Exponential decay of scattering coefficients

Irène Waldspurger [*]

This technical report is the fifth chapter of [Waldspurger, 2015].

1 Introduction

When trying to analyze or interpret a complex signal (image, audio signal, temporal series...), it is often difficult to directly work with the natural representation of the signal (an array of pixels, in the case of an image). On the contrary, using a more appropriate representation can render the task much easier to solve. The scattering transform, introduced by Mallat [2012], can serve as such a representation for various categories of signals and tasks.

Defined over $L^2(\mathbb{R}^d)$, the scattering transform consists in a cascade of wavelet transform modulus. It has the property to be invariant to translations, and stable to small deformations. In the last years, it has proven to be a very efficient tool in diverse settings [Andén and Mallat, 2011; Oyallon and Mallat, 2015; Hirn et al., 2015].

It belongs to the family of deep representations, but, contrarily to most other members of this family, it is not learned from data. Its theoretical properties are therefore easier to study, and shed some light on the behaviour of deep representations in general. In particular, through the scattering transform, we can hope to approach the following general question: what information contained in the original signal does a deep transform capture? This question is central for the understanding of all deep representations, but for learned ones, only empirical studies have until now been possible [Dosovitskiy and Brox, 2015; Simonyan et al., 2014; Mahendran and Vedaldi, 2015].

For the scattering transform, Bruna et al. [2015] have already shown that it characterizes important parameters of some classes of stationary processes. In the present work, we consider the scattering transform over $L^2(\mathbb{R})$, and partially describe the amount of information contained in each layer of the transform. Informally, we show that, for some $r > 0, a > 1$, the n-th layer of the scattering transform is almost insensitive to the frequencies of the signal that belong to the range $[-ra^n; ra^n]$.

---

[*]Previously at département d'informatique de l'ENS Paris; now at MIT Institute for Data, Systems and Society.

The exact statement of this result is given in Theorem 3.1. Some hypotheses on the wavelets are necessary, but there are much less restrictive than the ones used in [Mallat, 2012]. For simplicity, we restrict ourselves to the one-dimensional case, but we believe that the result could be extended to $L^2(\mathbb{R}^d)$ for $d \geq 1$.

The main two consequences of this result are the following.

- It generalizes the results of [Mallat, 2012, Section 2]: it allows to relax the *admissibility condition* on the wavelets.

- It explains why, in applications, high order scattering coefficients can be neglected. In cases where higher non-linearity orders seem desirable, this should incite us to redesign the definition of high order layers.

The organization is as follows. In Section 2, we precisely define the scattering transform and some known results about it. We also describe in more detail the propagation phenomenon. In Section 3, we state the theorem. In Section 4, we give the principle of the proof. Section 5 adapts the theorem to the scattering transform on stationary processes. Finally, Section 6 proves the lemmas necessary for the proof of the main theorem.

2 The scattering transform

As in the previous chapters, once a wavelet $\psi \in L^1 \cap L^2(\mathbb{R})$ (such that $\int_\mathbb{R} \psi = 0$) has been chosen, we define a family of wavelets $(\psi_j)_{j \in \mathbb{Z}}$ by:

$$\forall j \in \mathbb{Z}, t \in \mathbb{R} \qquad \psi_j(t) = 2^{-j}\psi(2^{-j}t)$$
$$\Longleftrightarrow \quad \forall j \in \mathbb{Z}, \omega \in \mathbb{R} \qquad \hat{\psi}_j(\omega) = \hat{\psi}(2^j\omega)$$

2.1 Definition

We follow the definition of [Mallat, 2012].

The scattering transform consists in a cascade of modulus of wavelet transforms. After each application of the modulus, the resulting functions are locally averaged. The set of averages constitutes the scattering transform.

The averaging is performed with a real-valued positive function $\phi \in L^1 \cap L^2(\mathbb{R})$ such that $\hat{\phi}(0) = 1$. We define:
$$\forall J \in \mathbb{Z}, t \in \mathbb{R} \qquad \phi_J(t) = 2^{-J}\phi(2^{-J}t)$$

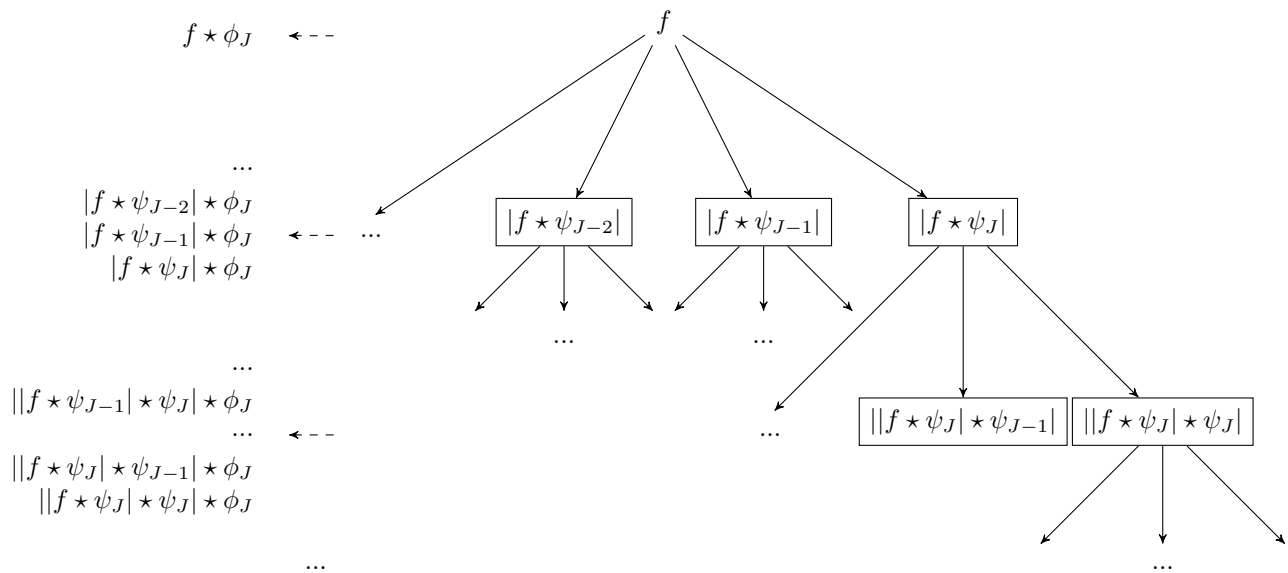The convolution with ϕ_J represents an average on an interval of characteristic size 2^J.

