A wavelet transform, even with \( \psi = (-1)^n \theta^{(n)} \) for \( \theta \) a Gaussian, may have a maxima line that converges to a point \( v \) even though \( f \) is regular at \( v \) (i.e., \( f \) is Lipschitz \( \alpha \) at \( v \) for \( \alpha > 1 \)); see Figure 23, and the maxima line that converges to \( v = 0.23 \). To distinguish such points from singular points it is necessary to measure the decay of the modulus maxima amplitude.

To interpret more easily the pointwise conditions (35) and (36) of Theorem 31, suppose that for \( s < s_0 \) all modulus maxima that converge to \( v \) are included in a cone \( C_v \) defined as:

\[
C_v = \{(u, s) \in \mathbb{R} \times (0, \infty) : |u - v| \leq Cs\}
\]

Figure 25 gives an illustration. In general this will not be true, in particular for functions \( f \) that have oscillations that accelerate in a neighborhood of \( v \) (e.g., \( f(t) = \sin(1/t) \) for \( v = 0 \)).

![Figure 25: The cone of influence \( C_v \) of an abscissa \( v \) consists of the time-scale points \( (u, s) \)](image)

Within the cone \( C_v \) we have \( |u - v|/s \leq C \), and so the conditions (35) and (36) of Theorem 31 can be written for these points as:

\[
|Wf(u, s)| \leq A's^{\alpha+1/2}, \quad \forall (u, s) \in C_v
\]

This is equivalent to:

\[
\log_2 |Wf(u, s)| \leq \log_2 A' + \left( \alpha + \frac{1}{2} \right) \log_2 s
\]

Thus the Lipschitz regularity at \( v \) is the maximum slope of \( \log_2 |Wf(u, s)| \) as a function of \( \log_2 s \) along the maxima line converging to \( v \). Figure 26 describes an example.

In practice variations in a signal \( f(t) \) may correspond to smooth singularities, for example due to blur or shadows in an image. In this case, points
Figure 23 revisited

Figure 26: Figure (b) plots $\log_2 |Wf(u, s)|$ as a function of $\log_2 s$ along two maxima lines. The solid line corresponds to the maxima line that converges to $v = 0.05$. It has a maximum slope of $\alpha + 1/2 \approx 1/2$, implying that $\alpha = 0$, which is expected since $f(t)$ is discontinuous at $t = 0.05$. The dashed line corresponds to the maxima line converging to $v = 0.42$. Here the maximum slope is $\alpha + 1/2 \approx 1$, indicating that the singularity is Lipschitz $1/2$. 
of rapid transition will technically be $C^\infty$. However, if the blurring effect is from a Gaussian kernel, we can still get precise measurements on the decay of the wavelet coefficients.

We suppose that in the neighborhood of a sharp transition $v$, $f(t)$ can be modeled as

$$f(t) = f_0 \ast g_\sigma(t)$$

where

$$g_\sigma(t) = \frac{1}{\sqrt{2\pi}\sigma}e^{-t^2/2\sigma^2}$$

If $f_0$ is uniformly Lipschitz $\alpha$ in a neighborhood of $v$, then we can relate the decay of the wavelet coefficients to $\alpha$ and $\sigma$ so long as $\psi = (-1)^n\theta^{(n)}$ for $\theta$ a Gaussian.

**Theorem 36.** Let $\psi = (-1)^n\theta^{(n)}$ with

$$\theta(t) = \lambda e^{-t^2/2\beta^2}$$

If $f = f_0 \ast g_\sigma$ and $f_0$ is uniformly Lipschitz $\alpha \leq n$ on $[v - \varepsilon, v + \varepsilon]$, then there exists $A > 0$ such that

$$|Wf(u, s)| \leq As^{\alpha+1/2}\left(1 + \frac{\sigma^2}{\beta^2s^2}\right)^{-(n-\alpha)/2}, \quad \forall (u, s) \in [v - \varepsilon, v + \varepsilon] \times (0, \infty)$$

