)) - h_j(x-\tau(x)-u+\tau(u))\Big)\,.$$

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

$$\tau(x) - \tau(u) = \nabla\tau(u)(x-u) + \alpha(u, x-u) \tag{153}$$

with

$$\alpha(u,z) = \int_0^1 t\, z\, H\tau(u + (1-t)z)\, z\, dt \ , \tag{154}$$

so

$$k_{j,2}(x,u) = -\det(\mathbf{1} - \nabla\tau(u)) \tag{155}$$
$$\int_0^1 \nabla h_j\Big((\mathbf{1} - t\,\nabla\tau(u))(x-u) + (1-t)\,(\tau(u) - \tau(x))\Big)\, \alpha(u, x-u)\, dt \ .$$

For $j \leq 0$, we prove that $\|K_{j,2}\|$ decays like 2^j. Observe that $|\det(\mathbf{1} - \nabla\tau(u))| \leq 2^d$. Since $\nabla h_j(u) = 2^{-j-dj}\nabla h(2^{-j}u)$, the change of variable $x' = 2^{-j}(x-u)$ in (155) gives

$$\int |k_{j,2}(x,u)|\, dx \leq 2^d$$
$$\int \left| \int_0^1 \nabla h\Big((\mathbf{1} - t\,\nabla\tau(u))x' + (1-t)\, 2^{-j}(\tau(u) - \tau(2^j x' + u))\Big)\, 2^{-j}\alpha(u, 2^j x')\, dt \right| dx' \ .$$

For any $0 \leq t \leq 1$

$$\Big| t\,(\mathbf{1} - \nabla\tau(u))x' + (1-t)\, 2^{-j}(\tau(2^j x' + u) - \tau(u)) \Big| \geq |x'|\,(1 - |\nabla\tau|_\infty) \geq |x'|\,\epsilon \ .$$

Equation (154) also implies that

$$|2^{-j}\,\alpha(u, 2^j x')| = 2^j \Big| \int_0^1 t\, x'\, H\tau(u + (1-t)2^j x')\, x'\, dt \Big| \leq 2^j\, |H\tau|_\infty\, \frac{|x'|^2}{2} \ . \tag{156}$$

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

$$\int |k_{j,2}(x,u)|\, dx \leq C\, 2^j\, |H\tau|_\infty \, . \tag{157}$$

For $j \geq 0$, we use a maximum error bound on the rest α of the Taylor approximation (153):

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

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

$$\int |k_{j,2}(x,u)|\, dx \leq C\, \min(2^j |H\tau|_\infty\, ,\, |\nabla\tau|_\infty) \ . \tag{158}$$

Similarly, we compute $\int |k_{j,2}(x,u)|\, du$ with the change of variable $u' = 2^{-j}(x-u)$ which leads to the same bound (158). It results from Schur lemma that

$$\|K_{j,2}\| \leq C\, \min(2^j |H\tau|_\infty\, ,\, |\nabla\tau|_\infty) \tag{159}$$

which finishes the proof of (150).

Let us now compute the upper bound (149) on $K_{j,1}$. Its kernel $k_{j,1}$ in (146) can be written $k_{j,1}(x,u) = 2^{-dj}\, g(u, 2^{-j}(x-u))$ with

$$g(u,v) = h(v) - h((\mathbf{1} - \nabla\tau(u))v)\, \det(\mathbf{1} - \nabla\tau(u)). \tag{160}$$

A first order Taylor decomposition of h gives

$$g(u,v) = (1 - \det(\mathbf{1} - \nabla\tau(u)))\, h((\mathbf{1} - \nabla\tau(u)v) + \int_0^1 \nabla h((1-t)v + t(\mathbf{1} - \nabla\tau(u))v) \cdot \nabla\tau(u) v\, dt. \tag{161}$$

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

$$|g(u,v)| \leq C\, |\nabla\tau|_\infty \Big(1 + \epsilon\, |v|\Big)^{-d-2}\ , \tag{162}$$

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

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

Schur lemma (123) proves that $\|K_{j,1}\| = O(|\nabla\tau|_\infty)$ and hence (149).

Let us now prove (152) when $\int h(x)\, dx = 0$. The kernel of the self-adjoint operator $Q_j = K_{j,1}^* K_{j,1}$ is:

$$\begin{aligned} \bar{k}_j(y,z) &= \int k_{j,1}^*(x,y)\, k_{j,1}(x,z)\, dx = \int 2^{-2dj} g^*(y, 2^{-j}(x-y))\, g(z, 2^{-j}(x-z))\, dx \\ &= \int 2^{-dj} g^*(y, x' + 2^{-j}(z-y))\, g(z, x'))\, dx'\ . \end{aligned} \tag{163}$$

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 [6] but not quite because it does not satisfy the decay condition $|\bar{k}(y,z) - \bar{k}(y,z')| \leq C|z'-z|^\alpha\, |z-y|^{-d-\alpha}$ for some $\alpha > 0$. We bound this operator with Cotlar's lemma [21] which proves that if Q_j satisfies

$$\forall j,l\, ,\ \|Q_j^*\, Q_l\| \leq |\eta(j-l)|^2 \ \text{ and } \ \|Q_j\, Q_l^*\| \leq |\eta(j-l)|^2 \tag{164}$$

then

$$\| \sum_j Q_j \| \leq \sum_j \eta(j) \ . \tag{165}$$

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

$$\bar{k}_{l,j}(y,z) = \int \bar{k}_j(z,u) \, \bar{k}_l(y,u) \, du. \tag{166}$$

An upper bound of $\|Q_l \, Q_j\|$ is obtained with Schur lemma (123) applied to $\bar{k}_{l,j}$. Inserting (163) in (166) gives

$$\int |\bar{k}_{l,j}(y,z)| \, dy = \int | \int \qquad g(u,x) \, g(u,x') \, 2^{-dl} \, g^*(y, x + 2^{-l}(u-y)) \\ 2^{-dj} \, g^*(z, x' + 2^{-j}(u-z)) \, dx \, dx' \, du \Big| \, dy \tag{167}$$

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

Since $\int h(x) \, dx = 0$ it results from (160) that $\int g(u,v) \, dv = 0$ for all x. One can thus write for $v = (v_n)_{n \leq d}$

$$g(u,v) = \frac{\partial \bar{g}(u,v)}{\partial v_1}$$

and (162) implies that

$$|\bar{g}(u,v)| \leq C \, |\nabla \tau|_\infty \Big(1 + |v| \, \epsilon \Big)^{-d-1} \ . \tag{168}$$