$f \star \phi_J \quad \leftarrow --$

...

$|f \star \psi_{J-2}| \star \phi_J$
$|f \star \psi_{J-1}| \star \phi_J \quad \leftarrow -- \quad ...$
$|f \star \psi_J| \star \phi_J$

...

$||f \star \psi_{J-1}| \star \psi_J| \star \phi_J$
... $\quad \leftarrow --$
$||f \star \psi_J| \star \psi_{J-1}| \star \phi_J$
$||f \star \psi_J| \star \psi_J| \star \phi_J$

...

f

$|f \star \psi_{J-2}|$ $\quad$ $|f \star \psi_{J-1}|$ $\quad$ $|f \star \psi_J|$

... $\quad$ ...

... $\quad$ $||f \star \psi_J| \star \psi_{J-1}|$ $\quad$ $||f \star \psi_J| \star \psi_J|$

...

Figure 1: Schematic illustration of the scattering transform: the tree on the right represents the cascade of modulus of wavelet transforms; the output scattering coefficients are on the left.

We now formally define the cascade of modulus of wavelet transforms. For any function $f \in L^2(\mathbb{R})$, we set:

$$U[\o]f = f$$

and iteratively define, for any n-uplet $(j_1, ..., j_n) \in \mathbb{Z}^n$, with $n \geq 1$:

$$U[(j_1, ..., j_n)]f = \left|U[(j_1, ..., j_{n-1})]f \star \psi_{j_n}\right|$$

For any $J \in \mathbb{Z}$, we set $\mathcal{P}_J = \{(j_1, ..., j_n), n \in \mathbb{N}, j_1, ..., j_n \in \{-\infty, ..., J\}\}$; we refer to the elements of $\mathcal{P}_J$ as *paths*. We denote the length (that is, the number of elements) of a path p by $|p|$.

For any $p \in \mathcal{P}_J$, we define:

$$S_J[p]f = U[p]f \star \phi_J$$

The *scattering coefficients* associated to f at scale J are the set $\{S_J[p]f\}_{p \in \mathcal{P}_J}$.

The computation of the scattering coefficients is schematized in Figure 1.

3

2.2   Norm preservation and energy propagation

When the wavelets are suitably chosen, the scattering transform preserves the norm [Mallat, 2012, Theorem 2.6]:

$$\sum_{p \in \mathcal{P}_J} ||S_J[p]f||_2^2 = ||f||_2^2 \tag{1}$$

In this paragraph, we briefly describe this result. To explain it, we introduce the informal idea that the repeated application of the modulus of the wavelet transform moves the energy contained in the high frequencies of the initial signal towards the low frequencies.

For the moment, we assume the wavelets to be analytical:

$$\hat{\psi}(\omega) = 0 \text{ if } \omega < 0$$

and, together with the low-pass filter ϕ, to satisfy a Littlewood-Paley condition:

$$\forall \omega \geq 0 \qquad |\hat{\phi}(\omega)|^2 + \frac{1}{2}\sum_{j \leq 0} |\hat{\psi}_j(\omega)|^2 = 1$$

This condition ensures that, for any scale $J \in \mathbb{Z}$, the wavelet transform with low-pass filter $f \to \{f \star \phi_J, \{f \star \psi_j\}_{j \leq J}\}$ is unitary over the set of real-valued functions:

$$\forall f \in L^2(\mathbb{R}, \mathbb{R}) \qquad ||f \star \phi_J||_2^2 + \sum_{j \leq J} ||f \star \psi_j||_2^2 = ||f||_2^2$$

Consequently, the application $W_J : f \to \{f \star \phi_J, \{|f \star \psi_j|\}_{j \leq J}\}$ preserves the norm. As the scattering transform is computed by recursively applying this operator to the input function, we can prove by iteration over n that, for any length $n \geq 0$:

$$\forall f \in L^2(\mathbb{R}, \mathbb{R}) \qquad \sum_{p \in \mathcal{P}_J, |p| < n} ||S_J[p]f||_2^2 + \sum_{p \in \mathcal{P}_J, |p| = n} ||U[p]f||_2^2 = ||f||_2^2$$

If we can prove that $\sum_{p \in \mathcal{P}_J, |p| = n} ||U[p]f||_2^2$ goes to zero when n goes to infinity, then we can prove Equation (1). Under an additional condition on the wavelets, this is done in [Mallat, 2012, Theorem 2.6]:

Theorem 2.1. *A family $(\phi, \{\psi_j\}_{j \in \mathbb{Z}})$ is said to be admissible if there exists $\eta \in \mathbb{R}$ and a positive real-valued function $\rho \in L^2(\mathbb{R})$ such that:*

$$\forall \omega \in \mathbb{R}, \quad |\hat{\rho}(\omega)| \leq |\hat{\phi}(2\omega)| \qquad and \qquad \hat{\rho}(0) = 1$$

and that the function:

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

satisfies:

$$\inf_{1 \leq \omega \leq 2} \sum_{j=-\infty}^{+\infty} \hat{\Psi}(2^j \omega) |\hat{\psi}_j(\omega)|^2 > 0 \tag{2}$$

If $(\phi, \{\psi_j\}_{j \in \mathbb{Z}})$ *is admissible, then, for any real-valued function* $f \in L^2(\mathbb{R})$ *and any scale* $J \in \mathbb{Z}$:

$$\sum_{p \in \mathcal{P}_J, |p|=n} ||U[p]f||_2^2 \to 0 \qquad \textit{when } n \to +\infty \tag{3}$$

which implies the norm preservation:

$$\sum_{p \in \mathcal{P}_J} ||S_J[p]f||_2^2 = ||f||_2^2$$

Intuitively, the property (3) holds because the iterative application of the modulus of the wavelet transform moves the energy carried by the high frequencies of the signal towards the low frequencies. At each step of the scattering transform, the energy in the lowest frequency bands is output by convolution with ϕ_J. The remaining part is shifted towards the lower frequencies by a new application of the modulus of the wavelet transform and so on.