**Proof.** Using Theorem 30 we write the wavelet transform as:

$$Wf(u, s) = s^n \frac{d^n}{du^n}(f \ast \theta_s)(u) = s^n \frac{d^n}{du^n}(f_0 \ast g_\sigma \ast \theta_s)(u)$$

Since $g_\sigma$ and $\theta$ are Gaussians, $g_\sigma \ast \theta_s$ is also a Gaussian and one calculate its scale as:

$$g_\sigma \ast \theta_s(t) = \sqrt{s/s_0} \theta_{s_0}(t), \quad s_0 = \sqrt{s^2 + \sigma^2/\beta^2}$$

Therefore we can rewrite the wavelet transform as

$$Wf(u, s) = s^n \sqrt{s/s_0} \frac{d^n}{du^n}(f_0 \ast \theta_{s_0})(u)$$

$$= \left(\frac{s}{s_0}\right)^{n+1/2} s_0^n \frac{d^n}{du^n}(f_0 \ast \theta_{s_0})(u)$$

$$= \left(\frac{s}{s_0}\right)^{n+1/2} Wf_0(u, s_0)$$
Since $f_0$ is uniformly Lipschitz $\alpha$ on $[v - \varepsilon, v + \varepsilon]$, Theorem 32 proves that there exists $A > 0$ such that

$$|Wf_0(u, s)| \leq As^{\alpha+1/2}, \quad \forall (u, s) \in [v - \varepsilon, v + \varepsilon] \times (0, \infty)$$

Therefore,

$$|Wf(u, s)| \leq \left(\frac{s}{s_0}\right)^{n+1/2} |Wf_0(u, s_0)|$$

$$\leq \left(\frac{s}{s_0}\right)^{n+1/2} As_0^{\alpha+1/2}$$

$$= As^{n+1/2}s_0^{-(n-\alpha)}$$

$$= As^{n+1/2}\left(s^2 + \frac{\sigma^2}{\beta^2}\right)^{-(n-\alpha)/2}$$

$$= As^{\alpha+1/2}\left(1 + \frac{\sigma^2}{\beta^2s^2}\right)^{-(n-\alpha)/2}$$

This theorem relates the wavelet transform decay expected by the Lipschitz $\alpha$ singularity versus what one observes due to the diffusion at the singularity. At large scales $s \gg \sigma/\beta$, the bound is essentially $|Wf(u, s)| \leq As^{\alpha+1/2}$ since the second term becomes nearly equal to one. In other words, the larger wavelets do not “feel” the blurring effect. However, for $s \leq \sigma/\beta$, the decay is more like $|Wf(u, s)| \leq As^{n+1/2}$, which depends upon the number of vanishing moments of the wavelet, not the regularity of the underlying singularity. This is because the blurred signal is in fact $C^\infty$, and thus the decay at fine scales will necessarily be limited by the finite number of vanishing moments. Figure 27 gives an example.

The windowed Fourier transform $Sf(u, \xi)$ and the wavelet transform $Wf(u, s)$ are examples of signal analysis operators, which can be put in a more general context via Frame theory. Frame theory will give us the mathematical foundation to consider general dictionaries of time frequency atoms. It will, additionally, give us the mathematical framework to synthesize signals, not just analyze them. This will be useful for, amongst other reasons, obtaining sparse compression of signals using just their wavelet modulus maxima coefficients. For now we leave wavelets to study frames, but we will return possessing the framework to not only complete this story, but also the tools
to chart the path forward into more general analysis of signal representations in dictionaries.

**Exercise 59.** Read Section 6.2.1 of *A Wavelet Tour of Signal Processing*.

**Exercise 60.** Read Section 6.4 of *A Wavelet Tour of Signal Processing* (we are not going to cover this section in class, but it is super interesting).

**Exercise 61.** [20 points] Implement a function to compute the devil’s staircase described on Example 6.10 of *A Wavelet Tour of Signal Processing*. Using your wavelet code from previous exercises, compute the wavelet transform of the devil’s staircase using $\psi = -\theta'$, where $\theta$ is a Gaussian. Estimate the modulus maxima points as well. Provide three plots: A plot of the devil’s staircase, a plot of the wavelet coefficients, and a plot of the maxima lines. How do your results compare to Figure 6.18 in *A Wavelet Tour of Signal Processing*?