Let us make an integration by part along the variable u_1 in (167). Since all first and second order derivatives of $h(x)$ have a decay which is $O((1 + |x|)^{-d-2})$, we derive from (160) that for any $u = (u_n)_{n \leq d} \in \mathbb{R}^d$ and $v = (v_n)_{n \leq d} \in \mathbb{R}^d$

$$\left| \frac{\partial g(u,v)}{\partial u_1} \right| \leq C \, |H\tau|_\infty \Big(1 + |v| \, (1 - |\nabla \tau|_\infty) \Big)^{-d-1} \ , \tag{169}$$

and from (161)

$$\left| \frac{\partial g(u,v)}{\partial v_1} \right| \leq C, |\nabla \tau|_\infty \Big(1 + |v| \, (1 - |\nabla \tau|_\infty) \Big)^{-d-1} \ . \tag{170}$$

In the integration by part, 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 (167) is obtained by inserting (162,168, 169,170), which prove that there exists C such that

$$\int |\bar{k}_{l,j}(y, z)|\, dy \leq C^2\, (2^j\, |\nabla\tau|_\infty^3\, |H\tau|_\infty + 2^{j-l}\, |\nabla\tau|_\infty^4) \leq C^2\, 2^{j-l}\, (|\nabla\tau|_\infty + |H\tau|_\infty)^4\,. \tag{171}$$

The same calculation proves the same bound on $\int |\bar{k}_{l,j}(y, z)|\, dz$ so Schur lemma (123) implies that

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

Applying Cotlar's lemma (164) with $\eta(j) = C\, 2^{-|j|/2}\, (|\nabla\tau|_\infty + |H\tau|_\infty)^2$ proves that

$$\| \sum_{j=-\infty}^{+\infty} K_{j,1}^* K_{j,1} \| = \| \sum_j Q_j \| \leq C\, (|\nabla\tau|_\infty + |H\tau|_\infty)^2\ , \tag{172}$$

which implies (152).

D Proof of Lemma 3

The proof of (36) shows that the scattering energy propagates towards larger scales, by measuring this propagation with an average arrival scale as the path length increases. The arrival scale index $j_{|p|}$ of $p = \{\lambda_n\}_{n \leq |p|}$ is the scale index of the last element $\lambda_{|p|} = (j_{|p|}, k_{|p|})$.

Let us write $e_m = \sum_{|p|=m} \|S(p)f\|^2$ and $e_{J,m} = \sum_{|p|=m} \|S_J(p)f\|^2$. The average arrival scale of paths of length m is

$$\bar{j}_m = e_m^{-1} \sum_{p,|p|=m} j_{|p|}\, \|S(p)f\|^2\ . \tag{173}$$

The following lemma shows that when the path length increases then the average arrival scale $\bar{j}_m$ increases by nearly $\alpha/2$.

Lemma 10 *If (31) is satisfied then*

$$\forall m > 0\ \ ,\ \ \frac{\alpha}{2}\, e_{m-1} \leq (J - \bar{j}_m) e_m - (J - \bar{j}_{m+1}) e_{m+1} + e_{m-1} - e_m\ . \tag{174}$$

We first show that (174) implies (36) and then prove this lemma. Summing over (174) gives

$$\frac{\alpha}{2}\sum_{k=0}^{m-1} e_k \leq (J-\bar{j}_1)e_1 - (J-\bar{j}_{m+1})e_{m+1} + e_0 - e_m \leq e_0 + (J-\bar{j}_1)e_1 \ . \quad (175)$$

For $m=1$, $p=(j,\gamma)$ so $\bar{j}_1\, e_1 = \sum_{j<J}\sum_{\gamma\in\Gamma} j\, \|W_{j,\gamma}f\|^2$. Moreover, $e_0 = \|f\|^2$ so

$$e_0 + (J-\bar{j}_1)e_1 = \|f\|^2 + \sum_{j<J}\sum_{\gamma\in\Gamma}(J-j)\,\|W_{j,\gamma}f\|^2 \ .$$

Inserting this in (175) proves (36).

Lemma 10 is proved by calculating the evolution of $\bar{j}_m$ as m increases. We consider the advancement of a path p of length $|p|=m-1$ with two steps $p+(j,\gamma)+(l,\beta)$, and write $f_p = S(p)f$. The average arrival scale $\bar{j}_m$ can be written as the average arrival scale of $S(p+(j,\gamma))f$ over all (j,γ) and all p with $|p|=m-1$:

$$\bar{j}_m\, e_m = \sum_{|p|=m-1}\sum_{j<J}\sum_{\gamma\in\Gamma} j\,\|f_p\star\psi_{j,\gamma}\|^2 \ . \quad (176)$$

After the second step, the average arrival scale index of $S\Big(p+(j,\gamma)+(l,\beta)\Big)f$ overall p with $|p|=m-1$, (j,γ) and (l,β) is $\bar{j}_{m+1}$:

$$\bar{j}_{m+1}\, e_{m+1} = \sum_{|p|=m-1}\sum_{j<J}\sum_{\gamma\in\Gamma}\sum_{l<J}\sum_{\gamma'\in\Gamma} l\,\||f_p\star\psi_{j,\gamma}|\star\psi_{l,\gamma'}\|^2 \ .$$

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

$$\|h\|^2 = \sum_{l<J}\sum_{\gamma'\in\Gamma}\|h\star\psi_{l,\gamma'}\|^2 + \|h\star\phi_J\|^2 \ .$$

Applied to each $h = f_p\star\psi_{j,\gamma}$ in (176) this relations, together with $e_{J,m} = \sum_{|p|=m-1}\sum_{j<J,\gamma\in\Gamma}\||f_p\star\psi_{j,\gamma}|\star\phi_J\|^2$, shows that $I = \bar{j}_{m+1}\, e_{m+1} - \bar{j}_m\, e_m + J\, e_{J,m}$ satisfies

$$\begin{aligned} I &= \sum_{|p|=m-1}\sum_{j<J}\sum_{\gamma=1}^{|\Gamma|}\Bigg(\sum_{l<J}\sum_{\gamma'=1}^{|\Gamma|}(l-j)\,\||f_p\star\psi_{j,\gamma}|\star\psi_{l,\gamma'}\|^2 \\ &\qquad + (J-j)\,\||f_p\star\psi_{j,\gamma}|\star\phi_J\|^2\Bigg) \ . \end{aligned}$$