For the displacement of the energy towards the low frequencies, the modulus is essential: for each signal f, $f \star \psi_j$ is a function whose energy is concentrated in a frequency band of characteristic size 2^{-j}, with a mean frequency also of the order of 2^{-j}. After application of the modulus, $|f \star \psi_j|$ tends to have its energy concentrated in a frequency band of characteristic size still equal to 2^{-j}, but now centered around zero. So the frequencies that it contains are globally lower than the frequencies of $f \star \psi_j$. An example of this phenomenon is displayed on Figure 2.

According to this simplistic reasoning, the modulus of the wavelet transform approximately moves the energy contained in a frequency band around 2^j to the frequency band $[-2^{j-1}, 2^{j-1}]$. By iteratively applying this argument, we expect the energy to arrive in the frequency band $[-2^J, 2^J]$ (and thus disappear) after a number of scattering steps proportional to j. The theorem of the next section will formalize this idea.

3 Theorem statement

The theorem of this section relies on similar ideas to the ones of Theorem 2.1 but improves it in two aspects:
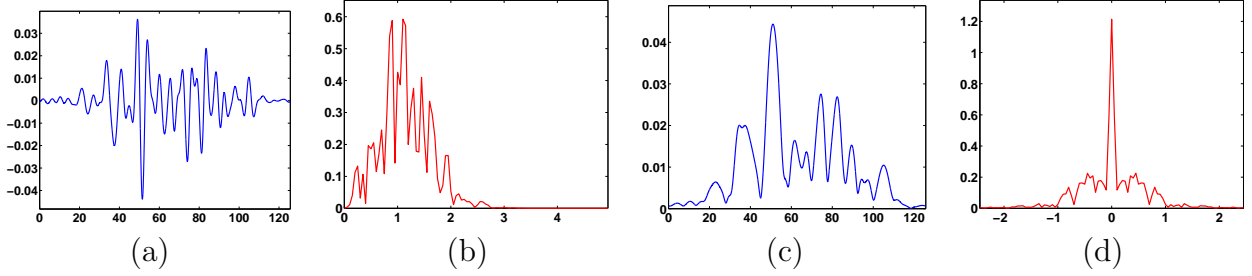
Figure 2: Illustration of the shift towards low frequencies due to the modulus. (a) $f \star \psi_j$ (real part) (b) $\widehat{f \star \psi_j}$ (in modulus) (c) $|f \star \psi_j|$ (d) $\widehat{|f \star \psi_j|}$. Observe that $\widehat{f \star \psi_j}$ and $\widehat{|f \star \psi_j|}$ are localized on frequency bands of the same width, but that, for $\widehat{|f \star \psi_j|}$, the band is centered around 0, and thus globally lower in frequency.

- It gives a precise bound on $\sum_{p \in \mathcal{P}_J, |p|=n} ||U[p]f||_2^2$, instead of simply ensuring that this quantity goes to zero. This bound formalizes the idea described in the end of the previous paragraph, that the energy of the signal carried by the frequencies around 2^j disappears after a number of scattering step proportional to j.

- It holds for a much more general class of wavelets that Theorem 2.1. In particular, it does not require the wavelets to be analytical. It notably applies to Morlet wavelets, and also to compactly-supported wavelets like Selesnick wavelets [Selesnick, 2001], two important choices for practical applications. The wavelets have to satisfy three simple conditions, which will be commented after the statement of the theorem.

For any $a > 0$, we denote by χ_a the Gaussian function:

$$\forall t \in \mathbb{R} \qquad \chi_a(t) = \sqrt{\pi} a \exp\left(-(\pi a)^2 t^2\right)$$

whose Fourier transform satisfies:

$$\forall \omega \in \mathbb{R} \qquad \hat{\chi}_a(\omega) = \exp(-(\omega/a)^2)$$

Theorem 3.1. *Let $(\psi_j)_{j \in \mathbb{Z}}$ be a family of wavelets. We assume the following Littlewood-Paley inequality to be satisfied:*

$$\forall \omega \in \mathbb{R} \qquad \frac{1}{2} \sum_{j \in \mathbb{Z}} \left(|\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2\right) \leq 1 \qquad (4)$$

We also assume that:

$$\forall j \in \mathbb{Z}, \omega > 0 \qquad |\hat{\psi}_j(-\omega)| \leq |\hat{\psi}_j(\omega)| \qquad (5)$$

6

and, for any ω, the inequality is strict for at least one value of j.

Finally, we assume that, for some $\epsilon > 0$:

$$\hat{\psi}(\omega) = O(|\omega|^{1+\epsilon}) \tag{6}$$

when $\omega \to 0$.

Then, for any $J \in \mathbb{Z}$, there exist $r > 0, a > 1$ such that, for all $n \geq 2$ and $f \in L^2(\mathbb{R}, \mathbb{R})$:

$$\sum_{p \in \mathcal{P}_J, |p|=n} ||U[p]f||_2^2 \leq ||f||_2^2 - ||f \star \chi_{ra^n}||_2^2 = \int_{\mathbb{R}} |\hat{f}(\omega)|^2 \left(1 - |\hat{\chi}_{ra^n}(\omega)|^2\right) d\omega \tag{7}$$

Before turning to the proof of this theorem, let us briefly comment the conditions (4), (5) and (6).

The Littlewood-Paley inequality (4) is equivalent to the fact that the wavelet transform is contractive over $L^2(\mathbb{R}, \mathbb{R})$. Without a condition of this kind, the wavelet transform can amplify some frequencies of the initial signal, and the energy contained in the long paths of length n will not necessarily decrease when n increases.

The condition (5) describes the fact that, although they do not need to be analytical, the wavelets must give more weight to the positive frequencies than to the negative ones. If this condition is not met, the phenomenon of energy shift towards the low frequencies may not happen. This is in particular the case if the wavelets are real.

The condition (6) states that the wavelets have a bit more than one zero momentum. It is necessary for our proof, but it is not clear whether the theorem stays true without it or not.

4   Principle of the proof

The proof proceeds by iteration over n.

In this paragraph, we show that, if a is close enough to 1, the fact that the property holds for n implies that it also holds for $n + 1$. The initialization, for $n = 2$, relies on the same principle, but is more technical; it is done in the paragraph 6.3.