**Exercise 62.** [20 points] Let $\psi = -\theta'$ where $\theta$ is a positive window with compact support. If $f$ is a devil’s staircase, prove directly without appealing to Theorem 33 that there exists sequences of modulus maxima that converge toward each singularity.

### 6 Frames

*Chapter 5 of A Wavelet Tour of Signal Processing.*

#### 6.1 Frames and Riesz Bases

*Section 5.1 of A Wavelet Tour of Signal Processing.*

#### 6.1.1 Stable Analysis and Synthesis Operators

*Section 5.1.1 of A Wavelet Tour of Signal Processing.*

Let $\mathcal{H}$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\|f\| = \sqrt{\langle f, f \rangle}$. The main examples we will want to keep in the back of our mind are the ones we have encountered thus far in the course, i.e., $L^2(\mathbb{R})$, $\ell^2$, and $\mathbb{R}^N$ or $\mathbb{C}^N$. Consider a dictionary

$$
\mathcal{D} = \{ \phi_\gamma \}_{\gamma \in \Gamma} \subset \mathcal{H}
$$
Figure 27: Top: Signal with two types of singularities, a jump discontinuity at $t = 0.35$ and a cusp at $t = 0.81$. Blurred versions of the same singularities are located at $t = 0.60$ and $t = 0.12$, respectively. (a) The wavelet transform $Wf(u, s)$ using a wavelet $\psi = \theta''$, where $\theta$ is a Gaussian with variance $\beta = 1$. (b) Modulus maxima lines. (c) Decay of $\log_2 |Wf(u, s)|$ along the maxima lines. The solid and dashed lines on the left correspond to the maxima lines converging to $t = 0.81$ and $t = 0.12$, respectively. The solid and dashed lines on the right correspond to the maxima lines converging to $t = 0.35$ and $t = 0.60$, respectively. Thus the solid lines correspond to the singularities, and the dashed lines correspond to the blurred singularities. Notice that the diffusion modifies the decay for $s \leq \sigma = 2^{-5}$. 

113
consisting of atoms $\phi_\gamma \in \mathcal{H}$, in which the index set $\Gamma$ is either finite or countable. The analysis operator associated to $\mathcal{D}$ is:

$$\Phi f(\gamma) = \langle f, \phi_\gamma \rangle, \quad \gamma \in \Gamma, \quad f \in \mathcal{H}$$

The dictionary $\mathcal{D}$ is a frame for $\mathcal{H}$ if there exist constants $0 < A \leq B < \infty$ such that

$$A\|f\|^2 \leq \sum_{\gamma \in \Gamma} |\langle f, \phi_\gamma \rangle|^2 \leq B\|f\|^2 \tag{46}$$

When $A = B$ the frame is tight. If the atoms in $\mathcal{D}$ are independent, then the frame is not redundant and it is called a Riesz basis. We shall see later that frames define invertible operators on $\text{image}(\Phi)$.

We remark that if a Hilbert space $\mathcal{H}$ admits a frame $\mathcal{D}$, then $\mathcal{H}$ must be separable. Indeed, suppose that $\langle f, \phi_\gamma \rangle = 0$ for all $\gamma \in \Gamma$. Then using the lower bound of (46), we obtain:

$$A\|f\|^2 \leq \sum_{\gamma \in \Gamma} |\langle f, \phi_\gamma \rangle|^2 = 0 \implies f = 0$$

Thus the only element of $\mathcal{H}$ orthogonal to every $\phi_\gamma \in \mathcal{D}$ is 0. It follows (with some work) that $\mathcal{D}$ must be a complete set of functions in $\mathcal{H}$. This means that for each $f \in \mathcal{H}$ and for each $\varepsilon > 0$ there exists an $N \in \mathbb{N}$, $\{\gamma_n\}_{n=1}^N \subset \Gamma$ and coefficients $\{c_n\}_{n=1}^N \subset \mathbb{C}$ such that

$$\left\| f - \sum_{n=1}^N c_n \phi_{\gamma_n} \right\| \leq \varepsilon$$

Since we can additionally take the coefficients $\{c_n\}_{n=1}^N$ to have rational real and imaginary parts, we have found a dense subset of $\mathcal{H}$.