A lower bound of I is calculated by dividing the sum on l for $l \leq j$ and $l > j$. In the $J - j - 1$ term for $l > j$, l is replaced by $j + 1$ and the convolution with ϕ_J is incorporated in the sum:

$$I \geq \sum_{|p|=m-1} \sum_{j<J} \sum_{\gamma=1}^{|\Gamma|} \left(\sum_{J>l>j} \left(\sum_{\gamma'=1}^{|\Gamma|} \| |f_p \star \psi_{j,\gamma}| \star \psi_{l,\gamma'} \|^2 \right) + \| |f_p \star \psi_{j,\gamma}| \star \phi_J \|^2 \right. \\ \left. - \sum_{l<j} \sum_{\gamma'=1}^{|\Gamma|} (j - l) \, \| |f_p \star \psi_{j,\gamma}| \star \psi_{l,\gamma'} \|^2 \right) . \tag{177}$$

Since wavelets satisfy the unitary property (10) for $\delta = 0$, for all $f \in \mathbf{L}^2(\mathbb{R}^d)$ and all $q \in \mathbb{Z}$

$$\sum_{q \leq l < J} \sum_{\gamma=1}^{|\Gamma|} \|f \star \psi_{l,\gamma}\|^2 + \|f \star \phi_J\|^2 = \|f \star \phi_q\|^2 . \tag{178}$$

Indeed (10) implies that

$$|\hat{\phi}(2^J\omega)|^2 + \sum_{q \leq l < J} \sum_{\gamma=1}^{|\Gamma|} \frac{1}{2} \Big(|\hat{\psi}_\gamma(2^l\omega)|^2 + |\hat{\psi}_\gamma(-2^l\omega)|^2 \Big) = |\hat{\phi}(2^q\omega)|^2 .$$

Multiplying this equation by $|\hat{f}(\omega)|^2$ and by integrating in ω proves (178) Inserting (178) in (177) gives

$$I \geq \sum_{|p|=m-1} \sum_{j<J} \sum_{\gamma=1}^{|\Gamma|} \Big(\| |f_p \star \psi_{j,\gamma}| \star \phi_{j+1} \|^2 \\ - \sum_{l<j} (j - l) \left(\| |f_p \star \psi_{j,\gamma}| \star \phi_l \|^2 - \| |f_p \star \psi_{j,\gamma}| \star \phi_{l+1} \|^2 \right) \Big) .$$

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

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

It results that

$$I \geq \sum_{|p|=m-1} \sum_{j<J} \sum_{\gamma=1}^{|\Gamma|} \Big(\| |f_p \star \psi_{j,\gamma}| \star \rho_{\gamma,j} \|^2 \\ - \sum_{l<j} (j - l) \left(\|f_p \star \psi_{j,\gamma}\|^2 - \| |f_p \star \psi_{j,\gamma}| \star \rho_{\gamma,l} \|^2 \right) \Big) .$$

Applying Lemma 2 for $h = \rho_{\gamma,l}$ and $\eta = 2^{-j}\xi_\gamma$ proves that

$$\| |f_p \star \psi_{j,\gamma}| \star \rho_{\gamma,l}\| \geq \|f_p \star \psi_{j,\gamma} \star \rho_{\gamma,l,\xi_\gamma}\| \quad \text{with} \quad \rho_{\gamma,l,j,\xi_\gamma}(x) = \rho_{\gamma,l}(x)\, e^{i2^{-j}\xi_\gamma x}$$

and $\hat\rho_{\gamma,l,j,\xi_\gamma}(\omega) = \hat\rho_\gamma(2^l\omega - 2^{l-j}\xi_\gamma)$. It results that

$$\begin{aligned} I \;\; \geq \;\; & \sum_{|p|=m-1} \sum_{j<J} \sum_{\gamma=1}^{|\Gamma|} \Big(\|f_p \star \psi_{j,\gamma} \star \rho_{\gamma,j,j,\xi_\gamma}\|^2 \\ & - \sum_{l<j} (j-l) \Big(\|f_p \star \psi_{j,\gamma}\|^2 - \|f_p \star \psi_{j,\gamma} \star \rho_{\gamma,l,j,\xi_\gamma}\|^2 \Big) \Big) \; . \end{aligned}$$

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

$$\begin{aligned} I \;\; \geq \;\; & \frac{1}{2} \sum_{|p|=m-1} \int |\hat f_p(\omega)|^2 \\ & \sum_{\gamma=1}^{|\Gamma|} \sum_{j<J} \Big(\Big(|\hat\psi_\gamma(2^j\omega)|^2\, |\hat\rho_\gamma(2^j\omega - \xi_\gamma)|^2 + |\hat\psi_\gamma(-2^j\omega)|^2\, |\hat\rho_\gamma(-2^j\omega - \xi_\gamma)|^2 \Big) \\ & \qquad - \sum_{l<j} (j-l) \Big(|\hat\psi_\gamma(2^j\omega)|^2\, (1 - |\hat\rho_\gamma(2^l\omega - 2^{l-j}\xi_\gamma)|^2) \\ & \qquad\qquad + \; |\hat\psi_\gamma(-2^j\omega)|^2\, (1 - |\hat\rho_\gamma(-2^l\omega - 2^{l-j}\xi_\gamma)|^2) \Big) \Big) \, d\omega \; . \end{aligned}$$

Inserting $\hat\Psi_\gamma$ defined in (30) by

$$\hat\Psi_\gamma(\omega) = |\hat\rho_\gamma(\omega - \xi_\gamma)|^2 - \sum_{k=1}^{+\infty} k\, (1 - |\hat\rho_\gamma(2^{-k}(\omega - \xi_\gamma))|^2)$$

with $k = j - l$ gives

$$I \geq \sum_{|p|=m-1} \int |\hat f_p(\omega)|^2 \sum_{j<J} a(2^j\omega)\, d\omega$$

with

$$a(\omega) = \sum_{\gamma=1}^{|\Gamma|} \Big(\hat\Psi_\gamma(\omega)\, |\hat\psi_\gamma(\omega)|^2 + \hat\Psi_\gamma(-\omega)\, |\hat\psi_\gamma(-\omega)|^2 \Big) \, .$$

Let us add to I

$$e_{J,m-1} = \sum_{|p|=m-1} \|f_p \star \phi_J\|^2 = \sum_{|p|=m-1} \int |\hat f_p(\omega)|^2\, |\hat\phi(2^J\omega)|^2\, d\omega \; .$$