Lemma 4.1. *For any $j \in \mathbb{Z}, x \in \mathbb{R}_+^*, \delta_j \in \mathbb{R}$:*

$$|| |f \star \psi_j| \star \chi_x||_2^2 \geq ||f \star \psi_j \star (\chi_x e^{2\pi i \delta_j \cdot})||_2^2 = \int_{\mathbb{R}} |\hat{f}(\omega)|^2 |\hat{\psi}_j(\omega)|^2 |\hat{\chi}_x(\omega - \delta_j)|^2 d\omega$$

This lemma is already present in [Mallat, 2012]. For sake of completeness, we prove it again in the paragraph 6.1.

If the property (7) holds for $n \geq 2$, then:

$$\sum_{p \in \mathcal{P}_J, |p|=n+1} ||U[p]f||_2^2 = \sum_{j \leq J} \left(\sum_{p \in \mathcal{P}_J, |p|=n} ||U[j,p]f||_2^2 \right)$$

$$= \sum_{j \leq J} \left(\sum_{p \in \mathcal{P}_J, |p|=n} ||U[p] \left(|f \star \psi_j| \right) ||_2^2 \right)$$

$$\leq \sum_{j \leq J} \left(|| \, |f \star \psi_j| \, ||_2^2 - || \, |f \star \psi_j| \star \chi_{ra^n} ||_2^2 \right) \qquad (8)$$

By Lemma 4.1, for any choice of $(\delta_j)_{j \leq J}$:

$$\sum_{p \in \mathcal{P}_J, |p|=n+1} ||U[p]f||_2^2 \leq \sum_{j \leq J} \left(||f \star \psi_j||_2^2 - ||f \star \psi_j \star (\chi_{ra^n} e^{2\pi i \delta_j})||_2^2 \right)$$

$$= \int_{\mathbb{R}} |\hat{f}(\omega)|^2 \left(\sum_{j \leq J} |\hat{\psi}_j(\omega)|^2 (1 - |\hat{\chi}_{ra^n}(\omega - \delta_j)|^2) \right) d\omega$$

We symmetrize the expression, by taking into account the fact that f is real, so $|\hat{f}(\omega)| = |\hat{f}(-\omega)|$:

$$\sum_{p \in \mathcal{P}_J, |p|=n+1} ||U[p]f||_2^2 \leq \int_{\mathbb{R}} |\hat{f}(\omega)|^2 \times \frac{1}{2} \left(\sum_{j \leq J} |\hat{\psi}_j(\omega)|^2 (1 - |\hat{\chi}_{ra^n}(\omega - \delta_j)|^2) \right.$$

$$\left. + |\hat{\psi}_j(-\omega)|^2 (1 - |\hat{\chi}_{ra^n}(-\omega - \delta_j)|^2) \right) d\omega$$

To conclude, it is thus sufficient to show that, if the δ_j's are well-chosen, then:

$$\frac{1}{2} \left(\sum_{j \leq J} |\hat{\psi}_j(\omega)|^2 (1 - |\hat{\chi}_{ra^n}(\omega - \delta_j)|^2) + |\hat{\psi}_j(-\omega)|^2 (1 - |\hat{\chi}_{ra^n}(-\omega - \delta_j)|^2) \right)$$

$$\leq 1 - |\hat{\chi}_{ar^{n+1}}(\omega)|^2 \qquad (\forall \omega \in \mathbb{R})$$

which is a direct consequence of the next lemma, proven in the paragraph 6.2.

Lemma 4.2. *For any $x > 0$, if $a > 1$ is close enough to 1, then there exist $(\delta_j)_{j \leq J}$ such that:*

$$\forall \omega \in \mathbb{R} \qquad \frac{1}{2} \left(\sum_{j \in \mathbb{Z}} |\hat{\psi}_j(\omega)|^2 (1 - |\hat{\chi}_x(\omega - \delta_j)|^2) \right.$$

$$\left. + \sum_{j \in \mathbb{Z}} |\hat{\psi}_j(-\omega)|^2 (1 - |\hat{\chi}_x(-\omega - \delta_j)|^2) \right) \leq 1 - |\hat{\chi}_{ax}(\omega)|^2$$

The proof is summarized in Figure 3.

8

Inductive hypothesis: the energy contained in the paths of length $n+1$ beginning by j is at most the energy of $f \star \psi_j$, multiplied by a high-pass filter (Equation (8)).

$$\sum_{|p|=n} ||U[j,p]f||_2^2 \;\leq\; \int$$

$|\hat{f}|^2$ $\qquad\qquad$ $|\hat{\psi}_j|^2$ $\qquad\qquad$ $1 - |\hat{\chi}_{ar^n}|^2$

Because of the modulus around $f \star \psi_j$, the high-pass filter can be arbitrarily shifted in frequency.

$$\sum_{|p|=n} ||U[j,p]f||_2^2 \;\leq\; \int$$

$|\hat{f}|^2$ $\qquad\qquad$ $|\hat{\psi}_j|^2$ $\qquad\qquad$ $1 - |\hat{\chi}_{ar^n}(.-\delta_j)|^2$

$$= \int$$

$|\hat{f}|^2$ $\qquad\qquad\qquad$ $|\hat{\psi}_j|^2(1 - |\hat{\chi}_{ar^n}(.-\delta_j)|^2)$

We consider real signals so we can symmetrize the filter.

$$= \int$$

$|\hat{f}|^2$ $\qquad$ $\frac{1}{2}\left(|\hat{\psi}_j|^2(1 - |\hat{\chi}_{ar^n}(.-\delta_j)|^2) + |\hat{\psi}_j(-.)|^2(1 - |\hat{\chi}_{ar^n}(-.-\delta_j)|^2)\right)$

Summing over j gives:

$$\sum_{|p|=n+1} ||U[p]f||_2^2 \;\leq\; \int$$

$|\hat{f}|^2$ $\qquad$ $\sum_{j \leq J}\frac{1}{2}\left(|\hat{\psi}_j|^2(1 - |\hat{\chi}_{ar^n}(.-\delta_j)|^2) + |\hat{\psi}_j(-.)|^2(1 - |\hat{\chi}_{ar^n}(-.-\delta_j)|^2)\right)$

$$\leq \int \qquad |\hat{f}|^2 \qquad \times \qquad 1 - |\hat{\chi}_{ar^{n+1}}|^2 \qquad \text{(shown in red)}$$

Figure 3: Schematic summary of the proof of Theorem 3.1

9

5 Adaptation of the theorem to stationary processes

In what follows, X is a real-valued stationary stochastic process, with continuous and integrable autocovariance R_X.