The analysis operator $\Phi$ analyzing a signal $f \in \mathcal{H}$ by testing it against the dictionary atoms $\phi_\gamma$. The adjoint of $\Phi$ defines a synthesis operator, which we now explain. Consider the space of $\ell^2$ sequences indexed by $\Gamma$:

$$\ell^2(\Gamma) = \{a : \|a\|^2 = \sum_{\gamma \in \Gamma} |a[\gamma]|^2 < \infty\}$$

Notice that the frame condition (46) guarantees that

$$\Phi : \mathcal{H} \rightarrow \ell^2(\Gamma)$$
Therefore $\Phi$ has an adjoint

$$\Phi^* : \ell^2(\Gamma) \to \mathcal{H}$$

which is defined through the following relation:

$$\langle \Phi^* a, f \rangle_{\mathcal{H}} = \langle a, \Phi f \rangle_{\ell^2(\Gamma)}$$

where the subscript on the inner products $\langle \cdot, \cdot \rangle$ is written to emphasize the space over which the inner product is computed (moving forward we will drop this subscript and infer the space from the context). Notice that

$$\langle a, \Phi f \rangle = \sum_{\gamma \in \Gamma} a[\gamma] \langle f, \phi_\gamma \rangle^*$$

$$= \sum_{\gamma \in \Gamma} a[\gamma] \langle \phi_\gamma, f \rangle$$

$$= \sum_{\gamma \in \Gamma} \langle a[\gamma] \phi_\gamma, f \rangle$$

$$= \left\langle \sum_{\gamma \in \Gamma} a[\gamma] \phi_\gamma, f \right\rangle$$

from which it follows that

$$\Phi^* a = \sum_{\gamma} a[\gamma] \phi_\gamma$$

We refer to $\Phi^*$ as the synthesis operator since it synthesizes signals in $\mathcal{H}$ from the sequence $a \in \ell^2(\Gamma)$.

Notice that the frame condition (46) can be rewritten as:

$$A \|f\|^2 \leq \|\Phi^* f\|^2 = \langle \Phi^* \Phi f, f \rangle \leq B \|f\|^2$$

where

$$\Phi^* \Phi f = \sum_{\gamma \in \Gamma} \langle f, \phi_\gamma \rangle \phi_\gamma$$

It follows that we can take $A$ and $B$ as:

$$A = \inf_{f \in \mathcal{H}} \frac{\langle \Phi^* \Phi f, f \rangle}{\|f\|^2}$$

$$B = \sup_{f \in \mathcal{H}} \frac{\langle \Phi^* \Phi f, f \rangle}{\|f\|^2}$$
This is just the infimum and supremum of the Rayleigh quotient of $\Phi^*\Phi$. In finite dimensions, this implies that $A$ is the smallest eigenvalue of $\Phi^*\Phi$ and $B$ is the largest eigenvalue of $\Phi^*\Phi$; note that the eigenvalues of $\Phi^*\Phi$ are the singular values of $\Phi$. The next theorem shows that if the frame analysis operator is stable (as defined by the frame condition (46)), then the frame synthesis operator obeys a similar stability condition.

**Theorem 37.** A dictionary $D = \{\phi_\gamma\}_{\gamma \in \Gamma}$ is a frame with bounds $0 < A \leq B < \infty$ if and only if

$$A \|a\|^2 \leq \left\| \sum_{\gamma \in \Gamma} a[\gamma] \phi_\gamma \right\|^2 \leq B \|a\|^2, \quad \forall a \in \text{image}(\Phi)$$

**Proof.** Note that

$$\left\| \sum_{\gamma \in \Gamma} a[\gamma] \phi_\gamma \right\|^2 = \langle \Phi^*a, \Phi^*a \rangle = \langle \Phi \Phi^*a, a \rangle$$