The wavelet unitary property (10) together with $\hat{\Psi}_\gamma(\omega) < 1$ implies that

$$|\hat{\phi}(2^J\omega)|^2 = \frac{1}{2} \sum_{j \geq J} \sum_{\gamma=1}^{|\Gamma|} \Big(|\hat{\psi}_\gamma(2^j\omega)|^2 + |\hat{\psi}_\gamma(-2^j\omega)|^2 \leq \frac{1}{2} \sum_{j \geq J} a(2^j\omega)$$

so

$$I + e_{J,m-1} \geq \frac{1}{2} \sum_{|p|=m-1} \int |\hat{f}_p(\omega)|^2 \sum_{j=-\infty}^{+\infty} a(2^j\omega)\, d\omega \ .$$

If $\alpha = \inf_{\pi \leq |\omega| \leq 2\pi} \sum_{j=-\infty}^{+\infty} a(2^j\omega)$ then $\sum_{j=-\infty}^{+\infty} a(2^j\omega) \geq \alpha$ for all $\omega \neq 0$. If the hypothesis (31) is satisfied and hence $\alpha > 0$ then

$$\begin{aligned} I + e_{J,m-1} &\geq \frac{\alpha}{2} \sum_{|p|=m-1} \int |\hat{f}_p(\omega)|^2\, d\omega = \frac{\alpha}{2} \sum_{|p|=m-1} \|f_p\|^2 \\ &= \frac{\alpha}{2} \sum_{|p|=m-1} \|S(p)f\|^2 = \frac{\alpha}{2}\, e_{m-1} \ . \end{aligned}$$

Inserting $I = \bar{j}_{m+1}\, e_{m+1} - \bar{j}_m\, e_m + J\, e_{J,m}$ proves that

$$\bar{j}_{m+1}\, e_{m+1} - \bar{j}_m\, e_m + J\, e_{J,m} + e_{J,m-1} \geq \frac{\alpha}{2}\, e_{m-1} \ . \tag{179}$$

Since $\overline{U}_J$ preserves the norm

$$e_m = e_{m+1} + e_{J,m} \ ,$$

indeed Proposition 5 shows that

$$\overline{U}_J \{S(p)f\}_{p \in P_J, |p|=m} = \{S(p)f\}_{p \in P_J, |p|=m+1} \cup \{S_J(p)f\}_{p \in P_J, |p|=m} \ .$$

Inserting $e_{J,m} = e_m - e_{m+1}$ and $e_{J,m-1} = e_{m-1} - e_m$ in (179) gives

$$\frac{\alpha}{2}\, e_{m-1} \leq (J - \bar{j}_m) e_m - (J - \bar{j}_{m+1}) e_{m+1} + e_{m-1} - e_m \ ,$$

which finishes the proof of Lemma 10.

E Proof of Lemma 4

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

$$\|[S_J\,,\, O]f\| \leq \|[\overline{U}_J\,,\, O]\|\, \|Sf\|_{1,P_J} = \|[\overline{U}_J\,,\, O]\| \sum_{n=0}^{\infty} \Big(\sum_{|p|=n} \|S(p)f\|^2 \Big)^{1/2} \ . \tag{180}$$

We define

$$U_J f = \{|W_\lambda f| = |f \star \psi_\lambda|\}_{\lambda \in \Lambda} \quad \text{with} \quad \Lambda = \{(j,\gamma) \ : \ j < J \ , \gamma \in \Gamma\}. \quad (181)$$

Iterating n times on this operator gives $U_J^n f = \{S(p) f\}_{|p|=n}$. To prove (180), we consider the restriction $S_{J,m}$ of S_J to paths of length $|p| < m$ and show that

$$[S_{J,m}, O] = \sum_{n=0}^{m} K_{m-n} U_J^n \ , \quad (182)$$

where $K_n = \{[A_J, O] \, , \, S_{J,n-1}[U_J, O]\}$ satisfies

$$\|K_n\| \leq \|[\overline{U}_J, O]\| \, . \quad (183)$$

This result implies that for any $f \in \mathbf{L}^2(\mathbb{R}^d)$

$$\|[S_{J,m}, O] f\| \leq \sum_{n=0}^{m} \|K_{m-n}\| \, \|U_J^n f\| \leq \|[\overline{U}_J, O]\| \sum_{n=0}^{m-1} \left(\sum_{|p|=n} \|S(p) f\|^2 \right)^{1/2} ,$$

and letting m tend to ∞ proves (180).

Property (182) is proved by first showing that

$$S_{J,m} O = \{O A_J \, , \, S_{J,m-1} O U_J\} + K_m \ , \quad (184)$$

where $K_m = \{[A_J, O] \, , \, S_{J,m-1}[U_J, O]\}$. Indeed, $A_J \, U_J^n f = \{S_J(p) f\}_{|p|=n}$ so $S_{J,m} = \{A_J \, U_J^n\}_{0 \leq n < m}$. It results that

$$\begin{aligned} S_{J,m} O &= \{A_J \, U_J^n O\}_{0 \leq n < m} \\ &= \{O A_J + [A_J, O] \, , \, A_J U_J^{n-1} O U_J + A_J U_J^{n-1} [U_J, O]\}_{1 \leq n < m} \\ &= \{O A_J \, , \, S_{J,m-1} O U_J\} + \{[A_J, O] \, , \, S_{J,m-1} [U_J, O]\} \\ &= \{O A_J \, , \, S_{J,m-1} O U_J\} + K_m \ , \end{aligned}$$

which proves (184). Iterating on (184) gives

$$S_{J,m} O = \{O A_J \, , \, O A_J U_J \, , \, S_{J,m-2} O U_J^2\} + K_{m-1} U_J + K_m \ .$$

With m iterations, we obtain

$$S_{J,m} O = \{O A_J U_J^n\}_{0 \leq n < m} + \sum_{n=0}^{m} K_{m-n} U_J^n = O S_{J,m} + \sum_{n=0}^{m} K_{m-n} U_J^n$$

which proves (182).

Let us now prove the norm upper bound (183) on $K_m = \{[A_J, O]\,,\, S_{J,m-1}[U_J, O]\}$. Since S_J is contracting, its restriction $S_{J,m}$ is also contracting and hence

$$\begin{aligned} \|K_m f\|^2 &= \|[A_J, O]f\|^2 + \|S_{J,m-1}[U_J, O]f\|^2 \\ &\leq \|[A_J, O]f\|^2 + \|[U_J, O]f\|^2 = \|[\overline{U}_J, O]f\|^2 \leq \|[\overline{U}_J, O]\|^2\, \|f\|^2 \end{aligned}$$

which proves (183).