The scattering transform of X is defined in the same way as for elements of $L^2(\mathbb{R})$:

$$U[\text{ø}]X = X$$
$$\forall n \geq 1, (j_1, ..., j_n) \in \mathbb{Z}^n \qquad U[(j_1, ..., j_n)]X = |U[(j_1, ..., j_{n-1})]X \star \psi_{j_n}|$$
$$\forall p \in \mathcal{P}_J \qquad S_J[p]f = U[p]f \star \phi_J$$

The results proven for the deterministic wavelet transform tend to be also valid for stationary processes, if one replaces the squared L^2-norms by the expectations of the squared modulus. In particular, Theorem 3.1 can be adapted to the case of stationary processes.

Theorem 5.1. *Let $(\psi_j)_{j \in \mathbb{Z}}$ be a family of wavelets, satisfying the same conditions* (4), (5) *and* (6) *as in Theorem 3.1.*

Then, for any $J \in \mathbb{Z}$, there exist $r > 0, a > 1$ such that, for all $n \geq 2$ and $f \in L^2(\mathbb{R}, \mathbb{R})$:

$$\sum_{p \in \mathcal{P}_J, |p| = n} \mathbb{E}(|U[p]X|_2^2) \leq \mathbb{E}(|X|^2) - \mathbb{E}(|X \star \chi_{ra^n}|^2) = \int_{\mathbb{R}} \hat{R}_X(\omega)(1 - |\hat{\chi}_{ra^n}(\omega)|^2)d\omega$$

The proof of this theorem is exactly the same as the one of Theorem 3.1. The only lines to be modified are the ones where we use the fact that the Fourier transform is an isometry of $L^2(\mathbb{R})$. We use instead the property according to which, for any stationary process with continuous and integrable autocovariance R_Y, and for any $h \in L^1 \cap L^2(\mathbb{R})$:

$$\mathbb{E}(|Y \star h|^2) = |\mathbb{E}(Y \star h)|^2 + \int_{\mathbb{R}} |\hat{h}(\omega)|^2 \hat{R}_Y(\omega)d\omega$$
$$= \left|\mathbb{E}(Y) \times \int_{\mathbb{R}} h(s)ds\right|^2 + \int_{\mathbb{R}} |\hat{h}(\omega)|^2 \hat{R}_Y(\omega)d\omega$$

6 Proof of Lemmas

6.1 Proof of Lemma 4.1

Lemma (4.1). *For any $j \in \mathbb{Z}, x \in \mathbb{R}_+^*, \delta_j \in \mathbb{R}$:*

$$|| |f \star \psi_j| \star \chi_x ||_2^2 \geq ||f \star \psi_j \star (\chi_x e^{2\pi i \delta_j \cdot})||_2^2 = \int_{\mathbb{R}} |\hat{f}(\omega)|^2 |\hat{\psi}_j(\omega)|^2 |\hat{\chi}_x(\omega - \delta_j)|^2 d\omega$$

Proof. For any t, because χ_x is a positive function:

$$\begin{aligned}
|f \star \psi_j| \star \chi_x(t) &= |(f \star \psi_j)e^{-2\pi i \delta_j \cdot}| \star \chi_x(t) \\
&= |(f \star \psi_j)e^{-2\pi i \delta_j \cdot}| \star |\chi_x|(t) \\
&\geq |((f \star \psi_j)e^{-2\pi i \delta_j \cdot}) \star \chi_x(t)| \\
&= |e^{-2\pi i \delta_j t}\big(f \star \psi_j \star (\chi_x \star e^{2\pi i \delta_j \cdot})\big)(t)| \\
&= |f \star \psi_j \star (\chi_x \star e^{2\pi i \delta_j \cdot})(t)|
\end{aligned}$$

This implies the inequality. The equality is a consequence of the unitarity of the Fourier transform. $\qquad\square$

6.2 Proof of Lemma 4.2

Lemma (4.2). *For any $x > 0$, if $a > 1$ is close enough to 1, then there exist $(\delta_j)_{j \leq J}$ such that:*

$$\forall \omega \in \mathbb{R} \qquad \frac{1}{2}\left(\sum_{j \in \mathbb{Z}}|\hat{\psi}_j(\omega)|^2(1 - |\hat{\chi}_x(\omega - \delta_j)|^2)\right.$$

$$\left. + \sum_{j \in \mathbb{Z}}|\hat{\psi}_j(-\omega)|^2(1 - |\hat{\chi}_x(-\omega - \delta_j)|^2)\right) \leq 1 - |\hat{\chi}_{ax}(\omega)|^2$$

Proof. We are going to look for δ_j's of the form $\delta_j = \delta 2^{-j}$.

$$\frac{1}{2}\left(\sum_{j \in \mathbb{Z}}|\hat{\psi}_j(\omega)|^2(1 - |\hat{\chi}_x(\omega - \delta 2^{-j})|^2) + \sum_{j \in \mathbb{Z}}|\hat{\psi}_j(-\omega)|^2(1 - |\hat{\chi}_x(-\omega - \delta 2^{-j})|^2)\right)$$

$$= \frac{1}{2}\left(\sum_{j \in \mathbb{Z}}|\hat{\psi}_j(\omega)|^2(1 - e^{2\left(-\frac{\omega^2}{x^2} + \frac{2\delta\omega 2^{-j}}{x^2} - \frac{\delta^2 2^{-2j}}{x^2}\right)}) + \sum_{j \in \mathbb{Z}}|\hat{\psi}_j(-\omega)|^2(1 - e^{2\left(-\frac{\omega^2}{x^2} - \frac{2\delta\omega 2^{-j}}{x^2} - \frac{\delta^2 2^{-2j}}{x^2}\right)})\right)$$

Let us define:

$$\forall \omega \in \mathbb{R}^* \qquad S(\omega) = \frac{1}{2}\left(\sum_{j \in \mathbb{Z}}|\hat{\psi}_j(\omega)|^2 + \sum_{j \in \mathbb{Z}}|\hat{\psi}_j(-\omega)|^2\right)$$

This function is never 0 if $\omega \neq 0$: if $\omega > 0$, there exist j such that $|\hat{\psi}_j(\omega)| > |\hat{\psi}_j(-\omega)| \geq 0$ so $S(\omega) > 0$; as S is even, we also have $S(\omega) > 0$ if $\omega < 0$.