The theorem will thus follow if we can show that

$$\inf_{f \in \mathcal{H}} \frac{\langle \Phi^*\Phi f, f \rangle}{\|f\|^2} = \inf_{a \in \text{image}(\Phi)} \frac{\langle \Phi \Phi^*a, a \rangle}{\|a\|^2} \quad (47)$$

and

$$\sup_{f \in \mathcal{H}} \frac{\langle \Phi^*\Phi f, f \rangle}{\|f\|^2} = \sup_{a \in \text{image}(\Phi)} \frac{\langle \Phi \Phi^*a, a \rangle}{\|a\|^2} \quad (48)$$

Let us first consider the case of a finite dimensional Hilbert space. Suppose that $\mathcal{D}$ is a frame and let $\lambda$ be an eigenvalue of $\Phi^*\Phi$ with eigenvector $f_\lambda \neq 0$. We claim that $\Phi f_\lambda$ is an eigenvector of $\Phi \Phi^*$ also with eigenvalue $\lambda$; indeed:

$$\Phi \Phi^*(\Phi f_\lambda) = \Phi \Phi^* \Phi f_\lambda = \lambda \Phi f_\lambda$$

Furthermore $\Phi f_\lambda \neq 0$ since the frame bounds (46) imply that $\|\Phi f_\lambda\|^2 \geq A \|f_\lambda\|^2$. Therefore the eigenvalues of $\Phi^*\Phi$ and $\Phi \Phi^*$ are identical and we conclude that (47) and (48) hold. For the converse, the proof proceeds along similar lines.

Now suppose that $\mathcal{H}$ is infinite dimensional and $\mathcal{D}$ is a frame for $\mathcal{H}$. From our previous discussion, we know that $\mathcal{H}$ is separable, which means that $\mathcal{H}$
has a countable orthonormal basis. Let $\mathcal{B} = \{e_1, e_2, \ldots\} \subset \mathcal{H}$ be such a basis. Define

$$\mathcal{H}_N = \text{span}\{e_1, \ldots, e_N\} \subset \mathcal{H}$$

Let $\Phi_N = \Phi|_{\mathcal{H}_N}$, that is $\Phi_N$ is the restriction of $\Phi$ to $\mathcal{H}_N$. Notice that $\lim_{N \to \infty} \mathcal{H}_N = \mathcal{H}$ and $\lim_{N \to \infty} \text{image}(\Phi_N) = \text{image}(\Phi)$. Using the proof for the finite dimensional case, we then have:

$$\inf_{f \in \mathcal{H}} \frac{\langle \Phi^* \Phi f, f \rangle}{\|f\|^2} = \lim_{N \to \infty} \inf_{f \in \mathcal{H}_N} \frac{\langle \Phi^* \Phi f, f \rangle}{\|f\|^2} = \lim_{N \to \infty} \inf_{a \in \text{image}(\Phi_N)} \frac{\langle \Phi \Phi^* a, a \rangle}{\|a\|^2} = \inf_{a \in \text{image}(\Phi)} \frac{\langle \Phi \Phi^* a, a \rangle}{\|a\|^2}$$

The proof for the supremum is identical, and the proof for the converse follows along similar lines. \[\Box\]

The operator $\Phi \Phi^* : \text{image}(\Phi) \to \text{image}(\Phi)$ is the Gram “matrix”. It is defined as:

$$\Phi \Phi^* a[\gamma] = \sum_{m \in \Gamma} a[m] \langle \phi_m, \phi_{\gamma} \rangle, \quad \forall a \in \text{image}(\Phi)$$

The next theorem shows that the redundancy of a finite frame in finite dimensions is easy to measure, and is the obvious answer.

**Theorem 38.** Let $\mathcal{D} = \{\phi_n\}_{n=1}^P$ be a finite frame for $\mathbb{R}^N$ or $\mathbb{C}^N$ in which $\|\phi_n\| = 1$ for all $1 \leq n \leq P$. Then the frame bounds satisfy:

$$A \leq \frac{P}{N} \leq B$$

and the frame is tight if and only if $A = B = P/N$.