F Proof of Lemma 6

This appendix proves Lemma 6

$$E\{|K_\tau F|^2\} \leq E\{\|K_\tau\|^2\}\, E\{|F|^2\}\ , \tag{185}$$

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

$$E\{|K_\tau F(x)|^2\} = \lim_{T\to\infty} \frac{E\{\|K_\tau F_T\|^2\}}{(2\,T)^d}\ . \tag{186}$$

Let first show how this result implies (185). The $\mathbf{L}^2(\mathbb{R}^d)$ operator norm definition implies

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

Since F and τ are independent processes

$$E\{\|K_\tau F_T\|^2\} \leq E\{\|K_\tau\|^2\}\, E\{|F|^2\}\, (2T)^d\ .$$

Applying (186) thus proves the lemma result (185).

To prove (186), we first compute

$$E\{|K_\tau F(x)|^2\} = E\Big\{\iint k_\tau(x,u)\, k_\tau^*(x,u')\, F(u)\, F^*(u') du\, du'\Big\}\ .$$

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

$$\begin{aligned} E\{|K_\tau F(x)|^2\} &= \iint \bar{k}_\tau(x-u,x-u')\,A_F(u-u')\,du\,du' \\ &= \iint \bar{k}_\tau(v,v')\,A_F(v-v')\,dv\,dv' \;, \end{aligned} \tag{187}$$

which proves that $E\{|K_\tau F(x)|^2\}$ does not depend upon x. Similarly

$$E\{|K_\tau F_T(x)|^2\} = \iint \bar{k}_\tau(v,v')\,A_F(v-v')\,\mathbf{1}_{I_T}(v-x)\,\mathbf{1}_{I_T}(v'-x)\,dv\,dv', \tag{188}$$

and integrating along x gives

$$(2T)^{-d}\,E\{\|K_\tau F_T\|^2\} = \iint \bar{k}_\tau(v,v')\,A_F(v-v')\,(1-\rho_T(v-v'))\,dv\,dv', \tag{189}$$

with

$$1-\rho_T(v-v') = (2T)^{-d}\int \mathbf{1}_{I_T}(v-x)\,\mathbf{1}_{I_T}(v'-x)\,dx = \prod_{i=1}^{d}\Big(1-\frac{|v_i-v'_i|}{2T}\Big)\,\mathbf{1}_{I_T}(v-v')$$

and hence

$$0 \le \rho_T(v) \le (2T)^{-1}\sum_{i=1}^{d}|v_i| \le d\,(2T)^{-1}|v| \;. \tag{190}$$

Inserting (187) in (189) proves that

$$(2\,T)^{-d}E\{\|K_\tau F_T\|^2\} = E\{|K_\tau F(x)|^2\} - \iint \bar{k}_\tau(v,v')\,A_F(v-v')\,\rho_T(v-v')\,dv\,dv' \;. \tag{191}$$

Since $\iint|\bar{k}_\tau(v,v')|\,|v-v'|\,dv\,dv' < \infty$ and $A_F(v-v') \le A_F(0) = E\{|F|^2\}$, it results from (191) and (190) that

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

which proves (186).

Lemma 6 is extended to sequences of operators $\overline{K}_\tau = \{K_{\tau,n}\}_{n\in I}$ with kernels $\{k_{\tau,n}\}_{n\in I}$, as follow. Let us write

$$|\overline{K}_\tau F|^2 = \sum_{n\in I}|K_{\tau,n}F|^2 \;\;\text{and}\;\; \|\overline{K}_\tau f\|^2 = \sum_{n\in I}\|K_{\tau,n}f\|^2 \;. \tag{192}$$

If each average bilinear kernel is stationary

$$E\{k_{\tau,n}(x,u)\, k^*_{\tau,n}(x,u')\} = \bar{k}_{\tau,n}(x-u, x-u') \tag{193}$$

and

$$\iint |\sum_{n\in I} \bar{k}_{\tau,n}(v,v')|\, |v-v'|\, dv\, dv' < \infty \ , \tag{194}$$

then

$$E\{|\overline{K}_\tau F|^2\} \leq E\{\|\overline{K}_\tau\|^2\}\, E\{|F|^2\} \ . \tag{195}$$

The proof of this extension follows the same derivations as the proof of (185) for a single operator. It just requires to replace 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 over sequence of $\mathbf{L}^2(\mathbb{R}^d)$ functions.

G Proof of Theorem 6

This appendix proves that

$$E\{|[S_J D_\tau]F|^2\} \leq C^2\, \|SF\|^2_{1,P_J}\, E\Big\{\Big(|\nabla\tau|_\infty (\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty\Big)^2\Big\} \tag{196}$$

with $\|S\,F\|_{1,P_J} = \sum_{n=0}^{+\infty} \Big(\sum_{p\in P_J, |p|=n} E\{|S(p)F|^2\}\Big)^{1/2}$.

For this purpose, we shall first prove that if for any stationary process F

$$E\{|[\overline{W}_J, D_\tau]F|^2\} \leq B(\tau)\, E\{|F|^2\} \tag{197}$$

where

$$E\{|[\overline{W}_J\, ,\, D_\tau]F|^2\} = E\{|[A_J\, ,\, D_\tau]F|^2\} + \sum_{\lambda\in\Lambda} E\{|[W_\lambda\, ,\, D_\tau]F|^2\}$$

then

$$E\{|[S_J\, ,\, D_\tau]F|^2\| \leq B(\tau)\, \|SF\|^2_{1,P_J} \ . \tag{198}$$

Since a modulus operator is contractive and commutes with D_τ, with the same argument as in the proof of (55), we derive from (197) that

$$E\{|[\overline{U}_J, D_\tau]F|^2\} \leq B(\tau)\, E\{|F|^2\} \ . \tag{199}$$

The proof of Proposition 9 also shows that $\overline{U}_J$ is contractive for the quadratic norm on processes. Since S_J is obtained by iterating on $\overline{U}_J$ it results that

$$E\{|[S_J\,,\,D_\tau]F|^2\} \leq B(\tau)\,\|SF\|^2_{1,P_J}\ .$$

The proof of this inequality follows the same derivations as in Appendix E, for $O = D_\tau$, by replacing f by F, $\|f\|^2$ by $E\{|F|^2\}$, $\|S(p)f\|^2$ by $E\{|S(p)F|^2\}$, and the $\mathbf{L}^2(\mathbb{R}^d)$ sup operator norm $\|[\overline{U_J},0]\|$ by $B(\tau)$ which satisfies (199) for all F.