The function $y \to 1 - e^{2\left(-\frac{\omega^2}{x^2} + y\right)}$ is concave so, for any $\omega \neq 0$:

$$\frac{1}{2}\left(\sum_{j \in \mathbb{Z}}|\hat{\psi}_j(\omega)|^2(1 - e^{2\left(-\frac{\omega^2}{x^2} + \frac{2\delta\omega 2^{-j}}{x^2} - \frac{\delta^2 2^{-2j}}{x^2}\right)}) + \sum_{j \in \mathbb{Z}}|\hat{\psi}_j(-\omega)|^2(1 - e^{2\left(-\frac{\omega^2}{x^2} - \frac{2\delta\omega 2^{-j}}{x^2} - \frac{\delta^2 2^{-2j}}{x^2}\right)})\right)$$

11

$$= S(\omega) \sum_{j \in \mathbb{Z}} \frac{\frac{1}{2}|\hat{\psi}_j(\omega)|^2}{S(\omega)} (1 - e^{2\left(-\frac{\omega^2}{x^2} + \frac{2\delta\omega 2^{-j}}{x^2} - \frac{\delta^2 2^{-2j}}{x^2}\right)}) + \sum_{j \in \mathbb{Z}} \frac{\frac{1}{2}|\hat{\psi}_j(-\omega)|^2}{S(\omega)} (1 - e^{2\left(-\frac{\omega^2}{x^2} - \frac{2\delta\omega 2^{-j}}{x^2} - \frac{\delta^2 2^{-2j}}{x^2}\right)})$$

$$\leq S(\omega) \left(1 - \exp\left(2\left(-\frac{\omega^2}{x^2} + \sum_{j \in \mathbb{Z}} \frac{\frac{1}{2}|\hat{\psi}_j(\omega)|^2}{S(\omega)} \left(\frac{2\delta\omega 2^{-j}}{x^2} - \frac{\delta^2 2^{-2j}}{x^2}\right)\right.\right.\right.$$

$$\left.\left.\left. + \sum_{j \in \mathbb{Z}} \frac{\frac{1}{2}|\hat{\psi}_j(-\omega)|^2}{S(\omega)} \left(-\frac{2\delta\omega 2^{-j}}{x^2} - \frac{\delta^2 2^{-2j}}{x^2}\right)\right)\right)\right)$$

$$= S(\omega) \left(1 - \exp\left(2\left(-\frac{\omega^2}{x^2} + \frac{2\delta\omega}{x^2} F_1(\omega) - \frac{\delta^2}{x^2} F_2(\omega)\right)\right)\right) \tag{9}$$

if we set:

$$F_1(\omega) = \sum_{j \in \mathbb{Z}} \left(\frac{1}{2} \frac{|\hat{\psi}_j(\omega)|^2 - |\hat{\psi}_j(-\omega)|^2}{S(\omega)}\right) 2^{-j}$$

$$F_2(\omega) = \sum_{j \in \mathbb{Z}} \left(\frac{1}{2} \frac{|\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2}{S(\omega)}\right) 2^{-2j}$$

The functions F_1 and F_2 are both continuous. Indeed, for any j, $\hat{\psi}_j$ is continuous, because $\psi \in L^1(\mathbb{R})$. The function S is also continuous, and lower-bounded by a strictly positive constant. Finally, the sums converge uniformly on every compact subset of $\mathbb{R}$, because of the assumption (6). This implies the continuity.

Because of the hypothesis (5), $F_1(\omega) > 0$ when $\omega > 0$. Moreover, for any $\omega > 0$, $F_1(2\omega) = 2F_1(\omega)$. By a compacity argument, there exist $c > 0$ such that:

$$\forall \omega > 0 \qquad F_1(\omega) \geq c\omega$$

which, as F_1 is odd, implies:

$$\forall \omega \in \mathbb{R} \qquad \omega F_1(\omega) \geq c\omega^2 \tag{10}$$

Similarly, there exist $C > 0$ such that:

$$\forall \omega \in \mathbb{R} \qquad F_2(\omega) \leq C\omega^2 \tag{11}$$

By combining (10) and (11) with (9), we get that, for all $\omega \neq 0$:

$$\frac{1}{2} \left(\sum_{j \in \mathbb{Z}} |\hat{\psi}_j(\omega)|^2 (1 - |\hat{\chi}_x(\omega - \delta 2^{-j})|^2) + \sum_{j \in \mathbb{Z}} |\hat{\psi}_j(-\omega)|^2 (1 - |\hat{\chi}_x(-\omega - \delta 2^{-j})|^2)\right)$$

12

$$\leq S(\omega) \left(1 - \exp\left(-2\frac{\omega^2}{x^2}(1 - 2c\delta + C\delta^2) \right) \right)$$

If we take $\delta = c/C$, this yields, for any a such that $1 < a \leq \frac{1}{\sqrt{1-c^2/C}}$:

$$\frac{1}{2} \left(\sum_{j\in\mathbb{Z}} |\hat{\psi}_j(\omega)|^2 (1 - |\hat{\chi}_x(\omega - \delta 2^{-j})|^2) + \sum_{j\in\mathbb{Z}} |\hat{\psi}_j(-\omega)|^2 (1 - |\hat{\chi}_x(-\omega - \delta 2^{-j})|^2) \right)$$

$$\leq S(\omega) \left(1 - \exp\left(-2\frac{\omega^2}{x^2}\left(1 - \frac{c^2}{C}\right) \right) \right)$$

$$\leq S(\omega) \left(1 - \exp\left(-2\frac{\omega^2}{(ax)^2} \right) \right)$$

$$\leq 1 - \exp\left(-2\frac{\omega^2}{(ax)^2} \right)$$

$$= 1 - |\hat{\chi}_{ax}(\omega)|^2$$

The bound $S(\omega) \leq 1$ comes from the Littlewood-Paley inequality (4).

The inequalities are also true for $\omega = 0$ because all terms are then equal to zero. $\square$

6.3   Initialization

In this paragraph, we prove that Theorem 3.1 holds for $n = 2$. More precisely, we prove that, for any $a > 1$, there exist $r > 0$ such that Equation (7) is valid for $n = 2$.