The proof is on page 157 of *A Wavelet Tour of Signal Processing* and is quite simple. Tight frames are easy to come up with by concatenating orthonormal bases. For $1 \leq k \leq K$, suppose that $\{\phi_{k,\gamma}\}_{\gamma \in \Gamma}$ is an orthonormal basis for $\mathcal{H}$. Since it is an orthonormal basis we have:

$$\sum_{\gamma \in \Gamma} |\langle f, \phi_{k,\gamma} \rangle|^2 = \|f\|^2$$

The dictionary

$$\mathcal{D} = \{\phi_{k,\gamma}\}_{\gamma \in \Gamma}, \; 1 \leq k \leq K$$
is a tight frame with $A = B = K$; indeed:

$$\sum_{k=1}^{K} \sum_{\gamma \in \Gamma} |\langle f, \phi_{k,\gamma} \rangle|^2 = \sum_{k=1}^{K} \|f\|^2 = K \|f\|^2$$

**Exercise 63.** Read Section 5.1.1 of *A Wavelet Tour of Signal Processing*.

**Exercise 64.** Prove that if $K \neq 0$, then

$$\mathcal{D} = \left\{ \phi_n(t) = e^{2\pi i nt/K} \right\}_{n \in \mathbb{Z}}$$

is a tight frame for $L^2[0,1]$. Compute the frame bound.

### 6.1.2 Dual Frame and Pseudo Inverse

*Section 5.1.2 of A Wavelet Tour of Signal Processing.*

If $\mathcal{D} = \{\phi_{\gamma}\}_{\gamma \in \Gamma}$ is a frame but not a Riesz basis, then the frame analysis operator $\Phi$ admits an infinite number of left inverses $M$ such that

$$M\Phi f = f, \quad \forall f \in \mathcal{H}$$

This is because of the redundancy of $\mathcal{D}$, which ensures that $\text{image}(\Phi)^\perp \neq \{0\}$, and so the left inverse is free to map $a \in \text{image}(\Phi)^\perp$ to any function $g \in \mathcal{H}$. The pseudo-inverse, written as $\Phi^\dagger$, is the left inverse $M$ that maps $\text{image}(\Phi)^\perp$ to 0:

$$\Phi^\dagger \Phi f = f, \quad \forall f \in \mathcal{H} \quad \text{and} \quad \Phi^\dagger a = 0, \quad \forall a \in \text{image}(\Phi)^\perp$$

The next theorem computes the pseudo-inverse explicitly.

**Theorem 39.** If $\mathcal{D} = \{\phi_{\gamma}\}_{\gamma \in \Gamma}$ is a frame then $\Phi^* \Phi$ is invertible and

$$\Phi^\dagger = (\Phi^* \Phi)^{-1} \Phi^*$$

**Proof.** First recall that we can rewrite the frame condition (46) as:

$$A\|f\|^2 \leq \langle \Phi^* \Phi f, f \rangle \leq B\|f\|^2$$

Thus

$$\Phi^* \Phi f = 0 \iff f = 0$$
and so $\Phi^*\Phi$ is invertible. It follows that

$$(\Phi^*\Phi)^{-1}(\Phi^*\Phi)f = f$$

which shows that $M = (\Phi^*\Phi)^{-1}\Phi^*$ is a left inverse for $\Phi$. Now we show that $M = \Phi^\dagger$.

We first show that $\text{null}(\Phi^*) = \text{image}(\Phi)^\perp$. Let $a \in \text{null}(\Phi^*)$ and $b \in \text{image}(\Phi)$ with $\Phi f = b$. Then:

$$\langle a, b \rangle = \langle a, \Phi f \rangle = \langle \Phi^* a, f \rangle = \langle 0, f \rangle = 0$$

Thus indeed $\text{null}(\Phi^*) = \text{image}(\Phi)^\perp$. But then

$$(\Phi^*\Phi)^{-1}\Phi^* a = 0, \quad \forall a \in \text{image}(\Phi)^\perp = \text{null}(\Phi^*)$$

and so we have $\Phi^\dagger = (\Phi^*\Phi)^{-1}\Phi^*$. \qed
References