The proof of (196) is ended by verifying that

$$E\{|[\overline{W}_J, D_\tau]F|^2\} \leq E\{C^2(\tau)\}\,E\{|F|^2\} \tag{200}$$

and hence $B(\tau) = E\{C^2(\tau)\}$ with

$$C(\tau) = C\left(|\nabla\tau|_\infty (\log\frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty\right)\ .$$

The inequality (200) is derived from Lemma 5 which proves that the $\mathbf{L}^2(\mathbb{R}^d)$ operator norm of the commutator $[\overline{W}_J, D_\tau]$ satisfies

$$\|[\overline{W}_J, D_\tau]\| \leq C(\tau)\ , \tag{201}$$

and by applying the extension (195) of Lemma 6 to $K_\tau = [\overline{W}_J, D_\tau] = \{[A_J, D_\tau], [W_{j,\gamma}, D_\tau]\}_{j<J,\gamma\in\Gamma}$. This extension proves that if the kernels of the wavelet commutator statisfy the conditions (193) and (194) then

$$E\{|[\overline{W}_J, D_\tau]F|^2\} \leq E\{\|[\overline{W}_J, D_\tau]\|^2\}\,E\{|F|^2\}.$$

Together with (201) it proves (200).

To finish the proof we verify the wavelet commutator kernels satisfy (193) and (194). 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 $[Z_j, D_\tau] = Z_j D_\tau - D_\tau Z_j$ is

$$k_{\tau,j}(x,u) = h_j(x-u-\tau(x)) - h_j(x-u-\tau(u+\tau(\eta(u))))\,|\det(\mathbf{1}-\nabla\tau(u+\tau(\eta(u))))|^{-1} \tag{202}$$

where η is defined by $\eta(x) = x + \tau(\eta(x))$. The kernels of $[A_J, D_\tau]$ is $k_{\tau,J}$ with $h = \phi$, and the kernel of $[W_\lambda, D_\tau]$ for $\lambda = (j,\gamma)$ is $k_{\tau,j}$ with $h = \psi_\gamma$. Since τ and $\nabla\tau$ are jointly stationary, the joint probability distribution of their values at x and $u + \tau(\eta(u))$ only depends upon $x - u$. It results that $E\{k_{\tau,j}(x,u)\,k_{\tau,j}(x,u')\} = \bar{k}_{\tau,j}(x-u,x-u')$ which proves the kernel stationarity (193) for wavelet commutators.

The second kernel hypothesis (194) is proved by showing that if $|h(x)| = O((1+|x|)^{-d-2})$ then

$$\iint |\sum_{j\le J} \bar{k}_{\tau,j}(v,v')|\,|v-v'|\,dv\,dv' < \infty\ .$$

Since $\bar{k}_{\tau,j}(v,v') = E\{k_{\tau,j}(x,x-v)\,k_{\tau,j}(x,x-v')\}$, it is sufficient to prove that there exists C such that for all x, with probability 1

$$I = \sum_{j\le J} \iint |k_{\tau,j}(x,x-v)|\,|k_{\tau,j}(x,x-v')|\,|v-v'|\,dv\,dv' \le C\ . \tag{203}$$

Since $h_j(x) = 2^{-dj}h(2^{-j}x)$ and $u+\tau(\eta(u)) = \eta(u)$, it results from (202) that $k_{\tau,j}(x,x-2^jw) = 2^{-dj}\,\tilde{k}_{\tau,j}(x,x-w)$ with

$$\tilde{k}_{\tau,j}(x,x-w) = h(w-2^{-j}\tau(x)) - h(w-2^{-j}\tau(\eta(x-2^jw)))\,|\det(\mathbf{1}-\nabla\tau(\eta(x-2^jw)))|^{-1}\ . \tag{204}$$

The change of variable $w = 2^{-j}v$ and $w' = 2^{-j}v'$ in (203) shows that $I = \sum_{j\le J} 2^j I_j$ with

$$I_j = \iint |\tilde{k}_{\tau,j}(x,x-w)|\,|\tilde{k}_{\tau,j}(x,x-w')|\,|w-w'|\,dw\,dw'\ .$$

Since $|h(w)| = O((1+|w|)^{-d-2})$ and $|\nabla\tau|_\infty < 1-\epsilon$ with probability 1 for $\epsilon > 0$, by computing separately the integrals of each of the four terms of the product $|\tilde{k}_{\tau,j}(x,x+w)|\,|k_{\tau,j}(x,x+w')|\,|w-w'|$, with change of variables, $y = w+2^{-j}\tau(x)$ and $z = w+2^{-j}\tau(\eta(x+2^jw))$, we verify that there exists C' such that $I_j \le C'$ and hence that $I = \sum_{j\le J} 2^j I_j \le 2^{J+1}C'$ with probability 1. It proves (203) and hence the second kernel hypothesis (194).

H Proof of Proposition 12

A scale increasing path whose first scale is j_1 is necessarily of length $|p| \le J - j_1$ since $j_n < j_{n+1} < J$. One can verify that there are $|\Gamma|^{|p|-1}\binom{J-j_1-1}{|p|-1}$ scale increasing paths of length $|p|$ whose first element is (j_1,γ_1). Applying (98) yields

$$\sum_{p\in\widetilde{P}_J} E\{|S_J(p)F - \overline{S_J(p)F}|^2\} \le E\{|F\star\phi_J - E\{F\}|^2\}$$

$$+\sum_{j_1=-\infty}^{J-1}\sum_{\gamma_1=1}^{|\Gamma|} E\{|F\star\psi_{j_1,\gamma_1}|^2\}\,2^{d(j_1-J)}\sum_{|p|=1}^{J-j_1}\binom{J-j_1-1}{|p|-1}|\Gamma|^{|p|-1}\alpha^{|p|-1}$$

$$\leq\ E\{|F\star\phi_J-E\{F\}|^2\}+\sum_{j_1=-\infty}^{J-1}\sum_{\gamma_1=1}^{|\Gamma|} E\{|F\star\psi_{j_1,\gamma_1}|^2\}\,2^{d(j_1-J)}\,(1+|\Gamma|\alpha)^{J-j_1-1}.$$

Since $E\{|F\star\psi_{j_1,\gamma_1}|^2\} = \int\hat R_F(\omega)\,|\hat\psi_{\gamma_1}(2^{j_1}\omega)|^2\,d\omega\ \leq\ 2^{-dj_1}\|\hat R_F\|_\infty^2\,\|\psi_{\gamma_1}\|^2$, by splitting the sum on j_1 in two parts we get

$$\sum_{p\in\widetilde P_J} E\{|S_J(p)F-\overline{S_J(p)F}|^2\}\leq\|\hat R_F\|_\infty^2 2^{-dJ}\|\phi\|^2$$

$$+\|\hat R_F\|_\infty^2\,2^{-dJ}\,(1+|\Gamma|\alpha)^J\sum_{j_1=1}^{J-1}\left(\sum_{\gamma_1=1}^{|\Gamma|}\|\psi_{\gamma_1}\|^2\right)(1+|\Gamma|\alpha)^{-j_1-1}$$

$$+2^{-Jd}\,(1+|\Gamma|\alpha)^J\sum_{j_1=-\infty}^{0}\sum_{\gamma_1=1}^{|\Gamma|}E\{|F\star\psi_{j_1,\gamma_1}|^2\}\,2^{dj_1}\,(1+|\Gamma|\alpha)^{-j_1-1}\ .$$