Proof. For any real-valued function $g \in L^2(\mathbb{R})$:

$$\sum_{j\leq J} \|g \star \psi_j\|_2^2 = \int_\mathbb{R} |\hat{g}(\omega)|^2 \left(\sum_{j\leq J} |\hat{\psi}_j(\omega)|^2 \right) d\omega$$

$$= \int_\mathbb{R} |\hat{g}(\omega)|^2 \left(\frac{1}{2}\sum_{j\leq J} |\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2 \right) d\omega \qquad (12)$$

If the following inequality held, for some $r > 0, a > 1$:

$$\frac{1}{2}\sum_{j\leq J} \left(|\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2 \right) \leq 1 - |\hat{\chi}_{ra}(\omega)|^2$$

then Theorem 3.1 would be valid for $n = 1$ and it could be proven for $n = 2$ in the exact same way as in Section 4. Unfortunately, this inequality is not necessarily valid (in particular, it is never the case if the wavelet transform is unitary and the wavelets are band-limited).

13

The next lemma (proven at the end of the paragraph) nevertheless shows that the inequality is satisfied, if one allows the Gaussian function to be replaced by a more general function.

Lemma 6.1. *There exist a real-valued positive function* $\phi \in L^1 \cap L^2(\mathbb{R})$ *such that:*

$$|\hat{\phi}(\omega)|^2 = 1 - O(\omega^2) \qquad \text{when } \omega \to 0 \tag{13}$$

and:

$$\forall \omega \in \mathbb{R} \qquad \frac{1}{2}\sum_{j \leq J}\left(|\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2\right) \leq 1 - |\hat{\phi}(\omega)|^2 \tag{14}$$

The rest of the proof consists in showing how to adapt Section 4, when the Gaussian function has been replaced by the ϕ of the lemma.

Because of (12), with ϕ defined as in Lemma 6.1, we have, for any real-valued function $g \in L^2(\mathbb{R})$:

$$\sum_{j \leq J}||g \star \psi_j||_2^2 \leq \int_{\mathbb{R}} |\hat{g}(\omega)|^2(1 - |\hat{\phi}(\omega)|^2)$$

$$= ||g||_2^2 - ||g \star \phi||_2^2$$

So:

$$\sum_{p \in \mathcal{P}_J, |p|=2} ||U[p]f||_2^2 = \sum_{j_1 \leq J}\sum_{j_2 \leq J}|| \, ||f \star \psi_{j_1}| \star \psi_{j_2}| \, ||_2^2$$

$$\leq \sum_{j \leq J}||f \star \psi_j||_2^2 - || \, |f \star \psi_j| \star \phi||_2^2 \tag{15}$$

Lemma 4.1 is still true when χ_x is replaced by ϕ, because the only property of χ_x which is really needed is its positivity. So for any $\delta \in \mathbb{R}$:

$$\forall j \in \mathbb{Z} \qquad || \, |f \star \psi_j| \star \phi||_2^2 \geq \int_{\mathbb{R}} |\hat{f}(\omega)|^2|\hat{\psi}_j(\omega)|^2|\hat{\phi}(\omega - \delta)|^2 d\omega$$

In Section 4, we used this same inequality, with a different δ for each value of j. Here, we do not need δ to vary as a function of j; however, we will need to consider different values of δ and to average the inequalities over these different values.

If we combine the last inequality with (15), we obtain:

$$\forall \delta \in \mathbb{R} \qquad \sum_{p \in \mathcal{P}_J, |p|=2}||U[p]f||_2^2 \leq \int_{\mathbb{R}} |\hat{f}(\omega)|^2(1 - |\hat{\phi}(\omega - \delta)|^2)\left(\sum_{j \leq J}|\hat{\psi}_j(\omega)|^2\right)d\omega$$

14

As it holds for any $\delta \in \mathbb{R}$, we have, for any positive $c \in L^1(\mathbb{R})$ whose integral over $\mathbb{R}$ is 1:

$$\sum_{p \in \mathcal{P}_J, |p|=2} ||U[p]f||_2^2 \leq \int_{\mathbb{R}} c(\delta) \int_{\mathbb{R}} |\hat{f}(\omega)|^2 (1 - |\hat{\phi}(\omega - \delta)|^2) \left(\sum_{j \leq J} |\hat{\psi}_j(\omega)|^2 \right) d\omega d\delta$$

$$\leq \int_{\mathbb{R}} |\hat{f}(\omega)|^2 (1 - |\hat{\phi}|^2 \star c(\omega)) \left(\sum_{j \leq J} |\hat{\psi}_j(\omega)|^2 \right) d\omega$$

We limit ourselves to the case where c is even. As $|\hat{\phi}|^2$ is also even, the last inequality yields, by using the fact that f is real:

$$\sum_{p \in \mathcal{P}_J, |p|=2} ||U[p]f||_2^2 \leq \int_{\mathbb{R}} |\hat{f}(\omega)|^2 (1 - |\hat{\phi}|^2 \star c(\omega)) \times \frac{1}{2} \left(\sum_{j \leq J} |\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2 \right) d\omega$$

The conclusion comes from a last lemma, proven at the end of the paragraph.

Lemma 6.2. *If we take $c(\delta) = \frac{1}{\sqrt{\pi}} \exp(-\delta^2)$, then there exists $x > 0$ such that:*

$$\forall \omega \in \mathbb{R} \qquad (1 - |\hat{\phi}|^2 \star c(\omega)) \times \frac{1}{2} \left(\sum_{j \leq J} |\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2 \right) \leq 1 - |\hat{\chi}_x(\omega)|^2$$

$\square$

We now give the proofs of Lemmas 6.1 and 6.2.

Proof of Lemma 6.1. Let $\gamma : \mathbb{R} \to \mathbb{R}$ be any even, rapidly decreasing and band-limited function such that $\int_{\mathbb{R}} \gamma^2 = 1$. We set $\phi_0 = \gamma^2$. This is also an even, rapidly decreasing and band-limited function.