Since $|\Gamma|\leq 2^d-1$, $1+|\Gamma|\alpha<2^d$, and hence the two geometric series converge. It results from (84) that

$$R_F(0)=E\{|F-E\{F\}|^2\}\geq\sum_{j_1=-\infty}^{0}\sum_{\gamma_1=1}^{|\Gamma|}E\{|F\star\psi_{j_1,\gamma_1}|^2\}.$$

Inserting this inequality proves that

$$\sum_{p\in\widetilde P_J} E\{|S_J(p)F-\overline{S_J(p)F}|^2\}\leq C\,(\|\hat R_F\|_\infty^2+R_F(0))\left(\frac{1+|\Gamma|\,\alpha}{2^d}\right)^J\ .$$

I Proof of Lemma 8

Let

$$\|\overline{W}_JD_\tau-D_\tau\overline{W}_{J,\beta}f\|^2=\|[A_J,D_\tau]f\|^2+\sum_{j<J}\int_G\|W_{j,\gamma}D_\tau f-D_\tau W_{j,\beta\gamma}f\|^2\ .$$

We want to prove that

$$\|\overline{W}_JD_\tau-D_\tau\overline{W}_{J,\beta}\|\leq C\left((|\overline{\nabla\tau}^\perp|_\infty+|\nabla\tau|_\infty^2)(\log\frac{|\tau|_\infty}{|\nabla\tau|_\infty}\vee 1)+|H\tau|_\infty\right)\quad(205)$$

with $\beta(x)=\exp(\overline{\nabla\tau}(x))\in G$. The proof of (205) is based on the following lemma.

Lemma 11 *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 $h_j(x) = 2^{-dj} h(2^{-j}x)$ and $Z_{j,\beta} f = f \star h_{j,\beta}$ with $h_{j,\beta}(x) = 2^{-dj} h(2^{-j}\beta(x)\, x)$. Suppose that $|\nabla\tau|_\infty < 1-\epsilon$ with $\epsilon > 0$ and $\beta(x) = \exp(\overline{\nabla\tau}(x))$.*
If $h(\gamma x) = h(x)$ for all $\gamma \in G$, then there exists $C > 0$ such that

$$\|[Z_j, D_\tau] f\| \leq C\, \|f\|\, (|\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|_\infty^2)\ . \tag{206}$$

If $\int h(x)\, dx = 0$ then there exists $C > 0$ such that

$$\sum_{j<J} \|Z_j D_\tau f - D_\tau Z_{j,\beta} f\|^2 \leq C\, \|f\|^2 \Big((|\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|_\infty^2)(\log \frac{|\tau|_\infty}{|\nabla\tau|_\infty} \vee 1) + |H\tau|_\infty \Big)^2 . \tag{207}$$

Equation (205) is derived from Lemma 11 by applying (206) for $j = J$ and $h = \phi$, and applying (207) for $h = \psi_\gamma$. The proof of Lemma 11 follows an approach similar to the proof of Lemma 9, which is described below.

Factorizing D_τ only affects the operator norm by a constant because Appendix C proves that $\|D_\tau\| \leq (1-\epsilon)^{-d}$:

$$Z_j D_\tau - D_\tau Z_{j,\beta} = K_j\, D_\tau \ \text{ with }\ K_j = Z_j - D_\tau\, Z_{j,\beta}\, D_\tau^{-1}$$

The kernel of K_j is

$$k_j(x,u) = h_j(x-u) - h_j(\beta(x)(x-\tau(x)-u+\tau(u)))\ \det(\mathbf{1}-\nabla\tau(u))\ . \tag{208}$$

Similarly, factorizing D_τ from $[Z_j, D_\tau]$ defines $[Z_j, D_\tau] D_\tau^{-1}$ whose kernel is

$$\tilde{k}_j(x,u) = h_j(x-u) - h_j(x-\tau(x)-u+\tau(u))\ \det(\mathbf{1}-\nabla\tau(u))\ .$$

If $h(\gamma x) = h(x)$ for all $\gamma \in G$ then $h_j(\beta(x)x) = h_j(x)$, so $\tilde{k}_j = k_j$ and hence $\|[Z_j, D_\tau]\| \leq \|K_j\|\, (1-\epsilon)^{-d}$. We shall prove that

$$\|K_j\| \leq C\, (|\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|_\infty^2), \tag{209}$$

which implies the first lemma result (206).

The second lemma result is proved by computing an upper bound of $\|\sum_{j=-\infty}^{\infty} K_j^*\, K_j\|$. As in (137), the sum over j is divided in three parts:

$$\|\sum_{j=-\infty}^{+\infty} K_j^*\, K_j\|^{1/2} \leq \|\sum_{j=-\infty}^{0} K_j^*\, K_j\|^{1/2} + \|\sum_{j=1}^{\eta} K_j^*\, K_j\|^{1/2} + \|\sum_{j=\eta+1}^{+\infty} K_j^*\, K_j\|^{1/2}. \tag{210}$$

An upper bound of the sum is obtained from a first bound at large scales

$$\| \sum_{j=\eta}^{+\infty} K_j^* K_j \|^{1/2} \leq C \left(|\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|_\infty^2 + 2^{-\eta}\, |\tau|_\infty + 2^{-\eta/2}\, |\tau|_\infty^{1/2}\, (|\overline{\nabla\tau}|_\infty + |\nabla\tau|_\infty^2)^{1/2} \right) , \tag{211}$$

and a second bound at intermediate scales derived from (209)

$$\| \sum_{j=1}^{\eta} K_j^*\, K_j \|^{1/2} \leq \eta\, \|K_j\| \leq C\, \eta \left(|\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|_\infty^2 \right) , \tag{212}$$

and a third bound at fine scales

$$\| \sum_{j=-\infty}^{0} K_j^*\, K_j \|^{1/2} \leq C \left(|\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|_\infty^2 + |H\tau|_\infty \right) . \tag{213}$$

Inserting (211), (212) and (213) in (210), and choosing $\eta = \max(\log \frac{|\tau|_\infty}{|\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|_\infty^2}, 1)$ yields

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

which proves the second lemma result (207).

The proof of (211) is almost identical to the proof of (138). It uses the fact that $|\beta(x)\tau| \leq |\tau|$ because $\beta(x) \in G$ which is a compact subgroup of $GL(\mathbb{R}^d)$ and hence $\beta(x)$ has a norm 1. It also uses the following upper bound:

$$|1 - \det(\mathbf{1} - \nabla\tau)| \leq d\, (|\overline{\nabla\tau}^{\perp}|_\infty + O(|\nabla\tau|_\infty^2)) .$$

This bound is proved by inserting in the determinant $\beta(x) = \exp(\overline{\nabla\tau}(x)) \in G$ which has a unit determinant and hence

$$\begin{aligned} |1 - \det(\mathbf{1} - \nabla\tau)| &= |1 - \det(\exp(\overline{\nabla\tau})\, (\mathbf{1} - \nabla\tau))| \\ &\leq |1 - (1 - |\overline{\nabla\tau}^{\perp}|_\infty - O(|\nabla\tau|_\infty^2))^{-d}| . \end{aligned}$$

Indeed

$$\beta(\mathbf{1} - \nabla\tau) = \mathbf{1} - \overline{\nabla\tau}^{\perp} + e \ \text{ with } \ |e| = O(|\nabla\tau|^2) . \tag{214}$$

To prove (209) and (213), we decompose $K_j = K_{j,1} + K_{j,2}$ by decomposing its kernel (208) into $k_j = k_{j,1} + k_{j,2}$ with

$$k_{j,1}(x,u) = h_j(x-u) - h_j(\beta(x)\, (\mathbf{1} - \nabla\tau(u))(x-u))\ \det(\mathbf{1} - \nabla\tau(u))$$

and

$$k_{j,2}(x,u) = \det(\mathbf{1} - \nabla\tau(u)) \quad \Big(\quad h_j(\beta(x)\,(\mathbf{1} - \nabla\tau(u))(x-u)) \\ -h_j(\beta(x)(x - \tau(x) - u + \tau(u)))\Big) \,.$$

These kernels have the same properties as the kernels defined in (146) and (147) but $\mathbf{1} - \nabla\tau$ is now replaced by $\beta\,(\mathbf{1} - \nabla\tau)$. Besides this modifications, all other calculations are identical. Similarly to (149) and (150), applying Schur lemma and taking into account (214) proves that

$$\|K_{j,1}\| \leq C \left(|\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|^2_\infty \right) \tag{215}$$

and

$$\|K_{j,2}\| \leq C \, \min(2^j |H\tau|_\infty \,,\, |\overline{\nabla\tau}^{\perp}|_\infty + |\nabla\tau|^2_\infty) \,, \tag{216}$$

which proves (209). As in (172) applying Cotlar lemma to $Q_j = K^*_{j,1} K_{j,1}$ and the same computations prove the second upper bound (213).

Acknowledgement I would like to thank Joan Bruna, Mike Glinsky and Grard Kerkyacharian for the many inspiring conversations in connection with image processing, physics and group theory.

References

[1] Y. Amit,, U. Grenander and M. Piccioni; "Structural image restoration through deformable templates". Journal of the American Statistical Association, 86 (414), 376? 387. 1991

[2] R. Bajcsy and S. Kovacic, Multi-resolution elastic matching, Comput. Vis., Graph. Image Processing, vol. 46, pp. 121, 1989.

[3] C. Bernard, "Discrete wavelets analysis: a new framework for fast optical flow computations", J. of Appl. and Comput. Harmonic Analysis, 11(1):32-63, 2001.

[4] J. Bouvrie, L. Rosasco, T. Poggio: "On Invariance in Hierarchical Models". NIPS 2009

[5] J. Bruna, S. Mallat, "Classification with scattering operators", CMAP Tech. Report, Ecole Polytechnique, http://arxiv.org/abs/1011.3023.

[6] G. David, J.-L. Journé, and S. Semmes, "Opérateurs de Calderòn Zygumund, fonctions para-accrétives et interpolation," Rev. Mat. Iberoameri. 1:1-56, 1985.

[7] D. Donoho, "Unconditional bases and bit-level compression", J. of Appl. and comput. Harmonic Analysis, 3: 388-392, 1996.

[8] D. Geller and I. Z. Pesenson, "Band-Limited Localized Parseval Frames and Besov Spaces on Compact Homogeneous Manifolds", J. of Geometric Analysis, July, 2010.

[9] M. Glinsky, private communication.

[10] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition", Proceedings of the IEEE, (86), 22782324, 1998.

[11] Y. LeCun, K. Kavukvuoglu and C. Farabet: "Convolutional Networks and Applications in Vision", Proc. International Symposium on Circuits and Systems (ISCAS'10), IEEE, 2010.

[12] W. Lohmiller and J.J.E. Slotine "On Contraction Analysis for Nonlinear Systems", Automatica, 34(6), 1998.

[13] S. Mallat, "A Wavelet Tour of Signal Processing", Academic Press, Elsevier, 2009.

[14] S. Mallat. "Recursive Interferometric Representation", Proc. of EUSICO conference, Danemark, August 2010.

[15] Y. Meyer, "Wavelets and Operators", Cambridge University Press, 1992.

[16] I. Rahman, I. Drori , V. Stodden , D. Donoho, "Multiscale Representations for Manifold-Valued Data," SIAM Multiscale Modelling and Simulation, vol. 4, no. 4, 2005, pp. 12011232.

[17] M. Riesenhuber, T. Poggio, "Hierarchical models of object recognition in cortex," Nature Neuroscience", 2: 10191025.

[18] B. Schlkopf, A. J. Smola, "Learning with Kernels", MIT Press, 2002.

[19] P. Simoncelli, and W. T. Freeman, "The Steerable Pyramid: A Flexible Architecture for Multi-Scale Derivative Computation", Proc. IEEE ICIP, Washington DC, October 1995.

[20] E. Stein, “Topics in harmonic analysis related to Littlewood-Paley theory”, Annals of Mathematical Studies, Princeton University Press, 1970.

[21] E. Stein, “Harmonic Analysis”, Princeton Mathematical Series, Princeton University Press, 1993.

[22] A.Trouve, L.Younes; “Local Geometry of Deformable Templates”; SIAM J. Math Anal. Volume 37, Issue 1, pp. 17-59; 2005

[23] Y. Yang, “Minimax Nonparametric Classification”, IEEE Trans. on Information Theory, 45: 2271-2284, 1999.