Because $\int_{\mathbb{R}} \gamma^2 = 1$, we have $\hat{\phi}_0(0) = 1$. As ϕ_0 is even, $\hat{\phi}_0'(0) = 0$. But the second derivative is strictly negative: $\hat{\phi}_0''(0) = -(2\pi)^2 \int_{\mathbb{R}} t^2 \phi_0(t) dt < 0$. We deduce from these relations that there exists $\alpha > 0$ such that:

$$|\hat{\phi}_0(\omega)|^2 = 1 - \alpha\omega^2 + o(\omega^2) \qquad \text{when } \omega \to 0 \tag{16}$$

Let us show that, for M large enough, $\phi : t \to M^{-1} \phi_0(M^{-1}t)$ satisfies the desired properties. By construction, it is a real-valued positive function. It is rapidly decreasing, so $\phi \in L^1 \cap L^2(\mathbb{R})$. The property (13) holds; only the inequality (14) is left to prove.

Because of Equation (16) and because $\hat{\phi}_0$ is compactly-supported, there exists $\tilde{\alpha} > 0$ such that for all $\omega \in \text{Supp}(\hat{\phi}_0)$:

$$|\hat{\phi}_0(\omega)|^2 \leq 1 - \tilde{\alpha}\omega^2$$

15

By the assumption (6), $|\hat{\psi}(\omega)| = O(|\omega|^{1+\epsilon})$ when $\omega \to 0$, for some $\epsilon > 0$. This implies that:

$$\frac{1}{2}\sum_{j \leq J} \left(|\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2 \right) = o(\omega^2) \qquad \text{when } \omega \to 0$$

Because this sum is moreover bounded by 1 on all $\mathbb{R}$, there exists $A > 0$ such that:

$$\forall \omega \in \mathbb{R} \qquad \frac{1}{2}\sum_{j \leq J} \left(|\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2 \right) \leq A\omega^2$$

If $M \geq \sqrt{A/\tilde{\alpha}}$, then, on the support of $\hat{\phi}$:

$$|\hat{\phi}(\omega)|^2 + \frac{1}{2}\sum_{j \leq J} \left(|\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2 \right)$$

$$= |\hat{\phi}_0(M\omega)|^2 + \frac{1}{2}\sum_{j \leq J} \left(|\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2 \right)$$

$$\leq 1 - \tilde{\alpha}M^2\omega^2 + A\omega^2$$

$$\leq 1$$

Outside the support of $\hat{\phi}$, the inequality is also true, because of the Littlewood-Paley condition (4). So Equation (14) holds. $\qquad\square$

Proof of Lemma 6.2. Let us define:

$$F(\omega) = (1 - |\hat{\phi}|^2 \star c(\omega)) \times \frac{1}{2} \left(\sum_{j \leq J} |\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2 \right)$$

We are going to prove that F has the following three properties:

1. $\forall \omega \in \mathbb{R}, F(\omega) < 1$

2. $F(\omega) = O(\omega^2)$ when $\omega \to 0$

3. There exists $x > 0$ such that $F(\omega) \leq 1 - |\hat{\chi}_x(\omega)|^2$ if $|\omega|$ is large enough.

These three properties imply that F is bounded by $1 - |\hat{\chi}_x|^2$ on all $\mathbb{R}$, for x small enough. This assertion relies on a compacity argument; as it is relatively straightforward, we do not prove it.

The first property is an immediate consequence of the Littlewood-Paley inequality (4) and of the fact that $|\hat{\phi}|^2 \star c > 0$ over $\mathbb{R}$.

16

The second one is a consequence of a fact that has been explained in the proof of Lemma 6.1:

$$\frac{1}{2}\left(\sum_{j\leq J}|\hat{\psi}_j(\omega)|^2 + |\hat{\psi}_j(-\omega)|^2\right) = o(\omega^2) \qquad \text{when } \omega \to 0$$

For the last one, we remark that, for any $\omega \in \mathbb{R}$ such that $\omega \geq 1$:

$$|\hat{\phi}|^2 \star c(\omega) = \int_{\mathbb{R}}|\hat{\phi}(\delta)|^2 c(\omega - \delta)d\delta$$

$$\geq \int_0^1 |\hat{\phi}(\delta)|^2 c(\omega - \delta)d\delta$$

$$\geq c(\omega)\int_0^1 |\hat{\phi}(\delta)|^2 d\delta$$

As all the functions are even, this is also true for $\omega \leq -1$. So when $|\omega|$ is large enough:

$$|\hat{\phi}|^2 \star c(\omega) \geq \frac{1}{\sqrt{\pi}}\exp(-\omega^2)\left(\int_0^1 |\hat{\phi}(\delta)|^2 d\delta\right) \geq \exp(-2\omega^2) = |\hat{\chi}_1(\omega)|^2$$

This proves the third property and concludes. □

Bibliography

J. Andén and S. Mallat. Multiscale scattering for audio classification. In <u>Proceedings of the International Society of Music Information Retrieval 2011 Conference</u>, pages 657–662, 2011.

J. Bruna, S. Mallat, E. Bacry, and J.-F. Muzy. Intermittent process analysis with scattering moments. <u>Annals of Statistics</u>, 43(1):323–351, 2015.

A. Dosovitskiy and T. Brox. Inverting convolutional networks with convolutional networks. Technical report, , 2015. http://arxiv.org/abs/1506.02753.

M. Hirn, N. Poilvert, and S. Mallat. Quantum energy regression using scattering transforms. <u>Submitted to the International Conference on Machine Learning</u>, 2015. http://arxiv.org/abs/1502.02077.

A. Mahendran and A. Vedaldi. Understanding deep image representations by inverting them. In <u>The IEEE Conference on Computer Vision and Pattern Recognition</u>, pages 5188–5196, 2015.

S. Mallat. Group invariant scattering. <u>Communications in Pure and Applied Mathematics</u>, 65 (10):1331–1398, 2012.

E. Oyallon and S. Mallat. Deep roto-translation scattering for object classification. In The IEEE Conference on Computer Vision and Pattern Recognition, pages 2865–2873, 2015.

I. W. Selesnick. Hilbert transform pairs of wavelet bases. IEEE Signal Processing Letters, 8, 2001.

K. Simonyan, A. Vedaldi, and A. Zisserman. Deep inside convolutionnal networks: visualising image classification models and saliency maps. In International Conference on Learning Representations Workshop, 2014.

I. Waldspurger. Wavelet transform modulus: phase retrieval and scattering. PhD thesis, École Normale Supérieure, Paris, 2015. URL http://www.di.ens.fr/~waldspurger/these.pdf.